nature chemical biology

SUPPLEMENTARY INFORMATION

https://doi.org/10.1038/s41589-020-0507-z

In the format provided by the authors and unedited.

New codons for efficient production of unnatural proteins in a semisynthetic organism

Emil C. Fischer[1], Koji Hashimoto[1], Yorke Zhang[1,2], Aaron W. Feldman[1], Vivian T. Dien[1], Rebekah J. Karadeema[1], Ramkrishna Adhikary[1], Michael P. Ledbetter[1], Ramanarayanan Krishnamurthy[1] and Floyd E. Romesberg[1,2] ✉

[1]Department of Chemistry, The Scripps Research Institute, La Jolla, CA, USA. [2]Synthorx, a Sanofi Company, La Jolla, CA, USA. ✉e-mail: fromesberg@synthorx.com

Supplementary Information

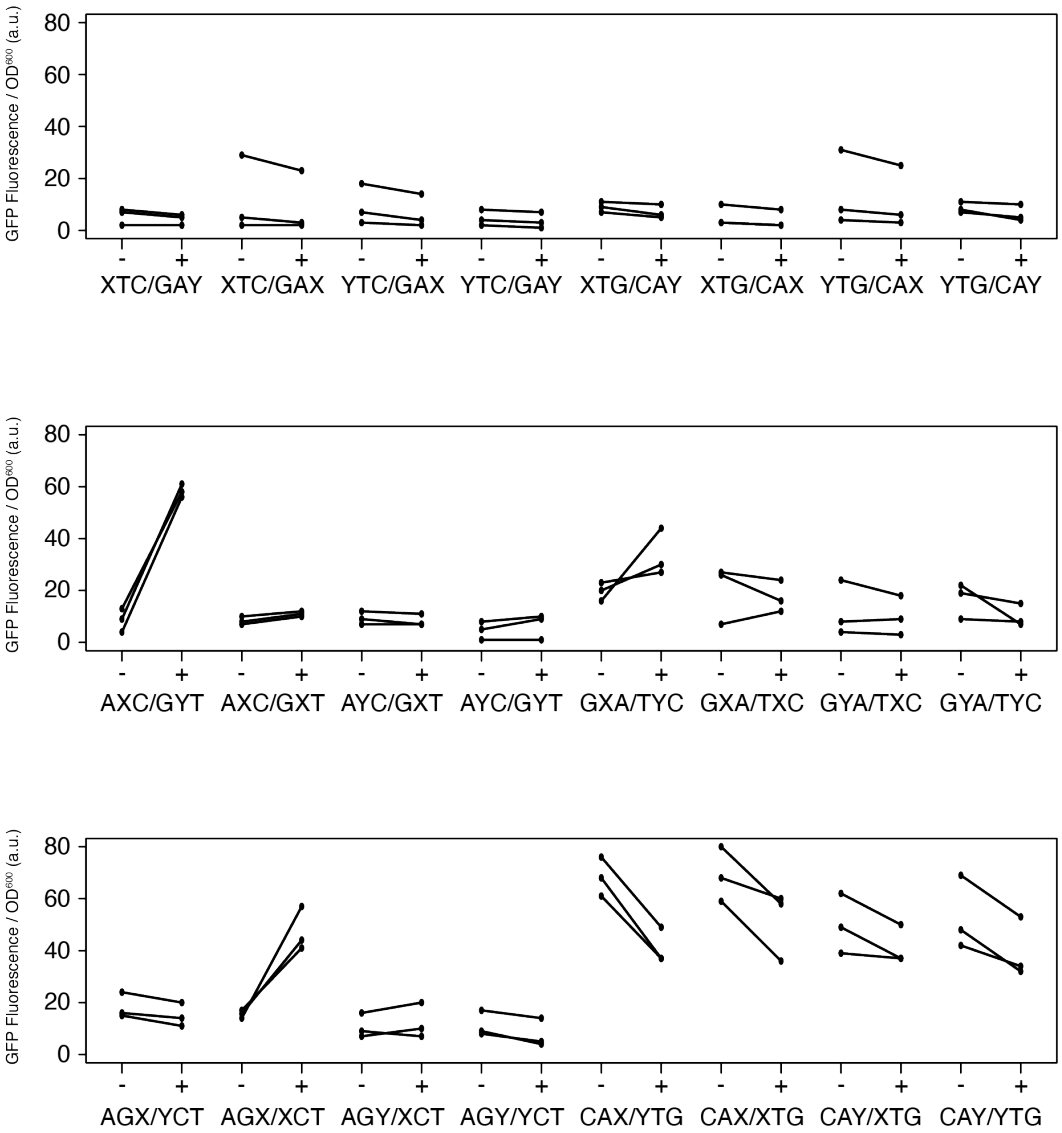
New codons for efficient production of unnatural proteins in a semi-synthetic organism

Emil C. Fischer, Koji Hashimoto, Yorke Zhang, Aaron W. Feldman, Vivian T. Dien, Rebekah J. Karadeema, Ramkrishna Adhikary, Michael P. Ledbetter, Ramanarayanan Krishnamurthy, Floyd E. Romesberg *

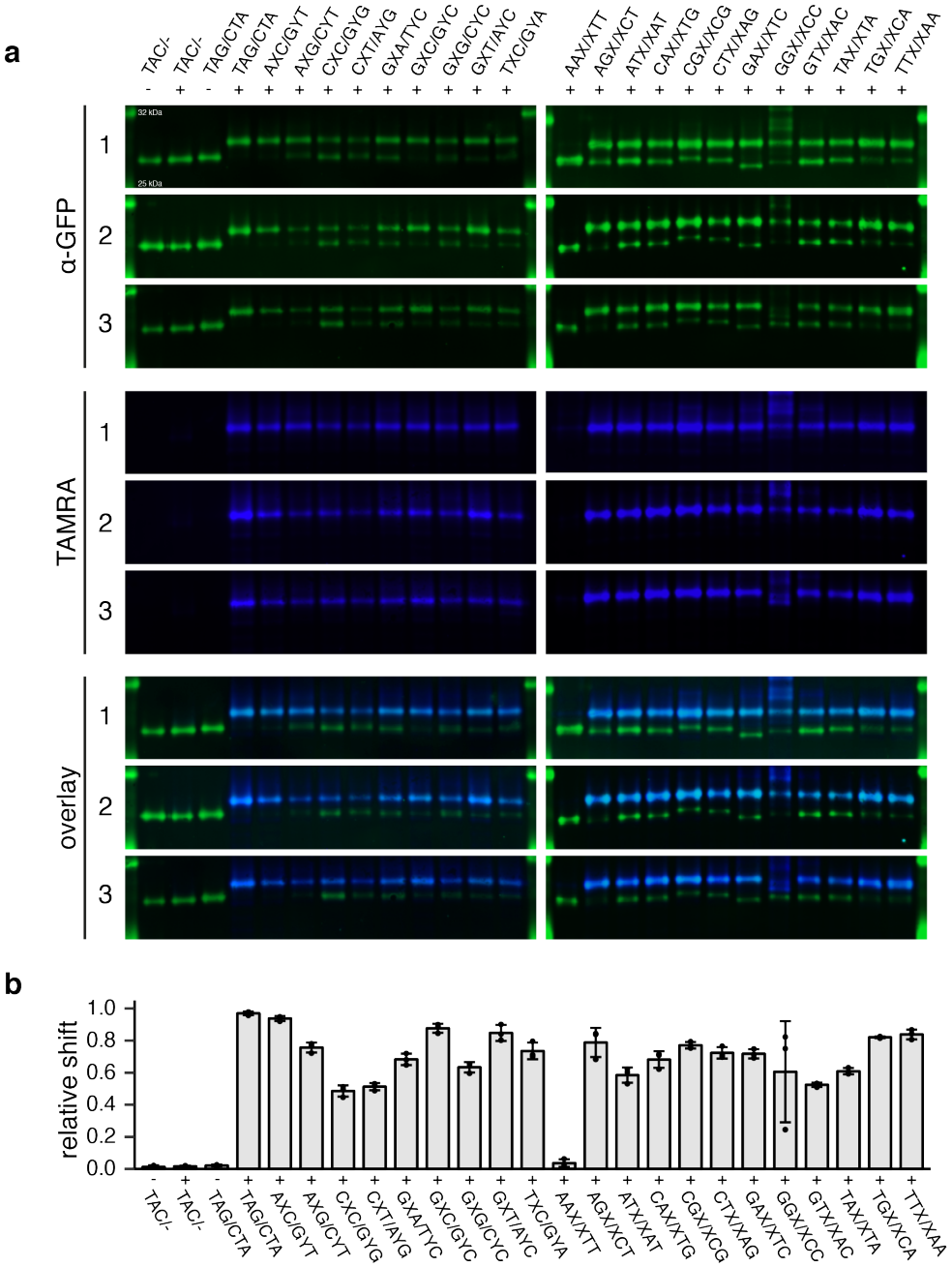
Department of Chemistry, The Scripps Research Institute, La Jolla, CA 92037 USA

Correspondence: fromesberg@synthorx.com

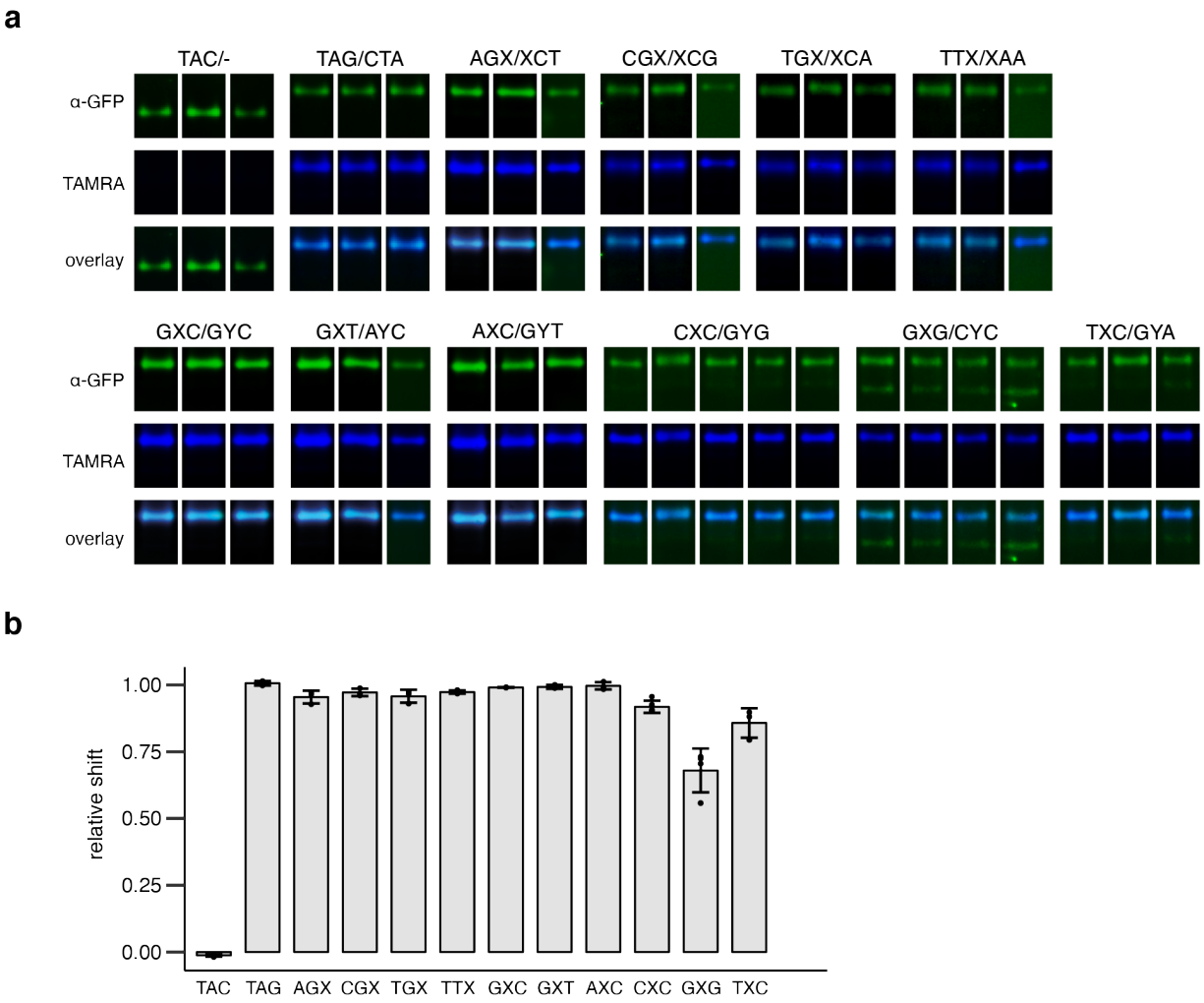
This PDF includes:



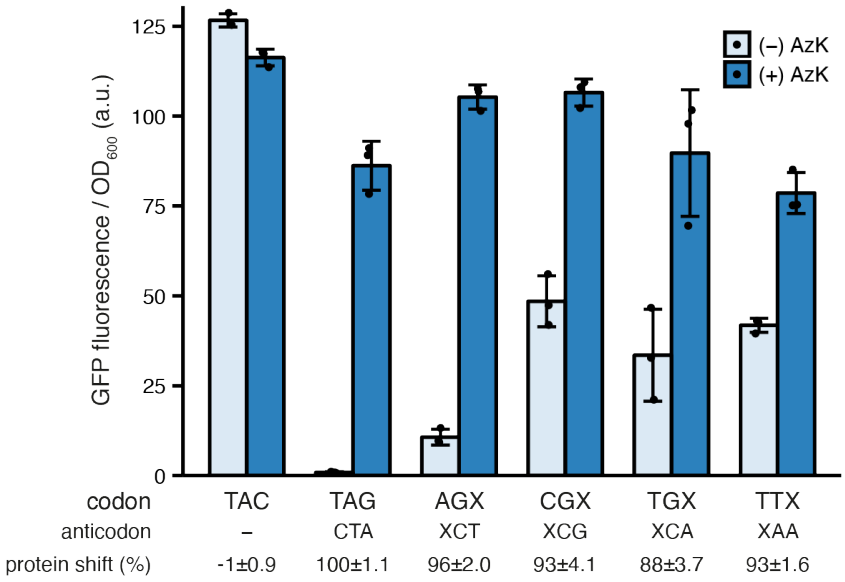
Supplementary Fig. 1 | Initial screen of unnatural codons in nonclonal SSOs. Paired strip charts of normalized fluorescence from SSO cells at the endpoint of protein expression (*i.e. t* = 180 min after aTc was supplemented) for select codon-anticodon pairs carrying the UBP in either first, second, or third position of the codon. Plus/minus denotes the addition of 20 mM AzK to the media. Each replicate derives from a different batch of competent SSO starter cells (*n* = *3*, biological replicates).



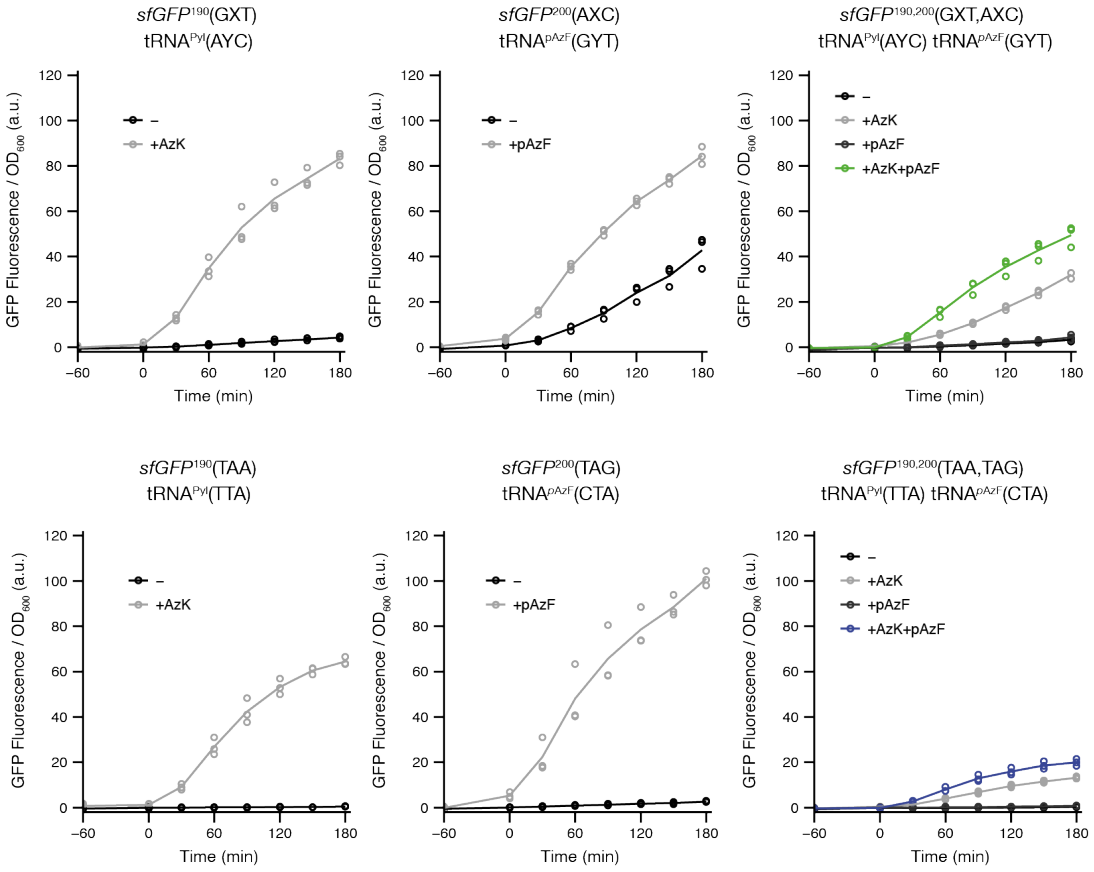
Supplementary Fig. 2 | Western blots and fluorescence scans for nonclonal SSO expression. **a**, Pseudocolored western blots (green) and TAMRA fluorescence scans (blue) of purified sfGFP from cultures in Fig. 1d with conjugation to TAMRA-PEG$_4$-DBCO by SPAAC. Plus/minus sign denotes if SPAAC was carried out. Three trials carried out (denoted 1, 2, 3; biological replicates). The three trial of each set (NXN/NYN and NNX/XNN) were processed in parallel. Uncropped fluorescence scans are shown in Supplementary Fig. 9. **b**, Quantifications of relative shift in western blots (in **a**) for specified codon-anticodon pairs (*i.e.* signal of the shifted band divided by the total signal of both shifted and unshifted bands). Plus/minus sign denotes if SPAAC was carried out. Mean ± standard deviation as well as individual data points shown ($n = 3$ biologically independent samples; *i.e.* biological replicates).



Supplementary Fig. 3 | Western blots and fluorescence scans for clonal SSO expression. a, Pseudocolored western blots (green) and TAMRA fluorescence scans (blue) of purified sfGFP from cultures in Fig. 2a with conjugation to TAMRA-PEG$_4$-DBCO by SPAAC. Displayed (cropped) area migrated in between 32 kDa and 25 kDa standard protein markers. Uncropped fluorescence scans are shown in Supplementary Fig. 10. **b**, Quantifications of relative shift in western blots (in **a**) for specified codons. Mean ± standard deviation as well as individual data points shown ($n = 3$ except $n_{CXC} = 5$ and $n_{GXG} = 4$; n biologically independent samples; *i.e.* biological replicates).

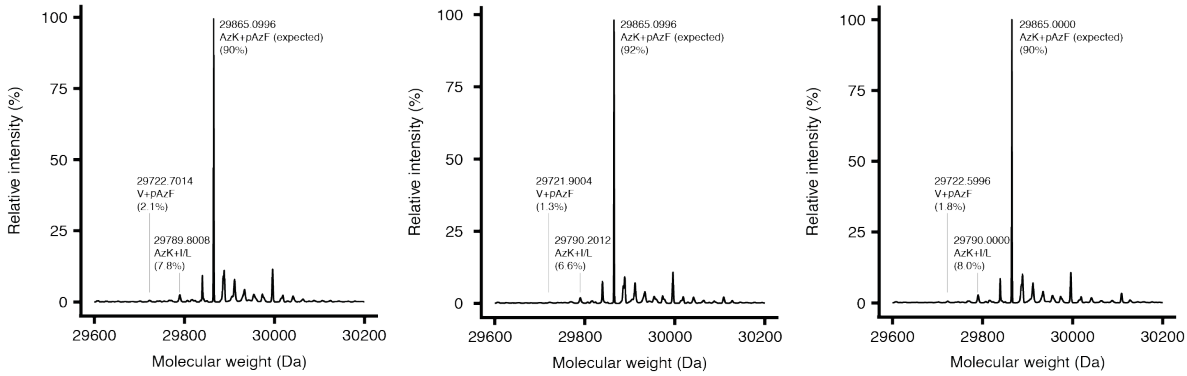


Supplementary Fig. 4 | Clonal SSO expressions in the absence of TPT3TP. Normalized fluorescence from clonal SSOs at the endpoint of protein expression (*i.e.* t = 180 min after aTc was supplemented) for the top four self-pairing codons/anticodons. Mean ± standard deviation shown for both fluorescence and quantified western blot protein shift (*i.e.* relative shift; gels not shown) as well as individual data points for fluorescence (n = *3* biologically independent samples; *i.e.* biological replicates).



Supplementary Fig. 5 | Controls for double codon expressions. Time-course plot of normalized fluorescence during *sfGFP* expressions of specified genotypes, with or without denoted ncAAs in the media. IPTG was added at $t = -60$ min and aTc was added at $t = 0$. Mean and individual data points shown ($n = 3$, biological replicates).

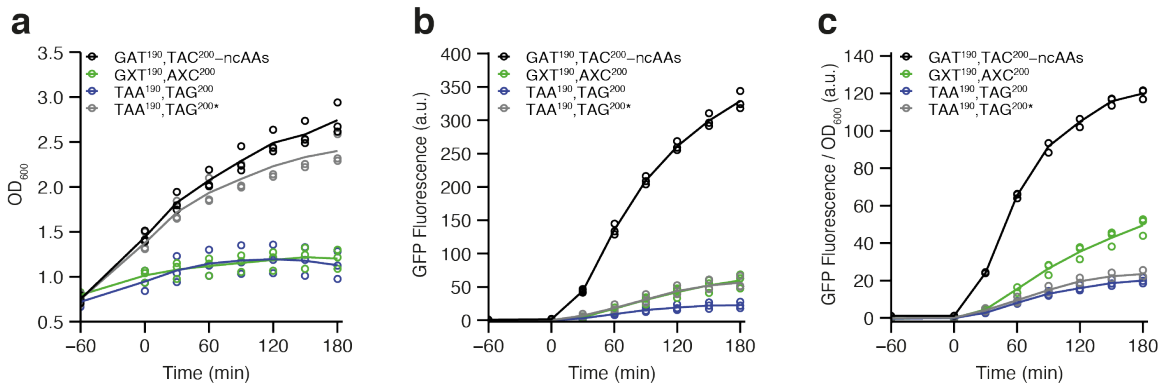
a



b

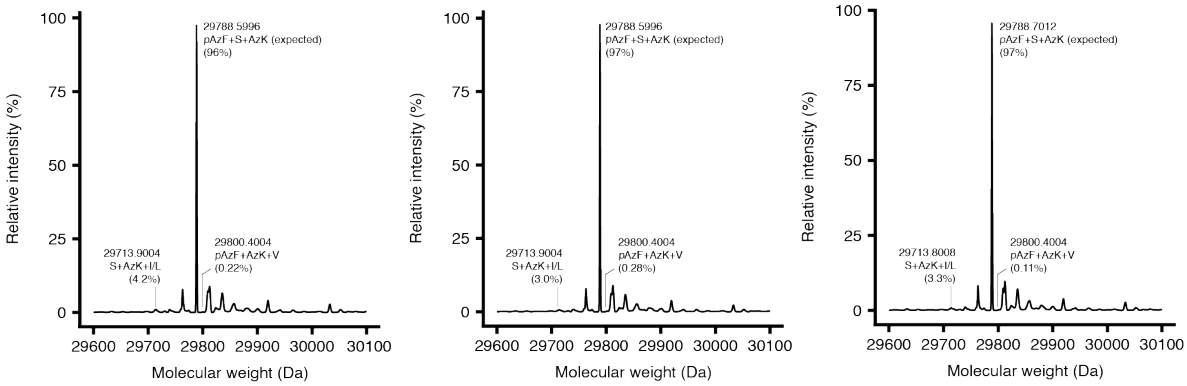
MW (Da)	ΔMW (rel. to reference; Da)	Assignment	Interpretation
29865.1	0.0	D190AzK Y200*p*AzF	expected product
29722.4	-142.7	D190V Y200pAzF	GXT→GTT
29790.0	-75.1	D190AzK Y200I/L	AXC→ATC
29839.4	-25.7	Oxidation of arginine	techninal
29888.6	23.5	Na adduct	techninal
29911.7	46.6	2×Na adduct	techninal
29933.5	68.4	3×Na adduct	techninal
29995.9	130.8	+Met	unprocessed N-terminal
30108.7	243.6	+Biotin	technical

Supplementary Fig. 6 | HRMS analysis of protein from double codon expression. HRMS analysis of intact sfGFP purified from SSOs expressing *sfGFP*^{190,200}(GXT,AXC), tRNA^{Pyl}(AYC), and tRNA^{pAzF}(GYT) with AzK and *p*AzF in the media, as shown in Fig. 3b (*n* = *3*, biological replicates). Standard single-letter amino acid code used. **a**, Deconvoluted spectra with annotation of relevant peaks and their relative abundance to each other. **b**, Peak assignment and interpretation.



Supplementary Fig. 7 | Growth and fluorescence for double codon expression compared to natural control. Time-course plot of OD_{600} (**a**), GFP fluorescence (**b**), and normalized fluorescence (**c**) during *sfGFP* expressions of specified genotypes with AzK and *p*AzF in the media and respective tRNAs (unless otherwise noted). Samples marked with asterisk (*) was carried without (d)NaMTP and (d)TPT3TP in the media. IPTG was added at $t = -60$ min and aTc was added at $t = 0$. Each replicate expression was carried out in cultures propagated from an individual colony ($n = 3$, biological replicates). Mean and individual data points shown.

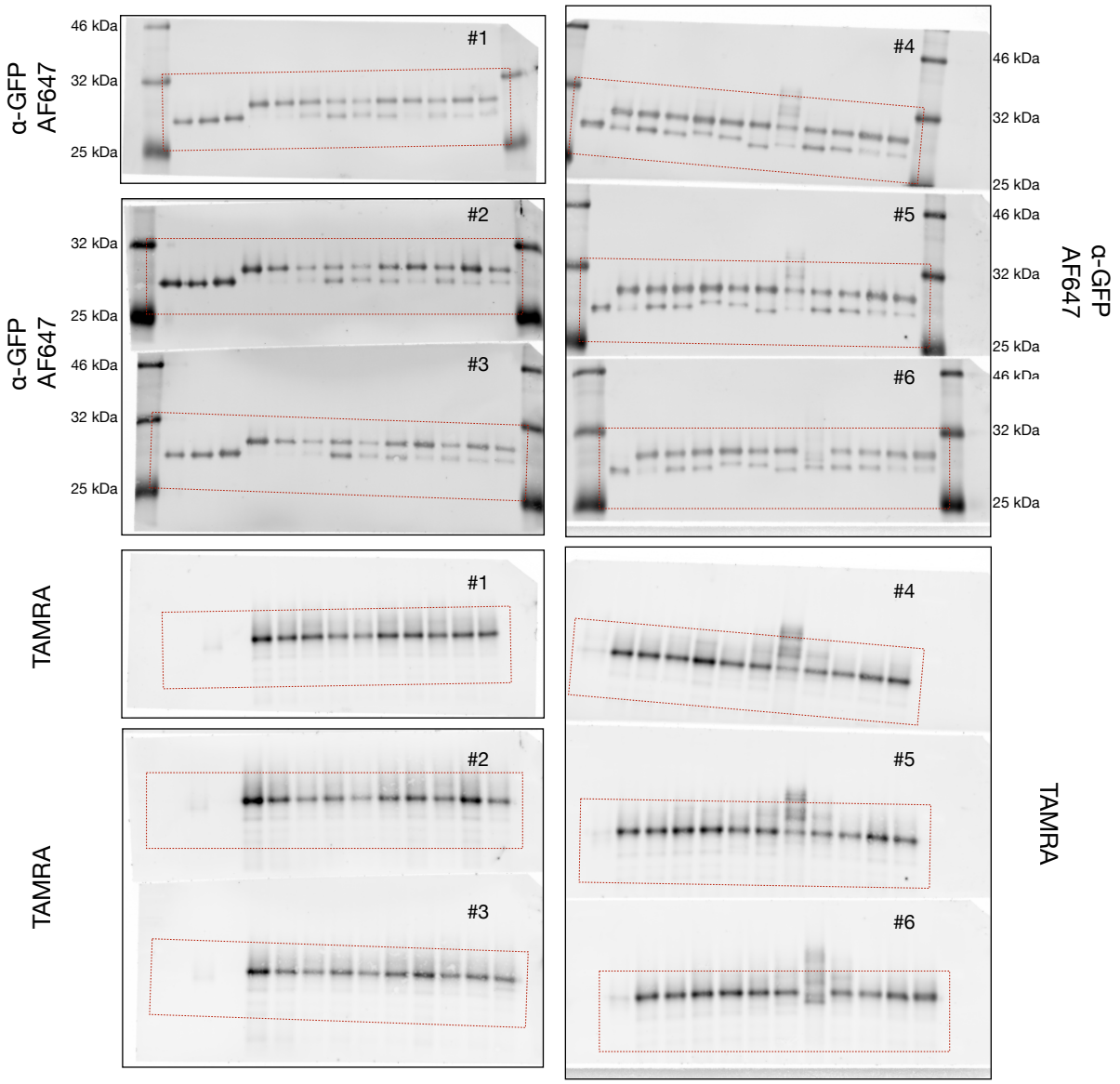
a



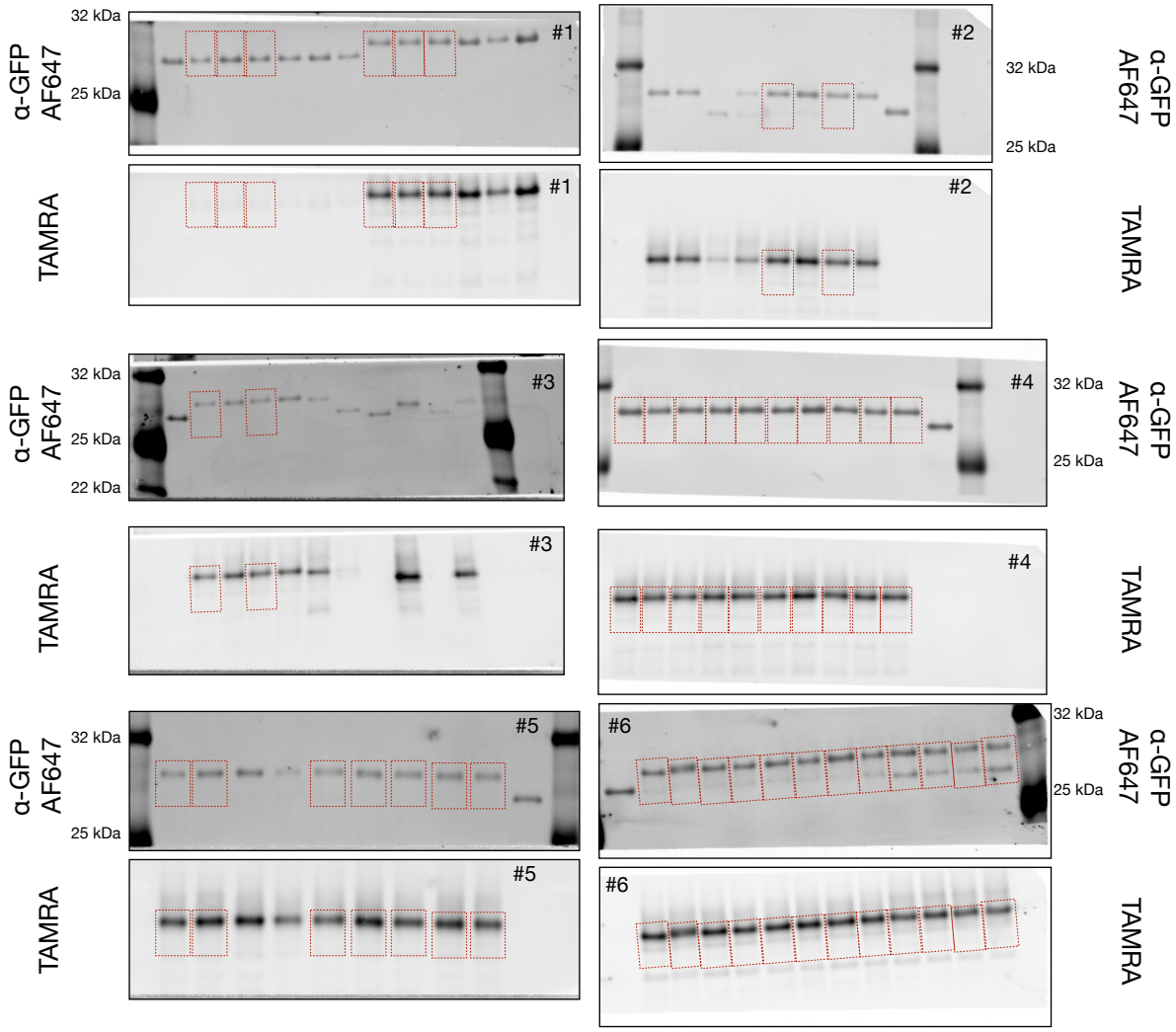
b

MW (Da)	ΔMW (rel. to reference; Da)	Assignment	Interpretation
29788.6	0.0	Y151pAzF D190S Y200AzK	expected product
29713.9	-74.7	Y151I/L D190S Y200AzK	AXC→ATC
29763.0	-25.6	Oxidation of arginine	techninal
29801.2	12.6	Y151pAzF D190V Y200AzK	GXT→GTT
29812.3	23.7	Na adduct	techninal
29835.6	47.0	2×Na adduct	techninal
29857.1	68.5	3×Na adduct	techninal
29919.4	130.8	+Met	unprocessed N-terminal
30033.0	244.4	+Biotin	technical

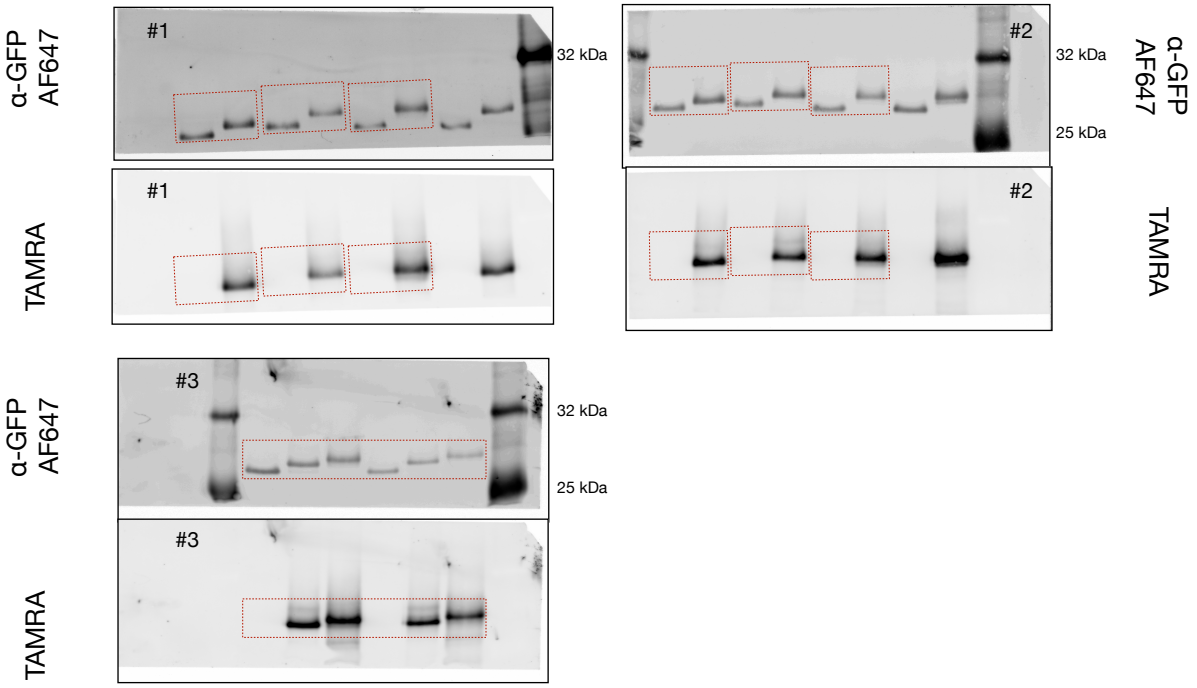
Supplementary Fig. 8 | HRMS analysis of protein from triple codon expression. HRMS analysis of intact sfGFP purified from SSOs expressing *sfGFP*^{151,190,200}(AXC,GXT,AGX), tRNA^{Pyl}(XCT), tRNA^{pAzF}(GYT), and tRNA^{Ser}(AYC) with AzK and *p*AzF in the media, as shown in Fig 4b (*n* = *3*, biological replicates). Standard single-letter amino acid code used. **a**, Deconvoluted spectra with annotation of relevant peaks and their relative abundance to each other. **b**, Peak assignment and interpretation.



Supplementary Fig. 9 | PVDF membranes for initial codon screen. Uncropped fluorescence scans without pseudocoloring for of six membranes presented in Supplementary Fig. 2a and represented in Fig. 1d. Approximate cropping for Supplementary Fig. 2a shown with red dashed line. Both scans for AlexaFlour 647 (α-GFP AF647) and TAMRA shown (see Online methods). Standard molecular weight markers indicated.



Supplementary Fig. 10 | PVDF membranes from clonal codon charaterization. Uncropped fluorescence scans without pseudocoloring for of six membranes presented in Supplementary Fig. 3a and represented in Fig. 2a. Approximate cropping for Supplementary Fig. 3a shown with red dashed line. Both scans for AlexaFlour 647 (α-GFP AF647) and TAMRA shown (see Online methods). Standard molecular weight markers indicated.



Supplementary Fig. 11 | PVDF membranes from simultaneous codon use. Uncropped fluorescence scans without pseudocoloring for of three membranes presented in Fig. 3d (membrane #1 and #2) and Fig. 3f (membrane #3). Approximate cropping shown with red dashed line. Both scans for AlexaFlour 647 (α-GFP AF647) and TAMRA shown (see Online methods). Standard molecular weight markers indicated.

Supplementary Table 1 | UBP retention in reported SSOs. Retention of UBPs based on the streptavidin-biotin shift assay (see Methods). Retention comprises relative shift (*i.e.* signal of shifted band divided by total signal of shifted and unshifted bands) normalized to relative shift of ssDNA template control, except for tRNA^{pAzF} and tRNA^{Ser} where no normalization could be done. Mean ± standard deviation shown (*n = 3* biologically independent samples; *i.e.* biological replicates).

Construct	Appears in	*n*	Codon(s)	UBP retention codon(s)	Anticodon(s)	UBP retention anticodon(s)
Single codon experiments						
sfGFP^{151} *M. mazei* tRNA^{Pyl}	Fig. 2a	3	AXC	94±3	GYT	92±4
sfGFP^{151} *M. mazei* tRNA^{Pyl}	Fig. 2a	3	GXC	94±3	GYC	96±5
sfGFP^{151} *M. mazei* tRNA^{Pyl}	Fig. 2a	3	GXT	99±1	AYC	99±1
sfGFP^{151} *M. mazei* tRNA^{Pyl}	Fig. 2a	3	AGX	89±3	XCT	61±18
sfGFP^{151} *M. mazei* tRNA^{Pyl}	Fig. 2a	3	CGX	89±3	XCG	83±8
sfGFP^{151} *M. mazei* tRNA^{Pyl}	Fig. 2a	3	TGX	91±2	XCA	78±13
sfGFP^{151} *M. mazei* tRNA^{Pyl}	Fig. 2a	3	TTX	95±3	XAA	76±37
sfGFP^{151} *M. mazei* tRNA^{Pyl}	Fig. 2a	5	CXC	67±8	GYG	91±4
sfGFP^{151} *M. mazei* tRNA^{Pyl}	Fig. 2a	4	GXG	58±2	CYC	60±10
sfGFP^{151} *M. mazei* tRNA^{Pyl}	Fig. 2a	3	TXC	87±6	GYA	94±11
sfGFP^{151} *M. mazei* tRNA^{Pyl}	Fig. 2a	3	AXT	97±3	AYT	95±1
sfGFP^{151} *M. mazei* tRNA^{Pyl}	Fig. 2b	3	AGX	91±1	AYC	101±1
sfGFP^{151} *M. mazei* tRNA^{Pyl}	Fig. 2b	3	AGX	92±1	GYT	99±6
sfGFP^{151} *M. mazei* tRNA^{Pyl}	Fig. 2b	3	AGX	82±3	XCT	100±4
sfGFP^{151} *M. mazei* tRNA^{Pyl}	Fig. 2b	3	AXC	96±3	AYC	99±2
sfGFP^{151} *M. mazei* tRNA^{Pyl}	Fig. 2b	3	AXC	98±1	GYT	94±8
sfGFP^{151} *M. mazei* tRNA^{Pyl}	Fig. 2b	3	AXC	99±2	XCT	84±12
sfGFP^{151} *M. mazei* tRNA^{Pyl}	Fig. 2b	3	GXT	99±4	AYC	97±2
sfGFP^{151} *M. mazei* tRNA^{Pyl}	Fig. 2b	3	GXT	100±1	GYT	100±1
sfGFP^{151} *M. mazei* tRNA^{Pyl}	Fig. 2b	3	GXT	99±1	XCT	101±1
Multicodon codons experiments (incl. controls)						
sfGFP^{190} *M. mazei* tRNA^{Pyl}	Fig. 3b	3	GXT	103±4	AYC	101±4
sfGFP^{200} *M. jannaschii* tRNA^{pAzF}	Fig. 3b	3	AXC	96±2	GYT	>94±1
sfGFP^{190,200} *M. mazei* tRNA^{Pyl} *M. jannaschii* tRNA^{pAzF}	Fig. 3b	3	GXT, AXC	98±3, 86±2	AYC, GYT	96±1, >88±1
sfGFP^{151,190,200} *M. mazei* tRNA^{Pyl} *M. jannaschii* tRNA^{pAzF} *E. coli* tRNA^{Ser}	Fig. 4b	3	AXC, GXT, AGX	92±1, 101±2, 96±3	XCT, GYT, AYC	93±3, >87±3, >94±2

Supplementary Table 2 | Protein yield of sfGFP expressions. All yields determined by sfGFP capture using excess Strep-Tactin XT beads during affinity purification. Yield normalized to final OD_{600} at t = 180 min of expression. Mean ± standard deviation shown (n = *3* biologically independent samples; *i.e.* biological replicates).

Construct	*n*	Codon(s)/anticodon(s)	Protein yield (µg/ml)	Norm. Protein yield (µg/ml/OD_{600})
sfGFP^{151} *M. mazei* $tRNA^{Pyl}$	3	TAC/−	66±13	23±1.9
sfGFP^{151} *M. mazei* $tRNA^{Pyl}$	3	TAG/CTA	52±11	18±3.0
sfGFP^{151} *M. mazei* $tRNA^{Pyl}$	3	AXC/GYT	28±6.3	19±2.1
sfGFP^{151} *M. mazei* $tRNA^{Pyl}$	3	GXC/GYC	31±0.32	18±2.9
sfGFP^{151} *M. mazei* $tRNA^{Pyl}$	3	GXT/AYC	29±3.3	21±0.22
sfGFP^{151} *M. mazei* $tRNA^{Pyl}$	3	AGX/XCT	34±4.7	19±1.7
sfGFP^{151} *M. mazei* $tRNA^{Pyl}$	3	CGX/XCG	29±2.8	19±5.2
sfGFP^{151} *M. mazei* $tRNA^{Pyl}$	3	TGX/XCA	27±3.2	18±4.8
sfGFP^{151} *M. mazei* $tRNA^{Pyl}$	3	TTX/XAA	27±4.1	19±4.6
sfGFP^{190,200} *M. mazei* $tRNA^{Pyl}$ *M. jannaschii* $tRNA^{pAzF}$	3	TAA,TAG/TTA,CTA	5.6±1.0	5.0±0.24
sfGFP^{190,200} *M. mazei* $tRNA^{Pyl}$ *M. jannaschii* $tRNA^{pAzF}$*	3	TAA,TAG/TTA,CTA	6.8±1.1	2.8±0.28
sfGFP^{190,200} *M. mazei* $tRNA^{Pyl}$ *M. jannaschii* $tRNA^{pAzF}$	3	GXT,AXC/AYC,GYT	16±3.2	13±1.6
sfGFP^{151,190,200} *M. mazei* $tRNA^{Pyl}$ *M. jannaschii* $tRNA^{pAzF}$ *E. coli* $tRNA^{Ser}$	3	AXC,GXT,AGX/XCT,GYT, AYC	12±1.9	7.8±1.1

* No (d)NaMTP or (d)TPT3TP added.

Supplementary Table 3 | Single-stranded DNA oligonucleotides. List of ssDNA oligonucleotides used as PCR primers in UBP PCR and streptavidin-biotin shift assay as well as UBP containing ssDNA PCR templates.

ID	Application	Sequence (5' to 3')
Primers for UBP PCR		
EFo309	sfGFP Y151 insert F	ATGGGTCTCACACAAACTCGAGTACAACTTTAACTCACAC
EFo310	sfGFP Y151 insert R	ATGGGTCTCGATTCCATTCTTTTGTTTGTCTGC
EFo296	sfGFP Y200 insert F	CATAATGGTCTCGCTGCTGCCCGATAACCAC
EFo297	sfGFP Y200 insert R	TGATATTGGTCTCGGTCTTTCGATAAAACACTCTGAGTAGAG
EFo311	*M. mazei* $tRNA^{Pyl}$ insert F	ATGGGTCTCGAAACCTGATCATGTAGATCGAACGG
EFo312	*M. mazei* $tRNA^{Pyl}$ insert R	ATGGGTCTCATCTAACCCGGCTGAACGG
EFo313	*M. jannaschii* $tRNA^{pAzF}$ insert F	ATGGGTCTCCGGTAGTTCAGCAGGGCAGAACG
EFo314	*M. jannaschii* $tRNA^{pAzF}$ insert R	ATGGGTCTCGGAGGGGATTTGAACCCCTGCCATG
EFo294	sfGFP D190 insert	ATATTCGGTCTCGTCAGCAGAATACGCCGATTGG
EFo295	sfGFP D190 insert	ACGCGTTGGTCTCGGTTATCGGGCAGCAGCACC
YZ401	*E. coli* $tRNA^{Ser}$ insert F	ATTGGTCTCGGCCGAGCGGTTGAAGGCAC
YZ403	*E. coli* $tRNA^{Ser}$ insert R	ATTGGTCTCTCTGGAACCCTTTCGGGTCG
Primers for streptavidin-biotin shift assay		
EFo251	Position Y151 F	CTCGAGTACAACTTTAACTCACAC
EFo252	Position Y151 R	GATTCCATTCTTTTGTTTGTCTGC
EFo294	Position D190 F	ATATTCGGTCTCGTCAGCAGAATACGCCGATTGG
EFo295	Position D190 R	ACGCGTTGGTCTCGGTTATCGGGCAGCAGCACC
EFo347	Position Y200 F	GCTGCTGCCCGATAACCAC
EFo348	Position Y200 R	GGTCTTTCGATAAAACACTCTGAGTAGAG
EFo343	*M. mazei* $tRNA^{Pyl}$ F	GAAACCTGATCATGTAGATCGAACGG
EFo344	*M. mazei* $tRNA^{Pyl}$ R	ATCTAACCCGGCTGAACGG
EFo313	*M. jannaschii* $tRNA^{pAzF}$ F	ATGGGTCTCCGGTAGTTCAGCAGGGCAGAACG
EFo305	*M. jannaschii* $tRNA^{pAzF}$ R	CCGCTGCCACTAGGAAGCTTATG
EFo119	*E. coli* $tRNA^{Ser}$ F	CCTCTAGAAAATCATTCCGGAAGTGTG
EFo162	*E. coli* $tRNA^{Ser}$ R	TAAACAAATAGGGGTTCCGCGCAC
Template for UBP PCR ([NNN] denotes any specified codon/anticodon triplet)		
GFP151_[NNN]	sfGFP Y151 insert	CTCGAGTACAACTTTAACTCACACAATGTA[NNN]ATCACGGCAGACAAACAAAAGAATGGAATC
GFP190_GXT	sfGFP D190 insert	CAGCAGAATACGCCGATTGGCGXTGGCCCGGTGCTGCTGCCCGATAACC
GFP200_AXC	sfGFP Y200 insert	GCTGCTGCCCGATAACCACAXCCTCTCTACTCAGAGTGTTTTATCGAAAGACC
GFP200_opt_AGX	sfGFP Y200 insert	GCTGCCCGATAACCACAGXTTGTCTACTCAGAGTGTTTTATCG
tRNA_Pyl_[NNN]	*M. mazei* $tRNA^{Pyl}$ insert	GAATCTAACCCGGCTGAACGGATT[NNN]AGTCCGTTCGATCTACATGATCAGG
tRNA_Mj_GYT	*M. mazei* $tRNA^{Pyl}$ insert	GATTTGAACCCCTGCCATGCGGATTAXCAGTCCGCCGTTCTGCCCTGCTGAA
tRNA_Eser_AYC	*E. coli* $tRNA^{Ser}$ insert	CTCTGGAACCCTTTCGGGTCGCCGGTTTGXTAGACCGGTGCCTTCAACCGCTCGGC

Supplementary Table 4 | Plasmids used in this study. All *sfGFP* reading frames are controlled by $P_{T7\text{-}tetO}$ and all tRNAs are controlled by $P_{T7\text{-}lacO}$. Backbone pSYN contain: ori(p15A) *bleoR*. Backbone pGEX contain: ori(pBR322) *ampR*. Golden gate destination sites (dest) are composed of recognition sequences BsaI-KpnI-BsaI. Genbank accession numbers are given in parentheses. Previous work refers to Zhang et al[1]. GB2sequin[2]was used for Genbank submission.

Backbone	Source	Application	Relevant properties
Superfolder GFP Expression plasmids			
pSYN	Previous work	Natural expression plasmid	*sfGFP*^{151}(TAG); *M. mazei* tRNA^{Pyl}(CTA)
pSYN	Previous work	Natural expression plasmid	*sfGFP*^{151}(TAC)
pSYN	MN882182	Natural expression plasmid	*sfGFP*^{190}(TAA) *M. mazei* tRNA^{Pyl}(TTA); opal stop codon
pSYN	MN882183	Natural expression plasmid	*sfGFP*^{200}(TAG) *M. jannaschii* tRNA^{pAzF}(CTA)
pSYN	MN882184	Natural expression plasmid	*sfGFP*^{190,200}(TAA,TAG) *M. mazei* tRNA^{Pyl}(TTA) *M. jannaschii* tRNA^{pAzF}(CTA); opal stop codon
pSYN	Previous work	UBP destination plasmid	*sfGFP*^{151}(dest) *M. mazei* tRNA^{Pyl}(dest)
pSYN	MN882185	UBP destination plasmid	*sfGFP*^{190}(dest) *M. mazei* tRNA^{Pyl}(dest)
pSYN	MN882186	UBP destination plasmid	*sfGFP*^{200}(dest) *M. jannaschii* tRNA^{pAzF}(dest)
pSYN	MN882187	UBP destination plasmid	*sfGFP*$^{190\text{-}200}$(dest) *M. mazei* tRNA^{Pyl}(dest) *M. jannaschii* tRNA^{pAzF}(dest)
pSYN	MN882188	UBP destination plasmid	*sfGFP*$^{151,190\text{-}200}$(dest, dest) *M. mazei* tRNA^{Pyl}(dest) *M. jannaschii* tRNA^{pAzF}(dest) *E. coli* tRNA^{Ser}(dest)
Accessory plasmids			
pGEX	MN882189	Accessory plasmid	P_{AmpR}-*tetR* P_{lacIq}-*lacI* $P_{tac\text{-}lacO}$-*chPylRS*^{IPYE}
pGEX	MN882190	Accessory plasmid	P_{AmpR}-*tetR* P_{lacIq}-*lacI* $P_{lacUV5\text{-}lacO}$-*MjpAzFRS* $P_{tac\text{-}lacO}$-*chPylRS*^{IPYE}

Supplementary References

1. Zhang, Y. et al. A semi-synthetic organism that stores and retrieves increased genetic information. *Nature* **551**, 644–647 (2017).

2 Lehwark, P. & Greiner, S. GB2sequin - A file converter preparing custom GenBank files for database submission. *Genomics* **111**, 759-761 (2019).