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

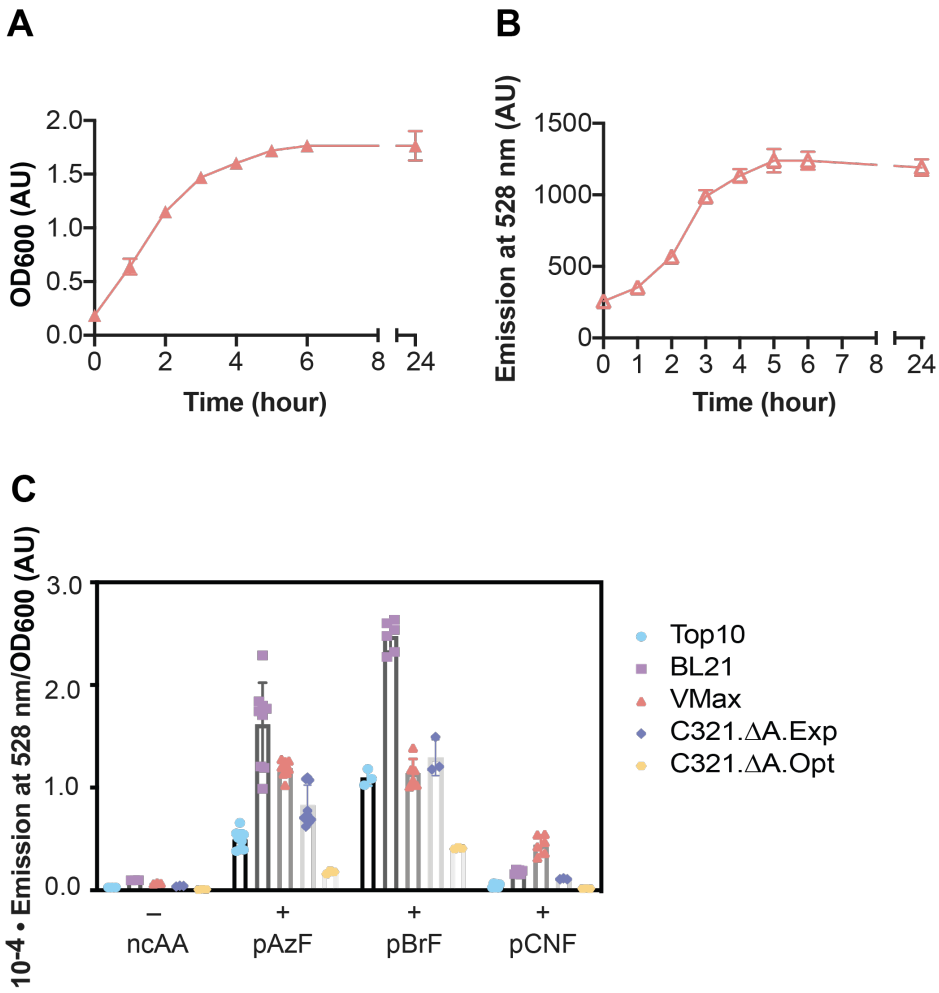
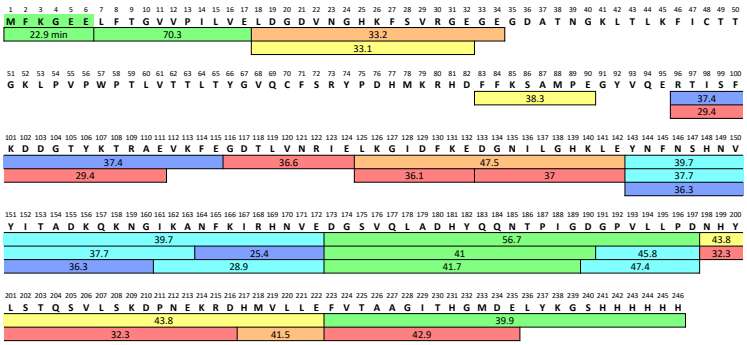
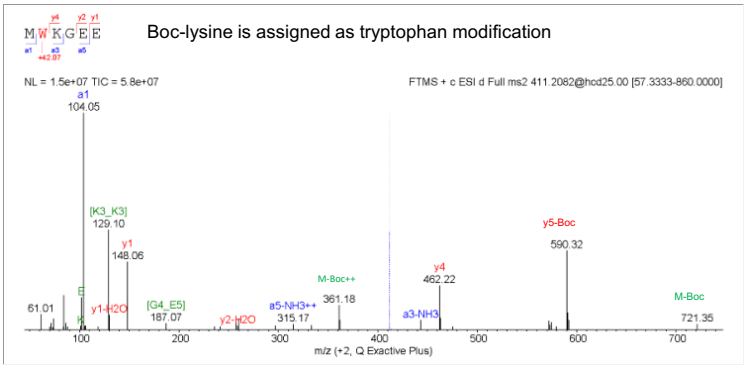


Figure S1. **Growth and sfGFP expression of Vmax X2 cells transformed with pET-S2TAGsfGFP and pEVOL-mmPyl.** Vmax X2 cells were transformed with pET-S2TAGsfGFP[1] and pEVOL-mmPyl[2] to induce expression of sfGFP bearing a ncAA at the second position of sfGFP. After induction, cells were incubated at 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).

A



B Vmax X2/BocK



C Vmax X2/pAzF

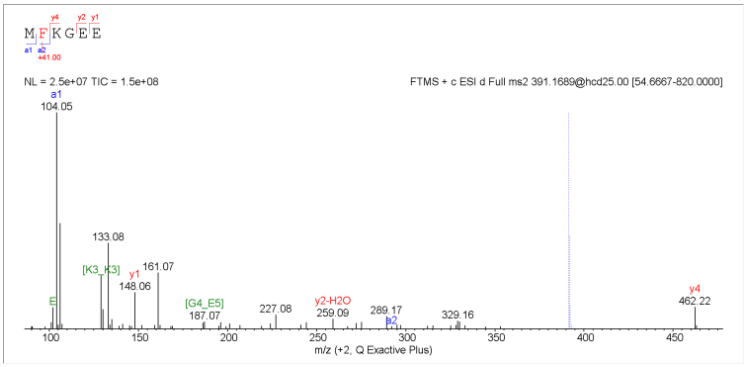


Figure S2. LC-MS/MS analysis of sfGFP expressed in Vmax X2. Vmax X2 cells were transformed with pET-S2TAGsfGFP[1] and either pEVOL-CNF[3] or pEVOL-mmPyl[2] to induce expression of sfGFP bearing a ncAA at position two. Cells were induced and incubated for 4 hours at 37°C in the presence of 0.5 mM pAzF (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%).

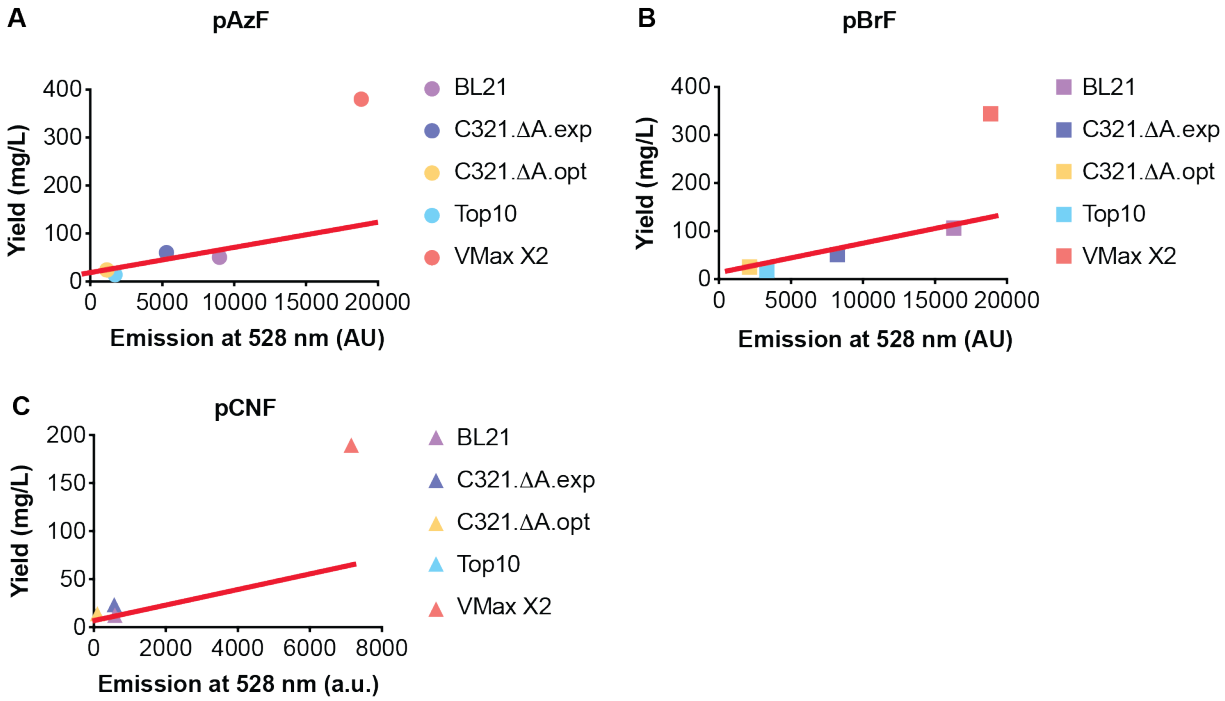


Figure S3. Isolated yield of sfGFP containing a ncAA at position 2 is higher than expected based on unit 528 nm emission. Plot of the isolated yield of sfGFP containing (A) pAzF; (B) pBrF; or (C) pCNF at position 2 versus the absorbance of the cell growth 4 h after induction. The red line shows the yield expected if yield correlated directly with unit 528 nm emission at 4 h.

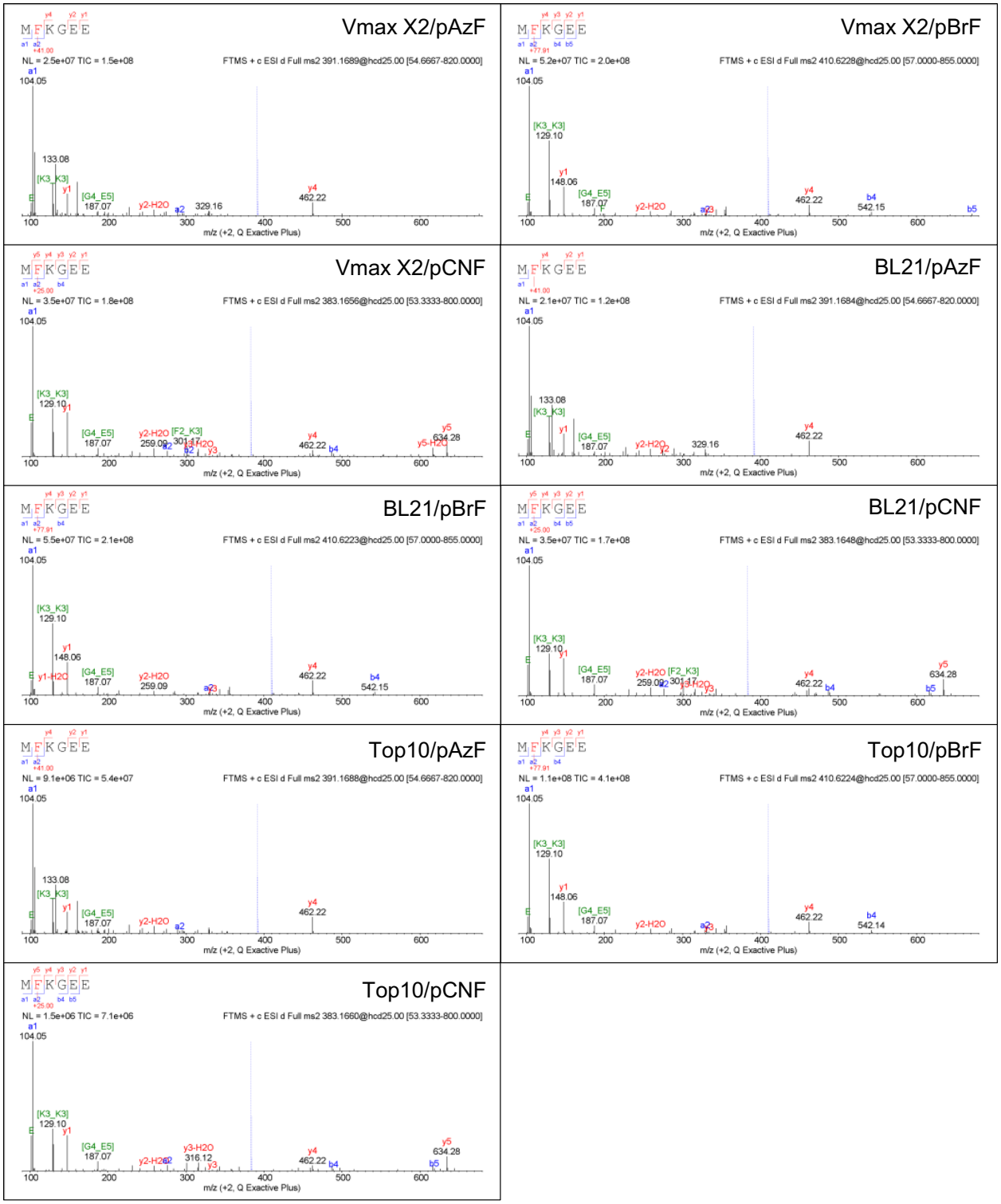


Figure S4. MS/MS identification of the N-terminal sfGFP peptide MXKGEE containing pAzF, pBrF, or pCNF at position 2 when produced in Vmax X2, BL21, or Top10 cells. The N-terminal peptide MXKGEE was generated by Glu-C digestions of sfGFP samples obtained from the indicated strain in growths containing the indicated ncAA.

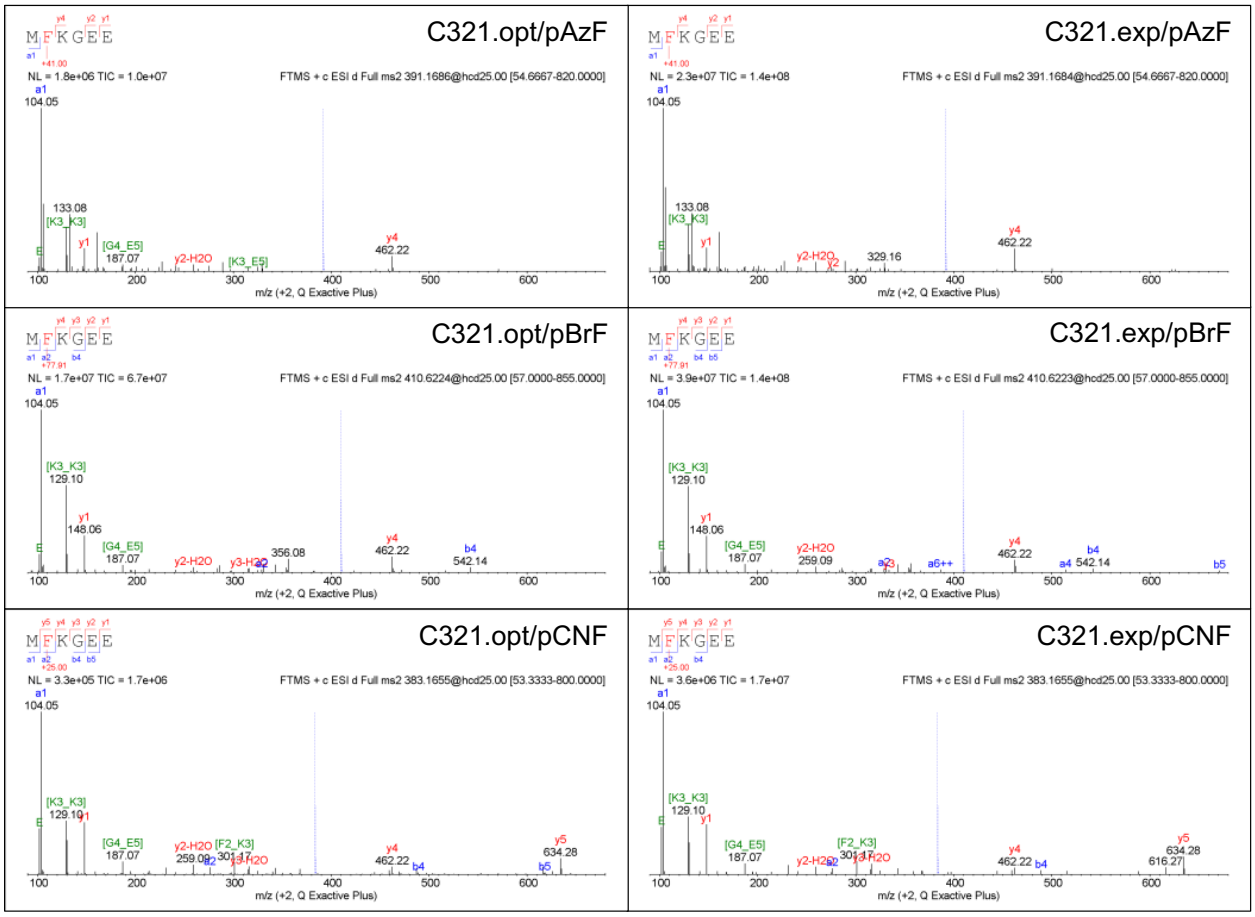


Figure S5. MS/MS identification of the N-terminal sfGFP peptide MXKGEE containing pAzF, pBrF, or pCNF at position 2 when produced in C321.ΔA.exp and C321.ΔA.opt cells. The N-terminal peptide MXKGEE was generated by Glu-C digestions of sfGFP samples obtained from the indicated strain in growths containing the indicated ncAA.

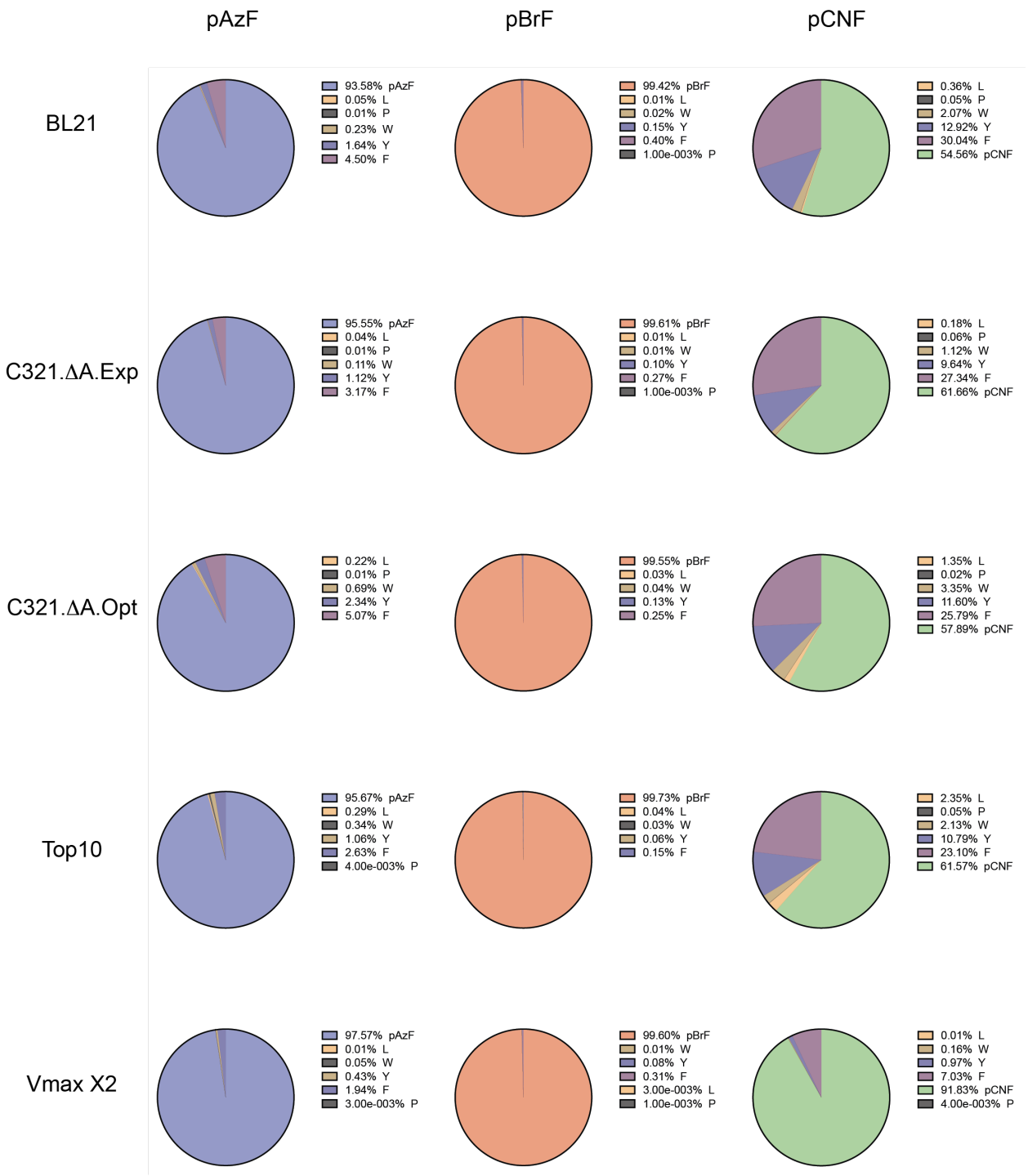


Figure S6. Pie charts plotting the distribution of amino acids incorporated at position 2 of sfGFP when expressed in the indicated strain. His-tagged proteins were isolated from growths of the indicated cells and digested with Glu-C. The relative abundance of the N-terminal sfGFP peptide fragments comprising the sequence MXKGEE were analyzed via LC-MS/MS to determine the identity of the amino acid at position 2. After sequence identification, relative

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

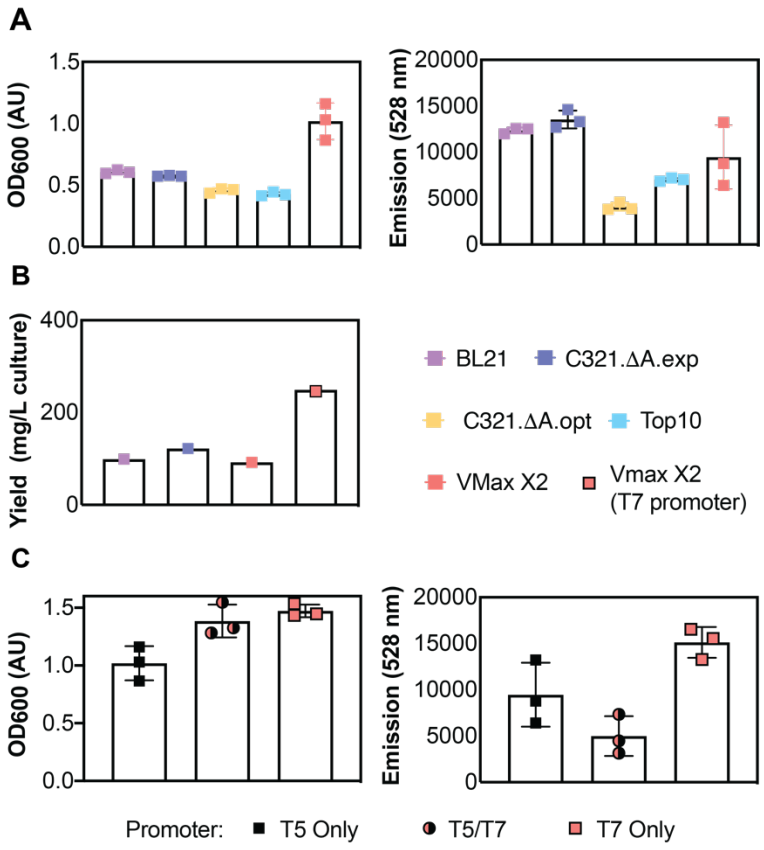


Figure S7. Comparison of GCE in Vmax X2 *versus* traditional (Top10, BL21) and genomically recoded (C321)[41,42] using plasmids under the control of T5 and/or T7 promoters. (A) Plot of the OD_{600} and emission at 528 nm of each cell growth at the 4 h timepoint. All cells were transformed with pET-22B-151TAG 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.

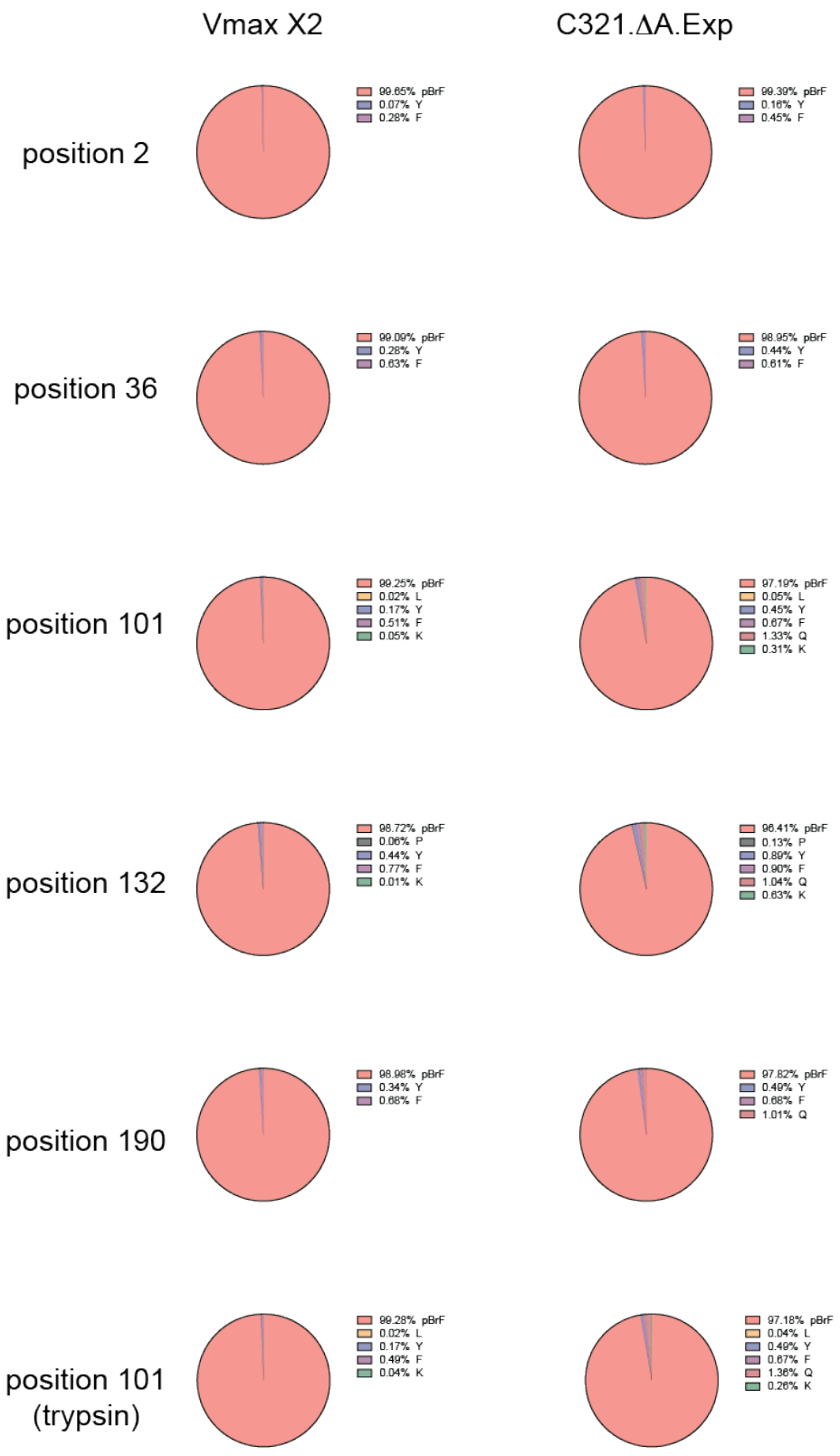


Figure S8. Pie charts plotting the distribution of amino acids incorporated at positions 2, 36, 101, 132, and 190 of sfGFP expressed in Vmax X2 or C321.ΔA.exp cells. His-tagged proteins were isolated from growths of the indicated cells and digested with Glu-C and/or trypsin.

II. Supplementary Tables

Table S1. 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

Table S2. 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		

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

Table S3. Distribution of amino acids incorporated at position 2 of sfGFP in the presence of pAzF (accounting for the reduction of the ncAA to pNH_2F).

Strain / **Amino Acid**	**BL21**	**C321.ΔA.Exp**	**C321.ΔA.Opt**	**Top10**	**Vmax X2**
pAzF	92.86	94.11	89.86	94.61	96.47
pNH_2F	0.58	1.50	1.82	0.82	1.11
L	0.04	0.04	0.22	0.28	0.01
P	0.01	0.01	0.01	0.00	0.00
W	0.23	0.11	0.68	0.05	0.05
Y	1.63	1.10	2.28	0.11	0.42
F	4.44	3.10	4.96	.27	1.91

III. Supplementary Methods

Safety Statement: No unexpected or unusually high safety hazards were encountered.

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-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 Spectrometry 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 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, peptide identification, retention time alignment, and relative quantitation in an automated fashion. Because the use of DTT and high temperatures can trigger the reduction of pAzF to pNH_2F, the peptide peaks containing the reduced azide were not included in the fidelity calculations shown in **Supplementary Figure 6**, but can be found in **Supplementary Table 3**.

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
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.