nature biotechnology

Article

https://doi.org/10.1038/s41587-024-02385-y

Efficient genetic code expansion without host genome modifications

Received: 8 November 2023

Accepted: 13 August 2024

Published online: 11 September 2024

Check for updates

Alan Costello[1,2], Alexander A. Peterson[1,2], David L. Lanster[1,2,3], Zhiyi Li[1,2,3], Gavriela D. Carver[4] & Ahmed H. Badran[1,2]

Supplementing translation with noncanonical amino acids (ncAAs) can yield protein sequences with new-to-nature functions but existing ncAA incorporation strategies suffer from low efficiency and context dependence. We uncover codon usage as a previously unrecognized contributor to efficient genetic code expansion using non-native codons. Relying only on conventional *Escherichia coli* strains with native ribosomes, we develop a plasmid-based codon compression strategy that minimizes context dependence and improves ncAA incorporation at quadruplet codons. We confirm that this strategy is compatible with all known genetic code expansion resources, which allowed us to identify 12 mutually orthogonal transfer RNA (tRNA)–synthetase pairs. Enabled by these findings, we evolved and optimized five tRNA–synthetase pairs to incorporate a broad repertoire of ncAAs at orthogonal quadruplet codons. Lastly, we extend these resources to an in vivo biosynthesis platform that can readily create >100 new-to-nature peptide macrocycles bearing up to three unique ncAAs. Our approach will accelerate innovations in multiplexed genetic code expansion and the discovery of chemically diverse biomolecules.

Genetic code expansion, where noncanonical amino acids (ncAAs) are added to the central dogma, has provided foundational tools to study and manipulate biological processes[1,2]. Pioneering labs have shown that genetic code expansion requires two components: heterologous transfer RNA (tRNA)–synthetase pairs engineered to accept researcher-defined ncAAs and 'blank' codons available for assignment to ncAAs during translation. Historically, the codon of choice has been the amber (UAG) stop and considerable effort has focused on improving UAG-dependent ncAA incorporation through strain and ribosome engineering[3–8]. While the linkage between a tRNA and the ncAA would later be addressed through the directed evolution of dedicated aminoacyl tRNA synthetases[9], the availability of unassigned codons has consistently been a key limitation[10]. Recent innovations in genome synthesis[11,12] and artificial nucleobases[13] have begun to address this fundamental issue by erasing redundant codons from the host genome or by creating de novo codons, respectively. However, genome synthesis requires a deep understanding of regulatory signatures embedded within genes and the use of artificial nucleobases relies on expertise in their synthesis and activation for use as substrates in vivo. Conversely, in vitro genetic code expansion can access chemically diversified peptides through precise control over tRNA aminoacylation and abundance[14] but can result in low protein yield and require extensive tRNA preparation. Combining the in vitro chemical diversity with the scalability and throughput of in vivo genetic code expansion could streamline the discovery of bioactive peptides and noncanonical polymers.

While many aspects of translation (for example, codon sequence, tRNA aminoacylation and ribosome components) have been investigated to address these limitations and improve ncAA incorporation outcomes[15–23], mRNA has received comparatively little attention. We set out to investigate the contribution of mRNA sequence to ncAA incorporation using quadruplet codon translation. Quadruplet-decoding tRNAs (qtRNAs) can 'read' a four-base codon complementary to an engineered four-base anticodon, resulting in a +1 frameshift and restoration of the open reading frame during translation[24]. While quadruplet codon

[1]Department of Chemistry, The Scripps Research Institute, La Jolla, CA, USA. [2]Department of Integrative Structural and Computational Biology, The Scripps Research Institute, La Jolla, CA, USA. [3]Doctoral Program in Chemical and Biological Sciences, The Scripps Research Institute, La Jolla, CA, USA. [4]Department of Molecular Biology, Princeton University, Princeton, NJ, USA. e-mail: ahbadran@scripps.edu

translation offers new codons for multiplexed ncAA incorporation, it is less efficient than triplet decoding[22]. Nonetheless, quadruplet decoding is compatible with native ribosomes because (1) qtRNAs are active in in vitro translation[25]; (2) qtRNAs can be evolved to improve their in vivo activities[18,22]; and (3) aminoacylation by engineered synthetases can occur in an anticodon-independent manner[26]. We hypothesized that two mRNA parameters may impact quadruplet codon translation. First, codon usage in natural genes can fine-tune the rate[27], fidelity[28] and efficiency[29] of translation. While amino acid[30,31] identities near UAG can affect ncAA incorporation[32,33], the impact of neighboring codons on quadruplet decoding remains unknown. Second, the unavoidable sequence overlap between triplet and quadruplet codons may sequester qtRNAs from their intended decoding events and inadvertently promote frameshift errors. In agreement with this, the most commonly used qtRNAs decode sequences that overlap with low-usage triplet codons[18,22,34]; however, no analysis of qtRNA-dependent mistranslation has been reported to date.

Here, we explored the impact of mRNA codon usage on quadruplet decoding, finding a preference for a high-usage triplet codon immediately downstream of a quadruplet codon. To improve on-target decoding and eliminate off-target activities in the same gene, we generalized this observation such that all plasmid-borne genes were recoded using a single representative codon for each of the 20 canonical amino acids (cAAs). We optimized virtually all parameters predicted to affect qtRNA-dependent decoding using this framework, including plasmid designs, copy numbers, ribosome-binding site (RBS) strengths, qtRNA sequences and synthetase identities. Our improved resources catalyze the efficient incorporation of multiple unique ncAAs into proteins without requiring host *Escherichia coli* cell alteration. We applied this approach toward the biosynthesis of peptide macrocycles encoding up to three unique ncAAs in living cells, bridging the gap between in vitro and in vivo genetic code expansion approaches. In summary, our simple recoding strategy was refined into a user-friendly 'plug-and-play' technology for expansive quadruplet codon translation.

Results

Local and remote codon usage influences quadruplet decoding

Local sequence context and codon usage can influence natural translation and UAG decoding[29,32,33,35]. To explore its impact on quadruplet decoding, we used a superfolder green fluorescent protein (sfGFP) reporter, where the permissive Y151 was replaced by a quadruplet codon[5], and varied the codon usage of five upstream and downstream residues (Fig. 1a). Using cAA-specific and ncAA-specific qtRNAs[22,34,36], we sorted cells on the basis of high versus low sfGFP expression (Supplementary Fig. 1). We repeated this analysis using a UAG codon at Y151 to ensure that our findings were specific to quadruplet decoding. Sorted clones showed a wider sfGFP distribution when decoding quadruplet codons versus UAG, suggesting that quadruplet decoding is more sensitive to local sequence contexts (Supplementary Fig. 4a,b). Next-generation sequencing (NGS) showed little change in codon preference between UAG bins (Fig. 1b and Supplementary Fig. 4c), whereas quadruplet decoding populations showed clear codon usage biases (Supplementary Fig. 4d–f). Specifically, high sfGFP bins consistently showed a strong preference for a high-usage codon immediately 3′ to the quadruplet codon (that is, +1 position; Fig. 1c–e). We also observed codon bias at other positions within our libraries, namely at the −4, −1, +2 and +3 positions with respect to the decoding site. However, these positions did not show a recurring trend for UAGA versus AGGA codons or for quadruplet versus triplet decoding. We hypothesized that these observations may, therefore, reflect context-dependent effects and we did not explore them further. We validated our NGS results by assaying the top sequences from the top 5% and top 1% pools (Fig. 1f) and confirmed amino acid incorporation using purified sfGFP from the naive library and all sorted pools using mass spectrometry (MS) (Supplementary Figs. 3 and 4). Lastly, we tested high-usage versus low-usage codon substitutions at the +1 position when decoding AGGA at other positions in sfGFP, finding that +1 recoding to high-usage codons is either beneficial or of no consequence (Fig. 1g).

Next, we speculated that qtRNAs likely do not distinguish between on-target (which corrects the translation frame) and off-target decoding (which creates de novo frameshift mutations during translation). Because quadruplet codons often include low-usage triplet codons at the first three positions[22], we hypothesized that restricting mRNAs to high-usage codons would eliminate off-target decoding and improve protein purity (Fig. 1h). We recoded our reporter genes, antibiotic resistance markers and plasmid replication proteins to use a single high-usage codon for each amino acid. We expected that this recoding may impair protein function in some cases but did not observe any issues with maintenance or growth of strains carrying recoded plasmids. Using different AGGA-decoding qtRNAs, we observed robust quadruplet decoding in our optimized circuit architecture lacking AGG triplet codons (Supplementary Fig. 5a,b), whereas reintroducing AGG in sfGFP reduced apparent decoding efficiency through presumed off-target frameshifting (Fig. 1i,j and Supplementary Fig. 5c). These results collectively implicate codon usage as an unrecognized yet important determinant of efficient on-target quadruplet decoding.

tRNA–synthetase pairs are active in recoded circuits

Introducing multiple unique ncAAs into a single protein requires mutually orthogonal tRNA–synthetase pairs[37]. To ensure that our recoding strategy can be extended to multiplexing studies, we first synthesized a *Methanosarcina barkeri* (*Mb*)-derived PylRS–$tRNA^{Pyl}_{CUA}$ using our high-usage codon set and evaluated its activity using a sensitized UAG luciferase reporter (Fig. 2a and Supplementary Fig. 6a)[22,38,39]. We optimized synthetase plasmid copy number, induction conditions and tRNA expression (Supplementary Fig. 6b–d), finding that our recoded components performed as effectively as their *E. coli* codon-optimized counterparts (Fig. 2b). Aiming to nominate the most robust components for multiplexed ncAA incorporation, we next evaluated representatives of all tRNA–synthetase pairs reported in *E. coli* to date using UAG decoding (for both cAA and ncAA incorporation)[15,18,34,39–51]. In some cases, we tested related synthetases with differing ncAA scopes, altered protein expression to improve host tolerance and optimized tRNA decoding efficiencies and specificities (Supplementary Figs. 7–9).

We assayed 25 representative synthetases alongside 32 cognate tRNAs (800 combinations) and quantified UAG decoding as a percentage of a wild-type luciferase reporter (Fig. 2c and Methods). We uncovered 69 tRNA–synthetase pairs with >20% decoding efficiency, which could be assigned to 12 mutually orthogonal clusters. Clusters 1–6 are ncAA specific and 7–12 are cAA specific (Fig. 2d–f). Cluster 1 contains TyrRS–$tRNA^{Tyr}_{UAG}$ pairs: *Af*TryRS, *Mj*TyrRS, *Ma*TyrRS and *Sc*TyrRS. Clusters 2–4 contain orthogonal PylRS–$tRNA^{Pyl}_{UAG}$ variants[41]: *Ma*PylRS, *G1*PylRS, *Rum*PylRS, *Mlum1*RS, *Mm*PylRS and *Mb*PylRS. Clusters 1–4 represent the current state of the art for mutually orthogonal tRNA–synthetase pairs in *E. coli*[34]. Cluster 5 contains *Sc*TrpRS with a new *Sc*$tRNA^{Trp(M13)}_{UAG}$ we developed (Supplementary Fig. 8). Cluster 6 contained two *Mm*SepRS–$tRNA^{Sep}_{UAG}$ variants[47,48], which required the L-serine kinase *Tk*SerK[52] and elongation factor EF-Sep[47] for phosphoserine incorporation (Supplementary Fig. 9). Clusters 7–12 correspond to six cAA-incorporating tRNA–synthetase pairs that may serve as starting points for further evolution: *Sc*PheRS[46], *Mm*SerRS[51], *In*GlnRS[45], *Sc*GlnRS[45], *Ph*LysRS[15] and *Mj*LeuRS[49] (Fig. 2f). In addition, this dataset uncovered 12 mutually orthogonal tRNA–synthetase pairs and confirmed that codon compression is compatible with cAA and ncAA incorporation.

qtRNA directed evolution improves quadruplet decoding

Multiplexed ncAA incorporation requires engineered qtRNAs that efficiently decode nonoverlapping quadruplet codons[22,34].

https://doi.org/10.1038/s41587-024-02385-y

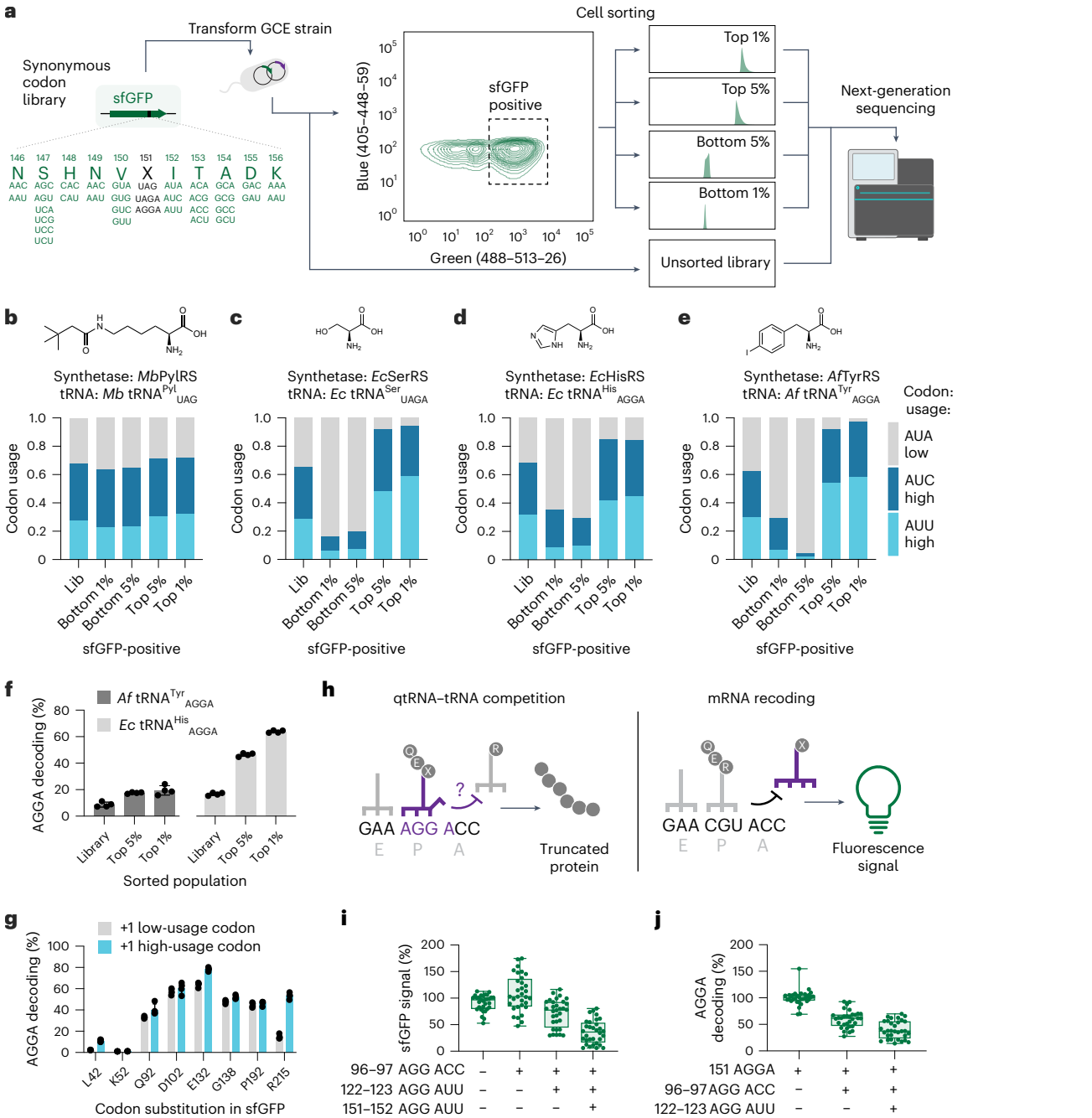


Fig. 1 | Local and remote codons influence apparent quadruplet decoding. **a**, Library design and screening approach to investigate the influence of codon usage on quadruplet decoding. Codon abundance before and after fluorescence-activated cell sorting is shown at the +1 position. I152 is encoded by three codons: AUA, AUC and AUU. AUA is a low-usage codon, while AUC and AUU are both high-usage codons in *E. coli*. **b**–**e**,The codon abundance in transcripts at the +1 (I152) position is shown for *Mb* tRNA$^{Pyl}_{UAG}$ with *Mb*PylRS incorporating N_6-(*tert*-butoxycarbonyl)-lysine (BocK) (**b**), *Ec* tRNA$^{Ser(evo2)}_{UAGA}$ incorporating serine (**c**), *Ec* tRNA$^{His}_{AGGA}$ incorporating histidine (**d**) and *Af* tRNA$^{Tyr(A01)}_{AGGA}$ with *Af*TyrRS(G5) incorporating *para*-iodophenylalanine (pIF) (**e**). **f**, Variants from the top 1% and 5% populations showed greater AGGA decoding than the library average using *Af*TyrRS(G5)-*Af* tRNA$^{Tyr(A01)}_{AGGA}$ ($n = 4$ biological replicates; error bars show the s.d.) and *Ec* tRNA$^{His}_{AGGA}$ ($n = 4$ biological replicates; error bars show the s.d.). **g**, Testing the influence of low-usage versus high-usage codons at the +1 position using multiple positions in sfGFP and *Ec* tRNA$^{His}_{AGGA}$. The codon changed to high or low usage is indicated on the *x* axis ($n = 4$ biological replicates; error bars show the s.d.). **h**, Schematic representation of putative competition between a qtRNA and tRNA whose codons overlap at nucleotides 1–3. Recoding the mRNA to remove triplet codons that overlap with quadruplet codons eliminates this competition at unintended sites. **i**, Using an all-triplet codon sfGFP reporter, introduction of in-frame AGG codons at positions R96, R122 and Y151 results in mistranslation by eight different AGGA-decoding qtRNAs (GltT, GlyU, MetT, ArgQ, ArgW, SerT, SerW and HisR) and reduced apparent GFP translation ($n = 4$ biological replicates). Box-and-whisker plot: minima and maxima are defined by the error bar, the center is the median and bounds of the box represent the upper and lower quartiles (75th and 25th percentiles, respectively). **j**, Apparent AGGA decoding at Y151 is similarly reduced when AGG codons are introduced in frame at positions R96 and R122, resulting in reduced apparent GFP translation with eight different AGGA-decoding qtRNAs (GltT, GlyU, MetT, ArgQ, ArgW, SerT, SerW and HisR) ($n = 4$ biological replicates). Box-and-whisker plot: minima and maxima are defined by the error bar, the center is the median and bounds of the box represent the upper and lower quartiles (75th and 25th percentiles, respectively).

https://doi.org/10.1038/s41587-024-02385-y

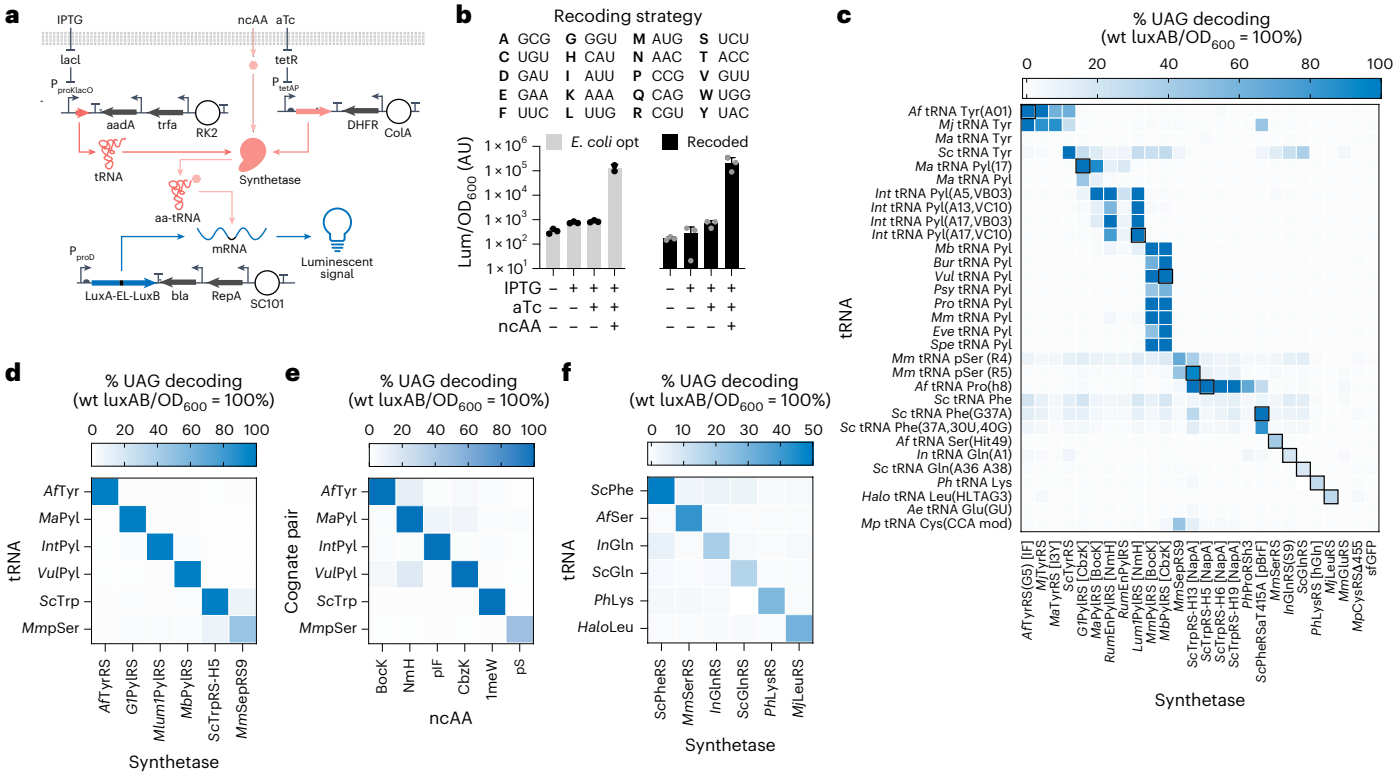


Fig. 2 | Twelve mutually orthogonal tRNA–synthetase pairs and validation in recoded circuits. a, Schematic representation of the tRNA–synthetase benchmarking assay. All tRNAs are expressed from the IPTG-controlled promoter (P_{proK}–lacO) and synthetase proteins are produced from the aTc-responsive promoter (P_{tetA}). The LuxAB reporter genes are separated by an engineered linker (EL) that contains an amber (UAG) codon. Origins of replication (white circles) and antibiotic resistance genes (gray arrows) are shown. **b**, All plasmid-encoded protein-coding genes were recoded to reflect a single high-usage triplet codon for each cAA as shown. The *Mb*PylRS–*Mb* $tRNA^{Pyl}_{UAG}$ was tested using this recoding strategy and compared to conventional *E. coli* codon-optimized genes. Luminescence values are normalized to OD_{600} and raw arbitrary units (AU) are shown ($n = 3$ biological replicates; error bars show the s.d.). **c**, Heterologous synthetase (*x* axis) and tRNA (*y* axis) genes were combinatorially tested using the UAG LuxAB reporter. All ncAAs shown in parentheses on the *x* axis were included in the growth medium at 1 mM each, whereas *Mm*SepRS was tested alongside EF-Sep and *Tk*SerK. This analysis yielded a priority list of 12 mutually orthogonal tRNA–synthetase pairs (black boxes). Reporter activity is normalized to wild-type (wt) LuxAB expression in all cases ($n = 4$ biological replicates). **d**, Validation of mutual orthogonality of the six identified ncAA-specific tRNA–synthetase pairs from the primary screen ($n = 4$ biological replicates). **e**, Validation of low crosstalk between ncAAs used by all six tRNA–synthetase pairs: BocK, *N*-methylhistidine (NmH), pIF, N_6-((benzyloxy)carbonyl)-lysine (CbzK), 1-methyltryptophan (1meW) and phosphoserine (pS). All ncAAs were included at 1 mM each, whereas *Mm*SepRS was tested alongside EF-Sep and *Tk*SerK ($n = 4$ biological replicates). **f**, Validation of mutual orthogonality of the six identified cAA-specific tRNA–synthetase pairs from the primary screen ($n = 4$ biological replicates). Wherever reported, the UAG decoding percentage was calculated by normalizing OD-corrected luminescence values to a wt LuxAB control.

Having established five mutually orthogonal tRNA–synthetase pairs for UAG decoding and ncAA incorporation, we next optimized their quadruplet-decoding activities. The tRNA–synthetase pairs *G1*PylRS, *Af*TyrRS, *Lum1*RS and *Mm*PylRS have been engineered to decode quadruplet codons[34], although their decoding efficiencies ranged from approximately 0% to 20% in our assays (Extended Data Fig. 1a). Reducing plasmid burden by condensing all components from three to two plasmids improved efficiencies to 8–100% depending on the tRNA–synthetase pair, albeit with high background activity in some cases (Extended Data Fig. 1b). These poor activities may restrict future multiplexing studies. Whereas the above ncAA qtRNAs were created by swapping anticodon stem-loop segments from established variants[53], we showed that directed evolution can improve qtRNA decoding efficiencies[22].

We used established chloramphenicol acyltransferase (CAT)-positive selections and barnase negative selections[54] to improve qtRNA activities (Fig. 3a). We systematically targeted the qtRNA anticodon stem loop, acceptor stem, Ψ-arm and/or D-arm with site saturation mutagenesis and subjected the libraries to iterative positive–negative selection campaigns. These efforts improved the qtRNA–synthetase pairs described above to decode orthogonal quadruplet codons with up to 100% efficiency in a two-plasmid system: *G1*PylRS (3.4-fold improvement in dynamic range using AUAG), *Af*TyrRS (1.6-fold; CUAG), *Lum1*RS (5.3-fold; AGGA) and *Mm*PylRS (9.8-fold; UAGA) (Fig. 3b–e, Extended Data Figs. 2–5 and Supplementary Tables 1–4). Seeking to supplement our ncAA repertoire with tryptophan analogs, we further developed CGGA as a new quadruplet codon for the *Sc*TrpRS–$tRNA^{Trp(M13)}$ pair (Extended Data Fig. 6) and used directed evolution to discover *Sc* $tRNA^{Trp}_{CGGA}$ variants with up to a 38.1-fold dynamic range in the three-plasmid circuit and fivefold dynamic range in the two-plasmid system (Fig. 3f,g and Supplementary Table 5). We confirmed these enhancements in quadruplet decoding between the starting and evolved qtRNAs through western blot (Supplementary Fig. 10) and MS analyses (Supplementary Fig. 11). Cumulatively, these campaigns yielded five ncAA-incorporating, mutually orthogonal qtRNA–synthetase pairs that effectively function in living cells.

Optimizing synthetase ncAA scope and qtRNA expression

Multiplex ncAA incorporation would benefit from synthetases with a broad substrate scope to enable the introduction of a large repertoire of ncAAs using only a handful of mutually orthogonal synthetase–tRNA pairs. Because synthetases are often reported to incorporate only a single ncAA, we screened active site variants and homologs

https://doi.org/10.1038/s41587-024-02385-y

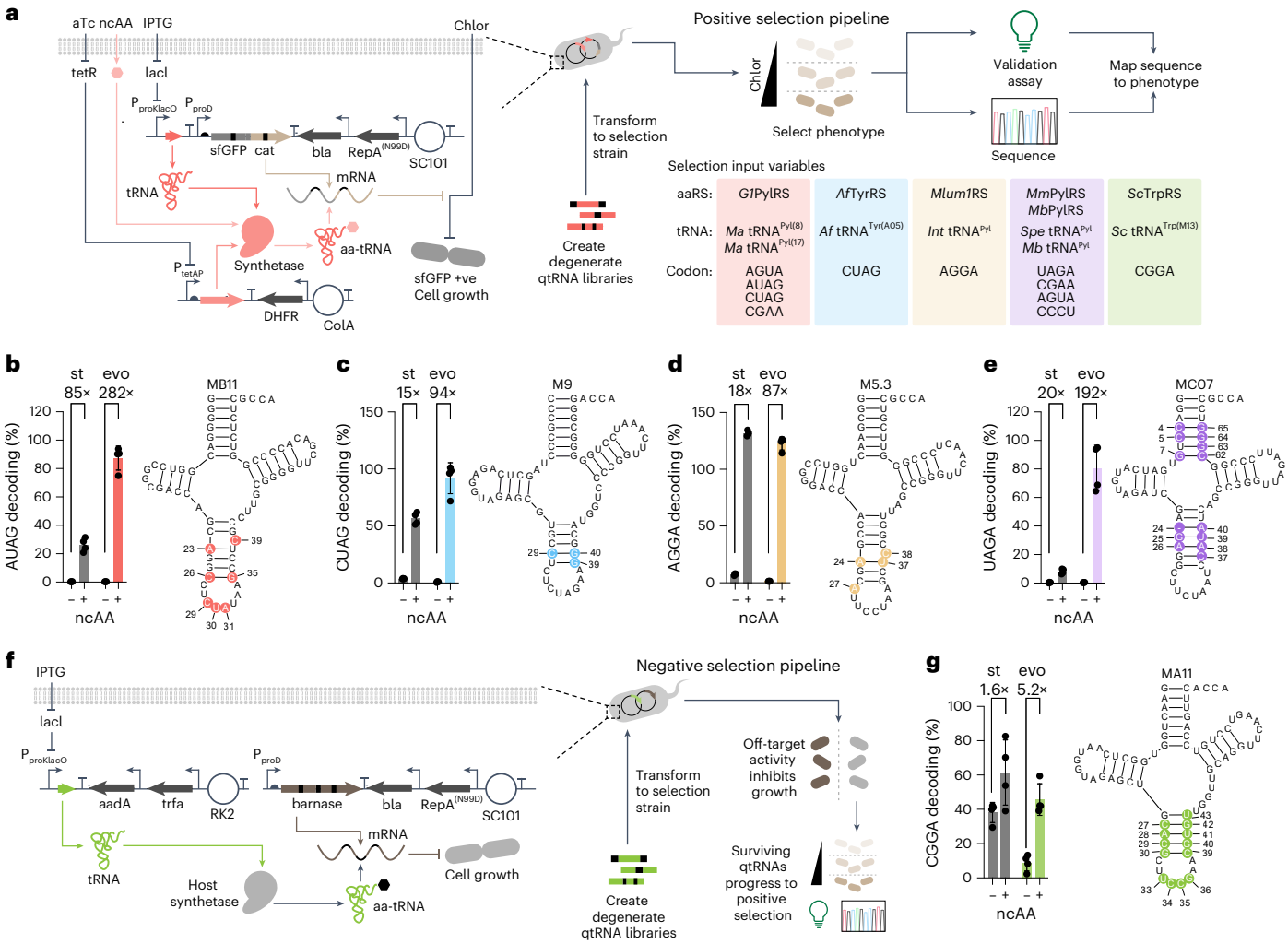


Fig. 3 | Directed evolution of highly efficient qtRNAs for ncAA incorporation. **a**, Schematic representation of the selection strategy to evolve improved qtRNA activities in the recoded framework: qtRNAs were subjected to positive selection by decoding quadruplet codons in an *sfGFP–cat* fusion gene. Starting (st) and evolved (evo) qtRNA genes were validated by decoding their cognate quadruplet codon at position Y151 in sfGFP. In each case, quadruplet decoding is reported for the final evolved variant and the discovered mutations are mapped onto the starting qtRNA scaffold. **b**, *Ma* $tRNA^{Pyl}_{AUAG}$ evolved to mutant (MB11), aminoacylated by *G1*PylRS ($n = 4$ biological replicates, error bars show the s.d.). **c**, *Af* $tRNA^{Tyr}_{CUAG}$ evolved to mutant (M9), aminoacylated by *Af*TyrRS ($n = 4$ biological replicates; error bars show the s.d.). **d**, *Int* $tRNA^{Pyl}_{AGGA}$ evolved to mutant (M5.3), aminoacylated by *Mlum1*PylRS ($n = 4$ biological replicates; error bars show the s.d.). **e**, *Spe* $tRNA^{Pyl}_{UAGA}$ evolved to mutant (MC07), aminoacylated by *Mm*PylRS ($n = 4$ biological replicates; error bars show the s.d.). **f**, In the case of *Sc* $tRNA^{Trp}_{CGGA}$, a negative selection was first implemented to eliminate ncAA-independent aminoacylation by host synthetases and then subjected to positive selection. **g**, *Sc* $tRNA^{Trp}_{CGGA}$ evolved to mutant (MA11), aminoacylated by *Sc*TrpRS ($n = 4$ biological replicates; error bars show the s.d.). Wherever reported, the quadruplet decoding percentage was calculated by normalizing OD-corrected fluorescence values to a wild-type sfGFP control.

of our prioritized tRNA–synthetase pairs for broadened ncAA scope beyond their cognate substrates (Supplementary Figs. 12–16 and Supplementary Tables 6–11). We prioritized synthetases that can access at least three ncAAs with >10% efficiency. These analyses uncovered variants of our five prioritized synthetases that together provide access to 47 unique ncAAs with diverse physicochemical properties (Fig. 4a).

Briefly, *Mj*TyrRS (Y32I, L65I, Q109M, D158G, L162V and V164G)[18] was developed to incorporate *para*-(2-tetrazole)phenylalanine and fortuitously uses diverse *para*-substituted and *para*/*meta*-disubstituted phenylalanine derivatives (Fig. 4b). We note that this *Mj*TyrRS also carried mutations that improve tolerance to qtRNA anticodon changes (Y230K, C231K, P232K, H283Q and D286S)[55]. *G1*PylRS (L124S, Y125F, Y204W, A221S and W237Y)[56] was developed to incorporate cyanopyridylalanine but is capable of charging a large repertoire of related *meta*-substituted and *para*-substituted phenylalanine and tyrosine derivatives (Fig. 4c). *Mlum1*RS (L125I, Y126F, M129G, V168F and Y205F) was grafted from a reported *Mb*PylRS[57] that uses histidine derivatives, which allows Lum1RS to charge derivatives of histidine, thiophene and naphthalene, as well as *meta*-substituted or *ortho*-substituted phenylalanines (Fig. 4d). *Mm*PylRS (Y271A and Y349F)[58] was developed to incorporate lysine analogs, which we reproduce here (Fig. 4e). *Sc*TrpRS (Y106V, E141P, T233C and I253C)[43] was evolved to use 6-methyltryptophan and we show that it can also tolerate related analogs (Fig. 4f).

Lastly, we optimized qtRNA production from a multicistronic cassette because this can improve multiplexed quadruplet decoding efficiency[22,34]. We used a reported 4-qtRNA cassette and mined the *E. coli* genome for inter-tRNA sequences (20–60 nucleotides) to use as spacers for our fifth qtRNA (Fig. 4g and Supplementary Table 2). Multiple spacers resulted in expression of the fifth qtRNA and robust quadruplet decoding (Extended Data Fig. 7). The best qtRNA expression cassette resulted in excellent quadruplet decoding for all five optimized qtRNAs in an ncAA-dependent manner (Fig. 4h). Together, these chosen synthetases and qtRNA expression constructs enable

https://doi.org/10.1038/s41587-024-02385-y

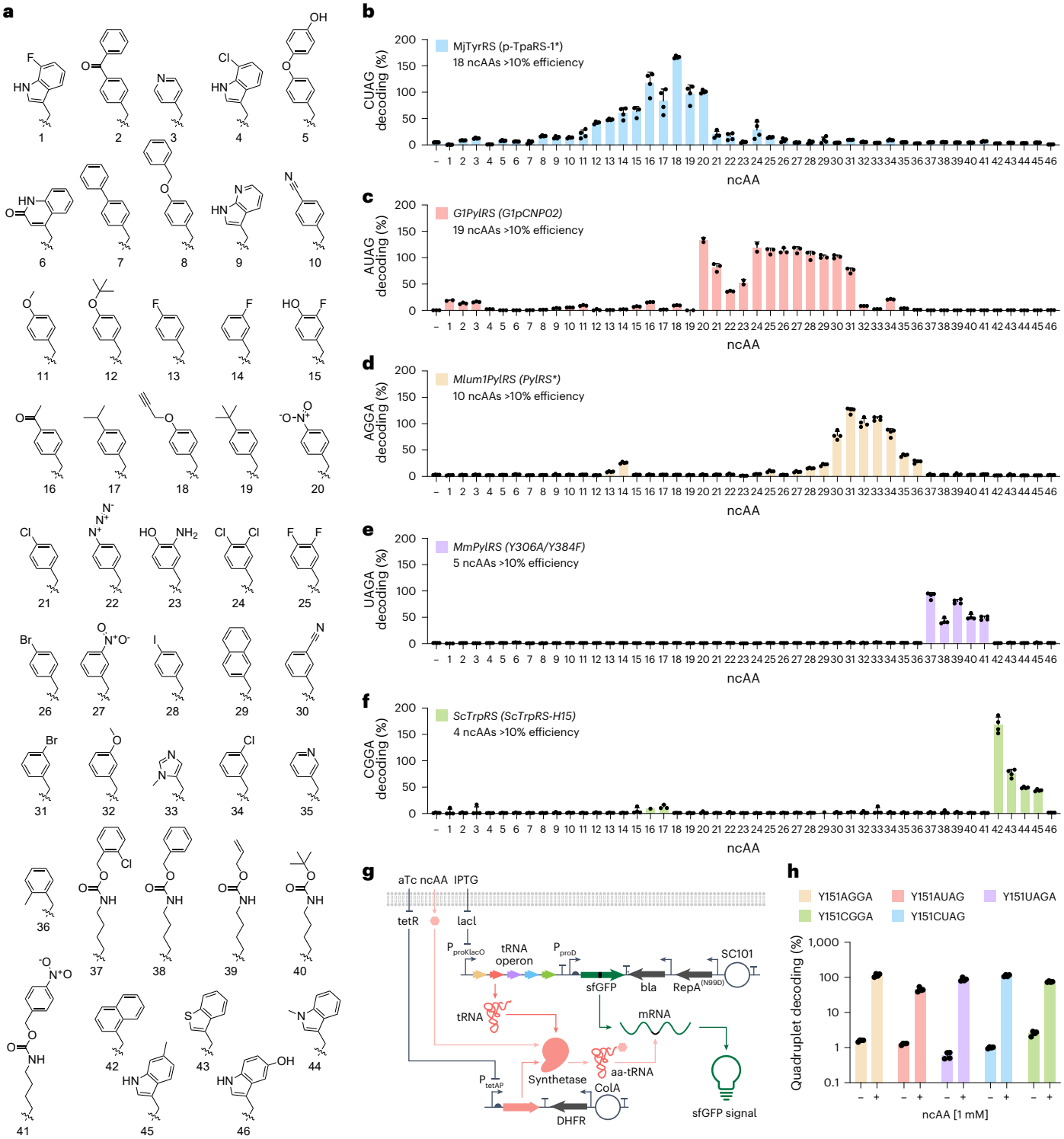


Fig. 4 | Validation of synthetases with broad, nonoverlapping ncAA substrate scope. a, Prioritized ncAAs for incorporation at quadruplet codons using optimized qtRNA–synthetase pairs. All shown ncAAs showed detectable signal using an sfGFP reporter using at least one qtRNA–synthetase pair. Detailed lists of tested synthetase variants (Supplementary Tables 6–10) and ncAAs (Supplementary Table 11) are provided. **b**–**f**, The ncAA substrate scope is shown for the representative synthetases: *Mj*TyrRS (p-TpaRS-1*) (**b**; $n = 4$ biological replicates; error bars show the s.d.); *G1*PylRS (G1pCNP02) (**c**; $n = 4$ biological replicates; error bars show the s.d.); *Mlum1*PylRS (PylHRS*) (**d**; $n = 4$ biological replicates; error bars show the s.d.); *Mm*PylRS (Y306A;N346A;C348A;Y384F), (**e**; $n = 4$ biological replicates; error bars show the s.d.); *Sc*TrpRS (ScTrpRS-H15) (**f**; $n = 4$ biological replicates; error bars show the s.d.). **g**, Schematic representation of a multicistronic qtRNA operon and testing using dedicated synthetases by decoding quadruplet codons at Y151 in sfGFP. **h**, Decoding efficiency of sfGFP Y151 mutated to a quadruplet codon for each evolved qtRNA encoded on an optimized multicistronic operon. Decoding was evaluated in the absence (−) and presence (+) of a cognate ncAA at 1 mM. Wherever reported, the quadruplet decoding percentage was calculated by normalizing OD-corrected fluorescence values to a wt sfGFP control ($n = 4$ biological replicates; error bars show the s.d.).

efficient decoding of quadruplet codons and incorporation of diverse ncAAs in conventional *E. coli* cells.

Biosynthesis of new-to-nature macrocycles with single ncAAs

Strategies to introduce ncAAs into peptide macrocycles have yielded natural product-inspired molecules[59,60] and probes of protein function[14,61]. Advances in in vitro tRNA aminoacylation and cell-free translation have catalyzed the discovery of chemically diverse bioactive macrocycles using mRNA display, usually enriched through binding to purified proteins[62,63]. While recent advances in genome rewriting have enabled the biosynthesis of macrocycles encoding two ncAAs or hydroxy acids, these methods are restricted to a single engineered bacterium and require in vitro cyclization of a cell-generated thioester intermediate[59,64]. Efforts to create chemically complex macrocycle libraries in vivo remain limited by poor ncAA incorporation and codon decoding efficiencies. Addressing these limitations would enable streamlined macrocycle discovery in vivo by increasing the chemical repertoire available to protein translation and potentially capitalizing on directed evolution paradigms to optimize attributes beyond binding.

We explored an intracellular biosynthesis strategy to generate macrocyclic peptides that encode diverse, researcher-assigned ncAAs. We adapted a reported split-intein circular ligation of peptides and proteins (SICLOPPS) system using the *Nostoc punctiforme* intein (*Npu*)[65] engineered to create cyclo-CLLFVY[66] and combined it with our optimized system for ncAA incorporation (Fig. 5a). We first confirmed that active *Npu* generated cyclo-CLLFVY with robust signal, whereas an active site mutant (C1A)[67] showed no detectable macrocycle production by liquid chromatography (LC)–MS (Fig. 5b). *Npu* plasmid copy number and cyclic peptide sequence both influenced product yield (Extended Data Fig. 8), in line with prior observations[68]. Next, we extended this approach to ncAA incorporation using quadruplet decoding in *Npu*-derived macrocycles. We co-transformed strains with each of the five optimized qtRNA–synthetase pairs and *Npu* genes encoding quadruplet codons at variable macrocycle positions, finding macrocycle products with robust ion counts comparable to cyclo-CLLFVY (Fig. 5c–g and Supplementary Data 1). In all cases, we observed the correct primary mass at M + 1. In these assays, incorporation of 3-bromophenylalanine (3BrF) was a key validation step because bromine has a unique isotopic pattern (M and M + 3 in roughly equal abundance) and is otherwise poorly abundant in *E. coli*. We observed this isotopic pattern for a 3BrF-containing macrocycle, providing confidence in the correct assignment of the target macrocycle from a heterogenous sample (Extended Data Fig. 9).

We note that apparent variability in macrocycle extracted ion chromatograms (XICs) may be attributed to splicing, ionization and/or protonation differences and cannot be directly compared. To provide insight into yield and corroborate our peak assignment, we compared our LC–MS profiles to authentic standards generated using solid-phase peptide synthesis (SPPS), which showed identical XIC retention times and confirmed correct peak assignment. By serially titrating each synthetic macrocycle, we could estimate cellular yield of select macrocycles. Across the five synthetase pairs, we observe up to 2 µM macrocycle yields following extraction (Extended Data Fig. 10 and Supplementary Data 4). In total, we generated 50 macrocycles encoding a single unique ncAA each, highlighting the scope of our optimized qtRNA–synthetase pairs.

New-to-nature macrocycles with up to three unique ncAAs

Prior work by our lab and others has shown that multiplexing quadruplet decoding events can often lead to poor overall protein production[22,34]. Bolstered by robust decoding efficiencies and broad ncAA scope in model macrocycles, we tested multiplexed incorporation of distinct ncAAs through quadruplet codon decoding. To our knowledge, no strategy has demonstrated the cellular biosynthesis of macrocycles with more than two distinct ncAAs or hydroxy acids. We first explored the incorporation of a single ncAA multiple times within the same macrocycle, finding that our approach resulted in excellent production of macrocycles containing up to three identical ncAAs (Fig. 6a,b and Supplementary Data 5).

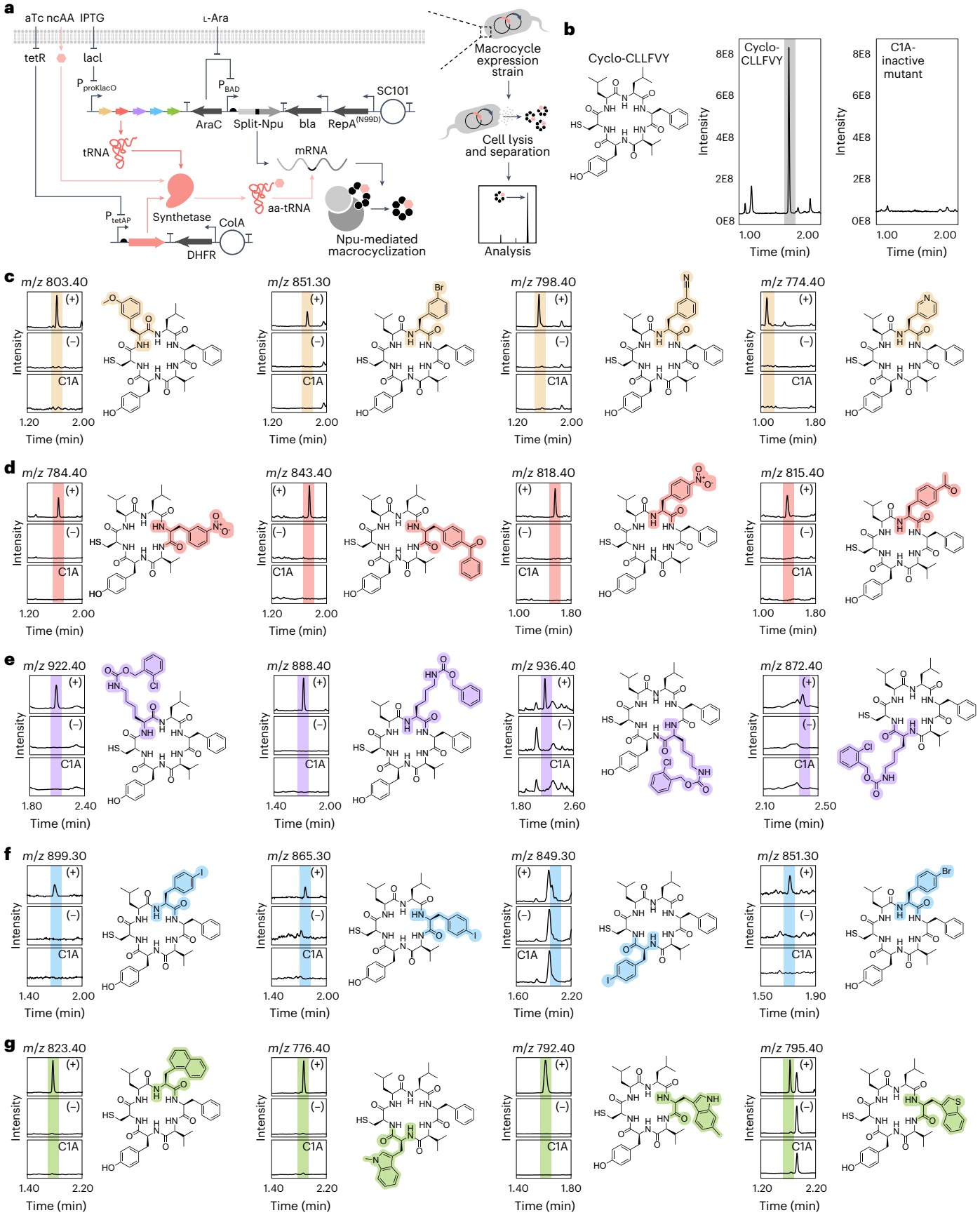
We then designed and validated genetic constructs to coexpress two or three of our prioritized synthetases (*Mlum1*RS + *Mm*PylRS, *Mlum1*RS + *Sc*TrpRS or *Mlum1*RS + *Mm*PylRS + *Sc*TrpRS; Fig. 6a and Supplementary Fig. 17). We substituted codons in the model macrocycle cyclo-CLLFVY with combinations of UAGA, CGGA and/or AGGA and monitored macrocycle biosynthesis by LC–MS. Using these components, we generated 72 unique macrocycles that encoded two or three ncAAs without extensive protein or macrocycle purification (Fig. 6b–d, Supplementary Fig. 18 and Supplementary Data 5). While poor processivity in decoding adjacent quadruplet codons can reduce protein yield[22,34], our evolved qtRNAs show robust processivity as evidenced by the biosynthesis of 17 macrocycles carrying three consecutive ncAAs. In summary, we showcase optimized resources for genetic code expansion capable of generating new-to-nature peptide macrocycles containing up to three distinct ncAA residues.

Discussion

To realize the promise of genetic code expansion, our lab and others recently developed new biotechnologies to improve qtRNAs, synthetases and the ribosome in high throughput[22,36,38,69]. Strategies that enhance the incorporation of diverse ncAAs can catalyze innovations in protein design and bioengineering that go beyond the constraints of the natural proteinogenic amino acids. In this report, we investigated how an underexplored parameter—mRNA codon composition—can influence this process. We found that local and global mRNA codon usage can have a notable impact on quadruplet decoding and ncAA incorporation efficiency. We develop new genetic circuits that obviate these issues, improve qtRNA decoding and expression and optimize synthetase dosing and ncAA scope. Together, our optimized resources were distilled into a streamlined plasmid series (Supplementary Table 14) that can incorporate up to three ncAAs across various protein contexts.

Fig. 5 | Intracellular biosynthesis of ncAA-encoding peptide macrocycles through optimized quadruplet decoding. a, Schematic representation of the macrocycle biosynthesis pipeline. The split *Npu* genes are expressed from an arabinose-controlled pBAD promoter, whereas the qtRNAs and synthetase are expressed from IPTG-inducible and aTc-inducible promoters, respectively. Desired macrocycle sequences (with or without quadruplet codons) are incorporated between the split *Npu* genes, resulting in splicing and macrocyclization after translation. Following expression, cells were lysed and organic soluble components were isolated for analysis by LC–MS. **b**, The model macrocycle cyclo-CLLFVY is robustly biosynthesized by *Npu*, whereas the C1A active site substitution ablates macrocycle generation. The peak corresponding to cyclo-CLLFVY is highlighted. This sequence was used as a template for all subsequent ncAA incorporation studies. **c**–**g**, Individual quadruplet codons were tested at positions 2–6of cyclo-CLLFVY (cysteine is required for intein-mediated peptide macrocyclization) and combined with the cognate synthetase plasmids to explore scope and positional tolerance for ncAA incorporation. Examples of XICs at the anticipated mass +H are shown (50 unique macrocycles found in Supplementary Information). In each case, XICs for the indicated mass are shown for the following conditions: +ncAA (+), −ncAA (−) and C1A inactivated intein + ncAA (C1A). Color highlights for the ncAAs indicate the used qtRNA–synthetase pairs consistent with Fig. 4: AGGA decoding in yellow by *Int* $tRNA^{Pyl}_{AGGA}$ M5.3 and *Mlum1*PylRS (PylHRS*) (**c**); AUAG decoding in red by *Ma* $tRNA^{Pyl}_{AUAG}$ MB11 and *GI*PylRS (G1pCNP02) (**d**); UAGA decoding in purple by *Spe* $tRNA^{Pyl}_{UAGA}$ MC07 and *Mm*PylRS (Y306A;N346A;C348A;Y384F) (**e**); CUAG decoding in blue by *Af* $tRNA^{Tyr}_{CUAG}$ M9 and *Mj*TyrRS (p-TpaRS-1*) (**f**); CGGA decoding in green by *Sc* $tRNA^{Trp}_{CGGA}$ MA11 and *Sc*TrpRS (ScTrpRS-H15) (**g**).

https://doi.org/10.1038/s41587-024-02385-y



https://doi.org/10.1038/s41587-024-02385-y

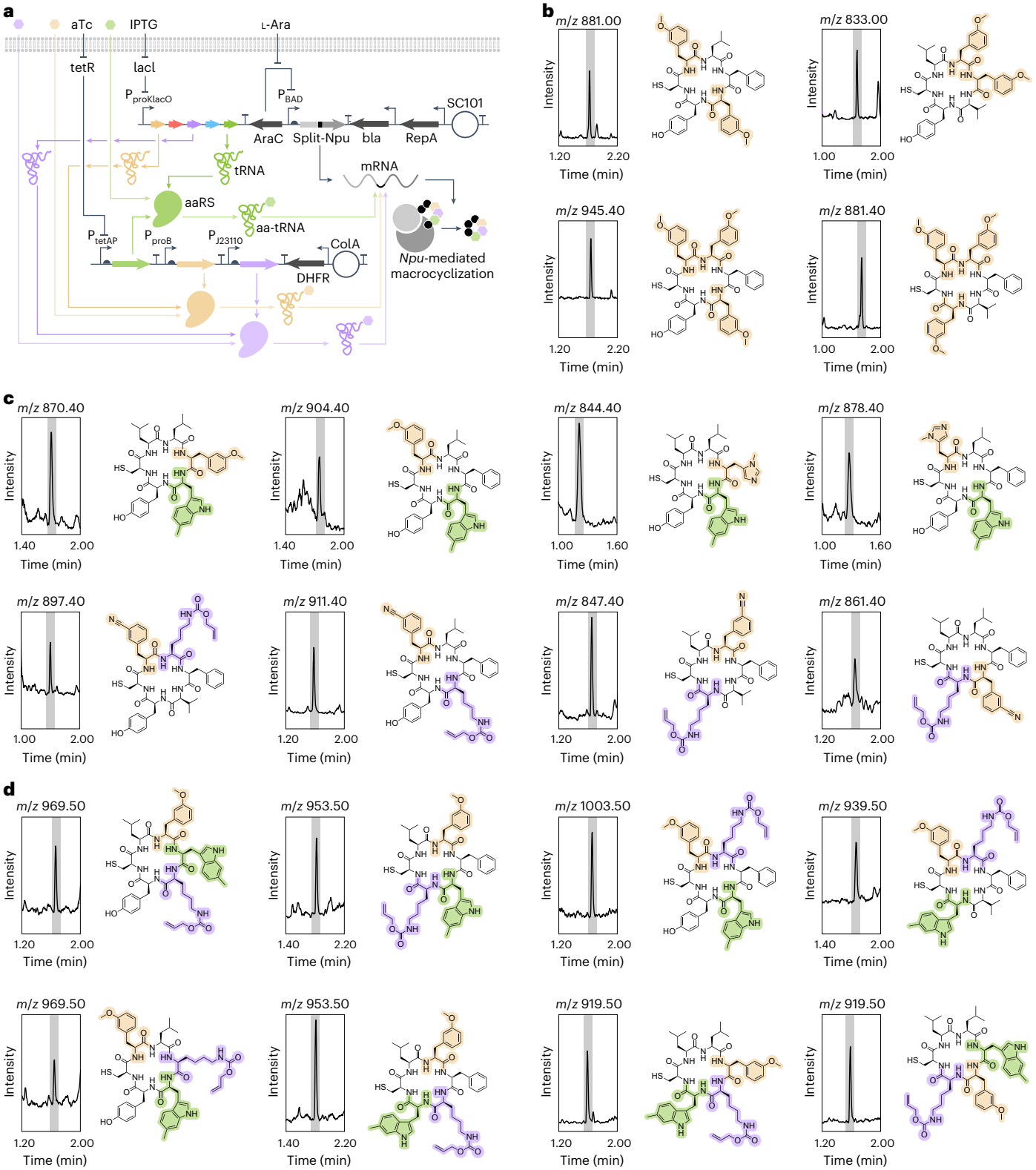


Fig. 6 | Biosynthesis of peptide macrocycles encoding multiple ncAAs. **a**, Schematic representation of the macrocycle biosynthesis pipeline for multiplexed ncAA incorporation. The inducible promoters and extraction strategy are identical to Fig. 5. **b**, Demonstration of processive ncAA incorporation into the model macrocycle cyclo-CLLFVY using 3-methoxy-L-phenylalanine. XICs at the anticipated mass +H are shown. **c**, Incorporation of two unique ncAAs in cyclo-CLLFVY using combinations of AGGA + CGGA or AGGA + UAGA codons. XICs at the anticipated mass +H are shown. In total, 36 macrocycles encoding two unique ncAAs were generated through quadruplet decoding (Supplementary Information). **d**, Incorporation of three unique ncAAs in cyclo-CLLFVY using combinations of AGGA + CGGA + UAGA codons. XICs at the anticipated mass +H are shown. In total, 17 macrocycles encoding three unique ncAA were generated through quadruplet decoding (Supplementary Information). High-resolution MS analysis for all samples is detailed in Supplementary Table 13.

We note, however, that our codon compression strategy assigns only a single codon for each of the 20 cAAs, which may be undesirable in some cases as this can affect protein folding or overall yield. Nonetheless, we found that all tested recoded proteins performed nearly identically to their *E. coli* codon-optimized counterparts, suggesting that many proteins may tolerate this approach. In cases where our focused codon compression strategy results in deleterious outcomes, intermediate versions of triplet codon compression may be required. For example, if implementing a single AGGA decoding event in a gene of interest, one could remove AGN triplet codons from the transcript, as well as alter the +1 position codon to a high-usage codon. In the context of multiplexed quadruplet codon decoding, more extensive codon compression may still be required. Furthermore, we speculate that future multiplexed ncAA incorporation efforts would benefit tremendously from the development of low-burden strategies for synthetase coexpression, which we found can have an unexpectedly large role in gene circuit behavior (Supplementary Fig. 17).

In summary, our approach here aims to combine the codon flexibility and chemical scope of in vitro macrocycle biosynthesis with the scalability and throughput of in vivo genetic code expansion. Accordingly, it provides key strengths beyond state-of-the-art methods. First, the chemical breadth provided by five mutually orthogonal qtRNA–synthetase pairs allows access to diverse macrocycles using a single DNA sequence. Future macrocycle discovery efforts may, therefore, benefit from smaller DNA libraries that can be easily elaborated using unique ncAA combinations, overcoming transformation efficiency bottlenecks that often hinder directed evolution campaigns. Second, our strategy does not require any host genome modifications to enable multiplexed ncAA incorporation, which will allow researchers to explore biosynthetic strategies using ncAAs in their favored *E. coli* strains. Third, our combined macrocycle biosynthesis and ncAA incorporation strategy is fully genetically encodable and does not require purification of any intermediates from cells to complete peptide cyclization. Our approach can, therefore, leverage design–build–test–learn cycles to accelerate small-molecule discovery through macrocycle-dependent selections. Taken together, we envision that these favorable properties will catalyze new innovations using in vivo genetic code expansion to build chemically diverse polymers in a programmable manner.

Online content

Any methods, additional references, Nature Portfolio reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at https://doi.org/10.1038/s41587-024-02385-y.

References

1. Chin, J. W. Expanding and reprogramming the genetic code of cells and animals. *Annu. Rev. Biochem.* **83**, 379–408 (2014).
2. Noren, C. J., Anthony-Cahill, S. J., Griffith, M. C. & Schultz, P. G. A general method for site-specific incorporation of unnatural amino acids into proteins. *Science* **244**, 182–188 (1989).
3. Wang, L., Brock, A., Herberich, B. & Schultz, P. G. Expanding the genetic code of *Escherichia coli*. *Science* **292**, 498–500 (2001).
4. Chin, J. W., Martin, A. B., King, D. S., Wang, L. & Schultz, P. G. Addition of a photocrosslinking amino acid to the genetic code of *Escherichia coli*. *Proc. Natl Acad. Sci. USA* **99**, 11020–11024 (2002).
5. Young, T. S., Ahmad, I., Yin, J. A. & Schultz, P. G. An enhanced system for unnatural amino acid mutagenesis in *E. coli*. *J. Mol. Biol.* **395**, 361–374 (2010).
6. Johnson, D. B. et al. *RF1* knockout allows ribosomal incorporation of unnatural amino acids at multiple sites. *Nat. Chem. Biol.* **7**, 779–786 (2011).
7. Chatterjee, A., Sun, S. B., Furman, J. L., Xiao, H. & Schultz, P. G. A versatile platform for single- and multiple-unnatural amino acid mutagenesis in *Escherichia coli*. *Biochemistry* **52**, 1828–1837 (2013).
8. Chin, J. W. Expanding and reprogramming the genetic code. *Nature* **550**, 53–60 (2017).
9. Dumas, A., Lercher, L., Spicer, C. D. & Davis, B. G. Designing logical codon reassignment—expanding the chemistry in biology. *Chem. Sci.* **6**, 50–69 (2015).
10. Ostrov, N. et al. Synthetic genomes with altered genetic codes. *Curr. Opin. Syst. Biol.* **24**, 32–40 (2020).
11. Fredens, J. et al. Total synthesis of *Escherichia coli* with a recoded genome. *Nature* **569**, 514–518 (2019).
12. Zurcher, J. F. et al. Continuous synthesis of *E. coli* genome sections and Mb-scale human DNA assembly. *Nature* **619**, 555–562 (2023).
13. Zhang, Y. et al. A semi-synthetic organism that stores and retrieves increased genetic information. *Nature* **551**, 644–647 (2017).
14. Goto, Y. & Suga, H. The RaPID platform for the discovery of pseudo-natural macrocyclic peptides. *Acc. Chem. Res.* **54**, 3604–3617 (2021).
15. Anderson, J. C. et al. An expanded genetic code with a functional quadruplet codon. *Proc. Natl Acad. Sci. USA* **101**, 7566–7571 (2004).
16. Rackham, O. & Chin, J. W. A network of orthogonal ribosome·mRNA pairs. *Nat. Chem. Biol.* **1**, 159–166 (2005).
17. Ohtsuki, T., Yamamoto, H., Doi, Y. & Sisido, M. Use of EF-Tu mutants for determining and improving aminoacylation efficiency and for purifying aminoacyl tRNAs with non-natural amino acids. *J. Biochem.* **148**, 239–246 (2010).
18. 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).
19. Agarwal, D., Kamath, D., Gregory, S. T., O'Connor, M. & Gourse, R. L. Modulation of decoding fidelity by ribosomal proteins S4 and S5. *J. Bacteriol.* **197**, 1017–1025 (2015).
20. Schmied, W. H. et al. Controlling orthogonal ribosome subunit interactions enables evolution of new function. *Nature* **564**, 444–448 (2018).
21. Aleksashin, N. A. et al. A fully orthogonal system for protein synthesis in bacterial cells. *Nat. Commun.* **11**, 1858 (2020).
22. Debenedictis, E. A., Carver, G. D., Chung, C. Z., Söll, D. & Badran, A. H. Multiplex suppression of four quadruplet codons via tRNA directed evolution. *Nat. Commun.* **12**, 5706 (2021).
23. Kim, D. S. et al. Three-dimensional structure-guided evolution of a ribosome with tethered subunits. *Nat. Chem. Biol.* **18**, 990–998 (2022).
24. Gamper, H., Masuda, I. & Hou, Y. M. Genome expansion by tRNA +1 frameshifting at quadruplet codons. *J. Mol. Biol.* **434**, 167440 (2022).
25. Hohsaka, T., Ashizuka, Y., Taira, H., Murakami, H. & Sisido, M. Incorporation of nonnatural amino acids into proteins by using various four-base codons in an *Escherichia coli* in vitro translation system. *Biochemistry* **40**, 11060–11064 (2001).
26. Chatterjee, A., Xiao, H. & Schultz, P. G. Evolution of multiple, mutually orthogonal prolyl-tRNA synthetase/tRNA pairs for unnatural amino acid mutagenesis in *Escherichia coli*. *Proc. Natl Acad. Sci. USA* **109**, 14841–14846 (2012).
27. Tuller, T. et al. An evolutionarily conserved mechanism for controlling the efficiency of protein translation. *Cell* **141**, 344–354 (2010).
28. Hooper, S. D. & Berg, O. G. Gradients in nucleotide and codon usage along *Escherichia coli* genes. *Nucleic Acids Res.* **28**, 3517–3523 (2000).

29. Gamble, C. E., Brule, C. E., Dean, K. M., Fields, S. & Grayhack, E. J. Adjacent codons act in concert to modulate translation efficiency in yeast. *Cell* **166**, 679–690 (2016).
30. Schinn, S. M. et al. Rapid in vitro screening for the location-dependent effects of unnatural amino acids on protein expression and activity. *Biotechnol. Bioeng.* **114**, 2412–2417 (2017).
31. Xu, H. et al. Re-exploration of the codon context effect on amber codon-guided incorporation of noncanonical amino acids in *Escherichia coli* by the blue–white screening assay. *ChemBioChem* **17**, 1250–1256 (2016).
32. Pott, M., Schmidt, M. J. & Summerer, D. Evolved sequence contexts for highly efficient amber suppression with noncanonical amino acids. *ACS Chem. Biol.* **9**, 2815–2822 (2014).
33. Bartoschek, M. D. et al. Identification of permissive amber suppression sites for efficient non-canonical amino acid incorporation in mammalian cells. *Nucleic Acids Res.* **49**, e62 (2021).
34. Dunkelmann, D. L., Oehm, S. B., Beattie, A. T. & Chin, J. W. A 68-codon genetic code to incorporate four distinct non-canonical amino acids enabled by automated orthogonal mRNA design. *Nat. Chem.* **13**, 1110–1117 (2021).
35. Fluitt, A., Pienaar, E. & Viljoen, H. Ribosome kinetics and aa-tRNA competition determine rate and fidelity of peptide synthesis. *Comput. Biol. Chem.* **31**, 335–346 (2007).
36. Bryson, D. I. et al. Continuous directed evolution of aminoacyl-tRNA synthetases. *Nat. Chem. Biol.* **13**, 1253–1260 (2017).
37. Melnikov, S. V. & Söll, D.Aminoacyl-tRNA synthetases and tRNAs for an expanded genetic code: what makes them orthogonal? *Int. J. Mol. Sci.* **20**, 1929 (2019).
38. Liu, F., Bratulic, S., Costello, A., Miettinen, T. P. & Badran, A. H. Directed evolution of rRNA improves translation kinetics and recombinant protein yield. *Nat. Commun.* **12**, 5638 (2021).
39. Willis, J. C. W. & Chin, J. W. Mutually orthogonal pyrrolysyl-tRNA synthetase/tRNA pairs. *Nat. Chem.* **10**, 831–837 (2018).
40. Chatterjee, A., Lajoie, M. J., Xiao, H., Church, G. M. & Schultz, P. G. A bacterial strain with a unique quadruplet codon specifying non-native amino acids. *ChemBioChem* **15**, 1782–1786 (2014).
41. Dunkelmann, D. L., Willis, J. C. W., Beattie, A. T. & Chin, J. W. Engineered triply orthogonal pyrrolysyl-tRNA synthetase/tRNA pairs enable the genetic encoding of three distinct non-canonical amino acids. *Nat. Chem.* **12**, 535–544 (2020).
42. Ikeda-Boku, A. et al. A simple system for expression of proteins containing 3-azidotyrosine at a pre-determined site in *Escherichia coli*. *J. Biochem.* **153**, 317–326 (2013).
43. Chatterjee, A., Xiao, H., Yang, P. Y., Soundararajan, G. & Schultz, P. G. A tryptophanyl-tRNA synthetase/tRNA pair for unnatural amino acid mutagenesis in *E.coli*. *Angew. Chem. Int. Ed. Engl.* **52**, 5106–5109 (2013).
44. Hughes, R. A. & Ellington, A. D. Rational design of an orthogonal tryptophanyl nonsense suppressor tRNA. *Nucleic Acids Res.* **38**, 6813–6830 (2010).
45. Cervettini, D. et al. Rapid discovery and evolution of orthogonal aminoacyl-tRNA synthetase–tRNA pairs. *Nat. Biotechnol.* **38**, 989–999 (2020).
46. Kwon, I., Wang, P. & Tirrell, D. A. Design of a bacterial host for site-specific incorporation of *p*-bromophenylalanine into recombinant proteins. *J. Am. Chem. Soc.* **128**, 11778–11783 (2006).
47. Park, H.-S. et al. Expanding the genetic code of *Escherichia coli* with phosphoserine. *Science* **333**, 1151–1154 (2011).
48. Rogerson, D. T. et al. Efficient genetic encoding of phosphoserine and its nonhydrolyzable analog. *Nat. Chem. Biol.* **11**, 496–503 (2015).
49. Anderson, J. C. & Schultz, P. G. Adaptation of an orthogonal archaeal leucyl-tRNA and synthetase pair for four-base, amber, and opal suppression. *Biochemistry* **42**, 9598–9608 (2003).
50. Liu, D. R. & Schultz, P. G. Progress toward the evolution of an organism with an expanded genetic code. *Proc. Natl Acad. Sci. USA* **96**, 4780–4785 (1999).
51. Zambaldo, C. et al. An orthogonal seryl-tRNA synthetase/tRNA pair for noncanonical amino acid mutagenesis in *Escherichia coli*. *Bioorg. Med. Chem.* **28**, 115662 (2020).
52. Makino, Y. et al. An archaeal ADP-dependent serine kinase involved in cysteine biosynthesis and serine metabolism. *Nat. Commun.* **7**, 13446 (2016).
53. Wang, K. et al. Optimized orthogonal translation of unnatural amino acids enables spontaneous protein double-labelling and FRET. *Nat. Chem.* **6**, 393–403 (2014).
54. Jaric, J. & Budisa, N. in *Hydrocarbon and Lipid Microbiology Protocols* (eds McGenity, T. et al.) 71–82 (Springer, 2015).
55. Wang, J. et al. A biosynthetic route to photoclick chemistry on proteins. *J. Am. Chem. Soc.* **132**, 14812–14818 (2010).
56. Abdelkader, E. H. et al. Genetic encoding of cyanopyridylalanine for in-cell protein macrocyclization by the nitrile–aminothiol click reaction. *Angew. Chem. Int. Ed. Engl.* **61**, e202114154 (2022).
57. Xiao, H. et al. Genetic incorporation of histidine derivatives using an engineered pyrrolysyl-tRNA synthetase. *ACS Chem. Biol.* **9**, 1092–1096 (2014).
58. Yanagisawa, T. et al. Structural basis for genetic-code expansion with bulky lysine derivatives by an engineered pyrrolysyl-tRNA synthetase. *Cell Chem. Biol.* **26**, 936–949 (2019).
59. Robertson, W. E. et al. Sense codon reassignment enables viral resistance and encoded polymer synthesis. *Science* **372**, 1057–1062 (2021).
60. Tianero, M. D., Donia, M. S., Young, T. S., Schultz, P. G. & Schmidt, E. W. Ribosomal route to small-molecule diversity. *J. Am. Chem. Soc.* **134**, 418–425 (2012).
61. Young, T. S. et al. Evolution of cyclic peptide protease inhibitors. *Proc. Natl Acad. Sci. USA* **108**, 11052–11056 (2011).
62. Vamisetti, G. B. et al. Selective macrocyclic peptide modulators of Lys63-linked ubiquitin chains disrupt DNA damage repair. *Nat. Commun.* **13**, 6174 (2022).
63. Katoh, T. & Suga, H. In vitro selection of foldamer-like macrocyclic peptides containing 2-aminobenzoic acid and 3-aminothiophene-2-carboxylic acid. *J. Am. Chem. Soc.* **144**, 2069–2072 (2022).
64. Spinck, M. et al. Genetically programmed cell-based synthesis of non-natural peptide and depsipeptide macrocycles. *Nat. Chem.* **15**, 61–69 (2023).
65. Townend, J. E. & Tavassoli, A. Traceless production of cyclic peptide libraries in *E.coli*. *ACS Chem. Biol.* **11**, 1624–1630 (2016).
66. Miranda, E. et al. A cyclic peptide inhibitor of HIF-1 heterodimerization that inhibits hypoxia signaling in cancer cells. *J. Am. Chem. Soc.* **135**, 10418–10425 (2013).
67. Cheriyan, M., Pedamallu, C. S., Tori, K. & Perler, F. Faster protein splicing with the *Nostoc punctiforme* DnaE intein using non-native extein residues. *J. Biol. Chem.* **288**, 6202–6211 (2013).
68. Stevens, A. J. et al. A promiscuous split intein with expanded protein engineering applications. *Proc. Natl Acad. Sci. USA* **114**, 8538–8543 (2017).
69. Kolber, N. S., Fattal, R., Bratulic, S., Carver, G. D. & Badran, A. H. Orthogonal translation enables heterologous ribosome engineering in *E.coli*. *Nat. Commun.* **12**, 599 (2021).

Methods

General methods

All DNA manipulations were performed using NEB Turbo cells (New England Biolabs) or Mach1F cells, which are Mach1 T1^R cells (Thermo Fisher Scientific) mated with S2057 F′ to constitutively provide TetR and LacI[69]. All assays were performed with *E. coli* S3489 as previously described[38]. Water was purified using a MilliQ water purification system (Millipore). All amplifications were carried out using Phusion U Hot Start DNA polymerase (Life Technologies) or repliQa HiFi ToughMix (Quantabio). A MinElute PCR purification kit (Qiagen) was used to purify all PCR products to a 10-µl final volume, which was quantified using a NanoDrop 1000 spectrophotometer (Thermo Fisher Scientific). All plasmids were constructed by USER (uracil-specific excision reagent; endonuclease VIII and uracil-DNA glycosylase) cloning[70].

USER cloning and sequencing

A single internal deoxyuracil base was included at 10–20 bases from the 5′ end of each primer used to amplify plasmid components. This region is described as the USER junction, which specifies the homology required for correct assembly. USER junctions were designed to contain minimal secondary structure, have 42 °C < T_m < 70 °C and begin with a deoxyadenosine and end with a deoxythymine (to be replaced by deoxyuridine). For USER assembly, equimolar ratios (up to 1 pmol each) of PCR products carrying complementary USER junctions were mixed in a 10-µl reaction containing 0.75 U of DpnI (New England Biolabs), 0.75 U of USER enzyme (New England Biolabs), 1 µl of CutSmart Buffer (50 mM potassium acetate, 20 mM Tris-acetate, 10 mM magnesium acetate and 100 µg ml^{-1} BSA at pH 7.9; New England Biolabs). The reactions were incubated at 37 °C for 20–45 min, followed by heating at 80 °C for 2 min and slow cooling to 12 °C at 0.1 °C s^{-1} in a thermocycler. The hybridized constructs were directly used for heat-shock transformation of chemically competent NEB Turbo *E. coli* cells or Mach1F *E. coli* cells. 2xYT (United States Biological) agar plates (1.8%) supplemented with the appropriate antibiotic(s) were used to select for transformants. In all cases, cloned plasmids were verified by Sanger sequencing using template generated by the TempliPhi 500 amplification kit (GE Life Sciences) according to the manufacturers protocol or by Nanopore sequencing of purified plasmid DNA (Primordium Labs). For high-throughput sequencing of synonymous codon library pools, plasmid DNA was isolated from sorted populations using conventional silica-based column purification. The region of interest was amplified, barcoded and sequenced using a MiSeq instrument (Illumina). For data analysis, a 33-bp or 34-bp insert between forward and reverse primers was extracted and the frequencies of silent codons was quantified. Data were filtered for only synonymous mutations to eliminate issues that may result from incorrect synthesis of the library primers. During our synonymous codon library generation efforts, we noted that creating a codon for serine at the −4 position required six codons but this could not be exclusively introduced using a degenerate codon approach. In this case, we elected to use the degenerate codon WSN (where W = A/T, S = C/G and N = A/C/G/T), which includes all six serine codons, as well as codons corresponding to arginine, threonine, cysteine and tryptophan. In our NGS analysis, we excluded the reads that did not match the intended amino acid sequence (that is, serine) as our goal was to only identify codon usage biases that affect quadruplet decoding and not local amino acid contexts.

Chemically competent cell preparation and transformation

An overnight culture was diluted 100-fold in 2xYT media containing maintenance antibiotics and grown at 37 °C with shaking at 350 r.p.m. to an optical density at 600 nm (OD_{600}) of 0.5–0.7. Cells were pelleted by centrifugation at 5,000*g* for 5 min at 0 °C. Supernatant was decanted and the pellet was resuspended in the residual medium, keeping it on ice for 10–15 min and gently shaking at short regular intervals. TSS (2xYT medium supplemented with 5% v/v DMSO, 10% w/v PEG 3350 and 20 mM $MgCl_2$) was added to the resuspend cells and mixed by gently swirling at a volume of 10% of the original culture. The cell suspension was then aliquoted, flash-frozen in liquid nitrogen and stored at −80 °C until needed. To transform cells, 100–400-µl aliquots of competent cells were thawed from −80 °C on ice for 15 min. An equal volume of KCM solution (100 mM KCl, 30 mM $CaCl_2$ and 50 mM $MgCl_2$) was added to the tube and mixed gently by tapping the tube. Plasmid DNA or USER reactions were mixed with TSS–KCM–cell mixtures and left on ice for 10–30 min. Then, cell–DNA mixtures were heat-shocked at 42 °C for 90 s. The transformation was then chilled on ice for 2 min and added to 1 ml of 2xYT medium. Cells were allowed to recover at 37 °C with shaking at 350 r.p.m. for 1 h before plating on 2xYT agar (1.5%) containing the appropriate antibiotics and incubated at 37 °C for 16–18 h.

Luminescence assays

Early log-phase (OD_{600} = 0.3–0.5) S3489 cells carrying the reporter plasmid (RP) grown in 2xYT (United States Biological) were made chemically competent, transformed with the desired tRNA–synthetase expression plasmids (EPs) and recovered for 2 h in 2xYT (United States Biological). Transformations were plated on 1.8% agar–2xYT plates (United States Biological) supplemented with carbenicillin (15 µg ml^{-1}), trimethoprim (3 µg ml^{-1}) and spectinomycin (30 µg ml^{-1}). The plates were incubated for 12–18 h in a 37 °C incubator. Colonies transformed with the appropriate EPs were picked the following day and grown in Davis Rich Medium (DRM) containing carbenicillin (15 µg ml^{-1}), trimethoprim (3 µg ml^{-1}) and spectinomycin (30 µg ml^{-1}) for 18 h. Following overnight growth of the EP/RP-carrying strains, cultures were diluted 250-fold into fresh DRM supplemented with carbenicillin (15 µg ml^{-1}), trimethoprim (3 µg ml^{-1}) and spectinomycin (30 µg ml^{-1}). The cultures were induced with anhydrotetracycline (aTc) (100 ng ml^{-1}) and IPTG (1 mM) and supplemented with relevant ncAAs (1 mM). After 4–6 h of growth in a 37 °C incubator shaking at 900 r.p.m., 150 µl of culture was transferred to a 96-well black-wall clear-bottom plate (Costar) and measured for OD_{600} and luminescence. In all cases, values were recorded using an Infinite M1000 Pro microplate reader (Tecan) or Spark plate reader (Tecan) running software (SPARKCONTROL version 2.3). Each plasmid combination was assayed in four biological replicates. Luminescence activities were tabulated at OD_{600} = 0.1–0.5 in all cases. Where luminescent signal is reported as a percentage of wild-type reporter activity, a control RP encoding a wild-type luciferase was grown in parallel and used for normalization across assays. Data analysis for mean and s.d. was performed in GraphPad Prism (version 9).

Fluorescence assays

Chemically competent S3489 cells were transformed with combinations of tRNA–synthetase EPs alongside sfGFP RPs. In all cases, transformations were recovered for 2 h in 2xYT (United States Biological) and then plated on 1.8% agar–2xYT plates (United States Biological) supplemented with carbenicillin (50 µg ml^{-1}) and trimethoprim (10 µg ml^{-1}) for one-plasmid or two-plasmid circuits or carbenicillin (15 µg ml^{-1}), trimethoprim (3 µg ml^{-1}) and spectinomycin (30 µg ml^{-1}) for three-plasmid circuits. Plates were incubated for 12–18 h in a 37 °C incubator. Colonies were picked the following day and grown in DRM containing appropriate antibiotics. After growth for 16–24 h at 37 °C shaking at 900 r.p.m., cultures were diluted 100–500-fold in DRM with antibiotics, inducers and ncAAs as above. Following another 16–24 h at 37 °C shaking at 900 r.p.m., 150 µl of culture was transferred to a 96-well black-wall clear-bottom plate (Costar) and measured for OD_{600} and fluorescence values. The reporter protein sfGFP was recorded by excitation at 485 nm and emission at 510 nm. In all cases, values were recorded using an Infinite M1000 Pro microplate reader (Tecan) or Spark plate reader (Tecan) running software (SPARKCONTROL version 2.3). Each variant was assayed in 4–8 biological replicates. Fluorescent protein yields were normalized to the culture OD_{600} in all cases. Where sfGFP yield is reported as a percentage of wild-type sfGFP, a control RP

encoding a wild-type sfGFP was grown in parallel and used for normalization across assays. Data analysis for mean and s.d. was performed in GraphPad Prism (version 9).

qtRNA directed evolution

qtRNA libraries were made by PCR with degenerate base primers to amplify the qtRNA of interest. Libraries were assembled by USER cloning (described above) and transformed into S3489 cells carrying an aaRS EP. Library transformations were recovered for 1 h in DRM with aaRS and tRNA induction and, where applicable, cognate ncAA (1 mM). Transformations were plated by spreading on DRM agar chloramphenicol selection plates (described below) and grown at 37 °C for up to 96 h. For DRM plates, DRM was combined 4:1 with a 5% molten agar solution. All maintenance antibiotics were added to the DRM agar: carbenicillin (50 μg ml^{-1}) and trimethoprim (10 μg ml^{-1}), along with aaRS and qtRNA induction using aTc (100 ng mL^{-1}) and IPTG (1 mM). For positive selections, cognate ncAA (1 mM) was added as a powder to molten DRM agar and mixed by stir bar. A titration of chloramphenicol from 32 to 0 μg ml^{-1} was made in molten DRM agar with all required small molecules. To validate colony resistance to chloramphenicol, colonies found to grow above the strain MCI were picked into DRM with maintenance antibiotics overnight. These cultures were used to inoculate liquid culture titration of chloramphenicol to validate the phenotype.

Protein purification and analysis

S3489 strains transformed with synthetase EPs and RPs (sfGFP with C-terminal His-tag) were grown with maintenance antibiotics, aTc (100 ng ml^{-1}), IPTG (1 mM) and, where relevant, ncAA (1 mM) in 4 ml of DRM. Overnight cultures were pelleted and the supernatant was removed. To each pellet, 0.5 ml of BugBuster (Millipore) was added before incubating for 30 min at room temperature with gentle rocking. Soluble protein was separated by centrifugation at 16,000*g* for 20 min and the supernatant was recovered (soluble total protein lysate). Next, 300 μl of soluble lysate was loaded onto a His-Spin Protein Mini-prep column (Zymo) and purified using the manufacturer's protocol with one change (centrifugation steps were conducted at 700*g* for 2 min). All samples were eluted in 150 μl of elution buffer. Soluble protein lysate or Ni-NTA-purified protein was visualized by mixing 10 μg of lysate or 5 μl of eluted protein from purification with 4 μl of NuPAGE LDS sample buffer (Invitrogen), 10 mM dithiothreitol and distilled H_2O up to 16 μl and running on a 4–12% Bis-Tris NuPAGE gel (Invitrogen).

Immunoblotting

Protein material was resolved by SDS–PAGE in 17-well 4–12% Bis-tris NuPAGE gels (Invitrogen) in 1× MES–SDS buffer (Invitrogen). Samples were transferred to a nitrocellulose membrane (Amersham Protran) using a semidry transfer apparatus (BioRad). Membranes were blocked for 1 h at room temperature with StartingBlock (Thermo Fisher Scientific) in PBS. Membranes were incubated overnight with shaking at 4 °C with primary antibody (anti-c-Myc monoclonal antibody 9E10) (Thermo Fisher Scientific) at a 1:1,000 dilution in TBST with 5% BSA. Membranes were exposed to fluorophore-conjugated secondary antibodies (IRDye 680RD donkey antimouse IgG secondary antibody) (Li-Cor) diluted 1:5,000 in TBST with 5% BSA at room temperature. Signals were recorded with a Li-Cor fluorescence imager using a Chemi-Doc instrument (BioRad).

Time-of-flight (TOF) MS analysis

The salt concentration of Ni-NTA-purified sfGFP purified products was reduced in samples by buffer exchange with Amicon (Millipore) column centrifugation. Protein samples were then subjected to LC–MS analysis. The analysis was performed on an LC–MS instrument consisting of a Waters I-Class LC and a Waters Xevo G2-XS TOF, which uses a LeuEnk lockmass and is calibrated against sodium formate clusters. An Acquity BEH C4 column (1.7 μm, 2.1 × 55 mm) was used with a 5–99% B gradient lasting 4.4 min, followed by a 99% B isocratic hold lasting 0.6 min (A, 0.1% aqueous formic acid; B, 0.06% formic acid in acetonitrile; flow rate, 0.4 ml min^{-1}; column temperature, 55 °C). Multiply-charged electrospray ionization data were deconvoluted using the Waters MaxEnt 1 algorithm (Masslynx, Waters). The theoretical molecular weights of proteins encoding ncAAs were calculated by first computing the theoretical molecular weight of wild-type protein using an online tool (Expasy ProtParam; http://web.expasy.org/protparam/) and then manually corrected for the theoretical molecular weight of ncAAs following chromophore maturation.

Macrocycle biosynthesis and purification

Chemically competent S3489 were transformed with plasmids encoding the synthetase, tRNA/qtRNA and desired macrocycle generating *Npu* intein. In all cases, transformations were recovered for 2 h in 2xYT (United States Biological). All transformations were plated on 1.8% agar–2xYT plates (United States Biological) supplemented with relevant antibiotics: carbenicillin (50 μg ml^{-1}) and trimethoprim (10 μg ml^{-1}) for two-plasmid circuits. The plates were incubated for 12–18 h in a 37 °C incubator. Colonies transformed with the appropriate EPs were picked the following day and grown in DRM containing carbenicillin (50 μg ml^{-1}) and trimethoprim (10 μg ml^{-1}). After growth for 16–24 h at 37 °C with 900 r.p.m. shaking overnight, cultures were diluted 200-fold in DRM with carbenicillin (50 μg ml^{-1}), trimethoprim (10 μg ml^{-1}), aTc (100 ng ml^{-1} or 15 ng ml^{-1} for triple suppression experiments), IPTG (1 mM), L-arabinose (10 mM) and chosen amino acid(s) (1 mM). Following 16–24 h at 37 °C with 900 r.p.m. shaking, overnight cultures were centrifuged at 3,000*g* for 5 min and the supernatant was discarded. The pellet was resuspended in 500 μl of PBS with 15% glycerol per ml of overnight culture and then centrifuged at 3,000*g* for 5 min. The supernatant was discarded. The cell pellet was vigorously vortexed and sonicated for 10 min and then subjected to 3–5 cycles of 10 s of liquid nitrogen, 2 min at 42 °C and 2 min of sonication. Following this, the pellet was resuspended in 50 μl of acetonitrile and sonicated for 20–30 min. Finally, this was filtered through a 96-well PVDF filter plate (Corning 3505) by centrifugation at 800*g* for 10 min. LC–MS analysis was conducted on filtered samples by the Scripps Research Institute Automated Synthesis Facility, performed on a Waters I-Class LC and a Waters SQD2 single-quadrupole MS with a Waters Cortecs C18 column (1.6 μm, 2.1 × 55 mm) using a 5–99% B gradient lasting 2.5 min (A, 0.1% aqueous formic acid; B, 0.06% formic acid in acetonitrile; flow rate, 0.8 ml min^{-1}; column temperature, 35 °C). XICs and mass spectra were analyzed, collected and exported using Waters Empower 3.7. SPPS and purified macrocycle standards were generated by a commercial manufacturer (GenScript).

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

All data supporting the findings of this study are available within the article and its Supplementary Information. Select representative plasmids and strains were deposited to Addgene. NGS data were uploaded to the National Center for Biotechnology Information Sequence Read Archive (PRJNA1111233). Source data are provided with this paper.

References

70. Geu-Flores, F., Nour-Eldin, H. H., Nielsen, M. T. & Halkier, B. A. USER fusion: a rapid and efficient method for simultaneous fusion and cloning of multiple PCR products. *Nucleic Acids Res.* **35**, e55 (2007).

Acknowledgements
We thank fellow Badran Lab members for their helpful discussions. We gratefully acknowledge H. Li for assistance with NGS analyses, M. L. Bulos for assistance with immunoblotting, B. Seegers, B. Monteverde and A. Owirka of the Scripps Research Flow Cytometry Core for their assistance with cell sorting and Q. Nguyen Wong, B. Sanchez, J. Lee and J. Chen of the Scripps Research Institute Automated Synthesis Facility for their assistance with mass spectral analysis. This work was supported by the Scripps Research Institute, the National Institutes of Health Director's Early Independence Award (DP5-OD024590 to A.H.B.), the Research Corporation for Science Advancement and Sloan Foundation (G-2023-19625 to A.H.B.), the Thomas Daniel Innovation Fund (627163_1 to A.H.B.), the Abdul Latif Jameel Water and Food Systems Lab Grand Challenge Award (GR000141-S6241 to A.H.B.), the Breakthrough Energy Explorer Grant (GR000056 to A.H.B.), the Foundation for Food and Agriculture Research New Innovator Award (28-000578 to A.H.B.), the Homeworld Collective Garden Grant (GR000129 to A.H.B.) and the Army Research Office Young Investigator Award (81341-BB-ECP to A.H.B.). A.A.P. is a Hope Funds for Cancer Research Fellow supported by the Hope Funds for Cancer Research Fellowship (HFCR-23-03-01). D.L.L. is supported by a Skaggs-Oxford Scholarship and a Fletcher Jones Foundation Fellowship.

Author contributions
A.C. designed the study, led the experimental work and analyzed the results. A.A.P. designed and led all macrocycle production studies. D.L.L. contributed the orthogonal aaRS–tRNA matrix and codon–anticodon discovery efforts. Z.L. investigated RP origin copy number changes to optimize quadruplet decoding. G.D.C. designed the mRNA synonymous codon libraries. A.H.B. conceptualized and designed the study, performed the experiments, analyzed the results and supervised the research. A.C. and A.H.B. wrote the paper with input from all authors.

Competing interests
A.H.B. and A.C. have filed a provisional patent application through The Scripps Research Institute on the sequences and activities of tRNAs, proteins, enzymes and bacterial strains described in this paper. The other authors declare no competing interests.

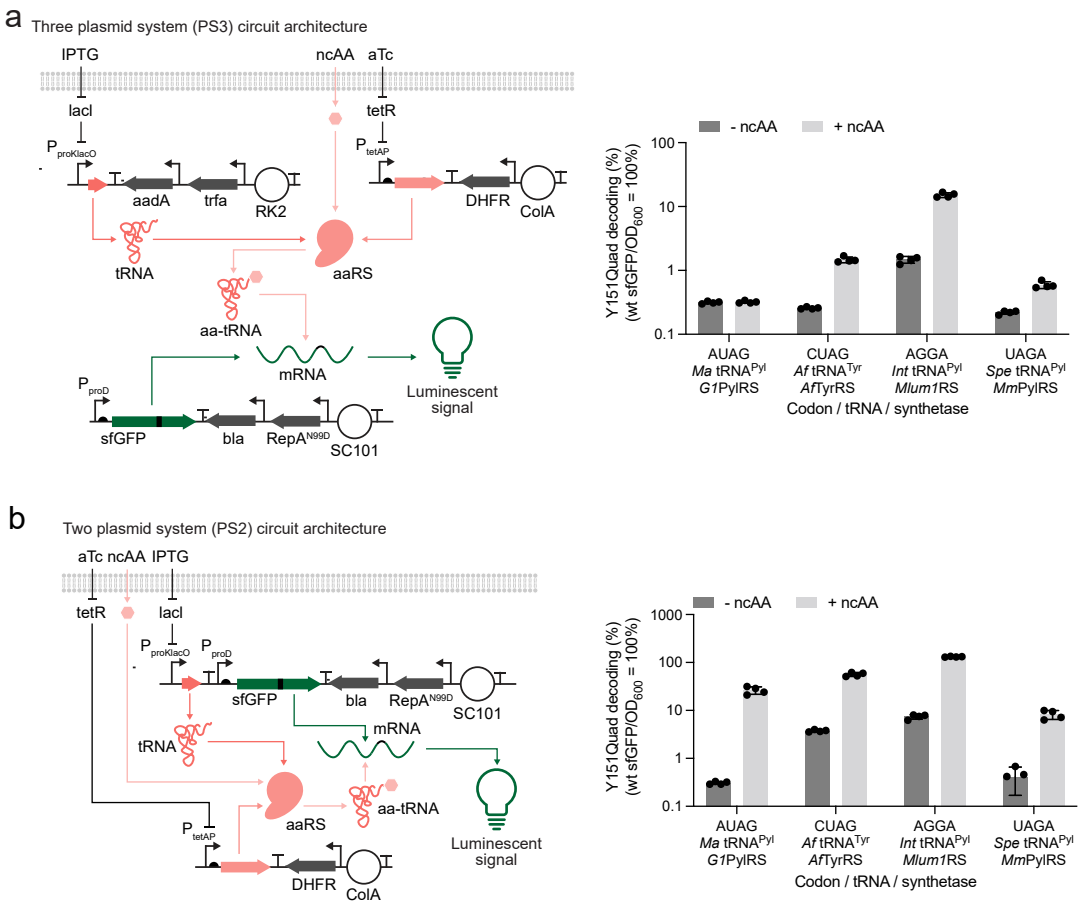
Additional information
Extended data is available for this paper at https://doi.org/10.1038/s41587-024-02385-y.

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41587-024-02385-y.

Correspondence and requests for materials should be addressed to Ahmed H. Badran.

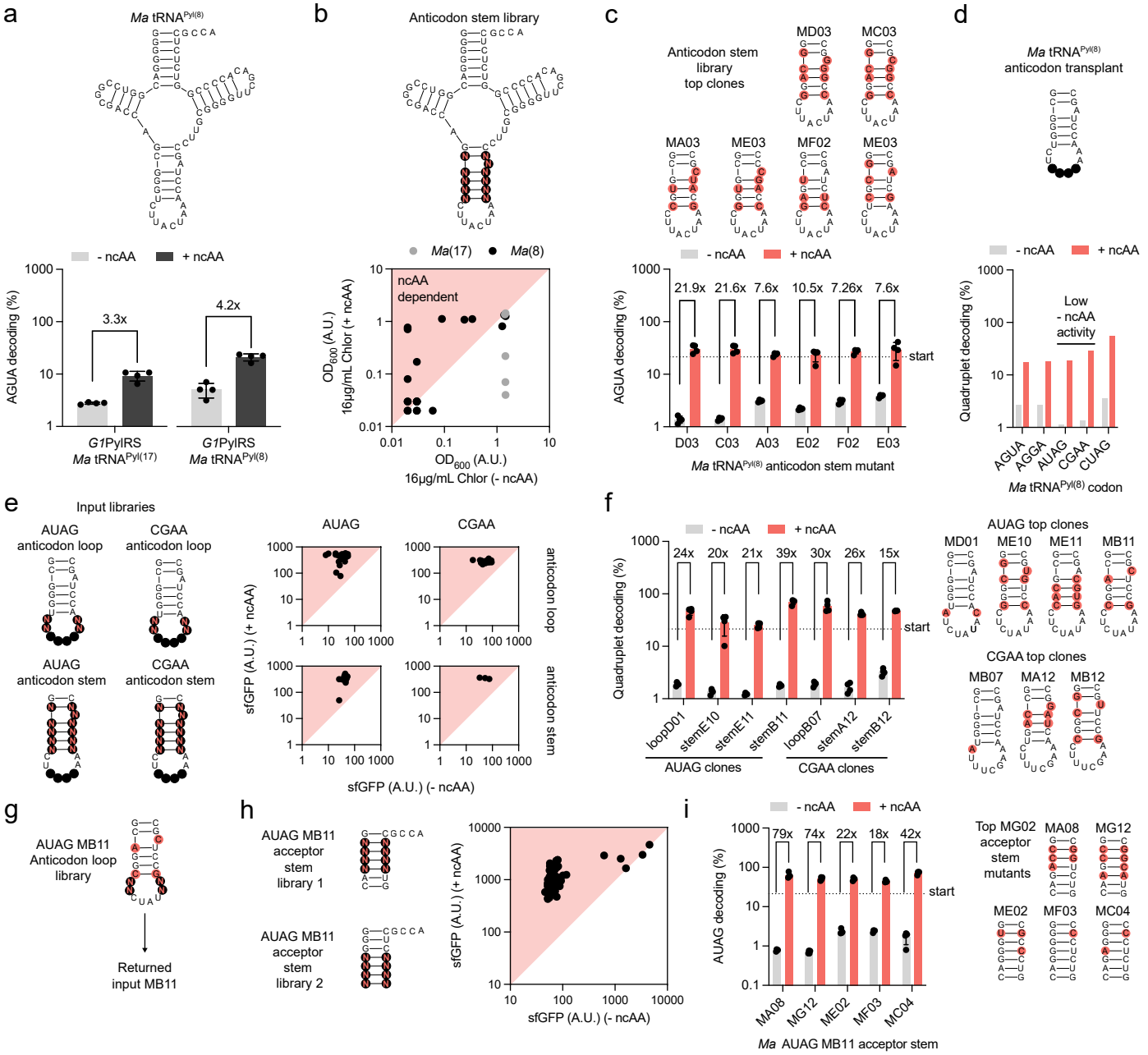
Peer review information *Nature Biotechnology* thanks the anonymous reviewers for their contribution to the peer review of this work.

https://doi.org/10.1038/s41587-024-02385-y



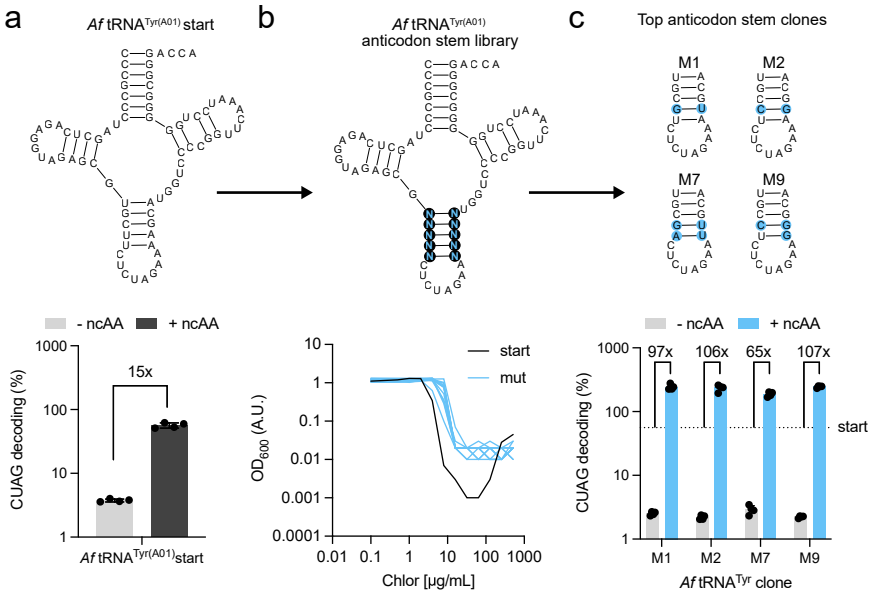
Extended Data Fig. 1 | Circuit Architecture Optimization for Efficient Quadruplet Decoding. a) Schematic representation and data of three-plasmid circuit architecture (PS3). aaRS and tRNA genes are encoded on isolated plasmids and used in combination to decode a quadruplet codon in the reporter plasmid. Y151 substituted sfGFP quadruplet decoding by *Ma* $tRNA^{Pyl(8)}_{AUAG}$, *Af* $tRNA^{Tyr(A01)}_{CUAG}$, *Int* $tRNA^{PylA17VB03}_{AGGA}$, and *Spe* $tRNA^{Pyl}_{UAGA}$ in PS3 circuit architectures. ncAAs are 3-cyano-L-phenylalanine, 4-iodo-L-phenylalanine, 3-methyl-L-histidine, and N6-Boc-L-Lysine respectively, (n = 4 biological replicates, error shows standard deviation). **b**) Circuit architecture was refined to two plasmids (PS2) by moving the tRNA expression cassette onto the reporter plasmid. Y151 substituted sfGFP quadruplet decoding by *Ma* $tRNA^{Pyl(8)}_{AUAG}$, *Af* $tRNA^{Tyr(A01)}_{CUAG}$, *Int* $tRNA^{PylA17VB03}_{AGGA}$, and *Spe* $tRNA^{Pyl}_{UAGA}$ in PS2 circuit architectures. ncAAs are 3-cyano-L-phenylalanine, 4-iodo-L-phenylalanine, 3-methyl-L-histidine, and N6-Boc-L-Lysine respectively, (n = 4 biological replicates, error shows standard deviation).

https://doi.org/10.1038/s41587-024-02385-y



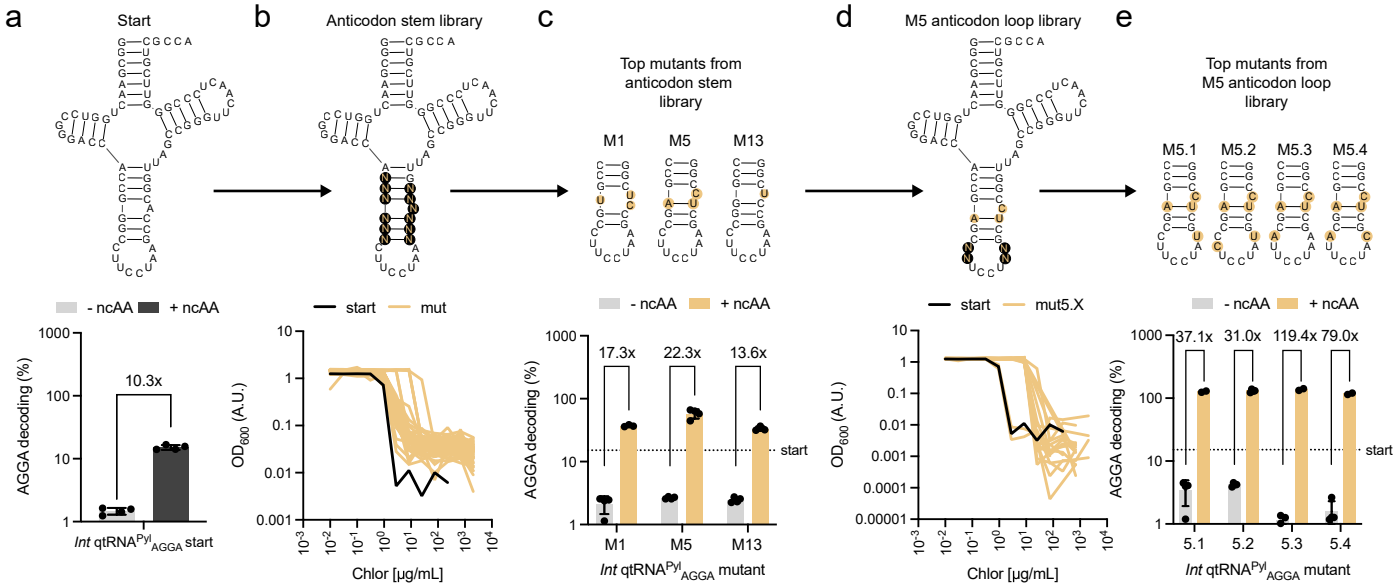
Extended Data Fig. 2 | Directed Evolution of Improved Quadruplet Decoding Using of G1PylRS–Ma qtRNAPylAUAG. a) The starting *Ma* tRNA$^{Pyl}_{AGUA}$ variants (8) and (17)[38] show poor AGUA decoding at Y151 in sfGFP using a two-plasmid system alongside *G1*PylRS and 1 mM H-Lys(Z)-OH (CbzK), four biological replicates, (n = 4 biological replicates, error shows standard deviation). **b**) Both qtRNA anticodon stems were randomized using degenerate oligonucleotides and subjected to positive selection on chloramphenicol agar plates. Single clones were assayed for resistance to 16 μg/mL chloramphenicol with and without 1 mM CbzK, yielding variants that catalyzed CbzK-dependent AGUA decoding (n = 1). **c**) Top chlor-resistant clones were further validated through AGUA decoding at Y151 in sfGFP. Start line indicates the highest activity achieved with the (8) variant, (n = 4 biological replicates, error shows standard deviation). Most clones showed a reduction in CbzK-independent decoding rather than an improvement in CbzK-dependent decoding. Subsequent selections and randomization at other positions did not further improve dynamic range (not shown). **d**) Alternative codons were tested through transplantation into the *Ma* tRNA$^{Pyl}_{AGUA}$ variant (8) scaffold. Cognate codon decoding was monitored using a dedicated sfGFP reporter at Y151 with and without 1 mM CbzK in each case (n = 1). AUAG and CGAA codons were prioritized due to low background. **e**) *Ma* tRNA$^{Pyl}_{AUAG}$ and *Ma* tRNA$^{Pyl}_{CGAA}$ anticodon stems and loops were randomized using degenerate oligonucleotides, and subjected to positive selection on chloramphenicol agar plates. Single clones were assayed for decoding at Y151 in sfGFP, showing robust CbzK dependence (n = 1). **f**) Top chlor-resistant clones were further validated through AUAG and CGAA decoding at Y151 in sfGFP. Start line indicates the highest activity achieved with the (8) variant, (n = 4 biological replicates, error shows standard deviation). AUAG decoding qtRNAs were prioritized due to their lower background without CbzK. **g**) An anticodon loop library for *Ma* tRNA$^{Pyl}_{CGAA}$ variant MB11 returned input sequence despite robust library coverage. **h**) *Ma* tRNA^{Pyl} AUAG variant MB11 acceptor stem was randomized using degenerate oligonucleotides, and subjected to positive selection on chloramphenicol agar plates. Single clones were assayed for decoding at Y151 in sfGFP, showing improved decoding and CbzK dependence (n = 1). **i**) Top chlor-resistant clones were further validated through AUAG decoding at Y151 in sfGFP. Start line indicates the highest activity achieved with the (8) variant, (n = 4 biological replicates, error shows standard deviation). Despite the improvements in dynamic range of these qtRNAs, mutant B11 (MB11) was designated as the best variant due its higher signal with 1 mM CbzK in **(f)** and used for all subsequent assays.

https://doi.org/10.1038/s41587-024-02385-y



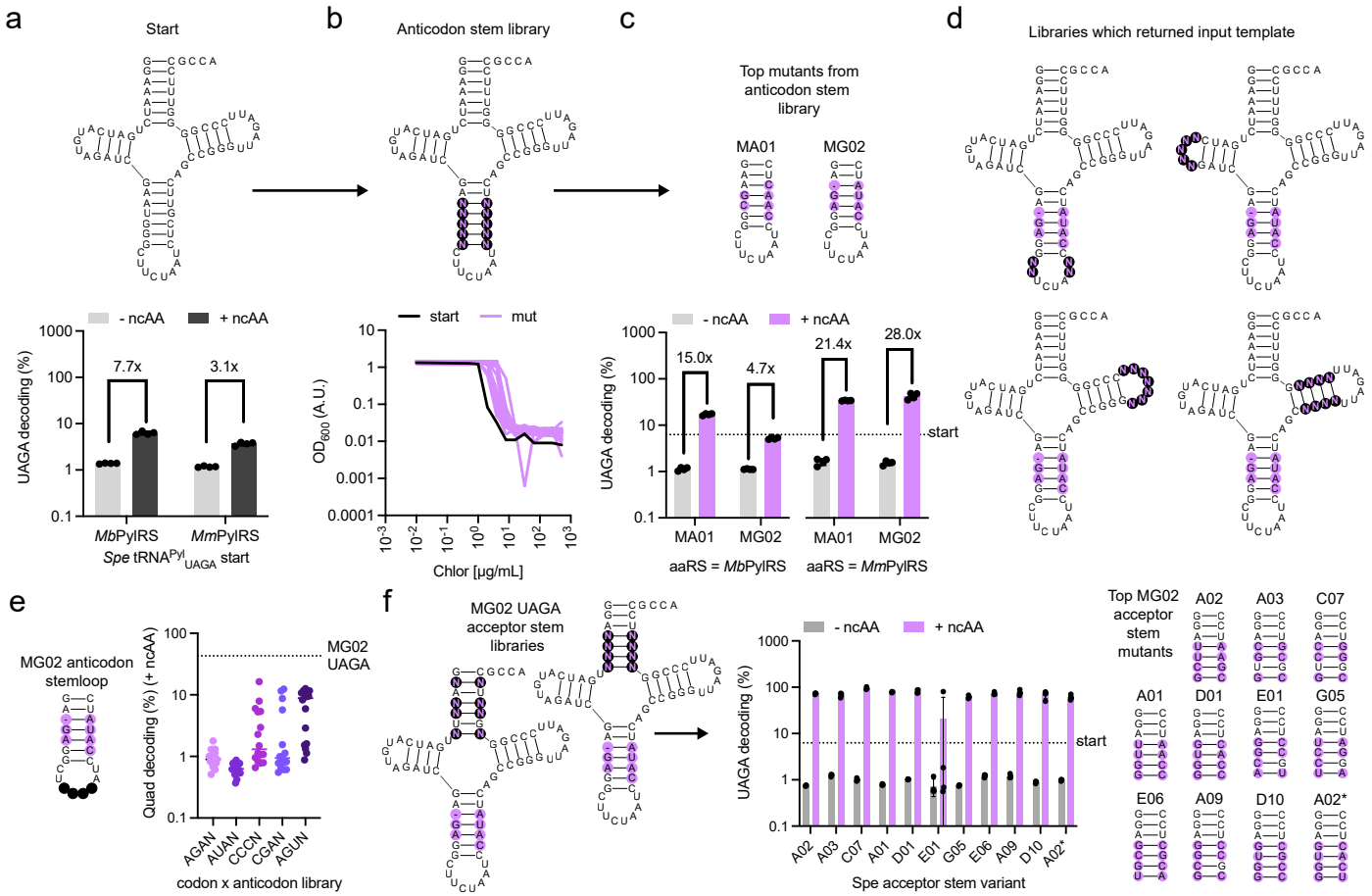
Extended Data Fig. 3 | Directed Evolution of Improved Quadruplet Decoding Using AfTyrRS–AfqtRNATyrCUAG. a) The starting *Af* qtRNA$^{Tyr(A01)}_{CUAG}$ variant[42] shows moderate CUAG decoding at Y151 in sfGFP using a two-plasmid system alongside *Af*TyrRS and 1 mM 4-iodo-L-phenylalanine, (n = 4 biological replicates, error shows standard deviation). **b**) The qtRNA anticodon stem was randomized using degenerate oligonucleotides, and subjected to positive selection on chloramphenicol agar plates. Evolved single clones showed robust chloramphenicol above the MIC using the starting qtRNA (n = 1). **c**) Top chloramphenicol-resistant clones were further validated through CUAG decoding at Y151 in sfGFP. Start line indicates the highest activity achieved with the *Af* qtRNA$^{Tyr(A01)}_{CUAG}$ variant, (n = 4 biological replicates, error shows standard deviation). Mutant 9 (M9) was designated as the best variant and used for all subsequent assays.

https://doi.org/10.1038/s41587-024-02385-y



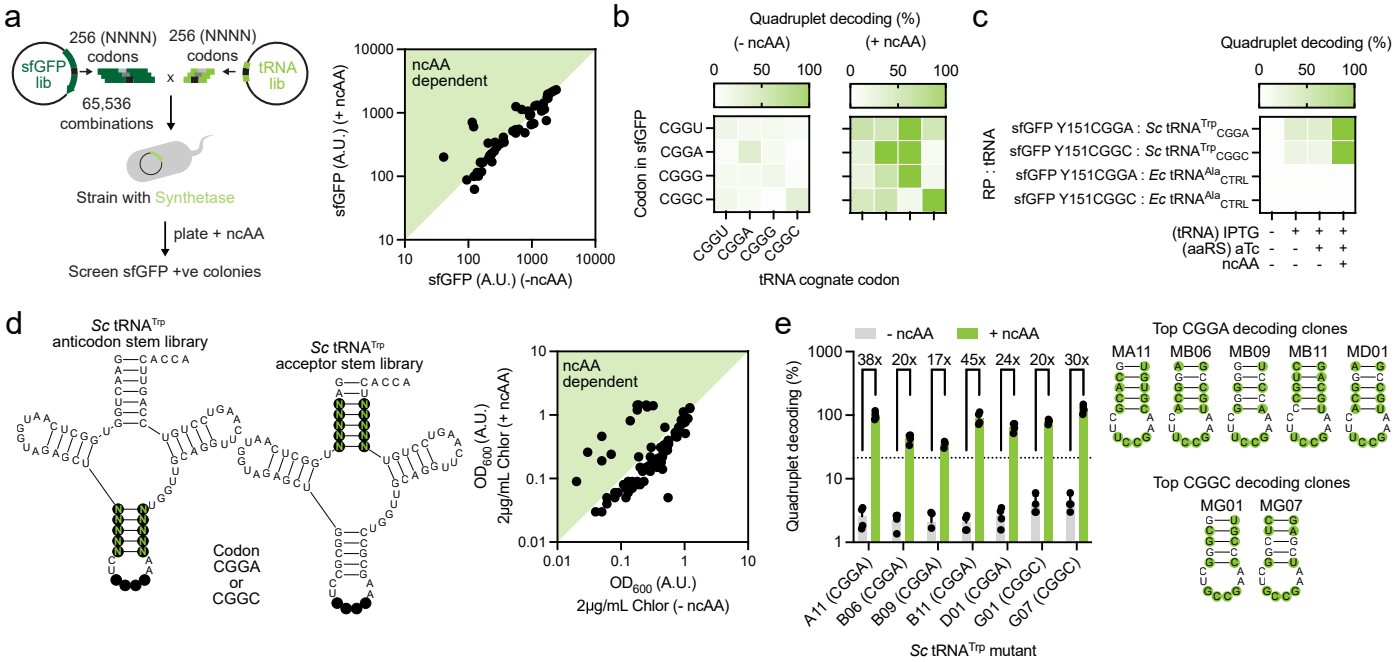
Extended Data Fig. 4 | Directed Evolution of Improved Quadruplet Decoding Using Mlum1RS–IntqtRNAPylAGGA. (**a**) The starting *Int* qtRNA$^{Pyl}_{AGGA}$ variant $^{A}17,^{V}B03$[38] shows poor AGGA decoding at Y151 in sfGFP using a three-plasmid system alongside *Mlum1*RS and 1 mM 3-methyl-L-histidine (NmH), (n = 4 biological replicates, error shows standard deviation). (**b**) The qtRNA anticodon stem was randomized using degenerate oligonucleotides, and subjected to positive selection on chloramphenicol agar plates. Evolved single clones showed robust chloramphenicol above the MIC using the starting qtRNA (n = 1). (**c**) Top chlor-resistant clones were further validated through AGGA decoding at Y151 in sfGFP. Start line indicates the highest activity achieved with the $^{A}17,^{V}B03$ variant, (n = 4 biological replicates, error shows standard deviation). (**d**) The top mutant (M5) was used as a scaffold for further randomized using targeting the anticodon loop, and subjected to another round of positive selection on chloramphenicol agar plates. Evolved single clones showed an even greater MIC to chloramphenicol as compared to the starting qtRNA (n = 1). (**e**) Top chlor-resistant clones were further validated through AGGA decoding at Y151 in sfGFP. Start line indicates the highest activity achieved with the $^{A}17,^{V}B03$ variant, (n = 4 biological replicates, error shows standard deviation). Mutant 5.3 (M5.3) was designated as the best variant and used for all subsequent assays.

https://doi.org/10.1038/s41587-024-02385-y



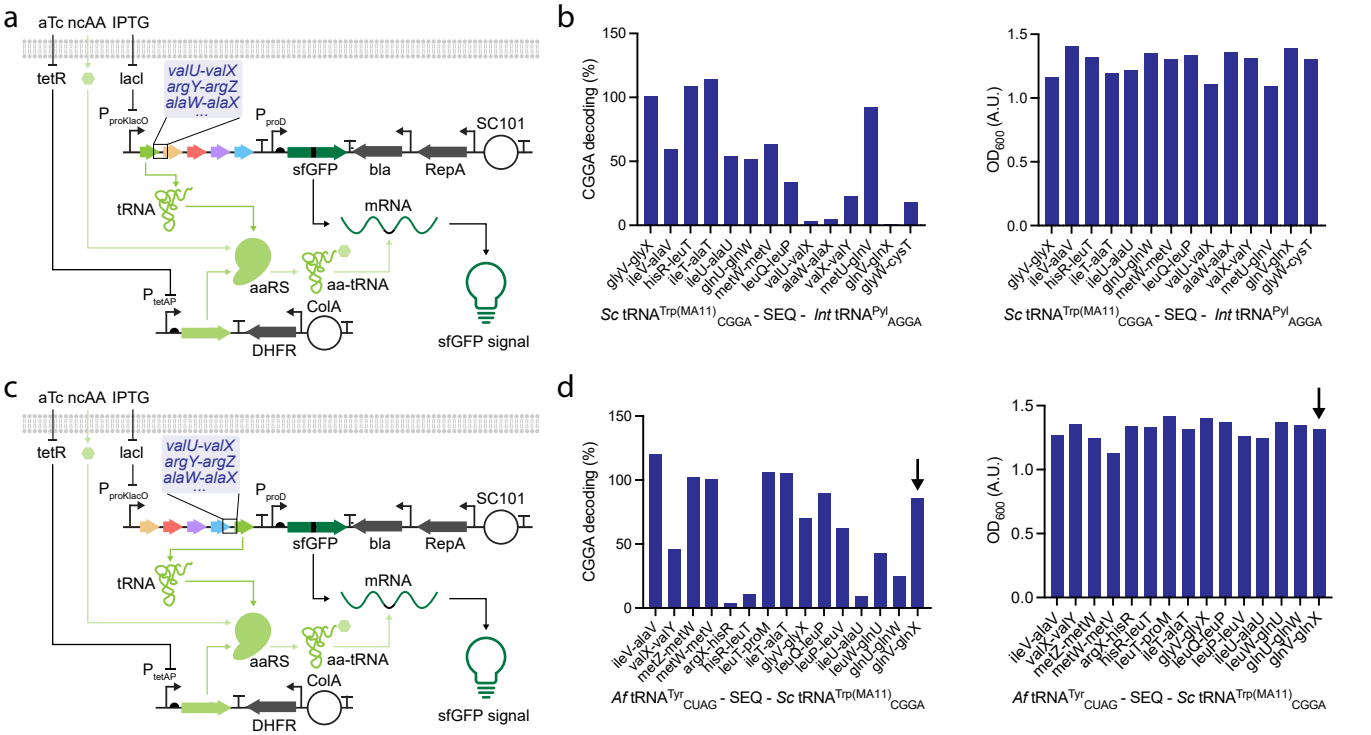
Extended Data Fig. 5 | Directed Evolution of Improved Quadruplet Decoding Using MmPylRS–SpetRNAPyl. a) The starting *Spe* qtRNA$^{Pyl}_{UAGA}$[38] shows poor UAGA decoding at Y151 in sfGFP using a two-plasmid system alongside both *Mb*PylRS and *Mm*PylRS and 1 mM N6-Boc-L-Lysine (BocK), (n = 4 biological replicates, error shows standard deviation). **b**) The qtRNA anticodon stem was randomized using degenerate oligonucleotides, and subjected to positive selection on chloramphenicol agar plates. Evolved single clones showed robust chloramphenicol above the MIC using the starting qtRNA (n = 1). Related qtRNA genes *Mb* qtRNA$^{Pyl}_{UAGA}$[15] and *Vul* qtRNA$^{Pyl}_{UAG}$[38] engineered to make *Vul* qtRNA$^{Pyl}_{UAGA}$, were subjected to the identical selection but did not result in chloramphenicol survival (not shown). **c**) Top chlor-resistant clones were further validated through UAGA decoding at Y151 in sfGFP using both *Mb*PylRS and *Mm*PylRS. Start line indicates the highest activity achieved with the initial *Spe* qtRNA$^{Pyl}_{UAGA}$ variant, (n = 4 biological replicates, error shows standard deviation). **d**) The starting *Spe* qtRNA$^{Pyl}_{UAGA}$, mutants A01 (MA01) and G02 (MG02) were used as scaffolds for further randomization targeting the anticodon loop, Ψ-arm, D-arm, and D-loop. In all cases, the selection returned *Spe* qtRNA$^{Pyl}_{UAGA}$ MG02 input suggesting maximal fitness. **e**) Alternative codons AGAN, AUAN, CCCN, CGAN, and AGUN were tested through transplantation into the *Spe* qtRNA$^{Pyl}_{UAGA}$ MG02 scaffold in a tRNA-mRNA selection using 4 μg/ mL chloramphenicol plates. Cognate codon decoding was validated by using dedicated sfGFP reporter at Y151 with and without 1 mM BocK. In each case, sixteen colonies tested for each codon-anticodon library, (n = 16). All variants were ultimately abandoned due to lower signal than *Spe* qtRNA$^{Pyl}_{UAGA}$ MG02. **f**) *Spe* qtRNA$^{Pyl}_{UAGA}$ variant MG02 acceptor stem was randomized using degenerate oligonucleotides, and subjected to positive selection on chloramphenicol agar plates. Top chlor-resistant clones were further validated through UAGA decoding at Y151 in sfGFP. Start line indicates the highest activity achieved with the initial *Spe* qtRNA$^{Pyl}_{UAGA}$ variant, (n = 4 biological replicates, error shows standard deviation). Mutant C07 (MC07) was designated as the best variant and used for all subsequent assays.

https://doi.org/10.1038/s41587-024-02385-y



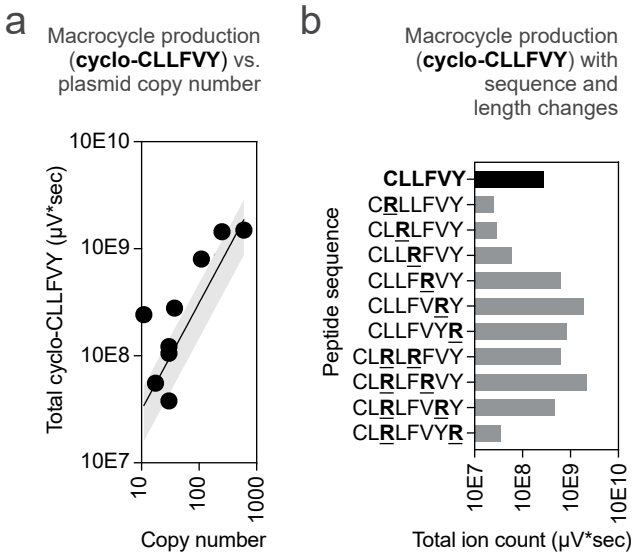
Extended Data Fig. 6 | Directed Evolution of Improved Quadruplet Decoding Using ScTrpRS–SctRNATrpCGGA. a) To define an appropriate quadruplet codon–anticodon pair, all possible pairs were evaluated using a library-cross-library approach. Position Y151 of sfGFP and the *Sc* $tRNA^{Trp(M13)}$ anticodon were randomized using degenerate oligonucleotides and co-transformed into S3489 *E. coli* cells carrying *Sc*TrpRS. Transformants were streaked on agar plates supplemented with 1 mM 5-hydroxy-L-tryptophan (5hW). Single colonies were then picked based on green fluorescence and evaluated with and without ncAA (n = 1). All 5hW-dependent clones carried CGGN codons and qtRNAs with the corresponding anticodons, including mismatches at the 4th position. **b**) Evaluation of all possible CGGN codon–anticodon pairs at position Y151 in sfGFP with and without 1 mM 5hW. *Sc* $tRNA^{Trp(M13)}_{CGGA}$ and *Sc* $tRNA^{Trp(M13)}_{CGGC}$ were prioritized in subsequent studies. **c**) Confirmation of ScTrpRS- and 5hW-dependent quadruplet decoding by comparison to a non-cognate tRNA, $AlaT_{GCA}$ (*Ec* $tRNA^{Ala}_{CTRL}$). We note a high 5hW-independent background translation. **d**) The starting *Sc* $tRNA^{Trp(M13)}_{CGGA}$ and *Sc* $tRNA^{Trp(M13)}_{CGGC}$ were randomized at the anticodon stem or acceptor stem, subjected to negative selection to eliminate aminoacylation by host synthetases, then subjected to positive selection on chloramphenicol agar plates. Evolved single clones showed robust 5hW-dependent growth at 2 µg/mL chloramphenicol (n = 1). **e**) Top chlor-resistant clones were further validated through CGGA or CGGC decoding at Y151 in sfGFP using *Sc*TrpRS. Start line indicates the highest activity achieved with the initial *Sc* $tRNA^{Trp(M13)}_{CGGA}$ variant, (n = 4 biological replicates, error shows standard deviation). Mutant A11 (MA11) was designated as the best variant and used for all subsequent assays.

https://doi.org/10.1038/s41587-024-02385-y



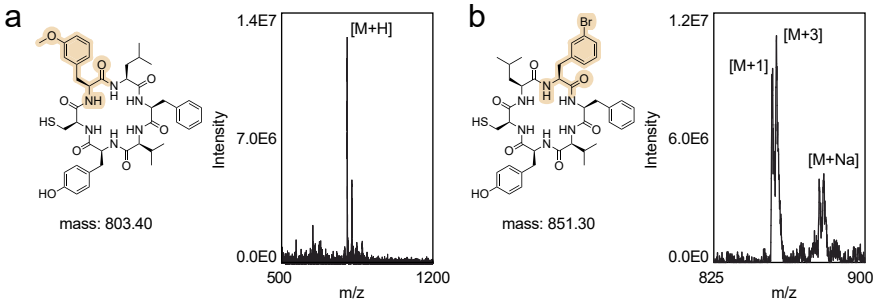
Extended Data Fig. 7 | Rational Engineering of a qtRNA Operon to Include Sc tRNATrp(M13)CGGA. a) Schematic representation of the genetic circuit for 5′ placement of *Sc* qtRNA$^{Trp(M13)}_{CGGA}$. *Sc* qtRNA$^{Trp(M13)}_{CGGA}$ is encoded 5′ of *Int* qtRNA$^{Pyl}_{AGGA}$, where the two qtRNAs are separated by *E. coli* derived inter-tRNA sequences (navy highlight). Functional *Sc* qtRNA$^{Trp(M13)}_{CGGA}$ production will yield green fluorescence when tested alongside *Sc*TrpRS and the cognate ncAA. **b**) *Sc* qtRNA$^{Trp(M13)}_{CGGA}$ production is dependent on the 5′ inter-qtRNA sequence as determined by CGGA decoding at Y151 in sfGFP (n = 1). OD_{600} values following overnight growth in the presence of 1 mM 5-hydroxy-L-tryptophan (5hW) are shown on the right (n = 1). **c**) Schematic representation of the genetic circuit for 3′ placement of *Sc* qtRNA$^{Trp(M13)}_{CGGA}$. *Sc* qtRNA$^{Trp(M13)}_{CGGA}$ is encoded 5′ of *Af* qtRNA$^{Tyr}_{CUAG}$, where the two qtRNAs are separated by *E. coli* derived inter-tRNA sequences (navy highlight). Functional *Sc* qtRNA$^{Trp(M13)}_{CGGA}$ production will yield green fluorescence when tested alongside *Sc*TrpRS and the cognate ncAA. **d**) *Sc* tRNA$^{Trp(M13)}_{CGGA}$ production is dependent on the 3′ inter-qtRNA sequence as determined by CGGA decoding at Y151 in sfGFP (n = 1). OD_{600} values following overnight growth in the presence of 1 mM 5hW are shown on the right (n = 1). Due to its lower error and tolerated cell growth, *glnV-glnX* inter-tRNA sequence was chosen for the 3′ spacer sequence, making *Sc* qtRNA$^{Trp(M13)}_{CGGA}$ the final qtRNA in the engineered operon.

https://doi.org/10.1038/s41587-024-02385-y



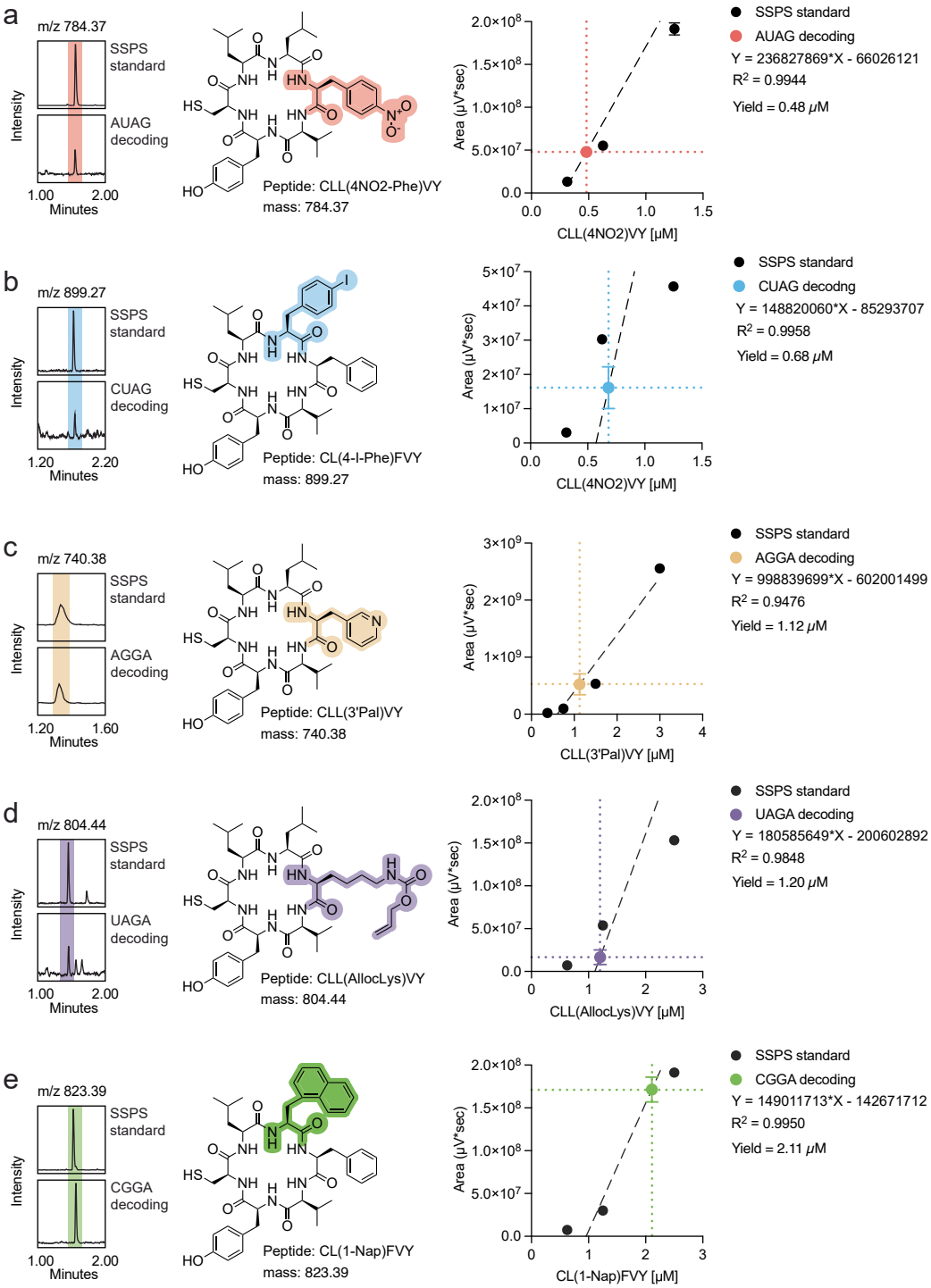
Extended Data Fig. 8 | Optimization of a Macrocycle Biosynthesis Platform for cyclo-CLLFVY. a) The split-*Npu* gene cassette driven by the pBAD promoter was cloned on plasmids bearing different origins of replication with variable copy number. Following expression, cells are lysed and organic soluble components isolated for analysis by LC-MS. Straight line is the logarithmic regression of these findings, and the grey area described the standard deviation of the regression. **b**) Arginine (R) substitutions and additions were made to the cyclo-CLLFVY macrocyclic peptide scaffold to explore tolerance to insertions and mutations. Following expression, cells were lysed and organic soluble components isolated for analysis by LC-MS.

https://doi.org/10.1038/s41587-024-02385-y



Extended Data Fig. 9 | Validation of LC-MS Analysis of E. coli Crude Lysate for Macrocycle Biosynthesis. Following expression, lysis, and organic solvent extraction of soluble components, LC-MS analysis was used to validate expression of macrocycle products. **a**) Mass spectrum for C[3OmeF]LFVY. The primary mass is M + H when analysing the peak indicated on the XIC trace from Fig. 5c. **b**) Characteristic isotopic pattern (M + 1 and M + 3) of bromine in the mass spectrum of this macrocycle product.

https://doi.org/10.1038/s41587-024-02385-y



Extended Data Fig. 10 | Quantifying Biosynthesis of Npu-Generated Macrocycles. Standard curves were derived using macrocycles generated by solid-phase peptide synthesis (SPPS standard) to quantify the biosynthetic yield of our platform. For each synthetase – tRNA – quadruplet codon set, we used a single representative macrocycle. We interpolated the Area (μV*sec) of SPPS standards over a serial dilution from 10 – 0.3 μM but changes depending on macrocycle (see Source Data File) and used this to calculate the yield of macrocycles generated by intracellular decoding of (**a**) AUGA with 4-nitro-L-phenylalanine, (**b**) CUAG with 4-iodo-L-phenylalanine, (**c**) AGGA with 3-pyridyl-L-alanine, (**d**) UAGA with N6-alloc-L-lysine, and (**e**) CGGA with 3-(1-naphthyl)-L-alanine. All SPPS and biosynthesized macrocycle samples were tested in biological triplicate (n = 3). LC traces and m/z spectra for all concentrations of SSPS synthesized samples and biosynthesized macrocycles can be found in (Supplemental Information).

nature portfolio

Corresponding author(s): Ahmed H. Badran

Last updated by author(s): Jul 28, 2024

Reporting Summary

Nature Portfolio wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Portfolio policies, see our Editorial Policies and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed
☐	☒ The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒ A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☒	☐ The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☒	☐ A description of all covariates tested
☒	☐ A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒ A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☒	☐ For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☒	☐ For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐ For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☒	☐ Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	Plate reader data was collected with Tecan Spark software SPARKCONTROL V2.3. For TOF-MS, the data was generated from an LC-MS consisting of a Waters I-Class LC and a Waters Xevo G2-XS TOF which uses a LeuEnk lockmass and is calibrated against sodium formate clusters.
Data analysis	Graph Pad Prism (version 9) was used for plotting and data analysis, including calculation of means, standard deviations, and regression lines. Clustal Omega server (https://www.ebi.ac.uk/jdispatcher/msa/clustalo) was used to generate alignments for sequence identities described in the manuscript. FlowJo V10.10.0 was used to analyze flow cytometry data. Analysis of TOF-MS samples: multiply-charged electrospray data (ESI+) were deconvoluted using the Waters' MaxEnt 1 algorithm (Masslynx; Waters). The theoretical molecular weights of proteins encoding ncAAs were calculated by first computing the theoretical molecular weight of wild-type protein using an online tool (Expasy ProtParam; http://web.expasy.org/protparam/) and then manually corrected for the theoretical molecular weight of ncAAs following chromophore maturation.

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Portfolio guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our policy

The data found in Figures 1-6 and Extended Data Figures 1-10 are available in the associated Source Data File. The data found in Supplementary Figures are available in the associated Supplementary Data File. Select representative plasmids and strains will be deposited to Addgene. NGS data has been uploaded to the NCBI SRA (PRJNA1111233).

Research involving human participants, their data, or biological material

Policy information about studies with human participants or human data. See also policy information about sex, gender (identity/presentation), and sexual orientation and race, ethnicity and racism.

Reporting on sex and gender	n/a
Reporting on race, ethnicity, or other socially relevant groupings	n/a
Population characteristics	n/a
Recruitment	n/a
Ethics oversight	n/a

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	For all assays, the sample size (generally n = 4) was determined by the dimensions of a 96-well plate. None of these experiments include animals and thus there are no individuals to examine. Sample size is not relevant but replication is important. The number of replicates is indicated in the figure legends.
Data exclusions	For fluorescence assays, any replicates that did not grow to an OD600>1.0 were removed from the subsequent analysis (it is not unusual for one of the four replicates to not grow).
Replication	For all experiments, the number of replicates is indicated in the figure legend.
Randomization	The nature of the experiments require addition of specific components to make the data meaningful and interpretable. Therefore randomization is not relevant.
Blinding	The nature of the experiments require addition of specific components to make the data meaningful and interpretable. Therefore blinding is not relevant.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☐	☒ Antibodies
☒	☐ Eukaryotic cell lines
☒	☐ Palaeontology and archaeology
☒	☐ Animals and other organisms
☒	☐ Clinical data
☒	☐ Dual use research of concern
☒	☐ Plants

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☐	☒ Flow cytometry
☒	☐ MRI-based neuroimaging

Antibodies

Antibodies used: Primary antibody (c-Myc Monoclonal Antibody 9E10) (Thermo Fisher).
Secondary antibodies (IRDye® 680RD Donkey anti-Mouse IgG Secondary Antibody) (Li-Cor).

Validation: We included a control in our Western Blot experiments: a c-Myc-tagged sfGFP bearing no quadruplet codon. This control was run in all experiments. We also run a blank lane in our experiments to show no binding of our antibody in the absence of the c-Myc tag.

Plants

Seed stocks: n/a

Novel plant genotypes: n/a

Authentication: n/a

Flow Cytometry

Plots

Confirm that:

☒ The axis labels state the marker and fluorochrome used (e.g. CD4-FITC).

☒ The axis scales are clearly visible. Include numbers along axes only for bottom left plot of group (a 'group' is an analysis of identical markers).

☒ All plots are contour plots with outliers or pseudocolor plots.

☒ A numerical value for number of cells or percentage (with statistics) is provided.

Methodology

Sample preparation: Bacterial cultures were diluted 100-fold in PBS solution prior to FACS.

Instrument: Beckman Coulter Astrios. Excitation/emission wavelength 488/513nm for sfGFP.

Software: FlowJo

Cell population abundance: For SCS library sample collection 1ES-1E6 cells were sorted for each enriched population "Top 1%, Top 5%, Bottom 5%, and Bottom 1%".

Gating strategy: The cells were gated first for all cells. Doublets were gated out of final analysis. Finally sfGFP positive cells were gated with respect to a negative control.

☒ Tick this box to confirm that a figure exemplifying the gating strategy is provided in the Supplementary Information.