Release Factor One is Nonessential in *Escherichia coli*

David B.F. Johnson[1,5], Chong Wang[1,5], Jianfeng Xu[1], Matthew D. Schultz[2,3], Robert J. Schmitz[2], Joseph R. Ecker[2,4], and Lei Wang[1,*]

Supporting Methods

Genomic Sequencing. Each strain was grown in 2xYT (yeast extract tryptone) medium to the stationary phase. Cells were pelleted, and genomic DNA was purified using a Qiagen DNeasy kit (Qiagen). One microgram of genomic DNA was used to prepare DNA libraries for sequencing. Genomic DNA was fractionated using the Covaris S2 System (Applied Biosystems) using the following parameters: cycle number = 6, duty cycle = 20%, intensity = 5, cycles/burst = 200 and time = 60 s. Fractionated DNA was purified using a Qiagen PCR minielute purification kit. Purified DNA was repaired using the End-It Repair Kit (Epicentre) and purified using a Qiagen minielute column. Purified DNA was A-tailed using dATP and Klenow (3'-5' exo-) from New England Biolabs (NEB) and then purified with a Qiagen minielute column. Purified DNA was then ligated overnight with Illumina genomic DNA adapters using T4 DNA Ligase from NEB and purified using a Qiagen minielute column. The ligated DNA was run on a 2% agarose gel and size selected to remove adapters. Gel extraction was performed on the gel slice using Qiagen minielute gel purification kit. Purified DNA was PCR amplified using 1 μL of ligated DNA and Phusion Taq from NEB and size selected from a 2% agarose gel.

Genomic DNA libraries were sequenced using the Illumina Genome Analyzer II or the HiSeq 2000 (Illumina) as per manufacturer's instructions. Sequencing of genomic DNA libraries was performed up to 101 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 (*1*) according to the documentation provided with the software. In brief, the parental *E. coli* 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. From this analysis the only deletion difference between the knockout and parental strains was the *prfA* gene in all cases. The REL606 parental strain obtained from the Coli Genetic Stock Center at Yale University contains a 3017bp deletion (from 3738169 to 3741185) compared to the published REL606 genomic sequence (*2*). Single nucleotide polymorphisms found in genomic sequencing were further verified using PCR amplification of the identified loci followed by standard DNA sequencing.

Western Analyses. *E. coli* cells containing EGFP expression plasmids were grown in 2xYT medium at 37 °C for 16 hr, harvested, washed twice 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). After incubating for 10 min at 95 °C, the samples were separated on 12% or 15% SDS polyacrylamide gel. After transfer to nitrocellulose membrane,

EGFP was detected using the HRP-conjugate penta-His antibody (Qiagen). All blots were developed using the pico chemiluminescence kit (Thermo Scientific) according to manufacturer's specifications.

Protein Purification. For EGFP preparations, 500 mL 2xYT cultures were grown for 16 hr with or without the Uaa. Cells were pelleted, resuspended in lysis buffer (10% glycerol, 50 mM Tris pH 8.0, 500 mM NaCl, 5 mg/mL lysozyme, DNase, and 10 mM β-mercaptoethanol) and sonicated for 4 cycles at 90% power with a duty cycle of 50. Cell lysate was collected after centrifugation at 12,000 g for 30 min. Lysate was then added to 500 μL of pre-equilibrated Ni-NTA resin (Qiagen), washed with 50 column volumes of wash buffer (50 mM Tris, 500 mM NaCl, 20 mM imidazole, pH 8), and then eluted in three 1 mL fractions of elution buffer (same as wash buffer except 250 mM imidazole). Purified EGFP was buffer exchanged to 50 mM Tris buffer (pH 8.0) containing 500 mM NaCl using Microcon Ultracel YM-10 spin columns (Millipore, Billerica, MA), and further purified using a Sephadex-200 size exclusion column on a UPC-900 FPLC (GE healthcare). Peak fractions were analyzed by SDS-PAGE and pooled for further analysis. All protein concentrations and total yields were determined using the Bio-Rad protein assay kit according to manufacturer's specifications.

Mass Spectrometry. Trypsin-digested protein was analyzed by LC electrospray ionization MS as described (*3*). Briefly, samples were loaded onto a capillary column with an integrated spray tip (75 μm I.D., 10 μm tip, New Objective), which was packed

in-house with C18 reversed phase material (Zorbax SB-C18, 5 μm particle size, Agilent) to a length of 10 cm. The reversed phase elution was achieved by a linear gradient of 0-60% acetonitrile in 0.1% formic acid within 60 min at a flow rate of 300 nL/min. The eluate was introduced into a Thermo LTQ-Orbitrap mass spectrometer (ThermoFisher) via a nano-spray source. Mass spectrometric analysis was conducted by recording precursor ion scans at a resolution of 60,000 in the Orbitrap Fourier-transform analyzer followed by MS/MS scans of the top 5 ions in the linear ion trap (cycle time approx. 1 s). An active exclusion window of 90 s was employed. Data were analyzed on a Sorcerer Solo system running Sorcerer-Sequest and by using the Mascot algorithm (V. 27 rev.11, Matrix Science, London, UK). Data were further analyzed and visualized using the Scaffold software package (v. 2.6, Proteome Software).

Supporting References:

1. Ossowski, S., Schneeberger, K., Clark, R. M., Lanz, C., Warthmann, N., and Weigel, D. (2008) Sequencing of natural strains of Arabidopsis thaliana with short reads, *Genome Res. 18*, 2024-2033.
2. Jeong, H., Barbe, V., Lee, C. H., Vallenet, D., Yu, D. S., Choi, S. H., Couloux, A., Lee, S. W., Yoon, S. H., Cattolico, L., Hur, C. G., Park, H. S., Segurens, B., Kim, S. C., Oh, T. K., Lenski, R. E., Studier, F. W., Daegelen, P., and Kim, J. F. (2009) Genome sequences of Escherichia coli B strains REL606 and BL21(DE3), *J. Mol. Biol. 394*, 644-652.

3. Schubert, D., Herrera, F., Cumming, R., Read, J., Low, W., Maher, P., and Fischer, W. H. (2009) Neural cells secrete a unique repertoire of proteins, *J. Neurochem. 109*, 427-435.

Table S1. Full-genome sequencing confirms the RF1 knockout in JX1.0 and reveals the following single nucleotide polymorphisms compared to the parental BL21(DE3).

Position	Reference[a]	Read	Number	Confidence	Gene	Mutation
colony 1						
1200795	T	C	382	1	[b]	
1200805	C	T	385	1	[b]	
1200849	C	T	409	1	[b]	
1200862	C	T	397	1	*ycgX*	silent
1200961	A	C	198	1	*ycgX*	silent
1200982	T	C	217	1	*ycgX*	silent
3708564	T	C	422	1	ECD_03520[c]	Ala→Val

[a.] Reference genome for BL21(DE3) has access number CP001509.

[b.] These mutations lie in the intergenic region between *ycgX* and *ycgW*, which both are of phage origin.

[c.] ECD_03520 is a gene involved in transport of hexuronate.

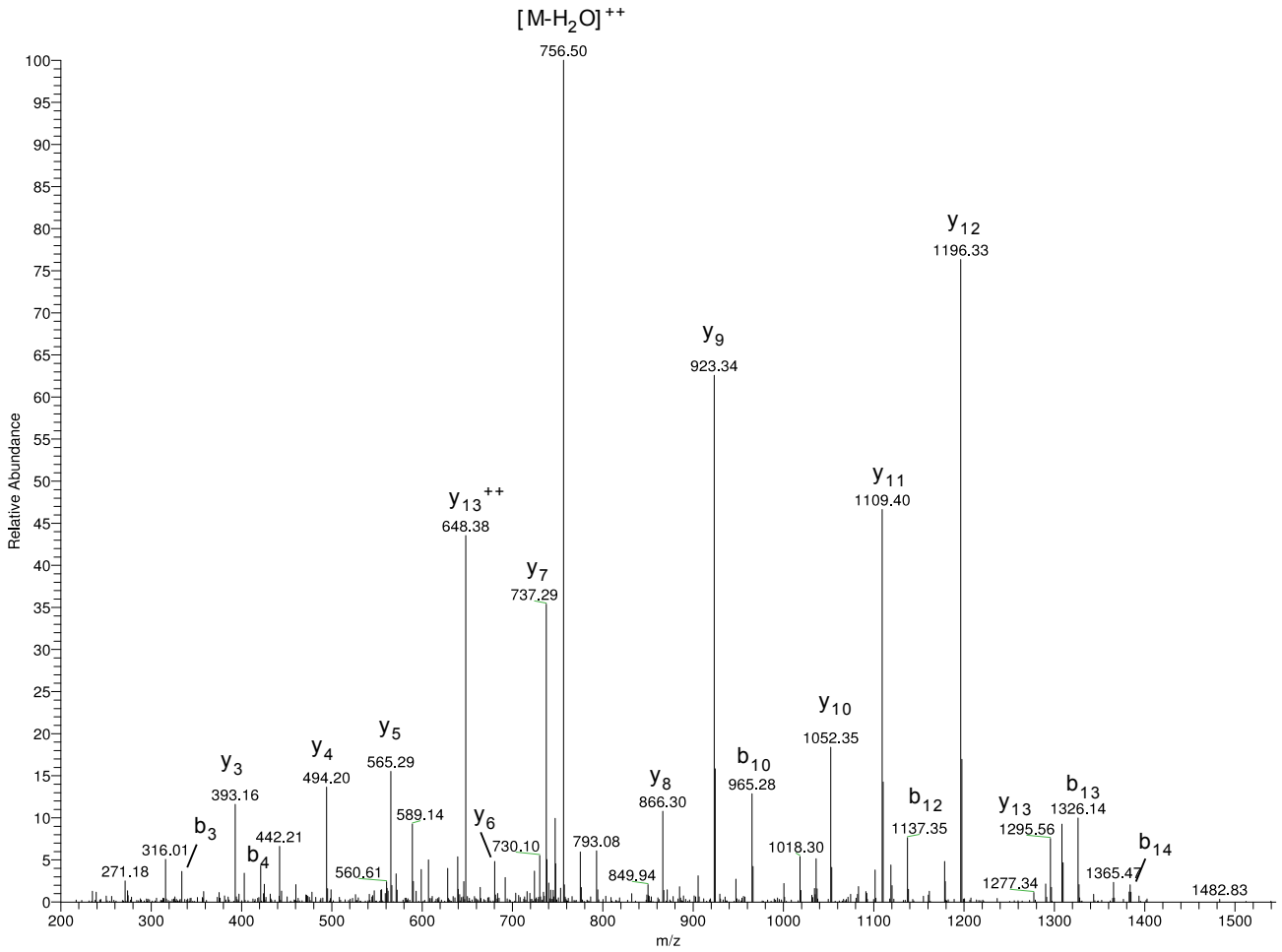

Figure S1a. Tandem mass spectrum of peptide FSVSGEGEGDAT**X**GK. **X** representing the UAG encoded site at position 39. The sequence of the peptide containing ActF at position **X** was assigned from the annotated b and y ion series.

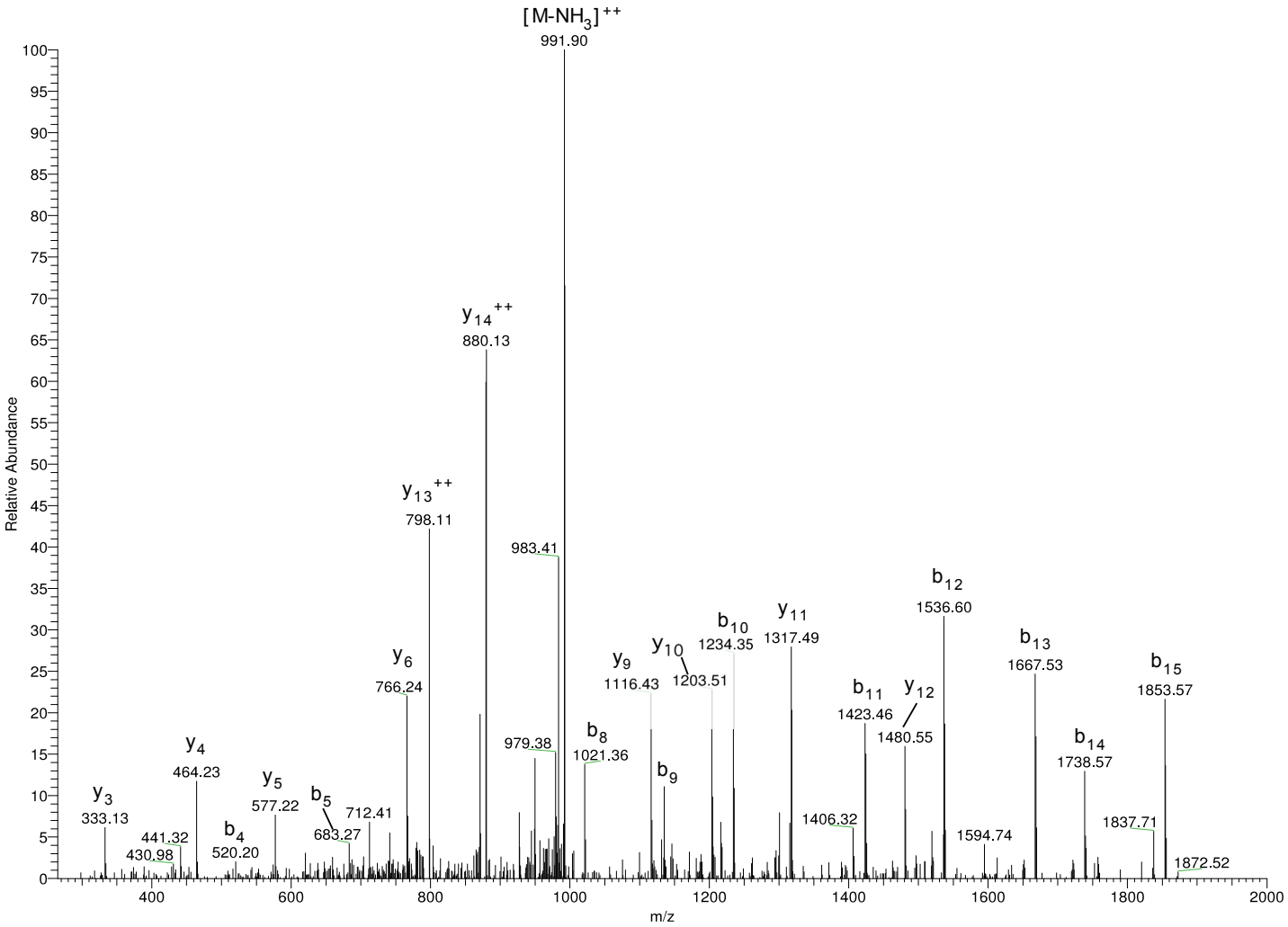

Figure S1b. Tandem mass spectrum of peptide LEYNYNSHNV**X**IMADK. **X** representing the UAG encoded site at position 151. The sequence of the peptide containing ActF at position **X** was assigned from the annotated b and y ion series.

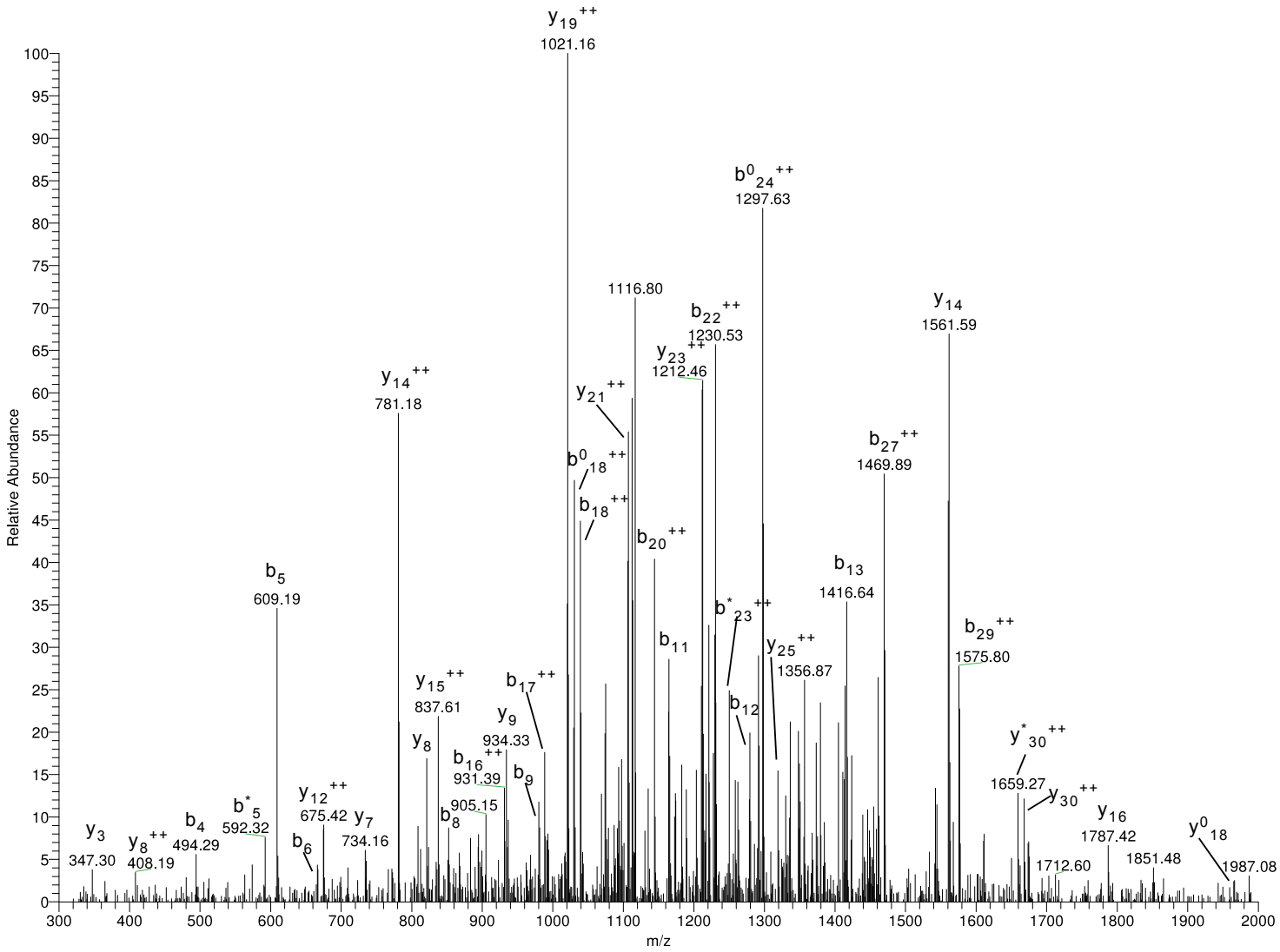

Figure S1c. Tandem mass spectrum of peptide HNIEDGSVQLADH**X**QQNTPIGDGPV LLPDNHYLSTQSALSK. **X** representing the UAG encoded site at position 182. The sequence of the peptide containing ActF at position **X** was assigned from the annotated b and y ion series.

coaD:
tag cgt tta tgc cgg atg gta tgc cat ccg gcg cgc atg aat tac ttc tgg cac tgc
cga caa taa aac gtt gcc cgc tgc gca tgt tta gtc gcc aca atc

hda:
tag ata agg tgt tta ttg tcg gat gcg gcg cga acg cct tat ccg aac tac gcg aag
tta atc cat agt att cgt agg ccg gat aag acg tgg caa gcg tcg cat ccg gca atg
tgc ttc aac tag cga tgg agc aca tct taa ccc acg att

hemA:
tag cag tac atc att ttc ttt ttt tac agg gtg cat tta cgc cta tga agc ctt cta
tcg ttg cca aac tgg aag ccc tgc atg aac gcc atg aag aag ttc

mreC:
tag tgg cga gct atc gta gcc agg gac gct ggg taa tct ggc tct ctt tcc tca ttg
cgc tgt tgc tgc aaa tca tgc cct ggc cgg ata acc tga ttg ttt

murF:
tag ttt ggc tgg ccg aac att tgg tca aat att att ccg gct tta acg tct ttt cct
atc tga cgt ttc gcg cca tcg tca gcc tgc tga ccg cgc tgt tca

lolA:
tag agg cac ctg agt gag caa tct gtc gct cga ttt ttc gga taa tac ttt tca acc
tct ggc cgc gcg tat gcg gcc aga aaa ttt agc aca gta tat cgg

lpxK:
tag tta cgc cgc ggc agc gtt cga ttg atg gag tca tga atg tcg ctg ccg cac ctc
tcc ctt gct gat gcg cgt aat ctt cac ctt gcc gca caa ggc ctg

Figure S2. The 7 essential genes ending with a TAG stop codon all have a different stop codon downstream in the reading frame. The TAG stop codon is shown at the beginning, and the second different stop codon is highlighted in yellow.