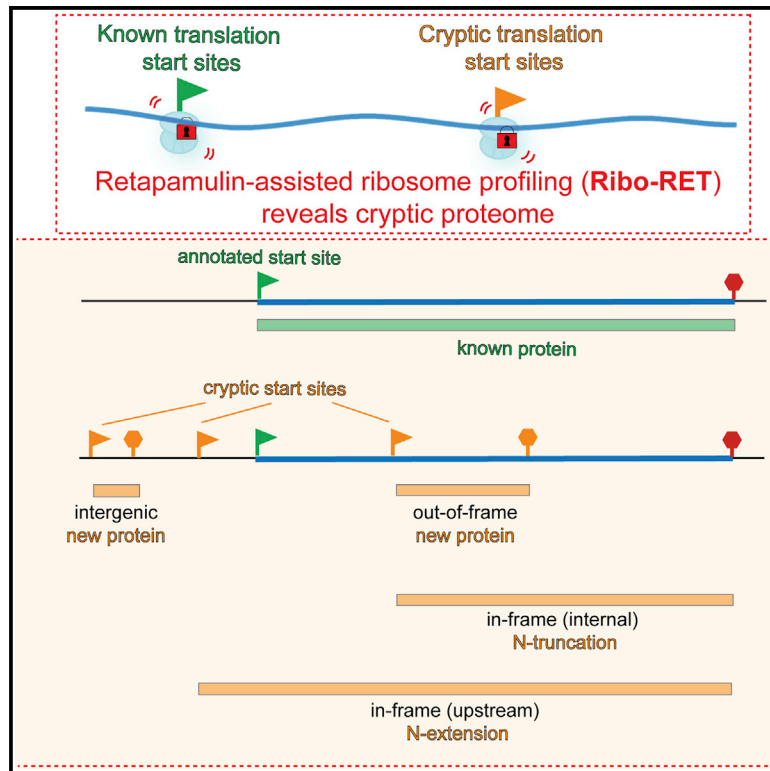# Molecular Cell

Retapamulin-Assisted Ribosome Profiling Reveals the Alternative Bacterial Proteome

Graphical Abstract

Authors

Sezen Meydan, James Marks, Dorota Klepacki, ..., Amira Kefi, Nora Vázquez-Laslop, Alexander S. Mankin

Correspondence

nvazquez@uic.edu (N.V.-L.),
shura@uic.edu (A.S.M.)

In Brief

Meydan et al. report that the antibiotic retapamulin arrests bacterial ribosomes specifically at start codons. Retapamulin-assisted ribosome profiling detects known start sites but also reveals multiple cryptic sites within or outside the coding regions. Translation from internal start sites may lead to generation of functional proteins or play a regulatory role.

Highlights

- Retapamulin arrests bacterial ribosomes specifically at translation start sites

- Retapamulin-assisted Ribo-seq reveals known and cryptic translation start sites

- Translation from start sites located within the ORFs may generate functional proteins

- Start-stop sites found within some genes may help to regulate gene expression

CellPress

Molecular Cell
Article

Retapamulin-Assisted Ribosome Profiling Reveals the Alternative Bacterial Proteome

Sezen Meydan,[1,7] James Marks,[1,8] Dorota Klepacki,[1] Virag Sharma,[2,3,4] Pavel V. Baranov,[5] Andrew E. Firth,[6] Tõnu Margus,[1,9] Amira Kefi,[1] Nora Vázquez-Laslop,[1,*] and Alexander S. Mankin[1,10,*]

[1]Center for Biomolecular Sciences-m/c 870, University of Illinois at Chicago, 900 S. Ashland Ave., Chicago, IL 60607, USA
[2]CRTD-DFG Center for Regenerative Therapies Dresden, Technische Universität Dresden, Dresden, Germany
[3]Paul Langerhans Institute Dresden of the Helmholtz Center Munich at University Hospital, Carl Gustav Carus Faculty of Medicine, Technische Universität Dresden, Dresden, Germany
[4]German Center for Diabetes Research, Munich, Neuherberg, Germany
[5]School of Biochemistry and Cell Biology, University College Cork, Cork T12 YN60, Ireland
[6]Division of Virology, Department of Pathology, University of Cambridge, Cambridge CB2 1QP, UK
[7]Present address: National Institute of Diabetes and Digestive and Kidney Diseases, National Institutes of Health, Bethesda, MD 20892, USA
[8]Present address: National Institute of Arthritis and Musculoskeletal and Skin Diseases, National Institutes of Health, Bethesda, MD 20892, USA
[9]Present address: Department of Molecular Biology, Umeå University, 90187 Umeå, Sweden
[10]Lead Contact
*Correspondence: nvazquez@uic.edu (N.V.-L.), shura@uic.edu (A.S.M.)
https://doi.org/10.1016/j.molcel.2019.02.017

SUMMARY

The use of alternative translation initiation sites enables production of more than one protein from a single gene, thereby expanding the cellular proteome. Although several such examples have been serendipitously found in bacteria, genome-wide mapping of alternative translation start sites has been unattainable. We found that the antibiotic retapamulin specifically arrests initiating ribosomes at start codons of the genes. Retapamulin-enhanced Ribo-seq analysis (Ribo-RET) not only allowed mapping of conventional initiation sites at the beginning of the genes, but strikingly, it also revealed putative internal start sites in a number of *Escherichia coli* genes. Experiments demonstrated that the internal start codons can be recognized by the ribosomes and direct translation initiation *in vitro* and *in vivo*. Proteins, whose synthesis is initiated at internal in-frame and out-of-frame start sites, can be functionally important and contribute to the "alternative" bacterial proteome. The internal start sites may also play regulatory roles in gene expression.

INTRODUCTION

A broader diversity of proteins with specialized functions can augment cell reproduction capacity, optimize its metabolism, and facilitate survival in the ever-changing environment. However, the fitness gain acquired by making a new protein is counterbalanced with the cost of expanding the size of the genome, a conundrum particularly onerous in bacteria whose genomes are highly streamlined.

Different strategies can be used for diversifying proteome without expanding genome size. For instance, ribosomes may initiate at a unique start codon of an open reading frame (ORF), but due to programmed ribosomal frameshifting or stop codon readthrough, some of them may produce a polypeptide whose sequence deviates from that encoded in the main ORF. Such recoding events lead to generation of more than one protein from a single gene (Atkins et al., 2016; Baranov et al., 2015).

Another possible way to produce diverse polypeptides from a single ORF is the utilization of alternative internally located start codons. Although translation of the majority of the bacterial genes is initiated at a unique translation initiation site (TIS), designated herein as primary (pTIS), several examples of genes with an additional internal TIS (iTIS) have been uncovered by detecting additional polypeptide products during the purification of the primary protein (reviewed in Meydan et al., 2018). In these genes, translation initiated at the pTIS results in production of the full-length (primary) protein, while ribosomes that initiate translation at the in-frame iTIS synthesize an alternative, N-terminally truncated polypeptide. Such primary and alternative proteins may have related but specialized functions. The products of in-frame internal initiation at several bacterial genes have been reported to participate in various cellular functions ranging from virulence to photosynthesis or antibiotic production, among others (reviewed in Meydan et al., 2018). In very few known cases, iTIS directs translation in a reading frame different from the primary ORF (D'Souza et al., 1994; Feltens et al., 2003; Yuan et al., 2018).

Most of the known examples of translation of a protein from an iTIS have been discovered serendipitously. Although several computational algorithms can predict the pTIS of many bacterial ORFs (Gao et al., 2010; Giess et al., 2017; Makita et al., 2007; Salzberg et al., 1998), iTIS prediction remains by far more challenging and has not even been pursued in most of those studies. The recent advent of new mass-spectrometry-based approaches have allowed the identification of N-terminal peptides

of a range of proteins expressed in bacteria (Bienvenut et al., 2015; Impens et al., 2017), including some whose translation was likely initiated at an iTIS. However, the success of the available techniques for identifying such proteins is intrinsically restricted by the stringent requirements for the chemical properties, size, and abundance of the peptides that can be detected by mass spectrometry. Therefore, the majority of the functional iTISs in the genomes likely remain overlooked.

Ribosome profiling (Ribo-seq), based on deep sequencing of ribosome-protected mRNA fragments ("ribosome footprints"), allows for genome-wide survey of translation (Ingolia et al., 2009). Ribo-seq experiments carried out with eukaryotic cells pre-treated with the translation initiation inhibitors harringtonine (Ingolia et al., 2011) and lactimidomycin (Gao et al., 2015; Lee et al., 2012) or with puromycin (Fritsch et al., 2012) showed specific enrichment of ribosome footprints at or near start codons of ORFs and facilitated mapping of TISs in eukaryotic genomes. These studies also revealed active translation of previously unknown short ORFs in the 5′ UTRs of many genes and identified several TISs within the genes that were attributed to leaky scanning through the primary start sites (Lee et al., 2012). Analogous studies, however, have been difficult to conduct in bacteria because of the paucity of inhibitors with the required mechanism of action. An inhibitor useful for mapping start sites should allow the assembly of the 70S translation initiation complex at a TIS but must prevent the ribosome from leaving the start codon. Unfortunately, most of the ribosomal antibiotics traditionally viewed as initiation inhibitors do not satisfy these criteria. Recently, tetracycline (TET), an antibiotic that prevents aminoacyl-tRNAs from entering the ribosomal A site and commonly known as an elongation inhibitor (Cundliffe, 1981), was used in conjunction with Ribo-seq to globally map pTISs in the E. coli genome (Nakahigashi et al., 2016). Although TET Ribo-seq data successfully revealed the pTISs of many of the actively translated genes, identification of iTISs was not feasible in that work because of the substantial number of footprints generated by elongating ribosomes. Furthermore, because TET can potentially bind to the ribosome at every round of elongation cycle, when the A-site is temporarily empty, it is impossible to distinguish whether the footprints within the ORFs represented elongating ribosomes or those initiating translation at an iTIS (Nakahigashi et al., 2016).

Here, we show that retapamulin (RET), an antibiotic of the pleuromutilin family, exclusively stalls ribosomes at the start codons of the ORFs. Brief pre-treatment of E. coli cells with RET dramatically rearranges the distribution of ribosomes along the ORFs, confining the ribosomal footprints obtained by Ribo-seq to the TISs of the genes. Strikingly, the application of the Ribo-seq/RET approach to the analysis of bacterial translation revealed that more than many E. coli genes contain actively used iTISs. In vitro and in vivo experiments confirmed initiation of translation at some of the discovered iTISs and show that internal initiation may lead to production of proteins with distinct functions. Our data show that initiation at alternative start sites is widespread in bacteria and reveal the possible existence of a previously cryptic fraction of the proteome.

RESULTS

RET Arrests the Initiating Ribosome at the Start Codons

Pleuromutilin antibiotics, including clinically used semi-synthetic RET, bind in the peptidyl transferase center of the bacterial ribosome, hindering the placement of the P- and A-site amino acids and thus preventing peptide bond formation (Figures S1A and S1B) (Davidovich et al., 2007; Poulsen et al., 2001; Schlünzen et al., 2004). In vitro studies have shown that presence of fMet-tRNA and RET in the ribosome are not mutually exclusive (Yan et al., 2006). Therefore, we reasoned that RET may allow the assembly of the 70S initiation complex at the start codon, but by displacing the aminoacyl moiety of the initiator fMet-tRNA and interfering with the placement of an elongator aminoacyl-tRNA in the A site, it could prevent formation of the first peptide bond.

The results of polysome analysis were compatible with RET being a selective translation initiation inhibitor, because treatment of E. coli cells with high concentrations of the drug, 100-fold over the minimal inhibitory concentration (MIC), rapidly converted polysomes into monosomes (Figure S1C). We then used toeprinting analysis (Hartz et al., 1988) to test whether RET captures ribosomes at start codons. When model genes were translated in an E. coli cell-free system (Shimizu et al., 2001), addition of RET stalled ribosomes exclusively at the ORFs start codons (Figure 1A, "RET" lanes), demonstrating that this antibiotic readily, and possibly specifically, inhibits translation initiation. In contrast, TET, which was used previously to map TISs in the E. coli genome (Nakahigashi et al., 2016), halted translation not only at the translation initiation sites but also at downstream codons of the ORFs (Figure 1A, "TET" lanes), confirming that this inhibitor interferes with both initiation and elongation of translation (Orelle et al., 2013).

The outcomes of the polysome and toeprinting analyses, along with the structural data showing the incompatibility of the nascent protein chain with RET binding (Figure S1B), encouraged us to assess whether RET would enable the use of Ribo-seq for mapping translation start sites in bacterial cells. Even a brief (2 min) exposure of the ΔtolC derivative of the E. coli strain BW25113 to a 32-fold MIC of RET nearly completely halts protein synthesis (Figure S1D). However, in Ribo-seq experiments, we exposed cells for 5 min to a 100-fold MIC of RET to ensure that elongating ribosomes complete translation of even long or slowly translated ORFs prior to cell harvesting. Analysis of the Ribo-seq data showed that the RET treatment led to a striking ribosome redistribution. The occupancy of the internal and termination codons of the expressed genes was severely reduced compared to that of the untreated control, whereas the ribosome density peaks at the start codons dramatically increased (Figures 1B and 1C). Although a generally similar trend can be observed in the metagene analysis of the Ribo-seq data in the RET-treated (this paper) and TET-treated cells (Nakahigashi et al., 2016), the start-codon peak in the TET experiments is smaller and broader compared to the peak of the RET-stalled ribosomes (Figures S1E and S1F), reflecting a higher potency of RET as initiation inhibitor. Filtered by fairly conservative criteria (see STAR Methods), Ribo-seq data revealed distinct peaks of ribosome density at the annotated start codons (pTISs) of 991 out of 1,153 (86%) E. coli genes expressed in the BW25113.

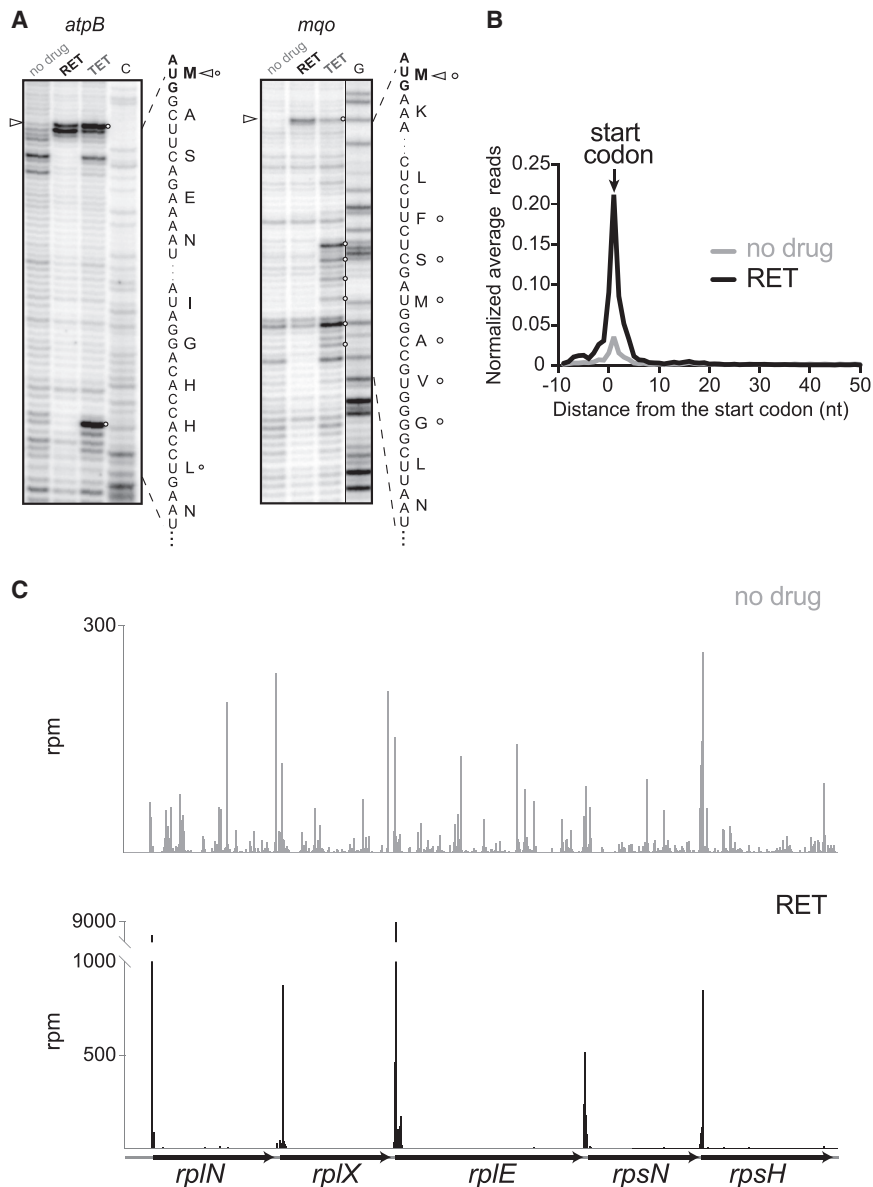**Figure 1. RET Specifically Arrests Ribosomes at Translation Initiation Sites**

(A) Toeprinting analysis showing retapamulin (RET)-induced (triangles) and tetracycline (TET)-induced (circles) translation arrest sites during cell-free translation of two model *E. coli* genes. "C" and "G" indicate sequencing lanes. Nucleotide and encoded amino acid sequences are shown.

(B) Metagene analysis plot representing normalized average relative density reads in the vicinity of the annotated start codons of the genes of *E. coli* cells treated or not with RET.

(C) Ribosome footprints density within the *spc* operon in cells treated or not with RET.

See also Figure S1.

The magnitude of the start codon peaks at the remaining 14% of the translated genes did not pass our threshold criteria (see STAR Methods), possibly reflecting changes in gene expression due to the RET treatment.

Taken together, our *in vitro* and *in vivo* results showed that RET acts as a specific inhibitor of translation initiation locking the ribosomes at the start codons, and in combination with Ribo-seq, can be used for mapping the pTISs of the majority of actively translated genes in bacterial genomes. We named the Ribo-seq/RET approach "Ribo-RET."

Ribo-RET Unmasks Initiation of Translation at Internal Codons of Many Bacterial Genes

Even though the majority of the ribosome footprints in the Ribo-RET dataset mapped to annotated pTISs, we also observed peaks at certain internal codons (Figure 2A). Hypothetically, the presence of internal Ribo-RET peaks could be explained by elongating ribosomes paused at specific sites within the ORF. Nonetheless, this possibility seems unlikely, because no substantial Ribo-RET peak was detected even at the most prominent programmed translation arrest site in the *E. coli* genome within the *secM* ORF (Nakatogawa and Ito, 2002) (Figure S1G). Similarly implausible was the origin of the internal RET peaks from context-specific elongation arrest observed with some other antibiotics (Kannan et al., 2014; Marks et al., 2016) because biochemical (Dornhelm and Högenauer, 1978) and structural (Davidovich et al., 2007) data strongly argue that RET cannot bind to the elongating ribosome (Figure S1B). We therefore concluded that the Ribo-RET peaks at internal sites within ORFs must represent ribosomes caught in the act of initiating translation.

Three *E. coli* genes, *infB*, *mcrB*, and *clpB*, were previously reported to encode two different polypeptide isoforms due to the iTIS presence: translation of the full-size protein is initiated at the pTIS while the shorter isoform is expressed from an iTIS (Broome-Smith et al., 1985; Park et al., 1993; Plumbridge et al., 1985). The Ribo-RET profile of these genes showed well-defined and highly specific ribosome density peaks (Figure 2B) at the known iTISs, thereby verifying the utility of Ribo-RET for mapping iTISs in bacterial genes.

Among the *E. coli* BW25113 genes expressed in our conditions, we identified 239 iTIS candidates. To further expand the systematic identification of iTISs in *E. coli* genes, we applied the Ribo-RET approach to the ΔtolC derivative of the *E. coli* strain BL21, a B-type strain that is genetically distinct from the K-strain BW25113 (Grenier et al., 2014; Studier et al., 2009). Ribo-RET analysis identified 620 iTISs in the
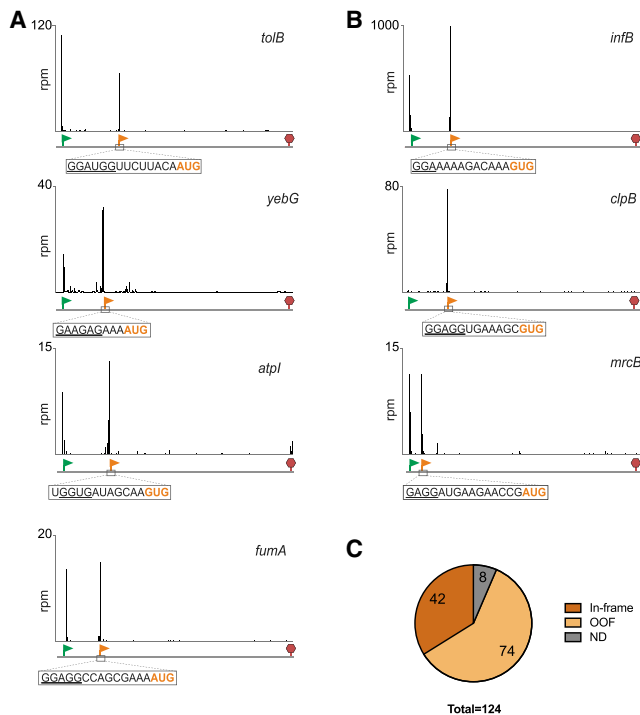
Figure 2. Ribo-RET Reveals the Presence of iTISs in Many Bacterial Genes

(A) Examples of Ribo-RET profiles of *E. coli* genes with newly detected iTISs. The annotated pTISs are marked with green flags and stop codons are shown with red stop signs; orange flags show the iTISs. iTIS start codons are highlighted in orange and the SD-like sequences are underlined.

(B) Ribo-RET profiles of *infB*, *clpB*, *mrcB*, the three *E. coli* genes where iTISs had been previously characterized.

(C) The iTISs common between the *E. coli* BW25113 and BL21 strains. ND (not determined): the internal Ribo-RET peaks not associated with the known start codons.

See also Table S1.

BL21 strain. Of these, 124 iTISs were common between the two strains (Table S1). While a notably higher number of iTIS in the BL21 remains somewhat puzzling, it may be related to the fact that more genes are expressed in this strain in comparison with BW25113 (1,990 genes with the identified pTISs in BL21 strain versus 1,554 such genes in BW25113), and to the sequence variations between the strains, as a result of which 244 BL21-specific iTISs did not have a perfect sequence match in the BW25113 strain. We limited our subsequent analysis to 124 iTISs conserved between the two strains, among which, 42 directed translation in frame with the main gene, whereas start codons of 74 iTISs were out of frame relative to the main ORF; for 8 iTISs the start site was not assigned (Figure 2C; Table S1). In the following sections, we consider the first two classes separately.

Internal Translation Initiation Sites that Are in-Frame with the Main ORF

The in-frame iTISs exploit various initiator codons that have been shown previously to be capable of directing translation initiation

in *E. coli* (Chengguang et al., 2017; Hecht et al., 2017), although similar to the pTISs, the AUG codon is the most prevalent (Figure 3A). A SD-like sequence could be recognized upstream of many of the in-frame internal start codons (Table S1).

Initiation at an in-frame iTIS would generate an N-terminally truncated form of the primary protein. The sizes of candidate proteins expressed from in-frame iTISs range from 6 to 805 amino acids in length (Table S1). Although the locations of in-frame iTISs are highly variable, the majority are clustered close to the beginning of the gene or are within the 3′ terminal quartile of the ORF length (Figure 3B). We examined the iTIS of the *arcB* gene as a 3′-proximal start site representative and that of *speA* as an example of the 5′-proximal iTIS.

A Protein with a Putative Specialized Function Is Translated from the 3′-Proximal iTIS of the *arcB* Gene

The gene *arcB* encodes the sensor kinase ArcB of the two-component signal transduction system ArcB/A that helps bacteria to sense and respond to changes in oxygen concentration (Alvarez and Georgellis, 2010) (Figures 3C and 3D). The ArcB protein consists of the transmitter, receiver, and phosphotransfer domains (Figure 3D). Under microaerobic conditions, ArcB undergoes a series of phosphorylation steps that eventually activate the response regulator ArcA that controls expression of nearly 200 genes (Alvarez and Georgellis, 2010; Salmon et al., 2005). The C-terminal ArcB-C domain is the ultimate receiver of the phosphoryl group within the ArcB membrane-anchored protein and serves as the phosphoryl donor for ArcA (Alvarez et al., 2016).

The Ribo-RET data showed a strong ribosome density peak at an iTIS in *arcB*, with the putative start codon GUG located precisely at the 5′ boundary of the segment encoding the ArcB-C domain (Figures 3C and 3D). A similarly located iTIS can be found in the *arcB* gene of several bacterial species (Figure S2B). Initiation of translation at the *arcB* iTIS could generate a diffusible ArcB-C polypeptide, detached from the membrane-bound ArcB kinase (Figure 3D). To test this possibility, we introduced the 3xFLAG-coding sequence at the 3′ end of the *arcB* gene, expressed it from a plasmid in *E. coli* cells, and analyzed the protein products. Expression of the tagged *arcB* resulted in the simultaneous production of the full-size ArcB and of a smaller protein with apparent molecular weight (MW) of 14 kDa, consistent with that of the FLAG-tagged ArcB-C (Figures 3E and S2A). Disruption of the iTIS by synonymous mutations did not affect the synthesis of the full-length ArcB but abrogated that of ArcB-C (Figure 3E), confirming that the ArcB-C polypeptide is produced via initiation of translation at the *arcB* iTIS. Previous *in vitro* experiments showed that the isolated ArcB-C domain could serve as a phosphoryl acceptor and donor for the ArcB-catalyzed phosphorylation reactions (Alvarez and Georgellis, 2010), suggesting that a self-standing ArcB-C protein is likely functional *in vivo*. In agreement with this possibility, under micro-aerobic conditions, *E. coli* cells with the operational *arcB* iTIS win over the mutant in which the iTIS is disrupted by synonymous mutations (Figures 3F and S2C). Diffusible ArcB-C may either facilitate the operation of the ArcB-ArcA signal transduction pathway or could enable a cross-talk with other signal transduction systems (Figure 3G).
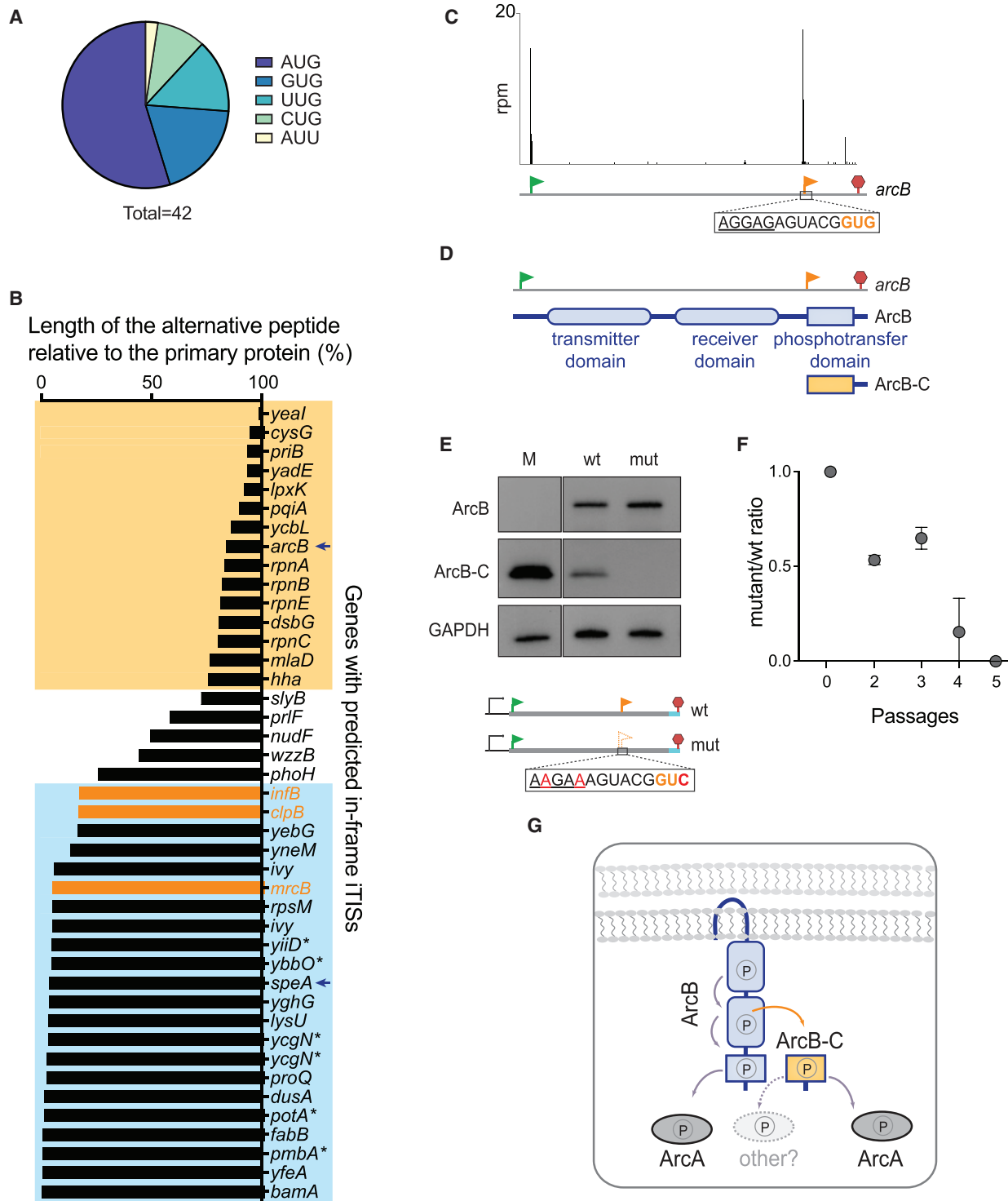
CellPress

Figure 3. In-Frame Internal Initiation Can Generate Functional N-Terminally Truncated Proteins

(A) The frequency of various putative start codons at the in-frame iTISs.

(B) The relative length of the predicted alternative proteins, products of internal initiation, in comparison with the main protein. The known examples of genes with in-frame iTISs (Figure 2B) are in orange. The genes with the iTISs located within the 3′ or 5′ quartile of the gene length are boxed in yellow or blue, respectively. Asterisks show genes with pTIS re-annotation proposed based on TET-assisted Ribo-seq (Nakahigashi et al., 2016). Arrows point at *arcB* and *speA* genes further analyzed in this work.

(C) Ribo-RET profile of *arcB* showing ribosomal density peaks at the pTIS (green flag) and the iTIS (orange flag).

(legend continued on next page)

CellPress

The expression of ArcB-C from the *arcB* iTIS is apparently quite efficient because *E. coli* and *Salmonella enterica* Ribo-seq datasets show a notable upshift in the ribosome density at the *arcB* codons located downstream of the iTIS (Baek et al., 2017; Kannan et al., 2014; Li et al., 2014) (Figures S2D–S2G). Curiously, the average ribosome occupancy of the *arcB* codons before and after the iTIS vary under different physiological conditions (Figures S2F and S2G), suggesting that utilization of the *arcB* pTIS and iTIS could be regulated.

Another remarkable example of in-frame 3'-proximal iTISs is found in the homologous *rpnA-E* genes of *E. coli*, encoding nucleases involved in DNA recombination (Kingston et al., 2017). Each of the *rpn* genes show Ribo-RET peaks at iTISs that appear to be their major initiation sites (Figure S2H) under the growth conditions of our experiments. Curiously, the polypeptide expressed from the *rpnE* iTIS is 98% identical to the product of the *ypaA* gene (Figure S2I), revealing a possible distinct functionality of the alternative products of the *rpn* gene family.

Translation initiated at a 5'-Proximal iTIS of a Gene May Generate Differentially Targeted Proteins

The *speA* gene encodes arginine decarboxylase (SpeA), an enzyme involved in polyamine production (Michael, 2016). SpeA has been found in the *E. coli* cytoplasmic and periplasmic fractions (Buch and Boyle, 1985) and was reported to be represented by two polypeptide isoforms, SpeA-74, with an apparent MW of 74 kDa, and a smaller one of ~70 kDa, SpeA-70, suggested to be a co-secretional maturation product of the full-length SpeA-74 (Buch and Boyle, 1985; Wu and Morris, 1973). Our analysis, however, revealed two Ribo-RET peaks in the *speA* ORF: one corresponding to the annotated pTIS and the second one mapped to an iTIS at codon Met-26 (Figure S3A). Initiation of translation at the pTIS and iTIS of *speA* would generate the 73,767 Da and 71,062 Da forms of SpeA, respectively, arguing that the SpeA-70 isoform is generated due to initiation of translation at the *speA* iTIS. In support of this conclusion, the peptide (M)SSQEASKMLR, which precisely corresponds to the N terminus of the short SpeA isoform defined by Ribo-RET, can be found in the database of the experimentally identified *E. coli* N-terminal peptides (Bienvenut et al., 2015).

It was suggested that SpeA-74 is targeted to the periplasm due to the presence of a putative N-terminal secretion signal sequence (Buch and Boyle, 1985). A segment of this signal sequence would be missing in the SpeA-70 isoform confining the shorter polypeptide to the cytoplasm (Figure S3B). Therefore, utilization of the 5'-proximal iTIS of *speA* could change compartmentalization of the encoded protein. The 5'-proximal iTISs identified in some other *E. coli* genes encoding proteins

with signal sequences (e.g., *bamA*, *ivy*, or *yghG*) may serve similar roles (Figure S3C). An analogous strategy for targeting polypeptide isoforms to different cellular compartments has been described for few other bacterial proteins (reviewed in Meydan et al., 2018).

Six of the 5'-proximal iTISs (marked by asterisks in Figure 3B) were detected by TET Ribo-seq; it was suggested that they represent incorrectly annotated pTISs (Nakahigashi et al., 2016). Lack of the Ribo-RET peaks at the annotated pTIS of some of these genes (Table S1) is generally consistent with this proposal, suggesting that assignment of the pTISs could be reassessed. However, such a conclusion should be drawn cautiously because the utilization of the upstream pTIS could depend on growth conditions.

Conservation Analysis of in-Frame iTISs

We analyzed alignments of bacterial genes homologous to the *E. coli* genes where internal in-frame start sites were detected by Ribo-RET. Sequence logos and codon conservation plots indicated preservation of in-frame potential initiation sites and locally enhanced synonymous site conservation at *phoH*, *speA*, *yebG*, *yfaD*, and *yadD* (Figure S4A). However, it remains to be seen whether these conserved regions are relevant to promoting iTIS usage or simply represent unrelated sequence requirements of these genes. The other iTISs identified by Ribo-RET in the *E. coli* genome show a lower degree of evolutionary conservation, indicating that many of them could be species- or strain-specific.

Ribo-RET Identified iTISs that Are out of Frame Relative to the Main ORF

Only two examples of a bacterial ORF nested in an alternative frame within another ORF had been previously described: *comS* within *srfAB* in *B. subtilis* and *rpmH* within *rnpA* in *Thermus thermophilus* (reviewed in Meydan et al., 2018). Our Ribo-RET analysis showed the presence of 74 OOF iTISs common between the examined *E. coli* BW25113 and BL21 strains (Figure S5A). The location of the OOF iTISs within the host genes varies significantly; the peptides generated by translation initiated at the OOF iTISs would range in size from 2 to 84 amino acids (Figures 4A, S5A, and S5B).

We tested two OOF iTIS candidates (found within the *birA* and *sfsA* genes) for their ability to direct initiation of translation. Initiation of translation at the OOF UUG internal start site (overlapping the *birA* gene Leu$_{300}$ codon) would yield a 5-amino-acid-long peptide, while translation initiated at the OOF AUG of the *sfsA* gene (overlapping the Leu$_{95}$ codon of the main ORF) would generate a 12-amino-acid peptide (Figure 4B). When the

---

(D) Schematic map of the functional domains of ArcB. The putative alternative ArcB-C protein would encompass the phosphotransfer domain.

(E) Western blot analysis of the C-terminally 3XFLAG-tagged translation products of the *arcB* gene expressed from a plasmid in *E. coli* cells. Inactivation of iTIS by the indicated mutations (mut) abrogates production of ArcB-C. Lane M: individually expressed marker protein ArcB-C-3XFLAG (see STAR Methods).

(F) Functional iTIS in *arcB* facilitates growth under low-oxygen conditions. BW25113 Δ*arcB E. coli* cells, expressing from a plasmid either WT *arcB* or mutant *arcB* with inactivated iTIS, were co-grown in low-oxygen conditions, and the ratio of mutant to WT cells was analyzed (see STAR Methods for details). Error bars represent deviation from the mean (n = 2).

(G) The phospho-relay across the WT ArcB domains results in the activation of the response regulator ArcA (Alvarez et al., 2016). Diffusible ArcB-C could amplify the signal capabilities of the ArcBA system and/or enable cross-talk with other response regulators.
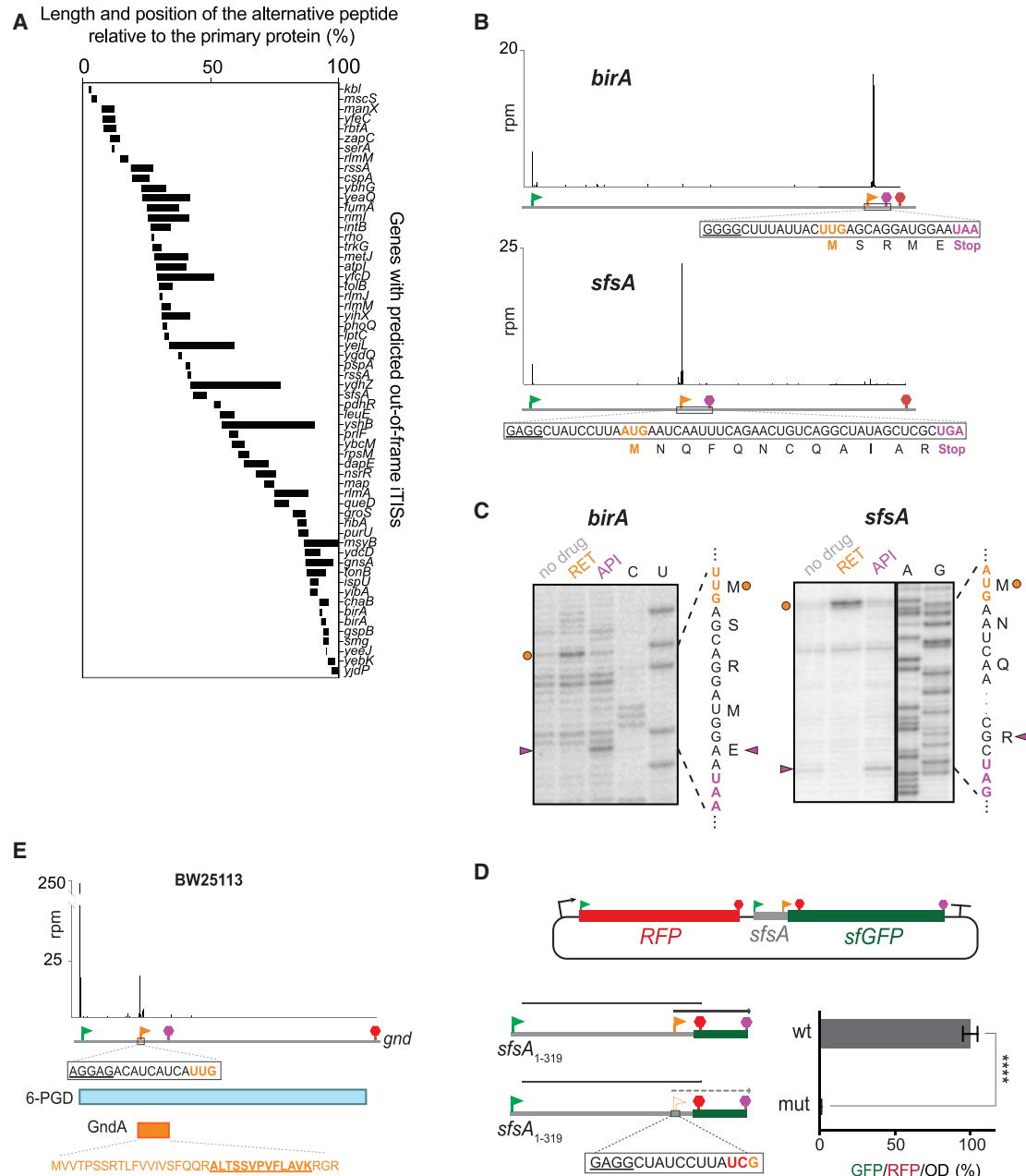
See also Figure S2.

Figure 4. OOF iTISs Revealed by Ribo-RET Can Direct Initiation of Translation

(A) The length and location of the alternative OOF protein-coding segments relative to the main ORF.

(B) Ribo-RET profiles of *birA* and *sfsA* genes with putative OOF iTISs. The iTIS OOF start codon and the corresponding stop codon are indicated by orange flag and stop sign, respectively. The sequences of the alternative ORFs are shown.

(C) Toeprinting analysis reveals that RET arrests ribosomes at the start codons (orange circles) of the alternative ORFs within the *birA* and *sfsA* genes; termination inhibitor Api137 arrests translation at the stop codons (purple arrowheads) of those ORFs. The nucleotide and amino acid sequences of the alternative ORFs are shown. Sequencing lanes are indicated.

(D) In the cell, translation is initiated at the *sfsA* OOF iTIS. Schematic map of the *RFP*/sf*GFP* reporter plasmid. The *rfp* and *sf-gfp* genes are co-transcribed. The sfGFP-coding sequence is expressed from the iTIS (orange flag) and is OOF relative to the *sfsA* pTIS start codon (green flag). The first stop codon in-frame with the pTIS (red) and the stop codon in-frame with the iTIS (orange) are indicated. The bar graph shows the change in relative green fluorescence in response to the iTIS start codon mutation (mut). The values represent the mean ± SD from technical replicates (n = 6). Two-tailed unpaired t test, ****p < 0.0001.

(E) Ribo-RET snapshots of the *gnd* gene revealing the putative location of the start codon of the alternative ORF. The amino acid sequence of the alternative ORF product GndA is shown in orange; the tryptic peptide identified by mass spectrometry (Yuan et al., 2018) is underlined.

See also Figure S5 and Table S1.

full-size *sfsA* and *birA* genes were translated *in vitro*, addition of RET resulted in the appearance of toeprint bands not only at the pTISs of the corresponding genes (Figures S5C and S5D) but also at the OOF iTIS start codons (Figure 4C, lanes ''RET,'' orange dots). The addition of the translation termination inhibitor Api137 (Florin et al., 2017) to the reactions generated toeprint bands at the stop codons of the OOF ORFs, indicating that the ribosomes not only bind to the OOF start sites but do translate the entire alternative ORFs (Figure 4C, lanes ''API,'' magenta triangles).

We then examined whether the alternative ORF in the *sfsA* gene is translated *in vivo*. For this purpose, we engineered a dual RFP/GFP reporter, where translation of the *gfp* gene is initiated at the *sfsA* OOF iTIS (Figure 4D). *E. coli* cells carrying such reporter construct actively expressed the GFP protein (Figure 4D, right), whereas *gfp* expression was abolished when the internal AUG start codon was mutated to UCG (Figure 4D). This result demonstrates that the OOF iTIS in *sfsA* is utilized for initiation of translation in the *E. coli*.

An independent validation of the functional significance of one of the Ribo-RET-identified OOF iTISs came from a recent study aimed at characterizing *E. coli* proteins activated by heat shock (Yuan et al., 2018). In that work, the sequence of one of the identified tryptic peptides mapped to the −1 frame of the *gnd* gene, although the location of the start codon from which translation of the alternative protein (named GndA) would initiate remained ambiguous (Yuan et al., 2018). Our Ribo-RET data not only validated those findings, but also suggested that translation of GndA initiates most likely at the UUG codon, which is preceded by a strong SD sequence (Figure 4E).

Expression of functional alternative proteins may be highly specialized because most of the OOF iTISs identified in the *E. coli* genome are not conserved. The strongest example that exhibits near-threshold significance of the OOF iTIS conservation is that of the *tonB* gene (Figure S4A). Furthermore, the internal initiation at this site is apparently sufficiently strong to be observed as an upshift of the local ribosome density in the Ribo-seq data collected from *E. coli* cells not treated with RET (Figure S4B).

Start-Stop Sites May Modulate Translation of the Primary Gene

Among the 74 OOF iTIS candidates, Ribo-RET revealed 14 unique sites where the start codon is immediately followed by a stop codon, and thus we called them start-stops (Table S1). Although start-stops have been identified in the 5′ UTRs of some viral and plant genes, where they likely play regulatory functions (Krummheuer et al., 2007; Tanaka et al., 2016), operational start-stops have not been reported within the bacterial genes.

We selected the identified start-stops within two genes, *yecJ* and *hslR* (Figure 5A), for further analysis. Consistently, *in vitro* studies, carried out using the full-length *yecJ* or *hslR* genes, showed that in both cases addition of initiation inhibitor RET or termination inhibitor Api137 caused the appearance of coincidental toeprint bands (either one of the inhibitors is expected to stall the ribosome at the initiation codon of the start-stop site) (Figure 5B). Thus, the iTISs of the start-stops nested in *yecJ*

and *hslR* genes can direct ribosome binding. For *in vivo* analysis, we fused the *gfp* reporter gene, devoid of its own start codon, immediately downstream of the AUG codon of the *yecJ* iTIS (stripped from its associated stop codon) (Figure 5C). GFP fluorescence derived from the resulting construct was readily detectable as long as the initiator codon of the start-stop site was intact but was significantly reduced when this AUG codon was mutated to ACG (Figure 5C). These results demonstrated that the start codon of the *yecJ* start-stop site is operational *in vivo*.

We surmised that functional OOF start-stops may carry out regulatory functions, possibly affecting the expression efficiency of the protein encoded in the main ORF. To test this hypothesis, we examined whether the presence of a functional start-stop affects the expression of the main ORF that hosts it. For that, we prepared a reporter construct where the *gfp* coding sequence was placed downstream of the *yecJ* start-stop but in frame with the *yecJ* pTIS (Figure 5D). Mutational analysis verified that expression of the YecJ-GFP fusion protein was directed by the *yecJ* pTIS (Figure 5D, wild-type [WT] versus pTIS(−) bars). Notably, when the start codon of the OOF start-stop was inactivated by mutation, the efficiency of expression of the YecJ-GFP reporter increased by ∼16% (Figure 5D, WT versus iTIS(−) bars). These results demonstrate that the presence of the active start-stop site within the *yecJ* gene attenuates translation of the main ORF, indicative of its possible regulatory function.

Interestingly, mutating the stop codon of the *yecJ* start-stop, that should lead to translation of a 14-codon internal ORF originating at the *yecJ* OOF iTIS, significantly reduced the expression of the YecJ-GFP reporter by ∼3-fold (the iSTOP(−) construct in Figure 5D). This result shows that active utilization of some of the OOF iTISs could significantly attenuate the expression of the main ORF whereas the position of the corresponding stop codon could modulate this effect.

Ribo-RET Reveals TISs Outside of the Known Annotated Coding Sequences

The ability of Ribo-RET to reveal the cryptic sites of translation initiation makes it a useful tool for identifying such sites located not only within the genes, but outside of the annotated protein coding regions. We have detected 6 upstream in-frame TISs (uTISs) in the *E. coli* strain BW25113 and 36 uTISs in the BL21 strain that would result in N-terminal extensions of the encoded proteins (Table S2). For one gene (*potB*), we did not observe any Ribo-RET peak at the annotated pTIS (Figure 6A), suggesting that either its start site has been mis-annotated or that the annotated pTIS is activated under growth conditions different from those used in our experiments. For several other genes (e.g., *yifN*), we detected Ribo-RET signals for both the annotated pTIS and the uTIS, indicating that two isoforms may be expressed (Figure 6B).

We also detected 36 TISs, common between the two analyzed strains, outside of the annotated genes likely delineating the translation start sites of unannotated short ORFs (Table S2) (analyzed in detail in another study [Weaver et al., 2019]). Although analysis of such ORFs was beyond the scope of our work, the ability to detect such ORFs underscores the utility of Ribo-RET as a general tool for the genome-wide identification of translation start sites in bacteria.
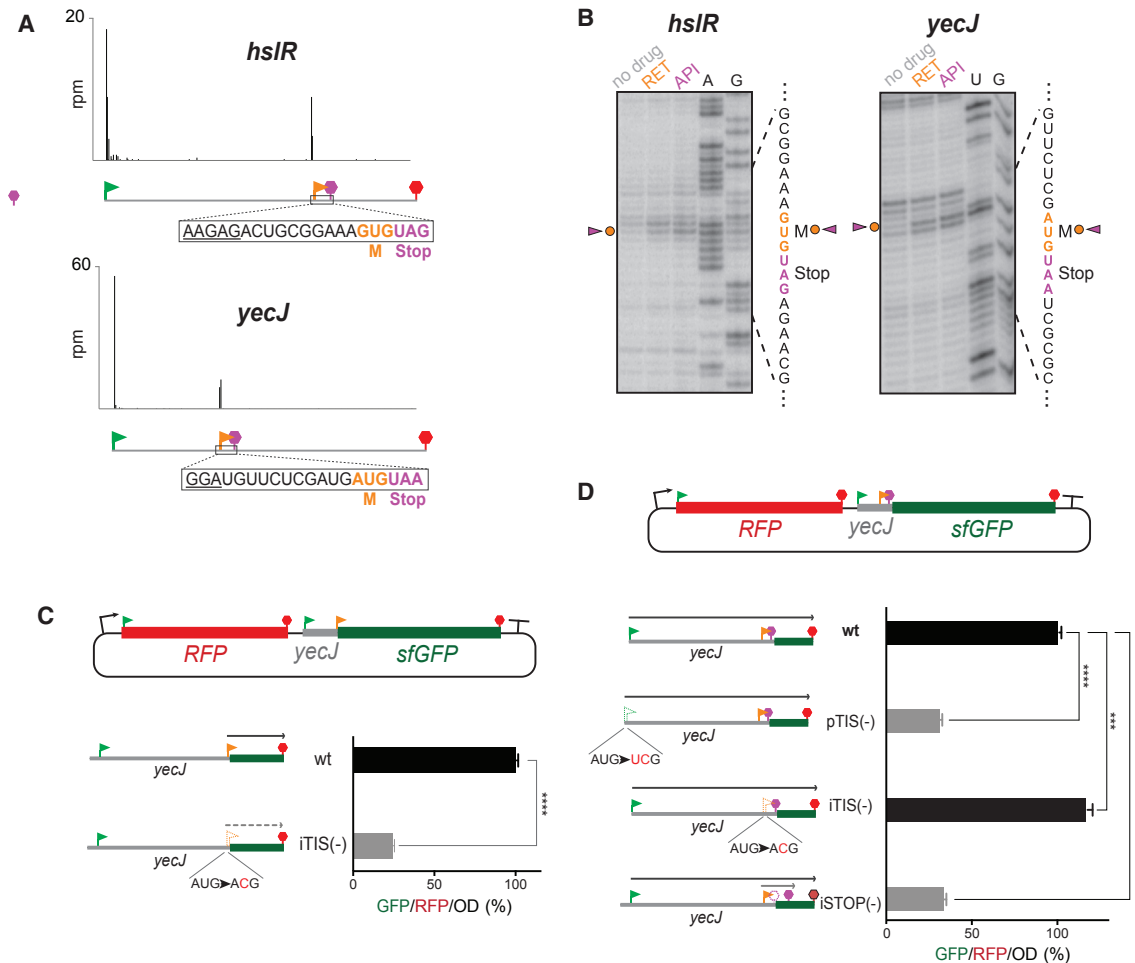
Figure 5. Start-Stops within *E. coli* Genes

(A) Representative Ribo-RET profiles revealing start-stops. The SD-like sequences are underlined.

(B) Toeprinting analysis shows ribosomes stalled at the start codons of the start-stop sites in response to the presence of initiation (RET) and termination (API) inhibitors. The start and stop codons of the start-stop sites are indicated by orange and purple characters, respectively.

(C) The start codon of the *yecJ* start-stop can direct initiation of translation *in vivo*. sfGFP expression in the RFP/sfGFP reporter is directed by the start codon of the *yecJ* start-stop (orange flag). The relative translation efficiency was estimated by measuring GFP/RFP/OD (%) ratio. The expression of *sf-gfp* is severely abrogated by a mutation that disrupts the start-stop initiation codon [iTIS (−)]. The values represent the SD from the mean in technical replicates (n = 3). Two-tailed unpaired t test, ****p < 0.0001.

(D) Start-stop impacts expression of the *yecJ* gene. The expression of the YecJ-GFP chimeric protein is controlled by *yecJ* pTIS (green flag) (*gfp* sequence is in 0-frame relative to pTIS). The reporter expression increases by ~16% when the start codon of the start-stop site is disrupted by a mutation [(iTIS(−)]. Mutating the stop codon of the start-stop site expands the length of the translated OOF coding sequence and results in severe inhibition of the main frame translation. The error bars represent SD from the mean in technical triplicates (n = 3). Two-tailed unpaired t test, ***p = 0.001, ****p < 0.0001.

See also Figure S5 and Table S2.

DISCUSSION

A genome-wide survey of TISs in two *E. coli* strains revealed translation initiation not only at the known start codons of the annotated genes, but also in the intergenic regions and, importantly, at over 100 mRNA sites nested within the currently recognized ORFs. Proteins, whose synthesis is initiated as such sites, may constitute a previously obscure fraction of the proteome and may play important roles in cell physiology. In addition, initiation of translation at internal codons may play a regulatory role

by influencing the efficiency of expression of the main protein product.

Mapping of the cryptic translation start sites was possible due to the action of RET as a highly specific inhibitor of translation initiation, arresting ribosomes at the mRNA start codons. It is this specificity of RET action that makes it possible to utilize the antibiotic for confidently charting not only the TISs at the beginning of the protein-coding sequences and also for mapping initiation-competent codons within the ORFs. Other antibiotics that exclusively bind to
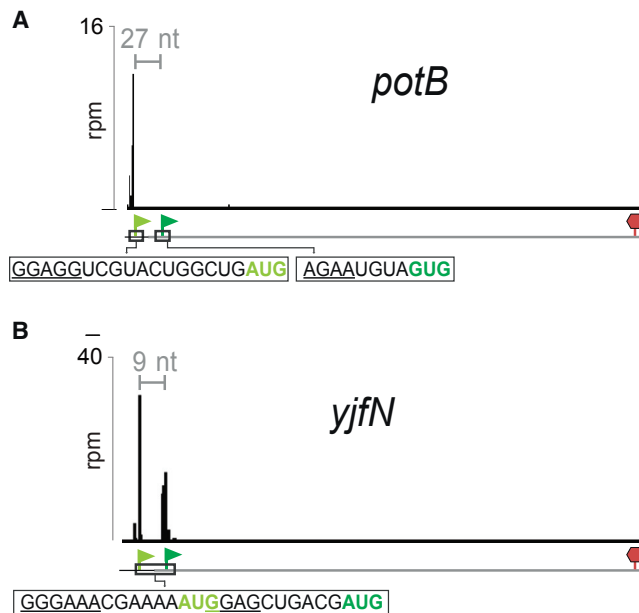
Figure 6. Examples of the Genes with TISs Identified outside of the Coding Regions

(A) No Ribo-RET peak at the annotated pTIS (dark green flag) of the *potB* gene, but a strong peak at an in-frame start codon 27 nt upstream (pale green flag).
(B) Ribo-RET peaks at the annotated pTIS (dark green flag) and at an in-frame upstream site (pale green flag) of the *yifN* gene. Putative SD sequences are underlined. See Table S2 for other cases of Ribo-RET signals outside of the coding regions.

the initiating ribosomes could also be explored for mapping TISs in bacteria (Weaver et al., 2019).

Our experiments were confined to the *E. coli* strains. However, we expect that the drug would exhibit a similar mode of action in other bacterial species. RET has limited activity against Gram-negative species, partly due to the active efflux of the drug (Jones et al., 2006). Therefore, in our experiments with RET we needed to use the *E. coli* strains lacking the TolC component of the multi-drug efflux pumps. Newer broad-spectrum pleuro-mutilins (Paukner and Riedl, 2017) could be likely used even more efficiently for mapping TISs in both Gram-positive and Gram-negative bacterial species.

Ribo-RET revealed the presence of internal start codons in over a hundred *E. coli* genes, dramatically expanding the number of putative cases of internal initiation in bacteria of which, before this work, only a few examples had been known (Meydan et al., 2018). Our findings suggest that inner-ORF initiation of translation is a much more widespread phenomenon. Although in most cases we have little knowledge about the possible functions of the alternative polypeptides encoded in the bacterial genes, one can envision several general scenarios:

(1) In-frame internal initiation generates a protein isoform with a distinct function. One such example is the ArcB-C polypeptide expressed from the iTIS within the *arcB* gene.
(2) The isoform expressed from an in-frame iTIS could partake in the hetero-complex formation with the main protein (reviewed in Meydan et al., 2018). Primary proteins encoded

by some of the iTIS-containing *E. coli* genes (e.g., *slyB*, *nudF*, *lysU*, and *wzzB*) are known to form homodimers and thus could be candidates for the formation of hetero-dimers with their N-terminally truncated isoforms.

(3) Translation from the in-frame 5′-proximal iTIS (e.g., *speA*, *bamA*, *ivy*, or *yghG*) can alter protein compartmentaliza-tion. Similarly, some of the iTISs identified in eukaryotic mRNAs and attributed to the "leaky scanning" may alter the subcellular localization of the alternative polypeptides (Kochetov, 2008; Lee et al., 2012).

(4) Because protein stability significantly depends on the nature of the N-terminal amino acid (Dougan et al., 2010), utilization of an alternative start site may alter the protein's half-life.

(5) The utilization of the OOF iTISs may generate polypep-tides with structure and function unrelated to those of the main protein.

The significance of some of the cryptic initiation sites, particularly the OOF iTISs, may reside in their regulatory role. In particular, the discovered start-stop sites within the *E. coli* genes may be utilized by the cell for fine-tuning the expression of proteins encoded in the host ORFs. However, a possible such mechanism is likely subtle, because no significant change in the ribosome density before and after the start-stop site in the Ribo-seq data collected with the untreated cells during fast growth was observed.

It is also likely that not all of the Ribo-RET-identified iTISs directly benefit bacterial cells, and a number of them could simply represent unavoidable noise of imprecise start codon recognition by the ribosome. Furthermore, the Ribo-RET peak shows only the potential of a codon to be used as a translation start site: by arresting the ribosomes at a pTIS while allowing the elongating ribosomes to run off the mRNAs, RET treatment leads to the generation of ribosome-free mRNAs, thereby allowing ribosome binding to the newly exposed putative iTISs. Nevertheless, several lines of evidence argue that many of the Ribo-RET-identified iTISs are recognized by the ribosomes even in the untreated cells: (1) for some genes (e.g., *arcB*) an increase in ribosome density downstream of the identified iTIS can be seen in the Ribo-seq data collected with the untreated cells; (2) we have experimentally demonstrated the functionality of iTISs in several genes (e.g., *arcB* and *sfsA*); and (3) the expression of an OOF ORF within the *gnd* gene was confirmed by proteomics (Yuan et al., 2018).

The mechanisms that control the relative utilization of pTISs and iTISs could operate at the level of translation, via modulating the activity of pTIS and iTIS, or at the level of transcription: some of the experimentally mapped transcription sites map reside between the pTIS and iTIS in some genes (see Table S1).

Besides iTISs, our Ribo-RET data revealed a number of the translation initiation sites outside of the annotated genes. Most of those sites delineate previously uncharacterized short genes. Proteins encoded in such ORFs may further expand the cryptic bacterial proteome (Storz et al., 2014; Weaver et al., 2019), whereas translation of the other ORFs could play regulatory roles.

CellPress

In conclusion, by mapping translation initiation landscape in bacteria, Ribo-RET unveils the hidden fraction of the bacterial proteome and offers insights into gene regulatory mechanisms.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- CONTACT FOR REAGENT AND RESOURCE SHARING
- METHOD DETAILS
  - Bacterial strains
  - Metabolic labeling of proteins
  - Ribo-seq experiments
  - Metagene analysis
  - Computational identification of translation initiation sites
  - Construction of ArcB-expressing plasmids
  - Western blot analysis of the FLAG-tagged ArcB
  - Growth competition under low oxygen conditions of cells expressing wild-type or mutant arcB genes
  - Toeprinting assay
  - Polysome analysis
  - Construction of the reporter plasmids
  - Fluorescence and cell density measurements
  - Evolutionary conservation analysis
  - Analysis of the conservation of arcB internal start site
- DATA AND SOFTWARE AVAILABILITY

SUPPLEMENTAL INFORMATION

Supplemental Information can be found with this article online at https://doi.org/10.1016/j.molcel.2019.02.017.

AUTHOR CONTRIBUTIONS

Conceptualization, S.M., J.M., N.V.-L., and A.S.M.; Methodology, S.M., N.V.-L., and A.S.M.; Software and Formal Analysis, J.M., V.S., P.V.B., A.E.F., T.M., and A.K.; Investigation, S.M. and D.K., Writing, S.M., N.V.-L., and A.S.M.; Supervision, Project Administration, and Funding Acquisition, N.V.-L. and A.S.M.

DECLARATION OF INTERESTS

The authors declare no competing interests.

SUPPORTING CITATIONS

The following references appear in the Supplemental Information: Arenz et al. (2016); Polikanov et al. (2014).

REFERENCES

Alvarez, A.F., and Georgellis, D. (2010). In vitro and in vivo analysis of the ArcB/A redox signaling pathway. Methods Enzymol. 471, 205–228.

Alvarez, A.F., Barba-Ostria, C., Silva-Jiménez, H., and Georgellis, D. (2016). Organization and mode of action of two component system signaling circuits from the various kingdoms of life. Environ. Microbiol. 18, 3210–3226.

Arenz, S., Bock, L.V., Graf, M., Innis, C.A., Beckmann, R., Grubmüller, H., Vaiana, A.C., and Wilson, D.N. (2016). A combined cryo-EM and molecular dynamics approach reveals the mechanism of ErmBL-mediated translation arrest. Nat. Commun. 7, 12026.

Atkins, J.F., Loughran, G., Bhatt, P.R., Firth, A.E., and Baranov, P.V. (2016). Ribosomal frameshifting and transcriptional slippage: From genetic steganography and cryptography to adventitious use. Nucleic Acids Res. 44, 7007–7078.

Baba, T., Ara, T., Hasegawa, M., Takai, Y., Okumura, Y., Baba, M., Datsenko, K.A., Tomita, M., Wanner, B.L., and Mori, H. (2006). Construction of Escherichia coli K-12 in-frame, single-gene knockout mutants: the Keio collection. Mol. Syst. Biol. 2, 2006.0008.

Baek, J., Lee, J., Yoon, K., and Lee, H. (2017). Identification of unannotated small genes in Salmonella. G3 (Bethesda) 7, 983–989.

Baranov, P.V., Atkins, J.F., and Yordanova, M.M. (2015). Augmented genetic decoding: global, local and temporal alterations of decoding processes and codon meaning. Nat. Rev. Genet. 16, 517–529.

Becker, A.H., Oh, E., Weissman, J.S., Kramer, G., and Bukau, B. (2013). Selective ribosome profiling as a tool for studying the interaction of chaperones and targeting factors with nascent polypeptide chains and ribosomes. Nat. Protoc. 8, 2212–2239.

Bienvenut, W.V., Giglione, C., and Meinnel, T. (2015). Proteome-wide analysis of the amino terminal status of Escherichia coli proteins at the steady-state and upon deformylation inhibition. Proteomics 15, 2503–2518.

Broome-Smith, J.K., Edelman, A., Yousif, S., and Spratt, B.G. (1985). The nucleotide sequences of the ponA and ponB genes encoding penicillin-binding protein 1A and 1B of Escherichia coli K12. Eur. J. Biochem. 147, 437–446.

Buch, J.K., and Boyle, S.M. (1985). Biosynthetic arginine decarboxylase in Escherichia coli is synthesized as a precursor and located in the cell envelope. J. Bacteriol. 163, 522–527.

Chengguang, H., Sabatini, P., Brandi, L., Giuliodori, A.M., Pon, C.L., and Gualerzi, C.O. (2017). Ribosomal selection of mRNAs with degenerate initiation triplets. Nucleic Acids Res. 45, 7309–7325.

Cundliffe, E. (1981). Antibiotic inhibitors of ribosome function. In The Molecular Basis of Antibiotic Action, E.F. Gale, E. Cundliffe, P.E. Reynolds, M.H. Richmond, and M.J. Waring, eds. (John Willey & Sons), pp. 402–545.

D'Souza, C., Nakano, M.M., and Zuber, P. (1994). Identification of comS, a gene of the srfA operon that regulates the establishment of genetic competence in Bacillus subtilis. Proc. Natl. Acad. Sci. USA 91, 9397–9401.

Datsenko, K.A., and Wanner, B.L. (2000). One-step inactivation of chromosomal genes in Escherichia coli K-12 using PCR products. Proc. Natl. Acad. Sci. USA 97, 6640–6645.

Davidovich, C., Bashan, A., Auerbach-Nevo, T., Yaggie, R.D., Gontarek, R.R., and Yonath, A. (2007). Induced-fit tightens pleuromutilins binding to ribosomes and remote interactions enable their selectivity. Proc. Natl. Acad. Sci. USA 104, 4291–4296.

Dornhelm, P., and Högenauer, G. (1978). The effects of tiamulin, a semisynthetic pleuromutilin derivative, on bacterial polypeptide chain initiation. Eur. J. Biochem. 91, 465–473.

Dougan, D.A., Truscott, K.N., and Zeth, K. (2010). The bacterial N-end rule pathway: expect the unexpected. Mol. Microbiol. 76, 545–558.

Feltens, R., Gossringer, M., Willkomm, D.K., Urlaub, H., and Hartmann, R.K. (2003). An unusual mechanism of bacterial gene expression revealed for the RNase P protein of *Thermus* strains. Proc. Natl. Acad. Sci. USA *100*, 5724–5729.

Firth, A.E. (2014). Mapping overlapping functional elements embedded within the protein-coding regions of RNA viruses. Nucleic Acids Res. *42*, 12425–12439.

Florin, T., Maracci, C., Graf, M., Karki, P., Klepacki, D., Berninghausen, O., Beckmann, R., Vázquez-Laslop, N., Wilson, D.N., Rodnina, M.V., and Mankin, A.S. (2017). An antimicrobial peptide that inhibits translation by trapping release factors on the ribosome. Nat. Struct. Mol. Biol. *24*, 752–757.

Fritsch, C., Herrmann, A., Nothnagel, M., Szafranski, K., Huse, K., Schumann, F., Schreiber, S., Platzer, M., Krawczak, M., Hampe, J., and Brosch, M. (2012). Genome-wide search for novel human uORFs and N-terminal protein extensions using ribosomal footprinting. Genome Res. *22*, 2208–2218.

Gao, T., Yang, Z., Wang, Y., and Jing, L. (2010). Identifying translation initiation sites in prokaryotes using support vector machine. J. Theor. Biol. *262*, 644–649.

Gao, X., Wan, J., Liu, B., Ma, M., Shen, B., and Qian, S.B. (2015). Quantitative profiling of initiating ribosomes in vivo. Nat. Methods *12*, 147–153.

Gibson, D.G., Young, L., Chuang, R.Y., Venter, J.C., Hutchison, C.A., 3rd, and Smith, H.O. (2009). Enzymatic assembly of DNA molecules up to several hundred kilobases. Nat. Methods *6*, 343–345.

Giess, A., Jonckheere, V., Ndah, E., Chyżyńska, K., Van Damme, P., and Valen, E. (2017). Ribosome signatures aid bacterial translation initiation site identification. BMC Biol. *15*, 76.

Grenier, F., Matteau, D., Baby, V., and Rodrigue, S. (2014). Complete Genome Sequence of *Escherichia coli* BW25113. Genome Announc. *2*, e01038-14.

Hartz, D., McPheeters, D.S., Traut, R., and Gold, L. (1988). Extension inhibition analysis of translation initiation complexes. Methods Enzymol. *164*, 419–425.

Hecht, A., Glasgow, J., Jaschke, P.R., Bawazer, L.A., Munson, M.S., Cochran, J.R., Endy, D., and Salit, M. (2017). Measurements of translation initiation from all 64 codons in *E. coli*. Nucleic Acids Res. *45*, 3615–3626.

Huerta-Cepas, J., Serra, F., and Bork, P. (2016). ETE 3: reconstruction, analysis, and visualization of phylogenomic data. Mol. Biol. Evol. *33*, 1635–1638.

Impens, F., Rolhion, N., Radoshevich, L., Bécavin, C., Duval, M., Mellin, J., García Del Portillo, F., Pucciarelli, M.G., Williams, A.H., and Cossart, P. (2017). N-terminomics identifies Prli42 as a membrane miniprotein conserved in Firmicutes and critical for stressosome activation in *Listeria monocytogenes*. Nat. Microbiol. *2*, 17005.

Ingolia, N.T., Ghaemmaghami, S., Newman, J.R., and Weissman, J.S. (2009). Genome-wide analysis in vivo of translation with nucleotide resolution using ribosome profiling. Science *324*, 218–223.

Ingolia, N.T., Lareau, L.F., and Weissman, J.S. (2011). Ribosome profiling of mouse embryonic stem cells reveals the complexity and dynamics of mammalian proteomes. Cell *147*, 789–802.

Jones, R.N., Fritsche, T.R., Sader, H.S., and Ross, J.E. (2006). Activity of retapamulin (SB-275833), a novel pleuromutilin, against selected resistant gram-positive cocci. Antimicrob. Agents Chemother. *50*, 2583–2586.

Kannan, K., Vazquez-Laslop, N., and Mankin, A.S. (2012). Selective protein synthesis by ribosomes with a drug-obstructed exit tunnel. Cell *151*, 508–520.

Kannan, K., Kanabar, P., Schryer, D., Florin, T., Oh, E., Bahroos, N., Tenson, T., Weissman, J.S., and Mankin, A.S. (2014). The general mode of translation inhibition by macrolide antibiotics. Proc. Natl. Acad. Sci. USA *111*, 15958–15963.

Kingston, A.W., Ponkratz, C., and Raleigh, E.A. (2017). Rpn (YhgA-Like) Proteins of *Escherichia coli* K-12 and their contribution to RecA-independent horizontal transfer. J. Bacteriol. *199*, e00787-16.

Kochetov, A.V. (2008). Alternative translation start sites and hidden coding potential of eukaryotic mRNAs. BioEssays *30*, 683–691.

Krummheuer, J., Johnson, A.T., Hauber, I., Kammler, S., Anderson, J.L., Hauber, J., Purcell, D.F., and Schaal, H. (2007). A minimal uORF within the

HIV-1 vpu leader allows efficient translation initiation at the downstream env AUG. Virology *363*, 261–271.

Lee, S., Liu, B., Lee, S., Huang, S.X., Shen, B., and Qian, S.B. (2012). Global mapping of translation initiation sites in mammalian cells at single-nucleotide resolution. Proc. Natl. Acad. Sci. USA *109*, E2424–E2432.

Li, G.W., Burkhardt, D., Gross, C., and Weissman, J.S. (2014). Quantifying absolute protein synthesis rates reveals principles underlying allocation of cellular resources. Cell *157*, 624–635.

Makita, Y., de Hoon, M.J., and Danchin, A. (2007). Hon-yaku: a biology-driven Bayesian methodology for identifying translation initiation sites in prokaryotes. BMC Bioinformatics *8*, 47.

Marks, J., Kannan, K., Roncase, E.J., Klepacki, D., Kefi, A., Orelle, C., Vázquez-Laslop, N., and Mankin, A.S. (2016). Context-specific inhibition of translation by ribosomal antibiotics targeting the peptidyl transferase center. Proc. Natl. Acad. Sci. USA *113*, 12150–12155.

Meydan, S., Vazquez-Laslop, N., and Mankin, A.S. (2018). Genes within genes in bacterial genomes. In Regulating with RNA in Bacteria and Archaea, G. Storz and K. Papenfort, eds. (ASM Press), pp. 133–154.

Michael, A.J. (2016). Biosynthesis of polyamines and polyamine-containing molecules. Biochem. J. *473*, 2315–2329.

Monk, J.W., Leonard, S.P., Brown, C.W., Hammerling, M.J., Mortensen, C., Gutierrez, A.E., Shin, N.Y., Watkins, E., Mishler, D.M., and Barrick, J.E. (2017). Rapid and inexpensive evaluation of nonstandard amino acid incorporation in *Escherichia coli*. ACS Synth. Biol. *6*, 45–54.

Nakahigashi, K., Takai, Y., Kimura, M., Abe, N., Nakayashiki, T., Shiwa, Y., Yoshikawa, H., Wanner, B.L., Ishihama, Y., and Mori, H. (2016). Comprehensive identification of translation start sites by tetracycline-inhibited ribosome profiling. DNA Res. *23*, 193–201.

Nakatogawa, H., and Ito, K. (2002). The ribosomal exit tunnel functions as a discriminating gate. Cell *108*, 629–636.

Notredame, C., Higgins, D.G., and Heringa, J. (2000). T-Coffee: A novel method for fast and accurate multiple sequence alignment. J. Mol. Biol. *302*, 205–217.

Orelle, C., Carlson, S., Kaushal, B., Almutairi, M.M., Liu, H., Ochabowicz, A., Quan, S., Pham, V.C., Squires, C.L., Murphy, B.T., and Mankin, A.S. (2013). Tools for characterizing bacterial protein synthesis inhibitors. Antimicrob. Agents Chemother. *57*, 5994–6004.

Park, S.K., Kim, K.I., Woo, K.M., Seol, J.H., Tanaka, K., Ichihara, A., Ha, D.B., and Chung, C.H. (1993). Site-directed mutagenesis of the dual translational initiation sites of the *clpB* gene of *Escherichia coli* and characterization of its gene products. J. Biol. Chem. *268*, 20170–20174.

Paukner, S., and Riedl, R. (2017). Pleuromutilins: potent drugs for resistant bugs-mode of action and resistance. Cold Spring Harb. Perspect. Med. *7*, a027110.

Plumbridge, J.A., Deville, F., Sacerdot, C., Petersen, H.U., Cenatiempo, Y., Cozzone, A., Grunberg-Manago, M., and Hershey, J.W. (1985). Two translational initiation sites in the *infB* gene are used to express initiation factor IF2 alpha and IF2 beta in *Escherichia coli*. EMBO J. *4*, 223–229.

Polikanov, Y.S., Steitz, T.A., and Innis, C.A. (2014). A proton wire to couple aminoacyl-tRNA accommodation and peptide-bond formation on the ribosome. Nat. Struct. Mol. Biol. *21*, 787–793.

Poulsen, S.M., Karlsson, M., Johansson, L.B., and Vester, B. (2001). The pleuromutilin drugs tiamulin and valnemulin bind to the RNA at the peptidyl transferase centre on the ribosome. Mol. Microbiol. *41*, 1091–1099.

Salmon, K.A., Hung, S.P., Steffen, N.R., Krupp, R., Baldi, P., Hatfield, G.W., and Gunsalus, R.P. (2005). Global gene expression profiling in *Escherichia coli* K12: effects of oxygen availability and ArcA. J. Biol. Chem. *280*, 15084–15096.

Salzberg, S.L., Delcher, A.L., Kasif, S., and White, O. (1998). Microbial gene identification using interpolated Markov models. Nucleic Acids Res. *26*, 544–548.

Schlünzen, F., Pyetan, E., Fucini, P., Yonath, A., and Harms, J.M. (2004). Inhibition of peptide bond formation by pleuromutilins: the structure of the

50S ribosomal subunit from *Deinococcus radiodurans* in complex with tiamulin. Mol. Microbiol. *54*, 1287–1294.

Shimizu, Y., Inoue, A., Tomari, Y., Suzuki, T., Yokogawa, T., Nishikawa, K., and Ueda, T. (2001). Cell-free translation reconstituted with purified components. Nat. Biotechnol. *19*, 751–755.

Sievers, F., Wilm, A., Dineen, D., Gibson, T.J., Karplus, K., Li, W., Lopez, R., McWilliam, H., Remmert, M., Söding, J., et al. (2011). Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. Mol. Syst. Biol. *7*, 539.

Storz, G., Wolf, Y.I., and Ramamurthi, K.S. (2014). Small proteins can no longer be ignored. Annu. Rev. Biochem. *83*, 753–777.

Studier, F.W., Daegelen, P., Lenski, R.E., Maslov, S., and Kim, J.F. (2009). Understanding the differences between genome sequences of *Escherichia coli* B strains REL606 and BL21(DE3) and comparison of the *E. coli* B and K-12 genomes. J. Mol. Biol. *394*, 653–680.

Tanaka, M., Sotta, N., Yamazumi, Y., Yamashita, Y., Miwa, K., Murota, K., Chiba, Y., Hirai, M.Y., Akiyama, T., Onouchi, H., et al. (2016). The Minimum Open Reading Frame, AUG-Stop, Induces Boron-Dependent Ribosome Stalling and mRNA Degradation. Plant Cell *28*, 2830–2849.

Weaver, J.S., Mohammad, F., Buskirk, Al., and Storz, G. (2019). Identifying small proteins by ribosome profiling with stalled initiation complexes. bioRxiv. https://doi.org/10.1101/511675.

Wilson, D., Pethica, R., Zhou, Y., Talbot, C., Vogel, C., Madera, M., Chothia, C., and Gough, J. (2009). SUPERFAMILY–sophisticated comparative genomics, data mining, visualization and phylogeny. Nucleic Acids Res. *37*, D380–D386.

Wistrand, M., and Sonnhammer, E.L. (2005). Improved profile HMM performance by assessment of critical algorithmic features in SAM and HMMER. BMC Bioinformatics *6*, 99.

Wu, W.H., and Morris, D.R. (1973). Biosynthetic arginine decarboxylase from *Escherichia coli*. Purification and properties. J. Biol. Chem. *248*, 1687–1695.

Yan, K., Madden, L., Choudhry, A.E., Voigt, C.S., Copeland, R.A., and Gontarek, R.R. (2006). Biochemical characterization of the interactions of the novel pleuromutilin derivative retapamulin with bacterial ribosomes. Antimicrob. Agents Chemother. *50*, 3875–3881.

Yuan, P., D'Lima, N.G., and Slavoff, S.A. (2018). Comparative membrane proteomics reveals a nonannotated *E. coli* heat shock protein. Biochemistry *57*, 56–60.

Zhou, J., and Rudd, K.E. (2013). EcoGene 3.0. Nucleic Acids Res. *41*, D613–D624.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
Anti-FLAG M2-Peroxidase	Sigma-Aldrich	#A8592; RRID: AB_439702
Anti-GAPDH	ThermoFisher	#MA5-15738-HRP; RRID: AB_10977387
Bacterial and Virus Strains		
BW25113 (lacI^q, rrnB$_{T14}$, ΔlacZWJ16, hsdR514, ΔaraBAD$_{AH33}$, ΔrhaBAD$_{LD78}$) ΔtolC	Kannan et al., 2012	N/A
BW25113ΔarcB (Keio collection strain)	Baba et al., 2006	JW3177
BL21, (F⁻, ompT, gal, dcm, lon, hsdS$_B$(r$_B$⁻m$_B$⁻)[malB⁺]$_{K-12}$ (λ^S)	NEB	#C2530H
BL21, (F⁻, ompT, gal, dcm, lon, hsdS$_B$(r$_B$⁻m$_B$⁻)[malB⁺]$_{K-12}$ (λ^S) ΔtolC	This paper	N/A
JM109	Promega	#P9751
Chemicals, Peptides, and Recombinant Proteins		
Retapamulin	Sigma-Aldrich	#CDS023386
L-[^{35}S]-methionine	MP Biomedicals	#MP0151001H5
Scintillation cocktail, Ultima Gold	Perkin Elmer	#6013321
RNase-free DNase I	Roche	#04716728001
SUPERase•In^{Tm} RNase inhibitor	Invitrogen	#AM2694
GMPPNP	Sigma-Aldrich	#G0635
MNase	Roche	#10107921001
B-PER Bacterial Protein Extraction Reagent	ThermoFisher	#78248
Clarity Western ECL Substrate	Bio-Rad	#170-5060
Tetracycline	Fisher scientific	#BP912-100
Api137	NovoPro Biosciences	N/A
RNase free DNase	Promega	#M610A
Kanamycin	Fisher scientific	#BP906-5
Critical Commercial Assays		
PUREexpress translation system	NEB	#E6800S
Deposited Data		
Raw and processed Ribo-seq data	This paper	GEO: GSE1221129
Unprocessed experimental data	This paper	https://doi.org/10.17632/hzk8t3w9w6.1
Oligonucleotides		
See Table S3	This paper	N/A
Recombinant DNA		
pKD46	Datsenko and Wanner, 2000	GenBank: MF287367.1
pTrc99A	N/A	GenBank: A13038.1
pArcB: pTrc99A-arcB-wt	This paper	N/A
pArcB(mut): pTrc99A-arcB-mutant	This paper	N/A
pArcB-C: pTrc99A-arcB-C	This paper	N/A
pRXG	Monk et al., 2017	Addgene#113643
pRXGSM	This paper	N/A
pRXGSM-sfsA	This paper	N/A
pRXGSM-yecJ	This paper	N/A

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Software and Algorithms		
Script1 and Script2 for initiation site assignment	This paper, Mendeley data	https://doi.org/10.17632/hzk8t3w9w6.1
Other		
35 mm 3MM paper discs	Whatman	#1030-025
TGX 4-20% gradient gel	Bio-Rad	#4561096
PVDF transfer pack	Bio-Rad	#1704156

CONTACT FOR REAGENT AND RESOURCE SHARING

Further information and requests for resources and reagents should be directed to and will be fulfilled by the Lead Contact, Alexander S. Mankin (shura@uic.edu).

METHOD DETAILS

Bacterial strains

Ribo-seq experiments were performed in two *E. coli* strains: the K12-type strain BW25113 (*lacI*^{q}, *rrnB*$_{T14}$, Δ*lacZ*$_{WJ16}$, *hsdR514*, Δ*araBAD*$_{AH33}$, Δ*rhaBAD*$_{LD78}$) that was further rendered Δ*tolC* (called previously BWDK Kannan, 2012 #81}) and the B-type strain, BL21, (F$^-$, *ompT*, *gal*, *dcm*, *lon*, *hsdS*$_B$(r$_B^-$m$_B^-$)[*malB*$^+$]$_{K-12}$ (λ^S) and was also rendered Δ*tolC* by recombineering (Datsenko and Wanner, 2000). For that, the kanamycin resistance cassette was PCR-amplified from BW25113 *tolC::kan* strain from the Keio collection (Baba et al., 2006) using the primers #P1 and P2 (Table S3). The PCR fragment was transformed into BL21 cells (NEB, #C2530H) carrying the Red recombinase expressing plasmid pKD46. After selection and verification of the BL21 *tolC::kan* clone, the kanamycin resistance marker was eliminated as previously described (Datsenko and Wanner, 2000). In the subsequent sections of STAR Methods we will refer to BW25113(Δ*tolC*) strain as 'K' strain and to BL21(Δ*tolC*) as 'B' strain.

Reporter plasmids were expressed in the *E. coli* strain JM109 (*endA1*, *recA1*, *gyrA96*, *thi*, *hsdR17* (r$_k^-$, m$_k^+$), *relA1*, *supE44*, Δ(*lac-proAB*), [F' *traD36*, *proAB*, *laqI*^{q}ZΔM15]) (Promega, #P9751).

Metabolic labeling of proteins

Inhibition of protein synthesis by RET was analyzed by metabolic labeling. Specifically, the B strain cells were grown overnight at 37°C in M9 minimal medium supplemented with 0.003 mM thiamine and 40 μg/mL of all 19 amino acids except methionine (M9AA-Met). Cells were diluted 1:200 into fresh M9AA-Met medium and grown at 37°C until the culture density reached A$_{600}$~0.2. Subsequent operations were performed at 37°C. The aliquots of cell culture (28 μL) were transferred to Eppendorf tubes that contained dried-down RET (Sigma-Aldrich, #CDS023386). The final RET concentration ranged from 1x MIC to 32x MIC (0.06 μg/mL to 2 μg/mL). After incubating cells with antibiotic for 3 min, the content was transferred to another tube containing 2 μL M9AA-Met medium supplemented with 0.3 μCi of L-[^{35}S]-methionine (specific activity 1,175 Ci/mmol) (MP Biomedicals). After 1 min incubation, 30 μL of 5% trichloracetic acid (TCA) was added to the cultures and this mixture was pipetted onto 35 mm 3MM paper discs (Whatman, Cat. No. 1030-025) pre-wetted with 25 μL of 5% TCA. The discs were then placed in a beaker with 500 mL 5% TCA and boiled for 5 min. TCA was discarded and this step was repeated one more time. Discs were rinsed in acetone, air-dried and placed in scintillation vials. After addition of 5 mL of scintillation cocktail (Perkin Elmer, Ultima Gold, #6013321) the amount of retained radioactivity was measured in a Scintillation Counter (Beckman, LS 6000). The data obtained from RET-treated cells were normalized to the no-drug control.

The time course of inhibition of protein synthesis by RET was monitored following essentially the same procedure except that antibiotic was added to a tube with the cells and 28 μL aliquots were withdrawn after specified time and added to tubes containing 2 μL M9AA-Met medium supplemented with 0.3 μCi of L-[^{35}S]-methionine. The rest of the steps were as described above.

Ribo-seq experiments

The Ribo-seq experiments were carried out following previously described procedures (Becker et al., 2013). The overnight cultures of *E. coli* grown in LB medium at 37°C were diluted to A$_{600}$~0.02 in 100 mL of fresh LB media sterilized by filtration and supplemented with 0.2% glucose. The cultures were grown at 37°C with vigorous shaking to A$_{600}$~0.5. RET was added to the final concentration of 100X MIC (12.5 μg/mL for the K strain or 5 μg/mL for the B strain) and incubated for 5 min (K strain) or 2 min (B strain). No antibiotic was added to the control no-drug cultures. Cells were harvested by rapid filtration, frozen in liquid nitrogen, cryo-lysed in 650 μL of buffer containing 20 mM Tris-HCl, pH 8.0, 10 mM MgCl$_2$, 100 mM NH$_4$Cl, 5 mM CaCl$_2$, 0.4% Triton X-100, 0.1% NP-40 and supplemented

with 65 U RNase-free DNase I (Roche, #04716728001), 208 U SUPERase·In^{Tm} RNase inhibitor (Invitrogen, #AM2694) and GMPPNP (Sigma-Aldrich, #G0635) to the final concentration of 3 mM. After clarifying the lysate by centrifugation at 20,000 g for 10 min at 4°C samples were subjected to treatment with ∼450 U MNase (Roche, #10107921001) per 25 A$_{260}$ of the cells for 60 min. The reactions were stopped by addition of EGTA to the final concentration of 5 mM and the monosome peak was isolated by sucrose gradient centrifugation. RNA was extracted and run on a 15% denaturing polyacrylamide gel. RNA fragments ranging in size from ∼28 to 45 nt were excised from the gel, eluted and used for library preparation as previously described (Becker et al., 2013). Resulting Ribo-seq data was analyzed using the GALAXY pipeline (Kannan et al., 2014). The reference genome sequences U00096.3 (BW25113, 'K' strain) and CP001509.3 (BL21, 'B' strain) were used to map the Ribo-seq reads. The first position of the P-site codon was assigned by counting 15 nucleotides from the 3′ end of the Ribo-seq reads. The Ribo-seq datasets were deposited under accession number GEO: GSE1221129.

Metagene analysis

The genes with the read counts ≥ 100 in both control and RET-treated samples were used for metagene analysis of K and B strains. The published tetracycline Ribo-seq data (Nakahigashi et al., 2016) were used to generate the corresponding metagene plot. The genes separated by less than 50 bp from the nearest neighboring gene were not included in the metagene analysis in order to avoid the 'overlapping genes' effects.

For every nucleotide of a gene, normalized reads were calculated by dividing reads per million (rpm) values assigned to a nucleotide by the total rpm count for the entire gene including 30 nt flanking regions. The metagene plot was generated by averaging the normalized reads for the region spanning 10 nt upstream and 50 nucleotides downstream of the first nucleotide of the start codon.

Computational identification of translation initiation sites

The assignment of RET peaks to the start codons was performed using the algorithm provided in Supplemental Information. Specifically, we searched for a possible start codon (AUG, GUG, CUG, UUG, AUU, AUC) within 3 nucleotides upstream or downstream of the Ribo-RET peak. All other codons associated with an internal RET peak were considered as "non-start" codons (Table S1).

For assessing whether Ribo-RET peaks in K strain match the annotated start codons in the genes expressed under no-drug conditions, we calculated the percentage of genes whose rpkm values were ≥ 100 in the no-drug conditions and whose corresponding pTIS Ribo-RET peak values were > 1 rpm. More stringent criteria were used for identification of alternative Ribo-RET peaks (rpm > 5). If the Ribo-RET peak matched an annotated TIS, it was classified as pTIS (Classification I (see schemes below) and Table S1). "Tailing peaks" (peaks within 10 nt downstream and upstream of the start codon) around the pTIS were considered as "near-annotated TIS" and merged with the pTISs after removing duplicates. All pTISs prior to duplicate removal are provided in Table S1 (the 'pTISs' tabs). The Ribo-RET peaks within coding regions were considered in Classification II (see schemes below) and were assigned as in-frame or out-of-frame iTISs depending on the position of the likely start codon (Table S1, the 'iTISs' tabs). Finally, the RET peaks outside of the coding regions were considered either as N-terminal extensions or unannotated ORFs (Classification III and Table S2). The criteria for each classification are detailed in the following schemes:
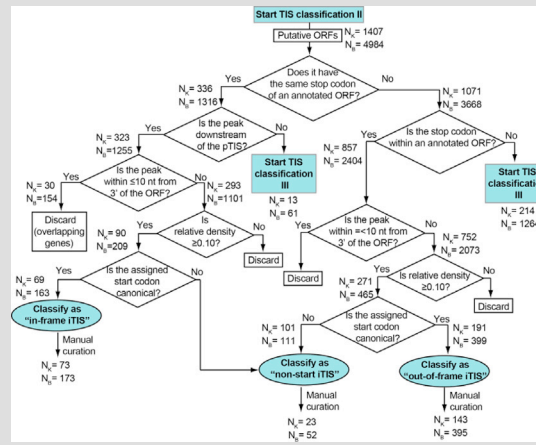
Flow chart scheme showing the criteria used to identify pTISs (Classification I)
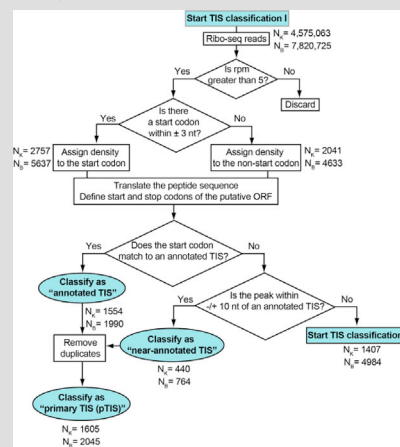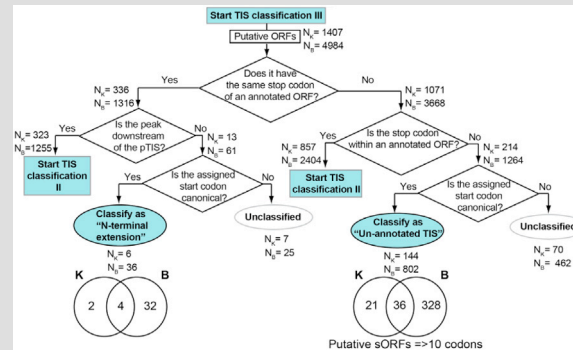
Continued

Flow chart scheme showing the criteria used to identify iTISs (Classification II)

Flow chart scheme showing the criteria used to identify N-terminal extensions and un-annotated TISs (Classification III)

Construction of ArcB-expressing plasmids

The plasmids carrying the WT *arcB* gene (pArcB) or its mutant variant pArcB(mut) (G1947A, G1950A, G1959C) were generated by Gibson assembly (Gibson et al., 2009). The PCR-generated fragments covering the length of WT or mutant *arcB* genes or of the ArcB-C coding *arcB* segment were introduced into *Nco*I and *Hind*III-cut pTrc99A plasmid. Three PCR fragments used for the assembly of WT *arcB* plasmid were generated by using primer pairs P3/P4, P5/P6 and P7/P8 (Table S3). To construct the pArcB(mut) plasmid, the PCR fragments were generated by using primer pairs P3/P9, P7/P8 and P10/P11. The plasmid pArcB-C expressing exclusively C-terminal domain of ArcB was prepared by acquiring the ArcB-C coding sequence as a gBlock (fragment #12 in Table S3) an dintroducing it into *Nco*I and *Hind*III-cut pTrc99A plasmid. All the plasmids were verified by Sanger sequencing of the inserts. The plasmids were introduced in the *E. coli* BW25113 or BW25113(*ΔarcB)* strains.

Western blot analysis of the FLAG-tagged ArcB

The BW25113 cells carrying either pArcB or pARcB(mut) plasmids (or the pArcB-C plasmid encoding the marker ArcB-C segment of ArcB) were grown overnight at 37°C in LB medium supplemented with ampicillin (final concentration of 50 μg/mL). The cultures were diluted 1:100 into 5 mL LB/ampicillin medium supplemented with 0.01 mM of isopropyl-β-D-1-thiogalactopyranoside (IPTG) and grown at 37°C until culture density reached $A_{600} \sim 0.5$. The cultures were harvested by centrifugation. Cells were resuspended in 300 μL of B-PER Bacterial Protein Extraction Reagent (Thermo Fisher, #78248) and centrifuged at 16,000 g for 10 min. Ten μL of the cell lysate were loaded on TGX 4%–20% gradient gel (Bio-Rad, #4561096). Resolved proteins were transferred to a PVDF membrane using PVDF transfer pack (Bio-Rad, #1704156) by electroblotting (Bio-Rad Trans-Blot SD. Semi-Dry Transfer Cell, 10 min at 25 V). Membrane was blocked by incubating in TBST (50 mM M Tris [pH 7.4], 150 mM NaCl, and 0.05% Tween-20) containing 5% non-fat dry milk and probed with Anti-FLAG M2-Peroxidase (Sigma-Aldrich, #A8592) and anti-GAPDH antibodies (Thermo Fisher, #MA5-15738-HRP) at 1:1000 dilution in TBST. The blot was developed using Clarity Western ECL Substrate (Bio-Rad, #170-5060) and visualized (Protein Simple, FluorChem R).

CellPress

Growth competition under low oxygen conditions of cells expressing wild-type or mutant arcB genes

Dependence of micro-aerobic cell growth on expression of *arcB* was initially verified by co-growing *E. coli* BW25113(*ΔarcB)* cells transformed with the empty vector pTrc99A or the vector carrying WT *arcB* gene (pArcB). Overnight cultures, grown in LB medium supplemented with 100 μg/mL of ampicillin, were diluted 1:100 into fresh LB supplemented with 100 μg/mL of ampicillin and 10 μM of IPTG, grown to A_{600}~0.5 and mixed in the proportion to provide equal number of pTrc99A- and pArcB cells. Plasmids from the mixed "0 passage" sample were isolated and stored. The 0 passage mix culture was diluted 1:1000 into two 14 mL culture tubes containing each 12.5 mL of fresh LB/amplicillin/IPTG medium. Tubes were tightly capped and grown vertically with no shaking at 37°C. Cell sedimentation was avoided by the slow rotation (~40 rpm) of a small magnet placed at the bottom of the tubes. After 24 h, cultures were diluted 1:1000 into tubes with fresh medium, while the rest of the cells were used for isolation of the total plasmid ("passage 1" sample). The same procedure was carried out for two more passages (passage 2 and 3). To assess the relative representation of cells with pTrc99A or pArcB the total plasmid from each of the passages 0-3 was linearized with HindIII. The 4176 bp pTrc99A DNA and 6570 bp pArcB DNA bands were resolved by agarose electrophoresis (Figure S2C).

This same low-oxygen experimental set-up was used for the growth competition of BW25113(*ΔarcB)* cells expressing WT or mutant *arcB*, from pArcB or pArcB(mut), respectively. In this case, 5 passages were performed (passages 0-5) and, instead of isolating plasmids from the cultures, cells from 500 μL aliquots of passages 0, 2, 3, 4, and 5 were collected and stored. After completing the passages, cells were resuspended in 200 μL H_2O, boiled for 10 min to lyse the cells and 1 μL of the lysate was used to PCR-amplify the segment of the *arcB* gene encompassing the iTIS region (primers P52 and P53, Table S3). The resulting PCR fragments were purified and subjected to capillary sequencing. The ratio of cells carrying WT and mutant *arcB* (that carried the G1947A, G1950A, G1959C mutations) genes was estimated by comparing the height of the sequencing peaks corresponding to the position G1959C.

Toeprinting assay

The DNA templates for toeprinting, were prepared by PCR amplification for the respective genes from the *E. coli* BW25113 genomic DNA. The following primer pairs were used for amplification of specific genes: *atpB*: P13/P14; *mqo*: P15/P16; *birA*: P18/P19; *hslR*: P30/P31; *yecJ*: P33/P34. Two point mutations were generated in *sfsA* in order to change the stop codon of the alternative ORF from TGA to TAG because in the PURE transcription-translation system, the termination inhibitor Api137 arrest termination at the TAG stop codon with a higher efficiency (Florin et al., 2017). This was achieved by first amplifying segments of the *sfsA* gene using pairs of primers P22/P23 and P24/P25 and then assembling the entire mutant *sfsA* sequence by mixing the PCR products together and re-amplifying using primers P26/P27. Toeprinting primer P17 was used with the *atpB* and *mqo* templates. Primers P20, P28, P32 and P36 were used for analysis of ribosome arrest at pTIS of *birA, sfsA*, *hslR* and *yecJ* templates, respectively. Primers P21, P29, P33 and P37 were used for the analysis of ribosome arrest at the iTISs of *birA*, *sfsA*, *hslR* and *yecJ* templates, respectively.

Transcription-translation was performed in 5 μL reactions of the PURExpress system (New England Biolabs, #E6800S) for 30 min at 37°C as previously described (Orelle et al., 2013). Final concentration of RET, tetracycline (Fisher scientific, #BP912-100) or Api137 (synthesized by NovoPro Biosciences, Inc.) was 50 μM. The primer extension products were resolved on 6% sequencing gels. Gels were dried, exposed overnight to phosphorimager screens and scanned on a Typhoon Trio phosphorimager (GE Healthcare).

Polysome analysis

For the analysis of the mechanism of RET action, the overnight culture of the K strain was diluted 1:200 in 100 mL of LB medium supplemented with 0.2% glucose.

The culture was grown at 37°C with vigorous shaking to A_{600} ~0.4 at which point RET was added to the final concentration of 100X MIC (12.5 μg/mL) (control culture was left without antibiotic). After incubation for 5 min at 37°C with shaking, cultures were transferred to pre-warmed 50 mL tubes and cells were pelleted by centrifugation in a pre-warmed 37°C Beckman JA-25 rotor at 8,000 rpm for 5 min. Pellets were resuspended in 500 μL of cold lysis buffer (20mM Tris-HCl, pH7.5, 15 mM $MgCl_2$), transferred to an Eppendorf tube and frozen in a dry ice/ethanol bath. Tubes were then thawed in an ice-cold water bath and 50 μL of freshly prepared lysozyme (10 mg/ml) was added. Freezing/thawing cycle was repeated two more times. Lysis was completed by addition of 15 μL of 10% sodium deoxycholate (Sigma, #D6750) and 2 μL (2U) of RQ1 RNase-free DNase (Promega, #M610A) followed by incubation on ice for 3 min. Lysates were clarified by centrifugation in a table-top centrifuge at 20,000 g for 15 min at 4°C. Three A_{260} of the lysate were loaded on 11 mL of 10%–40% sucrose gradient in buffer 20 mM Tris-HCl, pH 7.5, 10mM $MgCl_2$, 100 mM NH_4Cl_2, 2 mM β-mercaptoethanol. Gradients were centrifuged for 2 h in a Beckman SW-41 rotor at 39,000 rpm at 4°C. Sucrose gradients were fractionated using Piston Gradient Fractionator (Biocomp).

Construction of the reporter plasmids

The RFP/GFP plasmids were derived from the pRXG plasmid, kindly provided by Dr. Barrick (University of Texas). The vector was first reconstructed by cutting pRXG with EcoRI and SalI and re-assembling its backbone with 2 PCR fragments amplified from pRXG using primer pairs P38/P39 (*rfp* gene) and P40/P41 (*sf-gfp* preceded by a SD sequence). The resulting plasmid (pRXGSM), had RFP ORF with downstream *Spe*I sites flanking the SD-containing *sf*-GFP ORF. To generate pRXGSM-*sfsA* plasmids, *sfsA* sequences were PCR amplified from *E. coli* BW25113 genomic DNA with primers P42/P43, for the WT gene, or P42/P44, for the mutant variant, and assembled with the *Spe*I-cut pRXGSM plasmid. To generate the pRXGSM-*yecJ* reporter plasmids, the pRXGSM was cut with

*Spe*I and assembled with each of the PCR fragments generated using the following primer pairs: P45/P46 (iTIS-wt plasmid), P45/P47 (iTIS(-)); P45/P48 (pTIS-wt), P45/P50 (pTIS-iTIS(-)). pRXGSM-*yecJ*-pTIS(-) and pRXGSM-*yecJ*-pTIS-iStop(-) plasmids were generated by site directed mutagenesis of the pRXGSM-*yecJ*-pTIS-wt plasmid using primers P49 and P51, respectively.

Fluorescence and cell density measurements

E. coli JM109 cells carrying the reporter plasmids were grown overnight in LB medium supplemented with 50 μg/mL kanamycin (Fisher Scientific, #BP906-5). The cultures were then diluted 1:100 into fresh LB medium supplemented with kanamycin (50 μg/mL) and grown to A_{600} ~0.5-0.8. The cultures were diluted to the final density of A_{600} ~0.02 in fresh LB/kanamycin (50 μg/mL) medium supplemented with 0.1 mM IPTG and 120 μL were placed in the wells of a clear flat bottom 96 well microplate (Corning, #353072). The plates were placed in a Tecan Infinite M200 PRO plate reader, where they were incubated at 37°C with orbital shaking (duration: 1000 s; amplitude: 3 mm), and measurements of optical density (at 600 nm), 'green fluorescence' (excitation: 485 nm; emission: 520 nm) and 'red fluorescence' (excitation: 550 nm and emission: 675 nm) were acquired in real time.

Evolutionary conservation analysis

Protein sequences of the genes of interest were extracted from Ecogene database (Zhou and Rudd, 2013). Homologs for each gene were obtained by performing a tblastn search against the nr database. Briefly, the nr database was downloaded on 19/01/2018 to a local server and tblastn searches were performed for each gene (parameters -num_descriptions 1000000 -num_alignments 1000000 -evalue 0.0001). Only those tblastn hits which share a sequence identity of at least 45% with the query sequence and whose length is at least 75% of the query sequence were retained. Hits that contain in-frame stop codons were also discarded. Alignments for each gene of interest were generated by first translating the nucleotide sequences to protein sequences, aligning the protein sequences using Clustal-Omega (Sievers et al., 2011) and then back-translating the aligned protein sequence to their corresponding nucleotide sequence using T-coffee (Notredame et al., 2000). To analyze the conservation of internal start codons for each candidate gene, a 45-nucleotide region containing the internal start codon and 7 codons on either side of the start codon was extracted from each alignment. Sequence logos were built using this region of alignment and visualized to assess the conservation of the internal start codon. In order to determine if there is purifying selection at synonymous positions in the alignment, Synplot2 was used (Firth, 2014). Synplot2 was applied to each alignment (window size – 15 codons) and the resulting plots were visualized to assess the degree of synonymous site variability in the region of internal start codon.

Analysis of the conservation of arcB internal start site

Bacterial orthologs of *arcB* were retrieved from OrtholugeDB. Coordinates of h̲istidine-c̲ontaining p̲hosphotransfer domain (HPT domain) were determined with HMMSEARCH (Wistrand and Sonnhammer, 2005) using the model 0051674 retrieved from Superfamily database version 1.74 (Wilson et al., 2009). For predicting internal initiation site, 40 nt long fragments were extracted containing 12 codons upstream of the predicted beginning of HPT domain and 1 codon downstream. Potential SD-aSD interactions were estimated by scanning the fragment with aSD 5′-ACCUCCU-3′ using a ΔG threshold of −8.5. The presence of in-frame initiation codons (AUG, GUG, UUG) was checked. If initiation codon was found closer than 15 nt from SD-aSD sequence, the gene was reported to have in-frame iTIS. After removing redundancy for the strains of the same species, internal in-frame iTISs were identified for 26 bacterial species. Maximum Likelihood (ML) tree for 26 *arcB* sequences was computed using ETE command line tools by executing the command: "ete3 -w standard_fasttree -a arcB_protein_seq_in.fas -o ete_output" (Huerta-Cepas et al., 2016). Final figure (Figure S2B) was produced with function PhyloTree from ETE3 toolkit (Huerta-Cepas et al., 2016).

DATA AND SOFTWARE AVAILABILITY

Raw data and scripts used to identify translation initiation sites are available at Mendeley Data: https://doi.org/10.17632/hzk8t3w9w6.1. Ribo-seq datasets were deposited under accession number GEO: GSE1221129.