Plasmid construction

Previously described *gst-MalE* protein expression vectors p*gst-malE* and pO-*gst-malE*[9], are translated by wild type and orthogonal ribosomes respectively. These vectors were used as templates to construct variants containing one or two quadruplet codons in the linker region between the *gst* and *malE* open reading frame.

To create vectors containing a single AGGA quadruplet codon between *gst* and *malE* (p*gst*(*AGGA*)*malE and* pO-*gst*(*AGGA*)*malE*) the Tyr codon, TAC, in the linker between *gst* and *malE* was changed to AGGA by Quikchange mutagenesis (Stratagene), using the primers GMx1AGGAf and GMx1AGGAr (all primers used in this study are listed in **Supplementary Table 1**). For double AGGA mutants we additionally mutated the fourth codon in *malE* from GAA to AGGA by quick change PCR, with the primers GMx2AGGAf and GMx2AGGAr to create the vectors p*gst*(*AGGA*)$_2$*malE* and pO-*gst*(*AGGA*)$_2$*malE*. The vector pO-*gst-malE*(Y252AGGA) used for protein expression for mass spectrometry, in which the codon for Y17 of MBP was mutated to AGGA, was created by Quikchange mutagenesis (Stratagene) using the primers MBPY17AGGAf and MBPY17AGGAr.

To create vectors for constitutive production of the selected O-ribosomes the mutations in pRSF-OrDNA that confer the quadruplet decoding capacity on the orthogonal ribosome were transferred to pSC101 based O-rRNA expression vectors. pSC101*-ribo-X was used as a template and the mutations in 16S rDNA were introduced by enzymatic inverse PCR using the primers sc101Qr and sc101Q1f (for Ribo-Q1), sc101Q3f (forRibo-Q3) and sc101Q4f (for Ribo-Q4).

pDULE AzPheRS* tRNA$_{UCCU}$ (containing the gene for *Mj*tRNA$_{UCCU}$ and *Mj*AzPheRS*, each under the control of the lpp promoter) was created by changing the anticodon of the *Mj*tRNA$_{CUA}$ to UCCU by Quikchange and replacing the ORF of the *Mj*BPA-RS with *Mj*AzPheRS*-2 via ligation of the *Mj*AzPheRS*-2 gene, obtained by cutting pBK *Mj*AzPheRS*-2 with the restriction enzymes NdeI and StuI, into the same sites on pDULE *Mj*BPARS *Mj*tRNA$_{UCCU}$. pCDF PylST (a plasmid expressing *Mb*PylRS and *Mb*tRNA$_{CUA}$ from constitutive promoters) was created by cloning PCR products containing expression cassettes for *Mb*PylRS and *Mb*tRNA$_{CUA}$ into the BamHI and SalI or the SalI and NotI sites of pCDF DUET-1 (Novagen). The PCR products were obtained by amplifying the relevant regions of pBK PylRS and pREP PylT.

Plasmid encoding a fusion of GST and CaM were created by replacing the ORF of MBP in p-O-gst-malE with human CaM. The gene for CaM was amplified by PCR from pET3-CaM (a kind gift from K. Nagai) using primers CamEcof and CamH6Hindr (adding a C-terminal His_6-tag) and cloned into the EcoRI and HindIII sites of pO-*gst-malE*. Methionine-1 of CaM was mutated to AGGA by a subsequent round of Quikchange mutagenesis using primers CaM1aggaf and CaM1aggar (simultaneously removing part of the linker between GST and CaM). In a second round of mutagenesis an amber codon was introduced at position 149 using primers CaMK149TAGf and CaMK149TAGr. To create a sterically hindered control the amber codon was inserted at position 40 instead using primers CaM40tagf and CaM40tagr.

Construction of ribosome libraries and quadruplet decoding reporters.

11 different 16S rDNA libraries were constructed by enzymatic inverse PCR [8, 31] using pTrcRSF-O-ribo-X as a template. The resulting pRSF-O-rDNA libraries mutate between 7 and 13 nucleotides in defined regions on 16S rRNA and were constructed by multiple rounds of by enzymatic inverse PCR using the library construction primers in **Supplementary Table 1**. Each library has a diversity of greater than 10^9, ensuring more than 99% coverage. There is overlap in the nucleotides mutated in the 11 libraries and overall they cover the entire surface of decoding centre in the A site of the ribosome.

To create a reporter of quadruplet decoding by orthogonal ribosomes, we used a previously described *O-cat* (UAGA146)/tRNA(UAGA) vector as a template[9]. This vector contains a variant of *E. coli* $tRNA^{Ser2}$ on an *lpp* promoter and *rrnC* transcriptional terminator. The tRNA has an altered anticodon and selector codons for serine 146 in the chloramphenicol acetyl transferase (*cat*) gene downstream of an orthogonal ribosome-binding site. Ser146 is an essential and conserved catalytic serine residue that ensures the fidelity of incorporation. To create *O-cat* (AAGA 103 AAGA146)/tRNA(UCUU) the AAGA codon was introduced at position 146 and 103 and the anticodon of the tRNA was converted to UCUU by Quikchange mutagenesis using primers CAT146AGGAf, CAT146AGGAr and CAT103AGGAf, CAT103AGGAr. O-*cat* reporters containing the quadruplet codons AGGA, CCCU

(using primers CAT146CCCUf, CAT146CCCUr and CAT103CCCUf and CAT103CCUr) and the corresponding tRNAs (Ser2AGGAf, Ser2AGGAr, Ser2CCCUf and Ser2CCCUr) were also created by Quikchange mutagenesis. Reporters containing a single quadruplet selector codon were intermediates in the vector construction process. Vectors having the O-cat gene but lacking the tRNA were created using O-cat(UAGA146), which does not contain the tRNA cassette, as a template using Quik change primers CAT146AAGf, CAT146AGGAr, CAT103AGGAf, CAT103AGGAr, CAT146CCCUf, CAT146CCCUr, CAT103CCCUf and CAT103CCCUr that mutate the codons in O-*cat*.

Selection of orthogonal ribosomes with enhanced quadruplet decoding.

To select O-ribosomes with improved quadruplet decoding, each pRSF-O-rDNA library was transformed by electroporation into GeneHog *E. coli* (Invitrogen) cells containing O-*cat* (AAGA146). Transformed cells were recovered for 1 h in SOB medium containing 2% glucose and used to inoculate 200 ml of LB-GKT (LB medium with 2% glucose, 25 μg ml^{-1} kanamycin and 12.5 μg ml^{-1} tetracycline). After overnight growth (37 °C, 250 r.p.m., 16 h), 2 ml of the cells were pelleted by centrifugation (3,000*g*), and washed three times with an equal volume of LB-KT (LB medium with 12.5 μg ml^{-1} kanamycin and 6.25 μg ml^{-1} tetracycline). The resuspended pellet was used to inoculate 18 ml of LB-KT, and the resulting culture incubated (37 °C, 250 r.p.m. shaking, 90 min). To induce expression of plasmid encoded O-rRNA, 2 ml of the culture was added to 18 ml LB-IKT (LB medium with 1.1 mM isopropyl-D-thiogalactopyranoside (IPTG), 12.5 μg ml^{-1} kanamycin and 6.25 μg ml^{-1} tetracycline) and incubated for 4 h (37 °C, 250 r.p.m.). Aliquots (250 ml optical density at 600 nm (OD_{600}) = 1.5) were serial diluted and plated on LB-IKT agar (LB agar with 1 mM IPTG, 12.5 μg ml^{-1} kanamycin and 6.25 μg ml^{-1} tetracycline) supplemented with chloramphenicol of different concentrations (75 μg ml^{-1}, 100 μg ml^{-1}, 150 μg ml^{-1}, and 200 μg ml^{-1} respectively) and incubated (37 °C, 40 h).

Characterization of evolved orthogonal ribosomes with enhanced quadruplet decoding.

To separate selected pRSF-O-rDNA plasmids from the O-*cat* (AAGA146)/tRNA^{ser2}(UCUU) reporter plasmids, total plasmid DNA from selected

clones was purified and digested with *Not*I restriction endonuclease, and transformed into DH10B *E. coli*. Individual transformants were replica plated onto kanamycin agar and tetracycline agar and plasmid separation of pRSF-O-rDNA from the reporter confirmed by restriction digest and agarose gel analysis.

To quantify the quadruplet decoding activity of selected 16S rDNA clones, the selected pRSF-O-rDNA plasmids were cotransformed with O-*cat* (AGGA103, AGGA146) /tRNA^{ser2}(UCCU). Cells were recovered (SOB, 2% glucose, 1 h) and used to inoculate 10 ml of LB-GKT, which was incubated (16 h, 37 °C, 250 r.p.m.). We used 1 ml of the resulting culture to inoculate 9 ml of LB-KT, which was incubated (90 min, 37 °C, 250 r.p.m.). We used 1 ml of the LB-KT culture to inoculate 9 ml of LB-IKT medium, which was incubated (37 °C, 250 r.p.m., 4 h). Individual clones were transferred to a 96-well block and arrayed, using a 96-well pin tool, onto LB-IKT agar plates containing chloramphenicol at concentrations from 0 to 500 μg ml^{-1}. The plates were incubated (37°C, 16 h). We performed analogous experiments for other quadruplet codon-anticodon pairs.

To extract soluble cell lysates for *in vitro* CAT assays, 1 ml of each induced LB-IKT culture was pelleted by centrifugation at 3,000*g*. The cell pellets were washed three times with 500 μl Washing Buffer (40 mM Tris-HCl, 150 mM NaCl, 1 mM EDTA, pH 7.5) and once with 500 μl lysis buffer (250 mM Tris-HCl, pH 7.8). Cells were lysed in 200 μl Lysis Buffer by five cycles of flash-freezing in dry ice/ethanol, followed by rapid thawing in a 50 °C water bath. Cell debris was removed from the lysate by centrifugation (12,000*g*, 5 min) and the top 150 μl of supernatant frozen at -20 °C. To assay CAT activity in the lysates, 10 μl of soluble cell extract was mixed with 2.5 μl of FAST CAT Green (deoxy) substrate (Invitrogen) and preincubated (37 °C, 5 min). We added 2.5 μl of 9 mM acetyl-CoA (Sigma), and incubated (37 °C, 1 h). The reaction was stopped by the addition of ice-cold ethyl acetate (200 l, vortex 20 s). The aqueous and organic phases were separated by centrifugation (12,000*g*, 10 min) and the top 100 μl of the ethyl acetate layer collected. We spotted 1 μl of the collected solution onto a silica gel Thin-layer chromatography plate (Merck) for thin-layer chromatography in chloroform:methanol (85:15 vol/vol). The fluorescence of the spatially resolved substrate and product was visualized and quantified using a phosphorimager (Storm 860, Amersham Biosciences) with excitation and emission wavelengths of 450 nm and 520 nm, respectively.

Small scale expression and purification of *gst-malE* fusions.

E. coli containing the appropriate plasmid combinations were pelleted (3,000*g*, 10 min) from 50 ml overnight cultures, resuspended and lysed in 800 μl Novagen BugBuster Protein Extraction Reagent (supplemented with 1× protease inhibitor cocktail (Roche), 1 mM PMSF, 1 mg ml^{-1} lysozyme (Sigma), 1 mg ml^{-1} DNase I (Sigma)), and incubated (60 min, 25 °C, 1,000 r.p.m.). The lysate was clarified by centrifugation (6 min, 25,000*g*, 2 °C). GST containing proteins from the lysate were bound in batch (1 h, 4 °C) to 50 μl of glutathione sepharose beads (GE Healthcare). Beads were washed 3 times with 1 ml PBS, before elution by heating for 10 min at 80 °C in 60 μl 1× SDS gel-loading buffer. All samples were analyzed on 10% Bis-Tris gels (Invitrogen).

Measuring the translational fidelity of orthogonal quadruplet decoding ribosomes

^{35}S-cysteine misincorporation: *E. coli* containing either pO-*gst-malE* and pSC101*-O-ribosome, pO-*gst-malE* and pSC101*-ribo-X, pO-*gst-malE* and pSC101*-riboQ, or p*gst-malE* were resuspended in LB media (supplemented with ^{35}S-cysteine (1,000 Ci mmol^{-1}) to a final concentration of 3 nM, 750 μM methionine, 25 μg ml^{-1} ampicillin and 12.5 μg ml^{-1} kanamycin) to an OD_{600} of 0.1, and cells were incubated (3.5 h, 37°C, 250 r.p.m.). 10 ml of the resulting culture was pelleted (5,000*g*, 5 min), washed twice (1 ml PBS per wash), resuspended in 1 ml lysis buffer containing 1% Triton-X, incubated (30 min, 37°C, 1,000 r.p.m.) and lysed on ice by pipetting up and down. The clarified cell extract was bound to 100 μl of glutathione sepharose beads (1 h, 4°C) and the beads were pelleted (5,000*g*, 10 s) and washed twice in 1 ml PBS. The beads were added to 10 ml polypropylene column (Biorad) and washed (30 ml of PBS; 10 ml 0.5 M NaCl, 0.5x PBS; 30 ml PBS) before elution in 1 ml of PBS supplemented with 10 mM glutathione. Purified GST-MBP was digested with 12.5 units of thrombin for 1 h, to yield a GST fragment and an MalE fragment. The reaction was precipitated with 15% trichloroacetic acid and loaded onto an SDS-PAGE gel to resolve the GST, MBP and thrombin, and stained with InstantBlue (Expedeon). The ^{35}S activity in the GST and MBP protein bands were quantified by densitometry, using a Storm Phosphorimager (Molecular Dynamics) and ImageQuant

(GE Healthcare). The error frequency per codon for each ribosome examined was determined as follows: GST contains four cysteine codons, so the number of counts per second (c.p.s.) resulting from GST divided by four gives A, the cps per quantitative incorporation of cysteine. MBP contains no cysteine codons, but misincorporation at noncysteine codons gives B c.p.s. Because GST and MBP are present in equimolar amounts, (A/B 410, where 410 is the number of amino acids in the MBP containing thrombin cleavage fragment, gives the number of amino acids translated for one cysteine misincorporation C. Assuming the misincorporation frequency for all 20 amino acids is the same as that for cysteine the number of codons translated per misincorporation is C/20, and the error frequency per codon is given by $(C/20)^{-1}$.

Dual luciferase assays: The previously characterized pO-DLR contains a genetic fusion between a 5' *Renilla* luciferase (R-luc) and a 3' firefly luciferase (F-luc) on an orthogonal ribosome binding site [9]. pO-DLR, and its K529 codon variants, were transformed into *E. coli* cells with pSC101*-O-ribosome or pSC101*-ribo-Q1. Where indicated an additional *E. coli* Ser2A tRNA with a mutated anticodon, as specified in individual experiments, was supplied on plasmid p15A-tRNA-Ser2A. In this case 25 μg ml^{-1} tetracyclin was added to all culture media to maintain the additional plasmid. In experiments that used a suppressor tRNA recognizing AGGA codons a natural AGG codon, that is followed by a codon starting with an A, was removed from the linker region of pO-DLR by QuikChange using primers DLR952AAGxf and DLR953AGGxr.

Individual colonies were incubated (37°C, 250 r.p.m., 36 h) in 2 ml LB supplemented with ampicillin (50 μg ml^{-1}) and kanamycin (25 μg ml^{-1}), pelleted (5,000*g*, 5 min), washed with ice cold Millipore water and resuspended in 300 μl (1 mg ml^{-1} lysozyme, 1 mg ml^{-1} DNase I, 10 mM Tris (pH 8.0), 1 mM EDTA). Cells were incubated on ice for 20 min, frozen on dry ice, and thawed on ice. 10 μl samples of this extract were assayed for firefly (F-luc) and *Renilla* (R-luc) luciferase activity using the Dual-Luciferase Reporter Assay System (Promega). Each ribosome reporter combination was assayed from four independent cultures using an Orion microplate luminometer (Berthold Detection Systems) and the data analyzed as previously described. The error reported is the standard deviation.

Mass spectrometric characterization of p-azido-L-phenylalanine (2) incorporation by Ribo-Q1

E. coli DH10B containing p-O-*gst-malE*(Y252AGGA), pSC101*Ribo-Q1 and pDULE-AzPheRS*tRNA$_{UCCU}$ were used to produce protein for mass spectrometry. Protein was expressed in the presence of 2.5 mM **2** and purified on glutathione. The purified proteins were resolved by SDS-PAGE, stained with Instant Blue (Expedeon) and the band containing full length GST-MBP was excised for analysis by LC/MS/MS (NextGen Sciences). The samples were reduced with DTT at 60°C and alkylated with iodoacetamide after cooling to room temperature. The samples were then digested with trypsin (37°C, 4 h), and the reaction was stopped by the addition of Formic acid. The samples were analyzed by nano LC/MS/MS on a ThermoFisher LTQ Orbitrap XL. 30 μl of hydrolysate was loaded onto a 5 mm 75 μm ID C12 (Jupiter Proteo, Phenomenex) vented column at a flow-rate of 10 μl min^{-1}. Gradient elution was over a 15 cm 75 μm ID C12 column at 300 nl min^{-1} with a 1 hour gradient. The mass spectrometer was operated in data-dependent mode, and ions were selected for MS/MS. The Orbitrap MS scan was performed at 60,000 FWHM resolution. MS/MS data was searched using Mascot (www.matrixscience.com).

Evolution of a quadruplet decoding *Mj*AzPheRS

pBK *Mj*AzPheRS-7 [24] (a kanamycin resistant plasmid, which contains *Mj*AzPheRS-7 on a GlnRS promoter and terminator) was used as a template to create a library in the region of *Mj*AzPheRS that recognizes the anticodon. Codons for residues Y230, C231, P232, F261, H283 and D286 were randomized to NNK in two rounds of enzymatic inverse PCR, generating a library of 10^8 mutant clones. pREP JY(UCCU) was created by changing the anticodon of *Mj*tRNA$_{CUA}$ in pREP YC-JYCUA [32] from CUCUAAA to CUUCCUAA by QuikChange mutagenesis (Stratagene) and changing the amber codon in the chloramphenicol acetyltransferase gene to AGGA. *E. coli* DH10B harbouring this plasmid were transformed with the mutant library and grown in LB-KT (LB medium supplemented with 25 μg ml^{-1} kanamycin and 12.5 μg ml^{-1} tetracycline) supplemented with 1 mM **2**. 10^9 cells were plated on LB-KT plates containing 1 mM **2** and concentrations of chloramphenicol ranging from 50 to 250 μg ml^{-1}. After incubation (36 h, 37ºC) individual clones were tested for **2** dependent

growth on LB-KT plates with 0-250 μg ml^{-1} chloramphenicol with and without 1 mM **2**. The plasmid DNA from clones showing amino acid dependent growth was isolated and digested with HindIII to eliminate pREP JY(UCCU). After transformation and reisolation of the kanamycin resistant plasmid the DNA was sequenced.

To select quadruplet decoding pairs that incorporate other amino acids, the procedure above was repeated using the relevant starting template and unnatural amino acid.

Investigating the mutual orthogonality of *Mb*PylRS/*Mb*tRNA$_{CUA}$ and *Mj*TyrRS/*Mj*tRNA$_{CUA}$

To test the ability of *Mb*PylRS to aminoacylate *Mj*tRNA$_{CUA}$ *E. coli* DH10B were transformed with a pBK *Mb*PylRS encoding *Mb*PylRS under the control of a GlnRS promoter and terminator and pMyo4TAG-His$_6$, expressing sperm whale myoglobin with an amber codon at position 4 and *Mj*tRNA$_{CUA}$. The cells were grown overnight at 37ºC in LB-KT. Fresh LB-KT (50 ml) supplemented with 10 mM N6-[(tert.-butyloxy)carbonyl]-L-lysine (BocLys, **3**) was inoculated 1:50 with overnight culture. After 3 h at 37°C protein expression was induced by addition of 0.2% arabinose. After a further 3 h cells were harvested and washed with PBS. Proteins were extracted by shaking at 25ºC in 1 ml Ni-wash buffer (10 mM Tris/Cl, 20 mM imidiazole, 200 mM NaCl pH 8.0) supplemented with protease inhibitor cocktail (Roche), 1 mM PMSF, and approx. 1 mg ml^{-1} lysozyme and 0.1 mg ml^{-1} DNAse I. The extract was clarified by centrifugation (5 min, 25000 g, 4ºC), supplemented 50 μl Ni^{2+}-NTA beads and incubated with agitation for 1 h at 4ºC. Beads were washed in batch three times with 1 ml Ni-wash buffer and eluted in 100 μl sample buffer supplemented with 200 mM imidazole. To test the aminoacylation activity between the cognate pairs or between *Mj*TyrRS and *Mb*tRNA$_{CUA}$ analogous experiments were carried out as above using the relevant plasmids (pBK *Mj*TyrRS or pBK *Mb*PylRS and pMyo4TAG-His$_6$ or pMyo4TAG-His$_6$-PylT) and unnatural amino acids (**3** or none). Proteins were analysed by 4-12% SDS-PAGE and stained with Instant Blue.

Characterization of the quadruplet suppressing AzPheRS*

Expression and purification of myoglobin from pMyo4TAG-His$_6$ or pMyo4AGGA-His$_6$ was carried out as above using the relevant pBK plasmids and 2.5 mM **2**. Proteins were analysed by 4-12% SDS-PAGE.

Characterization of Myo4AzPhe produced with AzPheRS* from pMyo4AGGA-His$_6$ by ESI mass spectrometry

Myoglobin was expressed in *E. coli* DH10B using plasmids pBK AzPheRS* and pMyo4AGGA-His$_6$ essentially as described above but at 1 l scale. The protein was extracted by shaking at 25ºC in 30 ml Ni-wash buffer supplemented with protease inhibitor cocktail (Roche), 1 mM PMSF, 1 mg ml^{-1} lysozyme and 0.1 mg ml^{-1} DNAse I. The extract was clarified by centrifugation (15 min, 38000 g, 4ºC), supplemented 0.3 ml Ni^{2+}-NTA beads and incubated with agitation for 1 h at 4ºC. Beads were poured into a column and washed with 40 ml of Ni-wash buffer. Bound protein was eluted in 0.5 ml fractions of the same buffer containing 200 mM imidazole and immediately rebuffered to 10 mM ammonium carbonate pH 7.5 by dialysis. 50 μl of the sample was mixed 1:1 with 1% formic acid in 50% methanol and total mass determined on an LCT time-of-flight mass spectrometer with electrospray ionization (Micromass). The sample was injected at 10 μl min^{-1} and calibration performed in positive ion mode using horse heart myoglobin. 50 scans were averaged and molecular masses obtained by deconvoluting multiply charged protein mass spectra using MassLynx version 4.1 (Micromass). The theoretical mass of the wild-type myoglobin was calculated using Protparam (http://us.expasy.org/tools/protparam.html), and the theoretical mass for **2** adjusted manually.

MS/MS analysis of GST-MBP 234AzPhe 239CAK

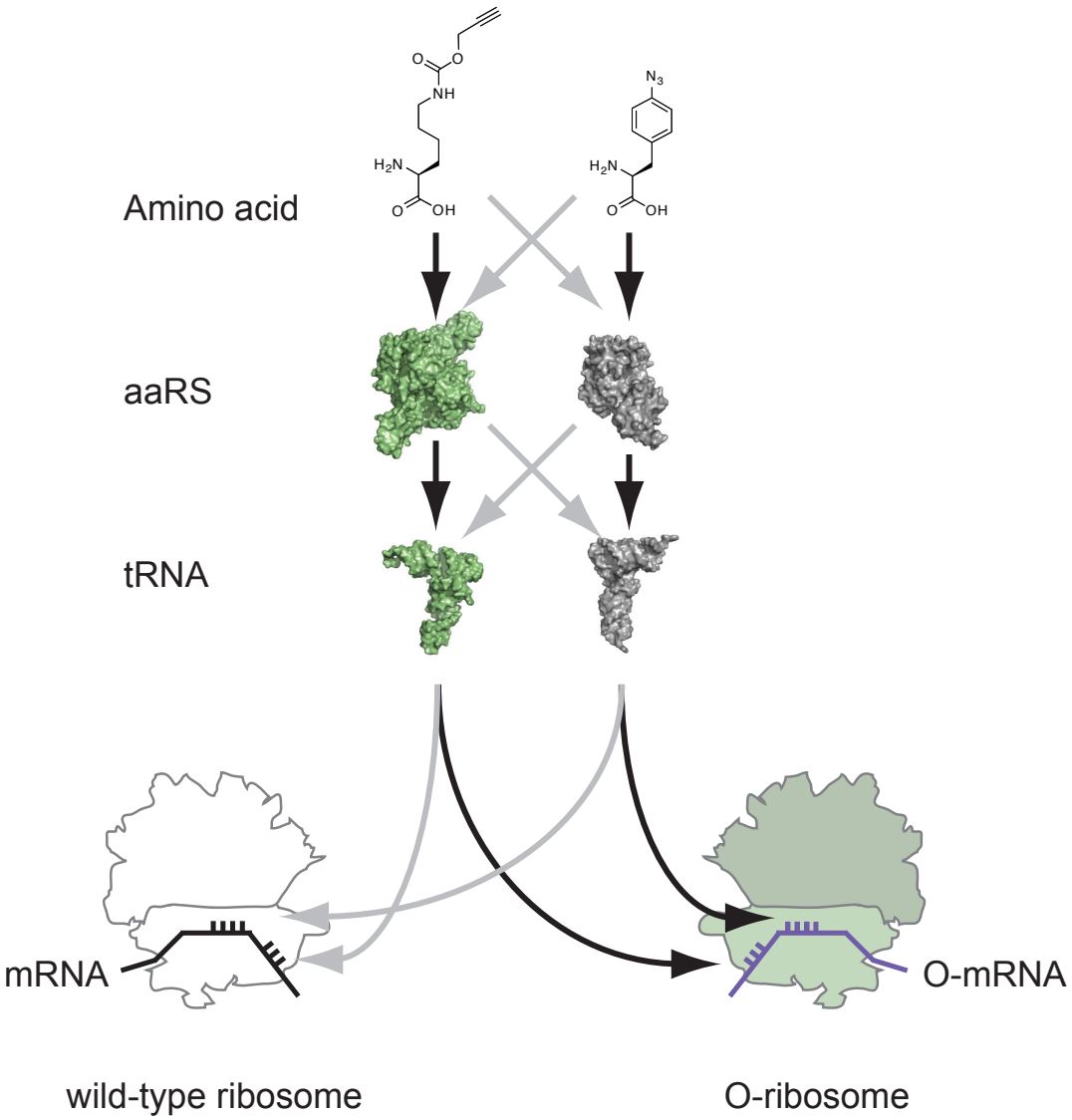
E. coli DH10B were transformed with pDULE AzPheRS*/tRNA$_{UCCU}$ and pCDF PylST and grown to logarithmic phase in LB-ST (25 μg ml^{-1} spectinomycin and 12.5 μg ml^{-1} tetracycline). Electrocompetent cells were prepared and transformed with a plasmid for the constitutive expression of an orthogonal ribosome (pSC101* Ribo-Q) and p-O-*gst*(234AGGA 239TAG)*malE*. The recovery of the transformation was used to inoculate LB-AKST (LB medium containing 50 μg ml^{-1} ampicillin, 12.5 μg ml^{-1} kanamycin, 25 μg ml^{-1} spectinomycin and 12.5 μg ml^{-1} tetracycline). The culture was grown to saturation at 37ºC and used to inoculate the main culture 1:50. Cells were grown overnight at 37ºC, harvested by centrifugation and stored at -20ºC. The GST-MBP protein was expressed at a scale of 100 ml using 2.5 mM of each AzPhe (**2**) and

CAK (**4**). Proteins were extracted and purified as above. After washing the beads with PBS the protein was eluted by heating in 100 μl 1x sample buffer containing 50 mM β-mercaptoethanol to 80°C for 5 min. The protein sample was analysed by 4-12% SDS-PAGE and stained with Instant Blue. The band containing full-length GST-MBP was excised and submitted for LC/MS/MS analysis (by NextGen Sciences).

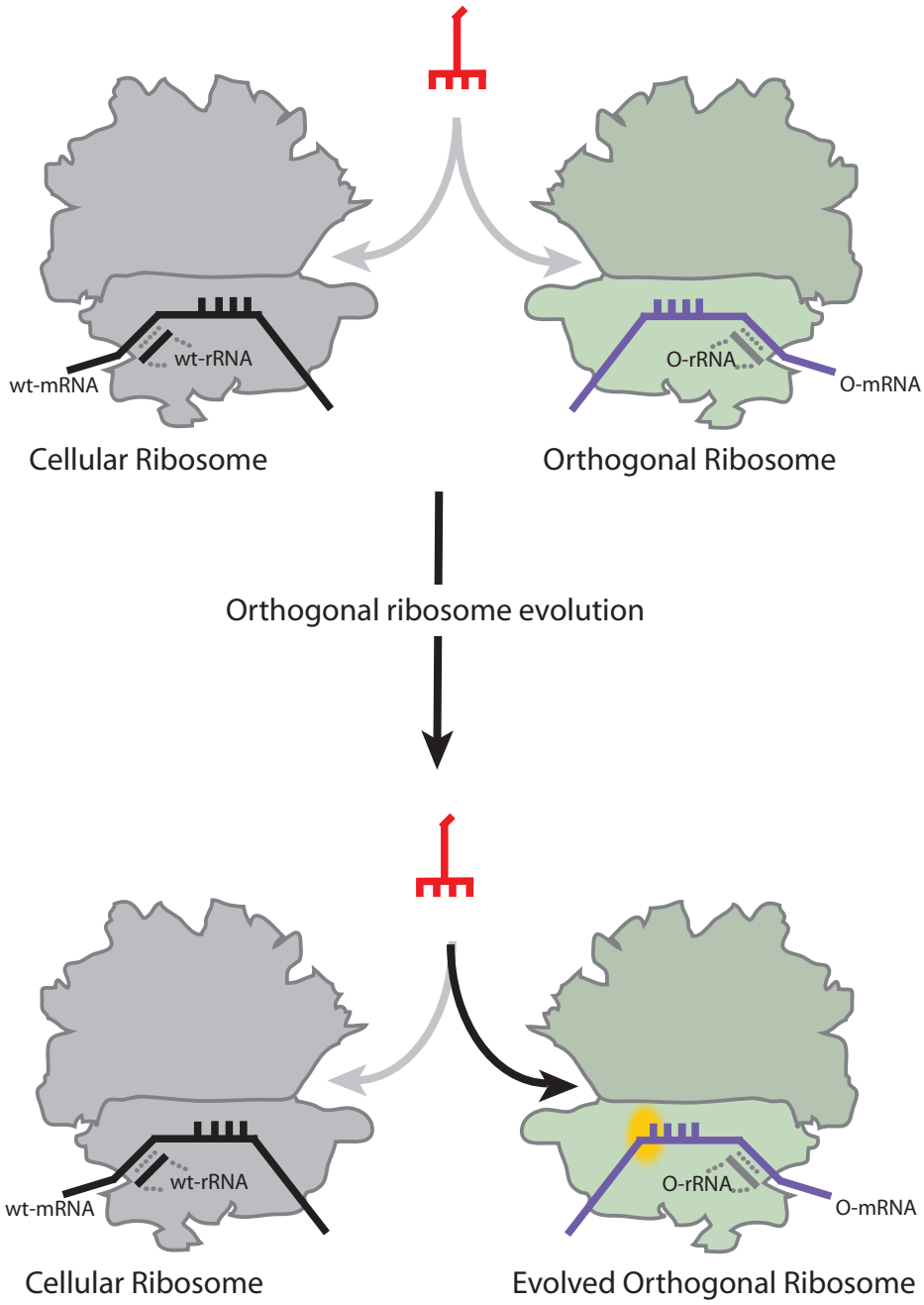
Cyclization of GST-CaM-His$_6$ 1AzPhe 149CAK

E. coli DH10B were transformed sequentially with four plasmids as described above using expression plasmids p-O-*gst-CaM*-His$_6$ 1AGGA 149UAG or p-O-*gst-CaM*-His$_6$ 1AGGA 40UAG. The protein was expressed at 0.5 L scale as described above using 5 mM **2** and 2.5 mM **4**. The cells were extracted and *GST-CaM-His$_6$* purified as described for myoglobin-His$_6$ and dialysed against 50 mM Na_2HPO_4 pH 8.3. To perform the cyclization reaction, 160 μl of protein sample was mixed with 40 μl of a fresh solution of 5 mM ascorbic acid, 5 mM $CuSO_4$ and 10 mM bathophenanthroline. The reaction was incubated at 4ºC and analysed by 4-12% SDS-PAGE.

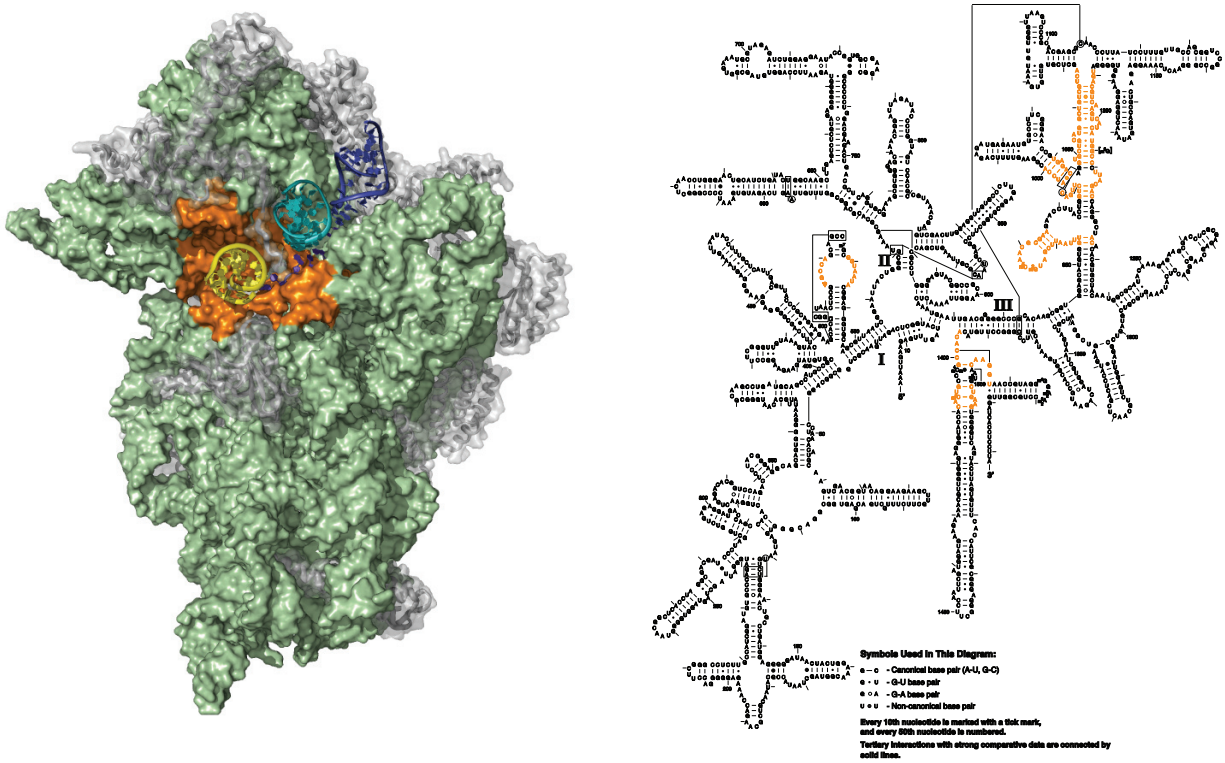
To analyze the cyclization product by mass spectrometry we introduced additional tryptic cleavage sites around the incorporation sites of unnatural amino acids to facilitate subsequent analysis. Therefore, the point mutations Q4K and M146K (numbering relative to the AGGA codon in p-O-*gst-CaM*-His$_6$ 1AGGA 149UAG) and a G_3K linker directly following the TAG codon were introduced by QuikChange. The protein was expressed, purified and cyclized as above with very similar yields. The cyclized protein was subsequently excised from an SDS-PAGE gel and submitted for mass spectrometric analysis (NextGen Sciences, Ann Arbor, USA).



Supplementary Figure 1. Strategy for the synthesis of an orthogonal genetic code. Combining the two mutually orthogonal pairs (*Mb*PylRS/*Mb*tRNA$_{CUA}$ and *Mj*AzPheRS*/tRNA$_{UCCU}$) with evolved orthogonal ribosomes (Ribo-Q) creates a system that is able to decode the UAG and AGGA codons on an orthogonal mRNA (O-mRNA) to produce a protein that contains two distinct unnatural amino acids at genetically encoded sites. UAG is decoded as **4** (CAK) or **3** (BocLys) by *Mb*PylRS/*Mb*tRNA$_{CUA}$ while AGGA is decoded as **2**.

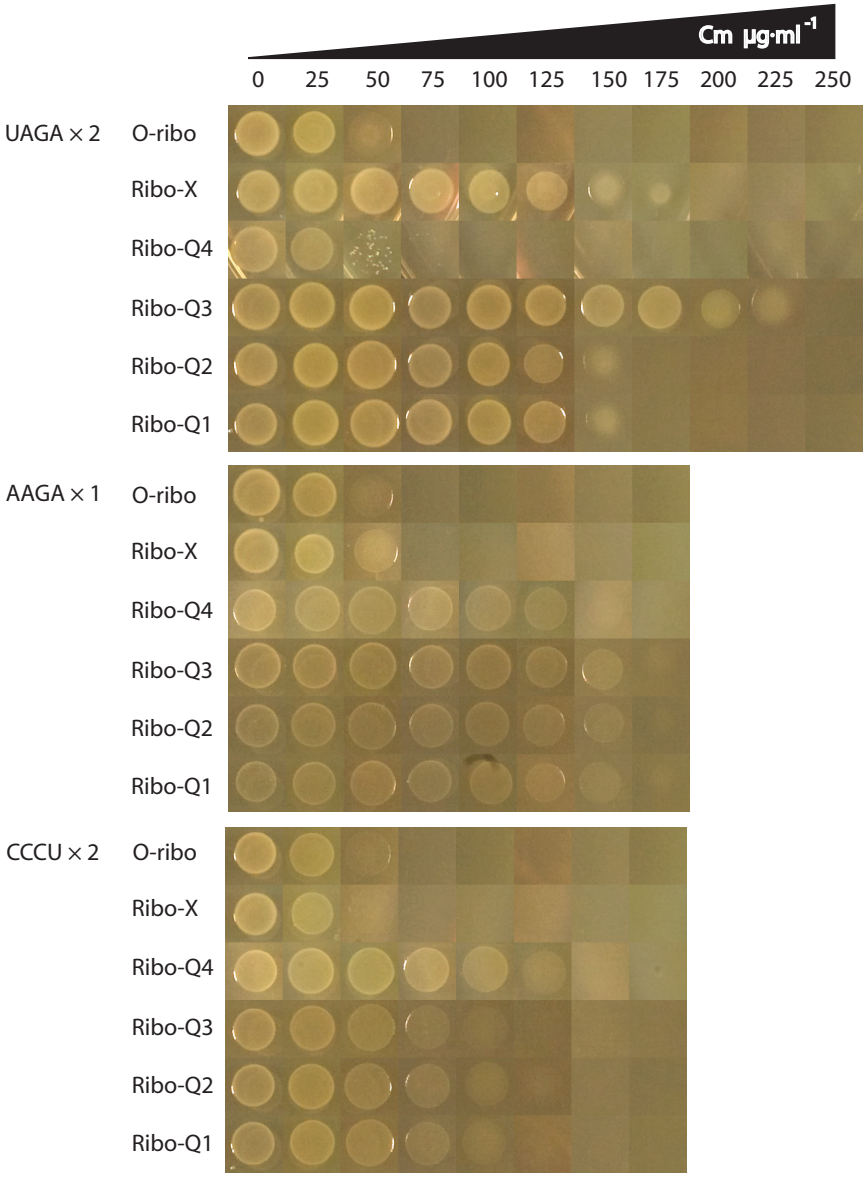


Supplementary Figure 2. Evolving an orthogonal quadruplet decoding ribosome. The natural ribosome (gray) and the progenitor orthogonal ribosome (green) utilize tRNAs with triplet anticodon to decode triplet codons in both wt- (black) and orthogonal- (purple) mRNAs, respectively. The decoding of quadruplet codons with extended anticodon tRNAs (red) is of low efficiency (light gray arrows) on both ribosomes. Synthetic evolution of the orthogonal ribosome leads to an evolved scenario in which a mutant (orange patch) orthogonal ribosome more efficiently decodes quadruplet codons on orthogonal mRNAs using extended anticodon tRNAs. Decoding of extended anticodon tRNAs on natural mRNAs is unaffected because the orthogonal ribosome does not read natural mRNAs and the natural ribosome is unaltered.



Supplementary Figure 3. Comprehensive mutagenesis of the ribosome decoding centre.

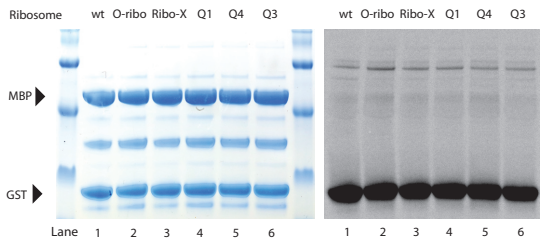
A. Structure of the ribosomal small subunit with bound tRNAs and mRNAs. tRNA anticodon stem loops are bound to A site (yellow), P site (cyan), and E site (dark blue). The mRNA is shown in purple. 16S ribosomal RNA is shown in green and ribosomal proteins in gray. The 118 residues in the decoding centre, targeted for mutation in the 11 libraries, are shown in orange (This figure was created using Pymol v0.99 (www.pymol.org) and PDB ID 2J00). **B.** Secondary structure of the *E. coli* 16S ribosomal RNA (www.rna.ccbb.utexas.edu). The nucleotides targeted for mutation are shown colored orange.



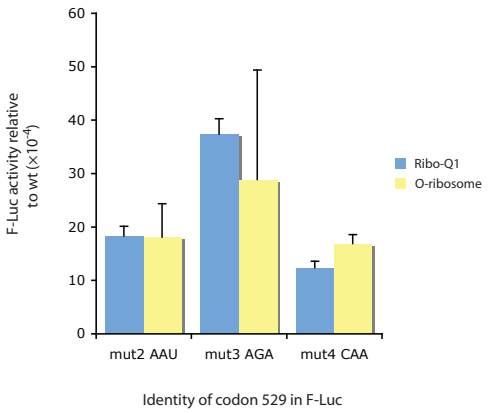
Supplementary Figure 4. Ribo-Q enhances the tRNA dependent decoding of different quadruplet codons. Ribo-X, Ribo-Q1-4 and the O-ribosome were produced from pRSF-O-rDNA vectors. The tRNAser2UCUA-dependent enhancement in decoding UAGA codons in the O-*cat* (UAGA103, UAGA146), the tRNAser2AGGG-dependent enhancement in decoding CCCU codons in the O-*cat* (CCCU103, CCCU146), and the tRNAser2UCUU-dependent enhancement in decoding AAGA codons in the O-*cat* (AAGA146) was measured by survival on increasing concentrations of chloramphenicol. pRSF-O-rDNA vectors and corresponding O-*cat* vectors were co-transformed into GeneHogs cells. Transformed cells were recovered for 1 h in SOB medium containing 2% glucose and used to inoculate 200 ml of LB-GKT (LB medium with 2% glucose, 25 μg ml^{-1} kanamycin and 12.5 μg ml^{-1} tetracycline). After overnight growth (37°C, 250 r.p.m., 16 h), 2 ml of the cells were

pelleted by centrifugation (3,000*g*), and washed three times with an equal volume of LB-KT (LB medium with 12.5 μg ml^{-1} kanamycin and 6.25 μg ml^{-1} tetracycline). The resuspended pellet was used to inoculate 18 ml of LB-KT, and the resulting culture incubated (37°C, 250 r.p.m. shaking, 90 min). To induce expression of plasmid encoded O-rRNA, 2 ml of the culture was added to 18 ml LB-IKT (LB medium with 1.1 mM isopropyl-D-thiogalactopyranoside (IPTG), 12.5 μg ml^{-1} kanamycin and 6.25 μg ml^{-1} tetracycline) and incubated for 4 h (37°C, 250 r.p.m.). Aliquots (250 μl optical density at 600 nm (OD600) = 1.5) were plated on LB-IKT agar (LB agar with 1 mM IPTG, 12.5 μg ml^{-1} kanamycin and 6.25 μg ml^{-1} tetracycline) supplemented with 50 μg ml^{-1} chloramphenicol and incubated (37°C, 40 h).

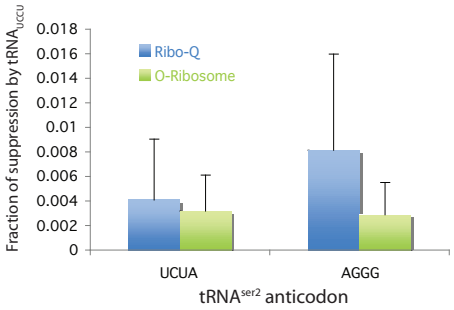
A



B



C

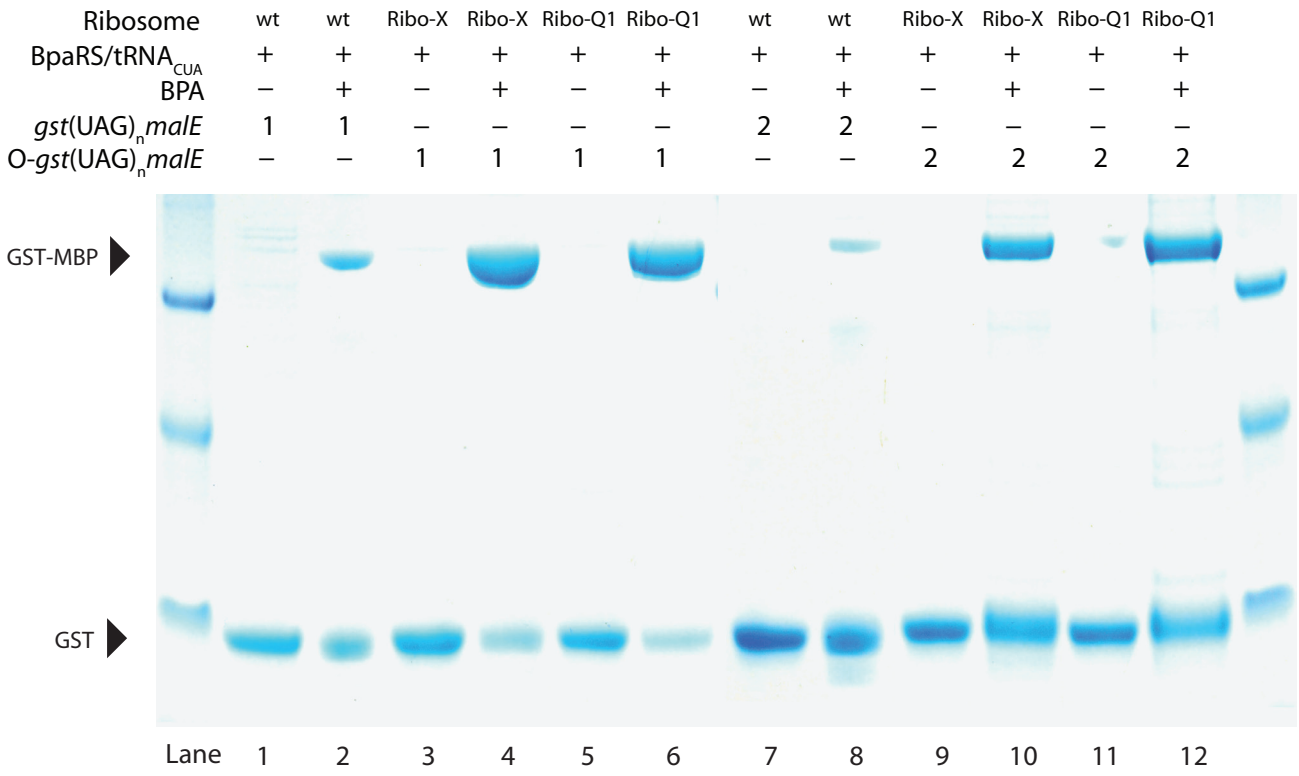


D

Ribosome	O-CAT codon	tRNA anticodon	Fractional Activity
O-ribosome	AGGA	UCCU	1.00
O-ribosome	AGGT	UCCU	0.08
O-ribosome	AGGG	UCCU	0.04
O-ribosome	AGGC	UCCU	0.08
Ribo-X	AGGA	UCCU	1.00
Ribo-X	AGGT	UCCU	0.07
Ribo-X	AGGG	UCCU	0.03
Ribo-X	AGGC	UCCU	0.07
Ribo-Q4	AGGA	UCCU	1.00
Ribo-Q4	AGGT	UCCU	0.07
Ribo-Q4	AGGG	UCCU	0.04
Ribo-Q4	AGGC	UCCU	0.05

Supplementary Figure 5: The translation fidelity of evolved ribosomes is comparable to that of the natural ribosome. **A.** The translational error frequency for triplet decoding as measured by ^{35}S-cysteine misincorporation is indistinguishable for ribo-Q1, ribo-Q3-Q4, ribo-X, the unevolved orthogonal ribosome and the wild-type ribosome. GST-MBP was synthesized by each ribosome in the presence of ^{35}S-cysteine, purified on glutathione sepharose and digested with thrombin. The left panel shows a Coomassie stain of the thrombin digest. The un-annotated bands result primarily from the thrombin preparation. The right panel shows ^{35}S labeling of proteins in the same gel, imaged using a Storm Phosphorimager. Lanes 1–6 show thrombin cleavage reactions of purified protein derived from cells containing the indicated ribosome (with the ribosomal RNA produced from pSC101* constructs that drive rRNA from a P1P2 promoter) and either pO-*gst-malE* (for orthogonal

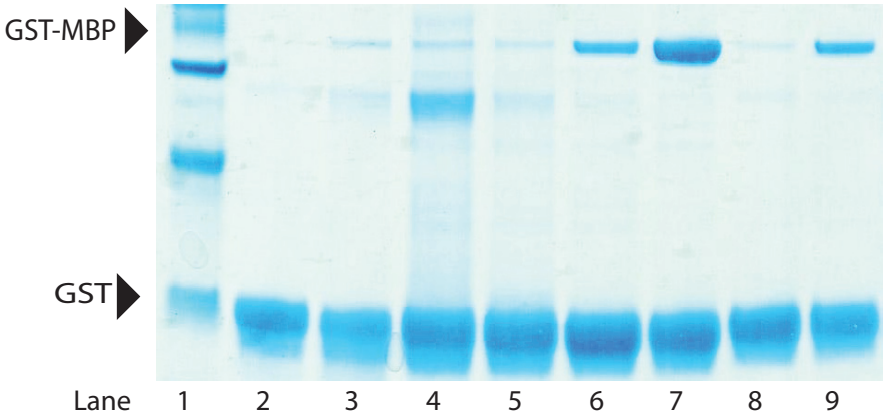
ribosomes) or p*gst-malE* (for wild-type ribosomes). The size markers are pre-stained standards (Bio-Rad 161-0305). The error frequency per codon translated by the ribo-Q ribosomes as measured by this method was less than $1x10^{-3}$. Control experiments with the progenitor orthogonal ribosome, ribo-X and the wild-type ribosome allowed us to put the same limit on their fidelity. This limit compares favourably with previous measurements of error frequency using ^{35}S mis-incorporation ($4x10^{-3}$ errors per codon) [33]**B.** The translational fidelity of ribo-Q1 in triplet decoding is comparable to that of the un-evolved ribosome, as measured by a dual-luciferase assay. In this system a C-terminal firefly luciferase is mutated at codon K529(AAA), which codes for an essential lysine residue. The extent to which the mutant codon is misread by $tRNA^{Lys}$(UUU) is determined by comparing the firefly luciferase activity resulting from the expression of the mutant gene to the wild-type firefly luciferase, and normalizing any variability in expression using the activity of the co-translated N-terminal *Renilla* luciferase. Previous work has demonstrated that measured firefly luciferase activities in this system result primarily from the synthesis of a small amount of protein that mis-incorporates lysine in response to the mutant codon [23], rather than a low activity resulting from the more abundant protein containing encoded mutations. In experiments examining the fidelity of ribo-Q1, lysate from cells containing pSC101*-ribo-Q1 and pO-DLR and its codon 529 variants were assayed. Control experiments used lysates from cells containing pSC101*-O-ribosome and pO-DLR and its codon 529 variants. **C.** The quadruplet decoding fidelity of ribo-Q is comparable to that of un-evolved ribosomes. Efficiencies were determined using a dual luciferase construct with an N-terminal Renilla and C-terminal Firefly luciferase (Ren-FF). The reporter was mutated to include a quadruplet AGGA codon in the linker between the two luciferases (Ren-AGGA-FF). Ren-AGGA-FF was transformed into DH10B cells along with a non-cognate anticodon Ser2A tRNA (UCUA or AGGG) and either ribo-Q or the O-ribosome. Readthrough efficiency for Ren-AGGA-FF was measured by taking the ratio of Firely luminescence/Renilla luminescence. This data was divided by the same Firefly/Renilla ratio when using the Ren-FF construct in the presence of tRNA (to normalize for effects of the tRNA on sites outside the AGGA codon under investigation). In order to obtain the level of decoding by these non-cognate tRNAs as a fraction of decoding by cognate tRNA, these data were compared with that obtained from the same experiment using a cognate Ser2A tRNA with the UCCU anti-codon. The data represent the average of at least 4 trials. The error bars represent the standard deviation. **D** Fourth base specificity in quadruplet decoding. *E. coli* DH10B expressing the indicated combination of an O-ribosome, a chloramphenicol acetyltransferase gene under the control of an orthogonal rbs with a quadruplet codon at a permissive site and *E. coli* Ser2A $tRNA_{UCCU}$ were scored for their ability to grow in the presence of increasing amounts of chloramphenicol. The fractional activity is the maximal Cm resistance of the cells relative to the combination containing a cognate codon in the mRNA and a particular o-ribosome.



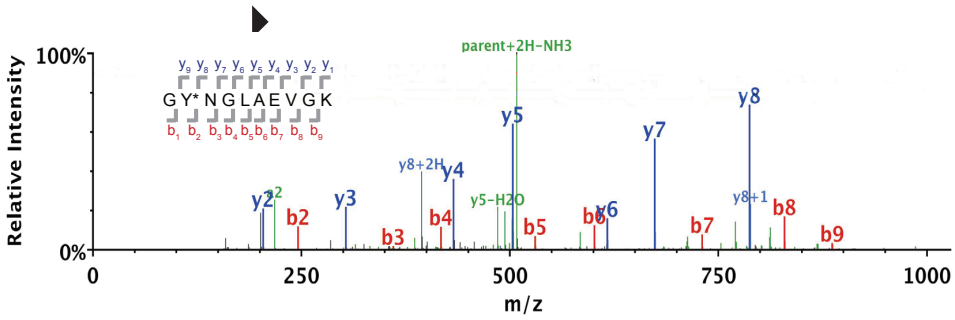
Supplementary Figure 6: Ribo-Q1 enhances the efficiency of BpaRS/tRNA$_{CUA}$–dependent unnatural amino acid incorporation in response to single and double UAG codons, maintaining the enhanced amber decoding of ribo-X. In each lane an equal volume of protein purified from glutathione sepharose under identical conditions is loaded. Orthogonal ribosomes are produced from pSC101*-ribo-X, pSC101*-ribo-Q1. Bpa, *p*-benzoyl-L-phenylalanine (**1**). The BpaRS/tRNA$_{CUA}$ pair is produced from pSUPBpa that contains six copies of *Mj*tRNA$_{CUA}$. (UAG)$_n$ describes the number of amber stop codons (n) between *gst* and *malE* in O-*gst*(UAG)$_n$*malE* or *gst*(UAG)$_n$*malE*. The ratio of GST-MBP to GST reflects the efficiency of amber suppression versus RF1 mediated termination. A part of this gel showing the band for full-length GST-MBP is shown in **Figure 2** of the main text.

A

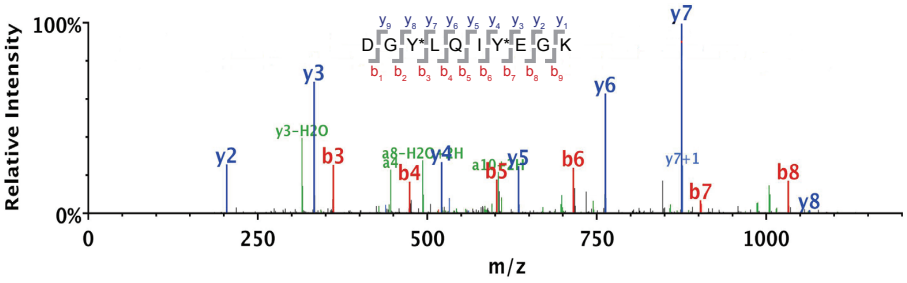
wt-ribosome	+	+	+	+	+	+	+	+
Ribo-Q1	-	-	-	-	+	+	+	+
AzPheRS/tRNA$_{UCCU}$	+	+	+	+	+	+	+	+
AzPhe	-	+	-	+	-	+	-	+
gst(AGGA)$_n$*malE*	1	1	2	2	-	-	-	-
O-*gst*(AGGA)$_n$*malE*	-	-	-	-	1	1	2	2



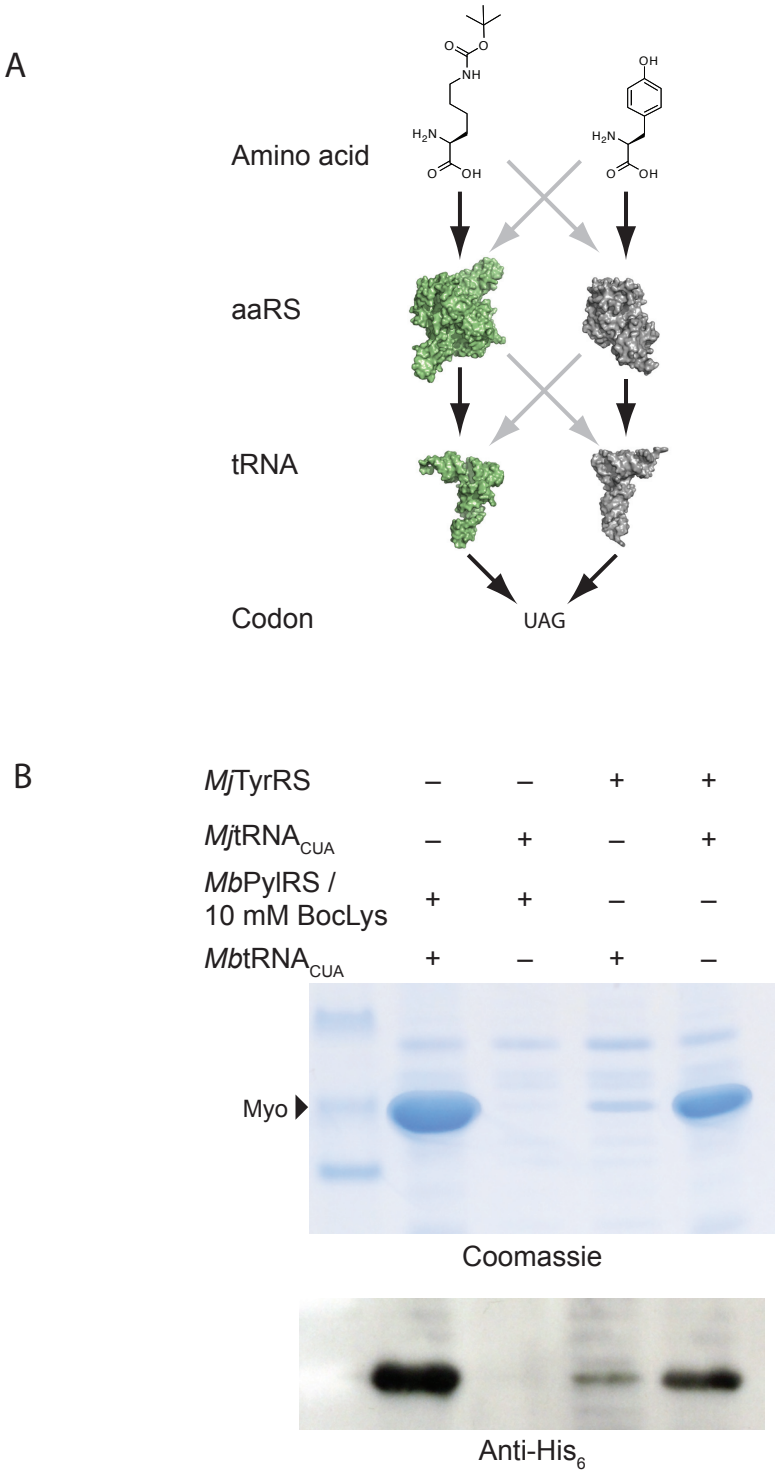
B



C

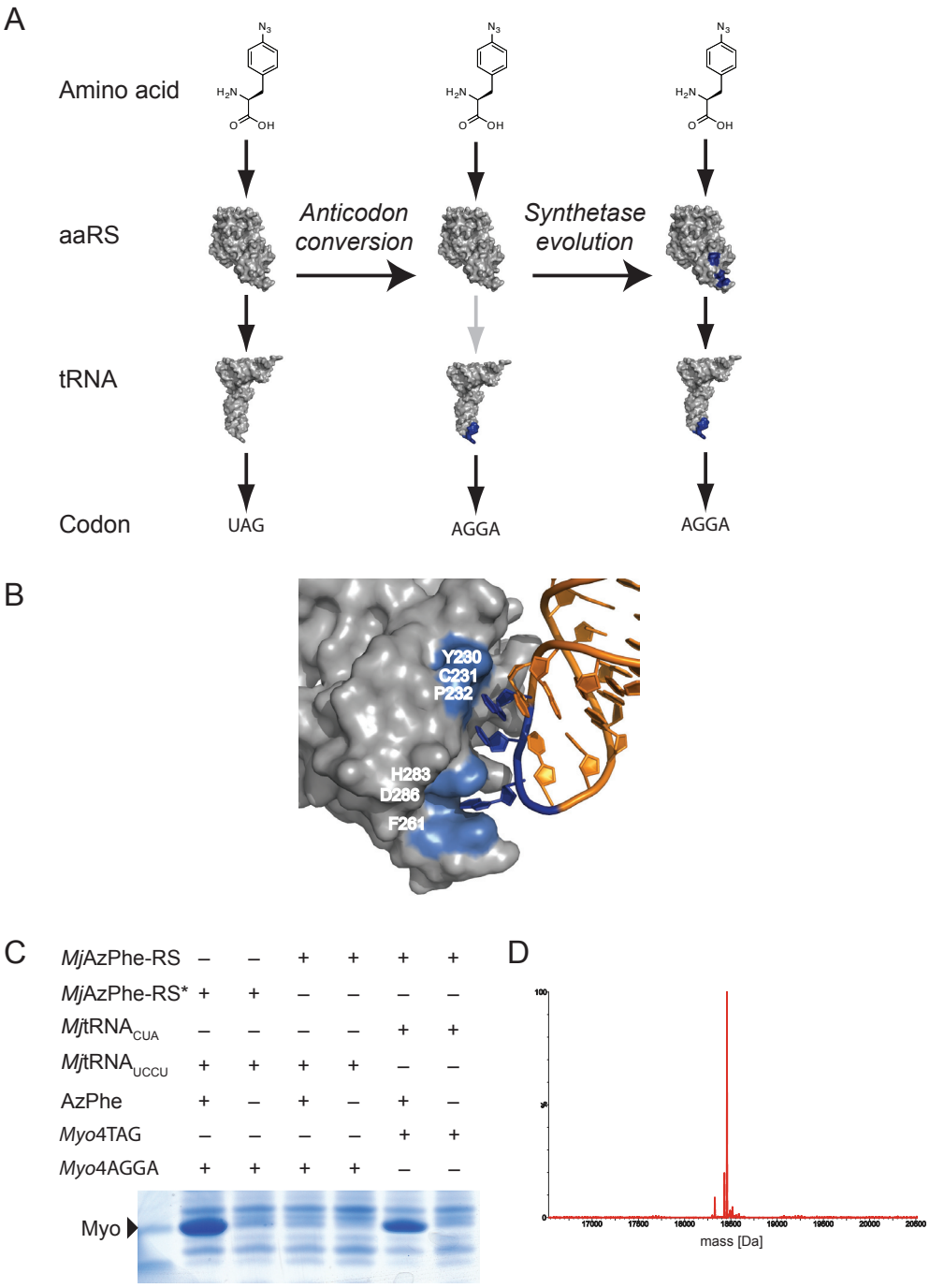


Supplementary Figure 7: Ribo-Q1 enhances the efficiency of AzPheRS*/tRNA$_{UCCU}$ unnatural amino acid incorporation in response to AGGA quadruplet codons. **A**. Ribo-Q1 is produced from pSC101*-ribo-Q1. AzPhe, 2.5 mM **2**. The AzPheRS*/tRNA$_{UCCU}$ pair is produced from pDULE AzPheRS*/tRNA$_{UCCU}$ that contains a single copy of *Mj*tRNA$_{UCCU}$. (AGGA)$_n$ describes the number of quadruplet codons (n) between *gst* and *malE* in O-*gst*(AGGA)$_n$*malE* or *gst*(AGGA)$_n$*malE*. The ratio of GST-MBP to GST reflects the efficiency of frameshift suppression. A part of this gel showing the bands for full-length GST-MBP is shown in **Figure 2** of the main text. B & C. MS/ MS spectra of tryptic fragments incorporating one or two AzPhes respectively.



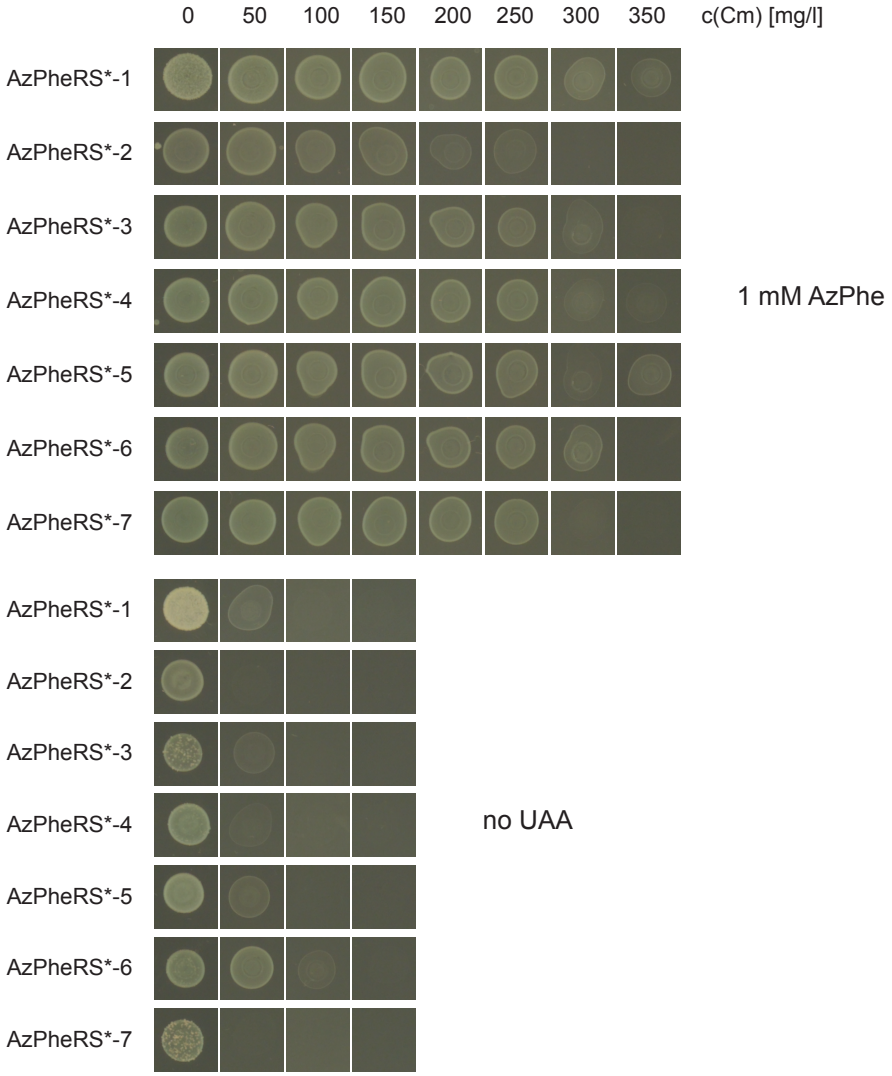
Supplementary Figure 8. *Mb*PylRS/*Mb*tRNA$_{CUA}$ and *Mj*TyrRS/tRNA$_{CUA}$ pairs are mutually orthogonal in their aminoacylation specificity. **A.** The decoding network of *Mb*PylRS/*Mb*tRNA$_{CUA}$ (lime) and *Mj*TyrRS/tRNA$_{CUA}$ (grey) and its unnatural amino acid incorporating derivatives. A unique unnatural amino acid is specifically recognized by each of the synthetases and used to aminoacylate its cognate tRNA. We asked whether the *Mb*PylRS/tRNA$_{CUA}$ pair [4, 5, 34]and *Mj*TyrRS/tRNA$_{CUA}$ pair are mutually orthogonal in their aminoacylation specificity. Our experiments demonstrate that there is no cross-acylation (grey arrows) between the two aminoacyl-tRNA synthetase/tRNA$_{CUA}$ pairs (as shown by decoding the amber codon in myo4TAGHis$_6$

using the different combinations of synthetases and tRNAs, see below). However, both tRNAs direct the incorporation of their amino acid in response to the amber codon. **B.** *E. coli* DH10B were transformed with pMyo4TAG-His_6, a plasmid holding the gene for sperm whale myoglobin with an amber codon at position 4 and a C-terminal hexahistidine tag and an expression cassette for either *Mb*$tRNA_{CUA}$ or *Mj*$tRNA_{CUA}$. *Mb*PylRS or *Mj*TyrRS were provided on pBKPylS or pBK*Mj*TyrRS, respectively. Cells expressing *Mb*PylRS received 10 mM **3** (BocLys) as a substrate for the synthetase. Myoglobin-His_6 produced by the cells was purified by Ni^{2+}-affinity chromatography, analysed by SDS-PAGE and detected with Coomassie stain or Western blot against the His_6-tag.



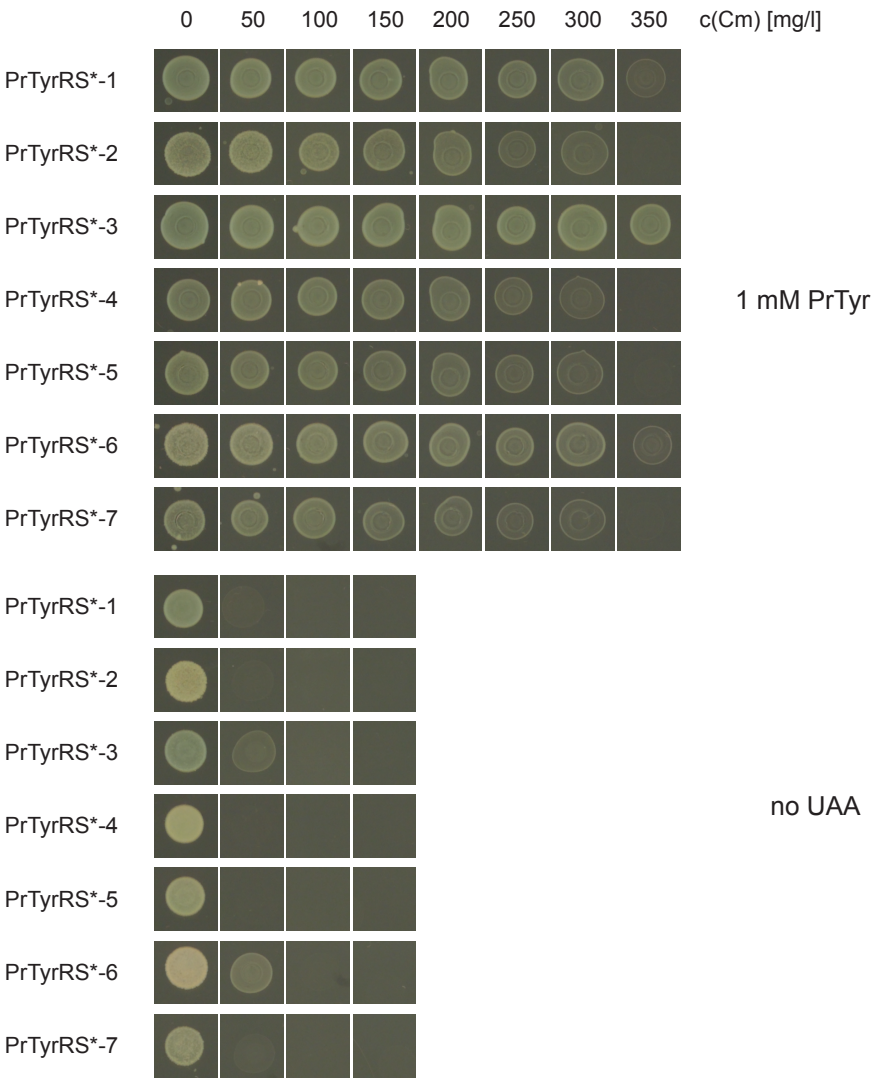
Supplementary Figure 9. Genetically encoding **2** in response to a quadruplet codon. **A.** *Mj*AzPheRS aminoacylates its cognate amber suppressor $tRNA_{CUA}$ with **2**. To differentiate the codons that the two mutually orthogonal tRNAs decode and to create a pair for the incorporation of an unnatural amino acid in response to a quadruplet codon, we altered the anticodon of *Mj*$tRNA_{CUA}$ from CUA to UCCU to create *Mj*$tRNA_{UCCU}$. After this, the resulting $tRNA_{UCCU}$ is no longer a substrate of the parent *Mj*AzPheRS. To create a version of *AzPheRS-7* that aminoacylates *Mj*$tRNA_{UCCU}$ we identified six residues (Y230, C231, P232, F261, H283, D286) in the parent synthetase that recognize the anticodon of the tRNA [35] and mutated these residues to all possible combinations, creating a library of 10^8 possible synthetase mutants. To select for *AzPhe*RS mutants that specifically aminoacylate *Mj*$tRNA_{UCCU}$ we created a chloramphenicol acetyl transferase reporter (pREP JY(UCCU), derived from pREP YC-JYCUA [32]), which contains the four base codon AGGA at position 111, a site

permissive to the incorporation of a range of amino acids. In the absence or presence of *AzPhe*RS/*Mj*tRNA$_{UCCU}$ this reporter confers resistance to chloramphenicol at low levels (30-50 μg ml^{-1}). We selected synthetase variants on 150 μg ml^{-1} of chloramphenicol that, in combination with *Mj*tRNA$_{UCCU}$, specifically direct the incorporation of **2** in response to the AGGA codon on pREP JY(UCCU). We characterized 24 synthetase/tRNA$_{UCCU}$ pairs by their chloramphenicol resistance in the presence of **2** and pREP JY(UCCU). The seven best synthetase/tRNA$_{UCCU}$ combinations confer a chloramphenicol resistance of 250-350 μg ml^{-1} on cells containing **2** and pREP JY(UCCU) (**Supplementary Figure 10**). In the absence of the **2**, we observe only background levels of resistance (30 μg ml^{-1}) for several synthetases indicating that the synthetase/*Mj*tRNA$_{UCCU}$ pairs specifically direct the incorporation of **2** in response to the quadruplet codon AGGA. Sequencing these seven clones revealed similar but non-identical mutations (**Supplementary Figure 10**). **B.** Library design. Structure of *Mj*TyrRS (grey) bound to its cognate tRNA (orange). Residues of the synthetase that recognize the anticodon and which are mutated in the library, as well as bases of the natural anticodon (G34, U35, A36) are shown in blue (Figure created using Pymol, www.pymol.org, and pdb-file 1J1U). **C.** The production of full-length myoglobin from myo4(AGGA)-his$_6$ by the AzPheRS*-2/*Mj*tRNA$_{UCCU}$ pair is dependent on the presence of **2**. In the remainder of the text we refer to *Mj*AzPheRS*-2 as *Mj*AzPheRS* for simplicity. *Mj*AzPheRS*/tRNA$_{UCCU}$ efficiently suppress an AGGA codon placed into the myoglobin gene. *E. coli* DH10B were transformed with pMyo4TAG-His$_6$ or pMyo4AGGA-His$_6$, a plasmid holding the gene for sperm whale myoglobin with an amber or an AGGA codon at position 4, respectively, and a C-terminal hexahistidine tag and an expression cassette for either *Mj*tRNA$_{CUA}$ or *Mj*tRNA$_{UCCU}$. *Mj*AzPheRS or *Mj*AzPheRS* were provided on pBK*Mj*AzPheRS or pBK*Mj*AzPheRS*, respectively. Cells received 2.5 mM **2** as a substrate for the synthetase. Myoglobin-His$_6$ produced by the cells was purified by Ni^{2+}-affinity chromatography, analysed by SDS-PAGE and detected with Coomassie stain. **D.** *Mj*AzPheRS*/tRNA$_{UCCU}$ decodes AGGA codons specifically with **2**. The incorporation of **2** into myoglobin-His$_6$ purified from cells expressing Myo4(AGGA) and *Mj*AzPheRS*/tRNA$_{UCCU}$ in the presence of 2.5 mM **2** was analysed by ESI-MS. The mass of the observed peak (18457.75 Da) corresponds to the calculated mass of myoglobin containing a single **2** (18456.2 Da).



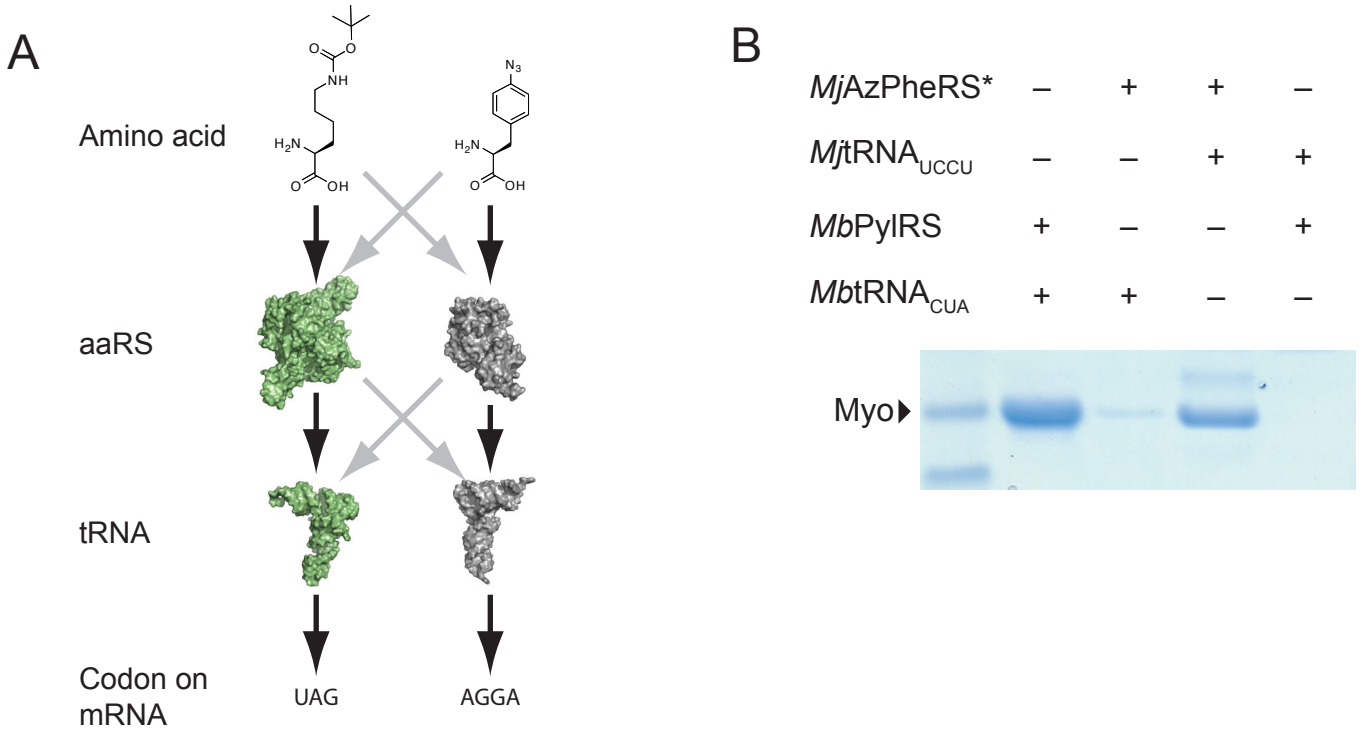
Name	Y230	C231	P232	F261	H283	D286
*Mj*AzPheRS*-1	P	R	R	F	G	G
*Mj*AzPheRS*-2	K	K	K	P	W	G
*Mj*AzPheRS*-3	H	P	P	G	G	G
*Mj*AzPheRS*-4	A	Q	N	G	W	E
*Mj*AzPheRS*-5	K	K	K	T	W	P
*Mj*AzPheRS*-6	L	S	L	P	I	S
*Mj*AzPheRS*-7	K	K	K	F	Q	S

Supplementary Figure 10: Amino acid dependent growth of selected *Mj*AzPheRS* variants. *E. coli* DH10B were co-transformed with isolates from a library built on pBK *Mj*AzPheRS-7 and pREP JY(UCCU) (coding for *Mj*tRNA$_{UCCU}$ and chloramphenicol acetyltransferase with an AGGA codon at position D111). Cells were grown in the presence or absence of 1 mM **2** for 5 h and pronged onto LB agar plates containing 25 μg ml^{-1} kanamycin, 12.5 μg ml^{-1} tetracycline and the indicated concentration of chloramphenicol with or without the unnatural amino acid. Plates were photographed after 18 h at 37ºC. Sequencing of mutations for incorporating tyrosine, **2** and propargyl-L-tyrosine (**Supplementary Figure 11**) in response to the AGGA codon reveals clones with common mutations Y230K, C231K and P232K, but divergent mutations at positions F261, H283 and D286. This suggests that amino acids 230, 231 and 232 confer affinity and specificity for the anticodon, and that 261, 283 and 286 may couple the identity of the anticodon to the amino acid identity.

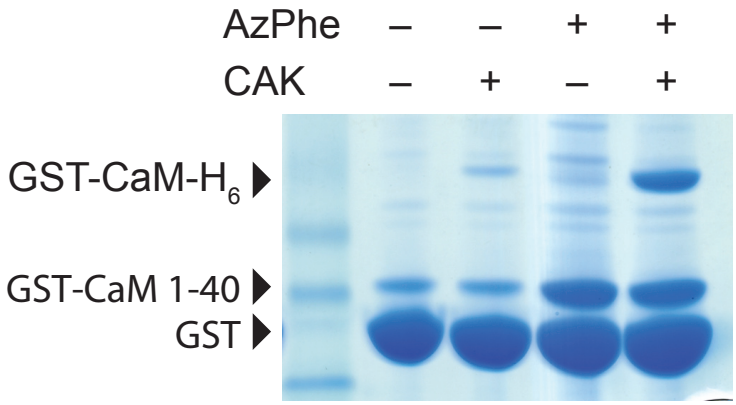


Name	Y230	C231	P232	F261	H283	D286
*Mj*PrTyrRS*-1	G	G	G	Y	A	L
*Mj*PrTyrRS*-2	G	G	G	Y	A	L
*Mj*PrTyrRS*-3	G	G	G	Y	A	L
*Mj*PrTyrRS*-4	M	S	G	F	G	G
*Mj*PrTyrRS*-5	T	R	K	P	T	C
*Mj*PrTyrRS*-6	K	K	K	G	Q	W
*Mj*PrTyrRS*-7	R	R	K	P	T	Q

Supplementary Figure 11: Amino acid dependent growth of selected *Mj*PrTyrRS* variants. *E. coli* DH10B transformed as in **Supplementary Figure 10** using isolates from a library built on *Mj*PrTyrRS and tested for unnatural amino acid dependent growth. Mutations relative to *Mj*PrTyrRS are given in the table below.



Supplementary Figure 12: The *Mb*PylRS/*Mb*tRNA$_{CUA}$ and *Mj*AzPheRS*/tRNA$_{UCCU}$ pairs incorporate distinct unnatural amino acids in response to distinct unique codons. **A.** The two orthogonal pairs (*Mb*PylRS/*Mb*tRNA$_{CUA}$ and *Mj*AzPheRS*/tRNA$_{UCCU}$) decode two distinct codons in the mRNA (UAG and AGGA) with two distinct amino acids (*N6*-[(tert.-butyloxy)carbonyl]-L-lysine and **2**). *Mb*PylRS does not aminoacylate *Mj*tRNA$_{UCCU}$ and *Mb*tRNA$_{CUA}$ is not a substrate for *Mj*AzPheRS*. **B.** Suppression of a cognate codon at position 4 in the gene of sperm whale myoglobin by different combinations of *Mb*PylRS/*Mb*tRNA$_{CUA}$ and *Mj*AzPheRS*/tRNA$_{UCCU}$. *E. coli* DH10B were transformed with pMyo4TAG-His$_6$ or pMyo4AGGA-His$_6$ as described in Figure 6C. Cells were provided with *Mb*PylRS (on pBKPylS) or *Mj*AzPheRS* (on pBK*Mj*PheRS*) and 2.5 mM *N6*-[(tert.-butyloxy)carbonyl]-L-lysine or 5 mM **2**, respectively. Myoglobin-His$_6$ produced by the cells was purified by Ni^{2+}-affinity chromatography, analysed by SDS-PAGE and detected with Coomassie stain. We see weak incorporation in response to the UAG codon using the *Mb*PylRS pair. This incorporation is independent of the presence of *Mj*AzPheRS* and results from a low level background acylation of the tRNA by *E. coli* synthetases in rich media, as previously observed.



Supplementary Figure 13: Encoding an azide and an alkyne in a single protein *via* orthogonal translation. **A.** Expression of GST-CaM-His_6 containing two unnatural amino acids. *E. coli* DH10B were transformed with four plasmids: pCDF PylST (expressing *Mb*PylRS and *Mb*$tRNA_{CUA}$), pDULE AzPheRS* $tRNA_{UCCU}$ (encoding *Mj*AzPheRS*/$tRNA_{UCCU}$), pSC101* ribo-Q1 and p-O-*gst-CaM*-His_6 1AGGA 40UAG (a GST-CaM-His_6 fusion translated by the orthogonal ribosome that contains an AGGA codon at position 1 and an amber codon at position 40 of calmodulin (CaM)). Cells were grown in LB medium containing antibiotics to maintain the plasmids and 2.5 mM **4** and/or 5 mM **2** as indicated. Cells were harvested, lysed and the protein purified on GSH-beads. Bound protein was eluted with 10 mM GSH in PBS and analysed by SDS-PAGE. A part of this gel is shown in **Figure 3** of the main text. Full-length protein was produced by this method with yields of upto 0.5 mg/L

	Supplementary Table 1
Name	Sequence 5'-3'
Lib1	**516-520 plus 529-535**
Lib1fN7	GGAAAGGTCTCACAGCCGCNNNNNNNCGGAGGGTGCAAGCGTTAATCGGAATTACTG
Lib1rN5	GGAAGGTCTCAGCTGCNNNNNCGGAGTTAGCCGGTGCTTCTTCTGCGGGTAACGTCAATG
Lib2	**1405-1409 plus 1491-1496**
Lib2a 1405-1409	
Lib2afN5	GGAAAGGTCTCACACCGCCCNNNNNACCATGGGAGTGGGTTGCAAAAGAAGTAGGTAGCTTAAC
Lib2ar	GGAAAGGTCTCTGGTGTGTACAAGGCCCGGGAACGTATTCACCGTGGCATTC
Lib2b 1491-1496	
Lib2bfN6	GGAAAGGTCTCACTGGGGTNNNNNNGTAACAAGGTAACCGTAGGGGAACCTGCGGTTGGATCATGGGATTAC
Lib2br	GGAAAGGTCTCTCCAGTCATGAATCACAAAGTGGTAAGCGCCCTCCCGAAG
Lib3	**1396-1401 plus 1501-1506**
Lib3a 1396-1401	
Lib3afN6	GGAAAGGTCTCACTTGTACNNNNNNCCCGTCACACCATGGGAGTGGGTTGCAAAAGAAGTAG
Lib3ar	GGAAAGGTCTCTCAAGGCCCGGGAACGTATTCACCGTGGCATTCTGATC
Lib3b 1501-1506	
Lib3bfN6	GGAAAGGTCTCAGTCGTAANNNNNNAACCGTAGGGGAACCTGCGGTTGGATCATGGGATTAC
Lib3br	GGAAAGGTCTCACGACTTCACCCCAGTCATGAATCACAAAGTGGTAAG
Lib4	**974-975 plus 954-961 plus 1225-1226**
Lib4ab 974+975 plus 954-961	
Lib4afN2	GGAAAGGTCTCACAACGCGNNGAACCTTACCTGGTCTTGACATCCACGGAAGTTTTC
Lib4brN6	GGAAAGGTCTCAGTTGCATCGNNNNNNNNNCACATGCTCCACCGCTTGTGCGGGCCCCCGTCAATTC
Lib4c 1225+1226	
Lib4cfN2	GGAAAGGTCTCACCAGGGCTNNACACGTGCTACAATGGCGCATACAAAGAGAAG
Lib4cr	GGAAAGGTCTCACTGGTCGTAAGGGCCATGATGACTTGACGTCATCCCCAC
Lib5	**962-973**
Lib5fN12	GGAAAGGTCTCTGTGGTTTAATTNNNNNNNNNNNNNAAGAACCTTACCTGGTCTTGACATCCACGGAAGTTTTC
Lib5r	GGAAAGGTCTCACCACATGCTCCACCGCTTGTGCGGGCCCCCGTCAATTC
Lib9	**1048-1055 plus 1205-1209**
Lib9a 1048-1055	
Lib9afN8	GGAAAGGTCTCTCGTGAGACAGNNNNNNNNTGGCTGTCGTCAGCTCGTGTTGTGAAATGTTG
Lib9ar	GGAAAGGTCTCTCACGGTTCCCGAAGGCACATTCTCATCTCTG
Lib9b 1205-1209	
Lib9bfN5	GGAAAGGTCTCACGTCAAGTCATCANNNNNCTTACGACCAGGGCTACACACGTGCTACAATG
Lib9br	GGAAAGGTCTCTGACGTCATCCCCACCTTCCTCCAGTTTATCAC
Lib10	**1058-1062 plus 1194-1199**
Lib10a 1194-1199	
Lib10afN6	GGAAAGGTCTCAGATGACGNNNNNNCATCATGGCCCTTACGACCAGGGCTACACACGTGCTAC
Lib10ar	GGAAAGGTCTCTCATCCCCACCTTCCTCCAGTTTATCACTGGCAGTCTC
Lib10b 1058-1062	
Lib10bfN5	GGAAAGGTCTCACTGCATGNNNNNCGTCAGCTCGTGTTGTGAAATGTTGGGTTAAGTC
Lib10br	GGAAAGGTCTCAGCAGCACCTGTCTCACGGTTCCCGAAGGCACATTC
Lib11	**1057-1058 plus 1199-1203**
Lib11a 1199-1203	
Lib11afN5	GGAAAGGTCTCTGACGTCAAGNNNNNATGGCCCTTACGACCAGGGCTACACACGTGCTACAATG
Lib11ar	GGAAAGGTCTCACGTCATCCCCACCTTCCTCCAGTTTATCACTGGCAGTCTC
Lib11b 1057-1058	
Lib11bfN2	GGAAAGGTCTCACTGCATNNCTGTCGTCAGCTCGTGTTGTGAAATGTTGGGTTAAGTC
Lib11br	GGAAAGGTCTCAGCAGCACCTGTCTCACGGTTCCCGAAGGCACATTC

Lib12	**1062-1066 plus 1190-1194**
Lib12a 1190-1194	
Lib12afN5	GGAAAGGTCTCAGTGGGGATNNNNNCAAGTCATCATGGCCCTTACGACCAGGGCTACACACGTGCTAC
Lib12ar	GGAAAGGTCTCTCCACCTTCCTCCAGTTTATCACTGGCAGTCTCCTTTG
Lib12b 1062-1066	
Lib12bfN5	GGAAAGGTCTCAATGGCTGNNNNNAGCTCGTGTTGTGAAATGTTGGGTTAAGTCCCGCAACGAG
Lib12br	GGAAAGGTCTCACCATGCAGCACCTGTCTCACGGTTCCCGAAGGCACATTC
Lib13	**1053-1058 plus 1199-1205**
Lib13a 1053-1058	
Lib13afN6	GGAAAGGTCTCAGGTGCTNNNNNNCTGTCGTCAGCTCGTGTTGTGAAATGTTGGGTTAAGTC
Lib13ar	GGAAAGGTCTCACACCTGTCTCACGGTTCCCGAAGGCACATTCTC
Lib13b 1199-1205	
Lib13bfN7	GGAAAGGTCTCTGACGTCAAGNNNNNNNGGCCCTTACGACCAGGGCTACACACGTGCTACAATG
Lib13br	GGAAAGGTCTCAGCAGCACCTGTCTCACGGTTCCCGAAGGCAC
SuLib	**964 965 967 968 970 971 1054 1194 1196 1197**
SuLiba 964 965 967 968 970 971 1054	
SuLibafN1	GGAAAGGTCTCTCAGAGATGAGAATGTGCCTTCGGGAACCGTGAGACAGGTGCTGNATGGCTGTCGTCAGCTCGTGTTGTGA AATGTTG
SuLibarN6	GGAAAGGTCTCATCTGAAAACTTCCGTGGATGTCAAGACCAGGTAAGGTTCTTCGNNTNNCNNCGAATTAAACCACATGCTC CACCGCTTGTGCG
SuLibb 1194 1196 1197	
SuLibbfN3	GGAAAGGTCTCAGATGATGNCNNGTCATCATGGCCCTTACGACCAGGGCTACACACGTGCTACAATG
SuLibbr	GGAAAGGTCTCTCATCCCCACCTTCCTCCAGTTTATCACTGGCAGTCTCCTTTG
CAT146AAGAf	TGTTCTTCGTCAAGAGCCAACCCGTGGGTCAGCTTC
CAT146AAGAr	ACGGGTTGGCTCTTGACGAAGAACATGTTTTCGATG
CAT146AGGAf	TGTTCTTCGTCAGGAGCCAACCCGTGGGTCAGCTTC
CAT146AGGAr	ACGGGTTGGCTCCTGACGAAGAACATGTTTTCGATG
CAT103AGGAf	GAAACCTTCAGGAAGCCTGTGGAGCGAATACCACGAC
CAT103AGGAr	CACAGGCTTCCTGAAGGTTTCGGTCTGTTCGTGGAAG
CAT146CCCUf	TGTTCTTCGTCCCCTGCCAACCCGTGGGTCAGCTTC
CAT146CCCUr	ACGGGTTGGCAGGGGACGAAGAACATGTTTTCGATG
CAT103CCCUf	GAAACCTTCCCCTAGCCTGTGGAGCGAATACCACGAC
CAT103CCCUr	CACAGGCTAGGGGAAGGTTTCGGTCTGTTCGTGGAAG
CAT146UAGAf	TGTTCTTCGTCTAGAGCCAACCCGTGGGTCAGCTTC
CAT146UAGAr	ACGGGTTGGCTCTAGACGAAGAACATGTTTTCGATG
CAT103UAGAf	GAAACCTTCTAGAAGCCTGTGGAGCGAATACCACGAC
CAT103UAGAr	CACAGGCTTCTAGAAGGTTTCGGTCTGTTCGTGGAAG
Ser2AAGAf	ACCGGTATTCTTACACCGGAGTAGGGGCAACTCTA
Ser2AAGAr	CTCCGGTGTAAGAATACCGGTCCGTTCAGCCGCTC
Ser2AGGAf	ACCGGTCTTCCTAAACCGGAGTAGGGGCAACTCTA
Ser2AGGAr	CTCCGGTTTAGGAAGACCGGTCCGTTCAGCCGCTC
Ser2CCCUf	ACCGGTGTAGGGTAACCGGAGTAGGGGCAACTCTA
Ser2CCCUr	CTCCGGTTACCCUACACCGGTCCGTTCAGCCGCTC
Ser2UAGAf	ACCGGTATTCTAACACCGGAGTAGGGGCAACTCTA
Ser2UAGAr	CTCCGGTGTUAGAATACCGGTCCGTTCAGCCGCTC

sc101Q1f	GGAAAGGTCTCAGATGATGTCGGGTCATCATGGCCCTTACGACCAGGGCTACACACGTGCTAC
sc101Q3f	GGAAAGGTCTCAGATGACGTTGGGTCATCATGGCCCTTACGACCAGGGCTACACACGTGCTAC
sc101Q4f	GGAAAGGTCTCAGATGACGTAGAGTCATCATGGCCCTTACGACCAGGGCTACACACGTGCTAC
sc101Qr	GGAAAGGTCTCTCATCCCCACCTTCCTCCAGTTTATCACTGGCAGTCTC
GMx1AGGAf	CGTGACGGGAGGACTCAAAATCGAAGAAGGTAAACTG
GMx1AGGAr	TCTTCGATTTTGAGTCCTCCCGTCACGATGAATTC
GMx2AGGAf	GACGGGAAGACTCAAAATCAGGAGAAGGTAAACTGGTAATCTGGATTAACG
GMx2AGGAr	TACCTTCTCTTGATTTTGAGTCCTCCCGTCACGATGAATTCCCGGGGATC
MBPY17AGGAf	CGATAAAGGCAGGAAACGGTCTCGCTGAAGTCCGGT
MBPY17AGGAr	CGAGACCGTTTCCTGCCTTTATCGCCGTTAATCCA
MBPS211tagf	ACACCGATTACTAGATCGCAGAAGCTGCCTTTAAT
MBPS211tagr	GCTTCTGCGATCTAGTAATCGGTGTCTGCATTCAT
MBPN234tagf	GGGCATGGTCCTAGATCGACACCAGCAAAGTGAAT
MBPN234tagr	CTGGTGTCGATCTAGGACCATGCCCACGGGCCGTT
GMK236Qf	ACGGGAGGACTCCAAATCGAATAGGGTAAACTGGT
GMK236Qr	CCTATTCGATTTGGAGTCCTCCCGTCACGATGAAT
GM238UAGf	CGAAAATCGAATAGGGTAAACTGGTAATCTGGATT
GM238UAGr	ACCAGTTTACCCTATTCGATTTTCGATCCTAGAGT
MjtYAGGA2f	CGGCGGACTTCCTAATCCGCATGTCGCTGGT
MjtYAGGA2r	ATGCGGATTAGGAAGTCCGCCGTTCTACCAG
CATp111AGGAf	AATACCACAGGAGATTTCCGGCAGTTTCTA
CATp111AGGAr	CGGAAATCTCCTGTGGTATTCACTCCAGAG
GMAGGAf	AAATCGAAAGGAGGTAAACTGGTAATCTGGATTAA
GMAGGAr	AGTTTACCTCCTTTCGATTTTGAGCTACCCGTCAC
Myo4AGGAf	TGGTTCTGAGGAGAAGGTGAATGGCAGCTGGTTCT
Myo4AGGAr	TCACCTTCTCCTCAGAACCATGGTTAATTCCTCCT
PylSBamf	CCAGGATCCTCGGGAGTTGTCAGCCTGTC
PylSSalr	ATGGTCGACCGCCGAACGCGGCGTTTTG
PylTSalf2	GCGGTCGACACAGATGTAGGTGTTCCACAG
PylTNotr2	TATGCGGCCGCCAGAACATATCCATCGCGTC
CamEcof	CGGGAATTCAAGCTGACCAACTGACAGAAGAG
CamH6Hindr	ACTAAGCTTAGTGATGGTGATGGTGATGCTTTGCTGTCATCATTTGTAC
CaM1aggaf	GCGTGGATCCAGGAGCTGACCAACTGACAGAAGAG
CaM1aggar	GTTGGTCAGCTCCTGGATCCACGCGGAACCAGATC
CaMK149TAGf	TGATGACAGCATAGCATCACCATCACCATCACTAA
CaMK149TAGr	ATGGTGATGCTATGCTGTCATCATTTGTACAAACT
CaM40tagf	TGAGGTCGCTTTAGCAAAACCCAACGGAAGCAGAA
CaM40tagr	GTTGGGTTTTGCTAAAGCGACCTCATAACGGTGCC

Supplementary Table 1: Oligonucleotides used in this study.

References

31. Stemmer, W. P. & Morris, S. K. Enzymatic inverse PCR: a restriction site independent, single-fragment method for high-efficiency, site-directed mutagenesis. Biotechniques 13, 214-20 (1992).
32. Santoro, S. W., Wang, L., Herberich, B., King, D. S. & Schultz, P. G. An efficient system for the evolution of aminoacyl-tRNA synthetase specificity. Nat Biotechnol 20, 1044-8 (2002).
33. Rice, J. B., Libby, R. T. & Reeve, J. N. Mistranslation of the mRNA encoding bacteriophage T7 0.3 protein. J Biol Chem 259, 6505-10 (1984).
34. Hao, B. et al. A new UAG-encoded residue in the structure of a methanogen methyltransferase. Science 296, 1462-6 (2002).
35. Kobayashi, T. et al. Structural basis for orthogonal tRNA specificities of tyrosyl-tRNA synthetases for genetic code expansion. Nat Struct Biol 10, 425-32 (2003).