

Title	Retapamulin-assisted ribosome profiling reveals the alternative bacterial proteome
Authors	Meydan, Sezen;Marks, James;Klepacki, Dorota;Sharma, Virag;Baranov, Pavel V.;Firth, Andrew E.;Margus, Tonu;Kefi, Amira;Vázquez-Laslop, Nora;Mankin, Alexander S.
Publication date	2019-03-20
Original Citation	Meydan, S., Marks, J., Klepacki, D., Sharma, V., Baranov, P. V., Firth, A. E., Margus, T., Kefi, A., Vázquez-Laslop, N. and Mankin, A. S. (2019) 'Retapamulin-assisted ribosome profiling reveals the alternative bacterial proteome', <i>Molecular Cell</i> , 74(3), pp. 481-493. doi: 10.1016/j.molcel.2019.02.017
Type of publication	Article (peer-reviewed)
Link to publisher's version	<a href="http://www.sciencedirect.com/science/article/pii/S1097276519301078">http://www.sciencedirect.com/science/article/pii/S1097276519301078</a> - 10.1016/j.molcel.2019.02.017
Rights	© 2019, Elsevier Inc. All rights reserved. This manuscript version is made available under the CC BY-NC-ND 4.0 license. - <a href="https://creativecommons.org/licenses/by-nc-nd/4.0/">https://creativecommons.org/licenses/by-nc-nd/4.0/</a>
Download date	2025-07-04 05:51:28
Item downloaded from	<a href="https://hdl.handle.net/10468/7933">https://hdl.handle.net/10468/7933</a>



# UCC

**University College Cork, Ireland**  
Coláiste na hOllscoile Corcaigh



37           **SUMMARY**

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

52

## 53           **INTRODUCTION**

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

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

67           Another possible way for producing diverse polypeptides from a single ORF  
68 is the utilization of alternative internally located start codons. Although translation  
69 of the majority of the bacterial genes is initiated at a unique TIS, designated herein  
70 as primary (pTIS), several examples of genes with an additional internal TIS (iTIS)  
71 have been uncovered by detecting additional polypeptide products during the  
72 purification of the primary protein (reviewed in (Meydan et al., 2018)). In these  
73 genes, translation initiated at the pTIS results in production of the full-length  
74 (primary) protein, while ribosomes that initiate translation at the in-frame iTIS  
75 synthesize an alternative, N-terminally truncated polypeptide. Such primary and

76 alternative proteins may have related but specialized functions. The products of in-  
77 frame internal initiation at several bacterial genes have been reported to participate  
78 in various cellular functions ranging from virulence, to photosynthesis, or antibiotic  
79 production among others (reviewed in (Meydan et al., 2018)). In very few known  
80 cases, iTIS directs translation in a reading frame different from the primary ORF  
81 (D'Souza et al., 1994; Feltens et al., 2003; Yuan et al., 2018).

82 Most of the known examples of translation of a protein from an iTIS have  
83 been discovered serendipitously. Although several computational algorithms can  
84 predict the pTIS of many bacterial ORFs (Gao et al., 2010; Giess et al., 2017;  
85 Makita et al., 2007; Salzberg et al., 1998), iTIS prediction remains by far more  
86 challenging and has not even been pursued in most of those studies. The recent  
87 advent of new mass-spectrometry-based approaches have allowed the  
88 identification of N-terminal peptides of a range of proteins expressed in bacteria  
89 (Bienvenut et al., 2015; Impens et al., 2017), including some whose translation  
90 was likely initiated at an iTIS. However, the success of the available techniques  
91 for identifying such proteins is intrinsically restricted by the stringent requirements  
92 for the chemical properties, size, and abundance of the peptides that can be  
93 detected by mass-spectrometry. Therefore, the majority of the functional iTISs in  
94 the genomes likely remain overlooked.

95 Ribosome profiling (Ribo-seq), based on deep sequencing of ribosome  
96 protected mRNA fragments (“ribosome footprints”), allows for genome-wide survey  
97 of translation (Ingolia et al., 2009). Ribo-seq experiments carried out with  
98 eukaryotic cells pre-treated with the translation initiation inhibitors harringtonine

99 (Ingolia et al., 2011) and lactimidomycin (Gao et al., 2015; Lee et al., 2012) or with  
100 puromycin (Fritsch et al., 2012), showed specific enrichment of ribosome footprints  
101 at or near start codons of ORFs and facilitated mapping of TISs in eukaryotic  
102 genomes. These studies also revealed active translation of previously unknown  
103 short ORFs in the 5' UTRs of many genes and identified several TISs within the  
104 genes that were attributed to leaky scanning through the primary start sites (Lee  
105 et al., 2012). Analogous studies, however, have been difficult to conduct in bacteria  
106 because of the paucity of inhibitors with the required mechanism of action. An  
107 inhibitor useful for mapping start sites should allow the assembly of the 70S  
108 translation initiation complex at a TIS but must prevent the ribosome from leaving  
109 the start codon. Unfortunately, most of the ribosomal antibiotics traditionally viewed  
110 as initiation inhibitors do not satisfy these criteria. Recently, tetracycline (TET), an  
111 antibiotic that prevents aminoacyl-tRNAs from entering the ribosomal A site and  
112 commonly known as an elongation inhibitor (Cundliffe, 1981), was used in  
113 conjunction with Ribo-seq to globally map pTISs in the *E. coli* genome  
114 (Nakahigashi et al., 2016). Although TET Ribo-seq data successfully revealed the  
115 pTISs of many of the actively translated genes, identification of iTISs was not  
116 feasible in that work because of the substantial number of footprints generated by  
117 elongating ribosomes. Furthermore, because TET can potentially bind to the  
118 ribosome at every round of elongation cycle, when the A-site is temporarily empty,  
119 it is impossible to distinguish whether the footprints within the ORFs represented  
120 elongating ribosomes or those initiating translation at an iTIS (Nakahigashi et al.,  
121 2016).

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

133

## 134 **RESULTS**

### 135 **RET arrests the initiating ribosome at the start codons**

136 Pleuromutilin antibiotics, including clinically-used semi-synthetic RET, bind in the  
137 peptidyl transferase center (PTC) of the bacterial ribosome, hindering the  
138 placement of the P- and A-site amino acids and thus preventing peptide bond  
139 formation (Figure S1A and S1B) (Davidovich et al., 2007; Poulsen et al., 2001;  
140 Schlunzen et al., 2004). In vitro studies have shown that presence of fMet-tRNA  
141 and RET in the ribosome are not mutually exclusive (Yan et al., 2006). Therefore,  
142 we reasoned that RET may allow the assembly of the 70S initiation complex at the  
143 start codon, but by displacing the aminoacyl moiety of the initiator fMet-tRNA and

144 interfering with the placement of an elongator aminoacyl-tRNA in the A site, it could  
145 prevent formation of the first peptide bond.

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

159 The outcomes of the polysome- and toeprinting analyses, along with the  
160 structural data showing the incompatibility of the nascent protein chain with RET  
161 binding (Figure S1B), encouraged us to assess whether RET would enable the  
162 use of Ribo-seq for mapping translation start sites in bacterial cells. Even a brief  
163 (2 min) exposure of the  $\Delta toIC$  derivative of the *E. coli* strain BW25113 to a 32-fold  
164 MIC of RET nearly completely halts protein synthesis (Figure S1D). However, in  
165 Ribo-seq experiments we exposed cells for 5 min to a 100-fold MIC of RET to  
166 ensure that elongating ribosomes complete translation of even long or slowly-

167 translated ORFs prior to cell harvesting. Analysis of the Ribo-seq data showed that  
168 the RET treatment led to a striking ribosome redistribution. The occupancy of the  
169 internal and termination codons of the expressed genes was severely reduced  
170 compared to that of the untreated control, whereas the ribosome density peaks at  
171 the start codons dramatically increased (Figure 1B and 1C). Although a generally  
172 similar trend can be observed in the metagene analysis of the Ribo-seq data in the  
173 RET- (this paper) and TET-treated cells (Nakahigashi et al., 2016), the start-  
174 codons peak in the TET experiments is smaller and broader compared to the peak  
175 of the RET-stalled ribosomes (Figure S1E and S1F), reflecting a higher potency of  
176 RET as initiation inhibitor. Filtered by fairly conservative criteria (see STAR  
177 Methods), Ribo-seq data revealed distinct peaks of ribosome density at the  
178 annotated start codons (pTISs) of 991 out of 1153 (86%) *E. coli* genes expressed  
179 in the BW25113. The magnitude of the start codon peaks at the remaining 14% of  
180 the translated genes did not pass our threshold criteria (see STAR Methods)  
181 possibly reflecting changes in gene expression due to the RET treatment.

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

187

188 **Ribo-RET unmasks initiation of translation at internal codons of many**  
189 **bacterial genes**

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

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

211 Among the *E. coli* BW25113 genes expressed in our conditions, we  
212 identified 239 iTIS candidates. To further expand the systematic identification of

213 iTISs in *E. coli* genes, we applied the Ribo-RET approach to the  $\Delta tolC$  derivative  
214 of the *E. coli* strain BL21, a B-type strain which is genetically distinct from the K-  
215 strain BW25113 (Grenier et al., 2014; Studier et al., 2009). Ribo-RET analysis  
216 identified 620 iTISs in the BL21 strain. Of these, 124 iTISs were common between  
217 the two strains (Table S1). While a notably higher number of iTIS in the BL21  
218 remains somewhat puzzling, it may be related to the fact that more genes are  
219 expressed in this strain in comparison with BW25113 (1990 genes with the  
220 identified pTISs in BL21 strain vs 1554 such genes in BW25113), and to the  
221 sequence variations between the strains, as a result of which 244 BL21-specific  
222 iTISs did not have a perfect sequence match in the BW25113 strain. We limited  
223 our subsequent analysis to 124 iTISs conserved between the two strains, among  
224 which, 42 directed translation in frame with the main gene, whereas start codons  
225 of 74 iTISs were out of frame relative to the main ORF; for 8 iTISs the start site  
226 was not assigned (Figure 2C and Table S1). In the following sections we consider  
227 the first two classes separately.

228

### 229 **Internal translation initiation sites that are in frame with the main ORF**

230 The in-frame iTISs exploit various initiator codons that have been shown previously  
231 to be capable of directing translation initiation in *E. coli* (Chengguang et al., 2017;  
232 Hecht et al., 2017), although similar to the pTISs, the AUG codon is the most  
233 prevalent (Figure 3A). A SD-like sequence could be recognized upstream of many  
234 of the in-frame internal start codons (Table S1).

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

242

243 **A protein with a putative specialized function is translated from the 3'-**  
244 **proximal iTIS of the *arcB* gene**

245 The gene *arcB* encodes the sensor kinase ArcB of the two-component  
246 signal transduction system ArcB/A that helps bacteria to sense and respond to  
247 changes in oxygen concentration (Alvarez and Georgellis, 2010) (Figure 3C and  
248 3D). The ArcB protein consists of the transmitter, receiver and phosphotransfer  
249 domains (Figure 3D). Under microaerobic conditions, ArcB undergoes a series of  
250 phosphorylation steps that eventually activate the response regulator ArcA that  
251 controls expression of nearly 200 genes (Alvarez and Georgellis, 2010; Salmon et  
252 al., 2005). The C-terminal ArcB-C domain is the ultimate receiver of the phosphoryl  
253 group within the ArcB membrane-anchored protein and serves as the phosphoryl  
254 donor for ArcA (Alvarez et al., 2016).

255 The Ribo-RET data showed a strong ribosome density peak at an iTIS in  
256 *arcB*, with the putative start codon GUG located precisely at the 5' boundary of the  
257 segment encoding the ArcB-C domain (Figure 3C, D). A similarly located iTIS can

258 be found in the *arcB* gene of several bacterial species (Figure S2B). Initiation of  
259 translation at the *arcB* iTIS could generate a diffusible ArcB-C polypeptide,  
260 detached from the membrane-bound ArcB kinase (Figure 3D). To test this  
261 possibility, we introduced the 3xFLAG-coding sequence at the 3' end of the *arcB*  
262 gene, expressed it from a plasmid in *E. coli* cells and analyzed the protein products.  
263 Expression of the tagged *arcB* resulted in the simultaneous production of the full-  
264 size ArcB and of a smaller protein with apparent molecular weight (MW) of 14 kDa,  
265 consistent with that of the FLAG-tagged ArcB-C (Figures 3E and S2A). Disruption  
266 of the iTIS by synonymous mutations did not affect the synthesis of the full-length  
267 ArcB but abrogated that of ArcB-C (Figure 3E) confirming that the ArcB-C  
268 polypeptide is produced via initiation of translation at the *arcB* iTIS. Previous in  
269 vitro experiments showed that the isolated ArcB-C domain could serve as a  
270 phosphoryl acceptor and donor for the ArcB-catalyzed phosphorylation reactions  
271 (Alvarez and Georgellis, 2010), suggesting that a self-standing ArcB-C protein is  
272 likely functional in vivo. In agreement with this possibility, under micro-aerobic  
273 conditions *E. coli* cells with the operational *arcB* iTIS win over the mutant in which  
274 the iTIS is disrupted by synonymous mutations (Figures 3F and S2C). Diffusible  
275 ArcB-C may either facilitate the operation of the ArcB-ArcA signal transduction  
276 pathway or could enable a cross-talk with other signal transduction systems  
277 (Figure 3G).

278         The expression of ArcB-C from the *arcB* iTIS is apparently quite efficient  
279 because *E. coli* and *Salmonella enterica* Ribo-seq datasets show a notable upshift  
280 in the ribosome density at the *arcB* codons located downstream from the iTIS

281 (Baek et al., 2017; Kannan et al., 2014; Li et al., 2014) (Figure S2D-S2G).  
282 Curiously, the average ribosome occupancy of the *arcB* codons before and after  
283 the iTIS vary under different physiological conditions (Figure S2F and S2G),  
284 suggesting that utilization of the *arcB* pTIS and iTIS could be regulated.

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

293

#### 294 **5'-proximal iTIS gene may generate differentially-targeted proteins**

295 The *speA* gene encodes arginine decarboxylase (SpeA), an enzyme  
296 involved in polyamine production (Michael, 2016). SpeA has been found in the *E.*  
297 *coli* cytoplasmic and periplasmic fractions (Buch and Boyle, 1985) and was  
298 reported to be represented by two polypeptide isoforms, SpeA-74, with an  
299 apparent MW of 74 kDa, and a smaller one of ~ 70 kDa, SpeA-70, suggested to  
300 be a co-secretional maturation product of the full-length SpeA-74 (Buch and Boyle,  
301 1985; Wu and Morris, 1973). Our analysis, however, revealed two Ribo-RET peaks  
302 in the *speA* ORF: one corresponding to the annotated pTIS and the second one  
303 mapped to an iTIS at codon Met-26 (Figure S3A). Initiation of translation at the

304 pTIS and iTIS of *speA* would generate the 73,767 Da and 71,062 Da forms of  
305 SpeA, respectively, arguing that the SpeA-70 isoform is generated due to initiation  
306 of translation at the *speA* iTIS. In support of this conclusion, the peptide  
307 (M)SSQEASKMLR, which precisely corresponds to the N-terminus of the short  
308 SpeA isoform defined by Ribo-RET, can be found in the database of the  
309 experimentally-identified *E. coli* N-terminal peptides (Bienvenut et al., 2015).

310 It was suggested that SpeA-74 is targeted to the periplasm due to the  
311 presence of a putative N-terminal secretion signal sequence (Buch and Boyle,  
312 1985). A segment of this signal sequence would be missing in the SpeA-70 isoform  
313 confining the shorter polypeptide to the cytoplasm (Figure S3B). Therefore,  
314 utilization of the 5'-proximal iTIS of *speA* could change compartmentalization of  
315 the encoded protein. The 5'-proximal iTISs identified in some other *E. coli* genes  
316 encoding secreted proteins (e.g. *bamA*, *ivy* or *yghG*), may serve similar (Figure  
317 S3C). Analogous strategy for targeting polypeptide isoforms to different cellular  
318 compartments has been described for few other bacterial proteins (reviewed in  
319 (Meydan et al., 2018)).

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

327

**328 Conservation analysis of in-frame iTISs**

329 We analyzed alignments of bacterial genes homologous to the *E. coli* genes  
330 where internal in-frame start sites were detected by Ribo-RET. Sequence logos  
331 and codon conservation plots indicated preservation of in-frame potential initiation  
332 sites and locally enhanced synonymous site conservation at *phoH*, *speA*, *yebG*,  
333 *yfaD* and *yadD* (Figure S4A-S4E). However, it remains to be seen whether these  
334 conserved regions are relevant to promoting iTIS usage or simply represent  
335 unrelated sequence requirements of these genes. The other iTISs identified by  
336 Ribo-RET in the *E. coli* genome show a lower degree of evolutionary conservation  
337 indicating that many of them could be species- or strain-specific.

338

**339 Ribo-RET identified iTISs that are out of frame relative to the main ORF**

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

348 We tested two OOF iTIS candidates (found within the *birA* and *sfsA* genes)  
349 for their ability to direct initiation of translation. Initiation of translation at the OOF

350 UUG internal start site (overlapping the *birA* gene Leu<sub>300</sub> codon) would yield a 5-  
351 amino acid long peptide, while translation initiated at the OOF AUG of the *sfsA*  
352 gene (overlapping the Leu<sub>95</sub> codon of the main ORF) would generate a 12 amino  
353 acid peptide (Figure 4B). When the full-size *sfsA* and *birA* genes were translated  
354 in vitro, addition of RET resulted in the appearance of toeprint bands not only at  
355 the pTISs of the corresponding genes (Figure S5C and S5D) but also at the OOF  
356 iTIS start codons (Figures 4C, lanes 'RET', orange dots). The addition of the  
357 translation termination inhibitor Api137 (Florin et al., 2017) to the reactions  
358 generated toeprint bands at the stop codons of the OOF ORFs, indicating that the  
359 ribosomes not only bind to the OOF start sites but do translate the entire alternative  
360 ORFs (Figure 4C, lanes 'API', magenta triangles).

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

368 An independent validation of the functional significance of one of the Ribo-  
369 RET identified OOF iTISs came from a recent study aimed at characterizing *E. coli*  
370 proteins activated by heat-shock (Yuan et al., 2018). In that work, the sequence of  
371 one of the identified tryptic peptides mapped to the -1 frame of the *gnd* gene,  
372 although the location of the start codon from which translation of the alternative

373 protein (named GndA) would initiate remained ambiguous (Yuan et al., 2018). Our  
374 Ribo-RET data not only validated those findings, but also suggested that  
375 translation of GndA initiates most likely at the UUG codon, which is preceded by a  
376 strong SD sequence (Figure 4E).

377       Expression of functional alternative proteins may be highly specialized  
378 since most of the OOF iTISs identified in the *E. coli* genome are not conserved.  
379 The strongest example that exhibits near-threshold significance of the OOF iTIS  
380 conservation is that of the *tonB* gene (Figure S4F). Furthermore, the internal  
381 initiation at this site is apparently sufficiently strong to be observed as an upshift of  
382 the local ribosome density in the Ribo-seq data collected from *E. coli* cells not  
383 treated with RET (Figure S4F).

384

### 385       **Start-Stop sites may modulate translation of the primary gene**

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

392       We selected the identified start-stops within two genes, *yecJ* and *hslR*  
393 (Figure 5A), for further analysis. Consistently, in vitro studies, carried out using the  
394 full-length *yecJ* or *hslR* genes, showed that in both cases addition of initiation  
395 inhibitor RET or termination inhibitor Api137 caused the appearance of

396 coincidental toeprint bands (either one of the inhibitors is expected to stall the  
397 ribosome at the initiation codon of the start-stop site) (Figure 5B). Thus, the iTISs  
398 of the start-stops nested in *yecJ* and *hsIR* genes can direct ribosome binding. For  
399 in vivo analysis we fused the *gfp* reporter gene, devoid of its own start codon,  
400 immediately downstream from the AUG codon of the *yecJ* iTIS (stripped from its  
401 associated stop codon) (Figure 5C). GFP fluorescence derived from the resulting  
402 construct was readily detectable as long as the initiator codon of the start-stop site  
403 was intact, but was significantly reduced when this AUG codon was mutated to  
404 ACG (Figure 5C). These results demonstrated that the start codon of the *yecJ*  
405 start-stop site is operational in vivo.

406 We surmised that functional OOF start-stops may carry out regulatory  
407 functions, possibly affecting the expression efficiency of the protein encoded in the  
408 main ORF. To test this hypothesis, we examined whether the presence of a  
409 functional start-stop affects the expression of the main ORF that hosts it. For that,  
410 we prepared a reporter construct where the *gfp* coding sequence was placed  
411 downstream of the *yecJ* start-stop but in frame with the *yecJ* pTIS (Figure 5D).  
412 Mutational analysis verified that expression of the YecJ-GFP fusion protein was  
413 directed by the *yecJ* pTIS (Figure 5D, wt vs. pTIS(-) bars). Notably, when the start  
414 codon of the OOF start-stop was inactivated by mutation, the efficiency of  
415 expression of the YecJ-GFP reporter increased by approximately 16% (Figure 5D,  
416 wt vs. iTIS(-) bars). These results demonstrate that the presence of the active start-  
417 stop site within the *yecJ* gene attenuates translation of the main ORF, indicative of  
418 its possible regulatory function.

419           Interestingly, mutating the stop codon of the *yecJ* start-stop, that should  
420 lead to translation of a 14-codon internal ORF originating at the *yecJ* OFF iTIS,  
421 significantly reduced the expression of the YecJ-GFP reporter by ~ 3-fold (the  
422 iSTOP(-) construct in Figure 5D). This result shows that active utilization of some  
423 of the OOF iTISs could significantly attenuate the expression of the main ORF  
424 whereas the position of the corresponding stop codon could modulate this effect.  
425

### 426           **Ribo-RET reveals TISs outside of the known annotated coding** 427 **sequences**

428           The ability of Ribo-RET to reveal the cryptic sites of translation initiation  
429 makes it a useful tool for identifying such sites located not only within the genes,  
430 but outside of the annotated protein coding regions. We have detected 6 upstream  
431 in-frame TISs (uTISs) in the *E. coli* strain BW25113 and 36 uTISs in the BL21  
432 strain that would result in N-terminal extensions of the encoded proteins (Table  
433 S2). For one gene (*potB*), we did not observe any Ribo-RET peak at the annotated  
434 pTIS (Figure S6A), suggesting that either its start site has been mis-annotated or  
435 that the annotated pTIS is activated under growth conditions different from those  
436 used in our experiments. For several other genes (e.g. *yifN*), we detected Ribo-  
437 RET signals for both the annotated pTIS and the uTIS, indicating that two isoforms  
438 may be expressed (Figure S6B).

439           We also detected 41 TISs, common between the two analyzed strains,  
440 outside of the annotated genes likely delineating the translation start sites of  
441 unannotated short ORFs (Table S2) (analyzed in detail in another study (Weaver,

442 2019)). Although analysis of such ORFs was beyond the scope of our work, the  
443 ability to detect such ORFs underscores the utility of Ribo-RET as a general tool  
444 for the genome-wide identification of translation start sites in bacteria.

445

## 446 **DISCUSSION**

447 Genome-wide survey of TISs in two *E. coli* strains revealed translation  
448 initiation not only at the known start codons of the annotated genes, but also in the  
449 intergenic regions and, importantly, at over 100 mRNA sites nested within the  
450 currently recognized ORFs. Proteins, whose synthesis is initiated at such sites,  
451 may constitute a previously obscure fraction of the proteome and may play  
452 important roles in cell physiology. In addition, initiation of translation at internal  
453 codons may play a regulatory role by influencing the efficiency of expression of the  
454 main protein product.

455 Mapping of the cryptic translation start sites was possible due to the action  
456 of RET as a highly-specific inhibitor of translation initiation, arresting ribosomes at  
457 the mRNA start codons. It is this specificity of RET action that makes it possible to  
458 utilize the antibiotic for confidently charting not only the TISs at the beginning of  
459 the protein-coding sequences and also for mapping initiation-competent codons  
460 within the ORFs. Other antibiotics that exclusively bind to the initiating ribosomes,  
461 could also be explored for mapping TISs in bacteria (Weaver, 2019).

462 Our experiments were confined to the *E. coli* strains. However, we expect  
463 that the drug would exhibit a similar mode of action in other bacterial species. RET  
464 has limited activity against Gram-negative species, partly due to the active efflux

465 of the drug (Jones et al., 2006). Therefore, in our experiments with we needed to  
466 use the *E. coli* strains lacking the TolC component of the multi-drug efflux pumps.  
467 Newer broad-spectrum pleuromutilins (Paukner and Riedl, 2017), could be likely  
468 used even more efficiently for mapping TISs in both Gram-positive and Gram-  
469 negative bacterial species.

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

477 1) In-frame internal initiation generates a protein isoform with a distinct  
478 function. One such example is the ArcB-C polypeptide expressed from the iTIS  
479 within the *arcB* gene.

480 2) The isoform expressed from an in-frame iTIS could partake in the hetero-  
481 complex formation with the main protein (reviewed in (Meydan et al., 2018)).  
482 Primary proteins encoded by some of the iTIS-containing *E. coli* genes (e.g. *slyB*,  
483 *nudF*, *lysU* and *wzzB*) are known to form homodimers and thus, could be  
484 candidates for the formation of heterodimers with their N-terminally truncated  
485 isoforms.

486 3) Translation from the in-frame 5'-proximal iTIS (e.g. *speA*, *bamA*, *ivy* or  
487 *yghG*) can alter proteins compartmentalization. Similarly, some of the iTISs

488 identified in eukaryotic mRNAs and attributed to the 'leaky scanning' may alter the  
489 subcellular localization of the alternative polypeptides (Kochetov, 2008; Lee et al.,  
490 2012).

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

494 5) The utilization of the OOF iTISs may generate polypeptides with structure  
495 and function unrelated to those of the main protein.

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

503 It is also likely that not all of the Ribo-RET-identified iTISs directly benefit  
504 bacterial cell and a number of them could simply represent unavoidable noise of  
505 imprecise start codon recognition by the ribosome. Furthermore, Ribo-RET peak  
506 shows only the *potential* of a codon to be used as a translation start site: by  
507 arresting the ribosomes at a pTIS while allowing the elongating ribosomes to run  
508 off the mRNAs, RET treatment leads to the generation of ribosome-free mRNAs,  
509 thereby allowing ribosome binding to the newly exposed putative iTISs.  
510 Nevertheless, several lines of evidence argue that many of the Ribo-RET-identified

511 iTISs are recognized by the ribosomes even in the untreated cells: i) for some  
512 genes (e.g. *arcB*) an increase in ribosome density downstream of the identified  
513 iTIS can be seen in the Ribo-seq data collected with the untreated cells; ii) we have  
514 experimentally demonstrated the functionality of iTISs in several genes (e.g. *arcB*,  
515 *sfsA*); iii) the expression of an OOF ORF within the *gnd* gene was confirmed by  
516 proteomics (Yuan et al., 2018).

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

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

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

529

## 530         **ACKNOWLEDGEMENTS**

531         We thank T. Florin for help with Ribo-seq experiments, G. Storz, J. Weaver  
532 and A. Buskirk for sharing their unpublished results and useful suggestions  
533 regarding the manuscript, J.E. Barrick for providing the pRXG plasmid, D.

534 Georgellis for advice with some experiments, Y. Polikanov and N. Aleksashin for  
535 help with some figures. This work was supported by the grant from the National  
536 Science Foundation MCB 1615851 (to ASM and NV-L). PVB is supported by SFI-  
537 HRB-Wellcome Trust Biomedical Research Partnership, grant no. 210692/Z/18/Z.  
538 AEF is supported by Wellcome Trust grant no. 106207.

539

#### 540 **AUTHOR CONTRIBUTIONS**

541 Conceptualization: S.M., J.M., N.V.-L., and A.S.M.; Methodology: S.M.,  
542 N.V.-L., and A.S.M.; Software and formal analysis: J.M., V.S., P.V.B., A.E.F., T.M.,  
543 and A.K.; Investigation: S.M. and D.K., Writing: S.M., N.V.-L., and A.S.M.;  
544 Supervision, Project Administration and Funding Acquisition: N.V.-L., and A.S.M.

545

#### 546 **DECLARATIONS OF INTEREST**

547 The authors declare no competing interests.

548

## 549 REFERENCES

550

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

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

556 Atkins, J.F., Loughran, G., Bhatt, P.R., Firth, A.E., and Baranov, P.V. (2016).  
557 Ribosomal frameshifting and transcriptional slippage: From genetic  
558 steganography and cryptography to adventitious use. *Nucleic Acids Res.* *44*,  
559 7007-7078.

560 Baba, T., Ara, T., Hasegawa, M., Takai, Y., Okumura, Y., Baba, M., Datsenko,  
561 K.A., Tomita, M., Wanner, B.L., and Mori, H. (2006). Construction of  
562 *Escherichia coli* K-12 in-frame, single-gene knockout mutants: the Keio  
563 collection. *Mol. Syst. Biol.* *2*, 2006 0008.

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

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

569 Becker, A.H., Oh, E., Weissman, J.S., Kramer, G., and Bukau, B. (2013). Selective  
570 ribosome profiling as a tool for studying the interaction of chaperones and  
571 targeting factors with nascent polypeptide chains and ribosomes. *Nat. Protoc.*  
572 *8*, 2212-2239.

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

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

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

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

585 Cundliffe, E. (1981). Antibiotic Inhibitors of Ribosome Function. In *The Molecular*  
586 *Basis of Antibiotic Action*, E.F. Gale, E. Cundliffe, P.E. Reynolds, M.H.  
587 Richmond, and M.J. Waring, eds. (London, New York, Sydney, Toronto: John  
588 Willey & Sons), pp. 402-545.

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

592 Datsenko, K.A., and Wanner, B.L. (2000). One-step inactivation of chromosomal  
593 genes in *Escherichia coli* K-12 using PCR products. *Proc. Natl. Acad. Sci. USA*  
594 *97*, 6640-6645.

- 595 Davidovich, C., Bashan, A., Auerbach-Nevo, T., Yaggie, R.D., Gontarek, R.R., and  
596 Yonath, A. (2007). Induced-fit tightens pleuromutilins binding to ribosomes and  
597 remote interactions enable their selectivity. *Proc. Natl. Acad. Sci. USA* *104*,  
598 4291-4296.
- 599 Dornhelm, P., and Hogenauer, G. (1978). The effects of tiamulin, a semisynthetic  
600 pleuromutilin derivative, on bacterial polypeptide chain initiation. *Eur. J.*  
601 *Biochem.* *91*, 465-473.
- 602 Dougan, D.A., Truscott, K.N., and Zeth, K. (2010). The bacterial N-end rule  
603 pathway: expect the unexpected. *Mol. Microbiol.* *76*, 545-558.
- 604 Feltens, R., Gossringer, M., Willkomm, D.K., Urlaub, H., and Hartmann, R.K.  
605 (2003a). An unusual mechanism of bacterial gene expression revealed for the  
606 RNase P protein of *Thermus* strains. *Proc. Natl. Acad. Sci. USA* *100*, 5724-  
607 5729.
- 608 Firth, A.E. (2014). Mapping overlapping functional elements embedded within the  
609 protein-coding regions of RNA viruses. *Nucleic Acids Res.* *42*, 12425-12439.
- 610 Florin, T., Maracci, C., Graf, M., Karki, P., Klepacki, D., Berninghausen, O.,  
611 Beckmann, R., Vazquez-Laslop, N., Wilson, D.N., Rodnina, M.V., Mankin, A.  
612 S. (2017). An antimicrobial peptide that inhibits translation by trapping release  
613 factors on the ribosome. *Nat. Struct. Mol. Biol.* *24*, 752-757.
- 614 Fritsch, C., Herrmann, A., Nothnagel, M., Szafranski, K., Huse, K., Schumann, F.,  
615 Schreiber, S., Platzer, M., Krawczak, M., Hampe, J., Brosch, M. (2012).  
616 Genome-wide search for novel human uORFs and N-terminal protein  
617 extensions using ribosomal footprinting. *Genome Res.* *22*, 2208-2218.
- 618 Gao, T., Yang, Z., Wang, Y., and Jing, L. (2010). Identifying translation initiation  
619 sites in prokaryotes using support vector machine. *J. Theor. Biol.* *262*, 644-649.
- 620 Gao, X., Wan, J., Liu, B., Ma, M., Shen, B., and Qian, S.B. (2015). Quantitative  
621 profiling of initiating ribosomes in vivo. *Nat. Methods* *12*, 147-153.
- 622 Gibson, D.G., Young, L., Chuang, R.Y., Venter, J.C., Hutchison, C.A., 3rd, and  
623 Smith, H.O. (2009). Enzymatic assembly of DNA molecules up to several  
624 hundred kilobases. *Nat. Methods* *6*, 343-345.
- 625 Giess, A., Jonckheere, V., Ndah, E., Chyzynska, K., Van Damme, P., and Valen,  
626 E. (2017). Ribosome signatures aid bacterial translation initiation site  
627 identification. *BMC Biol.* *15*, 76.
- 628 Grenier, F., Matteau, D., Baby, V., and Rodrigue, S. (2014). Complete Genome  
629 Sequence of *Escherichia coli* BW25113. *Genome Announc.* *2*, e01038-14.
- 630 Hartz, D., McPheeters, D.S., Traut, R., and Gold, L. (1988). Extension inhibition  
631 analysis of translation initiation complexes. *Methods Enzymol.* *164*, 419-425.
- 632 Hecht, A., Glasgow, J., Jaschke, P.R., Bawazer, L.A., Munson, M.S., Cochran,  
633 J.R., Endy, D., and Salit, M. (2017). Measurements of translation initiation from  
634 all 64 codons in *E. coli*. *Nucleic Acids Res.* *45*, 3615-3626.
- 635 Huerta-Cepas, J., Serra, F., and Bork, P. (2016). ETE 3: reconstruction, analysis,  
636 and visualization of phylogenomic data. *Mol. Biol. Evol.* *33*, 1635-1638.
- 637 Impens, F., Rolhion, N., Radoshevich, L., Becavin, C., Duval, M., Mellin, J., Garcia  
638 Del Portillo, F., Pucciarelli, M.G., Williams, A.H., and Cossart, P. (2017). N-  
639 terminomics identifies Prli42 as a membrane miniprotein conserved in

- 640 Firmicutes and critical for stressosome activation in *Listeria monocytogenes*.  
641 Nat. Microbiol. 2, 17005.
- 642 Ingolia, N.T., Ghaemmaghami, S., Newman, J.R., and Weissman, J.S. (2009).  
643 Genome-wide analysis in vivo of translation with nucleotide resolution using  
644 ribosome profiling. Science 324, 218-223.
- 645 Ingolia, N.T., Lareau, L.F., and Weissman, J.S. (2011). Ribosome profiling of  
646 mouse embryonic stem cells reveals the complexity and dynamics of  
647 mammalian proteomes. Cell 147, 789-802.
- 648 Jones, R.N., Fritsche, T.R., Sader, H.S., and Ross, J.E. (2006). Activity of  
649 retapamulin (SB-275833), a novel pleuromutilin, against selected resistant  
650 gram-positive cocci. Antimicrob. Agents Chemother. 50, 2583-2586.
- 651 Kannan, K., Kanabar, P., Schryer, D., Florin, T., Oh, E., Bahroos, N., Tenson, T.,  
652 Weissman, J.S., and Mankin, A.S. (2014). The general mode of translation  
653 inhibition by macrolide antibiotics. Proc. Natl. Acad. Sci. USA 111, 15958-  
654 15963.
- 655 Kingston, A.W., Ponkratz, C., and Raleigh, E.A. (2017). Rpn (YhgA-Like) Proteins  
656 of *Escherichia coli* K-12 and their contribution to RecA-independent horizontal  
657 transfer. J. Bacteriol. 199, e00787-16
- 658 Kochetov, A.V. (2008). Alternative translation start sites and hidden coding  
659 potential of eukaryotic mRNAs. Bioessays 30, 683-691.
- 660 Krummheuer, J., Johnson, A.T., Hauber, I., Kammler, S., Anderson, J.L., Hauber,  
661 J., Purcell, D.F., and Schaal, H. (2007). A minimal uORF within the HIV-1 vpu  
662 leader allows efficient translation initiation at the downstream env AUG.  
663 Virology 363, 261-271.
- 664 Lee, S., Liu, B., Lee, S., Huang, S.X., Shen, B., and Qian, S.B. (2012). Global  
665 mapping of translation initiation sites in mammalian cells at single-nucleotide  
666 resolution. Proc. Natl. Acad. Sci. USA 109, E2424-2432.
- 667 Li, G.W., Burkhardt, D., Gross, C., and Weissman, J.S. (2014). Quantifying  
668 absolute protein synthesis rates reveals principles underlying allocation of  
669 cellular resources. Cell 157, 624-635.
- 670 Makita, Y., de Hoon, M.J., and Danchin, A. (2007). Hon-yaku: a biology-driven  
671 Bayesian methodology for identifying translation initiation sites in prokaryotes.  
672 BMC Bioinformatics 8, 47.
- 673 Marks, J., Kannan, K., Roncase, E.J., Klepacki, D., Kefi, A., Orelle, C., Vazquez-  
674 Laslop, N., and Mankin, A.S. (2016). Context-specific inhibition of translation  
675 by ribosomal antibiotics targeting the peptidyl transferase center. Proc. Natl.  
676 Acad. Sci. USA 113, 12150-12155.
- 677 Meydan, S., Vazquez-Laslop, N., and Mankin, A.S. (2018). Genes within genes in  
678 bacterial genomes. In Regulating with RNA in Bacteria and Archaea, G. Storz,  
679 and K. Papenfort, eds. (Washington, DC: ASM Press ) pp.133-154).
- 680 Michael, A.J. (2016). Biosynthesis of polyamines and polyamine-containing  
681 molecules. Biochem. J. 473, 2315-2329.
- 682 Miller, M.J., and Wahba, A.J. (1973). Chain initiation factor 2. Purification and  
683 properties of two species from *Escherichia coli* MRE 600. J. Biol. Chem. 248,  
684 1084-1090.

- 685 Nakahigashi, K., Takai, Y., Kimura, M., Abe, N., Nakayashiki, T., Shiwa, Y.,  
686 Yoshikawa, H., Wanner, B.L., Ishihama, Y., and Mori, H. (2016).  
687 Comprehensive identification of translation start sites by tetracycline-inhibited  
688 ribosome profiling. *DNA Res.* **23**, 193-201.
- 689 Nakatogawa, H., and Ito, K. (2002). The ribosomal exit tunnel functions as a  
690 discriminating gate. *Cell* **108**, 629-636.
- 691 Notredame, C., Higgins, D.G., and Heringa, J. (2000). T-Coffee: A novel method  
692 for fast and accurate multiple sequence alignment. *J. Mol. Biol.* **302**, 205-217.
- 693 Orelle, C., Carlson, S., Kaushal, B., Almutairi, M.M., Liu, H., Ochabowicz, A.,  
694 Quan, S., Pham, V.C., Squires, C.L., Murphy, B.T., Mankin, A. S. (2013). Tools  
695 for characterizing bacterial protein synthesis inhibitors. *Antimicrob. Agents*  
696 *Chemother.* **57**, 5994-6004.
- 697 Park, S.K., Kim, K.I., Woo, K.M., Seol, J.H., Tanaka, K., Ichihara, A., Ha, D.B., and  
698 Chung, C.H. (1993). Site-directed mutagenesis of the dual translational  
699 initiation sites of the *clpB* gene of *Escherichia coli* and characterization of its  
700 gene products. *J. Biol. Chem.* **268**, 20170-20174.
- 701 Paukner, S., and Riedl, R. (2017). Pleuromutilins: potent drugs for resistant bugs-  
702 mode of action and resistance. *Cold Spring Harb. Perspect. Med.* **7**, a027110.
- 703 Plumbridge, J.A., Deville, F., Sacerdot, C., Petersen, H.S. A., Cenatiempo, Y.,  
704 Cozzone, A., Grunberg-Manago, M., and Hershey, J.W. (1985). Two  
705 translational initiation sites in the *infB* gene are used to express initiation factor  
706 IF2 alpha and IF2 beta in *Escherichia coli*. *EMBO J.* **4**, 223-229.
- 707 Poulsen, S.M., Karlsson, M., Johansson, L.B., and Vester, B. (2001). The  
708 pleuromutilin drugs tiamulin and valnemulin bind to the RNA at the peptidyl  
709 transferase centre on the ribosome. *Mol. Microbiol.* **41**, 1091-1099.
- 710 Salmon, K.A., Hung, S.P., Steffen, N.R., Krupp, R., Baldi, P., Hatfield, G.W., and  
711 Gunsalus, R.P. (2005). Global gene expression profiling in *Escherichia coli*  
712 K12: effects of oxygen availability and ArcA. *J. Biol. Chem.* **280**, 15084-15096.
- 713 Salzberg, S.L., Delcher, A.L., Kasif, S., and White, O. (1998). Microbial gene  
714 identification using interpolated Markov models. *Nucleic Acids Res.* **26**, 544-  
715 548.
- 716 Schlunzen, F., Pyetan, E., Fucini, P., Yonath, A., and Harms, J.M. (2004).  
717 Inhibition of peptide bond formation by pleuromutilins: the structure of the 50S  
718 ribosomal subunit from *Deinococcus radiodurans* in complex with tiamulin. *Mol.*  
719 *Microbiol.* **54**, 1287-1294.
- 720 Shimizu, Y., Inoue, A., Tomari, Y., Suzuki, T., Yokogawa, T., Nishikawa, K., and  
721 Ueda, T. (2001). Cell-free translation reconstituted with purified components.  
722 *Nat. Biotechnol.* **19**, 751-755.
- 723 Sievers, F., Wilm, A., Dineen, D., Gibson, T.J., Karplus, K., Li, W., Lopez, R.,  
724 McWilliam, H., Remmert, M., Söding, J., *et al.* (2011). Fast, scalable generation  
725 of high-quality protein multiple sequence alignments using Clustal Omega. *Mol.*  
726 *Syst. Biol.* **7**, 539.
- 727 Storz, G., Wolf, Y.I., and Ramamurthi, K.S. (2014). Small proteins can no longer  
728 be ignored. *Annual Review Biochem.* **83**, 753-777.
- 729 Studier, F.W., Daegelen, P., Lenski, R.E., Maslov, S., and Kim, J.F. (2009).  
730 Understanding the differences between genome sequences of *Escherichia coli*

- 731 B strains REL606 and BL21(DE3) and comparison of the *E. coli* B and K-12  
732 genomes. *J. Mol. Biol.* *394*, 653-680.
- 733 Tanaka, M., Sotta, N., Yamazumi, Y., Yamashita, Y., Miwa, K., Murota, K., Chiba,  
734 Y., Hirai, M.Y., Akiyama, T., Onouchi, H., Naito, S., Fujiwara, T. (2016). The  
735 Minimum Open Reading Frame, AUG-Stop, Induces Boron-Dependent  
736 Ribosome Stalling and mRNA Degradation. *Plant Cell* *28*, 2830-2849.
- 737 Weaver, J.S., Mohammad, F., Buskirk, Al., Storz, G. (2019). Identifying small  
738 proteins by ribosome profiling with stalled initiation complexes. *mBio*, *in press*.
- 739 Wilson, D., Pethica, R., Zhou, Y., Talbot, C., Vogel, C., Madera, M., Chothia, C.,  
740 and Gough, J. (2009). SUPERFAMILY-sophisticated comparative genomics,  
741 data mining, visualization and phylogeny. *Nucleic Acids Res.* *37*, D380-386.
- 742 Wistrand, M., and Sonnhammer, E.L. (2005). Improved profile HMM performance  
743 by assessment of critical algorithmic features in SAM and HMMER. *BMC*  
744 *Bioinformatics* *6*, 99.
- 745 Wu, W.H., and Morris, D.R. (1973). Biosynthetic arginine decarboxylase from  
746 *Escherichia coli*. Purification and properties. *J. Biol. Chem.* *248*, 1687-1695.
- 747 Yan, K., Madden, L., Choudhry, A.E., Voigt, C.S., Copeland, R.A., and Gontarek,  
748 R.R. (2006). Biochemical characterization of the interactions of the novel  
749 pleuromutilin derivative retapamulin with bacterial ribosomes. *Antimicrob.*  
750 *Agents Chemother.* *50*, 3875-3881.
- 751 Yuan, P., D'Lima, N.G., and Slavoff, S.A. (2018). Comparative membrane  
752 proteomics reveals a nonannotated *E. coli* heat shock protein. *Biochemistry* *57*,  
753 56-60.
- 754 Zhou, J., and Rudd, K.E. (2013). EcoGene 3.0. *Nucleic Acids Res.* *41*, D613-624.  
755  
756  
757  
758

759 **MAIN FIGURE TITLES AND LEGENDS**

760 **Figure 1 RET specifically arrests ribosomes at translation initiation sites**

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

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

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

770 See also Figures S1.

771

772 **Figure 2 Ribo-RET reveals the presence of iTISs in many bacterial genes**

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

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

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

782 See also Table S1.

783

784 **Figure 3 In-frame internal initiation can generate functional N-terminally**  
785 **truncated proteins**

786 (A) The frequency of various putative start codons at the in-frame iTISs. (B) The  
787 relative length of the predicted alternative proteins, products of internal initiation,  
788 in comparison with the main protein. The known examples of genes with in-frame  
789 iTISs (Figure 2B) are in orange. The genes with the iTISs located within the 3' or  
790 5' quartile of the gene length are boxed in yellow or blue, respectively. Asterisks  
791 show genes with pTIS re-annotation proposed based on TET-assisted Ribo-seq,  
792 (Nakahigashi et al., 2016). Arrows point at *arcB* and *speA* genes further analyzed  
793 in this work.

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

796 (D) Schematics of the functional domains of ArcB. The putative alternative ArcB-C  
797 protein would encompass the phosphotransfer domain.

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

802 (F) Functional iTIS in *arcB* facilitates growth under low oxygen conditions.  
803 BW25113  $\Delta arcB$  *E. coli* cells, expressing from a plasmid either wt *arcB* or mutant  
804 *arcB* with inactivated iTIS, were co-grown in low oxygen conditions and the ratio

805 of mutant to wt cells was analyzed (see Start Methods for details). Error bars  
806 represent deviation from the mean (n=2).

807 (G) The phosphorelay across the wt ArcB domains results in the activation of the  
808 response regulator ArcA (Alvarez et al., 2016). Diffusible ArcB-C could amplify the  
809 signal capabilities of the ArcBA system and/or enable cross-talk with other  
810 response regulators.

811 See also Figure S2.

812

813 **Figure 4 OOF iTISs revealed by Ribo-RET can direct initiation of translation**

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

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

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

824 (D) In the cell, translation is initiated at the *sfsA* OOF iTIS. Schematics of the  
825 *RFP/sfGFP* reporter plasmid. The *rfp* and *sf-gfp* genes are co-transcribed. The  
826 sfGFP-coding sequence is expressed from the iTIS (orange flag) and is OOF  
827 relative to the *sfsA* pTIS start codon (green flag). The first stop codon in-frame with

828 the pTIS (red) and the stop codon in-frame with the iTIS (orange) are indicated.  
829 The bar graph shows the change in relative green fluorescence in response to the  
830 iTIS start codon mutation (mut). The values represent the mean  $\pm$  standard  
831 deviation from technical replicates (n=6). Two-tailed unpaired t-test.

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

836 See also Figure S5 and Table S1.

837

### 838 **Figure 5 Start-Stops within *E. coli* genes**

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

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

845 (C) The start codon of the *yecJ* start-stop can direct initiation of translation in vivo.  
846 sfGFP expression in the RFP/sfGFP reporter is directed by the start codon of the  
847 *yecJ* start-stop (orange flag). The relative translation efficiency was estimated by  
848 measuring GFP/RFP/OD (%) ratio. The expression of *sf-gfp* is severely abrogated  
849 by a mutation that disrupts the start-stop initiation codon [iTIS (-)]. The values

850 represent the standard deviation from the mean in technical replicates (n=3). Two-  
851 tailed unpaired t-test.

852

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

861

862 See also Figure S5 and Table S2.

## 863 STAR METHODS

864

### 865 Bacterial strains

866 Ribo-seq experiments were performed in two *E. coli* strains: the K12-type strain  
867 BW25113 (*lac*<sup>f</sup>, *rrnB*<sub>T14</sub>,  $\Delta$ *lacZ*<sub>WJ16</sub>, *hsdR514*,  $\Delta$ *araBAD*<sub>AH33</sub>,  $\Delta$ *rhaBAD*<sub>LD78</sub>) that  
868 was further rendered  $\Delta$ *tolC* (called previously BWDK Kannan, 2012 #81}) and the  
869 B-type strain, BL21, (F<sup>-</sup>, *ompT*, *gal*, *dcm*, *lon*, *hsdS*<sub>B(rB<sup>-</sup>mB<sup>-</sup>)</sub>)[*malB*<sup>+</sup>]<sub>K-12</sub> ( $\lambda$ <sup>S</sup>) and  
870 was also rendered  $\Delta$ *tolC* by recombineering (Datsenko and Wanner, 2000). For  
871 that, the kanamycin resistance cassette was PCR-amplified from BW25113  
872 *tolC::kan* strain from the Keio collection (Baba et al., 2006) using the primers #P1  
873 and P2 (Table S3). The PCR fragment was transformed into BL21 cells (NEB,  
874 #C2530H) carrying the Red recombinase expressing plasmid pKD46. After  
875 selection and verification of the BL21 *tolC::kan* clone, the kanamycin resistance  
876 marker was eliminated as previously described (Datsenko and Wanner, 2000). In  
877 the subsequent sections of STAR Methods we will refer to BW25113( $\Delta$ *tolC*)  
878 strain as 'K' strain and to BL21( $\Delta$ *tolC*) as 'B' strain.

879 Reporter plasmids were expressed in the *E. coli* strain JM109  
880 (*endA1*, *recA1*, *gyrA96*, *thi*, *hsdR17* (r<sub>k</sub><sup>-</sup>, m<sub>k</sub><sup>+</sup>), *relA1*, *supE44*,  $\Delta$ (*lac-proAB*),  
881 [F' *traD36*, *proAB*, *laqI*<sup>q</sup>Z $\Delta$ M15]) (Promega, #P9751).

882

### 883 Metabolic labeling of proteins

884 Inhibition of protein synthesis by RET was analyzed by metabolic labeling.  
885 Specifically, the B strain cells were grown overnight at 37°C in M9 minimal medium  
886 supplemented with 0.003 mM thiamine and 40  $\mu$ g/mL of all 19 amino acids except

887 methionine (M9AA-Met). Cells were diluted 1:200 into fresh M9AA-Met medium  
888 and grown at 37°C until the culture density reached  $A_{600} \sim 0.2$ . Subsequent  
889 operations were performed at 37°C. The aliquots of cell culture (28  $\mu\text{L}$ ) were  
890 transferred to Eppendorf tubes that contained dried-down RET (Sigma-Aldrich,  
891 #CDS023386). The final RET concentration ranged from 1x MIC to 32x MIC (0.06  
892  $\mu\text{g}/\text{mL}$  to 2  $\mu\text{g}/\text{mL}$ ). After incubating cells with antibiotic for 3 min, the content was  
893 transferred to another tube containing 2  $\mu\text{L}$  M9AA-Met medium supplemented with  
894 0.3  $\mu\text{Ci}$  of L-[ $^{35}\text{S}$ ]-methionine (specific activity 1,175 Ci/mmol) (MP Biomedicals).  
895 After 1 min incubation, 30  $\mu\text{L}$  of 5% trichloroacetic acid (TCA) was added to the  
896 cultures and this mixture was pipetted onto 35 mm 3MM paper discs (Whatman,  
897 Cat. No. 1030-025) pre-wetted with 25  $\mu\text{L}$  of 5% TCA. The discs were then placed  
898 in a beaker with 500 mL 5% TCA and boiled for 5 min. TCA was discarded and  
899 this step was repeated one more time. Discs were rinsed in acetone, air-dried and  
900 placed in scintillation vials. After addition of 5 ml of scintillation cocktail (Perkin  
901 Elmer, Ultima Gold, #6013321) the amount of retained radioactivity was measured  
902 in a Scintillation Counter (Beckman, LS 6000). The data obtained from RET-  
903 treated cells were normalized to the no-drug control.

904         The time course of inhibition of protein synthesis by RET was monitored  
905 following essentially the same procedure except that antibiotic was added to a tube  
906 with the cells and 28  $\mu\text{L}$  aliquots were withdrawn after specified time and added to  
907 tubes containing 2  $\mu\text{L}$  M9AA-Met medium supplemented with 0.3  $\mu\text{Ci}$  of L-[ $^{35}\text{S}$ ]-  
908 methionine. The rest of the steps were as described above.  
909

**910 Ribo-seq experiments**

911 The Ribo-seq experiments were carried out following previously described  
912 procedures (Becker et al., 2013). The overnight cultures of *E. coli* grown in LB  
913 medium at 37°C were diluted to  $A_{600} \sim 0.02$  in 100 mL of fresh LB media sterilized  
914 by filtration and supplemented with 0.2% glucose. The cultures were grown at 37°C  
915 with vigorous shaking to  $A_{600} \sim 0.5$ . RET was added to the final concentration of  
916 100X MIC (12.5  $\mu\text{g}/\text{mL}$  for the K strain or 5  $\mu\text{g}/\text{mL}$  for the B strain) and incubated  
917 for 5 min (K strain) or 2 min (B strain). No antibiotic was added to the control no-  
918 drug cultures. Cells were harvested by rapid filtration, frozen in liquid nitrogen,  
919 cryo-lysed in 650  $\mu\text{L}$  of buffer containing 20 mM Tris-HCl, pH 8.0, 10 mM  $\text{MgCl}_2$ ,  
920 100 mM  $\text{NH}_4\text{Cl}$ , 5 mM  $\text{CaCl}_2$ , 0.4% Triton X100, 0.1% NP-40 and supplemented  
921 with 65 U RNase-free DNase I (Roche, #04716728001), 208 U SUPERase•In™  
922 RNase inhibitor (Invitrogen, #AM2694) and GMPPNP (Sigma-Aldrich, #G0635) to  
923 the final concentration of 3 mM. After clarifying the lysate by centrifugation at  
924 20,000 g for 10 min at 4°C samples were subjected to treatment with  $\sim 450$  U  
925 MNase (Roche, #10107921001) per 25  $A_{260}$  of the cells for 60 min. The reactions  
926 were stopped by addition of EGTA to the final concentration of 5 mM and the  
927 monosome peak was isolated by sucrose gradient centrifugation. RNA was  
928 extracted and run on a 15% denaturing polyacrylamide gel. RNA fragments  
929 ranging in size from  $\sim 28$  to 45 nt were excised from the gel, eluted and used for  
930 library preparation as previously described (Becker et al., 2013). Resulting Ribo-  
931 seq data was analyzed using the GALAXY pipeline (Kannan et al., 2014). The  
932 reference genome sequences U00096.3 (BW25113, 'K' strain) and CP001509.3

933 (BL21, 'B' strain) were used to map the Ribo-seq reads. The first position of the P-  
934 site codon was assigned by counting 15 nucleotides from the 3' end of the Ribo-  
935 seq reads. The Ribo-seq datasets were deposited under accession number  
936 GSE1221129.

937

### 938 **Metagene analysis**

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

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

950

### 951 **Computational identification of translation initiation sites**

952 The assignment of RET peaks to the start codons was performed using the  
953 algorithm provided in Supplemental Information. Specifically, we searched for a  
954 possible start codon (AUG, GUG, CUG, UUG, AUU, AUC) within 3 nucleotides

955 upstream or downstream of the Ribo-RET peak. All other codons associated with  
956 an internal RET peak were considered as “non-start” codons (Table S1).

957 For assessing whether Ribo-RET peaks in K strain match the annotated  
958 start codons in the genes expressed under no-drug conditions, we calculated the  
959 percentage of genes whose rpkms values were  $\geq 100$  in the no-drug conditions and  
960 whose corresponding pTIS Ribo-RET peak values were  $>1$  rpm. More stringent  
961 criteria were used for identification of alternative Ribo-RET peaks (rpm  $>5$ ). If the  
962 Ribo-RET peak matched an annotated TIS, it was classified as pTIS (Classification  
963 I, Scheme I and Table S1). “Tailing peaks” (peaks within 10 nt downstream and  
964 upstream of the start codon) around the pTIS were considered as “near-annotated  
965 TIS” and merged with the pTISs after removing duplicates. All pTISs prior to  
966 duplicate removal are provided in Table S1 (the ‘pTISs’ tabs). The Ribo-RET peaks  
967 within coding regions were considered in Classification II and were assigned as in-  
968 frame or out-of-frame iTISs depending on the position of the likely start codon  
969 (Table S1, the ‘iTISs’ tabs). Finally, the RET peaks outside of the coding regions  
970 were considered either as N-terminal extensions or unannotated ORFs  
971 (Classification III and Table S2). The criteria for each classification are detailed in  
972 Schemes I, II and III in Supplementary Information.

973

#### 974 **Construction of ArcB-expressing plasmids**

975 The plasmids carrying the wt *arcB* gene (pArcB) or its mutant variant pArcB(mut)  
976 (G1947A, G1950A, G1959C) were generated by Gibson assembly (Gibson et al.,  
977 2009). The PCR-generated fragments covering the length of wt or mutant *arcB*

978 genes or of the ArcB-C coding *arcB* segment were introduced into *NcoI* and  
979 *HindIII*-cut pTrc99A plasmid. Three PCR fragments used for the assembly of wt  
980 *arcB* plasmid were generated by using primer pairs P3/P4, P5/P6 and P7/P8  
981 (Table S3). To construct the pArcB(mut) plasmid, the PCR fragments were  
982 generated by using primer pairs P3/P9, P7/P8 and P10/P11. The plasmid pArcB-  
983 C expressing exclusively C-terminal domain of ArcB was prepared by acquiring  
984 the ArcB-C coding sequence as a gBlock (fragment #12 in Table S3) and  
985 introducing it into *NcoI* and *HindIII*-cut pTrc99A plasmid. All the plasmids were  
986 verified by Sanger sequencing of the inserts. The plasmids were introduced in the  
987 *E. coli* BW25113 or BW25113( $\Delta arcB$ ) strains.

988

#### 989 **Western blot analysis of the FLAG-tagged ArcB**

990 The BW25113 cells carrying either pArcB or pArcB(mut) plasmids (or the pArcB-  
991 C plasmid encoding the marker ArcB-C segment of ArcB) were grown overnight at  
992 37°C in LB medium supplemented with ampicillin (final concentration of 50 µg/mL).  
993 The cultures were diluted 1:100 into 5 mL LB/ampicillin medium supplemented with  
994 0.01 mM of isopropyl-β-D-1-thiogalactopyranoside (IPTG) and grown at 37°C until  
995 culture density reached  $A_{600} \sim 0.5$ . The cultures were harvested by centrifugation.  
996 Cells were resuspended in 300 µL of B-PER™ Bacterial Protein Extraction  
997 Reagent (Thermo Fisher, #78248) and centrifuged at 16,000 g for 10 min. Ten µL  
998 of the cell lysate were loaded on TGX 4-20% gradient gel (Bio-Rad, #4561096).  
999 Resolved proteins were transferred to a PVDF membrane using PVDF transfer  
1000 pack (Bio-Rad, #1704156) by electroblotting (Bio-Rad Trans-Blot SD Semi-Dry

1001 Transfer Cell, 10 min at 25 V). Membrane was blocked by incubating in TBST (50  
1002 mM M Tris [pH 7.4], 150 mM NaCl, and 0.05% Tween-20) containing 5% non-fat  
1003 dry milk and probed with Anti-FLAG M2-Peroxidase (Sigma-Aldrich, #A8592) and  
1004 anti-GAPDH antibodies (Thermo Fisher, #MA5-15738-HRP) at 1:1000 dilution in  
1005 TBST. The blot was developed using Clarity Western ECL Substrate (Bio-Rad,  
1006 #170-5060) and visualized (Protein Simple, FluorChem R).

1007

1008 **Growth competition under low oxygen conditions of cells expressing wild**  
1009 **type or mutant *arcB* genes.**

1010 Dependence of micro-aerobic cell growth on expression of *arcB* was initially  
1011 verified by co-growing *E. coli* BW25113( $\Delta arcB$ ) cells transformed with the empty  
1012 vector pTrc99A or the vector carrying wt *arcB* gene (pArcB). Overnight cultures,  
1013 grown in LB medium supplemented with 100  $\mu$ g/mL of ampicillin, were diluted  
1014 1:100 into fresh LB supplemented with 100  $\mu$ g/mL of ampicillin and 10  $\mu$ M of IPTG,  
1015 grown to  $A_{600} \sim 0.5$  and mixed in the proportion to provide equal number of pTrc99A-  
1016 and pArcB cells. Plasmids from the mixed “0 passage” sample were isolated and  
1017 stored. The 0 passage mix culture was diluted 1:1000 into two 14 ml culture tubes  
1018 containing each 12.5 ml of fresh LB/ampicillin/IPTG medium. Tubes were tightly  
1019 capped and grown vertically with no shaking at 37°C. Cell sedimentation was  
1020 avoided by the slow rotation ( $\sim 40$  rpm) of a small magnet placed at the bottom of  
1021 the tubes. After 24 h, cultures were diluted 1:1000 into tubes with fresh medium,  
1022 while the rest of the cells were used for isolation of the total plasmid (“passage 1”  
1023 sample). The same procedure was carried out for two more passages (passage 2

1024 and 3). To assess the relative representation of cells with pTrc99A or pArcB the  
1025 total plasmid from each of the passages 0-3 was linearized with HindIII. The 4176  
1026 bp pTrc99A DNA and 6570 bp pArcB DNA bands were resolved by agarose  
1027 electrophoresis (Figure S2C).

1028 This same low-oxygen experimental set-up was used for the growth competition  
1029 of BW25113( $\Delta arcB$ ) cells expressing wt or mutant *arcB*, from pArcB or pArcB(mut),  
1030 respectively. In this case, 5 passages were performed (passages 0-5) and, instead  
1031 of isolating plasmids from the cultures, cells from 500  $\mu$ L aliquots of passages 0,  
1032 2, 3, 4, and 5 were collected and stored. After completing the passages, cells were  
1033 resuspended in 200  $\mu$ L H<sub>2</sub>O, boiled for 10 min to lyse the cells and 1  $\mu$ L of the  
1034 lysate was used to PCR-amplify the segment of the *arcB* gene encompassing the  
1035 iTIS region (primers P52 and P53, Table S3). The resulting PCR fragments were  
1036 purified and subjected to capillary sequencing. The ratio of cells carrying wt and  
1037 mutant *arcB* (that carried the G1947A, G1950A, G1959C mutations) genes was  
1038 estimated by comparing the height of the sequencing peaks corresponding to the  
1039 position G1959C.

1040

#### 1041 **Toeprinting assay**

1042 The DNA templates for toeprinting, were prepared by PCR amplification for the  
1043 respective genes from the *E. coli* BW25113 genomic DNA. The following primer  
1044 pairs were used for amplification of specific genes: *atpB*: P13/P14; *mgo*: P15/P16;  
1045 *birA*: P18/P19; *hslR*: P30/P31; *yecJ*: P33/P34. Two point mutations were  
1046 generated in *sfsA* in order to change the stop codon of the alternative ORF from

1047 TGA to TAG because in the PURE transcription-translation system, the termination  
1048 inhibitor Api137 arrest termination at the TAG stop codon with a higher efficiency  
1049 (Florin et al., 2017). This was achieved by first amplifying segments of the *sfsA*  
1050 gene using pairs of primers P22/P23 and P24/P25 and then assembling the entire  
1051 mutant *sfsA* sequence by mixing the PCR products together and re-amplifying  
1052 using primers P26/P27. Toeprinting primer P17 was used with the *atpB* and *mgo*  
1053 templates. Primers P20, P28, P32 and P36 were used for analysis of ribosome  
1054 arrest at pTIS of *birA*, *sfsA*, *hslR* and *yecJ* templates, respectively. Primers P21,  
1055 P29, P33 and P37 were used for the analysis of ribosome arrest at the iTISs of  
1056 *birA*, *sfsA*, *hslR* and *yecJ* templates, respectively.

1057 Transcription-translation was performed in 5  $\mu$ L reactions of the PURExpress  
1058 system (New England Biolabs, #E6800S) for 30 min at 37°C as previously  
1059 described (Orelle et al., 2013). Final concentration of RET, tetracycline (Fisher  
1060 scientific, #BP912-100) or Api137 (synthesized by NovoPro Biosciences, Inc.) was  
1061 50  $\mu$ M. The primer extension products were resolved on 6% sequencing gels. Gels  
1062 were dried, exposed overnight to phosphorimager screens and scanned on a  
1063 Typhoon Trio phosphorimager (GE Healthcare).

1064

### 1065 **Polysome analysis**

1066 For the analysis of the mechanism of RET action, the overnight culture of the K  
1067 strain was diluted 1:200 in 100 mL of LB medium supplemented with 0.2% glucose.  
1068 The culture was grown at 37°C with vigorous shaking to  $A_{600} \sim 0.4$  at which point  
1069 RET was added to the final concentration of 100X MIC (12.5  $\mu$ g/mL) (control

1070 culture was left without antibiotic). After incubation for 5 min at 37°C with shaking,  
1071 cultures were transferred to pre-warmed 50 mL tubes and cells were pelleted by  
1072 centrifugation in a pre-warmed 37°C Beckman JA-25 rotor at 8,000 rpm for 5 min.  
1073 Pellets were resuspended in 500 µL of cold lysis buffer (20mM Tris-HCl, pH7.5, 15  
1074 mM MgCl<sub>2</sub>), transferred to an Eppendorf tube and frozen in a dry ice/ethanol bath.  
1075 Tubes were then thawed in an ice-cold water bath and 50 µL of freshly prepared  
1076 lysozyme (10 mg/ml) was added. Freezing/thawing cycle was repeated two more  
1077 times. Lysis was completed by addition of 15 µL of 10% sodium deoxycholate  
1078 (Sigma, #D6750) and 2 µL (2U) of RQ1 RNase-free DNase (Promega, #M610A)  
1079 followed by incubation on ice for 3 min. Lysates were clarified by centrifugation in  
1080 a table-top centrifuge at 20,000 g for 15 min at 4°C. Three A<sub>260</sub> of the lysate were  
1081 loaded on 11 ml of 10%-40% sucrose gradient in buffer 20 mM Tris-HCl, pH 7.5,  
1082 10mM MgCl<sub>2</sub>, 100 mM NH<sub>4</sub>Cl<sub>2</sub>, 2 mM β-mercaptoethanol. Gradients were  
1083 centrifuged for 2 h in a Beckman SW-41 rotor at 39,000 rpm at 4°C. Sucrose  
1084 gradients were fractionated using Piston Gradient Fractionator (Biocomp).

1085

#### 1086 **Construction of the reporter plasmids**

1087 The RFP/GFP plasmids were derived from the pRXG plasmid, kindly provided by  
1088 Dr. Barrick (University of Texas). The vector was first reconstructed by cutting  
1089 pRXG with *EcoRI* and *SaI* and re-assembling its backbone with 2 PCR fragments  
1090 amplified from pRXG using primer pairs P38/P39 (*rfp* gene) and P40/P41 (*sf-gfp*  
1091 preceded by a SD sequence). The resulting plasmid (pRXGSM), had RFP ORF  
1092 with downstream *SpeI* sites flanking the SD-containing *sf*-GFP ORF. To generate

1093 pRXGSM-*sfsA* plasmids, *sfsA* sequences were PCR amplified from *E. coli*  
1094 BW25113 genomic DNA with primers P42/P43, for the wt gene, or P42/P44, for  
1095 the mutant variant, and assembled with the *SpeI*-cut pRXGSM plasmid. To  
1096 generate the pRXGSM-*yecJ* reporter plasmids, the pRXGSM was cut with *SpeI*  
1097 and assembled with each of the PCR fragments generated using the following  
1098 primer pairs: P45/P46 (iTIS-wt plasmid), P45/P47 (iTIS(-)); P45/P48 (pTIS-wt),  
1099 P45/P50 (pTIS-iTIS(-)). pRXGSM-*yecJ*-pTIS(-) and pRXGSM-*yecJ*-pTIS-iStop(-)  
1100 plasmids were generated by site directed mutagenesis of the pRXGSM-*yecJ*-pTIS-  
1101 wt plasmid using primers P49 and P51, respectively.

1102

### 1103 **Fluorescence and cell density measurements**

1104 *E. coli* JM109 cells carrying the reporter plasmids were grown overnight in LB  
1105 medium supplemented with 50 µg/mL kanamycin (Fisher Scientific, #BP906-5).  
1106 The cultures were then diluted 1:100 into fresh LB medium supplemented with  
1107 kanamycin (50 µg/mL) and grown to  $A_{600} \sim 0.5-0.8$ . The cultures were diluted to the  
1108 final density of  $A_{600} \sim 0.02$  in fresh LB/kanamycin (50 µg/mL) medium  
1109 supplemented with 0.1 mM IPTG and 120 µL were placed in the wells of a clear  
1110 flat bottom 96 well microplate (Corning, #353072). The plates were placed in a  
1111 Tecan Infinite M200 PRO plate reader, where they were incubated at 37°C with  
1112 orbital shaking (duration: 1000 sec; amplitude: 3 mm), and measurements of  
1113 optical density (at 600 nm), 'green fluorescence' (excitation: 485 nm; emission: 520  
1114 nm) and 'red fluorescence' (excitation: 550 nm and emission: 675 nm) were  
1115 acquired in real time.

1116

**1117 Evolutionary conservation analysis**

1118 Protein sequences of the genes of interest were extracted from Ecogene database  
1119 (Zhou and Rudd, 2013). Homologs for each gene were obtained by performing a  
1120 tblastn search against the nr database. Briefly, the nr database was downloaded  
1121 on 19/01/2018 to a local server and tblastn searches were performed for each  
1122 gene (parameters -num\_descriptions 1000000 -num\_alignments 1000000 -evalue  
1123 0.0001). Only those tblastn hits which share a sequence identity of at least 45%  
1124 with the query sequence and whose length is at least 75% of the query sequence  
1125 were retained. Hits that contain in-frame stop codons were also discarded.  
1126 Alignments for each gene of interest were generated by first translating the  
1127 nucleotide sequences to protein sequences, aligning the protein sequences using  
1128 Clustal-Omega (Sievers et al., 2011) and then back-translating the aligned protein  
1129 sequence to their corresponding nucleotide sequence using T-coffee (Notredame  
1130 et al., 2000). To analyze the conservation of internal start codons for each  
1131 candidate gene, a 45-nucleotide region containing the internal start codon and 7  
1132 codons on either side of the start codon was extracted from each alignment.  
1133 Sequence logos were built using this region of alignment and visualized to assess  
1134 the conservation of the internal start codon. In order to determine if there is  
1135 purifying selection at synonymous positions in the alignment, Synplot2 was used  
1136 (Firth, 2014). Synplot2 was applied to each alignment (window size – 15 codons)  
1137 and the resulting plots were visualized to assess the degree of synonymous site  
1138 variability in the region of internal start codon.

1139

1140 **Analysis of the conservation of *arcB* internal start site**

1141 Bacterial orthologs of *arcB* were retrieved from OrtholugeDB. Coordinates of  
1142 histidine-containing phosphotransfer domain (HPT domain) were determined with  
1143 HMMSEARCH (Wistrand and Sonnhammer, 2005) using the model 0051674  
1144 retrieved from Superfamily database version 1.74 (Wilson et al., 2009). For  
1145 predicting internal initiation site, 40 nt long fragments were extracted containing 12  
1146 codons upstream of the predicted beginning of HPT domain and 1 codon  
1147 downstream. Potential SD-aSD interactions were estimated by scanning the  
1148 fragment with aSD 5'-ACCUCCU-3' using a  $\Delta G$  threshold of -8.5. The presence of  
1149 in-frame initiation codons (AUG, GUG, UUG) was checked. If initiation codon was  
1150 found closer than 15 nt from SD-aSD sequence, the gene was reported to have in-  
1151 frame iTIS. After removing redundancy for the strains of the same species, internal  
1152 in-frame iTISs were identified for 26 bacterial species. Maximum Likelihood (ML)  
1153 tree for 26 *arcB* sequences was computed using ETE command line tools by  
1154 executing the command: "ete3 -w standard\_fasttree -a arcB\_protein\_seq\_in.fas -  
1155 o ete\_output" (Huerta-Cepas et al., 2016). Final figure (Figure S2B) was produced  
1156 with function PhyloTree from ETE3 toolkit (Huerta-Cepas et al., 2016).

1157 **Table S1:** List of common and strain-specific iTISs identified by Ribo-RET in the  
1158 *E. coli* strains BW25113 and BL21, related to Figures 2 and 3.

1159

1160 **Table S2:** List of common and strain-specific TISs identified by Ribo-RET in the *E.*  
1161 *coli* strains BW25113 and BL21 outside of the annotated genes, related to Figure  
1162 5.

1163

Figure 1

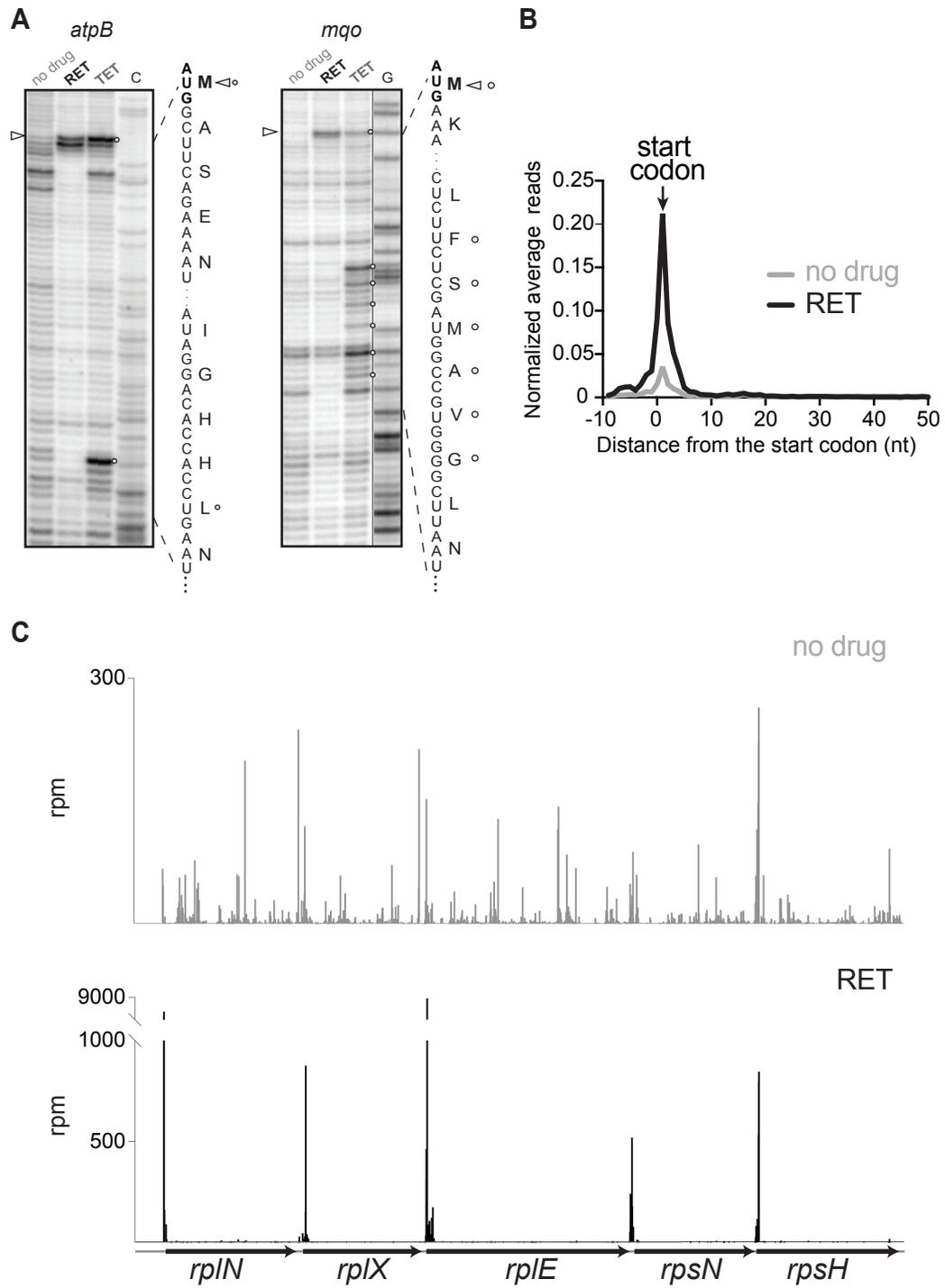


Figure 2

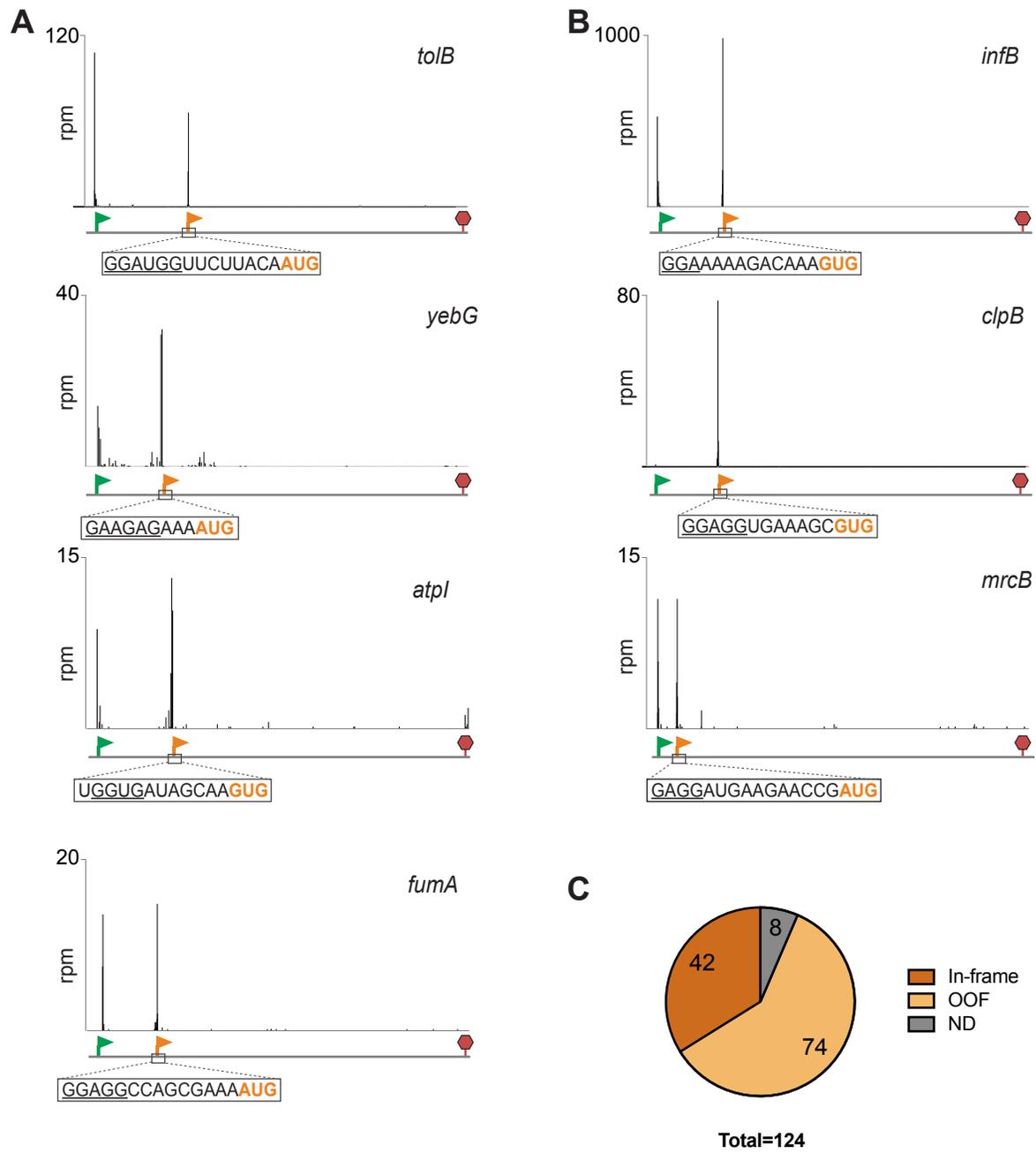


Figure 3

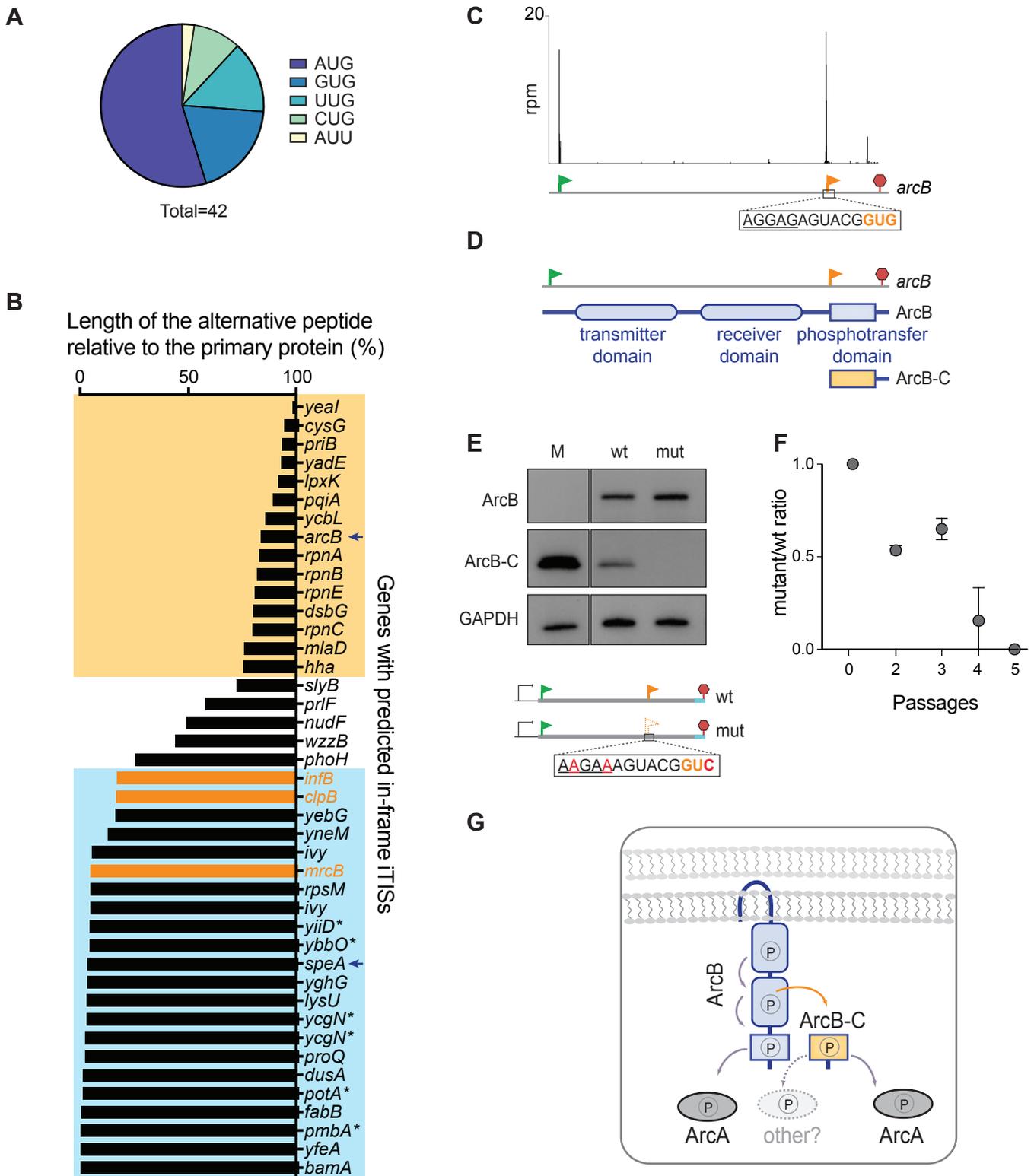


Figure 4

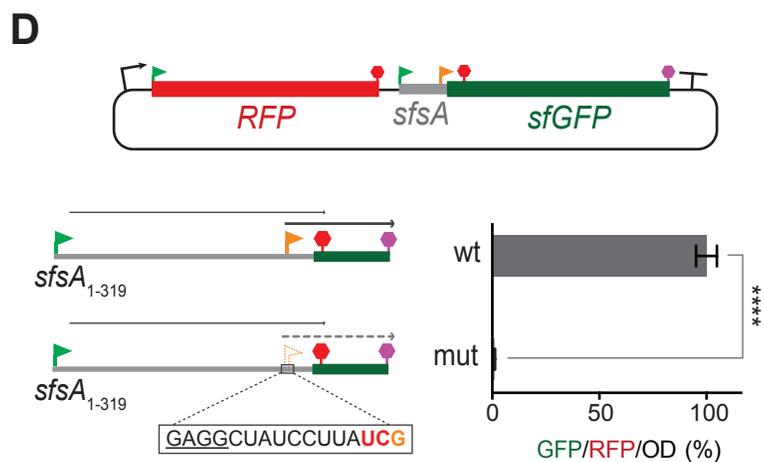
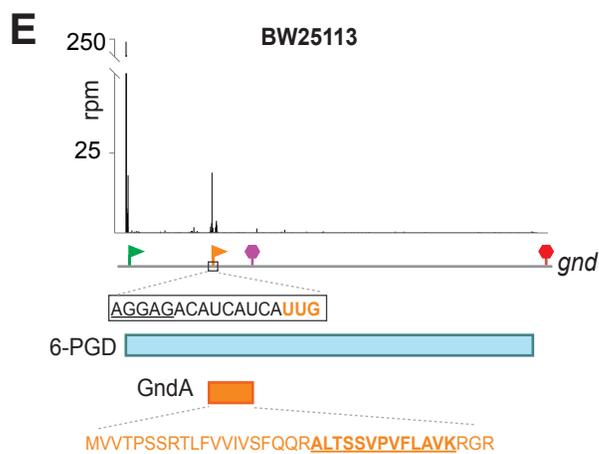
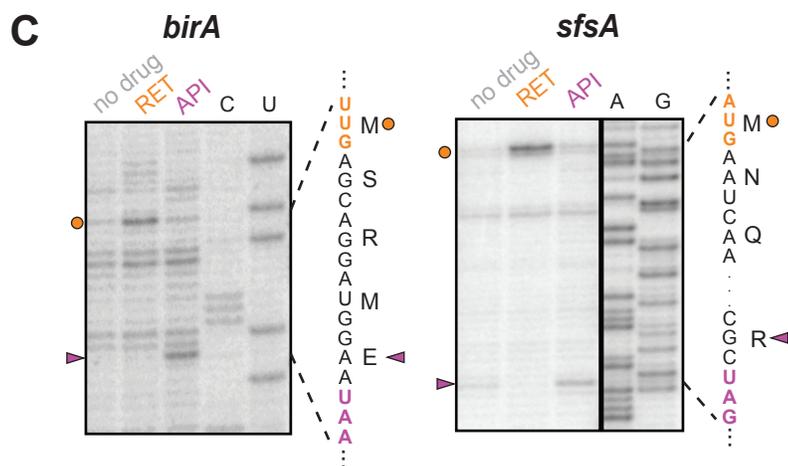
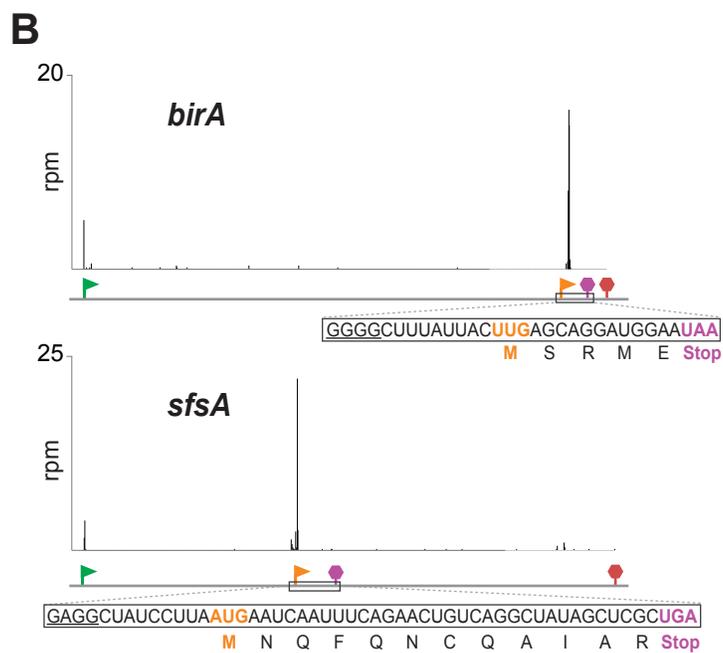
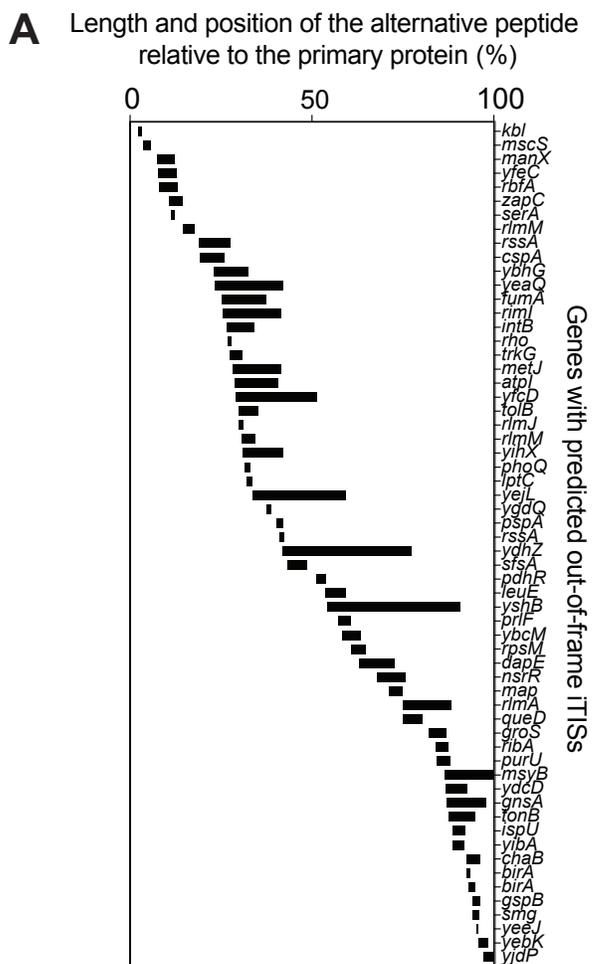
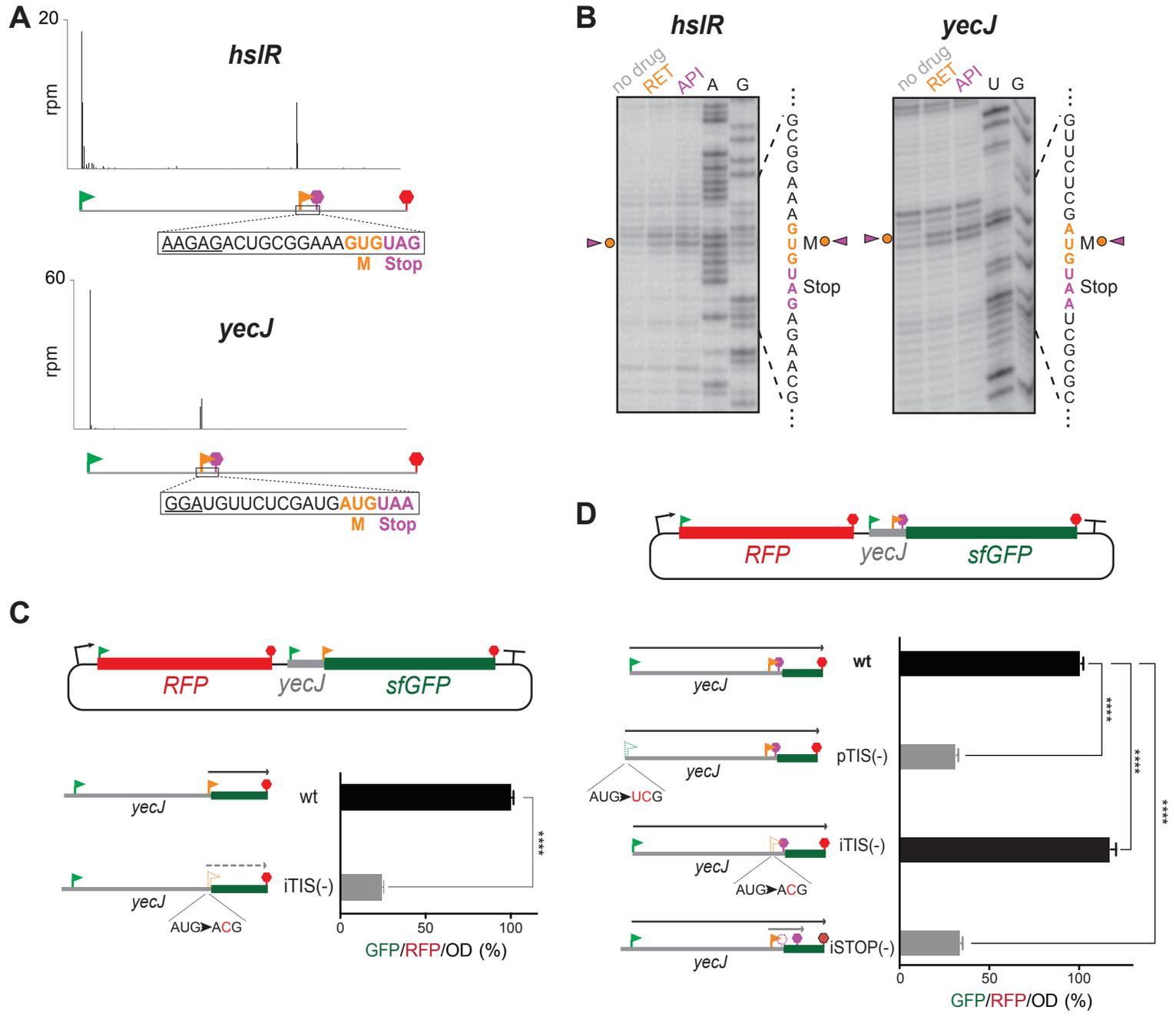


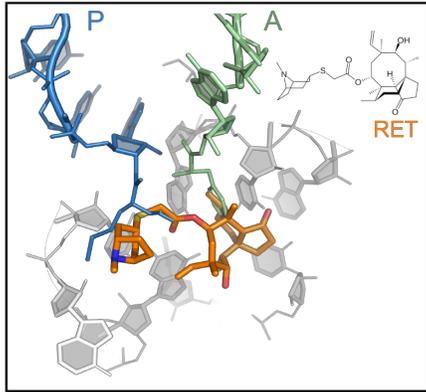
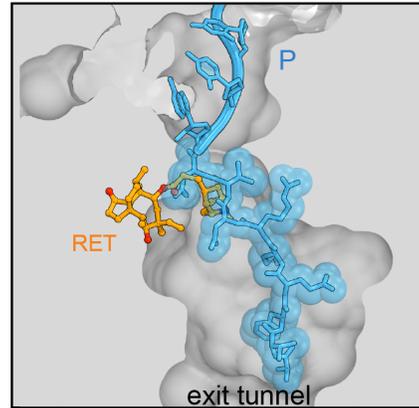
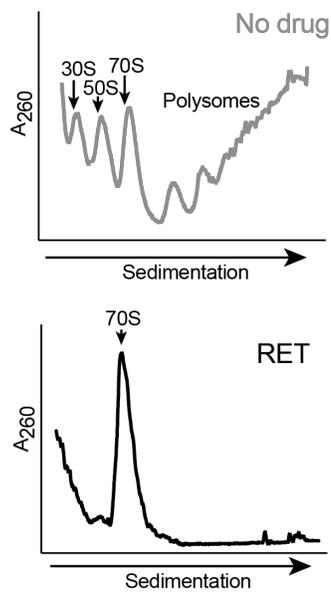
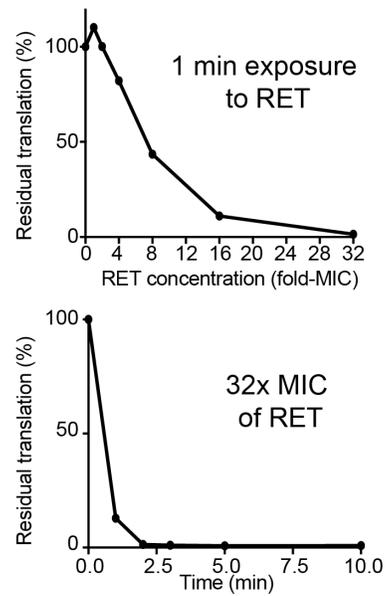
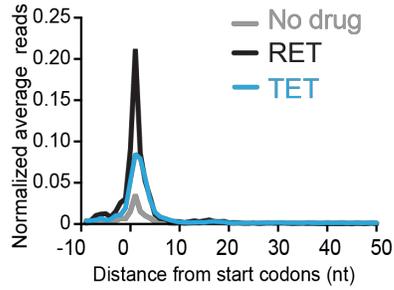
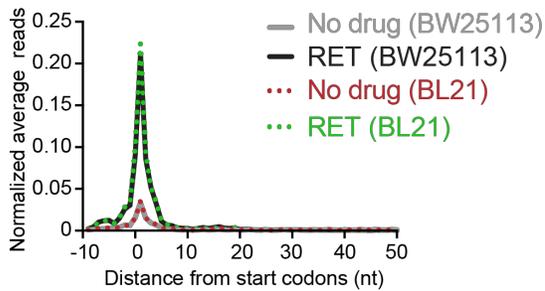
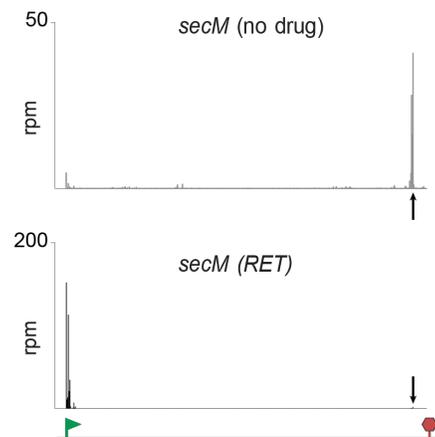
Figure 5



**Retapamulin-assisted ribosome profiling reveals the alternative bacterial  
proteome**

Sezen Meydan, James Marks, Dorota Klepacki, Virag Sharma, Pavel Baranov, Andrew Firth, Tõnu Margus, Amira Kefi, Nora Vázquez-Laslop and Alexander S. Mankin

**SUPPLEMENTARY INFORMATION**

**A****B****C****D****E****F****G**

**Figure S1 Retapamulin arrests ribosomes at initiation**, Related to Figure 1

(A) The chemical structure of the pleuromutilin antibiotic retapamulin (RET) bound at the PTC active site of the bacterial ribosome. The model is based on the structural alignment of the 50S ribosomal subunit of *Deinococcus radiodurans* (*Dr*) ribosomes in complex with RET (PDB 2OGO) (Davidovich et al., 2007) and *Thermus thermophilus* 70S ribosomes with fMet-tRNA bound in the P site and Phe-tRNA in the A site (PDB 1VY4) (Polikanov et al., 2014). Note that in the 70S initiation complex, the fMet moiety of the initiator tRNA has to be displaced from the PTC active site to allow for RET binding.

(B) RET cannot coexist with a nascent protein in the ribosome. Alignment of the structures of the *Dr* 50S RET complex with the *E. coli* 70S ribosome carrying ErmBL nascent peptide that esterifies P-site tRNA (PDB 5JTE) (Arenz et al., 2016).

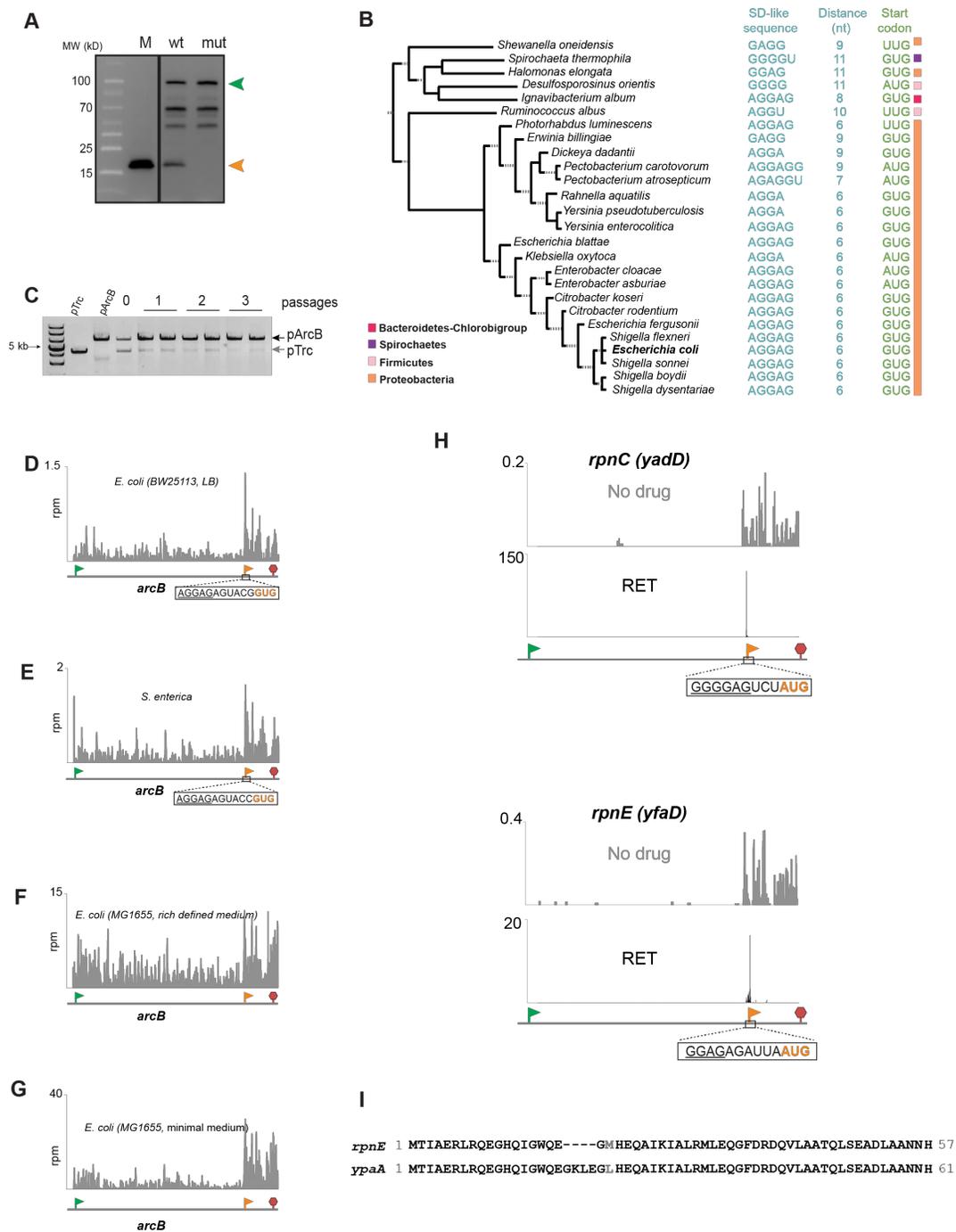
(C) Sucrose gradient analysis of polysome preparation from *E. coli* BW25113  $\Delta toIC$  cells untreated (top) or treated for 5 min with 12.5  $\mu\text{g}/\text{mL}$  (100X MIC) RET. The shown profiles represent cryo-lyzed preparations used in Ribo-seq experiments. Qualitatively similar results have been obtained in analytical experiments with the samples prepared by freezing-thawing (see STAR Methods).

(D) Residual protein synthesis in *E. coli* BL21  $\Delta toIC$  cells treated with RET, as estimated by incorporation of [ $^{35}\text{S}$ ]-methionine into the TCA-insoluble protein fraction, after 1 min exposure to increasing concentrations of RET (top) or treated with 2  $\mu\text{g}/\text{mL}$  of RET (32-fold MIC) for the indicated periods of time (bottom).

(E) Metagene plots comparing the normalized average relative density of ribosomal footprints in *E. coli* BW25113  $\Delta toIC$  cells untreated (gray trace) or treated 12.5  $\mu\text{g}/\text{mL}$  (100X MIC) of RET (black trace). Blue trace represents similar analysis of the publicly-available Ribo-seq data obtained with *E. coli* BW25113  $\Delta smpB$  cells exposed to tetracycline (TET) [the average of two replicates of Ribo-seq experiments reported in (Nakahigashi et al., 2016)].

(F) Metagene plots comparing the normalized average relative density of ribosomal footprints in the *E. coli* strains BW25113  $\Delta toIC$  cells or *E. coli* BL21  $\Delta toIC$  untreated or treated with RET.

(G) Snapshot of ribosomal footprints density in the *secM* gene of *E. coli* BW25113  $\Delta toIC$  cells untreated or treated with RET. The pTIS and stop codon of the gene are indicated by a green flag and red stop sign, respectively. The black arrow indicates the known site of translation arrest at the codon 165 of the 170-codon *secM* ORF (Nakatogawa and Ito, 2002).



**Figure S2 The utilization of an in-frame iTIS within the *arcB* gene leads to production of an alternative protein ArcB-C with a potential role in cell physiology, Related to Figure 3**

(A) The uncropped image of the immunoblot shown in Figure 3E, representing the bands corresponding to full-length ArcB-3X FLAG and internal initiation product ArcB-C-3XFLAG (marked with arrow heads). Protein size markers are shown. The origin of the bands marked with dots is unknown.

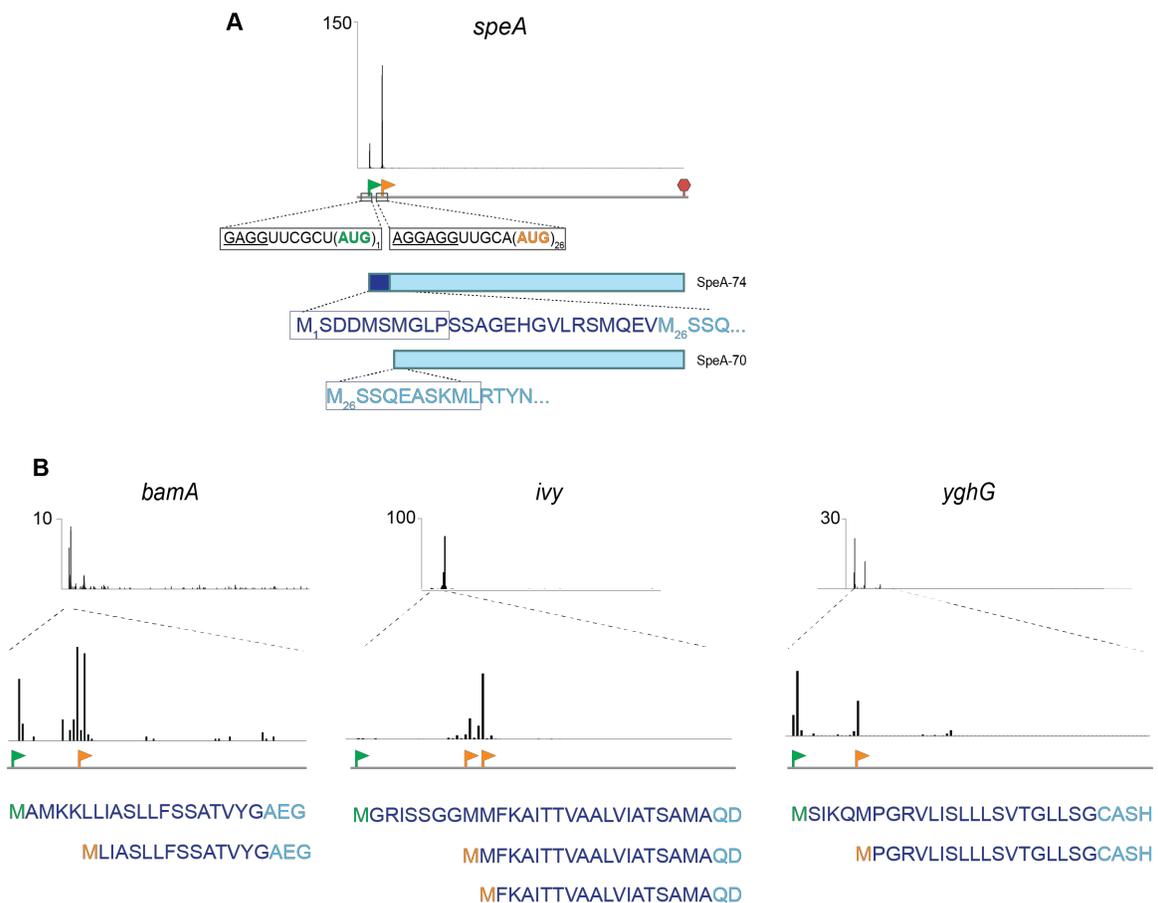
(B) The iTIS that directs translation of the ArcB-C protein is conserved in the *arcB* gene of diverse bacterial species. The putative start codons and the SD-like sequences are shown.

(C) Presence of *arcB* facilitates *E. coli* growth under low oxygen conditions. BW25113  $\Delta$ *arcB* *E. coli* cells carrying the empty vector pTrc99a or pArcB were co-grown in low oxygen conditions. Gel shows the HindIII-linearized plasmids, isolated from the co-growth cultures to determine fraction of cells with or without *arcB* in the mixture (see Start Methods for details). The '0' sample represents plasmids from the initial mixture containing equal number of pTrc99A and pArcB cells.

(D-G) The upshift of ribosomal footprints in the *arcB* segment encoding ArcB-C observed in the Ribo-seq profiles of untreated *E. coli* or *Salmonella enterica* cells (Baek et al., 2017; Kannan et al., 2014; Li et al., 2014). The pTIS and iTIS of *arcB* are marked with green and orange flags, respectively, and the stop codon is indicated by a red stop sign.

(G) Representative examples of Ribo-RET and Ribo-seq profiles of two out of five *E. coli* *rpn* genes.

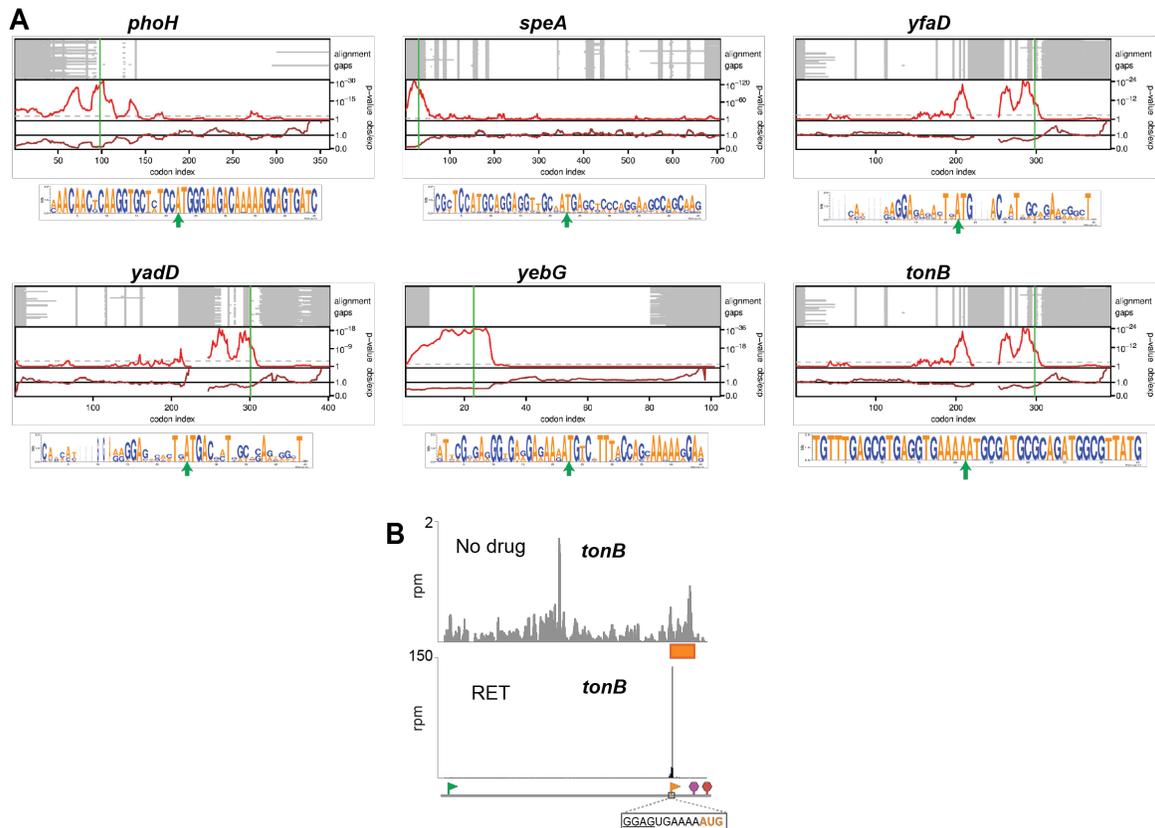
(H) Alignment of the amino acid sequence of the RpnE-C protein, translated from the iTIS within the *rpnE* gene and the protein encoded in an independent gene *ypaA*.



**Figure S3. Initiation at the 5'-end proximal iTISs could produce alternative products with incomplete N-terminal signal sequences, Related to Figure 4**

(A) Ribo-RET profile of the *speA* gene, showing peaks corresponding to pTIS (green flag) and iTIS (orange flag). The stop codon is indicated by a red stop sign. The putative signal sequence (indicated by dark blue letters) of SpeA-74 (Buch and Boyle, 1985) is lacking in the alternative product SpeA-70 whose translation is initiated at the iTIS. The SpeA isoforms, whose translation is initiated at the pTIS or the iTIS are expected to have different cellular localization. The peptides detected by N-terminomics are boxed (Bienvenut et al., 2015).

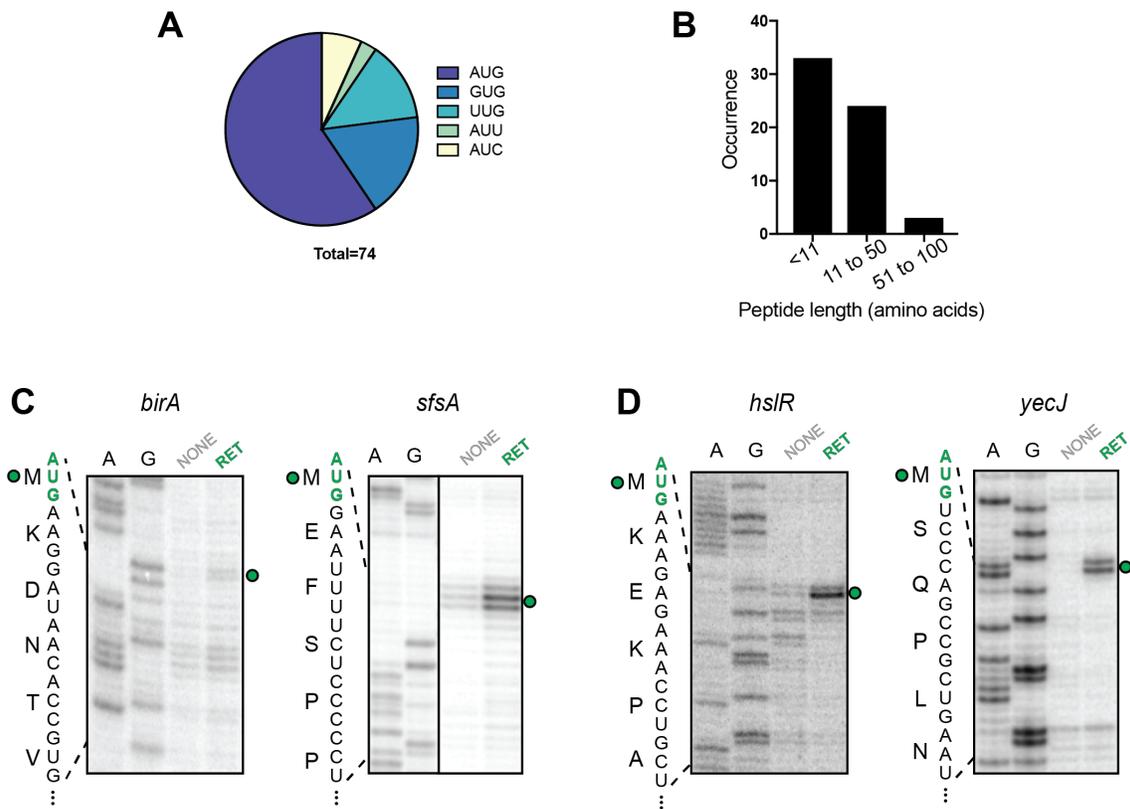
(B) Ribo-RET profiles of *bamA*, *ivy* and *yghG* genes. The N-terminal amino acid sequences of the primary and predicted alternative proteins are indicated. The reported signal sequences are shown in dark blue. The pTISs of the genes are marked by green flags; iTISs are indicated with orange flags.



**Figure S4. Synonymous site conservation for selected iTISs, Related to Figures 3-5**

(A) Synonymous site conservation plots and weblogs for genes with in-frame iTISs (*phoH*, *speA*, *yfaD*, *yadD*, *yebG*) and for the *tonB* gene with an OOF iTIS. Alignment gaps in each sequence are indicated in grey. The two panels show the synonymous substitution rate in a 15-codon sliding window, relative to the CDS average (observed/expected; brown line) and the corresponding statistical significance ( $p$ -value; red line). The horizontal dashed grey line indicates a  $p$ -value of  $0.05 / (\text{CDS length}/\text{window size})$  – an approximate correction for multiple testing within a single CDS.

(B) An upshift in the local density of ribosome footprints within the alternative frame defined by the *tonB* OOF iTIS (orange rectangle) in cells not exposed to antibiotic. Start codons of the pTIS and OOF iTIS are marked with green and orange flags, respectively, while the respective stop codons are indicated with red and purple stop signs. The start codon and SD-like sequence of the iTIS are shown.

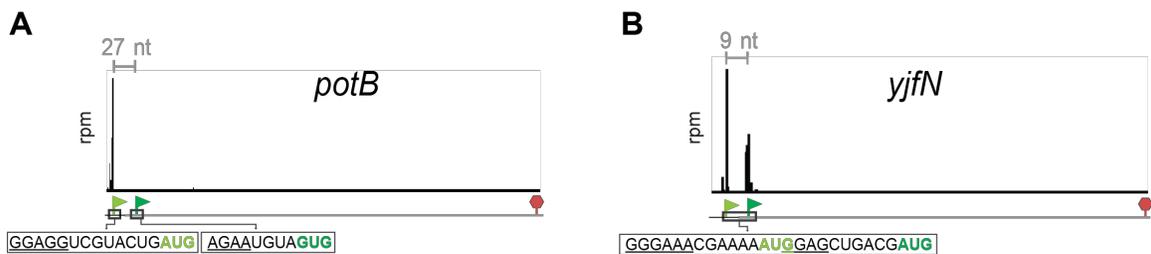


**Figure S5 Ribo-RET reveals OOF iTISs, Related to Figures 4 and 5**

(A) The distribution of start codons associated with OOF iTISs revealed by Ribo-RET.

(B) The length distribution of the putative alternative proteins whose translation is initiated at OOF iTISs.

(C) and (D) Toe-printing gels showing RET-induced ribosome stalling at the pTISs of *birA* and *sfsA* (shown in Figure 4) and *hsIR* and *yecJ* (shown in Figure 5) genes. Samples analyzed in the lanes marked NONE contained no antibiotics. Start codons of the pTISs are indicated in green. Sequencing lanes are shown.

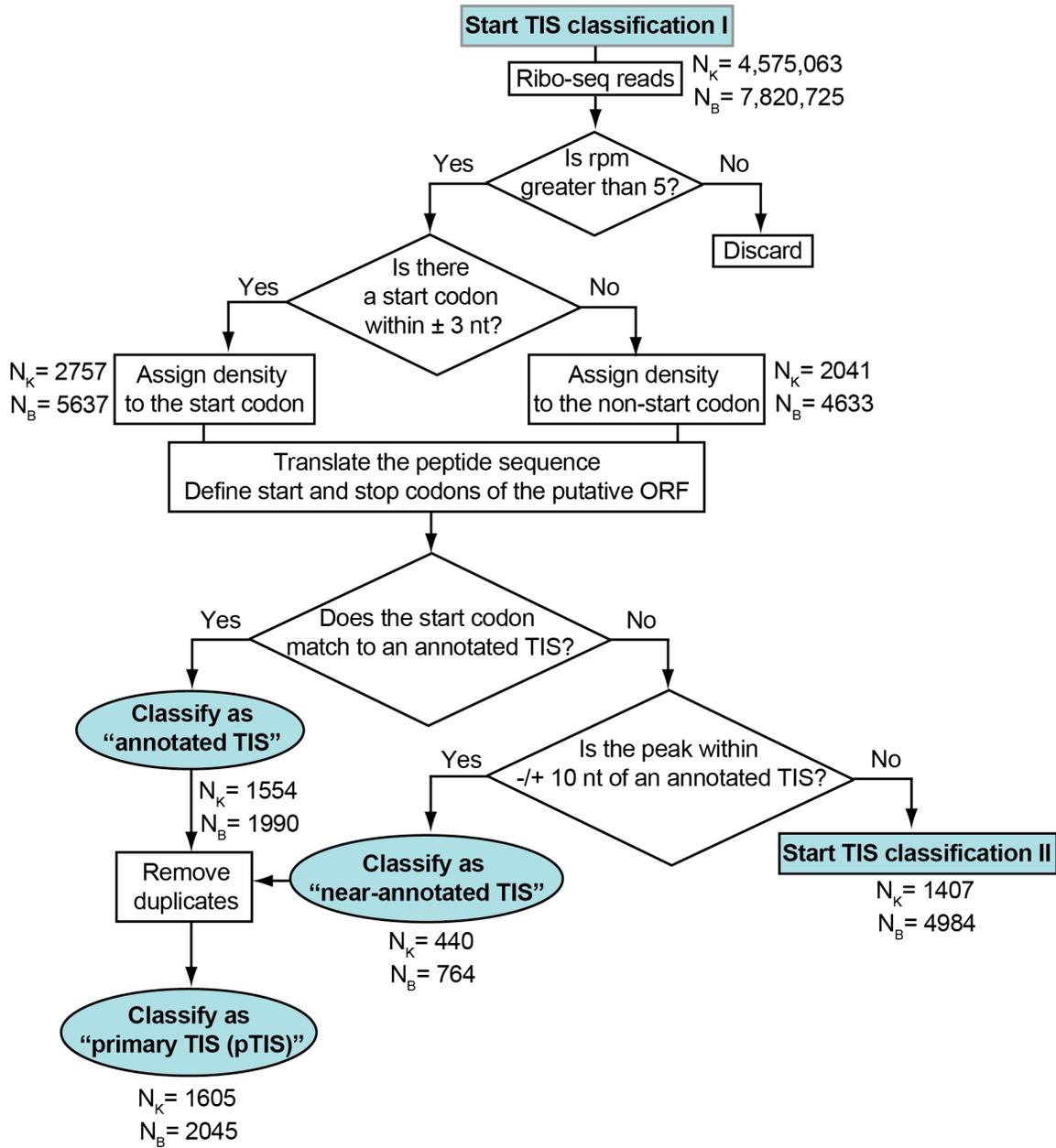


**Figure S6. Examples of the genes with Ribo-RET identified TISs outside of the coding regions, Related to Figure 1**

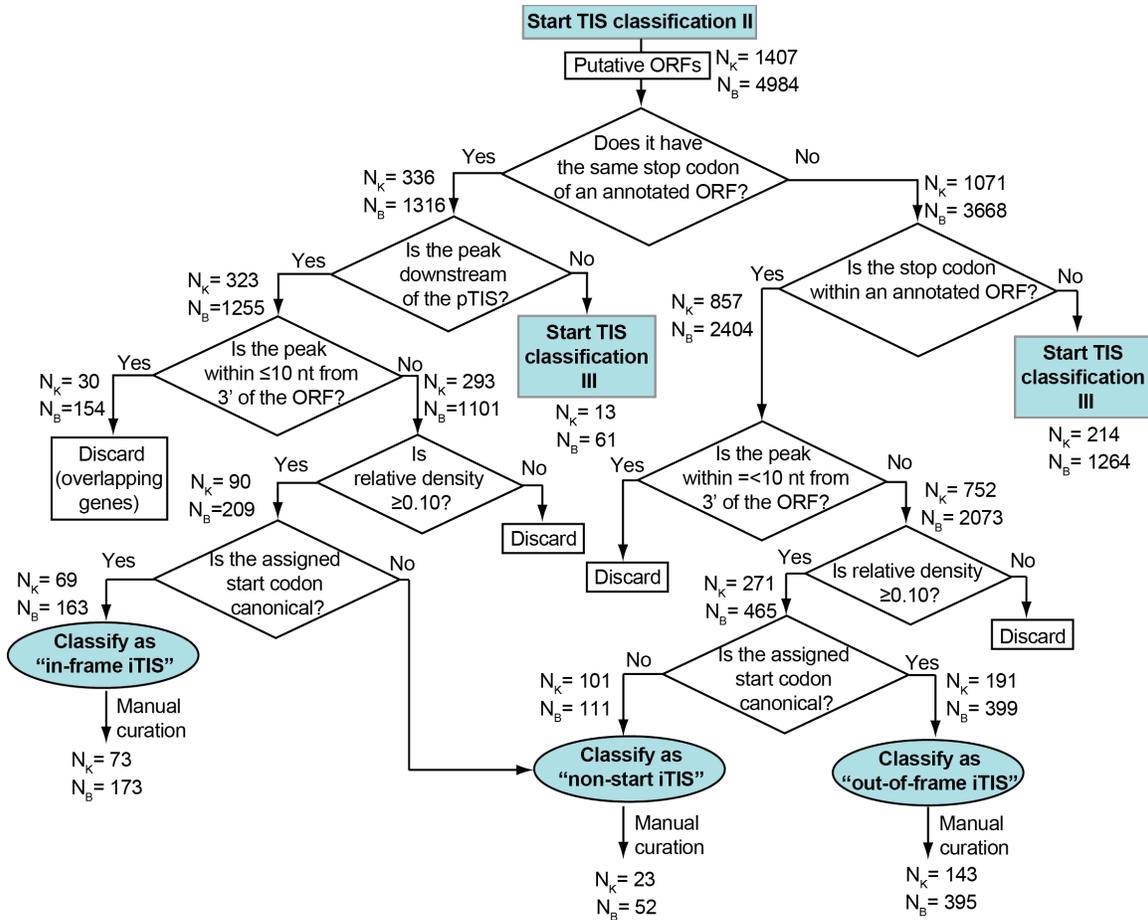
(A) Ribo-RET profile of the *potB* gene, shows no peak of the ribosome density at the start codon of the annotated pTIS (dark green flag), but instead reveals a strong peak at an in-frame start codon 27 nt upstream (pale green flag).

(B) In the *yjfN* gene, Ribo-RET reveals peak at the annotated pTIS (dark green flag) and an additional peak 9 nts upstream from it (marked with a pale green flag). The sequences surrounding the two TISs, including the SD-like regions (underlines) are shown.

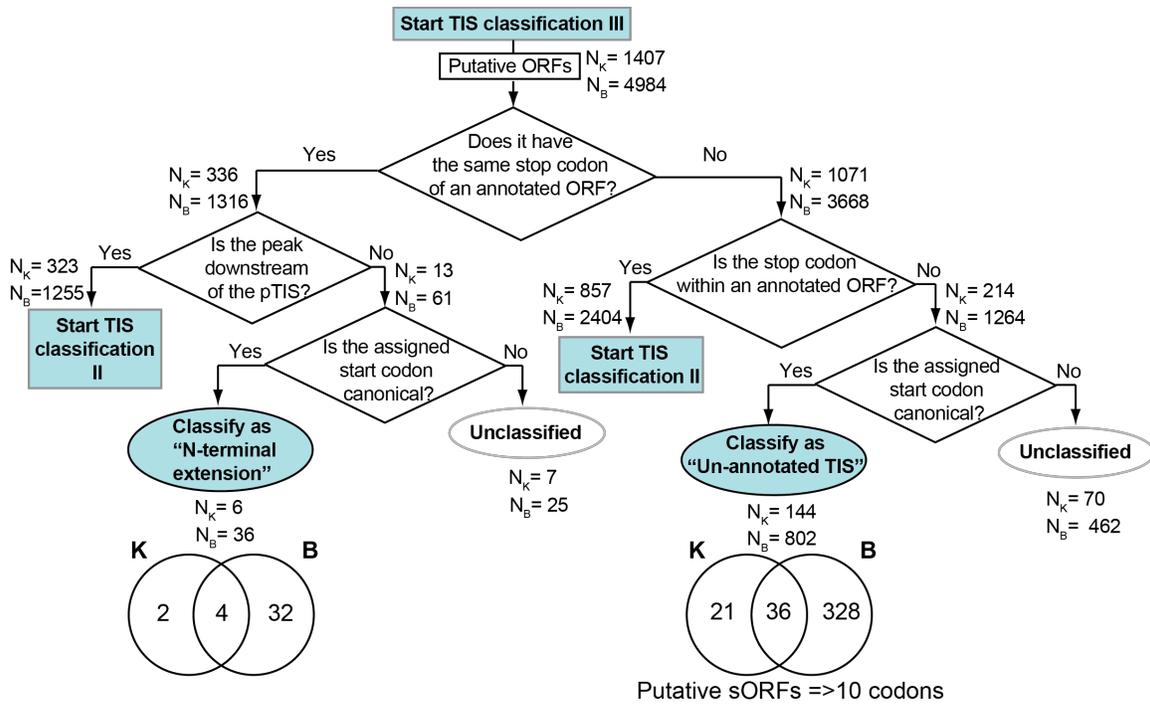
See Table S2 for other cases of Ribo-RET signals outside of the coding regions.



**Supplemental scheme 1:** Flow chart showing the criteria used to identify pTISs (Classification I), Related to STAR methods



**Supplemental scheme 2:** Flow chart showing the criteria used to identify iTISs (Classification II), Related to STAR methods



**Supplemental scheme 3:** Flow chart showing the criteria used to identify N-terminal extensions and un-annotated TISs (Classification III), Related to STAR methods

**Table S3:** List of primers and synthetic DNA fragments used in this study, Related to STAR methods

	Sequence (5' to 3')	Purpose
P1	TGTCCTGGCACTAATAGTGA	Forward primer for amplification of <i>tol::kan</i> cassette
P2	ACGATGCGTGGCGTATGG	Reverse primer for amplification of <i>tol::kan</i> cassette
P3	TTGTGAGCGGATAACAATTTACACAGG AAACAGACCATGGTGGGTATTATTGGGG CAGG	Forward primer for amplification of <i>arcB</i> -PCR 1-wt
P4	ACATAATACTGCGCCAGC	Reverse primer for amplification of <i>arcB</i> -PCR 1-wt
P5	AAGCAAATTCGTCTGCTGG	Forward primer for amplification of <i>arcB</i> -PCR 2-wt
P6	TGGGAATATCGAGCAATGCTT	Reverse primer for amplification of <i>arcB</i> -PCR 2-wt
P7	GAAGAGAACAGTAAATCAGAAGCATTG	Forward primer for amplification of <i>arcB</i> -PCR 3
P8	TCAGGCTGAAAATCTTCTCTCATCCGCC AAAACAGCCAAGCTTTCACCTTGTCATCGT CAT	Reverse primer for amplification of <i>arcB</i> -PCR 3
P9	TCCTGGGTATCCCAGAATTTTC	Reverse primer for amplification of <i>arcB</i> -PCR 1-mutant
P10	CGCTAACCGCGATGATCAAGAAATTCTG GGATACCCAGGATGATGAAGAAAGTACG GTCACGACAGAAGAG	Forward primer for amplification of <i>arcB</i> -PCR 2-mutant
P11	TGGGAATATCGAGCAATGCTTCTGATTTA CTGTTCTCTTCTGTCGTGACCGTACTTTC TTCATCATCCTGGGTATCCCA	Reverse primer for amplification of <i>arcB</i> -PCR 2-mutant
#12	AACAGACCATGGTACCCAGGATGATGAG GAGAGTACGGTGACGACAGAAGAGAAC AGTAAATCAGAAGCATTGCTCGATATTCC CATGCTGGAACAGTATCTCGAACTTGTA GGACCGAAGCTGATCACCGACGGGTTA GCGGTGTTTGAGAAGATGATGCCGGGCT ATGTCAGCGTGCTGGAGTCGAATCTGAC GGCGCAGGATAAAAAAGGCATTGTTGAG GAAGGACATAAAATTAAGGTGCGGCGG GGTCAGTGGGGTTACGCCATCTGCAACA GCTGGGTCAGCAAATTCAGTCTCCTGAC CTTCCGGCCTGGGAAGATAACGTCCGGTG AATGGATTGAAGAGATGAAAGAAGAGTG GCGTCACGACGTAGAAGTGCTGAAAGC GTGGGTGGCAAAGCCACTAAAAAAGAC TACAAAGACCATGACGGTGATTATAAAG ATCATGACATCGATTACAAGGATGACGA TGACAAGTGAAAGCTTGGCTGTT	gBlock for <i>arcB</i> -marker insert

P13	TAATACGACTCACTATAGGGCTGTAATTA ACAACAAAGGGT	Forward primer for amplification of <i>atpB</i>
P14	GGTTATAATGAATTTTGCTTATTAACCGA GAATGTACGCAGTTAGTCCAGCTGAAGG TT	Reverse primer for amplification of <i>atpB</i>
P15	TAATACGACTCACTATAGGGACTAAAAGT AAGGCATTAAC	Forward primer for amplification of <i>mgo</i>
P16	GGTTATAATGAATTTTGCTTATTAACCTG CTCCTCGGACGCTTATTTGCTTTTGCC GCC	Reverse primer for amplification of <i>mgo</i>
P17	GGTTATAATGAATTTTGCTTATTAAC	Reverse primer for toeprinting of <i>atpB</i> and <i>mgo</i>
P18	TAATACGACTCACTATAGGGAGCGCAGT GGAGACA	Forward primer for amplification of <i>birA</i>
P19	CTACGCAAATAATTTGCAGGG	Reverse primer for amplification of <i>birA</i>
P20	TTTCACCCAAGTCTC	Reverse primer for toeprinting of primary site of <i>birA</i>
P21	AATACTCCCCTTTCTTATTTTT	Reverse primer for toeprinting of internal site of <i>birA</i>
P22	TAATACGACTCACTATAGGGCAATAACAA GGATTGTCGCAATG	Forward primer for amplification of <i>sfsA</i> -PCR 1
P23	GCCGTATTTACTTCGCTTTCTAGCGAGC TATAGCCTGACAG	Reverse primer for amplification of <i>sfsA</i> -PCR 1
P24	CTGTCAGGCTATAGCTCGCTAGAAAGCG AAGTAAAATACGGC	Forward primer for amplification of <i>sfsA</i> -PCR 2
P25	CTACAATGTAACCGGCAGTG	Reverse primer for amplification of <i>sfsA</i> -PCR 2
P26	TAATACGACTCACTATAGGG	Forward primer for amplification of <i>sfsA</i> -g321a, a322g
P27	CTACAATGTAACCGGCAGTG	Reverse primer for amplification of <i>sfsA</i> -g321a, a322g
P28	CATCGGGTGTGATCAC	Reverse primer for toeprinting of primary site of <i>sfsA</i>
P29	CGATTTCACTTCAATATA	Reverse primer for toeprinting of internal site of <i>sfsA</i>
P30	TAATACGACTCACTATAGGGGCTAATGT GAAGGAGACGC	Forward primer for amplification of <i>hslR</i>
P31	TTATTTCACTGTCGCCGTG	Reverse primer for amplification of <i>hslR</i>
P32	GGGCCAGCGCGC	Reverse primer for toeprinting of primary site of <i>hslR</i>
P33	TTGTCCGGGCGTCCG	Reverse primer for toeprinting of internal site of <i>hslR</i>
P34	TAATACGACTCACTATAGGGAATGCTATC AGGAGTTTACGATG	Forward primer for amplification of <i>yecJ</i>
P35	TTAATGGGATTCACCCTGTGGG	Reverse primer for amplification of <i>yecJ</i>
P36	CATCCAGAATTTGTTTGATAAC	Reverse primer for toeprinting of primary site of <i>yecJ</i>

P37	GCGGCGGCGGGATGG	Reverse primer for toeprinting of internal site of <i>yecJ</i>
P38	GTGAGCGGATAACAATTTACACAGAAT TCATTAAAGAGGAGAAATTA ACTATGGCT	Forward primer for amplification of <i>RFP</i>
P39	ATATCTCCTTCTTAAAGTTAAACA ACTAG TCTATTCGCCAGAACCAGC	Reverse primer for amplification of <i>RFP</i>
P40	TTTAAGAAGGAGATATACATATGACTAGT GCATCCAAGGGCGA	Forward primer for amplification of <i>GFP</i>
P41	TCAGCTAATTAAGCTTGGCTGCAGGTCG ACCCGGGGTACCGAG	Reverse primer for amplification of <i>GFP</i>
P42	TCCGCTGCTGGTTCTGGCGAATAGACTA GTCAATAACAAGGATTGTCGCAATG	Forward primer for amplification of insert for pRXGSM-sfsA-wt
P43	AAAGAGCTCCTCGCCCTTGGATGCACTA GTGCGAGCTATAGCCTGAC	Reverse primer for amplification of insert for pRXGSM-sfsA-wt
P44	AAAGAGCTCCTCGCCCTTGGATGCACTA GTGCGAGCTATAGCCTGACAGTTCTGAA ATTGATTCGATAAGGATAGCCT	Reverse primer for amplification of insert for pRXGSM-sfsA-mutant
P45	GTTCTGGCGAATAGACTAGTAAATGCTAT CAGGAGTTTACG	Forward primer for amplification of insert for pRXGSM- <i>yecJ</i> derivatives
P46	AGCTCCTCGCCCTTGGATGCACTAGTCA TCGAGAACATCCAGAATTTG	Reverse primer for amplification of insert for pRXGSM- <i>yecJ</i> -iTIS-wt
P47	AGCTCCTCGCCCTTGGATGCACTAGTCG TCGAGAACATCCAGAATTTG	Reverse primer for amplification of insert for pRXGSM- <i>yecJ</i> -iTIS(-)
P48	AGAGCTCCTCGCCCTTGGATGCACTAGT TACATCGAGAACATCCAGAATTTG	Reverse primer for amplification of insert for pRXGSM- <i>yecJ</i> -pTIS-wt
P49	CTGCTGGTTCTGGCGAATAGACTAGTAA TGCTATCAAAAGTTTACGTCGTCGCCAGC CGCT	Primer for site directed mutagenesis to generate pRXGSM- <i>yecJ</i> -pTIS(-)
P50	AGCTCCTCGCCCTTGGATGCACTAGTTA CGTCGAGAACATCCAGAATTTG	Reverse primer for amplification of insert for pRXGSM- <i>yecJ</i> -pTIS-iTIS(-)
P51	AGCTCCTCGCCCTTGGATGCACTAGTGA CATCGAGAACATCCAGAATTTG	Primer for site directed mutagenesis to generate pRXGSM- <i>yecJ</i> -pTIS-iStop(-)
P52	GGCCTTAACCGCTAACGT	Direct primer for sequencing the iTIS region in the <i>arcB</i> gene
P53	TTTAATCTGTATCAGGCTGAAAATCTT	Reverse primer for sequencing the iTIS region in the <i>arcB</i> gene

## Supplementary Information References

- 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.
- Baek, J., Lee, J., Yoon, K., and Lee, H. (2017). Identification of unannotated small genes in *Salmonella*. *G3 (Bethesda)* **7**, 983-989.
- Bienvenut, W.V., Giglione, C., and Meinel, 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.
- 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.
- 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. U. S. A.* **104**, 4291-4296.
- 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. U. S. A.* **111**, 15958-15963.
- 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.
- 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.
- 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.