RF1 Knockout Allows Ribosomal Incorporation of Unnatural Amino Acids at Multiple Sites

David B.F. Johnson,[1,†] Jianfeng Xu,[1,†] Zhouxin Shen,[2] Jeffrey K. Takimoto,[1] Matthew D. Schultz,[3,4] Robert Schmitz,[3] Joseph R. Ecker,[3] Steven P. Briggs,[2] and Lei Wang[1,*]

Supplementary Methods

Strain construction

JX2.0 and JX3.0 were constructed as follows: first, a mutagenesis cassette was generated using overlapping PCRs. This cassette contained a mutated form of *prfB*^{f}, a chloramphenicol resistance (Cm^R) cassette flanked by two I-SceI cut sites, and homologous regions on both the 5' and 3' end to facilitate recombination. The mutant *prfB*^{f} had the in-frame premature TGA element removed[1], a Shine-Dalgarno like sequence mutated to a *Sac* II site, and the A246T mutation reverted back to alanine[2]. This cassette was electroporated into MDS42 cells harboring the pKD46 plasmid[3]. Cm^R colonies were screened using PCR to verify a correct knock-in, and then by *Sac* II digestion to verify that the mutant *prfB*^{f} was present. The resultant strain was transformed with the plasmid pATBSR, a derivative of pACBSR that has a tetracycline resistance (Tet) cassette in place of the original Cm^R gene[4]. Following induction with arabinose, cells were screened for removal of the Cm^R cassette using PCR and sequencing verification. Curing of the pATBSR plasmid resulted in the final JX2.0 strain.

JX3.0 was created using JX2.0 as the parental strain. JX2.0 cells harboring the pKD46 plasmid were electroporated with a PCR cassette to knock out the endogenous *prfA* gene. This

cassette contained a Cm^R gene flanked by 5' and 3' homologous overhangs to facilitate recombination. Cm^R colonies were again screened by PCR and sequencing verification. The resultant strain, JX3.0, contains an exact replacement of *prfA* with the Cm^R gene. A Tet^R derivative of JX33 strain was constructed from JX33 by replacing the Cm^R cassette with a Tet^R cassette, and was used to express histone H3a.

JX2.0 and JX33 derivatives containing an N-terminal FLAG-tag in the *yfiA* and *sufA* genes were created as follows: A PCR cassette was synthesized using overlapping PCRs to yield a construct containing a 5' homologous region followed by an I-*Sce*I flanked kanamycin resistance (Kan^R) cassette and an N-terminal FLAG-tag appended onto the target gene, which itself serves as the 3' overhang. In addition, immediately 5' of the Kan^R cassette is 75 bp of DNA that is perfectly homologous to 75 bp on the 3' of the Kan^R cassette. This repeat element will help to leave a scarless insertion of the N-terminal FLAG tag upon excision of the Kan^R cassette. These constructs were electroporated into JX2.0 or JX33 cells containing the pKD46 plasmid. Kanamycin resistant clones were screened for insertions using PCR and sequence verified for FLAG-tag insertion. Resultant strains were transformed with pATBSR, induced with arabinose, and screened for removal of the Kan^R cassette. Sequence verified clones were then used for further studies.

To construct an MRA8 derivative harboring a *prfB*^{f}*(A293E)* gene identical to that in JX33, a construct was synthesized using overlapping PCRs harboring a full-length copy of *prfB*^{f}*(A293E)* from JX33 with a Kan^R cassette on the 3' end. This construct was electroporated into MRA8 cells harboring the pKD46 plasmid, and kanamycin resistant clones were screened for insertion using PCR. Resultant strains were sequence verified and used for further analyses.

Western preparation and analyses

E. coli cells containing EGFP expression plasmids were grown at 37 ^{o}C for 16 hours, harvested, washed 2 times with PBS and diluted to an OD_{600} of 0.1 in PBS. One milliliter of cells was collected and resuspended in 100 µL Blue Juice (Qbiogene, Carlsbad, CA) and incubated for 10 minutes at 95 ^{o}C. Samples were separated by SDS-PAGE, transferred and probed with a penta-His antibody (Qiagen, Valencia, CA). *E. coli* cells containing Histone H3a were grown in large scale for Western analysis and prepared as described below.

For Western analysis of YfiA, a modified version of an established protocol was used[5]. Briefly, one liter of *E. coli* cells harboring an N-terminal FLAG tagged *yfiA* gene were grown for 16 hours at 37 ^{o}C. Cells were cold-shocked in ice-water for ten minutes followed by two hours of growth at 15 ^{o}C. Cells were harvested by centrifugation and frozen at -80 ^{o}C. SufA purification was also accomplished via an established procedure[6]. For *E. coli* cells harboring an N-terminal FLAG tagged *sufA* gene, a 50 mL culture was grown for 16 hours at 37 ^{o}C. Cells were diluted to an OD_{600} of 0.02 in one liter of fresh media. Once the OD_{600} reached 0.2, phenazine metholsulfate (Sigma, St. Louis, MO) was added to a final concentration of 0.1 mM. Cells were harvested by centrifugation and pellets were frozen at -80 ^{o}C after 90 minutes of growth at 37 ^{o}C. Protein from both cell types was extracted using BPER reagent (Thermo Scientific, Rockford, IL) and then applied to an Anti-FLAG agarose column (Sigma) to remove the vast majority of contaminating protein. Purified protein was then visualized using Western blotting with the monoclonal FLAG M2 antibody (Sigma).

Protein purification

For EGFP and GST preparations, 100 mL cultures were gown for 16 hours with or

without unnatural amino acid. Cells were pelleted and lysed for 10 minutes at room temperature in the appropriate volume of BPER (Thermo Scientific, Rockford, IL). Cell lysate was collected after centrifugation at 12,000 x g for 5 minutes. Ni-NTA resin (Qiagen, Valencia, CA) was added directly to the lysate and shaken for 1 hour at 4°C. The lysate was added to a column, washed with 10 column volumes of wash buffer (50 mM NaH_2PO_4, 300 mM NaCl, 10 mM imidazole, pH 7.5) and then eluted in six 1 mL fractions of elution buffer (Same as wash, but with 100 mM imidazole). Purified EGFP was buffer exchanged to 20 mM Tris buffer (pH 8.0) using Microcon Ultracel YM-10 spin columns (Millipore, Billerica, MA).

For histone H3a preparations, *E. coli* colonies transformed with plasmids pTak-tRNA-H3 and pBKt-ActKRS were picked and grown 16 hours. Cells were diluted 1:100 into fresh media containing 5 mM ActK or 1 mM *p*ActF. For all ActK preparations, nicotinamide was added to a final concentration of 5 mM to minimize deacetylation. When the OD_{600} reached 0.5, cells were induced with 0.4 mM of IPTG and grown for 4 hours at 37 °C. Cells were pelleted and frozen overnight to facilitate lysis. Pellets were thawed for ten minutes in a water bath, and then processed as described[7]. The final cell lysate was collected and applied to a Ni-NTA column pre-equilibrated with wash buffer (6 M guanidine HCl in PBS pH 7.6, 25 mM imidazole). Lysate was applied to the column 2 times, followed by 20 times the column volume of wash buffer. Fractions containing Histone H3a were eluted with elution buffer (wash buffer plus 250 mM imidazole) and analyzed by SDS-PAGE and Western blot.

Mass spectrometry

Intact protein analysis by ESI-MS: Purified *p*ActF-containing EGFP proteins were dissolved in 1% formic acid and infused into a LTQ Velos mass spectrometer at 1 μL/min by a

syringe pump. MS scans were collected for 1 minute. About 1,600 MS spectra were collected for each sample. Spectra were averaged and the charge states were de-convoluted using a freeware MagTran[8].

Tandem MS analysis: Purified *p*ActF-containing EGFP, SufA and YfiA proteins were solubilized in 50 mM Hepes (pH 7.2). The proteins were reduced and alkylated using 1 mM Tris(2-carboxyethyl)phosphine (Fisher, AC36383) at 95 °C for 5 minutes and 2.5 mM iodoacetamide (Fisher, AC12227) at 37 °C in dark for 30 minutes, respectively. *p*ActF-containing EGFP was digested with 1:50 chymotrypsin (Roche, 11418467001). SufA was digested with 1:50 trypsin (Roche, 03708969001) and YfiA was digested by both trypsin and Lys-C (Roche, 11420429001) at 37°C overnight. Automated 2D nanoflow LC-MS/MS analysis was performed using LTQ tandem mass spectrometer (Thermo Electron Corporation, San Jose, CA) employing automated data-dependent acquisition. The detailed LC-MS/MS method can be found in our published work[9,10]. Briefly, the peptides were fractionated by the on-line SCX column using a series of 7 salt gradients (10 mM, 20 mM, 30 mM, 50 mM, 70 mM, 100 mM, and 1M ammonium acetate for 20 minutes), followed by high resolution reverse phase separation using an acetonitrile gradient of 0 to 80% for 120 minutes.

The full MS scan range of 400-2000 m/z was divided into 3 smaller scan ranges (400-800, 800-1050, 1050-2000) to improve the dynamic range. Both CID (Collision Induced Dissociation) and PQD (Pulsed-Q Dissociation) scans of the same parent ion were collected for protein identification and quantitation. Each MS scan was followed by 4 pairs of CID-PQD MS/MS scans of the most intense ions from the parent MS scan. A dynamic exclusion of 1 minute was used.

The raw data was extracted and searched using Spectrum Mill (Agilent, version A.03.02). The CID and PQD scans from the same parent ion were merged together. MS/MS spectra with a sequence tag length of 1 or less were considered as poor spectra and discarded. The rest of the MS/MS spectra were searched against the NCBI (National Center for Biotechnology Information) RefSeq protein database (version 21, January 2007) limited to *E. coli* (16,324 sequences) plus the SufA and YfiA protein sequences with extended C-terminal sequences, as well as EGFP protein sequence. The enzyme parameter was limited to full chymotrypsin, tryptic or Lys-C peptides with a maximum miscleavage of 1. All other search parameters were set to SpectrumMill's default settings (carbamidomethylation of cysteines, +/- 2.5 Da for precursor ions, +/- 0.7 Da for fragment ions, and a minimum matched peak intensity of 50%). A variable modification of Gln to *p*ActF (+61 Da) was used for *p*ActF-containing EGFP database search. MS/MS spectra were validated using the following filtering criteria for database search results: cutoff score > 9 for 1+ peptide, > 9 for 2+ peptide, and > 12 for 3+ peptide.

Genomic sequencing of *E. coli* strains

Genomic DNA from JX2.0, JX31 and JX33 was harvested and purified using a Qiagen DNeasy kit. One μg of genomic DNA was used to prepare DNA libraries for sequencing. Genomic DNA was fractionated using the Covaris S2 System (Applied Biosystems, Foster City, CA) using the following parameters: cycle number = 6, duty cycle = 20%, intensity = 5, cycles/burst = 200 and time = 60 seconds. Fractionated DNA was purified using a Qiagen PCR minielute purification kit. Libraries were prepared using the NEBNext DNA Sample Prep Reagent Set 1 (New England Biolabs), following recommended protocols.

Genomic DNA libraries were sequenced using the Illumina Genome Analyzer II (Illumina, San Diego, CA) as per manufacturer's instructions. Sequencing of genomic DNA libraries was performed up to 82 cycles. Image analysis and base calling were performed with the standard Illumina pipeline (Firecrest v1.3.4 and Bustard v.1.3.4).

Sequence alignments and SNP analysis were performed using the SHORE package[11] according to the documentation provided with the software. In brief, the *E. coli* K-12 MG1655 reference genome was preprocessed into a SHORE acceptable format. Next, FASTQ files for each sample were converted to a SHORE flat file format. Reads were mapped using Genomemapper contained within the SHORE package using the following parameters –n 4, -g 3. Capitalizing on the large amount of coverage for this experiment we identified large deletions. We posited that any region of the reference genome where reads did not map were regions that were deleted in this strain. Therefore, we subtracted all positions from the reference genome that were covered by at least one read. The set of positions left over were the ones we called deleted. From this analysis the only deletion different between the JX2.0 and JX3.0 strains was the *prfA* gene. FASTQ files for each sample have been deposited to the Short Read Archives (SRA Accession# SRA016379.1).

Supplementary Results

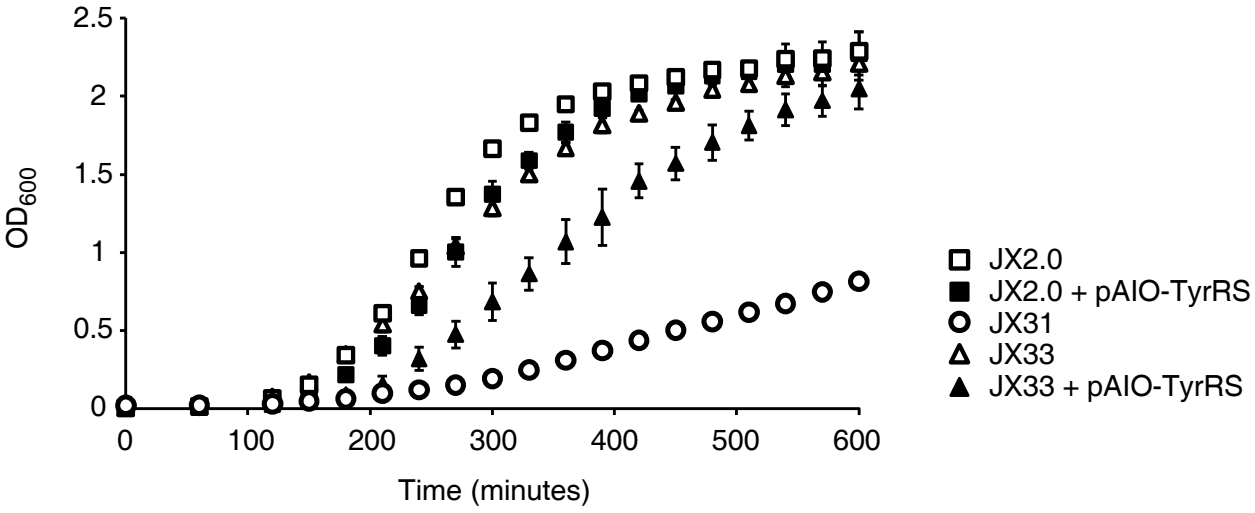

Figure 1. Growth rates of the JX2.0 and JX3.0 strains. Colony JX31 represents the JX3.0 strain; only a single colony JX33 was found to grow faster than other colonies. OD_{600} values were recorded at different time points for each strain. Growth rates of JX2.0 (□) and JX33 (Δ) were assayed in the absence (open) or presence (filled) of the pAIO-TyrRS plasmid (see **Figure 2**), which incorporates Tyr at the UAG codon. Shown is the average from three independent measurements with error bars representing s.e.m..

a

Strain	Postition	Reference	Mutation	Reads	Confidence	Gene	
JX33	2503033	G	A	246	0.984	*ypdE*	Silent
JX33	3033426	G	T	464	0.998	*prfB*	A293E

b

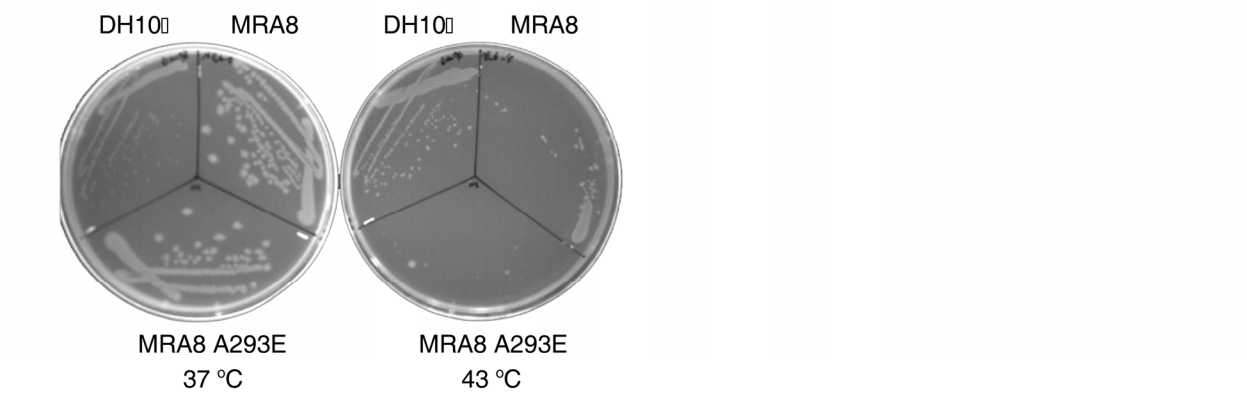

Figure 2. Full-genome sequencing of JX2.0, JX31 and JX33 confirms the knockout of RF1. (**a**) Compared to JX2.0, full-genome sequencing confirmed the deletion of *prfA* in JX31 and JX33. Two SNPs were found in JX33 but none in JX31. (**b**) The A293E mutation found in RF2 in JX33 could not rescue the RF1 temperature sensitive phenotype of the MRA8 strain. The MRA8 A293E strain has its *prfB* gene replaced by the *prfB*^{f}*(A293E)* gene from JX33. Cells were grown at both 37 and 43 ^{o}C to assay for any change in growth defects. The *prfB*^{f}*(A293E)* did not rescue the growth defect of MRA8 at 43 ^{o}C. The control strain DH10β had no such defect.

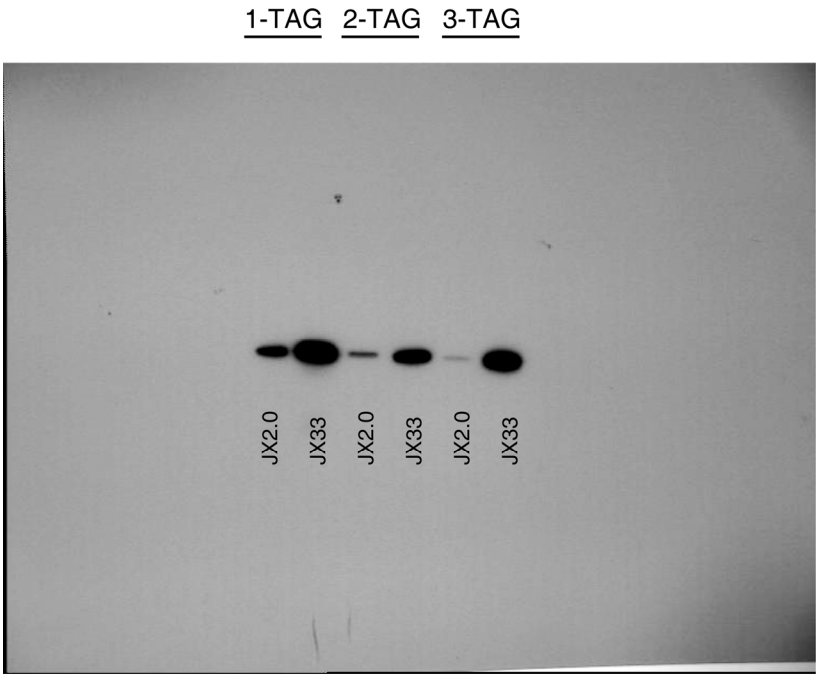

Figure 3. Full gel for Figure 2b, left panel.

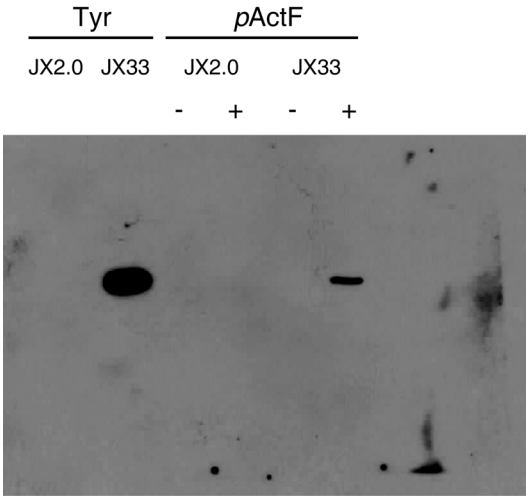

Figure 4. Full gel for Figure 2b, right panel and Figure 2d, right panel.

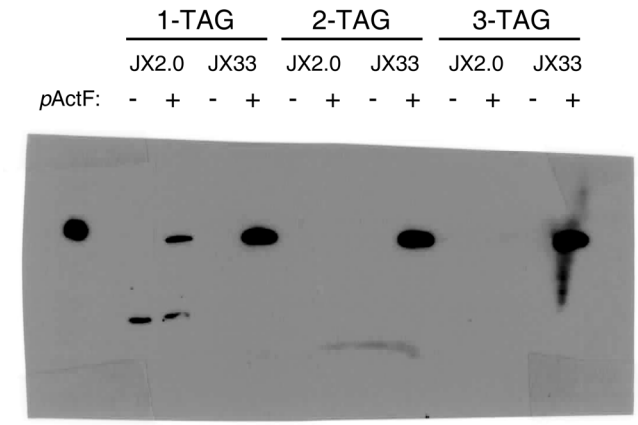

Figure 5. Full gel for Figure 2d, left panel.

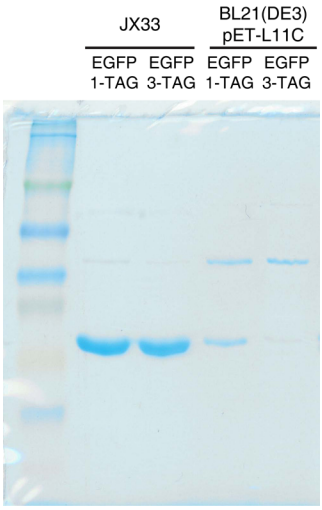

Figure 6. Full gel for Figure 2g.

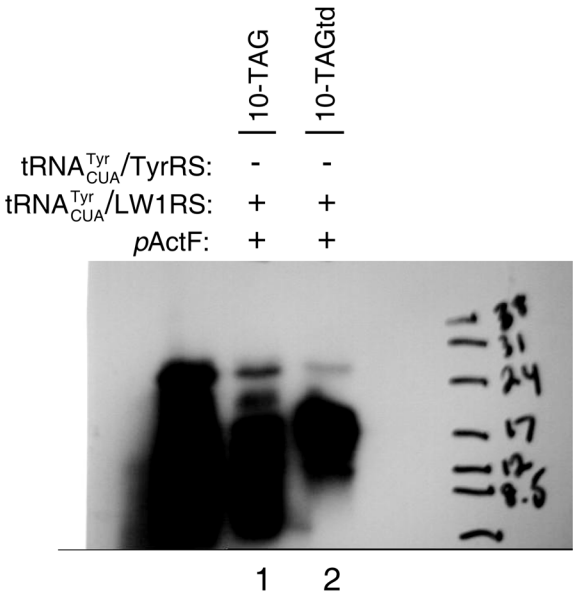

Figure 7. Full gel for Figure 4b, left panel.

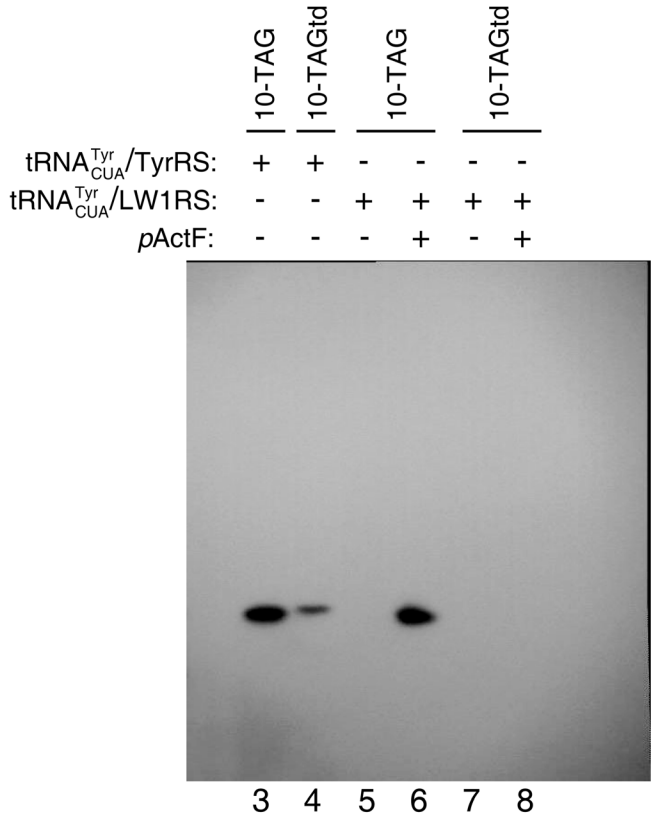

Figure 8. Full gel for Figure 4b, right panel.

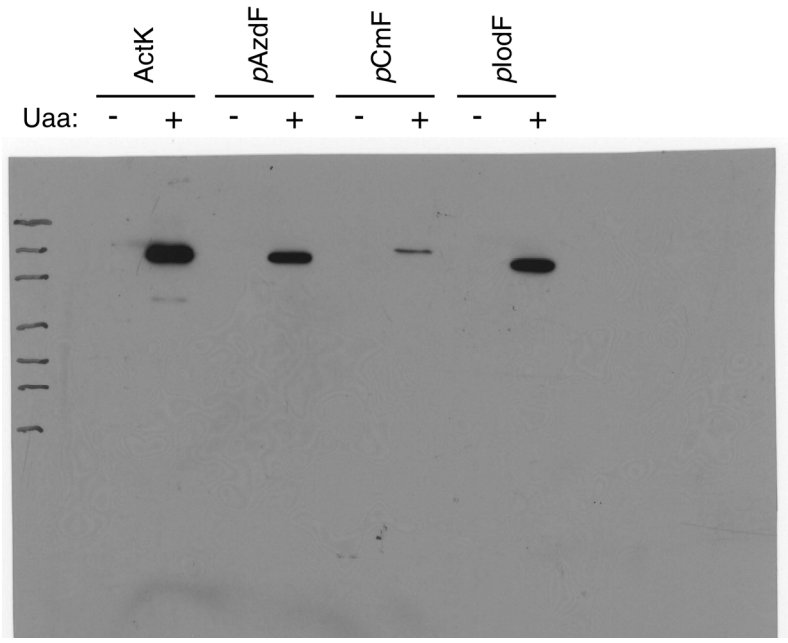

Figure 9. Full gel for Figure 5a.

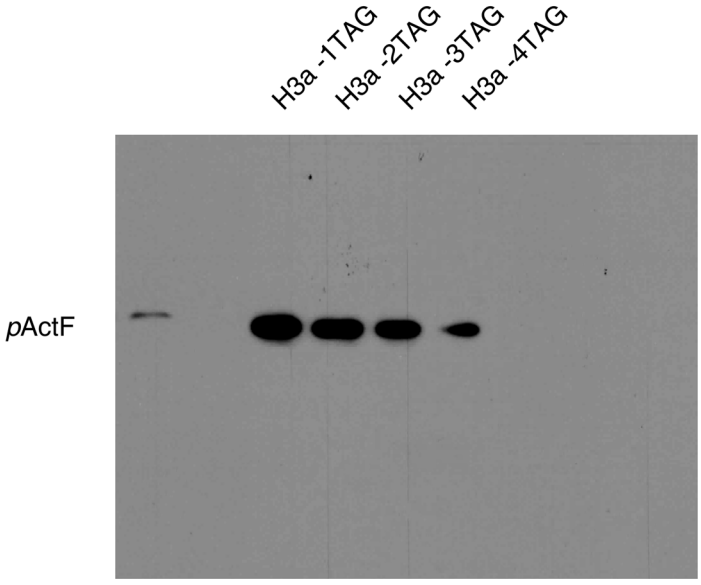

Figure 10. Full gel for Figure 5d, upper panel.

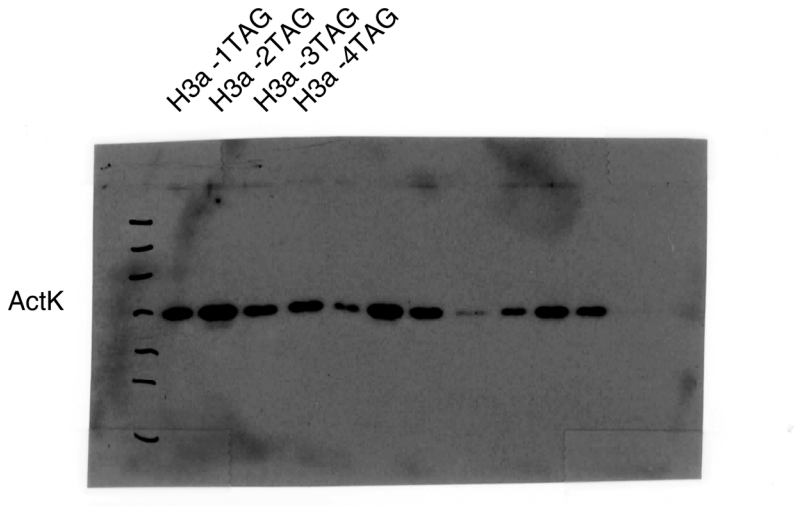

Figure 11. Full gel for Figure 5d, bottom panel.

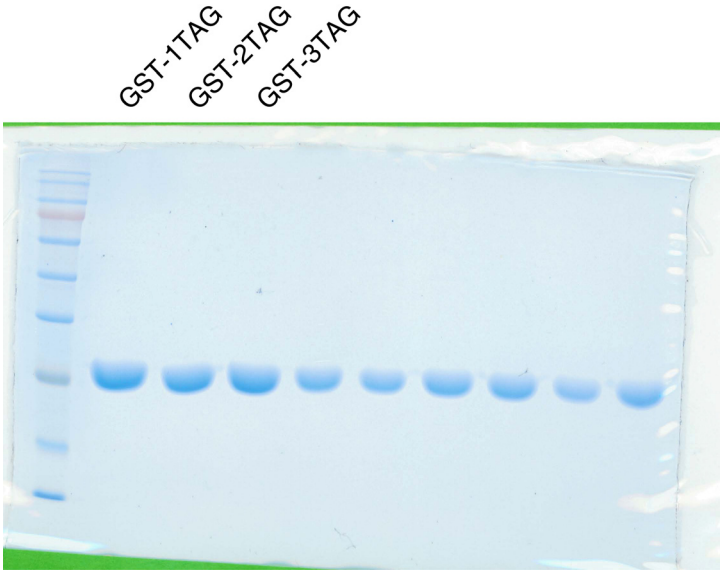

Figure 12. Full gel for Figure 5e.

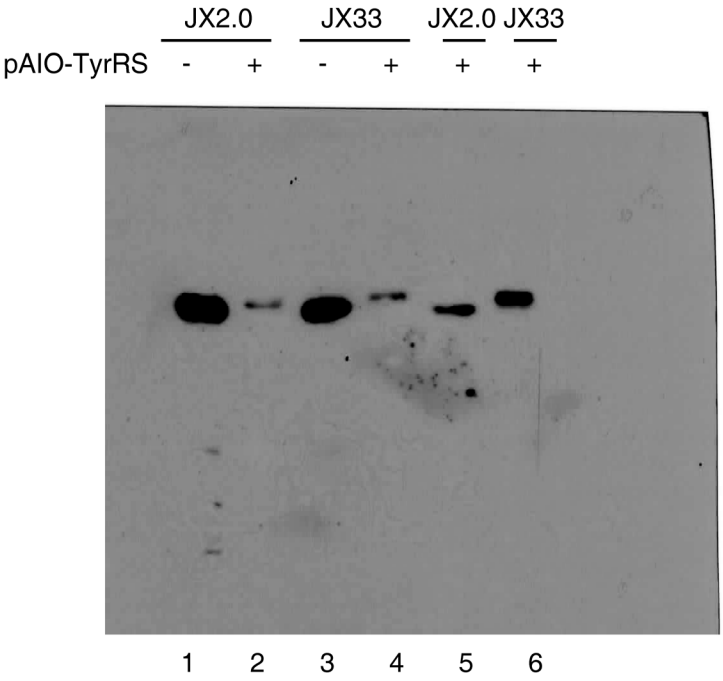

Figure 13. Full gel for Figure 6b.

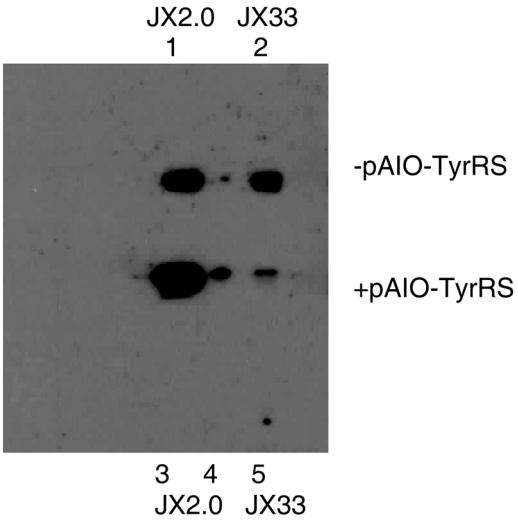

Figure 14. Full gel for Figure 6c, left and middle panels.

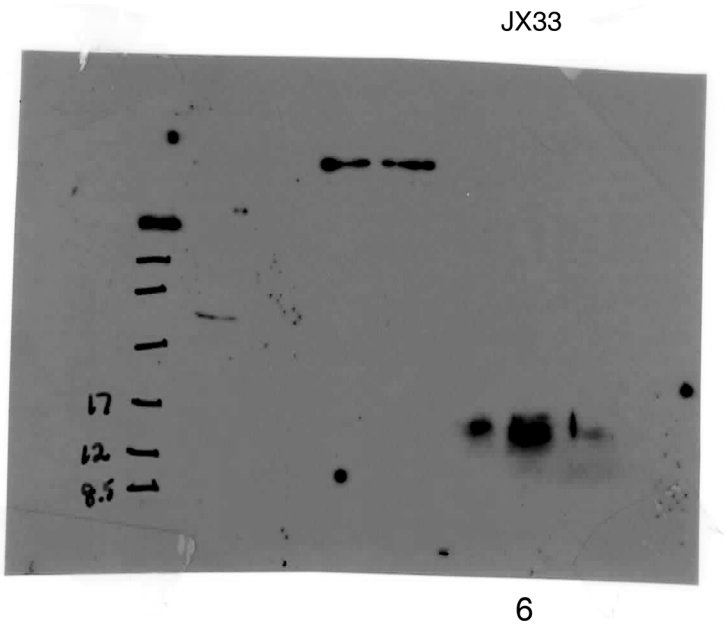

Figure 15. Full gel for Figure 6c, right panel.

Table 1. Protein yields of the various EGFP-TAG constructs expressed in JX2.0 and JX33.

Protein species	JX2.0 (mg/L)	JX33 (mg/L)
wild-type EGFP (no TAG)		14.9±0.5
6 - Tyr	N/D	6.8±0.3
10 - Tyr	N/D	0.4±0.1
1 - *p*ActF	1.8±0.1	3.5±0.1
2 - *p*ActF	N/D	3.5±0.1
3 - *p*ActF	N/D	5.4±0.2
6 - *p*ActF	N/D	0.5±0.1
10 - *p*ActF	N/D	0.5±0.1
3 - *p*AzdF	N/D	0.8±0.2
3 - ActK	N/D	1.0±0.2
3 - *p*CmF	N/D	0.2±0.1
3 - *p*IodF	N/D	0.9±0.2

Protein yields were determined from samples purified first with Ni-NTA chromatography followed by FPLC using an anion exchange column. FPLC purification is necessary to remove truncated protein products when the His6 tag is appended at the N-terminus. N/D: not determined due to too low yield; for these samples, no protein expression was detected in Western analyses. Errors represent s.e.m.

Table 2. Summary of all UAG site-containing peptides from the 1-, 2-, and 3-TAG EGFP mutants expressed in JX33.

Protein	UAG position	Gln-containing peptide intensity	pActF-containing peptide intensity	pActF incorporation
1-TAG EGFP	Y182	6.52×10^5	8.23×10^8	99.921%
2-TAG EGFP	Y39	1.92×10^5	7.11×10^9	99.997%
	Y182	1.25×10^7	6.42×10^9	99.806%
3-TAG EGFP	Y39	4.01×10^6	6.78×10^9	99.941%
	Y151	0	4.61×10^8	100%
	Y182	1.40×10^6	2.68×10^9	99.948%

Peptide intensities were determined by the area peak in the EIC and translated into the incorporation fidelity for *p*ActF

Table 3. Tandem MS analysis of SufA protein purified from JX33 harboring $tRNA^{Tyr}_{CUA}$ /TyrRS.

Peptide	No. of spectra	Intensity
(K)AQNECGCGESFGV	2	8.71×10^7
(K)AQNECGCGESFGVY	2	7.99×10^7
(K)AQNECGCGESFGVYAVL	1	2.91×10^5
(Y)AVLCLVILK	5	2.13×10^8
(V)LCLCILK	2	7.35×10^6
(K)QLTMSK	3	6.85×10^5

Peptide fragments confirmed that the protein was extended to the next in-frame UGA codon and a Tyr was incorporated at the UAG site. A small amount of non-extended wild type SufA C-terminal peptide (bolded) was also detected by the very sensitive MS but not on Western.

Table 4. Tandem MS analysis of YfiA purified from JX33 harboring $tRNA^{Tyr}_{CUA}$ /TyrRS.

Peptide	No. of spectra	Intensity
(K)DANFVEEVEE	1	8.21×10^4
(K)DANFVEEVEEE	7	5.54×10^6
(K)DANFVEEVEEEY	5	1.47×10^4
(K)DANFVEEVEEEYS	5	1.20×10^6
(K)DANFVEEVEEEYSF	2	2.29×10^4
(K)DANFVEEVEEEYSFIL	1	1.35×10^4
(K)DANFVEEVEEEYSFILS	16	6.40×10^5
(K)DANFVEEVEEEYSFILSPTR	1	6.85×10^4

Peptide fragments showed that no peptide was extended to the next UGA codon. Various extensions before the terminator hairpin structure were identified, suggesting an early ribosome drop-off. Peptides extended to the 0, 2nd and 6th amino acid after UAG were predominant and underlined.

Supplementary References

1. Craigen, W.J., Cook, R.G., Tate, W.P. & Caskey, C.T. Bacterial peptide chain release factors: conserved primary structure and possible frameshift regulation of release factor 2. *Proc. Natl. Acad. Sci. U. S. A.* **82**, 3616-20 (1985).
2. Uno, M., Ito, K. & Nakamura, Y. Functional specificity of amino acid at position 246 in the tRNA mimicry domain of bacterial release factor 2. *Biochimie* **78**, 935-43 (1996).
3. Datsenko, K.A. & Wanner, B.L. One-step inactivation of chromosomal genes in Escherichia coli K-12 using PCR products. *Proc. Natl. Acad. Sci. U. S. A.* **97**, 6640-5 (2000).
4. Herring, C.D., Glasner, J.D. & Blattner, F.R. Gene replacement without selection: regulated suppression of amber mutations in Escherichia coli. *Gene* **311**, 153-63 (2003).
5. Agafonov, D.E., Kolb, V.A. & Spirin, A.S. Ribosome-associated protein that inhibits translation at the aminoacyl-tRNA binding stage. *EMBO Rep.* **2**, 399-402 (2001).
6. Lee, J.H., Yeo, W.S. & Roe, J.H. Induction of the sufA operon encoding Fe-S assembly proteins by superoxide generators and hydrogen peroxide: involvement of OxyR, IHF and an unidentified oxidant-responsive factor. *Mol. Microbiol.* **51**, 1745-55 (2004).
7. Luger, K., Rechsteiner, T.J. & Richmond, T.J. Preparation of nucleosome core particle from recombinant histones. *Methods Enzymol.* **304**, 3-19 (1999).
8. Zhang, Z. & Marshall, A.G. A universal algorithm for fast and automated charge state deconvolution of electrospray mass-to-charge ratio spectra. *J. Am. Soc. Mass Spectrom.* **9**, 225-33 (1998).
9. Tanner, S. *et al.* Improving gene annotation using peptide mass spectrometry. *Genome Res.* **17**, 231-9 (2007).

10. Castellana, N.E. *et al.* Discovery and revision of Arabidopsis genes by proteogenomics. *Proc. Natl. Acad. Sci. U. S. A.* **105**, 21034-8 (2008).

11. Ossowski, S. *et al.* Sequencing of natural strains of Arabidopsis thaliana with short reads. *Genome Res.* **18**, 2024-33 (2008).