Novel small RNA-encoding genes in the intergenic regions of *Escherichia coli*

Liron Argaman*‡, Ruth Hershberg*‡, Jörg Vogel†‡, Gill Bejerano*, E. Gerhart H. Wagner†, Hanah Margalit* and Shoshy Altuvia*

Background: Small, untranslated RNA molecules were identified initially in bacteria, but examples can be found in all kingdoms of life. These RNAs carry out diverse functions, and many of them are regulators of gene expression. Genes encoding small, untranslated RNAs are difficult to detect experimentally or to predict by traditional sequence analysis approaches. Thus, in spite of the rising recognition that such RNAs may play key roles in bacterial physiology, many of the small RNAs known to date were discovered fortuitously.

Results: To search the *Escherichia coli* genome sequence for genes encoding small RNAs, we developed a computational strategy employing transcription signals and genomic features of the known small RNA-encoding genes. The search, for which we used rather restrictive criteria, has led to the prediction of 24 putative sRNA-encoding genes, of which 23 were tested experimentally. Here we report on the discovery of 14 genes encoding novel small RNAs in *E. coli* and their expression patterns under a variety of physiological conditions. Most of the newly discovered RNAs are abundant. Interestingly, the expression level of a significant number of these RNAs increases upon entry into stationary phase.

Conclusions: Based on our results, we conclude that small RNAs are much more widespread than previously imagined and that these versatile molecules may play important roles in the fine-tuning of cell responses to changing environments.

Addresses: *Department of Molecular Genetics and Biotechnology, The Hebrew University-Hadassah Medical School, Jerusalem 91120, Israel. †Institute of Cell and Molecular Biology, Biomedical Center, Uppsala University, Box 596, Uppsala 751 24, Sweden.

Correspondence: Shoshy Altuvia, Hanah Margalit, Gerhart Wagner
E-mail: shoshy@cc.huji.ac.il
hanah@md2.huji.ac.il
gerhart.wagner@icm.uu.se

‡These authors contributed equally to this work.

Received: **2 May 2001**
Revised: **21 May 2001**
Accepted: **23 May 2001**

Published: **26 June 2001**

Current Biology 2001, 11:941–950

0960-9822/01/$ – see front matter

Background

Small, untranslated RNAs are present in many different organisms, ranging from bacteria to mammals. These RNAs carry out a variety of biological functions. Many of them act as regulators of gene expression at a posttranscriptional level, either by acting as antisense RNAs, by binding to complementary sequences of target transcripts, or by interacting with proteins. Regulatory RNAs are involved in the control of a large variety of processes such as plasmid replication, transposition in pro- and eukaryotes, phage development, viral replication, bacterial virulence, global circuits in bacteria in response to environmental changes, or developmental control in lower eukaryotes [1–7]. The biological roles of other RNA species involve different aspects of metabolism. Such aspects include protein secretion, tRNA processing, splicing, and rRNA biogenesis [8–11].

Small, untranslated RNA species have been difficult to detect by experimental procedures or by traditional computational approaches. Indeed, the completion of the *Escherichia coli* genome sequence has led to the prediction of about 4290 protein-encoding genes but has added very little new information regarding possible RNA-encoding genes [12]. Thus, in spite of the rising recognition that RNAs are of major importance, many of these RNAs were discovered by chance while researchers were studying individual genetic systems. In the best-characterized bacterium, *E. coli*, only a limited number of chromosomally encoded small RNA (sRNA) molecules have so far been identified. The first bacterial sRNA, 6S RNA, was discovered more than three decades ago [13, 14], but its biological role was unknown until recently [15]. Since then, only eleven additional sRNAs have been identified, two of them during the last year [reviewed in 6, 16, 17]. Seven of these RNAs were fortuitously discovered either as genomic fragments whose expression modulated certain activities (RprA [17]; MicF [18]; DicF [19]; and DsrA [20]), under conditions suggesting possible functions (CsrB [21]), or by the use of DNA fragments that overlapped the RNA-encoding genes and simultaneous probing of the adjacent gene (GcvB [16]; OxyS [22]). The remaining four were discovered by orthophosphate labeling of total RNA (4.5S and 6S [13]) or as bands on two-dimensional gels (Spot 42, 10Sa (tmRNA), and 10Sb, now designated M1, the RNA component of RNase P [23]).

The availability of the complete sequence of the *E. coli* genome [12] prompted us to develop a systematic search for sRNA-encoding genes in *E. coli*. Because sRNAs are

untranslated, and since secondary-structural elements were shown to be insufficient for distinguishing nontranslated RNAs from random sequences [24], the computational screen for these genes had to incorporate different considerations based on the genomic and sequence features of the known molecules. We explored the location of the genes encoding the known sRNAs and found them to be located primarily in "empty" intergenic regions, with no other annotated genes on either strand. Another interesting feature of the known sRNAs emerged from phylogenetic comparisons. Most known sRNA sequences were found to be conserved in some of the closely related members of the *Enterobacteriaceae* whose genome sequences were available, e.g. *Salmonella typhimurium*, *Klebsiella pneumoniae*, and *Yersinia pestis*. In addition, since candidates for novel sRNAs cannot be identified by conventional searches for open reading frames, we focused on transcription signals by searching for promoter sequences within a short distance upstream of a terminator. We exploited the characteristic genomic features along with the transcription initiation and termination signals to develop a predictive algorithm to search for genes encoding sRNAs within the *E. coli* genome. The screen resulted in the prediction of 24 putative sRNA-encoding genes, of which 23 were tested experimentally. Here we report on the discovery of 14 genes encoding novel sRNA molecules and their expression patterns under a variety of physiological conditions.

Results

Computational approach

The best approach in developing a predictive scheme is to gain knowledge from already available data and to apply this knowledge in prediction. The limited number of already known sRNAs, ten in total at the time of the analysis, directed us to take a heuristic rather than an automatic machine-learning approach. We characterized the sequence and genomic features of these RNAs and incorporated the identified characteristics in a predictive scheme by employing the following principles: (1) We focused on "empty" intergenic regions, defined between annotated genes based on the Colibri database (http://genolist.pasteur.fr/Colibri/). (2) Within these regions, we searched for transcription initiation and termination signals that are widely used by the *E. coli* transcription machinery and that were also observed in the known sRNA genes. We focused on promoter DNA sequences recognized by the major *E. coli* RNA polymerase sigma factor, σ^{70}, and on Rho-independent terminators, in which the termination signal resides in specific sequence and structural features of the RNA. (3) Among the predicted sequences, we chose those in which the distance between the predicted promoter and terminator was 50–400 base pairs. (4) The predicted sequences obtained were compared to genome sequences of other bacteria, and only those that showed significant conservation were selected (see Materials and methods for details). The mutual incorporation of the different criteria narrowed down the list of candidates and resulted in the prediction of a total of 24 putative sRNA genes (Table 1). We denoted these candidate genes *psrA1-psrA24* (predicted small RNA). Two genes predicted to encode sRNAs by our algorithm (*psrA5* and *psrA11*) were, during the course of this study, reported in the literature as *rprA* and *gcvB*, respectively [16, 17].

Experimental identification of small RNAs

Guided by prediction, searches for the corresponding RNA species were carried out in *E. coli* K12 cells grown to different growth phases in either rich or minimal media supplemented with glycerol and in cells subjected to heat shock or cold shock treatment (see below). Isolated total RNA was separated on urea-polyacrylamide gels and transferred to nylon membranes. The RNAs in question were probed with end-labeled oligodeoxyribonucleotides designed to be complementary to the suitable regions within the RNAs. We tested 23 genes, predicted to encode sRNAs by the screen (Table 1), and discovered 14 novel sRNAs. The genes expressing sRNAs were denoted *sra* (A, B, etc.) for small RNA. Figure 1 shows the characterization of these RNAs with respect to approximate size, abundance, and expression pattern. The majority of the genes were expressed during stationary phase. Some of the RNAs, i.e., SraB, RprA, SraI, and SraL, were expressed in stationary phase only and reached their highest levels at 8 and 10 hr after dilution of the culture. Others, such as SraC, SraD, SraE, SraH, and SraK, were highly abundant in stationary phase, but low levels could be detected in exponentially growing cells as well (Figure 1). The transcript levels of *sraA* seemed to be invariable under all conditions. The *sraJ* and *gcvB* genes were expressed in early logarithmic phase, but their levels decreased with cell growth. SraG increased in late logarithmic phase, and steady-state levels remained high even 2 and 4 hr after dilution but decreased thereafter. The RNAs of three other genes (*psrA1*, *psrA6*, and *psrA9*) were found to be less abundant and were therefore excluded from further characterization.

Our computer search employed the consensus recognition sequence of the major sigma factor, σ^{70}. However, the regulation of gene expression may involve, in addition to σ^{70}, stress-specific transcription factors and/or alternative sigma factors. Therefore, we examined whether any of the predicted sRNA-encoding genes were induced by temperature-dependent stress conditions such as heat or cold shock. We found that the steady-state levels of SraF, SraG, and SraJ were increased by cold shock treatment; while *sraG* and *sraJ* expression was only slightly elevated, SraF RNA was almost exclusively present during cold shock. None of the other RNAs was affected by heat or cold shock treatment at early growth, neither those

Table 1

Small RNA-encoding genes predicted based on conservation and transcription signals.

Candidate	Adjacent genes	Strand*	5′ end†	3′ end‡	Length§ (nucleotides)
psrA1	*tsf/pyrH*	→ ← →	191793	191713	80
psrA2	*bolA/tig*	→ ← →	454262	454066	196
psrA3	*clpX/lon*	→ ← →	458008	457952	56
psrA4	*yceF/yceD*	← → →	1145859-1145923	1145977	54–118
psrA5¶	*ydiK/ydiL*	→ → →	1768291-1768414	1768502	88–211
psrA6¶	*yeaA/gapA*	← ← →	1860740-1860782	1860608	132–174
psrA7	*fadD/yeaY*	← → ←	1887849	1887959	110
psrA8¶	*pphA/yebY*	← → ←	1921078-1921118	1921360	242–282
psrA9¶	*cysK/ptsH*	→ ← →	2531601-2531608	2531422	179–186
psrA10¶	*ygaG/gshA*	← → ←	2812792	2812897	105
psrA11¶	*gcvA/ydgI*	← → ←	2940621-2940818	2940924	106–303
psrA12	*aas/galR*	← ← →	2974257	2974123	134
psrA13	*tktA/yggG*	← → →	3079665	3079899	234
psrA14	*ygjR/ygjT*	→ → →	3235947-3236013	3236205	192–258
psrA15¶	*pnp/rpsO*	← → ←	3308823-3308868	3309040	172–217
psrA16¶	*elbB/arcB*	← → ←	3348155-3348219	3348339	120–184
psrA17	*yheO/fkpA*	← ← ←	3474144	3474078	66
psrA18¶	*yhhX/yhhY*	← ← →	3578646-3578692	3578552	94–140
psrA19	*ivbL/ysdA*	← ← →	3850884-3850913	3850744	140–169
psrA20	*aslA/hemY*	← → ←	3983890-3984048	3984252	204–362
psrA21	*yihA/yihI*	← → →	4048554	4048858	304
psrA22¶	*yihA/yihI*	← ← →	4048893-4048916	4048823	70–93
psrA23¶	*rhaT/sodA*	← ← →	4098318	4098266	52
psrA24	*soxR/yjcD*	→ ← →	4275644-4275686	4275504	140–182

* The middle arrow represents the sRNA gene, while the flanking arrows indicate the orientation of the adjacent genes, respectively. Genes present on the strand given in the *E. coli* genome database are indicated by (→), and genes present on the complementary strand are indicated by (←).

† The position given is 7 bases downstream of the 3′ end of the −10 hexamer. For candidates with more than one putative promoter, the positions of the far-left and the far-right 5′ ends are given.

‡ Given is the position of the last uridine at the end of the terminator.

§ Range of possible lengths based on putative 5′ and 3′ ends.

¶ Conserved in *Salmonella, Klebsiella pneumoniae,* and *Yersinia pestis*. Unmarked candidates are conserved only in *Salmonella* and *Klebsiella pneumoniae*.

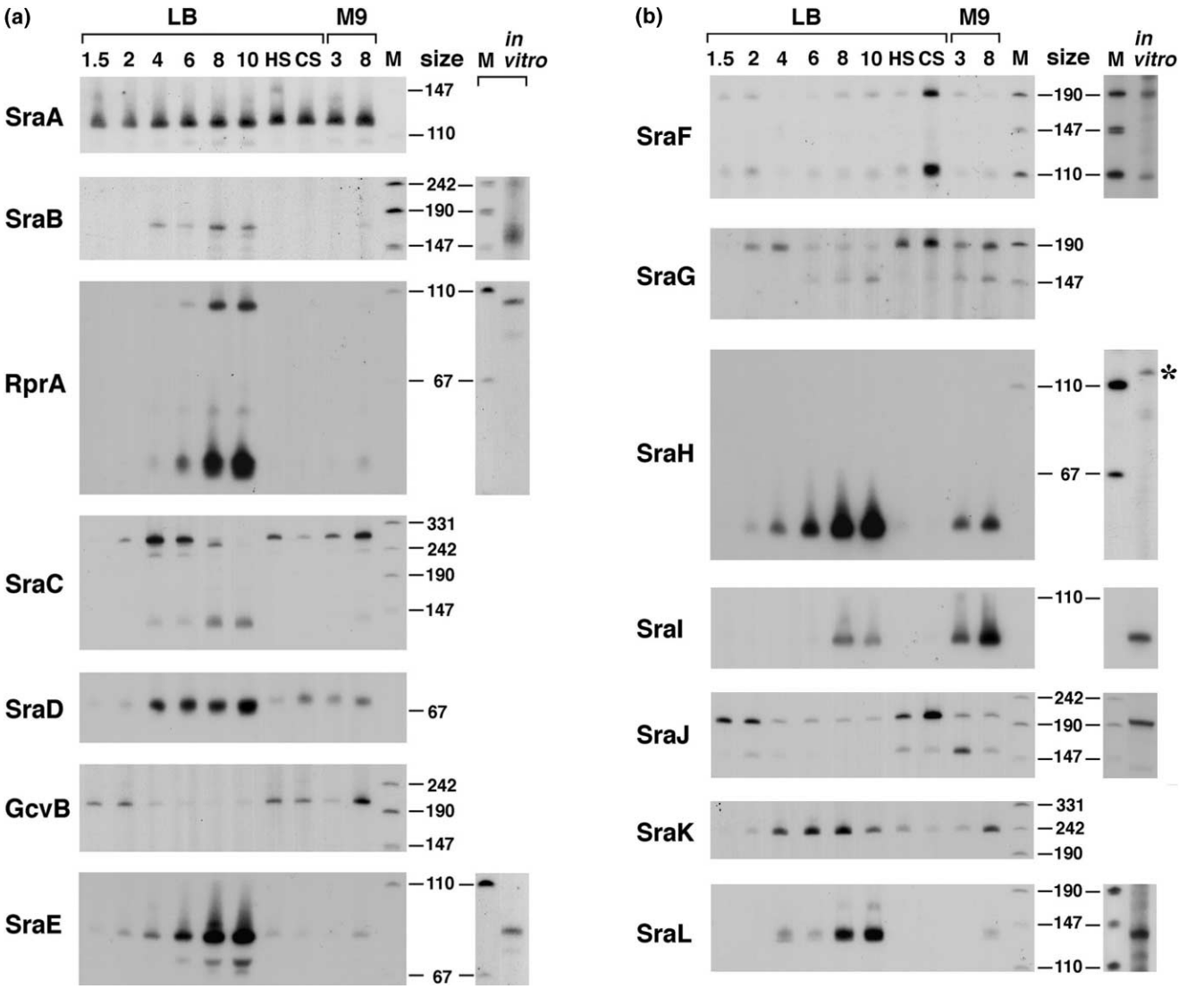
detected in logarithmic phase (SraC, SraD, GcvB, SraE, and SraK) nor those present predominantly in late stationary phase (SraB, RprA, SraH, SraI, and SraL; Figure 1).

Because a significant number of regulons in *E. coli* are subject to catabolite repression by glucose, we examined the expression of the predicted genes in cells grown in minimal medium supplemented with glycerol. None of the predicted genes was found to be specifically expressed under these conditions. Moreover, the expression pattern of the majority of the identified RNA genes was not affected upon growth in glycerol minimal medium (Figure 1). The only strong increase in this medium was observed for SraI. Minor increases in expression levels were detected for RNAs SraC and SraH in logarithmic phase and for GcvB in stationary phase. It is also noteworthy that the growth in minimal medium affected the ratio of processing of SraJ RNA (see below).

Mapping of the 5′ and 3′ ends of sRNAs

The observation that some sRNA genes display transcript lengths different from the ones predicted and that several sRNAs were present as multiple bands prompted us to determine the 5′ and 3′ ends of the respective RNAs. 5′ ends of the RNAs were determined by primer extension. In addition, to distinguish primary transcript 5′ ends from internal 5′ processing sites, we analyzed the RNAs by using 5′ RACE (rapid amplification of cDNA ends), with or without prior treatment by tobacco acid pyrophosphatase (TAP [25]). This enzyme converts the 5′ triphosphate of an RNA to a monophosphate and thereby enables the RNA to be ligated to an adapter, which allows specific amplification of the 5′-end sequence subsequent to reverse transcription. Duplicate samples of total RNA were either TAP- or mock-treated, followed by reverse transcription, PCR amplification, cloning, and sequence determination. Amplification products that were only obtained from TAP-treated samples indicated that they contained the transcription initiation point. 3′-RACE analysis was performed by the ligation of an adapter to the 3′ hydroxyl group of RNAs, followed by gene- and adapter-specific amplification. An example of such an analysis, carried out on SraL, is shown in Figure 2. The gene *sraL* is flanked by two reading frames, *soxR* and *yjcD*, both of which are transcribed in the direction opposite to that of *sraL*. The predicted terminator of *sraL* appears to be bidirectional and to encode stretches of uridines at both ends of the inverted repeats. The sequence characteristics

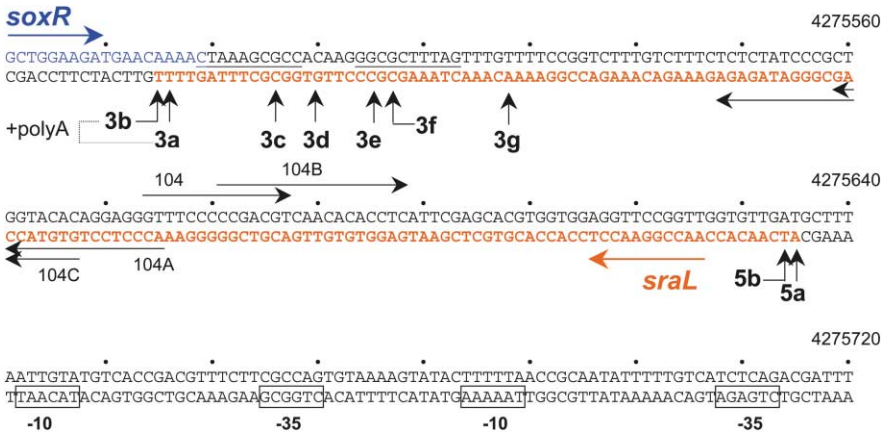
Figure 1



Detection of novel sRNAs by Northern analysis. Total RNA was isolated from *E. coli* K12 cells grown in either rich (LB; 1.5, 2, 4, 6, 8, and 10 hrs after dilution) or minimal media supplemented with glycerol (M9; 3 and 8 hr after dilution) and from cells subjected to heat shock (HS; 42°C for 15 min) or cold shock (CS; 15°C for 30 min) treatment. In vitro transcription of sRNA-encoding genes is presented on the right side of each panel. PCR-generated DNA fragments carrying the sRNA genes were transcribed in vitro by *E. coli* RNA polymerase, and the products were analyzed by Northern analysis. The full-length product of SraH, visible in the in vitro assay, is indicated by the asterisk. 5′ end-labeled, MspI-digested pUC19 DNA was used as a molecular-weight size marker (M).

and the position of the terminator suggest that this terminator functions for both *sraL* and *soxR*. Several possible promoters were predicted for the *sraL* gene positioned at a distance ranging from 140 to 182 base pairs upstream of the terminator (Table 1 and Figure 2). Experimental analysis of the SraL RNA mapped the 5′ end to two consecutive nucleotide positions corresponding to the promoter predicted closest to the terminator. Since 5′-RACE products could be amplified after TAP treatment only, we infer that they identify authentic transcription start sites. From 12 independent sequences obtained, 10 mapped to the A residue (labeled 5a) and two to the adjacent nucleotide (5b). Analysis of the major 3′-RACE fragment placed the 3′ ends at the two last uridines of the predicted terminator (3a, 3b, total of seven sequences). Surprisingly, all of these seven sequences contained nonencoded A residues, ranging from two to seven in number. This finding indicates that the RNA undergoes frequent polyadenylation. The slight size heterogeneity of SraL observed on Northern blots can thus be tentatively explained by the heterogeneous A tails detected in the 3′-RACE analysis. The sequencing of five independent clones of minor 3′-RACE fragments identified additional 3′ ends, scattered throughout the terminator stem loop region and

Figure 2



5′- and 3′-end mapping of SraL RNA. Red color indicates the DNA sequence of SraL RNA, whereas blue color indicates that of *soxR* mRNA. The position of the two sets of gene-internal primers (104, 104A–C) used in 5′- and 3′-RACE reactions are indicated by arrows. The lines between the top and bottom strands indicate the inverted repeat sequences of the bidirectional terminator. Several promoters were predicted; the −35 and −10 hexamers of the σ^{70} consensus of two such promoters, the far-left and the far-right putative promoters, are shown (boxed). 5′ ends, 5a and 5b, that were identified only in TAP-treated samples indicate the transcription initiation site. 3a and 3b define the transcription termination site. The additional 3′ ends (3c–g) that were identified in minor 3′-RACE products most likely represent degradation intermediates. Nonencoded poly(A) tails of different lengths were identified on transcripts terminated at 3a and 3b. The numbering indicates the position in the *E. coli* genome database. ORF *yjcD* is located downstream of the area shown and starts at 4276048.

most likely representing degradation intermediates (see 3c–g in Figure 2). Taken together, the transcript length of SraL, as calculated from the major 5′ and 3′ endpoints (5a, 3a), is of 139–141 nucleotides, consistent with its approximate size estimated from Northern blots, in vitro transcription, and prediction.

The same approach was used for most of the other RNAs. The major 5′ and 3′ ends determined by RACE analysis and primer extension are summarized in Table 2. In several cases, the RNA blots indicated the presence of other RNA sizes in addition to the full-length transcripts. For RprA, SraC, SraG, and SraJ, specific cleavage products were observed. The transcription start sites of RprA, SraC, GcvB, SraF, SraG, SraH, SraI, SraJ, and SraL were mapped to one out of the several possible promoters predicted by the algorithm (Tables 1 and 2). The results of the initiation point analyses were in agreement with the major full-length products observed except for SraH, in which the major RNA species is a stable cleavage product originating from a primary transcript of approximately 120 nucleotides. The full-length transcript was observed in overexposed autoradiograms (not shown) and in in vitro transcription assays (see below). A mapping of SraC indicated that a second low-level transcript originates from a second promoter located 36 nucleotides downstream of its major promoter. The transcripts of SraA, SraD, SraE, and SraK were found to be longer or shorter than predicted, and the mapping data assigned the initiation points to promoters with a less perfect match to the σ^{70} consensus sequence. More complex patterns were found for SraB and SraF. Transcription of the gene *sraF* resulted in two products of 110 and 189 nucleotides. The sequence of *sraF* seems to contain a weak transcription termination-like structure located approximately 110 nucleotides downstream of the start site. However, because the 110 nucleotide product was absent in an *rnc* (RNase III) mutant strain (data not shown), we suspect that this structure is recognized by RNase III and that the 110 nucleotide RNA is a cleavage product. The mapping of SraB RNA indicated that the sequence encoding the 168 nucleotide RNA carries an internal promoter that directs the synthesis of a 105 nucleotide RNA in both logarithmic- and stationary-phase cells (data not shown). In contrast to the 168 nucleotide product, under logarithmic growth conditions some of the transcripts of the 105 nucleotide RNA extend beyond the termination signal. This results in a longer RNA product, possibly representing a leader of the downstream RNA (not shown).

The 3′-end mapping of the identified sRNAs further confirmed that the RNA transcripts end at the predicted termination signal. Interestingly, some RNA molecules, mainly those that were subject to internal cleavage, such as SraC, SraH, SraJ, and SraL, exhibited 3′ ends at positions within or upstream of the terminator stem loops. This finding is consistent with 3′ exonucleolytic degradation. 3′-end analysis of SraK demonstrated that this RNA carries poly(A) tails of different lengths, as shown for SraL RNA. The sequence of *gcvB* was shown to contain two sites for transcription termination [16]. We found that most of the *gcvB* transcripts read through the first terminator and stop at the second one, and thus give an RNA product of 205 nucleotides (Figure 1 and Table 2).

Table 2

Summary of the newly identified small RNAs.

sRNA gene	Minutes	Adjacent genes	Strand[1]	5′ end[2]	3′ end[3]	Length[4,5,6]	Comments
sraA (*psrA3*)	9.9	*clpX/lon*	→ ← →			~120[5]	
sraB (*psrA4*)	24.7	*yceF/yceD*	← → →	1145812	1145961-1145980	149–168[4]	
rprA (*psrA5*)[7]	38.1	*ydiK/ydiL*	→ → →	1768396	1768500	105	Cleaved by endoribonuclease
sraC (*psrA8*)[7]	41.4	*pphA/yebY*	← → ←	1921090	1921323-1921338	234–249[4]	Cleaved by RNase III
sraD (*psrA10*)[7]	60.6	*ygaG/gshA*	← → ←	2812822		~70[5]	
gcvB (*psrA11*)[7]	63.4	*gcvA/ydgI*	← → ←	2940718	2940922	205	
sraE (*psrA12*)	64.1	*aas/galR*	← ← →	2974211	2974124	88	
sraF (*psrA14*)	69.8	*ygjR/ygjT*	→ → →	3236015	3236203	189	Cleaved by RNase III
sraG (*psrA15*)[7]	71.3	*pnp/rpsO*	← → ←	3308866	3309011-3309039	146–174[4,6]	3′ end with poly(A)
sraH (*psrA16*)[7]	72.2	*elbB/arcB*	← → ←	3348218	3348305-3348325	88–108[4]	Processed at 3348283 giving the observed 42nt RNA
sraI (*psrA18*)[7]	77.1	*yhhX/yhhY*	← ← →	3578647	3578554-3578557	91–94	
sraJ (*psrA20*)	85.9	*aslA/hemY*	← → ←	3984045	3984216	172[4]	Cleaved by RNase III
sraK (*psrA21*)	87.3	*yihA/yihI*	← → →	4048616	4048860	245[6]	3′ end with poly(A)
sraL (*psrA24*)	92.2	*soxR/yjcD*	→ ← →	4275645	4275506	140[6]	3′ end with poly(A)

[1] The middle arrow represents the sRNA gene, while the flanking arrows indicate the orientation of the adjacent genes, respectively. Genes present on the strand given in the *E. coli* genome database (→), and genes present on the complementary strand (←).
[2] Determined by primer extension or by both 5′ RACE and primer extension.
[3] Determined by 3′ RACE.
[4] The full-length RNA is shorter than expected due to 3′ end trimming. Given is the range of sizes obtained.
[5] RNA lengths estimated from Northern blots.
[6] The RNA length calculated based on the mapping does not include the non-encoded poly(A) tails.
[7] Conserved in *Salmonella, Klebsiella pneumoniae* and *Yersinia pestis*. Unmarked candidates are conserved only in *Salmonella* and *Klebsiella pneumoniae*.

In vitro transcription of the sRNA-encoding genes

The experimental identification of an RNA cannot guarantee that it is an independent transcript since an sRNA might be a leader of a downstream mRNA or a processed trailer of the preceding one, depending on the orientation of adjacent genes. In addition, the expression of a number of sRNAs in *E. coli* is induced by specific transcription factors (e.g., *oxyS*, *gcvB* [22, 16]). To examine the activity of the proposed promoters and to further establish transcription initiation and termination sites as opposed to processing, we assayed transcription of the genes in vitro. PCR-generated DNA fragments carrying eight of the sRNA genes, including promoter and terminator signals flanked by approximately 50 base pairs, were transcribed in vitro by *E. coli* RNA polymerase. The promoters of the genes *sraB*, *rprA*, *sraE*, *sraF*, *sraI*, *sraJ*, and *sraL* were found to be active in vitro, and transcript lengths similar to those observed in vivo were obtained. These results indicate that complete transcription units were contained within the DNA fragments used. In contrast, the template carrying *sraC* failed to yield an RNA product. Thus, transcription of this RNA may require a transcription activator or an alternative sigma factor. In vitro transcription of *sraH* resulted in a product of approximately 120 nucleotides, corresponding to the predicted full-length RNA. This finding further indicated that the short RNA observed in vivo is a cleavage product (see above). Interestingly, we also observed that the secondary transcription termination-like structure within gene *sraF*, which is subject to RNase III cleavage in vivo, is active as a terminator in vitro.

Discussion

The systematic genome search has led to the prediction of 24 genes encoding small RNA molecules in *E. coli*. We experimentally examined 23 putative genes and identified 14 new sRNA-encoding genes. Here we present a characterization of these novel RNA molecules. The experimental evidence suggests that they are independently expressed RNAs rather than the read-through products of preceding genes or processed leaders of downstream genes. Our findings strongly suggest that sRNAs play a prominent role in bacterial physiology. Given that the expression pattern of most of the newly discovered RNA molecules changes with bacterial growth conditions, it is conceivable that these sRNAs have regulatory roles rather than housekeeping functions.

The algorithm employed four different criteria to search for putative sRNA genes. Since not all known sRNAs exhibited all of the described characteristics in concert, the mutual incorporation of the different criteria in the predictive scheme automatically excluded a fraction of the known sRNAs. The screen has identified four out of the ten known genes encoding sRNAs. The remaining six escaped detection either because their primary transcripts were larger than 400 nucleotides (*ssrA*) or showed conservation to *Salmonella* only (*micF*) or because their transcription initiation (*csrB*, *dicF*) and/or termination signals (*dicF*, *ffs*, *ssrS*) could not be detected [6]. Thus, it is conceivable that additional sRNA genes would have been predicted with less stringent requirements. Also, it is possible that other classes of sRNAs have been overlooked. All cur-

rently known plasmid-borne sRNAs are encoded at the same genetic loci as the target genes (*cis*-encoded) and act as antisense RNAs [5]. Unlike these antisense RNAs, the bacterial RNAs are encoded at genetic loci other than those of the target genes (*trans*-encoded [5, 26]). The restriction of the computer search to intergenic regions may have excluded the class of *cis*-encoded antisense RNAs because their genes usually overlap part of the target gene. Similarly, sRNA genes that rely on Rho-dependent terminators have escaped detection. In addition, sRNAs processed from transcripts longer than 400 nucleotides would not have been included because they lack the Rho-independent termination signal within the examined range. Finally, sRNA genes that rely on alternative sigma factors using different recognition sequences escaped detection.

The sequence and genomic features used for prediction can now be investigated in view of the experimental mapping of the new sRNAs. Based on our knowledge of the known sRNAs, we focused the search on empty intergenic regions that cover 12% of the genome sequence and whose sizes range up to approximately 3500 base pairs. Notably, most of the newly identified sRNAs were clustered in relatively short intergenic regions (up to 600 base pairs in length). It is interesting to look at the conservation pattern of the identified sRNAs in comparison to their flanking regions. Such a comparison shows two types of conservation patterns; either the entire region within which the sRNA gene resides is conserved (including flanking sequences and genes), or the sRNA gene stands out within its genomic surrounding (i.e., the sRNA gene is conserved, while flanking regions are not, and conservation may resume in the surrounding genes). An isolated conservation as in the latter case was observed for 10 out of the 14 experimentally characterized sRNAs described in this study, and for 8 out of the 10 previously known sRNAs. We believe that this pattern of conservation lends further support to the proposal that these sRNAs are encoded by autonomous transcription units.

The prediction of promoters and terminators was carried out by heuristic approaches that were developed based on two data sets of experimentally determined promoters and terminators [27, 28]. The accuracy of the terminator prediction is impressive. In all but one case (*gcvB*), the algorithm predicted a single terminator for any given candidate sRNA. In each case, this terminator was confirmed experimentally. In contrast, the promoter prediction yielded more redundant results and predicted several putative promoters for each sRNA. In most cases, one of these correlated with the experimentally determined transcription start site. Thus, development of an algorithm that can faithfully select the correct promoter from several candidate sequences remains a major challenge.

Unlike the very unstable plasmid-borne sRNAs, most of the known bacterial sRNAs that are either involved in control of gene expression in response to stress, such as OxyS, MicF, and DsrA, or in housekeeping, such as 6S, 4.5S, and tmRNA, were found to be relatively stable and/or abundant [6, 15]. Similar to these previous observations, we found that most of the newly identified sRNAs were abundant. The majority of the previously known sRNAs were found through phenotypic effects under certain conditions, i.e., a biological effect was evident at the time of their discovery. Here, we have identified new sRNAs based on criteria that did not include a functional bias, and thus the biological effects of these RNAs were not evident. Yet, we noticed that the expression patterns of the vast majority of the sRNAs were intriguing, and they suggest a role in the regulation of physiological responses. RNA abundance changed with growth conditions and, for the majority, the transcript levels increased at entry into stationary phase. Two sRNAs, SraF and SraI, were strongly induced by cold shock and in glycerol medium, respectively. A few of the RNAs were preferentially expressed in exponentially growing cells, either in early or late logarithmic phase. Taken together, these findings suggest that most of these sRNAs are candidates for regulators involved in cellular responses to environmental changes and/or growth conditions. At this point, target genes of these putative regulators are unknown, and current work is being aimed at computational and experimental identification of targets.

The Northern blot analyses have shown that several of the RNAs undergo specific processing. This has been additionally supported by RACE experiments (data not shown). It has previously been observed that the maturation of stable RNA molecules such as 4.5S, 6S, tmRNA and M1 RNA is initiated by 3′ endonucleolytic cleavage followed by complete 5′ maturation and 3′ exonucleolytic trimming [29]. We found that several of the identified RNAs, RprA, SraC, SraF, SraG, SraH, and SraJ, showed band patterns that suggest specific cleavage. The cleavages of SraC, SraF, and SraJ were *rnc* dependent (data not shown), and some RACE products were consistent with 3′ exonucleolytic trimming. In addition, we observed that the processed products contained 3'-terminal stretches of adenosines which, in a few cases, carried internal G or C residues. Polyadenylation of stable RNA precursors has been demonstrated in bacterial cells deficient in exoribonucleases [30]. In these cells RNA maturation is delayed, and polyadenylated RNA precursors accumulate. Also, several unstable antisense RNAs, such as RNAI of ColE1, CopA of R1, and Sok RNA of a plasmid killer locus, were found to be polyadenylated [31–34]. Since mutations in the gene encoding poly(A) polymerase result in extended antisense RNA half-lives, polyadenylation is considered to destabilize these transcripts. We found that at least two of the newly discovered RNAs, SraL and SraK, although

very abundant, contained short A-tails at their 3′ ends. Interestingly, these poly(A) tails were detected in wild-type cells in spite of the presence of the exoribonuclease activity. Whether polyadenylation plays a role in destabilizing the new sRNAs described here needs further investigation.

The availability of genome sequences presents a challenge to discover genetic elements that are not easily identified by traditional methods. Such elements include genes encoding untranslated RNAs. An elegant study addressing this challenge has been reported by Lowe and Eddy [35]. Using probabilistic modeling based on known snoRNAs, these authors have discovered new members of the snoRNA gene family in the yeast *Saccharomyces cerevisiae*. A systematic search has also led to the discovery of brain-specific snoRNA genes in the mouse and humans [36]. We have demonstrated that the employment of biological principles in a computer algorithm can guide a systematic identification of new RNAs encoded within a genome sequence. Such computational approaches can be applied to other bacterial genomes and possibly to higher organisms, provided that the genomic and sequence features characteristic of the sRNAs can be defined in an explicit manner. Here, using rather restrictive criteria, we have identified 14 new sRNA-encoding genes. Others have found sRNA-encoding genes by using similar approaches [37].

Conclusions

Over a period of about 30 years, only four bona fide regulatory RNAs have been discovered in *E. coli*. Here we report on the discovery of 14 novel small RNA-encoding genes and their expression patterns under a variety of physiological conditions. The remarkably high number of RNA-encoding genes in *E. coli*, their diverse expression patterns, and their high transcript levels suggest that sRNAs are more widespread than previously imagined and that these small, likely regulatory RNA molecules may play important roles in integrating cellular responses to changing environments. The specific physiological roles of the newly discovered genes in the regulatory circuits in bacteria are presently under investigation.

Materials and methods

Determination of "empty" regions

Intergenic empty regions were determined as regions without gene annotations on either of the two strands. Annotations of known *E. coli* genes were based on the Colibri database (http://genolist.pasteur.fr/Colibri/). This database includes annotations of all of the open reading frames, as well as tRNA, rRNA and the ten previously known sRNA genes.

Promoter prediction

The computational identification of promoter sequences has been regarded as a difficult problem because the signal is rather weak (e.g., [38–40]). Since at present no promoter prediction algorithm has proven superior to others, we decided to adopt a heuristic strategy that combines consensus and weight matrix considerations based on the sequences upstream of experimentally determined transcription start sites [27]. This training set included 354 start sites. For each of these mRNA start sites, a promoter was determined by consensus considerations. σ^{70} promoters are defined by two consensus hexamers,TATAAT and TTGACA, located 10 and 35 base pairs upstream of the transcription start site, respectively. We searched the regions upstream of the mRNA start sites for sequences with no more than three mismatches to the consensus in any of the two hexamers. This process often yielded several candidates. These candidates were ranked by a hirerachy of heuristic criteria, such as the number of matches to the promoter consensus hexamers, the length of the spacer between the two promoter hexamers, the distance from the mRNA start site, etc. The sequences of the most probable candidates, based on these criteria, were then aligned, and they served as a basis for a weight matrix that provided scores for the four bases in each position of the promoter. The scores were determined by the formula $\log_2(P_{ij}/P_i)$, where P_{ij} is the frequency of base i at position j and P_i is 0.25 (i = A, C, G, T). The length of the spacer sequence was scored similarly. Five spacer lengths, 15–19, were considered, and their score was determined as $\log_2(P_s/0.2)$, where P_s is the frequency of a spacer of length s in the data. The overall score of a promoter sequence was determined as the sum of its position scores and spacer score. The training set of promoters based on the experimentally determined mRNA start sites was used to set a threshold of promoter scores. The average of promoter scores was 6.8 ± 3.7, and the threshold was set to 6. For predicting the promoters of the putative sRNA sequences, the consensus sequence was searched in the empty regions as above except that no more than four mismatches in total were allowed. The putative promoters were scored by the weight matrix. Only promoter sequences above the threshold were recorded. The average promoter scores of the verified sRNAs and of the full set of candidates were similar (9.1 ± 1.6 and 8.7 ± 1.6, respectively)

Prediction of Rho-independent terminators

As above, the prediction was based on 80 experimentally determined Rho-independent terminators, 75 from the compilation by d'Aubenton Carafa et al. [28], and five from the known sRNA genes, *oxyS, micF*, *dsrA, spf* (spot42), and *csrB*. We characterized the properties of these terminators and applied this knowledge in the prediction. It is well known that Rho-independent terminators form a stem loop structure. Therefore, the sequences of known terminators were folded by an RNA folding algorithm based on free-energy considerations [41], which provided both the predicted secondary structure and its stability. This was carried out by the Mfold program of GCG. Almost all of the known terminators formed a stem 5–10 base pairs in length with a loop of 3–8 bases. The stems were GC rich, and most of them had at least 60% GC base pairs. In most structures, the free energy was calculated to be below −7 kcal/mole. Another known feature of Rho-independent terminators is a uridine stretch that follows the stem. The average number of uridine residues in the known terminators was four. In most cases, this stretch was located immediately downstream of the end of the stem. Based on the above, the search of terminators in the "empty" regions involved several steps: (1) A search for sequences that could create a GC-rich stem with a loop, followed by at least four U residues. This was achieved by a search for two sequences that were the same length (5–10 bases) and were composed of at least 60% G/C and by the requirement that the two sequences be separated by 3–8 bases (the loop) and be followed by a stretch of at least four U residues; and (2) The candidate sequences were folded by the Mfold program, and structures with free energy values of at most −7 kcal/mole were selected. We reexamined these structures to validate that Mfold kept the characteristics of the structure determined in the first step. The average energy values of the verified sRNAs and that of the full set of candidates were −12.9 ± 3.1 kcal/mol and −12.3 ± 3.1 kcal/mol, respectively.

Conservation analysis

The region between promoter and terminator was compared by BLAST [42, 43] to the genomes of *Salmonella typhi*, *S. paratyphi*, and *S. typhimurium*. Predicted sequences with statistically significant alignment scores ($E \leq 0.001$) were further evaluated if one of the following criteria was satisfied: (1) The conserved region covered more than 70% of the

candidate sRNA, (2) the conserved region covered 50%–70% of the candidate sRNA, and either the promoter or terminator was predicted with high values, or (3) the conserved region covered only 30%–50% of the candidate sRNA, but both promoter and terminator were predicted with high values. Sequences that passed this screening stage were also compared to genome sequences of *Klebsiella pneumoniae* and *Yersinia pestis*. All predictions shown in Table 1 identified sequences that were at least conserved in *Salmonella* and *Klebsiella pneumoniae.* Conservation in *Yersinia pestis* provided further support to the prediction. A detailed summary of the conservation analysis is available at http://bioinfo.md.huji.ac.il/marg/small_rna. After the final list of predictions was obtained, candidate sRNAs appearing at multiple loci and those suspected as potential trailers of upstream genes were excluded.

Bacterial growth conditions

E. coli K12 cells (source: S. Kustu) diluted 1/100 in LB medium or M9 minimal medium supplemented with glycerol (0.4%) were grown at 37°C. Samples were taken at 1.5, 2, 4, 6, 8, and 10 hr after dilution in LB (at OD_{600} values of 0.3, 0.6, 1.4, 2.3, 2.4, and 2.5, respectively) and 3 and 8 hr after dilution in M9 medium (OD_{600} of 0.3 and 1.5, respectively). For heat shock treatment, cells grown at 30°C to an OD_{600} of 0.3 were transferred to 42°C for 15 min. For cold shock treatment, cultures at an OD_{600} of 0.3 were transferred from 37°C to 15°C for 30 min. To study the effect of RNase III, we isolated total RNA from K12 *rnc-14*::ΔTn*10* [44] cells grown in LB medium for 6 hr.

RNA isolation

Total RNA was isolated from cultured cells by acid-phenol extraction as described previously [45] or by use of the TriPure (Roche Diagnostics) reagent. For RNA isolation with TriPure, cell pellets were resuspended in 50 μl 10 mM Tris-HCl (pH 7.5) containing 1 mM EDTA. Lysozyme was added to 0.5 mg/ml, and samples were subjected to three freeze-thaw cycles. RNA was extracted according to the manufacturer's protocol (Boehringer) except that 1 ml of TriPure reagent was used for 6×10^9–12×10^9 cells. RNA samples used in RACE experiments were treated with DNaseI (Roche Diagnostics) twice for the elimination of DNA contamination.

Northern analysis

RNA samples (30 μg) were denatured for 5 min at 65°C in loading buffer containing 40% (final) formamide, separated on 6% urea-polyacrylamide gels and transferred to nylon membranes by electroblotting. The membranes were hybridized with specific [^{32}P] end-labeled primers except for SraF and SraA, which were detected with specific end-labeled, PCR-generated fragments.

In vitro transcription

Transcription mixtures (50 μl) contained 0.02 μM PCR-generated DNA fragment, 20 mM Tris-HCl (pH 7.6), 150 mM KCl, 2 mM DTT, 10 mM $MgCl_2$, 100 μM each of ATP, CTP, GTP, UTP, 20 units of RNase inhibitor (RNasin: Promega), and 1 unit of *E. coli* RNA polymerase (Boehringer). Transcription was carried out at 30°C for 20 min. Thereafter, the reactions were treated with 1 unit of DNaseI (Ambion) at 37°C for 15 min. The enzyme was heat inactivated for 10 min in the presence of 5 mM EDTA. Samples were run on denaturing gels, followed by electroblotting and hybridization as above. The transcripts of *sraF* and *sraH* were uniformly labeled by the inclusion of [α-^{32}P]GTP

Primer extension assays

RNA samples (30 μg) were annealed to the corresponding end-labeled primers (70°C for 5 min, followed by incubation for 20 min at 42°C and 10 min at room temperature) and then subjected to primer extension (at 42°C for 45 min) with 1 unit of AMV-RT (Promega or Roche Diagnostics) and dNTPs (0.5 mM each). The extension products were separated on 6% sequencing gels, alongside with sequencing reactions.

5′ and 3′ RACE

5′-RACE assays were carried out essentially as Bensing et al. described [25], with minor modifications. 5′ triphosphates were converted to monophosphates by treatment of 15 μg total RNA with 25 units of tobacco acid pyrophosphatase (Epicentre Technologies) at 37°C for 60 min in a total reaction volume of 50 μl containing 50 mM sodium acetate (pH 6.0), 10 mM EDTA, 1% β-mercapto-ethanol, and 0.1% Triton X-100. Control RNA was incubated under the same conditions in the absence of the enzyme. Reactions were stopped by phenol chloroform extraction, followed by ethanol sodium acetate precipitation. Precipitated RNAs were redissolved in water, mixed with 500 pmol of 5′ RNA adapter (A3, 5′-GAU AUG CGC GAA UUC CUG UAG AAC GAA CAC UAG AAG AAA-3′: Dharmacon Research), heat-denatured at 95°C for 5 min, then quick-chilled on ice. The adapter was ligated at 17°C for 12 hr with 50 units of T4 RNA ligase (New England Biolabs) in a buffer containing 50 mM Tris-HCl (pH 7.9), 10 mM $MgCl_2$, 4 mM DTT, 150 μM ATP, and 10% DMSO. Phenol chloroform-extracted, ethanol-precipitated RNA (5 μg) was then reverse-transcribed were gene-specific primers (2 pmol each) and the Thermoscript RT system (Gibco BRL) according to the manufacturer's instructions. Reverse transcription was performed in three subsequent 20 min steps at 55°C, 60°C, and 65°C. RNaseH treatment followed. The products of reverse transcription were amplified by the use of 1 μl aliquot of the RT reaction, 25 pmol of each gene-specific and adapter-specific primer, 250 μM of each dNTP, 1 unit of Hotstar *Taq*-DNA polymerase (Eurogentec, Belgium), and 1× *Taq* buffer containing 2.5 mM $MgCl_2$. Cycling conditions were as follows: 95°C/10 min; 40 cycles of 95°C/30 s, 52°C–55°C/40 s, 72°C/40 s; 72°C/10 min. Products were separated on 3% Nusieve agarose gels, and bands of interest were excised, gel-eluted (QIAex II: Qiagen), and cloned into pCR 2.1 TOPO-vector (Invitrogen). Bacterial colonies obtained after transformation were screened for the presence of inserts of appropriate size by colony PCR with REV and UNI primers. The PCR fragments were then purified on QIAquick spin columns (Qiagen) and sequenced with an ABI 373 automatic DNA sequencer (Applied Biosystems). 3′-RACE assays were carried out with RNA that had been dephosphorylated with calf intestine alkaline phosphatase (APB). Ligation was done as above with a 3′ RNA adapter (E1, 5′-phosphate-UUC ACU GUU CUU AGC GGC CGC AUG CUC-idT -3′ [Dharmacon Research]; idT, 3′ inverted deoxythymidine). Reverse transcription was carried out as described, but with 100 pmol of a single primer complementary to the E1 RNA adapter (E4). PCR amplification, cloning, and sequence analysis was done as described above. All enzymatic treatments of RNA were performed in the presence of 2 units of Super-RNase-Inhibitor (Ambion).

Oligodeoxyribonucleotides used

All RNAs were detected by the use of 5′ end-labeled oligodeoxyribonucleotides except for *sraA* and *sraF*, which were detected by the use of 5′ end-labeled PCR-generated fragments. Sequences of oligodeoxyribonucleotides used as probes and those used for PCR were as follows: *sraA* (PCR, 550, 5′-GCG CAA CAG GCA TCT G-3′ and 551, 5′-CCG CCA GGT AAT CAG AT-3′); *sraB* (456, 5′-CAC ATT GCG GGT TAC TGC-3′); *rprA* (499, 5′-CAA AGA CTA CAC ACA GCA A-3′); *sraC* (449, 5′-TCA GCT GAT GAC CAC CA-3′); *sraD* (534, 5′-GAT AAC AAA TGC GCG TC-3′); *gcvB* (447, 5′-GTC TGA ATC GCA GAC CA-3′); *sraE* (542, 5′-GTA CCG AAT AAT CTC ACC AA-3′); *sraF* (PCR, 511, 5′-TTG CCA AAG TAA AAC AGTG-3′ and 533, 5′-ATG ACG ATC GAC CGG CA-3′); *sraG* (539, 5′-AGG GTT GTC ATT AGT CG-3′); *sraH* (448, 5′-CGA ATA CTG CGC CAA C-3′); *sraI* (470, 5′-CTG GAA GCA ATG TGA GC-3′); *sraJ* (463, 5′-GTC AGT GGA CGA TAA GC-3′); *sraK* (474, 5′-TCT TCG CCT CCT GGC GC-3′); *sraL* (464, 5′-GGG TTT CCC CCG ACG TC-3′); *psrA1* (540, 5′-TCA GCT AAC CCT TGT GG-3′); *psrA6* (536, 5′-ACA CGA TTC CGC TTG AC-3′); *psrA9* (537, 5′-CCC CTC CTG GCA TTG AT-3′). The oligodeoxyribonucleotides used for SraL RNA 5′- and 3′-end mapping (RACE) were as follows: 5′-end mapping (104, 5′-G TTT CCC CCG ACG T-3′ and 104B, 5′-CCC GAC GTC AAC ACA C-3′); 3′-end mapping (104A, 5′-ACC CTC CTG TGT ACC AG-3′ and 104C, 5′-GTC CCA GCG GGA TAG AG-3′). The sequences of the oligodeoxyribo-

nucleotides used for primer extension, PCR-reactions, or RT-PCR/RACE are available at http://bioinfo.md.huji.ac.il/marg/small_rna.

Acknowledgements

The authors gratefully acknowledge the excellent technical assistance by Monica Tamasi and Anna Barladian. We also thank Julio Collado-Vides for providing the electronic list of terminators. This work was supported by the Human Frontier Science Program (G.W., H.M., and S.A.), the Swedish National Science Research Council (G.W.), the Israeli Ministry of Science (H.M., G.B.), Yeshaya Horowitz fellowship for distinction (L.A.), and THE ISRAEL SCIENCE FOUNDATION founded by The Academy of Sciences and Humanities – Centers of Excellence Program (S.A.).

References

1. Hildebrandt M, Nellen W: **Differential antisense transcription from the *Dictyostelium* EB4 gene locus: implications on antisense-mediated regulation of mRNA stability.** *Cell* 1992, **69:**197-204.
2. Lankenau S, Corces VG, Lankenau DH: **The *Drosophila* micropia retrotransposon encodes a testis-specific antisense RNA complementary to reverse transcriptase.** *Mol Cell Biol* 1994, **14:**1764-1775.
3. Morfeldt E, Taylor D, von Gabain A, Arvidson S: **Activation of alpha-toxin translation in *Staphylococcus aureus* by the *trans*-encoded antisense RNA, RNAIII.** *EMBO J* 1995, **14:**4569-4577.
4. Sharp TV, Schwemmle M, Jeffrey I, Laing K, Mellor H, Proud CG, *et al.*: **Comparative analysis of the regulation of the interferon-inducible protein kinase PKR by Epstein-Barr virus RNAs EBER-1 and EBER-2 and adenovirus VAI RNA.** *Nucleic Acids Res* 1993, **21:**4483-4490.
5. Wagner EGH, Simons RW: **Antisense RNA control in bacteria, phages, and plasmids**. *Annu Rev Microbiol* 1994, **48:**713-742.
6. Wassarman KM, Zhang A, Storz G: **Small RNAs in *Escherichia coli***. *Trends Microbiol* 1999, **7:**37-45.
7. Wightman B, Ha I, Ruvkun G: **Posttranscriptional regulation of the heterochronic gene lin-14 by lin-4 mediates temporal pattern formation in *C. elegans*.** *Cell* 1993, **75:**855-862.
8. Kirsebom LA: **RNase P–a 'Scarlet Pimpernel'**. *Mol Microbiol* 1995, **17:**411-420.
9. Luirink J, Dobberstein B: **Mammalian and *Escherichia coli* signal recognition particles.** *Mol Microbiol* 1994, **11:**9-13.
10. Tollervey D, Kiss T: **Function and synthesis of small nucleolar RNAs**. *Curr Opin Cell Biol* 1997, **9:**337-342.
11. Madhani HD, Guthrie C: **Dynamic RNA-RNA interactions in the spliceosome**. *Annu Rev Genet* 1994, **28:**1-26.
12. Blattner FR, Plunkett G, Bloch CA, Perna NT, Burland V, Riley M, *et al.*: **The complete genome sequence of *Escherichia coli* K-12**. *Science* 1997, **277:**1453-1474.
13. Hindley J: **Fractionation of ^{32}P-labelled ribonucleic acids on polyacrylamide gels and their characterization by fingerprinting.** *J Mol Biol* 1967, **30:**125-136.
14. Brownlee GG: **Sequence of 6S RNA of *E. coli***. *Nature New Biol* 1971, **229:**147-149.
15. Wassarman KM, Storz G: **6S RNA regulates *E. coli* RNA polymerase activity.** *Cell* 2000, **101:**613-623.
16. Urbanowski ML, Stauffer LT, Stauffer GV: **The *gcvB* gene encodes a small untranslated RNA involved in expression of the dipeptide and oligopeptide transport systems in *Escherichia coli*.** *Mol Microbiol* 2000, **37:**856-868.
17. Majdalani N, Chen SA, Murrow J, St John K, Gottesman S: **Regulation of RpoS by a novel small RNA: the characterization of RprA**. *Mol Microbiol* 2001, **39:**1382-1394.
18. Andersen J, Delihas N, Ikenaka K, Green PJ, Pines O, Ilercil O, *et al.*: **The isolation and characterization of RNA coded by the *micF* gene in *Escherichia coli***. *Nucleic Acids Res* 1987, **15:**2089-2101.
19. Faubladier M, Cam K, Bouche JP: ***Escherichia coli* cell division inhibitor DicF-RNA of the *dicB* operon. Evidence for its generation *in vivo* by transcription termination and by RNase III and RNase E-dependent processing.** *J Mol Biol* 1990, **212:**461-471.
20. Sledjeski DD, Gupta A, Gottesman S: **The small RNA, DsrA, is essential for the low temperature expression of RpoS during exponential growth in *Escherichia coli*.** *EMBO J* 1996, **15:**3993-4000.
21. Romeo T: **Global regulation by the small RNA-binding protein CsrA and the non-coding RNA molecule CsrB.** *Mol Microbiol* 1998, **29:**1321-1330.
22. Altuvia S, Weinstein-Fischer D, Zhang A, Postow L, Storz G: **A small, stable RNA induced by oxidative stress: role as a pleiotropic regulator and antimutator.** *Cell* 1997, **90:**43-53.
23. Ikemura T, Dahlberg JE: **Small ribonucleic acids of *Escherichia coli*. Characterization by polyacrylamide gel electrophoresis and fingerprint analysis.** *J Biol Chem* 1973, **248:**5024-5032.
24. Rivas E, Eddy SR: **Secondary structure alone is generally not statistically significant for the detection of noncoding RNAs.** *Bioinform* 2000, **16:**583-605.
25. Bensing BA, Meyer BJ, Dunny GM: **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* 1996, **93:**7794-7799.
26. Delihas N: **Regulation of gene expression by *trans*-encoded antisense RNAs.** *Mol Microbiol* 1995, **15:**411-414.
27. Hershberg R, Bejerano G, Santos-Zavaleta A, Margalit H: **PromEC: an updated database of *Escherichia coli* mRNA promoters with experimentally identified transcriptional start sites.** *Nucleic Acids Res* 2001, **29:**277.
28. d'Aubenton Carafa Y, Broody E, Thermes C: **Prediction of Rho-independent *Escherichia coli* transcription terminators. A statistical analysis of their RNA stem-loop structures.** *J Mol Biol* 1990, **216:**835-858.
29. Li Z, Pandit S, Deutscher MP: **3′ exoribonucleolytic trimming is a common feature of the maturation of small, stable RNAs in *Escherichia coli*.** *Proc Natl Acad Sci USA* 1998, **95:**2856-2861.
30. Li Z, Pandit S, Deutscher MP: **Polyadenylation of stable RNA precursors *in vivo***. *Proc Natl Acad Sci USA* 1998, **95:**12158-12162.
31. He L, Söderbom F, Wagner EGH, Binnie U, Binns N, Masters M: **PcnB is required for the rapid degradation of RNAI, the antisense RNA that controls the copy number of ColE1-related plasmids.** *Mol Microbiol* 1993, **9:**1131-1142.
32. Xu F, Lin-Chao S, Cohen SN: **The *Escherichia coli pcnB* gene promotes adenylylation of antisense RNAI of ColE1-type plasmids *in vivo* and degradation of RNAI decay intermediates.** *Proc Natl Acad Sci USA* 1993, **90:**6756-6760.
33. Dam Mikkelsen N, Gerdes K: **Sok antisense RNA from plasmid R1 is functionally inactivated by RNase E and polyadenylated by poly(A) polymerase I.** *Mol Microbiol* 1997, **26:**311-320.
34. Söderbom F, Wagner EGH: **Degradation pathway of CopA, the antisense RNA that controls replication of plasmid R1.** *Microbiology* 1998, **144:**1907-1917.
35. Lowe TM, Eddy SR: **A computational screen for methylation guide snoRNAs in yeast.** *Science* 1999, **283:**1168-1171.
36. Cavaillé J, Buiting K, Kiefmann M, Lalande M, Brannan CI, Horsthemke B, *et al*: **Identification of brain-specific and imprinted small nucleolar RNA genes exhibiting an unusual genomic organization**. *Proc Natl Acad Sci USA* 2000, **97:**14311-14316.
37. Wassarman KM, Repoila F, Rosenow C, Storz G, Gottesman S: **Identification of novel small RNAs using comparative genomics and microarrays.** *Genes Dev* 2001, in press.
38. Hertz GZ, Stormo GD: ***Escherichia coli* promoter sequences: analysis and prediction.** *Methods Enzymol* 1996, **273:**30-42.
39. Yada T, Nakao M, Totoki Y, Nakai K: **Modeling and predicting transcriptional units of *Escherichia coli* genes using hidden Markov models.** *Bioinform* 1999, **15:**987-993.
40. Craven M, Page D, Shavlik J, Bockhorst J, Glasner J: **A probabilistic learning approach to whole genome operon prediction**. *Proc Int Conf Intell Syst Mol Biol* 2000, **8:**116-127.
41. Zuker M: **On finding all suboptimal foldings of an RNA molecule.** *Science* 1989, **244:**48-52.
42. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ: **Basic local alignment search tool**. *J Mol Biol* 1990, **215:**403-410.
43. Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W, *et al.*: **Gapped BLAST and PSI-BLAST: a new generation of protein database search programs**. *Nucleic Acids Res* 1997, **25:**3389-3402.
44. Takiff HE, Baker T, Copeland T, Chen S, Court DL: **Locating essential *Escherichia coli* genes by using Mini-Tn*10* transposons: the pdxJ operon.** *J Bacteriol* 1992, **174:**1544-1553.
45. Storz G, Altuvia S: **OxyR regulon**. *Methods Enzymol* 1994, **234:**217-223.