



Published in final edited form as:

Mol Cell. 2017 July 06; 67(1): 30–43.e6. doi:10.1016/j.molcel.2017.05.025.

Natural RNA polymerase aptamers regulate transcription in *E. coli*

Nadezda Sedlyarova¹, Philipp Rescheneder², Andrés Magán¹, Niko Popitsch³, Natascha Rziha¹, Ivana Bilusic¹, Vitaly Epshtein⁷, Bob Zimmermann⁴, Meghan Lybecker⁵, Vitaly Sedlyarov⁶, Renée Schroeder¹, and Evgeny Nudler^{7,8,9}

¹Department of Biochemistry and Cell biology; Max F. Perutz Laboratories, University of Vienna. Dr. Bohrgasse 9/5, 1030 Vienna, Austria

²Center for Integrative Bioinformatics Vienna, Max F. Perutz Laboratories, University of Vienna & Medical University of Vienna, Vienna, Austria

³Children's Cancer Research Institute, St. Anna Kinderkrebsforschung, Vienna 1090, Austria

⁴Department of Molecular Evolution and Development, University of Vienna. Althanstrasse 14; 1090 Vienna, Austria

⁵University of Colorado, 1420 Austin Bluffs Pkwy, Colorado Springs, Colorado 80918, USA

⁶CeMM Research Center for Molecular Medicine of the Austrian Academy of Sciences, 1090 Vienna, Austria

⁷Department of Biochemistry and Molecular Pharmacology, New York University School of Medicine, New York, New York 10016, USA

⁸Howard Hughes Medical Institute, New York University School of Medicine, New York, New York 10016, USA

Summary

In search for RNA signals that modulate transcription via direct interaction with RNA polymerase (RNAP) we deep-sequenced an *E. coli* genomic library enriched for RNAP-binding RNAs. Many natural RNAP-binding aptamers, termed RAPs, were mapped to the genome. Over 60% of *E. coli* genes carry RAPs in their mRNA. Combining in vitro and in vivo approaches we characterized a subset of RAPs (iRAPs) that promote Rho-dependent transcription termination. A representative iRAP within the coding region of the essential gene, *nadD*, greatly reduces its transcriptional output in stationary phase and under oxidative stress, demonstrating that iRAPs control gene expression in response to changing environment. The mechanism of iRAPs involves active

Correspondence to: Renée Schroeder; Evgeny Nudler.

⁹Lead contact

Author Contribution: R.S. and E.N. conceptualized and supervised the study. R.S., N.S., I.B. and E.N. designed the experiments. N.S., A.M., N.R., I.B. and V.E. performed the experimental work. N.S., P.R., N.P., A.M., I.B. and M.L. discussed the results and commented on the manuscript. P.R., N.P., B.Z. and V.S. performed the bioinformatics analysis. N.S., R.S. and E.N. wrote the paper.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

uncoupling of transcription and translation, making nascent RNA accessible to Rho. iRAPs encoded in the antisense strand also promote gene expression by reducing transcriptional interference. In essence, our work uncovers a broad class of *cis*-acting RNA signals that globally control bacterial transcription.

Introduction

In bacteria all steps of transcription, from initiation to elongation and termination, are prone to regulation by *trans*-acting proteins and small RNAs (Belogurov and Artsimovitch, 2015; Browning and Busby, 2016; Decker and Hinton, 2013; Sedlyarova et al., 2016). Additionally, the nascent RNA itself can contain *cis*-acting regulatory signals. Riboswitches are *cis*-acting RNA elements mainly encoded in the 5' UTR of the nascent RNA that selectively bind small metabolites and ions, or respond to changes in temperature or pH, leading to regulation of transcription, translation, or RNA processing (Breaker, 2012; Mellin and Cossart, 2015; Serganov and Nudler, 2013). Transcription terminators are another example of bacterial regulatory RNA signals. They can be categorized into two major classes: intrinsic and factor-dependent (Nudler and Gottesman, 2002; Peters et al., 2011). GC-rich RNA hairpins followed by a stretch of uridines are intrinsic signals that cause termination independently of cofactors. Termination at factor-dependent signals requires the presence of protein factor Rho (Roberts, 1969), recognizing particular RNA regions, known as Rho utilization sites (*ruts*) (Hart and Roberts, 1991; Richardson, 2003). Nascent RNA can also increase the processivity of RNAP providing factor-dependent or -independent antitermination signals. For example, phage HK022 *put* RNA binds directly to *E. coli* RNAP throughout elongation rendering it resistant to termination (Komissarova et al., 2008; Sen et al., 2001). Such factor-independent antitermination has been unique to bacteriophages with one notable exception of the EAR enhancing transcript in *Bacillus subtilis* (Irnov and Winkler, 2010). *Cis*-acting sequences, boxA and boxB, is a classical example of a factor-dependent RNA signal that is responsible for processive antitermination in *E. coli* and coliphages (Gusarov and Nudler, 2001; Santangelo and Artsimovitch, 2011; Weisberg and Gottesman, 1999).

Considering the extensive molecular surface of RNAP exposed to solution and the proven ability of nascent RNA to interact with it to allosterically control transcription (Komissarova et al., 2008; Irnov and Winkler, 2010), it is surprising that the only two aforementioned examples are known that exhibit this mode of transcriptional regulation. To systematically search for bacterial RNA signals that regulate transcription through direct interaction with RNAP, we performed the evolution of ligands by exponential enrichment, SELEX (Tuerk and Gold, 1990), using a genomic library of *E. coli* as a source of RNA sequences and the *E. coli* RNAP holoenzyme as bait (Windbichler et al., 2008). Genomic SELEX generates short natural RNA aptamers representing functional domains within a cognate RNA that specify its binding affinity and potential regulatory properties (Ellington and Szostak, 1990; Lorenz et al., 2006; Singer et al., 1997). The RNAP-specific aptamer pool was subjected to next-generation sequencing, followed by in vivo and in vitro functional analysis. As a result, we established RNA polymerase-binding RNA aptamers (RAPs) as abundant natural signals on the nascent RNA that globally control transcription elongation acting *in cis*. We characterize

a broad class of pro-termination (inhibitory) RAPs, termed iRAPs, which actively uncouple transcription from translation, thereby allowing Rho to prematurely terminate transcription within the coding regions. We also characterized iRAPs encoded in the strand opposite to the annotated genes. Such *cis*-acting “antisense” iRAPs positively regulate gene expression by curbing transcriptional interference.

Results

Abundant natural RNA polymerase-binding aptamers across the *E. coli* genome

We examined the sequence space of RNA domains encoded within the *E. coli* genome that bind to bacterial RNAP with high affinity. The enriched sequences obtained under the most stringent selection conditions of genomic SELEX (see Methods and Resources; Windbichler et al., 2008) were deep-sequenced and ~1.0 million reads were mapped to the *E. coli* K12 MG1655 genome followed by peak calling using a custom algorithm (see Supplemental Information). Overall, we identified about 15,000 RAPs - RNA aptamers with high affinity to RNAP - in the *E. coli* genome (Table S1). The K_d of the total selected pool was shown to be below 10 nM (Windbichler et al., 2008). To classify the discovered elements, we developed a categorization system according to the genomic location of RAPs with respect to the currently annotated *E. coli* genes (Karp et al., 2014) (Figure 1A). The majority of identified RAPs (64.3%), termed antisense RAPs, map on the strand opposite to the annotated genes. Approximately 1/3 of RAPs (31.5%) is encoded within the mRNAs, including a subset of intragenic RAPs. UTR-associated and intergenic RAPs together account for 4% of all RAPs (Figure 1B). The positive control, a fragment of 6S noncoding RNA known to bind the RNAP holoenzyme (Cavanagh and Wassarman, 2014; Wassarman and Storz, 2000), was detected among the identified RAPs (RAP-10012, Table S1), validating the strategy for the genome-wide search of the RNAP-binding RNA domains. Interestingly, RAPs are not equally distributed among genes (Figure 1C). About 60% of all *E. coli* genes carry RAPs within their mRNAs (Figure 1C, left panel). We further analyzed RAP distribution within ORFs. RAPs are significantly overrepresented in the proximity of translation start sites and underrepresented in intergenic regions (Figure 1D). We also observed RAPs enrichment at the 3'-proximal part of ORFs on the opposite strand (antisense RAPs) (Figure 1D). We used different tools (MEME and Homer) to search for common sequence motifs within RAPs and could only identify a common CA-rich motif in 12% of the peaks (Figure S1A), suggesting that different RAPs interact with different parts of the extensive surface of RNAP. From these results we conclude that RNAP can bind a large variety of endogenous RNA domains.

We further looked at the conservation of RAP sequences among the representatives of related Enterobacteriaceae, including *Shigella*, *Citrobacter*, *Enterobacter*, *Klebsiella*, *Salmonella* and *Yersinia*. Remarkably, despite of the abundance and variety of encoded RAPs, we found these sequences to be significantly more conserved when compared to random genome fragments of the same size (Figure 1E; see also Method Details for the procedure description). Taken together, our data suggest that bacterial genomes encode numerous RAPs with a potential to affect RNAP functioning during elongation and termination.

A large subset of RAPs induces premature transcription termination

To elucidate a possible effect of RAPs on transcription, we tested the activity of several sense and antisense RAPs in transcriptional fusions upstream of the different reporter genes (Figure 2). We first evaluated the effect of RAPs on gene expression by monitoring changes in the downstream reporter. We identified a number of RAPs that strongly reduced the reporter output: for example, the antisense RAP-1086 and intragenic RAP-15 caused major reduction in GFP fluorescence (Figure 2A). This inhibitory effect was further confirmed by quantitative PCR (qRT-PCR; Figure 2 B-C). To ensure that the observed effect was not promoter- or context-dependent we tested these inhibitory RAPs in different reporter systems under control of different bacterial promoters (Figure 2 A-C).

To estimate the overall fraction of RAPs with transcription inhibitory properties we randomly chose 85 RAPs and found 24 (~28%) of them to exhibit this activity. A minimal decrease of 25% in gene expression in both exponential and stationary growth phases was taken as the primary criterion for defining a RAP as inhibitory (iRAP; Figure S2). Table 1 gives an overview of these iRAPs indicating their genetic location, category, sequence and the theoretical coverage (number of mapped bases divided by corresponding peak length) from deep sequencing. We found that ~40% of iRAPs are intragenic, representing RNA domains within protein-coding genes. Upon extrapolation, the obtained numbers suggest that about one-fourth of ~5000 sense RAPs might have transcription inhibitory activity.

To test the specificity of the observed RAP effect for bacterial RNAP, we cloned iRAPs in a similar reporter system with the phage T7 promoter. *E. coli* BL21(DE3)-derived strain was used to ensure transcription by IPTG-inducible phage T7 RNAP. In this case, we did not observe any significant reduction of transcript levels in response to RAPs (Figure 2C). Therefore, the inhibitory effect of RAPs is specific for bacterial RNAP and was not due to reduced RNA stability.

To gain further insight into the mechanism leading to transcription inhibition by iRAPs, we performed Northern blot analysis of RAP-containing transcripts. Total RNA from exponentially growing cells was probed with oligonucleotides annealing before and after the RAPs (probe location is indicated in Figure 2B). Hybridization with the downstream probe P2 (Figure 2D) confirmed our qRT-PCR data: virtually no RNA was detected for the RAP-15 and -1086. However, hybridization with the upstream probe P1 revealed the accumulation of stable truncated RNA products (Figure 2D). As judged by size, these transcripts were terminated approximately 30 nucleotides downstream of the inserted RAPs. Smearing of the corresponding bands also suggested the absence of a precise termination point. These data were complemented with 3'RACE experiments mapping the exact termination sites (Figure S3 A-B): the prevailing transcripts contained the complete RAP sequence and had the 3' ends mapped 27 to 38 nucleotides downstream of the RAP. To confirm that the inhibitory effect was due to the iRAP sequence, we introduced several point mutations into RAP-15, which abolished its inhibitory effect on transcription (Figure S3 C-E).

To estimate how many of the identified RAPs could cause transcriptional inhibition *in vivo*, we mapped natural 3'ends genome-wide using high-throughput sequencing and established

their proximity to RAPs. In total, we identified more than 20,000 stable 3'ends of *E. coli* transcripts in exponential phase (see Supplemental Information). 1697 RAPs (10.8%) were found to have at least 1 stable 3'end within their sequence extended 75 nt downstream (Table S2). From these termination-associated RAPs, 1511 are sense, accounting for ~27% of all discovered intragenic, intergenic, and UTR-associated RAPs (see examples in Figure S7 A-B). Notably, there are also 186 antisense RAPs associated with stable 3' ends, indicating that they were transcribed and promoted transcription termination *in vivo* in the exponential phase (see an example in Figure S7B).

Taken together, these results indicate that RAP-mediated transcription termination is a widespread phenomenon across the genome.

iRAP-mediated transcription termination is executed by Rho *in vivo* and *in vitro*

To elucidate the mechanism of iRAP-triggered transcription termination, we first searched for possible terminator hairpins formed between the RAP and downstream sequences in the reporter construct. As we did not find any potential intrinsic terminators, we hypothesized that iRAP-mediated termination is Rho-dependent. To test this, we utilized bicyclomycin (BCM) – a highly specific antibiotic that inhibits Rho (Zwiefka et al., 1993). Upon addition of BCM (25 µg/ml) to cells transformed with the RAP constructs described above, transcription inhibition by RAP-1086 and RAP-15 was abolished (Figure 3A-B). Transcription through RAP-2667 and RAP-7768, which had little inhibitory effect, was only mildly affected by BCM. We further observed strong response upon BCM addition for all other identified iRAPs (Figure S4). This suggests that RAP-mediated premature termination of transcription is Rho-dependent. These data were also complemented by 3'RACE experiments in the presence of BCM showing the accumulation of full-length RAP-containing transcripts (Figure S3 A-B).

To directly confirm Rho dependence of the RAP effects we performed single round *in vitro* transcription assays using synthetic DNA templates with selected RAPs and the strong T7A1 promoter (Figure 3A,C). Without Rho, the *in vitro* transcription with *E. coli* RNAP yielded full-length runoff transcripts (Figure 3C). The addition of Rho resulted in the formation of shorter RAP-containing transcripts (Figure 3C, lanes 2, 5 vs. lanes 8, 11). The termination products were further enhanced in the presence of NusG, a cofactor known to increase the efficiency of Rho-dependent termination (Figure 3C, lane 3, 6). *In vitro* transcription experiments showed that the addition of Rho factor to the reaction with the mutated RAP-15 (15 mut) template did not lead to Rho-dependent termination (Figure S3C-E). These data demonstrate that the inhibitory effect of iRAPs is due to Rho-dependent termination.

iRAP-15 from its native location within *nadD* controls transcription in response to stress

To address the physiological role of iRAP-mediated transcription termination we studied the activity of sense iRAP-15 in its natural genomic context. In *E. coli*, iRAP-15 is encoded within the ORF of the nicotinate-mononucleotide adenylyltransferase (*nadD*) gene. NadD is an essential enzyme required for *de novo* biosynthesis and salvage pathways of redox cofactors NAD⁺ and NADP⁺ (Mehl et al., 2000; Zhang et al., 2002). We hypothesized that iRAP-15, conserved among different Enterobacteriaceae (Figure S5 A-C), triggers

transcription termination within *nadD*. In this case, the steady state level of the *nadD* mRNA upstream of iRAP-15 sequence should be higher than just downstream of iRAP-15. We designed iRAP-15-flanking qRT-PCR primers to amplify the respective regions of the *nadD* mRNA and compared the amount of mRNA up- and downstream of iRAP-15 (Figure 4A). Total RNA was isolated from *E. coli* grown under several biologically relevant conditions and subjected to qRT-PCR. As shown in Figure 4B, during logarithmic phase, the RNA amount just downstream of RAP-15 was reduced to 60% compared to the amount upstream. The downstream RNA levels further decreased when cells reached the stationary phase (34%). Challenging cells from mid-exponential phase with 4 mM H₂O₂ for 20' resulted in further reduction of the downstream RNA (21%). Notably, after a brief exposure of exponentially growing cells to BCM (75 µg/ml), the premature termination was abolished, resulting in almost identical mRNA levels upstream and downstream of RAP-15. These experiments demonstrate that within its natural genomic context, RAP-15 inhibits the expression of its host *nadD* gene in response to stress via Rho-dependent transcription termination.

iRAPs enable Rho to act within the translated regions by promoting ribosome stalling

Strikingly, iRAP-15-mediated Rho termination occurs within the coding region of *nadD*. Because transcription and translation are coupled in bacteria (Proshkin et al., 2010), the leading ribosome normally follows RNAP closely to prevent Rho-dependent termination (Richardson, 1991). To verify that efficient translation does indeed take place during iRAP-15-mediated termination, we constructed *nad*(RAP-15)-GFP translational fusion, placing iRAP-15 in its natural genomic context in frame with the reporter gene (Figure 5A). A similar construct, *nad*(rev-15)-GFP, with a reverse complement sequence of RAP-15 was used as a control. The results of GFP plate assays confirmed that RAP-15 promotes transcription termination in spite of concomitant translation: fluorescence intensity of the RAP-15 construct was greatly reduced in comparison to control (Figure 5A-E). We also compared the fluorescence intensity between the two constructs during different stages of growth in liquid culture. While no difference in growth rates was observed (Figure 5C, left panel), substantial reduction in GFP intensity was detected in cells carrying the RAP-15 construct, starting from early exponential and becoming progressively more pronounced upon transition to the stationary phase (Figure 5C, right panel). Thus, we conclude that RAP-15 enables Rho to act within the coding region regardless of active translation.

To further correlate the ribosome occupancy with iRAP-mediated termination we compromised the strength of the corresponding RBS (Figure 5F). Whereas GFP production in the control *nad*(rev-15)-GFP reporter decreased only by ~1.7 fold, the GFP signal almost completely disappeared in the iRAP15-containing constructs (Figure 5F, see the brown bar for the control strain with no GFP expression). Thus, the ribosome occupancy is an important parameter in iRAP-mediated termination within the translated region.

A potential mechanism by which iRAP-15 enables Rho-mediated termination within the coding region is pausing of the ribosome, thereby temporarily uncoupling transcription from translation. Indeed, upon binding to the surface of RNAP, RAP-15 may create a physical barrier to ribosome progression (Figure 5H). To test this hypothesis, we collated our RAP

mapping results with the *E. coli* ribosomal profiling data (Li et al., 2012). Remarkably, a prominent ribosome pause appears just upstream of the RAP-15 sequence of *nadD* (Figure 4C). Similar co-localization between the sites of ribosomal stalling and iRAPs can be observed for the other identified iRAPs (Figure S6A), as well as all the RAPs encoded within translated regions (Figure 4D), indicating that iRAPs do indeed uncouple transcription from translation by stalling the ribosome to facilitate Rho-dependent termination (see also Figure S6). Our *in vivo* data on *nadD* transcript levels upon stress induction (conditions, known to affect translation progression (Deana, 2005)) support this model.

Active translation-transcription uncoupling distinguishes iRAPs from conventional Rho-utilization sites (*ruts*)

The ability of RAP-15 to trigger Rho termination within its ORF is remarkable and suggests a major operational difference between iRAPs and conventional *rut* signals. To directly compare the activities of iRAP and a *rut* site within the translated region, we replaced the sequence of RAP-15 with the canonical *rut* site (Ciampi, 2006; Krebs et al., 2014) in frame, not changing the surrounding sequence of *nadD* fragment (Figure 5D). In contrast to iRAP-15, the ORF-residing *rut* site failed to affect GFP expression, i.e. to stimulate Rho termination (Figure 5D,E). However, if the same *rut* site or iRAP-15 resided within the 5' UTR they both promoted Rho-mediated termination to a similar extent (Figure 3D,E).

Treatment with the sublethal concentration of erythromycin, known to interfere with aminoacyl translocation, did not decrease GFP expression much further for the RAP-15 construct (Figure 5G). However, such erythromycin-mediated uncoupling of translation from transcription resulted in a strong reduction in GFP expression for the *rut*-containing construct (Figure 5G). Taken together these results demonstrate that in addition to its *rut*-like capacity, iRAP is capable of uncoupling transcription from translation, thereby allowing Rho-mediated termination to occur within the actively translated regions.

Antisense iRAPs diminish transcriptional interference

The majority of RAPs map antisense to protein coding genes suggesting their potential involvement in controlling pervasive transcription via Rho-dependent termination (Peters et al., 2012; Wade and Grainger, 2014). To estimate how many of these “antisense” RAPs were actually transcribed, we correlated the antisense RAP dataset with our recent total transcriptome profile of *E. coli* (Sedlyarova et al., 2016). We identified more than 20% of all the antisense RAPs being expressed under normal growth conditions in rich medium (Figure 6A; Table S3). These antisense RAPs not only originate from the regions of the annotated transcripts (in case of inaccurate annotation of gene borders), but a large fraction of these expressed RAPs represent novel, previously unannotated transcripts (e.g. antisense iRAPs-11051 and -8398, Figure S7B and Figure 6D). We hypothesize that many more antisense RAPs will be expressed in response to stress and diverse growth conditions.

Strikingly, upon inhibiting Rho with BCM, almost 90% of the antisense RAPs were transcribed at relatively high levels, strongly supporting the involvement of antisense RAPs in widespread transcription inhibition on the strand opposite to the annotated genes.

Therefore, we hypothesized that iRAPs on the antisense strand are important for suppressing pervasive transcription (Wade and Grainger, 2014) and/or transcriptional interference control (Burgess, 2014).

To test the latter hypothesis, we designed a GFP-based reporter system to estimate the impact of RAPs on transcriptional interference (Figure 6C). Introduction of the strong convergent promoter within the 5' UTR (on the strand opposite to the reporter gene) causes transcriptional interference: GFP expression was considerably reduced (Figure 6B and C, compare pGFP and pGFP^{TI} values). To see the effect of iRAPs in this context we first placed our exemplary iRAP-15 downstream of the antisense promoter (on the strand opposite to the reporter gene). This effectively reduced transcriptional interference and restored GFP expression (Figure 6B and C). Next we examined the effect of three representative natural antisense RAPs (8398, 3129 and 15592) in the pGFP^{TI} setup. When encoded on the antisense strand all three RAPs efficiently inhibited transcription on the antisense strand *in cis*, thereby greatly activating GFP expression by eliminating transcriptional interference (Figure 6F and H, see also Figure S7C-D for endogenous levels). These experiments demonstrate the power of antisense iRAPs in preventing transcriptional interference and reinforce the role of Rho as a general protector against pervasive transcription.

Discussion

Numerous artificial aptamers have been generated by SELEX over the past decades for synthetic biology and therapeutic needs (Berens et al., 2015; Soldevilla et al., 2016). In contrast, natural RNA aptamers, such as riboswitches, have been already utilized by living cells to sense and efficiently respond to their environment (Breaker, 2012; Mellin and Cossart, 2015). When combined with the next-generation sequencing, genomic SELEX serves as a powerful tool for an unbiased identification of natural RNA aptamers independently of the actual levels of transcripts generated under particular conditions (Lorenz et al., 2006). Here we utilized this approach to explore genome-encoded RNA signals with potential to regulate transcription via direct interaction with their most proximal target -RNAP.

Our data demonstrate that chromosomally encoded aptamers specific to RNAP (RAPs) are widespread *cis*-acting RNA signals that affect transcription in bacteria in a controllable manner. A surprising outcome of our genome-wide approach is the diversity of these natural aptamers. We were unable to pinpoint a unique common consensus motif shared by all RAPs, which can be explained by the large surface and complexity of bacterial RNAP. This sequence diversity further suggests different types of RAP activities yet to be discovered under different growth and stress conditions and by utilizing reporter systems other than those used in the present study. Interestingly, despite their variety, RAP-coding sequences are significantly more conserved between the representatives of Enterobacteriaceae when compared to random genome fragments of the same size (Figure 1E).

Our *in vitro* and *in vivo* studies demonstrate that Rho-dependent transcription termination mediates the activity of inhibitory RAPs (iRAPs) - a subclass of RAPs characterized here in detail. Rho is a ring-shaped hexameric RNA helicase that travels with elongating RNAP

(Epshtein et al., 2010; Nudler and Gottesman, 2002; Skordalakes and Berger, 2003). To terminate transcription Rho binds to relatively unstructured nascent RNA (*rut* sites), categorized as ≈ 80 nt long regions with high cytidine/low guanine content (Boudvillain et al., 2013; Ciampi, 2006; Epshtein et al., 2010; Hart and Roberts, 1991; Peters et al., 2011; Salstrom et al., 1979). For iRAPs the C-content median value is 32.8%, whereas G-content median value is 15.4% (Figure S2D and E), suggesting that they can in principle facilitate Rho loading. However, the median length of the tested iRAP sequences does not exceed 47 nt, which is 1.5-2 - fold shorter than the presumed minimal length for a *rut* (at least 70-80 nucleotides) (Richardson, 2003; Zhu and von Hippel, 1998).

As iRAPs can stimulate Rho termination within UTRs (Figure 3B and S4), they must be either helping Rho loading onto RNA acting similarly to the conventional *rut* sequence and/or by modifying RNAP. The latter possibility is consistent with their ability to promote RNAP pausing (Figure S5D). iRAP's ability to pause RNAP should facilitate Rho-mediated termination due to kinetic coupling (Jin et al., 1992). The former possibility is supported by the fact that RAP-15 binds Rho better than its mutated version (Figure S3F). Therefore, it appears that iRAPs act via a combination of mechanisms to promote Rho-dependent termination. It is also likely that the contribution of each such mechanism varies among iRAPs, which constitute $>20\%$ of total RAPs. The unique regulatory and mechanistical features of iRAPs become apparent when they act within the translated region. Unlike conventional *rut* sites, which can recruit Rho only within UTRs, iRAPs do so within the coding region by directly uncoupling transcription from translation (Figure 5C,E). Therefore, we propose that iRAPs do not act as *bona fide rut* sites, but facilitate Rho loading via positioning the unstructured transcript in close proximity to Rho on the surface of RNAP, and, at the same time, preventing ribosome from interfering with this process (Figure 5H). iRAP-15 has a moderate affinity to Rho (Figure S3F) further facilitating Rho loading onto mRNA within this time window of opportunity. Curiously, the same interaction of iRAP with RNAP appears to cause both transcriptional and translational pausing. Transcriptional pausing must be the result of an allosteric effect of iRAP binding to RNAP after it emerges from the RNA exit channel. Translational pausing is likely to be the result of a topological interference imposed by the same interaction as shown schematically in Figure 5H.

The discovery of abundant iRAPs on the strand opposite to the coding sequence illuminates the mechanism of transcriptional “anti-interference”. We show that antisense iRAPs efficiently curb transcriptional interference by *cis*-acting as pro-termination signals, thereby greatly potentiating the expression of sense-encoded genes. The observed enrichment of antisense iRAPs opposite to the 3'-proximal part of annotated genes (Figure 1C) further supports the “guardian” role for iRAPs in preventing sporadic transcription on the antisense strand that can interfere with CDS expression.

The identification of RAPs as a widespread class of transcription-regulating RNA signals reveals that direct interaction of the nascent RNA with RNAP broadly impacts on the composition of the *E. coli* transcriptome. iRAP-mediated transcription termination represents a hitherto unknown type of cross-talk between the nascent RNA and the transcription machinery, in which RNA realizes its potential as a self-regulatory system.

Methods and Resources

Contact For Reagent and Resource Sharing

Further information and requests for reagents may be directed to, and will be fulfilled by the lead contact Evgeny Nudler (evgeny.nudler@nyumc.org).

Experimental Model and Subject Details

Bacterial Strains, Plasmids and Growth Conditions—All *E. coli* strains and main plasmids used in this study are listed in Key Resource Table. Tested arabinose-induced construct pBAD13-lacZ was derived from pBAD24 plasmid by first removing original RBS from MCS by NheI/EcoRI restriction, followed by DNA Polymerase I Large (Klenow) Fragment treatment and blunt ligation. Then new RBS-containing insert (+70 nt) was cloned between XbaI/PstI sites, followed by inserting *lacZ* sequence between PstI/HindIII sites. RAPs were tested in a series of pB13_RAP_lacZ constructs, derived from pBAD13_lacZ by introducing RAP sequence in the 5'UTR, between EcoRI and NcoI restriction sites. Additional plasmids - pB13_RAP_+99_lacZ and pB13_+49RAP_+99_lacZ – were used to ensure that the observed effect is not dependent on RAP 5'/3'end-flanking sequences.

Bacterial vector pET21_minRBS was derived from pET21a starting with removing the original MCS by XbaI/NdeI restriction and followed by DNA Polymerase I Large (Klenow) Fragment treatment and blunt ligation. RAP effect when transcribed by phage T7 polymerase was tested in pET21_RAP_lacZ series of constructs, derived from pET21_minRBS by cloning RAP_+99_lacZ insert (amplified from pB13_RAP_+99_lacZ, full length, +5 nt from TSS) between NheI/HindIII restriction sites.

pRAP#-GFP or pInsert-GFP (pWM3110-RAP or pWM3110-insert) are constructs for testing RAP/insert effects *in cis* in the 5'UTR, upstream of translated GFP. Derived from pWM3110 by cloning corresponding RAPs/ inserts between XhoI and SacI sites.

Transcriptional interference system pGFP^{TI} (pWM3150) harboring a convergent constitutive promoter downstream from the one controlling GFP expression (allowing transcription antisense to the 5'UTR of GFP) was derived from pGFP (pWM3110) construct by cloning 235nt-long insert (with added unique KpnI site) between EcoRI/SacI sites. pRAP#-GFP^{TI} (pWM3150-RAP) represents the series of constructs for testing RAPs in transcriptional interference-like system, on the strand opposite to the one encoding GFP. pRAP#-GFP^{TI} was derived from pGFP^{TI} (pWM3150) by cloning RAPs/inserts downstream of the inserted convergent promoters, antisense to the 5'UTR of reporter GFP.

For cloning, reporter protein assays and maintaining strains: cultures were grown in LB or LB-agar media at 37°C and, when required, supplemented with ampicillin (100 µg/ml; Sigma-Aldrich), kanamycin (30 µg/ml; Sigma-Aldrich), erythromycin (1.5 µg/ml; Sigma-Aldrich) and/or bicyclomycin (BCM; Astellas Corporation) (8-10 µg/ml – low concentration used for extended treatments, minimally affects bacterial growth). When grown in liquid media, cells were cultivated with constant aeration (shaking at 200 rpm). To elucidate RAP effects on transcription by bacterial RNAP, reporter constructs with constitutive bacterial RNAP promoter (pGFP-based vectors) and L-arabinose-inducible systems (pBAD-based

reporters) were used. *E. coli* DH5alpha cells carrying pBAD reporters were supplemented with 0.2% L-arabinose (Sigma-Aldrich) for early induction at $OD_{600} \approx 0.2-0.3$ and harvested later at the exponential phase. To check RAP effects on T7 transcription, *E. coli* BL21 (DE3) Tuner cells carrying described pET reporters were induced with low concentration of IPTG (50 μ M; Sigma-Aldrich) at $OD_{600} \approx 0.4-0.5$ and further cultivated for 15-20 minutes before harvesting.

For measurements of endogenous transcript levels: cells were harvested at $OD_{600} = 0.5-0.6$ (exponential phase) or $OD_{600} > 1.6$ (stationary phase). In experiments with oxidative stress induction, bacterial cells grown to early exponential phase ($OD_{600} \approx 0.4-0.5$) were treated with 4 mM H₂O₂ for 20 or 60 minutes (as indicated) prior to harvesting. In experiments with BCM treatment (followed by RNA extraction), bacterial cells grown to early exponential phase ($OD_{600} \approx 0.4-0.5$) were treated with BCM as indicated (25 / 50 / 75 μ g/mL; final concentrations selected based on previous studies (Dutta et al., 2011; Sedlyarova et al., 2016)) and cultivated for 15-20 minutes before harvesting.

Method Details

SELEX library deep sequencing—The RNA library from SELEX cycle 7 was first reverse transcribed to cDNA and amplified by 16 cycles of PCR with Phusion High-Fidelity DNA polymerase (New England Biolabs) using primers with fixed part and added barcode for multiplexing according to (Parameswaran et al., 2007), (Windbichler et al., 2008) (fixFW: TATAGGGGAATTCGGAGCGGG, fixRV_multi: TAGCCCGGGATCCTCGGGGCTG (Sigma)). Library preparation (adapter ligation step) and deep sequencing were performed at the CSF NGS unit <http://csf.ac.at/>. The SELEX library was multiplexed and sequenced with the Solexa technology on a GAIIX with paired-end 76 base-pair reads. Reads were trimmed using trimomatic (Bolger et al., 2014) and mapped against the reference genomes of the *E. coli* strain K12 (substrain MG1655, GenBank ID: U00096.3) using NextGenMap (Sedlazeck et al., 2013). Read mapping was conducted with permissive settings that allowed up to 20% sequence divergence in order to account for the considerable amounts of sequence alterations introduced by the SELEX procedure (Zimmermann et al., 2010a, 2010b). The mapped reads were then filtered for a minimum mapping quality of 20, which resulted 915,594 paired-end reads. Corresponding ORF annotations were downloaded from EcoGene 3.0 (Zhou and Rudd, 2013) and 5' and 3' UTR annotations - from RegulonDB (Zhou and Rudd, 2013).

Peak finding—Strand-specific coverage signals from K12 alignments were extracted and a custom peak-finding method was applied to these signals in order to identify the originating genomic regions of mRNA-fragments that were found to be enriched in the respective SELEX cycle. Briefly, our peak-calling method consists of the following steps: (1) signal smoothing using a moving Gaussian kernel, (2) approximation of the first derivative at each signal position by cubic Hermite interpolation, (3) calling peaks based on derivative sign-changes and defined minimum/maximum peak dimensions (minimum width: 10bp, maximum width: 500bp, minimum (smoothed) peak height: 8).

Interval annotation and categorization—Peak intervals were classified according to whether they overlap with annotated ORFs or UTRs. Intervals were annotated as being “antisense” when they mapped to the strand opposite to an existing ORF or UTR annotation. Based on the genomic location we classify them into several categories: 5'UTR (corresponding sequence aligns within any annotated 5'UTR), 3'UTR (corresponding sequence aligns within any annotated 3'UTR), intragenic (corresponding sequence aligns within the annotated gene, excluding cases of 5' - and 3'UTR), antisense (corresponding sequence aligns on the strand opposite to the annotated gene) and intergenic (corresponding sequence in between the annotated genes).

Furthermore, for each gene we counted the number of overlapping sense and antisense RAPs (Fig. 1A and B). If a RAP overlapped with more than one gene, it was counted only once for the gene with the biggest overlap. RAPs overlapping with a 5' UTR were assigned to the first gene of the respective operon.

In order to characterize the location of RAPs within a transcript, we calculated their relative positioning within annotated transcripts (Fig. 1C). Our data indicate enrichment of RAPs around the translation start site for both sense and antisense RAPs. Additionally, we observe a slight enrichment of RAPs around the midpoint of the transcripts and in the second ORF half for antisense RAPs.

Motif search—We searched for motifs in the RAP sequences using MEME algorithm (Bailey et al., 2009). Using default settings, we identified a (CAN)_n motif with a length of 29 bp in 1010 of the 15,724 RAPs (6.4%). We found that most (89%) of the RAPs containing this motif are anti-sense to an annotated ORF. Homer (Heinz et al., 2010), another motif discovery tool, confirmed the (CAN)_n motif when run with a motif length of 30 bp in 1909 of the 15,724 RAPs (12.14%) (Fig. S1E).

RiboSeq occupancy profiles—We downloaded two replicates of RiboSeq data (Li et al., 2012) from GEO (sample: GSM872393), converted them to BED, and used liftover procedure to convert the genomic coordinates from U00096.2 to U00096.3. Next we computed the mean of the two replicates for each position of the genome to retrieve a genome wide Riboseq occupancy profile. To visualize the effect of RAPs on ribosome occupancy, we computed local occupancy profiles for all RAPs by cutting out an interval starting 100 bp upstream and 200bp downstream of the RAP start and end site respectively.

Average occupancy profiles—To analyze the effect of RAPs on ribosome occupancy on a genome wide level we computed local occupancy profiles for all RAPs (located in sense) starting twice the RAP length upstream and ending twice the RAP length downstream of the RAP start and stop site respectively. We normalized the length of each local profile as well as the RiboSeq coverage to a minimum of 0 and a maximum 100. Next we inferred an average profile by computing the mean occupancy for each position of all normalized local profiles. To avoid edge effects, all intervals that do not have at least 95 % overlap with a gene were excluded from this analysis. The resulting average profile shows an enrichment of RiboSeq coverage at the start and a depletion at the stop sites of RAPs. To evaluate whether this effect was statistically significant we randomly shuffled all RAPs on the genome

(maintaining the length distribution and the overall number of sense and antisense RAPs) and then computed the average occupancy profile for the set of random RAPs. We repeated this process 10,000 times to get a null model for the average occupancy profile. From this null model we inferred the 0.025 and 97.5 percentile for each position of the profile giving us an upper and lower bound of what to expect from a random occupancy profile. Comparing the average profile from the RAP data to the null model showed that the enrichment and the depletion at the start and stop sites of the RAPs is not expected by chance.

Analysis of antisense RAP expression using RNA deep-sequencing data—We downloaded raw reads for two replicates of RNA-Seq data for wild-type *E. coli* K12 grown to exponential and stationary phases in rich media (SRA: SRP078327 (Sedlyarova et al., 2016)). Reads were mapped using NextGenMap (Sedlazeck et al., 2013), requiring a minimum sequence identify cutoff of 95%, and filtered for a mapping quality of 20. Strand specific read counts were computed for all RAPs using featurecount from the subread package (Liao et al., 2014). To identify the set of expressed RAPs we had to define the threshold. Therefore, we first computed RPKM value for regions that are less likely to be expressed. Since intergenic regions are very small in *E. coli*, we used all regions opposite to annotated genes instead. To reduce the effect of read-through and overlapping genes, we took into account only the middle 50% of the genes. To consider a RAP expressed we required its RPKM value to be higher than 90 % of the RPKM values we got from the non-expressed regions. This corresponded to a RPKM value of approximately 5 (threshold parameter for defining the antisense RAP as an expressed one).

Sequence conservation of *E.coli* RAPs—Genome sequences of several Enterobacteriaceae (*Citrobacter koseri* (NC_009792.1), *Enterobacter aerogenes* (NC_015663.1), *Klebsiella pneumoniae* (NZ_CP006659.2), *Salmonella enterica* (NC_003197.1), *Shigella dysenteriae* (NC_003197.1) and *Yersinia pestis* (NC_003143.1)) were downloaded from NCBI, Genbank (<https://www.ncbi.nlm.nih.gov/genbank/>). Those sequences were used to build local blast nucleotide database (Altschul et al., 1990). For each RAP nblast was executed and its output was processed using perl script.

15,724 fragments with the same length distribution as RAPs were randomly placed on *E. coli* genome. For each of the randomly placed fragments *nblast* search was performed against 6 selected closely related Enterobacteriaceae species. Fragment was considered conservative in a species if blast algorithm identifies a hit in the species genome with E-value less than 0.01 and query coverage not less than 60 %. In order to increase statistical test power, random shuffling was repeated 250 times. The conservation of each particular RAP or random fragment is represented by a number of species in which the element is conserved (integer from 0 to 6).

Generalized linear model on Poisson distribution was used to compare conservation distributions of RAPs versus random fragments. Data preprocessing, analysis and visualization were performed using R-project (<https://www.R-project.org/>).

Protein Reporter Assays—To monitor expression of *RAP-lacZ* fusions, *E. coli* cultures were grown overnight at 37°C in LB medium with constant aeration, and afterwards new cultures were started from 1:100 dilution in fresh medium. Cultures were grown to logarithmic phase ($OD_{600} = 0.5\text{--}0.6$). To determine β -galactosidase activity, corresponding β -galactosidase assays were performed in triplicates as described in (Soper et al., 2010).

GFP plate assay: To monitor GFP fluorescence in transformed *E. coli* with RAP-GFP fusions, cells were plated by streaking on LB-agar plates supplemented with kanamycin (30 mg/ml) and incubated for 16-20 hours at 37°C. Plates were photographed at GFP mode with excitation wavelength 460 nm (Fusion Fx7 Imager, PEQlab, Germany). To compare cell density, plates were imaged at visible light mode with excitation wavelength 510 nm (Fusion Fx7 Imager, PEQlab, Germany). Data processing was performed with ImageQuant TL or ImageJ softwares. *Growth monitoring and GFP assay in liquid culture*: transformed *E. coli* strains were grown in 10 ml LB media (supplemented with kanamycin, 30 mg /ml) up to early stationary phase ($OD_{600} = 1.2\text{--}1.4$), then diluted to $OD_{600} \approx 0.05$ and, and 150 μ l of diluted culture was transferred to a wells of a 96-well black plate (flat clear bottom) with a clear lid (Greiner, LOT 07290155). The plate was incubated at 37°C for >9 h with constant orbital shaking (amplitude – 1 mm) in Tecan Infinite F500. Measurements were performed from above and below every 10 min, with detection at 612 nm (Optical Density, 612 nm) and using a fixed signal gain of 30% with 485 nm and 535 nm detection for measuring GFP intensity. The obtained fluorescence intensity values were plotted either directly (Fluorescence Intensity, Relative Units - RU), or after normalization to the cell density (Fluorescence Intensity, normalized Units – nU). Measurements were performed minimum in three technical replicates for each biological replicate.

RNA Isolation—Total RNA was isolated using the hot phenol method as described in (Lybecker et al., 2014).

qRT-PCR—To measure the levels of transcripts, 0.5-2 μ g total DNA-free RNA was reverse transcribed using random oligo 9-mers (Sigma) and ProtoScript II First Strand cDNA Synthesis Kit (New England Biolabs) according to the manufacturer's protocol. Depending on the target of interest, cDNA was amplified with the corresponding primers (see List of Oligonucleotides Used in This Study). In each case the real-time PCR primer amplification efficiency was estimated using the standard curve method in one color detection system as described before (Pfaffl, 2004). qPCR was performed using Eppendorf Mastercycler® RealPlex2 and 5x HOT FIREPol® EvaGreen® qPCR Mix Plus (no ROX) from Medibena as suggested by the manufacturers. *gapA* gene transcript levels were used for normalization.

Northern Blot Analysis—For Northern blotting (NB) 10 μ g of total RNA was separated by gel electrophoresis using denaturing 8 % polyacrylamide-TBE-Urea (8M) gels in 1 \times TBE. RNA was loaded onto the gel after denaturing at 70°C (for 10 min) in 2 \times RNA load dye (Fermentas) followed by transferring on ice. Gel-separated RNA was transferred to HybondXL membranes (Ambion) by wet electro-blotting at 12 V for 1 hour in 0.5 \times TBE. The membranes were cross-linked by UV (150 mJ/cm) and probed with 5'-end γ^{32} P-labeled DNA oligonucleotide probes (see List of sequences) in ULTRAhyb®-Oligo Hybridization Buffer (Ambion) according to the manufacturer's instructions. The DNA labeling reaction

was performed with T4 PNK (New England Biolabs) according to the manufacturer's instructions.

3'RACE—To precisely determine the 3' end of transcripts, 550 pmol of 5'-phosphorylated RNA adapter (adapter sequence - 5'P -AAUGGACUCGUAUCACACCCGACAA - idT) was ligated to 6 µg of total RNA using T4 RNA Ligase 1 (ssRNA Ligase; New England Biolabs) according to the manufacturer's protocol overnight at 16°C. The reaction was mixed with the equal volume of phenol/chloroform/isoamyl alcohol (25:24:1, pH ca 4.0, AppliChem GmbH) and centrifuged for 5 min at 16,100 × g (4°C) in the Phase Lock Gel Heavy tube (5Prime). Then the aqueous phase was mixed with chloroform/isoamyl alcohol (24:1) followed by 5 min centrifugation at 16,100 × g (4°C). Nucleic acid was ethanol-precipitated from the aqueous phase (3 volumes ethanol, 1/10 volume of 3 M sodium acetate, 1/100 volume 0.5 M EDTA) and subjected to RT using SuperScript™ II Reverse Transcriptase (Life Technologies) followed by RNaseH treatment (Promega) according to the manufacturer's protocols. cDNA was amplified with Phusion High-Fidelity DNA Polymerase (NEB) using the forward primer to the reporter construct upstream RAP (5'-TTGGGCTAGAATTCTGTTGTTA-3') and the reverse primer to the ligated adapter (5'-TTGTCGGGTGTGATACGAGTCCATT-3'). The obtained PCR-amplified library was separated and checked by gel electrophoresis at 6% PAA/1×TBE gel. The obtained bands were gel-purified, TA-cloned using pGEM-T Easy Vector System (Promega, according to the manufacturer's instructions) and subjected to Sanger sequencing. ClustalW2 (see Multiple Sequence Alignment Tools) was used to generate the alignments.

cDNA library preparation for high-throughput 3'ends identification—Directional (strand-specific) 3' end enriched cDNA libraries were constructed as follows: total DNaseI treated RNA was depleted of ribosomal RNA using the Ribo-Zero™ RNA removal kit for Gram-negative bacteria (Epicentre). A 3' RNA adapter, based on the Illumina multiplexing adapter sequence (Oligonucleotide sequences © 2007-2014 Illumina, Inc. All rights reserved) blocked at the 3' end with an inverted dT (5'-GAUCGGAAGAGCACACGUCU[idT]-3'), was phosphorylated at the 5' end using T4 PNK (New England Biolabs) per the manufacturer's protocol. The 3' RNA adapter was ligated to the 3' ends of the rRNA depleted RNA using T4 RNA ligase I (New England Biolabs). 1.5 µg of RNA was incubated at 20°C for 6 hours in 1× T4 RNA ligase reaction buffer with 1 mM ATP, 30 µM 3' RNA adapter, 10 % DMSO, 10 U of T4 RNA ligase I, and 40 U of RNasin (Promega) in a 20 µl reaction. RNA was then fragmented in equivalents of 100 ng using the RNA fragmentation reagents (Ambion®) per the manufacturer's protocol at 70°C for 3 min and subsequently phosphorylated at the 5' ends using T4 PNK (New England Biolabs) per the manufacturer's protocol to allow for ligation of the 5' adapter. RNA was size-selected (≈ 150-300 nt) and purified over a denaturing 8 % polyacrylamide/8 M urea/TBE gel. Gel slices were incubated in RNA elution buffer (10 mM Tris-HCl, pH 7.5, 2 mM EDTA, 0.1 % SDS, 0.3 M NaOAc) with vigorous shaking at 4°C overnight. The supernatant was subsequently ethanol precipitated using glycogen as a carrier molecule. The Illumina small RNA 5' adapter (5'-GUUCAGAGUUCUACAGUCCGACGAUC-3') was ligated to the RNA as described before except the concentration of the adapter was 52 µM and 20 U of T4 RNA ligase I was used in total volume of 25 µl. The ligated RNAs were size-

selected (\approx 200-300 nt) and gel-purified over a denaturing 8 % polyacrylamide/8 M urea/TBE gel (as described above). The di-tagged RNA libraries were reverse-transcribed with SuperScript[®]II reverse transcriptase (Invitrogen) using random nonamers per the manufacturer's protocol. RNA was removed using RNase H (Promega) per the manufacturer's protocol and cDNA was amplified in PCR carried out using Phusion[®] High-Fidelity Polymerase (New England Biolabs). cDNA was amplified with modified designed Illumina-compatible PCR primers (3' library Forward 5' - CAAGCAGAAGACGGCATAACGACAGGTTTCAGAGTTCTACAGTCCGA-3'; Reverse 5' - AATGATACGGCGACCACCGAGATGTGACTGGAGTTCAGACGTGTGCTCTTCCGATC-3') by 18 cycles of PCR. The products were purified using Agencourt AMPure XP beads (Beckman) and analyzed on an Agilent 2100 Bioanalyzer. 3' end enriched cDNA libraries were sequenced on individual Genome Analyzer IIx lanes (36 bp, single-end) or on HiSeq 2000 lanes (50 bp, single-end) using primer based on Illumina Multiplexing Read 2 Sequencing Primer (5' -GTGACTGGAGTTCAGACGTGTGCTCTTCCGATC-3') at the VBCF NGS unit <http://ngs.vbcf.ac.at/>.

Computational analysis for the high-throughput 3'ends identification—The *E.coli* K12 genome was downloaded from Genbank (Benson et al., 2013), accession number NC_000913.3. The two replicate libraries, containing 180,677,983 and 103,186,381 50 bp single-end reads, were mapped to the genome with using bowtie2(Langmead and Salzberg, 2012) using default parameters. Alignments with bowtie2 mapping quality values lower than 40 were not retained for further analysis, leaving 128,413,654 and 76,508,335 reads. In order to distinguish between 3' ends of transient products of RNA metabolism and stable 3' ends, we developed an algorithm to call coverage peaks. The algorithm will be discussed in detail in an upcoming manuscript, and the scripts used are available upon request from the authors. In short, positions in the *E. coli* genome were considered in descending order of coverage and assigned a p-value based on a Poisson distribution parameterized by the mean coverage of all covered bases in the genome. Peaks were rejected if their p-values exceeded $1e-4$ or if there existed a >10 -bp window containing the peak in which all positions were within 2-fold coverage of the peak position. Parameters were chosen based on analysis of annotated 3' ends as well as qualitative analysis of peaks. This resulted in 20,019 peaks.

In vitro transcription—DNA templates containing strong T7A1 bacterial RNAP promoter (see Table S4 “List of oligonucleotide sequences used in this study”) for *in vitro* transcription experiments were synthesized by PCR with Phusion or Q5 High-Fidelity Polymerases (both from New England Biolabs) using reporter plasmids with different RAPs and appropriate synthetic primers (Sigma Aldrich). Transcription reactions were performed in solution using wild-type *E. coli* RNAP. To assemble the transcription initiation complexes, 2 pmol of RNAP were mixed with 1 pmol of DNA template in 20 μ l of Transcription Buffer 100 (10 mM MgCl₂, 40 mM Tris-HCl, pH 7.9, 100 mM NaCl) and incubated for 5 min at 37°C. Then AUC RNA primer (Dharmacon) was added up to 10 μ M together with GTP (25 μ M) and ATP (25 μ M) followed by incubation at 37°C for 5 minutes. Next, [α -32P]-CTP (800 Ci/mmol, 2 μ l, Hartmann Analytic) was added for another 2 min. To minimize RNA degradation 10 U of RNasin[®] Ribonuclease Inhibitor (Promega) were added to the reaction. For the reactions with Rho and NusG, the purified transcription factors were

added in a concentration of up to 0.5 μM to the transcriptional mixtures where indicated, followed by additional incubation at 25°C for 5 minutes. Chasing reaction was performed with all four NTPs added to a final concentration of 10-50 μM . To prevent re-initiation, rifampicin (Sigma) was added concurrently at this step to a final concentration of 10 μM . At the indicated time points, an aliquot of chase reaction was quenched with 3 \times Stop Solution, containing 100 mM EDTA, 8M UREA, 0.025% xylenthanol and 0.025% bromophenol blue. The products were separated by 6-8% denaturing TBE-PAGE (8M urea).

RAP sequence mapping within the transcript was performed by transcribing RAP-containing template and 25 μM of NTPs mixed with one of the four 3' dNTPs (3'dGTP, 3'dATP, 3'dUTP or 3'dCTP) at 3:1 ratio in four different transcription reactions for 10 min.

Quantification and Statistical Analysis

Tools used for statistical analysis at different steps of computational genome-wide data analysis are described in the corresponding chapters above.

Relative RT-qPCR quantification was performed as described in (Pfaffl, 2004). For each primer pair the efficiency of real-time PCR amplification was estimated using the standard curve method in a one color detection system (as described in (Pfaffl, 2004)).

Statistical parameters for each experiment are reported in the corresponding Figure Legends. The reported data were obtained from at least 3 biological replicates. The significance was tested using Student's t-test. In figures, asterisks indicate statistical significance between the compared datasets (specified by lines) as determined with mentioned tests where * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$; ns – non-significant

Statistical analysis was performed using R Statistical Software and Simple Interactive Statistical Analysis Tools (<http://www.quantitativeskills.com/sisa/index.htm>).

Data and Software Availability

Deep-sequencing data sets were uploaded to Gene Expression Omnibus database (GEO; <https://www.ncbi.nlm.nih.gov/geo/>) with corresponding accession number GSE98661.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We thank Arndt von Haeseler for helpful discussions on bioinformatical analysis. We also thank the Vienna BioCenter Core Facilities (VBCF) for NGS. N.S. and P.R. acknowledge support by the RNA-DK Biology (W1207-B09). This work was supported by the NIH grant R01 GM107329, by the Howard Hughes Medical Institute (E.N.), and by the Austrian Science Fund FWF Grants I538-B12, F4301 and F4308 (R.S.).

References

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. *J Mol Biol.* 1990; 215:403–410. [PubMed: 2231712]

- Bailey TL, Boden M, Buske FA, Frith M, Grant CE, Clementi L, Ren J, Li WW, Noble WS. MEME SUITE: tools for motif discovery and searching. *Nucleic Acids Res.* 2009; 37:W202–8. [PubMed: 19458158]
- Belogurov, Ga, Artsimovitch, I. Regulation of Transcript Elongation. *Annu Rev Microbiol.* 2015; 69:150701130309005.
- Benson DA, Cavanaugh M, Clark K, Karsch-Mizrachi I, Lipman DJ, Ostell J, Sayers EW. GenBank. *Nucleic Acids Res.* 2013; 41:D36–42. [PubMed: 23193287]
- Berens C, Groher F, Suess B. RNA aptamers as genetic control devices: The potential of riboswitches as synthetic elements for regulating gene expression. *Biotechnol J.* 2015; 10:246–257. [PubMed: 25676052]
- Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics.* 2014; 30:2114–2120. [PubMed: 24695404]
- Boudvillain M, Figueroa-Bossi N, Bossi L. Terminator still moving forward: expanding roles for Rho factor. *Curr Opin Microbiol.* 2013; 16:118–124. [PubMed: 23347833]
- Breaker RR. Riboswitches and the RNA world. *Cold Spring Harb Perspect Biol.* 2012; 4
- Browning DF, Busby SJW. Local and global regulation of transcription initiation in bacteria. *Nat Rev Microbiol.* 2016; 14:638–650. [PubMed: 27498839]
- Burgess DJ. Transcription: Interference from near and far. *Nat Rev Genet.* 2014; 15:144–144.
- Burmann BM, Schweimer K, Luo X, Wahl MC, Stitt BL, Gottesman ME, Rösch P. A NusE:NusG complex links transcription and translation. *Science.* 2010; 328:501–504. [PubMed: 20413501]
- Cavanagh AT, Wassarman KM. 6S RNA, A Global Regulator of Transcription in *Escherichia coli*, *Bacillus subtilis*, and Beyond. *Annu Rev Microbiol.* 2014
- Ciampi MS. Rho-dependent terminators and transcription termination. *Microbiology.* 2006; 152:2515–2528. [PubMed: 16946247]
- Deana A. Lost in translation: the influence of ribosomes on bacterial mRNA decay. *Genes Dev.* 2005; 19:2526–2533. [PubMed: 16264189]
- Decker KB, Hinton DM. Transcription Regulation at the Core: Similarities Among Bacterial, Archaeal, and Eukaryotic RNA Polymerases. *Annu Rev Microbiol.* 2013; 67:113–139. [PubMed: 23768203]
- Dutta D, Shatalin K, Epshtein V, Gottesman ME, Nudler E. Linking RNA polymerase backtracking to genome instability in *E. coli*. *Cell.* 2011; 146:533–543. [PubMed: 21854980]
- Ellington AD, Szostak JW. In vitro selection of RNA molecules that bind specific ligands. *Nature.* 1990; 346:818–822. [PubMed: 1697402]
- Epshtein V, Dutta D, Wade J, Nudler E. An allosteric mechanism of Rho-dependent transcription termination. *Nature.* 2010; 463:245–249. [PubMed: 20075920]
- Gusarov I, Nudler E. Control of intrinsic transcription termination by N and NusA: the basic mechanisms. *Cell.* 2001; 107:437–449. [PubMed: 11719185]
- Guzman LM, Belin D, Carson MJ, Beckwith J, Guzman L, Belin D, Carson MJ. Tight regulation, modulation, and high-level expression by vectors containing the arabinose PBAD promoter. These include: Tight Regulation, Modulation, and High-Level Expression by Vectors Containing the Arabinose P BAD Promoter. 1995; 177
- Hart CM, Roberts JW. Rho-dependent transcription termination Characterization of the requirement for cytidine in the nascent transcript. *J Biol Chem.* 1991; 266:24140–24148. [PubMed: 1721066]
- Heinz S, Benner C, Spann N, Bertolino E, Lin YC, Laslo P, Cheng JX, Murre C, Singh H, Glass CK. Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Mol Cell.* 2010; 38:576–589. [PubMed: 20513432]
- Imov I, Winkler WC. A regulatory RNA required for antitermination of biofilm and capsular polysaccharide operons in *Bacillales*. *Mol Microbiol.* 2010; 76:559–575. [PubMed: 20374491]
- Jin DJ, Burgess RR, Richardson JP, Gross CA. Termination efficiency at rho-dependent terminators depends on kinetic coupling between RNA polymerase and rho. *Proc Natl Acad Sci U S A.* 1992; 89:1453–1457. [PubMed: 1741399]
- Karp PD, Weaver D, Paley S, Fulcher C, Kubo A, Kothari A, Krummyacker M, Subhraveti P, Weerasinghe D, Gama-Castro S, et al. The EcoCyc Database. *EcoSal Plus.* 2014; 6

- Komissarova N, Velikodvorskaya T, Sen R, King RA, Banik-Maiti S, Weisberg RA. Inhibition of a transcriptional pause by RNA anchoring to RNA polymerase. *Mol Cell*. 2008; 31:683–694. [PubMed: 18775328]
- Krebs, JE., Lewin, B., Kilpatrick, ST., Goldstein, ES. *Lewin's genes XI*. Jones & Bartlett Learning; 2014.
- Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. *Nat Methods*. 2012; 9:357–359. [PubMed: 22388286]
- Li GW, Oh E, Weissman JS. The anti-Shine-Dalgarno sequence drives translational pausing and codon choice in bacteria. *Nature*. 2012; 484:538–541. [PubMed: 22456704]
- Li W, Cowley A, Uludag M, Gur T, McWilliam H, Squizzato S, Park YM, Buso N, Lopez R. The EMBL-EBI bioinformatics web and programmatic tools framework. *Nucleic Acids Res*. 2015; 43:W580–4. [PubMed: 25845596]
- Liao Y, Smyth GK, Shi W. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics*. 2014; 30:923–930. [PubMed: 24227677]
- Lorenz C, von Pelchrzim F, Schroeder R. Genomic systematic evolution of ligands by exponential enrichment (Genomic SELEX) for the identification of protein-binding RNAs independent of their expression levels. *Nat Protoc*. 2006; 1:2204–2212. [PubMed: 17406458]
- Lybecker M, Zimmermann B, Bilusic I, Tukhtubaeva N, Schroeder R. The double-stranded transcriptome of *Escherichia coli*. *Proc Natl Acad Sci U S A*. 2014; 111:3134–3139. [PubMed: 24453212]
- Mehl RA, Kinsland C, Begley TP. Identification of the *Escherichia coli* nicotinic acid mononucleotide adenyltransferase gene. *J Bacteriol*. 2000; 182:4372–4374. [PubMed: 10894752]
- Mellin JR, Cossart P. Unexpected versatility in bacterial riboswitches. *Trends Genet*. 2015; 31:150–156. [PubMed: 25708284]
- Nudler E, Gottesman ME. Transcription termination and anti-termination in *E. coli*. *Genes to Cells*. 2002; 7:755–768. [PubMed: 12167155]
- Parameswaran P, Jalili R, Tao L, Shokralla S, Gharizadeh B, Ronaghi M, Fire AZ. A pyrosequencing-tailored nucleotide barcode design unveils opportunities for large-scale sample multiplexing. *Nucleic Acids Res*. 2007; 35:e130. [PubMed: 17932070]
- Peters JM, Vangeloff AD, Landick R. Bacterial transcription terminators: the RNA 3'-end chronicles. *J Mol Biol*. 2011; 412:793–813. [PubMed: 21439297]
- Peters JM, Mooney RA, Grass JA, Jessen ED, Tran F, Landick R. Rho and NusG suppress pervasive antisense transcription in *Escherichia coli*. *Genes Dev*. 2012; 26:2621–2633. [PubMed: 23207917]
- Pfaffl, MW. Quantification strategies in real-time PCR. International University Line (IUL); La Jolla, CA, USA: 2004.
- Proshkin S, Rahmouni AR, Mironov A, Nudler E. Cooperation between translating ribosomes and RNA polymerase in transcription elongation. *Science*. 2010; 328:504–508. [PubMed: 20413502]
- Richardson JP. Preventing the synthesis of unused transcripts by Rho factor. *Cell*. 1991; 64:1047–1049. [PubMed: 2004415]
- Richardson JP. Loading Rho to terminate transcription. *Cell*. 2003; 114:157–159. [PubMed: 12887917]
- Roberts JW. Termination factor for RNA synthesis. *Nature*. 1969; 224:1168–1174. [PubMed: 4902144]
- Salstrom JS, Fiandt M, Szybalski W. N-Independent leftward transcription in coliphage lambda: Deletions, insertions and new promoters bypassing termination functions. *MGG Mol Gen Genet*. 1979; 168:211–230. [PubMed: 286866]
- Santangelo TJ, Artsimovitch I. Termination and antitermination: RNA polymerase runs a stop sign. *Nat Rev Microbiol*. 2011; 9:319–329. [PubMed: 21478900]
- Sedlazeck FJ, Rescheneder P, von Haeseler A. NextGenMap: fast and accurate read mapping in highly polymorphic genomes. *Bioinformatics*. 2013; 29:2790–2791. [PubMed: 23975764]
- Sedlyarova N, Shamovsky I, Bharati BK, Epshtein V, Chen J, Gottesman S, Schroeder R, Nudler E. sRNA-Mediated Control of Transcription Termination in *E. coli*. *Cell*. 2016; 167:111–121.e13. [PubMed: 27662085]

- Sen R, King RA, Weisberg RA. Modification of the properties of elongating RNA polymerase by persistent association with nascent antiterminator RNA. *Mol Cell*. 2001; 7:993–1001. [PubMed: 11389846]
- Serganov A, Nudler E. A decade of riboswitches. *Cell*. 2013; 152:17–24. [PubMed: 23332744]
- Sievers F, Wilm A, Dineen D, Gibson TJ, Karplus K, Li W, Lopez R, McWilliam H, Remmert M, Söding J, et al. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol Syst Biol*. 2011; 7:539. [PubMed: 21988835]
- Singer BS, Shtatland T, Brown D, Gold L. Libraries for genomic SELEX. *Nucleic Acids Res*. 1997; 25:781–786. [PubMed: 9016629]
- Skordalakes E, Berger JM. Structure of the Rho transcription terminator: mechanism of mRNA recognition and helicase loading. *Cell*. 2003; 114:135–146. [PubMed: 12859904]
- Soldevilla MM, Villanueva H, Pastor F. Aptamers: A Feasible Technology in Cancer Immunotherapy. *J Immunol Res*. 2016; 2016:1–12.
- Soper T, Mandin P, Majdalani N, Gottesman S, Woodson SA. Positive regulation by small RNAs and the role of Hfq. *Proc Natl Acad Sci U S A*. 2010; 107:9602–9607. [PubMed: 20457943]
- Tuerk C, Gold L. Systematic evolution of ligands by exponential enrichment: RNA ligands to bacteriophage T4 DNA polymerase. *Science*. 1990; 249:505–510. [PubMed: 2200121]
- Wade JT, Grainger DC. Pervasive transcription: illuminating the dark matter of bacterial transcriptomes. *Nat Rev Microbiol*. 2014:1–7.
- Wassarman KM, Storz G. 6S RNA regulates E. coli RNA polymerase activity. *Cell*. 2000; 101:613–623. [PubMed: 10892648]
- Weisberg RA, Gottesman ME. Processive Antitermination. *J Bacteriol*. 1999; 181:359–367. [PubMed: 9882646]
- Windbichler N, von Pelchrzim F, Mayer O, Csaszar E, Schroeder R. Isolation of small RNA-binding proteins from E. coli: evidence for frequent interaction of RNAs with RNA polymerase. *RNA Biol*. 2008; 5:30–40. [PubMed: 18388495]
- Zhang H, Zhou T, Kurnasov O, Cheek S, Grishin NV, Osterman A. Crystal structures of E. coli nicotinate mononucleotide adenylyltransferase and its complex with deamido-NAD. *Structure*. 2002; 10:69–79. [PubMed: 11796112]
- Zhou J, Rudd KE. EcoGene 3.0. *Nucleic Acids Res*. 2013; 41:D613–24. [PubMed: 23197660]
- Zhu AQ, von Hippel PH. Rho-dependent termination within the trp t' terminator. I. Effects of rho loading and template sequence. *Biochemistry*. 1998; 37:11202–11214. [PubMed: 9698366]
- Zimmermann B, Bilusic I, Lorenz C, Schroeder R. Genomic SELEX: a discovery tool for genomic aptamers. *Methods*. 2010a; 52:125–132. [PubMed: 20541015]
- Zimmermann B, Gesell T, Chen D, Lorenz C, Schroeder R. Monitoring genomic sequences during SELEX using high-throughput sequencing: neutral SELEX. *PLoS One*. 2010b; 5:e9169. [PubMed: 20161784]
- Zwiefka A, Kohn H, Widger WR. Transcription termination factor rho: the site of bicyclomycin inhibition in Escherichia coli. *Biochemistry*. 1993; 32:3564–3570. [PubMed: 8466900]

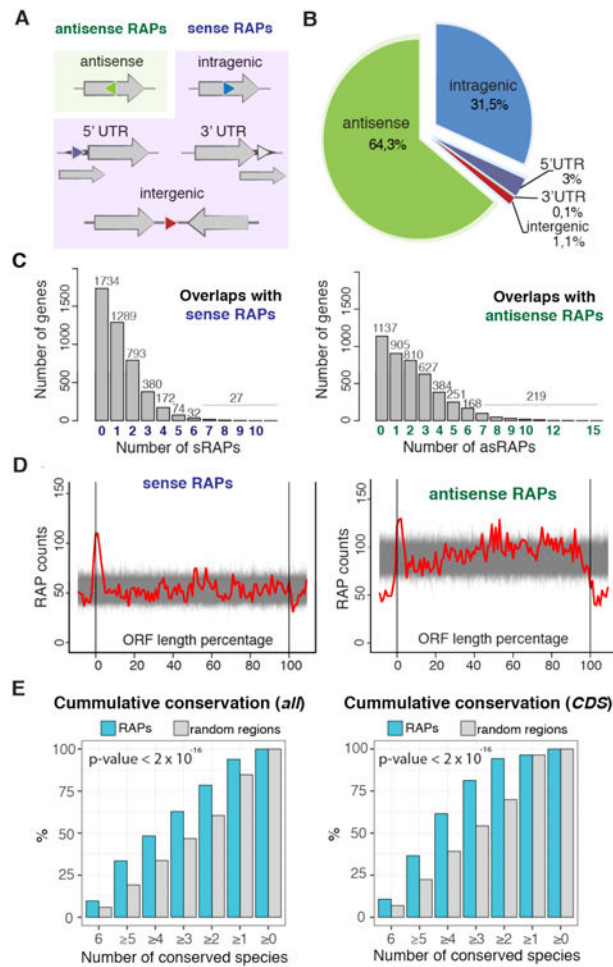


Figure 1. Distribution of RAPs within the *E. coli* genome

(A) RAP categorization according to their location in the *E. coli* genome. Annotated genes are shown in grey. RAP position is depicted as a colored arrow. Shaded areas highlight sense (violet) and antisense (green) groups of RAPs.

(B) Distribution of RAPs from different categories within the *E. coli* genome.

(C) Histogram showing the number of genes (y axis) that contain the indicated number of RAPs (x axis) in the same strand (sense RAPs – left panel) or in the opposite strand (antisense RAPs – right panel).

(D) Analysis of relative location of sense (left panel) and antisense (right panel) RAPs within annotated genes. The red line shows the counts of RAPs found in the respective region of all annotated ORFs (+10% up/downstream). The background (in grey) shows the respective signals from 1000 randomized runs.

(E) Cumulative number of RAP-encoding (cyan) versus random genome regions (grey) conserved between *E. coli* and different representatives of Enterobacteriaceae, including *Shigella*, *Citrobacter*, *Enterobacter*, *Klebsiella*, *Salmonella* and *Yersinia*. Cumulative conservation has been tested for all RAPs (left panel) and for RAPs located only within the coding sequence (CDS) of annotated genes (right panel). Y-axis shows the cumulative fraction of RAPs/random regions conserved. In both cases, RAPs are

significantly more conserved than random sequence elements (generalized linear model on Poisson distribution, p-value $< 2 \times 10^{-16}$).

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

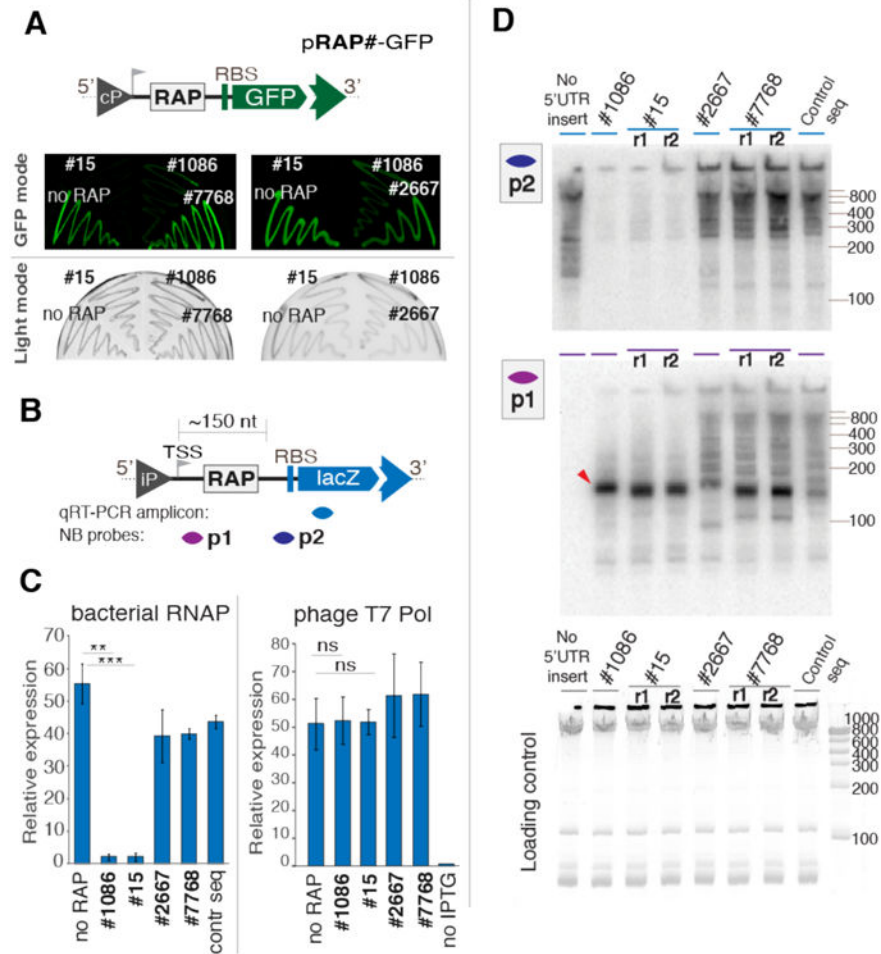


Figure 2. Inhibitory RAPs (iRAPs) promote transcription termination

(A) Upper panel: transcriptional GFP-based fusion, pRAP#-GFP, used to test the effect of RAPs (grey rectangle) *in cis*. “cP” (grey triangle) indicates the location of a constitutive promoter. Grey arrow depicts the transcription start site; RBS, the ribosome-binding site. Middle and lower panels: representative results from the GFP plate assays. DH5a cells transformed with pGFP (no RAP construct) or pRAP#-GFP grew on LB agar plate over 16 hours and the fluorescence intensity was measured (GFP mode, middle panel). The same plate was also captured under visible light to control bacterial density (Light mode, lower panel).

(B) Schematic of the *lacZ*-based reporter construct used to test the effect of RAPs (grey rectangle) *in cis*. “iP” (grey triangle) indicates the location of several promoters used for different experimental setups (inducible *E. coli* RNAP promoter or phage T7 promoter). TSS: transcription start site; RBS: ribosome binding site. The location of qRT-PCR amplicon and Northern Blot probes is depicted below (colored bars).

(C) Levels of reporter gene transcripts containing different RAPs upstream (indicated with their numbers; construct schematics from (B)). qRT-PCR data, normalized to the housekeeping *gapA* levels: on the left - transcription from RNAP promoter; on the right - transcription from phage T7 promoter upon induction with 50 μ M IPTG.

(D) Northern blot analysis of transcripts derived from reporter constructs shown in **(B)**. Numbering on the top indicates the tested RAP. Constructs with RAPs #15 and #7768 are shown in biological replicates (marked with r1 and r2 respectively). Upper panel “P2”: result of hybridization with the probe located downstream of the RAP. Middle panel “P1”: the result of hybridization with the probe upstream of the RAP. The red arrow indicates the stable transcript ~150 nts long. Lower panel: loading control (SYBR-stained denaturing Urea PAGE) to assess the amounts and integrity of the loaded RNA.

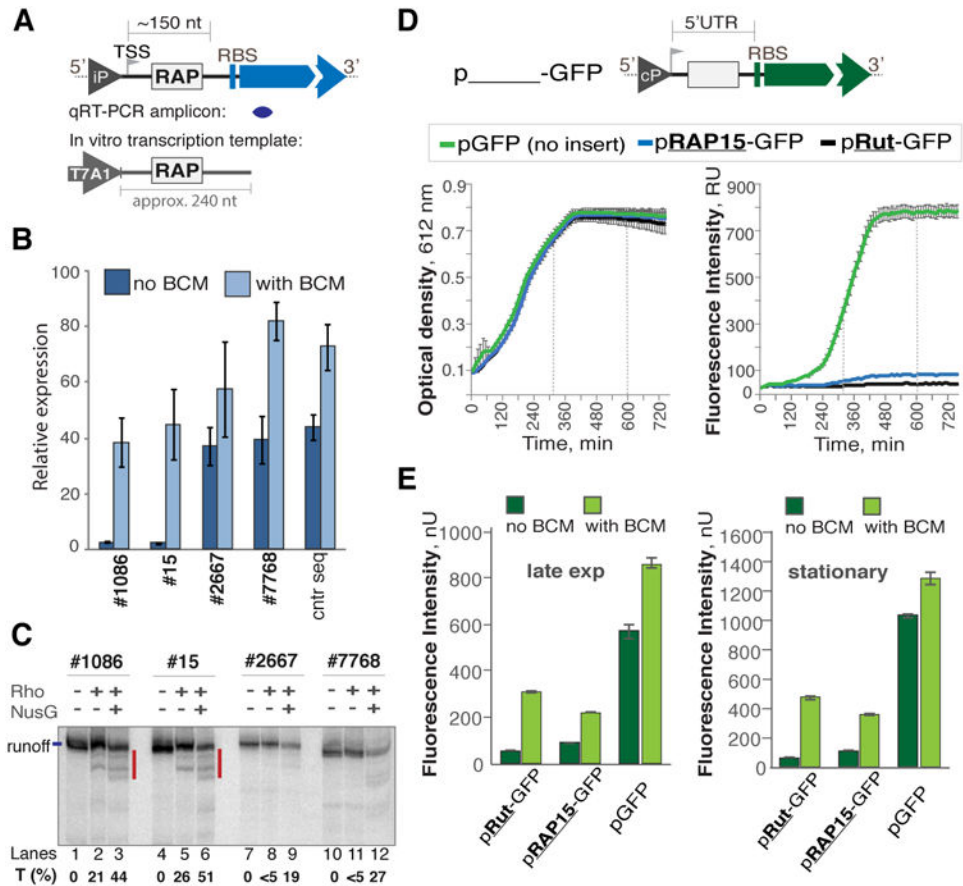


Figure 3. iRAPs cause Rho-mediated transcription termination *in vitro* and *in vivo*

(A) Upper panel: schematic of the *lacZ*-based reporter construct as in Figure 2B. Lower panel: schematic of the template used for *in vitro* transcription with tested RAPs (shown as a grey rectangle).

(B) Effect of BCM (Rho inhibitor) on reporter gene expression containing diverse RAPs. qRT-PCR data, normalized to the housekeeping *gapA* levels before and after BCM treatment (20' with final concentration 24 μ g/ml). Numbering on the bottom indicates the tested RAP.

(C) Representative run-off transcription assay. Pre-formed elongation complexes were chased with NTPs in the absence of Rho (lanes 1, 4, 7, 10), in the presence of Rho (lanes 2, 5, 8, 11) or with Rho and NusG (lanes 3, 6, 9, 12). Blue horizontal bar indicates the run-off product. The abundant products of Rho-dependent termination for RAPs-15 and -1086 are marked with red vertical bars. The percentage of termination (T%) was calculated for each reaction as a ratio between the amount of radioactivity in the bands corresponding to the termination products (highlighted with red bars) and the total radioactivity signal of the termination and readthrough bands. The observed termination efficiency *in vitro* correlates with the effect on reporter expression *in vivo* (see Fig. 2C-D).

(D) Upper panel: pInsert-GFP transcriptional fusion (see also Figure 2A) used to test insert effect (shown as grey rectangle) *in cis*. "cP" (gray triangle) indicates the location of the constitutive promoter. Grey arrow indicates the position of transcription start site; RBS, ribosome-binding site. Lower panels: inserts' effects on transcription as detected by real-time fluorescence. *E. coli* wild-type strains transformed with empty construct (pGFP), RAP-15-

containing vector (pRAP-15-GFP) or construct with characterized Rho-utilization site (pRut-GFP) grew in LB for more than 12 hours while cell optical density (left) and fluorescence intensities (right; shown in relative units, RU) were monitored simultaneously. The dashed vertical lines (in grey) indicate late exponential and stationary phases. For each time point the values represent means \pm SD, n = 3.

(E) Effect of BCM on GFP expression for the constructs in (D). Levels of fluorescent protein normalized to the amount of bacterial cells (Fluorescence Intensities, nU) for the compared constructs in the absence or presence of low concentration of BCM (10 μ g/ml) in two different growth phases – late exponential (after 5 hours growth; left panel) and stationary (after 10 hour growth; right panel).

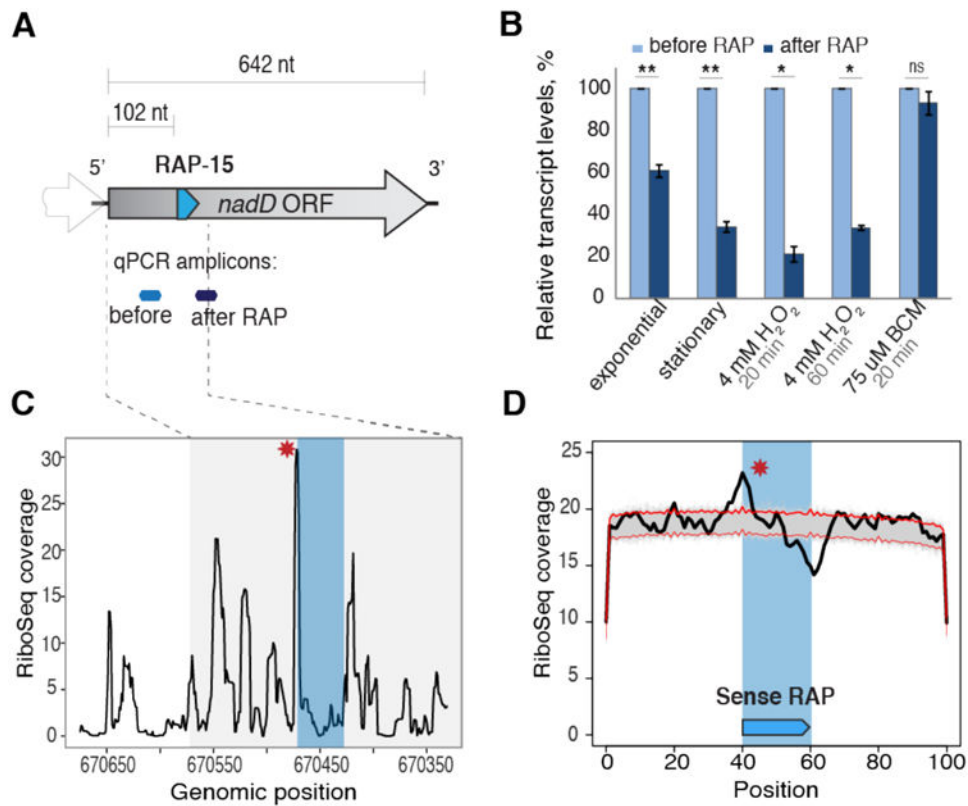


Figure 4. RAP-15 induces Rho-dependent termination within its host gene in response to stress

(A) Schematic of the *nadD* gene ORF with the location of the qRT-PCR primers used for measurements of endogenous *nadD* transcript levels upstream (light blue bar) and downstream (dark blue bar) the RAP-15 domain.

(B) qRT-PCR data, *nadD* mRNA levels upstream and downstream of RAP-15 under various growth conditions. Upstream mRNA levels are set to 100% and downstream levels are normalized to the upstream levels. Values represent means \pm SD, n = 3; ** P < 0.01; * P < 0.05; ns, not significant (tested with t-test).

(C) Ribosome occupancy within *nadD* (shaded in grey) with indicated location of RAP-15 (shaded in blue). The ribosome stalling (ribosome density) peak just upstream of RAP-15 (red asterisk) is the most prominent in the whole *nadD* gene (Li et al., 2012).

(D) Plot for average ribosome occupancy for all gene fragments containing sense iRAPs. RAP position is shown in blue shading. The prominent peak of maximal ribosome stalling right upstream of iRAPs is marked with red asterisk. The background (in grey) shows the respective signals from 1000 randomized runs.

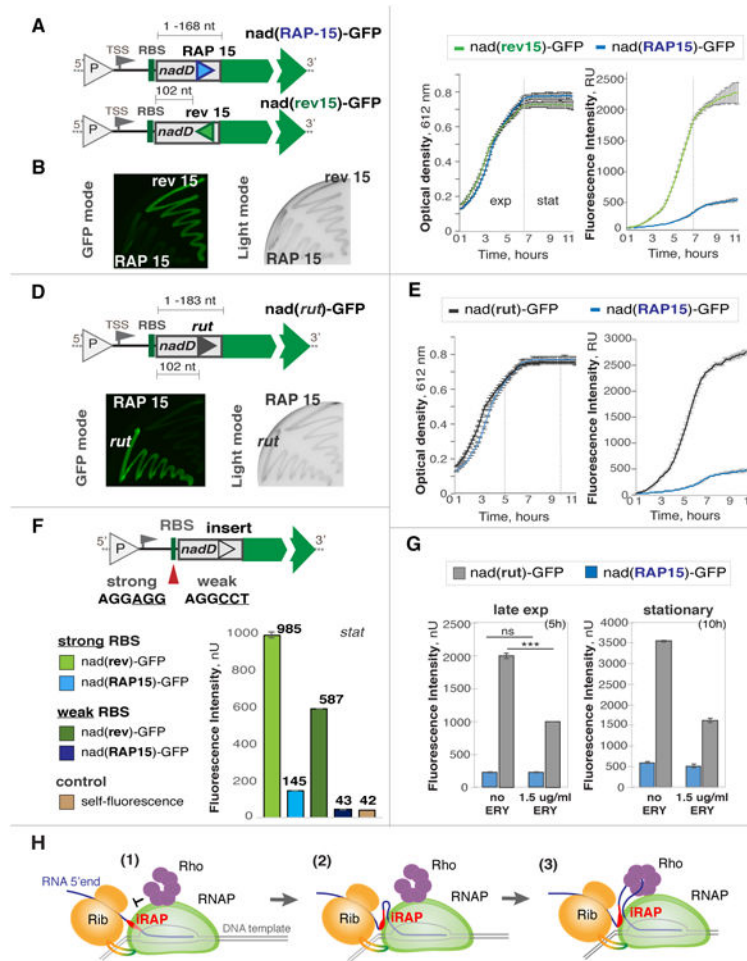


Figure 5. RAP-15 enables Rho to act within the translated region

(A) Schematics of the translational fusion reporter constructs used to test the inhibiting effect of RAP-15 within a translated region. Upper panel: RAP-15 (blue triangle) in its endogenous context (first 168 nt of *nadD* gene; see also Table S4) was fused with GFP reporter protein. Lower panel: control construct with reverse complement sequence of RAP-15 – rev 15 (green triangle). “P” (grey triangle) indicates the promoter location. TSS: transcription start site; RBS: ribosome binding site.

(B) *E. coli* GFP plate assay. Left panel: *E. coli* cells transformed with *nad*(RAP-15)-GFP or *nad*(rev-15)-GFP reporter plasmids were grown on LB agar plate followed by fluorescence intensity measurements (GFP mode). Right panel: corresponding plate was captured under visible light showing the bacterial density of strains with translational fusions (Light mode).

(C) *E. coli* growth experiment with real-time fluorescence detection. Cells transformed with *nad*(RAP-15)-GFP or *nad*(rev-15)-GFP reporter constructs (shown in blue and green) were grown in LB media with simultaneous cell density (left panel) and GFP intensities (right panel) measurements. Grey vertical lines indicate late exponential and stationary phases. At each time point values represent means \pm SD, $n = 3$.

(D) Comparison of RAP-15 and *rut* activity within the translated region. Upper panel: schematic of the translational reporter construct *nad*(*rut*)-GFP used to test the effect of

canonical Rho-utilization site (*rut*). Lower panels: *E. coli* GFP plate assay. Left: *E. coli* cells transformed with nad(RAP-15)-GFP or nad(*rut*)-GFP reporter plasmids were grown on LB agar plate followed by fluorescence intensity measurements (GFP mode). Right: corresponding plate was captured under visible light showing the bacterial density of strains with translational fusions (Light mode).

(E) *E. coli* growth experiment with real-time fluorescence detection. Cells transformed with nad(RAP-15)-GFP or nad(*rut*)-GFP reporter constructs (shown in blue and black) were grown in LB media with simultaneous cell density (left panel) and GFP intensities (right panel) measurements. Grey vertical lines indicate late exponential and stationary phases. At each time point values represent means \pm SD, n = 3.

(F) Ribosome occupancy effect on iRAP activity within ORF. Schematics of the translational fusion reporter nad(insert)-GFP used to test the inhibitory effect of RAP-15 or its reverse sequence rev-15 (location marked with a white triangle) within ORF. “P” (grey triangle) indicates the promoter location. Grey arrow - transcription start site (TSS). Stronger (AGGAGG) and weaker (AGGCCT) versions of the ribosome-binding site (RBS) are indicated. Fluorescence intensity signal normalized to the number of cells per sample (nU) measured after 10 hours in rich media (stationary phase). RBS weakening resulted in \sim 1.7 fold decrease of total GFP (green and dark-green bars; strong nad(*rev*)-GFP vs weak nad(*rev*)-GFP constructs). In the case of RAP-containing fusions, RBS weakening resulted in $>$ 3-fold reduction in GFP production, bringing the signal to the level of auto-fluorescence.

(G) Effect of macrolide antibiotic erythromycin on GFP expression for the constructs compared in (D). Levels of fluorescent protein normalized to the amount of bacterial cells (Fluorescence Intensities, nU) for the compared constructs in the absence or presence of low concentration of erythromycin (ERY, 1.5 μ g/ml) in two different growth phases – late exponential (after 5 hours growth; left panel) and stationary (after 10 hour growth; right panel).

(H) A model for iRAP-mediated transcription termination within the translated region. (1) Coupling between the leading ribosome and RNAP leaves no room for Rho to load on the nascent transcript (Burmam et al., 2010; Proshkin et al., 2010); (2) An iRAP emerging from the RNAP exit channel interacts with RNAP, resulting in ribosome pausing and RNA looping; (3) Naked unstructured RNA becomes available for Rho to load and terminate transcription.

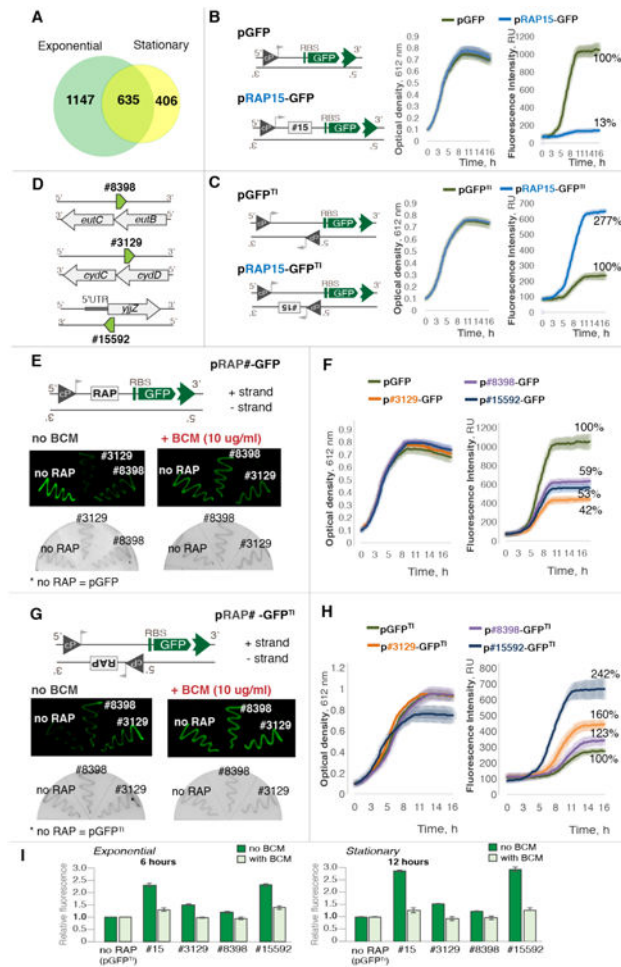


Figure 6. Antisense iRAPs control transcription interference

(A) Venn diagram shows the overlap of antisense RAPs identified as transcribed in exponential and stationary phases (Sedlyarova et al., 2016).

(B)-(C) Proof-of-principle system for studying the effect of iRAPs when encoded on the antisense strand. (B) Left panel: pGFP-based transcriptional fusion schematics used to test iRAP (here RAP-15) effect in *cis*. “cP” (gray triangle) indicates the location of the constitutive promoter. Grey arrow indicates the position of the transcription start site; RBS, ribosome-binding site. Middle and right panels: quantification of RAP-15 effect on transcription as detected by real-time fluorescence. *E. coli* DH5 α strains transformed with empty construct (pGFP) or RAP-15-containing vector (pRAP15-GFP) grew in LB for more than 12 hours while cell optical density (middle) and fluorescence intensities (right; shown in relative units, RU) were monitored simultaneously. Numbers on the right indicate GFP intensity relatively to pGFP construct after 16 hours growth. For each time point the values represent means \pm SD, $n = 3$. (C) Left panel: transcription interference system pGFP^{TI} for testing the effect on GFP expression when iRAP (grey rectangle) is expressed from the opposite strand. “cP” (gray triangles) indicate the locations of convergent constitutive promoters. Grey arrows indicate the position of transcription start sites; RBS, the ribosome-binding site. Middle and right panels: quantification of RAP-15 effect in the transcription

interference system as detected by real-time fluorescence. *E. coli* DH5a strains transformed with empty construct (pGFP^{TI}) or RAP-15-containing vector (pRAP15-GFP^{TI}) grew in LB for more than 12 hours while cell optical density (middle) and fluorescence intensities (right; shown in relative units, RU) were monitored simultaneously. Numbers on the right indicate GFP intensity relatively to pGFP^{TI} construct after 10 hours growth. For each time point the values represent means \pm SD, n = 3.

(D) Position of several antisense iRAPs (green arrows) and corresponding genes (grey arrows) encoded on the antisense strand.

(E) - (F) Effect of the antisense iRAPs when encoded on the same strand with reporter GFP:

(E) upper panel: schematic of pGFP-based plasmid (see also legend for (B)), lower panels: *E. coli* GFP plate assays for the transformed cells grown in the absence (left) and in the presence (right) of BCM; **(F)** Quantification of iRAP-mediated effect in pGFP constructs on transcription as detected by real-time fluorescence. Numbers on the right indicate GFP intensity relatively to that of pGFP construct after 16 hours growth. For each time point the values represent means \pm SD, n = 3.

(G) - (H) Effect of the antisense iRAPs when transcribed from the strand opposite to the one with reporter GFP: **(G)** upper panel: schematic of pGFP^{TI}-based transcription interference construct (see also legend for (C)), lower panels: *E. coli* GFP plate assays for the transformed cells grown in the absence (left) and in the presence (right) of BCM; **(H)** Quantification of iRAP-mediated effect in pGFP^{TI} constructs on transcription as detected by real-time fluorescence. Numbers on the right indicate GFP intensity relatively to pGFP^{TI} construct after 16 hours growth. For each time point the values represent means \pm SD, n = 3.

(I) Quantification of BCM effect on GFP expression in iRAP-containing transcription interference system pGFP^{TI} in exponential (left panel) and stationary phases (right panel). Relative fluorescence values from tested constructs normalized to no RAP (empty pGFP^{TI} vector) control values for cells grown in the absence (dark-green) and in the presence (light-green) of BCM (supplemented here to final concentration 10 μ g/ml). Values represent means \pm SD, n = 3.

Table 1

Representative inhibitory RAPs (iRAPs)

RAP ID	Gene	Strand	Category	Theoretical coverage	Inhibitory RAPs (iRAPs)	Tested sequence
15	nadD	-	intragenic	2042,3	GTCACAATCATCCCTAAATAATGTTCTCTCCGGCATCGTCCC	
683	glnD	-	intragenic	320,9	TTAAACGAAATGTCTGCATATATTTGGCGTATTCGGCTTTTGCCCTGCATCTG	
803	resF	-	intragenic	97,9	GTTCCATGTTAAGCAGATGCCCTGTGGAACCGGTTCAAAGCACTG	
1510	bolA	+	intragenic	139,3	CAACCCGTAATCTCTCGAAGTACTGGATGAAAGCTATCGTCACAA	
1800	gcl	+	intragenic	74,3	ACATGATCACCCGGCTCTAATTCGGCTTCTGTGATTCATTCCTAATCTGTG	
2136	uspG	-	intragenic	200,2	CAC TTCACCATCGATCCTTCCCGCATTAAACAAACATGTCCGT	
3930	flgL	+	intragenic	43,3	AACCCCTTCTGACGATCCCAITGCTGCATCACAAAGCCGTAGTTTC	
9559	ygdH	+	intragenic	22,6	TCGTTTTTCCC AACCTGAATCTCGACA ACTC	
10070	pgk	-	intragenic	10,6	CTAAACGTTCTGTGACCACTGCAACATCCC	
10243	glcB	-	intragenic	253,8	ACCAAACCAACGTACAGAGCGTACAAAGCCAACA	
11599	bfr	-	intragenic	20,8	TGCCAGATCAGAAGCAGCATTTCTCAACATCTTCACCAATGTTCA	
15592	yjZ	-	antisense	99,9	TACATCAGCCCTGCAATCAGCAATCCCGGCAGCAACACTCCCAGCCA	
1086	yahJ	-	antisense	6,4	GCGTTTTCAAATGGCGATCAACCTGCCACACTCCCACCA	
1436	pgpA	-	antisense	16,6	CAGCGCATGAGCGTGATCCACATACCAATAAATTCGTCCCAGACAATGCTGCCATGATCG	
1760	ybbP	-	antisense	1578,4	ACCAITCAACC ACTGACCGGACGATTAGCTTACGGCAGTTGGCGCTCGCCCTG	
4599	yciW	+	antisense	21,8	CGGGCATATTTGCGTGATTTGTGCCAACCGATGATTACTTTCCCCCTGC	
5039	yncJ	+	antisense	344,7	TCATGACGGAGCTGATGATCCACATTTTACCACACACAAATTCATGTCTCTTT	
6819	yeeO	+	antisense	342,0	ACGGCAAAACATTTCCCATCCAGACACC AACCCACACC	
9505	relA	+	antisense	24,6	CAATGTCCATACTTAATGTCGAGAGGATCTCCACCATCTCAAC	
9533	gudP	+	antisense	73,4	CAGCGCCATAAAGCCGATGATCATCCACTCAACGTTGACGTAGTTGCA GAACACCATCACC	
10822	tdcE	+	antisense	2064,3	GAAATACCCATCCAGCAGGCCGACAAAGGTTGGTTTTACGTACTGGATCTCTTTGGC	
15003	ulaG	+	antisense	129,1	CAAAGCCACCACATCGCAAATGTGCCAGGACCGACTGTCTTTGTTCAAATTTCT	
3129	eydD	+	antisense	128,4	CACACTTAATTCGGGTAACGTCCTTGTCTCAATAATCCGGCCATC	
8398	eutB	+	antisense	23,6	TAACAGCATCACAGCCCATAGATGCCCTCAGCTTGCCCATAAAGTGTCTTC	
RAPs with no inhibitory effect (several examples)						
7768	glpT	+	antisense	92,0	CACCATAGTAGCAGACCACACGGGGCCACCCCCAT	

RAP ID	Gene	Strand	Category	Theoretical coverage	Tested sequence
393	secA	-	antisense	1170,6	AGCCACTCGGCAATTGGCAAATCGAGGTCGAAAATCGTTCTTCAGAGGTTCCCTG
2160	citG	-	antisense	155,5	CACTGGTGATCAATCACGGCTTGGCCGCATTGCC
3052	poxB	-	antisense	47,4	CCACCCACAAGAGCTATTCCGGCAATGTAGTCACTATTGGCAGC
13901	pfkA	+	antisense	33,2	TCGACAACGACATCAAAAGGCCTGACTACACTATCGTTT
14168	coaA	-	antisense	129,9	TTATCCACACGATCCACATCATGTATTTGTT
15559	yjJL	+	antisense	86,8	CCCCATACCGATACGCGACCAACACGAACTGCGTAAAAGTT

Table shows RAP ID, categorization data, RAP encoding gene and strand, sequences used in reporter fusions, and theoretical coverage data (calculated as the ratio of number of mapped bases to corresponding peak length).