

# RNomics in *Escherichia coli* detects new sRNA species and indicates parallel transcriptional output in bacteria

Jörg Vogel<sup>1,2,\*</sup>, Verena Bartels<sup>3</sup>, Thean Hock Tang<sup>4</sup>, Gennady Churakov<sup>3</sup>,  
Jacoba G. Slagter-Jäger<sup>1</sup>, Alexander Hüttenhofer<sup>3,5</sup> and E. Gerhart H. Wagner<sup>1</sup>

<sup>1</sup>Institute of Cell and Molecular Biology, Biomedical Center, Uppsala University, Box 596, 75124 Uppsala, Sweden, <sup>2</sup>Department of Molecular Genetics and Biotechnology, The Hebrew University-Hadassah Medical School, PO Box 12272, 91120 Jerusalem, Israel, <sup>3</sup>Institut für Experimentelle Pathologie/Molekulare Neurobiologie, Universität Münster, Von-Esmarch-Str. 56, D-48149 Münster, Germany, <sup>4</sup>Department of Medical Microbiology and Parasitology, School of Medical Sciences, 16150 Kubang Kerian, Kelantan, Malaysia and <sup>5</sup>Institut für Molekularbiologie, Abt. Funktionelle Genomik, Universität Innsbruck, Peter-Mayr-Str. 4b, 6020 Innsbruck, Austria

Received September 2, 2003; Revised and Accepted September 30, 2003

## ABSTRACT

Recent bioinformatics-aided searches have identified many new small RNAs (sRNAs) in the intergenic regions of the bacterium *Escherichia coli*. Here, a shot-gun cloning approach (RNomics) was used to generate cDNA libraries of small sized RNAs. Besides many of the known sRNAs, we found new species that were not predicted previously. The present work brings the number of sRNAs in *E.coli* to 62. Experimental transcription start site mapping showed that some sRNAs were encoded from independent genes, while others were processed from mRNA leaders or trailers, indicative of a parallel transcriptional output generating sRNAs co-expressed with mRNAs. Two of these RNAs (SroA and SroG) consist of known (THI and RFN) riboswitch elements. We also show that two recently identified sRNAs (RyeB and SraC/RyeA) interact, resulting in RNase III-dependent cleavage. To the best of our knowledge, this represents the first case of two non-coding RNAs interacting by a putative antisense mechanism. In addition, intracellular metabolic stabilities of sRNAs were determined, including ones from previous screens. The wide range of half-lives (<2 to >32 min) indicates that sRNAs cannot generally be assumed to be metabolically stable. The experimental characterization of sRNAs analyzed here suggests that the definition of an sRNA is more complex than previously assumed.

## INTRODUCTION

In the traditional view, the transcriptional output of a genome comprises three major classes of RNAs, which function either

in genetic information transfer (mRNA) or in protein synthesis (rRNA and tRNA). A minor addition were a few small, untranslated RNAs with housekeeping or regulatory roles [reviewed in Wagner and Vogel (1) and Wassarman *et al.* (2)]. Recent systematic searches for such small non-coding RNAs (sRNAs) have significantly changed our view about their prevalence, since numerous representatives of these molecules are now known to be encoded by bacterial, archaeal and eukaryal genomes (3). In 2001, three groups searched the 'empty' spaces between known protein-coding regions, i.e. intergenic regions (IGRs), of *Escherichia coli* and discovered 31 new sRNAs (4–6). Bioinformatic analysis of sequence conservation among closely related bacteria and/or structural conservation at the RNA level proved to be powerful tools for identification of such sRNA genes. Additional strategies used were analyses of transcription initiation and termination features, or transcript detection on microarrays specifically in IGRs. Other recent approaches to predict sRNAs in *E.coli* employed neural networks to extract common features among known RNAs (7), relied on transcription features alone (8) or used whole genome arrays (9). A recent compilation sets the present number of *E.coli* sRNA genes at 55 and, furthermore, about 1000 non-redundant sRNA candidates have been proposed but are as yet unconfirmed (10).

Though sRNAs comprise a significant fraction of the overall transcriptional output in *E.coli*, the screens for these molecules have not been saturated. Experimental verification of sRNA candidates might have been limited by the growth conditions tested, and stable secondary structures of sRNAs could have hampered detection on oligoribonucleotide arrays [e.g. Wassarman *et al.* (6)]. Moreover, most screens (4,6,8) deliberately searched for sRNAs encoded by independent genes, although sRNAs can be generated by processing of longer transcripts, as in the case of *E.coli* 6S RNA (11).

In contrast to screens that rely on predictions, cDNA cloning of RNAs in the size range of 50–500 nt aims at identifying those sRNAs that are expressed in a given genome

\*To whom correspondence should be addressed Tel: +46 18 471 4579; Fax: +46 18 530 396; Email: joerg.vogel@icm.uu.se  
Correspondence may also be addressed to Alexander Hüttenhofer. Tel: +43 512 507 3630; Fax: +43 512 507 9880; Email: alexander.huettenhofer@uibk.ac.at

under a given set of conditions, irrespective of whether they are encoded independently or generated by processing. This approach, experimental RNomics, has been successfully employed to discover a vast number of non-coding RNAs in several eukaryotic organisms and in the archaeon *Archaeoglobus fulgidus* (12–17).

To complement previous studies of the overall sRNA output in *E. coli* by an entirely different methodology, an RNomics approach was here applied to a eubacterial genome for the first time. We present novel distinct sRNA species, some of which are processed from leader or trailer regions of mRNAs. We also report on their expression patterns, metabolic stability and precise genomic location, including analyses of some previously described sRNAs.

## MATERIALS AND METHODS

### Oligonucleotides

The complete list of RNA and DNA oligonucleotides used in RACE experiments and as probes in hybridization is provided as Supplementary Material available at NAR Online (Table S2).

### cDNA library construction

Overnight cultures of *E. coli* were diluted 1/100, grown at 37°C in LB medium, and harvested at OD<sub>600</sub> = 0.2, 0.6 and 1.5, representing lag phase, exponential/log phase and stationary phase, respectively. Total RNA was prepared by the TRIzol method (Gibco-BRL), and 200 µg were subsequently fractionated on denaturing 8% polyacrylamide gels (7 M urea, 1× TBE buffer). RNAs in the size range of ~50 to ~500 nt were eluted and ethanol precipitated. For a full description of the subsequent cDNA library construction steps, pre-screening on high density filter to exclude rRNA and tRNA fragments, and sequencing procedures, see Yuan *et al.* (17).

### Biocomputational analysis

Overlapping sequences were automatically sorted into contigs (LASERGENE sequence analysis program package) and mapped to the *E. coli* genome sequence (18) by BLASTN searches (NCBI; GenBank entry U00096). Annotations of *E. coli* genes were based on the Colibri database (<http://genolist.pasteur.fr/Colibri/>). MFOLD version 3.1 was used for RNA secondary structure predictions (19).

### RNA isolation, northern detection and RNA half-life determination

*Escherichia coli* K12 cells from overnight cultures were diluted 1/100 in LB medium or M9 minimal medium supplemented with glycerol (0.4%) and thiamine (0.5 mg/l), and subsequently grown at 37°C. Samples for RNA isolation were withdrawn 1.5, 2, 4.5, 6.5, 8.5 and 24 h after dilution in LB (OD<sub>600</sub> = 0.3, 0.6, 3, 3.9, 4 and 3.8, respectively) and 5, 15 and 24 h after dilution in M9 medium (OD<sub>600</sub> = 0.3, 1.05 and 1.1, respectively). The effect of RNase III was studied in the isogenic *rnc70* mutant obtained by P1 transduction from strain W3110 *rnc70::Tn10* (20), and sample time points were adjusted to compensate for slower growth. Total RNA was isolated by acid-phenol extraction and DNase treated (21).

RNA samples (~30 µg but normalized to equal 5S rRNA hybridization signals in final experiments) were denatured for 5 min at 65°C in loading buffer containing 50% formamide, separated on urea-polyacrylamide (8%) gels, and transferred to nylon membranes by electroblotting. Membranes were hybridized with gene-specific <sup>32</sup>P-end-labeled oligodeoxy-ribonucleotides, and hybridization signals were visualized on a Bio-Rad phosphorimager.

RNA half-lives were determined by treating cells with rifampicin (final concentration: 500 µg/ml) and isolation of RNA before (0 min) and 1, 2, 4, 8, 16, 32 and 64 min after rifampicin addition. Stability was determined in exponential (OD<sub>600</sub> = 0.3) and/or in stationary phase (5 h after 1:100 dilution; OD<sub>600</sub> = 3.5). To measure the half-life of the plasmid-borne CopA RNA, K12 cells carrying plasmid pKN177 (22) were used. Half-lives were calculated from lin-log graphs of time after rif treatment against RNA signal intensity.

### 5' and 3' RACE

5' RACE was carried out essentially as described by Bensing *et al.* (23), but with modifications. The detailed protocols for 5' and 3' RACE were published in Argaman *et al.* (4).

## RESULTS

### General strategy

To investigate the population of sRNAs in *E. coli* grown under standard laboratory conditions, three independent cDNA libraries were generated from strain MG1655 (18). Since growth rate-specific expression was observed for many *E. coli* sRNAs [e.g. Argaman *et al.* (4) and Wassarman *et al.* (6)], RNA was extracted from cells in lag phase, exponential phase and stationary phase. RNA was size selected (50–500 nt), C-tailed with poly(A) polymerase, reverse transcribed, and cDNA libraries were constructed by directional cloning (Materials and Methods). A total of 10 000 individual clones were pre-screened by hybridization on high density filters to exclude rRNA and tRNA sequences. Then 1000 cDNA clones from each growth phase, exhibiting the lowest hybridization scores, were sequenced, resulting in single or assembled overlapping sequences, which we refer to as contigs. Contigs were mapped to the *E. coli* genome, and candidates for unknown sRNAs, primarily from IGRs, were tested by northern blot. If signals for distinct sRNAs were observed, characterization was extended to a broader set of growth conditions in both rich and minimal medium. In addition, transcription initiation sites (TSS) and, when necessary, 3' ends were mapped, and the metabolic stability of sRNAs was determined in rifampicin run-out experiments.

### Collection of sRNA candidates from cDNA libraries from different growth phases

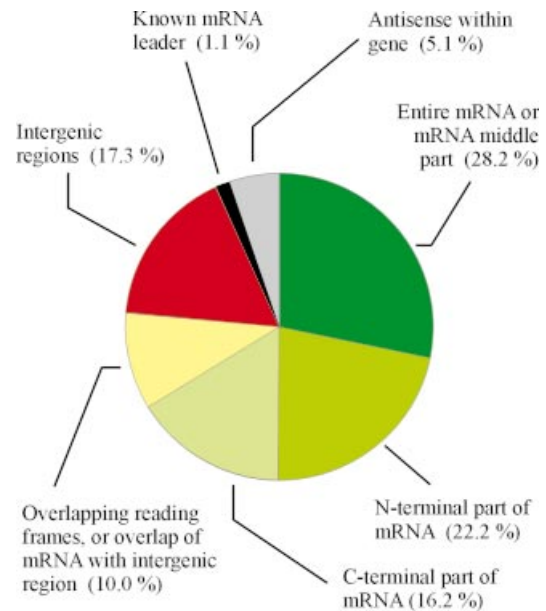
Before final analysis, cDNA sequences that either were too short (<15 bp), only partially matched genomic sequences, or represented tRNA or rRNA segments were removed. Of the final set of 451 contigs in the combined library (Table S1 in Supplementary Material), 77% were derived from within coding regions of known genes or open reading frames (ORFs) of unknown function. There was no strong bias towards

certain regions within mRNA-related fragments (Fig. 1). Surprisingly, 5% of the final contigs had matches within a gene or ORF but in antisense orientation, supporting previous experimental evidence for significant global antisense transcription (24). Fragments from intergenic, or, more precisely, from intercoding regions, accounted for 17%. Known regions of gene leader peptides comprised a further 1%. Since 89% of the *E. coli* DNA consists of protein-coding, tRNA and rRNA regions (18), these figures exhibit a slight bias towards IGRs in which most of the known sRNAs have been found to be located (10).

Comparing these with the list of 55 known *E. coli* sRNAs (10), 20 were found in our library. Significantly, the library contained many sRNAs expressed under standard growth conditions (refer to Supplementary Table S1 for details in this paragraph). Thirty-four candidate contigs, predominantly from IGRs, were selected for further analysis based on: (i) number of individual sequences assigned to a contig; (ii) sequence conservation in closely related enterobacteria and transcription features as in Argaman *et al.* (4); and/or (iii) matches to unverified candidates from the experimental sRNA screens published in 2001 (4–6), but excluding repetitive extragenic palindromic sequences (REP elements) (25). Northern analyses identified seven of these as sRNAs ranging in size from ~70 to 180 nt. The remaining contigs either showed hybridization to high molecular weight RNAs, suggesting them to be degradation intermediates of longer transcripts, or failed to yield a signal even with a second probe. The new sRNAs were named SroA–SroH (small RNomics RNA) according to their location on the *E. coli* genome map (Table 1), including tke1 (5) as SroF. Three sRNAs previously described by others, but whose precise genomic location remained undefined (RybB, RyeB and RygB) (6), were included in the experiments below.

### Expression features of sRNAs (northern blots and RACE mapping)

Northern blot analysis was performed using RNA sampled at different time points of growth in LB and M9 minimal medium, respectively (Fig. 2). As previously observed (4,6), the majority of sRNAs was upregulated upon entry into, or in,



**Figure 1.** Fractions of the final contigs listed in the Supplementary Material Table S1 with regard to their origin in coding or intergenic regions.

stationary phase, whereas SroA and SroG showed the opposite expression pattern. SroD showed a particularly narrow range of expression, peaking when cells grown in LB medium had reached a plateau. SroC reproducibly exhibited two expression peaks in early and in late stationary phase, respectively (most clearly seen in LB medium; Fig. 2). Most of these sRNAs were present as single size species. However, SroF and RybB exhibited growth-dependent size variations, indicative of RNA processing (see below).

While available cDNA sequences provide the genomic location of these sRNAs, they do not always pinpoint exact 5' and 3' ends. To this end, RACE experiments were conducted. For 5' end mapping, we chose a protocol that makes use of the enzyme tobacco acid pyrophosphatase [TAP; see Argaman *et al.* (4) and Bensing *et al.* (23) for details], which permits us to distinguish between 5' ends generated by cleavage/processing and 5' ends of primary transcripts (TSS). These

**Table 1.** Summary of positional information on sRNAs investigated in this work

RNA <sup>a</sup>	Contig	Candidate <sup>b</sup>	5' end <sup>c</sup>	3' end <sup>c</sup>	Size <sup>d</sup>	LAG <sup>e</sup>	LOG <sup>e</sup>	STAT <sup>e</sup>
<b>SroA</b>	406	tpe79	<b>75608</b>	75516	93	–	1	–
<b>SroB</b>	382	–	<b>506428</b>	506511	84	–	1	15
<b>SroC</b>	513	HB_314	686066	685904	163	–	1	3
<b>RybB</b>	485	p25	<b>887280</b>	887200	81	–	–	27
<b>SroD</b>	500	p24	1886126	1886041	86	–	–	1
<b>RyeB</b>	374	tpke79	<b>1921231</b> , 1921201 <sup>f</sup>	1921128	104, 74	–	1	59
<b>SroE</b>	519	k20	2638706	2638615	92	–	–	1
<b>SroF</b>	118	tke1	<b>2689360</b>	2689213, 2689180 <sup>f</sup>	181, 148	1	–	3
<b>RygB</b>	548	t59	<b>2974407</b> , 2974399 <sup>f</sup>	2974326	74, 82	–	1	–
<b>SroG</b>	71	HB_456/7	<b>3182734</b>	3182586	147	2	–	–
<b>SroH</b>	227	–	<b>4188065</b> , 4188041 <sup>f</sup>	4187905	161, 137	1	1	–

<sup>a</sup>For *rybB*, *ryeB* and *rygB*, the names were assigned by Wassarman *et al.* (6).

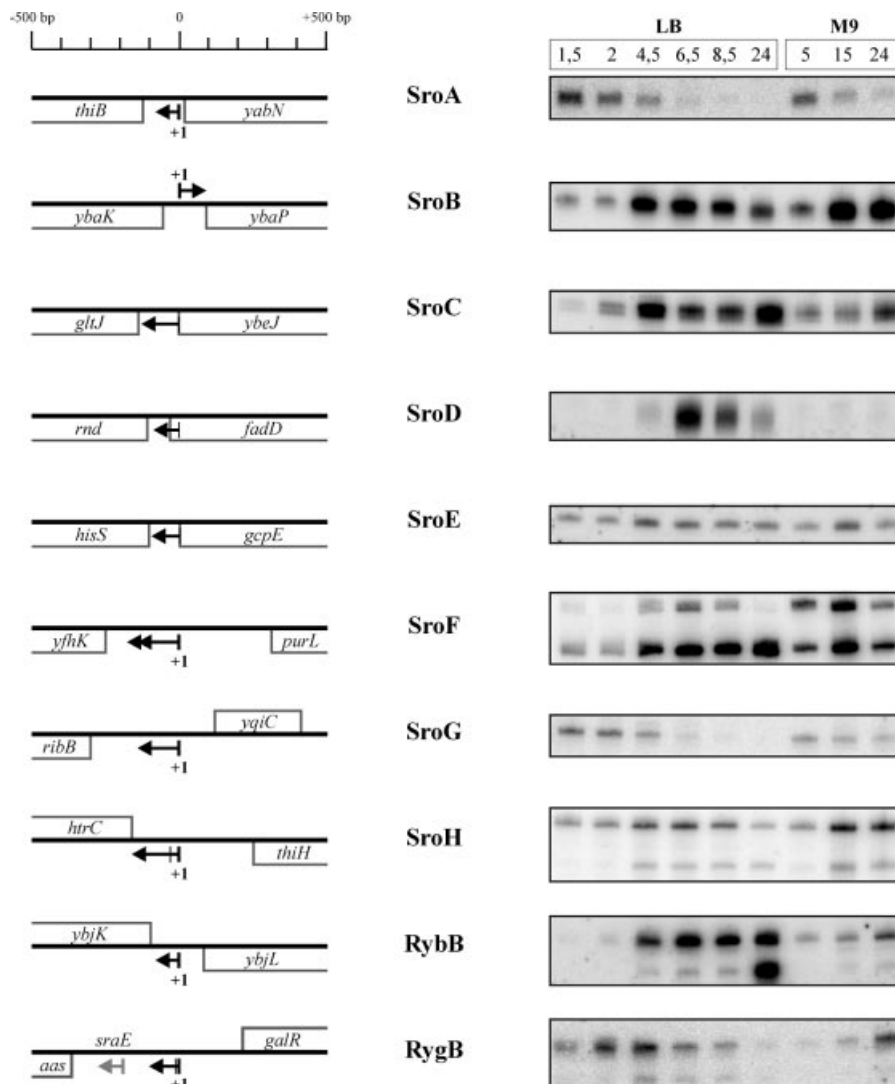
<sup>b</sup>Listed as candidate RNAs in (5, 7–9). For summary, see Hershberg *et al.* (10).

<sup>c</sup>Position according to *E. coli* MG1655 annotation. 5' end in bold if it coincides with TSS, i.e. a TAP-specific 5' RACE product. 3' ends derived from end of Rho-independent terminator (last U in consecutive stretch), by 3' RACE, or observed as homogenous end points in the library.

<sup>d</sup>In nucleotides.

<sup>e</sup>Number of individual clones in libraries from lag phase (LAG), exponential phase (LOG), and stationary phase (STAT).

<sup>f</sup>5' processing mapped in 5' RACE; 3' processing mapped in 3' RACE.



**Figure 2.** Genomic location and expression patterns of sRNAs. Left: location of sRNA region and neighboring genes (above line, + strand; below line, - strand). Thin vertical bars represent the mapped processed end of sRNA; the thick vertical bars and '+1' represent the mapped 5' triphosphate end (transcription initiation site; TSS). Right: autoradiograms of northern experiments. Time points represent hours of growth after 100-fold dilution of overnight cultures in LB or M9 medium (Materials and Methods). End-labeled MspI-digested pUC19 DNA was run as a size marker (not shown). SroA: the gene name *thiB* instead of *tbpA* was used in accordance with Rodionov *et al.* (26) and Webb *et al.* (27).

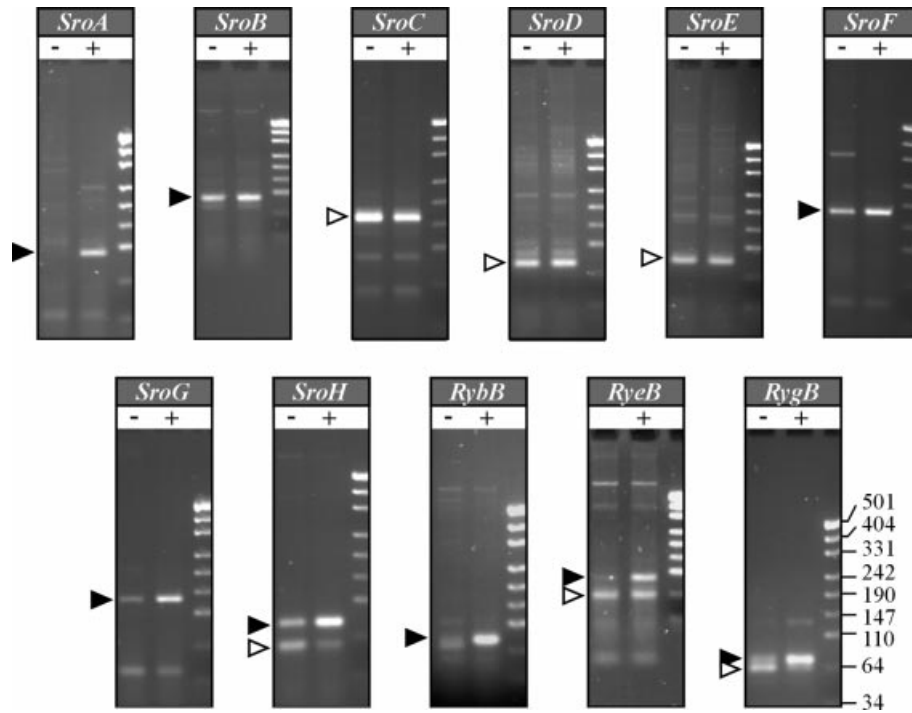
experiments suggested that eight of the 11 sRNAs investigated here were primary transcripts, originating from a promoter located immediately upstream (Figs 2 and 3). RNA processing sites that accounted for faster migrating bands were identified for SroH, RybB and RygB (Fig. 2; Table 1). 3' RACE was performed for some sRNAs that lacked a recognizable Rho-independent terminator at a distance from the 5' end consistent with the length of the RNA (data not shown). The results, which in all cases were consistent with transcript sizes on northern blots, are discussed along with some genomic features of the individual sRNAs as follows.

**SroA.** SroA is encoded in the 163 bp IGR between *thiB* and *yabN* and had been predicted by QRNA; however without specifying an orientation (5). The IGR is conserved between *E.coli* and *Salmonella* species, as are the adjoining genes. Transcription of SroA is initiated 36 bp downstream of the

*yabN* stop codon, and terminates in a U-run following an extended stem-loop region. The downstream *thiB* gene belongs to the *thiBPQ* transport operon whose regulation by thiamine derivatives involves an RNA sensor or riboswitch element (THI element) in the 5' untranslated region (UTR) of *thiB* (26,27). Since the *sroA* region covers this THI element, it is possible that SroA is generated by attenuation of *thiBPQ* transcription in rapidly growing cells (Fig. 2).

**SroB.** SroB is located in a 203 bp IGR, in the opposite orientation to the adjoining ORFs *ybaK* and *ybaP*, encoding proteins of unknown function. This region is conserved in *Salmonella* species, including its Rho-independent terminator and a promoter which is consistent with the mapped TSS (Fig. 3). In line with expression in stationary phase, a C residue at -13 shows the signature of  $\sigma^S$ -regulated promoters (28).





**Figure 3.** RACE mapping of sRNA 5' ends. Specific or increased PCR signals, upon treatment with tobacco acid pyrophosphatase (lane +) compared with mock treatment (lane -), are shown by filled triangles and indicate 5' triphosphate RNA ends. Open triangles: processed RNA 5' ends. Other control samples included untreated, 5' phosphorylated and total DNA (data not shown).

*SroC*. This covers almost the entire 169 bp IGR following the first gene (*ybeJ*) of a proposed alternative glutamate transport operon (29). The 5' end of *SroC* was mapped to the *ybeJ* stop codon and is unlikely to represent a TSS, indicated by RACE results (Fig. 3) and the absence of *E.coli* promoter motifs upstream. Similar alternative secondary structures predicted in *E.coli* and *Salmonella* species (data not shown) suggest that *SroC* is generated by transcription attenuation in front of the *gltJKL* genes.

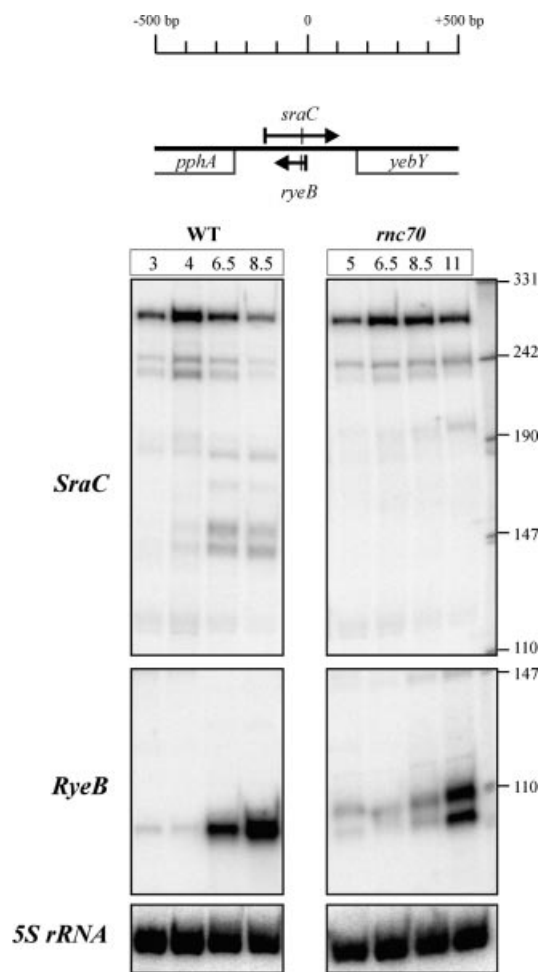
*SroD*. This sRNA extends from within the 3' region of *fadD* into the promoter of the *rnd* gene located downstream [(30); the *rnd* promoter was confirmed in RACE experiments; data not shown] and ends at a Rho-independent terminator that is implicated in regulation of *rnd* translation (30). The latter study failed to detect an sRNA ending at this terminator but, considering the narrow window for *SroD* detection (Fig. 2), this may be attributable to the investigated growth condition. RACE data suggest that *SroD* is processed from the upstream *fadD* mRNA.

*SroE*. *SroE* is yet another sRNA processed from a longer transcript, here from the upstream *gcpE* gene. Its 5' end was mapped to the UAA stop codon of *gcpE* (third nucleotide). *SroE* extends into the promoter region of the downstream *hisS* gene [(31); here confirmed as a TSS; data not shown] where it terminates in a U-run after an extended stem-loop. Both adjacent genes and the 110 bp IGR are conserved between *E.coli* and *Salmonella* species; the *SroE* sequences are predicted to fold into identical two-stem-loop structures with sequence variation confined to loops. Detection of *SroE* confirms a previous prediction by QRNA (5).

*SroF*. This is detectable in two different sizes, as observed by Rivas *et al.* (5), but abundance and molar ratio between the two species vary with growth phase. Both RNAs retain the same primary 5' end (TSS; Fig. 3). The 148 nt RNA is processed in the 3' region, whereas the longer 181 nt RNA ends at a Rho-independent terminator which overlaps the RBS of the downstream ORF *yfhK*, thus indicating *SroF* to be a putative leader/5' UTR. However, since different *yfhK* start codons were annotated in *E.coli* and *Salmonella* species, respectively, and the N-terminus of the YfhK protein is unknown as yet, the strong conservation of the RNA-coding region rather suggests *sroF* to be an independent sRNA gene.

*SroG*. *SroG* is derived from the leader of *ribB* which, like many genes involved in riboflavin biosynthesis, contains an RFN element (32). This riboswitch element senses flavin derivatives and regulates gene expression through transcription attenuation in some *rib* genes (33,34), and regulation of translation initiation in others, e.g. *E.coli ribB* (35). Intriguingly, RACE experiments (Fig. 3) suggest that *SroG* is created by cleavages within the *ribB* 5' UTR that faithfully and exclusively generate the entire RFN element. 3' RACE also showed significant polyadenylation of *SroG* (data not shown).

*SroH*. *SroH* is only present in *E.coli* K12 and OH157 strains. It is encoded within a long IGR (~400 bp) and shares a bidirectional Rho-independent terminator with the downstream *htrC* gene. Lack of conservation of *sroH* in other bacteria coincides with the absence of *htrC*, encoding a heat shock protein (36). In line with upregulation in stationary phase, *sroH* is preceded by a  $\sigma^S$ -like promoter (28).

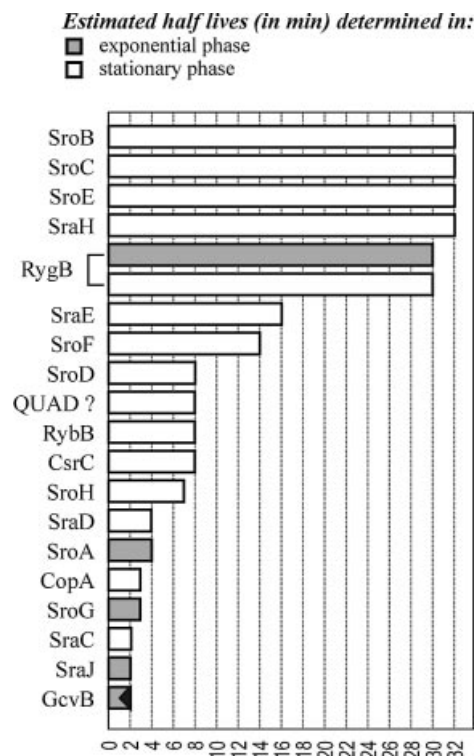


**Figure 4.** RyeB- and RNaseIII-dependent processing of SraC. Total RNA from isogenic strains (*rnc*<sup>+</sup> and *rnc70*) was analyzed by northern blot as in Figure 1. Lower panel: autoradiograms of membranes successively probed for SraC, RyeB and, as loading control, 5S rRNA. Upper panel: location of *sraC* and *ryeB* in *pphA*-*yebY* IGR.

**RybB and RygB.** These are two sRNAs originally described by others (6), but their precise location remained unknown. Neither of the two neighboring genes of *rybB* have Rho-independent terminators, which could suggest RybB to be generated by processing induced by base pairing of the overlapping transcripts in this IGR. However, RybB carries a 5' triphosphate (Fig. 3), and the sRNA region is, in some enterobacteria, only conserved with one of its neighboring genes, *ybjL* (10). Thus, we consider *rybB* to be a genuine sRNA gene. RybB accumulates as a shorter processed RNA species late in growth.

RygB has 77% sequence identity to SraE/RygA (4,6,10), both encoded from the IGR between *aas* and *galR*. Despite high sequence similarity, these two sRNAs exhibit an almost mutually exclusive expression pattern: RygB levels increase around the onset of stationary phase and decrease thereafter (Fig. 2), whereas SraE accumulates as stationary phase progresses (4).

**RyeB and SraC/RyeA** (4,6). These two sRNAs are encoded in opposite orientation from the same IGR, located between



**Figure 5.** Stability of new sRNAs. Half-lives (in min) were determined in rifampicin run-out experiments in exponentially growing cells (filled bars) or stationary phase (open bars). QUAD denotes probing with an oligodeoxyribonucleotide complementary to sRNA genes in QUAD repeats 1a, 1b and 1c (59), identified as *ryeC*, *ryeD* and *rygD* in Wassarman *et al.* (6).

*pphA* and *yebY*. The shorter one, RyeB (104 nt; minor species 74 nt; Table 1), is completely complementary to an internal segment of the longer one, SraC. Both full-length sRNAs are primary transcripts [Fig. 3 and Argaman *et al.* (4)]. Their biological roles are unknown. Interestingly, SraC/RyeA is present as an ~270 nt RNA in early growth and a shorter ~150 nt RNA in stationary phase (4,6). Increased abundance of RyeB upon entry into stationary phase coincides with the appearance of the shorter SraC, suggesting that base pairing between these RNAs could result in duplex-dependent RNase III processing. SraC and RyeB were analyzed on northern blots as a function of growth phase in a K12 strain and its *rnc70* mutant derivative (20). In the wild-type strain, RyeB became abundant at the 6.5 h time point and increased further (Fig. 4). Simultaneously, full-length SraC decreased, and two putative processing products accumulated. In contrast, SraC levels remained unchanged throughout growth in the *rnc70* strain, and the processing bands were absent. The reason for a slow mobility RyeB species in *rnc* mutant conditions is as yet unclear. SraC processing by RNase III in the presence of RyeB supports the first case of two sRNAs interacting by an antisense mechanism, which is additionally strengthened by recent SraC/RyeB binding studies *in vitro*, and analyses of SraC in a strain deficient in *ryeB* expression (F.Darfeuille, J.Vogel and E.G.H.Wagner, in preparation).

#### Intracellular stability of sRNAs

The intracellular concentration of an RNA depends on its rates of transcription and decay. In line with their functional

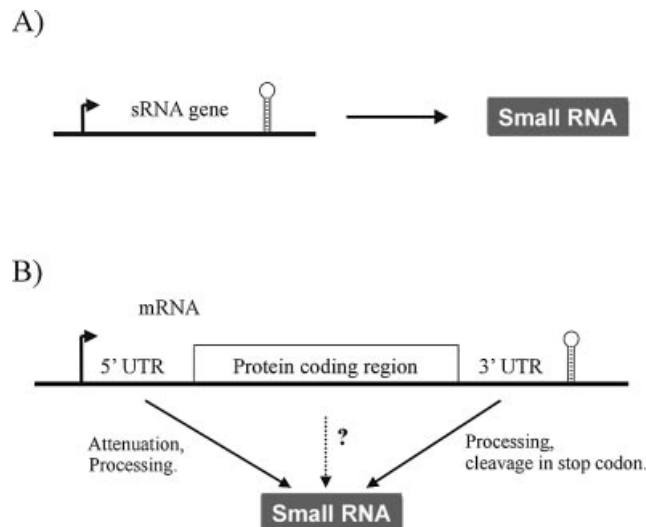
requirements, antisense RNAs of bacterial plasmids generally exhibit very short half-lives [ $<2$  min; (37)], whereas long half-lives (20–60 min) were reported for some housekeeping and regulatory sRNAs encoded by bacterial chromosomes [e.g. see references in Wagner and Vogel (1) and Wagner *et al.* (38)]. However, data are only available for a few sRNAs, and most half-lives were determined in exponentially growing cells, though many sRNAs are most abundant later in growth (4,6). Since little is known about how stationary phase physiology affects RNA decay, we probed, as a control, the antisense RNA CopA of plasmid R1 in a rifampicin run-out experiment in stationary phase cells. The half-life was determined to be  $\sim 3$  min, close to the value obtained in exponential growth (22). We then determined half-lives of 18 *E. coli* sRNAs (Fig. 5) in exponential or stationary phase, according to the condition of their highest abundance [this study and Argaman *et al.* (4)]. The half-life of RygB was determined in both growth conditions; no difference was found (Fig. 5).

Figure 5 reveals a wide spectrum of half-lives, ranging from  $<2$  min to  $>32$  min. The four most stable sRNAs could still be detected 64 min after transcription block (data not shown). Although 61% (11) of the sRNAs exhibited values within the 2–8 min range reported for the majority of *E. coli* mRNAs (39,40), half-lives exceeding 10 min were not uncommon. The group of the most stable sRNAs includes SroC ( $>32$  min), a putative processed fragment of the *ybeJ–gltJ* spacer, and SroB ( $>32$  min), an 84 nt sRNA which, according to predictions by MFOLD (19), almost entirely lacks stable secondary structure elements. On the other hand, GcvB, an sRNA that regulates genes of the major peptide transport systems by an unknown mechanism (41), is unstable ( $<2$  min), as is SraJ (RyiA), which is the most abundant sRNA in co-immunoprecipitates with the RNA-binding protein Hfq (6).

## DISCUSSION

Arguably, *E. coli* is one of the most extensively studied organisms in molecular biology. For this bacterium, six searches for new sRNAs have been carried out (4–9) and so far have extended the number of experimentally verified sRNAs to 55 (10). In addition, numerous candidate sRNAs were predicted but have not been tested so far. The rationale for carrying out yet another search lies in that bioinformatics-based approaches must build on patterns that require prior knowledge of searched-for features. RNAs that fail to meet these criteria may escape detection. RNomics approaches, used successfully in a number of organisms (12–17), identify sRNAs by cDNA shot-gun cloning of those RNA species that are present under a chosen set of conditions. Thus, RNomics is, in some respects, similar to the early approaches that identified distinct sRNAs by orthophosphate labeling and biochemical isolation [4.5S and 6S RNA (42), Spot 42, tmRNA and M1 RNA of RNase P (43)]. However, reverse transcription followed by cloning has since revolutionized isolation and sequence determination of numerous small RNAs in parallel, notably narrowly sized ( $\sim 22$  nt) subclasses of eukaryotic sRNAs, e.g. miRNAs and siRNAs (44–47).

In addition to 20 known sRNAs, our cDNA libraries of cloned sRNAs (or fragments thereof) contain a number of new sRNAs, some of which are absent from the current list of approximately 1000 non-redundant candidates [Table 1; (10)].



**Figure 6.** Definitions of sRNAs. (A) Origin of a distinct sRNA species in an own sRNA gene or (B) through parallel transcriptional output, in an mRNA gene.

Some well-characterized sRNAs were not found here, most probably due to specific expression patterns, e.g. OxyS [oxidative stress (48)] and DsrA [cold shock (49)]. Other RNAs might have escaped detection because of technical limitations, i.e. sRNAs may differ dramatically in their efficiency of C-tailing in cDNA library construction (17), e.g. due to structures near the 3' end. Thus, such technical biases may be responsible for the absence of some sRNAs, but improvement of the individual cloning steps and a greater sampling number might significantly reduce these problems in the future. The total number of experimentally verified sRNAs in *E. coli*, including the present work, is now 62. In the cDNA libraries obtained, 20 of the previously found sRNAs were represented, from single to 60 independent sequences in contigs (Table S1). So far, the biological roles of the recently identified *E. coli* sRNAs are unknown [exceptions are RyhB/FerA (50) and CsrC (51)], and so are their targets, presuming most of them to be riboregulators/antisense RNAs (52,53).

Given that functions have yet to be assigned, what can we learn from the present study? RACE experiments and positional information indicate that the definition of an sRNA is a matter of perspective. For example, distinct sRNA species are not always indicative of independently synthesized RNAs (independent gene, bordered by promoter and terminator). Some sRNAs are derived from leader (e.g. SroA and SroG) or trailer regions of mRNAs (e.g. SroD, SroE and SroC). An interesting observation concerns SroE, predicted to be encoded in the *hisS–gcpE* IGR by Rivas *et al.* (5). SroE is not a primary transcript (Fig. 3), as its 5' end is generated by cleavage within the stop codon of the preceding *gcpE* ORF. Intriguingly, a recent study showed that the RelE toxin, in poor growth conditions, arrests translation by cleaving ribosome-bound mRNAs within stop codons (54).

One can speculate that sRNAs may not only originate from independent genes, but may alternatively be generated and accumulated as independent functional units by processing from longer transcripts. Mattick and Gagen (55) suggested non-coding regulatory RNAs (in eukaryotes) to be part of a



parallel transcriptional output, e.g. they are processed out of mRNA introns and in turn play roles in regulatory cross-talk. Their location would ensure that they are produced along with the intron-containing mRNA. Similarly, bacterial sRNAs derived from mRNA leaders or trailers may have independent functions, exerted under conditions when the gene in question is active. SroA and SroG consist essentially exclusively of known riboswitch elements, THI and RFN, respectively (26,33,34,56). Riboswitch elements are aptamer-like binding sites for small molecules (here thiamine and flavin derivatives) that trigger secondary structure changes of the mRNA leader in response to ligand binding, resulting in translational or attenuation-type regulation of the genes they precede. Since both elements accumulate strongly as independent sRNA species with predicted binding affinity, it is conceivable that they could carry out additional, independent roles such as titration of the ligand.

Further complexity arises from the observation that one and the same RNA may function as both mRNA and regulatory RNA. *Staphylococcus aureus* RNAIII acts as both an mRNA encoding  $\delta$ -hemolysin, and an antisense RNA regulator of the  $\alpha$ -hemolysin mRNA (57). Whether such 'moonlighting' sRNAs are present in *E.coli* is unknown but remains an interesting possibility. Thus, in the absence of functional information, the current definition of an sRNA must rely on its mere presence in cells, and various themes are possible (Fig. 6). This may even include sRNAs derived from coding regions; some of the mRNA-derived contigs in the cDNA library represent such candidates and are currently being tested.

Riboregulators can be expected to be present in significant excess over their targets when regulation occurs. Characteristically, plasmid-encoded antisense RNAs are constitutively expressed, unstable, and yet abundant (37). Most chromosomally encoded sRNAs are induced, but intracellular stabilities are known for only a few. All RNAs tested were analyzed when they were most abundant, i.e. they can be assumed to be in excess over putative targets, so that target interaction-dependent degradation pathways (e.g. by RNase III or RNase E) at this point are expected to have little effect on the measured half-life. However, if/when target genes are induced, and the target RNA concentration approaches that of the riboregulator, second-order decay rates may increasingly dominate over target-independent (first order) decay rates. Thus, although very long half-lives in rifampicin run-out experiments may be questionable due to major changes in physiology at late time points, it is clear that sRNAs cover the entire range from very unstable to stable (Fig. 5).

In conclusion, this study used an RNomics approach to describe yet another set of new sRNAs in *E.coli*, adding to the previously found sRNAs (4–9). Since 62 sRNAs are now supported experimentally, it is clear that sRNAs (although the hallmarks of *bona fide* sRNAs may still be poorly defined) constitute a significant fraction of the transcriptional output. In a wider perspective, the ubiquitous presence of new sRNAs in eukaryotes and archaea (3,58) has only recently been recognized. The exciting and challenging task in the years to come will be the elucidation of the roles these sRNAs play in all kingdoms of life.

## SUPPLEMENTARY MATERIAL

Supplementary Material is available at NAR Online.

## ACKNOWLEDGEMENTS

We are grateful to Santanu Dasgupta for help with the *rnc70* mutant, and to Åsa Björklund and Kristina Jonas for excellent technical assistance. This work was supported by grants from the Human Science Frontier Program, the Swedish Research Council, the Wallenberg Foundation and the Swedish Foundation for Strategic Research to E.G.H.W., and by the German Human Genome Project through the BMBF (01KW9966) and an IZKF grant (Teilprojekt IKF3 G6, Münster) to A.H. J.V. acknowledges the support by an EMBO long-term fellowship.

## REFERENCES

- Wagner,E.G.H. and Vogel,J. (2003) Noncoding RNAs encoded by bacterial chromosomes. In Barciszewski,J. and Erdmann,V. (eds), *Noncoding RNAs*. Landes Bioscience, pp. 243–259.
- Wassarman,K.M., Zhang,A. and Storz,G. (1999) Small RNAs in *Escherichia coli*. *Trends Microbiol.*, **7**, 37–45.
- Storz,G. (2002) An expanding universe of noncoding RNAs. *Science*, **296**, 1260–1263.
- Argaman,L., Hershberg,R., Vogel,J., Bejerano,G., Wagner,E.G.H., Margalit,H. and Altuvia,S. (2001) Novel small RNA-encoding genes in the intergenic regions of *Escherichia coli*. *Curr. Biol.*, **11**, 941–950.
- Rivas,E., Klein,R.J., Jones,T.A. and Eddy,S.R. (2001) Computational identification of noncoding RNAs in *E.coli* by comparative genomics. *Curr. Biol.*, **11**, 1369–1373.
- Wassarman,K.M., Repoila,F., Rosenow,C., Storz,G. and Gottesman,S. (2001) Identification of novel small RNAs using comparative genomics and microarrays. *Genes Dev.*, **15**, 1637–1651.
- Carter,R.J., Dubchak,I. and Holbrook,S.R. (2001) A computational approach to identify genes for functional RNAs in genomic sequences. *Nucleic Acids Res.*, **29**, 3928–3938.
- Chen,S., Lesnik,E.A., Hall,T.A., Sampath,R., Griffey,R.H., Ecker,D.J. and Blyn,L.B. (2002) A bioinformatics based approach to discover small RNA genes in the *Escherichia coli* genome. *Biosystems*, **65**, 157–177.
- Tjaden,B., Saxena,R.M., Stolyar,S., Haynor,D.R., Kolker,E. and Rosenow,C. (2002) Transcriptome analysis of *Escherichia coli* using high-density oligonucleotide probe arrays. *Nucleic Acids Res.*, **30**, 3732–3738.
- Hershberg,R., Altuvia,S. and Margalit,H. (2003) A survey of small RNA-encoding genes in *Escherichia coli*. *Nucleic Acids Res.*, **31**, 1813–1820.
- Hsu,L.M., Zagorski,J., Wang,Z. and Fournier,M.J. (1985) *Escherichia coli* 6S RNA gene is part of a dual-function transcription unit. *J. Bacteriol.*, **161**, 1162–1170.
- Cavaille,J., Buiting,K., Kieffmann,M., Lalande,M., Brannan,C.I., Horsthemke,B., Bachelier,J.P., Brosius,J. and Hüttenhofer,A. (2000) Identification of brain-specific and imprinted small nuclear RNA genes exhibiting an unusual genomic organization. *Proc. Natl Acad. Sci. USA*, **97**, 14311–14316.
- Hüttenhofer,A., Kieffmann,M., Meier-Ewert,S., O'Brien,J., Lehrach,H., Bachelier,J.P. and Brosius,J. (2001) RNomics: an experimental approach that identifies 201 candidates for novel, small, non-messenger RNAs in mouse. *EMBO J.*, **20**, 2943–2953.
- Marker,C., Zemann,A., Terhorst,T., Kieffmann,M., Kastenmayer,J.P., Green,P., Bachelier,J.P., Brosius,J. et al. (2002) Experimental RNomics. Identification of 140 candidates for small non-messenger RNAs in the plant *Arabidopsis thaliana*. *Curr. Biol.*, **12**, 2002–2013.
- Tang,T.H., Bachelier,J.P., Rozhdestvensky,T., Bortolin,M.L., Huber,H., Drungowski,M., Elge,T., Brosius,J. et al. (2002) Identification of 86 candidates for small non-messenger RNAs from the archaeon *Archaeoglobus fulgidus*. *Proc. Natl Acad. Sci. USA*, **99**, 7536–7541.
- Tang,T.H., Rozhdestvensky,T.S., d'Orval,B.C., Bortolin,M.L., Huber,H., Charpentier,B., Branlant,C., Bachelier,J.P. et al. (2002) RNomics in



- Archaea reveals a further link between splicing of archaeal introns and rRNA processing. *Nucleic Acids Res.*, **30**, 921–930.
17. Yuan, G., Klämbt, C., Bachellerie, J.P., Brosius, J. and Hüttenhofer, A. (2003) RNomics in *Drosophila melanogaster*: identification of 66 candidates for novel non-messenger RNAs. *Nucleic Acids Res.*, **31**, 2495–2507.
  18. Blattner, F.R., Plunkett, G., III, Bloch, C.A., Perna, N.T., Burland, V., Riley, M., Collado-Vides, J., Glasner, J.D. *et al.* (1997) The complete genome sequence of *Escherichia coli* K-12. *Science*, **277**, 1453–1474.
  19. Zuker, M. (2003) Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res.*, **31**, 3406–3415.
  20. Dasgupta, S., Fernandez, L., Kameyama, L., Inada, T., Nakamura, Y., Pappas, A. and Court, D.L. (1998) Genetic uncoupling of the dsRNA-binding and RNA cleavage activities of the *Escherichia coli* endoribonuclease RNase III—the effect of dsRNA binding on gene expression. *Mol. Microbiol.*, **28**, 629–640.
  21. Blomberg, P., Wagner, E.G.H. and Nordström, K. (1990) Control of replication of plasmid R1: the duplex between the antisense RNA, CopA and its target, CopT, is processed specifically *in vivo* and *in vitro* by RNase III. *EMBO J.*, **9**, 2331–2340.
  22. Söderbom, F., Binnie, U., Masters, M. and Wagner, E.G.H. (1997) Regulation of plasmid R1 replication: PcnB and RNase E expedite the decay of the antisense RNA, CopA. *Mol. Microbiol.*, **26**, 493–504.
  23. Bensing, B.A., Meyer, B.J. and Dunny, G.M. (1996) Sensitive detection of bacterial transcription initiation sites and differentiation from RNA processing sites in the pheromone-induced plasmid transfer system of *Enterococcus faecalis*. *Proc. Natl Acad. Sci. USA*, **93**, 7794–7799.
  24. Selinger, D.W., Cheung, K.J., Mei, R., Johansson, E.M., Richmond, C.S., Blattner, F.R., Lockhart, D.J. and Church, G.M. (2000) RNA expression analysis using a 30 base pair resolution *Escherichia coli* genome array. *Nature Biotechnol.*, **18**, 1262–1268.
  25. Stern, M.J., Ames, G.F., Smith, N.H., Robinson, E.C. and Higgins, C.F. (1984) Repetitive extragenic palindromic sequences: a major component of the bacterial genome. *Cell*, **37**, 1015–1026.
  26. Rodionov, D.A., Vitreschak, A.G., Mironov, A.A. and Gelfand, M.S. (2002) Comparative genomics of thiamin biosynthesis in procaryotes. New genes and regulatory mechanisms. *J. Biol. Chem.*, **277**, 48949–48959.
  27. Webb, E., Claas, K. and Downs, D. (1998) *thiBPQ* encodes an ABC transporter required for transport of thiamine and thiamine pyrophosphate in *Salmonella typhimurium*. *J. Biol. Chem.*, **273**, 8946–8950.
  28. Hengge-Aronis, R. (2002) Stationary phase gene regulation: what makes an *Escherichia coli* promoter sigmaS-selective? *Curr. Opin. Microbiol.*, **5**, 591–595.
  29. Hosie, A.H. and Poole, P.S. (2001) Bacterial ABC transporters of amino acids. *Res. Microbiol.*, **152**, 259–270.
  30. Zhang, J.R. and Deutscher, M.P. (1989) Analysis of the upstream region of the *Escherichia coli* *rnd* gene encoding RNase D. Evidence for translational regulation of a putative tRNA processing enzyme. *J. Biol. Chem.*, **264**, 18228–18233.
  31. Freedman, R., Gibson, B., Donovan, D., Biemann, K., Eisenbeis, S., Parker, J. and Schimmel, P. (1985) Primary structure of histidine-tRNA synthetase and characterization of hisS transcripts. *J. Biol. Chem.*, **260**, 10063–10068.
  32. Mironov, M.S., Mironov, A.A., Jomantas, J., Kozlov, Y.I. and Perumov, D.A. (1999) A conserved RNA structure element involved in the regulation of bacterial riboflavin synthesis genes. *Trends Genet.*, **15**, 439–442.
  33. Mironov, A.S., Gusarov, I., Rafikov, R., Lopez, L.E., Shatalin, K., Kreneva, R.A., Perumov, D.A. and Nudler, E. (2002) Sensing small molecules by nascent RNA: a mechanism to control transcription in bacteria. *Cell*, **111**, 747–756.
  34. Winkler, W.C., Cohen-Chalamish, S. and Breaker, R.R. (2002) An mRNA structure that controls gene expression by binding FMN. *Proc. Natl Acad. Sci. USA*, **99**, 15908–15913.
  35. Vitreschak, A.G., Rodionov, D.A., Mironov, A.A. and Gelfand, M.S. (2002) Regulation of riboflavin biosynthesis and transport genes in bacteria by transcriptional and translational attenuation. *Nucleic Acids Res.*, **30**, 3141–3151.
  36. Raina, S. and Georgopoulos, C. (1990) A new *Escherichia coli* heat shock gene, *htrC*, whose product is essential for viability only at high temperatures. *J. Bacteriol.*, **172**, 3417–3426.
  37. Wagner, E.G.H. and Brantl, S. (1998) Kissing and RNA stability in antisense control of plasmid replication. *Trends Biochem. Sci.*, **23**, 451–454.
  38. Wagner, E.G.H., Altuvia, S. and Romby, P. (2002) Antisense RNAs in bacteria and their genetic elements. *Adv. Genet.*, **46**, 361–398.
  39. Selinger, D.W., Saxena, R.M., Cheung, K.J., Church, G.M. and Rosenow, C. (2003) Global RNA half-life analysis in *Escherichia coli* reveals positional patterns of transcript degradation. *Genome Res.*, **13**, 216–223.
  40. Bernstein, J.A., Khodursky, A.B., Lin, P.H., Lin-Chao, S. and Cohen, S.N. (2002) Global analysis of mRNA decay and abundance in *Escherichia coli* at single-gene resolution using two-color fluorescent DNA microarrays. *Proc. Natl Acad. Sci. USA*, **99**, 9697–9702.
  41. Urbanowski, M.L., Stauffer, L.T. and Stauffer, G.V. (2000) The *gcvB* gene encodes a small untranslated RNA involved in expression of the dipeptide and oligopeptide transport systems in *Escherichia coli*. *Mol. Microbiol.*, **37**, 856–868.
  42. Hindley, J. (1967) Fractionation of <sup>32</sup>P-labelled ribonucleic acids on polyacrylamide gels and their characterization by fingerprinting. *J. Mol. Biol.*, **30**, 125–136.
  43. Ikemura, T. and Dahlberg, J.E. (1973) Small ribonucleic acids of *Escherichia coli*. I. Characterization by polyacrylamide gel electrophoresis and fingerprint analysis. *J. Biol. Chem.*, **248**, 5024–5032.
  44. Elbashir, S.M., Lendeckel, W. and Tuschl, T. (2001) RNA interference is mediated by 21- and 22-nucleotide RNAs. *Genes Dev.*, **15**, 188–200.
  45. Lagos-Quintana, M., Rauhut, R., Lendeckel, W. and Tuschl, T. (2001) Identification of novel genes coding for small expressed RNAs. *Science*, **294**, 853–858.
  46. Lau, N.C., Lim, L.P., Weinstein, E.G. and Bartel, D.P. (2001) An abundant class of tiny RNAs with probable regulatory roles in *Caenorhabditis elegans*. *Science*, **294**, 858–862.
  47. Lee, R.C. and Ambros, V. (2001) An extensive class of small RNAs in *Caenorhabditis elegans*. *Science*, **294**, 862–864.
  48. Altuvia, S., Weinstein-Fischer, D., Zhang, A., Postow, L. and Storz, G. (1997) A small, stable RNA induced by oxidative stress: role as a pleiotropic regulator and antimutator. *Cell*, **90**, 43–53.
  49. Sledjeski, D. and Gottesman, S. (1995) A small RNA acts as an antisilencer of the H-NS-silenced *rcaA* gene of *Escherichia coli*. *Proc. Natl Acad. Sci. USA*, **92**, 2003–2007.
  50. Massé, E. and Gottesman, S. (2002) A small RNA regulates the expression of genes involved in iron metabolism in *Escherichia coli*. *Proc. Natl Acad. Sci. USA*, **99**, 4620–4625.
  51. Weilbacher, T., Suzuki, K., Dubey, A.K., Wang, X., Gudapaty, S., Morozov, I., Baker, C.S., Georgellis, D. *et al.* (2003) A novel sRNA component of the carbon storage regulatory system of *Escherichia coli*. *Mol. Microbiol.*, **48**, 657–670.
  52. Gottesman, S. (2002) Stealth regulation: biological circuits with small RNA switches. *Genes Dev.*, **16**, 2829–2842.
  53. Wagner, E.G.H. and Flärdh, K. (2002) Antisense RNAs everywhere? *Trends Genet.*, **18**, 223–226.
  54. Pedersen, K., Zavialov, A.V., Pavlov, M.Y., Elf, J., Gerdes, K. and Ehrenberg, M. (2003) The bacterial toxin RelE displays codon-specific cleavage of mRNAs in the ribosomal A site. *Cell*, **112**, 131–140.
  55. Mattick, J.S. and Gagen, M.J. (2001) The evolution of controlled multitasked gene networks: the role of introns and other noncoding RNAs in the development of complex organisms. *Mol. Biol. Evol.*, **18**, 1611–1630.
  56. Winkler, W., Nahvi, A. and Breaker, R.R. (2002) Thiamine derivatives bind messenger RNAs directly to regulate bacterial gene expression. *Nature*, **419**, 952–956.
  57. Novick, R.P., Ross, H.F., Projan, S.J., Kornblum, J., Kreiswirth, B. and Moghazeh, S. (1993) Synthesis of staphylococcal virulence factors is controlled by a regulatory RNA molecule. *EMBO J.*, **12**, 3967–3975.
  58. Eddy, S.R. (2001) Non-coding RNA genes and the modern RNA world. *Nature Rev. Genet.*, **2**, 919–929.
  59. Rudd, K.E. (1999) Novel intergenic repeats of *Escherichia coli* K-12. *Res. Microbiol.*, **150**, 653–664.