Identifying Small Proteins by Ribosome Profiling with Stalled Initiation Complexes

Jeremy Weaver,[a] Fuad Mohammad,[b] Allen R. Buskirk,[b] Gisela Storz[a]

[a]Division of Molecular and Cellular Biology, Eunice Kennedy Shriver National Institute of Child Health and Human Development, Bethesda, Maryland, USA
[b]Department of Molecular Biology and Genetics, Johns Hopkins School of Medicine, Baltimore, Maryland, USA

ABSTRACT Small proteins consisting of 50 or fewer amino acids have been identified as regulators of larger proteins in bacteria and eukaryotes. Despite the importance of these molecules, the total number of small proteins remains unknown because conventional annotation pipelines usually exclude small open reading frames (smORFs). We previously identified several dozen small proteins in the model organism *Escherichia coli* using theoretical bioinformatic approaches based on sequence conservation and matches to canonical ribosome binding sites. Here, we present an empirical approach for discovering new proteins, taking advantage of recent advances in ribosome profiling in which antibiotics are used to trap newly initiated 70S ribosomes at start codons. This approach led to the identification of many novel initiation sites in intergenic regions in *E. coli*. We tagged 41 smORFs on the chromosome and detected protein synthesis for all but three. Not only are the corresponding genes intergenic but they are also found antisense to other genes, in operons, and overlapping other open reading frames (ORFs), some impacting the translation of larger downstream genes. These results demonstrate the utility of this method for identifying new genes, regardless of their genomic context.

IMPORTANCE Proteins comprised of 50 or fewer amino acids have been shown to interact with and modulate the functions of larger proteins in a range of organisms. Despite the possible importance of small proteins, the true prevalence and capabilities of these regulators remain unknown as the small size of the proteins places serious limitations on their identification, purification, and characterization. Here, we present a ribosome profiling approach with stalled initiation complexes that led to the identification of 38 new small proteins.

KEYWORDS Ribo-seq, small protein, alternate ORFs, antisense, genome annotation, leader peptide

Protein-protein interactions play an essential role in a variety of cellular processes, such as signal transduction and gene regulation. Small proteins, here considered to be 50 amino acids or fewer and encoded by small open reading frames (smORFs), have been shown to interact with and modulate the functions of larger proteins (reviewed in references 1 to 3). These regulators have been identified in organisms spanning the phylogenetic tree of life, and important roles have been characterized for small proteins in bacteria and eukaryotes. In *Escherichia coli*, for example, the absence of the 49-amino-acid protein AcrZ renders cells more susceptible to specific antibiotics (4), and cells lacking the 31-amino-acid protein MgtS are sensitive to low magnesium concentrations (5, 6). In humans and other mammals, the small proteins myoregulin, sarcolipin, and phospholamban regulate muscle activity by affecting calcium transport (7, 8).

Despite the possible importance of small proteins, the true numbers of these regulators remain unknown, as their small size limits their identification. ORF-finding algorithms traditionally employ a size limit for the scoring of genes (9) and apply a

penalty for overlapping other ORFs (10). Their small size also often prevents these proteins from being accurately detected with protein gels, as they may run in the dye front and be poorly bound by SDS or protein dye (11). Traditional methods of purification are also biased against small proteins (12, 13), which have insufficient charge to bind ion-exchange columns and insufficient size to interact with non-reverse-phase hydrophobic columns or be retained during dialysis. Additionally, small membrane proteins can bind nonspecifically to many column matrices due to their hydrophobicity. Finally, the few peptide fragments generated by proteolysis of small proteins limit detection by shotgun proteomics (14). These challenges have stifled the detection of this class of proteins by standard methods.

As the importance of small proteins is being recognized, more focused searches for these proteins are being carried out (reviewed in reference 15). Early genome-wide studies in *E. coli* utilized conservation of intergenic DNA sequences and the strength of ribosome binding sites as a starting point for finding new small genes (16, 17). Similar approaches have been applied in eukaryotic organisms (18, 19), though the computational methods are more difficult, as both the increased size of the genome and the use of alternative splicing can mask small protein genes. In addition to the smORFs found in intergenic regions, there is growing recognition that transcripts can encode proteins in more than one ORF in the same region (reviewed in references 20 and 21); these alternative ORFs (altORFs) generally code for smaller proteins than the originally annotated ORF, with some reported altORF-encoded proteins as small as 14 amino acids (22). Despite the success of computational methods in identifying new smORFs, it is likely that many small proteins have been missed (false-negative results). Conversely, it is critical that the synthesis of predicted small proteins be verified to avoid false-positive results.

Integration of data from large transcriptome analyses can improve the success of computational searches for smORFs. Ribosome profiling, a method that involves deep sequencing of ribosome-protected mRNA fragments, reveals the position of ribosomes throughout the transcriptome, clarifying which smORFs are translated under the conditions examined (reviewed in reference 23). This approach has led to the identification of many small proteins (24–28), but again, there are limitations. Signals corresponding to altORFs encoded inside the confines of other genes can be swamped by the signal of the annotated gene. Another issue is that ribosome binding to an RNA does not prove that it leads to the production of a polypeptide (29). In eukaryotes, several signatures of profiling data that argue for translation are the presence of strong start and stop codon peaks, as well as three nucleotide periodicity arising from the translocation of ribosomes one codon at a time (30). In bacteria, however, these signatures are weaker and more variable due to the lower resolution of the method, further complicating the discrimination of which transcripts are translated and which are merely bound to the ribosome.

Although peaks in ribosome density at start and stop codons are the most useful in identifying new ORFs, the vast majority of ribosome-protected footprints in profiling data correspond to elongating ribosomes. In eukaryotes, the antibiotics harringtonine and lactimidomycin have been used to trap newly initiated 80S ribosomes at start codons and identify initiation sites (31, 32); elongating ribosomes are not inhibited by these antibiotics and continue elongation, terminating normally at stop codons. However, these compounds do not work in bacteria. Mori and coworkers (33) found that treating *E. coli* cultures with tetracycline, an antibiotic that blocks tRNA binding in the ribosomal A site, leads to the accumulation of ribosome density at start codons. Using ribosome profiling of tetracycline-treated cells, they were able to reannotate the N termini of many known ORFs and discover candidate smORFs in intergenic regions (33). However, tetracycline traps ribosomes imperfectly at start codons. Only half the ribosomes on genes map to their start sites, blurring the signal. One promising alternative, Onc112, prevents initiation complexes from entering into the elongation phase (34, 35). Another promising substitute, retapamulin, a small molecule member of the pleuromutilin class, was previously shown to have a similar ability to specifically inhibit the

first steps of elongation (36). The recent application of retapamulin in profiling experiments showed strong ribosome density at known start codons and little density attributable to elongating ribosomes; these data allowed the identification of start codons of altORFs within the coding sequences of other genes (37).

Here we present a strategy for identifying small protein genes in *E. coli* by combining traditional ribosome elongation data with information about initiation sites gleaned from profiling experiments conducted with Onc112 and retapamulin. We sought to verify the synthesis of a subset of the predicted small proteins by assays to detect tagged derivatives and observed expression of 38 of the 41 genes tested. These results demonstrate that ribosome profiling with stalled initiation complexes provides a high-confidence prediction of new small proteins in bacteria. Finally, the presence and location of these new smORFs reveal the density of information encoded by bacterial genomes.

RESULTS

Onc112 traps ribosomes at start codons but does not interfere with elongating ribosomes. The identification of initiation sites in eukaryotes has been aided by the use of antibiotics that enrich for ribosome density at start codons in ribosome profiling experiments. Since such antibiotics have not been available for bacteria, we tested a promising candidate, Onc112, a proline-rich antimicrobial peptide (PrAMP) that binds in the exit tunnel and blocks aminoacyl-tRNA binding in the ribosomal A site (34, 38). Toeprinting analyses showed that Onc112 traps ribosomes at start codons, blocking elongation (35). We hypothesized that Onc112 should be selective for newly initiated 70S complexes because elongating ribosomes contain a nascent polypeptide that should prevent antibiotic binding. To test this possibility, we performed ribosome profiling on an untreated *E. coli* culture, as well as one treated with 50 μM Onc112 for 10 min. As shown in Fig. 1a, ribosome density on the highly expressed *lpp* gene is spread across the coding sequence in the untreated sample but is found almost exclusively at or near the start codon in the Onc112-treated sample. This effect holds genome-wide; in plots of ribosome density averaged over thousands of genes aligned at their start codons, a strong peak appears at the start codon, while there is little or no density attributable to elongating ribosomes within coding sequences (Fig. 1b). These data show that like harringtonine and lactimidomycin in eukaryotes, Onc112 specifically traps ribosomes at start codons, while allowing elongating ribosomes to complete protein synthesis and terminate normally.

Ribosome profiling signals for Onc112 and retapamulin are similar. A recent study used retapamulin and ribosome profiling to identify sites of noncanonical initiation within annotated ORFs (37). Like Onc112, retapamulin traps newly initiated 70S ribosomes at start codons while allowing elongating ribosomes to complete protein synthesis, such that ribosome density is strongly enriched at start codons (Fig. 1a). In the Vázquez-Laslop and Mankin study (37), 12.5 μg/ml retapamulin was added to a culture 5 min prior to harvesting the cells; ribosome profiling was then performed following the standard protocol (39). Since the gene encoding the efflux pump TolC was deleted in the strain assayed (BW25113), 12.5 μg/ml retapamulin corresponds to approximately 100 times the MIC. To compare the effects of Onc112 and retapamulin treatment, we calculated the intensity of start codon peaks on annotated ORFs, finding 3,020 genes with any detected ribosome density in the start codon region in both samples. There is a strong correlation between start peak intensity in these two antibiotic-treated samples (Spearman's $r = 0.83$), arguing that both methods capture initiating ribosomes in a reproducible way (Fig. 1c).

Subtle differences may arise from variations in gene expression due to the different culture conditions; the retapamulin-treated sample was cultured in LB media, whereas the Onc112-treated sample was obtained from a culture in complete synthetic MOPS media (see Fig. S1 in the supplemental material). However, the primary relevant difference in the sample preparation is that our protocol with Onc112 includes an additional step not found in the standard protocol: the samples are pelleted over a
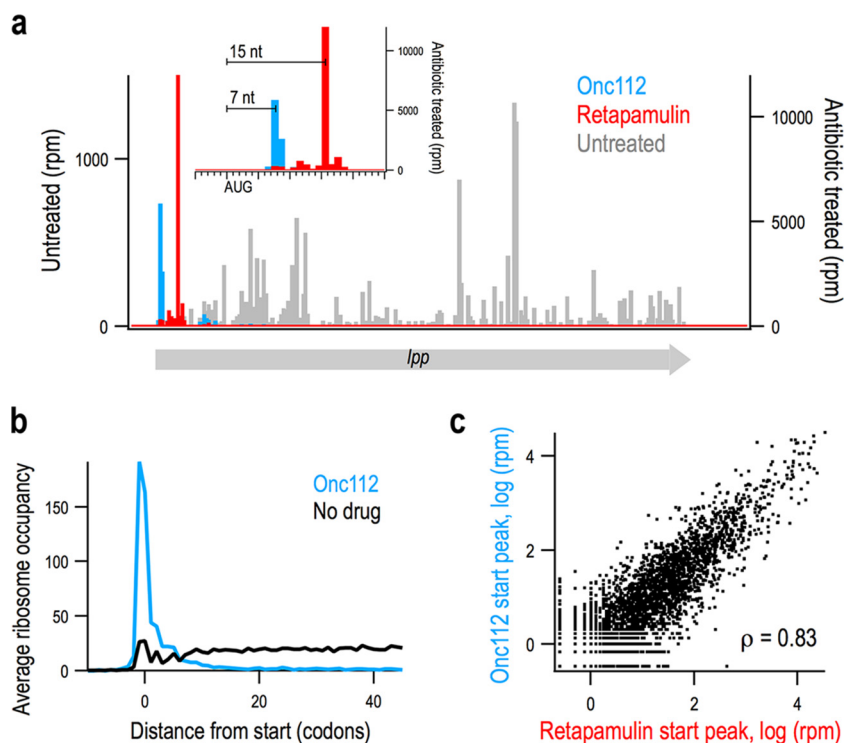
FIG 1 Onc112 and retapamulin similarly trap ribosomes at start codons. (a) Ribosome density on the *lpp* gene from Onc112-treated (blue), retapamulin-treated (red) samples and an untreated control (gray). (Inset) Close-up view of the start site of *lpp*. (b) Average ribosome density at many genes aligned at their start sites in a sample treated with Onc112 and an untreated control. (c) Scatter plot of density at start sites in annotated genes in samples treated with Onc112 or retapamulin. The Spearman rank correlation is reported.

sucrose cushion prior to nuclease treatment. This step depletes tRNAs from the ribosomal A site, allowing nucleases to cleave within the ribosome, thus shortening ribosome footprints. As a result, while the distance from the P-site codon to the 3′ boundary of the ribosome is reliably 15 nucleotides (nt) in the retapamulin-treated library, it is more variable in our Onc112-treated library. Most often, the peak is 6 to 10 nt downstream of the start codon; it is 7 nt in the *lpp* example (Fig. 1a). This difference is useful in annotating novel initiation sites: a start codon 6 to 10 nt upstream of density in the Onc112 data and 15 nt upstream of density in the retapamulin data has a high chance of being a bona fide start site and not a sequencing artifact.

Onc112 and retapamulin can be used to identify putative translated smORFs. Given the ability of Onc112 and retapamulin to trap ribosomes at start codons at most annotated ORFs, we combined the information from these data sets to create a high-fidelity screening method for identifying new smORFs likely to be translated (Fig. 2a). We first generated a list of 160,995 smORFs of eight codons or longer whose start codons (AUG, GUG, or UUG) are 18 nt or more away from annotated coding regions (either protein-coding sequences or functional RNA genes).

We computed the ribosome density associated with each start site, including ribosome footprints 0 to 18 nt downstream of the first nt in the start codon. A broad window (18 nt) was used because the distance from the 3′ end of footprints to the start codon can vary depending on the sequence context and the manner in which libraries were prepared, and we wanted to capture all the relevant footprints. A plot of the cumulative distribution function (CDF) of these data is shown in Fig. 2b (left *y* axis); the *y* value reflects the percentage of predicted smORFs that have a start peak less than or equal to the *x* value. This plot shows that ~96% of the putative smORFs have no associated density at their start sites ($x = 0$). This means that ~4% have start peaks greater than zero. Only 0.25% of the predicted smORFs had more than 5 reads per
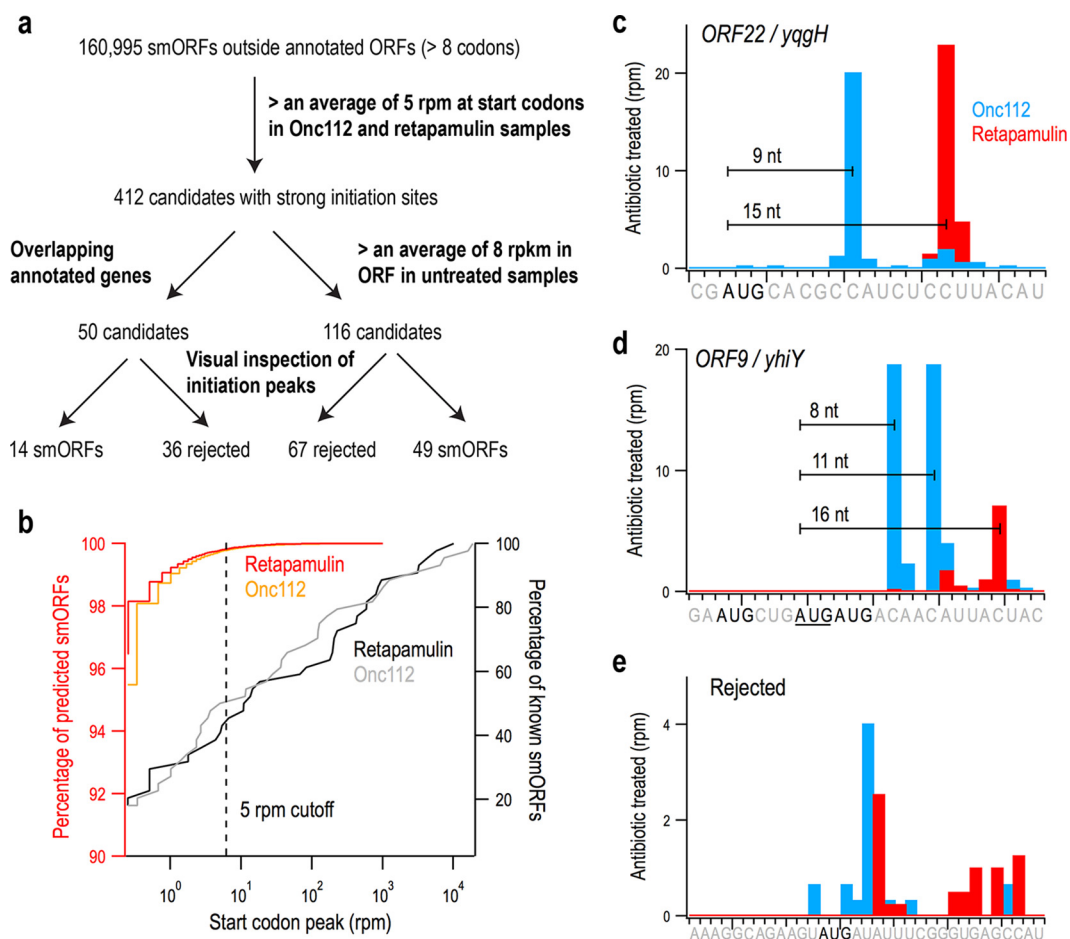
FIG 2 Using ribosome profiling data to discover new smORFs. (a) Flow chart showing the criteria used to identify smORFs in intergenic regions. (b) CDF plot showing the percentage of known, annotated smORFs ($n = 44$) (right y axis, black and gray) on the y axis less than or equal to the ribosome density near the start site (x axis) compared with candidate smORFs ($n = 160,995$) (left y axis, red and orange). Candidates with an average of >5 rpm were selected for further screening (broken line). (c) The proper spacing of ribosome density at start codons in treated samples helps to identify bona fide small protein-coding genes such as *ORF22/yqgH*. (d) In cases where several start codons could explain the ribosome density, spacing helps determine the correct site. *ORF9/yhiY* likely initiates with the second AUG codon of the three shown. (e) Many candidates were rejected because the start site does not align properly with the density observed.

million mapped reads (rpm), as delineated by the broken line in Fig. 2b. Thus, the vast majority of the putative smORFs likely are not translated.

To calibrate our method for identifying new candidate smORFs, we examined the ribosome density on the start codons of smORFs previously shown to encode proteins. The test set included two different groups. The first group was comprised of 44 small proteins annotated initially together with small proteins identified by sequence conservation and strong matches to ribosome binding site models (16). The ribosome density after Onc112 or retapamulin treatment varied by 4 orders of magnitude (see Table S1 in the supplemental material): ~80% of this group had detected signal at start sites and ~60% of known smORFs had start peaks above 5 rpm (Fig. 2b, right y axis). The second group of proteins had less conservation and weaker matches to ribosome binding site models but were shown to be synthesized as tagged derivatives in a recent study (17). Of the 36 proteins in the second set, ~70% showed signal but only 20% had Onc112 or retapamulin reads above 5 rpm at the start site (Table S1), possibly due to the lower level of expression of these smORFs. Given that the majority of the small proteins in the first set of annotated smORFs have start peaks 5 rpm or higher (Fig. 2b), we used this threshold to eliminate false-positive results in our list of putative smORFs; 412 novel smORFs above this threshold were selected for further consideration.

An important caveat in treating cells with Onc112 and retapamulin is that these antibiotics could enhance ribosome density on some initiation sites that are not normally used. The antibiotics dramatically increase the concentration of free 30S and 50S subunits given that they allow elongating ribosomes to complete protein synthesis and be recycled but block entry into the elongation cycle. The recycled subunits are free to initiate at less optimal start codons, where they will be trapped by the antibiotics. To remove these false-positive results, we used traditional ribosome profiling data (from untreated cells) to capture elongating ribosomes along the entire ORF. Of the 412 smORFs with strong start codon peaks, 116 had traditional ribosome profiling density above 8 reads per kilobase per million mapped reads (rpkm).

We next examined the 116 most promising candidates on a genome browser. In our screen for ribosome density at initiation sites (Fig. 2a), we summed the reads from 0 to 18 nt downstream of the first nucleotide in the start codon, an intentionally broad range. In our visual inspection, we searched for retapamulin peaks ~15 nt and Onc112 peaks 6 to 10 nt downstream of the first nucleotide in the start codon as seen for *lpp* (Fig. 1a). The same spacing is observed for the most promising candidates (e.g., *ORF22/yqgH* in Fig. 2c). In some cases of multiple possible start codons, we were able to readily predict the correct start based on distance (e.g., *ORF9/yhiY* [Fig. 2d]). For most of the candidates that were rejected, the predicted start site did not align with the Onc112 or retapamulin ribosome density (Fig. 2e). Another source of false-positive results were smORFs close to highly translated genes, such as ribosomal proteins, where the noise is high enough to pass the cutoff for start peaks and normal profiling density (data not shown). Based on these criteria, 67 candidates were rejected, leaving 49 candidates. Visual inspection proved helpful in refining the data, but in the future, our algorithms can be further developed to incorporate additional criteria for large-scale screens for candidate smORFs.

We also inspected 50 additional smORFs with strong start peaks (>5 rpm) for which we were unable to calculate rpkm values for elongating ribosomes because the smORFs overlap an annotated gene and the ribosome density cannot be assigned to one gene or the other. Upon inspection, 36 of these smORFs were rejected due to incorrect start site selection or high levels of noise due to adjacent highly translated genes, leaving 14 of interest. In addition to these 14 candidates and the 49 discussed above, another three were discovered as the correct start sites for candidates that were rejected, and two more were discovered in a preliminary screen using similar cutoffs but a different collection of traditional ribosome profiling data.

Together, this workflow yielded 68 candidate smORFs with high start codon peaks and some level of traditional ribosome profiling data, including both independent genes and altORFs (Table S3). Initially, the smORFs were assigned numbers but were renamed if we obtained evidence of small protein synthesis (see below). As expected, the majority of these candidates start with AUG codons ($n = 50$), although GUG ($n = 9$) and UUG ($n = 9$) codons were also observed. A histogram of the predicted protein lengths is shown in Fig. S2: the majority of the predicted small proteins are 40 amino acids or fewer, although the analysis also identified seven candidates that were longer than 50 residues. A few of the candidates are overlapping in that they correspond to different possible start codons in the same frame.

The majority of predicted small proteins are synthesized. To validate that the corresponding small proteins are synthesized, a sequential peptide affinity (SPA) tag (comprised of the 3× FLAG tag and calmodulin binding protein, adding 8 kDa [41]) was integrated into the chromosome upstream of the stop codon of the 38 putative smORF genes with the highest ribosome density in the presence of the inhibitors and deemed the strongest candidates by the visual inspection. The tag allowed immunoblot analysis on the basis of the 3× FLAG epitope (Fig. 3). While the exposure needed to detect the small proteins varied significantly (as reflected in different levels of the background bands), 36 of the 38 tagged small proteins were detected in cells grown to exponential
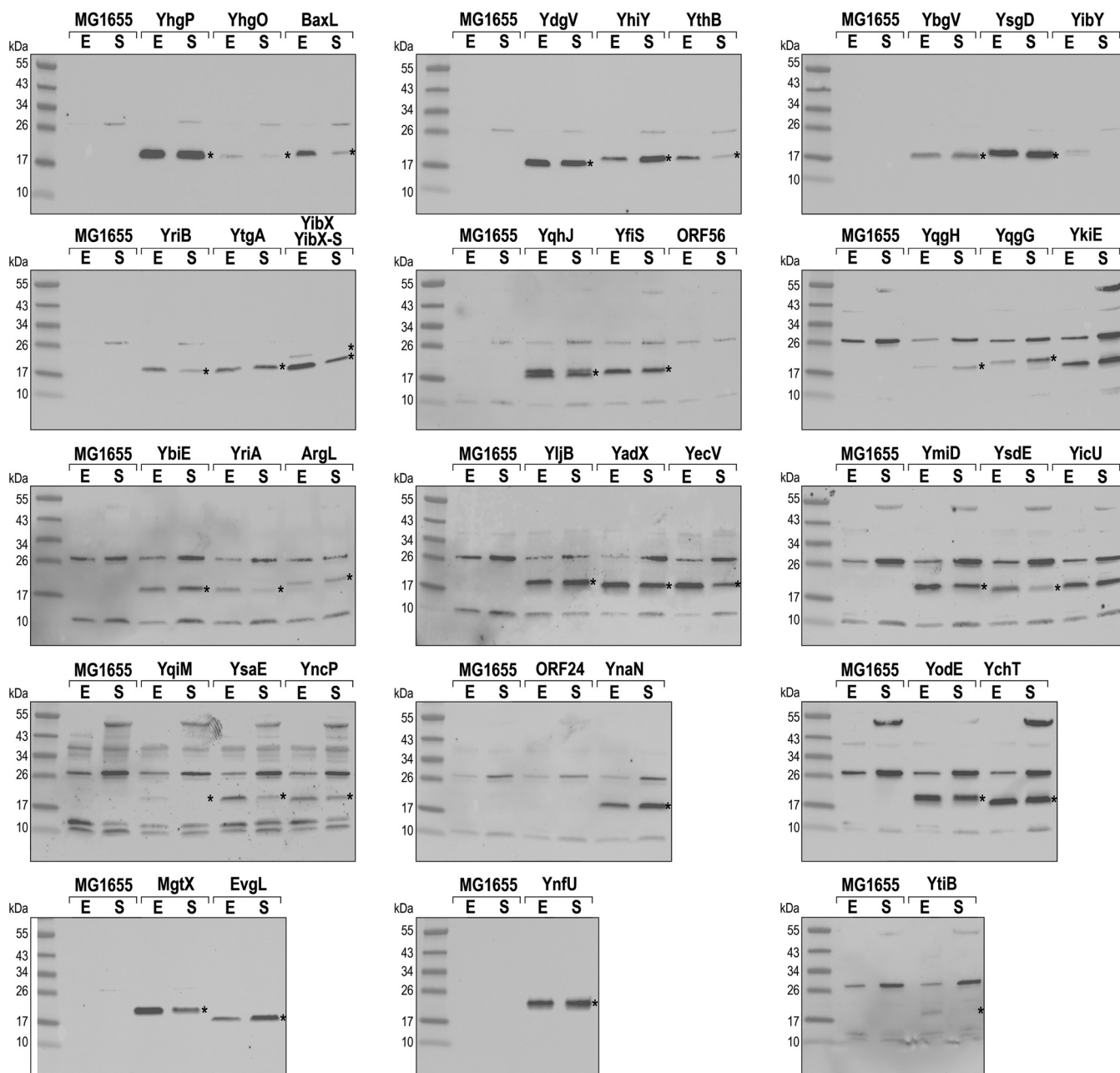
FIG 3 Western analysis confirms synthesis of 95% of predicted small proteins tested. *E. coli* MG1655 strains with chromosomally tagged, putative smORFs were grown to exponential (E) and stationary (S) phase in rich media (LB). Gel samples were prepared to load equivalent numbers of cells based on OD_{600}. Immunoblot analysis was conducted against the 3× FLAG motif included in the SPA tag using HRP-conjugated, anti-FLAG antibodies. Wild-type MG1655 was included as a negative control. Blots requiring a longer exposure to show tagged proteins have more background bands. Bands corresponding to small proteins are marked with an asterisk.

or stationary phase in LB at 37°C, conditions comparable to those used in the ribosome profiling experiments. The inability to detect the remaining two chromosomally tagged smORFs (*ORF24* and *ORF56*, as well as *ORF33* as shown below) could stem from these smORFs yielding false-positive results in the screen or from the degradation of the tagged derivatives. Nonetheless, we have observed the expression of the majority of the predicted genes, validating the predictive capability of utilizing multiple ribosome profiling data sets.

Several previously detected small proteins are expressed only under very specific growth conditions (17, 42). As shown in Fig. 3, we observe that 10 newly detected small

proteins are present at >2-fold higher levels in exponential phase and four are present at >2-fold higher levels in stationary phase. The majority of the small proteins appear at roughly equal levels during both of these growth phases but may be induced under other conditions.

The levels of tagged small proteins span a wide range. As indicated above, the ability to detect the small proteins varied. To directly compare the overall levels of the proteins, both among themselves and with previously identified small proteins, we analyzed stationary-phase samples of several examples of each group of proteins (Fig. 4). Among the newly identified proteins, the levels of YnfU are highest, but these levels fall between the levels of the characterized multidrug efflux pump regulator AcrZ (diluted fivefold in Fig. 4a) and the uncharacterized protein YoaK, which, respectively, are among the better- and worse-expressed small proteins identified in initial searches (16). The levels of the remaining small proteins cover a wide range, as is seen when comparing the samples loaded on two different gels as a reference, YsgD in Fig. 4a and b and YthB on Fig. 4b and c. These blots also show that most of the other newly identified small proteins are expressed at levels below the level of YoaK under the conditions tested.

We also compared the levels of the new small proteins to five (YnaM, YnfS, YgbU, YddY, and YmjE) of the 36 small proteins identified more recently (17). Three of the proteins (YnaM, YnfS, and YbgU) are observed at levels comparable to most of the newly identified small proteins, while two (YddY and YmjE) are more comparable to the least-abundant small proteins identified in this study (Fig. 4c). It is interesting to note that YnaM, which had no ribosome density at start codons in the presence of Onc112 or retapamulin, was detected at higher levels than most of the newly detected small proteins, while YnfS, which has strong start peaks in both antibiotic-treated samples, was detected at lower levels.

Some small proteins are encoded antisense to genes encoding expressed proteins. Given that antisense transcription in bacteria frequently is a means of gene silencing (reviewed in references 43 and 44), we were surprised to note that eight of the newly detected proteins are encoded antisense to annotated protein-coding genes (Table 1). Additionally, one predicted smORF, *yoaM*, could not be tagged as it is found antisense to the operon of the essential *nrdA* and *nrdB* genes (encoding ribonucleoside-diphosphate reductase 1) (Fig. 5a). To test for expression of YoaM, we generated a translational fusion at the *lacZ* locus. Consistent with translation of this antisense-encoded small protein, we detect higher β-galactosidase expression for the *yoaM-lacZ* fusion than for an out-of-frame control fusion (Fig. 5e). Given that a clear transcriptional start was noted 174 nucleotides upstream of the YoaM start codon (45), it is possible that the synthesis of this protein is under posttranscriptional regulation.

We wanted to determine whether annotated proteins and the newly identified small proteins encoded by transcripts on opposing strands are both synthesized. We therefore introduced chromosomal SPA tags upstream of the stop codons of the previously annotated genes *yqgC* (antisense to *yqgG*) (Fig. 5b), *yghE* (antisense to *yqhJ*) (Fig. 5c), and *waaL* (antisense to *yibX* and *yibY*) (Fig. 5d). For YqgC (a protein of unknown function), the gene does not have any associated ribosome density in either treated or untreated cells, and the corresponding tagged protein is not observed under these conditions (Fig. 5f). YghE (another protein of unknown function), while detected, appears to be present at lower levels than YqhJ (Fig. 5g), consistent with its low levels of normal ribosome density (not visible at the scale used in Fig. 5c). WaaL (an O-antigen ligase) was clearly detected under the same growth conditions as YibX and YibX-S (Fig. 5h). We suggest the appearance of a smear for WaaL may be due to bound oligosaccharide substrates. In general, our results confirm that proteins can be encoded by both strands of the same region of DNA and expressed under the same growth conditions.
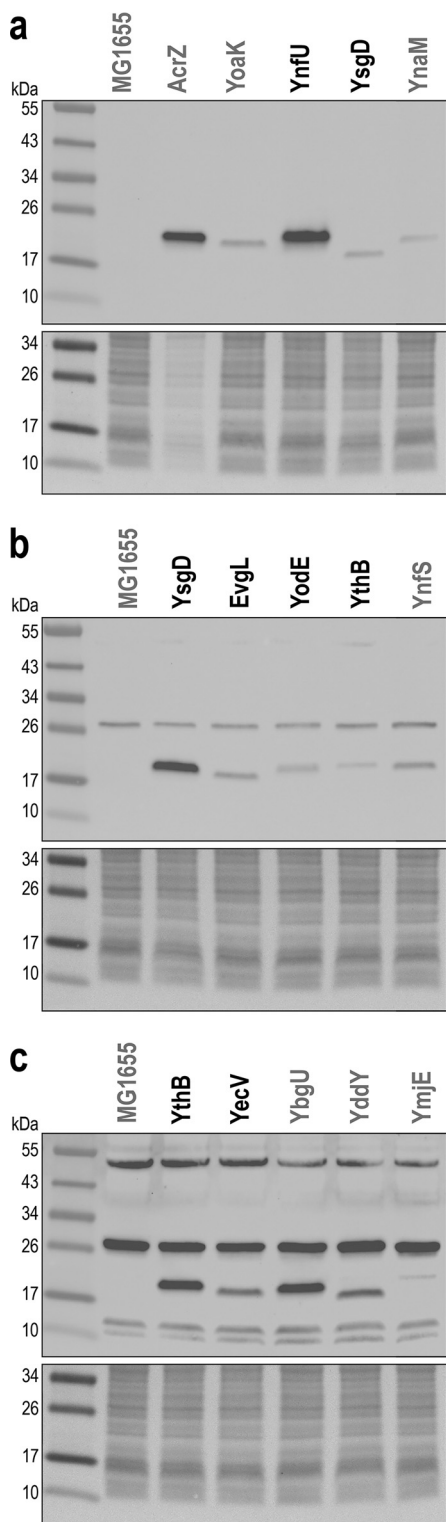
FIG 4 Observed small protein levels span several orders of magnitude. Stationary-phase samples grown in LB from Fig. 3 (black) were compared to each other and to similarly prepared samples of previously detected small proteins (gray) with the same chromosomal tag (17, 42). Immunoblot analysis for cells grown to stationary phase was conducted as described in the legend to Fig. 3 with *E. coli* MG1655 as a negative control. All samples are in the MG1655 background and equally loaded, except for AcrZ, where the sample was diluted 1:5. Ponceau S staining for the same region is shown below each immunoblot.

TABLE 1 New small proteins detected[a]

Left	Right	Strand	Minute	Gene name	Amino acid sequence	Length	Last codon	Start 1st codon	Reta	Onc112	Start avg	R untreated	O untreated	Untreated avg	ORF	Adjacent genes	Orientation
175048	175095	+	3.771243514	yadX	MERASKMPSSYLYDQ*	15	TAA	ATG	5	5	5	clcA	clcA	clcA	ORF15	heml-clcA	<>
290275	290370	–	6.253700191	argL	MNNYTYKVNFNSISGVRHARIKCPIYTKNTF*	31	TGA	ATG	18	15	17	argF	argF	argF	ORF51	argF-ykgS	<(<
391592	391648	–	8.436479081	yitB	MIGINVAIFTNVQRRTVS*	18	TAA	ATG	12	5	9	157	34	95	ORF14	yaiT-insF1	>(<)<
754658	754690	–	16.25839249	ybgV	MVDKFNNSDC*	10	TAA	GTG	49	22	35	44	31	38	ORF3	gltA-sdhC	<<>
837503	837562	+	18.04320962	ybiE	MRLSANRQGNTCAGRSVNA*	19	TGA	ATG	4	8	6	16	6	11	ORF31	ybiC-ybiJ	>>>
922740	922814	+	19.87956012	yljB	MNQKFEAVNAIDRNVTDVADANDR*	24	TGA	ATG	11	1	6	19	3	11	ORF35	cspD-clpS	<<>
1298514	1298567	+	27.97525536	ychT	MLSKGVLSARLFNLIYG*	17	TAA	ATG	186	25	106	93	10	51	ORF16	adhE-ychE	<<>
1344647	1344775	+	28.96914719	ymiD	MQKLPLKEKCLTATANYHPGIRYIMTGYSAKYIYSSTYARFR*	42	TAG	ATG	12	10	11	1	4	3	ORF37	yciZ-pdeN/gmr	<<>
1407989	1408036	+	30.33379064	ynaN	MLRNRARRLDYKIMR*	15	TGA	ATG	25	16	21	ydaN	ydaN	ydaN	ORF29	ydaM-ydaN	<<>
1491876	1491944	+	32.14105668	yncP	MNLLVKCAGKIPALALTWTCRP*	22	TGA	ATG	77	532	304	45	15	30	ORF21	ydcA-hokB	>>)<
1622741	1622845	+	34.96041926	mgtT	MNGDNPSPNRPLVTVVYKGPDFYDGEKKPPVNRR*	34	TGA		By eye			By eye			ORF61	mgtS-mgrR	>(>)<
1640805	1640975	+	35.34959105	ynfU	MSERKNSKSRRNYLVKCSCPNCTQESEHSFSRVQKGALLICPHCN KVFQTNLKAVA*	56	TGA	ATG	8	28	18	9	18	14	ORF38	essQ-cspB	<<>
1673608	1673709	–	36.05630064	ydgV	MDNLFRTLFSTFTHLRTSSTILLVGEQHWRNAL*	33	TAG	ATG	7	11	9	367	70	219	ORF4	mdtU-tqsA	<<>
1961837	1961881	–	42.2659217	yecV	MSIFRIHLDGNKKA*	14	TGA	ATG	19	20	20	14	0	7	ORF49	argS-yecT	>>>
2347350	2347385	–	50.57143448	yoaM	MSVSCWEFNAG*	11	TAA	ATG	15	31	23	14	0	7	ORF48	Antisense to nrdB	(<)>
2470497	2470643	–	53.22452006	yodE	MGNDAFKLSSADRGDITINNESGHLIVNTAILSGDIVTLRGGEIRLVL*	48	TAG	ATG	29	1	15	1329	66	697	ORF1	gtrS-tfaS	><>
2483729	2483758	+	53.50959098	evgL	MLHCKGNNL*	9	TGA	ATG	25	21	23	evgA	evgA	evgA	ORF47	emrK-evgA	<>)>
2722702	2722743	+	58.65803813	pssL	MNREEMHCDVVKI*	13	TAA	ATG	362	198	280	pssA	pssA	pssA	ORF34	pka-pssA	>>)>
2725925	2725987	+	58.72747461	yfsS	MEVGKLGKPYPLLNLAYVGV*	20	TAA	ATG	15	14	15	39	12	25	ORF26	kgtP-rrlG	<<>
3086145	3086288	–	66.44807364	yagG	MNRCLLLNLSHRSGEDSFPALCISALHTCRCYTHLGASQDSRAGYSY*	47	TGA	ATG	43	128	85	73	130	101	ORF2	antisense to yagC	<<)>
3109317	3109400	–	66.98729246	yagH	MHAISLHSLAYRRFARNVSLSMLLLMR*	27	TGA	ATG	33	28	31	19	8	14	ORF22	speC-yagA	<<>
3111266	3111325	+	67.02928182	yqhJ	MKYLAFLIKGKTHHFDVGK*	19	TGA	GTG	4	29	16	316	78	197	ORF7	antisense to yghE	(>)<
3247668	3247751	+	69.96793383	yqiM	MLMYQTRRTYQNSNNIAVVHLLKPAWR*	27	TAA	GTG	16	5	11	23	37	30	ORF12	exuR-yqiA	<>)<
3573660	3573701	–	76.99112299	yhgO	MIYRELNRALAIL*	13	TAA	ATG	31	30	31	162	3	83	ORF6B	glgB-asd	<<)<<
3573679	3573708	–	76.99153233	yhgP	MFHDLPGVK*	9	TAG	ATG	29	59	44	168	17	92	ORF6A	glgB-asd	<(<<
3630665	3630778	–	78.21924177	yhiY	MMTTLLPVFTKPSPLALNALRAGRICRFLLIPDGRIR*	37	TGA	ATG	11	47	29	4	12	8	ORF9	yhiI-yhiJ	<<<
3640612	3640686	–	78.43354047	yriA	MLYWLGILIAYADRRPDKVFTLIR*	24	TAA	ATG	27	6	17	54	2	28	ORF50	uspA-dtpB	>(<)<>
3640643	3640696	–	78.43420834	yriB	MVLNVVLAGDFDCVRRP*	17	TAG	ATG	5	5	14	36	2	19	ORF52	uspA-dtpB	><(<>
3720400	3720471	+	80.15249743	ysaE	MRNAVSKAGIISRRRLLLFQFAG*	23	TAA	ATG	183	639	411	18	16	17	ORF32	cspA-hokA	>>)<
3737177	3737203	–	80.51394202	baxL	MIKSDQET*	8	TGA	ATG	343	171	257	bax	bax	bax	ORF46	bax-malS	<(<>
3797772	3797846	+	81.81940395	yibX-S	MENIFKPISYIASLAPREVTLLALV*	24	TAA	ATG	107	21	64	71	29	50	ORF11	Antisense to waaL	<<)>
3797772	3798014	+	81.81940395	yibX	MEKIPIRPFSDPASIISLCRCTLENSNAPLSLLCSATNKFILSARDSAD LNEKGDFMNIFKPISYIASLAPREVTLLALV*	80	TAA	GTG	10	10	10	43	22	32	ORF18	Antisense to waaL	<<)>
3798039	3798149	–	81.82515622	yibY	MIIGMLRAHMITSLSPIPTMPSVNINKPKARFLYAI*	36	TAA	ATG	13	1	7	7	14	10	ORF62	Antisense to waaL	(<<)>
3838642	3838701	–	82.69990943	yicU	MTSDYRVSPIKKSRPAMSK*	19	TAG	ATG	2	40	21	8	2	5	ORF44	Antisense to yicC	(<)>
3853766	3853840	–	83.0257417	ysdE	MNVSQIYARNGELFSGRICKQKRQ*	24	TAA	ATG	6	7	6	24	10	17	ORF43	tisB-emrD	><>
4001218	4001274	+	86.20245551	ysgD	MDTPSRYWLTILSSRINS*	18	TAA	TTG	1,008	924	966	333	102	218	ORF54	yigE-corA	>>>
4486105	4486155	+	96.64888708	yjgA	MSIIILATISTTQENK*	16	TGA	TTG	58	17	38	29	10	19	ORF57	pepA-lptF	<>>
4518372	4518440	+	97.34404966	ykiE	MGWSLCFWHVSVRMTHDVVCYY*	22	TGA	ATG	11	9	10	71	2	36	ORF63	fecI-insA7	<>>
4536597	4536695	+	97.73668944	ythB	MNKLPAHLSRQNCKIASTNLSEIIPRRAAVLK*	32	TGA	ATG	61	100	81	38	27	33	ORF20	Antisense to nanS	>>)<

[a]For each gene, we list the genomic coordinates (left and right for the left and right ends of the smORF on the genome), systematic gene name, minute position on the genome in E. coli MG1655 genome NC_000913.3), sense strand (+, plus; −, minus), start peak intensity in retapamulin (Reta)- or Onc112-treated samples (in rpm), candidate ORF number, peptide sequence and length, start and stop codons, and the names and orientation of neighboring genes (parentheses indicate overlap with the adjacent gene). The levels of normal ribosome profiling in untreated controls matched with the retapamulin (R)- and Onc112 (O)-treated samples (in rpkm). The levels of normal ribosome profiling in untreated controls cannot be determined if the smORF overlaps an annotated gene; in these cases, the name of the overlapping gene is given instead of an rpkm value. MgtT was discovered by visual inspection of a genome browser.

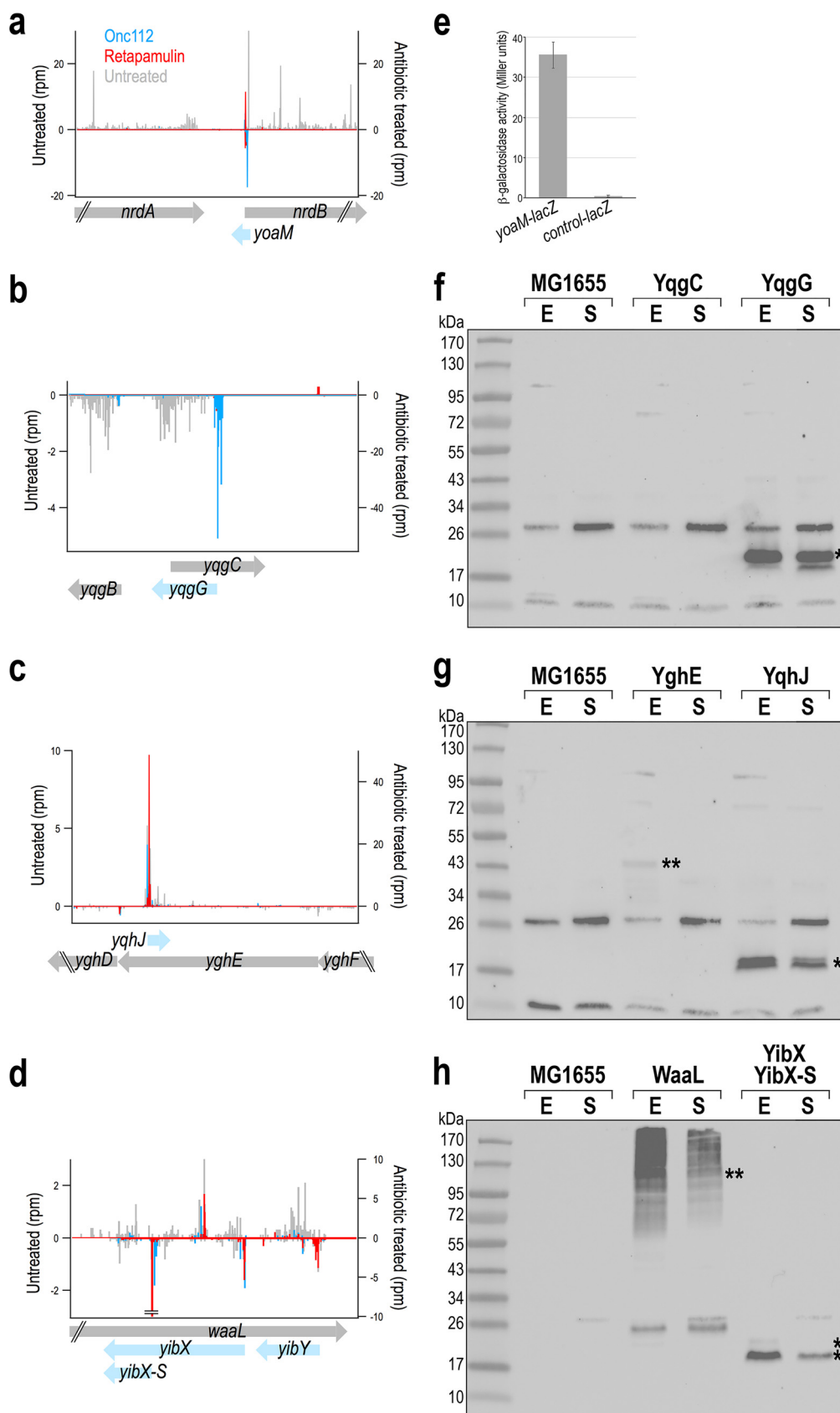**FIG 5** Novel smORFs (blue) are encoded antisense to known genes (gray). (a to d) Gene organization for the *nrdB-yoaM* (a), *yqgC-yqgG* (b), *yghE-yqhJ* (c), and *waaL-yibX-yibY* (d) loci. β-Galactosidase activity was assayed for cells carrying

YibX is translated as two isoforms. The *yibX* gene was also interesting as the profiling data suggested translation could initiate from two different start codons. While most bacterial ORFs encode a single protein, there are some examples where different isoforms of the same protein are generated by different translation starts in the same frame, as has been found for the *E. coli* proteins ClpB, IF-2, and MrcB (46–48). Frequently, the longer polypeptide is expressed at higher levels than the shorter isoform. A broad peak near the start codon for the ribosome profiling data suggests that several small proteins are potentially translated as different isoforms. Although most of the potential isoforms vary by only a few codons and would be indistinguishable on immunoblots, the YibX alternative start sites lead to proteins of substantially different sizes. The stronger signal corresponds to the 24-amino-acid (aa) YibX-S protein, while a second signal at a GTG codon upstream and in frame with YibX-S yields an 80-aa protein, adding ~6.1 kDa (Fig. 5h). Both bands are detected in Fig. 3 and 5, but in contrast to other known primary isoforms, the 80-amino-acid protein is detected at lower levels than the shorter isoform. A second protein for which there are possible isoforms is YqhJ, which shows two bands in Fig. 5g. YqhJ initiates at a GTG codon and is 19 residues long; initiation at a downstream TTG codon would yield a 13-residue protein (Table S3). Ribosome density in Onc11- and retapamulin-treated samples is consistent with both of these initiation sites being used (data not shown).

Multiple smORFs are encoded by different, overlapping frames. There are a growing number of bacterial examples where more than one protein is encoded in the same region in different frames, as has been found for *rzoD* encoded within *rzpD*, which are homologous to the rz/rz1 lysis cassette of bacteriophage λ (49, 50). A similar gene arrangement of nested start codons and substantial overlap is also found for two sets of newly identified small proteins: YhgO/YhgP (Fig. 6a) and YriA/YriB (Fig. 6b). Additionally, the smORFs encoding two other new proteins, YbgV and MgtT, overlap the 3′ ends of the previously identified smORFs *ybgU* (Fig. 6c) and *mgtS* (Fig. 6d), respectively. We sought to compare the levels of the paired small proteins under the same conditions by assaying cells with one or the other smORF tagged (Fig. 6e to h). Although there generally appears to be limited correlation between ribosome density and observed protein levels, for each of these pairs, the small protein corresponding to the smORF with the higher ribosome density with either Onc112 or retapamulin treatment (YhgP, YriB, YbgV, and MgtS) was present at higher levels. Perhaps there is a better correlation between ribosome density and observed protein levels for cotranscribed genes.

smORFs overlap the 5′ ends of larger protein-coding genes. The genes of several new small proteins detected by immunoblot analysis (Fig. 3) were found to overlap the 5′ ends of annotated larger genes in a different frame including *baxL-baxA*, *evgL-evgA*, and *arglL-argF*. Two additional smORFs predicted by ribosome profiling, *ORF33* and *pssL*, also overlap the 5′ end of the neighboring gene in a different frame, but we were unable to SPA tag these predicted proteins because the downstream genes, *accD* (acetyl-CoA carboxyltransferase subunit β) (Fig. 7a) and *pssA* (phosphatidylserine synthase) (Fig. 7b), are essential. To investigate the expression of ORF33 and PssL, the 5′ UTR and the first few codons of the smORFs were translationally fused to *lacZ* on the chromosome (40). While there was no measurable β-galactosidase activity for the *ORF33-lacZ* fusion (Fig. 7f), there was clear expression of the *pssL-lacZ* fusion, which was diminished by the introduction of a stop codon at the start codon position (Fig. 7g).

FIG 5 Legend (Continued)
chromosomal fusions of the 5′ UTR and initial codons of *yoaM* fused to *lacZ* as well as out-of-frame control fusion (e), which were grown in rich media (LB) with 0.2% arabinose. (f to h) Protein levels for chromosomally SPA-tagged *yqgC* and *yqgG* (f), *yghE* and *yqhJ* (g) and *waaL* and *yibX* (h) genes. Gel samples were prepared from MG1655 strains grown to exponential (E) and stationary (S) phase in LB. Immunoblot analysis was conducted as described in the legend to Fig. 3 with MG1655 as a negative control. Bands corresponding to small proteins are marked with an asterisk, and bands corresponding to antisense-encoded larger proteins are marked with two asterisks.
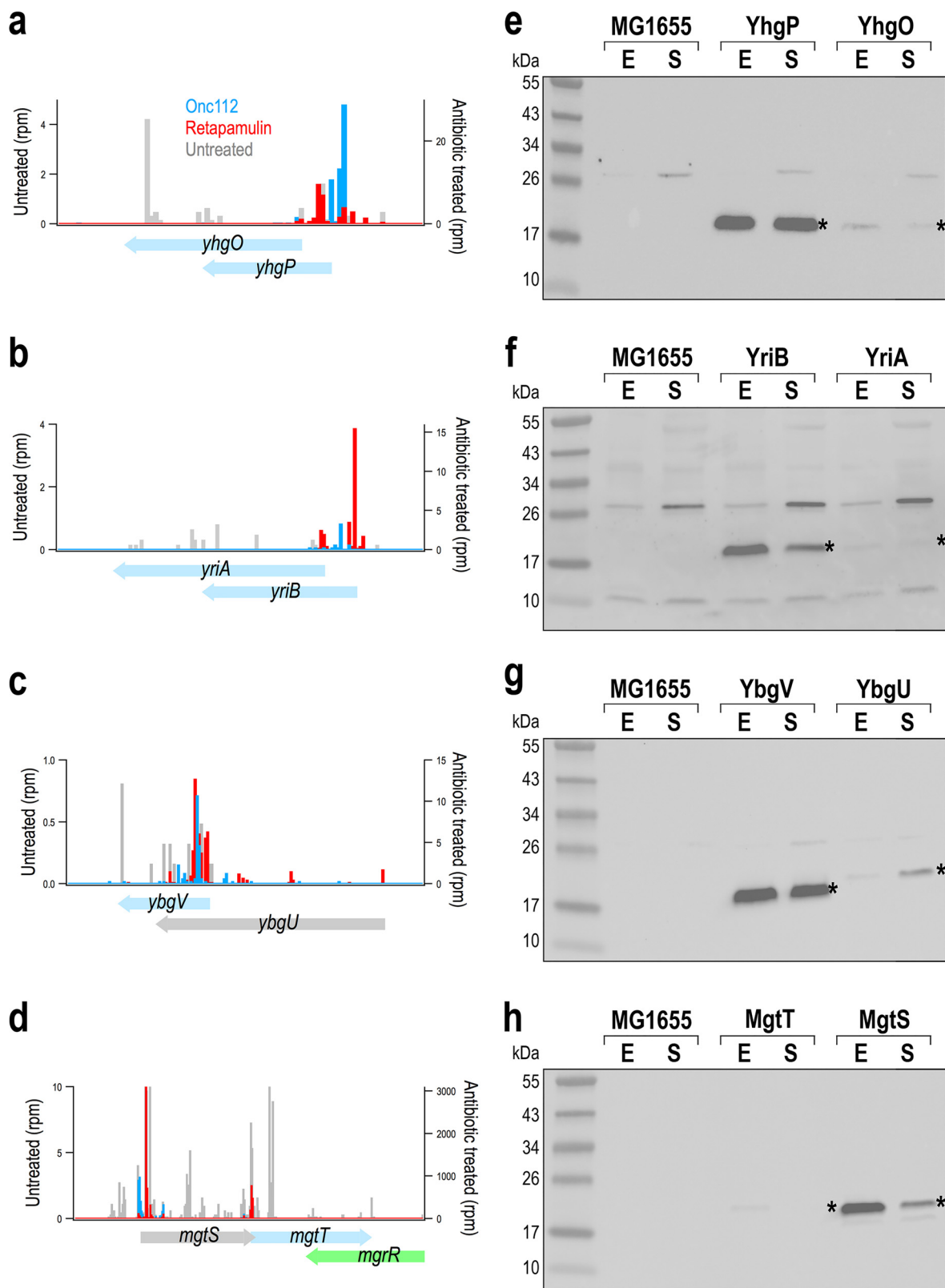
FIG 6 smORFs are found in complex gene arrangements. (a to d) Gene organization for *yhgO/yhgP* (a), *yriA/yriB* (b), *ybgU/ybgV* (c), and *mgtS/mgtT* (d), with previously identified small protein genes in gray, newly identified small protein genes in blue, and small RNA gene *mgrR* in green. (e to h) Levels of corresponding proteins. Gel samples were prepared from MG1655 strains grown to exponential (E) and stationary (S) phase in LB. Immunoblot analysis was conducted as described in the legend to Fig. 3 with MG1655 as a negative control. Bands corresponding to small proteins are marked with an asterisk.
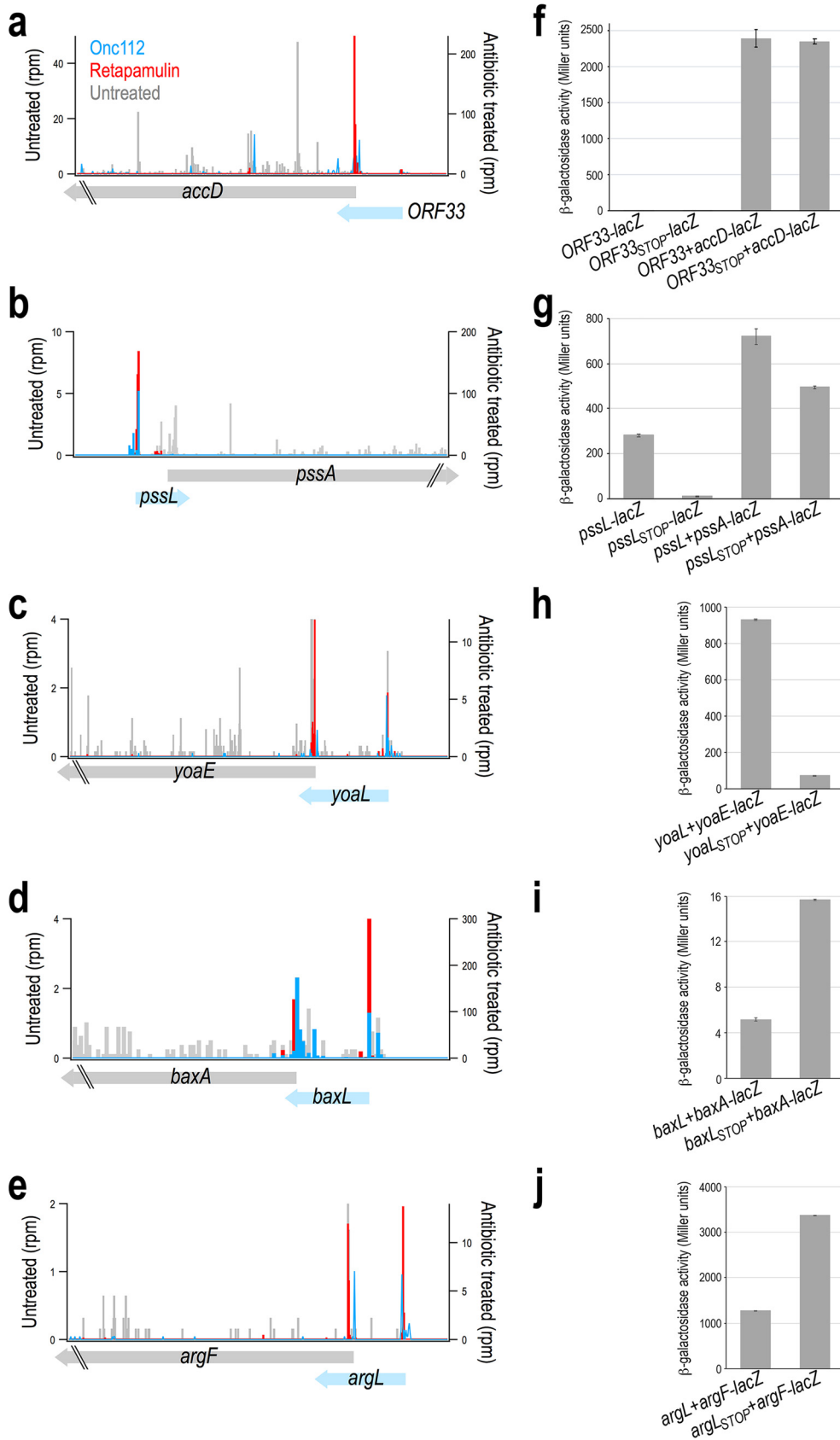
FIG 7 smORFs regulate expression of downstream genes. (a to e) Organization of smORFs (blue) in 5′ UTRs of known genes (gray). β-Galactosidase activity was assayed for cells carrying chromosomal fusions of the 5′ UTR and initial

(Continued on next page)

These results indicate that although we could not construct a *pssL-SPA* fusion at the endogenous location of the genome, the protein is translated.

Role of smORFs regulating expression of larger protein encoded downstream. Given other examples where smORFs overlapping downstream genes serve as leader peptides involved in modulating the translation of the larger gene (51; reviewed in reference 52), we next sought to investigate whether translation of the smORFs overlapping larger ORFs described above affects translation of the downstream ORF. To test this, the entire 5′ UTR, including the smORF together with the first codons of the downstream gene was fused to *lacZ* at the endogenous *lacZ* locus. We also generated a second version of these constructs by introducing amber or ochre stop codons into the smORF as a replacement for the start codon. If translation of the two ORFs is coupled, the stop codon, which blocks the expression of the upstream smORF, should impact translation of the downstream gene. In the case of the *ORF33-accD* pair, for which we did not see any expression of ORF33, the stop codon had no impact on *accD-lacZ* expression (Fig. 7f). In contrast, introduction of a stop codon into *pssL* led to a 30% decrease in the expression of the *pssA-lacZ* fusion (Fig. 7g), while introduction of a stop codon into *yoaL*, a recently identified smORF (17), led to strongly decreased expression of *yoaE-lacZ* (Fig. 7h). An increase in the expression of the downstream gene is observed when stop codons are introduced into *baxL* (Fig. 7i) and *argL* (Fig. 7j). Together, these results indicate that translation of these upstream smORFs may be playing a regulatory role.

DISCUSSION

Fundamentally, the challenge of identifying expressed small proteins stems from the great number of putative smORFs, with ~161,000 possible smORFs in intergenic regions of *E. coli* alone. The key question is how best to identify and validate candidate smORFs in a manner that prevents the annotation of uncorroborated genes. Rather than relying solely on bioinformatic approaches, as has been done previously, we demonstrated that an approach that utilizes multiple ribosome profiling data sets can identify translated smORFs with a high degree of accuracy. The expression of 36 of these smORFs was verified by immunoblot analysis of the chromosomally tagged genes, and the expression of two other genes that could not be tagged at the endogenous loci was observed as chromosomal *lacZ* fusions. We noted a number of interesting gene arrangements, including small proteins encoded on the strand opposite larger, annotated proteins, as well as smORFs in the 5′ UTRs of known genes.

Limitations of approach. While we were able to identify many new small proteins, we are cognizant of some limitations. One important caveat is that start codon peak intensity in profiling experiments is not a truly quantitative measure of initiation rates, given that reads at a single site are prone to sequence-specific artifacts (53). Examination of the profiling data of previously identified smORFs illustrates this limitation, as there is not a strong correlation between ribosome density in the presence of the initiation complex inhibitors and the band intensity observed by immunoblot analysis. While the degradation of some tagged small proteins may explain ribosome density without corresponding protein bands, other smORFs yield strong bands without any sequencing reads in the profiling experiments. Determining the factors that contribute to the perceived mismatch between ribosome density and observed protein levels would allow for a more accurate prediction of expression. It also must be considered that, although only occurring for a short duration, treatment with Onc112 or retapamulin represents stress on the bacteria that can cause changes in the expression profile.

FIG 7 Legend (Continued)
codons of *ORF33* (f) and *pssL* (g) fused to *lacZ*. β-Galactosidase activity was assayed for cells carrying *lacZ* chromosomal fusions to the 5′ UTR and initial codons of the downstream gene with a wild-type start codon for the upstream smORF or with a stop codon replacing the start codon (f to j). For all β-galactosidase assays, cells were grown in LB with 0.2% arabinose.

One other major limitation regarding the general application of this approach is that the microbes must be susceptible to these initiation complex inhibitors. Retapamulin is a member of the pleuromutilin class of antibiotics that show activity against a broad spectrum of Gram-positive bacteria, though some derivatives show activity against Gram-negative bacteria as well (54, 55). To increase susceptibility to retapamulin, the group of Vázquez-Laslop and Mankin (37) used a *tolC* mutant strain of *E. coli*, an approach that may need to be employed in other bacteria. Onc112, a member of the PrAMP family of peptide antibiotics, is actively transported into Gram-negative bacteria by proteins such as the SbmA transporter (56). It may be possible to extend the range of compounds like Onc112 by exogenous expression of transporters such as SbmA in bacteria that otherwise lack them.

Advantages of approach. Despite the possible limitations, the ability to identify start codons through ribosome profiling with inhibitors is a powerful approach with broad applications. As shown here, translated smORFs are more prevalent than previously believed and are found in contexts that would be difficult to distinguish by other methods, including bioinformatic approaches that have been successfully employed previously (16, 17). While traditional ribosome profiling can guide the prediction (57, 58) or, in conjunction with experiments to verify protein synthesis, even support the annotation of intergenic smORFs in bacteria (27), ribosome profiling with stalled initiation complexes allows for the identification of protein-coding sequences in contexts that are generally ignored, including within or overlapping other genes as shown here and by the group of Vázquez-Laslop and Mankin (37). These new, internal altORFs may represent new classes of functional and regulatory proteins that comprise an ever-expanding proteome.

Interestingly, we noted relatively poor overlap between our predicted smORFs and those reported in the other ribosome profiling studies (27, 33), suggesting that many small proteins remain to be discovered. Of the 328 smORFs predicted by Mori and coworkers (33) in intergenic regions in *E. coli* based on ribosome enrichment at start codons after treatment with tetracycline, only 20 overlap with our list of 68 likely candidates (Table S3). The fact that Onc112 and retapamulin are more specific than tetracycline for newly initiated ribosomes, providing higher resolution for start codon identification, may partially explain the limited overlap with our predicted smORFs. We also looked for overlap between our 68 likely candidates and the 130 smORFs predicted in *Salmonella enterica* in a recent study using traditional ribosome profiling (27). Only one exact match and three close matches were found between these related species.

In addition to facilitating the identification of new smORFs, the profiling data with inhibitors provide valuable information about known ORFs and suggest the need to reannotate some genes (see Table S1 in the supplemental material). One example is the smORF *ymiA*, which is annotated both as beginning with MLISDGDYMRLAMPSGNQEP (59) and as beginning with the third methionine at MPSGNQEP (16) but likely initiates with MRLAMPSGNQEP (Fig. S3). Another example is the *ymdG* protein. Although it is annotated as 40 residues, our data show that a later start codon is used and that the smORF is only eight codons long (Fig. S3). Finally, *yoaL*, which was herein examined for function as a leader peptide (Fig. 7), was originally annotated as initiating on a methionine 13 codons upstream of its likely start site (17) (Fig. S3). Our data provide the first experimental evidence of where translation begins in these three smORFs, but we cannot rule out the possibility that alternate start sites are used under different biological conditions.

Small protein function. Many of the small proteins are expected to have functions that involve the binding to other, larger proteins. However, the primary structures of the small proteins are often too short for bioinformatic tools to identify motifs or domains that may offer insights into their functions in the cell. Of the newly identified small proteins, only YnfU, which is encoded within the Qin prophage region of the *E. coli* genome, had an identifiable motif. The protein contains a pair of zinc knuckles, a

motif with two copies of the CPXC sequence that together chelate a zinc ion (reviewed in reference 60). Homology modeling of YnfU using PSIPRED (61) also revealed a moderate match to the zinc-binding domain of PA0128, a protein of unknown function from *Pseudomonas aeruginosa*.

Although motif identification is often not available for smORFs, multiple previously identified proteins were predicted to contain transmembrane helices and were later experimentally shown to localize to the cellular membrane (16). When we examined the sequences of the new smORFs using the Phobius or ExPASy TMpred algorithms (62, 63), none of the newly identified proteins were predicted to contain transmembrane helices. This analysis shows that the skew toward hydrophobic α-helices overall is not as strong as observed for the small *E. coli* proteins identified in the first systematic search for these proteins (16). In general, the next challenge will be to determine functions for the large numbers of newly identified proteins.

Four new smORFs and one previously annotated smORF were examined for possible roles as leader peptides, as these small protein genes overlap the downstream coding sequences of larger proteins in alternate frames (Fig. 7). For each of the expressed genes, either an increase or decrease in the translation of the downstream gene was observed when the upstream smORF was not translated. For genes where expression decreases, this drop may stem from a loss of translational coupling from the upstream gene, while for genes with improved expression, translation of the smORF may impede translation of the downstream gene. It is interesting to note that a mutation (*pssR1*) that leads to increased expression of *pssA* mapped to the anti-Shine-Dalgarno sequence of the 16S rRNA encoded by *rrnC* (64). Further characterization will be required to distinguish smORFs that are simply translated in operons versus those that specifically serve to control the translation of downstream genes and to elucidate the regulatory mechanisms.

Complex gene organization. Beyond the expanded presence of smORFs as possible upstream leaders of other genes, our analysis also pointed to other forms of complex gene organization. We found several smORFs that overlap other new or known smORFs. We also discovered small proteins encoded antisense to larger proteins, as well as at least one small protein that is translated as two isoforms. We hypothesize that the pairs of bacterial genes encoded in overlapping regions have related functions.

Since we think we have not yet identified the complete set of small protein genes, we suggest that antisense genes and translational regulation by upstream smORFs may be far more prevalent than currently thought. Full annotation of translated regions of the chromosome will be required to obtain a more comprehensive picture of cellular regulation. Additionally, more complete annotation of translation will provide a better understanding of the roles of the many seemingly orphan transcription start sites observed in transcriptome data (45). The use of ribosome profiling with initiation complex inhibitors revealed 38 new protein-coding genes in *E. coli*, an organism already known to express nearly 100 small proteins. For less-well-characterized bacteria, the ability to define the small proteome accurately and in an unbiased manner opens new doors to uncovering the regulation that allows the growth and survival of these organisms.

MATERIALS AND METHODS

Onc112 ribosome profiling. A culture of *E. coli* MG1655 was grown overnight at 37°C in MOPS EZ Rich Defined media (Teknova) with 0.2% glucose, diluted 1:100 into 150 ml of fresh medium, and grown to an optical density at 600 nm (OD$_{600}$) of 0.3. The culture was treated with 50 μM Onc112 for 10 min and harvested by rapid filtration and freezing in liquid nitrogen. Ribosome profiling libraries were prepared and sequenced as previously described (65) with the following modifications. Normally, the standard lysis buffer contains chloramphenicol to arrest translation in the lysate. We omitted chloramphenicol and added 1 M NaCl to the lysis buffer because we have found that high salt concentrations arrest translation better than chloramphenicol. Use of high-salt buffers necessitates a buffer exchange prior to nuclease digestion: 25 AU of RNA in the lysate was pelleted over a 1-ml sucrose cushion (20 mM Tris [pH 7.5], 500 mM NH$_4$Cl, 0.5 mM EDTA, 1.1 M sucrose) using a TLA 100.3 rotor at 65,000 rpm for 2 h. Pellets were resuspended in 200 μl of the standard lysis

buffer, and the RNA was digested with MNase following the standard protocol. We anticipate that the standard protocol for harvesting cells and preparing libraries would give equally good results after Onc112 or retapamulin treatment.

Analysis of ribosome profiling data. Raw reads were filtered and trimmed using Skewer v0.2.2. Reads were mapped uniquely to the *E. coli* MG1655 genome NC_000913.3 (allowing two mismatches) using Bowtie v 0.12.7 after reads mapping to tRNA and rRNA were discarded. Ribosome density was assigned to the 3′ ends of reads. We identified novel open reading frames eight sense codons or longer starting with ATG, GTG, or TTG codons at least 18 nt away (on either side) from any annotated genes. For each potential site, the ribosome density in Onc112- or retapamulin-treated samples was summed 0 to 18 nt downstream of the first nucleotide in the start codon to calculate the initiation peak intensity. Note that a single peak of Onc112 or retapamulin density may correspond to multiple start codons in the 18-nt window; this redundancy was eliminated by inspecting the top candidates in a genome browser and looking for the optimal spacing as described in the text and Fig. 2. We also calculated rpkm values for normal ribosome profiling data for each candidate smORF unless any part of it comes within 15 nt of an annotated gene. The candidate smORFs and their scores are reported in Table S2 in the supplemental material. The retapamulin treatment data can be found at GSE122129 and the Onc112 treatment data can be found at GSE123675. Our code is available at https://github.com/greenlabjhmi.

Strain construction. All strains generated for this study are listed in Table S4 together with the sequences of the oligonucleotides used to construct the strains. smORFs were tagged on the chromosome following published procedures (66). In short, an SPA-kan cassette was inserted at the C-terminal end of each ORF using the λ Red recombination system in *E. coli* NM400 and moved into *E. coli* MG1655 by P1 transduction. All insertions were verified by sequencing.

Construction of the *lacZ* reporter strains followed a published procedure (40). Briefly, DNA including the 5′ UTR and several codons of each ORF, along with flanking homology regions, were transformed into *E. coli* PM1205, which utilizes the λ Red-mediated recombination system, and selected for sucrose resistance. All insertions were verified by sequencing.

Immunoblot analysis. For all expression experiments, Luria broth (LB) was inoculated 1:200 with overnight culture of various strains and grown at 37°C. One milliliter of culture was taken during exponential growth (2 h postinoculation, OD_{600} of 0.5 to 0.7) and during stationary phase (3.5 h postinoculation, OD_{600} of 2.5 to 3). To normalize for total cells (number/density/count), the cell pellet collected for each sample was resuspended according to the OD_{600}. Samples were analyzed by SDS-PAGE, transferred to nitrocellulose membranes, and blotted using anti-FLAG(M2)-HRP (Sigma).

Assays of β-galactosidase activity. For all experiments, LB with 0.2% arabinose was inoculated 1:200 with overnight culture of PM1205 strains carrying various *lacZ* fusions. These cultures were grown at 37°C for 2.25 h (OD_{600} of 0.75 to 1.0). Culture (10, 50, or 100 μl depending on the sample) was added directly to Z buffer (800-μl total volume) in 1.5-ml microcentrifuge tubes. SDS (0.00184%) and chloroform (3.5% vol/vol) were added, and samples were vortexed for 30 s. The samples were incubated at 28°C for 15 min before the addition of ortho-nitrophenyl-β-galactoside (ONPG) (0.875 mg/ml). Incubation at 28°C continued until a visible color change occurred, at which time sodium carbonate (353 mM) was added to quench the reaction. All reactions were quenched by 75 min, even if no color change was observed. Samples were centrifuged at maximum speed in a table-top microcentrifuge ($\sim$21,000 $\times$ *g*) for 2 min. The absorbance at 550 nm and 420 nm was measured for 1 ml of supernatant, and Miller units were calculated using the established formula (67).

SUPPLEMENTAL MATERIAL

Supplemental material for this article may be found at https://doi.org/10.1128/mBio .02819-18.

FIG S1, TIF file, 1.5 MB.
FIG S2, TIF file, 2 MB.
FIG S3, JPG file, 0.2 MB.
TABLE S1, XLSX file, 0.02 MB.
TABLE S2, XLSX file, 9.5 MB.
TABLE S3, XLSX file, 0.04 MB.
TABLE S4, XLSX file, 0.02 MB.

REFERENCES

1. Storz G, Wolf YI, Ramamurthi KS. 2014. Small proteins can no longer be ignored. Annu Rev Biochem 83:753–777. https://doi.org/10.1146/annurev-biochem-070611-102400.
2. Andrews SJ, Rothnagel JA. 2014. Emerging evidence for functional peptides encoded by short open reading frames. Nat Rev Genet 15:193–204. https://doi.org/10.1038/nrg3520.
3. Saghatelian A, Couso JP. 2015. Discovery and characterization of smORF-encoded bioactive polypeptides. Nat Chem Biol 11:909–916. https://doi.org/10.1038/nchembio.1964.
4. Hobbs EC, Yin X, Paul BJ, Astarita JL, Storz G. 2012. Conserved small protein associates with the multidrug efflux pump AcrB and differentially affects antibiotic resistance. Proc Natl Acad Sci U S A 109:16696–16701. https://doi.org/10.1073/pnas.1210093109.
5. Wang H, Yin X, Wu Orr M, Dambach M, Curtis R, Storz G. 2017. Increasing intracellular magnesium levels with the 31-amino acid MgtS protein. Proc Natl Acad Sci U S A 114:5689–5694. https://doi.org/10.1073/pnas.1703415114.
6. Yin X, Wu Orr M, Wang H, Hobbs EC, Shabalina SA, Storz G. 2019. The small protein MgtS and small RNA MgrR modulate the PitA phosphate symporter to boost intracellular magnesium levels. Mol Microbiol 111:131–144. https://doi.org/10.1111/mmi.14143.
7. Bal NC, Maurya SK, Sopariwala DH, Sahoo SK, Gupta SC, Shaikh SA, Pant M, Rowland LA, Bombardier E, Goonasekera SA, Tupling AR, Molkentin JD, Periasamy M. 2012. Sarcolipin is a newly identified regulator of muscle-based thermogenesis in mammals. Nat Med 18:1575–1579. https://doi.org/10.1038/nm.2897.
8. Anderson DM, Anderson KM, Chang CL, Makarewich CA, Nelson BR, McAnally JR, Kasaragod P, Shelton JM, Liou J, Bassel-Duby R, Olson EN. 2015. A micropeptide encoded by a putative long noncoding RNA regulates muscle performance. Cell 160:595–606. https://doi.org/10.1016/j.cell.2015.01.009.
9. Goffeau A, Barrell BG, Bussey H, Davis RW, Dujon B, Feldmann H, Galibert F, Hoheisel JD, Jacq C, Johnston M, Louis EJ, Mewes HW, Murakami Y, Philippsen P, Tettelin H, Oliver SG. 1996. Life with 6000 genes. Science 274:546–567. https://doi.org/10.1126/science.274.5287.546.
10. Burge CB, Karlin S. 1998. Finding the genes in genomic DNA. Curr Opin Struct Biol 8:346–354. https://doi.org/10.1016/S0959-440X(98)80069-9.
11. Congdon RW, Muth GW, Splittgerber AG. 1993. The binding interaction of Coomassie blue with proteins. Anal Biochem 213:407–413. https://doi.org/10.1006/abio.1993.1439.
12. Scopes RK. 1994. Protein purification principles and practice, 3rd ed. Springer-Verlag, New York, NY.
13. Burgess R, Deutscher M (ed). 2009. Guide to protein purification. Academic Press, San Diego, CA.
14. Slavoff SA, Mitchell AJ, Schwaid AG, Cabili MN, Ma J, Levin JZ, Karger AD, Budnik BA, Rinn JL, Saghatelian A. 2013. Peptidomic discovery of short open reading frame-encoded peptides in human cells. Nat Chem Biol 9:59–64. https://doi.org/10.1038/nchembio.1120.
15. Pueyo JI, Magny EG, Couso JP. 2016. New peptides under the s(ORF)ace of the genome. Trends Biochem Sci 41:665–678. https://doi.org/10.1016/j.tibs.2016.05.003.
16. Hemm MR, Paul BJ, Schneider TD, Storz G, Rudd KE. 2008. Small membrane proteins found by comparative genomics and ribosome binding site models. Mol Microbiol 70:1487–1501. https://doi.org/10.1111/j.1365-2958.2008.06495.x.
17. VanOrsdel CE, Kelly JP, Burke BN, Lein CD, Oufiero CE, Sanchez JF, Wimmers LE, Hearn DJ, Abuikhdair FJ, Barnhart KR, Duley ML, Ernst SEG, Kenerson BA, Serafin AJ, Hemm MR. 2018. Identifying new small proteins in Escherichia coli. Proteomics 18:e1700064. https://doi.org/10.1002/pmic.201700064.
18. Ladoukakis E, Pereira V, Magny EG, Eyre-Walker A, Couso JP. 2011. Hundreds of putatively functional small open reading frames in Drosophila. Genome Biol 12:R118. https://doi.org/10.1186/gb-2011-12-11-r118.
19. Kessler MM, Zeng Q, Hogan S, Cook R, Morales AJ, Cottarel G. 2003. Systematic discovery of new genes in the Saccharomyces cerevisiae genome. Genome Res 13:264–271. https://doi.org/10.1101/gr.232903.
20. Landry CR, Zhong X, Nielly-Thibault L, Roucou X. 2015. Found in translation: functions and evolution of a recently discovered alternative proteome. Curr Opin Microbiol 32:74–80. https://doi.org/10.1016/j.sbi.2015.02.017.
21. Mouilleron H, Delcourt V, Roucou X. 2016. Death of a dogma: eukaryotic mRNAs can code for more than one protein. Nucleic Acids Res 44:14–23. https://doi.org/10.1093/nar/gkv1218.
22. Vanderperre B, Lucier JF, Bissonnette C, Motard J, Tremblay G, Vanderperre S, Wisztorski M, Salzet M, Boisvert FM, Roucou X. 2013. Direct detection of alternative open reading frames translation products in human significantly expands the proteome. PLoS One 8:e70698. https://doi.org/10.1371/journal.pone.0070698.
23. Hellens RP, Brown CM, Chisnall MA, Waterhouse PM, Macknight RC. 2016. The emerging world of small ORFs. Trends Plant Sci 21:317–328. https://doi.org/10.1016/j.tplants.2015.11.005.
24. Neuhaus K, Landstorfer R, Simon S, Schober S, Wright PR, Smith C, Backofen R, Wecko R, Keim DA, Scherer S. 2017. Differentiation of ncRNAs from small mRNAs in Escherichia coli O157:H7 EDL933 (EHEC) by combined RNAseq and RIBOseq – ryhB encodes the regulatory RNA RyhB and a peptide, RyhP. BMC Genomics 18:216. https://doi.org/10.1186/s12864-017-3586-9.
25. Bazzini AA, Johnstone TG, Christiano R, Mackowiak SD, Obermayer B, Fleming ES, Vejnar CE, Lee MT, Rajewsky N, Walther TC, Giraldez AJ. 2014. Identification of small ORFs in vertebrates using ribosome footprinting and evolutionary conservation. EMBO J 33:981–993. https://doi.org/10.1002/embj.201488411.
26. Aspden JL, Eyre-Walker YC, Phillips RJ, Amin U, Mumtaz MA, Brocard M, Couso JP. 2014. Extensive translation of small open reading frames revealed by Poly-Ribo-Seq. Elife 3:e03528. https://doi.org/10.7554/eLife.03528.
27. Baek J, Lee J, Yoon K, Lee H. 2017. Identification of unannotated small genes in Salmonella. G3 7:983–989. https://doi.org/10.1534/g3.116.036939.
28. Ingolia NT, Brar GA, Stern-Ginossar N, Harris MS, Talhouarne GJ, Jackson SE, Wills MR, Weissman JS. 2014. Ribosome profiling reveals pervasive translation outside of annotated protein-coding genes. Cell Rep 8:1365–1379. https://doi.org/10.1016/j.celrep.2014.07.045.
29. Guttman M, Russell P, Ingolia NT, Weissman JS, Lander ES. 2013. Ribosome profiling provides evidence that large noncoding RNAs do not encode proteins. Cell 154:240–251. https://doi.org/10.1016/j.cell.2013.06.009.
30. Fields AP, Rodriguez EH, Jovanovic M, Stern-Ginossar N, Haas BJ, Mertins P, Raychowdhury R, Hacohen N, Carr SA, Ingolia NT, Regev A, Weissman JS. 2015. A regression-based analysis of ribosome-profiling data reveals a conserved complexity to mammalian translation. Mol Cell 60:816–827. https://doi.org/10.1016/j.molcel.2015.11.013.
31. Ingolia NT, Lareau LF, Weissman JS. 2011. Ribosome profiling of mouse embryonic stem cells reveals the complexity and dynamics of mammalian proteomes. Cell 47:789–802. https://doi.org/10.1016/j.cell.2011.10.002.
32. Lee S, Liu B, Lee S, Huang SX, Shen B, Qian SB. 2012. Global mapping of translation initiation sites in mammalian cells at single-nucleotide resolution. Proc Natl Acad Sci U S A 109:E2424–E2432. https://doi.org/10.1073/pnas.1207846109.
33. Nakahigashi K, Takai Y, Kimura M, Abe N, Nakayashiki T, Shiwa Y, Yoshikawa H, Wanner BL, Ishihama Y, Mori H. 2016. Comprehensive identification of translation start sites by tetracycline-inhibited ribosome profiling. DNA Res 23:193–201. https://doi.org/10.1093/dnares/dsw008.
34. Seefeldt AC, Nguyen F, Antunes S, Pérébaskine N, Graf M, Arenz S, Inampudi KK, Douat C, Guichard G, Wilson DN, Innis CA. 2015. The proline-rich antimicrobial peptide Onc112 inhibits translation by blocking and destabilizing the initiation complex. Nat Struct Mol Biol 22:470–475. https://doi.org/10.1038/nsmb.3034.
35. Gagnon MG, Roy RN, Lomakin IB, Florin T, Mankin AS, Steitz TA. 2016. Structures of proline-rich peptides bound to the ribosome reveal a common mechanism of protein synthesis inhibition. Nucleic Acids Res 44:2439–2450. https://doi.org/10.1093/nar/gkw018.
36. Dornhelm P, Högenauer G. 1978. The effects of tiamulin, a semisynthetic pleuromutilin derivative, on bacterial polypeptide chain initiation. Eur J Biochem 91:465–473. https://doi.org/10.1111/j.1432-1033.1978.tb12699.x.
37. Meydan S, Marks J, Klepacki D, Sharma V, Baranov PV, Firth AE, Margus T, Kefi A, Vázquez-Laslop N, Mankin AS. Retapamulin-assisted ribosome profiling reveals the alternative bacterial proteome. Mol Cell, in press.

38. Roy RN, Lomakin IB, Gagnon MG, Steitz TA. 2015. The mechanism of inhibition of protein synthesis by the proline-rich peptide oncocin. Nat Struct Mol Biol 22:466–469. https://doi.org/10.1038/nsmb.3031.

39. Li GW, Oh E, Weissman JS. 2012. The anti-Shine-Dalgarno sequence drives translational pausing and codon choice in bacteria. Nature 484: 538–541. https://doi.org/10.1038/nature10965.

40. Mandin P, Gottesman S. 2009. A genetic approach for finding small RNAs regulators of genes of interest identifies RybC as regulating the DpiA/DpiB two-component system. Mol Microbiol 72:551–565. https://doi.org/10.1111/j.1365-2958.2009.06665.x.

41. Zeghouf M, Li J, Butland G, Borkowska A, Canadien V, Richards D, Beattie B, Emili A, Greenblatt JF. 2004. Sequential Peptide Affinity (SPA) system for the identification of mammalian and bacterial protein complexes. J Proteome Res 3:463–468. https://doi.org/10.1021/pr034084x.

42. Hemm MR, Paul BJ, Miranda-Rios J, Zhang A, Soltanzad N, Storz G. 2010. Small stress response proteins in Escherichia coli: proteins missed by classical proteomic studies. J Bacteriol 192:46–58. https://doi.org/10.1128/JB.00872-09.

43. Sesto N, Wurtzel O, Archambaud C, Sorek R, Cossart P. 2013. The excludon: a new concept in bacterial antisense RNA-mediated gene regulation. Nat Rev Microbiol 11:75–82. https://doi.org/10.1038/nrmicro2934.

44. Georg J, Hess WR. 2018. Widespread antisense transcription in prokaryotes. Microbiol Spectr 6(4):RWR-0029-2018. https://doi.org/10.1128/microbiolspec.RWR-0029-2018.

45. Thomason MK, Bischler T, Eisenbart SK, Förstner KU, Zhang A, Herbig A, Nieselt K, Sharma CM, Storz G. 2015. Global transcriptional start site mapping using differential RNA sequencing reveals novel antisense RNAs in Escherichia coli. J Bacteriol 197:18–28. https://doi.org/10.1128/JB.02096-14.

46. Squires CL, Pedersen S, Ross BM, Squires C. 1991. ClpB is the Escherichia coli heat shock protein F84.1. J Bacteriol 173:4254–4262. https://doi.org/10.1128/jb.173.14.4254-4262.1991.

47. Sacerdot C, Dessen P, Hershey JW, Plumbridge JA, Grunberg-Manago M. 1984. Sequence of the initiation factor IF2 gene: unusual protein features and homologies with elongation factors. Proc Natl Acad Sci U S A 81:7787–7791. https://doi.org/10.1073/pnas.81.24.7787.

48. Ross TK, Achberger EC, Braymer HD. 1989. Nucleotide sequence of the McrB region of Escherichia coli K-12 and evidence for two independent translational initiation sites at the mcrB locus. J Bacteriol 171:1974–1981. https://doi.org/10.1128/jb.171.4.1974-1981.1989.

49. Zhang N, Young R. 1999. Complementation and characterization of the nested Rz and Rz1 reading frames in the genome of bacteriophage lambda. Mol Gen Genet 262:659–667. https://doi.org/10.1007/s004380051128.

50. Toba FA, Thompson MG, Campbell BR, Junker LM, Rueggeberg KG, Hay AG. 2011. Role of DLP12 lysis genes in Escherichia coli biofilm formation. Microbiology 157:1640–1650. https://doi.org/10.1099/mic.0.045161-0.

51. Park H, McGibbon LC, Potts AH, Yakhnin H, Romeo T, Babitzke P. 2017. Translational repression of the RpoS antiadapter IraD by CsrA is mediated via translational coupling to a short upstream open reading frame. mBio 8:e01355-17. https://doi.org/10.1128/mBio.01355-17.

52. Gollnick P, Babitzke P. 2002. Transcription attenuation. Biochim Biophys Acta 1577:240–250. https://doi.org/10.1016/S0167-4781(02)00455-4.

53. Gao X, Wan J, Liu B, Ma M, Shen B, Qian SB. 2015. Quantitative profiling of initiating ribosomes in vivo. Nat Methods 12:147–153. https://doi.org/10.1038/nmeth.3208.

54. Goethe O, Heuer A, Ma X, Wang Z, Herzon SB. 2019. Antibacterial properties and clinical potential of pleuromutilins. Nat Prod Rep 36:220–247. https://doi.org/10.1039/c8np00042e.

55. Novak R. 2011. Are pleuromutilin antibiotics finally fit for human use? Ann N Y Acad Sci 1241:71–81. https://doi.org/10.1111/j.1749-6632.2011.06219.x.

56. Mattiuzzo M, Bandiera A, Gennaro R, Benincasa M, Pacor S, Antcheva N, Scocchi M. 2007. Role of the Escherichia coli SbmA in the antimicrobial activity of proline-rich peptides. Mol Microbiol 66:151–163. https://doi.org/10.1111/j.1365-2958.2007.05903.x.

57. Hücker SM, Ardern Z, Goldberg T, Schafferhans A, Bernhofer M, Vestergaard G, Nelson CW, Schloter M, Rost B, Scherer S, Neuhaus K. 2017. Discovery of numerous novel small genes in the intergenic regions of the Escherichia coli O157:H7 Sakai genome. PLoS One 12:e0184119. https://doi.org/10.1371/journal.pone.0184119.

58. Friedman RC, Kalkhof S, Doppelt-Azeroual O, Mueller SA, Chovancová M, von Bergen M, Schwikowski B. 2017. Common and phylogenetically widespread coding for peptides by bacterial small RNAs. BMC Genomics 18:553. https://doi.org/10.1186/s12864-017-3932-y.

59. UniProt Consortium. 2015. UniProt: a hub for protein information. Nucleic Acids Res 43:D204–D212. https://doi.org/10.1093/nar/gku989.

60. Krishna SS, Majumdar I, Grishin NV. 2003. Structural classification of zinc fingers: survey and summary. Nucleic Acids Res 31:532–550. https://doi.org/10.1093/nar/gkg161.

61. Lobley A, Sadowski MI, Jones DT. 2009. pGenTHREADER and pDomTHREADER: new methods for improved protein fold recognition and superfamily discrimination. Bioinformatics 25:1761–1767. https://doi.org/10.1093/bioinformatics/btp302.

62. Ikeda M, Arai M, Okuno T, Shimizu T. 2003. TMPDB: a database of experimentally-characterized transmembrane topologies. Nucleic Acids Res 31:406–409. https://doi.org/10.1093/nar/gkg020.

63. Kall L, Krogh A, Sonnhammer EL. 2004. A combined transmembrane topology and signal peptide prediction method. J Mol Biol 338: 1027–1036. https://doi.org/10.1016/j.jmb.2004.03.016.

64. Bartoli J, My L, Belmudes L, Couté Y, Viala JP, Bouveret E. 2017. The long hunt for pssR—looking for a phospholipid synthesis transcriptional regulator, finding the ribosome. J Bacteriol 199:e00202-17. https://doi.org/10.1128/JB.00202-17.

65. Woolstenhulme CJ, Guydosh NR, Green R, Buskirk AR. 2015. High-precision analysis of translational pausing by ribosome profiling in bacteria lacking EFP. Cell Rep 11:13–21. https://doi.org/10.1016/j.celrep.2015.03.014.

66. Yu D, Ellis HM, Lee EC, Jenkins NA, Copeland NG, Court DL. 2000. An efficient recombination system for chromosome engineering in Escherichia coli. Proc Natl Acad Sci U S A 97:5978–5983. https://doi.org/10.1073/pnas.100127597.

67. Miller JH. 1992. A short course in bacterial genetics: a laboratory manual and handbook for Escherichia coli and related bacteria. Cold Spring Harbor Laboratory Press, Plainview, NY.

68. Wada A, Sako T. 1987. Primary structures of and genes for new ribosomal proteins A and B in Escherichia coli. J Biochem 101:817–820. https://doi.org/10.1093/jb/101.3.817.

69. Hansen FG, Hansen EB, Atlung T. 1982. The nucleotide sequence of the dnaA gene promoter and of the adjacent rpmH gene, coding for the ribosomal protein L34, of Escherichia coli. EMBO J 1:1043–1048. https://doi.org/10.1002/j.1460-2075.1982.tb01294.x.

70. Wassarman KM, Repoila F, Rosenow C, Storz G, Gottesman S. 2001. Identification of novel small RNAs using comparative genomics and microarrays. Genes Dev 15:1637–1651. https://doi.org/10.1101/gad.901001.

71. Chen S, Lesnik EA, Hall AT, Sampath R, Griffey RH, Ecker DJ, Blyn LB. 2002. A bioinformatics based approach to discover small RNA genes in the Escherichia coli genome. Biosystems 65:157–177. https://doi.org/10.1016/S0303-2647(02)00013-8.

72. Kawano M, Reynolds AA, Miranda-Rios J, Storz G. 2005. Detection of 5′ and 3′ UTR-derived small RNAs and cis-encoded antisense RNAs in Escherichia coli. Nucleic Acids Res 33:1040–1050. https://doi.org/10.1093/nar/gki256.

73. Muller MM, Webster RE. 1997. Characterization of the tol-pal and cyd region of Escherichia coli K-12: transcript analysis and identification of two new proteins encoded by the cyd operon. J Bacteriol 179: 2077–2080. https://doi.org/10.1128/jb.179.6.2077-2080.1997.

74. Guan Z, Wang X, Raetz CR. 2011. Identification of a chloroform-soluble membrane miniprotein in Escherichia coli and its homolog in Salmonella typhimurium. Anal Biochem 409:284–289. https://doi.org/10.1016/j.ab.2010.10.035.

75. Baba T, Ara T, Hasegawa M, Takai Y, Okumura Y, Baba M, Datsenko KA, Tomita M, Wanner BL, Mori H. 2006. Construction of Escherichia coli K-12 in-frame, single-gene knockout mutants: the Keio collection. Mol Syst Biol 2:2006.0008. https://doi.org/10.1038/msb4100050.

76. Wong RS, McMurry LM, Levy SB. 2000. 'Intergenic' blr gene in Escherichia coli encodes a 41-residue membrane protein affecting intrinsic susceptibility to certain inhibitors of peptidoglycan synthesis. Mol Microbiol 37:364–370. https://doi.org/10.1046/j.1365-2958.2000.01998.x.

77. Gassel M, Möllenkamp T, Puppe W, Altendorf K. 1999. The KdpF subunit is part of the K⁺-translocating Kdp complex of Escherichia coli and is responsible for stabilization of the complex in vitro. J Biol Chem 274: 37901–37907. https://doi.org/10.1074/jbc.274.53.37901.

78. Blattner FR, Plunkett G, III, Bloch CA, Perna NT, Burland V, Riley M, Collado-Vides J, Glasner JD, Rode CK, Mayhew GF, Gregor J, Davis NW, Kirkpatrick HA, Goeden MA, Rose DJ, Mau B, Shao Y. 1997. The complete genome sequence of Escherichia coli K-12. Science 277:1453–1462. https://doi.org/10.1126/science.277.5331.1453.

79. Kato A, Tanabe H, Utsumi R. 1999. Molecular characterization of the PhoP-PhoQ two-component system in Escherichia coli K-12: identifica-

tion of extracellular Mg²⁺-responsive promoters. J Bacteriol 181: 5516–5520.

80. Izutsu K, Wada C, Komine Y, Sako T, Ueguchi C, Nakura S, Wada A. 2001. *Escherichia coli* ribosome-associated protein SRA, whose copy number increases during stationary phase. J Bacteriol 183:2765–2673. https://doi.org/10.1128/JB.183.9.2765-2773.2001.

81. Dassa J, Marck C, Boquet PL. 1990. The complete nucleotide sequence of the *Escherichia coli* gene *appA* reveals significant homology between pH 2.5 acid phosphatase and glucose-1-phosphatase. J Bacteriol 172: 5497–5500. https://doi.org/10.1128/jb.172.9.5497-5500.1990.

82. Tjaden B, Goodwin SS, Opdyke JA, Guillier M, Fu DX, Gottesman S, Storz G. 2006. Target prediction for small, noncoding RNAs in bacteria. Nucleic Acids Res 34:2791–2802. https://doi.org/10.1093/nar/gkl356.

83. Cam K, Béjar S, Gil D, Bouché JP. 1988. Identification and sequence of gene *dicB*: translation of the division inhibitor from an in-phase internal start. Nucleic Acids Res 16:6327–6338. https://doi.org/10.1093/nar/16.14.6327.

84. Bishop RE, Leskiw BK, Hodges RS, Kay CM, Weiner JH. 1998. The entericidin locus of *Escherichia coli* and its implications for programmed bacterial cell death. J Mol Biol 280:583–596. https://doi.org/10.1006/jmbi.1998.1894.

85. Wadler CS, Vanderpool CK. 2007. A dual function for a bacterial small RNA: SgrS performs base pairing-dependent regulation and encodes a functional polypeptide. Proc Natl Acad Sci U S A 104:20454–20459. https://doi.org/10.1073/pnas.0708102104.

86. Kitagawa R, Mitsuki H, Okazaki T, Ogawa T. 1996. A novel DnaA protein-binding site at 94.7 min on the *Escherichia coli* chromosome. Mol Microbiol 19:1137–1147. https://doi.org/10.1046/j.1365-2958.1996.453983.x.