nature portfolio

https://doi.org/10.1038/s41557-021-00764-5

Supplementary information

A 68-codon genetic code to incorporate four distinct non-canonical amino acids enabled by automated orthogonal mRNA design

In the format provided by the authors and unedited

Supplementary Information

A 68-codon genetic code to incorporate four distinct non-canonical amino acids enabled by automated orthogonal mRNA design

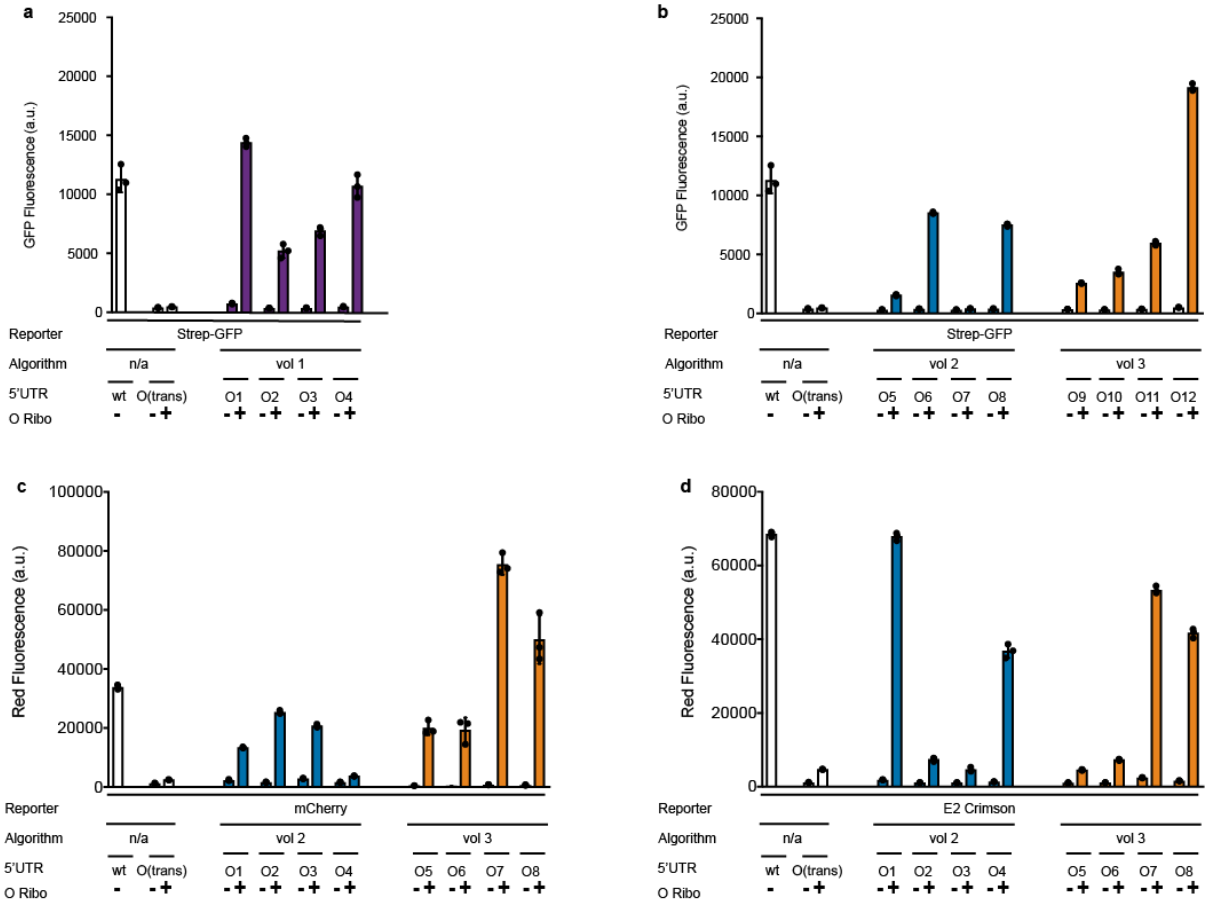
Daniel L. Dunkelmann[1, †], Sebastian B. Oehm[1, †], Adam T. Beattie[1], Jason W. Chin[1,*]

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

[†]These authors contributed equally.

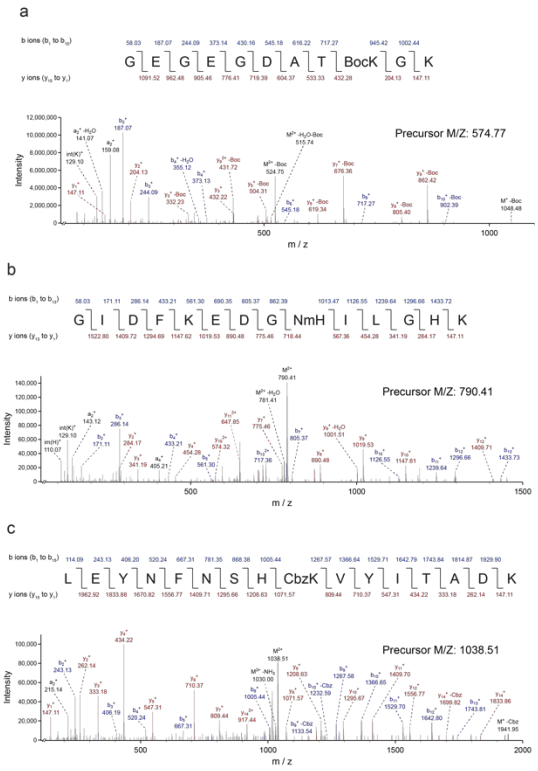
[*]Correspondence: chin@mrc-lmb.cam.ac.uk

Supplementary Figures



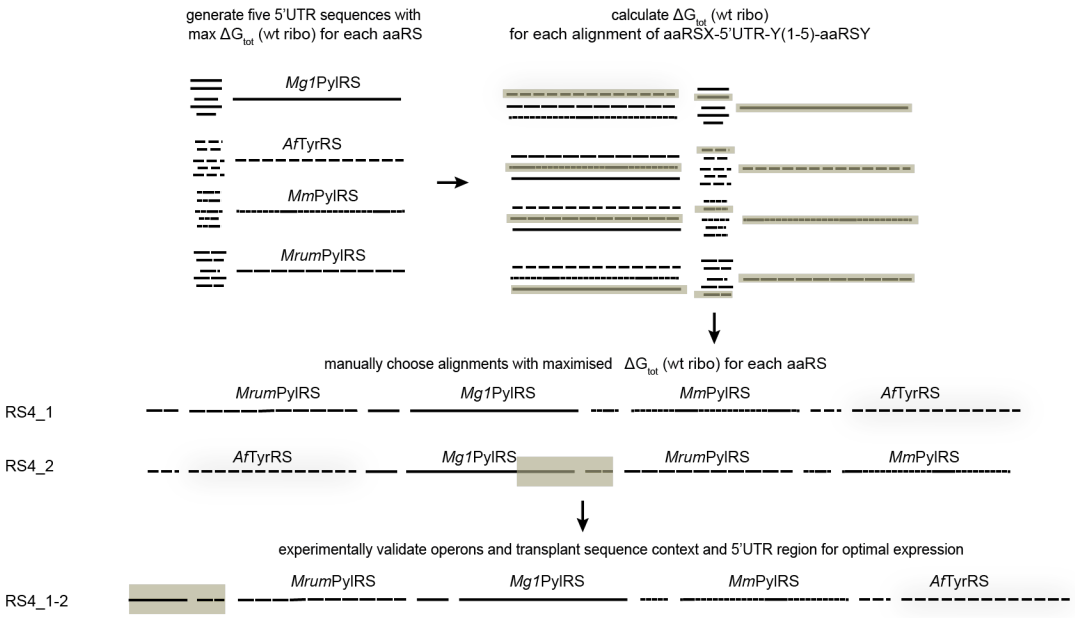
Supplementary Figure 1

Fluorescence measurements of reporter protein production by from O-mRNAs generated by the indicated algorithm. We cloned the sequences (O1-O12 $_{strep}GFP_{His6}$ for $_{strep}GFP_{His6}$ (**a** and **b**), O1-O8 E2Crimson for *E2Crimson* (**c**) as well as O1-O8 mCherry for *mCherry* (**d**) into a standardised p15A reporter construct, and produced proteins in the presence of a plasmid encoding either the O-ribosome or an additional copy of the wt ribosome. Control experiments used a construct with a 5' UTR and RBS commonly used in our lab (wt), and a construct with the O(trans) 5' UTR previously used for highly efficient O-GST-CaM production[1, 2]. Bars represent the mean of three biological replicates ± s. d.. Dots represent individual experiments. All numerical values are provided in **Supplementary Table 1**.



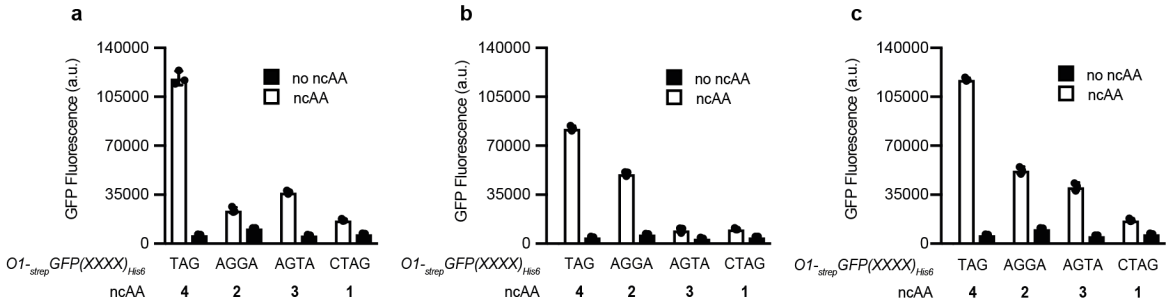
Supplementary Figure 2

MS/MS spectra of ncAA-containing peptides obtained following tryptic digest of $_{strep}$GFP(40BocK, 136NmH, 150CbzK)$_{His6}$. The precursor ions confirm the incorporation of the ncAAs. Fragmentation of each peptide is predicted to yield a series of b ions (blue) and a series of y ions (red), as well as ions corresponding to the loss of the lysine protecting groups in the fragmentation process (**a** and **c)**. Ion peaks were assigned manually; along with precursor ion masses, these confirmed the incorporation of each ncAA at its expected position. The mass spectrometry analysis was performed three times with similar results. **a**, MS/MS spectra confirming BocK **1** incorporation at position 40. **b**, MS/MS spectra confirming NmH **2** incorporation at position 136. **c**, MS/MS spectra confirming CbzK **3** incorporation at position 150.



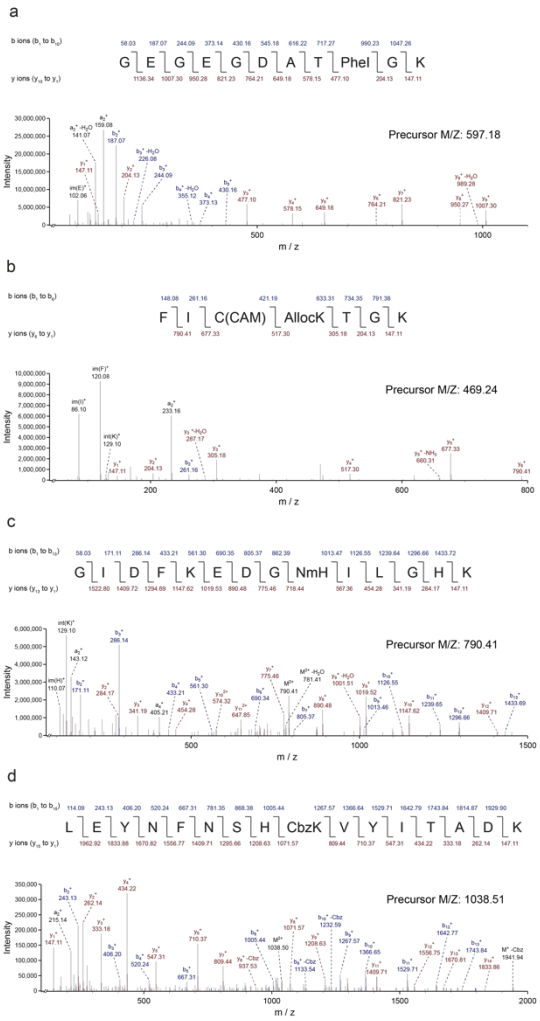
Supplementary Figure 3

The assembly pipeline for the generation of polycistronic operons containing the genes for four mutually orthogonal aaRSs (*Af*TyrRS(PheI), *Mrum*PylRS(NmH), *Mg1*PylRS(CbzK) and *Mm*PylRS). For each synthetase, five 5' UTR sequences were generated using the online RBS calculator[3-7] optimised for max ΔG_{tot} (wt ribo). Then, ΔG_{tot} (wt ribo) for each alignment of the form aaRSX-5'_UTR(Y1-Y5)-aaRSY (where X and Y refer to any combination of two out of the four synthetases) was calculated using the online tool[3-7]. Finally, all four synthetases were manually aligned in a way that guaranteed a high ΔG_{tot} (wt ribo) for each synthetase. Two independent solutions yielded similar results. After experimental validation, the favorable sequence context of one synthetase was copied into the other operon yielding the final construct (all 5' UTR sequences and ΔG_{tot} (wt ribo) are given in **Supplementary Table 3**).



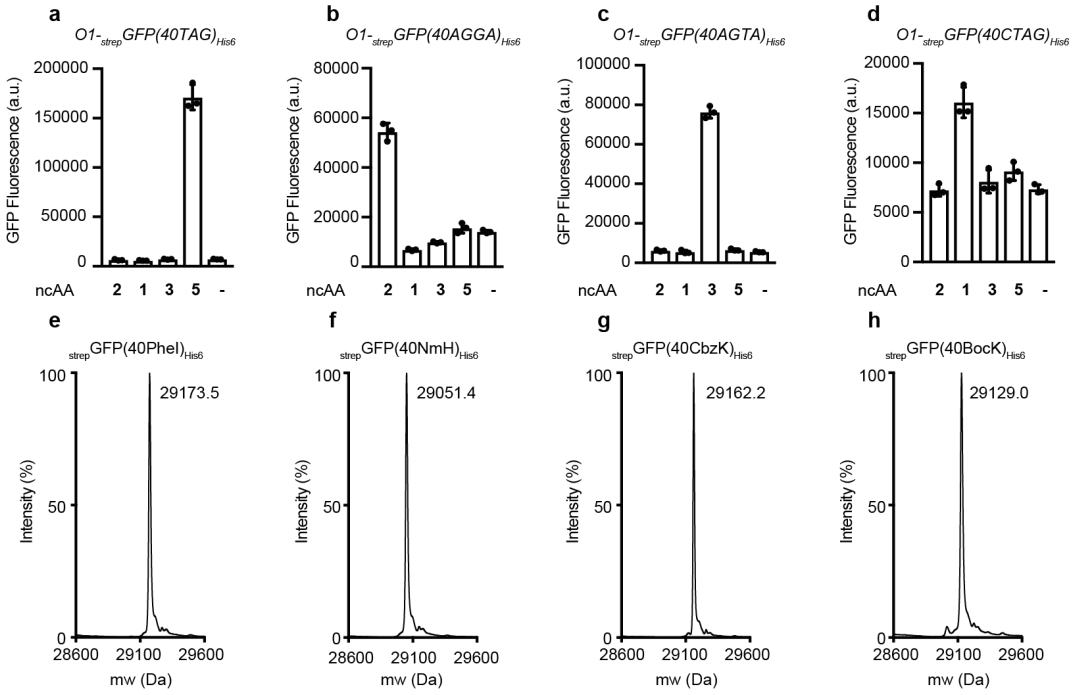
Supplementary Figure 4

Fluorescence from cells containing *O1-$_{strep}$GFP(XXXX)$_{His6}$*, with XXXX being either TAG, CTAG, AGGA or AGTA. *E. coli* also contained O-riboQ1 and *Mm*PylRS/*Mspe*PyltRNA$_{CUAG}$, *Mrum*PylRS(NMH)/*Mint*PyltRNA(^{A}17,^{V}C10)$_{UCCU}$, *Af*TyrRS/*Af*RNA$_{CUA}$ and *Mg1*PylRS(CbzK)/*Malv*PyltRNA(8)$_{UACU}$ and one of the ncAAs: NmH **2**, CbzK **3**, BocK **1** or PheI **5**. Synthetases were initially either arranged in operons RS4_1/tRNA4 or RS4_2/tRNA4 (see Supplementary Figure 3 and Supplementary Table 3). RS4_1/tRNA4 (**a**) yielded better results for the decoding of TAG, CTAG and AGTA; however, AGGA was only decoded with half of the efficiency as in RS4_2/tRNA4 (**b**). Therefore 150 nt region upstream of *Mrum*PylRS was copied into RS4_1/tRNA4 yielding operon 1 RS4_1-2/tRNA4 (**c**) leading to a 2.6 higher activity of *Mrum*PylRS. Bars represent the mean of three biological replicates ± s. d.. Dots represent individual experiments. All numerical values are provided in **Supplementary Table 5**.



Supplementary Figure 5

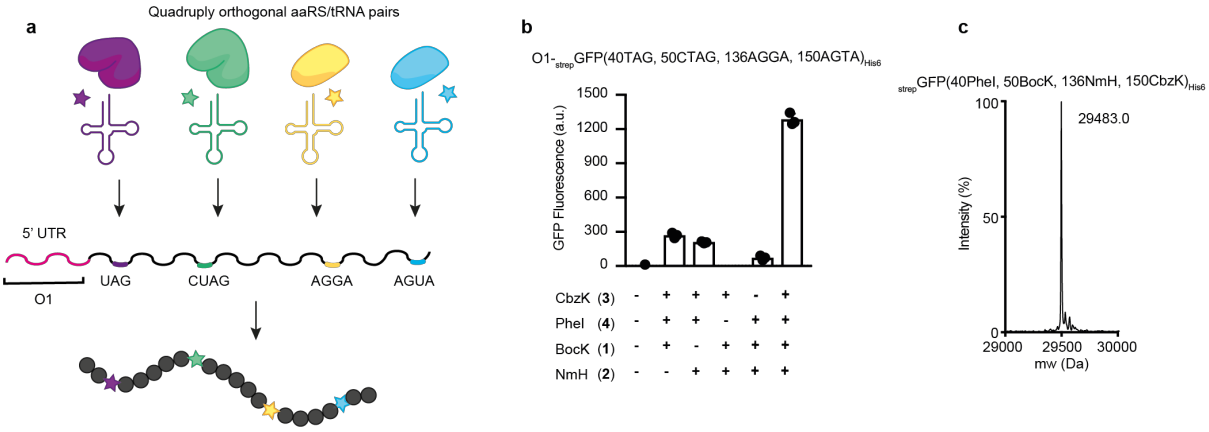
MS/MS spectra of ncAA-containing peptides obtained following tryptic digest of $_{strep}$GFP(40PheI, 50AllocK, 136NmH, 150CbzK)$_{His6}$. The precursor ions confirm the incorporation of the ncAAs. Fragmentation of each peptide is predicted to yield a series of b ions (blue) and a series of y ions (red), as well as ions corresponding to the loss of the lysine protecting groups in the fragmentation process (**d**). Ion peaks were assigned manually; along with precursor ion masses, these confirmed the incorporation of each ncAA at its expected position. The mass spectrometry analysis was performed three times with similar results. **a**, MS/MS spectra confirming PheI **5** incorporation at position 40. **b**, MS/MS spectra confirming AllocK **4** incorporation at position 50. **c**, MS/MS spectra confirming NmH **2** incorporation at position 136. **d**, MS/MS spectra confirming CbzK **3** incorporation at position 150.



Supplementary Figure 6

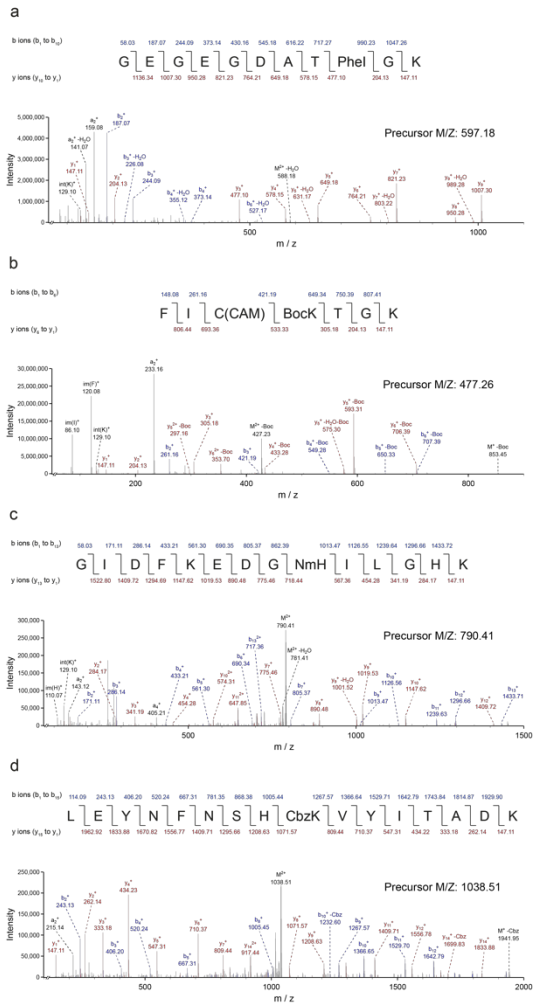
Four orthogonal aaRS/tRNA pairs, decoding one amber codon and three orthogonal quadruplet codons are expressed from aaRS operons and computationally generated tRNA operons. These pairs are mutually orthogonal in their aminoacylation specificity, recognize distinct ncAAs, and decode distinct orthogonal codons. **a-d,** Fluorescence from cells containing *O1-*$_{strep}$*GFP(40XXXX)*$_{His6}$, with XXXX being the codon at position 40 in sfGFP: TAG, CTAG, AGGA or AGTA. *E. coli* also contained ribo-Q1 and the aaRS and tRNA operons (aaRS4_1-2/tRNA4); these operons expressed *Mm*PylRS/*Mspe*tRNA$^{Pyl\text{-}evol}_{CUAG}$, *Mrum*PylRS(NMH)/*Mint*tRNA$^{Pyl\text{-}A17VC10}_{UCCU}$, *Af*TyrRS(PheI)/*Af*tRNA$^{Tyr\text{-}A01}_{CUA}$ and *Mg1*PylRS(CbzK)/*Malv*tRNA$^{Pyl\text{-}8}_{UACU}$. The indicated ncAAs: N^{π}-methyl-*L*-histidine (NmH) **2**, N^{6}-((benzyloxy)carbonyl)-*L*-lysine (CbzK) **3**, N^{6}-(*tert*-butoxycarbonyl)-*L*-lysine (BocK) **1**, (*S*)-2-amino-3-(4-iodophenyl)propanoic acid (PheI) **5** were added to cells or omitted (-). Each codon was only efficiently decoded in the presence of cognate ncAA of the aaRS/tRNA pair assigned to the respective quadruplet codon: (**a**) *O1-*$_{strep}$*GFP(TAG)*$_{His6}$ decoded by *Af*TyrRS(PheI)/*Af*tRNA$^{Tyr\text{-}A01}_{CUA}$, (**b**) *O1-*$_{strep}$*GFP(AGGA)*$_{His6}$ decoded by *Mrum*PylRS(NMH)/*Mint*tRNA$^{Pyl\text{-}A17VC10}_{UCCU}$, (**c**) *O1-*$_{strep}$*GFP(AGTA)*$_{His6}$ decoded by *Mg1*PylRS(CbzK)/ *Malv*tRNA$^{Pyl\text{-}8}_{UACU}$, and (**d**) *O1-*$_{strep}$*GFP(CTAG)*$_{His6}$ decoded by *Mm*PylRS/*Mspe*tRNA$^{Pyl\text{-}evol}_{CUAG}$. Bars represent the mean of three biological replicates ± s. d.. Dots

represent individual experiments. All numerical values are provided in **Supplementary Table 5**. **e-h,** Positive electrospray TOF-MS of nickel-NTA-purified $_{strep}GFP_{His6}$, expressed from *O1-*$_{strep}$*GFP(40XXXX)*$_{His6}$, with XXXX being either TAG (**e**), AGGA (**f**), AGTA (**g**) or CTAG (**h**), in the presence of NmH **2**, CbzK **3**, BocK **1**, PheI **5**. Cells also contained O-riboQ1 and operon aaRS4_2-1/tRNA4. $_{strep}GFP(40PheI)_{His6}$ mass predicted 29174.03 mass found 29174.2. $_{strep}GFP(40BocK)_{His6}$ mass predicted 29129.4 mass found 29129.0. $_{strep}GFP(40NmH)_{His6}$ mass predicted 29052.1 mass found 29052.5. $_{strep}GFP(40CbzK)_{His6}$ mass predicted 29163.3 mass found 29164.2. $_{strep}GFP(40BocK)_{His6}$ mass predicted 29129.4 mass found 29129.0. The mass spectrometry analysis was performed three times with similar results.



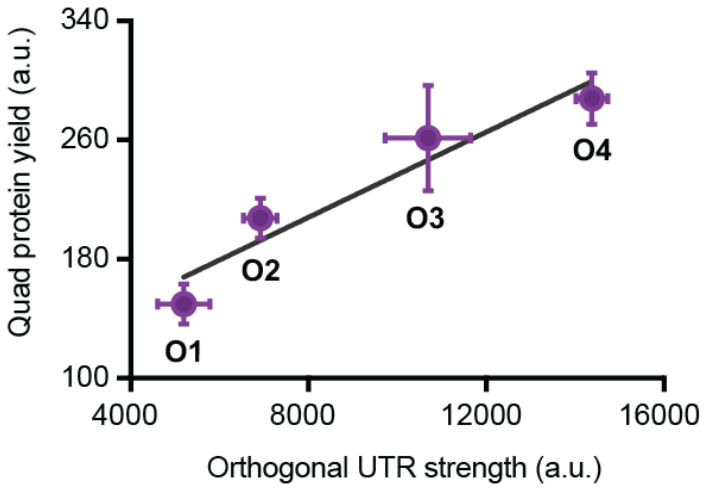
Supplementary Figure 7

Genetically encoding four distinct ncAAs into a protein in response to an amber codon and three distinct quadruplet codons. **a**, Schematic representation of four mutually orthogonal aaRS/tRNA pairs used for the incorporation of four distinct ncAAs in response to an amber codon and three distinct quadruplet codons. **b**, Efficient production of full length $_{strep}$GFP(40PheI, 50AllocK, 136NmH, 150CbzK)$_{His6}$ was dependent upon the addition of all four ncAAs (BocK **1**, NmH **2**, CbzK **3** and PheI **5**). Fluorescence from cells containing *O1-$_{strep}$GFP(40TAG, 50CTAG, 136AGGA, 150AGTA)$_{His6}$*, O-riboQ1, operon aaRS4_1-2/tRNA4 (encoding *Mm*PylRS/*Mspe*tRNA$^{Pyl\text{-}evol}_{CUAG}$, *Mrum*PylRS(NMH)/*Mint*tRNA$^{Pyl\text{-}A17VC10}_{UCCU}$, *Af*TyrRS(PheI)/*Af*tRNA$^{Tyr\text{-}A01}_{CUA}$ and *Mg1*PylRS(CbzK)/*Malv*tRNA$^{Pyl\text{-}8}_{UACU}$) in presence or absence of a combination of BocK **1,** NmH **2,** CbzK **3,** PheI **5.** Bars represent the mean of three biological replicates ± s. d.. Dots represent individual experiments. All numerical values are provided in **Supplementary Table 5**. **c,** TOF-MS ES+ of purified $_{strep}$GFP(40PheI, 50BocK, 136NmH, 150CbzK)$_{His6}$ purified from cells containing *O1-$_{strep}$GFP(40TAG, 50CTAG, 136AGGA, 150AGTA)$_{His6}$*, O-riboQ1 and operon RS4_1-2/tRNA4 in presence of 8 mM BocK **1**, 4 mM NmH **2**, 2 mM PheI **5** and 2 mM CbzK **3**. Mass predicted 29482.0, mass found 29483.0. The mass spectrometry analysis was performed three times with similar results.



Supplementary Figure 8

MS/MS spectra of ncAA-containing peptides obtained following tryptic digest of $_{strep}$GFP(40PheI, 50BocK, 136NmH, 150CbzK)$_{His6}$. The precursor ions confirm the incorporation of the ncAAs. Fragmentation of each peptide is predicted to yield a series of b ions (blue) and a series of y ions (red), as well as ions corresponding to the loss of the lysine protecting groups in the fragmentation process (**b** and **d**). Ion peaks were assigned manually; along with precursor ion masses, these confirmed the incorporation of each ncAA at its expected position. The mass spectrometry analysis was performed three times with similar results. **a**, MS/MS spectra confirming PheI **5** incorporation at position 40. **b**, MS/MS spectra confirming BocK **1** incorporation at position 50. **c**, MS/MS spectra confirming NmH **2** incorporation at position 136. **d**, MS/MS spectra confirming CbzK **3** incorporation at position 150.



Supplementary Figure 9

The yield of protein containing four distinct ncAAs (Quad protein yield, y axis) from translation of O-mRNAs containing four distinct quadruplet codons by O-ribosomes is positively correlated with the yield of protein from the corresponding O-mRNA without quadruplet codons (Orthogonal UTR strength, x axis). We cloned the 5' UTR sequences O1-O4 $_{strep}GFP_{His6}$ for $_{strep}\textit{GFP}_{His6}$ and $_{strep}\textit{GFP(40CTAG, 50TAGA, 136AGGA, 150AGTA)}_{His6}$ into a standardised p15A reporter. We produced proteins from $_{strep}\textit{GFP}_{His6}$ in presence of the O-ribosome and proteins from $_{strep}\textit{GFP(40TAGA, 50CTAG, 136AGGA, 150AGTA)}_{His6}$ in the presence of O-RiboQ1 and RS4_1-2/tRNA4(quad) as well as ncAAs NmH **2**, CbzK **3**, AllocK **4** and PheI **5**. We measured in vivo fluorescence. Data points represent the mean of three biological replicates ± standard deviation. Orthogonal UTR strength correlates with quadruplet incorporation yield (Pearson's $R^2 = 0.919$, P (two-tailed) = 0.041). All numerical values are given in **Supplementary Table 2**.

Supplementary Tables

All supplementary tables are provided as separate excel files.

Supplementary Table 1

Containing all ORBS calculations, 5' UTR sequences and optimised ORFs.

Supplementary Table 2

Containing all $_{strep}GFP_{His6}$ yield data including all primary data for linear regressions used for isolated yield determination by fluorescence measurements as well as all primary data for the correlation of quadruplet incorporation yield with UTR strength.

Supplementary Table 3

Containing all calculated 5' UTR sequences and ΔG_{tot} (wt ribo) used for aaRS operon assembly.

Supplementary Table 4

Annotated list of all plasmids used in this work.

Supplementary Table 5

Source data of **Supplementary Fig. 5, 6** and **7**.

1. Wang, K., Neumann, H., Peak-Chew, S.Y. & Chin, J.W. Evolved orthogonal ribosomes enhance the efficiency of synthetic genetic code expansion. *Nat. Biotechnol.* **25**, 770–777 (2007).
2. Neumann, H., Wang, K., Davis, L., Garcia-Alai, M. & Chin, J.W. Encoding multiple unnatural amino acids via evolution of a quadruplet-decoding ribosome. *Nature* **464**, 441–444 (2010).
3. Salis, H.M., Mirsky, E.A. & Voigt, C.A. Automated design of synthetic ribosome binding sites to control protein expression. *Nat. Biotechnol.* **27**, 946–950 (2009).
4. Salis, H.M. in Methods in Enzymology, Vol. 498 19–42 (Academic Press, Cambridge, MA, USA; 2011).
5. Espah Borujeni, A., Channarasappa, A.S. & Salis, H.M. Translation rate is controlled by coupled trade-offs between site accessibility, selective RNA unfolding and sliding at upstream standby sites. *Nucleic Acids Res.* **42**, 2646–2659 (2014).
6. Espah Borujeni, A. & Salis, H.M. Translation Initiation is Controlled by RNA Folding Kinetics via a Ribosome Drafting Mechanism. *J. Am. Chem. Soc.* **138**, 7016–7023 (2016).
7. Espah Borujeni, A. et al. Precise quantification of translation inhibition by mRNA structures that overlap with the ribosomal footprint in N-terminal coding sequences. *Nucleic Acids Res.* **45**, 5437–5448 (2017).