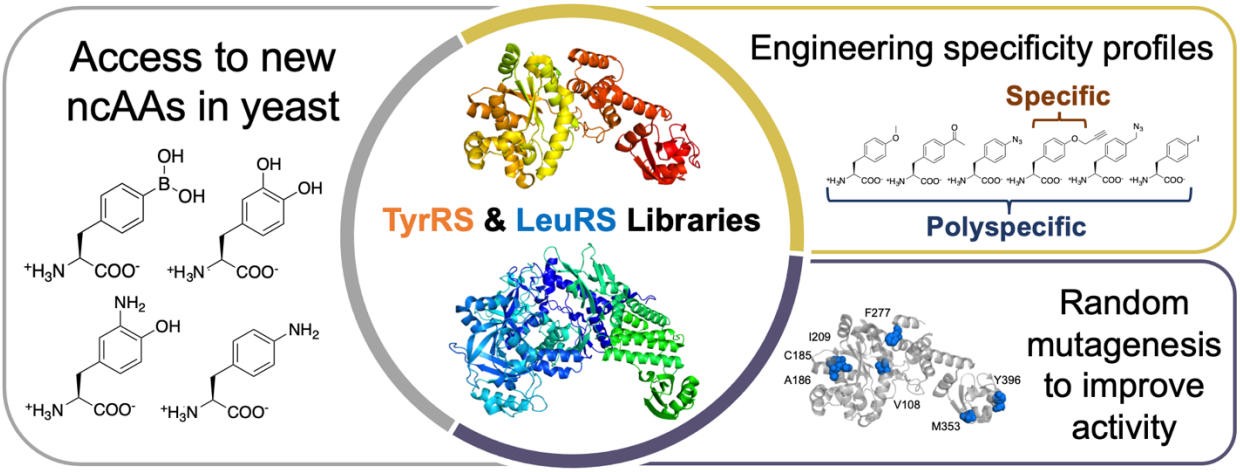
High-throughput aminoacyl-tRNA synthetase engineering for genetic code expansion in yeast



Authors/Affiliations

Jessica T. Stieglitz[†] and James A. Van Deventer[*, †, ‡]

[†]Chemical and Biological Engineering Department, Tufts University, Medford, Massachusetts 02155, USA
‡Biomedical Engineering Department, Tufts University, Medford, Massachusetts 02155, USA

*Correspondence: James.Van_Deventer@tufts.edu

Keywords

Noncanonical amino acids, genetic code manipulation, amber suppression, yeast display, aminoacyl-tRNA synthetase, polyspecificity, fluorescence-activated cell sorting

Abstract

Protein expression with genetically encoded noncanonical amino acids (ncAAs) benefits a broad range of applications, from the discovery of biological therapeutics to fundamental biological studies. A major factor limiting the use of ncAAs is the lack of orthogonal translation systems (OTSs) that support efficient genetic code expansion at repurposed stop codons. Aminoacyl-tRNA synthetases (aaRSs) have been extensively evolved in *E. coli* but are not always orthogonal in eukaryotes. In this work, we use a yeast display-based ncAA incorporation reporter platform with fluorescence-activated cell sorting (FACS) to screen libraries of aaRSs in high throughput for 1) incorporation of ncAAs not previously encoded in yeast; 2) improvement of the performance of an existing aaRS; 3) highly selective OTSs capable of discriminating between closely related ncAA analogs; and 4) OTSs exhibiting enhanced polyspecificity to support translation with structurally diverse sets of ncAAs. The number of previously undiscovered aaRS variants we report in this work more than doubles the total number of translationally active aaRSs available for genetic code manipulation in yeast. The success of myriad screening strategies has important implications related to the fundamental properties and evolvability of aaRSs. Furthermore, access to OTSs with diverse activities and

specific/polyspecific properties are invaluable for a range of applications within chemical biology, synthetic biology, and protein engineering.

Introduction

Diversification of protein function with genetically encoded noncanonical amino acids supports a growing number of applications in chemical and synthetic biology (ncAAs; also called unnatural amino acids, nonnatural amino acids, or nonstandard amino acids). Examples include preparation of proteins with precisely defined post-translational modifications,[1, 2] dissection of protein-protein interactions,[3, 4] discovery of leads for biological therapeutics,[5, 6] and numerous other applications.[7-10] Genetic code expansion depends on the availability of orthogonal translation systems (OTSs), comprised of aminoacyl-tRNA synthetases (aaRSs) that facilitate charging of ncAAs onto tRNAs without interfering with the host organism's native translation machinery.[11] The ncAA-charged tRNA is then translated at the ribosome in response to a stop codon or other codon that does not conventionally encode amino acids.

AaRSs have evolved to maintain the fidelity of the genetic code—that is, to precisely join canonical amino acids (cAAs) with their cognate tRNAs while discriminating against other substrates.[12] Modifying the activities of aaRSs to enable charging of tRNAs (usually suppressor tRNAs such as $tRNA_{CUA}$) with ncAAs facilitates the generation of OTSs.[11, 13] These orthogonal aaRS/tRNA pairs have been extensively engineered to improve the efficiency of ncAA incorporation in proteins in both prokaryotic and eukaryotic hosts, with the vast majority of efforts taking place in *Escherichia coli*.[14-17] Alteration of aaRS activities is commonly achieved using diversification of residues in the aminoacylation active site and a series of positive and negative selections to identify candidates that support incorporation of the target ncAA while discriminating against cAAs. While these approaches are powerful, the most direct and tailorable methods for the discovery and evolution of aaRSs with desirable aminoacylation activities remain unclear. For example, the extent to which aaRSs can be engineered to exhibit specificity toward a single ncAA or to enhance amino acid polyspecificity to support efficient translation with several ncAAs is not known. More broadly, there remain significant questions regarding 1) which selection or screening strategies are most effective for discovering aaRSs that support genetic code expansion; and 2) the extent to which these strategies can be applied to genetic code expansion in organisms other than *E. coli*.

Previous work has shown that selection strategies facilitate the discovery of aaRSs capable of supporting protein translation with ncAAs. Some methods for evolving aaRSs utilize life-or-death assays.[8, 18-20] In *E. coli*, positive selections in the presence of an ncAA utilize an antibiotic resistance gene containing one or more codons to be suppressed to facilitate growth in the presence of an antibiotic with an ncAA-mediated suppression event. Negative selections to enrich for aaRS candidates that discriminate against cAAs commonly utilize a toxic gene containing the codon to be suppressed, resulting in cell death if aaRSs charge cAAs to the tRNA.[18, 21] In yeast, an analogous selection strategy has been used for identifying aaRS variants with improved activity toward ncAAs: *Saccharomyces cerevisiae* strain MaV203 is transformed with a plasmid encoding a GAL4 transcriptional activator gene containing two TAG codons. Suppression events drive the expression of three reporter genes (*HIS3*, *URA3*, and *lacZ*) that can be used to identify aaRS variants with desired readthrough properties.[19, 20] These selection strategies are straightforward to implement in the laboratory and have yielded a number of aaRSs exhibiting altered amino acid substrate preferences. However, these selections tend to lead to only moderately active variants that have unknown and uncontrolled specificity profiles.[22] Moreover, few studies have subjected isolated variants to further mutagenesis and selection (or screening) to improve the function of the initially isolated mutants. Careful adaptations of these

selections have addressed some of these limitations using bioinformatics approaches[23] and counter-selections to refine aaRS specificity[24] in *E. coli*, but these approaches remain in their infancies.

A number of alternative approaches to identifying aaRSs with engineered amino acid preferences have proven to be advantageous over selection-based methodologies. By subjecting aaRSs to continuous mutagenesis and selection over many generations, phage-assisted continuous evolution (PACE) has led to the isolation of aaRS variants that outperform variants isolated via conventional selections.[25] Additionally, several strategies using flow cytometry-based screening have demonstrated the promise of these approaches in discovering and engineering aaRSs. Early efforts described by Tirrell and coworkers demonstrated that aaRS activity and discrimination against cAAs can be evaluated using reporters based on either GFP or on a combination of *E. coli* surface display and bioorthogonal "click" chemistry.[26-29] While these findings were applied to the residue-specific replacement of methionine, later findings evaluating OTSs for site-specific applications appear to be consistent with these earlier studies.[30-33] Eppinger, Rueping, and coworkers evolved *Methanosarcina barkeri* pyrrolysyl-tRNA synthetases (PylRSs) using a high-throughput FACS screen in *E. coli*, identifying a polyspecific variant capable of supporting stop codon readthrough with up to 31 aliphatic, cyclic, or fluorinated ncAAs.[33] Isaacs and coworkers used multiplex automated genome engineering (MAGE) to evolve chromosomally-encoded OTSs with diversity at 12 residues in the amino acid binding pocket and five residues in the tRNA anticodon recognition domain of the *Methanocaldococcus jannaschii* tyrosyl-tRNA synthetase (MjTyrRS). Using a GFP reporter and FACS, they were able to isolate MjTyrRSs with 25-fold improvement of activity for two ncAAs, one of which they further evolved for specificity of a target ncAA and exclusion of 237 non-target ncAAs.[34] Previous work from Chin and coworkers used deep sequencing in combination with high-throughput parallel screens of pyrrolysyl-tRNA synthetases (PylRSs) in search of variants that support efficient protein translation as well as discrimination between two ncAAs.[23] This methodology used only conventional selections, but their large datasets revealed statistically significant differences in aaRS sequences to identify clones with different frequencies between the two sets of selections. Collectively, these advances have resulted in a number of aaRSs suitable for use with genetic code manipulation in *E. coli*. In addition, the mutual orthogonality of some aaRS/tRNA pairs in *E. coli* and mammalian cells, particularly PylRSs, has resulted in a substantial toolkit for genetic code manipulation in higher order eukaryotes. However, an ongoing challenge is that not all aaRSs that can be engineered in *E. coli* can be used in other organisms. One example of this is the MjTyrRS, which is commonly used to encode ncAAs in *E. coli*, but it is not orthogonal in yeast or mammalian cells. Even the heavily utilized PylRSs that can be engineered in *E. coli* and transferred with minimal loss of activity to higher order eukaryotes[35] are not accessible in every organism.

In contrast to *E. coli*, comparatively little work has been done to engineer OTSs in the *S. cerevisiae*, despite its critical role as both a model biological organism and as a key host in many biotechnology applications. *S. cerevisiae* is an ideal organism for evolving *E. coli* aaRSs, as these are naturally orthogonal, and evolution of *E. coli* aaRSs in *E. coli* requires dramatic genomic modifications.[36] AaRSs evolved in yeast may also be transferrable to other eukaryotes, albeit with potential changes in activity.[37] To the best of our knowledge, all yeast-based campaigns to identify aaRSs for ncAA incorporation have utilized the same set of positive and negative selections. Characterizations of selected aaRSs with advanced reporter systems in our group indicate that the performance of aaRSs available in yeast is highly variable, and the available OTSs cover only a limited range of chemical functionalities.[31, 38] Expanding the toolkit for genetic code manipulation in yeast has the potential to lead to fundamental insights into eukaryotic biology and new biotechnologies for engineering proteins and cells. Moreover,

engineering OTSs in yeast may lead to a broader range of OTSs for use in other organisms that cannot currently be accessed using engineering in *E. coli*.

In this work, we established a flow cytometry-based, high-throughput screening platform for discovering aaRSs with a wide range of properties in yeast. To investigate the utility of this platform, we prepared libraries of aaRSs based on both saturation mutagenesis and random mutagenesis approaches. We found that our platform supports several positive and negative screening strategies that facilitate identification of aaRSs exhibiting diverse properties, including 1) aaRSs capable of incorporating ncAAs new to yeast; 2) improved activity of an existing aaRS using random mutagenesis; 3) selective aaRSs that can discriminate between similar ncAA analogs; and 4) aaRSs that support translation with structurally diverse sets of ncAAs (i.e., polyspecific aaRSs). Use of both the *E. coli* tyrosyl-tRNA synthetase (EcTyrRS) and *E. coli* leucyl-tRNA synthetase (EcLeuRS) revealed insights into the range of aaRS performance levels and substrate preferences that can be accessed with appropriate screening conditions. AaRS mutants capable of supporting protein translation with numerous ncAAs were isolated, including variants supporting translation with amino acids not previously reported to be genetically encoded in yeast, including 3,4-dihydroxy-L-phenylalanine, 4-borono-L-phenylalanine, and 3-amino-L-tyrosine. Screens with decreasing concentrations of ncAA in the induction media revealed variants capable of supporting moderately efficient protein translation at ncAA concentrations an order of magnitude lower than typically used in the genetic code expansion field. The expansive set of variants we isolated highlights the range of aaRS substrate preferences that can be engineered, including the discovery of highly specific EcTyrRS variants that discriminate between closely related aromatic ncAAs. On the other hand, screening for aaRSs with enhanced polyspecificity led to the isolation of many variants with broadened substrate preferences. The high plasticity of the EcLeuRS active site was especially noteworthy, with some individual variants supporting efficient translation with both aromatic and aliphatic ncAAs. Overall, our screening platform supports the identification of aaRSs exhibiting a broad scope of properties for genetic code manipulation in yeast. When combined with the powerful protein engineering, genetic, and genomic resources available in yeast, this expanded toolkit is expected to lead to new opportunities in chemical biology, synthetic biology, protein engineering, and many other related disciplines.

Results & Discussion

***E. coli* aaRS saturation mutagenesis library design.** In order to engineer aminoacyl-tRNA synthetases (aaRSs) capable of charging diverse ncAAs to suppressor tRNAs ($tRNA_{CUA}$), we constructed saturation mutagenesis libraries of the *E. coli* tyrosyl-tRNA synthetase (EcTyrRS) and *E. coli* leucyl-tRNA synthetase (EcLeuRS) (Fig. 1a–c). Previous work has shown that EcTyrRS and EcLeuRS are orthogonal to the native translation machinery present in *S. cerevisiae* (i.e. they do not interact with the endogenous yeast machinery), and that they can be used to encode ncAAs in response to the amber stop codon.[14] Using a combination of literature precedents and the crystal structures of the aaRSs, we identified several residues in the substrate binding pocket of each aaRS for mutagenesis.[14] For the EcTyrRS library, degenerate codons at positions Y37, L71, Q179, D182, F183, L186, and Q195 were designed to reduce the size of residues in the active site that were expected to interact with the ncAA to facilitate aminoacylation of the cognate $tRNA_{CUA}^{Tyr}$ (Fig. 1, SI Tables 1 and 2). For the EcLeuRS library, positions M40, L41, S496, Y499, Y527, and H537 were randomized in addition to a T252A mutation in the editing domain known to reduce instances of $tRNA_{CUA}^{Leu}$ being charged with leucine (Fig. 1, SI Tables 3 and 4).[39, 40] The permitted residues in each active site were chosen with the intention of allowing ncAAs with bulkier or longer side chains to better access aaRS substrate binding pockets to be charged to suppressor tRNAs.

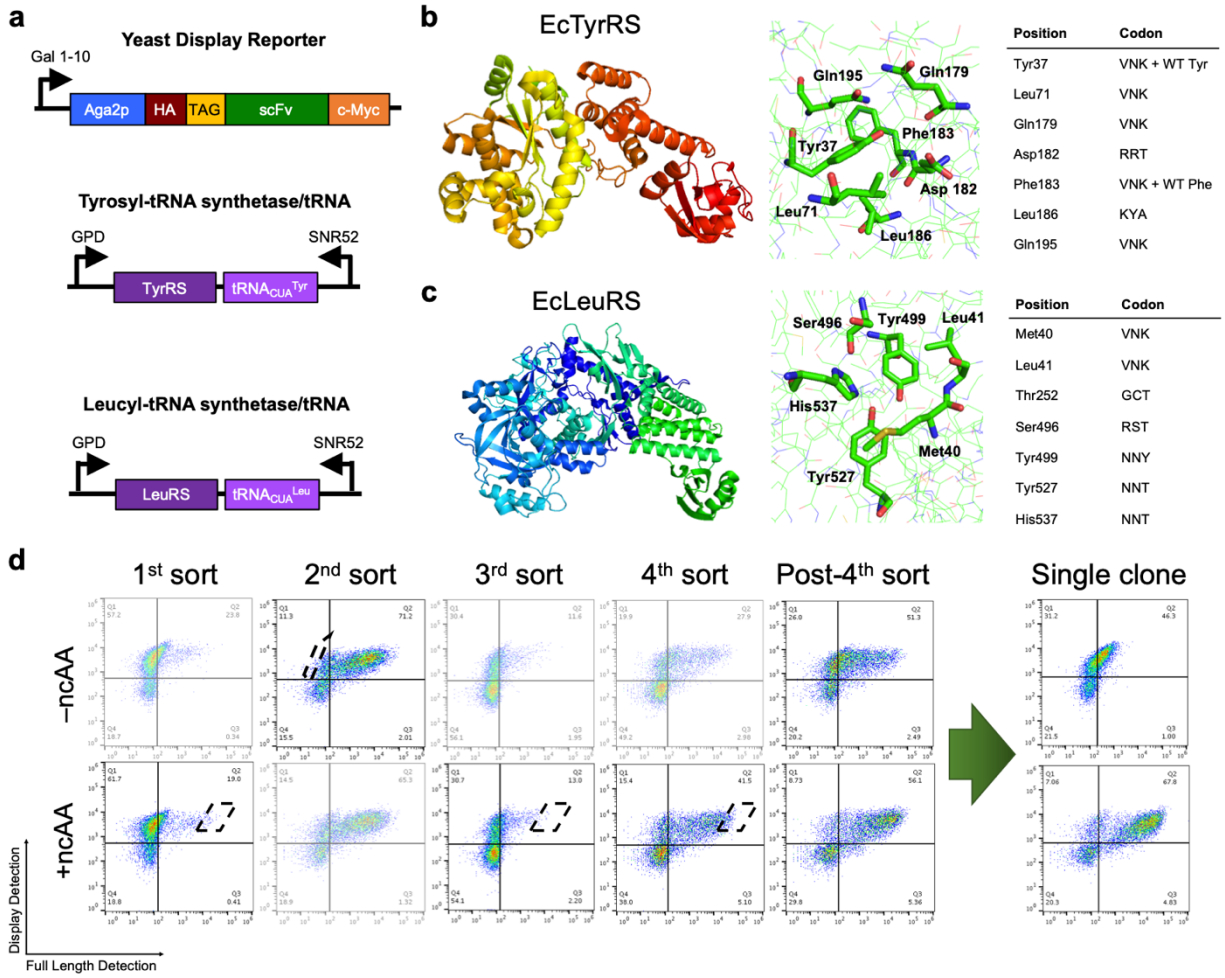


Figure 1. EcTyrRS and EcLeuRS library construction and FACS. **a**, Schematic representations of the genes encoding the ncAA incorporation reporter[31] and the EcTyrRS and EcLeuRS and their respective $tRNA_{CUA}$ molecules. **b**, EcTyrRS crystal structure (PDB 6HB5) with highlighted active site residues chosen for mutation. Residues and degenerate codons with library mutations are listed in the table on the right. Wildtype (WT) EcTyrRS residues were also included at positions 37 and 183. Degenerate codon VNK encodes the following residues: Leu, Pro, His, Gln, Arg, Ile, Met, Thr, Asn, Lys, Ser, Val, Ala, Asp, Glu, and Gly. Degenerate codon RRT encodes the following residues: Asn, Ser, Asp, and Gly. Degenerate codon KYA encodes the following residues: Leu, Ser, Val, and Ala. **c**, EcLeuRS crystal structure (PDB 4CQN) with highlighted active site residues chosen for mutation. Residues and degenerate codons with library mutations are listed in the table on the right. Degenerate codon RST encodes the following residues: Thr, Ala, Ser, and Gly. Degenerate codons NNY and NNT encode the following residues: Phe, Ser, Tyr, Cys, Leu, Pro, His, Arg, Ile, Thr, Asn, Val, Ala, Asp, and Gly. **d**, Example FACS rounds of the pooled EcTyrRS and EcLeuRS libraries for ncAA LysN3. Screens with flow cytometry-based analysis between rounds allowed for progression of screening with either a positive or negative round as needed until a population with low cAA and high ncAA incorporation was attained.

***E. coli* aaRS saturation mutagenesis library construction.** Libraries based on EcTyrRS and EcLeuRS were constructed in *S. cerevisiae* RJY100[41] using homologous recombination and

evaluated for sequence diversity (SI Tables 5 and 6). Each library was co-transformed with a yeast display-based ncAA incorporation reporter that enables simultaneous detection of epitope tags located at the N- and C-termini of the reporter protein.[31] Based on previous studies in which yeast-displayed protein N- and C-terminus detection was evaluated in high throughput, this format was expected to support high-throughput screening.[42] The theoretical diversity of the EcTyrRS library was 1.3×10^8, and the actual number of transformants was calculated to be 1×10^7. All 10 randomly sequenced clones were unique. Sanger sequencing was used to evaluate all naïve and sorted aaRSs; establishing a deep sequencing platform capable of covering the full sequences of genes encoding aaRSs was beyond the scope of this work. The theoretical diversity of the EcLeuRS library was 1.9×10^7, and the calculated number of transformants was 3×10^6. Sequence characterization of the initial preparation of the EcLeuRS library revealed that out of 10 clones, the last two positions chosen for mutation (Y527 and H537) were disproportionately WT residues (SI Table 6). To correct the lack of mutations in those active site residues, a second library was constructed with slightly modified primers and was determined to contain 1×10^7 transformants (SI Table 7). Sequence characterization of nine clones from the reconstructed EcLeuRS library revealed eight unique clones with expected mutations at all positions in the active site (SI Table 8). Since the primary goal of this work was to discover aaRSs with a range of properties, we pooled libraries encoding both EcTyrRS and EcLeuRS variants prior to sorting (SI Fig. 1). To distinguish between screens conducted with the first EcLeuRS library and the reconstructed EcLeuRS library, the original library with predominantly WT residues at positions Y527 and H537 was named Library A, and the reconstructed library was named Library B.

Screening aaRS libraries. Combined libraries were screened via fluorescence-activated cell sorting (FACS) against several aromatic and aliphatic ncAA targets in search of aaRSs capable of charging these ncAAs to suppressor tRNAs (Fig. 2, SI Fig. 1). Each ncAA track was subjected to a combination of positive and negative screens. Positive sorts were conducted following induction in the presence of 1 mM ncAA and cells exhibiting high levels of full-length protein display were recovered (Fig. 1d). Negative sorts were conducted following induction in the absence of ncAAs and cells exhibiting no full-length protein display were recovered (Fig. 1d). Following several rounds of positive and negative screens, we identified enriched populations exhibiting moderate to high levels of full-length protein display following induction in the presence of ncAA and little to no full-length protein display in the absence of ncAAs (Fig. 3, SI Tables 9 and 10). Unique clones from these populations were individually evaluated for relative readthrough efficiency (RRE) and maximum misincorporation frequency (MMF)—quantitative metrics that describe the efficiency and fidelity of the aaRSs, respectively.[30, 31, 38] In general, a high RRE value corresponds to more efficient ncAA incorporation with 0 corresponding to almost all reporter proteins truncated at the TAG codon and 1 corresponding to WT protein translation efficiency. MMF is a measure of the cAA misincorporation of the aaRS at the TAG codon but does not directly indicate a percentage of reporter proteins that contain a cAA at the TAG codon. Rather, MMF provides an approximation of the highest possible cAA misincorporation in the worst-case scenario when there is no ncAA present.

For screens with *O*-methyl-L-tyrosine (OmeY, **1**) conducted with pooled Library A, two positive sorts and one negative sort yielded several individual clones capable of charging OmeY while discriminating against cAAs. The best clone, A-OmeRS-7, exhibited an RRE value of 0.26 ± 0.04 with an MMF value of 0.025 ± 0.004. An ncAA for which no previous studies have demonstrated translation in yeast (to the best of our knowledge) is 3,4-dihydroxy-L-phenylalanine (DOPA, **7**). We isolated five unique clones from pooled Library A that supported DOPA incorporation after four positive and one negative screens, including one (A-DOPARS-4) with an RRE value of 0.14 ± 0.025 and MMF of 0.045 ± 0.01. Screening proceeded similarly for

3-amino-L-tyrosine (ATyr, **11**), (*S*)-2-amino-6-((2-azidoethoxy)carbonylamino)hexanoic acid (LysN3, **15**), and *p*-propargyloxy-L-phenylalanine (OPG, **4**). Clones A-ATyrRS-1, A-LysN3RS-1, and B-OPGRS-L6 were isolated after five positive and two negative screens (pooled Library A), three positive and one negative screens (pooled Library A), and two positive and one negative screens (pooled Library B), respectively. One out of two unique ATyrRS clones, two out of five unique LysN3RS, and three out of four unique OPGRSs successfully charged the ncAA of interest while excluding cAAs (data not shown). For 4-borono-L-phenylalanine (BPhe, **8**), three positive and negative sorts of pooled Library A led to isolation of an aaRS, A-BPheRS-2, that was able to charge BPhe at detectable levels. To our knowledge, no prior reports have demonstrated the incorporation of BPhe in proteins in yeast. Using a selection scheme in genomically engineered *E. coli*, Chatterjee and coworkers isolated an EcTyrRS variant that supports translation with BPhe in both *E. coli* and in HEK293T cells.[36] While this variant exhibited some translation activity when cloned into the yeast expression system used in this work, it appears to be less efficient than A-BPheRS-2 (SI Fig. 2). Sequences and comparisons of aaRSs isolated from these screens can be found in more detail in a later section.

Interestingly, three out of five unique APheRS clones from pooled library A sorts showed some 4-amino-L-phenylalanine (APhe, **12**) incorporation and no cAA misincorporation in the qualitative flow cytometry dot plots, but corresponding RRE values were calculated to be zero (SI Fig. 3). This suggests that quantitative assessments of ncAA incorporation with the stringent RRE metric may not capture aaRS activity levels in cases where partial readthrough is evident. To further characterize the ncAA-containing proteins, several promising aaRS mutants were co-transformed with a secreted reporter protein, purified using Protein A column chromatography, and evaluated via MALDI mass spectrometry analysis. Masses for each ncAA-containing peptide were consistent with the expected masses (SI Fig. 4, SI Table 11).

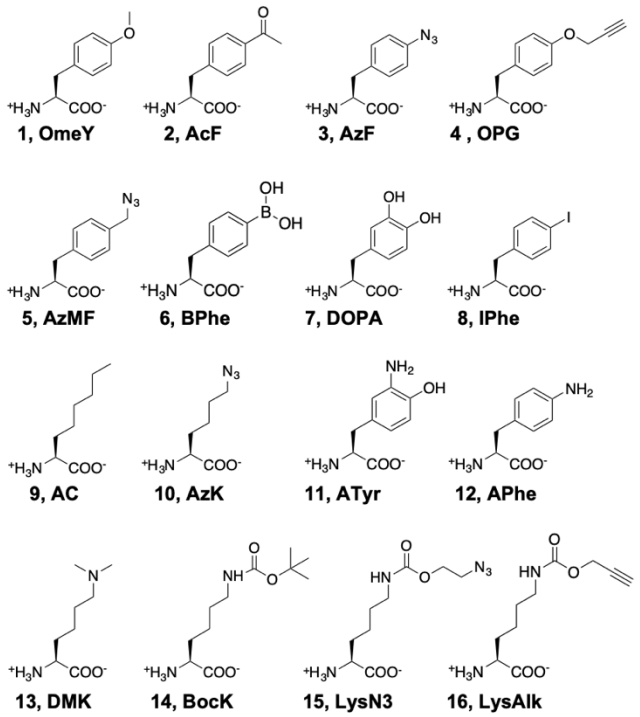


Figure 2. Noncanonical amino acids used in this study. **1**: *O*-methyl-L-tyrosine (OmeY); **2**: *p*-acetyl-L-phenylalanine (AcF); **3**: *p*-azido-L-phenylalanine (AzF); **4**: *p*-propargyloxy-L-

phenylalanine (OPG); **5**: 4-azidomethyl-L-phenylalanine (AzMF); **6**: 4-borono-L-phenylalanine (BPhe); **7**: 3,4-dihydroxy-L-phenylalanine (DOPA); **8**: 4-iodo-L-phenylalanine (IPhe); **9**: L-α-aminocaprylic acid (AC); **10**: N^{ε}-azido-L-lysine (AzK); **11**: 3-amino-L-tyrosine (ATyr); **12**: 4-amino-L-phenylalanine (APhe); **13**: Boc-L-lysine (BocK); **14**: (*S*)-2-amino-6-((2-azidoethoxy)carbonylamino)hexanoic acid (LysN3); **15**: 2-amino-6-(prop-2-ynoxycarbonylamino)hexanoic acid (LysAlk); **16**: N^{ε}, N^{ε}-dimethyl-L-lysine (DMK).

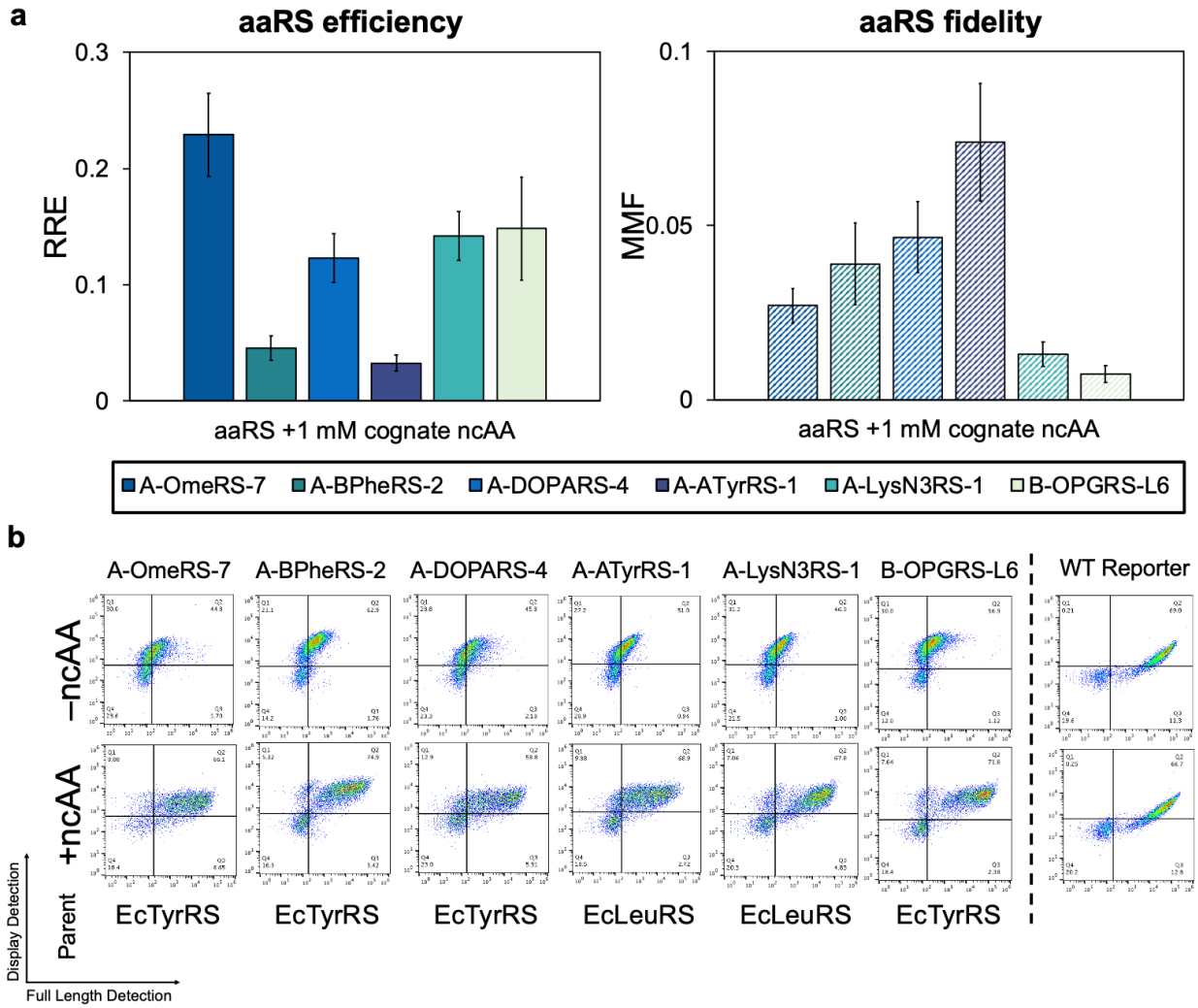


Figure 3. Isolated aaRSs that support translation with unique ncAAs. **a**, RRE and MMF for aaRSs that exhibit activity with target ncAAs OmeY, BPhe, DOPA, ATyr, LysN3, and OPG. **b**, Flow cytometry dot plots for each of the aaRSs in panel **a**. Only one of three biological replicates is shown for each condition (–ncAA and +ncAA). An example of WT display (no TAG codon in reporter sequence) is shown for comparison. The *E. coli* parent enzyme (EcTyrRS or EcLeuRS) is also indicated. All RRE and MMF values were derived from samples evaluated in biological triplicate. Error bars represent the standard deviation of the samples that was propagated during RRE and MMF calculations.

EpPCR library construction and screening. While we successfully isolated aaRSs capable of charging new ncAAs from the EcTyrRS and EcLeuRS libraries during initial sorts, some clones exhibited only modest stop codon readthrough. We sought to further improve the activity of one of these aaRSs with the use of random mutagenesis mediated by error-prone

PCR (epPCR). Some prior work has demonstrated the utility of epPCR in enhancing aaRS activity in *E. coli*.[43-48] However, this approach has never been explored in yeast, and, more broadly, this approach remains underexplored in aaRS engineering. In this case, an EcTyrRS mutant that charges DOPA, A-DOPARS-8, was PCR-amplified with two concentrations of mutagenic dNTPs to introduce random point mutations. Two epPCR libraries, DOPARS-1X and DOPARS-5X, were constructed in *S. cerevisiae* RJY100 with 1.6×10^7 transformants for DOPARS-1X and 2×10^7 transformants for DOPARS-5X. Based on sequence characterization of 12 clones per library, the average number of point mutations per gene in DOPARS-1X was determined to be 4, and in DOPARS-5X the average number of mutations per gene was determined to be 17; additionally, all clones sequenced from the naïve libraries were unique. Both naïve libraries were pooled and screened via FACS following reporter induction in the presence of 1 mM DOPA. Inductions in subsequent positive screening rounds utilized progressively lower concentrations of DOPA (0.1 and 0.05 mM). After four positive screens and one negative screen, several variants were identified with improved ability to support protein display following induction in the presence of 0.1 mM or 1 mM DOPA in comparison to the parent clone (Fig. 4, SI Table 12, SI Fig. 5). Given that most studies with aaRSs utilize ncAA concentrations of 1 mM or higher, the substantial improvement of activity at 0.1 mM is noteworthy.

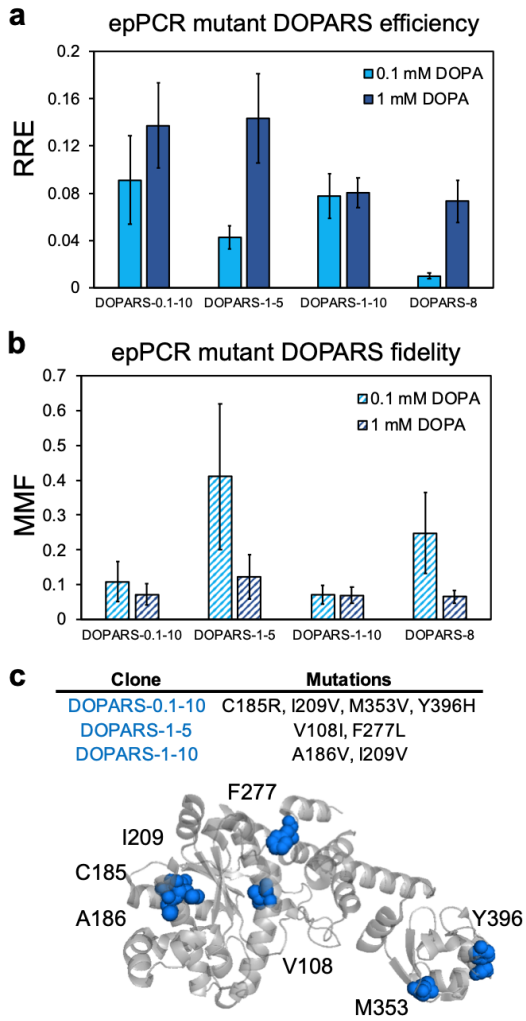


Figure 4. Performance of aaRS mutants isolated from random mutagenesis libraries. **a**, RRE of three epPCR mutants that outperform the parent DOPARS at 0.1 mM, 1 mM, or both concentrations of DOPA. **b**, MMF of three epPCR mutants that are comparable to the parent DOPARS fidelity of DOPA incorporation at both 0.1 and 1 mM DOPA. **c**, Crystal structure of EcTyrRS with mutated residues identified following random mutagenesis and screening highlighted in blue. The crystal structure in this figure was derived from PDB 6HB5. All RRE and MMF values were derived from samples in biological triplicate and error bars represent the standard deviation of the samples that was propagated during RRE and MMF calculations.

Many unique epPCR DOPARSs were identified from sequence characterization, with no clear enrichment for one particular sequence. However, several trends emerged regarding the positions of mutations. Most commonly, clones contained mutations directly adjacent to the active site or in the tRNA binding domain (Fig. 4c, SI Table 12). Clones DOPARS-0.1-2, -5, -9, -10 and DOPARS-1-7, -8, -9, -10, -11, and -12 all contained one or more of the following mutations in the active site: Q18R, T37G, M179N, C185Y/R, A186T/V, or I209V (based on the sequence of A-DOPARS-8). In total, half of the epPCR-derived mutants sequenced had one or more mutations in the active site. Additionally, many variants contained mutations in the $tRNA^{Tyr}$ recognition domain. Key residues in the EcTyrRS tRNA binding domain have been previously reported elsewhere[49] and recently a DOPARS variant of the MjTyrRS was reported by Ellington and coworkers.[44] One of the epPCR clones, DOPARS-0.1-9, had a mutation at a known tRNA recognition site: K377E. Several additional clones had mutations that may be involved with tRNA recognition based on previous structural characterizations, but these mutations are not at locations where direct effects on tRNA recognition have been experimentally demonstrated.[50, 51] While these mutations are not located in the active site or at known tRNA recognition sites, their occurrence in multiple unique clones suggests that they may be important for enhancing EcTyrRS activity. Further investigations of clones possessing these mutations may lead to direct biochemical insights into the individual roles of these mutations. In any case, our findings here indicate that a single round of epPCR followed by increasingly stringent screens substantially improves readthrough activity, even at reduced ncAA concentrations.

Modified screens for isolation of aaRSs with desired specificity profiles. Having validated positive and negative screens with our FACS-based approach, we sought to investigate alternative induction conditions prior to FACS in an attempt to isolate highly specific aaRSs—aaRSs that charge a single ncAA to the cognate $tRNA_{CUA}$ and do not mischarge structurally related ncAAs. Previous selections in yeast[19, 52-54] and the sorts described above did not attempt to control ncAA substrate specificity preferences other than the preference for the target; such conditions can lead to polyspecific aaRSs.[33, 55, 56] For these experiments, we used a group of six structurally similar aromatic ncAAs: OmeY, *p*-acetyl-L-phenylalanine (AcF, **2**), *p*-azido-L-phenylalanine (AzF, **3**), OPG, 4-azidomethyl-L-phenylalanine (AzMF, **5**), and 4-iodo-L-phenylalanine (IPhe, **6**). Based on our previous characterizations of EcTyrRS variant TyrAcFRS[57] and EcLeuRS variant LeuOmeRS,[58] we expected that similarities between these ncAAs would make it difficult for aaRS variants to selectively charge only a single ncAA without also charging others.[31]

We investigated the possibility of isolating aaRSs capable of charging OPG but not the other five ncAAs of this set of six ncAAs with our FACS-based screens. Starting from the pooled EcTyRS and EcLeuRS B libraries, we modified the negative screens and added 1 mM of all five non-target ncAAs during induction, while standard positive screens remain unchanged (1 mM OPG was added during induction for positive sorts). After two positive and three negative rounds of screening, RRE and MMF evaluation of individual clones isolated from these specificity sorts reveal several clones that exhibit substantial translational activity with OPG but

nearly undetectable levels of translation with the other ncAAs in the set (Fig. 5a). Three unique clones, SpecOPGRS-3, -7, and -9, were able to charge OPG at RRE values of 0.05–0.20, with RRE <0.01 for each of OmeY, AcF, AzF, AzMF, and IPhe. In contrast, clone B-OPGRS-L6, which was isolated using conventional positive and negative (no off-target ncAAs) screens, exhibited high levels of polyspecificity, supporting protein translation with four off-target ncAAs (AcF, AzF, OmeY, and IPhe) (Fig. 6a, SI Fig. 6a; see below for further descriptions of characterizations). Out of the four unique OPGRS clones from conventional sorts only three were active and sequence similarity between these clones and clones isolated from polyspecificity sorts (see below) suggested that the OPGRSs would non-specifically encode ncAAs other than OPG (This was confirmed for B-OPGRS-L6 but not for the other two active, unique OPGRSs). There are marked differences between the specificity profiles of aaRSs isolated using specificity screening methods and aaRSs isolated using more standard screening approaches. Tailoring evolutionary pressures can drastically improve the substrate specificity of an aaRS isolated during screens.

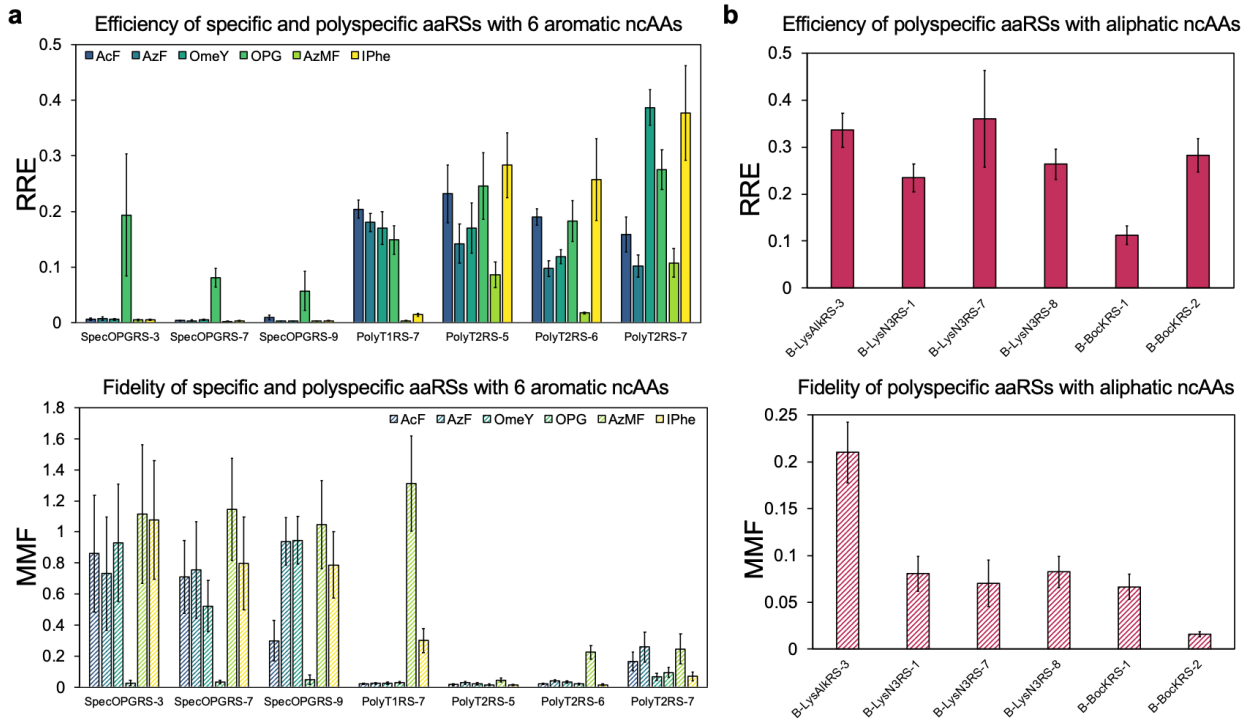


Figure 5. NcAA incorporation efficiency and fidelity of specific and polyspecific aaRSs isolated via FACS. **a**, Evaluation of RRE and MMF of three specific aaRSs (SpecOPGRS-3, SpecOPGRS-7, and SpecOPGRS-9) and four polyspecific aaRSs (PolyT1RS-7, PolyT2RS-5, PolyT2RS-6, and PolyT2RS-7) from Track 1 and Track 2 polyspecificity sorts with a series of six aromatic ncAAs used during screening. For Track 1 sorts, all six ncAAs were added during the induction step prior to positive sorts. For Track 2 sorts, only one ncAA was added during induction prior to a positive sort and subsequent positive sorts were performed with a distinct ncAA from the group of six. **b**, Evaluation of RRE and MMF of six aaRSs originating from a Track 1 polyspecificity screen that encode aliphatic ncAAs LysAlk, LysN3, and BocK. AaRSs were evaluated here for translation activity with the ncAA included in the name of the clone (e.g., B-LysAlkRS-3 was evaluated for translation activity with LysAlk). The concentration of ncAA used during induction was 1 mM for all samples in this figure. All RRE and MMF values were derived from samples evaluated in biological triplicate with error bars representing the standard deviation of the samples that was propagated during RRE and MMF calculations.

a

Relative readthrough efficiency of aaRSs with aromatic ncAAs

	–ncAA	AcF	AzF	OmeY	OPG	AzMF	IPhe	DOPA	BPhe	ATyr	APhe
A-OmeRS-7	0.0017	0.49	0.46	0.40	0.16	0.0030	0.075	0.0051	0.0044	0.0032	0.0037
DOPARS-0.1-10	0.0035	0.011	0.0064	0.0074	0.0051	0.0027	0.0017	0.21	0.0064	0.0062	0.0040
A-BPheRS-2	0.0010	0.0028	0.0035	0.0026	0.0017	0.0012	0.0011	0.0021	0.13	0.0019	0.0019
B-OPGRS-L6	0.0015	0.18	0.17	0.17	0.18	0.029	0.27	0.0037	0.0033	0.0026	0.0079
SpecOPGRS-3	0.0025	0.0040	0.0043	0.0031	0.10	0.0017	0.0019	0.0044	0.0034	0.0028	0.0024
PolyT2RS-5	0.0012	0.15	0.16	0.16	0.22	0.036	0.23	0.0024	0.0020	0.0017	0.0080
PolyT2RS-7	0.011	0.21	0.075	0.32	0.18	0.045	0.41	0.022	0.014	0.016	0.011
B-LysAlkRS-3	0.048	0.058	0.087	0.010	0.094	0.14	0.30	0.066	0.068	0.055	0.045
B-BocKRS-2	0.0023	0.35	0.36	0.34	0.35	0.064	0.38	0.0010	0.0017	0.0021	0.0019

Maximum misincorporation frequency of aaRSs with aromatic ncAAs

	AcF	AzF	OmeY	OPG	AzMF	IPhe	DOPA	BPhe	ATyr	APhe
A-OmeRS-7	0.0035	0.0037	0.0042	0.011	0.56	0.023	0.34	0.39	0.53	0.46
DOPARS-0.1-10	0.32	0.55	0.47	0.68	1.3	2.0	0.017	0.55	0.56	0.86
A-BPheRS-2	0.36	0.29	0.39	0.61	0.86	0.92	0.49	0.008	0.53	0.54
B-OPGRS-L6	0.0083	0.0087	0.0086	0.0080	0.051	0.0055	0.39	0.45	0.57	0.19
SpecOPGRS-3	0.61	0.58	0.81	0.025	1.5	1.3	0.57	0.72	0.87	1.0
PolyT2RS-5	0.0079	0.0074	0.0075	0.0054	0.033	0.0051	0.48	0.60	0.69	0.15
PolyT2RS-7	0.053	0.15	0.034	0.060	0.24	0.027	0.49	0.77	0.67	1.0
B-LysAlkRS-3	0.83	0.55	0.48	0.51	0.35	0.16	0.73	0.71	0.88	1.1
B-BocKRS-2	0.0064	0.0063	0.0067	0.0064	0.035	0.0060	0.23	1.3	1.1	1.2

b

RRE of aaRSs with aliphatic ncAAs

	–ncAA	LysN3	LysAlk	BocK	AzK	DMK	AC
PolyT2RS-5	0.0025	0.0019	0.0015	0.0013	0.0021	0.0018	0.0015
PolyT2RS-7	0.018	0.039	0.012	0.010	0.021	0.016	0.12
B-LysAlkRS-3	0.048	0.42	0.30	0.10	0.042	0.052	0.10
B-BocKRS-2	0.0024	0.054	0.024	0.20	0.006	0.011	0.041
B-LysN3RS-7	0.013	0.31	0.012	0.012	0.012	0.013	0.016
LeuOmeRS	0.0017	0.0020	0.0011	0.0010	0.0017	0.0014	0.0019
A-OmeRS-7	0.0013	0.0030	0.0027	0.0023	0.0022	0.0017	0.0018
B-OPGRS-L6	0.0014	0.0033	0.0031	0.0024	0.0019	0.0019	0.0017

MMF of aaRSs with aliphatic ncAAs

	LysN3	LysAlk	BocK	AzK	DMK	AC
PolyT2RS-5	1.3	1.7	2.0	1.2	1.4	1.7
PolyT2RS-7	0.45	1.4	1.7	0.83	1.1	0.14
B-LysAlkRS-3	0.11	0.16	0.49	1.1	0.92	0.46
B-BocKRS-2	0.044	0.10	0.012	0.42	0.21	0.06
B-LysN3RS-7	0.042	1.1	1.1	1.1	1.0	0.81
LeuOmeRS	0.86	1.5	1.9	1.0	1.3	0.90
A-OmeRS-7	0.44	0.49	0.57	0.61	0.75	0.72
B-OPGRS-L6	0.41	0.44	0.56	0.71	0.70	0.78

Scale: 0.5, 0.4, 0.3, 0.2, 0.1, 0.0

Scale: 2.0, 1.5, 1.0, 0.8, 0.6, 0.3, 0.0

Figure 6. Comprehensive evaluation of aaRS activity with aromatic and aliphatic ncAAs. **a**, RRE and MMF for nine aaRSs with 10 aromatic ncAAs. **b**, RRE and MMF for eight aaRSs with six aliphatic ncAAs. The concentration of ncAA used during induction was 1 mM for all samples in this figure. All RRE and MMF values were derived from samples evaluated in biological triplicate and error can be found in SI Figure 6.

Modified screens for isolation of aaRSs with desired polyspecificity profiles. We also sought to investigate whether we could enhance aaRS polyspecificity using altered screening methods via two strategies: Track 1 and Track 2. For Track 1, we added all six of the same group of aromatic ncAAs (OmeY, AcF, AzF, OPG, AzMF, and IPhe) at 1 mM to the induction media for positive sorts. In a Track 1 series of screens consisting of three consecutive positive sorts followed by a single negative sort, characterization yielded four unique EcTyrRS mutants and one EcLeuRS mutant (SI Tables 9 and 10). For Track 2, only one of the six aromatic ncAAs was added during induction per positive sort round, with a different ncAA added for each subsequent positive sort. The B pooled libraries were screened first using AzF, followed by a positive screen with AcF, a negative screen, and two consecutive positive screens for AzMF due to our observation of reduced incorporation of AzMF in comparison to the other five ncAAs during flow cytometry characterization of the sorted populations (SI Fig. 7). Out of 11 individual clones isolated from the Track 2 sorted library population, seven were unique: five EcTyrRS variants and two EcLeuRS variants (SI Tables 9 and 10).

Unique clones from Track 1 and 2 polyspecificity sorts were first evaluated for incorporation of each of the six aromatic ncAAs and misincorporation of cAAs. Evaluation of the efficiency (RRE) and fidelity (MMF) of the most active polyspecific aaRSs demonstrates that the aaRSs were able to encode several of the group of six ncAAs, and also reveals a key difference in outcome between the two sort tracks (Fig. 5a). For Track 2 sorts, we performed two consecutive AzMF positive sorts to ultimately isolate aaRSs that encoded all six ncAAs, as flow cytometry characterizations between sorts indicated reduced AzMF incorporation compared to the other five ncAAs. In the case of Track 1, the most active clone (PolyT1RS-7) encodes only four out of the six ncAAs, which was known during flow cytometry characterizations between sort rounds but was not controllable since all six ncAAs were added during all inductions. Interestingly, an identical sequence to known polyspecific enzyme TyrAcFRS[31, 57] was found seven times in the Track 1 population and two times in the Track 2 population. Each polyspecificity sort strategy has its advantages and shortcomings. Track 1 is straightforward but does not provide control over which ncAA is being encoded by the aaRSs for any given sort round. Track 2 requires careful monitoring of individual ncAA incorporation by the sorted library populations after each sort but allows for more control of individual ncAA incorporation

as needed. By making simple modifications to the induction step prior to screening, aaRSs with polyspecific characteristics can be isolated for use in applications where a single aaRS that can encode multiple ncAAs is desirable. Furthermore, both strategies could be readily applied to other types of screens for isolation of aaRSs with tailored specificity profiles in *E. coli* or other organisms, and with screens that use methods other than FACS.

Characterizing polyspecific aaRSs for activity with aliphatic ncAAs. To further evaluate the extent of polyspecificity of the final screened Track 1 library population, we tested the population for incorporation of 21 ncAAs (SI Fig. 8). Interestingly, while the Track 1 aaRSs were only screened against a group of aromatic ncAAs, they also appeared to support translation with several aliphatic ncAAs at low levels. To investigate if aaRSs from the Track 1 library population were polyspecific for ncAAs beyond the original six aromatic ncAAs, we further screened the library for three aliphatic ncAAs: Boc-L-lysine (BocK, **14**), LysN3, and 2-amino-6-(prop-2-ynoxycarbonylamino)hexanoic acid (LysAlk, **16**). After one positive screen for incorporation of BocK, nine clones were sequenced. All clones were determined to be EcLeuRS variants, six of which were unique (one sequence appeared four times). For the LysAlk sorts, sequence characterization of 11 clones yielded all EcLeuRS clones, seven of which were unique (one sequence appeared five times). The consensus sequence for the LysAlk population was identical to a clone that appeared multiple times in the BocK population. Sequence characterization of 12 clones in the LysN3 population revealed eight unique EcLeuRS variants. One of the eight unique EcLeuRS clones appeared three times, and two other clones appeared twice each, including a sequence that appeared in both the LysAlk and BocK populations (SI Tables 9 and 10). We evaluated the efficiency and fidelity of ncAA incorporation for several clones from each population and discovered that some were able to encode the aliphatic ncAA target at levels comparable to the Track 1 and Track 2 RRE values for the original six aromatic ncAAs with similarly low misincorporation of cAAs (Fig. 5b). It is noteworthy that while very few EcLeuRS variants were identified during characterization of the original Track 1 and 2 screens, all of the aaRSs isolated from the aliphatic ncAA sorts were derived from EcLeuRS. Overall, these results demonstrate that polyspecificity screens can lead to the identification of aaRSs that support translational activity with structurally diverse ncAAs spanning aromatic and aliphatic side chains.

Trends in aaRS active site residues. We compared the sequences of EcTyrRS and EcLeuRS variants isolated via the various screening efforts and identified several trends regarding which active site mutations supported translational activity with particular ncAAs. For example, 11 out of the 12 EcTyrRS variants isolated from non-specificity sorts for OPG contained mutations D182G, F183M, and L186A, and maintained WT residues at positions Q179 and Q195—from herein referred to as “QGMAQ”. This is particularly interesting when compared to the OPGRSs isolated from specificity screens, for which the active site residues varied significantly more (SI Tables 9 and 10). It is possible that QGMAQ improves the efficiency of charging OPG to the $tRNA_{CUA}^{Tyr}$ but also allows OPG analogs to undergo aminoacylation and translation and was consequently removed from the library population during specificity screens. QGMAQ appears in all of the unique EcTyrRS clones from both Track 1 and Track 2 polyspecificity sorts. The only positions that differ between the polyspecific aaRSs derived from EcTyrRS are at positions Y37, L71, and V72. Position V72 was not included in the original set of active site residues chosen for mutation and did not appear in the initial EcTyrRS library characterization. However, mutations at V72 appeared in several EcTyrRS clones isolated from different screens and may be attributable to a mutation introduced during PCR amplification prior to library construction. For the polyspecific aaRSs, the Y37, L71, and V72 positions generally only contained mutations to leucine, isoleucine, and valine. These trends in active site residues suggest that QGMAQ contributes to polyspecific

aminoacylation of some ncAAs, and that small, hydrophobic residues may further alter the active site conformation in a way that reduces aaRS selectivity.

EcTyrRS variants supporting incorporation of LysN3 showed a similar trend at positions Y37 and L71, where a majority of residues were alanine, valine, leucine, and methionine (all hydrophobic). However, QGMAQ did not appear, and all seven unique clones retained the WT aspartic acid at position 182. LysN3RSs isolated from EcLeuRS Library A revealed a disproportionately high rate of glycine and threonine appearing across all active site positions. The other group of LysN3RSs were variants of EcLeuRS Library B and were derived from polyspecificity Track 1. All 12 of the aaRSs isolated from this population were EcLeuRS variants and contained the expected T252A mutation. For the six active site positions diversified in the original library construction, some very interesting trends were discovered: M40, L41, S496, Y499, and H537 were primarily mutated to glycine, proline, or alanine with a few instances of serine and isoleucine residues. In particular, virtually all of the clones were mutated to glycine at both positions S496 and H537. As S496 has not, to our knowledge, been included in EcLeuRS saturation mutagenesis library design, these results are intriguing. Previously screened EcLeuRS libraries containing diversity at H537 have also noted convergence to glycine at this position.[59] Brustad et al postulated that as H537 is directly positioned over the leucine side chain when it is in the EcLeuRS active site, the H537G mutation may simultaneously increase the size of the active site cavity while reducing interaction with the native leucine substrate. Position Y527 showed the most variability with residues such as cysteine, histidine, and threonine appearing most often. The frequency of glycine and alanine mutations, particularly in clones B-T1-LysN3RS-1, -7, and -8 (all of which support high levels of protein translation with LysN3) raises the possibility that these mutations alter the active site specificity to facilitate aminoacylation of large, aliphatic ncAAs. This notion facilitate is further supported by comparison of aaRS sequences capable of charging the ncAAs LysAlk and BocK. With one exception, all 24 clones isolated from sorts for BocK and LysAlk encoded glycine at the H537 position of EcLeuRS. All clones had the expected T252A mutation, and similarly to the LysN3 mutants, primarily glycine, proline, or alanine at positions M40, L41, and Y499. Residue S496 was virtually always a glycine or serine and residue H527 mainly converged to threonine. Perhaps the most surprising observation from the LysAlkRS clones was that in the best-performing LysAlkRS identified, B-T2-LysAlkRS-3, residue M40 was unmutated and residue L41 was mutated to histidine. This clone was the only EcLeuRS variant isolated from any screen in this work that contained an L41H mutation.

Additionally, we investigated trends between the ncAA screening targets and whether more EcTyrRS or EcLeuRS variants were isolated from the final library populations after screening. For sorts where a single ncAA of interest was the target, aaRSs isolated tended to predominantly originate in the EcTyrRS if the ncAA contained a benzyl ring, or EcLeuRS if the ncAA was aliphatic (SI Tables 9 and 10). Clonal characterizations revealed that aaRSs sorted for the ability to charge OmeY, DOPA, BPhe, OPG, APhe, and the polyspecificity sorts for groups of aromatic ncAAs consisted of 80-100% EcTyrRS clones. Similarly, for sorts for the incorporation of aliphatic ncAAs, clones of aaRSs for all three sorts that started as polyspecificity sorts against aromatic ncAAs but were then steered toward the aliphatic ncAAs BocK, LysAlk, and LysN3 were all derived from EcLeuRS. There are two notable exceptions: LysN3 from sorts with EcLeuRS Library A and ATyr. For both of these sets, EcTyrRS and EcLeuRS clones appeared in approximately equal numbers in sequence characterizations. For LysN3, this may be attributable to the predominance of WT residues tyrosine and histidine at positions 527 and 537, respectively. Despite this predominance, the EcLeuRS variants isolated from screens with Library A were all found to be mutated at position 537, which could be an indication of the importance of mutating H537 to accommodate LysN3 in the active site. For

ATyr, it is not immediately clear why half of the isolated aaRSs were found to be EcLeuRS variants, where even the highest performing variant was derived from EcLeuRS. Further aaRS screens could be used to more systematically determine the range of chemical diversity that each of the EcTyrRS and EcLeuRS are capable of charging to suppressor tRNAs to support protein translation. The aaRS libraries constructed here were able to encode many unique ncAAs by one or both of the parent aaRSs, indicating that the library designs were sufficiently diverse and could be used to screen for incorporation of additional ncAAs beyond those investigated in this work.

Comparison between aaRSs sorted using conventional versus modified methods. In order to directly compare the performance of aaRSs isolated using distinct screening methods, we determined the efficiency and fidelity of nine aaRSs with 10 aromatic ncAAs and eight aaRSs with six aliphatic ncAAs (Fig. 6). For aaRSs evolved using conventional positive and negative screens, clones exhibited variable levels of specificity or polyspecificity. For example, A-BPheRS-2 and DOPARS-0.1-10 both exhibited specificity towards the ncAAs BPhe and DOPA, respectively, whereas A-OmeRS-7 and B-OPGRS-L6 were able to encode five of the group of 10 aromatic ncAAs at moderate to high levels (Fig. 6a, SI Fig. 6a). Conversely, for aaRSs isolated from intentionally polyspecific or specific sorts, the behavior of the aaRSs were consistent with types of screens employed. Specificity clone SpecOPGRS-3 only supports stop codon suppression with OPG and not any of the other aromatic ncAAs evaluated here. Polyspecificity clones PolyT2RS-5 and PolyT2RS-7 perform well with five out of the original six aromatic ncAAs they were screened for incorporation with and support translation with the sixth ncAA, AzMF, at low but detectable levels (AzMF incorporation is apparent in the flow cytometry dot plots, but here the stringency of the RRE metric resulted in the calculation of background-level values; see SI Fig. 3 for dot plots). Interestingly, these polyspecificity clones do not appear to support high levels of incorporation with any of the other four aromatic ncAAs tested here.

The two aaRSs originally identified using Track 1 polyspecific methods and then isolated after subsequent screens with aliphatic ncAAs LysAlk and BocK both exhibit highly polyspecific behavior toward the original set of six aromatic ncAAs, with notably lower AzMF incorporation. We noted that APhe also supports some degree of translation with PolyT2RS-5 based on flow cytometry dot plots, although again RRE calculations resulted in values close to background levels (SI Fig. 3). We attribute the low RRE values that occur, despite the presence of detectable readthrough events in flow cytometry plots, to the following: cells that have lost the suppression machinery and are incapable of reading through the TAG codon are included in the determinations of N- and C-terminal display levels, which lowers the overall median fluorescence intensity (MFI). With enough reduction, the result is near-zero RRE values that do not capture the readthrough events that are apparent in flow cytometry plots. This effect is most noticeable for aaRSs exhibiting highly variable cell-to-cell activity levels or generally low activity levels. Despite near-zero RRE values, the corresponding MMF values were still low in comparison to aaRSs with no qualitative readthrough and zero RRE values. This raises the possibility that low MMF values may be a way to identify translational activity in cases of low or variable translational activity, but further investigations are needed to thoroughly investigate this conjecture. In addition, modifications to gating strategies or the use of mean or geometric fluorescence intensities in place of median fluorescence intensities for calculating RRE may lead to values that better capture low but nonzero aaRS activity levels.

We also evaluated several aaRSs against a panel of aliphatic ncAAs (Fig. 6b, SI Fig. 6b). Polyspecificity clone PolyT2RS-5 showed virtually no incorporation of any of the aliphatic ncAAs tested, but PolyT2RS-7 incorporated L-α-aminocaprylic acid (AC, **9**) with an RRE value

of 0.12 ± 0.025. LeuOmeRS, A-OmeRS-7, and B-OPGRS-L6, all of which were isolated following screens or selections for incorporation of a single ncAA, did not support stop codon readthrough with any of the aliphatic ncAAs tested. The highest performing aliphatic ncAA-encoding aaRSs (that were screened for aromatic ncAA incorporation prior to screens for incorporation of individual aliphatic ncAAs) were B-LysAlkRS-3, B-BocKRS-2, and B-LysN3RS-7. B-LysN3RS-7 encoded LysN3 efficiently with an RRE value of 0.31 ± 0.051 but did not encode any other aliphatic ncAAs tested. B-BocKRS-2 similarly encoded BocK well (RRE of 0.20 ± 0.033) but showed little to no activity for any other aliphatic ncAAs. Unlike its BocK and LysN3 counterparts, B-LysAlkRS-3 not only supported translation with LysAlk efficiently (RRE of 0.30 ± 0.024), but also was translationally active with LysN3, BocK, and AC. Notably, B-LysAlkRS-3 exhibits a slightly higher background level of cAA misincorporation, as determined by MMF measurements. The ability to take a polyspecificity screen designed to encode aromatic ncAAs and isolate aaRSs additionally capable of charging aliphatic ncAAs further demonstrates the extent of polyspecificity achievable using slightly modified induction steps prior to aaRS screens.

Another interesting trend we observed was differences in polyspecificity between EcTyrRS and EcLeuRS variants isolated from various screens. The highest performing EcTyrRS variants isolated from screens for aminoacylation of OmeY and OPG exhibited polyspecificity with aromatic ncAAs AcF, AzF, OmeY, OPG, and IPhe (as well as AzMF in the case of B-OPGRS-L6), but these variants were not able to support translation with any of the aliphatic ncAAs tested (Fig. 6). On the other hand, some EcTyrRS clones such as DOPARS-0.1-10 and A-BPheRS-2 showed excellent substrate specificity for their cognate ncAAs. While PolyT2RS-5 efficiently produced proteins containing five out of the six aromatic ncAAs it was originally screened against, and demonstrated detectable translational activity with APhe, it was not able to charge the other three aromatic or six aliphatic ncAAs evaluated. However, PolyT2RS-7 was able to support translation with the same aromatic ncAAs as PolyT2RS-5 as well as LysN3 and AC at low but detectable levels. Similarly, EcLeuRS clones LysAlkRS-3 and BocKRS-2 both supported stop codon suppression with some aromatic ncAAs, as well as multiple aliphatic ncAAs. In particular, LysAlkRS-3 charged BocK and AC at moderate levels and LysAlk and LysN3 at high levels. Conversely, B-LysN3RS-7 did not charge any other aliphatic ncAAs tested. These results suggest that EcLeuRS variants are capable of exhibiting broader levels of polyspecificity than EcTyrRS variants. Taken together, the EcTyrRS and EcLeuRS variants isolated in this work exhibit a range (poly)specificity profiles, some of which were only accessible after tailoring screening conditions to maximize (poly)specificity. These findings raise interesting questions about the limits of aaRS substrate specificity and polyspecificity and point to strategies for systematically investigating such questions.

Conclusions

In this work, we employed flow cytometry-based screens to discover aaRSs exhibiting a wide range of properties for genetic code expansion in yeast. Isolation of clones from saturation mutagenesis EcTyrRS and EcLeuRS libraries led to numerous variants supporting protein translation with a broad set of ncAAs, including DOPA, BPhe, and other ncAAs that had not previously been genetically encoded in yeast. However, we note that not all ncAAs we attempted discovery with yielded functional aaRSs, suggesting that further library design based on our findings may be necessary to expand the substrates that can be accepted by engineered EcTyRSs and EcLeuRSs. With this platform, we were also able to improve the properties of an existing aaRS via error-prone PCR. Random mutagenesis of a DOPARS variant followed by increasingly stringent screening led to identification of clones that support improved protein translation with DOPA even at reduced ncAA concentrations. The facile

discovery of improved variants in a single round of mutagenesis suggests the possibility that this platform will facilitate more extensive, multi-round aaRS discovery and mutagenesis campaigns in the future—a generally underexplored route to modifying aaRS activity. Moreover, these findings highlight the strong potential to enhance OTS performance by broadly exploring aaRS diversification strategies beyond aaRS aminoacylation active sites; this observation is consistent with the findings of other recent work in this area.[25, 43-45]

By tailoring the conditions under which positive and negative screens for ncAA incorporation were performed, we were able to isolate aaRSs with a wide range of specificity and polyspecificity profiles. Inclusion of off-target ncAAs during negative screens enabled us to identify aaRS variants that incorporate OPG into proteins while discriminating against five closely related aromatic ncAAs. In the absence of these ncAAs, we were only able to isolate polyspecific aaRSs that support translation with OPG, in line with previous reports.[53] On the other hand, when attempting to identify clones exhibiting higher levels of polyspecificity, both induction in the presence of multiple ncAAs (Track 1) and induction in the presence of different ncAAs in subsequent FACS rounds (Track 2) led to the isolation of clones exhibiting enhanced polyspecificity. While each of these approaches was successful, we note some important differences between the approaches. Track 1 is straightforward because the same set of ncAAs is utilized during induction for each positive screen, but this approach does not allow for control over which individual ncAAs are incorporated into reporter proteins during screening. Track 2 requires careful (and somewhat laborious) monitoring of individual ncAA incorporation after each sort, but these evaluations allow for the inductions prior to each sort to be precisely modified to maximize polyspecificity across a set of ncAAs. Intriguingly, by taking a population already enriched for polyspecific aromatic ncAA incorporation and screening against a set of aliphatic ncAAs, we identified aaRSs capable of supporting efficient protein translation with both aromatic and aliphatic ncAAs. Access to both highly specific aaRSs and highly polyspecific aaRSs are both desirable in different applications of genetic code manipulation.

Our powerful screening platform for engineering aaRSs in yeast generated many mutants capable of incorporating structurally diverse ncAAs into proteins. The breadth of aaRS properties accessed in this work underscores the excellent plasticity and “evolvability” of these enzymes. We attribute the functional diversity of the mutants reported here primarily to the carefully controlled screening conditions, with both flow cytometry gating strategies and well-defined induction conditions playing key roles in biasing screening outcomes. There are several ways in which the findings described here could be extended further in future work. First, detailed sequence-activity relationships for the aaRSs investigated here may be attainable, especially if deep sequencing methodologies can be applied.[23] Second, understanding the relationship between the observed translation properties reported here and underlying OTS properties (e.g., kinetic constants of aaRSs, expression levels of OTS components, and expression conditions) could lead to better understanding of how best to efficiently prepare ncAA-containing proteins in high yields and purities. AaRS kinetics have the potential to provide critical insights into how best to enhance the properties of the protein translation machinery used to manipulate the genetic code. Third, the availability of highly specific OTSs has the potential to facilitate genetic code expansion to include multiple ncAAs in the same protein, even when the two ncAAs of interest are similar in structure. Finally, polyspecific aaRSs have potential utility in applications where protein medicinal chemistry to systematically explore the effects of different ncAA side chains on protein properties is desirable.[60, 61] Our findings begin to explore the most efficient way to select sets of ncAAs that lead to tightly controlled aaRS activity using modified specificity profile screens. Additional work using various sets of ncAAs may lead to maximally polyspecific aaRSs that are still able to discriminate against cAAs.

Beyond changing the ncAA substrate preferences of aaRSs, our findings indicate that there remain substantial opportunities to enhance aaRS activities by exploring mutations beyond aminoacylation active sites. Here, a single round of random mutagenesis led to the discovery of aaRS variants possessing a diverse set of mutations exhibiting improved translational activity. These findings are consistent with screens performed by other groups in *E. coli*, where random mutagenesis and screening led to rapid improvement of aaRS properties by the introduction of mutations at positions outside of aminoacylation active sites. Further studies using random mutagenesis, deep mutational scanning,[62] or other approaches to facilitate more comprehensive exploration of the sequence spaces surrounding known aaRSs represents a major opportunity for understanding and engineering aaRSs. In addition to advancing discovery of highly active, next-generation OTSs, such studies are expected to lead to fundamental insights into the biochemical properties of aaRSs. Overall, the availability of a high-throughput screening platform for aaRSs in yeast broadens opportunities for generating versatile aaRSs suitable for use in genetic code manipulation in yeast, mammalian cells, and other eukaryotes. Such tools are expected to facilitate dissection of basic biological and biochemical phenomena as well as myriad applications at the interface of chemical biology, synthetic biology, and protein engineering.

Materials and Methods

All restriction enzymes used for molecular biology were from New England Biolabs (NEB). Synthetic oligonucleotides for cloning and sequencing were purchased from Eurofins Genomics or GENEWIZ. All sequencing in this work was performed by Eurofins Genomics (Louisville, KY) or Quintara Biosciences (Cambridge, MA). Epoch Life Science GenCatch™ Plasmid DNA Mini-Prep Kits were used for plasmid DNA purification from *E. coli*. Yeast chemically competent cells and subsequent transformations were prepared using Zymo Research Frozen-EZ Yeast Transformation II kits. NcAAs were purchased from the indicated companies: *p*-acetyl-L-phenylalanine (SynChem), *p*-azido-L-phenylalanine (Chem-Impex International), *O*-methyl-L-tyrosine (Chem-Impex International), *p*-propargyloxy-L-phenylalanine (Iris Biotech), 4-azidomethyl-L-phenylalanine (SynChem), 4-iodo-L-phenylalanine (AstaTech), 3,4-dihydroxy-L-phenylalanine (Alfa Aesar), 4-borono-L-phenylalanine (Acros Organics), 3-Amino-L-tyrosine (Bachem), 4-Amino-L-phenylalanine (Bachem), (*S*)-2-amino-6-((2-azidoethoxy)carbonylamino)hexanoic acid (Iris Biotech), (*S*)-2-amino-6-(((prop-2-yn-1-yloxy)carbonyl)amino)hexanoic acid (AstaTech), N^{ε}-Boc-L-lysine (Chem-Impex International), N^{ε}-azido-L-lysine (Chem-Impex International), N^{ε},N^{ε} -dimethyl-L-lysine (Chem-Impex International), and L-α-aminocaprylic acid (Acros Organics).

Media preparation and yeast strain construction

The preparation of liquid and solid media was performed as described previously.[31] Unless otherwise noted, all SD-SCAA and SG-SCAA media used here were prepared without tryptophan (TRP), leucine (LEU) or uracil (URA). The strain RJY100 was constructed using standard homologous recombination approaches as described previously.[41]

Preparing ncAA acid liquid stocks

All ncAA stocks were prepared at a final concentration of 50 mM L-isomer. DI water was added to the solid ncAA to approximately 90% of the final volume needed to make the stock, and 6.0 N NaOH was used as needed to fully dissolve the ncAA powder in the water. Water was added to the final volume and the solution was sterile filtered through a 0.2 micron filter. OmeY was pH adjusted to 7 prior to sterile filtering. No pH adjustment of additional ncAAs was performed unless otherwise noted. Filtered solutions were stored at 4 °C for up to four weeks for less labile

ncAAs; for more labile ncAAs (AzF, BPhe, DOPA), 50 mM stocks were made immediately prior to induction.

Reporter and secretion plasmid construction
The pCTCON2-FAPB2.3.6 and pCTCON2-FAPB2.3.6L1TAG reporter constructs have been previously described.[31] pCTCON2-FAPB2.3.6 was used as a WT control to compare TAG-containing samples against, and was not used for library construction or sorting. The construction of reporter constructs pCHA-Donkey1.1-TAA and pCHA-Donkey1.1-H54TAG-TAA has also been reported previously.[63]

EcTyrRS and EcLeuRS Library construction and characterization
Construction of plasmids pRS315-KanRmod-AcFRS[63] and pRS315-EcLeuRS[31] has been described previously. The EcLeuRS gene and cognate $tRNA_{CUA}$, as well as the constitutive promoters for each gene, were PCR-amplified from pRS315-EcLeuRS and cloned into pRS315-KanRmod-AcFRS digested with SacI and PstI via Gibson assembly. The resulting plasmid was sequence verified and named pRS315-KanRmod-EcLeuRS.

Primers containing degenerate codons were used to amplify the aaRS genes from parent plasmids pRS315-KanRmod-EcLeuRS (EcLeuRS library) or pRS315-KanRmod-AcFRS (EcTyrRS library). Seven positions in the EcTyrRS active site were chosen for mutation: Y37, L71, Q179, N182, F183, L186, and Q195 (SI Tables 1 and 2). A separate primer with only the WT tyrosine codon at position Y37 was also used. The AcFRS gene contained a preexisting D165G mutation. An additional mutation, I7M, was inadvertently introduced when a primer containing that mutation was received and used for PCR amplification of the gene. However, a side-by-side comparison of AcFRS with and without the I7M mutations showed that the activity of AcFRS was not significantly affected by the presence of the mutation (SI Fig. 9). pRS315-KanRmod-AcFRS was digested with restriction enzymes NcoI and NdeI and the PCR-amplified EcTyrRS gene fragments were concentrated using Pellet Paint® NF Co-precipitant according to the manufacturer's protocols. Similarly, seven positions in the EcLeuRS active site were chosen for mutation: M40, L41, T252, S496, Y499, Y527, and H537, with only an alanine mutation at position T252 (SI Tables 3, 4, and 7). pRS315-KanRmod-EcLeuRS was digested with restriction enzymes NcoI and NdeI and the PCR-amplified EcLeuRS gene fragments were concentrated using Pellet Paint® NF Co-precipitant according to the manufacturer's protocols. A control DNA sample for each aaRS library that contained only doubly digested vector and no insert DNA was prepared similarly.

Electrocompetent *S. cerevisiae* RJY100 were prepared as described previously and electroporation protocols for transforming the library DNA were also followed as previously described.[64] Electroporated cells were plated to determine transformation efficiency on SD-SCAA (−TRP −LEU −URA),[64] and the remainder of cells were recovered in 100 mL SD-SCAA (−TRP −LEU −URA) supplemented with penicillin-streptomycin. One electroporation was performed for each library (EcTyrRS, EcLeuRS Library A, and EcLeuRS Library B). Dilutions plated on solid media to evaluate transformation efficiency were grown at 30 °C for 3–4 days and colonies were counted from multiple dilutions to approximate the number of individual transformants.[64] The remainder of the libraries were grown at 30 °C with shaking at 300 rpm to saturation, then expanded into 1 L SD-SCAA (−TRP −LEU −URA) and grown at 30 °C overnight. The 1 L cultures were centrifuged at 3,214 rcf for 30 min and the supernatant was decanted. The cell pellets were resuspended in 60% glycerol to a final concentration of 15% glycerol, aliquoted to cryogenic vials, and stored at –80 °C. At least 5×10^9 cells were stored per vial. A portion of each library was passaged separately in 5 mL SD-SCAA (−TRP −LEU −URA) cultures to characterize the naïve libraries using flow cytometry and to isolate and

sequence the aaRS plasmids from several individual colonies (see "AaRS characterization post-FACS" for details).

Error-prone PCR (epPCR) library construction and characterization
EcTyrRS mutant A-DOPARS-8 was used as a template for epPCR. epPCR was performed by combining 5 µL 10X ThermoPol Buffer, 1 µL 10 mM dNTPs, 5 µL 20 µM or 100 µM dPTP, 5 µL 20 µM or 100 µM 8-oxo-dGTP, 1 µL Taq polymerase, 1 µL DNA template (1 ng total), and 2.5 µL of each forward and reverse primer at 10 µM to amplify across the entire aaRS gene, as well as 27 µL sterile water to bring the total volume to 50 µL. Two concentrations (20 µM or 100 µM) of mutagenic dNTPs were used to vary the number of mutations made across the aaRS. Reactions were run on the thermal cycler at 95 °C for 500 s followed by 16 cycles of 95 °C for 45 s, 60 °C for 30 s, 72 °C for 135 s. Once cycles were complete, samples underwent a 10 min 72 °C final extension and hold at 4 °C until they were removed from the thermal cycler.

Following PCR with mutagenic dNTPs, each gene was amplified again via PCR at a higher volume to prepare enough DNA for electroporation into yeast. PCR was performed by combining 20 µL 10X ThermoPol Buffer, 4 µL 10 mM dNTP, 4 µL Taq polymerase, 10 µL epPCR-mutated DNA template, and 2 µL of each forward and reverse primer at 100 µM to amplify across the entire aaRS gene, as well as 158 µL sterile water to bring the total volume to 200 µL. Reactions were run on the thermal cycler at 95 °C for 180 s followed by 30 cycles of 95 °C for 45 s, 55 °C for 30 s, 72 °C for 135 s. Once cycles were completed, samples underwent a 10 min 72 °C final extension and hold at 4 °C until they were removed from the thermal cycler.

Digested pRS315-KanRmod-AcFRS vectors (still containing $tRNA_{CUA}^{Tyr}$) were prepared in the same manner as for the EcTyrRS saturation mutagenesis library (see above). For each of the DOPARS epPCR libraries, the following masses of DNA were concentrated using the Pellet Paint procedure: 4 µg epPCR-amplified aaRS, 1 µg double digested pRS315-KanRmod vector, and 1 µg pCTCON2-FAPB2.3.6L1TAG. Preparation of electrocompetent cells, electroporations, and subsequent characterization proceeded in the same manner as construction of EcTyrRS and EcLeuRS libraries (see above).

Yeast transformations, propagation, and induction
Reporter plasmids pCTCON2-FAPB2.3.6L1TAG or pCTCON2-FAPB2.3.6 (TRP marker) and pRS315 plasmids from which the aaRS/tRNA pairs are expressed (LEU marker) were co-transformed into Zymo competent RJY100 cells, plated on solid SD-SCAA media (−TRP −LEU −URA), and grown at 30°C until colonies appeared (3 days). WT controls containing only pCTCON2-FAPB2.3.6 were transformed similarly into Zymo competent RJY100 cells, plated on solid SD-SCAA media (−TRP −URA), and grown at 30°C until colonies appeared (3 days). Inoculation and propagation in liquid SD media and induction in SG media have been described in detail elsewhere.[31] Briefly: Three separate transformant colonies (biological triplicates) were inoculated from each plate except the WT control, where only one colony was inoculated, in 5 mL SD media of the same composition as the plates from transformations. All liquid cultures were supplemented with penicillin-streptomycin to prevent bacterial contamination. Liquid cultures were grown to saturation and then diluted to OD_{600} = 1 in 5 mL of the same media. The diluted cultures were grown to OD_{600} 2–5 (4–6 h at 30 °C with shaking) and then induced in 2 mL SG media at OD_{600} 1. Induction cultures with no ncAA and 1 mM final concentration of each respective ncAA were prepared for each replicate. The WT control was only induced with no ncAA. In the case of the epPCR aaRSs, induction cultures with 0.1 mM final concentration of the respective ncAA were also prepared. Induced cultures were incubated at 20 °C with shaking at 300 rpm for 16 h.

For library propagation and induction, the steps were identical but in 100 mL media inoculated from a 2 mL glycerol stock of the library with propagation and inductions in 100 mL in order to preserve the full diversity of the library.

Flow cytometry data collection and analysis

Freshly induced samples were labeled in 1.7 mL microcentrifuge tubes or 96-well V-bottom plates. Flow cytometry was performed on an Attune NxT flow cytometer (Life Technologies) at the Tufts University Science and Technology Center. Detailed protocols describing the antibody labeling process have been described elsewhere.[31] Primary and secondary antibody names and dilutions can be found in SI Table 13. Flow cytometry data analysis was performed using FlowJo and Microsoft Excel. Detailed descriptions of the calculations for RRE and MMF with corresponding error propagation have been described previously.[31, 32]

Fluorescence-activated cell sorting (FACS)

Aminoacyl-tRNA synthetase library populations were induced and labeled using identical methods to the "Flow cytometry data collection and analysis" section.[31] For naïve library screens, 6×10^7 cells were used for antibody labeling and antibody/PBSA volumes for primary and secondary labeling were adjusted accordingly. A number of cells was used that was at minimum ten times larger than the library population being sorted for all subsequent screens. Cell pellets were resuspended in ice-cold PBSA immediately prior to sorting. Samples were sorted using a FACSAria™ III (Becton, Dickinson and Company) flow cytometer at the Tufts University Flow Cytometry Core. Some first round sorts of Pooled Library A were performed on a MoFlo (Beckman-Coulter) and sorted in all subsequent rounds on the FACSAria™ III. Sorted samples were collected in 14 mL culture tubes containing 1 mL SD-SCAA (−TRP −LEU −URA) supplemented with penicillin-streptomycin. Following sorting, the sides of the culture tubes were washed with an additional 1 mL SD-SCAA (−TRP −LEU −URA), then transported back to the main laboratory facilities. An additional 3 mL SD-SCAA (−TRP −LEU −URA) was added to each sample and cultures were then grown at 30 °C with shaking at 300 rpm until saturated (2–3 days). Subsequent flow cytometry characterization was performed on each sorted population before the following round of screening (see above for details).

AaRS characterization post-FACS

Once library populations exhibiting low stop codon readthrough in the presence of only cAAs and higher stop codon readthrough in the presence of ncAA(s) were isolated, aaRS plasmid DNA was purified using a Zymoprep Yeast Plasmid Miniprep II kit with slightly modified manufacturer's protocols. 500 µL of 5 mL cultures library populations that had been previously propagated for flow cytometry characterization were diluted into 4.5 mL of the same media, supplemented with penicillin-streptomycin. Cultures were grown for 4 h at 30 °C with shaking at 300 rpm. 1 mL of each culture was removed to a microcentrifuge tube and pelleted at 13,000 rcf for 30 s. Supernatant was aspirated and each pellet was resuspended in 200 µL Solution I with 6 µL reconstituted zymolase from the Zymo kit. Each sample was vortexed briefly and then incubated at 37 °C with shaking at 300 rpm overnight or up to 24 h. 200 µL of Solution II from the Zymo kit was added and tubes were inverted to mix. 400 µL of Solution III was added and tubes were inverted to mix. Samples were pelleted at 13,000 rcf for 30 min to separate out cell debris. The supernatant was transferred to Epoch Life Science *E. coli* DNA purification columns and purified using the manufacturer's protocols. DNA was eluted in 40 µL sterile water and then transformed into chemically competent *E. coli* DH5alphaZ1 cells and plated on LB media with 34 µg/mL kanamycin. 10-12 individual colonies were inoculated into separate 5 mL LB cultures with 50 µg/mL kanamycin and grown overnight at 37 °C with shaking at 300 rpm. Cultures were miniprepped using an Epoch *E. coli* GenCatch™ Plasmid DNA Mini-Prep Kit and submitted for sequencing.

Secreted protein expression
Secretion plasmids pCHA-Donkey1.1-TAA and pCHA-Donkey1.1-L93TAG-TAA (constructed and described previously by Van Deventer and coworkers[63]) were transformed into chemically competent RJY100 using same method as described in the "Yeast transformations, propagation, and induction" section. Plasmid pCHA-Donkey1.1-TAA was not transformed with a second plasmid, while pCHA-Donkey1.1-L93TAG-TAA was co-transformed with pRS315 plasmids expressing the aaRS/$tRNA_{CUA}$ pairs to be evaluated. All cells transformed with the WT construct were grown on or in SD-CAA during propagation and induced in YPG. All other cells cotransformed with pCHA constructs were grown on or in SD-SCAA (−TRP −LEU −URA) for growth and induced in YPG. All liquid cultures for growth and induction were supplemented with penicillin-streptomycin. Transformations, selection, and initial propagation in liquid media were essentially identical to pCTCON2 transformations in the "Yeast transformations, propagation, and induction" section. Following the initial inoculation of a single transformant colony into 5 mL SD selective media, cultures were grown at 30 °C for 24–48 h with shaking at 300 rpm. Each 5 mL culture was diluted into 45 mL SD media and grown for 24 h at 30 °C with shaking at 300 rpm. 50 mL cultures were then diluted into 450 mL SD media and grown for 24 h at 30 °C with shaking at 300 rpm. For inductions, 500 mL cultures were transferred to Nalgene centrifuge bottles and pelleted for 10 min at 2400 rcf. Supernatant was decanted and pellets were resuspended in 1 L YPG media with 1 mM final concentration of the corresponding ncAA and induced for 96 h at 20 °C with shaking at 300 rpm.

Secreted protein purification
Induced cultures were centrifuged in Nalgene centrifuge bottles at 3,214 rcf for 30 min and the supernatant was sterile filtered along with a volume of 10X PBS buffer (1.37 M NaCl, 27 mM KCl, 100 mM Na_2HPO_4, and 18 mM KH2PO4 at pH 7.4) to bring the filtrate to a final concentration of 1X PBS. Protein purification columns were washed with 25 mL 1X PBS pH 7.0 before adding 2 mL Protein A resin (GenScript) and then washed with 25 mL 1X PBS. The equilibrated supernatant was added to each column and passed over the resin twice to improve protein capture efficiency. The resin was then washed three times with 1X PBS. Proteins were eluted in 7 mL 100 mM glycine pH 3.0 into a conical tube containing 0.7 mL 1M Tris pH 8.5 for immediate neutralization. Elution fractions were buffer exchanged into sterile water chilled to 4 °C using 15 mL 30 kDA molecular weight cutoff devices (Millipore Sigma). Buffer-exchanged proteins were flash frozen in 50% v/v glycerol and stored at –80 °C.

Tryptic digests and sample preparation for mass spectrometry
Frozen glycerol stocks of secreted proteins (Donkey1.1) were thawed in a water bath at room temperature and buffer exchanged into sterile water chilled to 4 °C using 0.5 mL 30 kDA molecular weight cutoff devices (Millipore Sigma). A volume of protein corresponding to 10 µg total Donkey1.1 was boiled at 100 °C for 5 min and then cooled to room temperature on the benchtop. 1 µg Trypsin Gold (Promega) was added and samples were incubated at 37 °C overnight. Following tryptic digests, all samples were desalted using ZipTip C18 (Millipore Sigma) protocols, flash frozen in liquid nitrogen, and then shipped on dry ice to a mass spectrometry core facility.

Acknowledgments
This work was supported in part by a grant from the Army Research Office (W911NF-16-1-0175), in part by a grant from the National Institute of General Medical Sciences of the National Institutes of Health (1R35GM133471), and in part by Tufts University startup funds (to J.A.V.). J.T.S. was supported in part by an NSF Graduate Research Fellowship (ID: 2016231237). The content is solely the responsibility of the authors and does not necessarily represent the official

views of the Army Research Office, the National Institutes of Health, the National Science Foundation, or Tufts University. We would like to thank Stephen Kwok and Allen Parmelee for their expertise and assistance with FACS and Michael Ream for his help with error-prone PCR experiments. Additionally, we would like to thank Dr. Richard Cook, Heather Amoroso, and Alla Leshinsky at the Koch Institute Biopolymers and Proteomics Core for their assistance with mass spectrometry data collection. Finally, we thank Dr. Rafael Alcala-Toraño and Rebecca Hershman for their invaluable editing contributions and feedback.

Author Contributions

Conceptualization, J.T.S. and J.A.V.; Methodology, J.T.S. and J.A.V.; Investigation, J.T.S.; Writing – Original Draft, J.T.S.; Writing – Review & Editing, J.T.S. and J.A.V.; Funding Acquisition, J.A.V.; Supervision, J.A.V.

Declaration of Interests

Authors J.T.S. and J.A.V. have submitted a patent application for the aaRS sequences identified in this work: U.S. Provisional Application No. 63/190,336. A version of this manuscript is included in the PhD thesis of J.T.S.

References

[1] Venkat, S., Sturges, J., Stahman, A., Gregory, C., Gan, Q., and Fan, C. (2018) Genetically Incorporating Two Distinct Post-translational Modifications into One Protein Simultaneously, *ACS Synth Biol 7*, 689-695.

[2] Liu, W. R., Wang, Y. S., and Wan, W. (2011) Synthesis of proteins with defined posttranslational modifications using the genetic noncanonical amino acid incorporation approach, *Mol Biosyst 7*, 38-47.

[3] Zheng, Y., Gilgenast, M. J., Hauc, S., and Chatterjee, A. (2018) Capturing Post-Translational Modification-Triggered Protein-Protein Interactions Using Dual Noncanonical Amino Acid Mutagenesis, *ACS Chem Biol 13*, 1137-1141.

[4] Coin, I. (2018) Application of non-canonical crosslinking amino acids to study protein-protein interactions in live cells, *Curr Opin Chem Biol 46*, 156-163.

[5] Rezhdo, A., Islam, M., Huang, M., and Van Deventer, J. A. (2019) Future prospects for noncanonical amino acids in biological therapeutics, *Curr Opin Biotechnol 60*, 168-178.

[6] Kim, C. H., Axup, J. Y., and Schultz, P. G. (2013) Protein conjugation with genetically encoded unnatural amino acids, *Curr Opin Chem Biol 17*, 412-419.

[7] Davis, L., and Chin, J. W. (2012) Designer proteins: applications of genetic code expansion in cell biology, *Nat Rev Mol Cell Biol 13*, 168-182.

[8] Liu, C. C., and Schultz, P. G. (2010) Adding new chemistries to the genetic code, *Annu Rev Biochem 79*, 413-444.

[9] Neumann, H. (2012) Rewiring translation - Genetic code expansion and its applications, *FEBS Lett 586*, 2057-2064.

[10] Wang, Q., Parrish, A. R., and Wang, L. (2009) Expanding the genetic code for biological studies, *Chem Biol 16*, 323-336.

[11] Soye, B. J. D., Patel, J. R., Isaacs, F. J., and Jewett, M. C. (2015) Repurposing the translation apparatus for synthetic biology, *Curr Opin Chem Biol 28*, 83-90.

[12] Sankaranarayanan, R., and Moras, D. (2001) The fidelity of the translation of the genetic code, *Acta Biochim Pol 48*, 323-335.

[13] Liu, C. C., Jewett, M. C., Chin, J. W., and Voigt, C. A. (2018) Toward an orthogonal central dogma, *Nat Chem Biol 14*, 103-106.

[14] Wiltschi, B. (2016) Incorporation of non-canonical amino acids into proteins in yeast, *Fungal Genet Biol 89*, 137-156.

[15] Dumas, A., Lercher, L., Spicer, C. D., and Davis, B. G. (2015) Designing logical codon reassignment - Expanding the chemistry in biology, *Chem Sci 6*, 50-69.
[16] Italia, J. S., Zheng, Y., Kelemen, R. E., Erickson, S. B., Addy, P. S., and Chatterjee, A. (2017) Expanding the genetic code of mammalian cells, *Biochem Soc Trans 45*, 555-562.
[17] Chin, J. W. (2017) Expanding and reprogramming the genetic code, *Nature 550*, 53-60.
[18] Wals, K., and Ovaa, H. (2014) Unnatural amino acid incorporation in E. coli: current and future applications in the design of therapeutic proteins, *Front Chem 2*, 15.
[19] Chin, J. W., Cropp, T. A., Anderson, J. C., Mukherji, M., Zhang, Z., and Schultz, P. G. (2003) An expanded eukaryotic genetic code, *Science 301*, 964-967.
[20] Wang, L., and Schultz, P. G. (2004) Expanding the genetic code, *Angew Chem Int Ed Engl 44*, 34-66.
[21] Santoro, S. W., Wang, L., Herberich, B., King, D. S., and Schultz, P. G. (2002) An efficient system for the evolution of aminoacyl-tRNA synthetase specificity, *Nat Biotechnol 20*, 1044-1048.
[22] Young, D. D., Young, T. S., Jahnz, M., Ahmad, I., Spraggon, G., and Schultz, P. G. (2011) An evolved aminoacyl-tRNA synthetase with atypical polysubstrate specificity, *Biochemistry 50*, 1894-1900.
[23] Zhang, M. S., Brunner, S. F., Huguenin-Dezot, N., Liang, A. D., Schmied, W. H., Rogerson, D. T., and Chin, J. W. (2017) Biosynthesis and genetic encoding of phosphothreonine through parallel selection and deep sequencing, *Nat Methods 14*, 729-736.
[24] Sungwienwong, I., Hostetler, Z. M., Blizzard, R. J., Porter, J. J., Driggers, C. M., Mbengi, L. Z., Villegas, J. A., Speight, L. C., Saven, J. G., Perona, J. J., Kohli, R. M., Mehl, R. A., and Petersson, E. J. (2017) Improving target amino acid selectivity in a permissive aminoacyl tRNA synthetase through counter-selection, *Org Biomol Chem 15*, 3603-3610.
[25] Bryson, D. I., Fan, C., Guo, L. T., Miller, C., Soll, D., and Liu, D. R. (2017) Continuous directed evolution of aminoacyl-tRNA synthetases, *Nat Chem Biol 13*, 1253-1260.
[26] Van Deventer, J. A., Yuet, K. P., Yoo, T. H., and Tirrell, D. A. (2014) Cell surface display yields evolvable, clickable antibody fragments, *Chembiochem 15*, 1777-1781.
[27] Beatty, K. E., Liu, J. C., Xie, F., Dieterich, D. C., Schuman, E. M., Wang, Q., and Tirrell, D. A. (2006) Fluorescence visualization of newly synthesized proteins in mammalian cells, *Angew Chem Int Ed Engl 45*, 7364-7367.
[28] Mahdavi, A., Szychowski, J., Ngo, J. T., Sweredoski, M. J., Graham, R. L., Hess, S., Schneewind, O., Mazmanian, S. K., and Tirrell, D. A. (2014) Identification of secreted bacterial proteins by noncanonical amino acid tagging, *Proc Natl Acad Sci U S A 111*, 433-438.
[29] Yoo, T. H., and Tirrell, D. A. (2007) High-throughput screening for methionyl-tRNA synthetases that enable residue-specific incorporation of noncanonical amino acids into recombinant proteins in bacterial cells, *Angew Chem Int Ed Engl 46*, 5340-5343.
[30] Monk, J. W., Leonard, S. P., Brown, C. W., Hammerling, M. J., Mortensen, C., Gutierrez, A. E., Shin, N. Y., Watkins, E., Mishler, D. M., and Barrick, J. E. (2017) Rapid and Inexpensive Evaluation of Nonstandard Amino Acid Incorporation in Escherichia coli, *ACS Synth Biol 6*, 45-54.
[31] Stieglitz, J. T., Kehoe, H. P., Lei, M., and Van Deventer, J. A. (2018) A Robust and Quantitative Reporter System To Evaluate Noncanonical Amino Acid Incorporation in Yeast, *ACS Synth Biol 7*, 2256-2269.
[32] Potts, K. A., Stieglitz, J.T., Lei, M., Van Deventer, J.A. (2020) Reporter system architecture affects measurements of noncanonical amino acid incorporation efficiency and fidelity, *Mol. Syst. Des. Eng. 5*, 573-588.

[33] Hohl, A., Karan, R., Akal, A., Renn, D., Liu, X., Ghorpade, S., Groll, M., Rueping, M., and Eppinger, J. (2019) Engineering a Polyspecific Pyrrolysyl-tRNA Synthetase by a High Throughput FACS Screen, *Sci Rep 9*, 11971.
[34] Amiram, M., Haimovich, A. D., Fan, C., Wang, Y. S., Aerni, H. R., Ntai, I., Moonan, D. W., Ma, N. J., Rovner, A. J., Hong, S. H., Kelleher, N. L., Goodman, A. L., Jewett, M. C., Soll, D., Rinehart, J., and Isaacs, F. J. (2015) Evolution of translation machinery in recoded bacteria enables multi-site incorporation of nonstandard amino acids, *Nat Biotechnol 33*, 1272-1279.
[35] Chen, P. R., Groff, D., Guo, J. T., Ou, W. J., Cellitti, S., Geierstanger, B. H., and Schultz, P. G. (2009) A Facile System for Encoding Unnatural Amino Acids in Mammalian Cells, *Angewandte Chemie-International Edition 48*, 4052-4055.
[36] Italia, J. S., Latour, C., Wrobel, C. J. J., and Chatterjee, A. (2018) Resurrecting the Bacterial Tyrosyl-tRNA Synthetase/tRNA Pair for Expanding the Genetic Code of Both E. coli and Eukaryotes, *Cell Chem Biol*.
[37] Hendel, S. J., and Shoulders, M. D. (2021) Directed evolution in mammalian cells, *Nat Methods 18*, 346-357.
[38] Potts, K. A., Stieglitz, J. T., Lei, M., and Van Deventer, J. A. (2020) Reporter system architecture affects measurements of noncanonical amino acid incorporation efficiency and fidelity, *Mol Syst Des Eng 5*, 573-588.
[39] Mursinna, R. S., and Martinis, S. A. (2002) Rational design to block amino acid editing of a tRNA synthetase, *J Am Chem Soc 124*, 7286-7287.
[40] Mursinna, R. S., Lincecum, T. L., and Martinis, S. A. (2001) A conserved threonine within Escherichia coli leucyl-tRNA synthetase prevents hydrolytic editing of leucyl-tRNA(Leu), *Biochemistry 40*, 5376-5381.
[41] Van Deventer, J. A., Kelly, R. L., Rajan, S., Wittrup, K. D., and Sidhu, S. S. (2015) A switchable yeast display/secretion system, *Protein Eng Des Sel 28*, 317-325.
[42] Hackel, B. J., Ackerman, M. E., Howland, S. W., and Wittrup, K. D. (2010) Stability and CDR composition biases enrich binder functionality landscapes, *J Mol Biol 401*, 84-96.
[43] Owens, A. E., Grasso, K. T., Ziegler, C. A., and Fasan, R. (2017) Two-Tier Screening Platform for Directed Evolution of Aminoacyl-tRNA Synthetases with Enhanced Stop Codon Suppression Efficiency, *Chembiochem 18*, 1109-1116.
[44] Thyer, R., d'Oelsnitz, S., Blevins, M. S., Klein, D. R., Brodbelt, J. S., and Ellington, A. D. (2021) Directed Evolution of an Improved Aminoacyl-tRNA Synthetase for Incorporation of L-3,4-Dihydroxyphenylalanine (L-DOPA), *Angew Chem Int Ed Engl*.
[45] Truong, F., Yoo, T. H., Lampo, T. J., and Tirrell, D. A. (2012) Two-strain, cell-selective protein labeling in mixed bacterial cultures, *J Am Chem Soc 134*, 8551-8556.
[46] Schwark, D. G., Schmitt, M. A., and Fisk, J. D. (2021) Directed Evolution of the Methanosarcina barkeri Pyrrolysyl tRNA/aminoacyl tRNA Synthetase Pair for Rapid Evaluation of Sense Codon Reassignment Potential, *Int J Mol Sci 22*.
[47] Kwok, H. S., Vargas-Rodriguez, O., Melnikov, S. V., and Soll, D. (2019) Engineered Aminoacyl-tRNA Synthetases with Improved Selectivity toward Noncanonical Amino Acids, *ACS Chem Biol*.
[48] Kuhn, S. M., Rubini, M., Fuhrmann, M., Theobald, I., and Skerra, A. (2010) Engineering of an orthogonal aminoacyl-tRNA synthetase for efficient incorporation of the non-natural amino acid O-methyl-L-tyrosine using fluorescence-based bacterial cell sorting, *J Mol Biol 404*, 70-87.
[49] Bedouelle, H. (1990) Recognition of tRNA(Tyr) by tyrosyl-tRNA synthetase, *Biochimie 72*, 589-598.
[50] Labouze, E., and Bedouelle, H. (1989) Structural and kinetic bases for the recognition of tRNATyr by tyrosyl-tRNA synthetase, *J Mol Biol 205*, 729-735.

[51] Hountondji, C., Lederer, F., Dessen, P., and Blanquet, S. (1986) Escherichia coli tyrosyl- and methionyl-tRNA synthetases display sequence similarity at the binding site for the 3'-end of tRNA, *Biochemistry 25*, 16-21.
[52] Chin, J. W., Cropp, T. A., Chu, S., Meggers, E., and Schultz, P. G. (2003) Progress toward an expanded eukaryotic genetic code, *Chem Biol 10*, 511-519.
[53] Deiters, A., Cropp, T. A., Mukherji, M., Chin, J. W., Anderson, J. C., and Schultz, P. G. (2003) Adding amino acids with novel reactivity to the genetic code of Saccharomyces cerevisiae, *J Am Chem Soc 125*, 11782-11783.
[54] Liu, D. R., and Schultz, P. G. (1999) Progress toward the evolution of an organism with an expanded genetic code, *Proc Natl Acad Sci U S A 96*, 4780-4785.
[55] Vargas-Rodriguez, O., Sevostyanova, A., Soll, D., and Crnkovic, A. (2018) Upgrading aminoacyl-tRNA synthetases for genetic code expansion, *Curr Opin Chem Biol 46*, 115-122.
[56] Young, D. D., Jockush, S., Turro, N. J., and Schultz, P. G. (2011) Synthetase polyspecificity as a tool to modulate protein function, *Bioorg Med Chem Lett 21*, 7502-7504.
[57] Van Deventer, J. A., Le, D. N., Zhao, J., Kehoe, H. P., and Kelly, R. L. (2016) A platform for constructing, evaluating, and screening bioconjugates on the yeast surface, *Protein Eng Des Sel 29*, 485–494.
[58] Wu, N., Deiters, A., Cropp, T. A., King, D., and Schultz, P. G. (2004) A genetically encoded photocaged amino acid, *J Am Chem Soc 126*, 14306-14307.
[59] Brustad, E., Bushey, M. L., Brock, A., Chittuluru, J., and Schultz, P. G. (2008) A promiscuous aminoacyl-tRNA synthetase that incorporates cysteine, methionine, and alanine homologs into proteins, *Bioorg Med Chem Lett 18*, 6004-6006.
[60] Lieblich, S. A., Fang, K. Y., Cahn, J. K. B., Rawson, J., LeBon, J., Ku, H. T., and Tirrell, D. A. (2017) 4S-Hydroxylation of Insulin at ProB28 Accelerates Hexamer Dissociation and Delays Fibrillation, *J Am Chem Soc 139*, 8384-8387.
[61] Dougherty, D. A., and Van Arnam, E. B. (2014) In vivo incorporation of non-canonical amino acids by using the chemical aminoacylation strategy: a broadly applicable mechanistic tool, *Chembiochem 15*, 1710-1720.
[62] Faber, M. S., and Whitehead, T. A. (2019) Data-driven engineering of protein therapeutics, *Curr Opin Biotechnol 60*, 104-110.
[63] Islam, M., Kehoe, H. P., Lissoos, J. B., Huang, M., Ghadban, C. E., Berumen Sanchez, G., Lane, H. Z., and Van Deventer, J. A. (2021) Chemical Diversification of Simple Synthetic Antibodies, *ACS Chem Biol*.
[64] Van Deventer, J. A., and Wittrup, K. D. (2014) Yeast surface display for antibody isolation: library construction, library screening, and affinity maturation, *Methods Mol Biol 1131*, 151-181.