Article

Automated prototyping of genetic codes

https://doi.org/10.1038/s41586-026-10949-y

Received: 6 June 2025

Accepted: 24 July 2026

Published online: 26 August 2026

Open access

Check for updates

Felix Radford[1,2]✉, Nayan Sapers[1], Hana M. Burgess[1,3], Lucy Ort[1], Bogdan Budnik[2] & George M. Church[1,2]✉

The standard genetic code uses 64 codons to encode 20 canonical amino acids across domains of life. New-to-nature genetic codes enable new chemistries, therapeutics and ecosystem engineering, but recoding the genome of an organism is exceptionally challenging[1–6]. Here we describe automated genetic tRNA expansion (AGENTEX) for multiplexed robotic prototyping of genetic codes in cell-free translation systems. Two Watson–Crick interactions in the ribosomal large subunit (LSU) mediate recognition of the 3′ CCA end of tRNAs, preventing tRNAs with alternative 3′ sequences from being accommodated during translation[7,8]. Building on these interactions, we investigated the extent to which non-CCA-3′ tRNAs (otRNAs) would be aminoacylated by natural aminoacyl tRNA synthetases (aaRSs), allowing pools of otRNAs to specify unique genetic codes using ribosomes with altered LSU. We developed multiplexed and automated methods to read aminoacylation in libraries of synthetic tRNAs. We discovered that the tRNA 3′ end shows remarkable flexibility to mutation, allowing aminoacylation of most otRNAs by all *Escherichia coli* aaRSs. Building on our discovery, we developed cell-free translation systems enabling compressed genetic codes of 34 aaRSs for 34 codons. Using AGENTEX, we evaluated two genetic codes alongside the standard genetic code, with non-standard amino acid incorporation and reassignment of up to three codons. Our findings have implications for the design of radically new translation systems, the synthesis of biopolymers with several instances of non-standard monomers, and understanding of possible past and future genetic codes.

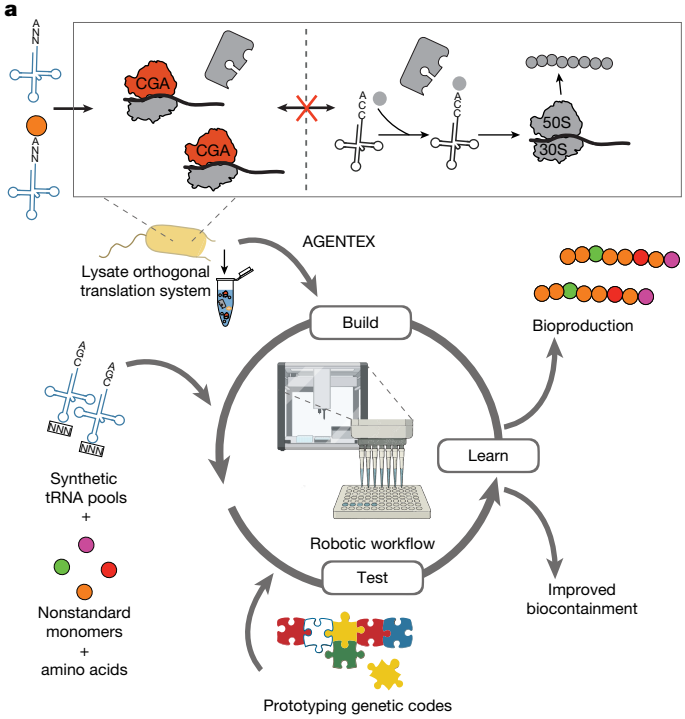
Synthetic genetic codes[1,2,4,5] have enabled construction of organisms with new properties such as increased virus resistance and reduced horizontal gene transfer[2,9,10], as well as production of genetically encoded materials[11,12], therapeutics[13] and biocatalysts[14]. Using engineered ribosomes[15–17], tRNAs[18] and aaRSs[18,19], genetic code expansion has enabled the introduction of more than 400 non-standard amino acids (nsAAs) into proteins in living cells[20] and synthesis of sequence-defined polymers[21–23]. However, reassigning more than a handful of sense codons across an entire genome[3,6] remains challenging, as does incorporation of multiple nsAAs into polypeptides with high purity. Much more radical redesign of translation systems is required to achieve complete control, from genetic sequence space to sequence-defined chemistry.

Given the challenges of genome synthesis and editing, a fully automated workflow to design and test new genetic codes can accelerate genome engineering, push genomes beyond what nature has evolved and allow unprecedented control over sequence-defined polymer synthesis (Fig. 1a). Compressing the 64-codon near-universal genetic code to only 20 sense codons, one start codon and one stop codon liberates 14 codons for non-standard building blocks (Fig. 1b and Extended Data Fig. 1). We predicted a modular, automated, data-driven redesign cycle. By combining custom lysate-based translation systems with synthetic pools of tRNAs carrying engineered anticodons, we can rapidly prototype alternative genetic codes and evaluate their performance in polypeptide synthesis (Fig. 1a).

Cell-free translation platforms are divided into lysate-derived (lysate) and PURExpress (PURE) systems. Lysate is produced from crude cell extracts and retains both translational machinery and enzymes for RNA and post-translational modifications[24]. PURE uses individually purified components, including ribosomes, aaRSs, tRNAs and translation factors, combined in defined ratios[25]. Lysate offers several advantages for genetic code prototyping, including support for native tRNA modifications that enhance codon–anticodon pairing[26,27], scalability for bioproduction[24,28,29] and similarity to in vivo systems[30]. However, lysates also contain the complete tRNA set of the host, which can exhibit crosstalk with reassigned sense codons and obscure synthetic genetic code designs. Consequently, proof-of-concept genetic code expansion studies and directed evolution have largely been performed in PURE[31,32], although PURE is less suitable for preparative-scale bioproduction.

To overcome this limitation, we proposed engineering the universal Watson–Crick interaction between the 3′ CCA sequence of all tRNAs and bases G2251 and G2553 in the 23S rRNA of the *Escherichia coli* ribosome[7,8], which ensures specific tRNA accommodation during translation. Purified ribosomes containing G2251C and G2553C mutations were previously shown in PURE to accommodate tRNAs with 3′ CGA ends for short peptide translation[33]. This, however, relied on flexizyme charging and purification of each unique tRNA species used[33,34], preventing multiplexed testing of entire genetic codes. We asked whether

[1]Department of Genetics, Harvard Medical School, Boston, MA, USA. [2]Wyss Institute for Biologically Inspired Engineering, Harvard University, Boston, MA, USA. [3]Present address: Weill Cornell Medicine Graduate School of Medical Sciences, New York, NY, USA. ✉e-mail: felix_radford@hms.harvard.edu; gchurch@genetics.med.harvard.edu



b

	U	C	A	G	
U	UUC F UUG 1	UCC S UCG 5	UAC Y UAA STOP UAG 9	UGC C UGA 10 UGG W	U C A G
C	CUC 2 CUG L	CCC 6 CCG P	CAC H CAG Q	CGC R CGG 11	U C A G
A	AUC I AUA 3 AUG M/START	ACC T ACG 7	AAC N AAA K	AGC 12 AGG 13	U C A G
G	GUC 4 GUA V	GCC 8 GCG A	GAC D GAA E	GGC G GGG 14	U C A G

Fig. 1 | AGENTEX enables multiplexed testing of genetic codes and sequence-defined biopolymer synthesis. a, Ribosomes containing G2251C and G2553C mutations in the large subunit (CGA ribosomes) specifically accommodate tRNAs with 3′ CGA ends (otRNAs) and do not accept natural tRNAs for translation. Natural aaRSs can aminoacylate all otRNAs in a one-pot reaction, allowing their use in translation. A cell-free translation system (lysate) is generated by lysing cells containing CGA ribosomes and used for subsequent translation tests in an automated workflow. Each genetic code is synthesized as a pool of otRNAs that are robotically produced and introduced into the lysate. For example, a single tRNA for each amino acid can be provided, producing a compressed genetic code. Similarly, mRNA templates for polypeptide translation are automatically synthesized and purified, and all reaction steps are automated. The otRNAs are continuously aminoacylated by aaRSs in the lysate and specifically interact with the CGA ribosome while avoiding crosstalk with the natural translation system. The effect of each genetic code, specified in the composition of an otRNA pool, can be analysed and used for rapid generation of the next genetic code experiments. The makeup of this system ensures biocontainment, as the components of AGENTEX are not functional if transferred to living organisms in the environment. **b**, Design of compressed genetic codes. Given 64 otRNA triplet codons and normal wobble rules, 20 canonical amino acids can each be decoded by a single tRNA and 14 codons available for reassignment, indicated in red. Alternatively, all 34 codons could be used for making peptides analogous to non-ribosomal peptide synthase using a mixture of canonical and other amino acids. Further details on codon choice and design of open codons are provided in Extended Data Fig. 1.

aaRSs could instead aminoacylate an entire tRNA pool in one reaction, thus enabling one-pot synthesis of biopolymers using engineered genetic codes. By supplementing engineered ribosomes with only tRNAs carrying non-CCA 3′ ends (otRNAs) into a lysate already containing the complete set of aaRSs, we predicted a seamless, high-throughput workflow for evaluating expanded genetic codes.

A longstanding assumption has held, however, that the 3′ CCA sequence is required for successful aminoacylation of most tRNAs[35–38] by aaRSs because of its proximity to the site of aminoacylation, and universality in tRNAs charged by both class I and Class II aaRSs. Earlier reports of in vitro aminoacylation assays using tRNAs with mutated CCA ends found that they could not be aminoacylated by purified aaRSs[39–41]. However, a previous study reported weak aminoacylation of in-vitro-transcribed $tRNA^{Val}$ carrying CUA or CAA 3′ ends by *E. coli* valyl-tRNA synthetase, although a CGA-ending $tRNA^{Val}$ was a poor substrate for aminoacylation[40]. We therefore proposed that other tRNAs might tolerate greater variation at their 3′ ends and sought to determine the extent of this flexibility. Understanding why the CCA sequence is universally conserved may also provide insights into the evolution of the translation system and the origins of life.

Here, we developed multiplexed, automated methods to quantify tRNA aminoacylation and applied them to study the full sequence variation of engineered pools of *E. coli* tRNAs with diversified 3′ terminal trinucleotides. Using cell-free translation, robotics, next-generation sequencing and analytical chemistry, we identified permissive 3′ tRNA sequences that function as substrates for all aaRSs and tested them in translation. We also measured aminoacylation of all *E. coli* tRNA isoacceptors in their unmodified state compared with modification conditions. Building on our discovery, we developed AGENTEX to construct and test genetic codes in a multiplexed, automated workflow while avoiding the crosstalk and laborious manipulations that limited previous approaches. Using AGENTEX, we built and tested minimal genetic codes for polypeptide translation and nsAA incorporation. This work provides insights into the evolution of the translation system and enables further evolution and expansion of the genetic code.

Development and optimization of tSCAN

As we sought to study the aminoacylation of otRNAs lacking the universally conserved 3′ CCA sequence, we anticipated challenges that would require new methods. tRNAs undergo the most extensive covalent modifications of any cellular RNA[42]. Moreover, tRNA biogenesis involves multi-step maturation in vivo, and it is challenging to produce intact tRNAs with non-natural architectures such as non-CCA ends in living cells[38,43]. As tRNA modifications mediate tRNA–aaRS interactions, they may affect aminoacylation, especially if the CCA end is altered.

For this reason, we required a multiplexed method to measure aminoacylation of synthetic tRNAs in a complex pool while avoiding tRNA maturation challenges. Furthermore, we sought to modify tRNAs on demand to causally determine whether the modification status of each tRNA affected its aminoacylation. To our knowledge, no such method was available. Previously, periodate tRNA sequencing was developed to quantify in vivo tRNA repertoires with next-generation sequencing (NGS)[44]. Instead of cloning libraries of tRNA genes into cells and lysing populations to sequence their tRNAs[44], we reasoned we could use lysate-based *E. coli* cell-free translation systems to recapitulate charging of in vitro produced tRNAs by the aaRS repertoire of the cell present in the lysate. We called this method tRNA sequencing of charging by automated NGS (tSCAN) (Fig. 2a and Supplementary Discussion 1).

We optimized tSCAN to use magnetic beads for all tRNA purification steps and automated the workflow on an open-source robotic platform, the Opentrons OT2 (Fig. 2b, Supplementary Discussion 1 and Supplementary Figs. 1–5), enabling a multiplexed pipeline to build and test non-natural tRNA designs. Aminoacylated tRNAs are protected from periodate oxidation and retain the CCA 3′ end, whereas uncharged

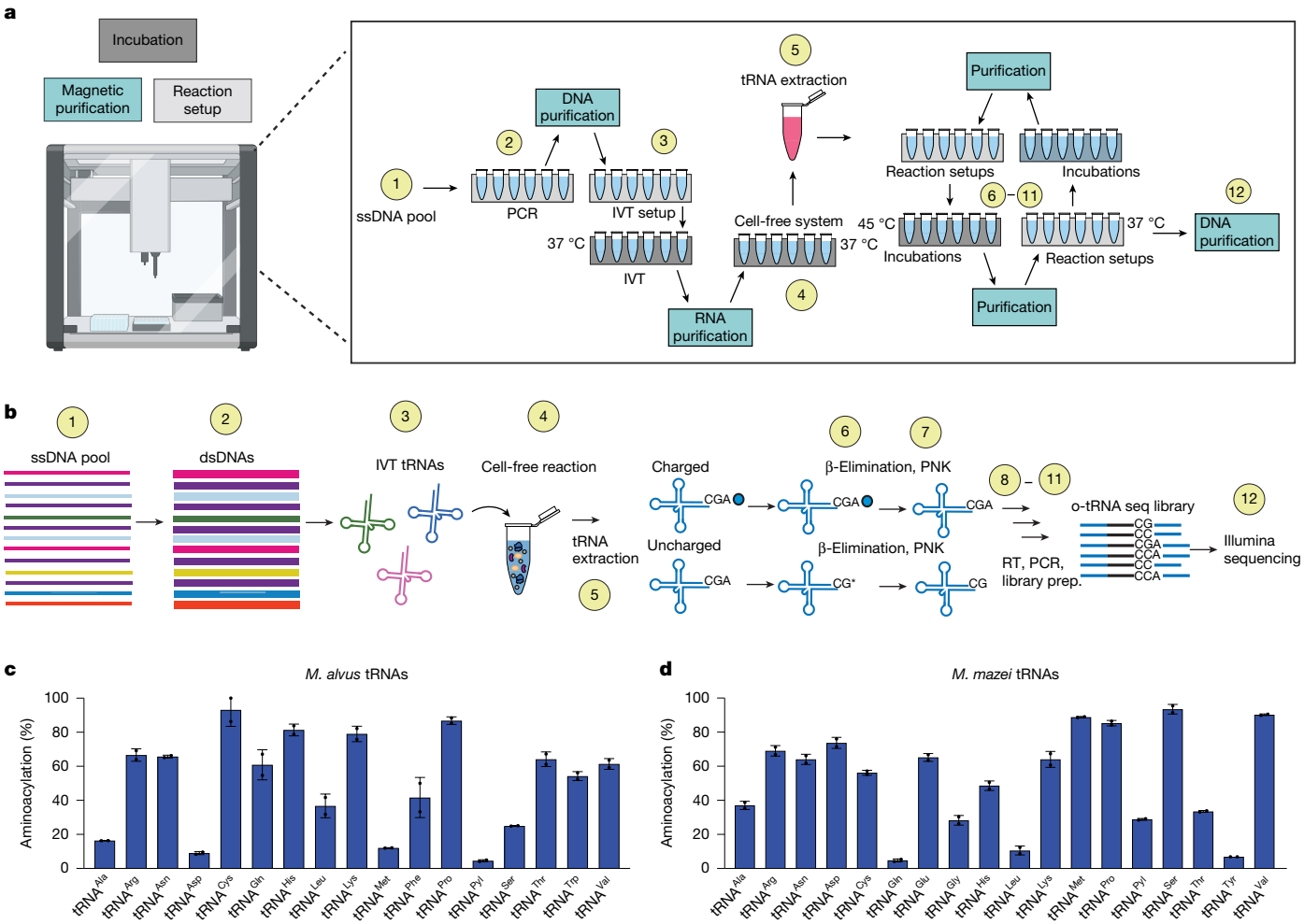


Fig. 2 | Automated reading of aminoacylation with tSCAN. a, The tSCAN procedure is divided into unit operations on the Opentrons OT-2 robotic system. All procedures consist of incubations, magnetic purifications or reaction setups. Each step of the procedure is inherited from one of these base programs (numbered as in **b**). **b**, The steps within the tSCAN protocol (Supplementary Discussion 1). Chip-synthesized oligonucleotide pools (1) are PCR-amplified into dsDNA (2) and in-vitro-transcribed into tRNAs (3). Cell-free reactions are incubated with the tRNAs (4), and the tRNA is extracted (5). Then, periodate oxidation, β-elimination (6) and dephosphorylation (7) are carried out. The RNA is purified at each step using magnetic beads. Then 3′ adapter ligation is performed, followed by reverse transcription and 5′ adapter ligation. The cDNA is PCR-amplified to add Illumina sequencing barcodes (8–11) (Supplementary Fig. 4), and the library is read by NGS (12). **c**,**d**, tSCAN was used to profile the aminoacylation of tRNAs from *M. alvus* (**c**) and *M. mazei* (**d**). We confirmed minimal charging of known orthogonal $tRNA^{Pyl}$ from both species and determined cross-reactivity of additional tRNAs from these organisms with the *E. coli* translation system. All data represent $n = 2$ independent replicates.

tRNAs lose the terminal adenosine. By quantifying ratios of intact to truncated reads for each species, we quantify aminoacylation. We examined orthogonality[22] of exogenous tRNAs to the *E. coli* system, because native *E. coli* tRNAs would be indistinguishable from wild-type (WT) tRNAs in lysate. We synthesized all tRNAs from *Methanomethylophilus alvus* and *Methanosarcina mazei*, two species used for genetic code expansion[45–47], as chip-synthesized oligo pools[48] and in-vitro-transcribed and purified them (Methods). After incubation in *E. coli* lysate, we found that as expected the $tRNA^{Pyl}$ from *M. alvus* and *M. mazei* showed low aminoacylation efficiency, consistent with minimal cross-reactivity with the *E. coli* system (Fig. 2c,d). Other tRNAs also showed low aminoacylation, including $tRNA^{Ala}$ in *M. alvus* (Fig. 2c) and $tRNA^{Leu}$ in *M. mazei* (Fig. 2d), although not previously reported as orthogonal and may represent additional orthogonal tRNAs in *E. coli*. Most of the *M. alvus* and *M. mazei* tRNAs were aminoacylated by the *E. coli* lysate (Fig. 2c,d).

All aaRSs aminoacylate non-CCA-tRNAs

After validation of tSCAN, we next tested a library of *E. coli* tRNAs in which the second base of the CCA sequence was degenerate (CNA) to determine whether *E. coli* synthetases aminoacylate tRNAs with this change. The extent of aaRS promiscuity to otRNAs was unknown, and knowledge of which aaRSs could accept otRNAs would be an important starting point for the development of AGENTEX. We produced tRNAs through in vitro transcription (IVT) having an A, U or G substitution at the penultimate position of the tRNA and incubated these tRNAs in an *E. coli* cell-free protein synthesis system (Methods) for 2 h at 37 °C. We observed that *E. coli* aaRSs can charge most non-CCA-tRNAs (Fig. 3 and Extended Data Fig. 2a–c), with CGA-tRNAs having the best overall charging levels. As previous work examining tRNA aminoacylation with non-CCA ends had been performed with in vitro purified aaRSs[40], we chose to benchmark our lysate-based assays using PURExpress Δ (aa, tRNA) kit, in which all components have been affinity-purified, and instead of a whole-cell lysate[25], native tRNAs are excluded. In contrast to the lysate-based cell-free translation system, we expected that no tRNA modifications would be added to the tRNAs produced through IVT in the PURE system. The IVT tRNA pool was incubated for 2 h at 37 °C. We observed much lower levels of aminoacylation by aaRSs for most of the tRNAs (Extended Data Fig. 3). Only about 62% of otRNAs with a CGA end were charged by *E. coli* aaRSs. Meanwhile, approximately 57% of CUA

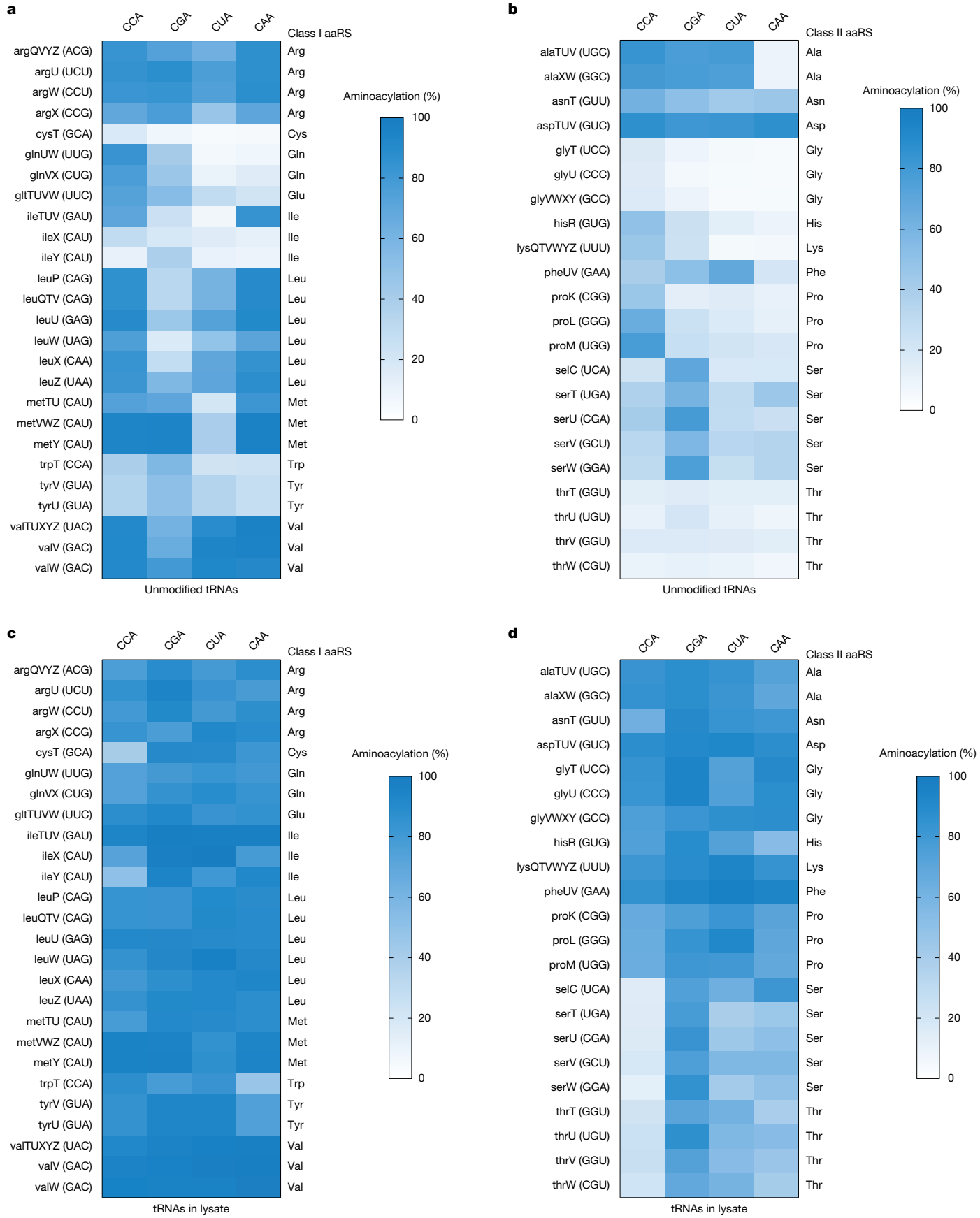


Fig. 3 | Analysis of aminoacylation levels of tRNAs containing CCA, CGA, CUA and CAA 3′ ends without or with RNA modifications. a–**d**, Aminoacylation data collected using tSCAN in PURExpress reconstituted cell-free translation system (**a**,**b**) or *E. coli* lysate-based translation system (**c**,**d**) was divided into tRNAs known to be aminoacylated by class I (**a**,**c**) or class II (**b**,**d**) aaRSs to compare aminoacylation levels by each aaRS class. All data represent $n = 2$ independent replicates. These data are also represented in Extended Data Figs. 2–4 as bar graphs.

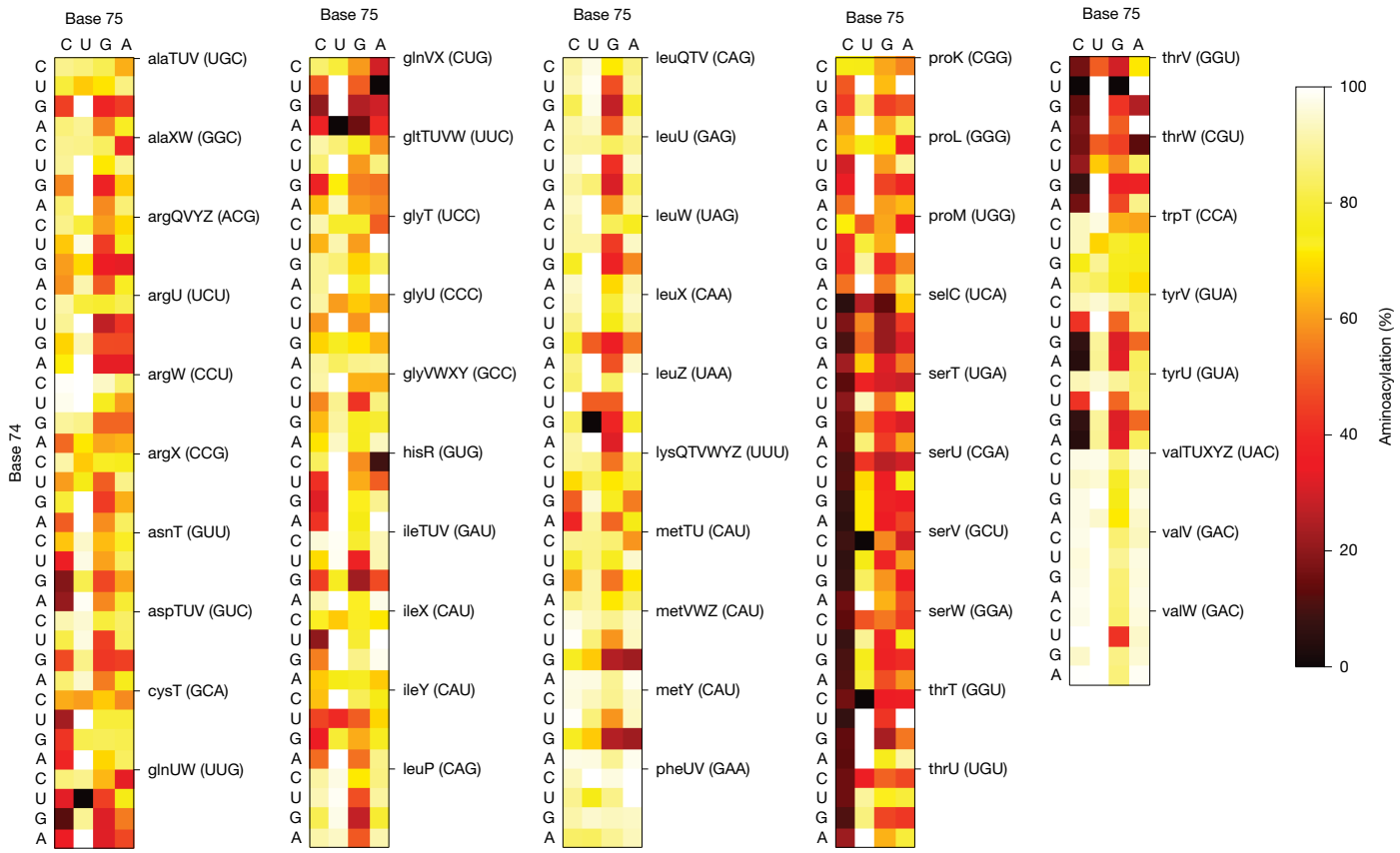


Fig. 4 | *E. coli* tRNAs with full variation of the CCA 3′ end are aminoacylated by natural *E. coli* aaRSs. A library of all *E. coli* isoacceptor tRNAs was synthesized as a ssDNA pool and amplified and transcribed into tRNAs. Bases 74 and 75 comprising the 3′ CCA sequence were degenerate in all tRNA isoacceptors. IVT tRNAs were added to the *E. coli* cell-free translation system, and aminoacylation was quantified with tSCAN. The aminoacylation percentage of all *E. coli* tRNAs was measured and plotted as a heatmap. The rows represent the identity of the nucleotide of the third-to-last position on the tRNA and the columns represent the identity of the nucleotide in the penultimate position on the tRNA. All data represent $n = 2$ independent replicates.

otRNAs were charged and around 42% of CAA otRNAs were charged (Fig. 3a,b and Extended Data Fig. 3a–c). The comparable otRNAs in the lysate-based system all showed much higher charging levels (Fig. 3c,d and Extended Data Fig. 2a–c).

Impact of tRNA modifications on charging

We proposed that differences between lysate and PURE translation systems arise from native tRNA modifications affecting tRNA–aaRS interactions, because RNA-modifying enzymes are present in lysate but absent in PURE. We produced IVT tRNAs for all *E. coli* natural tRNAs (3′ CCA) and compared aminoacylation in PURExpress Δ (aa, tRNA) and lysate systems under identical conditions (Fig. 3, Extended Data Fig. 4 and Supplementary Discussion 2), matching assays for CNA tRNAs. Aminoacylation levels in lysate agreed with previous measurements of *E. coli* isoacceptors, including low serine and threonine charging in vivo[49], supporting that tSCAN captures biologically meaningful differences. Average aminoacylation was 56% for CCA-tRNAs in PURE, 47% for CGA-tRNAs in PURE, and 67% for CCA-tRNAs in lysate (Extended Data Figs. 3 and 4). Although CCA-tRNAs were generally more highly aminoacylated than CGA-tRNAs in PURE (Extended Data Fig. 4a), several CCA-tRNAs showed reduced aminoacylation in PURE compared with lysate, including glyT and lysQ, which were about 50% lower in PURE (Extended Data Fig. 4b).

tRNAs charged by class II aaRSs generally showed lower aminoacylation than class I, especially serine and threonine tRNAs (Fig. 3b,d), consistent with previous reports[49] in *E. coli*. Altering the WT CCA end to CGA, CUA or CAA in lysate-incubated serine and threonine tRNAs increased aminoacylation (Fig. 3d), but this was not seen for most unmodified tRNAs (Fig. 3b). Overall, unmodified tRNAs showed reduced aminoacylation compared with lysate, particularly for class II aaRSs. Structural modelling (AlphaFold3) suggested steric hindrance differences between CCA and CGA ends, including π–π stacking between A75 of tRNA glyT and Y80 of GlyRS (Extended Data Fig. 7 and Supplementary Discussion 2).

Collectively, lysate improved aminoacylation of most tRNAs. Given that aaRS concentration in PURE is more than 1,000-fold lower than tRNA concentration[50] and comparable to lysate ratios[51], these differences probably reflect meaningful system-specific aaRS–tRNA interactions (that is, lysate compared with PURE). To further validate tSCAN aminoacylation patterns in lysate and PURE, we analysed aminoacylation with liquid chromatography–mass spectrometry (LC–MS) (Supplementary Discussion 2). We performed in vitro aminoacylation reactions on purified aaRS enzymes and IVT tRNAs, followed by detection of the aminoacylated amino acid by LC–MS (Methods, Extended Data Fig. 5a, top, and Supplementary Discussion 2). Analysis of leuW, proK and valT tRNAs (with 3′ CCA or 3′ CGA ends) with their cognate aaRSs matched the tRNA aminoacylation patterns in PURE, as IVT tRNAs were unmodified in both experimental systems. We observed aminoacylation of all tRNAs with 3′ CCA ends, but only valT was aminoacylated among tRNAs with 3′ CGA ends (Extended Data Fig. 5b), corresponding to tSCAN data (Fig. 3a,b).

To examine the identity of amino acids charged to tRNAs directly in the lysate, we required a new method that would allow incubation of IVT tRNAs in lysate and distinguish tRNAs differing by even a single nucleotide (that is, 3′ CGA or CCA ends). We developed tRNA-species

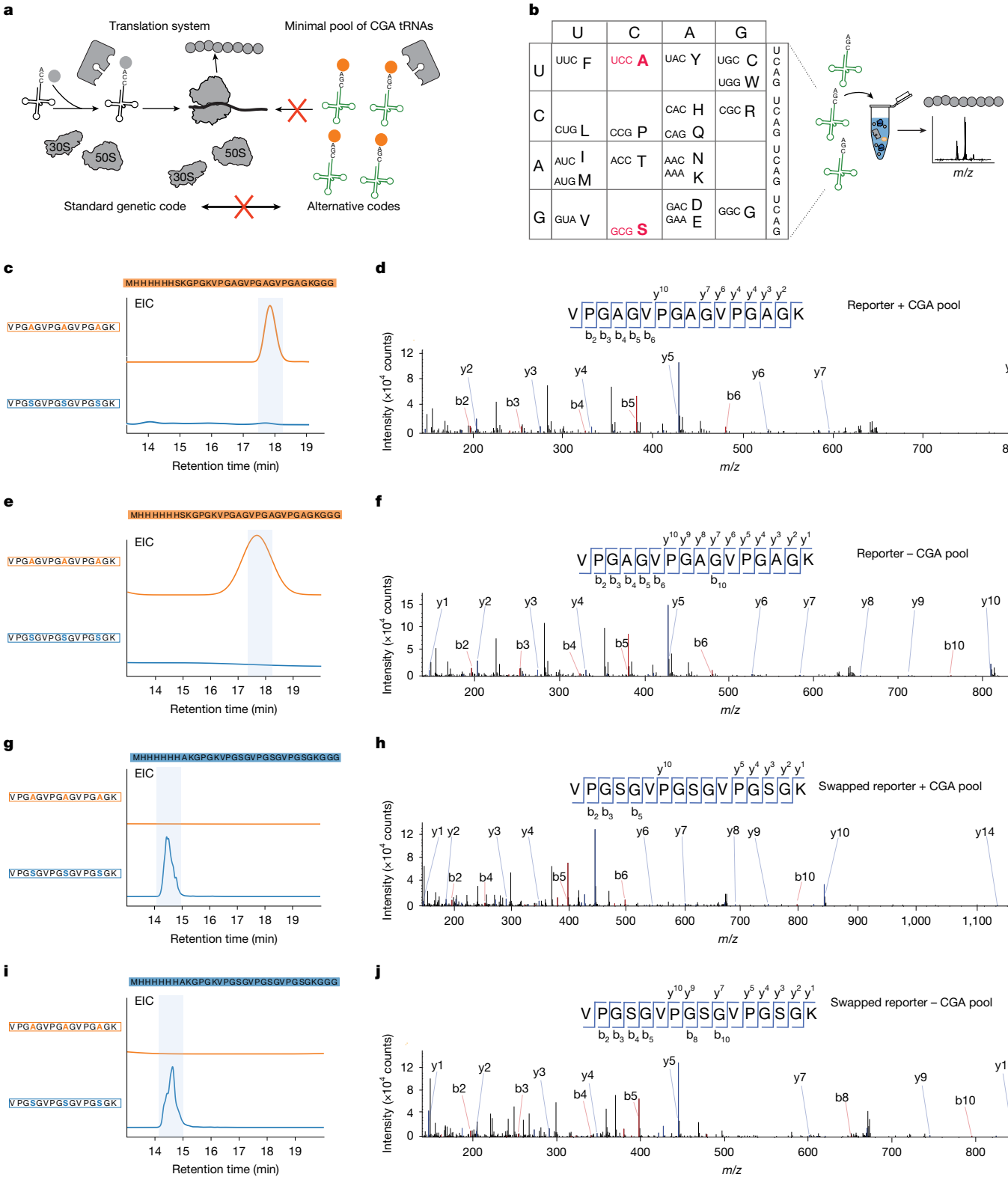


Fig. 5 | See next page for caption.

charging analysis by mass (tSCAN-M), which uses isotopically labelled tRNAs incubated in lysate and subsequent digestion, isolation and LC–MS detection of the charged amino acid species (Extended Data Fig. 6a and Supplementary Discussion 2). We first validated tSCAN-M by in vitro transcribing natural Tyr and Trp tRNAs using N^{15}-ATP during the IVT reaction and incubating them in lysate for aminoacylation by endogenous aaRSs. After bulk digestion and precipitation of proteins, we confirmed both endogenous and IVT Tyr and Trp tRNAs were aminoacylated with their target amino acids (Extended Data Fig. 6b and Supplementary Discussion 2). Using tSCAN-M, we analysed leuW, proK and valT tRNAs with 3′ CCA or CGA ends in lysate. We found that all of these tRNAs were aminoacylated with both 3′ end sequences (Extended Data Fig. 5c), again confirming our findings with tSCAN in lysate (Fig. 3c,d).

Fig. 5 | The ribosome and 3′ end of tRNAs prevent crosstalk between genetic codes. a, We propose that the ribosome, and not aaRSs, maintains tight control over the genetic code by interactions with the tRNA 3′ end, given that aaRSs can readily acylate both CCA and non-CCA-tRNAs. This prevents alternative genetic codes that could operate with tRNAs having non-CCA 3′ sequences. **b**, To determine whether CGA-tRNAs can still be accommodated at low levels by WT ribosomes, we used a cell-free translation system with WT ribosomes and introduced dsDNA templates for reporter peptides that can be identified by LC–MS/MS. We performed IVT of a library of CGA-otRNAs representing a minimal code of all 20 amino acids and methionine initiator (MP1) and introduced the MP1 pool to a cell-free translation system also containing *E. coli* native-charged tRNAs and *E. coli* aaRSs. The alanine and serine anticodons were swapped so that we could determine whether a CGA-tRNA was incorporated at an alanine codon in the reporter peptide instead of the native serine tRNA. **c**–**j**, In cell-free translation reactions, we used dsDNA templates for peptide MHHHHHHSKVPGAGVPGAGVPGAGKGGG (ELP-AAA) or MHHHHHHAKVPGSG VPGSGVPGSGKGGG (ELP-SSS), in which the trypsin-digested product, VPGSG VPGSGVPGSGK or VPGAGVPGAGVPGAGK, respectively, can be robustly detected by LC–MS/MS[10]. The elution times of the respective peptides is indicated with shading. The ELP-AAA peptide was detected and confirmed by LC–MS/MS using a Q Exactive Orbitrap mass spectrometer both in the presence (**c**,**d**) or absence (**e**,**f**) of the MP1 pool. In neither case was there evidence of misincorporation of otRNAs. Similarly, the ELP-SSS reporter peptide was detected and confirmed by LC–MS/MS both in the presence (**g**,**h**) or absence (**i**,**j**) of the MP1 pool. In neither case was there evidence of misincorporation of otRNAs.

Landscape of acceptor stem charging

Following from our surprising discovery of the flexibility of *E. coli* aaRSs to charge otRNAs having CGA, CUA and CAA ends, we next asked whether the position two bases away from the terminal adenosine could also be diversified and be tolerated by aaRSs as a substrate for aminoacylation. We synthesized an oligonucleotide library of all *E. coli* isoacceptor tRNAs with the two terminal bases, 74 and 75, before the final adenosine was degenerate (Fig. 4a). We performed tSCAN within a lysate-based *E. coli* translation system. We plotted the analysis of the tSCAN results as a heatmap in which the rows correspond to the identity of base 74 and the columns correspond to the identity of base 75 in each terminal tRNA sequence. Many positions were found to be permissive to substitutions (Fig. 4b), such as A or U being the first base of the 3′ sequence. Also, some tRNAs exhibited a generally larger flexibility to substitution, such as valT, valV, valU and pheV. GCA exhibited reduced aminoacylation in about 60% of tRNAs, in contrast to other terminal sequences that were mostly permitted. This is intriguing, as it was previously shown[33] that the WT *E. coli* ribosome can accept a few tRNAs with 3′ GCA end that were charged with an amino acid using flexizymes but with less efficiency than natural tRNAs with 3′ CCA end. However, accommodation of tRNAs with 3′ CGA ends by WT ribosomes was not observed[33]. This difference between GCA and CGA 3′ ends may suggest translation regulation of non-CCA-tRNAs at the aminoacylation stage if ribosome selection is not effective.

Ribosomes prevent translation crosstalk

Having determined that there is wide flexibility for charging otRNAs, we next investigated if otRNAs could participate in translation in the native translation system (that is, using natural ribosomes without G2251C and G2553C mutations in the 23S rRNA) (Supplementary Discussion 3). We proposed that the WT ribosome, by interactions with the tRNA acceptor stem, would selectively accept tRNAs only with CCA ends, even if otRNAs aminoacylated by *E. coli* aaRSs are present (Fig. 5a). We began by testing translation using a compressed genetic code with two swapped codons to measure crosstalk with the natural genetic code. Based on our tSCAN data, we synthesized an IVT pool of otRNAs with 3′ CGA ends, including 20 tRNAs for sense codons and one initiator methionine tRNA (Fig. 5b and Extended Data Fig. 8a). Anticodons of alanine and serine were swapped (pool MP1) to determine whether a native tRNA or otRNA was used in decoding a reporter peptide messenger RNA (mRNA). We designed and validated translation of peptide reporters (ELP-AAA and ELP-SSS) that can be detected by LC–MS/MS[10] to determine whether alanine or serine was incorporated at three positions within the peptide (Supplementary Fig. 9a,b and Supplementary Discussion 3).

Next, we tested translation in the presence or absence of the swapped compressed genetic code MP1 to determine the levels of accommodation of charged CGA-otRNAs by WT *E. coli* ribosomes (Supplementary Discussion 3). With translation carried out using dsDNA for the ELP-AAA peptide as a template, we detected ELP-AAA (Fig. 5c,d). We saw equivalent results in the no MP1 pool addition control (Fig. 5e, f). With translation carried out using dsDNA for the swapped ELP-SSS reporter template, we detected ELP-SSS but not ELP-AAA or any alanine substitutions into the peptide (Fig. 5g,h). We saw equivalent results in the no MP1 pool addition control (Fig. 5i,j). In summary, we could not detect evidence of non-specific accommodation of otRNAs containing a CGA end into the WT ribosome for translation. Thus, although otRNAs are aminoacylated by all *E. coli* native aaRSs, the WT ribosome acts as a filter to prevent otRNA accommodation during translation. These findings were highly encouraging for us to proceed with testing AGENTEX with the addition of a mutant ribosome into a lysate system, which would specifically interact with otRNAs while not having crosstalk with the standard genetic code.

Compressed genetic codes with AGENTEX

As we found the otRNAs do not have detectable crosstalk with the native translation system, we next deployed AGENTEX to test a compressed genetic code and apply it to translate polypeptides with an expanded genetic alphabet. We tested a 22-codon genetic code for polypeptide translation with AGENTEX. We synthesized a minimal tRNA pool (MP2) with otRNAs containing 3′ CGA ends, consisting of an initiator methionine otRNA and 20 otRNAs for all canonical amino acids (Fig. 6a and Supplementary Discussion 3). The codons of serine and alanine were swapped. The *Methanocaldococcus jannaschii* $tRNA^{Tyr}_{CUA}$ with CGA 3′ end was included to decode UAG and enable para-azido phenylalanine (pazF) incorporation by its aaRS[52] (Extended Data Fig. 8b).

As before, we designed automated procedures for AGENTEX with custom Python code controlling all unit operations in the OT-2 robot and that separated data representation from robotic execution (Supplementary Fig. 10 and Supplementary Discussion 3). Compressed genetic codes were generated to specify a minimal set of codons for the canonical amino acids and the ability to modify anticodons of tRNAs to maximize codon space.

We developed a high-sensitivity luminescence reporter assay for rapid testing of translation of recoded genetic codes with AGENTEX. Our reporter is based on the HiBiT peptide, which with LgBiT reconstitutes luciferase function, enabling quantitative measurement of its production by luminescence (Fig. 6b and Supplementary Discussion 3). We constructed an engineered peptide (h-TAG-1), in which a short peptide (MGVGXGVG, where X is a UAG codon) precedes the HiBiT sequence. We tested translation using AGENTEX and a 22-codon code using the MP2 tRNA pool (Extended Data Fig. 8b). Successful translation produces a full HiBiT peptide and luminescence on LgBiT complementation. If MP2 tRNAs are not accommodated, translation terminates early because of lack of suppression by *M. jannaschii* $otRNA^{Tyr}_{CUA}$, and HiBiT contains two incorrect alanines instead of serines. We performed translation with or without pazF in lysates containing WT or CGA ribosomes. With CGA ribosomes, there was approximately six-fold higher HiBiT production with MP2 and pazF (+pazF, +pazFRS) than WT lysate, and about 10-fold higher than the no-template control (Fig. 6c). This suggests correct message translation is more successful with CGA ribosomes and MP2

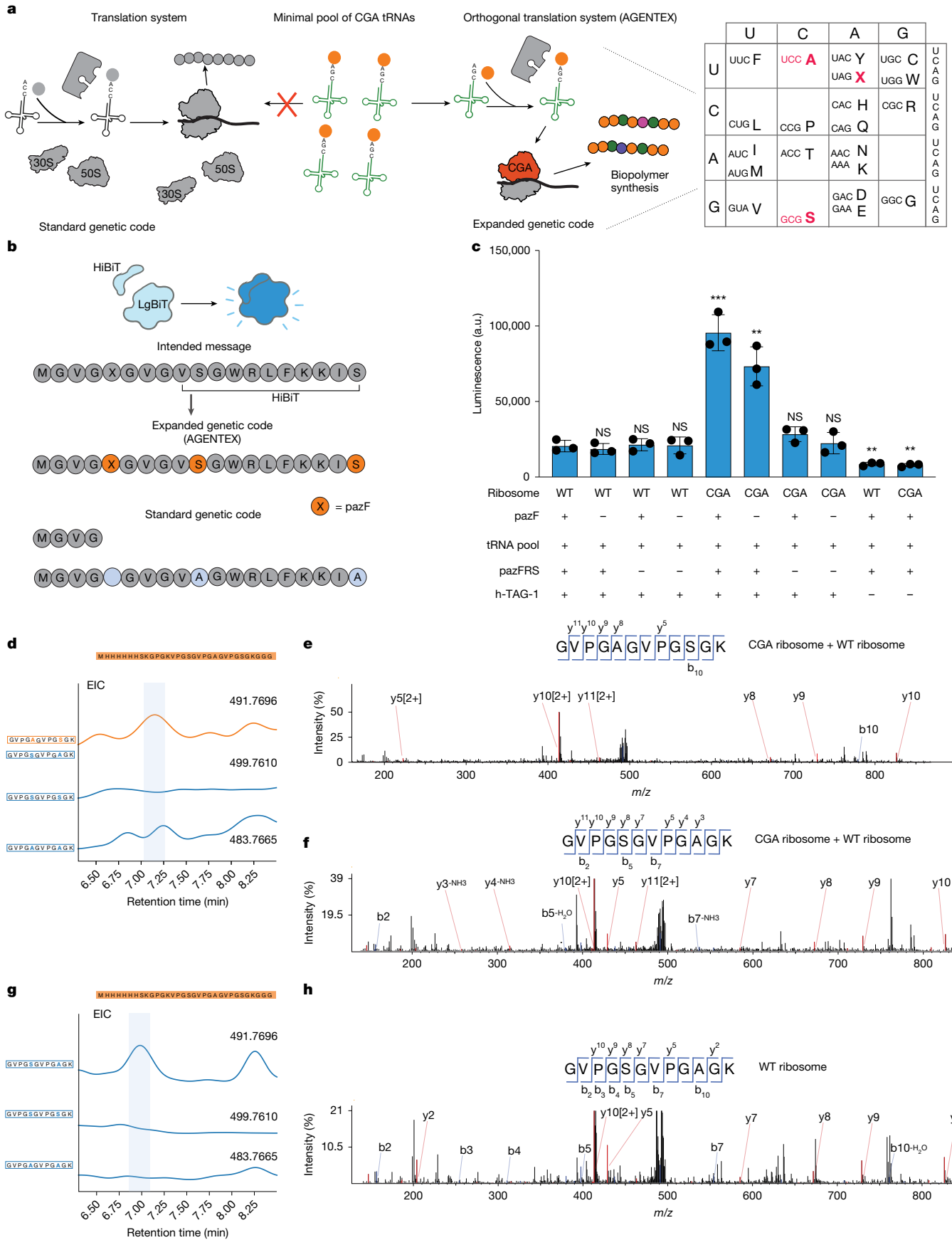


Fig. 6 | See next page for caption.

Fig. 6 | Translation of polypeptides with an expanded genetic code using sequential decoding of otRNAs and mutant ribosome. a, We tested translation of compressed genetic code MP2 with AGENTEX using mutant CGA ribosomes (G2251C/G2553C in 23S rRNA) and an MP2 orthogonal tRNA (otRNA) pool. The compressed genetic code should be specifically translated by CGA ribosomes and is orthogonal to the natural translation system. The MP2 tRNA pool decodes each canonical amino acid with a single tRNA, with serine and alanine anticodons swapped. *M. jannaschii* $tRNA^{Tyr}_{CUA}$ decodes the UAG codon with para-azido phenylalanine (pazF). **b**, Translation efficiency and fidelity were evaluated using the h-TAG-1 reporter, which contains a TAG codon for pazF and swapped Ser/Ala codons. Standard-code translation produces truncated or mistranslated h-TAG-1 with reduced luminescence, whereas the compressed code yields full-length h-TAG-1, which reconstitutes luciferase luminescence by full-length HiBiT (Supplementary Discussion 3). **c**, h-TAG-1 was translated in AGENTEX using WT or CGA ribosome lysates, with or without pazF, pazFRS and the reporter template. Luminescence was measured in an endpoint assay. Data represent $n = 3$ independent replicates (mean ± s.d.). A Student's *t*-test was used to calculate statistical significance relative to the WT ribosome condition and presence of all variables (leftmost bar), where NS denotes not significant, $P > 0.05$, $^{*}P < 0.05$, $^{**}P < 0.01$, $^{***}P < 0.001$ and $^{****}P < 0.0001$. **d**–**h**, To verify peptide products and assess translational crosstalk, we used the ELP-SAS reporter, containing three codons whose identities depend on the genetic code and are readily distinguished by LC–MS/MS (Supplementary Fig. 12). With CGA ribosomes and the MP1 pool, both swapped and standard-code products were detected, but no mixed alanine/serine crosstalk products; MS/MS confirmed swapped (**e**) and standard-code (**f**) products. With WT ribosomes, only the standard-code product was detected (**g**), confirmed by MS/MS (**h**).a.u., arbitrary units.

tRNAs (Supplementary Discussion 3). When translating in WT and CGA lysates without functional pazFRS (+pazF, −pazFRS), luminescence was at background levels (Fig. 6c), indicating the high signal (+pazF, +pazFRS) with CGA ribosomes reflects correct decoding (about six-fold over background, $P < 0.001$).

To determine whether otRNAs were accommodated by CGA ribosomes and whether there was any crosstalk with native ribosomes, we used LC–MS/MS for peptide detection (Supplementary Discussion 3). We used a ZenoTOF7600+ mass spectrometer with femtomolar sensitivity, up to 100-fold higher than Q Exactive Orbitrap[53], to improve detection of rare crosstalk events (Methods). We tested AGENTEX using a 21-codon code with MP1 tRNA pool (Extended Data Fig. 8a) for production of peptide ELP-SAS (Fig. 6d–h and Supplementary Fig. 12) in WT compared with CGA lysates. We used ELP-SAS because it has three codons whose amino acid identity would depend on which genetic code was used for its translation and which is robustly identified by LC–MS/MS. As ELP-SAS is 32 amino acids in length, it would also be a test of the ability for consecutive decoding of a peptide of this length solely by otRNAs (Supplementary Discussion 3). We monitored GVPG**J**GVPG**J**GK across all samples, where J = A or S. The extracted ion chromatograms (EICs) for the expected masses of the peptide (Extended Data Table 1) were analysed and from this generated MS/MS spectra. In CGA lysates, both GVPG**A**GVPG**S**GK (expanded genetic code) and GVPG**S**GVPG**A**GK (standard genetic code) were detected and confirmed (Fig. 6d–f), which would be expected as two translation systems are present in this sample, with higher abundance of GVPG**S**GVPG**A**GK (Supplementary Fig. 13). We did not detect GVPG**S**GVPG**S**GK or GVPG**A**GVPG**A**GK, indicating no crosstalk (Fig. 6d). A non-specific eluent with mass matching GVPG**A**GVPG**S**GK (Extended Data Table 1) was not confirmed by MS/MS and was a false-positive hit. In analysis of all peptides matching our library search and manual inspection of MS/MS spectra, no other alanine/serine substitutions were found, indicating no detectable crosstalk. In WT lysates, only GVPG**S**GVPG**A**GK was detected (Fig. 6g,h), which matched our previous analysis of WT ribosomes in the presence of otRNAs (Fig. 5). We did not detect any other peptides that would indicate translation crosstalk (Fig. 6g). These findings are in agreement with the increased HiBiT activity observed with CGA ribosomes in the presence of the MP2 otRNA pool. Collectively, these data demonstrate translation of peptides with expanded genetic alphabet and two compressed genetic codes. When PTC mutations in the ribosome, such as G2251C and G2553C exist, otRNAs are aminoacylated by natural aaRSs and are selectively and efficiently accommodated into polypeptides.

Discussion

We developed AGENTEX and demonstrated its abilities for the testing of multiple genetic codes for polypeptide production with an expanded genetic alphabet, using a robotic workflow from dsDNA assembly and tRNA production to polypeptide translation and purification. We determined that we could avoid detectable crosstalk with the native translation system, and prototyped multiple compressed genetic codes. Furthermore, we developed new methods, tSCAN and tSCAN-M, to quantify aminoacylation of synthetic pools of tRNAs in cell-free translation systems and identify their esterified amino acids. In contrast to previous work, with tSCAN we can rapidly test multiple exogenous tRNA designs using laboratory automation, with control over translation system composition, and determine the identity of amino acids charged to tRNA. Using this technology, we studied orthogonal tRNAs from *M. alvus* and *M. mazei*, characterized the biochemical flexibility of one of the most conserved sequences in nature, the CCA end of tRNA, and investigated aminoacylation of all *E. coli* isoacceptor tRNAs in conditions in which RNA modifications were absent or could occur. tRNAs are central to the biology of all organisms, and mutations to tRNAs can cause diseases in humans such as hearing loss and diabetes mellitus[54]. Many studies of tRNAs and construction of artificial genetic codes could benefit from the ability to read tRNA aminoacylation alongside tRNA modification status and identity of esterified amino acids.

Our results indicate that the tRNA CCA end, despite its essential role in tRNA function[35–38], can exhibit remarkable flexibility in its sequence without compromising its ability to interact with aaRSs. By comprehensively varying the first two bases of the 3′ CCA end of all *E. coli* tRNAs and testing their aminoacylation efficiency, we determined which sequence patterns were tolerated by aaRSs. Many non-CCA-tRNA variants retain substantial aminoacylation efficiency, suggesting that the CCA sequence is more tolerant to mutation than previously anticipated. Moreover, this flexibility may have been underexplored because previous studies examined aminoacylation using purified aaRSs and tRNAs in vitro, lacking native RNA modifications. We show that the full cellular context (that is, in lysate), not only purified translation system components (that is, in PURE or in vitro aminoacylation reactions), is necessary for aminoacylation of most tRNAs with diverging 3′ ends, which we propose reflects the requirement for proper modifications[42,55,56]. Our comparison of CCA-tRNA charging in lysate and PURE system conditions (Fig. 3 and Extended Data Fig. 4) highlights that even CCA-tRNAs are preferentially aminoacylated under native conditions. This pattern appears more pronounced for tRNAs with diverging 3′ ends. We also identified using purified aaRSs in the PURE system that only a portion of aaRSs charged non-CCA-tRNAs, suggesting that tRNA modifications modulate recognition and aminoacylation (Supplementary Discussion 4).

Our findings further help to revise the view of the translation system as a linear progression from genome to proteome in genetic code maintenance (Extended Data Fig. 9a) to a multilayered system in which the genetic code is actively maintained by the translation apparatus (Extended Data Fig. 9b and Supplementary Discussion 4). Although aaRSs can aminoacylate tRNAs with altered 3′ CCA sequences, the ribosome serves as the primary filter, and ribosomes with matching mutations can accommodate these variant tRNAs, raising the possibility of parallel genetic codes within a single cell. This feature

may originate before the last universal common ancestor, in an RNA world preceding DNA genomes[57] or aaRS proteins[58,59] (Supplementary Discussion 4). We also found that tRNA modifications influence aaRS recognition by class I and class II aaRSs, with altered aminoacylation becoming more pronounced when the CCA sequence is mutated (Extended Data Fig. 9b). These results indicate that tRNA modifications can provide a second layer of regulation that helps preserve the standard genetic code. Together, we propose that coordinated changes in ribosomes, tRNAs and their modification status could enable alternative genetic codes, but that strong selective pressures normally prevent their emergence.

On the basis of these discoveries, we successfully tested compression of the 64-codon standard genetic code into a reduced genetic code specified by 22 tRNAs using AGENTEX (21 tRNAs with sense anticodons and one tRNA with a stop anticodon), allowing a theoretical genetic code of 20 canonical amino acids and 14 open codons (34-codon genetic code). Using this code, we translated polypeptides with an nsAA and three reassigned codons. This work exemplifies how AGENTEX enables rapid exploration of genetic codes not found in nature (Supplementary Discussion 5). Moreover, we demonstrate two genetic codes operating in parallel in the same translation system: our compressed genetic code and the standard genetic code. We anticipate that these insights will enable construction of in vivo genetic codes with up to 14 open codons for non-standard monomer incorporation.

Collectively, this work provides a foundation for rapid testing of genetic codes far removed from their natural counterparts. Future applications of this work will facilitate the construction of new organisms with both recoded genomes and in vivo parallel genetic codes (relying on orthogonal ribosomes, tRNAs, aaRSs and mRNAs). Our technology presents a generalized solution for evaluating the full dynamics of a genetic code in vitro at the level of mRNA, ribosomes, tRNAs and polypeptides, and with biocontainment built in (Supplementary Discussion 5). The cell-free nature of this technology enables rapid and safe testing of highly divergent genetic codes. Many dynamics of genetic codes are consequential at this level even without the context of a full genome. Future developments to this work may expand the mechanistic assays that can be used in vitro when genetic codes are evaluated with AGENTEX, including changes in mRNA secondary structure of multiple genes translated in parallel, variation of coding sequences, ribosomal occupancy during translation, translational rates and translation accuracy under diverse genetic codes. However, we also recognize that construction of new genomes entails higher-order interactions beyond genetic code considerations, such as chromosome organization and entire metabolisms[60]. Similar to the rapid design–build–test cycle used here, we anticipate that AGENTEX will efficiently evaluate many new genetic code designs before strain construction, substantially derisking future genome engineering efforts. These abilities will be instrumental for constructing organisms with radically divergent properties from extant life and facilitating cellular synthesis of biopolymers composed of non-standard monomers.

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/s41586-026-10949-y.

1. Gibson, D. G. et al. Complete chemical synthesis, assembly, and cloning of a *Mycoplasma genitalium* genome. *Science* **319**, 1215–1220 (2008).
2. Lajoie, M. J. et al. Genomically recoded organisms expand biological functions. *Science* **342**, 357–360 (2013).
3. Ostrov, N. et al. Design, synthesis, and testing toward a 57-codon genome. *Science* **353**, 819–822 (2016).
4. Fredens, J. et al. Total synthesis of *Escherichia coli* with a recoded genome. *Nature* **569**, 514–518 (2019).
5. Grome, M. W. et al. Engineering a genomically recoded organism with one stop codon. *Nature* **639**, 512–521 (2025).
6. Robertson, W. E. et al. *Escherichia coli* with a 57-codon genetic code. *Science* **390**, eady4368 (2025).
7. Moazed, D. & Noller, H. F. Sites of interaction of the CCA end of peptidyl-tRNA with 23S rRNA. *Proc. Natl Acad. Sci. USA* **88**, 3725–3728 (1991).
8. Berk, V., Zhang, W., Pai, R. D. & Cate, J. H. D. Structural basis for mRNA and tRNA positioning on the ribosome. *Proc. Natl Acad. Sci. USA* **103**, 15830–15834 (2006).
9. Mandell, D. J. et al. Biocontainment of genetically modified organisms by synthetic protein design. *Nature* **518**, 55–60 (2015).
10. Nyerges, A. et al. A swapped genetic code prevents viral infections and gene transfer. *Nature* **615**, 720–727 (2023).
11. Israeli, B., Vaserman, L. & Amiram, M. Multi-site incorporation of nonstandard amino acids into protein-based biomaterials. *Isr. J. Chem.* **60**, 1118–1128 (2020).
12. Amiram, M. et al. Evolution of translation machinery in recoded bacteria enables multi-site incorporation of nonstandard amino acids. *Nat. Biotechnol.* **33**, 1272–1279 (2015).
13. Rezhdo, A., Islam, M., Huang, M. & Van Deventer, J. A. Future prospects for noncanonical amino acids in biological therapeutics. *Curr. Opin. Biotechnol.* **60**, 168–178 (2019).
14. Drienovská, I. & Roelfes, G. Expanding the enzyme universe with genetically encoded unnatural amino acids. *Nat. Catal.* **3**, 193–202 (2020).
15. Orelle, C. et al. Protein synthesis by ribosomes with tethered subunits. *Nature* **524**, 119–124 (2015).
16. Schmied, W. H. et al. Controlling orthogonal ribosome subunit interactions enables evolution of new function. *Nature* **564**, 444–448 (2018).
17. Radford, F., Elliott, S. D., Schepartz, A. & Isaacs, F. J. Targeted editing and evolution of engineered ribosomes in vivo by filtered editing. *Nat. Commun.* **13**, 180 (2022).
18. Park, H.-S. et al. Expanding the genetic code of *Escherichia coli* with phosphoserine. *Science* **333**, 1151–1154 (2011).
19. Kunjapur, A. M. et al. Engineering posttranslational proofreading to discriminate nonstandard amino acids. *Proc. Natl Acad. Sci. USA* **115**, 619–624 (2018).
20. Icking, L.-S. et al. iNClusive: a database collecting useful information on non-canonical amino acids and their incorporation into proteins for easier genetic code expansion implementation. *Nucleic Acids Res.* **52**, D476–D482 (2024).
21. Connor, R. E. & Tirrell, D. A. Non-canonical amino acids in protein polymer design. *J. Macromol. Sci. C Polym. Rev.* **47**, 9–28 (2007).
22. Chin, J. W. Expanding and reprogramming the genetic code. *Nature* **550**, 53–60 (2017).
23. Dunkelmann, D. L. et al. Adding α,α-disubstituted and β-linked monomers to the genetic code of an organism. *Nature* **625**, 603–610 (2024).
24. Hunt, A. C. et al. Cell-free gene expression: methods and applications. *Chem. Rev.* **125**, 91–149 (2024).
25. Tuckey, C., Asahara, H., Zhou, Y. & Chong, S. Protein synthesis using a reconstituted cell-free system. *Curr. Protoc. Mol. Biol.* **108**, 16.31.1–16.31.22 (2014).
26. Fujino, T. et al. Ser/Leu-swapped cell-free translation system constructed with natural/in vitro transcribed-hybrid tRNA set. *Nat. Commun.* **15**, 4143 (2024).
27. Agris, P. F., Narendran, A., Sarachan, K., Väre, V. Y. P. & Eruysal, E. The importance of being modified: the role of RNA modifications in translational fidelity. *Enzymes* **41**, 1–50 (2017).
28. Rasor, B. J. et al. Toward sustainable, cell-free biomanufacturing. *Curr. Opin. Biotechnol.* **69**, 136–144 (2021).
29. Didovyk, A., Tonooka, T., Tsimring, L. & Hasty, J. Rapid and scalable preparation of bacterial lysates for cell-free gene expression. *ACS Synth. Biol.* **6**, 2198–2208 (2017).
30. Moore, S. J., MacDonald, J. T. & Freemont, P. S. Cell-free synthetic biology for in vitro prototype engineering. *Biochem. Soc. Trans.* **45**, 785–791 (2017).
31. Hibi, K. et al. Reconstituted cell-free protein synthesis using in vitro transcribed tRNAs. *Commun. Biol.* **3**, 350 (2020).
32. Goto, Y. & Suga, H. The RaPID platform for the discovery of pseudo-natural macrocyclic peptides. *Acc. Chem. Res.* **54**, 3604–3617 (2021).
33. Terasaka, N., Hayashi, G., Katoh, T. & Suga, H. An orthogonal ribosome-tRNA pair via engineering of the peptidyl transferase center. *Nat. Chem. Biol.* **10**, 555–557 (2014).
34. Goto, Y., Katoh, T. & Suga, H. Flexizymes for genetic code reprogramming. *Nat. Protoc.* **6**, 779–790 (2011).
35. Zhou, X.-L. et al. Role of tRNA amino acid-accepting end in aminoacylation and its quality control. *Nucleic Acids Res.* **39**, 8857–8868 (2011).
36. Chládek, S. & Sprinzl, M. The 3′-end of tRNA and its role in protein biosynthesis. *Angew. Chem. Int. Ed. Engl.* **24**, 371–391 (1985).
37. Sprinzl, M. & Cramer, F. in *Progress in Nucleic Acid Research and Molecular Biology* Vol. 22 (ed. Cohn, W. E.) 1–69 (Academic Press, 1979).
38. Wellner, K., Betat, H. & Mörl, M. A tRNA's fate is decided at its 3′ end: collaborative actions of CCA-adding enzyme and RNases involved in tRNA processing and degradation. *Biochim. Biophys. Acta Gene Regul. Mech.* **1861**, 433–441 (2018).
39. Schulman, L. H. & Pelka, H. Structural requirements for aminoacylation of *Escherichia coli* formylmethionine transfer RNA. *Biochemistry* **16**, 4256–4265 (1977).
40. Tamura, K., Nameki, N., Hasegawa, T., Shimizu, M. & Himeno, H. Role of the CCA terminal sequence of tRNA (Val) in aminoacylation with valyl-tRNA synthetase. *J. Biol. Chem.* **269**, 22173–22177 (1994).
41. Roy, K. L. & Tener, G. M. Inhibition of aminoacyl transfer ribonucleic acid synthetases by modified transfer ribonucleic acids. *Biochemistry* **6**, 2847–2852 (1967).
42. Björk, G. R. & Hagervall, T. G. Transfer RNA modification: presence, synthesis, and function. *EcoSal Plus* **6**, https://doi.org/10.1128/ecosalplus.esp-0007-2013 (2014).
43. Xiong, Y. & Steitz, T. A. A story with a good ending: tRNA 3′-end maturation by CCA-adding enzymes. *Curr. Opin. Struct. Biol.* **16**, 12–17 (2006).
44. Evans, M. E., Clark, W. C., Zheng, G. & Pan, T. Determination of tRNA aminoacylation levels by high-throughput sequencing. *Nucleic Acids Res.* **45**, e133–e133 (2017).
45. Fischer, J. T., Söll, D. & Tharp, J. M. Directed evolution of *Methanomethylophilus alvus* pyrrolysyl-tRNA synthetase generates a hyperactive and highly selective variant. *Front. Mol. Biosci.* **9**, 850613 (2022).

46. Des Soye, B. J., Patel, J. R., Isaacs, F. J. & Jewett, M. C. Repurposing the translation apparatus for synthetic biology. *Curr. Opin. Chem. Biol.* **28**, 83–90 (2015).
47. Wan, W., Tharp, J. M. & Liu, W. R. Pyrrolysyl-tRNA synthetase: an ordinary enzyme but an outstanding genetic code expansion tool. *Biochim. Biophys. Acta* **1844**, 1059–1070 (2014).
48. Kosuri, S. et al. Scalable gene synthesis by selective amplification of DNA pools from high-fidelity microchips. *Nat. Biotechnol.* **28**, 1295–1299 (2010).
49. Avcilar-Kucukgoze, I. et al. Discharging tRNAs: a tug of war between translation and detoxification in *Escherichia coli*. *Nucleic Acids Res.* **44**, 8324–8334 (2016).
50. Shimizu, Y. et al. Cell-free translation reconstituted with purified components. *Nat. Biotechnol.* **19**, 751–755 (2001).
51. Jakubowski, H. & Goldman, E. Quantities of individual aminoacyl-tRNA families and their turnover in *Escherichia coli*. *J. Bacteriol.* **158**, 769–776 (1984).
52. Chin, J. W. et al. Addition of *p*-azido-L-phenylalanine to the genetic code of *Escherichia coli*. *J. Am. Chem. Soc.* **124**, 9026–9027 (2002).
53. Wheeler, A. M., Orsburn, B. C. & Bumpus, N. N. Biotransformation of efavirenz and proteomic analysis of cytochrome P450s and UDP-glucuronosyltransferases in mouse, macaque, and human brain-derived in vitro systems. *Drug Metab. Dispos.* **51**, 521–531 (2023).
54. Abbott, J. A., Francklyn, C. S. & Robey-Bond, S. M. Transfer RNA and human disease. *Front. Genet.***5**, 158 (2014).
55. Lorenz, C., Lünse, C. E. & Mörl, M. tRNA modifications: impact on structure and thermal adaptation. *Biomolecules* **7**, 35 (2017).
56. de Crécy-Lagard, V. & Jaroch, M. Functions of bacterial tRNA modifications: from ubiquity to diversity. *Trends Microbiol.* **29**, 41–53 (2021).
57. Agmon, I. in *Genesis-In the Beginning: Precursors of Life, Chemical Models and Early Biological Evolution* (ed. Seckbach, J.) 653–668 (Springer, 2012).
58. Crick, F. H. C. The origin of the genetic code. *J. Mol. Biol.* **38**, 367–379 (1968).
59. Noller, H. F. Evolution of protein synthesis from an RNA world. *Cold Spring Harb. Perspect. Biol.* **4**, a003681 (2012).
60. Mitchell, L. A. et al. Synthesis, debugging, and effects of synthetic chromosome consolidation: synVI and beyond. *Science* **355**, eaaf4831 (2017).

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by-nc-nd/4.0/.

© The Author(s) 2026

Methods

tRNA in vitro production

Oligonucleotides containing tRNA under a T7 promoter and followed by an HDV ribozyme[61] were synthesized by IDT as high-fidelity DNA microchip oligo pools from 254–275 nucleotides in length, and dsDNA was amplified in a single PCR reaction for each pool. These tRNA pools included 48 unique *E. coli* isoacceptor tRNAs with all combinations of base 75, all combinations of bases 74 and 75 (degeneracy introduced with machine mixing), and all *M. alvus* and *M. mazei* isoacceptor tRNAs. IVT reactions were carried out at 37 °C using the amplified tRNA dsDNA as template using NEB HiScribe T7 High Yield RNA Synthesis Kit, with addition of ATP, CTP, GTP, UTP, reaction buffer, 200 ng dsDNA template and T7 RNA polymerase. After reverse transcription, polynucleotide kinase treatment was performed to dephosphorylate the tRNA 3′ terminus[62]. tRNA produced by IVT was purified and size-selected using SPRIselect beads. The tRNA was refolded by heating to 80 °C for 5 min and slow cooling to room temperature. At 50 °C, $MgCl_2$ was added to a final concentration of 10 mM.

For amplification of individual tRNAs from a tRNA pool, a unique primer was designed for each tRNA covering the T7 promoter and 5′ unique sequence of the tRNA. The same reverse primer was used for all tRNAs as for the pooled amplification. Alternatively, individual tRNA sequences were synthesized and amplified.

tRNA aminoacylation in cell-free lysates

A total of 40–120 µg of IVT tRNA libraries were incubated in 200 µl NEBExpress Cell-free *E. coli* Protein Synthesis System (New England Biolabs), or a custom cell lysate translation system we prepared (see below), and incubated for 2 h at 37 °C. Alternatively, 40–120 µg of IVT tRNA libraries were incubated in the NEB PURExpress Δ(aa, tRNA) kit and incubated for 2 h at 37 °C.

tRNA extraction

Cell pellets were suspended in 1 ml TRIzol (Thermo Fisher) and frozen at −80 °C. 1-Bromo-3-chloropropane (1/10 volume) was added, and the samples were vortexed and centrifuged at 15,000 × *g* for 15 min at 4 °C. The aqueous phase was transferred to a new tube containing 400 µl of 70% ethanol. Short RNAs were then isolated using a modified RNeasy MinElute Cleanup Kit protocol (Qiagen) to size-select for RNAs <200 nt in length. Samples were centrifuged through MinElute spin column at 12,000 × *g* for 5 min at room temperature. The flow-through was added to 450 µl of 100% ethanol and centrifuged in a new MinElute spin column at 12,000 × *g* for 1 min at room temperature. The column was washed three times with 80% ethanol in 50 mM sodium acetate and dried with open caps at 12,000 × *g* for 5 min. Samples were eluted in 50 mM sodium acetate and 1 mM EDTA.

Periodate oxidation and β-elimination

Periodate oxidation was carried out with 10 µg total RNA isolated above in 10 mM sodium acetate and 50 mM $NaIO_4$. The reaction was incubated at 22 °C for 30 min and quenched with 100 mM D-glucose for 5 min. The samples were size-selected using SPRIselect beads (Beckman Coulter) and eluted in RNase-free water. β-Elimination and deacylation were carried out in 60 mM sodium tetraborate at 45 °C for 90 min. RNA was purified and size-selected using SPRIselect beads. The 3′ phosphoryl group was removed using New England Biolabs T4 Polynucleotide Kinase kit following the protocol of the manufacturer. Dephosphorylated RNA was purified using SPRIselect beads.

tRNA library preparation and sequencing

tRNAs were ligated to a 3′ pre-adenylated adapter oligo using New England Biolabs T4 RNA Ligase 2, truncated KQ kit to ligate adenylated 5′ end of a DNA oligo to the 3′ OH on RNA samples. The reaction was carried out according to the protocol of the manufacturer for 2 h at 25 °C and quenched with 0.5 M EDTA and subsequently size-selected with SPRIselect. Primer-dependent reverse transcription reaction was carried out using Maxima H Minus Reverse Transcriptase[63]. A total of 50 ng of 3′-adapter-ligated samples were incubated with 6 pM RT primer and 20 mM dNTP mixture at 75 °C for 5 min to denature the primer and were added to Maxima kit components according to the protocol of the manufacturer. After incubation, alkaline hydrolysis was carried out through the addition of NaOH and incubation at 95 °C for 3 min, and the solution was neutralized with HCl. The samples were purified using SPRIselect. The 5′ adapter was ligated using New England Biolabs Thermostable 5′ App DNA/RNA Ligase according to the protocol of the manufacturer, and the adapter-ligated samples were size-selected with SPRIselect. The barcodes were added to the 3′ and 5′ ends of cDNA through single-step PCR with KAPA Hifi HotStart ReadyMix using primers, including the Illumina P5 and P7 regions, and barcodes adapted from Illumina TruSeq. Libraries were purified using SPRIselect and quantified using QuBit. Libraries were prepared for sequencing using the Illumina MiSeq protocol and sequenced on an Illumina MiSeq or NovaSeq instrument.

Robotic procedure for tSCAN

We carried out the same procedure as above, with all pipetting and magnetic bead purification steps fully automated using the OT-2 platform. We used an Opentrons OT-2 equipped with two Magnetic Modules (GEN2), a P300 8-Channel Pipette (GEN2), a P20 Single-Channel Pipette (GEN2) and two Temperature Modules with 96-well aluminium temperature blocks. We used 96-well plates and 12-well reservoirs for input of samples and reagents.

We used a Jupyter notebook of OT-2 to control all protocols. For each consumable in each deck position, we initially calibrated the *x*, *y* and *z* positions and stored these objects for reuse in all procedures. We also calibrated the *x*, *y* and *z* positions of the 8-Channel within the wells for aspiration and resuspension steps. For all purifications, we specified the number of columns to be purified at the beginning of the protocol, and the multichannel pipette was used to carry out the procedure. All buffers used during the procedure were provided in 12-well reservoirs with quantities calculated by the protocol at time of initialization. All reaction steps were inherited from a mixing script that separated sample data representation from machine control. Three.xlsx spreadsheets were provided containing layouts of two source plates as well as reactions. A CSV file was automatically generated that specified all operations the instrument would perform. The machine control protocol would then control the hardware and execute all steps.

Analysis of tRNA aminoacylation from NGS data

The .fastq files obtained from sequencing were analysed with custom Python scripts. After quality filtering, reads were aligned with MMseq to a library of all *E. coli* tRNAs. Next, the aligned data were searched for each tRNA sequence and adapter sequence along with each combination of terminal 3′ nucleotides. All reads were quantified for the presence or absence of terminal adenosine as charged or uncharged tRNAs, respectively. All charged and uncharged percentages were quantified for each isoacceptor tRNA with all 3′ trinucleotide combinations and output as .csv files.

In-lysate determination of tRNA aminoacylation using isotopically labelled ATP with tSCAN-M

Each tRNA species being assayed for aminoacylation was in-vitro-transcribed as described above, except that isotopically labelled ATP (Adenosine-$^{15}N_5$ 5′-triphosphate, Adenosine-$^{13}C_{10}$,$^{15}N_5$ 5′-triphosphate or Adenosine-$^{13}C_{10}$ 5′-triphosphate) replaced the ATP supplied in the HiScribe kit. The tRNA was purified and refolded as before.

In each cell-free reaction, 1–2 µg of in-vitro-transcribed tRNA being assayed for aminoacylation was added into a 10 µl volume of

NEBExpress Cell-free *E. coli* Protein Synthesis System (New England Biolabs) and incubated at 37 °C for 2 h. Then 10 µl of RNase A digestion solution was added, consisting of 200 mM sodium acetate (pH 5.2) and 1.5 U μl^{-1} RNAse A and incubated for 20 min at room temperature. The protein was precipitated by addition of 1% formic acid, and the reactions were then frozen at −80 °C for 30 min to 1 day. After precipitating protein at −80 °C for 30 min, the insoluble material was removed by centrifugation at 16,000 × *g* for 15 min at 4 °C. The soluble fraction was then transferred to autosampler vials, kept on ice until immediately before high-resolution LC–MS analysis and returned to ice immediately afterwards.

aaRS protein production and purification

E. coli LeuRS, ProRS and ValRS were subjected to PCR amplification from the genome of *E. coli* K12 MG1655 and cloned with His-tag into a ColE1 backbone with a bla resistance marker and transformed into *E. coli* BL21(DE3) expression strain. The aaRSs were overexpressed in *E. coli* cells with an optical density (OD) of 0.6. Protein production was induced by adding 0.5 mM isopropyl-b-D-1-thiogalactopyranoside (Millipore Sigma), and the cells were moved to a 33 °C incubator and grown at 250 rpm for 2.5 h. The cells were harvested at 6,000 × *g* for 15 min at 4 °C, washed with 1× PBS buffer, and stored at −80 °C. The frozen cell pellet was thawed in lysis buffer (100 mM HEPES pH 7.2, 500 mM NaCl, 5 mM BME). The cell paste was suspended in 15 ml of lysis buffer (50 mM Tris (pH 7.5), 300 mM NaCl, 20 mM imidazole) and lysed by sonication. The crude extract was centrifuged at 30,000 × *g* for 30 min at 4 °C. The soluble fraction was loaded onto a column containing 2 ml of Ni-NTA resin (Qiagen) previously equilibrated with 20 ml lysis buffer. The column was washed with 20 ml lysis buffer, and the bound protein was then eluted with 2 ml of 50 mM Tris (pH 7.5), 300 mM NaCl and 300 mM imidazole. The purified proteins were dialysed with 10 mM Tris (pH 7.5), 0.5 M NaCl, 1 mM DTT and 50% glycerol, and stored at −80 °C for further studies.

In vitro aminoacylation activity assay

A 20 µl aminoacylation reaction contained the following components: 50 mM Tris-HCl (pH 7.2), 10 mM MgCl2, 10 mM ATP, 2 mM amino acids, 500 nM aaRS and 20 µg tRNA. Aminoacylation reactions were incubated at 1 h at 37 °C. The reactions were stopped by the adding 20 µl digestion solution consisting of 200 mM sodium acetate (pH 5.2) and 1.5 U μl^{-1} RNase A, and incubated for 20 min at room temperature. The protein was precipitated by addition of 1% formic acid, and the reactions were then frozen at −80 °C for 30 min to 1 day. After precipitating protein at −80 °C for 30 min, the insoluble material was removed by centrifugation at 16,000 × *g* for 15 min at 4 °C. The soluble fraction was then transferred to autosampler vials, kept on ice until immediately before high-resolution LC–MS analysis and returned to ice immediately afterwards.

Mass spectrometry analysis of tRNA aminoacylation

Quantitative mass spectrometry data were collected using an Agilent 6530 Quadrupole Time-of-Flight (QTOF) MS with an electrospray ionization (ESI) source, coupled to an Agilent Infinity 1290 ultrahigh-performance liquid chromatography (UHPLC) system with an Agilent Poroshell 120 EC-C18 2.7 µm, 2.1 ×50 mm column. The solvents used were water and 0.1% formic acid (solvent A) and acetonitrile and 0.1% formic acid (solvent B). Mass spectra were gathered using Dual Agilent Jet Stream (AJS) ESI in positive mode. The mass range was set from 100 *m*/*z* to 1,700 *m*/*z* with a scan speed of 3 scans per second. The capillary and nozzle voltages were set at 3,500 V and 1,000 V, respectively. The source parameters were set with a gas temperature of 325 °C and a flow rate of 12 l min^{-1}, nebulizer at 35 psi and sheath gas temperature at 350 °C at a flow rate of 11 l min^{-1}. MS data were acquired with MassHunter Workstation Data Acquisition (v.B.06.01, Agilent Technologies) and analysed using MassHunter Qualitative Analysis (v.10.0, Agilent Technologies). All amino acid adenylate masses were monitored, and the EICs were plotted for each ion of interest.

Production of custom *E. coli* cell-free translation system

We cloned a plasmid containing the ribosomal 23S rRNA with G2251C and G2553C mutations and transformed it into BL21(DE3) competent cells to produce a lysate with CGA ribosomes. As a control, we cloned a plasmid containing the WT ribosomal 23S rRNA and transformed it into BL21(DE3). For both plasmids, we used a ColE1 backbone with a design as previously reported[17] to have a copy number of 50–100 within each cell and increase the quantity of CGA ribosomes in the lysate (compared with the seven copies of native ribosomes in the *E. coli* genome).

BL21(DE3) competent cells containing ribosome plasmids were grown in 1 l of 2xYT medium (16 g l^{-1} tryptone, 10 g l^{-1} yeast extract, 5 g l^{-1} NaCl, 7 g l^{-1} K_2HPO4, 3 g l^{-1} KH_2PO4, pH 7.2) supplemented with 100 µg ml^{-1} of carbenicillin and incubated at 37 °C at 250 rpm. At an OD_{600} of 0.6, cultures were inoculated with 1 mM IPTG and grown to an OD of 3.0. The cells were then centrifuged for 15 min at 5,000 × *g* and 4 °C. The pellets were washed three times with ice-cold S30 buffer (10 mM tris-acetate pH 8.2, 14 mM magnesium acetate, 60 mM potassium acetate, 2 mM dithiothreitol) and frozen at −80 °C.

The frozen pellets were thawed and resuspended in 0.8 ml g^{-1} of pellet mass and sonicated in a QSonica sonicator at 50% amplitude in an ice-water bath with a cycle of 45 s on and 59 s off, with a total of about 600 J delivered per ml of sample. DTT (3 ml) was added per ml of the sample immediately after sonication. The lysate was centrifuged at 18,000 × *g* at 4 °C for 15 min, and the supernatant underwent a run-off reaction with incubation at 37 °C and 250 rpm for 1 h. The samples were then centrifuged again at 10,000 × *g* and 4 °C for 10 min, and the supernatant was aliquoted and frozen at −80 °C.

Peptide production in cell-free translation systems

In vitro translation reactions were set up with NEBExpress Cell-free *E. coli* Protein Synthesis System (New England Biolabs) or custom-produced *E. coli* cell-free translation systems. Briefly, the reactions were set up consisting of 12 µl S30 extract, 25 µl of the NEBxpress synthesis buffer, 5 µg of linear DNA, 1 µl RNase inhibitor, 1 µl T7 polymerase, 1 µl GamS inhibitor and water to 50 µl. Parallel reactions were also set up to increase yields of peptide production and then pooled before purification. Linear DNA was amplified with T7 promoter and terminator from a double-stranded DNA template and purified with AMPure XP magnetic beads (Thermo). The reactions were incubated for 18 h at 30 °C. The translation reactions were affinity-purified using Dynabeads His-Tag Isolation and Pulldown beads (Thermo). Samples for WT ribosome crosstalk experiments were washed in 10 mM imidazole in 1× PBS and eluted with 500 mM imidazole in 1× PBS. The purified samples were then size-filtered using 10 kDa filters (Pall) and run on a Q Exactive Orbitrap mass spectrometer (see below). The other samples were incubated on the Dynabeads His-Tag Isolation and Pulldown beads and directly digested on beads before proteomic analysis (see below).

Compressed genetic code generation and AGENTEX workflow

As mentioned previously, we used an Opentrons OT-2 equipped with two Magnetic Modules (GEN2), a P300 8-Channel Pipette (GEN2), a P20 Single-Channel Pipette (GEN2) and two Temperature Modules with 96-well aluminium temperature blocks. We used 96-well plates and 12-well reservoirs for input of samples and reagents.

All protocols were run using the Jupyter notebook of OT-2. The protocol was divided into dsDNA production, DNA purification, tRNA production, tRNA purification, polypeptide production and polypeptide purification steps (polypeptide purification was optional and used for LC–MS/MS analysis but not for luminescence experiments). For all purifications, we specified the number of columns to be purified at the beginning of the protocol, and the multichannel pipette was used to carry out the procedure.

The tRNA pool sequences were automatically generated depending on the specified number of codons to be used for translation. The peptide DNA sequence was also generated with compressed genetic code from a provided amino acid sequence. All DNA sequences were ordered as chip-synthesized oligonucleotides and amplified and purified by machine protocols. All of the tRNAs in each tRNA pool were transcribed in one reaction, and all pools were purified in parallel. When an orthogonal aaRS, as in pazFRS used for *M. jannaschii* $tRNA^{tyr}_{CUA}$, was required, its dsDNA was also included in the translation reaction mixture. All reaction steps (dsDNA production, tRNA production, cell-free translation reaction) were inherited from a mixing script that separated sample data representation from machine control. Three .xlsx spreadsheets were provided containing layouts of two source plates as well as reactions. A .csv file was automatically generated that specified all operations the instrument would perform. The machine control protocol would then control the hardware and execute all steps.

Proteomic analysis of peptides produced in cell-free translation systems

Quantitative analysis was performed by a Q Exactive Orbitrap mass spectrometer (Thermo) equipped with an Evosep (Odense) nano-pump. The samples were run on 30 SPD (samples per day) method by Evosep. Flow-through that was size-filtered by 10 kDa filters was further digested in 50 mM TEAB buffer with trypsin for 3 h. The resulting digests were loaded directly onto Evotips (Evosep). The peptides were eluted from Evotips directly to 15 cm PepSep C18 (Bruker) column for chromatographic separation. The Q Exactive instrument was run in WWA mode (wide window isolation) with MS data-dependent acquisition. The data were searched by Thermo Proteome Discoverer 3.1.1.93 with Chimerys search engine. The data were searched against the general *E. coli* K12 database from Uniprot and a custom-built library database that included all predicted amino acid changes. We allowed variable modification on Ala (+4 Da) in cases in which isotopically labelled alanine was included in the cell-free translation system. All quantitation of the presence of such heavy amino acids was done based on MS spectra peak volumes that were automatically extracted by software from raw data. The results were kept at 0.1% FDR level on both peptide and protein level by Percolator[64].

For analysis of AGENTEX crosstalk, quantitative analysis of pulldown samples from magnetic beads was performed with ZenoTOF7600+ (SCIEX) coupled with Evosep (Odense) nanoHPLC system. The samples were digested directly from the magnetic beads in 50 mM TEAB buffer at 37 °C over 3 h. The digested peptides were loaded directly into Evosep tips for analysis. The peptides were eluted from Evotips directly to 15 cm PepSep C18 (Bruker) column for chromatographic separation. The instrument was running in DIA mode with a 400–800 Da and automatic window isolation with 20 ms accumulation time for MS/MS acquisition. Searches were done with the PEAKS 13 database (BSI). All searches were done with 1% FDR at protein and peptide levels. The data were searched against the general *E. coli* K12 database from Uniprot and a custom-built library database that included all predicted amino acid changes.

Reporting summary

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

Data availability

All data generated or analysed during this study are included in this published article and the Supplementary Information. All raw NGS data generated in this study have been deposited in the European Nucleotide Archive under accession number PRJEB120976. All proteomics data generated in this study have been deposited in MassIVE at https://doi.org/10.25345/C5Z31P24Q. Any additional information needed is available from the corresponding authors upon reasonable request. Source data are provided with this paper.

Code availability

All code used for the tSCAN pipeline is available at the associated GitHub repository (https://github.com/FelixRadford/tSCAN_v1.0). A permanent archive of the version used in this study is available at Zenodo[65] (https://doi.org/10.5281/zenodo.21321596). All code for AGENTEX is available at the associated GitHub repository (https://github.com/FelixRadford/AGENTEX). A permanent archive of the version used in this study is available at Zenodo[66] (https://doi.org/10.5281/zenodo.21321702).

61. Schürer, H., Lang, K., Schuster, J. & Mörl, M. A universal method to produce in vitro transcripts with homogeneous 3′ ends. *Nucleic Acids Res.* **30**, e56 (2002).
62. Mörl, M., Lizano, E., Willkomm, D. K. & Hartmann, R. K. in *Handbook of RNA biochemistry* (eds. Hartmann, R. K. et al.) 22–35 (Wiley, 2005).
63. Lucas, M. C. et al. Quantitative analysis of tRNA abundance and modifications by nanopore RNA sequencing. *Nat. Biotechnol.* **42**, 72–86 (2024).
64. The, M., MacCoss, M. J., Noble, W. S. & Käll, L. Fast and accurate protein false discovery rates on large-scale proteomics data sets with Percolator 3.0. *J. Am. Soc. Mass. Spectrom.* **27**, 1719–1727 (2016).
65. Radford, F. tSCAN_v1.0 [Computer software]. *Zenodo*. https://doi.org/10.5281/zenodo.21321596 (2026).
66. Radford, F. AGENTEX [Computer software]. *Zenodo*. https://doi.org/10.5281/zenodo.21321702 (2026).

Acknowledgements We thank the members of the Church lab for generous discussions and support. We thank the anonymous reviewers for their feedback, which has improved this paper. We thank K. Jurcic and the team at Bioinformatics Solutions for providing to us PEAKS 13 software and help with comparative MS/MS peak intensity analysis.

Author contributions G.M.C. and F.R. conceptualized the study. F.R. conducted the formal analysis. F.R., B.B., N.S., H.-M.B., L.O. and G.M.C. devised the methodology. F.R. conducted the investigation. F.R., B.B. and N.S. developed the software. F.R., B.B., N.S. and H.-M.B. performed the visualization. G.M.C. and F.R. helped with funding acquisition. G.M.C. and F.R. administered the project. F.R. wrote the original draft, and all authors reviewed and edited the paper.

Funding This work was funded by the National Science Foundation grant 2123243 (G.M.C.), the Department of Energy grant DE-FG02-02ER63445 (G.M.C.) and the Harvard Medical School Dean's Innovation Award for the Use of Artificial Intelligence in Education, Research and Administration (F.R. and G.M.C.).

Competing interests F.R. and G.M.C. are listed as inventors of a provisional patent application filed through Harvard University (PCT 017868). G.M.C. is a founder of companies with related financial interests: GRO Biosciences, EnEvolv (Ginkgo Bioworks) and Pearl Bio. Other relevant financial interests of G.M.C. are listed at http://arep.med.harvard.edu/gmc/tech.html. All other authors declare no competing interests.

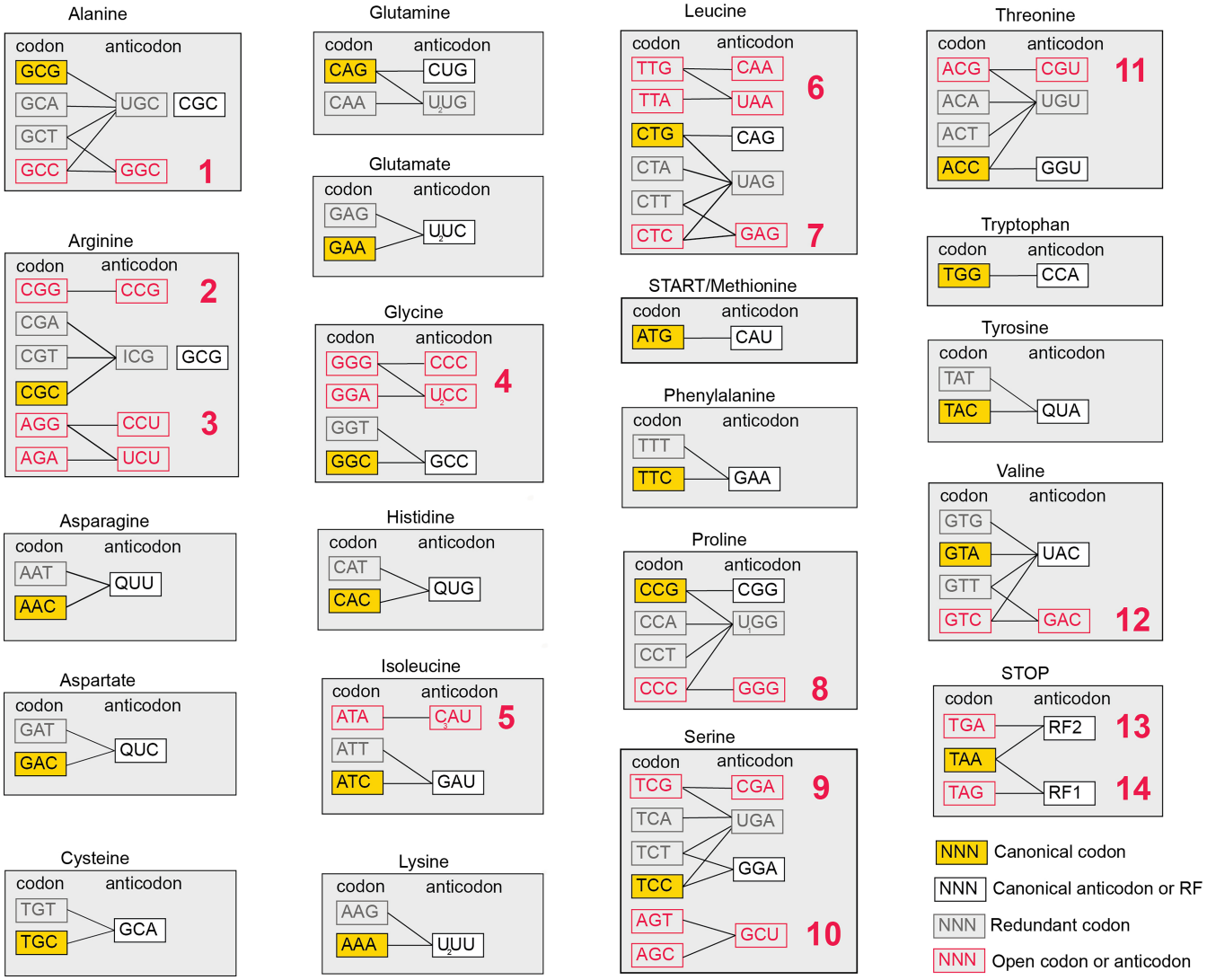
Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41586-026-10949-y.

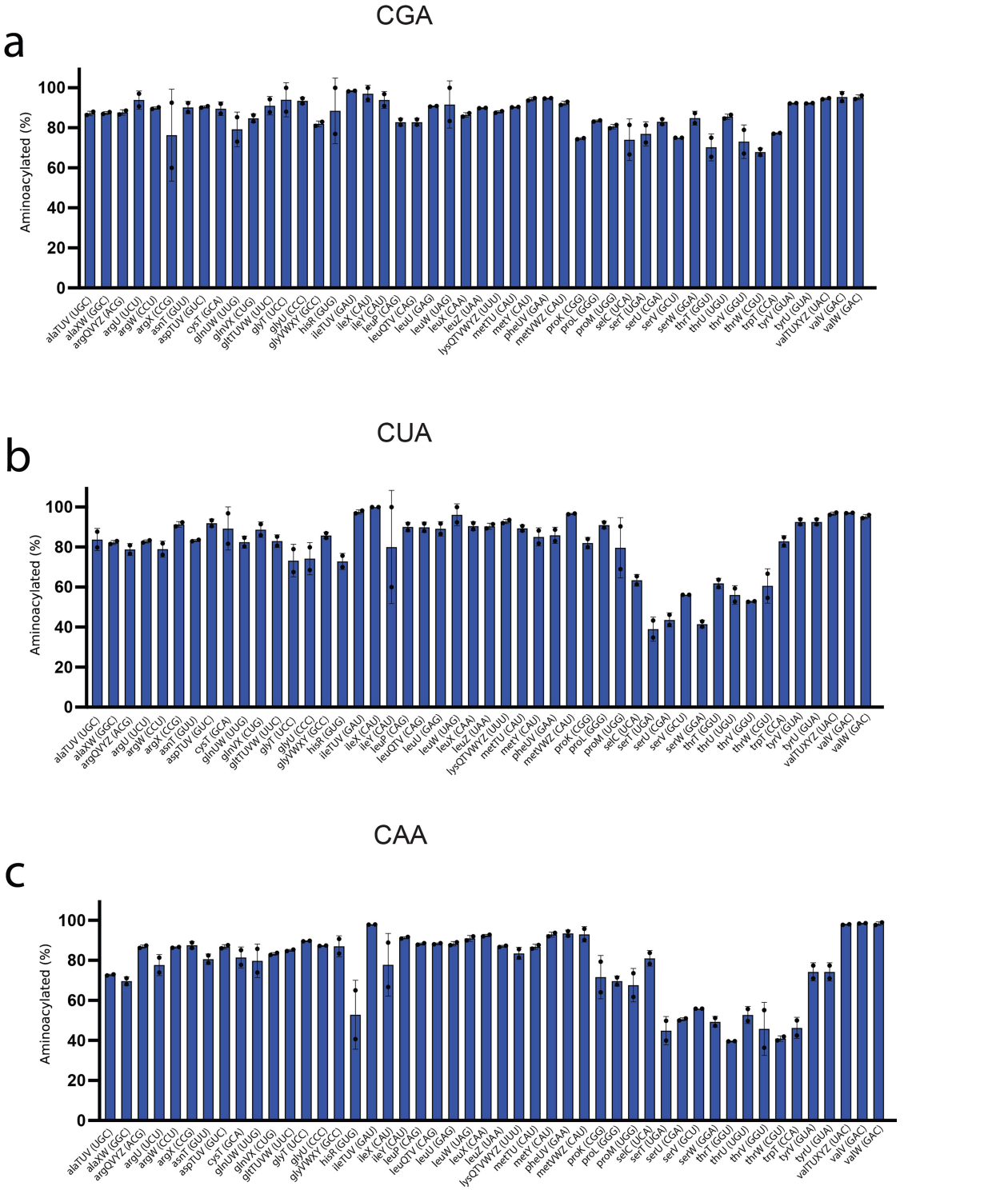
Correspondence and requests for materials should be addressed to Felix Radford or George M. Church.

Peer review information *Nature* thanks Christopher Francklyn, Patrick Mullen and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

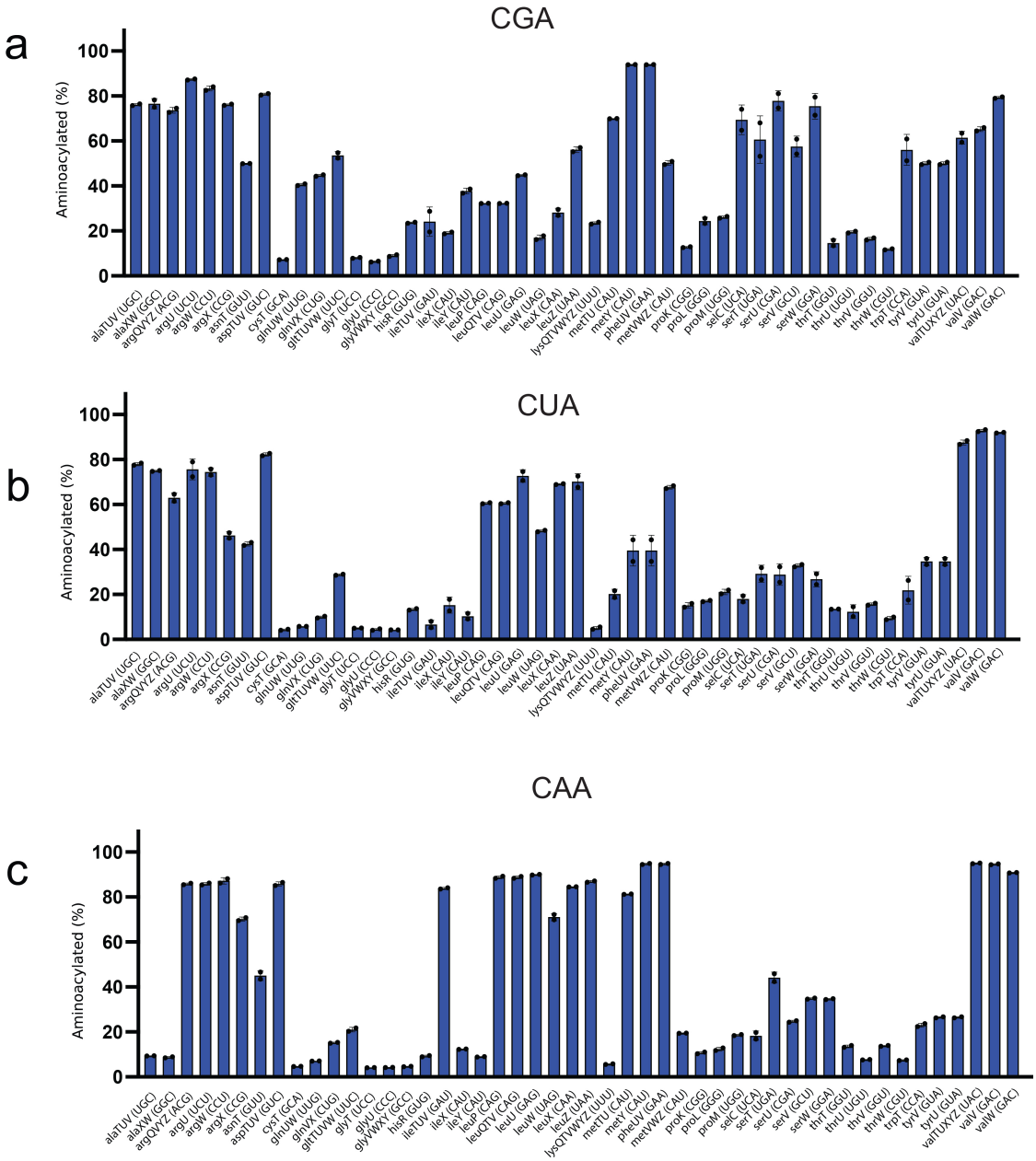
Reprints and permissions information is available at http://www.nature.com/reprints.



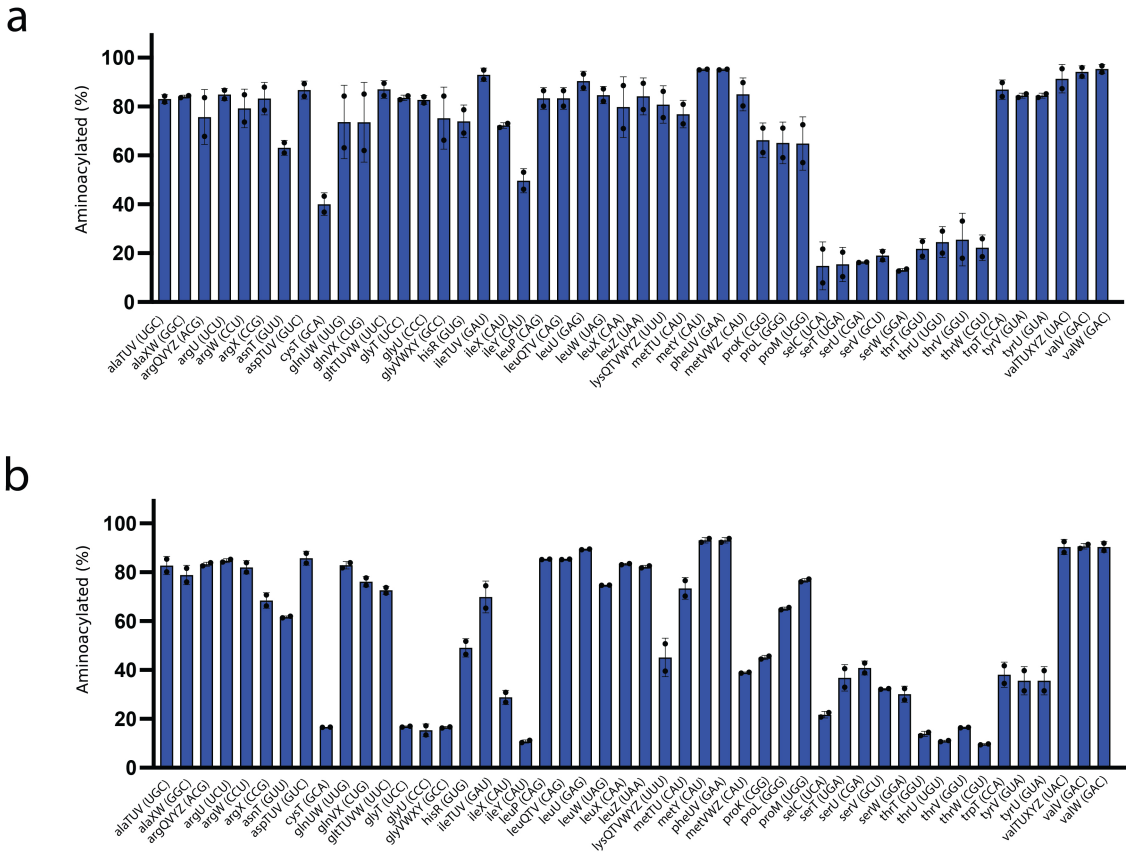
Extended Data Fig. 1 | Design of compressed genetic codes. Given 64 otRNA triplet codons and normal wobble rules, 20 canonical amino acids can each be decoded by a single tRNA and 14 codons available for reassignment, indicated in red and numbered. The codons used to decode the 20 canonical amino acids are indicated with yellow, and anticodons of tRNAs or release factors (RFs) used to decode the canonical amino acids or for start or stop codons indicated in white. Gray rectangles indicate redundant codons and red rectangles indicate open codons or anticodons. RNA modifications are indicated as: Q = Queuosine, U_1 = xo^{5}U, C_3 = K^{2}C, I = Inosine, U2 = xm^5(s^2)U.



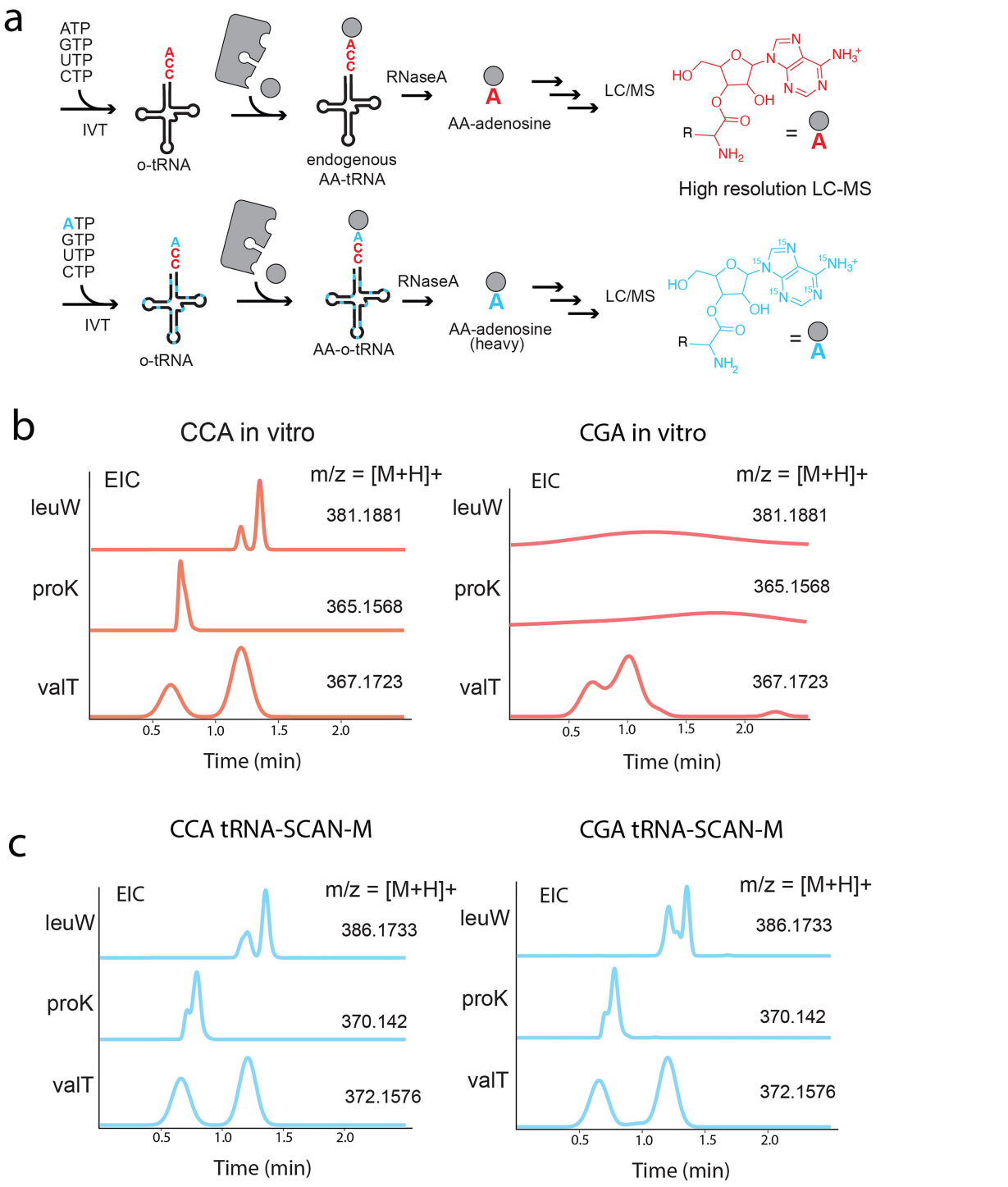
Extended Data Fig. 2 | Measurement of aminoacylation of tRNAs with non-CCA ends incubated in Lysate. *E. coli* tRNAs were synthesized (**a**) having a CGA 3′ end, (**b**) a CUA 3′ end, or (**c**) a CAA 3′ end were incubated in Lysate cell-free translation and analyzed by tSCAN to determine their levels of aminoacylation. All data represent n = 2 independent replicates. Note controls in Extended Data Fig. 4a with a synthetic pool of 3′ CCA tRNAs that were conducted under identical conditions and constitute a single panel with these experiments. These data are also represented in heatmap format in Fig. 3.



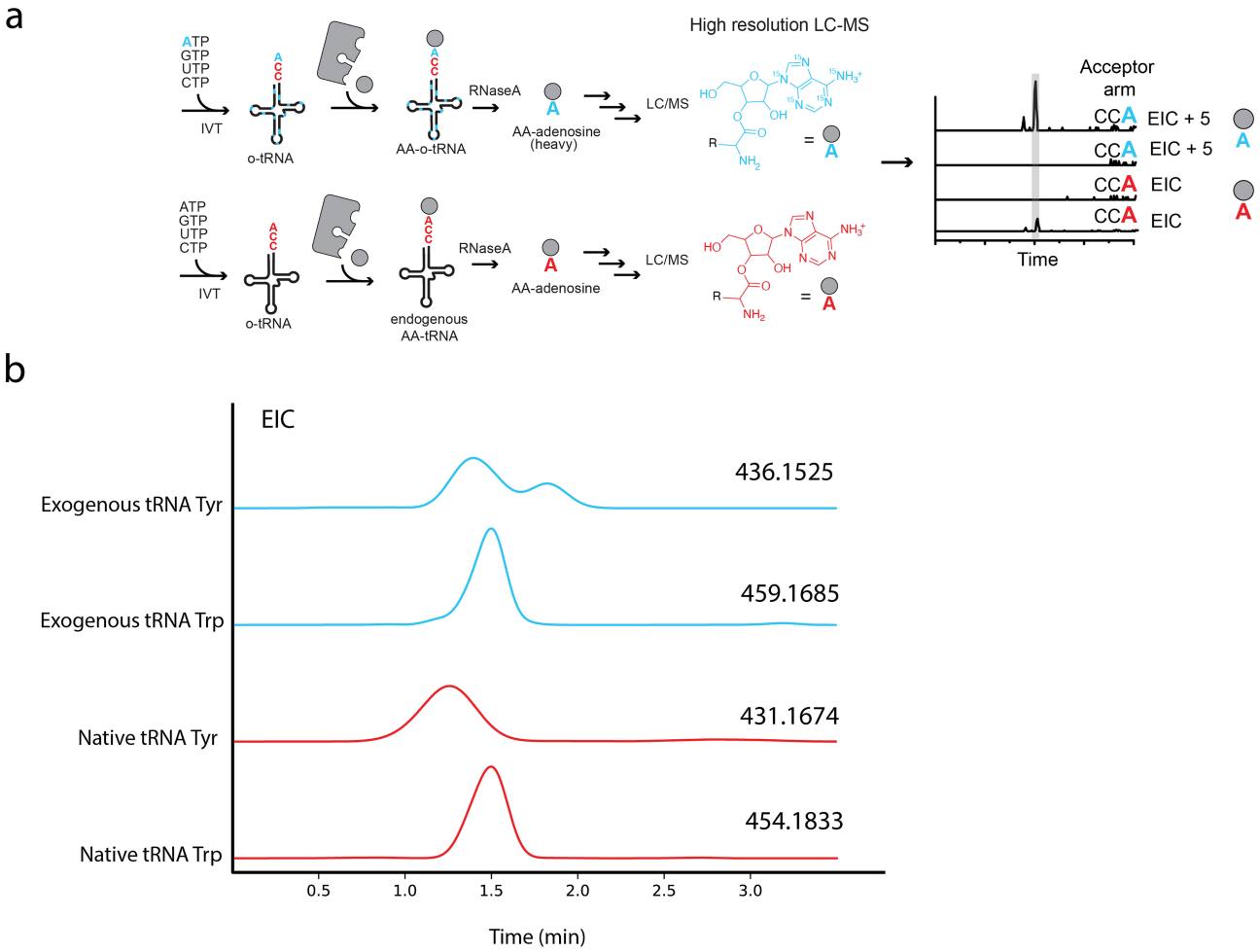
Extended Data Fig. 3 | Measurement of aminoacylation of tRNAs with non-CCA ends incubated in PURExpress reconstituted cell-free translation system. *E. coli* tRNAs were synthesized (**a**) having a CGA 3′ end, (**b**) a CUA 3′ end, or (**c**) a CAA 3′ end were incubated in PURExpress cell-free translation and analyzed by tSCAN to determine their levels of aminoacylation. All data represent n = 2 independent replicates. Note controls in Extended Data Fig. 4b with a synthetic pool of 3′ CCA tRNAs that were conducted under identical conditions and constitute a single panel with these experiments. These data are also represented in heatmap format in Fig. 3.



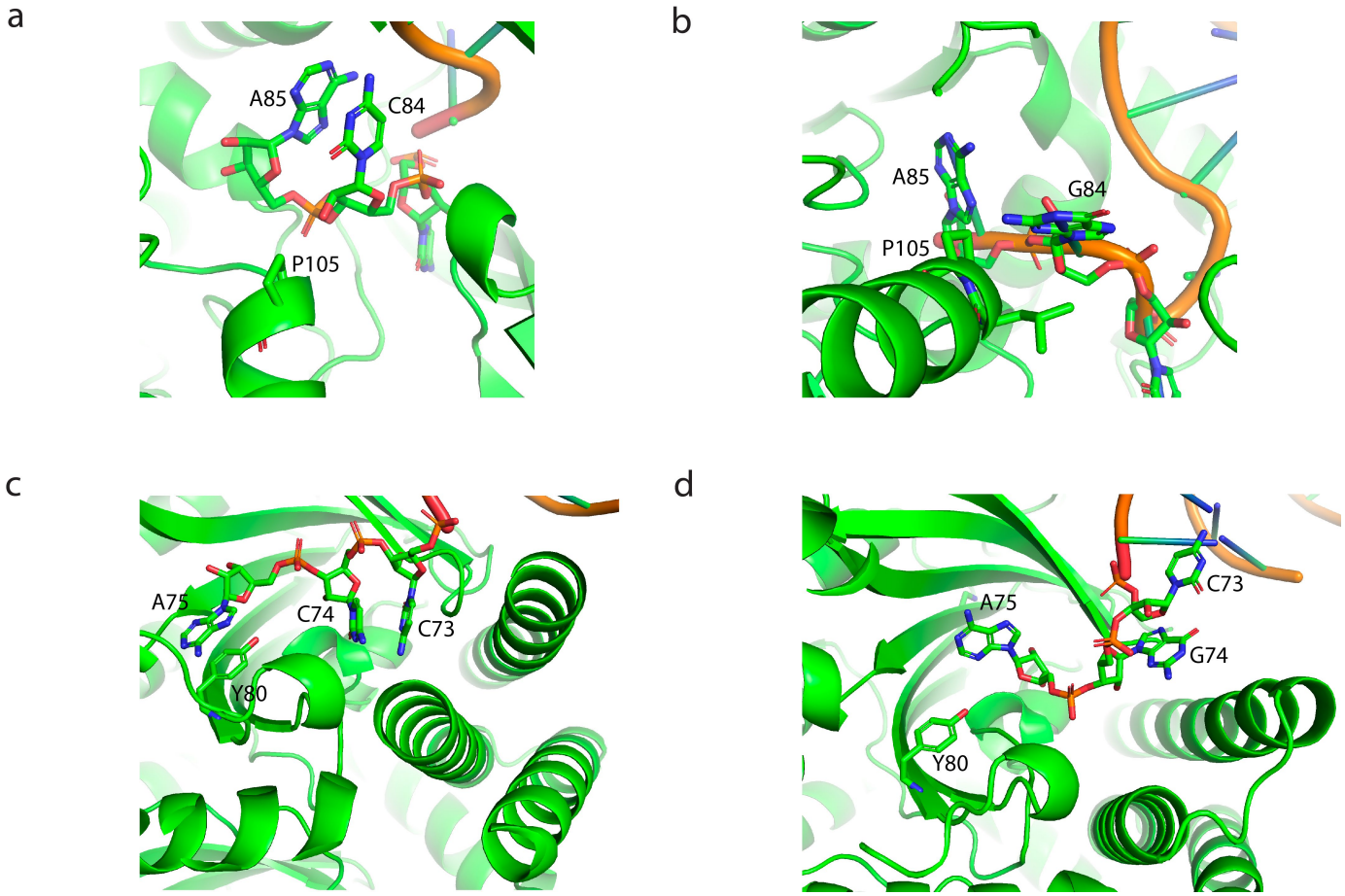
Extended Data Fig. 4 | Measurement of the aminoacylation levels of tRNAs with 3′ CCA-ends using tSCAN. A synthetic pool of all *E. coli* tRNA isoacceptors was either incubated in (**a**) *E. coli* lysate-based cell free translation system or (**b**) *E. coli* PURExpress reconstituted cell free translation system to determine the effects of the native translation system compared to purified aaRSs on the efficiency of aminoacylation of synthetic tRNAs. All data represent n = 2 independent replicates. These data are also represented in heatmap format in Fig. 3.



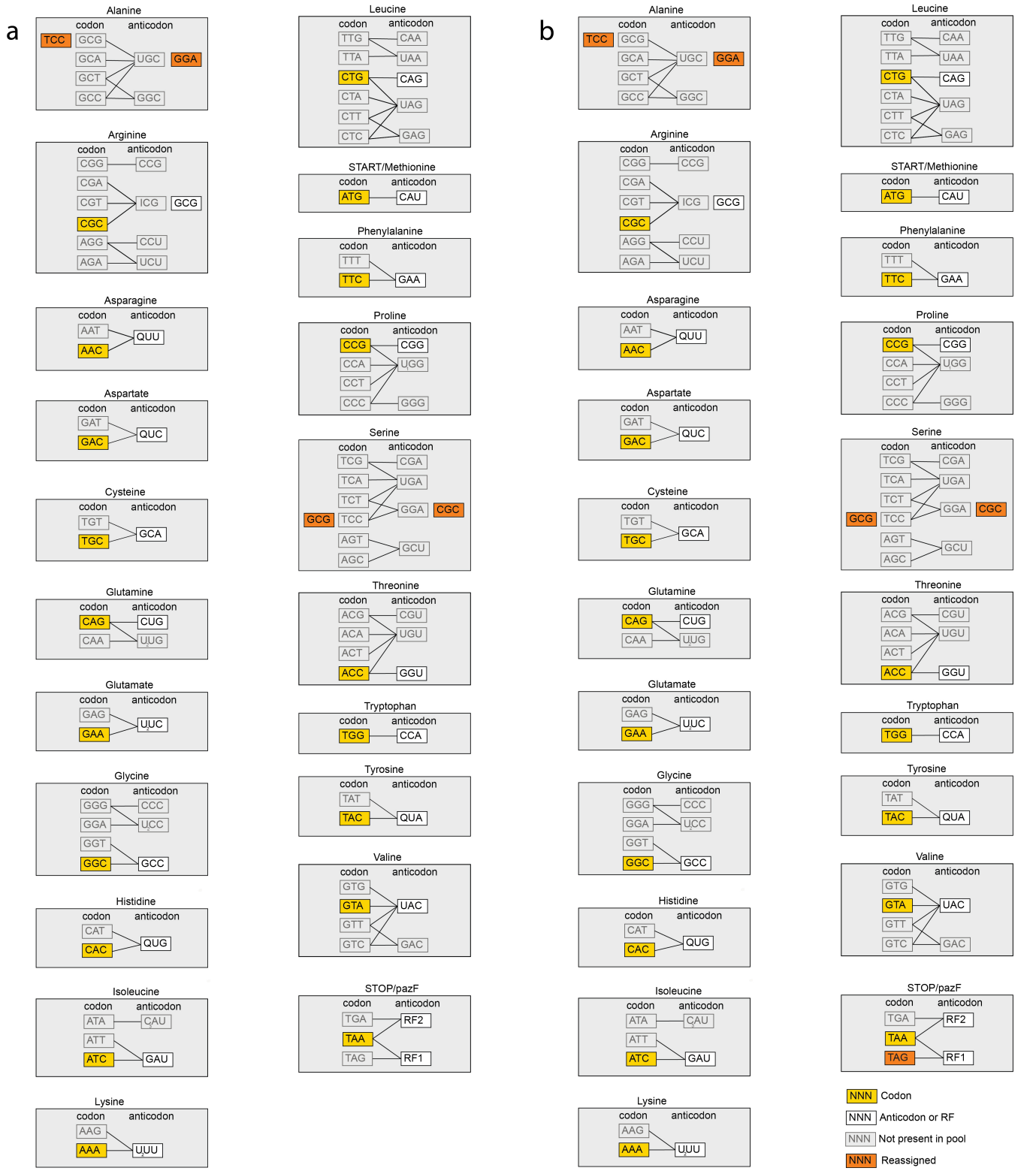
Extended Data Fig. 5 | Natural tRNA modifications are required for otRNA aminoacylation by aaRSs. (**a**) A schematic representation of in vitro determination of tRNA aminoacylation with a specific amino acid using liquid chromatography mass spectrometry (LC-MS), top, and determination of the charging of a specific tRNA directly within a cell-free translation system using tSCAN-M, bottom. In tSCAN-M, isotopically-labeled ATP is used for in vitro transcription of a specified tRNA which is then added into cell-free lysate. RNaseA digestion yields an isotopically labeled amino acid adenylate which can be distinguished from the native tRNA repertoire. (**b**) In vitro aminoacylation was performed on $tRNA^{Leu}$, $tRNA^{Pro}$, and $tRNA^{Val}$ with CCA or CGA 3′ ends, respectively. (**c**) tSCAN-M was performed on $tRNA^{Leu}$, $tRNA^{Pro}$, and $tRNA^{Val}$ with CCA or CGA 3′ ends, respectively directly within an *E. coli* cell-free lysate system, showing that aminoacylation of tRNAs with a CGA 3′ end was restored compared to an in vitro aminoacylation reaction (**b**). All data show gaussian smoothed plots of raw extracted ion chromatograms, showing one dataset in each subplot (relative scale in each subplot).



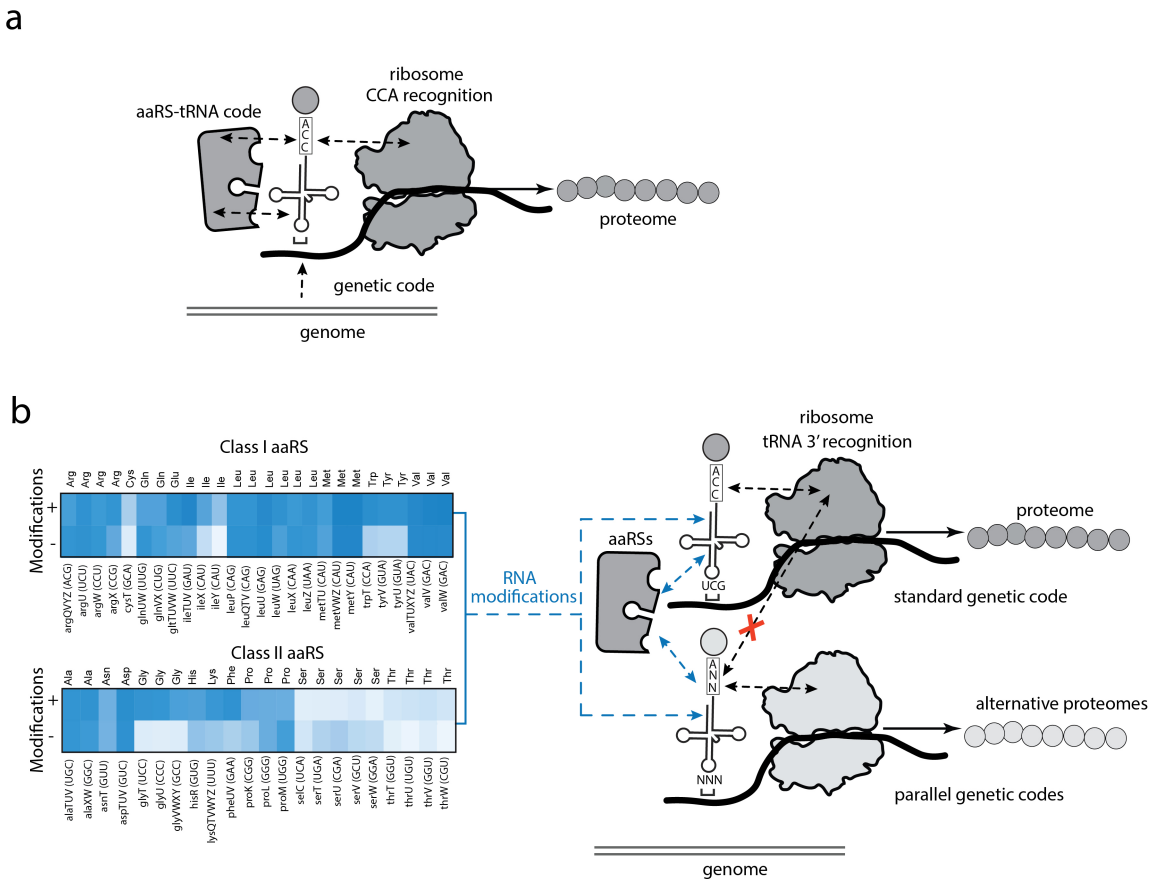
Extended Data Fig. 6 | tSCAN-M validation with *E. coli* tRNA^{Tyr} and tRNA^{Trp}. (**a**) The tSCAN-M workflow. IVT tRNAs containing isotopically labeled adenosines are introduced into an *E. coli* cell lysate, where they are subject to tRNA modifying enzymes and aaRS charging. In-solution RNAase A digestion digests both native and exogenous tRNAs to single nucleotides. Formic acid precipitation removes proteins and large macromolecules from the reaction mixture. The aminoacyl adenylates from exogenous tRNAs can be distinguished from native tRNAs by LC-MS by a 5 Da mass difference, and the identity of the charged amino acid can be determined. (**b**) tSCAN-M was performed on exogenous *E. coli* tRNA^{Tyr} and tRNA^{Trp} introduced into a cell free reaction of *E. coli* lysate. After incubation, RNase A was added directly to the lysates. Samples were precipitated and run on QTOF LC-MS. The aminoacyl adenylates of the exogenous tRNAs are expected to have a 5 Da greater mass than their corresponding native tRNA aminoacyl adenylates. Both native and exogenous tRNAs were detected from the *E. coli* translation system using tSCAN-M.



Extended Data Fig. 7 | Modeling of class I and class II aaRSs to understand differences in CGA vs. CCA tRNA reactivity. E. coli aaRSs were modeled with AlphaFold3 in complex with their cognate tRNAs containing a CCA or CGA 3′ end, respectively. (**a**) LeuRS, a class I aaRS, and tRNA leuW-CCA, or (**b**) leuW-CGA. (**c**) GlyRS, a class II aaRS, and tRNA glyT-CCA or (**d**) tRNA glyT-CGA. (A-B) When tRNA leuW contains a CGA 3′ end, there is a steric hinderance between A85 and G84 with P105. This hinderance is not present in the WT leuW tRNA where cytosine is at position 84. (**c-d**) π-π stacking is present between A75 of WT tRNA glyT and Y80 of glyRS. In glyT-CGA, this interaction is not observed, which may destabilize the tRNA-aaRS interaction for proper aminoacylation.



Extended Data Fig. 8 | Compressed genetic codes MP1 and MP2. (**a**) The compressed genetic code corresponding to otRNA pool MP1, with alanine and serine codons swapped. AGENTEX was used to test MP1 for translation. (**b**) The compressed genetic code corresponding to otRNA pool MP2, with TAG codon used for pazF incorporation, and alanine and serine codons swapped. AGENTEX was used to test MP2 for translation. The codons included in each pool are indicated with yellow, and anticodons of tRNAs included in the pool designated with white. Gray rectangles indicate that the codons and anticodons were removed from the code. Orange rectangles indicate codons that have been reassigned to a different amino acid, or stop codon(TAG) reassigned for nsAA incorporation. RNA modifications are indicated as: Q = Queuosine, U_1 = xo^5U, C_3 = K^2C, I = Inosine, U2 = $xm^5(s^2)U$.



Extended Data Fig. 9 | A revised model of genetic code maintenance. (**a**) The currently accepted model of the genetic code and the translation system. The CCA end of tRNA interacts both with aaRSs and the ribosome to maintain aminoacylation and accommodation. The identity elements of each tRNA are further used for recognition by its cognate aaRS. The genetic code is linearly specified via the genome through mRNA and the tRNA molecule connects all aspects of the translation system to allow the instructions from DNA to be translated into protein. (**b**) The revised model of the translation system from this work highlighting the balance between aaRS-tRNA interactions and tRNA-ribosome interactions that can maintain the present genetic code. The tRNA molecule is extensively modified, and Class I and Class II aaRSs display differences in recognition of tRNA modifications. The efficiency of aminoacylation further differs depending on whether the 3′ CCA sequence is mutated (Fig. 3). The large subunit of the ribosome interacts with the 3′ CCA end of tRNA and prevents tRNAs with non-CCA ends from being accommodated (Fig. 5). However, if the CCA sequence is altered, and if the ribosome has a complimentary mutation, it can accept that tRNA for translation (Fig. 6). There is the possibility of new genetic codes arising purely from co-evolution of tRNA CCA ends and the ribosomal large subunit. As this could result in alternative genetic codes, there is selective pressure against this.

Extended Data Table 1 | Peptides and masses in MS/MS analysis of translation crosstalk

Peptide	m/z	z	Retention time (min)
VPGAGVPGAGVPGAGK	645.3582	2	17.85 *
VPGSGVPGSGVPGSGK	669.3506	2	14.44 *
PGSGVPGAGVPGSGK	612.6840	2	7.9373
GVPGAGVPGSK	491.7696	2	7.1291
GVPGSGVPGAK	491.7696	2	7.1085
GVPGSGVPGSK	499.7610	2	ND
GVPGAGVPGAK	483.7665	2	ND

* Run on Q Exactive Orbitrap mass spectrometer (Thermo, MA).
ND = Not detected.

nature portfolio

Corresponding author(s): Felix Radford; George Church

Last updated by author(s): Jul 13, 2026

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	Code used for AGENTEX and tSCAN experiments is deposited in Zenodo at https://doi.org/10.5281/zenodo.21321596 and https://doi.org/10.5281/zenodo.21321702
Data analysis	GraphPad Prism 10, Python version 3.10, MMSeqs2

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

All data generated or analyzed during this study are included in this published article (and its supplementary information files). Any additional information needed will be available from the corresponding authors upon reasonable request.

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	*Use the terms sex (biological attribute) and gender (shaped by social and cultural circumstances) carefully in order to avoid confusing both terms. Indicate if findings apply to only one sex or gender; describe whether sex and gender were considered in study design; whether sex and/or gender was determined based on self-reporting or assigned and methods used.* *Provide in the source data disaggregated sex and gender data, where this information has been collected, and if consent has been obtained for sharing of individual-level data; provide overall numbers in this Reporting Summary. Please state if this information has not been collected.* *Report sex- and gender-based analyses where performed, justify reasons for lack of sex- and gender-based analysis.*
Reporting on race, ethnicity, or other socially relevant groupings	*Please specify the socially constructed or socially relevant categorization variable(s) used in your manuscript and explain why they were used. Please note that such variables should not be used as proxies for other socially constructed/relevant variables (for example, race or ethnicity should not be used as a proxy for socioeconomic status).* *Provide clear definitions of the relevant terms used, how they were provided (by the participants/respondents, the researchers, or third parties), and the method(s) used to classify people into the different categories (e.g. self-report, census or administrative data, social media data, etc.)* *Please provide details about how you controlled for confounding variables in your analyses.*
Population characteristics	*Describe the covariate-relevant population characteristics of the human research participants (e.g. age, genotypic information, past and current diagnosis and treatment categories). If you filled out the behavioural & social sciences study design questions and have nothing to add here, write "See above."*
Recruitment	*Describe how participants were recruited. Outline any potential self-selection bias or other biases that may be present and how these are likely to impact results.*
Ethics oversight	*Identify the organization(s) that approved the study protocol.*

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	Sample size was determined by the number of tRNAs tested or translation reactions performed.
Data exclusions	We did not exclude any data.
Replication	Experiments were repeated in triplicate for all assays. The principle next generation sequencing experiments were repeated in duplicate, with > 1000000 measurements per experiment.
Randomization	Randomized mutations were made in libraries of 3' CNA and NNA tRNAs in order to determine patterns of aminoacylation of non-CCA tRNAs. Samples were treated under identical conditions in other experiments, so randomization was not used.
Blinding	Blinding was not relevant to this study. All samples were treated identically, and reported quantitative outputs to molecular changes.

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

Plants

Seed stocks

Report on the source of all seed stocks or other plant material used. If applicable, state the seed stock centre and catalogue number. If plant specimens were collected from the field, describe the collection location, date and sampling procedures.

Novel plant genotypes

Describe the methods by which all novel plant genotypes were produced. This includes those generated by transgenic approaches, gene editing, chemical/radiation-based mutagenesis and hybridization. For transgenic lines, describe the transformation method, the number of independent lines analyzed and the generation upon which experiments were performed. For gene-edited lines, describe the editor used, the endogenous sequence targeted for editing, the targeting guide RNA sequence (if applicable) and how the editor was applied.

Authentication

Describe any authentication procedures for each seed stock used or novel genotype generated. Describe any experiments used to assess the effect of a mutation and, where applicable, how potential secondary effects (e.g. second site T-DNA insertions, mosiacism, off-target gene editing) were examined.