

INRI-seq enables global cell-free analysis of translation initiation and off-target effects of antisense inhibitors

Jens Hör¹, Jakob Jung¹, Svetlana Đurica-Mitić¹, Lars Barquist^{2,3} and Jörg Vogel^{1,2,3,*}

¹Institute for Molecular Infection Biology, University of Würzburg, D-97080 Würzburg, Germany, ²Helmholtz Institute for RNA-based Infection Research (HIRI), Helmholtz Center for Infection Research (HZI), D-97080 Würzburg, Germany and ³Faculty of Medicine, University of Würzburg, D-97080 Würzburg, Germany

Received April 28, 2022; Revised September 11, 2022; Editorial Decision September 12, 2022; Accepted September 19, 2022

ABSTRACT

Ribosome profiling (Ribo-seq) is a powerful method for the transcriptome-wide assessment of protein synthesis rates and the study of translational control mechanisms. Yet, Ribo-seq also has limitations. These include difficulties with the analysis of translation-modulating molecules such as antibiotics, which are often toxic or challenging to deliver into living cells. Here, we have developed *in vitro* Ribo-seq (INRI-seq), a cell-free method to analyze the translational landscape of a fully customizable synthetic transcriptome. Using *Escherichia coli* as an example, we show how INRI-seq can be used to analyze the translation initiation sites of a transcriptome of interest. We also study the global impact of direct translation inhibition by antisense peptide nucleic acid (PNA) to analyze PNA off-target effects. Overall, INRI-seq presents a scalable, sensitive method to study translation initiation in a transcriptome-wide manner without the potentially confounding effects of extracting ribosomes from living cells.

INTRODUCTION

Protein synthesis is one of the most energy-consuming processes in living cells, making its precise regulation a crucial matter of cellular economy (1,2). While mRNA levels determined by RNA-sequencing (RNA-seq) are often used as a proxy for protein synthesis in global gene expression analysis (3), final protein abundance does not always correlate with mRNA levels (4–7). This is due to a multitude of regulatory mechanisms, for example, direct control of the translational machinery (1) and post-transcriptional control of mRNAs by base-pairing small RNAs (sRNAs) (8) or intrinsic mRNA structure (9).

Over the past decade, ribosome profiling (Ribo-seq) has become a primary method to more directly measure protein synthesis in a transcriptome-wide manner (10–12). Ribo-seq is based on RNA-seq analysis of ribosome-protected fragments (RPFs), which are ribosome-bound mRNA fragments that survive nuclease treatment after cell lysis because they are covered and protected by translating ribosomes (13–15). In *Escherichia coli*, RPFs typically are 15–45 nt in length (16). Since each RPF represents the position of one ribosome, their sum not only reveals which mRNAs but also which parts of the coding sequence (CDS) of these mRNAs were being translated at the point of sampling. Thus, Ribo-seq can map the landscape of actively translating ribosomes *in vivo* to provide global information about translational pausing (16), stalling (17) and start site usage (18–20), as well as an estimate of protein copy numbers (6).

While Ribo-seq has greatly advanced the study of translation-related processes, the method has not been without limitations. Coverage of weakly expressed genes remains challenging, preventing many genes from being detected in common study designs. This includes certain gene classes with notoriously low expression under standard growth conditions, such as toxins whose expression is only triggered by specific stresses that lead to cell death or inhibition of growth (21). Similarly, Ribo-seq of microbes from important ecological habitats such as the human gut (22) remains difficult since many of them cannot be cultured in the laboratory, though recent efforts have started to tackle this challenge (7).

On the mechanistic level, Ribo-seq-based studies of molecules affecting translation can be hampered by cellular responses. For instance, the antibiotic retapamulin (RET) acts right after translation initiation when the ribosome is still at the start codon, which has enabled the global annotation of translation initiation sites (TISs) in *Escherichia coli* (18,23). However, this study required genetic inactivation of the ABC transporter TolC to prevent export of the antibiotic prior to it exerting its effect. While this was possible in

*To whom correspondence should be addressed. Tel: +49 931 3182576; Email: joerg.vogel@uni-wuerzburg.de
Present address: Jens Hör, Department of Molecular Genetics, Weizmann Institute of Science, Rehovot 7610001, Israel.

E. coli, a lack of either genetic tools or knowledge of transporters would preclude similar studies in many other organisms. Lastly, since Ribo-seq is performed on living cells, it can be difficult to dissect direct and indirect effects on translation. This is exemplified by antisense antibiotics (24–26), whose import via carrier peptides broadly affects gene expression in addition to the desired antisense inhibition of the targeted gene of interest (27,28).

To overcome some of these limitations, we have developed *in vitro* Ribo-seq (INRI-seq) for the global study of translation in a cell-free manner. INRI-seq uses the commercially available PURExpress *in vitro* translation system combined with an *in vitro*-synthesized, fully customizable transcriptome for better control of individual mRNA levels. INRI-seq obviates the need for translation-modulating compounds to traverse cellular membranes and the extraction of ribosomes from a large number of living cells. As proof of concept, we apply INRI-seq to a synthetic *E. coli* transcriptome and show that the method faithfully validates known TISs and can be used to predict new TISs. In addition, we use the system to study the fidelity of translational inhibition by antisense peptide nucleic acid (PNA), demonstrating on-target specificity and defining base-pairing criteria that influence the off-target effects of these short antisense oligomers. INRI-seq bears great potential as a scalable alternate method to study translation control mechanisms and translation-modulating compounds in other organisms and organismal communities, including eukaryotes.

MATERIALS AND METHODS

Design and synthesis of the synthetic transcriptome

To obtain a synthetic variant of the *E. coli* MG1655 transcriptome containing the first 51 codons of each gene, a pool of single-stranded DNA oligonucleotides was designed by extracting the sequence of the first 153 nt of each annotated open reading frame (start codon + 50 codons). At the 5' end, 30 nt upstream of each start codon was added as 5' UTR and extended by addition of a T7 RNA polymerase promoter (GTTTTTTTTAATACGACTCACTATAGGG). At the 3' end, a TAA stop codon was added and extended by the first 28 nt of the 3' UTR of *E. coli hns* (TCTTTTGTAGATTGCACTTGCTTAAAAT). The final pool of 4386 DNA oligonucleotides (JVOpool-001) was ordered from Integrated DNA Technologies at a scale of 1 pmol/oligo and is listed in Supplementary Table S6.

Using the KAPA HiFi HotStart PCR kit (Roche), double-stranded DNA was generated from JVOpool-001. Per 50 μ l of PCR reaction, 20 ng of JVOpool-001, 10 μ l HF buffer, 1 μ l dNTPs, 1 μ l DNA polymerase and 2 μ l each of the oligos JVO-18582 and JVO-18583 (10 μ M each) were used. The PCR was run with the following protocol: 95°C for 3 min; 10 cycles of 95°C for 20 s followed by 60°C for 20 s and 72°C for 15 s; 72°C for 2 min. A total of 400 μ l of this PCR was run and purified using column-based cleanup (Macherey-Nagel).

The RNA of the synthetic transcriptome was obtained by *in vitro* transcription of 500 ng each of the double-stranded DNA pool in two 40 μ l reactions using the MEGascript

T7 transcription kit (Thermo Fisher) according to the manufacturer's instructions, except that the reactions were performed overnight. The next day, the reactions were treated with 2 μ l each of TURBO DNase (Thermo Fisher) at 37°C for 15 min. The RNA was denatured by addition of 40 μ l of 2 \times GLII loading buffer (95% formamide, 18 mM EDTA, pH 8, 0.02% (w/v) SDS, 0.025% (w/v) bromophenol blue, 0.025% (w/v) xylene cyanol) and incubation at 95°C for 5 min, then placed on ice. To purify the RNA, it was separated by 6% denaturing PAGE with 7 M urea in 1 \times TBE, stained with EtBr, cut from the gel and eluted with 750 μ l RNA elution buffer (0.1 M NaOAc, pH 5.4, 0.1% (w/v) SDS, 10 mM EDTA, pH 8) at 4°C overnight. The next day, the RNA-containing supernatant was mixed with 800 μ l acidic phenol-chloroform-isoamylalcohol and centrifuged at 4°C for 15 min. The aqueous phase was collected, split into two tubes and precipitated with 1 ml each of ice-cold EtOH. Following precipitation at –20°C for at least 1 h, the samples were centrifuged at 4°C for at least 30 min, the supernatants discarded, the pellets washed with 400 μ l ice-cold 70% EtOH and centrifuged again at 4°C for 15 min. Finally, the supernatants were discarded and the pellets were dried at room temperature for 5 min and resuspended in 25 μ l water. The resuspended pellets were pooled into one tube to obtain the final synthetic transcriptome. RNA purity and integrity were tested by denaturing urea PAGE.

In vitro Ribo-seq (INRI-seq)

To globally investigate translation *in vitro*, the PURExpress *In Vitro* Protein Synthesis Kit (New England Biolabs) was used to translate the synthetic transcriptome described above. To denature the synthetic transcriptome, it was incubated at 95°C for 2 min, then placed on ice. For each sample, a 25 μ l reaction containing 10 μ l solution A (PURExpress), 7.5 μ l solution B (PURExpress), 5 μ l water and 2.5 μ l of the denatured synthetic transcriptome (stock concentration 10 μ M; final concentration 1 μ M) was incubated at 37°C for 15 min in order to allow the ribosomes to start unhindered translation. Then, 1.25 μ l of retapamulin (stock concentration 100 μ g/ml in DMSO; final concentration 5 μ g/ml) was added to block translation and the incubation continued at 37°C for 30 min. To stop the reaction, 175 μ l of ice-cold stop buffer (20 mM Tris–HCl, pH 8, 100 mM NH₄Cl, 10 mM MgCl₂, 5 mM CaCl₂, 1 mM DTT, 0.4% Triton X-100, 0.1% NP-40, 200 U/ml RNase-inhibitor, 5 μ g/ml retapamulin) was added and the reaction put on ice.

When the effect of PNAs on translation was investigated, the same protocol was followed with the following exceptions: The PNA stock solutions (10 \times the concentrations of the respective final concentrations) were incubated at 55°C for 5 min. Then, 2.5 μ l of the PNA stock solution and 2.5 μ l water were added to 2.5 μ l of the denatured synthetic transcriptome and the mixture annealed at 37°C for 5 min, then placed on ice. The resulting 7.5 μ l were then mixed with 10 μ l solution A and 7.5 μ l solution B and the translation carried out as above.

RPFs were generated by addition of 1.4 μ l of MNase (420 U in total; Thermo Fisher) and incubation at 25°C for 1 h. The reaction was quenched with 2 μ l of 0.5 M EGTA, pH 8 and placed on ice. To isolate 70S monosomes contain-

ing the desired RPFs, 185 μ l of the sample was loaded on a 10–55% sucrose gradient (20 mM Tris–HCl, pH 7.5, 100 mM NH_4Cl , 10 mM MgCl_2 , 5 mM CaCl_2 , 1 mM DTT) formed in an open-top polyclear ultracentrifugation tube (Seton Scientific). The gradient was centrifuged at 4°C and 35 000 rpm for 2.5 h using an SW 40 Ti rotor (Beckman Coulter). Afterward, the 70S monosome fraction (\sim 1 ml) was collected using a Biocomp Model 153 gradient station and snap-frozen in liquid N_2 .

The frozen monosome fraction was thawed on ice and split into two tubes. To each tube, 800 μ l of acidic phenol-chloroform-isoamylalcohol was added, the samples vortexed for 15 s and then centrifuged at 4°C for 15 min. The aqueous phases were collected and precipitated by addition of 1 μ l GlycoBlue (Thermo Fisher) and 1.4 ml of ice-cold precipitation mix (30:1 EtOH:3 M NaOAc, pH 6.5) at -20°C for at least 1 h. The samples were centrifuged at 4°C for at least 30 min, the supernatants discarded, the pellets washed with 400 μ l ice-cold 70% EtOH and centrifuged again at 4°C for 15 min. Finally, the supernatants were discarded and the pellets were dried at room temperature for 5 min and one pellet resuspended in 25 μ l water. The solution was then transferred to the other tube to resuspend the corresponding second pellet.

30 μ l of 2 \times GLII loading buffer were added and the sample denatured at 95°C for 5 min, then placed on ice. As ladder, 10 μ l of microRNA Marker (New England Biolabs) were mixed with 2.5 μ l Low Range ssRNA Ladder (New England Biolabs), 25 μ l 2 \times GLII loading buffer and 22.5 μ l water and then denatured like the sample. The RNA samples containing the RPFs and the ladder were separated by 15% denaturing PAGE with 7 M urea in 1 \times TBE. The RPFs were visualized by staining the gel with SybrGold (Thermo Fisher) and cut from the gel in the range of 15–45 nt. Gel extraction was performed as described above, except that 1 μ l GlycoBlue was added during ethanol precipitation. Finally, the pellets were dried at room temperature for 5 min and one pellet resuspended in 25 μ l water. The solution was then transferred to the other tube to resuspend the corresponding second pellet.

cDNA library preparation

The purified RPFs were subjected to library preparation for next-generation sequencing (vertis Biotechnologie). The RNA 5' ends were phosphorylated with T4 polynucleotide kinase (New England Biolabs). Adapters were ligated to the 5' and 3' ends of the RNA and first-strand cDNA synthesis was performed using M-MLV reverse transcriptase. The cDNA was then amplified using a high fidelity DNA polymerase and purified with Agencourt AMPure XP beads (Beckman Coulter). The cDNA samples were pooled in equimolar amounts and the rDNA within the pool was depleted targeting 24 unique sequences (Supplementary Table S1) using a CRISPR–Cas9 protocol similar to DASH (29). The final pool was subjected to sequencing on an Illumina NextSeq 500 system using 75 nt single-end read length.

Quantification of INRI-seq data

Reads of the INRI-seq experiments were preprocessed and mapped against the oligo pool using tools from the BBtools

suite (<https://sourceforge.net/projects/bbmap/>). To remove sequencing adapters from raw reads and low-quality bases (Phred quality score <10), BBduk was used. The resulting reads were mapped against the oligo pool using BBMap (v38.79). Aligned reads were then assigned genes and quantified using the featureCounts (v2.0.1) method of the Subread package (30).

To facilitate the visualization of the 3' ends of reads in coverage plots, wiggle files were created with the bamCoverage (v3.4.3) tool from the deepTools package (31). Here, the counts per million (CPM) normalization option was used to normalize for read depth per library. To get exact positional information on the ribosome position, only the last base of the 3' end of each aligned read was profiled using the -Offset option of bamCoverage. The resulting coverages of the wiggle files were then visualized using integrative genomics viewer (IGV) (32) and used for the identification of translation initiation sites.

Analysis of translation inhibition by PNA

R packages from the tidyverse-suite and edgeR (v3.30.0) were used to analyze the *in vitro* translation of the oligo pool (33,34). A raw count table of quantified reads was imported into the edgeR environment. Oligos with <4.21 CPM were removed. This cutoff was calculated by dividing 10/L, where L is the minimum library size in millions, as proposed by Chen *et al.* (35). Next, read counts were normalized with edgeR's trimmed mean of M values normalization method (36). Differential translation was measured by first estimating the quasi-likelihood dispersions with the glmFit function and then comparing conditions with the glmQLFTest function. Transcripts with false discovery rate-adjusted (37) *P*-values <0.001 and \log_2 fold changes >1 were considered to be differentially translated. In order to screen for possible off-targets of the *acpP*-PNA in the oligo-pool, the PNA sequence was mapped against the whole oligo pool using SeqMap, accepting alignments with up to one mismatch (38).

Metagene analysis

3' End-aligned INRI-seq data was compared to an *in vivo* ribosome profiling dataset of *E. coli* pre-treated with retapamulin (18). The raw coverage files were downloaded from GEO: GSE122129 and adjusted manually so that both datasets show the 3' end of ribosomal footprints. To generate metagene plots of Figure 2A and Supplementary Figure S2F, footprints from -10 to $+55$ and from -30 to $+180$ relative to the start codon were extracted, respectively. The footprints were then normalized against the total read depth of these reads. Then plots were created using functions of the R packages ggplot2 and the dplyr (34).

Computational analysis of translation initiation sites

To identify annotated TISs, an algorithm was implemented using custom python scripts. The algorithm scans through each of the oligo sequences base by base and looks for peaks. For annotated TISs, it searches for peaks in the region of 15 nt (\pm 3 nt) inside the CDS, because that is where

the 3' end of the reads is located during translation initiation, when a ribosome is attached (39). If a peak was identified (CPM normalized depth > 5) in at least two replicates, it was considered a TIS in the INRI-seq dataset (Supplementary Figure S6A, Supplementary Table S2).

In addition to the confirmation of annotated TISs, *in vivo* retapamulin-treated ribosome profiling had been used previously to search for alternative TISs, which do not coincide with annotated start codons (18). Custom python scripts were run to find alternative TISs and to compare them to the *in vivo* data (Supplementary Figure S6B). Each oligo was screened for peaks with > 5 CPM outside its annotated TIS. Only peaks with a relative density (reads of peak divided by the total reads for the respective oligo) > 10% were considered. Additionally, only peaks with a start codon (ATG, TTG, GTG, CTG, ATC and ATT) 15 nt (± 3 nt) upstream of the peak were considered. To prevent the identification of more than one alternative TIS for the same site, peaks in close proximity (up to 5 nt in distance) were merged and the highest respective peak was selected. Next, the reading frame of the alternative TIS relative to the annotated reading frame was noted and a stop codon downstream of the alternative TIS was searched for, designating the alternative TIS as in-frame or out-of-frame. Finally, the *in vivo* Ribo-RET data (18) was searched for the alternative TIS using the same settings.

Start codon analysis

The upstream sequences (37 nt upstream) from our identified TIS peaks were screened for potential start codons (ATG, GTG, TTG, CTG, ATC, ATT), Shine-Dalgarno (SD) motifs (AGG, GGA) and A-rich (AAA) motifs by counting these motifs using a custom R script. Then the frequency of occurrence was visualized as \log_2 counts for annotated TISs detected by INRI-seq, annotated TISs not detected by INRI-seq, alternative TISs detected by INRI-seq, and peaks without called TIS in Supplementary Figure S3A–D using the ComplexHeatmap R package (v4.2) (40).

Additionally, the compositions of start codons were compared between annotated TISs and alternative TISs for both INRI-Seq and Ribo-RET. For this, the abundances of start codons were divided by the total number of the respective TIS type and visualized as percentages (Supplementary Figure S3G).

Prediction of RNA secondary structure

To investigate the strength of secondary structures of our mRNAs around their TISs, regions between -30 and $+15$ relative to the start codon were extracted and the minimum free energy (MFE) was predicted using RNAfold (v2.4.14) (41). As control, RNA sequences of the same length were generated using the dinucleotide content of the *E. coli* genome. Shuffling was done with the esl-shuffle command of the HMMER suite (v3.3.2) (42). Resulting MFE predictions were then plotted in combined beeswarm and boxplots (Figures S2G and S3E). Wilcoxon signed rank tests were applied to test for significant differences between categories.

In vitro translation followed by western blotting

The *acpP::gfp* fusion transcript was obtained as previously described (43). *E. coli* genomic DNA was PCR amplified with JVO-18305 and JVO-18306 (Supplementary Table S7) and cut with NheI and NsiI. Plasmid pXG-10 (43) was also cut with NheI and NsiI and ligated with the *acpP* insert. After transformation and plasmid isolation, the plasmid containing the *acpP::gfp* fusion template was PCR amplified with JVO-18200 and revSF to obtain the dsDNA for *in vitro* transcription. After *in vitro* transcription and purification as described above, *in vitro* translation of the *acpP::gfp* fusion transcript was carried out as for INRI-seq with the following exceptions: Translation reaction volumes were scaled down to a total of 10 μ l, the final transcript concentration was 100 nM, no RET was added to the reactions and translation was carried out for 2 h. After the incubation, 2.5 μ l of 5 \times protein loading buffer (300 mM Tris-HCl, pH 6.8, 50% (v/v) glycerol, 10% (w/v) SDS, 500 mM DTT, 0.05% (w/v) bromophenol blue) was added and the samples incubated at 95°C for 2 min and placed at room temperature. The samples were then separated by 12% SDS-PAGE followed by blotting onto a PVDF membrane (GE Healthcare). The membrane was stained with Ponceau S (Sigma-Aldrich) as loading control and incubated with a mouse α -GFP antibody (Roche). An HRP-coupled goat α -mouse antibody (Thermo Fisher) was finally used to develop the blot.

RESULTS

Design and workflow of INRI-seq

To obtain a synthetic transcriptome for INRI-seq, we designed a single-stranded DNA oligonucleotide pool consisting of 4386 unique oligonucleotides covering all annotated CDSs of *E. coli* K-12 MG1655 (NC.000913.3). For each CDS, we included the natural, *i.e.* annotated, start codon (generally ATG, GTG or TTG) followed by the first 150 nt of the open reading frame (ORF) (Figure 1). We also included the ribosome binding site by addition of 30 nt of the respective 5' untranslated region (UTR), which is the average length of 5' UTRs in *E. coli* (44). At the 3' end, we added a strong TAA stop codon (45) and the 3' UTR of the *E. coli hns* gene for uniform termination of translation. Importantly, this 3' UTR doubled as primer binding site to generate a double-stranded DNA (dsDNA) pool. Finally, a T7 RNA polymerase promoter was added to the 5' end of each oligonucleotide to facilitate *in vitro* transcription, while at the same time representing the second primer binding site for pool amplification. The sequence in between the flanking regions (T7 promoter and 3' UTR) is completely flexible, and can be replaced with the sequence of any gene of interest. In sum, each oligonucleotide in the pool was 242 nt long (Figure 1), which is below the 250 nt limit for oligonucleotide pool production offered by some companies. Furthermore, the included 52 codons should allow all the steps of translation (initiation, elongation, and termination) to occur in close-to-natural manner.

Following a limited-cycle PCR of the oligonucleotide pool, we used the resulting dsDNA as template for T7 *in vitro* transcription. RNA-seq analysis of the obtained syn-

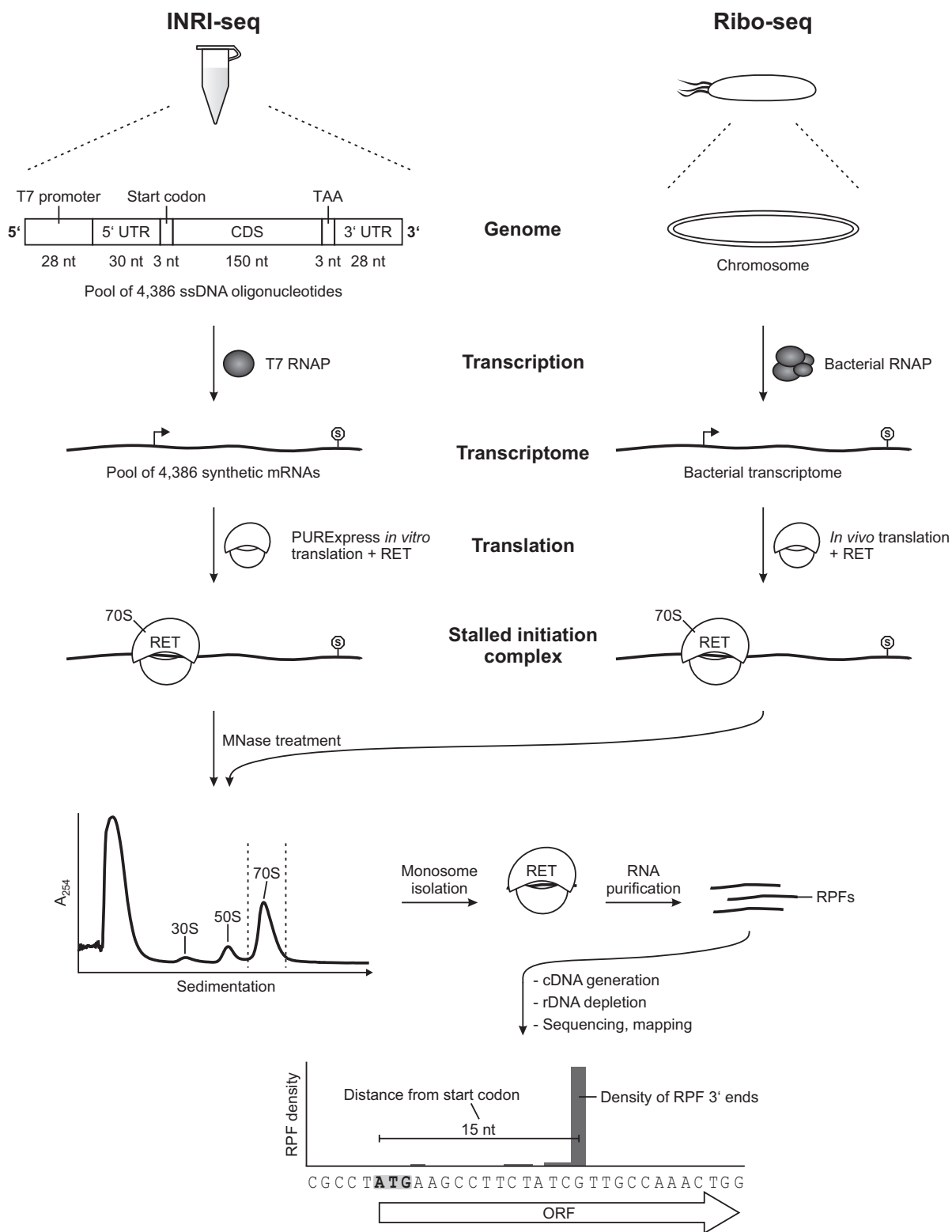


Figure 1. Workflow of INRI-seq. INRI-seq employs a synthetic transcriptome, for the generation of which one ssDNA oligonucleotide is synthesized for each gene of *E. coli* MG1655. Every gene contains a T7 RNA polymerase promoter, 30 nt of its natural 5' UTR, its natural start codon, the first 150 nt of its CDS, a TAA stop codon, and a 28-nt 3' UTR. This pool of oligonucleotides is converted to dsDNA by limited cycle PCR and then transcribed *in vitro* to obtain the synthetic transcriptome. Following *in vitro* translation of the synthetic transcriptome, RET is added and the translation continued, leading to ribosomes stalled at the start codons of the transcripts. RNA not protected by ribosomes is cleaved with MNase and the reaction is sedimented on a sucrose gradient. After 70S monosome isolation, RPFs are extracted from the ribosomes and size-selected on a gel. The purified RPFs are converted to cDNA, rDNA is depleted and the resulting sample is sequenced. Attributing the read density to the 3' ends of the sequenced reads allows identification of the ribosome positions and with it analysis of TISs.

thetic transcriptome detected 4225 of the 4386 synthetic mRNAs at an abundance of >10 reads per million (median abundance: 125 reads per million), showing almost complete (~96%) coverage of the ORFs included in the original DNA pool design (Supplementary Figure S1A).

Next, we applied the INRI-seq protocol to study the TISs of the synthetic *E. coli* transcriptome *in vitro*. We used RET, an antibiotic known to stall the ribosome at the start codon immediately after initiation (Figure 1) (18,23). Since an initiating ribosome protects 14–16 nt of the mRNA downstream of the first nucleotide of the start codon (39), the 3' ends of RPFs will identify the TIS of the translated CDS following Ribo-seq (20). Using toeprinting, a previous study showed that RET-induced stalling also occurs in the *in vitro* system applied for INRI-seq, suggesting that *in vitro* TIS identification would be feasible (18). Since RET does not inhibit translation elongation (18), its addition to translating ribosomes causes polysomes to collapse into monosomes if translation is continued. We found that under our experimental conditions, addition of RET at a final concentration of 5 $\mu\text{g/ml}$ for 30 min after initial translation for 15 min causes polysome collapse (Supplementary Figure S1B). This is considerably longer than the 5 min of RET treatment necessary *in vivo* (18), suggesting that *in vitro* translation of the shorter CDSs used here might be less efficient.

To purify RET-stalled monosomes, the associated transcripts were cleaved with MNase and the digested samples were subsequently run on sucrose gradients, as is common for bacterial Ribo-seq (Figure 1) (16). After collection of the 70S peak, RPFs were extracted with acidic phenol-chloroform followed by size selection (15–45 nt) of RNA in a denaturing polyacrylamide gel. These isolated RPFs can directly be used for library preparation and sequencing (Figure 1). Preliminary sequencing showed 24 fragments of ribosomal RNAs (rRNA; 9 fragments for 23S rRNA, 13 for 16S rRNA, and 2 for 5S rRNA) to be strongly enriched in this short RNA fraction (Supplementary Figure S1C). In the final INRI-seq protocol, these 24 rRNA fragments are depleted from the final cDNA library using a Cas9-based DASH protocol (29), which reduces rRNA reads from >90% to <1% (Supplementary Figure S1D, Supplementary Table S1). This *in vitro* Ribo-seq protocol enables us to globally study translation initiation in a synthetic *E. coli* transcriptome.

INRI-seq identifies annotated TISs

We sequenced five independent INRI-seq libraries (~8 million reads per sample) to study the TISs of *E. coli*, as described above. All libraries correlated well, indicating high reproducibility between experiments (Supplementary Figure S2A). Of the 4225 transcripts present in the transcriptome (Supplementary Figure S1A), 4149 were detected in all five replicates (Supplementary Figure S2B). By assigning the read densities to the 3' ends of the sequenced RPFs, we observed an average distance of 15 nt from the first nucleotide of the annotated start codons, in line with the expected ribosome protection (Figure 2A, Supplementary Figure S2C). Crucially, this distribution was fully consistent with the peak distribution of sequenced RPFs in an *in vivo* study using a similar RET-based protocol (Ribo-RET)

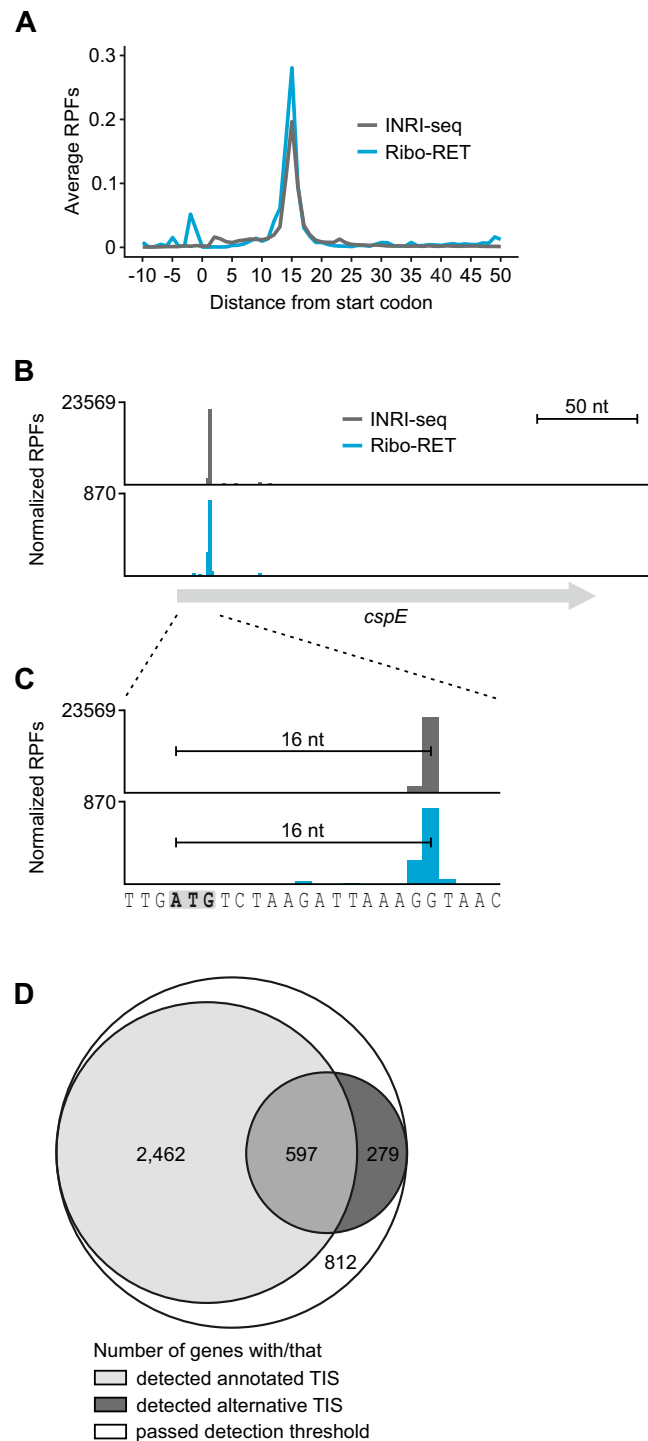


Figure 2. INRI-seq verifies annotated TISs. (A) The average distance of RPF peak density from annotated start codons is 15 nt, in agreement with previous *in vivo* Ribo-RET data (18). (B) INRI-seq detects one RPF peak in the *cspE* gene. (C) INRI-seq RPF peak density is located exactly 16 nt downstream of the annotated start codon (gray and bold) agreeing with previous *in vivo* Ribo-RET data (18). (D) Euler diagram showing the genes passing the thresholds for TIS detection. For 3059 of these, a peak at the annotated start codon was detected, whereas one or more alternative TISs were detected for 876 of them.

(Pearson's correlation $\rho = 0.966$) (18). Manual inspection of the data further revealed that the position of RPF density generally overlapped very well between the *in vitro* and the *in vivo* data, as exemplified by the *ospE* transcript, where the only detected density was close to the start codon (Figure 2B). Indeed, in both the INRI-seq and Ribo-RET datasets, the distance of the RPF peak density to the *ospE* start codon was exactly 16 nt (Figure 2C). The overlap was also present for transcripts with unexpected density distributions such as *hfq*, for which two peaks 13 and 16 nt downstream of the start codon were detected by both methods (Supplementary Figure S2D, E).

Globally, INRI-seq confirmed the annotated TISs of 3059 out of 4386 (~70%) genes using stringent cut-off criteria (RPF peak with counts per million mapped reads (CPM) ≥ 5 within 12–18 nt of the start codon in at least two of the five replicates; Figure 2D, Supplementary Table S2). Using the same criteria, *in vivo* Ribo-RET verified only 780 (~18%) annotated TISs (18) (Supplementary Table S2), suggesting that INRI-seq is more sensitive, likely due to more even transcript abundance in the pool. No TIS was detected for 812 transcripts that passed our detection thresholds (Figure 2D). For 99 of them, a peak was detected in only one of the five replicates, whereas the others did not show a clear peak that could be attributed to a TIS (Supplementary Figure S2F, Supplementary Table S2). This could be due to a variety of reasons, for example the requirement of translation activation by sRNAs or RNA folding, as in the case of certain riboswitches. We therefore predicted the strength of RNA secondary structures, which can hinder translation (9,46,47), from -30 to $+15$ nt with respect to the corresponding start codons of the annotated TISs (Supplementary Figure S2G). The predicted secondary structures around the 1091 annotated TISs INRI-seq did not detect (Figure 2D) was indeed stronger than those of the 3059 detected ones, which could explain why we failed to detect some of them (Supplementary Figure S2G). Furthermore, pseudogenes that are part of the NC_000913.3 annotation and that might not be translated, are present in our transcriptome. We also cannot exclude misannotation of some of the start codons. We observed no major differences in the start codon sequences of the annotated TISs that we identified and the ones that we missed (Supplementary Figure S3A, B). Similarly, we detected the expected enrichment of Shine–Dalgarno (SD)-motifs upstream of these start codons, as well as an enrichment of A-rich sequences, which were recently shown to promote translation initiation (48).

INRI-seq confirms alternative TISs

Stalling ribosomes at start codons using antibiotics is a powerful way of analyzing where on a transcript translation initiation takes place (18–20). In addition to validating known TISs, these kinds of datasets can also be searched for new TISs, revealing alternative in-frame (Figure 3A) or out-of-frame TISs (Figure 3B) within annotated genes. To investigate whether INRI-seq is able to identify alternative TISs, we compared our data to *in vivo* Ribo-RET data (18). Of the 64 alternative TISs identified *in vivo*, we detected 51 (Supplementary Table S3). Most of the alternative TISs that

were missed in the INRI-seq data set were >150 nt downstream of the annotated start codons and therefore not part of our synthetic transcriptome.

Next, we inspected some examples in more detail. The gene encoding arginine decarboxylase, *speA*, was previously shown to contain an in-frame alternative TIS that shortens the secretion signal-containing N-terminus of SpeA by 26 amino acids (aa) and leads to cytoplasmic rather than periplasmic localization of this isoform (18). In accordance with the *in vivo* Ribo-RET data, INRI-seq displayed a strong peak of RPF density 16 nt downstream of this alternative start codon (Figure 3C). Importantly, INRI-seq also detected the start codon of the 43-aa short *yggB* gene encoded upstream of *speA*, which was not possible *in vivo*, most likely due to its low abundance (Figure 3C). Different from the alternative in-frame TIS in *speA*, the *oppA* gene encoding an oligopeptide uptake protein as well as the OppX RNA sponge (49,50) contains an out-of-frame TIS for an alternative 7-aa ORF (18). Our INRI-seq data supports this prediction, exhibiting the same RPF peaks as the *in vivo* data (Figure 3D). While it is unclear if this 7-aa peptide has a biological function, translation of this alternative frame could impact the translation of *oppA*, similarly to upstream ORFs that can regulate translation of their downstream genes (51). In conclusion, these examples illustrate how INRI-seq not only captures annotated TISs but also reports TISs that were not detected by *in vivo* Ribo-seq.

INRI-seq identifies putative new TISs

After confirming that INRI-seq is able to detect both annotated and known alternative TISs with high sensitivity, we asked whether this method also enables the detection of putative new TISs. To do so, we searched for peaks with a relative density of $\geq 10\%$ of the total reads identified for a given gene. These were further filtered to only include peaks of ≥ 5 CPM, which had to be present in at least two of the five replicates. Using these criteria, we identified 918 putative new TISs, 279 of which could be assigned to genes for which INRI-seq did not detect the annotated TIS (Figure 2D, Supplementary Table S3). Similar to the annotated TISs (Supplementary Figure S3A, B), the putative new TISs showed an enrichment of ATG start codons, although alternative start codons such as GTG or TTG occurred in higher frequency (Supplementary Figure S3C). In contrast, no enrichment of SD-motifs upstream of our putative new TISs was observed. We did, however, notice a strong enrichment of A-rich sequences upstream of these TISs, which also promote translation initiation (48), indicating that these putative new TISs might be translated in absence of classic SD-motifs. Finally, we analyzed the INRI-seq peaks (relative density of $\geq 10\%$ of the total reads, ≥ 5 CPM in at least two of the five replicates) for which no TIS was called (Supplementary Figure S3D). As expected, no start codons were enriched 12 to 18 nt upstream of these peaks. There was, however, an enrichment in ATG start codons and SD-motifs closer to the peaks, indicating that our distance cut-offs were too stringent to call these potential TISs.

Stable mRNA secondary structures around TISs negatively influence translation initiation (9,46,47). We therefore analyzed the predicted free energy from -30 to $+15$ nt

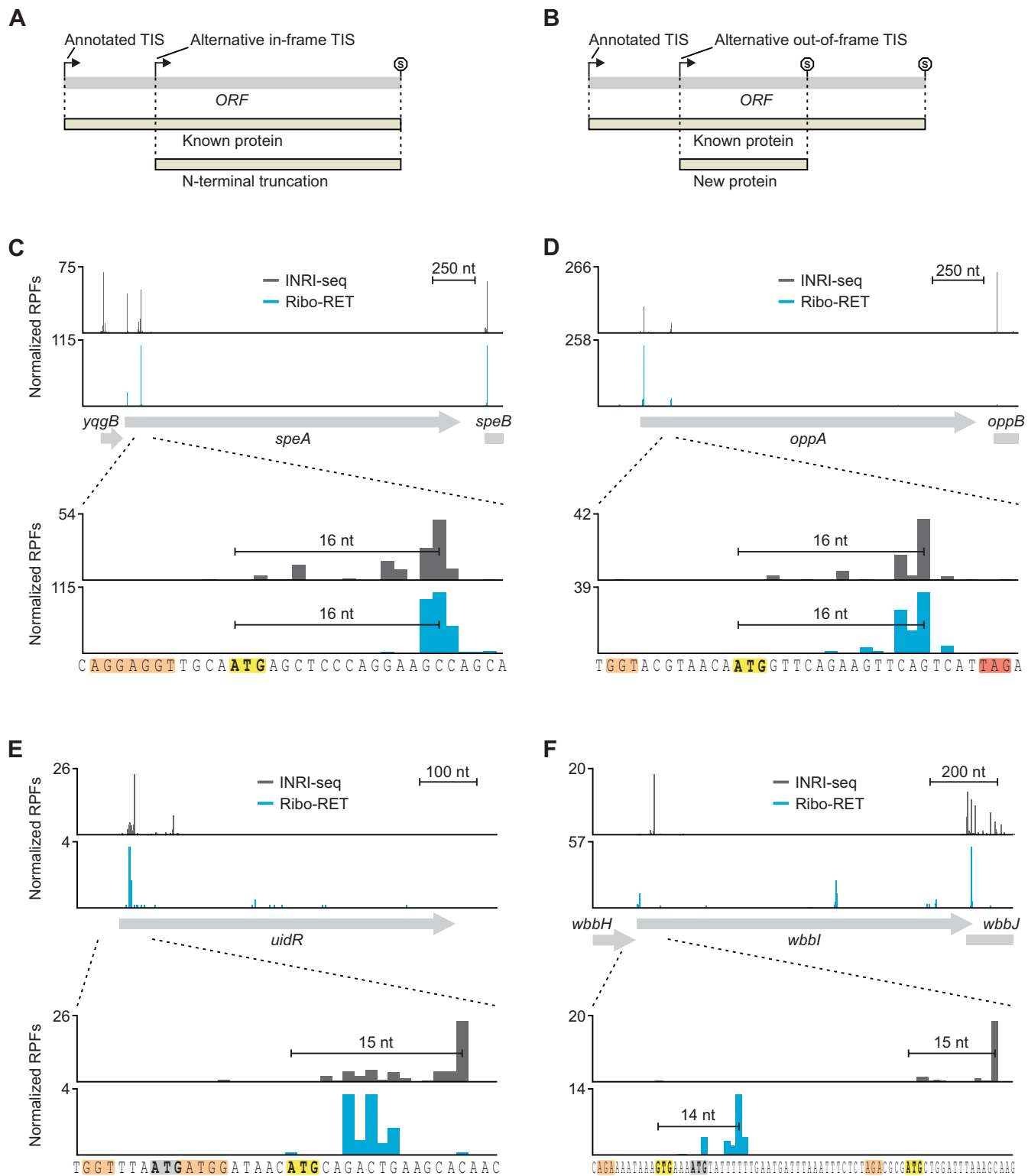


Figure 3. INRI-seq verifies known alternative TISs and identifies new putative TISs. (A) Alternative in-frame TISs lead to N-terminally truncated versions of known proteins. (B) Alternative out-of-frame TISs generate new proteins. (C) INRI-seq detects two RPF peaks at the 5' end of *speA*. The second RPF peak identified in *speA* belongs to a known, in-frame alternative start codon (18). (D) INRI-seq detects two RPF peaks at the 5' end of *oppA*. The second RPF peak identified in *oppA* belongs to a known, out-of-frame alternative start codon (18). (E) INRI-seq detects an RPF peak at the 5' end of *uidR*. The detected RPF peak cannot be attributed to the annotated start codon of *uidR* but rather derives from an in-frame alternative start codon. (F) INRI-seq detects an RPF peak toward the 5' end of *wbbI*. The detected RPF peak cannot be attributed to the annotated start codon of *wbbI* but rather derives from an out-of-frame alternative start codon. Gray and bold, annotated start codon. Yellow and bold, alternative start codon. Orange, SD sequence. Red, stop codon.

with respect to the corresponding start codons of the putative new TISs. While the putative new TISs showed a slightly stronger predicted secondary structure when compared to annotated TISs, the structures were significantly weaker than the ones of random sequences (Supplementary Figure S3E). This suggests that the putative TISs might mediate translation initiation.

INRI-seq detected the majority of TISs reported in the *in vivo* Ribo-RET data, only missing the annotated TISs of 104 genes that could be verified *in vivo* (Supplementary Figure S3F). For only five genes, both datasets detected alternative TISs without a signal for their annotated TISs, whereas both the annotated and the same alternative TISs were identified for 39 genes in both studies. Finally, we examined the distribution of used start codons among the annotated *versus* the alternative TISs (Supplementary Figure S3G). While ~88% of annotated *E. coli* TISs use ATG start codons, the alternative TISs of both INRI-seq and *in vivo* Ribo-RET identified considerably more non-ATG start codons: >50% of the alternative TISs use GTG, TTG, CTG, ATC or ATT as the start codon. While these less common start codons were less frequent than ATG among the alternative TISs, they still occurred more frequently 12–18 nt upstream of INRI-seq peaks than other trinucleotides, supporting their role in translation initiation (Supplementary Figure S3C). Overall, these data indicate that INRI-seq is able to detect the majority of annotated and alternative TISs previously reported *in vivo*. We then studied two interesting candidates of putative new TISs in more detail.

UidR is a transcriptional repressor of the *uid* operon, which is involved in transport and degradation of β -glucosides (52). INRI-seq showed a clear RPF density peak toward the 5' end of *uidR*, but downstream of the annotated start codon (Figure 3E). The distance of the identified peak agreed well with an in-frame ATG four codons downstream of the annotated start codon, suggesting that the UidR protein might be 4 aa shorter than annotated (Figure 3E). Indeed, this shorter N-terminus is supported by a recent mass spectrometry study of the N-termini of *E. coli* proteins, which detected the exact peptide (MQTEAQPTR) that corresponds to the TIS identified by INRI-seq (53). By contrast, *in vivo* Ribo-RET (18) showed only minimal density in this position, leaving the TIS of *uidR* ambiguous.

Another interesting alternative TIS candidate identified by INRI-seq is present in *wbbI* (Figure 3F), the gene encoding β -1,6-galactofuranosyltransferase (54). As for *uidR*, both INRI-seq and Ribo-RET did not detect RPF density corresponding to the annotated start codon of *wbbI* (Figure 3F). While *in vivo* Ribo-RET identified an alternative GTG start codon two codons upstream of the annotated start codon, INRI-seq detected a peak corresponding to an out-of-frame ATG close to the 5' end of the gene. This internal ORF encodes for a 31-aa long peptide, for which we could not find any homologous domains or sequences using Pfam or PHMMER queries, respectively (55,56). Follow-up *in vivo* studies will be necessary to reveal whether this peptide is produced within the cell. Interestingly, although *in vivo* Ribo-RET did not identify this alternative TIS, it revealed another out-of-frame internal ORF encoding a 47-aa peptide within *wbbI* (18). Due to its distance from the annotated start codon, INRI-seq was unable to detect this

additional alternative TIS. Yet, it is intriguing that a single gene might contain two different out-of-frame ORFs. Overall, INRI-seq is able to validate known alternative TISs and allows identification of putative new TISs, which should facilitate follow-up *in vivo* studies.

Analysis of *in vitro* translation inhibition by PNAs

Translation inhibition by antisense oligomers, specifically peptide nucleic acid (PNA), is an upcoming field of antibiotics research. PNAs have the potential for species-specific killing, only targeting the pathogen while leaving beneficial microbiota untouched (24–26). Similar to bacterial sRNAs (8), antisense oligomers inhibit translation by blocking ribosome access to the ribosome binding site of their targets (57). Yet, the high sample volumes required by standard Ribo-seq studies pose a challenge for studies of translation control and potential off-targeting by such antisense antibiotics *in vivo* because of the high manufacturing costs of PNAs. We therefore decided to exploit the small-volume, cell-independent nature of INRI-seq to quantitatively study the effects of varying PNA concentrations on translation in a transcriptome-wide manner.

To test whether PNA-mediated translation inhibition can be analyzed *in vitro*, we adopted the well-established translation inhibition by a 10-mer antisense PNA that sequesters the start codon region of the *E. coli* *acpP* mRNA, which encodes the essential acyl carrier protein (Figure 4A) (28,58,59). To this end, we created a translational fusion transcript containing 50 nt of the 5' UTR and the first 8 codons of *acpP* fused to the CDS of *gfp*. Upon addition of increasing concentrations of *acpP*-PNA, *in vitro* synthesis of the AcpP::GFP fusion protein was strongly reduced (Figure 4B). Importantly, this was not the case when a scrambled version of the PNA (*acpP*-PNA-scr), *i.e.* the same nucleobases as in *acpP*-PNA but in a randomly shuffled order, was added to the reaction. At an equimolar ratio of *acpP*-PNA to *acpP*::*gfp* (100 nM), translation was inhibited by ~17%, whereas a 5-fold excess of *acpP*-PNA caused a ~74% drop in translation (Figure 4C). These results are consistent with a previous *in vitro* study analyzing PNA efficiency in a less-defined reticulocyte extract using DNA as template (60) and demonstrate that our *in vitro* system can be used to study the effect of antisense oligomers.

Global analysis of PNA target inhibition using INRI-seq

Next, we set out to use INRI-seq to study the global effects of *acpP*-PNA on translation of its intended target as well as potential off-targets. We employed the same protocol that we used for the TIS analysis, except that we pre-annealed the synthetic transcriptome with varying concentrations of *acpP*-PNA or *acpP*-PNA-scr before starting the translation reactions. Similarly to the observed repression of the *acpP*::*gfp* fusion above (Figure 4B, C), INRI-seq reported a clearly reduced RPF density on the synthetic *acpP* mRNA with increasing *acpP*-PNA concentrations (Figure 5A). No inhibition of *acpP* translation was detected when *acpP*-PNA-scr was added, showing the specificity of PNA targeting. While the estimated concentration of *acpP* in our synthetic transcriptome was ~500 pM (Supplementary

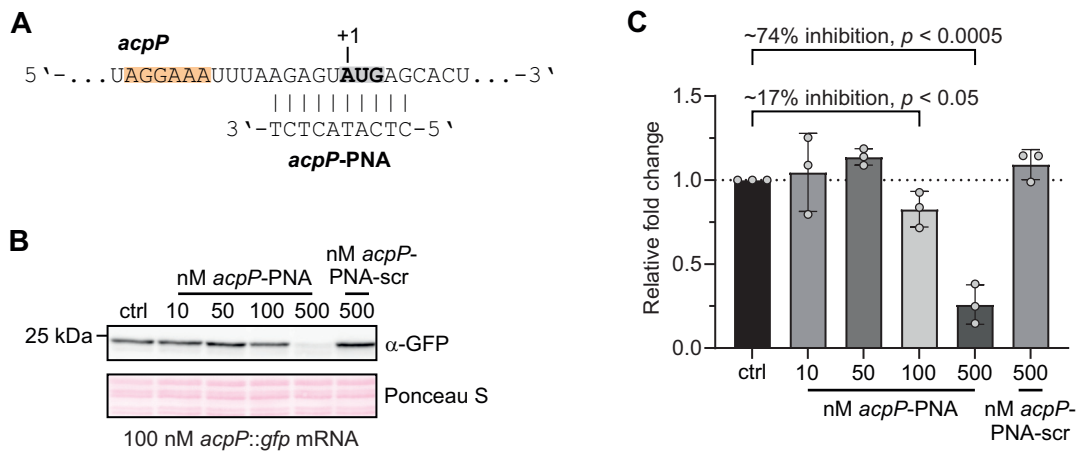


Figure 4. *In vitro* translational inhibition by antisense PNA. (A) Region of the *acpP* transcript targeted by *acpP*-PNA. Gray and bold, start codon. Orange, SD sequence. (B) Western blotting of an *in vitro* translation of an *acpP::gfp* fusion transcript reveals reduced GFP protein levels with increasing concentrations of *acpP*-PNA. Ponceau S staining was used as loading control. (C) Quantification of three independent western blots as shown in (B). Error bars show mean with SD. *P*-values were calculated using an unpaired, two-tailed *t*-test.

Table S4), the actual concentration was likely higher: During the establishment of INRI-seq, we sequenced a test sample without addition of RET, revealing high levels of reads coming from the 3' region of *acpP*, which was not part of the synthetic transcriptome (Supplementary Figure S4A). This suggested that the *in vitro* translation kit probably contains traces of highly abundant RNAs such as *acpP*. This hampers accurate estimations of ratios between *acpP*-PNA and its targets. Still, even low *acpP*-PNA concentrations showed a strong effect on *acpP* translation.

In the *E. coli* genome, the *acpP* mRNA is co-transcribed with the downstream gene *fabF* (61). *acpP*-PNA has been shown to lead to rapid decay of both cistrons following translation inhibition in cells, even though *fabF* is not a target of the antisense oligomer (28). In agreement with the notion that *fabF* is down-regulated *in vivo* because its transcript is coupled to *acpP*, INRI-seq did not detect any change in translation of *fabF* *in vitro* (where the two transcripts are uncoupled), suggesting that INRI-seq faithfully reports only direct PNA-induced translational changes (Figure 5B).

INRI-seq identifies direct PNA off-targets

We then analyzed the full spectrum of *acpP*-PNA-regulated targets. While *acpP* clearly was the primary transcript affected by the antisense oligomer, translation of a few other transcripts was also reduced at higher *acpP*-PNA concentrations (Figure 5C). One of them was *yqjF*, which has a stretch of 9 complementary bases to *acpP*-PNA within its translation initiation region (Figure 5D). Only the very 5' nucleotide of the PNA is a mismatch to *yqjF* (C:U). This suggests that at high concentration, *acpP*-PNA will also repress mRNAs that harbor a mismatch to the PNA sequence. Crucially, this was true for the additional off-targets *gpmM* and *ugpQ*. The translation initiation regions of both genes are complementary to *acpP*-PNA with only one mismatch at the 3' end of the PNA (Figure 5C, Supplementary Figure S4B, C). Finally, translation of the *gpp* transcript, which has complementarity to the antisense oligomer around its start

codon, but with a T-G pair formed by the PNA's second position (Supplementary Figure S4D), was downregulated at 1 μ M *acpP*-PNA as well, although it did not meet our strict cut-off conditions ($p_{\text{adj}} < 0.001$, $\log_2 \text{FC} > 11$; Figure 5C).

The downregulated off-target *arnC* exhibits full complementarity to the antisense oligomer (Figure 5C, Supplementary Table S5). However, to our surprise, the *acpP*-PNA binds ~ 120 nt downstream of the start codon of *arnC* (Figure 5E) where antisense repression of translation initiation is unlikely (57,62,63). Interrogating our INRI-seq TIS dataset for alternative TISs within *arnC*, we indeed detected a peak in RPF density corresponding to an alternative in-frame TTG start codon, located a few nucleotides upstream of the *acpP*-PNA binding site (Supplementary Figure S5A). RPF coverage around the annotated TIS was not affected by PNA addition (Supplementary Figure S5B), while the alternative in-frame TIS clearly showed the expected depletion in RPFs when PNA was added (Supplementary Figure S5C). Since no TISs were identified for *arnC* by *in vivo* Riboret (18), it remains unclear whether this alternative TIS is used *in vivo*.

Overall, these results suggest that, while target mismatches within the PNA sequence are generally thought to render the PNA inactive (58,64), mismatches at the 5' or 3' ends of the PNA are tolerated, leading to target regulation at high PNA concentration. In particular, of the 16 genes that harbor an *acpP*-PNA complementary sequence with one mismatch within their translation initiation region, six have the mismatch in position 1 or 10 of the PNA (Supplementary Table S5). Among those, the off-targets *gpmM*, *ugpQ* and *yqjF* are significantly downregulated, as discussed above. Of the other three, *upp* and *yodC* showed a tendency to be downregulated at the highest *acpP*-PNA concentrations, though they missed our statistical cut-offs (Figure 5C), while *araH* was not regulated at all. The ten genes whose translation initiation regions include a complementary sequence to *acpP*-PNA with one mismatch in positions 2–8, showed no significant regulation in presence of the antisense oligomer. The same was true for all potential off-targets with two or more mismatches.

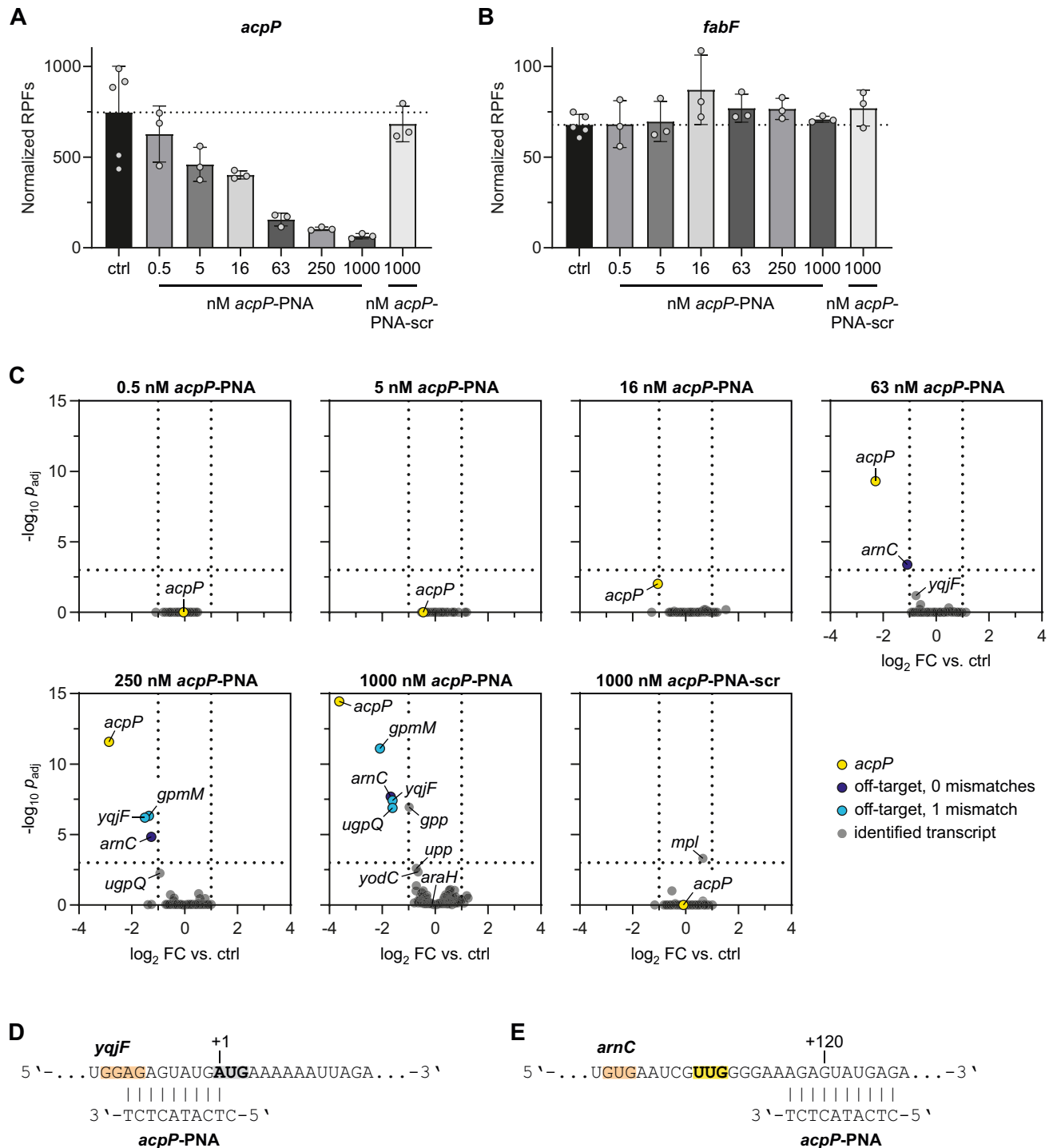


Figure 5. INRI-seq globally identifies targets of translation inhibition by PNA. (A) *acpP*-PNA inhibits translation of *acpP* in a concentration-dependent manner. Error bars show mean with SD. (B) *acpP*-PNA does not inhibit translation of *fabF*. Error bars show mean with SD. (C) Volcano plots showing calculated changes in translation as FDR-adjusted p -value against fold change. Cut-offs were set to $p_{adj} < 0.001$, $\log_2 FC > 1$ and are depicted as dotted lines. (D) Region of the *yqjF* transcript targeted by *acpP*-PNA. Gray and bold, start codon. Orange, SD sequence. (E) Region of the *arnC* transcript targeted by *acpP*-PNA. Yellow and bold, alternative start codon (compare to Supplementary Figure S5A). Orange, SD sequence.

This analysis suggests that INRI-seq is able to accurately determine direct transcriptome-wide PNA-mediated inhibition of translation in response to defined PNA concentrations in a small reaction volume. Thus, it not only allows quantitative analysis of the direct effects of the PNA on the intended target transcript, but also enables the identification of PNA off-targets.

DISCUSSION

In this study, we established INRI-seq, a cell-free method to globally study translation of a fully customizable synthetic transcriptome. Presenting two applications for this method, we identified known and putative new *E. coli* TISs and determined the fidelity of translation inhibition by antisense oligomers. Although optimized here for these purposes, the INRI-seq protocol is easily adaptable to other translation-related questions. For example, our protocol for PNA analysis can also be applied to the identification of the global targetome of sRNAs, including transcripts that have low abundance *in vivo*. This would allow the expansion of regulatory networks of known sRNAs and the investigation of the targets of currently understudied sRNAs. For single sRNA targets, *in vitro* translation has been successfully used to investigate sRNA-mediated regulation, indicating that a global study based on INRI-seq is feasible (63,65,66). Importantly, INRI-Seq allows to include within the synthetic transcriptome targets that carry mutations in the putative sRNA binding site based on *in silico* predictions (67). This enables testing of targeting specificity in the same experiment.

Since INRI-seq uses a fully synthetic transcriptome, users can analyze any desired sequence compilation. For example, the method can be used to compare translation of transcriptomes of related bacterial species or to analyze translation within complex microbial communities such as the human gut microbiome. INRI-seq also enables the analysis of translation of phage transcripts independent of its host. It also facilitates the study of overlapping CDSs—a common feature in the complex genome organization of phages (68)—by disentangling them into single transcripts. Furthermore, mutational studies, in which a single or multiple transcripts are present in tens or even hundreds of different variants, could be designed to investigate the influence of, for example, synonymous mutations on translation.

The design of the synthetic transcriptome is only limited by the length of the oligonucleotides: currently, up to 350 nt is commercially possible. This restriction necessitates the disruption of bacterial operon structures, which are known to be an important factor in translation regulation (9,69,70). Yet, this property can also be exploited to discriminate between direct and indirect effects: for example, we show that *acpP*-PNA directly inhibits translation of *acpP*, but not of *fabF* (Figure 5A, B). In contrast, *in vivo* the PNA influences the stability of the entire dicistronic *acpP-fabF* mRNA (27,28), which makes it harder to assess its direct effect on translation.

The number of oligonucleotides in a pool (*i.e.* the available sequence space) is less restricted than the length of the oligonucleotides, with several commercial providers able to synthesize extensive oligo pools. This enables the study of

more complex synthetic transcriptomes, including eukaryotic ones. In this case, eukaryotic cell-free translation systems, such as those commercially available for human or rabbit (71), should be used in order to obtain faithful results for these organisms.

INRI-seq allows the study of translation in a quantitative manner, which is supported by the even coverage of transcripts in the synthetic transcriptome (half of the transcripts have concentrations within a ~2-fold range of the mean concentration) (Supplementary Figure S1A) and read counts in the INRI-seq data (Supplementary Figure S2B). Concentration-dependent analysis of translation-modulating molecules becomes feasible, because their exact concentration in the reaction is defined. This is not true in an *in vivo* setup, where the intracellular concentration is unknown. In addition, modulators can be added at high concentrations, which might otherwise lead to premature cell lysis when working with living cells. The small volume of the INRI-seq reactions (25 μ l) is another advantage, especially if the translation modulator is limited or particularly expensive, like PNAs or antibiotic lead structures. Moreover, since there is no cell barrier or other resistance mechanisms to consider, comparative analysis of different types of antisense oligomers becomes possible, although there is no general mechanism for the cellular delivery of these compounds (24,72–75). This enables the study of the direct targets of such molecules, which is often difficult to disentangle in an *in vivo* setup due to the simultaneous occurrence of direct and indirect effects.

In this study, we globally analyzed TISs of a synthetic *E. coli* transcriptome. Compared to a technically similar *in vivo* study (18), we validated almost four times more annotated TISs (3059 vs. 780), demonstrating the high sensitivity of INRI-seq. In addition, we detected 51 out of the 64 alternative TISs (Supplementary Table S3) detected by *in vivo* Ribo-RET (18). INRI-seq further identified 918 additional putative alternative TISs for 876 genes, of which 279 are the only detected TISs for the respective genes (Figure 2D). These results are in agreement with a recent study suggesting that 5–12% of prokaryotic genes might have misannotated TISs (76). One reason for TIS misannotation is that gene prediction algorithms generally rely on SD-motifs upstream of the start codon (76,77). Yet, translation of transcripts lacking SD-motifs is relatively frequent (77). For example, it was recently shown in *E. coli* that A-rich sequences promote translation initiation, whereas SD-motifs influence initiation efficiency, but are not necessary to determine the site of initiation (48). For the putative new TISs we detected in this study, we do not observe enrichment of SD-motifs; instead, we find enrichment of A-rich motifs (Supplementary Figure S3C), which indicates that these TISs might be translated in absence of classic SD-motifs. Still, *in vivo* validation is required to verify whether these TISs are used within the cell, for example, by mass spectrometry. Analogous to our application of RET to map TISs, we expect INRI-seq to be a useful tool to study the effects of other translation-inhibiting drugs, such as macrolides, lincosamides or tetracyclines. INRI-seq further complements inverse toeprinting, a related *in vitro* method that enables the high throughput analysis of ribosome stalling sites and their upstream arrest sequences (78).

To exploit the cell-free, low-volume nature of INRI-seq, we globally analyzed the influence of an antisense PNA on translation. In line with *in vivo* data (27,28), we found that *acpP*-PNA is specific for its target, *acpP*. Down-regulation of PNA off-targets that harbor mismatches in the complementary sequence requires substantially higher PNA concentrations. In addition, mismatches are only tolerated in the terminal nucleotides of the PNA. Gratifyingly, this position-dependent effect of mismatches on PNA-mediated translation inhibition was also recently shown for *Salmonella* using a combination of minimum inhibitory concentration and RNA-Seq experiments (79). Further, a recent study investigating the effects of *acpP*-PNA on global transcript levels in a uropathogenic *E. coli* (UPEC) strain observed very similar off-target effects as the ones identified by INRI-seq: Of the four off-targets identified by INRI-seq, *gpmM* and *ugpQ* were among the top regulated transcripts upon addition of *acpP*-PNA (27). These findings should aid future PNA design by providing a framework to weigh the predicted specificity of PNAs according to their predicted off-targets.

Limitations of the study

Despite the clear value of INRI-seq for the *in vitro* analysis of translation, as discussed above, there are several caveats to consider. The commercial *in vitro* translation kit (PURE-Express, New England Biolabs) used in the INRI-seq protocol is based on the PURE system (80). It consists of individually purified *E. coli* components required for protein synthesis, such as initiation, elongation and termination factors as well as aminoacyl-tRNA synthetases, tRNAs, amino acids and ribosomes. This reaction mix allows efficient translation of most transcripts. Nevertheless, the lack of factors like the translation elongation factor EF-P, which is necessary for efficient translation of polyproline stretches (81,82), could hamper the study of specific transcripts. Likewise, the concentrations of the loosely associated ribosomal proteins S1 and bL31, which promote translation of certain mRNAs and the association of the ribosomal subunits, respectively (83,84), might be unnaturally low due to loss during ribosome purification. This might lead to less efficient translation compared to the *in vivo* situation.

RNA-binding proteins such as Hfq, which facilitates base pairing between sRNAs and their targets (85), are not part of the PURE system, either. Therefore, these proteins need to be added to the reaction mixture when analyzing the effects of sRNAs, as previously done in an analysis of the sRNA GlmZ (66). sRNAs that activate or repress translation of certain mRNAs are also lacking from the kit, which can influence the translation of particular transcripts (8). Similarly, the concentration of the single transcripts in our pool does not reflect the ratios present *in vivo*. This might lead to unnatural competition for ribosome binding, favoring translation of certain transcripts compared to the *in vivo* situation. Finally, the presence of contaminating RNAs (Supplementary Figure S1C), which are likely co-purified with the individual proteins of the kit, has to be kept in mind when designing concentration-sensitive, quantitative experiments. Despite these drawbacks, the PURE system works very well for the applications tested here. It enabled INRI-seq to globally analyze TISs of *E. coli* as well as the direct

effects of PNA with high sensitivity. By using other cell-free translation systems, such as self-generated S30 extracts from bacteria other than *E. coli* (86), we expect INRI-seq to be applicable to more distantly related organisms such as the Gram-positive model *Bacillus subtilis*.

The necessary shortening of the transcripts in the synthetic transcriptome has disadvantages as well. It may influence mRNA folding, which in turn can impact translational efficiency (9) by, for example, affecting start codon recognition or diverting ribosomes to alternative TISs. The shortened transcripts may also engage in aberrant interactions, which could lead to sequestration of ribosome binding sites. Further, the shortened ORFs present in our transcriptome will result in the translation of unphysiological, truncated proteins, which might influence translation in an unpredictable manner. Nevertheless, INRI-seq provides a rapid, flexible *in vitro* protocol to investigate a wide range of translation-related questions that cannot be easily addressed *in vivo*. We expect this method to be an attractive alternative to *in vivo* Ribo-seq, particularly when availability or delivery of a translation-modulating molecule such as an antisense oligonucleotide of interest is limiting.

DATA AVAILABILITY

The sequencing data have been deposited in NCBI's Gene Expression Omnibus (87) and are accessible through GEO Series accession number GSE190954 (<https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE190954>). The custom code used to analyze the sequencing data can be found on GitHub (https://github.com/BarquistLab/oligopool_riboseq).

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

We thank the members of the Vogel lab for splendid ideas and discussions, Lydia Hadjeras and Sarah Svensson for their invaluable help to set up ribosome profiling, and Anke Sparmann for editing the manuscript. We thank the Core Unit SysMed at the University of Würzburg for excellent technical support and RNA-seq data generation.

Author contributions: Conceptualization: J.H., J.V. Methodology: J.H., J.J., L.B., J.V. Software: J.J., L.B. Validation: J.H., J.J., S.D.-M. Formal analysis: J.H., J.J. Investigation: J.H. Resources: L.B., J.V. Data curation: J.J., L.B. Writing – Original Draft: J.H., J.V. Writing – Review & Editing: J.H., J.J., S.D.-M., L.B., J.V. Visualization: J.H. Supervision: L.B., J.V. Project administration: L.B., J.V. Funding acquisition: L.B., J.V.

FUNDING

Bavarian Bayresq.net project Rbiotics (to L.B., J.V.); German Research Council Leibniz Award [DFG Vo875-18 to J.V.]. Funding for open access charge: Library, University of Würzburg.

Conflict of interest statement. None declared.

REFERENCES

1. Tollerson, R. 2nd and Ibba, M. (2020) Translational regulation of environmental adaptation in bacteria. *J. Biol. Chem.*, **295**, 10434–10445.
2. Hu, X.P., Dourado, H., Schubert, P. and Lercher, M.J. (2020) The protein translation machinery is expressed for maximal efficiency in *Escherichia coli*. *Nat. Commun.*, **11**, 5260.
3. Hör, J., Gorski, S.A. and Vogel, J. (2018) Bacterial RNA biology on a genome scale. *Mol. Cell*, **70**, 785–799.
4. Bartholomäus, A., Fedyunin, I., Feist, P., Sin, C., Zhang, G., Valleriani, A. and Ignatova, Z. (2016) Bacteria differently regulate mRNA abundance to specifically respond to various stresses. *Philos. Trans. A Math. Phys. Eng. Sci.*, **374**, 20150069.
5. Ishihama, Y., Schmidt, T., Rappsilber, J., Mann, M., Hartl, F.U., Kerner, M.J. and Frishman, D. (2008) Protein abundance profiling of the *Escherichia coli* cytosol. *BMC Genomics*, **9**, 102.
6. 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.
7. Fremin, B.J., Sberro, H. and Bhatt, A.S. (2020) MetaRibo-Seq measures translation in microbiomes. *Nat. Commun.*, **11**, 3268.
8. Hör, J., Matera, G., Vogel, J., Gottesman, S. and Storz, G. (2020) Trans-acting small RNAs and their effects on gene expression in *Escherichia coli* and *Salmonella enterica*. *EcoSal Plus*, **9**, ESP-0030–2019.
9. Burkhardt, D.H., Rouskin, S., Zhang, Y., Li, G.W., Weissman, J.S. and Gross, C.A. (2017) Operon mRNAs are organized into ORF-centric structures that predict translation efficiency. *Elife*, **6**, e22037.
10. Ingolia, N.T., Ghaemmhami, S., Newman, J.R. and Weissman, J.S. (2009) Genome-wide analysis in vivo of translation with nucleotide resolution using ribosome profiling. *Science*, **324**, 218–223.
11. Li, G.W., Oh, E. and Weissman, J.S. (2012) The anti-Shine-Dalgarno sequence drives translational pausing and codon choice in bacteria. *Nature*, **484**, 538–541.
12. Oh, E., Becker, A.H., Sandikci, A., Huber, D., Chaba, R., Gloge, F., Nichols, R.J., Typas, A., Gross, C.A., Kramer, G. *et al.* (2011) Selective ribosome profiling reveals the cotranslational chaperone action of trigger factor in vivo. *Cell*, **147**, 1295–1308.
13. Ingolia, N.T. (2014) Ribosome profiling: new views of translation, from single codons to genome scale. *Nat. Rev. Genet.*, **15**, 205–213.
14. Ingolia, N.T., Hussmann, J.A. and Weissman, J.S. (2019) Ribosome profiling: global views of translation. *Cold Spring Harb. Perspect. Biol.*, **11**, a032698.
15. Brar, G.A. and Weissman, J.S. (2015) Ribosome profiling reveals the what, when, where and how of protein synthesis. *Nat. Rev. Mol. Cell Biol.*, **16**, 651–664.
16. Mohammad, F., Green, R. and Buskirk, A.R. (2019) A systematically-revised ribosome profiling method for bacteria reveals pauses at single-codon resolution. *Elife*, **8**, e42591.
17. Tesina, P., Lessen, L.N., Buschauer, R., Cheng, J., Wu, C.C., Berninghausen, O., Buskirk, A.R., Becker, T., Beckmann, R. and Green, R. (2020) Molecular mechanism of translational stalling by inhibitory codon combinations and poly(A) tracts. *EMBO J.*, **39**, e103365.
18. 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. *Mol. Cell*, **74**, 481–493.
19. 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.
20. Weaver, J., Mohammad, F., Buskirk, A.R. and Storz, G. (2019) Identifying small proteins by ribosome profiling with stalled initiation complexes. *Mbio*, **10**, e02819–e02818.
21. Harms, A., Brodersen, D.E., Mitarai, N. and Gerdes, K. (2018) Toxins, targets, and triggers: an overview of toxin-antitoxin biology. *Mol. Cell*, **70**, 768–784.
22. Nayfach, S., Shi, Z.J., Seshadri, R., Pollard, K.S. and Kyrpides, N.C. (2019) New insights from uncultivated genomes of the global human gut microbiome. *Nature*, **568**, 505–510.
23. Meydan, S., Klepacki, D., Mankin, A.S. and Vázquez-Laslop, N. (2021) Identification of translation start sites in bacterial genomes. *Methods Mol. Biol.*, **2252**, 27–55.
24. Kole, R., Krainer, A.R. and Altman, S. (2012) RNA therapeutics: beyond RNA interference and antisense oligonucleotides. *Nat. Rev. Drug Discov.*, **11**, 125–140.
25. Pifer, R. and Greenberg, D.E. (2020) Antisense antibacterial compounds. *Transl Res.*, **223**, 89–106.
26. Vogel, J. (2020) An RNA biology perspective on species-specific programmable RNA antibiotics. *Mol. Microbiol.*, **113**, 550–559.
27. Popella, L., Jung, J., Do, P.T., Hayward, R.J., Barquist, L. and Vogel, J. (2022) Comprehensive analysis of PNA-based antisense antibiotics targeting various essential genes in uropathogenic *Escherichia coli*. *Nucleic Acids Res.*, **50**, gkab362.
28. Popella, L., Jung, J., Popova, K., Đurica-Mitić, S., Barquist, L. and Vogel, J. (2021) Global RNA profiles show target selectivity and physiological effects of peptide-delivered antisense antibiotics. *Nucleic Acids Res.*, **49**, gkab242.
29. Prezza, G., Heckel, T., Dietrich, S., Homberger, C., Westermann, A.J. and Vogel, J. (2020) Improved bacterial RNA-seq by Cas9-based depletion of ribosomal RNA reads. *RNA*, **26**, 1069–1078.
30. Liao, Y., Smyth, G.K. and Shi, W. (2014) featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics*, **30**, 923–930.
31. Ramírez, F., Dündar, F., Diehl, S., Grüning, B.A. and Manke, T. (2014) deepTools: a flexible platform for exploring deep-sequencing data. *Nucleic Acids Res.*, **42**, W187–W191.
32. Robinson, J.T., Thorvaldsdóttir, H., Winckler, W., Guttman, M., Lander, E.S., Getz, G. and Mesirov, J.P. (2011) Integrative genomics viewer. *Nat. Biotechnol.*, **29**, 24–26.
33. Robinson, M.D., McCarthy, D.J. and Smyth, G.K. (2010) edgeR: a bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics*, **26**, 139–140.
34. Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L., François, R., Grolemund, G., Hayes, A., Henry, L., Hester, J. *et al.* (2019) Welcome to the tidyverse. *J. Open Source Softw.*, **4**, 1686.
35. Chen, Y., Lun, A.T. and Smyth, G.K. (2016) From reads to genes to pathways: differential expression analysis of RNA-Seq experiments using rsubread and the edgeR quasi-likelihood pipeline. *F1000Res*, **5**, 1438.
36. Robinson, M.D. and Oshlack, A. (2010) A scaling normalization method for differential expression analysis of RNA-seq data. *Genome Biol.*, **11**, R25.
37. Benjamini, Y. and Hochberg, Y. (1995) Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. R. Stat. Soc. B*, **57**, 289–300.
38. Jiang, H. and Wong, W.H. (2008) SeqMap: mapping massive amount of oligonucleotides to the genome. *Bioinformatics*, **24**, 2395–2396.
39. Hartz, D., McPheeters, D.S., Traut, R. and Gold, L. (1988) Extension inhibition analysis of translation initiation complexes. *Methods Enzymol.*, **164**, 419–425.
40. Gu, Z., Eils, R. and Schlesner, M. (2016) Complex heatmaps reveal patterns and correlations in multidimensional genomic data. *Bioinformatics*, **32**, 2847–2849.
41. Lorenz, R., Bernhart, S.H., Höner Zu Siederdisen, C., Tafer, H., Flamm, C., Stadler, P.F. and Hofacker, I.L. (2011) ViennaRNA package 2.0. *Algorithms Mol Biol*, **6**, 26.
42. Mistry, J., Finn, R.D., Eddy, S.R., Bateman, A. and Punta, M. (2013) Challenges in homology search: HMMER3 and convergent evolution of coiled-coil regions. *Nucleic Acids Res.*, **41**, e121.
43. Urban, J.H. and Vogel, J. (2007) Translational control and target recognition by *Escherichia coli* small RNAs in vivo. *Nucleic Acids Res.*, **35**, 1018–1037.
44. Kim, D., Hong, J.S., Qiu, Y., Nagarajan, H., Seo, J.H., Cho, B.K., Tsai, S.F. and Palsson, B. (2012) Comparative analysis of regulatory elements between *Escherichia coli* and *Klebsiella pneumoniae* by genome-wide transcription start site profiling. *PLoS Genet.*, **8**, e1002867.
45. Korkmaz, G., Holm, M., Wiens, T. and Sanyal, S. (2014) Comprehensive analysis of stop codon usage in bacteria and its correlation with release factor abundance. *J. Biol. Chem.*, **289**, 30334–30342.

46. Hall, M.N., Gabay, J., Débarbouillé, M. and Schwartz, M. (1982) A role for mRNA secondary structure in the control of translation initiation. *Nature*, **295**, 616–618.
47. de Smit, M.H. and van Duin, J. (1990) Secondary structure of the ribosome binding site determines translational efficiency: a quantitative analysis. *Proc. Natl. Acad. Sci. U.S.A.*, **87**, 7668–7672.
48. Saito, K., Green, R. and Buskirk, A.R. (2020) Translational initiation in *E. coli* occurs at the correct sites genome-wide in the absence of mRNA-rRNA base-pairing. *Elife*, **9**, e55002.
49. Matera, G., Altuvia, Y., Gerovac, M., El Mouali, Y., Margalit, H. and Vogel, J. (2022) Global RNA interactome of salmonella discovers a 5' UTR sponge for the MicF small RNA that connects membrane permeability to transport capacity. *Mol. Cell*, **82**, 629–644.
50. Klepsch, M.M., Kovermann, M., Löw, C., Balbach, J., Permentier, H.P., Fusetti, F., de Gier, J.W., Slotboom, D.J. and Berntsson, R.P. (2011) Escherichia coli peptide binding protein OppA has a preference for positively charged peptides. *J. Mol. Biol.*, **414**, 75–85.
51. Dever, T.E., Ivanov, I.P. and Sachs, M.S. (2020) Conserved upstream open reading frame nascent peptides that control translation. *Annu. Rev. Genet.*, **54**, 237–264.
52. Blanco, C. (1987) Transcriptional and translational signals of the uidA gene in *Escherichia coli* K12. *Mol. Gen. Genet.*, **208**, 490–498.
53. Bienvenut, W.V., Giglione, C. and Meinzel, T. (2015) Proteome-wide analysis of the amino terminal status of escherichia coli proteins at the steady-state and upon demethylation inhibition. *Proteomics*, **15**, 2503–2518.
54. Wing, C., Errey, J.C., Mukhopadhyay, B., Blanchard, J.S. and Field, R.A. (2006) Expression and initial characterization of WbbI, a putative D-Galf:alpha-D-Glc beta-1,6-galactofuranosyltransferase from escherichia coli K-12. *Org. Biomol. Chem.*, **4**, 3945–3950.
55. Mistry, J., Chuguransky, S., Williams, L., Qureshi, M., Salazar, G.A., Sonnhammer, E.L.L., Tosatto, S.C.E., Paladin, L., Raj, S., Richardson, L.J. et al. (2021) Pfam: the protein families database in 2021. *Nucleic Acids Res.*, **49**, D412–D419.
56. Potter, S.C., Luciani, A., Eddy, S.R., Park, Y., Lopez, R. and Finn, R.D. (2018) HMMER web server: 2018 update. *Nucleic Acids Res.*, **46**, W200–W204.
57. Dryselius, R., Aswasti, S.K., Rajarao, G.K., Nielsen, P.E. and Good, L. (2003) The translation start codon region is sensitive to antisense PNA inhibition in *Escherichia coli*. *Oligonucleotides*, **13**, 427–433.
58. Good, L., Awasthi, S.K., Dryselius, R., Larsson, O. and Nielsen, P.E. (2001) Bactericidal antisense effects of peptide-PNA conjugates. *Nat. Biotechnol.*, **19**, 360–364.
59. Yavari, N., Goltermann, L. and Nielsen, P.E. (2021) Uptake, stability, and activity of antisense Anti-acpP PNA-Peptide conjugates in escherichia coli and the role of sbmA. *ACS Chem. Biol.*, **16**, 471–479.
60. Mologni, L., leCoutre, P., Nielsen, P.E. and Gambacorti-Passerini, C. (1998) Additive antisense effects of different PNAs on the in vitro translation of the PML/RARalpha gene. *Nucleic Acids Res.*, **26**, 1934–1938.
61. Conway, T., Creecy, J.P., Maddox, S.M., Grissom, J.E., Conkle, T.L., Shadid, T.M., Teramoto, J., San Miguel, P., Shimada, T., Ishihama, A. et al. (2014) Unprecedented high-resolution view of bacterial operon architecture revealed by RNA sequencing. *Mbio*, **5**, e01442-14.
62. Bouvier, M., Sharma, C.M., Mika, F., Nierhaus, K.H. and Vogel, J. (2008) Small RNA binding to 5' mRNA coding region inhibits translational initiation. *Mol. Cell*, **32**, 827–837.
63. Pfeiffer, V., Papenfort, K., Lucchini, S., Hinton, J.C. and Vogel, J. (2009) Coding sequence targeting by MicC RNA reveals bacterial mRNA silencing downstream of translational initiation. *Nat. Struct. Mol. Biol.*, **16**, 840–846.
64. Goltermann, L., Yavari, N., Zhang, M., Ghosal, A. and Nielsen, P.E. (2019) PNA length restriction of antibacterial activity of peptide-PNA conjugates in escherichia coli through effects of the inner membrane. *Front. Microbiol.*, **10**, 1032.
65. Sharma, C.M., Darfeuille, F., Plantinga, T.H. and Vogel, J. (2007) A small RNA regulates multiple ABC transporter mRNAs by targeting C/A-rich elements inside and upstream of ribosome-binding sites. *Genes Dev.*, **21**, 2804–2817.
66. Urban, J.H. and Vogel, J. (2008) Two seemingly homologous noncoding RNAs act hierarchically to activate glmS mRNA translation. *PLoS Biol.*, **6**, e64.
67. Mann, M., Wright, P.R. and Backofen, R. (2017) IntaRNA 2.0: enhanced and customizable prediction of RNA-RNA interactions. *Nucleic Acids Res.*, **45**, W435–W439.
68. Wright, B.W., Molloy, M.P. and Jaschke, P.R. (2022) Overlapping genes in natural and engineered genomes. *Nat. Rev. Genet.*, **23**, 154–168.
69. Chemla, Y., Peeri, M., Heltberg, M.L., Eichler, J., Jensen, M.H., Tuller, T. and Alfanta, L. (2020) A possible universal role for mRNA secondary structure in bacterial translation revealed using a synthetic operon. *Nat. Commun.*, **11**, 4827.
70. Lim, H.N., Lee, Y. and Hussein, R. (2011) Fundamental relationship between operon organization and gene expression. *Proc. Natl. Acad. Sci. U.S.A.*, **108**, 10626–10631.
71. Chong, S. (2014) Overview of cell-free protein synthesis: historic landmarks, commercial systems, and expanding applications. *Curr. Protoc. Mol. Biol.*, **108**, 16.30.11–16.30.11.
72. Barkowsky, G., Lemster, A.L., Pappesch, R., Jacob, A., Krüger, S., Schröder, A., Kreikemeyer, B. and Patenge, N. (2019) Influence of different cell-penetrating peptides on the antimicrobial efficiency of PNAs in streptococcus pyogenes. *Mol. Ther. Nucleic Acids*, **18**, 444–454.
73. Geller, B.L., Marshall-Batty, K., Schnell, F.J., McKnight, M.M., Iversen, P.L. and Greenberg, D.E. (2013) Gene-silencing antisense oligomers inhibit acinetobacter growth in vitro and in vivo. *J. Infect. Dis.*, **208**, 1553–1560.
74. Greenberg, D.E., Marshall-Batty, K.R., Brinster, L.R., Zarembek, K.A., Shaw, P.A., Mellbye, B.L., Iversen, P.L., Holland, S.M. and Geller, B.L. (2010) Antisense phosphorodiamidate morpholino oligomers targeted to an essential gene inhibit *Burkholderia cepacia* complex. *J. Infect. Dis.*, **201**, 1822–1830.
75. Howard, J.J., Sturge, C.R., Moustafa, D.A., Daly, S.M., Marshall-Batty, K.R., Felder, C.F., Zamora, D., Yabe-Gill, M., Labandeira-Rey, M., Bailey, S.M. et al. (2017) Inhibition of pseudomonas aeruginosa by peptide-conjugated phosphorodiamidate morpholino oligomers. *Antimicrob. Agents Chemother.*, **61**, e01938–e01916.
76. Korandla, D.R., Wozniak, J.M., Campeau, A., Gonzalez, D.J. and Wright, E.S. (2020) AssessORF: combining evolutionary conservation and proteomics to assess prokaryotic gene predictions. *Bioinformatics*, **36**, 1022–1029.
77. Nakagawa, S., Niimura, Y. and Gojobori, T. (2017) Comparative genomic analysis of translation initiation mechanisms for genes lacking the shine-dalgarno sequence in prokaryotes. *Nucleic Acids Res.*, **45**, 3922–3931.
78. Seip, B., Sacheau, G., Dupuy, D. and Innis, C.A. (2018) Ribosomal stalling landscapes revealed by high-throughput inverse toeprinting of mRNA libraries. *Life Sci. Alliance*, **1**, e201800148.
79. Jung, J., Popella, L., Do, P.T., Pfau, P., Vogel, J. and Barquist, L. (2022) Predicting off-target effects of antisense oligomers targeting bacterial mRNAs with the MASON webserver. bioRxiv doi: <https://doi.org/10.1101/2022.05.24.492283>, 25 May 2022, preprint: not peer reviewed.
80. Ohashi, H., Kanamori, T., Shimizu, Y. and Ueda, T. (2010) A highly controllable reconstituted cell-free system—a breakthrough in protein synthesis research. *Curr. Pharm. Biotechnol.*, **11**, 267–271.
81. Doerfel, L.K., Wohlgemuth, I., Kothe, C., Peske, F., Urlaub, H. and Rodnina, M.V. (2013) EF-P is essential for rapid synthesis of proteins containing consecutive proline residues. *Science*, **339**, 85–88.
82. Ude, S., Lassak, J., Starosta, A.L., Kraxenberger, T., Wilson, D.N. and Jung, K. (2013) Translation elongation factor EF-P alleviates ribosome stalling at polyproline stretches. *Science*, **339**, 82–85.
83. Chadani, Y., Niwa, T., Izumi, T., Sugata, N., Nagao, A., Suzuki, T., Chiba, S., Ito, K. and Taguchi, H. (2017) Intrinsic ribosome destabilization underlies translation and provides an organism with a strategy of environmental sensing. *Mol. Cell*, **68**, 528–539.
84. Delvillani, F., Papiani, G., Dehò, G. and Briani, F. (2011) S1 ribosomal protein and the interplay between translation and mRNA decay. *Nucleic Acids Res.*, **39**, 7702–7715.
85. Updegrove, T.B., Zhang, A. and Storz, G. (2016) Hfq: the flexible RNA matchmaker. *Curr. Opin. Microbiol.*, **30**, 133–138.
86. Moore, S.J., MacDonald, J.T., Wienecke, S., Ishwarbhai, A., Tsipa, A., Aw, R., Kyllis, N., Bell, D.J., McClymont, D.W., Jensen, K. et al. (2018) Rapid acquisition and model-based analysis of cell-free transcription-translation reactions from nonmodel bacteria. *Proc. Natl. Acad. Sci. U.S.A.*, **115**, E4340–E4349.
87. Edgar, R., Domrachev, M. and Lash, A.E. (2002) Gene expression omnibus: NCBI gene expression and hybridization array data repository. *Nucleic Acids Res.*, **30**, 207–210.