bioRxiv preprint doi: https://doi.org/10.1101/2021.06.22.449487; this version posted June 22, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Genetic code expansion in the engineered organism Vmax X2: High yield and exceptional fidelity

Sebasthian Santiago[a], Omer Ad[b], Bhavana Shah[c], Zhongqi Zhang[c], Xizi Zhang[a], Abhishek Chatterjee[d], & Alanna Schepartz[aef]*

Author affiliations:

[a]Department of Chemistry, University of California, Berkeley, CA 94720
[b]Department of Chemistry, Yale University, New Haven, CT 06511
[c]Process Development, Attribute Sciences, Amgen Inc., Thousand Oaks, CA 91320
[d]Department of Chemistry, Boston College, Chestnut Hill, MA 02467
[e]Department of Molecular and Cellular Biology, University of California, Berkeley, CA 94720
[f]California Institute for Quantitative Biosciences (QB3), University of California, Berkeley, CA 94720

Abstract:

We report that the recently introduced commercial strain of *V. natriegens* (Vmax X2) supports robust unnatural amino acid mutagenesis, generating exceptional yields of soluble protein containing up to 5 non-canonical α-amino acids (ncAA). The isolated yields of ncAA-containing superfolder green fluorescent protein (sfGFP) expressed in Vmax X2 are up to 25-fold higher than those achieved using commercial expression strains (Top10 and BL21) and more than10-fold higher than those achieved using two different genomically recoded *E. coli* strains that lack endogenous UAG stop codons and release factor 1 and have been optimized for improved fitness and preferred growth temperature (C321.ΔA.opt and C321.ΔA.exp). In addition to higher yields of soluble protein, Vmax X2 cells also generate proteins with significantly lower levels of mis-incorporated natural α-amino acids at the UAG-programmed position, especially in cases where the ncAA is an imperfect substrate for the chosen orthogonal aminoacyl tRNA synthetase (aaRS). This increase in fidelity implies that use of Vmax X2 cells as the expression host can obviate the need for time-consuming directed evolution experiments to improve specific activity of highly desirable but imperfect ncAA substrates.

bioRxiv preprint doi: https://doi.org/10.1101/2021.06.22.449487; this version posted June 22, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Introduction

Non-canonical amino acid mutagenesis, often referred to as genetic code expansion (GCE), is a powerful tool for introducing unique chemical functionality or reactivity into an expressed protein.[1–3] Literally hundreds of diverse non-canonical α-amino acids (ncAA), including those that support protein labeling or conjugation, can be introduced into proteins[3–6] biosynthesized in laboratory *E. coli* strains, pathogenic[7] and soil bacteria,[8] yeast,[9] mammalian cells,[9–11] and whole organisms.[12–14] Unnatural amino acid mutagenesis *in vivo* demands co-expression of one or more aminoacyl-tRNA synthetase (aaRS)/tRNA pairs that are orthogonal in the expression host. Commonly used pairs include variants of pyrrolysyl aminoacyl-tRNA synthetase (PylRS)/pylT from Methanosarcina,[15–19] tyrosyl aminoacyl-tRNA synthetase (TyrRS)/$tRNA^{Tyr}$ from *Methanococcus jannaschii*[20–23] and others;[24–26] newly identified orthogonal pairs include those from *Lumatobacter nonamiensi*, *Sorangium cellulosum*, and *Archaeoglobus fulgidus*.[24]

In addition to its utility in basic research, non-canonical amino acid mutagenesis has significant and growing importance in the pharmaceutical and biotechnological industries. The ability to efficiently introduce reactive bioorthogonal functionality into a therapeutic antibody provides a streamlined route to homogeneous antibody-drug conjugates with high (~95%) conjugation efficiency.[27–29] Incorporation of non-canonical α-amino acids can be leveraged to confer favorable therapeutic properties, such as increased circulation half-life and improved bioactivity.[30–32] Finally, GCE expands the chemical space that is accessible for the development of novel macrocycles and therapeutic peptides.[33–35]

For almost all genetic code expansion applications, yield and purity are paramount.[36] One factor that can limit the yield of a protein carrying one or more ncAAs is competition between the mis-acylated suppressor tRNA and release factor 1 (RF1), both of which recognizes the amber UAG stop codon. Recognition of the mis-acylated suppressor tRNA leads to incorporation of the ncAA, whereas recognition by RF1 triggers early translation termination.[37,38] Another factor that can limit yields is that the orthogonality of an aaRS/tRNA pair is rarely absolute, resulting in suppressor tRNAs that are acylated incorrectly with one or more α-amino acids. To circumvent this issue, researchers have developed *E. coli* strains that lack RF1, including those that have been genomically recoded to eliminate all or some of the 321 endogenous UAG stop codons in *E. coli*.[39–41] While these strains can improve expression yields, genomically recoded organisms

bioRxiv preprint doi: https://doi.org/10.1101/2021.06.22.449487; this version posted June 22, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

that lack RF1 suffer from fitness defects as well as higher levels of misincorporation events when utilizing suboptimal aaRS/tRNA pairs.[42–45] Cell-free translation systems offer the opportunity to omit RF1 as well as tune the individual levels of near-cognate tRNAs to potentially increase incorporation fidelity, but these systems are significantly more costly than cellular bioproduction.[46,47]

Further challenges arise when high-yield expression of the target protein in *E. coli* demands low temperatures. Many orthogonal synthetases are derived from thermophilic organisms and exhibit minimal activity at temperatures below 25°C.[48,49] The requirement for higher expression temperatures can lower the yield of target proteins that are unstable and/or insoluble under these conditions. *Vibrio natriegens*, a bacteria isolated originally from salt marsh mud, is the fastest growing organism on record and expresses many recombinant proteins in exceptional yields and at a variety of temperatures (20°C – 37°C).[50–56] Previous research has revealed considerable compatibility between *E. coli* and *V. natriegens* in terms of commonly used genetic elements (i.e. promoters, ribosome binding sites, etc.) and plasmids.[52,57] This compatibility facilitates the use of numerous extensively optimized *E. coli* GCE plasmid systems with little or no plasmid modification. Vmax X2 cells also possess advantages for the production of protein destined for used in animals, where endotoxin contamination remains a persistent concern, generating endotoxin titers even lower than those seen in cells like ClearColi® cells.[58]

Here we demonstrate that the recently introduced commercial strain of *V. natriegens* (Vmax X2) supports robust unnatural amino acid mutagenesis, generating exceptional yields of soluble protein containing up to 5 ncAAs.[59] Yields are especially high when ncAA are introduced using the *M. jannaschii* tyrosyl aminoacyl-tRNA synthetase (TyrRS)/$tRNA^{Tyr}$ variant pCNFRS.[20,21] The isolated yields of ncAA-containing superfolder green fluorescent protein (sfGFP) expressed in Vmax X2 are up to 25-fold higher than those achieved using commercial expression strains (Top10 and BL21) and more than 10-fold higher than those achieved using two different genomically recoded *E. coli* strains that lack endogenous UAG stop codons and have been optimized for improved fitness and preferred growth temperature (C321.ΔA.opt and C321.ΔA.exp, Addgene strains #87359 and #49018).[39,40] The rapid doubling time of Vmax X2 (~10-14 min)[52] also translates into a highly convenient three-day workflow for protein expression, as opposed to the traditional 4-day workflow for protein expression using traditional *E. coli* strains. In addition to high yields, Vmax X2 cells also generate proteins with significantly lower levels of mis-incorporated natural α-amino acids at the UAG-programmed position,

bioRxiv preprint doi: https://doi.org/10.1101/2021.06.22.449487; this version posted June 22, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

especially in cases where the ncAA is an imperfect substrate for the chosen orthogonal aminoacyl tRNA synthetase (aaRS). Thus, use of Vmax X2 can obviate the need for time-consuming directed evolution experiments to improve the specific activity of highly desirable but imperfect aaRS substrates.

Results

To evaluate Vmax X2 as a host organism for unnatural amino acid mutagenesis, we first asked if it would support the incorporation of a single ncAA into sfGFP using two orthogonal translation systems (OTS) that are used commonly in *E. coli*. The first is the *p*-cyano-L-phenylalanyl aminoacyl-tRNA synthetase (pCNFRS)-$tRNA_{CUA}^{Tyr}$ pair[20,21] derived from *M. jannaschii*, while the second is the pyrrolysyl-tRNA synthetase (PylRS)-$tRNA_{CUA}^{Pyl}$ pair[15] derived from *M. mazei*. Both of these aaRS/tRNA pairs support the incorporation of chemically diverse ncAA into proteins in *E. coli* but only the activity of the PylRS)-$tRNA_{CUA}^{Pyl}$ pair[15] has been tested in *Vibrio natriegens*.[59] Specifically, we asked whether either of these aaRS/tRNA pairs would support the incorporation of *p*-azido-L-phenylalanine (using pCNFRS) or Boc-L-Lysine (using PylRS) at position 2 of sfGFP (Figure 1A). Vmax X2 cells were transformed with either pEVOL-mmPyl or pEVOL-CNF[60] (along with pET-S2TAG sfGFP[61]) and grown in Vmax-optimized media supplemented with either 0.5 mM 4-azido-L-phenylalanine (pAzF) or 10 mM Boc-L-lysine (BocK)[62] at a temperature of 37°C. Vmax X2 cells grew quickly under these conditions, reaching saturation after approximately 4 hours (**Figure 2A** and **Supplementary Figure 1**).

bioRxiv preprint doi: https://doi.org/10.1101/2021.06.22.449487; this version posted June 22, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

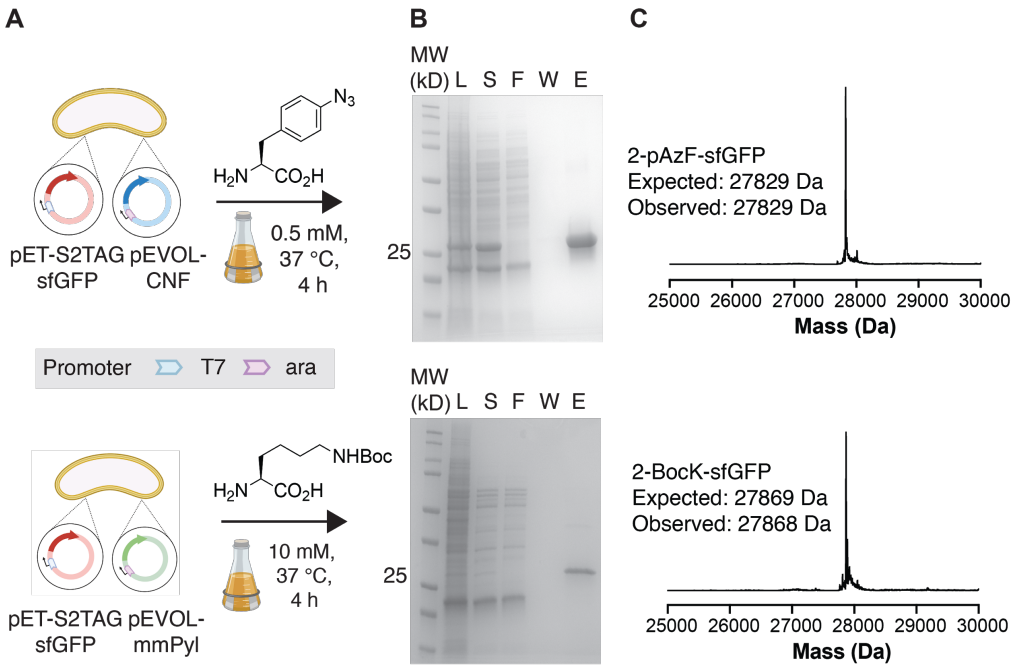


Figure 1. Genetic code expansion in Vmax X2. (A) Vmax X2 cells were transformed with pET-S2TAGsfGFP[61] and either pEVOL-CNF[60] or pEVOL-mmPyl to induce expression of sfGFP bearing a ncAA at the second position of sfGFP. Cells were induced and incubated for 4 hours at 37°C in the presence of 0.5 mM *p*-azido-L-phenylalanine (pAzF) (*p*CNFRS) or 10 mM L-Boc-lysine (BocK) (MmPylRS). (B) SDS-PAGE gels illustrate proteins produced in Vmax X2 cells transformed with pET-S2TAGsfGFP and either pEVOL-CNF or pEVOL-mmPyl. L = lysate; S = supernatant; F = flow-through; W = wash; E = elution. (C) Intact protein mass spectra of 2TAG sfGFP variants purified from Vmax X2 cells co-expressing *p*CNFRS (top) or MmPylRS (bottom) in the presence of pAzF or BocK, respectively.

bioRxiv preprint doi: https://doi.org/10.1101/2021.06.22.449487; this version posted June 22, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

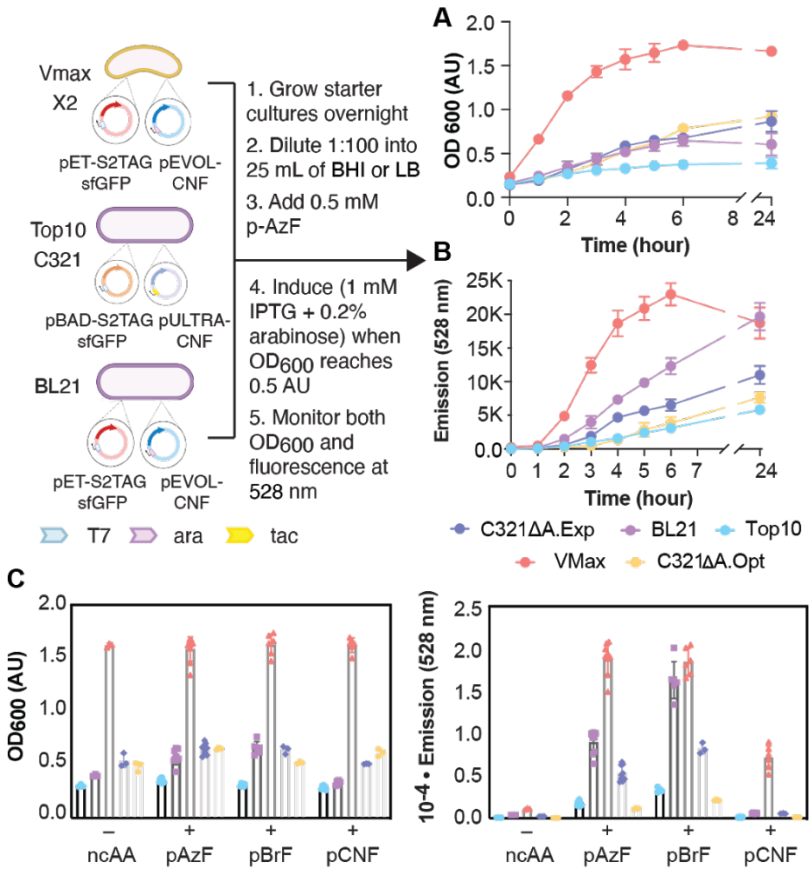


Figure 2. Growth and sfGFP expression in Vmax X2 versus traditional (Top10, BL21) and genomically recoded (C321)[39,40] *E. coli* strains. (A) Plot of the OD_{600} of each cell growth as a function of time. Vmax X2 and BL21 cells were transformed with pET-S2TAGsfGFP and pEVOL-CNF, whereas Top10 and C321 cells were transformed with pBAD-S2TAGsfGFP and pULTRA-CNF (Top10, C321) to induce expression of sfGFP bearing a ncAA at the second position of sfGFP. After induction, cells were grown for 24 hours at 37°C (Vmax X2, BL21, Top10, C321.ΔA.exp) or 34°C (C321.ΔA.opt) in the presence of 0.5 mM pAzF. (B) Plot of the emission of each cell growth at 528 nm as a function of time. (C) Plots comparing the OD_{600} and 528 nm fluorescence of each growth at the 4 h time point in the presence or absence of pAzF, *p*-bromo-L-phenylalanine (pBrF), or *p*-cyano-L-phenylalanine (pCNF).

The 2TAG sfGFP variants expressed in Vmax X2 during a 4-hour incubation were isolated using immobilized metal affinity chromatography (IMAC) and their purity and identity assessed using SDS-PAGE and mass spectrometry (**Figure 1B and C**). In each case, SDS-PAGE evaluation of the isolated protein products revealed a prominent band just above the 25 kDa marker as expected, whose intact mass spectrum was consistent with incorporation of a single copy of either BocK or pAzF. To verify that the non-canonical α-amino acid was introduced into the expected (2^{nd}) position, the proteins produced in Vmax X2 cells were digested with the endoproteinase Glu-C and analyzed further by LC-MS/MS (**Supplementary Figure 2**). Sequence matching of the digested peptides confirmed the incorporation of BocK and pAzF at position 2 of sfGFP. These data indicate that the orthogonal synthetases pCNFRS and PylRS are expressed and active in Vmax X2 cells, acylate their cognate tRNAs with the provided ncAA, and that the charged tRNA is utilized by the Vmax X2 translational machinery. The yield

bioRxiv preprint doi: https://doi.org/10.1101/2021.06.22.449487; this version posted June 22, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

of sfGFP containing an ncAA at position 2 was 8.8 mg (BocK) and 387.2 mg (pAzF) per liter of culture. The yield of sfGFP containing BocK at position 2 is more than 8-fold higher than that observed previously (~1 mg/L), likely due to the use of a stronger promoter for the sfGFP transcript, as well as ncAA-dependent effects.[59]

Next, we set out to evaluate how the yield of ncAA-containing sfGFP produced in Vmax X2 cells compared to those obtained in several *E. coli* strains used for protein expression and genetic code expansion. The strains evaluated included Top10 (a broad utility strain related to DH10B™), BL21 (optimized for protein expression from T7 promoters), as well as two genomically recoded strains, C321.ΔA.exp[39] and C321.ΔA.opt,[40] in which the 321 endogenous UAG stop codons are replaced by the alternative stop codon UAA. Both GROs are derived from the original C321.ΔA**;** the former has been engineered to lower the rate of spontaneous mutagenesis and can be grown at 37 °C; the latter carries additional mutations that improve doubling time. Because C321 strains lack T7 RNA polymerase,[39] sfGFP expression in these strains is under control of the commonly used pBAD promoter. All strains were grown under their own optimized conditions and in the presence of 0.5 mM pAzF; both OD_{600} and fluorescence at 528 nm (λ_{max} for sfGFP) were monitored as a function of time (**Figure 2A**).

As expected, the Vmax X2 cultures grew faster than all others and reached saturation at an OD_{600} of 1.5 approximately 4 hours after induction. All other strains required more than 24 h to reach an OD_{600} of 1.0. As judged by the emission value at 528 nm, sfGFP expression in Vmax X2 reached a maximal value after 6 hours of expression and then decreased slightly. In the case of all other strains, the signal at 528 nm increased linearly over time over the entire course of the experiment (24 h). The greatest difference in OD_{600} and fluorescence occurred at the 4h time point (**Figure 2A**).

bioRxiv preprint doi: https://doi.org/10.1101/2021.06.22.449487; this version posted June 22, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Examination of the OD_{600} and emission at 528 nm after 4 h incubation in the presence of three different pCNFRS substrates (pAzF, pBrF, and pCNF) reveals several trends (**Figure 2C**). First, as expected, the Vmax X2 growth rate exceeded that of any other strain in the absence of a ncAA or in the presence of 0.5 mM pAzF, pBrF, or pCNF. The presence or the identity of the ncAA had little or no effect on the growth rate of any strain examined. The changes in fluorescence at 528 nm show more significant changes and greater dependence on ncAA identity. For example, although the values for 528 nm emission in the presence of pAzF mirrored the OD_{600} values across all strains, the values in the presence of pBrF were higher than expected in BL21. The values for 528 nm emission in the presence of pCNF were low in all strains other than Vmax X2. Despite these differences, two overarching trends emerge: in all cases, the 528 nm emission values suggest that higher levels of sfGFP are produced in Vmax X2 than in any other strain tested, and that the yields in BL21 cells exceed those obtained in either C321.ΔA.exp[39] or C321.ΔA.opt.[40] A plot of 528 nm fluorescence/OD_{600} (**Supplementary Figure 1C**) shows that Vmax X2 cells often express less sfGFP per cell, except in the case where pCNF was incorporated, compared to BL21 and C321.ΔA.exp.

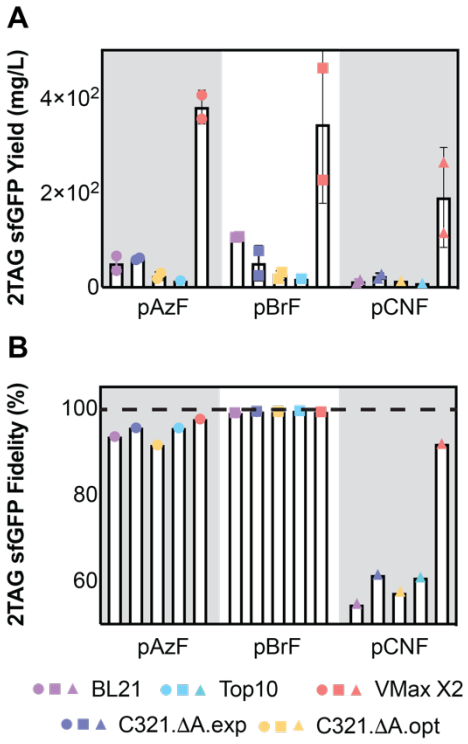


Figure 3. Yield and purity of sfGFP containing a single ncAA when expressed in Vmax X2 versus *E. coli* strains. (A) Isolated yield (mg/L) and (B) fidelity (%) of sfGFP containing the indicated ncAA at position 2 when expressed in the indicated strain. Cells were grown for 4 h, lysed *via* sonication, and sfGFP was isolated using IMAC. Yields were determined using the Pierce 660 nm Protein Assay (Thermo Scientific) and a BSA standard curve. Fidelity was determined from LC-MS/MS data as the fraction of the Glu-C-generated N-terminal peptide MXKGEE containing the desired ncAA (X) at position 2 relative to all other detectable amino acids at that position. The dashed line indices 100% fidelity. Additional LC-MS/MS data is found in **Supplementary Figures 4 and 5**.

Although fluorescence emission at 528 nm is often taken as a measure of sfGFP expression, for most applications it is the isolated, purified protein yield that matters more, and isolated yields can be affected negatively if aggregation occurs at high protein concentration.[63] To evaluate whether the isolated yields would parallel fluorescence at 528 nm, we isolated sfGFP from each

bioRxiv preprint doi: https://doi.org/10.1101/2021.06.22.449487; this version posted June 22, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

strain after a 4 h growth (**Figure 3A** and **Supplementary Table 1**). The unoptimized yield of isolated sfGFP produced in Vmax X2 cells was at least 2-fold and in some cases as much as 25-fold higher than the yield obtained in any *E. coli* strain tested. In the case of pAzF, the yield of sfGFP produced in Vmax was 25-fold higher than the yield obtained in either Top10 or C321.ΔA.opt cells. In fact, although proteins yields can be affected negatively by high expression titers, for Vmax X2 the opposite trend is observed: significantly more soluble protein is obtained per unit 528 nm absorbance in Vmax X2 than in any other strain, across all ncAA examined (**Supplementary Figure 3**).

One strain-dependent complication of GCE is misincorporation of an endogenous α-amino acid in place of the desired ncAA.[42–44] Many factors can influence the misincorporation rate, including inefficient transport of the ncAA into cells, imperfect specificity of the aaRS enzyme for the ncAA over natural α-amino acids, differences in EF-Tu-mediated delivery of the mis-acylated tRNA to the ribosome, and slower rates of accommodation, elongation, or translocation of mis-acylated tRNAs.[43,64,65] Misincorporation at an amber stop codon can also arise from near-cognate suppression, in which natural (and correctly acylated) tRNAs pair imperfectly with a UAG codon and direct the incorporation of one or more natural α-amino acids in place of the ncAA. The frequency of near-cognate suppression events depends on many factors, including the concentration of charged suppressor tRNA relative to near-cognate tRNAs and the presence of RF1, with higher occurrence of near-cognate suppression when the level of charged tRNA is low and RF1 is absent.[42] It has been found that G/U mismatch as well as the 3rd-base wobble mismatch are the most common errors during codon recognition, followed by some other single-base mismatches (A/C, U/U and U/C).[66,67] It follows that the aminoacyl-tRNA of tyrosine (both codons), glutamine (CAG codon), tryptophan, lysine (AAG codon), and glutamic acid (GAG codon) are likely near-cognate suppressors of UAG, in agreement with observations here and elsewhere.[68] Additionally, phenylalanine and tyrosine misincorporations are also expected due to tRNA mischarging for their structural similarities to the ncAAs (pAzF, pBrF, and pCNF).

We used high-resolution LC-MS/MS to evaluate and compare the strain-dependent fidelity of ncAA incorporation at position 2 of sfGFP using pCNFRS (**Figure 3B** and **Supplementary Figure 6**). In each case, sfGFP samples were isolated from each strain after 4 h using IMAC, denatured, reduced and alkylated with iodoacetamide, and finally digested with Glu-C. The peptide fragments so-generated were analyzed on an LC-MS/MS system composed of an Agilent 1290-II HPLC and a Thermo Fisher Q Exactive Biopharma mass spectrometer. The mass spectrometry data were searched against the sfGFP sequence using MassAnalyzer (an

bioRxiv preprint doi: https://doi.org/10.1101/2021.06.22.449487; this version posted June 22, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

in-house developed program, available in Biopharma Finder from Thermo Fisher) for potential amino acid substitutions.[66]

The LC-MS/MS data revealed that the yield of sfGFP containing the desired ncAA at position 2 was both strain and ncAA dependent (**Supplementary Figure 6**). As expected from the reported substrate preferences of pCNFRS, the highest and lowest fidelity were observed with pBrF and pCNF, respectively.[21] The yield of sfGFP containing pBrF at position 2 ranged from a high of 99.7% (in Top10 cells) to a low of 99.4% (BL21 cells). With pAzF, the range was slightly greater, from a high of 97.6% (Vmax X2) to a low of 91.7% (C321.ΔA.opt). The lowest fidelity was observed for the incorporation of pCNF. Here, the fidelity was low (between 54.5% and 61.6%) in all strains except Vmax X2, where the fidelity was 91.8%. It is notable that the most significant impurities, regardless of strain, contained Phe in place of the desired ncAA, although in certain cases a significant amount of Tyr was also detected (**Supplementary Figure 6)**. The improved fidelity of sfGFP containing a ncAA at position 2 when produced in Vmax X2 could be due to many factors, including differences in ncAA permeability, aaRS activity, tolerance of Vmax X2 EF-Tu to mis-acylated tRNAs, even the relative activity of Vmax X2 RF1. Nevertheless, if the goal is a homogeneous and uniquely modified ncAA-containing protein in high yield and purity, Vmax X2 outperformed the other strain, even those that funnel GCE through a single UAG channel. This statement is especially true in the case of ncAA possessing moderate to low specific activity, which could obviate the need for directed evolution to improve specific activity further.

We note that in the previous experiments, the plasmids used to express the aaRS/tRNA pair and sfGFP in Vmax X2 cells were identical to the plasmids used in BL21, but those used in C321 and Top10 cells were different. In the case of Vmax X2 and BL21, the aaRS/tRNA pair was under the control of an arabinose-dependent promoter and sfGFP expression was under control of the T7 promoter. However, as C321 and Top10 cells lack T7 RNA polymerase, in this case, aaRS/tRNA expression was under control of a tac promoter and sfGFP was under control of an arabinose promoter. We wondered about the extent to which these differences in promoter/inducer identity affected the relative yield of sfGFP in each cell line (**Supplementary Figure 7**). To evaluate this question, we made use of a plasmid in which the expression of sfGFP and the aaRS/tRNA pair expression were under control of T5 promoter and arabinose promoters, respectively - both of which are compatible with all cell types examined here.[45] In this case, the plasmid encoded sfGFP carried a UAG codon at position 151. When under the

bioRxiv preprint doi: https://doi.org/10.1101/2021.06.22.449487; this version posted June 22, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

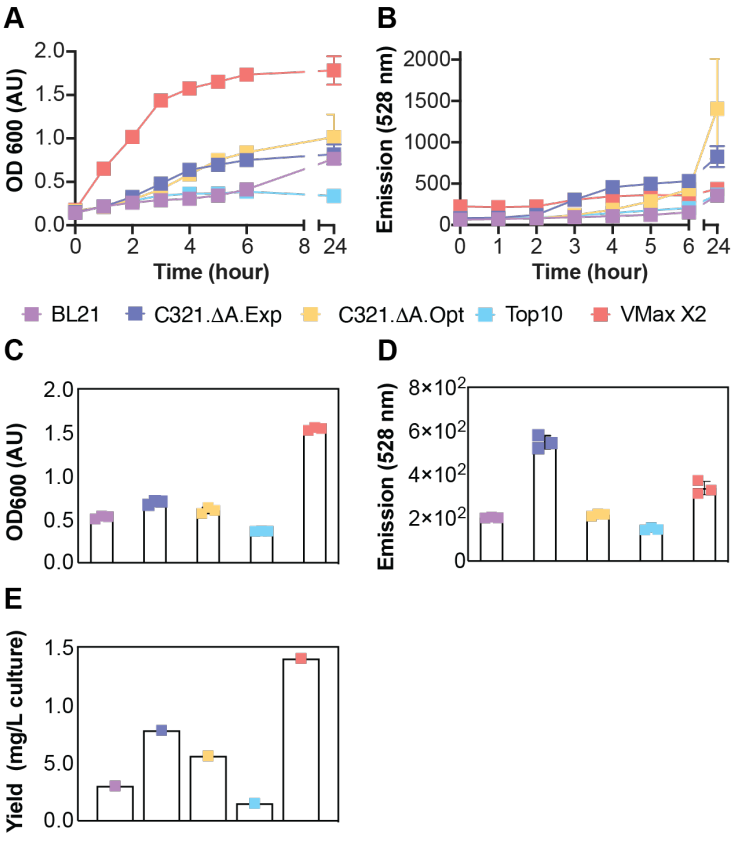
control of a T5 promoter and in the presence of pBrF, although the OD_{600} of Vmax X2 cells was highest, the value of 528 nm emission at the 4 h timepoint was highest for BL21 and C321.ΔA.exp cells, lowest for C321.ΔA.opt and Top10 cells, and intermediate for Vmax X2 cells (**Supplementary Figure 7A**). The isolated yield of 151TAG sfGFP was comparable whether expression was performed in BL21, C321.ΔA.exp, or Vmax X2 – approximately 100 mg/L of culture. These yields were at least 2-fold lower than those obtained in Vmax X2 cells when sfGFP expression was under control of a T7 promoter (**Supplementary Figure 7B and 7C).**

For many ncAA applications relating to fundamental research or the preparation of homogeneous antibody-drug conjugates (ADCs), one ncAA per polypeptide chain is generally sufficient. For others, such as the design of sequence-defined protein materials, multiple copies of one (or more) ncAA may be desired.[69] These applications push the limits of genetic code expansion, as the isolated yields of such materials from standard *E. coli* strains can be low due to the increased frequency of RF1-mediated termination events.[45] Strains that lack RF1, such as B95 and C321 derivatives, have been reported to support greatly improved yields of model proteins containing multiple copies of a single ncAA such as pAzF.[39,41,45] Likewise, cell-free expression systems derived from these strains have been utilized.[55,70,71] These systems are able to produce appreciable amounts of material containing up to 40 non-canonical amino acids,[70] but fidelity is evaluated only rarely.

We wondered whether Vmax X2 would also provide advantages in yield or purity for expression of proteins whose coding sequences contained multiple UAG codons. Thus, we examined the strain-dependent yield and purity of sfGFP containing pBrF at five positions with sfGFP (S2, D36, K101, E132, and D190). All of these positions have been shown previously to accept one of more ncAAs.[71,72] As before, all strains were grown under their own optimized conditions and in the presence of 0.5 mM pBrF; both OD_{600} and fluorescence at 528 nm were monitored as a function of time (**Figure 4A and B**). As observed previously, Vmax X2 cells grew fastest (**Figure 4A**) and there was no direct correlation between the 528 nm emission value at 4 h and the isolated yield. Although the average 528 nm emission of C321.ΔA.exp cells was higher than that of the Vmax X2 growth (**Figure 4D)**, the isolated yield of sfGFP from Vmax X2 was significantly higher (**Figure 4E**). The fidelity of sfGFP containing five copies of pBrF was also higher in Vmax X2 (>98.7%) than in C321.ΔA.exp (**Supplementary Figure 7**). Other detected amino acids include phenylalanine, glutamine (C321 strain only), tyrosine, lysine, leucine/isoleucine, and

bioRxiv preprint doi: https://doi.org/10.1101/2021.06.22.449487; this version posted June 22, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

proline (**Supplementary Figure 8**). None of the other E. coli strains generated sufficient material for detailed LC-MS/MS analysis.



Legend for Figure 4. Yield of sfGFP containing five copies of pBrF ncAA when expressed in Vmax X2 versus *E. coli* strains. (A) Time-dependent growth curves; (B) Time-dependent increase in 528 nm fluorescence; (C) OD_{600}; (D) 528 nm fluorescence; and (E) isolated yields after 4 h. Vmax X2 and BL21 cells were transformed with pET-5XTAG-sfGFP and pEVOL-CNF, whereas Top10 and C321 cells were transformed with pET-5XTAG-sfGFP and pULTRA-CNF (Top10, C321), to induce expression of sfGFP bearing a ncAA at five positions with sfGFP (S2, D36, K101, E132, and D190). After induction, cells were grown for 4 hours at 37°C (Vmax X2, BL21, Top10, C321.ΔA.exp) or 34°C (C321.ΔA.opt) in the presence of 0.5 mM pBrF.

In summary, here we show that Vmax X2 is capable of producing exceptional yields of soluble protein containing up to 5 ncAAs. The yields realized with Vmax X2 are up to 25-fold higher than those achieved using commercial expression strains (Top10 and BL21) and more than 10-fold higher than those achieved using two different genomically recoded *E. coli* strains that lack endogenous UAG stop codons and have been optimized for improved fitness and preferred growth temperature (C321.ΔA.opt and C321.ΔA.exp, Addgene strains #87359 and #49018).[39,40] In addition to high yields, Vmax X2 cells also generate proteins with significantly lower levels of mis-incorporated natural α-amino acids at the UAG-programmed position, especially in cases in which the ncAA is only a moderate substrate for the chosen aaRS. Thus, use of Vmax X2 can obviate the need for time-consuming directed evolution experiments to improve specific activity of highly desirable substrates.

bioRxiv preprint doi: https://doi.org/10.1101/2021.06.22.449487; this version posted June 22, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

References cited

(1) Chin, J. W. Expanding and Reprogramming the Genetic Code. *Nature* **2017**, *550* (7674), 53–60. https://doi.org/10.1038/nature24031 10.1038/nature24031.

(2) Young, D. D.; Schultz, P. G. Playing with the Molecules of Life. *ACS Chem Biol* **2018**, *13* (4), 854–870. https://doi.org/10.1021/acschembio.7b00974 10.1021/acschembio.7b00974. Epub 2018 Mar 2.

(3) Mukai, T.; Lajoie, M. J.; Englert, M.; Soll, D. Rewriting the Genetic Code. *Annu Rev Microbiol* **2017**, *71*, 557–577. https://doi.org/10.1146/annurev-micro-090816-093247 10.1146/annurev-micro-090816-093247. Epub 2017 Jul 11.

(4) Liu, D. R.; Magliery, T. J.; Pastrnak, M.; Schultz, P. G. Engineering a TRNA and Aminoacyl-TRNA Synthetase for the Site-Specific Incorporation of Unnatural Amino Acids into Proteins in Vivo. *Proc. Natl. Acad. Sci.* **1997**, *94* (19), 10092–10097. https://doi.org/10.1073/pnas.94.19.10092.

(5) Wang, L.; Magliery, T. J.; Liu, D. R.; Schultz, P. G. A New Functional Suppressor TRNA/Aminoacyl−tRNA Synthetase Pair for the in Vivo Incorporation of Unnatural Amino Acids into Proteins. *J. Am. Chem. Soc.* **2000**, *122* (20), 5010–5011. https://doi.org/10.1021/ja000595y.

(6) Chin, J. W. Expanding and Reprogramming the Genetic Code of Cells and Animals. *Annu. Rev. Biochem.* **2014**, *83* (1), 379–408. https://doi.org/10.1146/annurev-biochem-060713-035737.

(7) Gan, Q.; Lehman, B. P.; Bobik, T. A.; Fan, C. Expanding the Genetic Code of Salmonella with Non-Canonical Amino Acids. *Sci Rep* **2016**, *6*, 39920. https://doi.org/10.1038/srep39920 10.1038/srep39920.

(8) Lopatniuk, M.; Myronovskyi, M.; Luzhetskyy, A. Streptomyces Albus: A New Cell Factory for Non-Canonical Amino Acids Incorporation into Ribosomally Synthesized Natural Products. *ACS Chem Biol* **2017**, *12* (9), 2362–2370. https://doi.org/10.1021/acschembio.7b00359 10.1021/acschembio.7b00359. Epub 2017 Aug 16.

(9) Hancock, S. M.; Uprety, R.; Deiters, A.; Chin, J. W. Expanding the Genetic Code of Yeast for Incorporation of Diverse Unnatural Amino Acids via a Pyrrolysyl-TRNA Synthetase/TRNA Pair. *J Am Chem Soc* **2010**, *132* (42), 14819–14824. https://doi.org/10.1021/ja104609m 10.1021/ja104609m.

(10) Brown, K. A.; Deiters, A. Genetic Code Expansion of Mammalian Cells with Unnatural Amino Acids. *Curr Protoc Chem Biol* **2015**, *7* (3), 187–199. https://doi.org/10.1002/9780470559277.ch150038 10.1002/9780470559277.ch150038.

(11) Beranek, V.; Reinkemeier, C. D.; Zhang, M. S.; Liang, A. D.; Kym, G.; Chin, J. W. Genetically Encoded Protein Phosphorylation in Mammalian Cells. *Cell Chem Biol* **2018**, *25* (9), 1067-1074 e5. https://doi.org/10.1016/j.chembiol.2018.05.013 10.1016/j.chembiol.2018.05.013. Epub 2018 Jun 21.

(12) Greiss, S.; Chin, J. W. Expanding the Genetic Code of an Animal. *J Am Chem Soc* **2011**, *133* (36), 14196–14199. https://doi.org/10.1021/ja2054034 10.1021/ja2054034. Epub 2011 Aug 22.

(13) Ernst, R. J.; Krogager, T. P.; Maywood, E. S.; Zanchi, R.; Beránek, V.; Elliott, T. S.; Barry, N. P.; Hastings, M. H.; Chin, J. W. Genetic Code Expansion in the Mouse Brain. *Nat Chem Biol* **2016**, *12* (10), 776–778. https://doi.org/10.1038/nchembio.2160.

(14) Brown, W.; Liu, J.; Deiters, A. Genetic Code Expansion in Animals. *ACS Chem Biol* **2018**, *13* (9), 2375–2386. https://doi.org/10.1021/acschembio.8b00520.

bioRxiv preprint doi: https://doi.org/10.1101/2021.06.22.449487; this version posted June 22, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

(15) Wan, W.; Tharp, J. M.; Liu, W. R. Pyrrolysyl-TRNA Synthetase: An Ordinary Enzyme but an Outstanding Genetic Code Expansion Tool. *Biochim Biophys Acta* **2014**, *1844* (6), 1059–1070. https://doi.org/10.1016/j.bbapap.2014.03.002.

(16) Srinivasan, G.; James, C. M.; Krzycki, J. A. Pyrrolysine Encoded by UAG in Archaea: Charging of a UAG-Decoding Specialized TRNA. *Science* **2002**, *296* (5572), 1459–1462. https://doi.org/10.1126/science.1069588.

(17) Willis, J. C. W.; Chin, J. W. Mutually Orthogonal Pyrrolysyl-TRNA Synthetase/TRNA Pairs. *Nat Chem* **2018**, *10* (8), 831–837. https://doi.org/10.1038/s41557-018-0052-5.

(18) Suzuki, T.; Miller, C.; Guo, L.-T.; Ho, J. M. L.; Bryson, D. I.; Wang, Y.-S.; Liu, D. R.; Söll, D. Crystal Structures Reveal an Elusive Functional Domain of Pyrrolysyl-TRNA Synthetase. *Nat. Chem. Biol.* **2017**, *13* (12), 1261–1266. https://doi.org/10.1038/nchembio.2497.

(19) Neumann, H.; Peak-Chew, S. Y.; Chin, J. W. Genetically Encoding Nε-Acetyllysine in Recombinant Proteins. *Nat. Chem. Biol.* **2008**, *4* (4), 232–234. https://doi.org/10.1038/nchembio.73.

(20) Wang, L.; Brock, A.; Herberich, B.; Schultz, P. G. Expanding the Genetic Code of Escherichia Coli. *Science* **2001**, *292* (5516), 498–500. https://doi.org/10.1126/science.1060077.

(21) Young, D. D.; Young, T. S.; Jahnz, M.; Ahmad, I.; Spraggon, G.; Schultz, P. G. An Evolved Aminoacyl-TRNA Synthetase with Atypical Polysubstrate Specificity. *Biochemistry* **2011**, *50* (11), 1894–1900. https://doi.org/10.1021/bi101929e.

(22) Krahn, N.; Tharp, J. M.; Crnković, A.; Söll, D. Engineering Aminoacyl-TRNA Synthetases for Use in Synthetic Biology. In *The Enzymes*; Elsevier, 2020; Vol. 48, pp 351–395. https://doi.org/10.1016/bs.enz.2020.06.004.

(23) Stokes, A. L.; Miyake-Stoner, S. J.; Peeler, J. C.; Nguyen, D. P.; Hammer, R. P.; Mehl, R. A. Enhancing the Utility of Unnatural Amino Acid Synthetases by Manipulating Broad Substrate Specificity. *Mol. Biosyst.* **2009**, *5* (9), 1032. https://doi.org/10.1039/b904032c.

(24) Cervettini, D.; Tang, S.; Fried, S. D.; Willis, J. C. W.; Funke, L. F. H.; Colwell, L. J.; Chin, J. W. Rapid Discovery and Evolution of Orthogonal Aminoacyl-TRNA Synthetase–TRNA Pairs. *Nat. Biotechnol.* **2020**, *38* (8), 989–999. https://doi.org/10.1038/s41587-020-0479-2.

(25) Chatterjee, A.; Xiao, H.; Yang, P.-Y.; Soundararajan, G.; Schultz, P. G. A Tryptophanyl-TRNA Synthetase/TRNA Pair for Unnatural Amino Acid Mutagenesis in *E. Coli*. *Angew. Chem. Int. Ed.* **2013**, *52* (19), 5106–5109. https://doi.org/10.1002/anie.201301094.

(26) Italia, J. S.; Addy, P. S.; Wrobel, C. J. J.; Crawford, L. A.; Lajoie, M. J.; Zheng, Y.; Chatterjee, A. An Orthogonalized Platform for Genetic Code Expansion in Both Bacteria and Eukaryotes. *Nat. Chem. Biol.* **2017**, *13* (4), 446–450. https://doi.org/10.1038/nchembio.2312.

(27) Axup, J. Y.; Bajjuri, K. M.; Ritland, M.; Hutchins, B. M.; Kim, C. H.; Kazane, S. A.; Halder, R.; Forsyth, J. S.; Santidrian, A. F.; Stafin, K.; Lu, Y.; Tran, H.; Seller, A. J.; Biroc, S. L.; Szydlik, A.; Pinkstaff, J. K.; Tian, F.; Sinha, S. C.; Felding-Habermann, B.; Smider, V. V.; Schultz, P. G. Synthesis of Site-Specific Antibody-Drug Conjugates Using Unnatural Amino Acids. *Proc. Natl. Acad. Sci.* **2012**, *109* (40), 16101–16106. https://doi.org/10.1073/pnas.1211023109.

(28) Hallam, T. J.; Wold, E.; Wahl, A.; Smider, V. V. Antibody Conjugates with Unnatural Amino Acids. *Mol. Pharm.* **2015**, *12* (6), 1848–1862. https://doi.org/10.1021/acs.molpharmaceut.5b00082.

(29) Beck, A.; Goetsch, L.; Dumontet, C.; Corvaïa, N. Strategies and Challenges for the next Generation of Antibody–Drug Conjugates. *Nat. Rev. Drug Discov.* **2017**, *16* (5), 315–337. https://doi.org/10.1038/nrd.2016.268.

bioRxiv preprint doi: https://doi.org/10.1101/2021.06.22.449487; this version posted June 22, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

(30) Cho, H.; Daniel, T.; Buechler, Y. J.; Litzinger, D. C.; Maio, Z.; Putnam, A.-M. H.; Kraynov, V. S.; Sim, B.-C.; Bussell, S.; Javahishvili, T.; Kaphle, S.; Viramontes, G.; Ong, M.; Chu, S.; Gc, B.; Lieu, R.; Knudsen, N.; Castiglioni, P.; Norman, T. C.; Axelrod, D. W.; Hoffman, A. R.; Schultz, P. G.; DiMarchi, R. D.; Kimmel, B. E. Optimized Clinical Performance of Growth Hormone with an Expanded Genetic Code. *Proc. Natl. Acad. Sci.* **2011**, *108* (22), 9060–9065. https://doi.org/10.1073/pnas.1100387108.

(31) Lieblich, S. A.; Fang, K. Y.; Cahn, J. K. B.; Rawson, J.; LeBon, J.; Ku, H. T.; Tirrell, D. A. 4 *S* -Hydroxylation of Insulin at ProB28 Accelerates Hexamer Dissociation and Delays Fibrillation. *J. Am. Chem. Soc.* **2017**, *139* (25), 8384–8387. https://doi.org/10.1021/jacs.7b00794.

(32) Rezhdo, A.; Islam, M.; Huang, M.; Van Deventer, J. A. Future Prospects for Noncanonical Amino Acids in Biological Therapeutics. *Curr. Opin. Biotechnol.* **2019**, *60*, 168–178. https://doi.org/10.1016/j.copbio.2019.02.020.

(33) Katoh, T.; Sengoku, T.; Hirata, K.; Ogata, K.; Suga, H. Ribosomal Synthesis and de Novo Discovery of Bioactive Foldamer Peptides Containing Cyclic β-Amino Acids. *Nat. Chem.* **2020**, *12* (11), 1081–1088. https://doi.org/10.1038/s41557-020-0525-1.

(34) Zambaldo, C.; Luo, X.; Mehta, A. P.; Schultz, P. G. Recombinant Macrocyclic Lanthipeptides Incorporating Non-Canonical Amino Acids. *J. Am. Chem. Soc.* **2017**, *139* (34), 11646–11649. https://doi.org/10.1021/jacs.7b04159.

(35) Frost, J. R.; Jacob, N. T.; Papa, L. J.; Owens, A. E.; Fasan, R. Ribosomal Synthesis of Macrocyclic Peptides *in Vitro* and *in Vivo* Mediated by Genetically Encoded Aminothiol Unnatural Amino Acids. *ACS Chem. Biol.* **2015**, *10* (8), 1805–1816. https://doi.org/10.1021/acschembio.5b00119.

(36) Huang, Y.; Liu, T. Therapeutic Applications of Genetic Code Expansion. *Synth. Syst. Biotechnol.* **2018**, *3* (3), 150–158. https://doi.org/10.1016/j.synbio.2018.09.003.

(37) Zhou, J.; Korostelev, A.; Lancaster, L.; Noller, H. F. Crystal Structures of 70S Ribosomes Bound to Release Factors RF1, RF2 and RF3. *Curr. Opin. Struct. Biol.* **2012**, *22* (6), 733–742. https://doi.org/10.1016/j.sbi.2012.08.004.

(38) Capecchi, M. R. Polypeptide Chain Termination in Vitro: Isolation of a Release Factor. *Proc. Natl. Acad. Sci.* **1967**, *58* (3), 1144–1151. https://doi.org/10.1073/pnas.58.3.1144.

(39) Lajoie, M. J.; Rovner, A. J.; Goodman, D. B.; Aerni, H. R.; Haimovich, A. D.; Kuznetsov, G.; Mercer, J. A.; Wang, H. H.; Carr, P. A.; Mosberg, J. A.; Rohland, N.; Schultz, P. G.; Jacobson, J. M.; Rinehart, J.; Church, G. M.; Isaacs, F. J. Genomically Recoded Organisms Expand Biological Functions. *Science* **2013**, *342* (6156), 357–360. https://doi.org/10.1126/science.1241459.

(40) Kuznetsov, G.; Goodman, D. B.; Filsinger, G. T.; Landon, M.; Rohland, N.; Aach, J.; Lajoie, M. J.; Church, G. M. Optimizing Complex Phenotypes through Model-Guided Multiplex Genome Engineering. *Genome Biol.* **2017**, *18* (1), 100. https://doi.org/10.1186/s13059-017-1217-z.

(41) Mukai, T.; Hoshi, H.; Ohtake, K.; Takahashi, M.; Yamaguchi, A.; Hayashi, A.; Yokoyama, S.; Sakamoto, K. Highly Reproductive Escherichia Coli Cells with No Specific Assignment to the UAG Codon. *Sci. Rep.* **2015**, *5* (1), 9699. https://doi.org/10.1038/srep09699.

(42) Beyer, J. N.; Hosseinzadeh, P.; Gottfried-Lee, I.; Van Fossen, E. M.; Zhu, P.; Bednar, R. M.; Karplus, P. A.; Mehl, R. A.; Cooley, R. B. Overcoming Near-Cognate Suppression in a Release Factor 1-Deficient Host with an Improved Nitro-Tyrosine TRNA Synthetase. *J. Mol. Biol.* **2020**, *432* (16), 4690–4704. https://doi.org/10.1016/j.jmb.2020.06.014.

(43) Aerni, H. R.; Shifman, M. A.; Rogulina, S.; O'Donoghue, P.; Rinehart, J. Revealing the Amino Acid Composition of Proteins within an Expanded Genetic Code. *Nucleic Acids Res* **2015**, *43* (2), e8. https://doi.org/10.1093/nar/gku1087.

(44) George, S.; Aguirre, J. D.; Spratt, D. E.; Bi, Y.; Jeffery, M.; Shaw, G. S.; O'Donoghue, P. Generation of Phospho-Ubiquitin Variants by Orthogonal Translation Reveals Codon

bioRxiv preprint doi: https://doi.org/10.1101/2021.06.22.449487; this version posted June 22, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Skipping. *FEBS Lett.* **2016**, *590* (10), 1530–1542. https://doi.org/10.1002/1873-3468.12182.

(45) Zheng, Y.; Lajoie, M. J.; Italia, J. S.; Chin, M. A.; Church, G. M.; Chatterjee, A. Performance of Optimized Noncanonical Amino Acid Mutagenesis Systems in the Absence of Release Factor 1. *Mol. Biosyst.* **2016**, *12* (6), 1746–1749. https://doi.org/10.1039/C6MB00070C.

(46) Gan, Q.; Fan, C. Increasing the Fidelity of Noncanonical Amino Acid Incorporation in Cell-Free Protein Synthesis. *Biochim Biophys Acta Gen Subj* **2017**, *1861* (11 Pt B), 3047–3052. https://doi.org/10.1016/j.bbagen.2016.12.002.

(47) Claassens, N. J.; Burgener, S.; Vögeli, B.; Erb, T. J.; Bar-Even, A. A Critical Comparison of Cellular and Cell-Free Bioproduction Systems. *Curr. Opin. Biotechnol.* **2019**, *60*, 221–229. https://doi.org/10.1016/j.copbio.2019.05.003.

(48) Vieille, C.; Zeikus, G. J. Hyperthermophilic Enzymes: Sources, Uses, and Molecular Mechanisms for Thermostability. *Microbiol Mol Biol Rev* **2001**, *65* (1), 1–43. https://doi.org/10.1128/MMBR.65.1.1-43.2001.

(49) Hu, L.; Qin, X.; Huang, Y.; Cao, W.; Wang, C.; Wang, Y.; Ling, X.; Chen, H.; Wu, D.; Lin, Y.; Liu, T. Thermophilic Pyrrolysyl-TRNA Synthetase Mutants for Enhanced Mammalian Genetic Code Expansion. *ACS Synth. Biol.* **2020**, *9* (10), 2723–2736. https://doi.org/10.1021/acssynbio.0c00257.

(50) *Vmax™ X2 Chemically Competent Cells*; SGI-DNA: La Jolla, CA, 2020.

(51) Lee, H. H.; Ostrov, N.; Wong, B. G.; Gold, M. A.; Khalil, A. S.; Church, G. M. Functional Genomics of the Rapidly Replicating Bacterium Vibrio Natriegens by CRISPRi. *Nat Microbiol* **2019**, *4* (7), 1105–1113. https://doi.org/10.1038/s41564-019-0423-8.

(52) Weinstock, M. T.; Hesek, E. D.; Wilson, C. M.; Gibson, D. G. Vibrio Natriegens as a Fast-Growing Host for Molecular Biology. *Nat Methods* **2016**, *13* (10), 849–851. https://doi.org/10.1038/nmeth.3970.

(53) Becker, W.; Wimberger, F.; Zangger, K. *Vibrio Natriegens* : An Alternative Expression System for the High-Yield Production of Isotopically Labeled Proteins. *Biochemistry* **2019**, *58* (25), 2799–2803. https://doi.org/10.1021/acs.biochem.9b00403.

(54) Schleicher, L.; Muras, V.; Claussen, B.; Pfannstiel, J.; Blombach, B.; Dibrov, P.; Fritz, G.; Steuber, J. Vibrio Natriegens as Host for Expression of Multisubunit Membrane Protein Complexes. *Front. Microbiol.* **2018**, *9*, 2537. https://doi.org/10.3389/fmicb.2018.02537.

(55) Des Soye, B. J.; Davidson, S. R.; Weinstock, M. T.; Gibson, D. G.; Jewett, M. C. Establishing a High-Yielding Cell-Free Protein Synthesis Platform Derived from *Vibrio Natriegens*. *ACS Synth. Biol.* **2018**, *7* (9), 2245–2255. https://doi.org/10.1021/acssynbio.8b00252.

(56) Wiegand, D. J.; Lee, H. H.; Ostrov, N.; Church, G. M. Establishing a Cell-Free *Vibrio Natriegens* Expression System. *ACS Synth. Biol.* **2018**, *7* (10), 2475–2479. https://doi.org/10.1021/acssynbio.8b00222.

(57) Tschirhart, T.; Shukla, V.; Kelly, E. E.; Schultzhaus, Z.; NewRingeisen, E.; Erickson, J. S.; Wang, Z.; Garcia, W.; Curl, E.; Egbert, R. G.; Yeung, E.; Vora, G. J. Synthetic Biology Tools for the Fast-Growing Marine Bacterium. *ACS Synth Biol* **2019**, *8* (9), 2069–2079. https://doi.org/10.1021/acssynbio.9b00176.

(58) Mamat, U.; Woodard, R. W.; Wilke, K.; Souvignier, C.; Mead, D.; Steinmetz, E.; Terry, K.; Kovacich, C.; Zegers, A.; Knox, C. Endotoxin-Free Protein Production—ClearColi™ Technology. *Nat. Methods* **2013**, *10* (9), 916–916. https://doi.org/10.1038/nmeth.f.367.

(59) Ozer, E.; Alfonta, L. Genetic Code Expansion of Vibrio Natriegens. *Front. Bioeng. Biotechnol.* **2021**, *9*, 95. https://doi.org/10.3389/fbioe.2021.594429.

(60) Young, T. S.; Ahmad, I.; Yin, J. A.; Schultz, P. G. An Enhanced System for Unnatural Amino Acid Mutagenesis in E. Coli. *J Mol Biol* **2010**, *395* (2), 361–374. https://doi.org/10.1016/j.jmb.2009.10.030.

bioRxiv preprint doi: https://doi.org/10.1101/2021.06.22.449487; this version posted June 22, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

(61) Kwok, H. S.; Vargas-Rodriguez, O.; Melnikov, S. V.; Söll, D. Engineered Aminoacyl-TRNA Synthetases with Improved Selectivity toward Noncanonical Amino Acids. *ACS Chem. Biol.* **2019**, *14* (4), 603–612. https://doi.org/10.1021/acschembio.9b00088.

(62) Dunkelmann, D. L.; Willis, J. C. W.; Beattie, A. T.; Chin, J. W. Engineered Triply Orthogonal Pyrrolysyl–TRNA Synthetase/TRNA Pairs Enable the Genetic Encoding of Three Distinct Non-Canonical Amino Acids. *Nat. Chem.* **2020**, *12* (6), 535–544. https://doi.org/10.1038/s41557-020-0472-x.

(63) Lebendiker, M.; Danieli, T. Production of Prone-to-Aggregate Proteins. *FEBS Lett.* **2014**, *588* (2), 236–246. https://doi.org/10.1016/j.febslet.2013.10.044.

(64) Haruna, K.; Alkazemi, M. H.; Liu, Y.; Söll, D.; Englert, M. Engineering the Elongation Factor Tu for Efficient Selenoprotein Synthesis. *Nucleic Acids Res.* **2014**, *42* (15), 9976–9983. https://doi.org/10.1093/nar/gku691.

(65) Antonczak, A. K.; Morris, J.; Tippmann, E. M. Advances in the Mechanism and Understanding of Site-Selective Noncanonical Amino Acid Incorporation. *Curr. Opin. Struct. Biol.* **2011**, *21* (4), 481–487. https://doi.org/10.1016/j.sbi.2011.04.004.

(66) Zhang, Z.; Shah, B.; Bondarenko, P. V. G/U and Certain Wobble Position Mismatches as Possible Main Causes of Amino Acid Misincorporations. *Biochemistry* **2013**, *52* (45), 8165–8176. https://doi.org/10.1021/bi401002c.

(67) Wong, H. E.; Huang, C.-J.; Zhang, Z. Amino Acid Misincorporation Propensities Revealed through Systematic Amino Acid Starvation. *Biochemistry* **2018**, *57* (49), 6767–6779. https://doi.org/10.1021/acs.biochem.8b00976.

(68) Roy, B.; Leszyk, J. D.; Mangus, D. A.; Jacobson, A. Nonsense Suppression by Near-Cognate TRNAs Employs Alternative Base Pairing at Codon Positions 1 and 3. *Proc. Natl. Acad. Sci.* **2015**, *112* (10), 3038–3043. https://doi.org/10.1073/pnas.1424127112.

(69) Israeli, B.; Vaserman, L.; Amiram, M. Multi-Site Incorporation of Nonstandard Amino Acids into Protein-Based Biomaterials. *Isr. J. Chem.* **2020**, *60* (12), 1118–1128. https://doi.org/10.1002/ijch.201900043.

(70) Des Soye, B. J.; Gerbasi, V. R.; Thomas, P. M.; Kelleher, N. L.; Jewett, M. C. A Highly Productive, One-Pot Cell-Free Protein Synthesis Platform Based on Genomically Recoded Escherichia Coli. *Cell Chem. Biol.* **2019**, *26* (12), 1743-1754.e9. https://doi.org/10.1016/j.chembiol.2019.10.008.

(71) Hong, S. H.; Ntai, I.; Haimovich, A. D.; Kelleher, N. L.; Isaacs, F. J.; Jewett, M. C. Cell-Free Protein Synthesis from a Release Factor 1 Deficient *Escherichia Coli* Activates Efficient and Multiple Site-Specific Nonstandard Amino Acid Incorporation. *ACS Synth. Biol.* **2014**, *3* (6), 398–409. https://doi.org/10.1021/sb400140t.

(72) Johnson, D. B. F.; Xu, J.; Shen, Z.; Takimoto, J. K.; Schultz, M. D.; Schmitz, R. J.; Xiang, Z.; Ecker, J. R.; Briggs, S. P.; Wang, L. RF1 Knockout Allows Ribosomal Incorporation of Unnatural Amino Acids at Multiple Sites. *Nat. Chem. Biol.* **2011**, *7* (11), 779–786. https://doi.org/10.1038/nchembio.657.

bioRxiv preprint doi: https://doi.org/10.1101/2021.06.22.449487; this version posted June 22, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Genetic code expansion in the engineered organism Vmax X2: High yield and exceptional fidelity

Sebasthian Santiago[a], Omer Ad[b], Bhavana Shah[c], Zhongqi Zhang[c], Xizi Zhang[a], Abhishek Chatterjee[d], & Alanna Schepartz[aef]*

Author affiliations:

[a]Department of Chemistry, University of California, Berkeley, CA 94720

[b]Department of Chemistry, Yale University, New Haven, CT 06511

[c]Process Development, Attribute Sciences, Amgen Inc., Thousand Oaks, CA 91320

[d]Department of Chemistry, Boston College, Chestnut Hill, MA 02467

[e]Department of Molecular and Cellular Biology, University of California, Berkeley, CA 94720

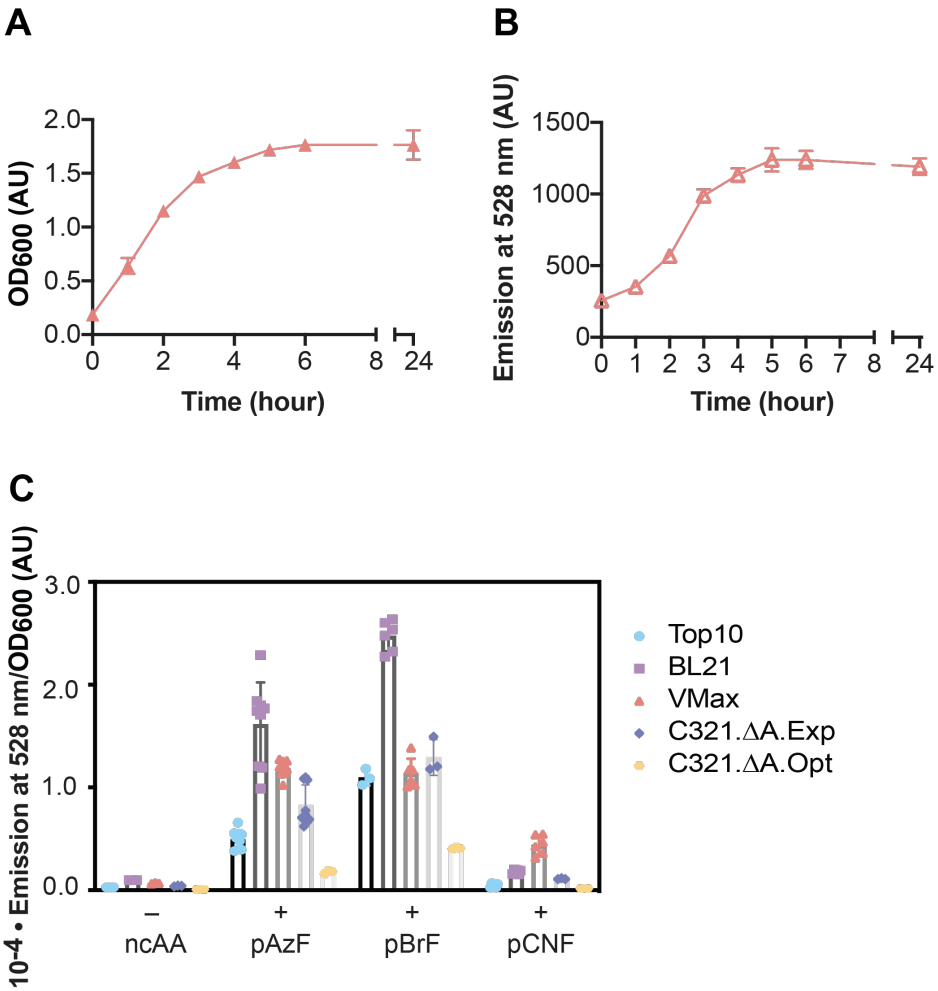
[f]California Institute for Quantitative Biosciences (QB3), University of California, Berkeley, CA 94720

Contents:

I. **Supplementary Figures 1-8**
II. **Supplementary Tables**
III. **Supplementary Methods**
IV. **Supplemental References**

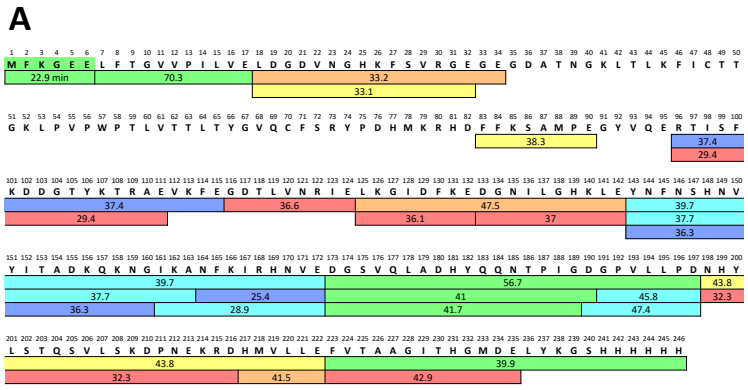
bioRxiv preprint doi: https://doi.org/10.1101/2021.06.22.449487; this version posted June 22, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

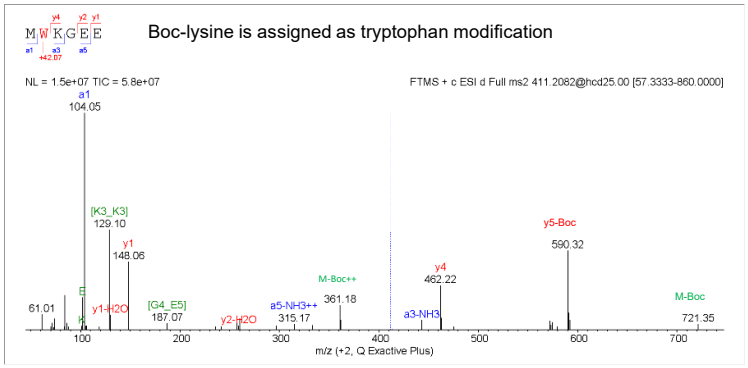
I. Supplementary Figures

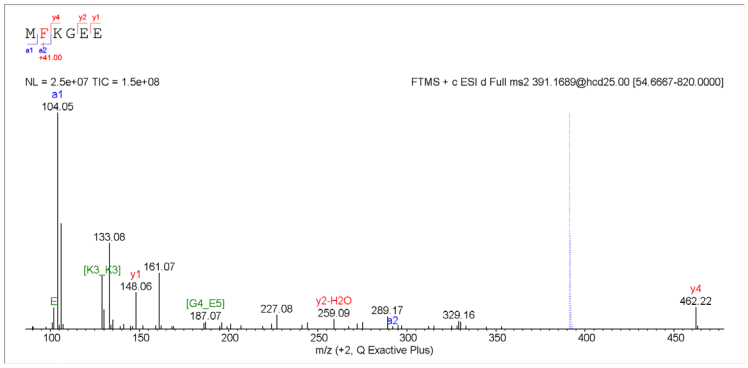


Supplementary Figure 1. **Growth and sfGFP expression of Vmax X2 cells transformed with pET-S2TAGsfGFP and pEVOL-mmPyl.** Vmax X2 cells were transformed with pET-S2TAGsfGFP[1] and pEVOL-mmPyl[2] to induce expression of sfGFP bearing a ncAA at the second position of sfGFP. After induction, cells were incubated at 37°C in Brain-Heart Infusion broth supplemented with V2 salts for 24 hours in the presence of 10 mM BocK. Time-dependent changes in (A) OD_{600} and (B) fluorescence emission at 528 nm. (C) Plots comparing the OD-normalized 528 nm fluorescence of each strain at the 4 h time point in the presence or absence of pAzF, *p*-bromo-L-phenylalanine (pBrF), or *p*-cyano-L-phenylalanine (pCNF).

bioRxiv preprint doi: https://doi.org/10.1101/2021.06.22.449487; this version posted June 22, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

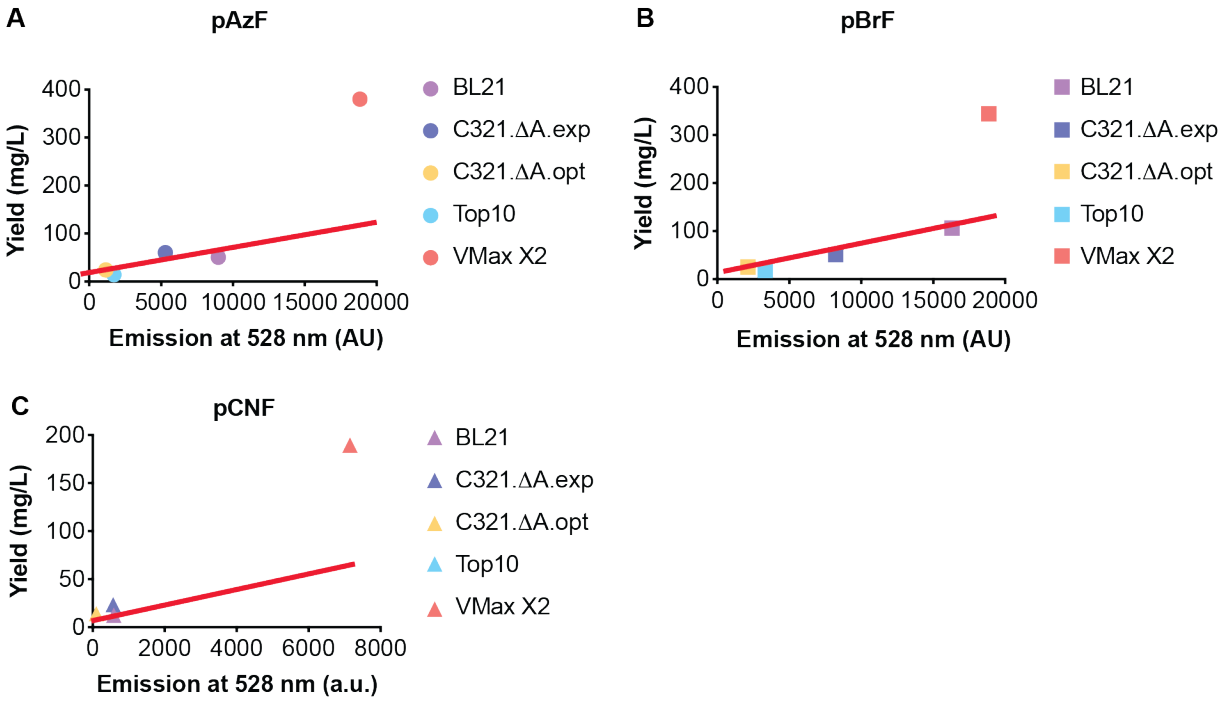






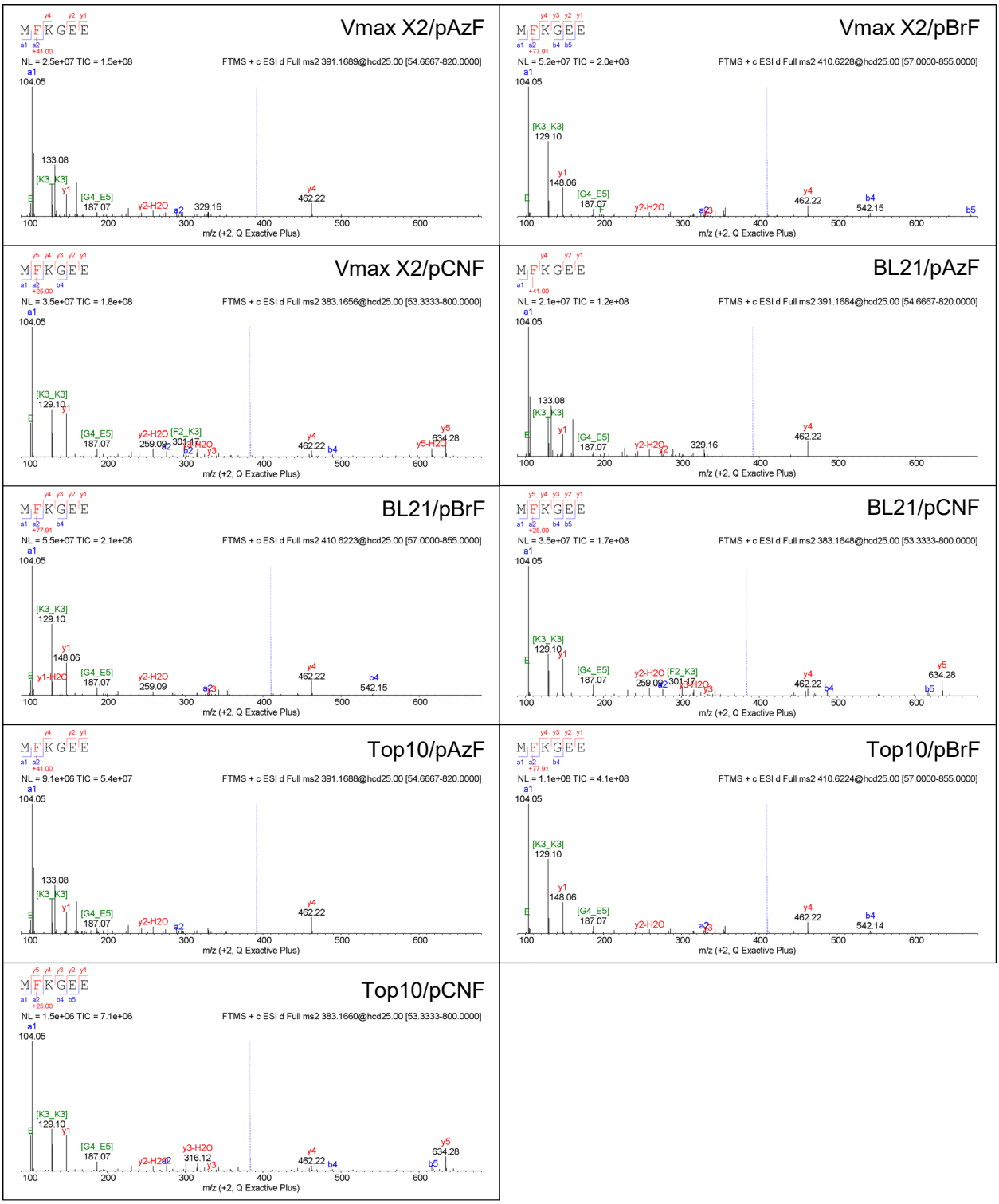
Supplementary Figure 2. LC-MS/MS analysis of sfGFP expressed in Vmax X2. Vmax X2 cells were transformed with pET-S2TAGsfGFP[1] and either pEVOL-CNF[3] or pEVOL-mmPyl[2] to induce expression of sfGFP bearing a ncAA at position two. Cells were induced and incubated for 4 hours at 37°C in the presence of 0.5 mM pAzF (pCNFRS) or 10 mM BocK (MmPylRS), purified using IMAC, and submitted for LC-MS/MS analysis**.** (A) Sequence of wild type sfGFP illustrating the peptide fragments obtained after digestion with Glu-C and their retention times. Colors from red to blue represent decreasing signal intensity. (B) MS/MS identification of the major N-terminal peptide derived from sfGFP isolated from Vmax X2 cells transformed with pET-S2TAGsfGFP and pEVOL-mmPyl and incubated for 4 h in the presence of 10 mM BocK. (C) MS/MS identification of the major N-terminal peptide derived from sfGFP isolated from Vmax X2 cells transformed with pET-S2TAGsfGFP and pEVOL-CNF and incubated for 4 h in the presence of 0.5 mM 4-azido-L-phenylalanine (pAzF). The fidelity of BocK incorporation at position 2 in Vmax X2 cells was ~99%. Other amino acids that could be detected at position 2 include W (0.46%), L/I (0.31%), and Y (0.06%).

bioRxiv preprint doi: https://doi.org/10.1101/2021.06.22.449487; this version posted June 22, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.



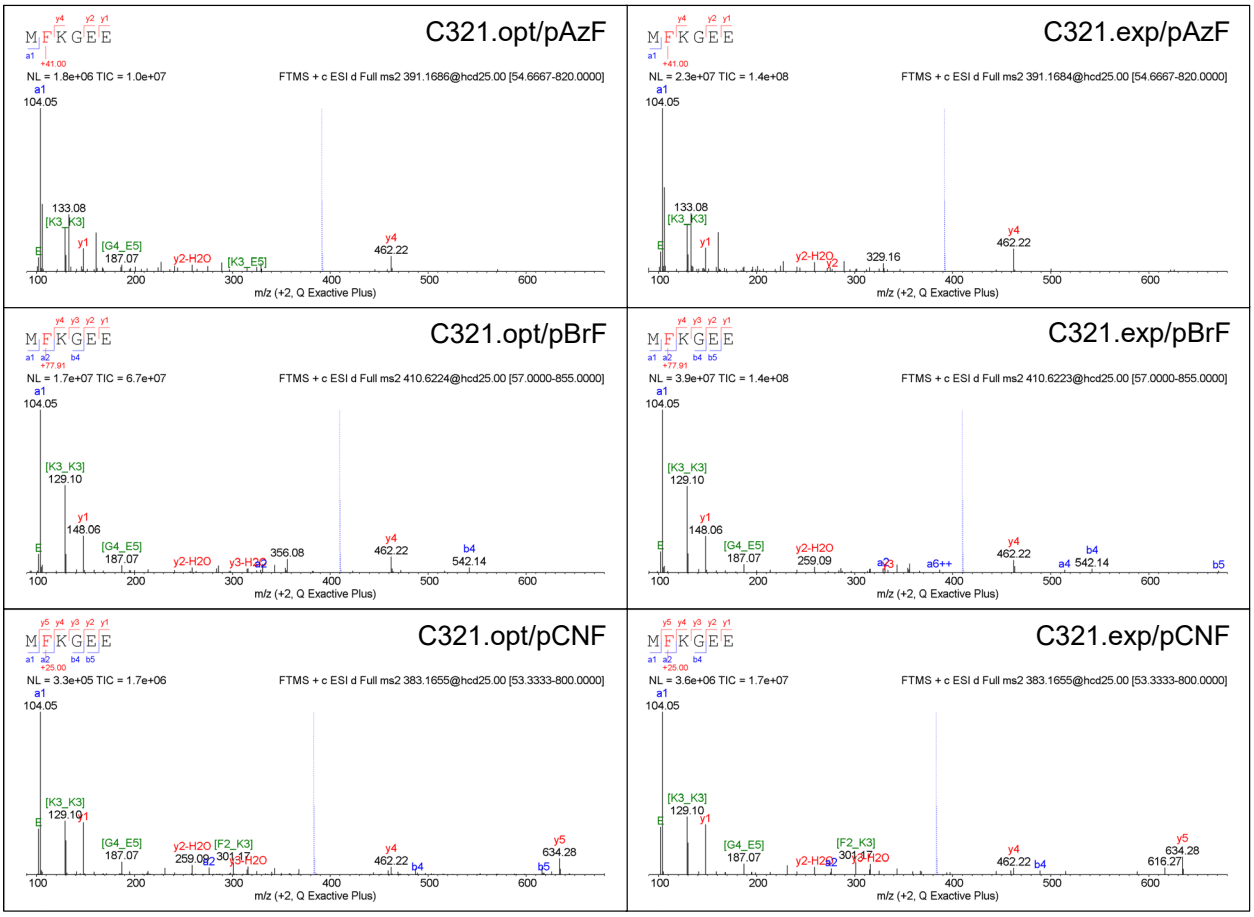
Supplementary Figure 3. Isolated yield of sfGFP containing a ncAA at position 2 is higher than expected based on unit 528 nm emission. Plot of the isolated yield of sfGFP containing (A) pAzF; (B) pBrF; or (C) pCNF at position 2 versus the absorbance of the cell growth 4 h after induction. The red line shows the yield expected if yield correlated directly with unit 528 nm emission at 4 h.

bioRxiv preprint doi: https://doi.org/10.1101/2021.06.22.449487; this version posted June 22, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.



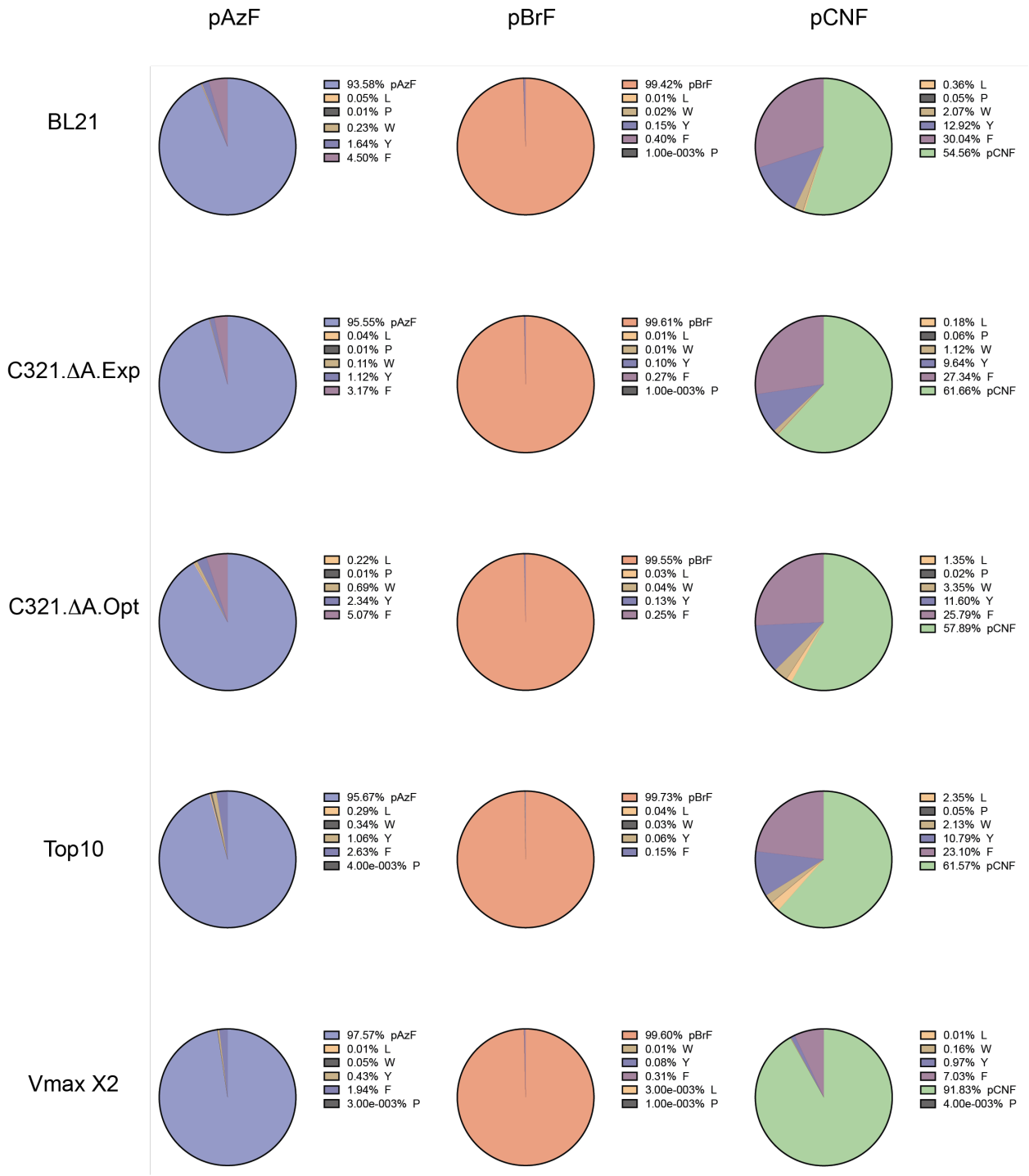
Supplementary Figure 4. MS/MS identification of the N-terminal sfGFP peptide MXKGEE containing pAzF, pBrF, or pCNF at position 2 when produced in Vmax X2, BL21, or Top10 cells. The N-terminal peptide MXKGEE was generated by Glu-C digestions of sfGFP samples obtained from the indicated strain in growths containing the indicated ncAA.

bioRxiv preprint doi: https://doi.org/10.1101/2021.06.22.449487; this version posted June 22, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.



Supplementary Figure 5. MS/MS identification of the N-terminal sfGFP peptide MXKGEE containing pAzF, pBrF, or pCNF at position 2 when produced in C321.ΔA.exp and C321.ΔA.opt cells. The N-terminal peptide MXKGEE was generated by Glu-C digestions of sfGFP samples obtained from the indicated strain in growths containing the indicated ncAA.

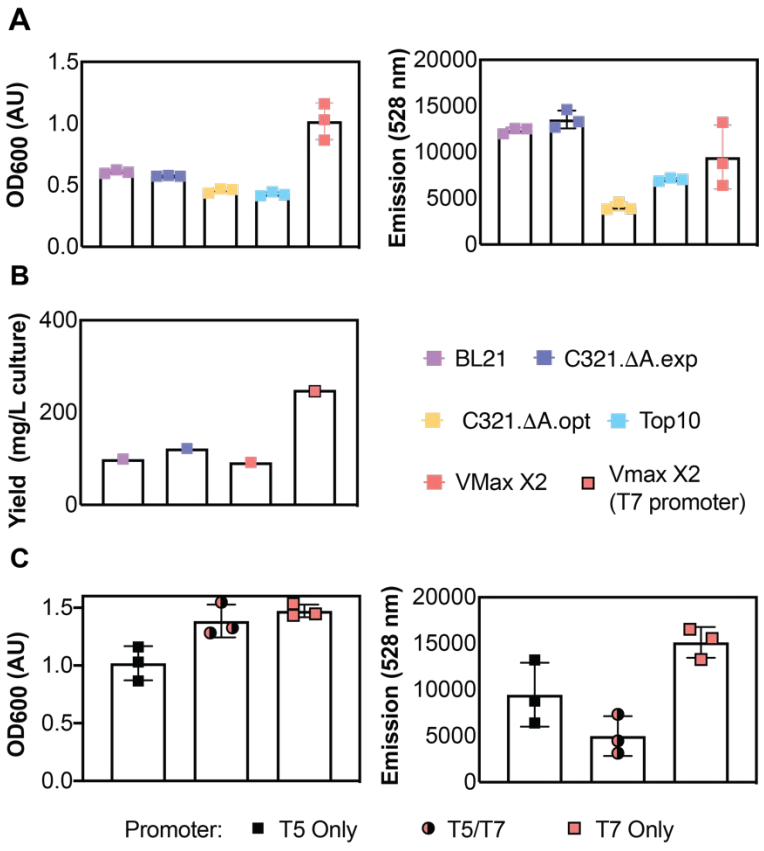
bioRxiv preprint doi: https://doi.org/10.1101/2021.06.22.449487; this version posted June 22, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.



Supplementary Figure 6. Pie charts plotting the distribution of amino acids incorporated at position 2 of sfGFP when expressed in the indicated strain. His-tagged proteins were isolated from growths of the indicated cells and digested with Glu-C. The relative abundance of the N-terminal sfGFP peptide fragments comprising the sequence MXKGEE were analyzed via LC-MS/MS to determine the identity of the amino acid at position 2. After sequence identification,

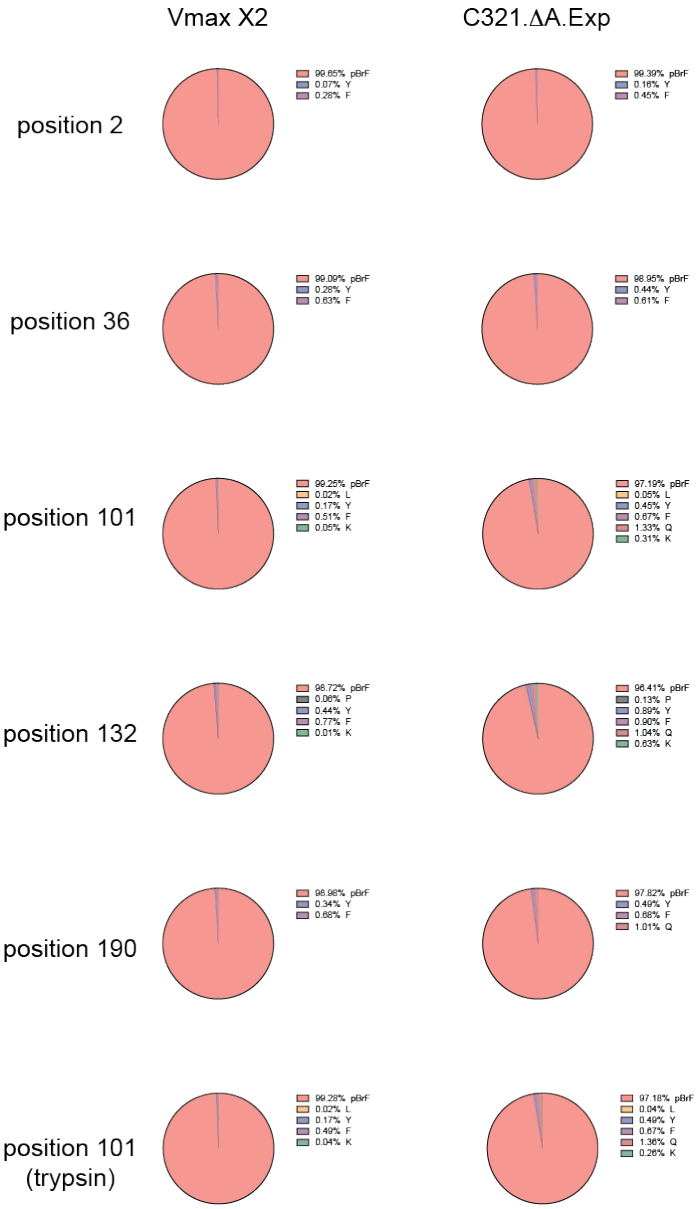
bioRxiv preprint doi: https://doi.org/10.1101/2021.06.22.449487; this version posted June 22, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

relative amounts were calculated by integrating the area under the peak for each extracted ion chromatogram.



Supplementary Figure 7. Comparison of GCE in Vmax X2 *versus* traditional (Top10, BL21) and genomically recoded (C321)[41,42] using plasmids under the control of T5 and/or T7 promoters. (A) Plot of the OD_{600} and emission at 528 nm of each cell growth at the 4 h timepoint. All cells were transformed with pET-22B-151TAG sfGFP and pEVOL-CNF to induce expression of sfGFP bearing a ncAA at position 151 of sfGFP. After induction, cells were grown for 4 hours at 37°C (Vmax X2, BL21, Top10, C321.ΔA.exp) or 34°C (C321.ΔA.opt) in the presence of 0.5 mM pBrF. (B) Plot of the isolated yield of sfGFP obtained from each growth after 4 h incubation. The isolated yield of 2TAG sfGFP when expression is under control of the T7 promoter is shown for comparison. (C) Plots comparing the OD_{600} and 528 nm fluorescence Vmax X2 cells grown 0.5 mM pBrF as a function of promoter identity.

bioRxiv preprint doi: https://doi.org/10.1101/2021.06.22.449487; this version posted June 22, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.



Supplementary Figure 8. Pie charts plotting the distribution of amino acids incorporated at positions 2, 36, 101, 132, and 190 of sfGFP expressed in Vmax X2 or C321.DA.exp cells. His-tagged proteins were isolated from growths of the indicated cells and digested with Glu-C and/or trypsin.

bioRxiv preprint doi: https://doi.org/10.1101/2021.06.22.449487; this version posted June 22, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

II. Supplementary Tables

Supplementary Table 1. Isolated yields (mg/L) of ncAA-containing sfGFP isolated from Vmax X2, BL21, Top10, C321.ΔA.exp, and C321.ΔA.opt cells after 4 h incubation. VMax X2 and BL21 cells expressed sfGFP and pCNFRS under the control of T7 and arabinose promoters, respectively. Top10, C321.ΔA.exp, and C321.ΔA.opt cells expressed sfGFP and pCNFRS under the control of arabinose and tac promoters, respectively.

strain/ ncAA	VMax X2 (ratio)	BL21	C321.ΔA.exp	C321.ΔA.opt	Top10	Promoter (sfGFP/aaRS)
pAzF	355 (5-fold to 25-fold)	66				T7/Arabinose
		42	62	18	14	Arabinose/Tac
pBrF	226 (2-fold to 12-fold)	107				T7/Arabinose
			25	18	18	Arabinose/Tac
	92	99	122			T5/Arabinose
pCNF	115 (8-fold to 13-fold)	9				T7/Arabinose
			19	14	9	Arabinose/Tac

bioRxiv preprint doi: https://doi.org/10.1101/2021.06.22.449487; this version posted June 22, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Supplementary Table 2. Plasmids used in this study

Plasmid Name	Resistance	Ori	ORF 1		ORF 2	
			Promoter	Gene	Promoter	Gene
pET-S2TAG-sfGFP	Carb	pBR322	T7	S2TAG-sfGFP		
pBAD-S2TAG-sfGFP	Carb	pBR322	P_{BAD}	S2TAG-sfGFP		
pET22b-T5-151TAG-sfGFP	Carb	pBR322	T5	151TAG-sfGFP		
pET22b-T5/7-151TAG-sfGFP	Carb	pBR322	T5/T7	151TAG		
pET22b-T7-151TAG-sfGFP	Carb	pBR322	T7	151TAG		
pET-5xTAG-sfGFP	Carb	pBR322	T7	5xTAG-sfGFP		
pBAD-5xTAG-sfGFP	Carb	pBR322	P_{BAD}	sfGFP		
pULTRA-CNF	Spec	CloDF13	tac	pCNFRS		

bioRxiv preprint doi: https://doi.org/10.1101/2021.06.22.449487; this version posted June 22, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

pEVOL-CNF	Cm	p15A	P_{BAD}	pCNFRS	glnS	pCNFRS

III. Supplementary Methods

Bacterial Strains

Vmax X2, BL21 (DE3), and Top10 cells were purchased from Codex DNA, NEB (Catalog # C2527), and ThermoFisher (Catalog # C404010) respectively. C321.ΔA.Opt and C321.ΔA.Exp were gifts from George Church (Addgene plasmids #87359 and #49018).

Amino Acids

pAzF, pBrF, and pCNF were purchased from Chem-Impex International (Catalog # 06162, 04086, and 04110). BocK was purchased from Sigma-Aldrich (SKU 349661).

General Methods

The following antibiotic concentrations were used: carbenicillin, 100 µg/mL (*E. coli)* or 12.5 µg/mL (VMax X2); chloramphenicol 25 µg/mL; spectinomycin 50 µg/mL. Additionally, C321.ΔA.Opt and C321.ΔA.Exp starter cultures were grown in the presence of 15 µg/mL gentamicin or 25 µg/mL Zeocin™ (ThermoFisher) respectively.

Transformation protocols

Vmax X2 and *E. coli* (BL21, C321, Top10) cells were transformed in accordance with manufacturer protocols with some modifications as follows. Frozen stocks were thawed on ice. Upon thawing, 1 µL of plasmid (see below) encoding sfGFP and the orthogonal synthetase was added. After a 30 min incubation on ice, cells were heat shocked for 45 (Vmax X2) or 90 s (BL21, C321, Top10) and put back on ice for 2 minutes. 550 µL of Vmax recovery media (Vmax X2) or SOB media was added and cells were recovered at 34°C (C321.ΔA.opt) or 37°C for one (*E. coli*) or four hours (Vmax X2) before plating. For 2TAG-sfGFP expression, Vmax X2 and BL21 cells were transformed with pEVOL-CNF and pET-S2TAG-sfGFP, while C321 and Top10 cells were transformed with pULTRA-CNF and pBAD-S2TAG-sfGFP. pEVOL-mmPyl was used in place of pEVOL-CNF for expression of sfGFP containing BocK at the 2nd position. For 5XTAG-sfGFP expression Vmax X2 and BL21 cells were transformed with pEVOL-CNF and pET-5XTAG-sfGFP, C321 and Top10 cells were transformed with pULTRA-CNF and pBAD-

bioRxiv preprint doi: https://doi.org/10.1101/2021.06.22.449487; this version posted June 22, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

5XTAG-sfGFP. For 151TAG-sfGFP expression, all strains were transformed with pEVOL-CNF and pET22b-151TAG-sfGFP.

Expression of sfGFP variants

Starter cultures were grown for 3 hours (Vmax X2) or overnight (*E. coli*) in 25 mL of BHI (Teknova, Catalog #B9505) + v2salts (204 mM NaCl, 23.2 mM $MgCl_2$, 4.2 mM KCl) or 10 mL of LB Miller (AmericanBio, Catalog #AB01201) supplemented with antibiotics at 34°C (C321.A.opt) or 37°C. To maximize aeration growth rate, starter cultures for Vmax X2 cells were grown in a baffled flask. Following the initial incubation period, starter cultures were diluted 1:100 into 25 mL BHI + v2salts or LB supplemented with 0.5 mM (pAzF, pBrF, pCNF) or 10 mM (BocK) ncAA. Once cultures reached an OD of 0.5, protein expression was induced by addition of 1 mM IPTG and 0.2% arabinose.

Purification of sfGFP variants

Cultures were centrifuged for 15 minutes at 10,000 x g and 4°C. Pellets were resuspended in 15 mL of lysis buffer (50 mM sodium phosphate (pH 8), 300 mM NaCl, 20 mM imidazole,) supplemented with 1 tablet cOmplete, mini EDTA-free protease inhibitor cocktail (Sigma-Aldrich, St. Louis, MO) and sonicated for 5 minutes total (30s on, 30s off) at 30% duty cycle. Following sonication, the soluble fraction was isolated by centrifugation of the lysate for 25 minutes at 10,000 x g and 4°C. The supernatant was isolated and incubated with 500 µL of Ni-NTA (Qiagen, Catalog # 30230) resin for an hour at 4°C. Slurry was poured onto a gravity flow column and the resin was washed with 15 mL of lysis buffer following drainage of the flowthrough. Bound protein was then eluted by the addition of 3.5 mL of elution buffer (50 mM sodium phosphate (pH 8), 250 mM imidazole). For quantification and MS analysis, the eluent was buffer exchanged into 50 mM sodium phosphate utilizing a PD-10 column.

Intact Protein Mass Spectomery Analysis

LC/MS analysis was performed on an Agilent 1290 Infinity II HPLC connected to an Agilent 6530B QTOF AJS-ESI. 1 µg of protein was injected onto a Poroshell 300SB-C8 column (2.1 x 75 mm, 5-Micron, room temperature, Agilent) using a linear gradient from 5 to 75% acetonitrile over 9.5 minutes with 0.1% formic acid as the aqueous phase after an initial hold at 5% acetonitrile for 0.5 min (0.4 mL/min). The following parameters were used during acquisition: Fragmentor voltage 175 V, gas temperature 300ºC, gas flow 12 L/min, sheath gas temperature 350ºC, sheath gas flow 12 L/min, nebulizer pressure 35 psi, skimmer voltage 65 V, Vcap 5000

bioRxiv preprint doi: https://doi.org/10.1101/2021.06.22.449487; this version posted June 22, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

V, 3 spectra/s. Intact protein masses were obtained via deconvolution using the Maximum Entropy algorithm in Mass Hunter Bioconfirm (V10, Agilent).

Monitoring of cell growth and sfGFP expression over time

25 mL cultures were grown in 250 mL baffled flask and expression sfGFP variants and pCNFRS was induced as described previously. At each timepoint, 100 uL aliquots of each culture were transferred onto a black, clear bottom, 96-well plate. The OD_{600} and emission at 528 nm (λ_{ex}= 485 nm) was measured in BioTek Synergy H1 microplate reader.

Fidelity of ncAA incorporation by LC-MS/MS.

To determine the fidelity of amino acid incorporation at position 2 of sfGFP, isolated sfGFP (13 to 72 µg, most at ~25 µg) was denatured with 6 M guanidine in a 0.15 M Tris buffer at pH 7.5, followed by disulfide reduction with 8 mM dithiothreitol (DTT) at 37 °C for 30 min. The reduced sfGFP was alkylated in the presence of 14 mM iodoacetamide at 25 °C for 25 min, and then quenched using 6 mM DTT. The reduced/alkylated protein was exchanged into ~50 µL of 0.1 M Tris buffer at pH 7.5 using a Microcon 10-KDa membrane, followed by addition of 2.5 to 7.0 µg endoproteinase Glu-C (in a 0.5 µg/µL solution) directly to the membrane to achieve an enzyme-to-substrate ratio of at least 1:10. After 3 hours at 37°C, the digestion was quenched with an equal volume of 0.25 M acetate buffer (pH 4.8) containing 6 M guanidine. Peptide fragments were collected by spinning down through the membrane and subjected to LC-MS/MS analysis.

To determine the fidelity amino acid incorporation at the remaining 4 positions, isolated sfGFP was also digested by trypsin, with the same procedure as described above, except trypsin was used in place of Glu-C, and digested was allowed to proceed for 1 hour instead of 3 hours.

LC-MS/MS analysis was performed on an Agilent 1290-II HPLC directly connected to a Thermo Fisher Q Exactive high-resolution mass spectrometer. Peptides were separated on a Waters HSS T3 reversed-phase column (2.1 × 150 mm) at 50°C with a 70-min acetonitrile gradient (0.5% to 35%) containing 0.1% formic acid in the mobile phase, and a total flow rate of 0.25 mL/min. The MS data were collected at 70,000 resolution, followed by data-dependent higher-energy collision dissociation (HCD) MS/MS at a normalized collision energy of 25%. Proteolytic peptides were identified and quantified on MassAnalyzer, an in-house developed program (available in Biopharma Finder from Thermo Fisher). The program performs feature extraction,

bioRxiv preprint doi: https://doi.org/10.1101/2021.06.22.449487; this version posted June 22, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

peptide identification, retention time alignment, and relative quantitation in an automated fashion.

Sequences

*Denotes a stop codon

2TAG-sfGFP:

M*KGEELFTGVVPILVELDGDVNGHKFSVRGEGEGDATNGKLTLKFICTTGKLPVPWPTLVTTL
TYGVQCFSRYPDHMKRHDFFKSAMPEGYVQERTISFKDDGTYKTRAEVKFEGDTLVNRIELKG
IDFKEDGNILGHKLEYNFNSHNVYITADKQKNGIKANFKIRHNVEDGSVQLADHYQQNTPIGDG
PVLLPDNHYLSTQSVLSKDPNEKRDHMVLLEFVTAAGITHGMDELYKGSHHHHHH

5XTAG-sfGFP:

M*KGEELFTGVVPILVELDGDVNGHKFSVRGEGEG*ATNGKLTLKFICTTGKLPVPWPTLVTTLT
YGVQCFSRYPDHMKRHDFFKSAMPEGYVQERTISF*DDGTYKTRAEVKFEGDTLVNRIELKGID
FK*DGNILGHKLEYNFNSHNVYITADKQKNGIKANFKIRHNVEDGSVQLADHYQQNTPIG*GPVL
LPDNHYLSTQSVLSKDPNEKRDHMVLLEFVTAAGITHGMDELYKGSHHHHHH

151TAG-sfGFP:

MSKGEELFTGVVPILVELDGDVNGHKFSVRGEGEGDATNGKLTLKFICTTGKLPVPWPTLVTTL
TYGVQCFSRYPDHMKRHDFFKSAMPEGYVQERTISFKDDGTYKTRAEVKFEGDTLVNRIELKG
IDFKEDGNILGHKLEYNFNSHNV*ITADKQKNGIKANFKIRHNVEDGSVQLADHYQQNTPIGDGP
VLLPDNHYLSTQSVLSKDPNEKRDHMVLLEFVTAAGITHGMDELYKGSHHHHHH

pCNFRS:

MDEFEMIKRNTSEIISEEELREVLKKDEKSALIGFEPSGKIHLGHYLQIKKMIDLQNAGFDIIIVLAD
LHAYLNQKGELDEIRKIGDYNKKVFEAMGLKAKYVYGSEWMLDKDYTLNVYRLALKTTLKRAR
RSMELIAREDENPKVAEVIYPIMQVNGAHYLGVDVAVGGMEQRKIHMLARELLPKKVVCIHNPV
LTGLDGEGKMSSSKGNFIAVDDSPEEIRAKIKKAYCPAGVVEGNPIMEIAKYFLEYPLTIKRPEKF
GGDLTVNSYEELESLFKNKELHPMDLKNAVAEELIKILEPIRKRL

mmPylRS

MDKKPLNTLISATGLWMSRTGTIHKIKHHEVSRSKIYIEMACGDHLVVNNSRSSRTARALRHHK
YRKTCKRCRVSDEDLNKFLTKANEDQTSVKVKVVSAPTRTKKAMPKSVARAPKPLENTEAAQA

bioRxiv preprint doi: https://doi.org/10.1101/2021.06.22.449487; this version posted June 22, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

QPSGSKFSPAIPVSTQESVSVPASVSTSISSISTGATASALVKGNTNPITSMSAPVQASAPALTK
SQTDRLEVLLNPKDEISLNSGKPFRELESELLSRRKKDLQQIYAEERENYLGKLEREITRFFVDR
GFLEIKSPILIPLEYIERMGIDNDTELSKQIFRVDKNFCLRPMLAPNLYNYLRKLDRALPDPIKIFEI
GPCYRKESDGKEHLEEFTMLNFCQMGSGCTRENLESIITDFLNHLGIDFKIVGDSCMVYGDTLD
VMHGDLELSSAVVGPIPLDREWGIDKPWIGAGFGLERLLKVKHDFKNIKRAARSESYYNGISTN
L

IV. Supplemental References

(1) Kwok, H. S.; Vargas-Rodriguez, O.; Melnikov, S. V.; Söll, D. Engineered Aminoacyl-TRNA Synthetases with Improved Selectivity toward Noncanonical Amino Acids. *ACS Chem. Biol.* **2019**, *14* (4), 603–612. https://doi.org/10.1021/acschembio.9b00088.

(2) Wang, Y. S.; Fang, X.; Wallace, A. L.; Wu, B.; Liu, W. R. A Rationally Designed Pyrrolysyl-TRNA Synthetase Mutant with a Broad Substrate Spectrum. *J Am Chem Soc* **2012**, *134* (6), 2950–2953. https://doi.org/10.1021/ja211972x.

(3) Young, D. D.; Young, T. S.; Jahnz, M.; Ahmad, I.; Spraggon, G.; Schultz, P. G. An Evolved Aminoacyl-TRNA Synthetase with Atypical Polysubstrate Specificity. *Biochemistry* **2011**, *50* (11), 1894–1900. https://doi.org/10.1021/bi101929e.