Electronic Supplementary Material (ESI) for Molecular BioSystems.

Supplementary information for:

Performance of optimized noncanonical amino acid mutagenesis systems in the absence of release factor 1

Yunan Zheng,[a] Marc J. Lajoie,[bc] James S. Italia,[a] Melissa Chin,[a] George M. Church,[b] Abhishek Chatterjee*[a]

[a] Department of Chemistry, Boston College, 2609 Beacon Street, Chestnut Hill, MA 02467, USA. Email: abhishek.chatterjee@bc.edu; Tel: +1-617-552-1778

[b] Department of genetics, Harvard Medical School, 77 Avenue Louis Pasteur, Boston, MA 02115, USA

[c] Department of Biochemistry, University of Washington, Seattle, WA 98195, USA.

Materials and methods:

For cloning and plasmid propagation, the DH10B (Life Technologies, Carlsbad, CA) strain of *E. coli* was used. Polymerase chain reaction (PCR) was performed using the Phusion Hot Start II DNA polymerase (Fisher Scientific, Waltham, MA) using manufacturer's protocol. For purification of DNA (plasmid as well as PCR products, etc.) spin columns from Epoch Life Science (Fort Bend County, TX) were used. Restriction enzymes and T4 DNA ligase were obtained from NEB (Ipswich, MA). DNA oligomers for PCR were purchased from IDT (Coralville, IA). Verification of cloned DNA by Sanger sequencing was performed by Eton Biosciences. Antibiotics, isopropyl β-D-1- thiogalactopyranoside (IPTG), and L-arabinose were purchased from Sigma-Aldrich (St. Louis, MO) or Fisher Scientific (Waltham, MA). Components of media were obtained from Fisher Scientific (Waltham, MA). Bacteria were grown on LB/agar plates and LB liquid medium with the following antibiotic concentrations unless otherwise mentioned: 50 μg/mL spectinomycin, 50 μg/mL chloramphenicol, and 100 μg/mL ampicillin. C321.ΔRF1 strain, where the λ-prophage was replaced with a zeocin resistance cassette was used in our experiments. Various ncAAs were obtained from following vendors: O-methyltyrosine (1), *p*-iodophenylalanine (2) and *p*-boronophenylalanine (6) were purchased from Fisher Scientific (Waltham, MA), *p*-azidophenylalanine (3) and O-sulfotyrosine (7) from Bachem (Bubendorf, Switzerland), *p*-acetylphenylalanine (4) from Chem Impex International (Wood Dale, IL). O-propargyltyrosine was synthesized as reported earlier.[1]

Construction of reporter plasmids: The construction of the pET22b-*t5.lac*-sfGFP reporter plasmid has been described previously.[2] To construct the pET22b-*t5.lac*-sfGFP plasmids, containing 1, 2, 4, 6, or 10 contiguous N-terminal UAG stop codons, sfGFP was amplified using the primers listed below, and the previously reported pET22b-*t5.lac*-sfGFP as the template. The PCR products were digested with NdeI/HindIII to replace the wild-type sfGFP gene in the aforementioned plasmid. The gene encoding the sfGFP construct, where 8 surface exposed threonine residues are replaced with UAG codons, was synthesized as a gBlock® by IDT (Coralville, IA). This synthetic DNA was cloned into the pET22b-*t5.lac* vector using the NdeI/HindIII site analogous to the contiguous mutants.

List of primers:

1TAG-sfGFP-NdeI-F:
AATAATCATATGGCTTAGAGCAAAGGAGAAGAACTTTTCACTGGAGTTG
2TAG-sfGFP-NdeI-F:
AATAATCATATGGCTTAGTAGAGCAAAGGAGAAGAACTTTTCACTGGAGTTGTCCC
4TAG-sfGFP-NdeI-F:
AATAATCATATGGCTTAGTAGTAGTAGAGCAAAGGAGAAGAACTTTTCACTGGAGTTGTC
6TAG-sfGFP-NdeI-F:
AATAATCATATGGCTTAGTAGTAGTAGTAGTAGAGCAAAGGAGAAGAACTTTTCACTGGAG
TTG
10TAG-sfGFP-NdeI-F:
AATAATCATATGGCTTAGTAGTAGTAGTAGTAGTAGTAGTAGAGCAAAGGAGAAGAAC
TTTTCACTGGAGTTG
sfGFP-HindIII-R: AATAATAATTAAGCTTTTAGTGGTGGTGGTGGTGG

sfGFP Expression:

Various strains, carrying the expression plasmid (pET22-*t5.lac* vector encoding wild type or mutant sfGFP), and the desired pUltra or pEVOL suppressor plasmid, were grown in LB medium supplemented with ampicillin and spectinomycin (pUltra) or chloramphenicol (pEVOL) at 37 °C until the OD_{600} reach between 0.5 to 0.6, at which point IPTG (1 mM) was added to induce the expression of sfGFP and the synthetase (for pUltra). When pEVOL suppression vectors were used, the aaRS expression was also induced at this point by adding 0.02% L-arabinose. For high-throughput analysis, the induced cultures were transferred to sterilized deep 96-well polypropylene plates (500 µL per well). Various non-canonical amino acids were added to the final concentration of 1 mM at this point. Cell growth was allowed to continue for 16 h at 30 °C with shaking. To evaluate sfGFP expression, cells from 100 µL of culture were resuspended in 100 µL PBS and transferred to a 96-well clear–bottom assay plate, and the fluorescence was measured by using a SpectraMAX M5 (Molecular Devices) (Excitation at 480 nm and Emission at 510 nm). Fluorescence of each sample was normalized using its OD_{600}.

To isolate protein for ESI-MS characterization, 5-25 mL of cultures of the appropriate strains, carrying the reporter and the suppression vectors, in terrific broth medium was used. Protein expression was performed as described above, and the overnight expression cultures were centrifuged and resuspended in lysis buffer: B-PER Bacterial Protein Extraction Reagent (Thermo Scientific) + 1X Halt Protease Inhibitor Cocktail (Thermo Scientific) + 0.01% Pierce Universal Nuclease (Thermo Scientific). After a 30 min incubation on ice, the lysate was clarified by centrifuging at 18000 xg for 10 min. The C-terminally polyhistidine tagged soluble sfGFP in the supernatant was purified by using HisPur Ni-NTA Resin (Thermo Scientific) following manufacture's protocol. The purity of the protein was confirmed by SDS-PAGE analysis, and the molecular weight of the purified protein was confirmed by ESI-MS (Agilent Technologies 1260 Infinity ESI-TOF).

Sequence of contiguous sfGFP reporters:

ATGGCT$(TAG)_n$AGCAAAGGAGAAGAACTTTTCACTGGAGTTGTCCCAATTCTTGTTGAATTAG
ATGGTGATGTTAATGGGCACAAATTTTCTGTCCGTGGAGAGGGTGAAGGTGATGCTACAAAC
GGAAAACTCACCCTTAAATTTATTTGCACTACTGGAAAACTACCTGTTCCGTGGCCAACACT
TGTCACTACTCTGACCTATGGTGTTCAATGCTTTTCCCGTTATCCGGATCACATGAAACGGCA
TGACTTTTTCAAGAGTGCCATGCCCGAAGGTTATGTACAGGAACGCACTATATCTTTCAAAG
ATGACGGGACCTACAAGACGCGTGCTGAAGTCAAGTTTGAAGGTGATACCCTTGTTAATCGT
ATCGAGTTAAAGGGTATTGATTTTAAAGAAGATGGAAACATTCTTGGACACAAACTCGAGTA
CAACTTTAACTCACACAATGTATACATCACGGCAGACAAACAAAAGAATGGAATCAAAGCT
AACTTCAAAATTCGCCACAACGTTGAAGATGGTTCCGTTCAACTAGCAGACCATTATCAACA
AAATACTCCAATTGGCGATGGCCCTGTCCTTTTACCAGACAACCATTACCTGTCGACACAAT
CTGTCCTTTCGAAAGATCCCAACGAAAAGCGTGACCACATGGTCCTTCTTGAGTTTGTAACT
GCTGCTGGGATTACACATGGCATGGATGAGCTCTACAAAGGATCCCACCACCACCACCACCA
CTAA

Sequence of sfGFP reporter with 8 TAG codons replacing surface exposed threonines:

ATGAGCAAAGGAGAAGAACTTTTCACTGGAGTTGTCCCAATTCTTGTTGAATTAGATGGTGA
TGTTAATGGGCACAAATTTTCTGTCCGTGGAGAGGGTGAAGGTGATGCTtagAACGGAAAACT
CtagCTTAAATTTATTTGCtagACTGGAAAACTACCTGTTCCGTGGCCAACACTTGTCACTACTC
TGACCTATGGTGTTCAATGCTTTTCCCGTTATCCGGATCACATGAAACGGCATGACTTTTTCA
AGAGTGCCATGCCCGAAGGTTATGTACAGGAACGCtagATATCTTTCAAAGATGACGGGtagTA
CAAGACGCGTGCTGAAGTCAAGTTTGAAGGTGATtagCTTGTTAATCGTATCGAGTTAAAGGG
TATTGATTTTAAAGAAGATGGAAACATTCTTGGACACAAACTCGAGTACAACTTTAACTCAC
ACAATGTATACATCtagGCAGACAAACAAAAGAATGGAATCAAAGCTAACTTCAAAATTCGC
CACAACGTTGAAGATGGTTCCGTTCAACTAGCAGACCATTATCAACAAAATtagCCAATTGGC
GATGGCCCTGTCCTTTTACCAGACAACCATTACCTGTCGACACAATCTGTCCTTTCGAAAGAT
CCCAACGAAAAGCGTGACCACATGGTCCTTCTTGAGTTTGTAACTGCTGCTGGGATTACACA
TGGCATGGATGAGCTCTACAAAGGATCCCATCATCACCATCACCATTAA

Supplementary figures:

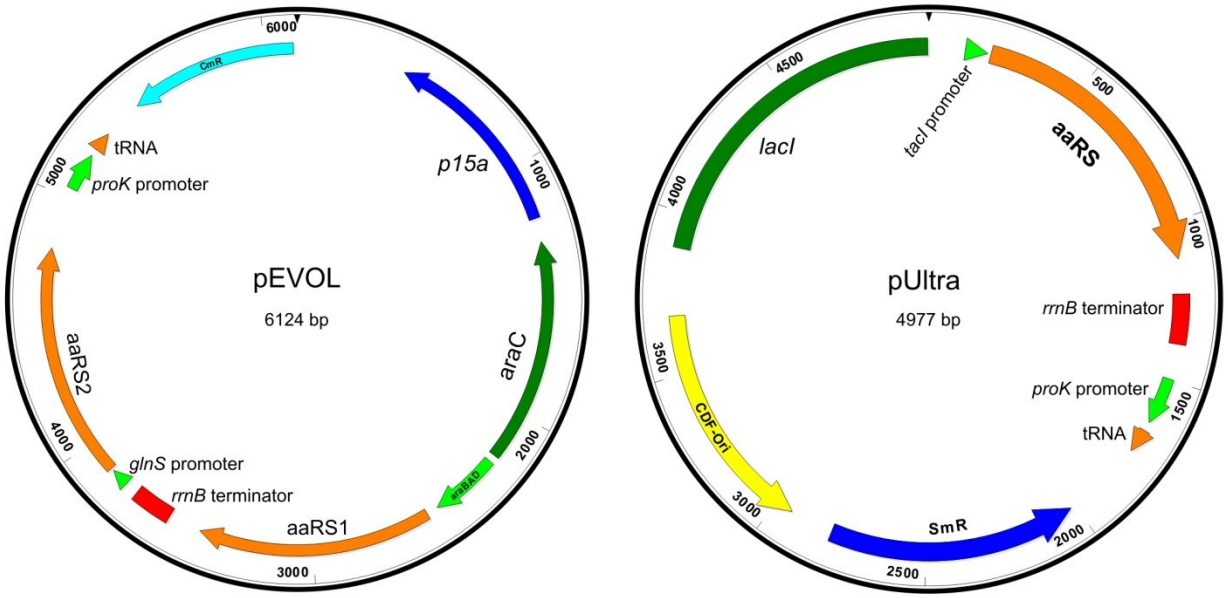

Figure S1: Vector maps of pEVOL and pUltra

Figures S2-S8: These figures show the expression levels of sfGFP reporters (harboring increasing numbers of contiguous N-terminal UAG codons) in strains with or without a functional RF1. Individual figure legends describe the suppression vector (pEVOL or pUltra) used, the orthogonal tRNA/aaRS pair encoded therein, and the ncAA incorporated in response to the UAG codons. The number of contiguous UAG stop codons present in the reporter is indicated along the x-axis, and the expression levels in the presence or the absence of the ncAA is denoted by '+' and '-', respectively.

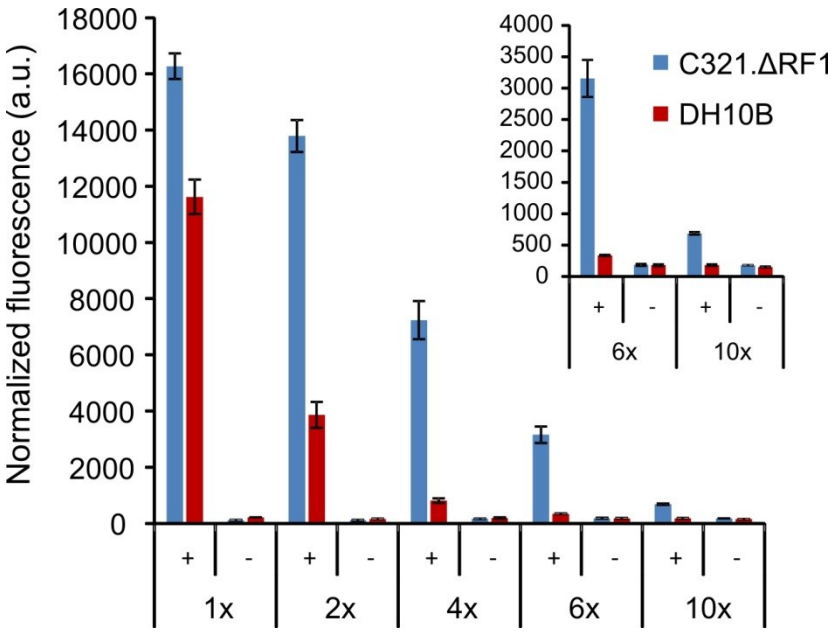

Figure S2: Expression of sfGFP reporters with increasing N-terminal UAG codons using a pEVOL plasmid encoding an *M. jannaschii*-derived polyspecific tyrosyl tRNA/aaRS pair. O-methyltyrosine (1, Fig 3C) was used as a substrate. Data for reporters with 6 or 10 UAG codons are magnified at inset.

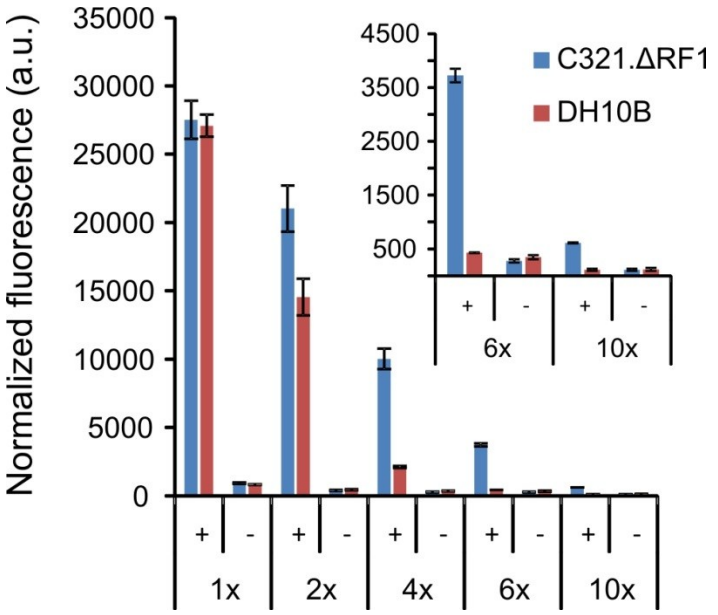

Figure S3: Expression of sfGFP reporters with increasing N-terminal UAG codons using a pUltra plasmid encoding an *M. jannaschii*-derived polyspecific tyrosyl tRNA/aaRS pair. *p*-iodo-phenylalanine (2, Fig 3C) was used as a substrate. Data for reporters with 6 or 10 UAG codons are magnified at inset.

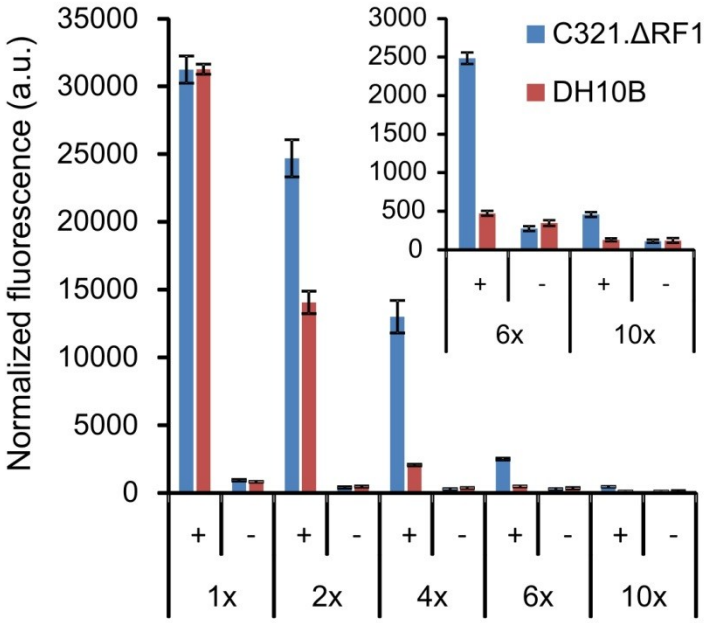

Figure S4: Expression of sfGFP reporters with increasing N-terminal UAG codons using a pUltra plasmid encoding an *M. jannaschii*-derived polyspecific tyrosyl tRNA/aaRS pair. *p*-azido-phenylalanine (**3**, Fig 3C) was used as a substrate. Data for reporters with 6 or 10 UAG codons are magnified at inset.

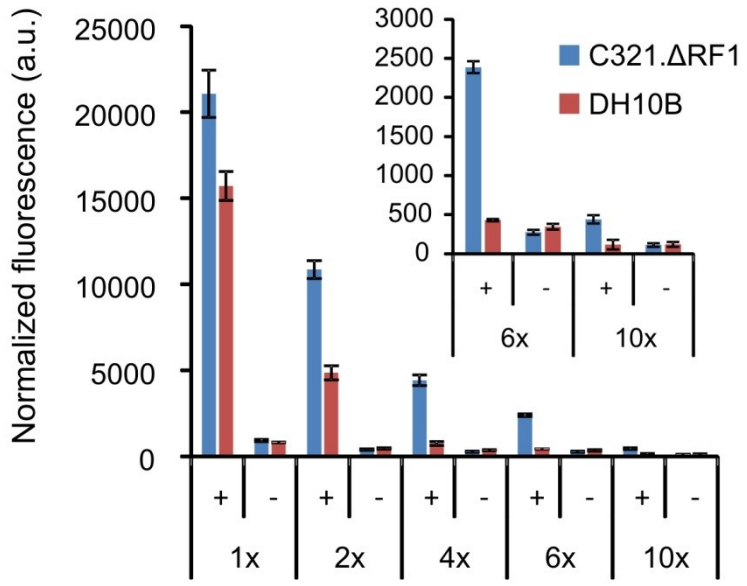

Figure S5: Expression of sfGFP reporters with increasing N-terminal UAG codons using a pUltra plasmid encoding an *M. jannaschii*-derived polyspecific tyrosyl tRNA/aaRS pair. *p*-acetyl-phenylalanine (**4**, Fig 3C) was used as a substrate. Data for reporters with 6 or 10 UAG codons are magnified at inset.

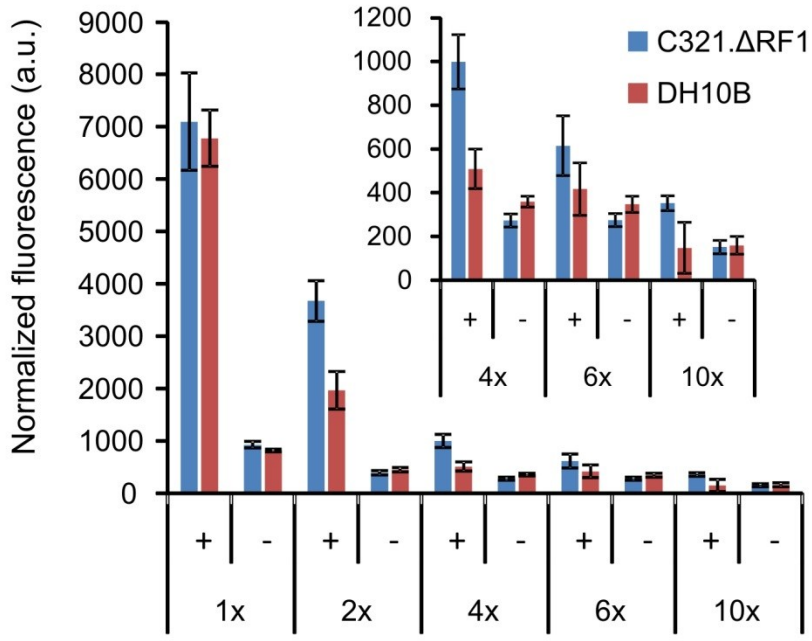

Figure S6: Expression of sfGFP reporters with increasing N-terminal UAG codons using a pUltra plasmid encoding an *M. jannaschii*-derived polyspecific tyrosyl tRNA/aaRS pair. O-propargyl-tyrosine (**5**, Fig 3C) was used as a substrate. Data for reporters with 4, 6 or 10 UAG codons are magnified at inset.

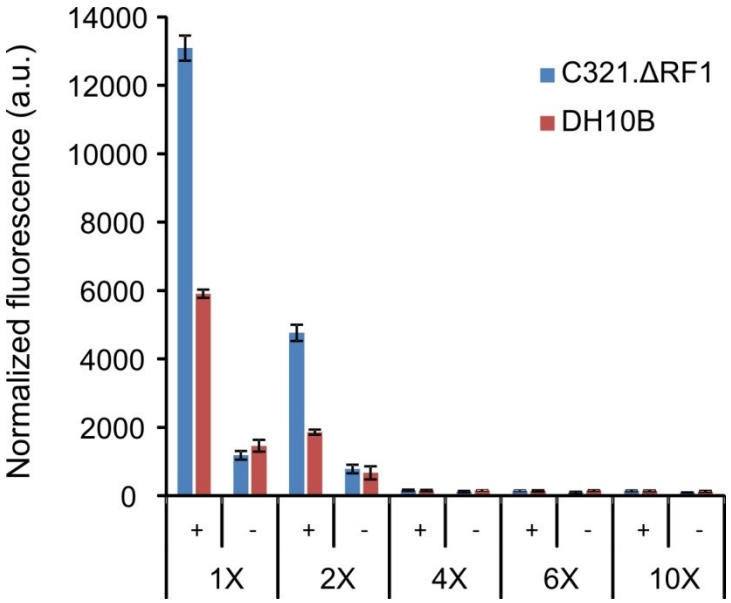

Figure S7: Expression of sfGFP reporters with increasing N-terminal UAG codons using a pEVOL plasmid encoding an *M. jannaschii*-derived O-sulfotyrosine (**7**, Fig 3C) specific tRNA/aaRS pair. O-sulfotyrosine was used as a substrate.

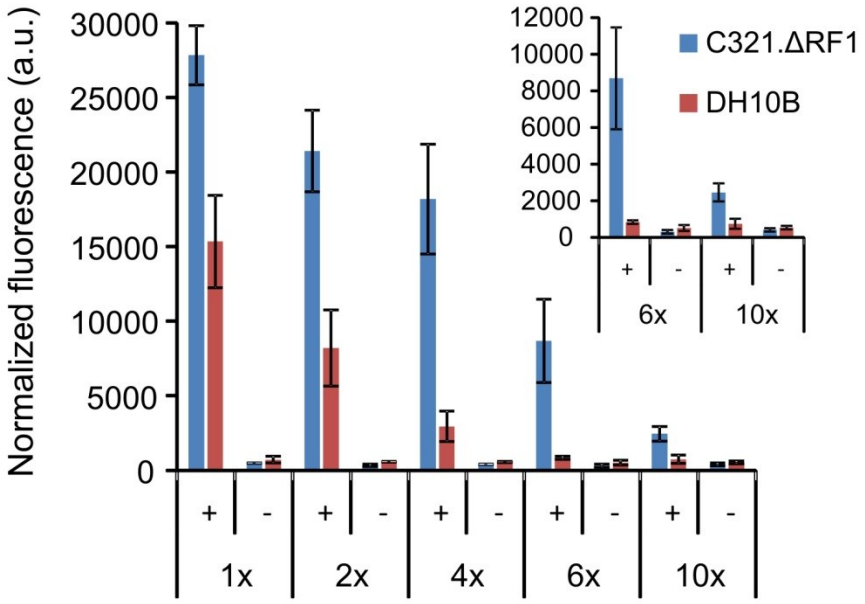

Figure S8: Expression of sfGFP reporters with increasing N-terminal UAG codons using a pEVOL plasmid encoding an *M. jannaschii*-derived *p*-borono-phenylalanine (**6**, Fig 3C) specific tRNA/aaRS pair. *p*-borono-phenylalanine was used as a substrate. Data for reporters with 6 or 10 UAG codons are magnified at inset.

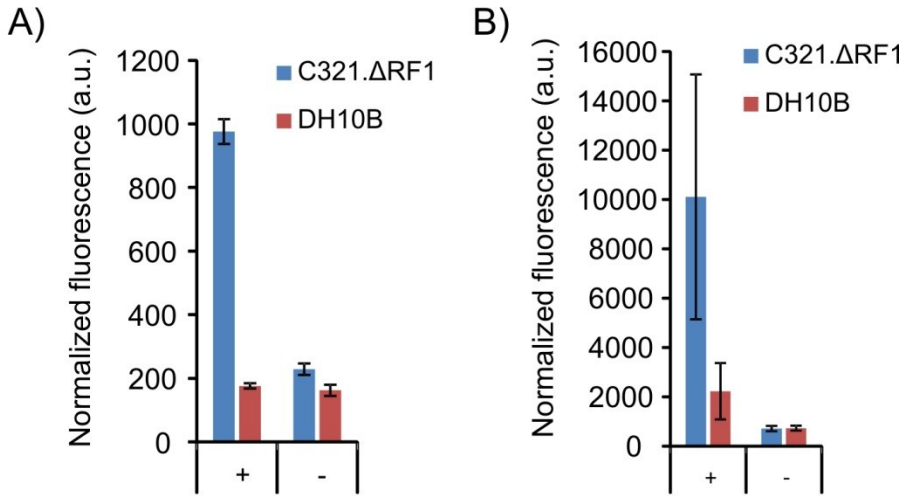

Figure S9: Expression of a sfGFP reporter harboring 8 UAG codons (replacing all surface exposed threonine residues) in strains with or without a functional RF1, using pEVOL suppressor plasmids harboring *M. jannaschii*-derived tRNA/aaRS pairs specific for A) O-sulfotyrosine (**7**, Fig 3C), or B) *p*-borono-phenylalanine (**6**, Fig 3C). Expression levels in the presence or the absence of the relevant ncAA in the media are indicated as "+" or "-", respectively.

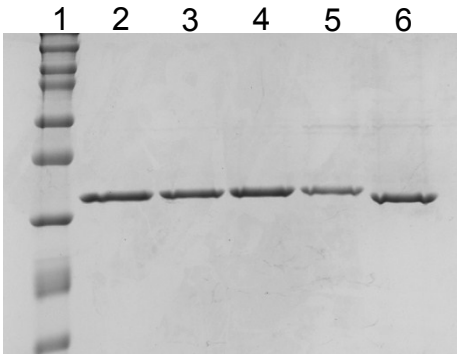

Figure S10: Representative SDS-PAGE analysis of purified reporter proteins followed by Coomassie-staining. All reporters here were expressed using the polyspecific *M. jannaschii*-derived tyrosyl-tRNA/aaRS pair (encoded in a pUltra vector) in the presence of 1 mM O-methyl-tyrosine (1, Figure 3C). Lanes 1: molecular weight marker; 2: $(UAG)_1$-sfGFP; 3: contiguous-$(UAG)_2$-sfGFP; 4: contiguous-$(UAG)_4$-sfGFP; 5: contiguous-$(UAG)_6$-sfGFP; 6: scattered-$(UAG)_8$-sfGFP. The last reporter migrates slightly lower than the rest, even though its MS analysis revealed the correct mass (Table S1).

Table S1: ESI-MS analysis of reporter proteins:

Reporter	ncAA	expected mass	observed mass	note
1x-contiguous-sfGFP*	OMeY, 1	27708	27708	Reporter expressed using the pUltra suppression vector
2x-contiguous-sfGFP*	OMeY, 1	27885	27885	Reporter expressed using the pUltra suppression vector
4x-contiguous-sfGFP	OMeY, 1	28376	28376	Reporter expressed using the pUltra suppression vector
6x-contiguous-sfGFP	OMeY, 1	28731	28731	Reporter expressed using the pUltra suppression vector
1x-contiguous-sfGFP*	O-SY, 7	27774	27774	Reporter expressed using the pEVOL suppression vector
2x-contiguous-sfGFP*	O-SY, 7	28017	28017	Reporter expressed using the pEVOL suppression vector
1x-contiguous-sfGFP*	pBoF, 6	27721	27721	Reporters with higher numbers of pBoF show complex patterns due to oxidation, dehydration, etc., of the pBoF side chain
8x-scattered-sfGFP	OMeY, 1	28206	28206	Reporter expressed using the pUltra suppression vector
8x-scattered-sfGFP	pIF, 2	28972	28972	Reporter expressed using the pUltra suppression vector
8x-scattered-sfGFP	pAzF, 3	28293	28293, 28267, 28241, 28215	The additional peaks correspond to the reduction product of one, two or three pAzF side chains. Reporter expressed using the pUltra suppression vector
8x-scattered-sfGFP	pAcF, 4	28302	28301	Reporter expressed using the pUltra suppression vector
8x-scattered-sfGFP	OPrY, 5	28398	28398	Reporter expressed using the pUltra suppression vector
8x-scattered-sfGFP	O-SY, 7	28733	28733, 28654, 28574	The additional peaks correspond to desulfation product at one or two O-SY residues; Reporter expressed using the pEVOL suppression vector

* These reporters harbor a penta-histidine tag, while others harbor a hexa-histidine tag at the C-terminus.

References:

1. A. Deiters, T. A. Cropp, M. Mukherji, J. W. Chin, J. C. Anderson, P. G. Schultz, *J Am Chem Soc*, 2003, **125**, 11782-3
2. A. Chatterjee, M. J. Lajoie, H. Xiao, G. M. Church and P. G. Schultz, *Chembiochem*, 2014, **15**, 1782-1786.