Published in final edited form as:
Nat Chem. 2018 August ; 10(8): 831–837. doi:10.1038/s41557-018-0052-5.

Mutually orthogonal pyrrolysyl-tRNA synthetase/tRNA pairs

Julian C. W. Willis[1] and **Jason W. Chin**[1,2]

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

Abstract

Genetically encoding distinct non-canonical amino acids (ncAAs) into proteins synthesized in cells requires mutually orthogonal aminoacyl-tRNA synthetase (aaRS)/tRNA pairs. The pyrrolysyl-tRNA synthetase/^{Pyl}tRNA pair from *M. mazei* (*Mm*) has been engineered to incorporate diverse ncAAs and is commonly considered an ideal pair for genetic code expansion. However, finding new aaRS/tRNA pairs that share the advantages of the *Mm*PylRS/*Mm*^{Pyl}tRNA pair and are orthogonal to both endogenous aaRS/tRNA pairs and the *Mm*PylRS/*Mm*^{Pyl}tRNA pair has proved challenging. Here we demonstrate that several ΔNPylRS/^{Pyl}tRNA$_{CUA}$ pairs, in which PylRS lacks an N-terminal domain, are active, orthogonal and efficiently incorporate ncAAs in *E. coli*. We create new PylRS/^{Pyl}tRNA pairs that are mutually orthogonal to the *Mm*PylRS/*Mm*^{Pyl}tRNA pair and show that transplanting mutations that reprogram the ncAA specificity of *Mm*PylRS into the new PylRS reprograms its substrate specificity. Finally we show that distinct PylRS/^{Pyl}tRNA derived pairs can function in the same cell, decode distinct codons, and incorporate distinct ncAAs.

Genetically encoding the co-translational incorporation of multiple distinct ncAAs into proteins facilitates strategies for synthesizing proteins bearing combinations of biophysical probes (for example, Förster resonance energy transfer (FRET) probes) to follow protein conformational change1–3, protein stabilization through cyclization1, 4, and decoding the combinatorial effects of post-translational modifications. Moreover, strategies for encoding the co-translational incorporation of multiple ncAAs may provide a foundation for the encoded cellular synthesis of non-canonical biopolymers5.

Encoding multiple ncAAs into proteins synthesized in cells requires (1) the creation of additional codons or the repurposing of triplet codons5, (2) the creation or discovery of aminoacyl-tRNA synthetase/tRNA pairs that are both orthogonal to the synthetases and tRNAs used by the host organism for natural translation, and mutually orthogonal with

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

Data availability statement
Source data for all Figures are available from the corresponding author upon reasonable request.

Author contributions
J.C.W.W. performed all experiments. J.C.W.W and J.W.C. conceived and designed experiments, analyzed the data and wrote the paper.

Competing Financial Interests
The authors declare no competing financial interests.

Europe PMC Funders Author Manuscripts

respect to other orthogonal aminoacyl-tRNA synthetases and tRNAs6, 7, and (3) the reprogramming of these mutually orthogonal pairs to selectively use distinct ncAAs4, 5, 8.

The *Methanosarcina mazei* (*Mm*) pyrrolysyl-tRNA synthetase (PylRS encoded by *PylS*)/*Mm*^{Pyl}tRNA$_{CUA}$ (encoded by *MmPylT*) pair, along with the homologous pair from *Methanosarcina barkeri* (*Mb*), has been extensively developed for the co-translational incorporation of ncAAs into proteins via genetic code expansion9. These orthogonal pairs do not recognize the canonical 20 amino acids, and their active sites have been evolved to accommodate numerous ncAAs, including those used in this study (Fig. 1). Moreover, *Mm*PylRS and *Mb*PylRS do not recognize the anticodon of their cognate ^{Pyl}tRNA$_{CUA}$s10: a feature that facilitates the decoding of diverse codons by these pairs, through mutation of their anticodons1, 11. The *Mm* or *Mb* PylRS/^{Pyl}tRNA pairs have been used in combination with derivatives of the *Methanocaldococcus janaschii* (*Mj*) TyrRS/^{Tyr}tRNA$_{CUA}$ pair, to direct the co-translational incorporation of several pairs of ncAAs into proteins in *E. coli*1. However, *Mj*TyrRS recognizes the anticodon of its cognate tRNA, which restricts the codons this pair can be easily altered to efficiently decode, and this pair has primarily been used to incorporate aromatic ncAAs related to phenylalanine12. There is a pressing need to discover new highly active and mutually orthogonal synthetase/tRNA pairs with comparable flexibility to the *Mm*PylRS/*Mm*^{Pyl}tRNA$_{CUA}$ pair. Such pairs would provide a foundation for expanding the diversity of chemical structures that can be simultaneously incorporated into proteins and, in combination with emerging advances in creating or repurposing codons4, 13–15, may provide a foundation for increasing the number of monomers that can be simultaneously encoded in cellular translation.

*Mm*PylRS and *Mb*PylRS are composed of an N-terminal domain, that binds to the T-arm and variable loop of their cognate tRNAs16, 17, linked to a C-terminal catalytic domain (PylRSΔ184 in *M. mazei*)18. While the catalytic domain can (inefficiently) aminoacylate its cognate tRNA *in vitro*, the full-length protein is absolutely required for measureable amber suppression activity in cells16, 19. Another group of PylRS enzymes, commonly exemplified by *Desulfitobacterium hafniense (Dh)*, have separate genes encoding the N-terminal domain (*Dh*PylSn) and the C-terminal catalytic domain (*Dh*PylSc) of PylRS as distinct polypeptides. These polypeptides are believed to assemble to create a functional synthetase *in vivo*. In *E. coli* the C-terminal protein, *Dh*PylSc, is reported to have less than 1% of the activity of *Mb*PylRS19.

Here we demonstrate that several ΔNPylRS/^{Pyl}tRNA$_{CUA}$ pairs, in which PylRS lacks an N-terminal domain, are active and orthogonal in *E. coli* and efficiently incorporate ncAAs. We create new PylRS/^{Pyl}tRNA pairs that are mutually orthogonal to the *Mm*PylRS/*Mm*^{Pyl}tRNA pair and show that transplanting mutations that reprogram the ncAA specificity of *Mm*PylRS into the new PylRS reprograms its substrate specificity in a predictable manner. Finally we show that two PylRS/^{Pyl}tRNA derived pairs can function in the same cell, decode distinct codons, and incorporate distinct ncAAs.

Results

Identifying active and orthogonal ΔNPylRS/^{Pyl}tRNA pairs

To identify the full complement of PylRS genes we searched genomes for genes that contain the C-terminal region of PylRS, via a BLAST search for sequence similarity to *Mm*PylRSΔ184. We then searched the resulting genomes for sequence similarity to *Dh*PylSn. This search identified genomes that contain both a *Dh*PylSn-like sequence and the C-terminal catalytic domain, either in a single gene (as in *M. mazei*) or in separate genes (as in *D. hafniense*). Importantly the search also identified 21 PylRS sequences that contained the C-terminal catalytic domain, but for which we did not identify a *Dh*PylSn-like sequence in the corresponding genome (Fig. 2a); we designated this class of PylRS variants as ΔNPylRS. In some cases the N-terminal domain may not be identified because of an incomplete genome sequence. However, phylogenetic analysis suggested that most of these PylRS sequences that do not contain a *Dh*PylSn-like sequence in their genomes come from organisms that form a distinct clade (Supplementary Fig. 1). Intriguingly, the organisms within this clade are part of a recently described seventh order of methanogens (Methanomassiliicoccales) that have an energy metabolism distinct from other methanogens20 and are divergent from the order of methanogens that contains *Mm*PylRS and *Mb*PylRS (Methanosarcinales)21. For five of the PylRS sequences (*Methanomethylophilus alvus* (*Ma*) PylRS, *Methanogenic archaeon ISO4-G1* (*G1*) PylRS, *Methanogenic archaeon ISO4-H5* (*H5*) PylRS, *Methanonatronarchaeum termitum* (*Mt*) PylRS, and *Methanomassiliicoccus luminyensis* (*Ml*) PylRS) that do not contain a *Dh*PylSn-like sequence in their genome (Supplementary Fig. 2) we were able to identify a *PylT*-like sequence in the genome. The predicted *MaPylT* sequence and *MlPylT* sequence have previously been identified21 and BLAST searches, using these sequences as a reference, allowed us to identify predicted sequences for *G1*, *H5* and *Mt PylT* in the relevant genomes. The identified tRNAs are predicted to fold into clover-leaf structures with a similar length of D-arm, T-arm, acceptor stem and anticodon stem to *Mm*^{Pyl}tRNA$_{CUA}$ and *Dh* ^{Pyl}tRNA$_{CUA}$ (Supplementary Fig. 3). Interestingly, the majority of tRNA sequences identified contained nucleotide loops or bulges in the anticodon stem; to the best of our knowledge this is a hitherto uncharacterized feature in cellular tRNAs.

The five (*Ma*, *G1*, *H5*, *Mt*, *Ml*) PylRS/^{Pyl}tRNA$_{CUA}$ pairs lack the N-terminal domain commonly considered to be required for *in vivo* activity, and we were curious whether these pairs are sufficient to support co-translational read-through of the amber codon *in vivo*. To investigate this possibility we cloned synthetic genes for each synthetase/tRNA pair into a pKW vector for expression in *E. coli* and tested their ability to produce GFPHis$_6$ from *GFP(150TAG)His$_6$* in the presence and absence of Nε-((tert-butoxy)carbonyl)-L-lysine (BocK, **1,** Fig. 1). We anticipated that BocK would be a good substrate for the five synthetases because their active sites are homologous to those of *Mb*PylRS and *Mm*PylRS (Supplementary Fig. 2), both of which efficiently incorporate BocK22.

Comparing the GFPHis$_6$ produced by each pair in the presence and absence of BocK to the GFPHis$_6$ produced by the *Mm*PylRS/*Mm*^{Pyl}tRNA$_{CUA}$ pair in the presence and absence of BocK allowed us to assess the activity and orthogonality of the pairs relative to a system

commonly used for genetic code expansion (Fig. 2b). The activity of the *Ma*PylRS/*Ma*^{Pyl}tRNA$_{CUA}$ pair with BocK was greater than that of the *Mm*PylRS/*Mm*^{Pyl}tRNA$_{CUA}$ pair with BocK, and the two pairs had comparably low levels of GFP production in the absence of BocK. We conclude that the *Ma*PylRS/*Ma*^{Pyl}tRNA$_{CUA}$ pair is active and that *Ma*^{Pyl}tRNA$_{CUA}$ is functionally orthogonal with respect to the aminoacyl-tRNA sythetases in *E. coli*.

Electrospray ionization mass spectrometry of GFP-His6 produced from *GFP(150TAG)His$_6$* by the *Mm*PylRS/*Mm*^{Pyl}tRNA$_{CUA}$ pair in the presence of BocK reveals a single peak corresponding to the incorporation of BocK in response to the amber stop codon (Fig. 2c, Supplementary Fig. 4), and confirms that *Mm*PylRS is functionally orthogonal with respect to the set of tRNAs in *E. coli*. Similarly, the *Ma*PylRS/*Ma*^{Pyl}tRNA$_{CUA}$ pair incorporates a single BocK into GFP in response to the amber codon (Fig. 2d, Supplementary Fig. 4) indicating that *Ma*PylRS is functionally orthogonal in *E. coli*. Taken together our observations reveal that the *Ma*PylRS/*Ma*^{Pyl}tRNA$_{CUA}$ pair is a new extremely active and orthogonal aminoacyl-tRNA synthetase/tRNA pair in *E. coli*. Similarly, we find that the *G1*PylRS/*G1*^{Pyl}tRNA$_{CUA}$ pair is an orthogonal, albeit slightly less active, orthogonal pair in *E. coli (*Fig. 2b,e, Supplementary Fig. 4*)*. The *H5*PylRS/ *H5*^{Pyl}tRNA$_{CUA}$ pair, *Ml*PylRS/ *Ml*^{Pyl}tRNA$_{CUA}$ pair and *Mt*PylRS/*Mt*^{Pyl}tRNA$_{CUA}$ pair are less active than the *Mm*PylRS/*Mm*^{Pyl}tRNA$_{CUA}$ pair. The *H5*^{Pyl}tRNA$_{CUA}$ and *Ml*^{Pyl}tRNA$_{CUA}$ are orthogonal while the *Mt*PylRS/*Mt*^{Pyl}tRNA$_{CUA}$ pair facilitates read through of the amber stop codon in the absence of BocK. Because of the lower activity of these three pairs we did not investigate them further. Our experiments identified the *Ma*PylRS/*Ma*^{Pyl}tRNA$_{CUA}$ pair and *G1*^{Pyl}RS/*G1*^{Pyl}tRNA$_{CUA}$ pair as orthogonal, and very active, pairs for further investigation. In additional experiments we demonstrated that the anticodon of the *Ma*PylRS/ *Ma*^{Pyl}tRNA$_{CUA}$ pair can be altered to facilitate the decoding of additional codons (Supplementary Fig. 5).

The high activity of the *Ma*PylRS/*Ma*^{Pyl}tRNA$_{CUA}$ pair and the low activity of the *Ml*^{Pyl}RS/*Ml*^{Pyl}tRNA$_{CUA}$ pair are consistent with the higher percentage of amber stop codons predicted to be used for termination in the *Ml* genome (11%) versus the *Ma* genome (1.6%), the number of amber stop codons in open reading frames predicted to incorporate pyrrolysine in *Ml* (3) versus *Ma* (19), and the predicted methylamine availability in the evolutionary history of the two organisms21.

Mutual orthogonality amongst natural PylRS/^{Pyl}tRNA pairs

Next we investigated the orthogonality of *Mm*PylRS, *Ma*PylRS, and *G1*PylRS with respect to *Mm*^{Pyl}tRNA$_{CUA}$, *Ma*^{Pyl}tRNA$_{CUA}$, and *G1*^{Pyl}tRNA$_{CUA}$ by measuring the ability of each of the nine synthetase/tRNA combinations to produce GFPHis$_6$ from *GFP(150TAG)His$_6$* in the presence of BocK (Fig. 3). We found that *Mm*PylRS functions with *Ma*^{Pyl}tRNA$_{CUA}$ about as efficiently as with its cognate *Mm*^{Pyl}tRNA$_{CUA}$. In contrast, *Mm*PylRS functioned 24 times less efficiently with *G1*^{Pyl}tRNA$_{CUA}$ than with its cognate tRNA. *Ma*PylRS does not function with *Mm*^{Pyl}tRNA$_{CUA}$ but does aminoacylate *G1*^{Pyl}tRNA$_{CUA}$. *G1*PylRS functions with *Ma*^{Pyl}tRNA$_{CUA}$ but not with *Mm*^{Pyl}tRNA$_{CUA}$.

Thus, the *Ma*PylRS/*Ma*^{Pyl}tRNA$_{CUA}$ pair and the *G1*PylRS/*G1*^{Pyl}tRNA$_{CUA}$ pair are not mutually orthogonal, as each synthetase functions with its non-cognate tRNA. And the *Mm*PylRS/*Mm*^{Pyl}tRNA$_{CUA}$ pair and the *Ma*PylRS/*Ma*^{Pyl}tRNA$_{CUA}$ pair are not mutually orthogonal because *Mm*PylRS functions with *Ma*^{Pyl}tRNA$_{CUA}$.

However, the *Mm*PylRS/*Mm*^{Pyl}tRNA$_{CUA}$ pair and the *G1*PylRS/*G1*^{Pyl}tRNA$_{CUA}$ pairs are mutually orthogonal, though there is clearly a low level of activity of *Mm*PylRS with *G1*^{Pyl}tRNA$_{CUA}$. This surprising observation reveals that there is sufficient divergence between PylRS/^{Pyl}tRNA$_{CUA}$ recognition within methanogenic archea to generate mutually orthogonal pairs. We are not aware of any prior examples of natural mutual orthogonality between synthetase/tRNA pairs used for the same amino acid within a domain of life.

Evolving mutually orthogonal PylRS/^{Pyl}tRNA$_{CUA}$ pairs

Next we wanted to create a pair that was optimized for both activity and mutual orthogonality with respect to the *Mm*PylRS/*Mm*^{Pyl}tRNA$_{CUA}$ pair. We considered either improving both the orthogonality and the activity of the *G1* pair, or improving the orthogonality of the *Ma*^{Pyl}tRNA$_{CUA}$ with respect to *Mm*PylRS, while maintaining the activity of the *Ma*PylRS/*Ma*^{Pyl}tRNA$_{CUA}$ pair. We decided to address the later challenge as we had a clear hypothesis about how to improve the orthogonality of *Ma*^{Pyl}tRNA$_{CUA}$ with minimal effect on the activity of the *Ma*PylRS/*Ma*^{Pyl}tRNA$_{CUA}$ pair. The N-terminal domain of *Mm*PylRS, which is absent in *Ma*PylRS, binds to the T-arm and variable loop regions of its cognate tRNA16, 17. We therefore reasoned that it might be possible to create variants of *Ma*^{Pyl}tRNA$_{CUA}$ that are not recognized by *Mm*PylRS or endogenous synthetases but still function with *Ma*PylRS, by altering the variable loop of *Ma*^{Pyl}tRNA$_{CUA}$.

To discover variants of *Ma*^{Pyl}tRNA$_{CUA}$ that are orthogonal to *Mm*PylRS and function efficiently with *Ma*PylRS we preformed a positive selection followed by a negative screen on *Ma*^{Pyl}tRNA$_{CUA}$ libraries (Fig. 4a). We first created a library of 64 *Ma*^{Pyl}tRNA$_{CUA}$ mutants by randomising nucleotide positions 41, 42 and 43 in its variable loop to all possible combinations of the four common bases (Fig. 4b). We selected *Ma*^{Pyl}tRNA$_{CUA}$ variants that functioned with *Ma*PylRS and enabled cells to grow on 100 μg ml^{-1} of chloramphenicol in the presence of BocK, by facilitating read through of an amber codon at position 111 of a *chloramphenicol acetyl transferase* reporter, *cat(111TAG)*. Next we performed a negative screen on the selected *Ma*^{Pyl}tRNA$_{CUA}$ variants, to identify tRNAs that do not function with *Mm*PylRS or any endogenous synthetases. Cells bearing *GFP(150TAG)His$_6$*, *Mm*PylRS and each *Ma*^{Pyl}tRNA$_{CUA}$ variant were provided with BocK and screened for the absence of GFPHis$_6$ expression.

This serial positive selection and negative screen identified four unique sequences, which vary nucleotides 41 and 42 but conserve G43 (Fig. 4c). Expression of *GFP(150TAG)His$_6$* in the presence of each evolved *Ma*^{Pyl}tRNA$_{CUA}$ variant, BocK, and either *Mm*PylRS or *Ma*PylRS revealed that the selected sequences do not function with *Mm*PylRS but do function with *Ma*PylRS (Fig. 4d). However, the activity of these evolved *Ma*PylRS/*Ma*^{Pyl}tRNA$_{CUA}$ pairs did not exceed 47% that of the parent *Ma*PylRS/*Ma*^{Pyl}tRNA$_{CUA}$ pair.

Next we investigated whether *Ma*PylRS/*Ma*^{Pyl}tRNA$_{CUA}$ pairs that are more active but still mutually orthogonal to the *Mm*PylRS/*Mm*^{Pyl}tRNA$_{CUA}$ pair can be created by introducing further diversity in the variable loop of *Ma*^{Pyl}tRNA$_{CUA}$. We generated libraries in which the length of the variable loop was expanded from 3 nucleotides to 4, 5 or 6 randomised nucleotides (Fig. 4b). We performed the serial positive selection and negative screen for each of the expanded libraries and characterized the resulting *Ma*^{Pyl}tRNA$_{CUA}$ clones. From the expanded libraries we obtained evolved *Ma*^{Pyl}tRNA$_{CUA}$ sequences with expanded variable loops and non-programmed mutations in either the T-loop or D-loop (e.g. *Ma*^{Pyl}tRNA(6)$_{CUA}$, *Ma*^{Pyl}tRNA(10)CUA, *Ma*^{Pyl}tRNA(14)CUA and *Ma*^{Pyl}tRNA(23)CUA (Fig. 4c)). These tRNAs form *Ma*PylRS/*Ma*^{Pyl}tRNA$_{CUA}$ pairs that are as active as the *Mm*PylRS/*Mm*^{Pyl}tRNA$_{CUA}$ pair, and do not function with *Mm*PylRS (Fig. 4d,e). These experiments demonstrated that the *Mm*PylRS/*Mm*^{Pyl}tRNA$_{CUA}$ pair and several of the evolved *Ma*PylRS/*Ma*^{Pyl}tRNA$_{CUA}$ pairs, with expanded variable loops and high activity, are mutually orthogonal. The activity of the evolved *Ma*PylRS/*Ma*^{Pyl}tRNA(6)$_{CUA}$ pair is 91% of the parent *Ma*PylRS/*Ma*^{Pyl}tRNA$_{CUA}$ and is slightly greater than that of the *G1*PylRS/*G1*^{Pyl}tRNA$_{CUA}$ pair. The evolved *Ma*PylRS/*Ma*^{Pyl}tRNA$_{CUA}$ pairs are more orthogonal than the *G1*PylRS/*G1*^{Pyl}tRNA$_{CUA}$ pair, with respect to the *Mm*PylRS/*Mm*^{Pyl}tRNA$_{CUA}$ pair, since *G1*^{Pyl}tRNA$_{CUA}$ shows a low level of activity with *Mm*PylRS (Fig. 3).

Engineering *Ma*PylRS for selective ncAA incorporation

*Mm*PylRS and *Mb*PylRS have been used extensively for genetic code expansion and the active sites of these enzymes naturally accept several ncAA substrates, but exclude others9, 12. Structure-guided mutagenesis and directed evolution of the active site of *Mb*PylRS and *Mm*PylRS has expanded the repertoire of ncAAs that can be incorporated23, 24, and created enzymes that are selective for some ncAAs with respect to others25. Mutants discovered in *Mb*PylRS can commonly be transplanted to *Mm*PylRS and vice versa.

In light of the extensive homology in the amino acid-binding pockets of *Mm*PylRS and *Ma*PylRS (Supplementary Fig. 2), we investigated whether the substrate specificity of *Ma*PylRS could be expanded through the introduction of mutations identified in the active sites of *Mm*PylRS or *Mb*PylRS. The Y306A, Y384F double mutant of *Mm*PylRS (*Mm*PylRS-AF) expands its substrate specificity to facilitate the incorporation of several additional ncAAs23. Using an alignment of *Mm*PylRS and *Ma*PylRS (Supplementary Fig. 2) we identified Y126 and Y206 in *Ma*PylRS as the residues corresponding to Y306 and Y384 in *Mm*PylRS, and created the corresponding Y126A and Y206F mutations in *Ma*PylRS (*Ma*PylRS-AF). The *Mm*PylRS-AF/*Mm*^{Pyl}tRNA$_{CUA}$ pair facilitates the incorporation of a bicyclononyne-containing amino acid (BCNK, **2**), which is poorly incorporated by *Mm*PylRS/*Mm*^{Pyl}tRNA$_{CUA}$ pair (Fig. 5a)26, 27. Similarly we find that the *Ma*PylRS/*Ma*^{Pyl}tRNA$_{CUA}$ pair inefficiently incorporates BCNK, and that the efficiency of BCNK incorporation is substantially improved by using the *Ma*PylRS-AF/*Ma*^{Pyl}tRNA$_{CUA}$ pair (Fig. 5a). These experiments demonstrated that the mutations identified in *Mm*PylRS can be transferred to *Ma*PylRS to expand its substrate specificity.

Next we asked whether we could create *Ma*PylRS and *Mm*PylRS variants with mutually orthogonal ncAA substrate specificity. We recently demonstrated that parallel positive

selections on a *Mb*PylRS library in the presence of either Nε - benzyloxycarbonyl-L-lysine (CbzK, **3**) or *N*ε-(((2-methylcycloprop-2-en-1-yl)methoxy)carbonyl)-L-lysine (CypK**, 4**) or in the absence of ncAA, coupled to deep sequencing and comparative sequence analysis, enables the direct identification of *Mb*PylRS variants that incorporate CbzK, but not CypK or any natural amino acids, and vice versa25. *Mb*PylRS-MutRS1 (Y271M, L274G, C313T) selectively incorporates CbzK, but not CypK; we introduced the corresponding mutations into *Ma*PylRS, creating *Ma*PylRS-MutRS1 (Y126M, M129G,V168T). We found that the *Ma*PylRS-MutRS1/*Ma*^{Pyl}tRNA$_{CUA}$ pair directs the incorporation of CbzK but not CypK (Fig. 5b). *Mb*PylRS-MutRS2 (A267S) selectively incorporates CypK, but not CbzK. We found that the *Mm*PylRS-MutRS2 (A302S)/*Mm*^{Pyl}tRNA$_{CUA}$ pair directs the incorporation of CypK but not CbzK (Supplementary Fig. 6); we subsequently found that the wild-type *Mm*PylRS/*Mm*^{Pyl}tRNA$_{CUA}$ pair directs the incorporation of CypK but not CbzK with comparable efficiency and specificity to the mutant, and used this pair for further experiments (Fig. 5b). These experiments demonstrated that active sites of *Ma*PylRS and *Mm*PylRS can be diverged for mutually orthogonal ncAA incorporation.

Encoding distinct ncAAs with mutually orthogonal PylRS/^{Pyl}tRNA pairs

Next we demonstrated that the *Ma*PylRS-MutRS1/*Ma*^{Pyl}tRNA(6) and *Mm*PylRS/*Mm*^{Pyl}tRNA pairs can function as mutually orthogonal pairs in the same cell. We created a single vector encoding both the *Ma*PylRS-MutRS1/*Ma*^{Pyl}tRNA(6)$_{CUA}$ pair and the *Mm*PylRS/*Mm*^{Pyl}tRNA$_{UCCU}$ pair, and introduced this into cells along with *GFP(150TAG)His$_6$*. Addition of the *Ma*PylRS-MutRS1 substrate, CbzK, led to robust GFPHis$_6$ synthesis, while addition of the *Mm*PylRS substrate, CypK, led to minimal synthesis of GFPHis$_6$, and addition of both CbzK and CypK led to robust GFPHis$_6$ synthesis (Fig. 6a, Supplementary Fig. 7). Mass spectrometry demonstrates that GFPHis$_6$ produced from *GFP(150TAG)His$_6$* in the presence of the *Ma*PylRS-MutRS1/*Ma*^{Pyl}tRNA(6)$_{CUA}$ and *Mm*PylRS/*Mm*^{Pyl}tRNA$_{UCCU}$ pairs and both CbzK and CypK incorporates CbzK but not CypK (Fig 6b, Supplementary Fig. 8). In complementary experiments we introduced the *Ma*PylRS-MutRS1/*Ma*^{Pyl}tRNA(6)$_{UACU}$ and *Mm*PylRS/*Mm*^{Pyl}tRNA$_{CUA}$ pairs, along with *GFP(150TAG)His$_6$*, into cells. Addition of the *Mm*PylRS substrate, CypK, led to robust GFPHis$_6$ synthesis, while addition of the *Ma*PylRS-MutRS1 substrate, CbzK, led to minimal synthesis of GFPHis$_6$, and addition of both CbzK and CypK led to robust GFPHis$_6$ synthesis (Fig. 6c, Supplementary Fig. 7). Mass spectrometry demonstrates that GFPHis$_6$ produced from *GFP(150TAG)His$_6$* in the presence of the *Ma*PylRS-MutRS1/*Ma*^{Pyl}tRNA(6)$_{UACU}$ and *Mm*PylRS/*Mm*^{Pyl}tRNA$_{CUA}$ pairs and both CbzK and CypK incorporates CypK, but not CbzK (Fig 6d, Supplementary Fig. 8).

These experiments demonstrated the modular combination of mutations in *Ma*^{Pyl}tRNA(6)$_{CUA}$, that confer orthogonality with respect to *Mm*PylRS, with the mutations in *Ma*PylRS-MutRS1, that re-program the ncAA specificity of this enzyme, to produce a pair that combines the properties conferred by each set of mutations. Moreover, these experiments demonstrated that the *Ma*PylRS-MutRS1/*Ma*^{Pyl}tRNA(6) and *Mm*PylRS/*Mm*^{Pyl}tRNA pairs can function independently in the same cell.

Finally we demonstrated that we can encode distinct ncAAs into a single polypeptide using the mutually orthogonal PylRS/^{Pyl}tRNA pairs. We introduced the *Ma*PylRS-MutRS1/*Ma*^{Pyl}tRNA(6)$_{CUA}$ pair and *Mm*PylRS/*Mm*^{Pyl}tRNA$_{UCCU}$ pair into cells that produce ribo-Q, an orthogonal ribosome that is directed to an orthogonal message and facilitates decoding of quadruplet and amber codons by cognate tRNAs4. Cells also contained *o-GST-CaM(1UAG, 40AGGA*), an orthogonal message for a *glutathione-S-transferase-calmodulin* fusion that is read by ribo-Q1 and contains amber and quadruplet codons that are complementary to the anticodons of *Ma*^{Pyl}tRNA(6)$_{CUA}$ and *Mm*^{Pyl}tRNA$_{UCCU}$. Addition of both CbzK and CypK to cells was required for efficient full length GST-CaM synthesis, with an unoptimized yield of approximately 0.5 mg per L of culture, from the orthogonal message (Fig. 6e, Supplementary Fig. 7). Additional experiments and mass spectrometry confirmed the programmed incorporation of both ncAAs into a single polypeptide (Supplementary Fig. 9). These experiments demonstrated that the *Ma*PylRS-MutRS1/*Ma*^{Pyl}tRNA(6)$_{CUA}$ and *Mm*PylRS/*Mm*^{Pyl}tRNA$_{UCCU}$ pairs can function in the same cell to incorporate distinct substrates into a single polypeptide in response to distinct codons.

Discussion

The N-terminal domain of PylRS is commonly believed to be essential for robust activity *in vivo*. We have discovered PylRS variants that lack the N-terminal domain but are exceptionally active and orthogonal in *E. coli* and can be used to direct the incorporation of ncAAs. These single domain PylRS variants may provide an improved starting point for engineering new substrate specificity. We have shown that the *G1*PylRS/*G1*^{Pyl}tRNA pair is mutually orthogonal with the *Mm*PylRS/*Mm*^{Pyl}tRNA pair. Our results reveal that there is sufficient divergence between PylRS/^{Pyl}tRNA$_{CUA}$ recognition within methanogenic archea to generate mutually orthogonal pairs. We are not aware of any prior examples of natural mutual orthogonality between synthetase/tRNA pairs used for the same amino acid within a domain of life and our results suggest it will be interesting to further explore the mutual orthogonality of aaRS/tRNA pairs from distinct archeal orders.

We have created exceptionally active variants of the *Ma*PylRS/*Ma*^{Pyl}tRNA pair that have a high level of mutual orthogonality with respect to the *Mm*PylRS/*Mm*^{Pyl}tRNA pair. Moreover we have shown that the active sites of the *Ma*PylRS/*Ma*^{Pyl}tRNA and *Mm*PylRS/*Mm*^{Pyl}tRNA pairs can be diverged to recognize distinct substrates, and that the pairs can be used together to program distinct ncAAs into proteins in *E. coli*. We anticipate that the *Ma*PylRS/*Ma*^{Pyl}tRNA(6) pair and its derivatives may be used to further expand the combinations of ncAAs that can be encoded into polypeptides synthesized in *E. coli*. It will be interesting to explore whether the new PylRS/tRNA pairs reported herein are also orthogonal in eukaryotic cells and organisms.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

This work was supported by the Medical Research Council, UK (MC_U105181009 and MC_UP_A024_1008), BBSRC (BB/M000842/1, for automation) and the ERC Advanced Grant (SGCR), all to J.W.C. J.C.W.W. was supported by the EPSRC Nanoscience and Nanotechnology CDT at Cambridge University.

References

1Wang K, et al. Optimized orthogonal translation of unnatural amino acids enables spontaneous protein double-labelling and FRET. Nature chemistry. 2014; 6:393–403.

2Sachdeva A, Wang K, Elliott T, Chin JW. Concerted, rapid, quantitative, and site-specific dual labeling of proteins. Journal of the American Chemical Society. 2014; 136:7785–7788. [PubMed: 24857040]

3Chatterjee 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

4Neumann 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. [PubMed: 20154731]

5Chin JW. Expanding and reprogramming the genetic code. Nature. 2017; 550:53–60. [PubMed: 28980641]

6Neumann H, Slusarczyk AL, Chin JW. De NovoGeneration of Mutually Orthogonal Aminoacyl-tRNA Synthetase/tRNA Pairs. Journal of the American Chemical Society. 2010; 132:2142–2144. [PubMed: 20121121]

7Chatterjee A, Xiao H, Schultz PG. Evolution of multiple, mutually orthogonal prolyl-tRNA synthetase/tRNA pairs for unnatural amino acid mutagenesis in Escherichia coli. P Natl Acad Sci USA. 2012; 109:14841–14846.

8Anderson JC, et al. An expanded genetic code with a functional quadruplet codon. P Natl Acad Sci USA. 2004; 101:7566–7571.

9Chin JW. Expanding and Reprogramming the Genetic Code of Cells and Animals. Annu Rev Biochem. 2014; 83:379–408. [PubMed: 24555827]

10Ambrogelly A, et al. Pyrrolysine is not hardwired for cotranslational insertion at UAG codons. Proc Natl Acad Sci U S A. 2007; 104:3141–3146. [PubMed: 17360621]

11Elliott TS, et al. Proteome labeling and protein identification in specific tissues and at specific developmental stages in an animal. Nature biotechnology. 2014; 32:465–472.

12Dumas A, Lercher L, Spicer CD, Davis BG. Designing logical codon reassignment - Expanding the chemistry in biology. Chem Sci. 2015; 6:50–69. [PubMed: 28553457]

13Wang K, et al. Defining synonymous codon compression schemes by genome recoding. Nature. 2016; 539:59–64. [PubMed: 27776354]

14Ostrov N, et al. Design, synthesis, and testing toward a 57-codon genome. Science. 2016; 353:819–822. [PubMed: 27540174]

15Zhang Y, et al. A semi-synthetic organism that stores and retrieves increased genetic information. Nature. 2017; 551:644–647. [PubMed: 29189780]

16Jiang R, Krzycki JA. PylSn and the homologous N-terminal domain of pyrrolysyl-tRNA synthetase bind the tRNA that is essential for the genetic encoding of pyrrolysine. J Biol Chem. 2012; 287:32738–32746. [PubMed: 22851181]

17Suzuki T, et al. Crystal structures reveal an elusive functional domain of pyrrolysyl-tRNA synthetase. Nat Chem Biol. 2017; 13:1261–1266. [PubMed: 29035363]

18Kavran JM, et al. Structure of pyrrolysyl-tRNA synthetase, an archaeal enzyme for genetic code innovation. Proc Natl Acad Sci U S A. 2007; 104:11268–11273. [PubMed: 17592110]

19Herring S, et al. The amino-terminal domain of pyrrolysyl-tRNA synthetase is dispensable in vitro but required for in vivo activity. FEBS letters. 2007; 581:3197–3203. [PubMed: 17582401]

20Sollinger A, et al. Phylogenetic and genomic analysis of Methanomassiliicoccales in wetlands and animal intestinal tracts reveals clade-specific habitat preferences. FEMS Microbiol Ecol. 2016; 92

21Borrel 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. [PubMed: 25124552]

22Virdee S, Ye Y, Nguyen DP, Komander D, Chin JW. Engineered diubiquitin synthesis reveals Lys29-isopeptide specificity of an OTU deubiquitinase. Nat Chem Biol. 2010; 6:750–757. [PubMed: 20802491]

23Yanagisawa T, et al. Multistep Engineering of Pyrrolysyl-tRNA Synthetase to Genetically Encode Nε-(o-Azidobenzyloxycarbonyl) lysine for Site-Specific Protein Modification. Chemistry & Biology. 2008; 15:1187–1197. [PubMed: 19022179]

24Neumann H, Peak-Chew SY, Chin JW. Genetically encoding Nε-acetyllysine in recombinant proteins. Nat Chem Biol. 2008; 4:232–234. [PubMed: 18278036]

25Zhang MS, et al. Biosynthesis and genetic encoding of phosphothreonine through parallel selection and deep sequencing. Nat Methods. 2017; 14:729–736. [PubMed: 28553966]

26Borrmann A, et al. Genetic Encoding of a Bicyclo[6.1.0]nonyne-Charged Amino Acid Enables Fast Cellular Protein Imaging by Metal-Free Ligation. ChemBioChem. 2012; 13:2094–2099. [PubMed: 22945333]

27Lang K, et al. Genetic Encoding of Bicyclononynes and trans-Cyclooctenes for Site-Specific Protein Labeling in Vitro and in Live Mammalian Cells via Rapid Fluorogenic Diels–Alder Reactions. Journal of the American Chemical Society. 2012; 134:10317–10320. [PubMed: 22694658]

Europe PMC Funders Author Manuscripts

Summary

Pyrrolysyl-tRNA synthetase(PylRS)/^{Pyl}tRNA$_{CUA}$ pairs that lack the N-terminal domain but are active and orthogonal are discovered, and pairs that are mutually orthogonal to existing PylRS/^{Pyl}tRNA$_{CUA}$ pairs are developed. Mutually orthogonal PylRS/^{Pyl}tRNA pairs are combined to genetically encode the incorporation of distinct ncAAs into proteins synthesized in *E. coli.*

Europe PMC Funders Author Manuscripts

Europe PMC Funders Author Manuscripts

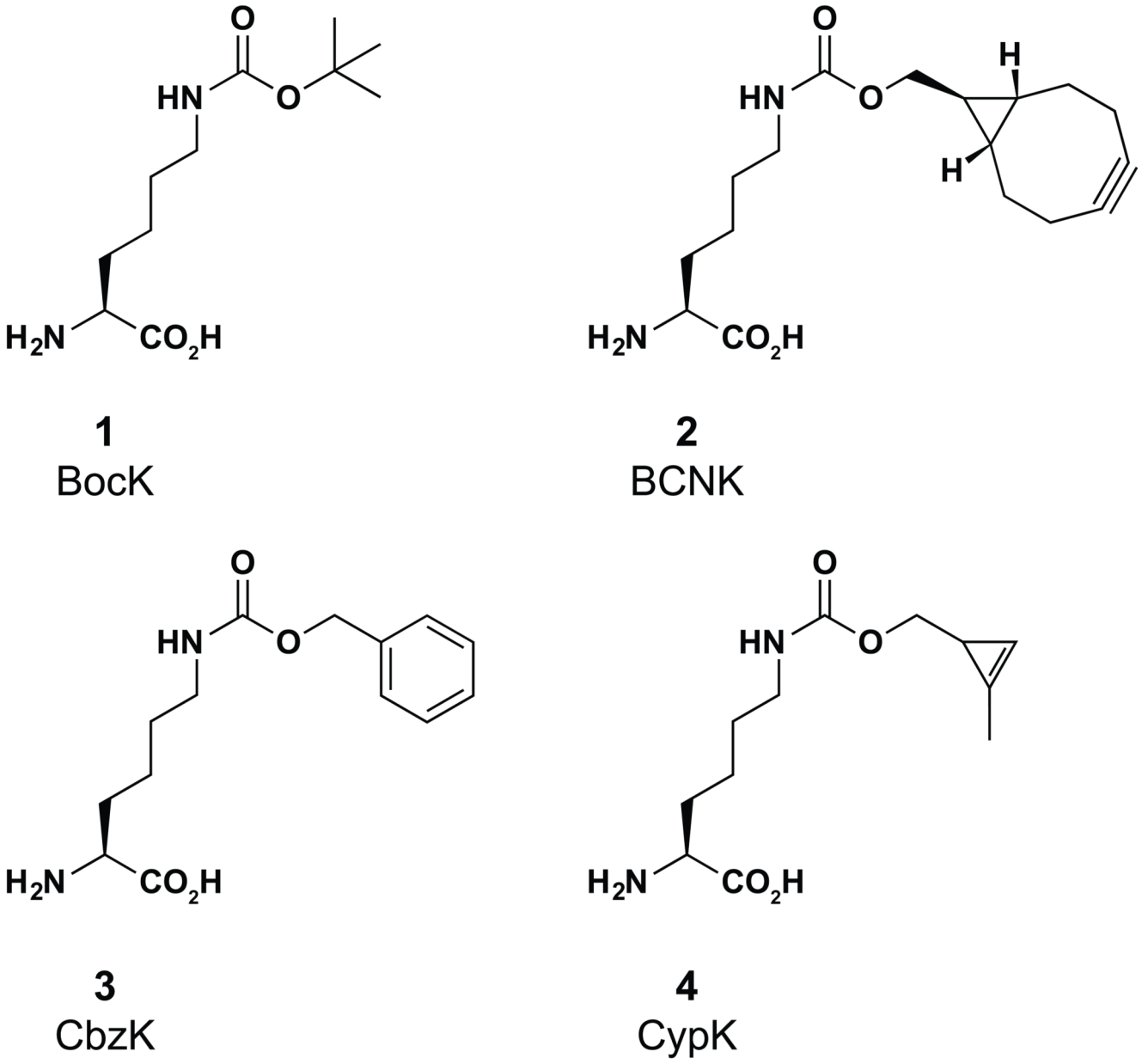

Figure 1. The structures of the amino acids used in this work.

1, Nε-((tertbutoxy)carbonyl)-L-lysine; **2** Nε-(((1R,8S)-bicyclo[6.1.0]non-4-yn-9-ylmethoxy)carbonyl)-L-lysine, exo isomer shown, 4:1 exo/endo mixture used; **3**, Nε-benzyloxycarbonyl-L-lysine; **4**, Nε-(((2-methylcycloprop-2-en-1-yl)methoxy)carbonyl)-L-lysine.

Europe PMC Funders Author Manuscripts

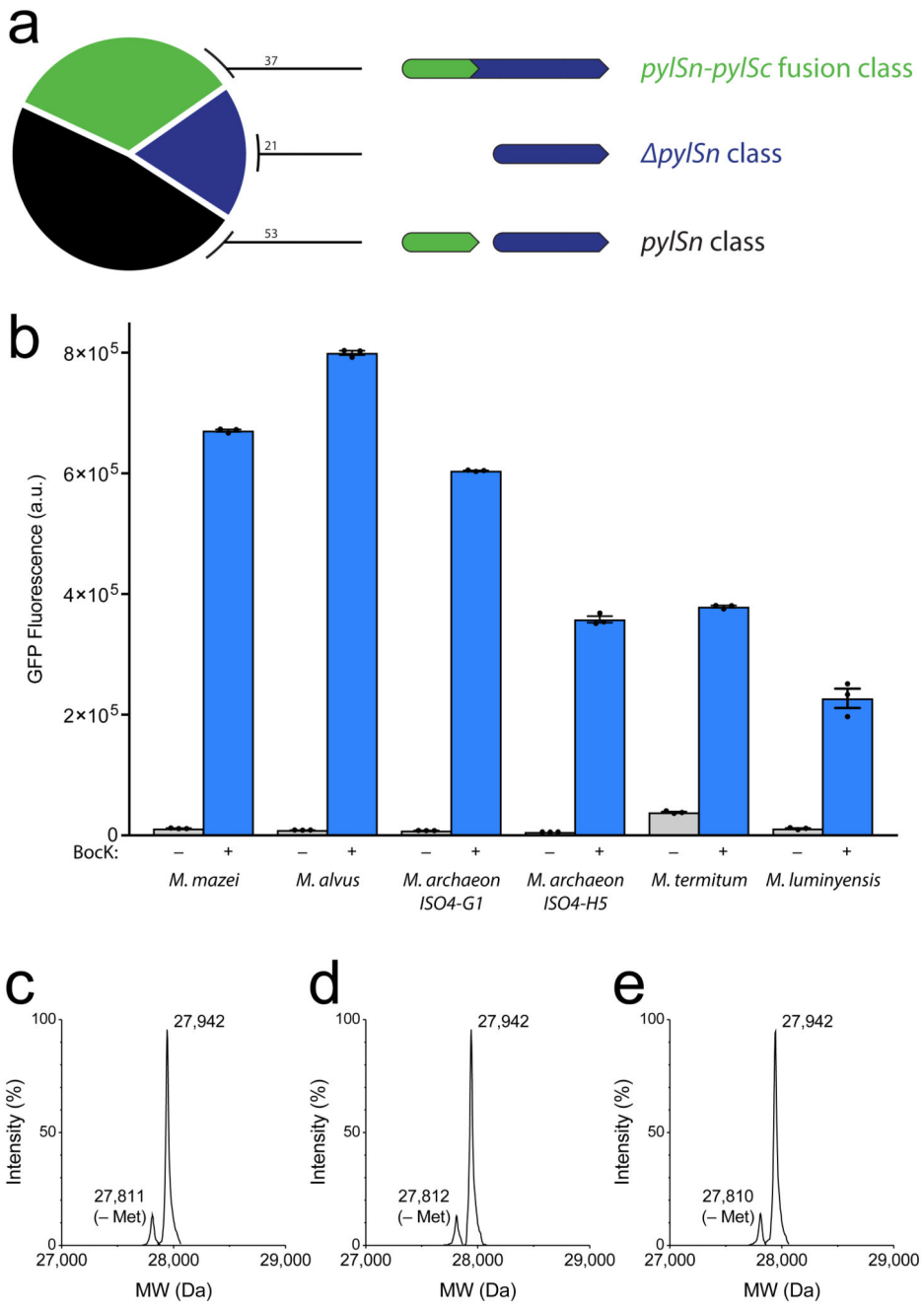

Figure 2. Identifying ΔNPylRS/^{Pyl}tRNA pairs are active and orthogonal in *E. coli*

(a) Classification of identified PylRS sequences according to the presence of a domain homologous to *D. hafniense* PylSn either within the same gene (*pylSn-pylSc* fusion class), present within a separate gene in the same genome (*pylSn* class), or absent entirely (*ΔpylSn* class). (b) *In vivo* amber suppression activity assay using *E. coli* DH10B bearing pBAD GFP(150TAG)His_6 and the corresponding pKW PylRS/^{Pyl}$tRNA_{CUA}$ plasmid in the presence and absence of BocK (**1**) demonstrates the activity of ΔNPylRS/^{Pyl}$tRNA_{CUA}$ pairs and the orthogonality of several ^{Pyl}tRNAs with respect to *E. coli* aminoacyl-tRNA sythetases. Each

data point shows the mean of three technical replicates that form one biological replicate; the error bars show the mean and SEM of three independent biological replicates. (c), (d), (e) Confirmation of specific BocK incorporation into purified GFPHis$_6$ by *Mm*PylRS/*Mm*^{Pyl}tRNA$_{CUA}$, *Ma*PylRS/*Ma*^{Pyl}tRNA$_{CUA}$ and *G1*PylRS/*G1*^{Pyl}tRNA$_{CUA}$ analysed by electrospray ionisation mass spectrometry (ESI-MS, predicted mass 27,942 Da, observed mass 27,942 Da) reveals the functional orthogonality of each PylRS with respect to *E. coli* tRNAs. The ESI-MS experiments in (c), (d), (e) were performed once. Raw ESI-MS spectra are provided in Supplementary Fig. 4.

Europe PMC Funders Author Manuscripts

Europe PMC Funders Author Manuscripts

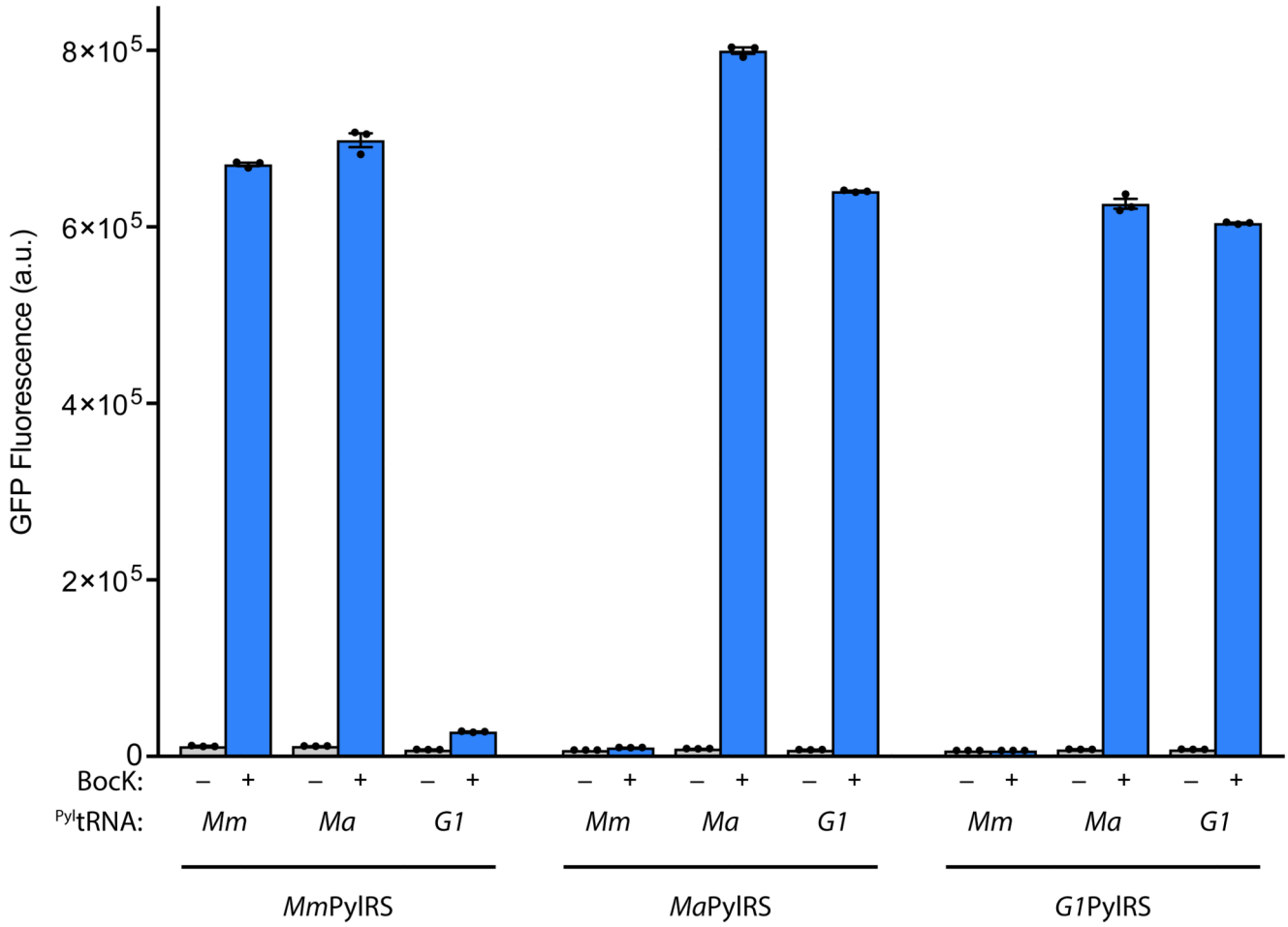

Figure 3. Mutual orthogonality amongst natural PylRS/^{Pyl}tRNA pairs.

In vivo amber suppression activity assay using *E. coli* DH10B bearing pBAD GFP(150TAG)His_6 and the corresponding pKW PylRS/$^{Pyl}tRNA_{CUA}$ plasmid in the presence and absence of BocK (**1**) demonstrates the activity of the indicated PylRS with cognate and non cognate $^{Pyl}tRNA_{CUA}$. Each data point shows the mean of three technical replicates that form one biological replicate; the error bars show the mean and SEM of three independent biological replicates.

Europe PMC Funders Author Manuscripts

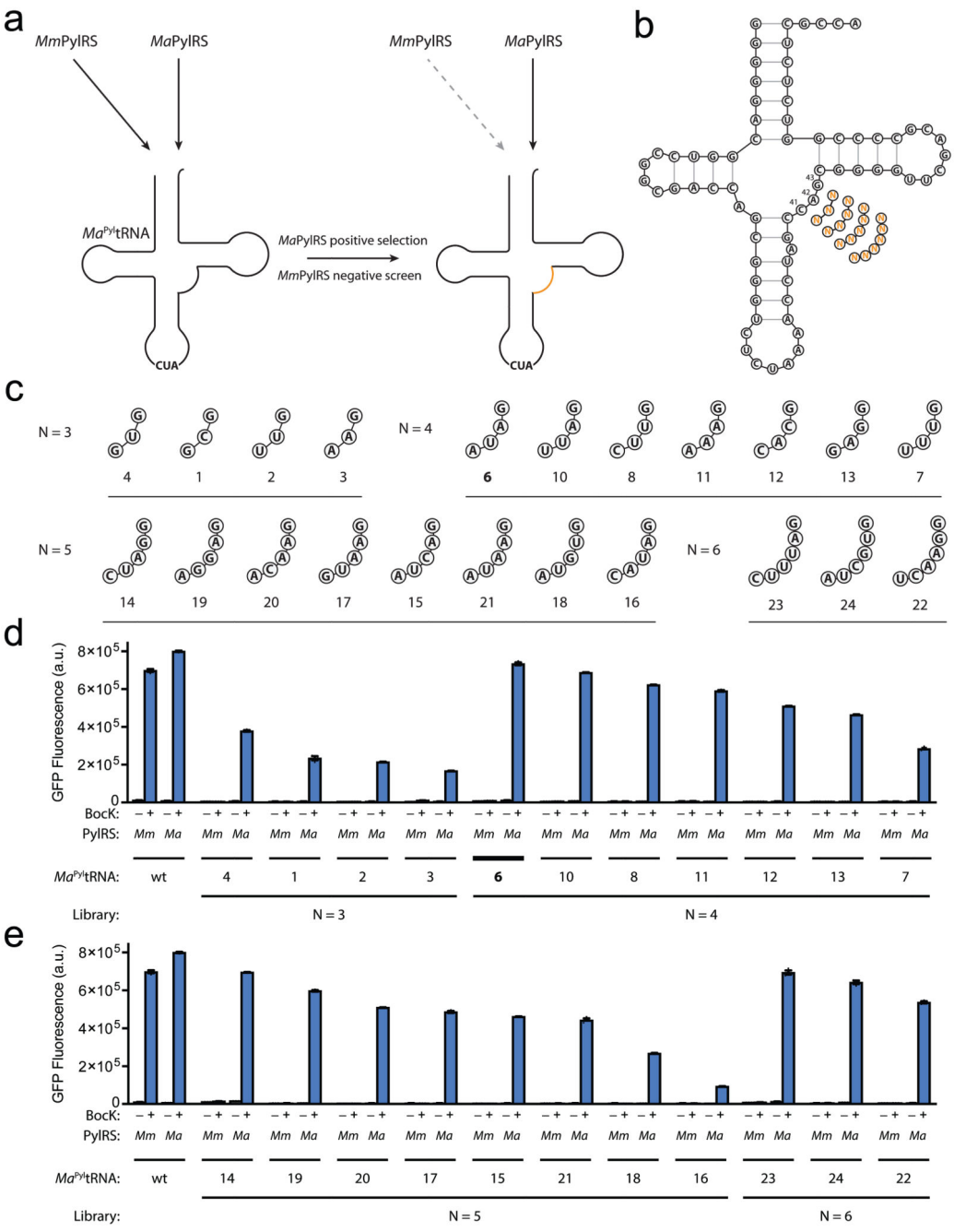

Figure 4. Evolution of *Ma*^{Pyl}tRNA to create mutually orthogonal PylRS/^{Pyl}tRNA pairs.

(a) Evolution *Ma*^{Pyl}tRNA to abolish its function with *Mm*PylRS while preserving its function with *Ma*PylRS, via positive selection in the presence of *Ma*PylRS followed by a negative screen in the presence of *Mm*PylRS. (b) Libraries of *Ma*^{Pyl}tRNA created by randomising or expanding the length of the variable loop to 4, 5 or 6 randomised nucleotides. (c) Variable loop sequences for the *Ma*^{Pyl}tRNA hits identified. Hits are grouped according to the library from which they were identified. Hits 6, 8, 10, 12, 15, 16, 17, 18 and 20 also contained a G55A mutation; hits 14, 19, 22, 23 and 24 also contained a G55T

mutation; hits 11, 13 and 21 also contained a G55C mutation; hit 7 also contained a C15T mutation. (d), (e) *In vivo* amber suppression activity assay (libraries N = 3 and 4 (d) and libraries N = 5 and 6 (e)). Each data point shows the mean of three technical replicates that form one biological replicate; the error bars show the mean and SEM of three independent biological replicates. wt, wild-type

Europe PMC Funders Author Manuscripts

Europe PMC Funders Author Manuscripts

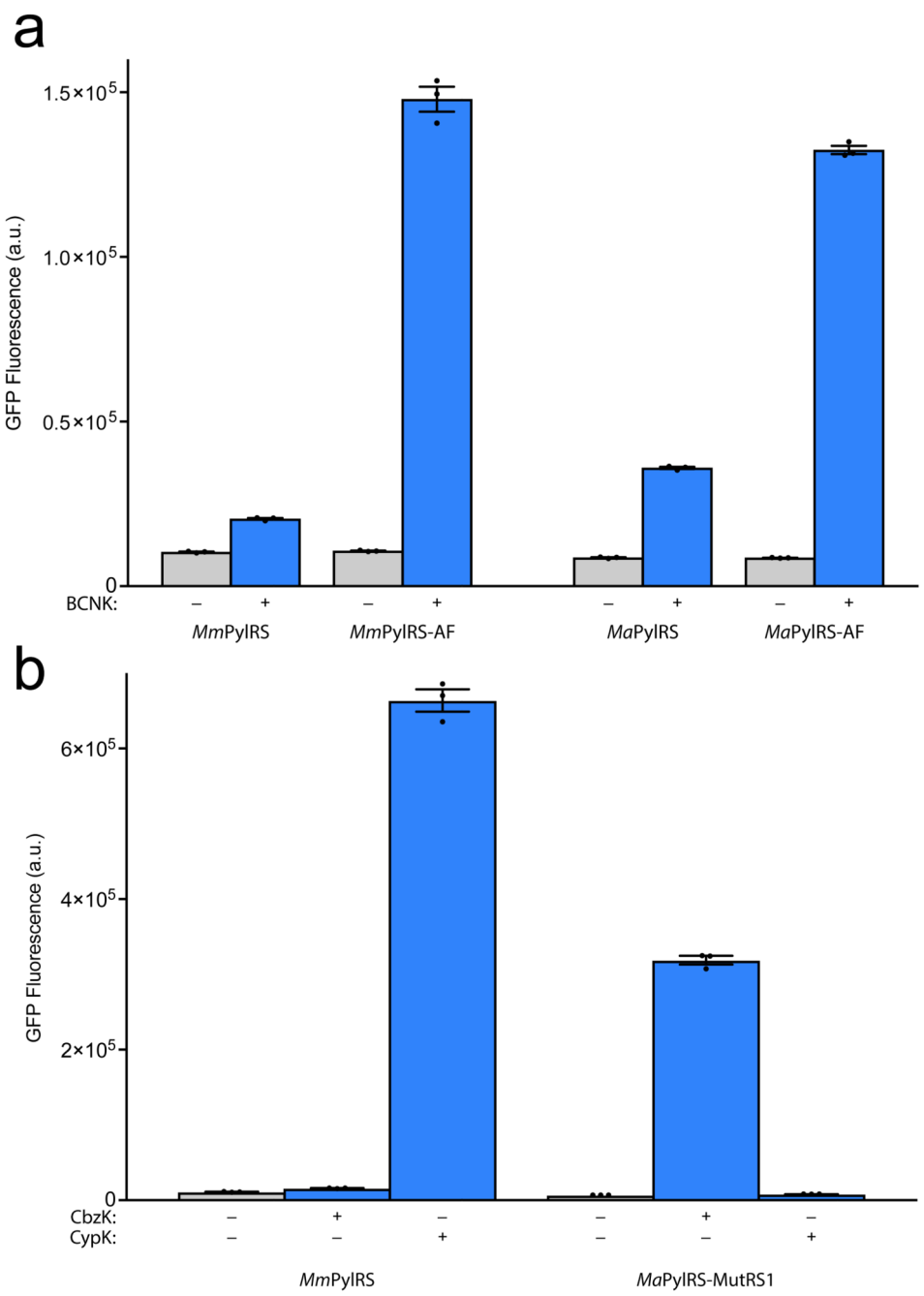

Figure 5. Engineering the active site of *Ma*PylRS for selective ncAA incorporation.

(a) *In vivo* amber suppression activity assay using *E. coli* DH10B bearing pBAD GFP(150TAG)His_6 and the corresponding pKW PylRS/$^{Pyl}tRNA_{CUA}$ plasmid in the presence and absence of BCNK (**2**) demonstrates the transferability of the *Mm*PylRS-AF mutations into *Ma*PylRS to facilitate improved incorporation of BCNK using *Ma*PylRS-AF. Each data point shows the mean of three technical replicates that form one biological replicate; the error bars show the mean and SEM of three independent biological replicates. (b) *In vivo* amber suppression activity assay using *E. coli* DH10B bearing pBAD GFP(150TAG)His_6

and the corresponding pKW PylRS/^{Pyl}tRNA$_{CUA}$ plasmid in the presence and absence of CbzK (**3**) and CypK (**4**) demonstrates the selective incorporation of CypK by *Mm*PylRS and the selective incorporation of CbzK by *Ma*PylRS-MutRS1. Each data point shows the mean of three technical replicates that form one biological replicate; the error bars show the mean and SEM of three independent biological replicates.

Europe PMC Funders Author Manuscripts

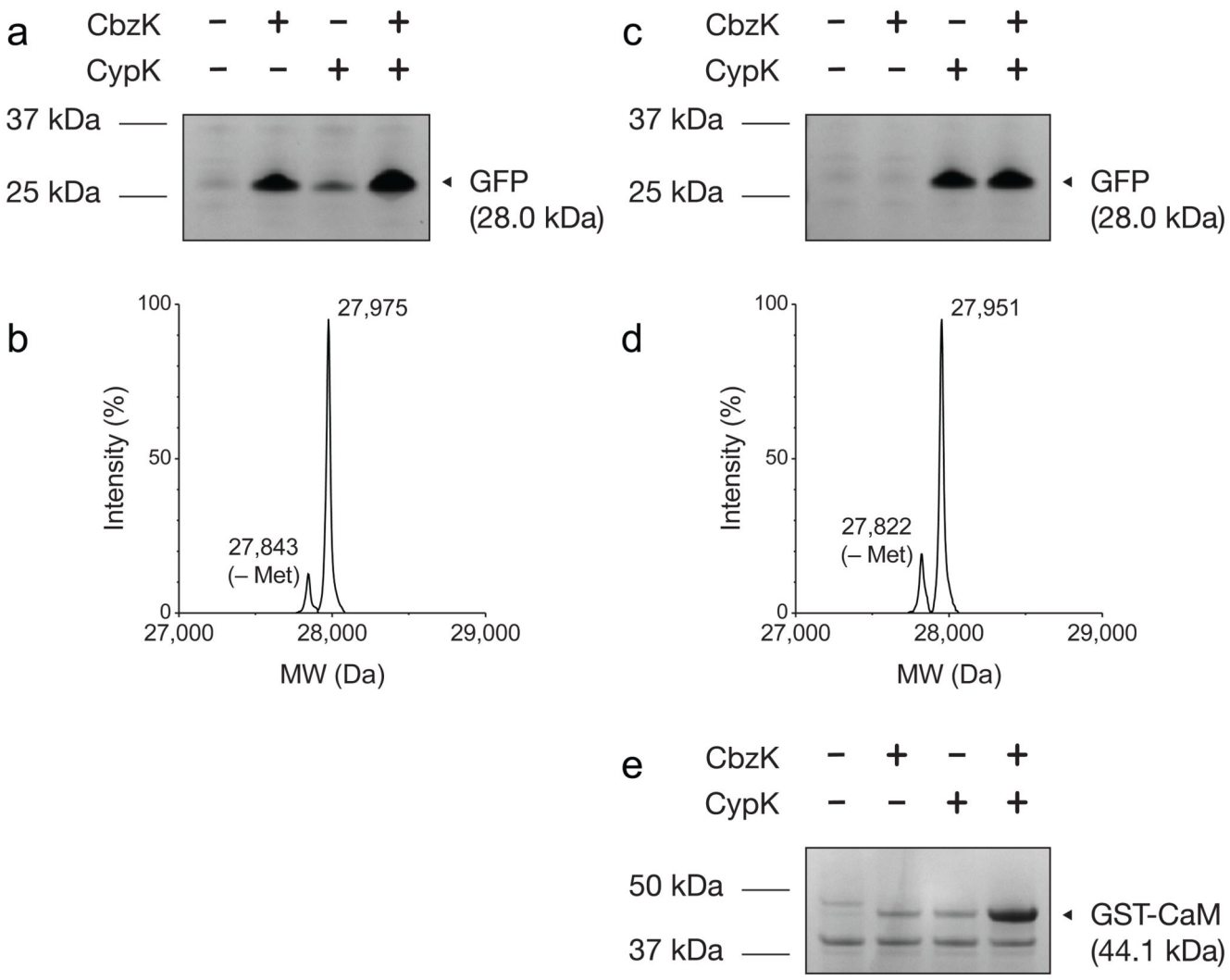

Figure 6. Encoding distinct ncAAs using mutually orthogonal PylRS/^{Pyl}tRNA pairs.

(a) GFPHis$_6$ purified from *E. coli* containing GFP(150TAG)His$_6$, *Mm*PylRS/*Mm*^{Pyl}tRNA$_{UCCU}$ and *Ma*PylRS-MutRS1/*Ma*^{Pyl}tRNA(6)$_{CUA}$ and analyzed by SDS-PAGE. Two independent experiments were performed with similar results. (b) Electrospray ionization mass spectrometry of GFPHis$_6$ purified from *E. coli* containing GFP(150TAG)His$_6$, *Mm*PylRS/*Mm*^{Pyl}tRNA$_{UCCU}$, *Ma*PylRS-MutRS1/*Ma*^{Pyl}tRNA(6)$_{CUA}$ in the presence of both CbzK and CypK. The peak corresponds to CbzK incorporation (predicted mass 27,976 Da, observed mass 27,975 Da). ESI-MS was performed once. (c) GFPHis$_6$ purified from *E. coli* containing GFP(150TAG)His$_6$, *Mm*PylRS/*Mm*^{Pyl}tRNA$_{CUA}$, *Ma*PylRS-MutRS1/*Ma*^{Pyl}tRNA(6)$_{UACU}$ and analyzed by SDS-PAGE. Two independent experiments were performed with similar results. (d) Electrospray ionization mass spectrometry of GFPHis$_6$ purified from *E. coli* containing GFP(150TAG)His$_6$, *Mm*PylRS/*Mm*^{Pyl}tRNA$_{CUA}$, *Ma*PylRS-MutRS1/*Ma*^{Pyl}tRNA(6)$_{UACU}$ in the presence of both CbzK and CypK. The peak corresponds to CypK incorporation (predicted mass 27,952 Da, observed mass 27,951 Da). ESI-MS was performed once. (e) GST-CaM purifications from *E. coli* containing ribo-Q1, o-GST-CaM(1TAG, 40AGGA), *Mm*PylRS/*Mm*^{Pyl}tRNA$_{UCCU}$,

*Ma*PylRS-MutRS1/*Ma*^{Pyl}tRNA(6)$_{CUA}$ grown in presence and absence of the indicated ncAAs. Samples were analyzed by SDS-PAGE. Two independent experiments were performed with similar results.

Europe PMC Funders Author Manuscripts