Biol. Chem., Vol. 386, pp. 1219–1238, December 2005 • Copyright © by Walter de Gruyter • Berlin • New York. DOI 10.1515/BC.2005.140

Review

How to find small non-coding RNAs in bacteria

Jörg Vogel* and Cynthia Mira Sharma

Max Planck Institute for Infection Biology, RNA Biology, Schumannstr. 21/22, D-10117 Berlin, Germany

* Corresponding author
e-mail: vogel@mpiib-berlin.mpg.de

Abstract

Small non-coding RNAs (sRNAs) have attracted considerable attention as an emerging class of gene expression regulators. In bacteria, a few regulatory RNA molecules have long been known, but the extent of their role in the cell was not fully appreciated until the recent discovery of hundreds of potential sRNA genes in the bacterium *Escherichia coli*. Orthologs of these *E. coli* sRNA genes, as well as unrelated sRNAs, were also found in other bacteria. Here we review the disparate experimental approaches used over the years to identify sRNA molecules and their genes in prokaryotes. These include genome-wide searches based on the biocomputational prediction of non-coding RNA genes, global detection of non-coding transcripts using microarrays, and shotgun cloning of small RNAs (RNomics). Other sRNAs were found by either co-purification with RNA-binding proteins, such as Hfq or CsrA/RsmA, or classical cloning of abundant small RNAs after size fractionation in polyacrylamide gels. In addition, bacterial genetics offers powerful tools that aid in the search for sRNAs that may play a critical role in the regulatory circuit of interest, for example, the response to stress or the adaptation to a change in nutrient availability. Many of the techniques discussed here have also been successfully applied to the discovery of eukaryotic and archaeal sRNAs.

Keywords: bacteria; gene expression regulation; Hfq; non-coding RNA; post-transcriptional regulation; riboregulation; RNA processing; RNomics; small RNA.

Introduction

In addition to the triumvirate of tRNA, rRNA and mRNA genes, bacterial genomes are now also known to harbor many, perhaps several hundred, loci that encode non-canonical regulatory RNAs. These RNAs are often referred to as small non-coding RNAs (sRNAs or ncRNAs) because they range in length from approximately 50 to ca. 400 nt and are not translated into proteins. Such RNA molecules were first observed in *E. coli* four decades ago (Griffin, 1971; Ikemura and Dahlberg, 1973a), but neither the genes encoding them nor their functional role was established. Soon after the discovery in the early 1980s that bacterial phages, transposons, and plasmids use small antisense RNAs to control their life cycle or copy number, a seemingly similar antisense RNA, MicF, was found to be encoded by the *E. coli* chromosome (Mizuno et al., 1983, 1984). Unlike the *bona fide* *cis*-antisense RNAs of mobile elements, however, MicF RNA was not transcribed from the DNA strand opposite its target gene, *ompF* (encoding outer membrane protein F). Moreover, MicF exhibited only partial and imperfect sequence complementarity to *ompF* mRNA, yet its binding to the *ompF* message near the start codon strongly inhibited the translation of this mRNA.

It is now clear that MicF was only the first of an ever-growing class of *trans*-encoded antisense RNAs. Two other major outer membrane proteins, OmpA and OmpC, have recently been shown to be regulated at the translational level by their cognate sRNAs, MicA and MicC (Chen et al., 2004; Udekwu et al., 2005). More generally, most of the bacterial sRNAs known to date target mRNAs via imperfect sequence complementarity (see Figure 1 for examples). Binding may result in either the blockage of ribosome entry (translation repression), or the melting of inhibitory secondary structures that sequester the ribosome entry site of the mRNA (translation activation; for a review of sRNAs that modulate translation see Storz et al., 2004). Regulation is frequently accomplished by nuclease-mediated cleavage of the mRNA, e.g., RNase E cleavage of *sodB* mRNA upon RyhB binding (Massé et al., 2003), and RNase III cleavage of *tisAB* mRNA upon IstR-1 binding (Vogel et al., 2004).

Messenger RNAs are not the only targets of sRNAs. Three *E. coli* RNAs, and their homologs in other bacteria, interact with cellular proteins to modulate their activities. 6S RNA, which is highly conserved in prokaryotes and accumulates in stationary phase (Barrick et al., 2005), interacts with σ^{70}-RNA polymerase and induces a change in the holoenzyme's promoter recognition specificity (Wassarman and Storz, 2000; Trotochaud and Wassarman, 2004, 2005). The CsrB and CsrC RNAs form a regulatory feedback loop with CsrA protein, a global post-transcriptional regulator, in which the two RNAs act to antagonize CsrA, thereby tightly controlling the active pool of that protein (Romeo, 1998; Weilbacher et al., 2003).

In addition to regulatory sRNAs that interact with mRNAs or proteins, the chromosome of *E. coli* encodes three sRNAs that serve specialized housekeeping functions. M1 RNA forms the catalytic subunit of RNase P, which is required for tRNA 5′-end maturation (Stark et al., 1978). A second housekeeping RNA, tmRNA (transfer messenger RNA, SsrA, or 10Sa RNA), has unique properties. It has one domain that mimics a tRNA and a sec-

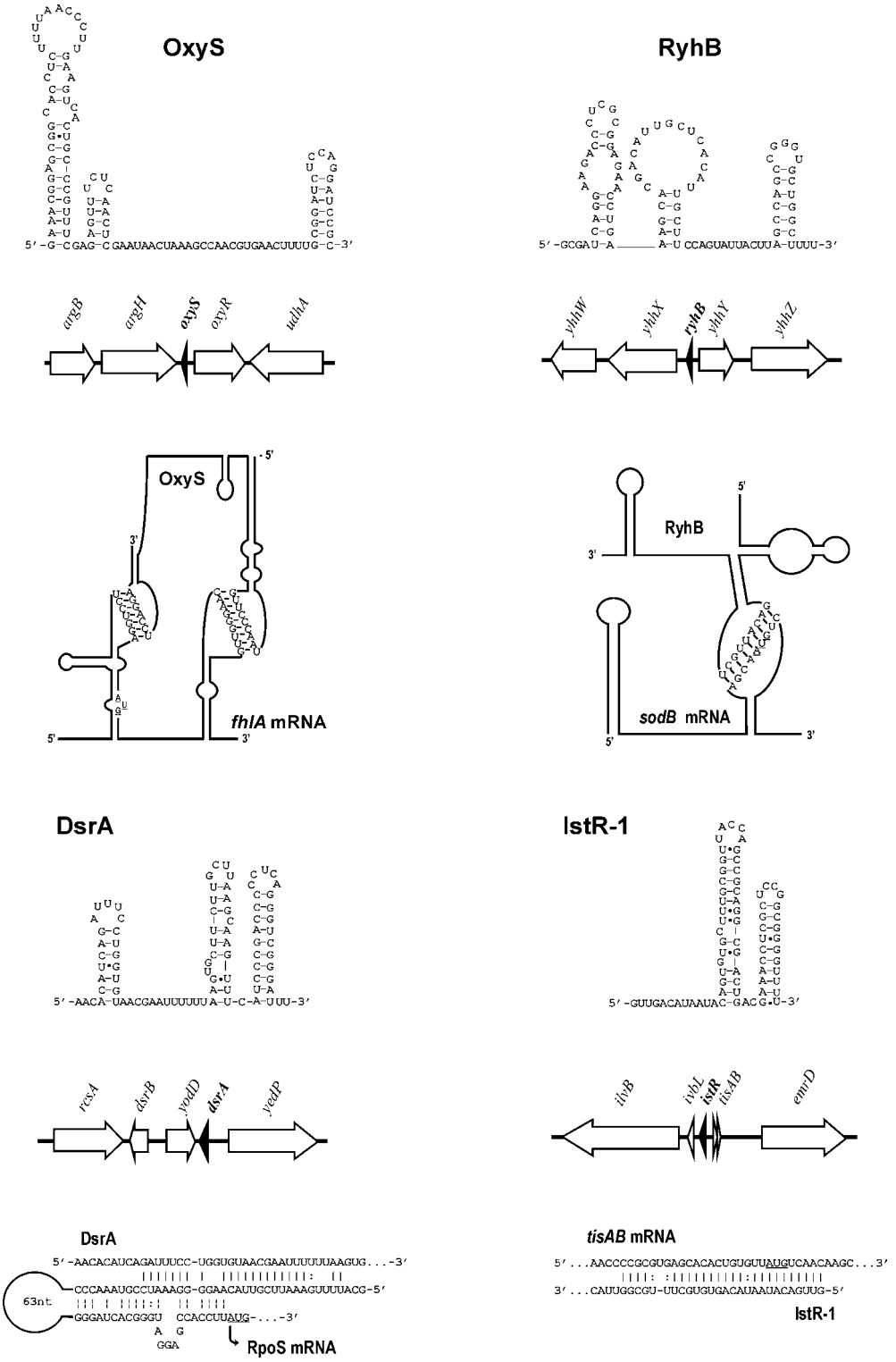

Figure 1 Secondary structure, genomic location, and target interactions of four *E. coli* sRNAs.
OxyS: secondary structure based on structural probing (Altuvia et al., 1997). Like many bacterial sRNAs, OxyS is encoded by a free-standing gene oriented in opposite fashion to both flanking genes. OxyS represses translation of *fhlA* mRNA by the formation of a loop-loop kissing complex (Argaman and Altuvia, 2000). DsrA: proposed secondary structure based on nuclease footprinting and phylogenetic data (Lease and Belfort, 2000). In contrast to the other examples, interaction of DsrA RNA with *rpoS* mRNA leads to translational activation by dissolving the fold-back structure of *rpoS* mRNA in which the ribosome binding site is masked (Majdalani et al., 1998). RyhB: secondary structure based on nuclease footprinting experiments. RyhB interacts with its target *sodB* mRNA by base pairing across the start codon and thereby inhibits translation (Geissmann and Touati, 2004). The interaction between the two RNAs is mediated by Hfq, which induces a structural rearrangement of the *sodB* 5′-UTR. IstR-1: secondary structure as determined by *in vitro* RNA structure probing (F. Darfeuille, personal communication). IstR-1 inactivates *tisAB* mRNA for translation by an antisense interaction in the *tisA* start codon region; RNase III-mediated cleavage of the interacting RNAs is required (Vogel et al., 2004). The start codons of respective target mRNAs are underlined.

ond domain that encodes a short peptide tag. When ribosomes stall on a damaged mRNA, the nascent peptide is transferred to tmRNA and translation resumes – *in trans* – on the mRNA segment within tmRNA. This has the dual effect of freeing the locked ribosome and tagging the protein for degradation (Keiler et al., 1996). A third sRNA, 4.5S RNA, is part of a ribonucleoprotein complex that plays an essential role in protein secretion. This bacterial signal recognition particle, which appears to be a simplified version of its eukaryotic counterpart (Ribes et al., 1990; Poritz et al., 1990), recognizes signal sequences on nascent peptide chains emerging from the ribosome. Genes encoding all three of these housekeeping RNAs have been identified in virtually all sequenced bacterial genomes.

The regulatory properties of bacterial sRNAs are as diverse as the circumstances that led to their discovery. Following the initial discovery of MicF RNA, a dozen regulatory non-coding RNAs were identified serendipitously in *E. coli* and in other bacteria in the 1990s (Wassarman et al., 1999). Many of these sRNAs were stumbled upon while analyzing the transcriptional regulation of neighboring protein-coding genes. Other sRNAs were isolated in genetic studies involving multicopy plasmid screens, which were aimed at investigating a certain phenotype. The year 2001, however, marked a turning point in sRNA identification when several laboratories undertook systematic genome-wide searches for new sRNA genes in *E. coli* (Argaman et al., 2001; Rivas et al., 2001; Wassarman et al., 2001). These searches revealed more than 50 new sRNAs, and generated an even longer list of sRNA candidate loci. Alerted by the sheer abundance of these molecules in *E. coli*, several groups have since taken a closer look at other bacterial genomes and reported a plethora of sRNAs from remotely related organisms such as cyanobacteria (Axmann et al., 2005).

In parallel, functional analyses have revealed that sRNAs are the 'missing links' in well-studied bacterial regulons. Newly realized physiological roles have filled gaps in our understanding of the regulation of bacterial iron homeostasis (Massé and Gottesman, 2002), sugar metabolism (Møller et al., 2002; Vanderpool and Gottesman, 2004), and growth-dependent outer membrane protein expression (Chen et al., 2004; Udekwu et al., 2005). An additional finding has suggested hidden roles for non-coding RNAs in prokaryotes. The bacterial Sm-like protein, Hfq, long known as a host factor for Qβ RNA bacteriophage replication in *E. coli* (Franze de Fernandez et al., 1972), has been shown to bind with high affinity to approximately one-third of the known *E. coli* sRNAs (Zhang et al., 2003). Hfq is often required for both sRNA stability and interaction with target mRNAs. As *hfq* deletion strains display complex phenotypes in many bacteria, ranging from a difficulty in coping with diverse stresses to the reduced or attenuated virulence of pathogenic species (Valentin-Hansen et al., 2004), it is now reasonable to speculate that such changes actually result from the loss of function of specific sRNAs. For instance, the requirement for Hfq in *rpoS* expression (*rpoS* encodes the major stress sigma factor of *E. coli*) can now partly be explained by the fact that two Hfq-binding sRNAs interact with this mRNA under various growth and stress conditions (Muffler et al., 1996; Zhang et al., 1998; Sledjeski et al., 2001). A second argument can be found in a recent study of *quorum sensing* signal cascades in *Vibrio cholerae*. Here, the altered expression of a central signal transducer in a *hfq* mutant was found to be linked to the action of four novel sRNAs (Lenz et al., 2004).

With these new molecules and their functions in hand, the following question arises: what is the hallmark of a bacterial sRNA? Non-coding RNAs of eukaryotes and archaea, e.g., snoRNAs or microRNAs, often share common sequence or structural elements. In prokaryotes, however, few shared features have emerged with respect to location in the genome, GC content, or transcript length (Hershberg et al., 2003). If there is a prominent class of bacterial sRNAs, it is those molecules that are encoded by freestanding genes in the 'empty' intergenic regions (IGRs) of bacterial chromosomes. Such sRNAs are transcribed from their own promoters, and transcription most often terminates at a strong Rho-independent terminator. In general, if the sRNA sequence is directly downstream or upstream of a reading frame, its conservation, including promoter and terminator sequences, in other bacteria can be taken as strong evidence for an autonomous transcription unit. Indeed, conservation of transcription signals and primary sequence within intergenic regions between closely related species was the key to success in many of the recent computational sRNA searches (see below). Moreover, many of the known sRNA genes are encoded in the direction opposite to that of both flanking genes (see *dsrA* and *oxyS* genes in Figure 1, and Hershberg et al., 2003) and for that reason cannot possibly be leader or trailer sequences of these. In this regard, comparison of the genomic location of a given sRNA gene in different organisms may be particularly helpful. For example, the gene for RyhB RNA is immediately upstream of an open reading frame (*yhhX*) in *E. coli*, and hence could be a *yhhX* mRNA leader sequence; however, the location of *ryhB* homologs in other enterobacteria strongly indicates that *ryhB* is an independent gene (Figure 2).

Even though single genes flanked by independent promoters and terminators represent the strongest sRNA candidates, not all functional sRNAs meet these criteria. 6S RNA, one of the most abundant and conserved RNAs in bacteria, was first detected by *in vivo* RNA labeling experiments (Hindley, 1967) and subsequently sequenced by enzyme digestion (Brownlee, 1971), but its gene was identified much later (Hsu et al., 1985). Interestingly, it lacks a strong Rho-independent terminator and requires processing from a dicistronic transcript that includes the downstream *ygfA* gene (Hsu et al., 1985; Kim and Lee, 2004). Because the 3′ end of 6S RNA and the start codon of *ygfA* are only approximately 70 bp apart, this important sRNA could easily have been dismissed as a processed leader fragment in sRNA predictions that were strictly based on conserved transcription features.

In an effort to avoid the bias inherent to all predictions, several groups based their screens on cloning size-fractionated small RNAs (Vogel et al., 2003; Kawano et al., 2005). This approach detected many new, often abundant, RNA species that derive from 5′ or 3′ UTRs of

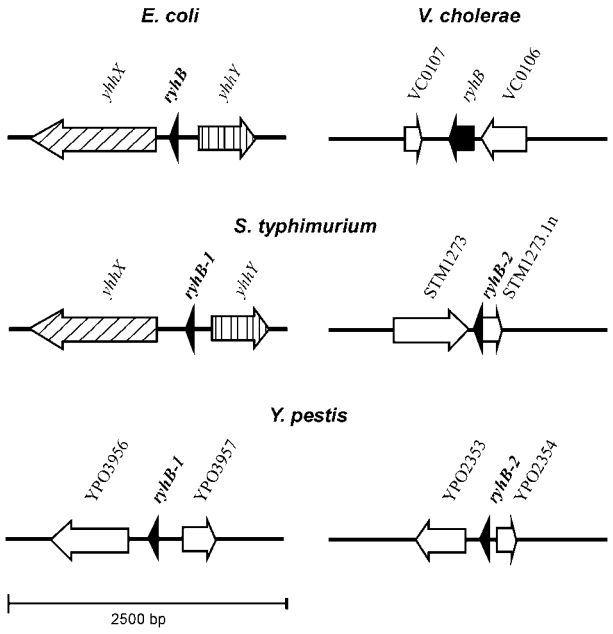

Figure 2 Genomic context of *ryhB* and its orthologs in various bacteria.
In *E. coli*, RyhB is encoded immediately upstream of *yhhX* and could thus be considered as a leader of *yhhX* mRNA. However, while *ryhB* is flanked by orthologs of *E. coli yhhX* and *yhhY* in *Salmonella typhimurium*, it is neighbored by different genes in *Yersinia pestis* and *Vibrio cholerae*. This strongly indicates that *ryhB* is an autonomous gene, which is also supported by significant conservation of *ryhB* promoter and terminator sequences among these organisms (data not shown). Note that *Salmonella* and *Yersinia pestis* each carry an additional *ryhB*-like gene encoded by a distant locus. *V. cholerae* RyhB is considerably longer, carrying ca. 60 extra nucleotides at either end as compared to *E. coli* RyhB.

mRNAs. Whether these molecules are regulatory sRNAs that affect the expression of *trans*-encoded target genes remains to be determined by functional analysis.

Regulatory RNAs do not necessarily have to be non-coding RNAs. Some bacteria possess bifunctional RNAs that act as both mRNA and riboregulator. The highly structured RNAIII of *Staphylococcus aureus* both encodes the peptide δ-hemolysin and modulates the expression of two other virulence genes, *spa* and *hla*, through base pairing of its non-coding regions with the mRNAs of *spa* and *hla* (Novick et al., 1993; Huntzinger et al., 2005). Similar cases of virulence-controlling RNAs, either bifunctional or non-coding, have recently been reported in studies of *Clostridium perfringens* (Shimizu et al., 2002) and streptococcal species (Kreikemeyer et al., 2001; Mangold et al., 2004).

The aim of this review is to critically summarize the various experimental approaches taken over the years to identify sRNA molecules and their genes in prokaryotes (Table 1). For further reading on biological and regulatory functions, as well as mechanisms of regulation, we recommend several other recent reviews (Wassarman et al., 1999; Johansson and Cossart, 2003; Wagner and Vogel, 2003; Gottesman, 2004; Storz et al., 2004; Vogel and Wagner, 2005). *cis*-encoded antisense RNAs also play important roles in prokaryotes, but appear to be mainly confined to mobile elements such as plasmids, transposons, and bacteriophages. For a comprehensive review of *cis*-antisense RNAs, see Wagner et al. (2002).

Labeling and staining of abundant sRNAs: what you see is what you get

The first bacterial small RNAs, other than tRNAs and 5S rRNA, were found by gel fractionation of metabolically labeled *E. coli* total RNA (Hindley, 1967; Griffin, 1971; Ikemura and Dahlberg, 1973a,b). These studies made use of radiolabeled orthophosphate ($^{32}PO_4^{3-}$) which is readily taken up by growing bacteria and incorporated into nucleic acids. Following such treatment, total cellular RNA was isolated, resolved by 1-D or 2-D PAGE, and analyzed by autoradiography. Selected bands or spots were then recovered from the gel and sequenced by digestion with nucleases (fingerprinting). Many of the sRNAs discovered in this way have since been shown to carry out important housekeeping or regulatory functions, including M1 RNA of RNase P, tmRNA, 4.5S RNA, 6S RNA, and Spot 42 RNA.

Provided that the labeling time is long enough to compensate for the slow turnover of particularly stable sRNAs, this method has the advantage that the band or spot intensity of a given RNA is directly related to its abundance in the cell. By reducing the labeling time, or pulse labeling, this approach is also suitable for identifying RNAs with the highest synthesis rate under certain growth or stress conditions. For example, OxyS, a 109-nt regulatory RNA in the oxidative stress regulon of *E. coli*, is barely detectable under standard growth conditions. However, upon oxidative stress induction by treatment with H_2O_2, this RNA accumulates to approximately 4500 molecules per cell (Altuvia et al., 1997). This concentration is comparable to that of abundant *E. coli* RNAs (13 000 molecules per cell of tmRNA is typical), and thus it would be expected that OxyS is readily detectable by metabolic labeling under conditions of oxidative stress.

Such direct labeling approaches certainly have disadvantages, and they have been used to a much lesser extent in recent years. However, some of the drawbacks initially associated with this method can now be minimized. Potential health risks associated with handling radiolabeled bacterial cultures have become less of a concern since more efficacious RNA isolation techniques and higher detection sensitivity allow for significant reductions in the input of radiolabeled orthophosphate. Electrophoresis techniques have also improved greatly, and provide better separation without the need to perform 2D-PAGE. Total RNA may be pre-fractionated on affinity or gel filtration columns to enrich the preparation in sRNAs. This would reduce the background signal and hence the amount of labeled RNA to be analyzed. Finally, the early studies used tedious nuclease-fingerprinting assays to determine the sequence of isolated RNAs. Today, this step can be replaced with rapid cDNA cloning of gel-extracted labeled RNA molecules of interest, which is routine in many laboratories. With these improvements,

Table 1 Brief overview of strategies for sRNA identification in bacteria (see text for references).

Strategy		Advantages (⊕) and disadvantages (⊖)
RNA labeling and staining	⊕	Most abundant sRNAs and/or sRNAs with highest synthesis rate under a given growth condition are readily visualized; does not require prior knowledge of sRNA characteristics in the organism of interest; allows detection of species-specific sRNAs; points to the mature form of the sRNA identified
	⊖	Does not distinguish between sRNAs and abundant processed fragments of rRNAs or tRNAs; can require handling of highly radiolabeled bacterial cultures (orthophosphate labeling)
Functional genetic screens	⊕	May immediately pinpoint a functional role of the identified sRNA; could build on mutant strains and methods already established in genetic studies
	⊖	Difficult if sRNA is either essential or toxic when overexpressed; sRNAs acting under special conditions may not be identified; labor-intensive
Biocomputational searches	⊕	Rapidly generates a list of many potential sRNA candidates; allows phylogenetic comparison with genomes of related bacteria
	⊖	Requires prior knowledge of sRNA characteristics and validation of many candidate loci
Microarray detection	⊕	Yields transcriptional profiles for many sRNA genes in parallel; rapid detection of condition-dependent sRNA expression patterns; allows detection of species-specific sRNA transcripts
	⊖	Requires microarrays that cover intergenic regions; expensive; often yields inconsistent sRNA detection results compared to Northern blot signals
Shotgun cloning (RNomics)	⊕	Should allow detection of all RNAs of a certain size range that are expressed at a given time point; does not require prior knowledge of sRNA characteristics; can be automated; can detect processed, species-specific and non-canonical sRNAs; permits detection of primary transcripts
	⊖	Expensive (sequencing); labor-intensive (screening and evaluation of non-canonical candidates); cDNA synthesis may be biased against highly structured sRNAs
Co-purification with proteins	⊕	Could indicate specific interactions with proteins and the active form of the sRNA
	⊖	RNA has to remain tightly associated with the protein throughout purification; co-immunoprecipitation requires highly specific antibodies; limited to a subclass of sRNAs

metabolic labeling could again prove to be a valuable tool in obtaining an initial glimpse of the most abundant or most actively synthesized sRNAs.

In vitro labeling of extracted total RNA at the 5′ or 3′ terminus provides an alternative route to metabolic or *in vivo* labeling. This approach uses either T4 polynucleotide kinase and γ-[^{32}P]ATP for labeling the 5′ terminus, or T4 RNA ligase and [^{32}P]pCp for labeling the 3′ end. It should be noted that labeling efficiency can vary significantly for the two termini in a given RNA pool, and can thus skew the quantitative representation of individual sRNAs. Figure 3 illustrates such differences between 5′ and 3′ labeling as observed for total RNA from *Helicobacter pylori*. Dramatic differences were also observed for RNA from the cyanobacterium *Synechococcus* (Watanabe et al., 1997). As with *Helicobacter*, the pattern of labeled bands was almost mutually exclusive between the two end-labeling reactions. In *E. coli*, however, the sRNA profile does not differ much between 5′ and 3′ labeling (K.M. Wassarman, personal communication).

What is the nature of these differences? Secondary RNA structure can affect the accessibility of the 5′ or 3′ end to be labeled, and so can the functional group at the 5′ end of sRNAs. 5′ RACE (rapid amplification of cDNA ends) experiments have suggested that many primary sRNA transcripts retain a 5′ triphosphate that, unless removed, will preclude labeling (Argaman et al., 2001; Vogel et al., 2003). The 3′ end of processed sRNAs may be less problematic, since all bacterial ribonucleases known to be involved in sRNA processing generate 5′ phosphate and 3′ hydroxyl groups (RNases E, P and III; cf. Table 1 in Wassarman et al., 1999).

The *in vitro* labeling approach has so far not been exploited for sRNA identification on a larger scale. In the *Synechococcus* sRNA study mentioned above, nine of the 11 most strongly 5′ labeled bands were identified as rRNA fragments. The remaining two fragments originated from the tmRNA homolog of this organism (Watanabe et al., 1998), and an abundant 185-nt RNA (6Sa RNA) that is now regarded as a 6S RNA homolog (Watanabe et al., 1997; Barrick et al., 2005). Neither of these two RNAs was efficiently labeled at its 3′ terminus. However, 3′ labeling was successful in visualizing sRNAs co-immunoprecipitated with Hfq and RNA polymerase in *E. coli* and *Bacillus subtilis*, respectively (Wassarman and Storz, 2000; Wassarman et al., 2001; Trotochaud and Wassarman, 2005).

Some abundant sRNAs were detectable by various staining protocols for total RNA after separation on polyacrylamide gels. The BS190 and BS203 RNAs of *Bacillus subtilis*, both of which are now considered *E. coli* 6S RNA homologs, could be visualized after treatment with ethidium bromide, as could 4.5S RNA (Ando et al., 2002; Suzuma et al., 2002). Silver staining allowed detection of the MP200 RNA(s) of *Mycoplasma pneumoniae* and an MP170 RNA homolog from *M. genitalium* (Göhlmann et al., 2000).

After visualization and gel extraction of RNA species, numerous routes can be taken to determine the sequences or genes of these molecules. In the classic method, the labeled RNA fragment is subjected to enzymatic sequencing. Provided that the full genome of the organism is available, the determined sequence may be used directly in BLASTN searches (Trotochaud and Was-

sarman, 2005). However, since nuclease digests often yield partial or ambiguous sequences, Watanabe et al. (1997) designed oligonucleotide probes to screen a *Synechococcus* genomic phage library by plaque hybridization. To identify the genes of *M. pneumoniae* MP200 RNAs, the extracted RNAs were 5′ labeled and used as probes to screen a cosmid DNA library of the organism (Göhlmann et al., 2000). Differently, gel-extracted BS190 and BS203 RNAs were reverse-transcribed into cDNA, cloned, sequenced, and mapped by searching the *B. subtilis* genome sequence (Ando et al., 2002; Suzuma et al., 2002).

Genetic screens: phenotypes tracked down to small RNAs

Unarguably, the labeling and staining of abundant RNA molecules opened the door to a new sRNA world, but it gave few hints as to the functions of these molecules. Riboregulation was first conceptualized with the discovery of sRNAs in genetic analyses of protein factors that modulated certain physiological activities. For example, when studying the genetic basis for regulation of the two *E. coli* outer membrane proteins, OmpC and OmpF (Mizuno et al., 1984), it was observed that multiple-copy plasmids carrying a 300-bp DNA segment of the *ompC* promoter blocked OmpF expression. As this fragment had partial sequence complementarity with the 5′ end of *ompF* mRNA, it was speculated that it encoded an *ompF* antisense RNA. Further analysis revealed a 93-nt transcript without protein-coding potential, MicF RNA, which interfered with ribosome entry of *ompF* RNA. MicF was novel in being the first *trans*-encoded regulatory antisense RNA identified, in contrast to the then recently discovered *cis*-encoded antisense RNAs.

A mucoid phenotype led to the discovery of *E. coli* DsrA RNA (87 nt). When studying factors involved in capsular synthesis, such as the positive regulator RcsA, it was found that multi-copy plasmids carrying a region downstream of the *rcsA* gene caused capsule overproduction. Subcloning of this region resulted in the isolation of the non-coding *dsrA* gene. DsrA was further shown to antagonize *hns* mRNA translation by an antisense mechanism, which finally explained the mucoid phenotype of multi-copy *dsrA* plasmids: under normal conditions, the histone-like protein, H-NS, silences the *rcsA* gene. Overproduction of DsrA decreases H-NS levels, which abrogates *rcsA* repression and leads to elevated capsule polysaccharide synthesis (Sledjeski and Gottesman, 1995; Lease et al., 1998, and references therein).

In *E. coli*, DsrA not only acts as a repressor, but also activates translation of the major stress and stationary phase sigma factor, RpoS, at low growth temperatures (Sledjeski et al., 1996; Majdalani et al., 1998). The *rpoS* mRNA features an extraordinarily long 5′ UTR (ca. 600 nt), which adopts a more translation-competent structure upon base-pairing with DsrA RNA. Intriguingly, this was the first observation of a regulatory RNA that activated its target gene, and it prompted the Gottesman group to search for sRNAs that would modulate RpoS expression. The result was the identification of the 105-nt RprA RNA

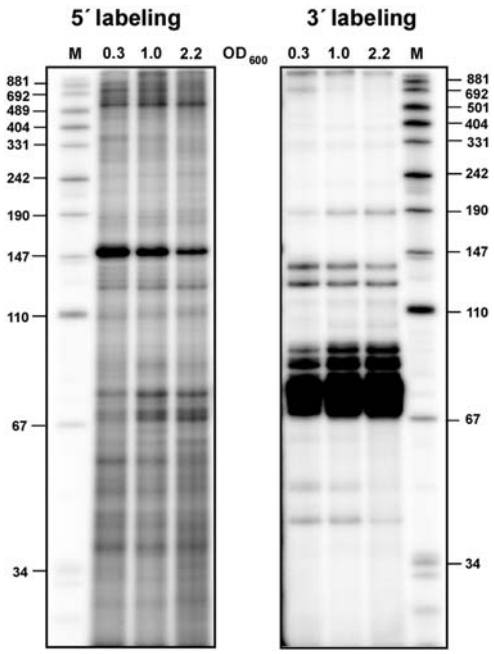

Figure 3 *In vitro* labeling of bacterial total RNA.
Total RNA of *Helicobacter pylori* was radiolabeled at either the 5′ terminus by T4 polynucleotide kinase and [γ-^{32}P]ATP, or at the 3′ terminus by T4 RNA ligase and [^{32}P]pCp. Equal amounts of labeled RNAs from three different growth phases (OD_{600}=0.3, 1, or 2.2) were separated on a 6% denaturing PAA gel and visualized with a phosphorimager. Labeling efficiency varies significantly for the two termini.

(Majdalani et al., 2001). This case bears further description. Majdalani et al. (2001) introduced a pBR322-based plasmid library of *E. coli* genomic DNA fragments ranging in size from 1.5 to 5 kb (Ulbrandt et al., 1997) into a strain that harbored a *rpoS::lacZ* reporter gene (translational fusion) and a mutated *dsrA* locus. By screening 25 000 colonies on MacConkey lactose plates, 12 plasmids were isolated from colonies exhibiting enhanced β-galactosidase activity (red colonies). Of these plasmids, eight mapped to the genomic region in which the *rprA* gene was found (the other four plasmids contained unrelated genomic segments). The smallest of these *rprA*-related fragments induced six-fold higher *rpoS::lacZ* reporter activity in the Δ*dsrA* strain, but supported only two-fold upregulation if the *dsrA* gene was intact. In conclusion, it seems advisable to inactivate an sRNA gene before screening for additional sRNAs that could regulate the same target.

Multi-copy plasmid screens may pick up not only chromosomally encoded sRNAs, but also extrachromosomal sRNA loci, as made apparent by the discovery of the 92-nt UptR RNA (Guigueno et al., 2001). The *uptR* gene resides on the *E. coli* F plasmid, and was identified as a suppressor of export toxicity, a phenomenon caused by proteins that fail to fold properly when passing the membrane. Here, the reporter was an 'unfoldable' DsbA′-PhoA hybrid protein. If encoded by pBR322 plasmids in a protease-deficient *E. coli* strain (K10 Δ*degP*), DsbA′-PhoA confers lethality. The authors first used phage MudII to clone a random library of host genome fragments. Subsequently these cells were transformed with the toxic reporter plasmid and viable colonies containing

putative suppressor genes were picked from selective plates. Of the 18 phagemids that relieved export toxicity, 11 simply restored protease deficiency by having picked up the *degP* gene, while another five phagemids had *recG* inserts, which lowered the copy number of the toxic pBR322 plasmids. It was the remaining two plasmids that carried an unrelated DNA region that was subsequently narrowed down to the *uptR* locus (unfolded protein toxicity-relieving factor).

Colony color also played a role in the identification of CsrC RNA. In *E. coli*, the two small RNAs, CsrB and CsrC, act as antagonists of CsrA, a major regulator of carbon storage genes. CsrA modulates (usually inhibits) the translation of certain target mRNAs, including some that are involved in glycogen biosynthesis. Overproduction of glycogen is easily scored by iodine staining of mutant colonies, which tend to have a darker appearance. A library of 2–10-kb chromosomal DNA fragments cloned in a low-copy plasmid, pGL339, was screened for effects on glucan biosynthesis by iodine staining of colonies (Romeo et al., 1991). Positive clones were subsequently tested for altered expression of a *glgC::lacZ* fusion gene. By subcloning one of the active inserts, the Romeo group arrived at a 360-bp region that promoted *glgC::lacZ* activation, yet did not contain a reading frame. Instead, it harbored the *csrC* gene, encoding a 245-nt RNA (Weilbacher et al., 2003). It is interesting to note that CsrB, the regulatory counterpart of CsrC that was already known to be the primary regulator of *glgC* mRNA, had earlier been identified by direct cloning of RNA molecules bound to CsrA protein (see below).

CsrA is present in many bacterial species, as are homologs of its two regulatory sRNAs, CsrB and CsrC. For example, PrrB RNA, the functional CsrB homolog in the biocontrol strain *Pseudomonas fluorescens* F113, was found in a multi-copy plasmid screen for genes that were able to restore secondary metabolite production in a *gacS*/*gacA*-deficient (two-component system) mutant (Aarons et al., 2000). Here, a single plasmid carrying a 5.4-kb fragment from a genomic library suppressed the mutant phenotype. Further analysis of a subfragment in which all putative ORFs lacked identifiable ribosome binding sites revealed the *prrB* gene, which encodes a 132-nt RNA. It should be noted that all the CsrB homologs described from a variety of organisms have little similarity at the primary sequence level, which is also true of *E. coli* CsrB and CsrC. However, all these RNAs do share significant similarity at the secondary structure level, and in the frequency of occurrence of GGA repeats, which are required for CsrA/RsmA recognition (cf. Romeo, 1998; Valverde et al., 2004). CsrB-like RNAs are often functional when expressed in other species. The latter observation was exploited to identify new *rsmB* (*csrB*) genes in certain *Erwinia* species. An *rsmB* gene had been found in *Erwinia carotovora*, and is responsible for elevated activities of extracellular enzymes when overproduced (Liu et al., 1998). *E. carotovora* was then successfully used in a functional screen with plasmid libraries from two related plant pathogens, *E. herbicola* and *E. amylovora* (Ma et al., 2001), to identify the *rsmB* genes of the latter two species. Such cross-species approaches may be particularly useful for organisms with unknown genomes, i.e., without the possibility of homology searches at the primary sequence level.

In summary, multi-copy plasmid libraries are a valuable tool for identifying sRNAs, and often provide an immediate link to a physiological function. However, it should be noted that few of the screens listed above were tailored for sRNA genes; on the contrary, the libraries often contained DNA fragments long enough to encode even large proteins or operons. To enrich libraries in sRNA genes, it may be best to clone small DNA fragments, which could be generated by frequent-cutter restriction enzymes or mechanical shearing of genomic DNA, followed by size fractionation. As the majority of known sRNAs are >200 nt, DNA fragments ranging from 300 to 400 bp should ensure that the sRNA genes are cloned as autonomous transcription units. For example, the 195-bp minimal fragment that expressed the 92-nt UptR RNA covered approximately 100 bp upstream of the *uptR* transcription start site (Guigueno et al., 2001). Note, however, that when building small insert libraries, a correspondingly higher number of colonies must be screened to ensure full genome coverage.

The plasmid type used for library construction should also be considered carefully. To the best of our knowledge, all of the screens listed above used plasmids of moderate copy number (>30 copies per cell). On the one hand, some sRNA genes may remain repressed under normal growth conditions, even when present in a few more copies. This could be due to autoregulation of the sRNA gene, titration of a positive regulator, or to transcription activation/derepression occurring only under a certain stress condition. Here, cloning DNA fragments under the control of plasmid-borne inducible or constitutive promoters would facilitate expression. (Note, however, that transcription of a random fragment may not always give a functional RNA, even if such an insert contains the entire sRNA sequence. This is because transcription of additional 5′ sequences may affect proper folding of the sRNA.) On the other hand, very high copy numbers may result in pleiotropic effects, inactivate an sRNA gene by mutation, or simply render it lethal to its host. We have indeed observed that some *E. coli* sRNAs are toxic when present in high copy-number plasmids or when driven by a strong plasmid-borne promoter (J. Vogel and E.G.H. Wagner, unpublished results). Thus, constructing the same library with both low and high copy-number plasmids may be more fruitful.

Other than elevated gene dosage, chromosomal inactivation by random transposon insertion mutagenesis can help to identify sRNA genes responsible for a certain phenotype, or that increase host viability under a given stress condition. Fortunately, few sRNAs seem to be essential, namely *E. coli* M1 RNA of RNase P and 4.5S RNA (cf. Wassarman et al., 1999), and *Neisseria gonorrhoeae* tmRNA (Huang et al., 2000). Our laboratory has constructed >30 *Salmonella* strains that have single deletions of sRNA genes that are conserved between *E. coli* and *Salmonella*. All of these strains are viable and none shows a discernible phenotype on standard LB agar plates (V. Pfeiffer and J. Vogel, unpublished results). Regarding specialized growth or stress conditions, however, the DNA damage-induced *istR* locus of *E. coli*,

encoding two regulatory sRNAs, cannot be deleted in SOS constitutive strains (Vogel et al., 2004); therefore, insertions in this gene would result in death, rather than a measurable phenotype, under the most relevant assay condition. Hence, the essential nature of an sRNA gene may become manifest only under certain stress conditions. Another caveat regarding gene disruption-based sRNA screening is that, because of the comparatively small size of sRNA genes, a transposon is five- to ten-fold more likely to disrupt a protein-coding region than an sRNA gene, assuming an average size of 1000 bp for bacterial ORFs (ca. 960 nt in *E. coli*; Blattner et al., 1997). In addition, most protein genes are organized as polycistrons, in which disruption of a single gene often compromises expression of the remaining operon genes (polarity). In contrast, most sRNA genes are autonomous transcription units, and therefore must be disrupted independently. In spite of this statistical bias, transposon insertion mutagenesis in *Bradyrhizobium japonicum* led to the identification of the *sra* gene, which encodes a 213-nt sRNA that is essential for symbiotic nodule development (Ebeling et al., 1991).

Shotgun cloning (RNomics): all you can clone

Shotgun cloning of RNA within a defined size range is an approach that transcends the tedious isolation and sequence determination of individual abundant RNAs. This approach, termed experimental RNomics, led to the discovery of hundreds of non-coding RNAs in several eukaryotes and archaebacteria (see, for example, Hüttenhofer et al., 2001; Marker et al., 2002; Tang et al., 2002, 2005; Yuan et al., 2003). Similarly, the first large-scale assignments of miRNAs and siRNAs, two abundant classes of eukaryotic small RNAs, were also based on shotgun cloning of RNAs of a defined size, in this case approximately 22 nt (Elbashir et al., 2001; Lagos-Quintana et al., 2001; Lau et al., 2001; Lee and Ambros, 2001).

Typical RNomics protocols include an initial size-fractionation of total RNA on polyacrylamide gels. The gel-extracted RNA is subjected to directional cDNA cloning and the resulting libraries are sequenced (Figure 4). This method of randomly cloning as many small RNA fragments as possible aims to comprehensively identify RNAs that are expressed by a given genome under a given set of conditions, irrespective of whether they are primary or processed transcripts.

In the two shotgun-cloning studies conducted in *E. coli* to date, the RNA size range was either 50–500 nt (Vogel et al., 2003) or 30–65 nt (Kawano et al., 2005). Following size fractionation, individual cDNA libraries representing two or three distinct growth phases were constructed, based on earlier observations that many *E. coli* sRNAs are expressed in a growth rate-specific manner (Argaman et al., 2001; Wassarman et al., 2001). In addition to their focus on different RNA sizes, these two studies differed in certain details of cDNA cloning and library screening. For example, Vogel et al. (2003) first C-tailed the extracted RNA with poly(A) polymerase, then constructed cDNA libraries following reverse transcription (Figure 4, left panel). Individual library clones (10 000) were spotted on high-density filters and hybridized with a mix of rRNA and tRNA probes to exclude such clones from further study. cDNA clones that passed this test (ca. 1000 from each growth phase) were sequenced. In contrast, the cloning strategy of Kawano et al. (2005) was more similar to that used to discover eukaryotic miRNAs (Figure 4, right panel). Here, extracted small-sized RNA fragments were ligated to specific 5′ and 3′ RNA adapter molecules, reverse-transcribed, and PCR-amplified. Prior to cloning, the amplification products were concatenated to increase the sequence information per individual library clone.

Following inspection of the cDNA sequences, oligonucleotide probes were designed for the purpose of validating sRNA candidates by Northern hybridizations. It should be noted that by the time these two studies were performed, biocomputation-driven screens of *E. coli* IGRs had already discovered a great number (>50) of sRNAs and predicted even more (ca. 1000) putative sRNA loci. Many of these sRNAs or fragments thereof were represented in the cDNA libraries (20 out of 55 sRNAs known at the time; Vogel et al., 2003). Nonetheless, the two cloning-based screens reported new sRNAs, both from IGRs and, more importantly, from 5′ leader and 3′ trailer regions of mRNAs, i.e., regions that had been deliberately excluded in the other studies. Such UTR-derived RNA species included SroA, a 93-nt RNA that is processed from the 5′ leader of *thiB* mRNA, and the 147-nt RNA SroG, which corresponds to the 5′ end of *ribB* mRNA (Vogel et al., 2003). Expression of *thiB* and *ribB* is controlled at the post-transcriptional level by riboswitches, i.e., metabolite-sensing RNA elements, that are located in the 5′ UTR of these mRNAs (Winkler et al., 2002; Rodionov et al., 2002; Vitreschak et al., 2002; Mironov et al., 2002). The sequences of SroA and SroG RNAs correspond to the riboswitch elements of *thiB* and *ribB* mRNA, respectively. This suggests that these small RNA species may be aptamers; that is, they too may be able to bind the metabolites that control the expression of their parental mRNAs (thiamine or riboflavin derivatives). If such aptamer function were demonstrated *in vivo*, our understanding of sRNA-mediated regulation in bacteria would be greatly enhanced. UTR-derived sRNAs were also reported by Kawano et al. (2005), including a very abundant 35-nt *fimA*-derived RNA species. Interestingly, a number of 3′ UTR-derived small RNAs showed different expression from their parental mRNAs, which suggests that these molecules could have independent functions. It is unlikely that such diverse processed sRNA species would have been found by other methods.

Another strength of cloning-based approaches is their ability to identify sRNAs from intergenic regions that are not conserved in species related to *E. coli*, e.g., *Salmonella*, since such candidates would not rank highly in screens having sRNA gene conservation as the main criterion. For example, the 161-nt RNA, SroH, is encoded within the long *htrC–thiH* IGR of *E. coli*, sharing a bidirectional transcription terminator sequence with the *htrC* gene (Vogel et al., 2003). The small heat-shock gene, *htrC*, is not found in other enterobacteria besides *E. coli* OH157 strains, and neither is *sroH*. In addition to the conservation criterion, many global screens that were

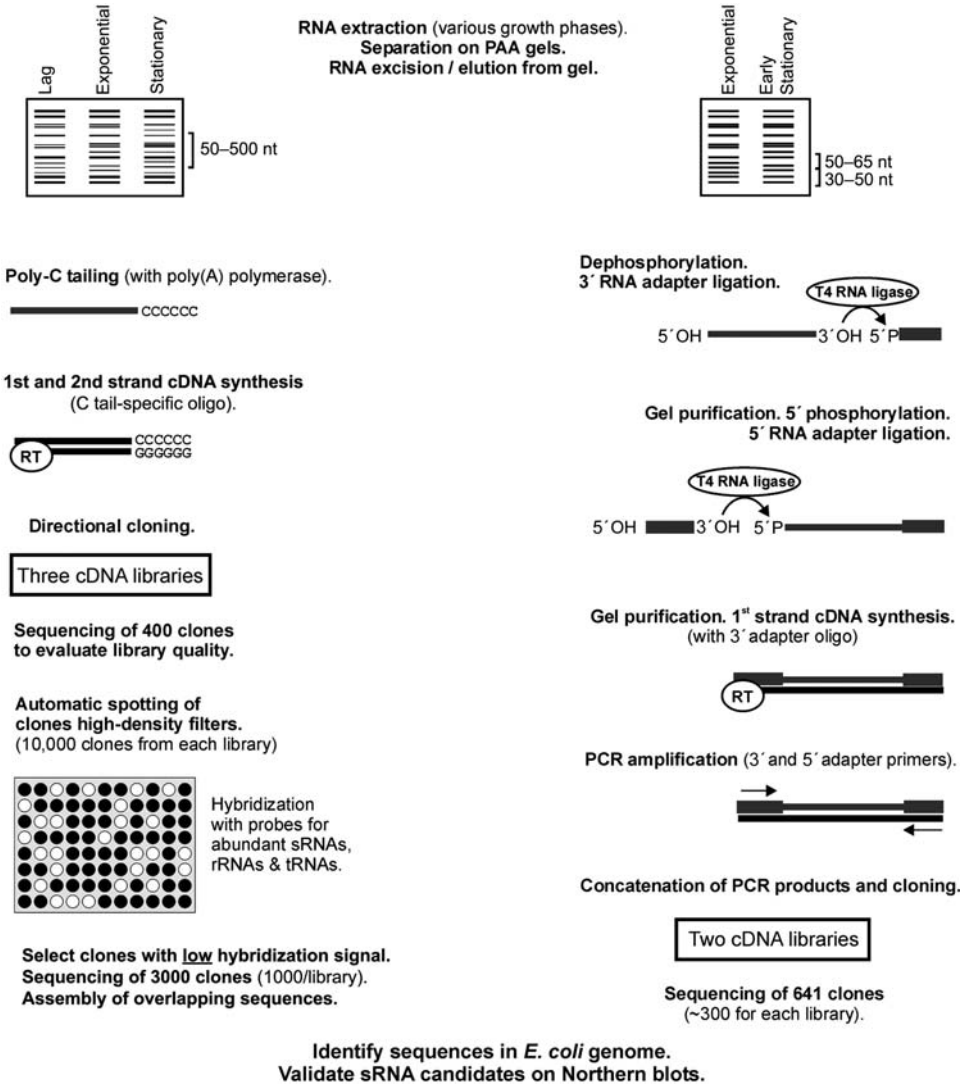

Figure 4 Shotgun cloning (RNomics) of small-sized RNAs in *E. coli.*
The two RNomics screens (left panel: Vogel et al., 2003; right panel: Kawano et al., 2005) in *E. coli* differed in the size range of cloned RNAs and the methods for cDNA cloning and library screening. RNAs were extracted from a size range of 50–500 nt (Vogel et al., 2003) or 30–65 nt (Kawano et al., 2005). Furthermore, the first approach used C-tailing followed by reverse transcription in cDNA library construction. In the second approach, the extracted RNAs were first ligated to specialized 5′ and 3′ RNA adapter molecules, then reverse-transcribed and PCR-amplified. The PCR products were then concatenated before cloning to reduce the library size. In addition, Vogel et al. (2003) pre-selected cDNA clones prior to sequencing by spotting on high-density filters and hybridization with rRNA and tRNA probes.

based on sRNA gene prediction set a minimum length of IGRs to be inspected, typically 180 bp (Wassarman et al., 2001). However, the cloning-based approaches identified sRNAs from smaller IGRs, including the 140-nt RyfD RNA that originates in the 129-bp IGR between *clpB* and *yfiH* (Kawano et al., 2005). This abundant RNA partially overlaps the *clpB* 5′ UTR and coding region, yet it seems to be expressed independently. *Cis*-encoded antisense RNAs were another prominent class of small RNA molecules found by Kawano et al. (2005). Among these were many known RNAs associated with repetitive elements, such as LDR or *hok/sok* (Pedersen and Gerdes, 1999; Kawano et al., 2002), as well as new antisense RNAs, such as RyiC, which may control expression of the damage-inducible *yjiW* gene.

Generally speaking, which bacterial RNA fragments can we expect to find in random cloning approaches? Closer inspection of the *E. coli* cDNA sequences, with regard to their genomic origin, reveals interesting patterns. Vogel et al. (2003) sequenced a total of approximately 3000 clones (1000 clones from each of three growth phases; see above). Following removal of short sequences (ca. 50%), those remaining were automatically assembled into 451 contigs, that is groups of sequences that overlap and map to the same region of the *E. coli* genome. Of these, 78% were fragments of mRNAs of known genes and annotated ORFs of unknown function, including 5′ and 3′ UTRs. Because cDNA cloning was directional, a further 5% of the contigs were classified as antisense transcripts of mRNA coding regions. Since these antisense fragments were excluded from further analysis, it is still unclear whether they are bona fide antisense RNAs, or should be considered as ‘noise’ resulting from the global low-level antisense tran-

scription observed by others in *E. coli* (Selinger et al., 2000). IGR-derived fragments accounted for 17%. Considering that less than 11% of the annotated *E. coli* genome sequence lies *outside* protein, tRNA and rRNA coding regions (Blattner et al., 1997), the library clearly exhibited a slight bias towards sequences from the intergenic space. In contrast, Kawano et al. (2005) observed a high proportion of tRNA and rRNA fragments (ca. 39%) in their two libraries. This difference may be attributed to the fact that rapid pre-screening of the libraries by filter hybridization (see above) could not be performed because the individual cDNA clones contained concatenated sequences corresponding to more than one RNA fragment.

When weighing the advantages of an RNomics screen, namely identification of non-conventional RNAs from UTRs, against the disadvantages, such as the low percentage of actual sRNA clones in the libraries, it should also be considered that the methodology could be considerably improved. In addition, with an ever-increasing capacity for sequencing and automated handling of cDNA clones, the costs of library screening will certainly be reduced to a level at which RNomics becomes a rapid and affordable way to obtain an initial glimpse of the RNome of a bacterium. Regarding cDNA library construction, not all RNAs are equally amenable to linker ligation or C-tailing, and this is expected to skew the representation of some abundant but highly structured sRNAs in the cDNA library. For example, 6S RNA was poorly represented, even in the stationary phase library, corresponding to the growth phase in which 6S RNA accumulates to approximately 10 000 molecules per cell (Wassarman and Storz, 2000; Vogel et al., 2003). Optimizing the individual enzymatic steps, e.g., by using a thermostable reverse transcriptase to perform cDNA synthesis at elevated temperature, may overcome the limitations imposed by stable RNA structures. We have also observed that many sRNA transcripts retain a 5′ triphosphate (Argaman et al., 2001; Vogel et al., 2003). This suggests that if cDNA cloning involves 5′ adapter ligation, conversion of the non-linkable 5′ triphosphates to 5′ monophosphates with tobacco acid pyrophosphatase may increase the retention of primary sRNA transcripts. Further improvement can be expected by affinity-based removal of rRNA from the total RNA pool prior to cloning. As noted above, in spite of RNA extraction from gels (lower cut-off of 50 nt), approximately half of the cDNA inserts sequenced by Vogel et al. (2003) were too small to be mapped unambiguously to a single *E. coli* locus. To increase the number of meaningful cDNAs, it could be helpful to clone the RNA only after repeated gel extraction, or to include another pre-fractionation step, e.g., on columns or in sucrose gradients.

Expression of most sRNAs is limited to specific growth conditions, and this was largely reflected in the representation of individual sRNAs in the libraries prepared from different growth phases. For example, the stationary phase-specific RNA, RyeB, was found in 0, 1, and 59 clones from lag, exponential, and stationary phase libraries, respectively, whereas sRNAs that are specifically stress-induced, such as OxyS RNA, were entirely absent from these libraries (Vogel et al., 2003). However, considering that such regulators may be highly expressed upon induction, e.g., OxyS RNA with 4500 copies per cell (Altuvia et al., 1997), these molecules may be greatly enriched in cDNA libraries prepared from the relevant stress conditions.

Beyond *E. coli*, the only other bacterial RNomics screen conducted thus far was of *Aquifex aeolicus*, a hyperthermophilic bacterium often referred to as being at the origin of the eubacterial phylogenetic tree (Willkomm et al., 2005). The initial motivation for this work came from the perplexing observation that a gene encoding RNase P RNA, which is essential in all bacteria investigated to date, could not be found in this genome. Shotgun cloning allowed the detection of approximately half a dozen sRNA candidates, some from the intergenic space and some that were antisense RNAs. With the exception of housekeeping sRNAs, these candidates were the first to be described in hypothermophilic eubacteria, and included the *A. aeolicus* 6S RNA homolog. Although the *A. aeolicus* genome shows a relatively low, and otherwise unbiased, GC content of ca. 44%, some of the sRNAs identified from IGRs exhibit high GC values similar to those of stable RNAs (rRNA, tmRNA, tRNA: >65%). For example, 6S RNA has a GC content of 60%, indicating that this feature may be used to search for new sRNAs in thermophilic eubacteria. In stark contrast to the *E. coli* study by Vogel et al. (2003) that used the same cloning procedure, 6S RNA clones were highly represented in the *A. aeolicus* library (approx. 10% of the total, including tRNA and rRNA clones). Both 6S RNA species share a tight rod-like structure, and the *A. aeolicus* molecule has an even higher GC content than that of *E. coli*, hence ruling out the possibility of impaired cDNA synthesis of *E. coli* 6S RNA. Whether the reason for this discrepancy originates in an extremely high copy number of *A. aeolicus* 6S RNA or stems from possible overrepresentation of individual abundant RNAs encoded by small genomes (*A. aeolicus*, 1.6 Mb vs. *E. coli*, 4.6 Mb) remains unclear. Nevertheless, the existence of an *A. aeolicus* 6S RNA gene is certainly an exciting discovery and may have further implications for clarifying the evolutionary history of this organisms. If there was an entirely unexpected result of this global screen, it is that an *A. aeolicus* RNase P RNA still remains at large.

Biocomputational screens: bits made the news

Computer-based prediction of protein-coding genes is a standard procedure for annotating bacterial genome sequences. In addition to searching for the longest possible reading frame, such predictions of mRNA genes are frequently supported by the existence of orthologs in related bacteria, and by the occurrence of putative ribosome-binding sites in the vicinity of the predicted start codon. As discussed above (see Introduction), sRNA genes are seldom so luxuriously appointed with identifiers. Fortunately, the ever-increasing number of completed bacterial genome sequences has been paving the way for computer-based sRNA searches at the genomic level.

In 2001, three biocomputation-driven studies quadrupled overnight the number of known *E. coli* sRNAs (Argaman et al., 2001; Rivas et al., 2001; Wassarman et al., 2001). The common denominator was that these groups

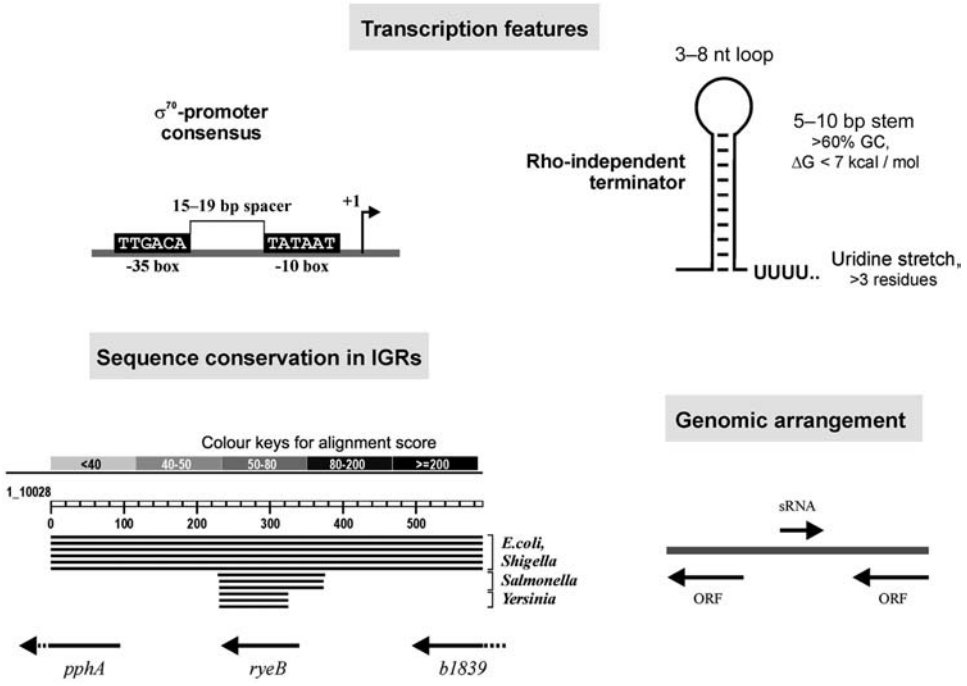

Figure 5 Biocomputational prediction of *E. coli* sRNAs by Argaman et al. (2001).
This sRNA prediction strategy was based on the following criteria: transcription features of non-coding RNA genes; sequence conservation of IGRs; and the genomic location of putative sRNA genes. IGRs of *E. coli* were scanned for sequence matches with the σ^{70} promoter consensus (upper left) and Rho-independent terminator sequences (GC-rich stem-loops followed by at least four U residues; upper right). Sequences containing a predicted promoter and terminator, separated by 50–400 bp, were compared to the genomes of *Salmonella*, *Yersinia*, and *Klebsiella*. High conservation of the predicted sRNA gene in other bacterial genomes and less conservation of the flanking regions was used as an additional criterion for a good sRNA candidate (lower left). Furthermore, the genomic arrangement of putative sRNA candidates was evaluated for the purpose of excluding conserved mRNA leaders or trailers from the final sRNA candidate list (lower right).

based their sRNA gene predictions on sequence conservation between *E. coli* IGRs and those of the closely related enterobacteria *Salmonella typhimurium* and *Yersinia pestis*, for which complete genome sequences had been published. A partial genome sequence was available for *Klebsiella pneumoniae*, and it also supported these predictions. The precise approach, however, differed and each is discussed in more detail below.

Argaman et al. (2001) extracted IGRs from the annotated *E. coli* genome and subjected them to a predictive scheme with four main criteria (Figure 5) derived from the characteristics of the 10 *E. coli* sRNA genes known at the time (Wassarman et al., 1999). First, promoters were predicted that would match the consensus sequence recognized by the vegetative sigma factor, σ^{70}. Second, the IGRs were inspected for strong Rho-independent termination signals, defined as GC-rich ($\geq$60%) hairpin structures with a stem of 5–10 bp and a loop of 3–8 nt followed by a stretch of $\geq$4 uridines. Third, *E. coli* IGRs that contained a promoter and a terminator on the same strand, and within a distance of 50–400 bp, were selected and compared to the genomes of the bacteria listed above by BLASTN searches. Conserved IGRs were extracted based on statistically significant alignment scores ($E \leq 0.001$). An upper limit of 400 bp was set according to the size of the longest *E. coli* sRNAs known at that time (M1 RNA, 377 nt; CsrB, 369 nt; tmRNA, 363 nt). Finally, the genomic locations of sRNA candidates were evaluated: putative sRNA genes that were oriented in the opposite direction to both adjacent genes scored higher because these could not be conserved mRNA leaders or trailers. Application of all these criteria resulted in the prediction of 24 putative sRNA genes. The promoter algorithm based on the σ^{70} consensus gave redundant results; that is, multiple promoters were predicted for most candidate genes, but the majority of these promoters found no support in subsequent biochemical analysis. Thus, the final candidate list was assembled by assigning higher weighting coefficients to the terminator sequence, degree of conservation and gene orientation. In Northern hybridizations with RNA from 10 different growth points and/or conditions, 14 of 23 candidates were shown to be new sRNAs. As an aside, a high-ranking candidate gene that failed in this first validation round was later shown to encode the SOS response-related IstR RNAs (Vogel et al., 2004). The high number of true positives in this screen may be attributed to the rather stringent application of prediction criteria. In other words, the list of candidates to be tested was deliberately kept short and free of seemingly non-canonical sRNAs. For example, most candidates genes were free-standing and possessed a Rho-independent terminator, so that many sRNAs with 3′ ends generated either by processing, e.g., 6S RNA, or by Rho-dependent termination would not easily be selected in this approach.

Wassarman et al. (2001) took a similar route by extracting all 1087 *E. coli* IGR sequences longer than 180 bp, comparing these to *Salmonella* and *Klebsiella* IGRs, and evaluating transcription signals and sRNA gene orientation for those with a high degree of sequence conser-

vation. These predictions were supplemented by experiments in which putative sRNA transcripts were detected by means of *E. coli* microarrays. Finally, 59 sRNA candidates were listed, and 23 of these were confirmed on Northern blots. Of these, 17 were considered to be new sRNA genes, while the remaining six were reclassified as small protein-coding genes, based on reading frame conservation and the presence of putative Shine-Dalgarno sequences.

Rivas et al. (2001) also relied on sRNA conservation across related genomes, but introduced a conceptual change by scoring conservation of RNA secondary structure rather than of primary sequence. Interestingly, the authors had previously observed that secondary structure alone does not generally suffice for the prediction of non-coding RNAs in a single genome (Rivas and Eddy, 2000). Thus, they then combined structure prediction with comparative analysis of *E. coli*, *Salmonella* and *Klebsiella* genomes. The program implemented, QRNA, searches for mutational patterns in pairwise sequence alignments that would distinguish conserved RNA secondary structure from the background of other conserved sequence elements, such as transcription factor-binding sites. In contrast to the patterns of synonymous codon substitutions in conserved protein-coding regions, structural RNAs are revealed through compensatory mutations that are consistent with maintaining predicted secondary structure elements. QRNA predictions in *E. coli* yielded a total of 275 candidate RNAs after removal of known regulatory and repetitive elements. Of these final candidate genes, 49 were assayed experimentally and 11 of these were found to express small transcripts. In this case, Northern analysis was limited to a single growth condition, the exponential phase. Because many sRNAs are known to be expressed in stationary phase or under specific stress conditions, a broader set of growth conditions (Argaman et al., 2001; Wassarman et al., 2001) may be likely to increase the number of confirmed candidates. On the other hand, this method does not distinguish independent sRNA genes from conserved *cis*-regulatory mRNA structures, and for this reason the authors cautiously pointed out that the high number of sRNA candidates is preliminary and requires further validation.

In the main, the three biocomputational screens discussed so far required that sRNAs be conserved. How important is the conservation criterion? Researchers interested in bacteria that are more distantly related to *E. coli* will not always have complete genome sequences of closely related species available. In this regard, another sRNA search in *E. coli* that relied solely on transcription signal prediction offered much insight (Chen et al., 2002). Here, the search strategy was confined to predicting intergenic pairs of σ^{70}-type promoter/Rho-independent terminator pairs. Such pairs were required to lie on the same strand within 45–350 bp of one another. Starting with 227 sRNA candidates, 51 were subsequently removed as putative short protein-coding genes. It should be noted that approximately 50% of these ORFs were missing in the complete genome annotation, whereas the remaining ones were annotated, albeit inconsistently, in various *E. coli* databases. A further 32 loci were filtered out as orphan tRNA genes, short leaders, tRNA/rRNA operon fragments, or known sRNA genes. Of 144 final candidates, eight were experimentally tested by Northern analysis, and seven were found to be new sRNA species. Interestingly, a mere 10 of the 40 sRNAs known at the time were recognized by the search algorithm. However, novel sRNA genes were found and most of them had no overlap with the untested predictions of previous global screens. Disregarding the conservation aspect also seems to have strengthened the prediction of sRNA genes that are specific to *E. coli* K12 and closely related pathogenic *E. coli* strains. A note of caution should be inserted here: in the absence of support from phylogenetic conservation, experimental validation should be even more rigorous before it can be assumed that such species-specific sRNAs are functional. In fact, two of these sRNAs, since renamed GadY and MicC, have recently been assigned regulatory roles (Chen et al., 2004; Opdyke et al., 2004).

An sRNA search in *Vibrio* genomes integrated many features of the prior *E. coli* screens, and added yet another layer (Lenz et al., 2004). The Bassler group had observed that the quorum-sensing master regulator, LuxR, was controlled post-transcriptionally by Hfq. Consequently, they hypothesized that this specific Hfq effect reflected the action of one or more unknown sRNAs. Further results indicated that such sRNAs would be activated by the sigma factor, σ^{54}. Therefore, a computer-based method was developed to scan *Vibrio cholerae* IGRs for pairs of σ^{54} binding sites and Rho-independent terminators, followed by conservation analysis in two other *Vibrio* species. The result was the discovery of four novel sRNAs that are almost identical and are conserved in all three *Vibrio* species investigated. The extent of functional redundancy among these RNAs came as a surprise: to eliminate Hfq-directed quorum-sensing repression, all four genes had to be deleted simultaneously.

Similarly, functional evidence that pointed to the involvement of a hidden sRNA led to the discovery of *Pseudomonas* homologs of the *E. coli* RNA, RyhB (Wilderman et al., 2004). Iron homeostasis is regulated in many prokaryotes by the Fur protein (ferric uptake regulator). Fur acts as a transcriptional repressor unless iron becomes scarce, in which case Fur-repressed genes, e.g., those involved in iron acquisition, are upregulated. It had long remained a mystery as to why iron depletion and the concomitant alleviation of Fur repression also caused downregulation of a number of genes. Identification of the Fur-regulated sRNA, RyhB, of *E. coli* provided an explanation for this apparent paradox (Massé and Gottesman, 2002). Upon iron depletion, RyhB is derepressed and then acts as an antisense RNA to a number of mRNAs. The *ryhB* gene, along with its promoter and Fur-binding site, is well conserved in enterobacteria, but could not be found in *Pseudomonas aeruginosa*, another organism in which positive regulation by Fur had been observed and left unexplained (Ochsner et al., 2002). Encouraged by the obvious parallel between the iron regulatory processes in *Pseudomonas* and *E. coli*, a pattern-based search was derived to scan *Pseudomonas* IGRs for Fur-regulated sRNA genes. Wilderman et al. (2004) used the pattern search program PATSCAN to scan for the following characteristics of the *E. coli ryhB*

gene: a Fur-box consensus (searching for an identity match of ≥14/19 bp), a potential Rho-independent terminator (7–12-bp stem closed by a 5–9-nt loop and immediately followed by a run of ≥3 uridines), and a ≤200-bp spacer separating these two elements. This pattern search yielded three candidate genes, two of which, *prrF1* and *prrF2*, are identical in all but five residues, and are expressed in a Fur-dependent manner. As with *E. coli ryhB* deletion, a *prrF1-prrF2* double deletion offsets *sodB* mRNA downregulation under low iron conditions. In addition, PrrF1 and PrrF2 RNAs show a stretch of complementarity with the translation initiation region of *sodB* mRNA, further suggesting that these newly discovered sRNAs are functional homologs of RyhB.

An automated sRNA screening procedure for the extraction, selection and visualization of candidate IGRs has now been implemented in the software package Intergenic Sequence Inspector, or ISI (Pichon and Felden, 2003). This program filters IGRs according to variable input parameters, including length or GC content, and can select those with significant sequence conservation among phylogenetically related bacteria. In addition to a multiple sequence alignment output, secondary structure predictions and known locations of putative promoters and terminators can be included in the visualization. Besides all the previously characterized *E. coli* sRNAs, ISI is reported to have identified additional candidates. Subsequent characterization of one candidate, which is conserved in 21 strains of *Escherichia*, *Salmonella*, and *Shigella,* revealed that this novel RNA, RydC, folds into a pseudoknot and regulates expression of an ABC transporter gene (Antal et al., 2005). Interestingly, RydC RNA was independently identified by microarray analysis of RNAs that co-immunoprecipitate with Hfq (Zhang et al., 2003).

Two additional bioinformatic approaches to identify sRNAs in *E. coli* have to be mentioned, although both studies primarily focused on the development of novel algorithms and included little or no experimental validation of the resulting predictions. In one of these studies (Carter et al., 2001), a machine learning approach that made use of neural networks and support vector machines was developed to extract the shared features of known sRNAs for the prediction of new candidates in several prokaryotic and archaeal genomes. Similar to QRNA (Rivas et al., 2001), this approach seems to be less dependent on prior knowledge of the specific RNA gene features of a given organism. The underlying algorithm uses both compositional parameters (nucleotide and dinucleotide composition) and structural motif parameters to discriminate functional RNAs from random non-coding sequences. The output of this screen took the form of 562 sequence windows, each of which was 80 bp long and was likely to encode a functional RNA. Disregarding consecutive windows and predictions on both strands, ca. 370 novel sRNA candidates in the *E. coli* genome were predicted. More recently, boosted genetic programming was used to create sRNA classifiers to select non-coding functional RNA sequences from intergenic sequences (Saetrom et al., 2005). This prediction covered 152 of the previously known or predicted sRNAs and yielded 135 novel candidates. The authors sought to validate 16 candidates experimentally: preliminary results of low-stringency Northern hybridizations indicate that some of these candidates are indeed sRNAs.

Following the aforementioned studies in *E. coli*, biocomputational approaches were taken to scan the genomes of entirely unrelated bacteria. These included four marine cyanobacteria of the *Prochlorococcus-Synechococcus* lineage, two of which had compact and small genomes of approximately 1.7 Mb (Axmann et al., 2005). It had previously been assumed that certain features of single functional RNAs, e.g., folding free energy, did not suffice as a statistically significant criterion for finding non-coding RNAs in whole genome sequences (Rivas and Eddy, 2000). However, it was later shown that thermodynamic stability values derived from the consensus folding of aligned related sequences allows effective prediction of functional RNAs (Washietl and Hofacker, 2004; Washietl et al., 2005). Using this strategy, Axmann et al. (2005) first performed BLASTN searches to detect local sequence conservation in IGRs of ≥50 nt among three *Prochlorococcus* genomes and one *Synechococcus* genome. Sequences that overlapped by ≥85% in the four genomes were extracted and subjected to further rounds of alignment using BLASTN and CLUSTALW (Thompson et al., 1994). The resulting alignments were then scored by ALIFOLDZ, with structures being calculated at the approximate habitat temperature (24°C) of the cyanobacteria studied. Expression analysis of the highest-scoring candidate regions under various growth and stress conditions confirmed seven new sRNAs in *Prochlorococcus marinus*, several of which had homologs in the other three strains. Four of these new sRNAs that are highly similar in structure constitute a rapidly evolving gene family with different numbers and locations of genes in these four genomes. In addition, these searches also uncovered new cyanobacterial 6S RNA orthologs, i.e., in addition to the 6S RNA-like genes previously reported in other *Synechococcus* strains (Watanabe et al., 1997).

Borrelia burgdorferi, which causes Lyme disease, also has a small genome (1.6 Mb). However, application of four different search methods only yielded two new sRNAs in addition to the previously annotated tmRNA and RNase P RNA genes (Östberg et al., 2004). First, searches for homologs of 62 *E. coli* sRNAs suggested a single candidate gene (by similarity to *micC*), the expression of which could not be confirmed on Northern blots. Likewise, comparative analysis of *B. burgdorferi* IGRs of ≥150 bp with all available microbial genomes yielded candidates that later failed to show transcripts in Northern hybridizations. The authors then resorted to a novel method in which they searched for sequence complementary between *Borrelia* IGRs and orthologs of genes that are sRNA targets in other bacteria. As discussed previously, *E. coli rpoS* mRNA is regulated by several sRNAs, of which RprA and DsrA bind to *rpoS* 5′ UTR with imperfect sequence complementarity. Reasoning that *Borrelia rpoS* may be regulated in a similar fashion, BLASTN searches were performed with the goal of detecting similarity with intergenic sequences. Out of four candidate IGRs, one was found to express a ca. 200-nt

RNA, RrbA, which may act in antisense fashion to the *rpoS* message during exponential phase. Lastly, there were a number of loci that had been annotated in the *Borrelia* genome as redundant homologs of well-known RNA genes. Reinspection of these loci, however, revealed that several annotated RNase P RNA genes may in fact be repetitive elements, whereas one locus that had been postulated as a secondary tmRNA gene may actually express a >200-nt RNA (BsrA) of unrelated function.

The fact that so few sRNAs were identified in *B. burgdorferi* is surprising. One explanation offered by the authors (Östberg et al., 2004) is that this organism also lacks two proteins, Hfq and RNase E, that facilitate sRNA function in other bacteria. However, a clear correlation between the presence of *hfq* and sRNA genes was not observed in the cyanobacteria study by Axmann et al. (2005). Furthermore, *Helicobacter pylori*, another bacterium with an equally small genome (1.6 Mb), does seem to possess a number of small RNAs despite having neither Hfq nor RNase E (own unpublished results). It has been hypothesized that a disproportionately large number of sRNAs reflects increased regulatory complexity in organisms with larger genomes (Mattick, 2001; Mattick and Gagen, 2001). Conversely, bacteria with small genomes may rely less on sRNAs as regulators. Hence, future screens will also have to clarify the intriguing question of whether sRNA genes are truly underrepresented in streamlined bacterial genomes.

Homology searches: 'O Brother, Where Art Thou?'

Since many of the global *E. coli* sRNA screens took advantage of sequence conservation in related enterobacteria (Argaman et al., 2001; Rivas et al., 2001; Wassarman et al., 2001), these studies en passant set the blueprint for sRNA identification in species such as *Salmonella* and *Yersinia*. This extended pool of enterobacterial sRNAs may give valuable leads as to how to improve predictions of sRNAs across subdivisions of bacteria. To this end, Hershberg et al. (2003) sought to extract the shared features of 55 sRNA genes of *E. coli* and their orthologs in other bacteria. All of these genes were located in IGRs, biased towards longer IGRs (300–900 nt), and typically occurred once per IGR. However, it is important to note that the majority of these sRNAs had been found in IGR-specific screens, and this may add to the observed bias. GC-content analysis showed that all sRNA genes are more GC-rich than the average intergenic space, but have lower GC values than tRNAs and rRNAs. Exceptions include certain abundant and/or housekeeping RNAs (tmRNA, 6S RNA, 4.5 S RNA, RNase P RNA), for which GC values exceed the average for tRNAs and rRNAs. Conservation analysis by means of BLASTN searches of more than 100 completed bacterial genomes did not yield significant sequence similarity for most of these 55 *E. coli* sRNAs beyond *Yersinia pestis*. No sequence similarity was found in archaea. Conservation of sRNA flanking genes decreases with phylogenetic distance, that is, it is highest in closely related *Shigella* and *Salmonella* species and lower in the more distant *Yersinia*. In turn, this suggests that an sRNA gene is more likely to be found in distantly related bacteria when both flanking genes are also conserved. The correlation observed between sRNA gene conservation and the evolutionary distance of compared organisms was subsequently confirmed by an independent study (Zhang et al., 2004). Interestingly, sRNA and protein-coding genes were found to evolve at the same rate in bacteria. In contrast, tRNA genes tend to be more conserved than other genomic regions, and loci of unknown function evolve much faster than the average. Zhang et al. (2004) also made use of this statistical bias in calculating the total number of *E. coli* sRNAs; their estimate was close to 200.

A recent study in *Vibrio cholerae* (Davis et al., 2005) illustrates the degree of sRNA sequence deviation in loosely related bacteria. Of the 55 *E. coli* sRNAs investigated, only seven were found to have homologs in *V. cholerae*. One of the these, RyhB, was analyzed in more detail. In *E. coli*, the *ryhB* gene is transcribed into a 90-nt RNA in a Fur-dependent manner (Massé and Gottesman, 2002), and RyhB RNA abundance is dependent on Hfq (Massé et al., 2003). Both of these features also apply to *Vibrio* RyhB RNA, but Northern analysis showed that it is a considerably longer (approx. 225 nt) transcript. Sequence inspection revealed that the *Vibrio ryhB* gene carries a ca. 60-nt extension at each end compared with its *E. coli* counterpart, and the actual sequence similarity between *E. coli* and *Vibrio* is limited to a central 28-bp region. Sequence identity is also low among related *Vibrio* species, and comes down to a ca. 60-bp stretch. Several of the genes that are regulated by RyhB in *E. coli* were found to be differentially expressed in a *ryhB* mutant of *Vibrio*. Thus, despite limited sequence similarity, the two RNAs can be assumed to be functional homologs.

If primary sequence conservation turns out to be a poor guide for finding a long-lost sRNA relative, what will be the alternative? Functional RNAs often maintain a particular secondary structure that is not necessarily reflected by primary sequence conservation. For example, homologous sequences of *E. coli* 6S RNA were, until recently, only known in the γ-subdivision of proteobacteria. 6S RNA folds into a rod-like structure that is required for binding to the σ^{70}-containing RNAP holoenzyme (Wassarman and Storz, 2000; Trotochaud and Wassarman, 2005). This extended hairpin structure is thought to mimic the DNA template of the open promoter complex of RNAP (Barrick et al., 2005; Trotochaud and Wassarman, 2005). A comprehensive screen recently identified multiple 6S RNA homologs in >100 bacterial species of diverse eubacterial lineages (Barrick et al., 2005). This screen was based on a covariance model of 6S RNA structure, a probabilistic model for flexible description of secondary structure and primary sequence consensus of an RNA family, which was used as a query in sequence databases searches for additional members of the RNA family of interest. The search also indicated that a number of abundant *Bacillus* and *Synechococcus* RNAs of previously unknown function are in fact 6S RNA homologs (Watanabe et al., 1997; Ando et al., 2002; Suzuma et al., 2002). These *Bacillus* RNAs were inde-

pendently confirmed as 6S RNA homologs through biochemical analysis by others (Trotochaud and Wassarman, 2005). Strikingly, primary sequence similarity between *E. coli* 6S RNA and each of the two *Bacillus* species is as low as ~46%, yet they qualify as structural homologs because key secondary structure elements and functionally important residues are conserved. Multiple iterations of alignments of high-scoring 6S RNA candidate loci, as well as inclusion of candidates with slightly deviating structures, significantly improved the covariance model and led to a final curated alignment that contained 121 sequences. The final covariance model predicts hundreds of additional 6S RNA sequences in microbial genomes (Barrick et al., 2005).

Other than conserved RNA structure, regulatory aspects of an sRNA gene could give hints as to how to find homologous genes in either the same or a different organism. We have already discussed that a Fur box upstream of *ryhB* helped to identify the PrrF RNAs in *Pseudomonas aeruginosa* (see previous section). Similarly, a short palindromic element played a role in finding the second RNA that counteracted the global regulator, RsmA, in *Pseudomonas* strains (Valverde et al., 2003). Following the observation that inactivation of the primary *rsmZ* gene was not enough to abrogate repression of the RsmA target genes, the existence of an additional, hidden RsmZ-like RNA was hypothesized. Several *Pseudomonas* genomes carry a well-conserved thermoregulatory region (TRR) that shares striking similarity with the palindromic upstream activating region of the *rsmZ* promoter. In addition, the TRR locus contains a number of GGA repeats that are the hallmark of CsrA/RsmA-antagonizing RNAs (Romeo, 1998; Valverde et al., 2004). Probing of the TRR locus in *P. fluorescens* CHAO revealed the ca. 120-nt RsmY RNA, which was subsequently shown to act in concert with RsmZ on RsmA protein (Valverde et al., 2003).

Finally, homology searches should take into consideration the fact that, like proteins, some regulatory RNAs carry functional domains that can become scrambled or even split into individual pieces that are then encoded by separate loci. For example, circularly permuted tmRNA genes were recently reported to exist in *Caulobacter* and cyanobacteria. In these cases, a segment normally found at the 3′ end of tmRNA genes is located upstream of the segment normally at the 5′ end. The two conserved tmRNA segments are linked by a short non-conserved sequence (Keiler et al., 2000; Gaudin et al., 2002) that is excised post-transcriptionally to yield the standard tRNA acceptor stem terminus, thus resulting in a two-piece mature tmRNA.

Microarray detection: sRNAs lighten up

Microarrays have become the method of choice for monitoring mRNA expression profiles at the genome level. They could also represent a means for studying sRNA expression or even for finding new sRNA transcripts. The main caveat in this case, however, seems to be current microarray design: most of the commercially available arrays are limited to ORFs and at best include tRNA and rRNA genes. In other words, transcripts from IGRs, where most of the sRNA genes reside, will not be detected. This aspect of microarrays is being improved. Selinger et al. (2000) introduced a high-density oligonucleotide probe array for *E. coli* that not only carries strand-specific probes for all mRNA, tRNA, and rRNA regions, but also covers IGRs of >40 bp; the latter regions have one probe for every six bases on both strands. Although this study primarily focused on the sensitivity and reproducibility of mRNA-level profiling, it provided preliminary data on some intergenic and antisense RNAs that were detected en passant. Subsequently, Wassarman et al. (2001) supplemented their biocomputational sRNA prediction by using the same type of array and specifically analyzing intergenic transcriptional output.

These authors discussed several problems associated with reliable detection of small structured RNAs on microarrays, including probe preparation. Frequently, only a few oligonucleotides within the range of a given sRNA transcript region yield a signal peak, even though the same sRNA locus gives a strong and distinct band on Northern blots (compare Figures 2 and 3 in Wassarman et al., 2001). These technical difficulties notwithstanding, microarray signals paralleled Northern blot results for at least one-third of the confirmed sRNAs. In another study that used the same *E. coli* high-density arrays, transcripts were occasionally detected on the strand opposite an experimentally validated sRNA region (Tjaden et al., 2002). Whether these signals account for as yet unknown sRNA antisense transcripts or represent experimental noise remains undetermined.

The power of these microarrays cannot be overstated. The extraordinarily high probe density may facilitate detection of 3′ or 5′ UTR RNA fragments that accumulate after the processing of mRNA transcripts (Tjaden et al., 2002). The same study, however, emphasized the need to validate initial microarray results by independent experimental techniques such as Northern hybridization or RNase protection assays. Nine new intergenic sRNAs were predicted based on their transcript intensity and growth-dependent expression. There also appeared to be some sequence homology of these *E. coli* loci with corresponding regions in *Salmonella typhimurium*. However, to create sRNA deletion strains in *Salmonella*, we revisited these nine candidates and failed to identify statistically significant sequence similarity for six of them (C0614, C0465, C0299, C0293, C0719, C0067), while another two seem to represent highly repetitive elements (C0664, C0362).

We nevertheless expect that microarrays will become a standard tool for both the identification and expression profiling of bacterial sRNAs. Species-specific arrays with at least some IGR coverage (>150 bp) are now available for other model bacteria, including *B. subtilis*, *Pseudomonas aeruginosa*, and *Staphylococcus aureus* (C. Rosenow, Affymetrix, personal communication). In addition, the problems inherent to labeling and cDNA synthesis of small structured RNAs to be used as samples on microarrays can be bypassed if alternative RNA detection methods are used. In such one approach, Zhang et al. (2003) immunoprecipitated Hfq protein (see below) and

directly detected the sRNAs bound to this protein on microarrays with antibodies specific for RNA:DNA hybrids. The highly improved sensitivity of this method is demonstrated by the detection OxyS RNA, which is present in very low concentrations under the growth conditions used in this study. Unfortunately, as of the time of writing, it is unclear whether these antibodies will become commercially available.

Co-purification with proteins: looking for intimate pairs

In a cellular environment, many sRNAs can be assumed to form complexes with proteins at various points in their life cycle. RNA-binding proteins may help an sRNA fold into its active conformation, shield it from nucleases prior to reaching a target, or promote its annealing with target mRNAs. Other sRNAs interact with proteins to directly regulate their activity. With these scenarios in mind, purification of ribonucleoprotein complexes via the protein partners should facilitate the isolation of new sRNAs. Indeed, several sRNAs have been discovered by co-purification with the abundant RNA-binding proteins CsrA, Hfq, and with RNA polymerase. In *E. coli*, His-tagged CsrA protein was observed to purify as a globular ribonucleoprotein complex that, in addition to 18 CsrA subunits, contained a 366-nt RNA (Liu et al., 1997). This unknown RNA, later named CsrB, was poly(A)-tailed, cloned as cDNA, sequenced and mapped to the *E. coli* genome by BLASTN searches. Similarly, RsmZ RNA of *Pseudomonas fluorescens* was later identified through co-purification with the CsrA homolog of this organism, RsmA (Heeb et al., 2002). The two RNAs, CsrB and RmsZ, are functional homologs since their primary function lies in antagonizing CsrA and RsmA, respectively. As mentioned before (see genetic approaches), the genes of the CsrB-like RNA family show poor sequence similarity. Hence, co-purification with interacting proteins could be an alternative strategy for finding sRNAs with primary sequences that have diverged to such an extent that they are unidentifiable by BLASTN searches.

To date, the Sm-like protein, Hfq, is the best characterized sRNA binder and has been suggested to interact with more than one-third of the sRNAs known in *E. coli* (Zhang et al., 2003; Valentin-Hansen et al., 2004). Immunoprecipitation of Hfq followed by microarray detection of the RNAs bound to the protein was recently used as a new screening approach. In this way, half a dozen new sRNAs were found in *E. coli* (Zhang et al., 2003). Similar approaches for finding Hfq-binding sRNAs have been taken for the bacterial pathogens *Listeria monocytogenes* and *Pseudomonas aeruginosa* (U. Bläsi and B. Kallipolitis, personal communication). Interestingly, *hfq* deletion strains of these bacteria show reduced virulence (Sonnleitner et al., 2003; Christiansen et al., 2004). There is also an earlier report that antibodies against *eukaryotic* snRNP (small nuclear ribonucleoprotein) immunoprecipitated small RNAs (140–240 nt) in the cyanobacterium *Synechococcus leopoliensis* and in *B. subtilis* (Kovacs et al., 1993). It is tempting to speculate that these antibodies actually precipitated *Synechococcus* and *Bacillus* Hfq homologs, which may share epitopes with the Sm proteins of eukaryotic snRNPs. Somewhat weakening this speculation is the fact that no sRNAs were co-immunoprecipitated in *E. coli* extracts. Unfortunately, this interesting observation does not seem to have been followed up, and the RNAs remain unidentified.

Since species-specific Hfq antibodies may not always be available, affinity tags such as FLAG and $(His)_6$ provide epitopes that can be targeted by specific antibodies. Alternatively, RNAs may be cloned directly from Hfq protein purified via its affinity tag. As a proof of principle, *Vibrio cholera* RyhB RNA was observed to remain bound to His-tagged Hfq in the course of affinity purification on nickel agarose columns (Davis et al., 2005).

All of the aforementioned studies isolated Hfq-binding RNAs from bacterial extracts, and hence required that the sRNAs be expressed under the condition sampled. The Schroeder laboratory, however, opted for an *in vitro* SELEX approach that would in principle cover all sRNAs that are encoded by a given genome (C. Lorenz and R. Schroeder, personal communication). Here, a representative library of the *E. coli* genome was constructed from random sequences of 50–500 bp (Singer et al., 1997). These fragments were *in vitro* transcribed with T7 RNA polymerase, incubated with Hfq, and selected for Hfq binding on filters. Taking the standard SELEX route, the retained RNA was converted into cDNA and subjected to additional (eight) rounds of selection, resulting in a pool of RNAs that bound Hfq with K_d values of 5–50 nM. Subsequently, specific Hfq interaction of the thus enriched RNAs was determined *in vivo* using a yeast three-hybrid screen (Bernstein et al., 2002). Preliminary results suggest that these experiments identified a number of novel Hfq-binding RNAs, including antisense RNAs and candidate sRNAs from intergenic regions.

Aside from CsrA and Hfq, there are many more RNA-binding proteins that could also be exploited for the purpose of identifying new sRNAs. Specifically, *Bacillus* homologs of 6S RNA were identified by means of antibodies that targeted either the α subunit of RNAP or the primary sigma factor, σ^A (Trotochaud and Wassarman, 2005). 6S RNA had previously been shown to associate with RNAP in *E. coli* (Wassarman and Storz, 2000). Examples of other abundant proteins that form complexes with RNA and/or DNA and, like Hfq, are present in $\geq$10 000 copies per cell, include the RNA chaperone StpA, the histone-like protein HU, and the transcriptional regulator H-NS (Ali et al., 1999). In addition, CspA and CspA-like cold shock proteins comprise a large family of RNA-binding proteins that are involved in post-transcriptional regulation in bacteria. Whether all these proteins have intimate relationships with small RNAs remains to be determined. However, this strategy holds promise: when photolyases (proteins involved in DNA damage repair) of *Vibrio cholerae* were purified, it was observed that one of these proteins invariably associates with ~70-nt RNA(s) of yet unknown sequence and function (Worthington et al., 2003).

Conclusions

In many respects, sRNA screens in the model bacterium *Escherichia coli* have set a blueprint for the global and

functional identification of sRNAs in the many microbes that have yet to be studied. Which approach to use when embarking on a new global search, or when hunting for a specific sRNA, will depend on a researcher's taste and laboratory design. Considering that significant overlap exists in the methodology used so far, researchers interested in bacteria should also keep an eye on the techniques being developed to identify non-coding RNA regulators in eukaryotes and archaebacteria. For example, the base composition analysis that successfully identified non-coding RNAs in AT-rich archaebacteria (Schattner, 2002; Klein et al., 2002) may well be applicable to hyperthermophilic eubacteria. We also expect that the rapidly growing list of completed bacterial genomes will fill the phylogenetic gaps between distantly related organisms, and will thereby greatly facilitate sRNA discovery using conservation criteria. Hopefully, future genome annotations will increasingly take into consideration the fact that conserved genomic loci with poor coding potential are likely to harbor genes that encode small RNAs. Along similar lines, it will be interesting to observe how many of the hitherto annotated species-specific ORFs of unknown function in fact express non-coding RNAs.

Much effort is now devoted to dissecting the functional roles of the recently discovered sRNAs. By studying the details of how bacterial small RNAs exert their regulatory functions at the molecular level, we will learn more about structure-function relationships, which could help to improve sRNA predictions.

For the moment, a unified nomenclature for bacterial sRNAs remains an unresolved issue. Many *E. coli* sRNAs were discovered in parallel in different laboratories, and hence named independently (for examples, see Table 1 in Hershberg et al., 2003). Introducing a nomenclature similar to the system established for eukaryotic microRNAs (Ambros et al., 2003) could be one way to get around the confusion that is often caused by multiple names for the same sRNA. This is also desirable as we anticipate the discovery of many more sRNAs by global screens in other model bacteria. As for *E. coli*, Wassarman et al. (2001) introduced a nomenclature to name sRNAs of unknown function in a manner similar to the protein-coding ORFs of unidentified function. The fact that genomic annotation of sRNA genes and candidate loci is lagging is both good and bad. One argument in favor of a delay is that the vast majority of predicted sRNAs indeed require further analysis before being regarded as functional RNAs. However, if a genomic locus with no apparent protein-coding potential emerges in a functional screen unrelated to non-coding RNAs, it would certainly help if there were a quick link to the list of the ca. 1000 potential sRNA candidates. Outside of standard genome annotations, the Rfam database (http://www.sanger.ac.uk/Software/Rfam/; Griffiths-Jones et al., 2003) keeps track of bacterial sRNAs.

Within the last 5 years, the number of small non-coding RNAs known in bacteria has increased in a way that can be likened to an explosion. Adding to this is the diversity of regulatory roles that old and new sRNAs have been found to play. Nevertheless, our understanding of sRNA-mediated regulation in bacteria is still limited and this field promises to keep us busy searching for new sRNA molecules and their functions in the years to come.

Acknowledgments

We thank many colleagues for helpful discussions and sharing unpublished data, and Allison A. Henry and Alexandra Sittka for critical comments on the manuscript. Cynthia Sharma is supported by DFG grant VO 875/1-1 (Deutsche Forschungsgemeinschaft).

References

Aarons, S., Abbas, A., Adams, C., Fenton, A., and O'Gara, F. (2000). A regulatory RNA (PrrB RNA) modulates expression of secondary metabolite genes in *Pseudomonas fluorescens* F113. J. Bacteriol. *182*, 3913–3919.

Ali, A.T., Iwata, A., Nishimura, A., Ueda, S., and Ishihama, A. (1999). Growth phase-dependent variation in protein composition of the *Escherichia coli* nucleoid. J. Bacteriol. *181*, 6361–6370.

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.

Ambros, V., Bartel, B., Bartel, D.P., Burge, C.B., Carrington, J.C., Chen, X., Dreyfuss, G., Eddy, S.R., Griffiths-Jones, S., Marshall, M., et al. (2003). A uniform system for microRNA annotation. RNA *9*, 277–279.

Ando, Y., Asari, S., Suzuma, S., Yamane, K., and Nakamura, K. (2002). Expression of a small RNA, BS203 RNA, from the *yocI–yocJ* intergenic region of *Bacillus subtilis* genome. FEMS Microbiol. Lett. *207*, 29–33.

Antal, M., Bordeau, V., Douchin, V., and Felden, B. (2005). A small bacterial RNA regulates a putative ABC transporter. J. Biol. Chem. *280*, 7901–7908.

Argaman, L. and Altuvia, S. (2000). *fhlA* repression by OxyS RNA: kissing complex formation at two sites results in a stable antisense-target RNA complex. J. Mol. Biol. *300*, 1101–1112.

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.

Axmann, I.M., Kensche, P., Vogel, J., Kohl, S., Herzel, H., and Hess, W.R. (2005). Identification of cyanobacterial non-coding RNAs by comparative genome analysis. Genome Biol. *6*, R73.

Barrick, J.E., Sudarsan, N., Weinberg, Z., Ruzzo, W.L., and Breaker, R.R. (2005). 6S RNA is a widespread regulator of eubacterial RNA polymerase that resembles an open promoter. RNA *11*, 774–784.

Bernstein, D.S., Buter, N., Stumpf, C., and Wickens, M. (2002). Analyzing mRNA-protein complexes using a yeast three-hybrid system. Methods *26*, 123–141.

Blattner, F.R., Plunkett, G. III, Bloch, C.A., Perna, N.T., Burland, V., Riley, M., Collado-Vides, J., Glasner, J.D., Rode, C.K., Mayhew, G.F., et al. (1997). The complete genome sequence of *Escherichia coli* K-12. Science *277*, 1453–1474.

Brownlee, G.G. (1971). Sequence of 6S RNA of *E. coli*. Nat. New Biol. *229*, 147–149.

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.

Chen, S., Zhang, A.X., Blyn, L.B., and Storz, G. (2004). MicC, a second small-RNA regulator of Omp protein expression in *Escherichia coli*. J. Bacteriol. *186*, 6689–6697.

Christiansen, J.K., Larsen, M.H., Ingmer, H., Sogaard-Andersen, L., and Kallipolitis, B.H. (2004). The RNA-binding protein Hfq of *Listeria monocytogenes*: role in stress tolerance and virulence. J. Bacteriol. *186*, 3355–3362.

Davis, B.M., Quinones, M., Pratt, J., Ding, Y., and Waldor, M.K. (2005). Characterization of the small untranslated RNA RyhB and its regulon in *Vibrio cholerae*. J. Bacteriol. *187*, 4005–4014.

Ebeling, S., Kundig, C., and Hennecke, H. (1991). Discovery of a rhizobial RNA that is essential for symbiotic root nodule development. J. Bacteriol. *173*, 6373–6382.

Elbashir, S.M., Lendeckel, W., and Tuschl, T. (2001). RNA interference is mediated by 21- and 22-nucleotide RNAs. Genes Dev. *15*, 188–200.

Franze de Fernandez, M.T., Hayward, W.S., and August, J.T. (1972). Bacterial proteins required for replication of phage Q ribonucleic acid. Purification and properties of host factor I, a ribonucleic acid-binding protein. J. Biol. Chem. *247*, 824–831.

Gaudin, C., Zhou, X., Williams, K.P., and Felden, B. (2002). Two-piece tmRNA in cyanobacteria and its structural analysis. Nucleic Acids Res. *30*, 2018–2024.

Geissmann, T.A. and Touati, D. (2004). Hfq, a new chaperoning role: binding to messenger RNA determines access for small RNA regulator. EMBO J. *23*, 396–405.

Göhlmann, H.W.H., Weiner, J., Schön, A., and Herrmann, R. (2000). Identification of a small RNA within the *pdh* gene cluster of *Mycoplasma pneumoniae* and *Mycoplasma genitalium*. J. Bacteriol. *182*, 3281–3284.

Gottesman, S. (2004). The small RNA regulators of *Escherichia coli*: roles and mechanisms. Annu. Rev. Microbiol. *58*, 303–328.

Griffin, B.E. (1971). Separation of ^{32}P-labelled ribonucleic acid components. The use of polyethylenimine-cellulose (TLC) as a second dimension in separating oligoribonucleotides of '4.5 S' and 5 S from *E. coli*. FEBS Lett. *15*, 165–168.

Griffiths-Jones, S., Bateman, A., Marshall, M., Khanna, A., and Eddy, S.R. (2003). Rfam: an RNA family database. Nucleic Acids Res. *31*, 439–441.

Guigueno, A., Dassa, J., Belin, P., and Boquet, P.L. (2001). Oversynthesis of a new *Escherichia coli* small RNA suppresses export toxicity of DsbA′-PhoA unfoldable periplasmic proteins. J. Bacteriol. *183*, 1147–1158.

Heeb, S., Blumer, C., and Haas, D. (2002). Regulatory RNA as mediator in GacA/RsmA-dependent global control of exoproduct formation in *Pseudomonas fluorescens CHA0*. J. Bacteriol. *184*, 1046–1056.

Hershberg, R., Altuvia, S., and Margalit, H. (2003). A survey of small RNA-encoding genes in *Escherichia coli*. Nucleic Acids Res. *31*, 1813–1820.

Hindley, J. (1967). Fractionation of ^{32}P-labelled ribonucleic acids on polyacrylamide gels and their characterization by fingerprinting. J. Mol. Biol. *30*, 125–136.

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.

Huang, C., Wolfgang, M.C., Withey, J., Koomey, M., and Friedman, D.I. (2000). Charged tmRNA but not tmRNA-mediated proteolysis is essential for *Neisseria gonorrhoeae* viability. EMBO J. *19*, 1098–1107.

Huntzinger, E., Boisset, S., Saveanu, C., Benito, Y., Geissmann, T., Namane, A., Lina, G., Etienne, J., Ehresmann, B., Ehresmann, C., et al. (2005). *Staphylococcus aureus* RNAIII and the endoribonuclease III coordinately regulate spa gene expression. EMBO J. *24*, 824–835.

Hüttenhofer, A., Kiefmann, M., Meier-Ewert, S., O'Brien, J., Lehrach, H., Bachellerie, 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.

Ikemura, T. and Dahlberg, J.E. (1973a). Small ribonucleic acids of *Escherichia coli*. 1. Characterization by polyacrylamide-gel electrophoresis and fingerprint analysis. J. Biol. Chem. *248*, 5024-5032.

Ikemura, T. and Dahlberg, J.E. (1973b). Small ribonucleic acids of *Escherichia coli*. 2. Noncoordinate accumulation during stringent control. J. Biol. Chem. *248*, 5033–5041.

Johansson, J. and Cossart, P. (2003). RNA-mediated control of virulence gene expression in bacterial pathogens. Trends Microbiol. *11*, 280–285.

Kawano, M., Oshima, T., Kasai, H., and Mori, H. (2002). Molecular characterization of long direct repeat (LDR) sequences expressing a stable mRNA encoding for a 35-amino acid cell-killing peptide and a cis-encoded small antisense RNA in *Escherichia coli*. Mol. Microbiol. *45*, 333–349.

Kawano, M., Reynolds, A.A., Miranda-Rios, J., and Storz, G. (2005). Detection of 5′- and 3′-UTR-derived small RNAs and *cis*-encoded antisense RNAs in *Escherichia coli*. Nucleic Acids Res. *33*, 1040–1050.

Keiler, K.C., Waller, P.R., and Sauer, R.T. (1996). Role of a peptide tagging system in degradation of proteins synthesized from damaged messenger RNA. Science *271*, 990–993.

Keiler, K.C., Shapiro, L., and Williams, K.P. (2000). tmRNAs that encode proteolysis-inducing tags are found in all known bacterial genomes: a two-piece tmRNA functions in Caulobacter. Proc. Natl. Acad. Sci. USA *97*, 7778–7783.

Kim, K.S. and Lee, Y. (2004). Regulation of 6S RNA biogenesis by switching utilization of both sigma factors and endoribonucleases. Nucleic Acids Res. *32*, 6057–6068.

Klein, R.J., Misulovin, Z., and Eddy, S.R. (2002). Noncoding RNA genes identified in AT-rich hyperthermophiles. Proc. Natl. Acad. Sci. USA *99*, 7542–7547.

Kovacs, S.A., Oneil, J., Watcharapijarn, J., Moekirvan, C., Vijay, S., and Silva, V. (1993). Eubacterial components similar to small nuclear ribonucleoproteins – identification of immunoprecipitable proteins and capped RNAs in a cyanobacterium and a Gram-positive eubacterium. J. Bacteriol. *175*, 1871–1878.

Kreikemeyer, B., Boyle, M.D., Buttaro, B.A., Heinemann, M., and Podbielski, A. (2001). Group A streptococcal growth phase-associated virulence factor regulation by a novel operon (Fas) with homologies to two-component-type regulators requires a small RNA molecule. Mol. Microbiol. *39*, 392–406.

Lagos-Quintana, M., Rauhut, R., Lendeckel, W., and Tuschl, T. (2001). Identification of novel genes coding for small expressed RNAs. Science *294*, 853–858.

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.

Lease, R.A. and Belfort, M. (2000). A trans-acting RNA as a control switch in *Escherichia coli*: DsrA modulates function by forming alternative structures. Proc. Natl. Acad. Sci. USA *97*, 9919–9924.

Lease, R.A., Cusick, M.E., and Belfort, M. (1998). Riboregulation in *Escherichia coli*: DsrA RNA acts by RNA:RNA interactions at multiple loci. Proc. Natl. Acad. Sci. USA *95*, 12456–12461.

Lee, R.C. and Ambros, V. (2001). An extensive class of small RNAs in *Caenorhabditis elegans*. Science *294*, 862–864.

Lenz, D.H., Mok, K.C., Lilley, B.N., Kulkarni, R.V., Wingreen, N.S., and Bassler, B.L. (2004). The small RNA chaperone Hfq and multiple small RNAs control quorum sensing in *Vibrio harveyi* and *Vibrio cholerae*. Cell *118*, 69–82.

Liu, M.Y., Gui, G., Wei, B., Preston, J.F. III, Oakford, L., Yuksel, U., Giedroc, D.P., and Romeo, T. (1997). The RNA molecule CsrB binds to the global regulatory protein CsrA and antagonizes its activity in *Escherichia coli*. J. Biol. Chem. *272*, 17502–17510.

Liu, Y., Cui, Y., Mukherjee, A., and Chatterjee, A.K. (1998). Characterization of a novel RNA regulator of *Erwinia carotovora* ssp. *carotovora* that controls production of extracellular enzymes and secondary metabolites. Mol. Microbiol. *29*, 219–234.

Ma, W.L., Cui, Y., Liu, Y., Dumenyo, C.K., Mukherjee, A., and Chatterjee, A.K. (2001). Molecular characterization of global regulatory RNA species that control pathogenicity factors in *Erwinia amylovora* and *Erwinia herbicola* pv. *gypsophilae*. J. Bacteriol. *183*, 1870–1880.

Majdalani, N., Cunning, C., Sledjeski, D., Elliott, T., and Gottesman, S. (1998). DsrA RNA regulates translation of RpoS message by an anti-antisense mechanism, independent of its action as an antisilencer of transcription. Proc. Natl. Acad. Sci. USA *95*, 12462–12467.

Majdalani, N., Chen, S., Murrow, J., John, K.S., and Gottesman, S. (2001). Regulation of RpoS by a novel small RNA: the characterization of RprA. Mol. Microbiol. *39*, 1382–1394.

Mangold, M., Siller, M., Roppenser, B., Vlaminckx, B.J., Penfound, T.A., Klein, R., Novak, R., Novick, R.P., and Charpentier, E. (2004). Synthesis of group A streptococcal virulence factors is controlled by a regulatory RNA molecule. Mol. Microbiol. *53*, 1515–1527.

Marker, C., Zemann, A., Terhorst, T., Kiefmann, M., Kastenmayer, J.P., Green, P., Bachellerie, J.P., Brosius, J., and Hüttenhofer, A. (2002). Experimental RNomics: identification of 140 candidates for small non-messenger RNAs in the plant *Arabidopsis thaliana*. Curr. Biol. *12*, 2002–2013.

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.

Massé, E., Escorcia, F.E., and Gottesman, S. (2003). Coupled degradation of a small regulatory RNA and its mRNA targets in *Escherichia coli*. Genes Dev. *17*, 2374–2383.

Mattick, J.S. (2001). Non-coding RNAs: the architects of eukaryotic complexity. EMBO Rep. *2*, 986–991.

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.

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.

Mizuno, T., Chou, M.Y., and Inouye, M. (1983). Regulation of gene expression by a small RNA transcript (MicRNA) in *Escherichia coli* K-12. Proc. Jpn. Acad. Ser. B Phys. Biol. Sci. *59*, 335–338.

Mizuno, T., Chou, M.Y., and Inouye, M. (1984). A unique mechanism regulating gene expression – translational inhibition by a complementary RNA transcript (micRNA). Proc. Natl. Acad. Sci. USA *81*, 1966–1970.

Møller, T., Franch, T., Udesen, C., Gerdes, K., and Valentin-Hansen, P. (2002). Spot 42 RNA mediates discoordinate expression of the *E. coli* galactose operon. Genes Dev. *16*, 1696–1706.

Muffler, A., Fischer, D., and Hengge-Aronis, R. (1996). The RNA-binding protein HF-I, known as a host factor for phage Qb RNA replication, is essential for rpoS translation in *Escherichia coli*. Genes Dev. *10*, 1143–1151.

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.

Ochsner, U.A., Wilderman, P.J., Vasil, A.I., and Vasil, M.L. (2002). GeneChip expression analysis of the iron starvation response in *Pseudomonas aeruginosa*: identification of novel pyoverdine biosynthesis genes. Mol. Microbiol. *45*, 1277–1287.

Opdyke, J.A., Kang, J.G., and Storz, G. (2004). GadY, a small-RNA regulator of acid response genes in *Escherichia coli*. J. Bacteriol. *186*, 6698–6705.

Östberg, Y., Bunikis, I., Bergström, S., and Johansson, J. (2004). The etiological agent of Lyme disease, *Borrelia burgdorferi*, appears to contain only a few small RNA molecules. J. Bacteriol. *186*, 8472–8477.

Pedersen, K. and Gerdes, K. (1999). Multiple hok genes on the chromosome of *Escherichia coli*. Mol. Microbiol. *32*, 1090–1102.

Pichon, C. and Felden, B. (2003). Intergenic sequence inspector: searching and identifying bacterial RNAs. Bioinformatics *19*, 1707–1709.

Poritz, M.A., Bernstein, H.D., Strub, K., Zopf, D., Wilhelm, H., and Walter, P. (1990). An *E. coli* ribonucleoprotein containing 4.5S RNA resembles mammalian signal recognition particle. Science 250, 1111–1117.

Ribes, V., Romisch, K., Giner, A., Dobberstein, B., and Tollervey, D. (1990). *E. coli* 4.5S RNA is part of a ribonucleoprotein particle that has properties related to signal recognition particle. Cell *63*, 591–600.

Rivas, E. and Eddy, S.R. (2000). Secondary structure alone is generally not statistically significant for the detection of noncoding RNAs. Bioinformatics *16*, 583–605.

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.

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.

Romeo, T. (1998). Global regulation by the small RNA-binding protein CsrA and the non-coding RNA molecule CsrB. Mol. Microbiol. *29*, 1321–1330.

Romeo, T., Moore, J., and Smith, J. (1991). A simple method for cloning genes involved in glucan biosynthesis: isolation of structural and regulatory genes for glycogen synthesis in *Escherichia coli*. Gene *108*, 23–29.

Saetrom, P., Sneve, R., Kristiansen, K.I., Snøve, O. Jr., Grünfeld, T., Rognes, T., and Seeberg, E. (2005). Predicting non-coding RNA genes in *Escherichia coli* with boosted genetic programming. Nucleic Acids Res. *33*, 3263–3270.

Schattner, P. (2002). Searching for RNA genes using base-composition statistics. Nucleic Acids Res. *30*, 2076–2082.

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. Nat. Biotechnol. *18*, 1262–1268.

Shimizu, T., Yaguchi, H., Ohtani, K., Banu, S., and Hayashi, H. (2002). Clostridial VirR/VirS regulon involves a regulatory RNA molecule for expression of toxins. Mol. Microbiol. *43*, 257–265.

Singer, B.S., Shtatland, T., Brown, D., and Gold, L. (1997). Libraries for genomic SELEX. Nucleic Acids Res. *25*, 781–786.

Sledjeski, D. and Gottesman, S. (1995). A small RNA acts as an antisilencer of the H-NS-silenced *rcsA* gene of *Escherichia coli*. Proc. Natl. Acad. Sci. USA *92*, 2003–2007.

Sledjeski, D.D., Gupta, A., and Gottesman, S. (1996). The small RNA, DsrA, is essential for the low temperature expression of RpoS during exponential growth in *Escherichia coli*. EMBO J. *15*, 3993–4000.

Sledjeski, D.D., Whitman, C., and Zhang, A. (2001). Hfq is necessary for regulation by the untranslated RNA DsrA. J. Bacteriol. *183*, 1997–2005.

Sonnleitner, E., Hagens, S., Rosenau, F., Wilhelm, S., Habel, A., Jager, K.E., and Bläsi, U. (2003). Reduced virulence of a *hfq* mutant of *Pseudomonas aeruginosa* O1. Microb. Pathog. *35*, 217–228.

Stark, B.C., Kole, R., Bowman, E.J., and Altman, S. (1978). Ribonuclease P: an enzyme with an essential RNA component. Proc. Natl. Acad. Sci. USA *75*, 3717–3721.

Storz, G., Opdyke, J.A., and Zhang, A.X. (2004). Controlling mRNA stability and translation with small, noncoding RNAs. Curr. Opin. Microbiol. *7*, 140–144.

Suzuma, S., Asari, S., Bunai, K., Yoshino, K., Ando, Y., Kakeshita, H., Fujita, M., Nakamura, K., and Yamane, K. (2002). Identification and characterization of novel small RNAs in the

aspS–yrvM intergenic region of the *Bacillus subtilis* genome. Microbiology *148*, 2591–2598.

Tang, T.H., Bachellerie, J.P., Rozhdestvensky, T., Bortolin, M.L., Huber, H., Drungowski, M., Elge, T., Brosius, J., and Hüttenhofer, A. (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., Polacek, N., Zywicki, M., Huber, H., Brugger, K., Garrett, R., Bachellerie, J.P., and Hüttenhofer, A. (2005). Identification of novel non-coding RNAs as potential antisense regulators in the archaeon *Sulfolobus solfataricus*. Mol. Microbiol. *55*, 469–481.

Thompson, J.D., Higgins, D.G., and Gibson, T.J. (1994). CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res. *22*, 4673–4680.

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.

Trotochaud, A.E. and Wassarman, K.M. (2004). 6S RNA function enhances long-term cell survival. J. Bacteriol. *186*, 4978–4985.

Trotochaud, A.E. and Wassarman, K.M. (2005). A highly conserved 6S RNA structure is required for regulation of transcription. Nat. Struct. Mol. Biol. *12*, 313–319.

Udekwu, K.I., Darfeuille, F., Vogel, J., Reimegård, J., Holmqvist, E., and Wagner, E.G.H. (2005). Hfq-dependent regulation of OmpA synthesis is mediated by an antisense RNA. Genes Dev. *19*, 2355–2366.

Ulbrandt, N.D., Newitt, J.A., and Bernstein, H.D. (1997). The *E. coli* signal recognition particle is required for the insertion of a subset of inner membrane proteins. Cell *88*, 187–196.

Valentin-Hansen, P., Eriksen, M., and Udesen, C. (2004). The bacterial Sm-like protein Hfq: a key player in RNA transactions. Mol. Microbiol. *51*, 1525–1533.

Valverde, C., Heeb, S., Keel, C., and Haas, D. (2003). RsmY, a small regulatory RNA, is required in concert with RsmZ for GacA-dependent expression of biocontrol traits in *Pseudomonas fluorescens CHA0*. Mol. Microbiol. *50*, 1361–1379.

Valverde, C., Lindell, M., Wagner, E.G.H., and Haas, D. (2004). A repeated GGA motif is critical for the activity and stability of the riboregulator RsmY of *Pseudomonas fluorescens*. J. Biol. Chem. *279*, 25066–25074.

Vanderpool, C.K. and Gottesman, S. (2004). Involvement of a novel transcriptional activator and small RNA in post-transcriptional regulation of the glucose phosphoenolpyruvate phosphotransferase system. Mol. Microbiol. *54*, 1076–1089.

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.

Vogel, J. and Wagner, E.G.H. (2005). RNA mining. In: Handbook of RNA Biochemistry, R. Hartmann, R. Bindereif, A. Schön, and E. Westhof, eds. (Weinheim, Germany: Wiley-VCH), pp. 595–613.

Vogel, J., Bartels, V., Tang, T.H., Churakov, G., Slagter-Jäger, J.G., Hüttenhofer, A., and Wagner, E.G.H. (2003). RNomics in *Escherichia coli* detects new sRNA species and indicates parallel transcriptional output in bacteria. Nucleic Acids Res. *31*, 6435–6443.

Vogel, J., Argaman, L., Wagner, E.G.H., and Altuvia, S. (2004). The small RNA IstR inhibits synthesis of an SOS-induced toxic peptide. Curr. Biol. *14*, 2271–2276.

Wagner, E.G.H. and Vogel, J. (2003). Noncoding RNAs encoded by bacterial chromosomes. In: Noncoding RNAs, J. Barciszewski and V. Erdmann, eds. (Georgetown, TX, USA: Landes Bioscience), pp. 243–259.

Wagner, E.G.H., Altuvia, S., and Romby, P. (2002). Antisense RNAs in bacteria and their genetic elements. Adv. Genet. *46*, 361–398.

Washietl, S. and Hofacker, I.L. (2004). Consensus folding of aligned sequences as a new measure for the detection of functional RNAs by comparative genomics. J. Mol. Biol. *342*, 19–30.

Washietl, S., Hofacker, I.L., and Stadler, P.F. (2005). Fast and reliable prediction of noncoding RNAs. Proc. Natl. Acad. Sci. USA *102*, 2454–2459.

Wassarman, K.M. and Storz, G. (2000). 6S RNA regulates *E. coli* RNA polymerase activity. Cell *101*, 613–623.

Wassarman, K.M., Zhang, A., and Storz, G. (1999). Small RNAs in *Escherichia coli*. Trends Microbiol. *7*, 37–45.

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.

Watanabe, T., Sugiura, R., and Sugita, M. (1997). A novel small stable RNA, 6Sa RNA, from the cyanobacterium *Synechococcus* sp. strain PCC6301. FEBS Lett. 416, 302–306.

Watanabe, T., Sugita, M., and Sugiura, M. (1998). Identification of 10Sa RNA (tmRNA) homologues from the cyanobacterium *Synechococcus* sp. strain PCC6301 and related organisms. Biochim. Biophys. Acta *1396*, 97–104.

Weilbacher, T., Suzuki, K., Dubey, A.K., Wang, X., Gudapaty, S., Morozov, I., Baker, C.S., Georgellis, D., Babitzke, P., and Romeo, T. (2003). A novel sRNA component of the carbon storage regulatory system of *Escherichia coli*. Mol. Microbiol. *48*, 657–670.

Wilderman, P.J., Sowa, N.A., FitzGerald, D.J., FitzGerald, P.C., Gottesman, S., Ochsner, U.A., and Vasil, M.L. (2004). Identification of tandem duplicate regulatory small RNAs in *Pseudomonas aeruginosa* involved in iron homeostasis. Proc. Natl. Acad. Sci. USA *101*, 9792–9797.

Willkomm, D.K., Minnerup, J., Hüttenhofer, A., and Hartmann, R.K. (2005). Experimental RNomics in *Aquifex aeolicus*: identification of small non-coding RNAs and the putative 6S RNA homolog. Nucleic Acids Res. *33*, 1949–1960.

Winkler, W., Nahvi, A., and Breaker, R.R. (2002). Thiamine derivatives bind messenger RNAs directly to regulate bacterial gene expression. Nature *419*, 952–956.

Worthington, E.N., Kavakli, I.H., Berrocal-Tito, G., Bondo, B.E., and Sancar, A. (2003). Purification and characterization of three members of the photolyase/cryptochrome family blue-light photoreceptors from *Vibrio cholerae*. J. Biol. Chem. *278*, 39143–39154.

Yuan, G.Z., Klambt, 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.

Zhang, A., Altuvia, S., Tiwari, A., Argaman, L., Hengge-Aronis, R., and Storz, G. (1998). The OxyS regulatory RNA represses *rpoS* translation and binds the Hfq (HF-I) protein. EMBO J. *17*, 6061–6068.

Zhang, A., Wassarman, K.M., Rosenow, C., Tjaden, B.C., Storz, G., and Gottesman, S. (2003). Global analysis of small RNA and mRNA targets of Hfq. Mol. Microbiol. *50*, 1111–1124.

Zhang, Y., Zhang, Z., Ling, L., Shi, B., and Chen, R. (2004). Conservation analysis of small RNA genes in *Escherichia coli*. Bioinformatics *20*, 599–603.