Small regulatory RNAs in Archaea

Julia Babski1, Lisa-Katharina Maier2, Ruth Heyer2, Katharina Jaschinski1, Daniela Prasse3, Dominik Jäger3, Lennart Randau4, *Correspondence to: Jörg Soppa; Email: soppa@bio.uni-frankfurt.de

Small regulatory RNAs (sRNAs) are universally distributed in all three domains of life, Archaea, Bacteria, and Eukaryotes. In bacteria, sRNAs typically function by binding near the translation start site of their target mRNAs and thereby inhibit or activate translation. In eukaryotes, miRNAs and sRNAs typically bind to the 3′-untranslated region (3′-UTR) of their target mRNAs and influence translation efficiency and/or mRNA stability. In archaea, sRNAs have been identified in all species investigated using bioinformatic approaches, RNomeics, and RNA-Seq. Their size can vary significantly between less than 50 to more than 500 nucleotides. Differential expression of sRNA genes has been studied using northern blot analysis, microarrays, and RNA-Seq. In addition, biological functions have been unraveled by genetic approaches, i.e., by characterization of designed mutants. As in bacteria, it was revealed that archaial sRNAs are involved in many biological processes, including metabolic regulation, adaptation to extreme conditions, stress responses, and even in regulation of morphology and cellular behavior. Recently, the first target mRNAs were identified in archaea, including one sRNA that binds to the 5′-region of two mRNAs in Methanosarcina mazei Gs, and a few sRNAs that bind to 3′-UTRs in Sulfolobus solfataricus, three Pyrobaculum species, and Haloferax volcanii, indicating that archaial sRNAs appear to be able to target both the 5′-UTR or the 3′-UTR of their respective target mRNAs. In addition, archaea contain sRNA-derived fragments (sRFs), and one tRF has been identified as a major ribosome-binding sRNA in archaea. sRFs, and one tRF have been identified as a major ribosome-binding RNA in H. volcanii, which downregulates translation in response to stress. Besides regulatory sRNAs, archaea contain further classes of sRNAs, e.g., CRISPR RNAs (crRNAs) and snoRNAs.

Introduction

About 35 y ago it has been proposed that archaea form a third domain of life, phylogenetically separated from the bacteria and eukaryotes.1 This proposal was based on partial 16S/18S rRNA sequences of very few methanogenic archaea, bacteria, and lower eukaryotes, but has been proven to be true after the analysis of thousands additional species and sequencing of hundreds of archaical and bacterial genomes.2-4 Like bacteria, archaea do not contain a nucleus and are thus prokaryotes. In addition, archaea and bacteria share a much higher metabolic diversity compared with eukaryotes, including a large variety of anaerobic energy yielding pathways. On the other hand, archaea and eukaryotes share homologous proteins in many central biological processes to the exclusion of the bacteria, including DNA packaging, replication, transcription, translation, and cell cycle. Given this dichotomy, we find it attractive to compare the results obtained about small RNAs in archaea with the knowledge on sRNAs in bacteria and eukaryotes.

The sRNAs are universally distributed in all three domains of life: archaea, bacteria, and eukaryotes. In 1984, the first sRNA was identified in the genome of Escherichia coli, but it was long thought to be exceptional. However, in recent years it became clear that bacteria typically contain 200–400 sRNAs that can target more than 1000 mRNAs. Typically, bacterial sRNAs are 50–300 nucleotides in length and fulfill numerous important regulatory functions, including stress response, and regulation of virulence genes, carbon source uptake, and metabolism (for reviews see refs. 6–10). Bacterial sRNAs are often encoded in intergenic regions in trans to their target genes and typically function by imperfect base-pairing interactions with their target mRNAs near the ribosomal binding site (RBS). This interaction can repress translation by masking the RBS or can induce translation by making the RBS accessible. Also, additional molecular mechanisms of sRNA functions in bacteria have been described, including destabilization or stabilization of the target mRNAs or the action via specific binding to a protein. Another class of sRNAs are antisense sRNAs (asRNAs), which are encoded in cis on the opposite strand of their target gene, and thus, have full complementarity to their target mRNA. The fraction of genes regulated by asRNAs varies widely in different bacterial species and can be higher than 48%.11 It has been suggested that bacterial sRNAs are responsible for fine-tuning of gene regulation due to the fact that sRNA gene deletion mutants often have no obvious phenotype compared with the parent strain or the phenotype is very mild (reviewed in refs. 8 and 12). Many sRNAs in gram-negative bacteria need the Hfq protein for function, which is required for target mRNA recognition and complex formation (reviewed in ref. 13).

The best-studied eukaryotic sRNAs are microRNAs (miRNAs), small interfering RNAs (siRNAs), and piwi-associated RNAs (piRNAs), which are approximately 20–30 nt in length and are thus
much smaller than bacterial sRNAs. Currently, it is thought that in higher eukaryotes, e.g., humans, more than half of all mRNAs is regulated by these sRNAs in their translation and/or stability.14 In animals, miRNAs typically repress translation of their target mRNAs via imperfect base pairing between miRNAs and the 3′-UTRs. In eukaryotes, the lack of a miRNA often leads to severe phenotypic defects. In humans, miRNA dysfunction or deficiency is associated with a variety of diseases, developmental defects, and cancer formation (reviewed in refs. 15–18). A second class of eukaryotic sRNAs, siRNAs, is often processed from exogenous or endogenous RNA molecules that enter the cell. Typically, siRNAs bind to the 3′-UTRs of target mRNAs, have the ability to form perfect hybrids, and binding results in degradation of the target mRNA. Thus, siRNAs are thought to have evolved as a defense system against RNA viruses (reviewed in refs. 19 and 20). The piwi-interacting RNAs (piRNAs) form complexes with Piwi proteins and are involved in the silencing of transposons during spermatogenesis (reviewed in ref. 21). A different class of eukaryotic sRNAs, the tRNA-derived fragments (tRFs), is generated via the processing of mature tRNAs or precursor tRNAs. Recently, it became clear that tRFs are not only degradation intermediates, but can have important regulatory functions (reviewed in ref. 22). An additional class of eukaryotic sRNAs comprises small nuclear RNAs (snRNAs), which are localized in the nucleolus and guide site-specific modifications of the rRNAs. Based on specific sequence motifs, they are divided into the two groups of C/D box and H/ACA box snoRNAs (for detailed reviews, see refs. 23–25).

Surprisingly, the first sRNAs identified in archaea were C/D box snoRNAs, which had not been expected in a prokaryote lacking a nucleolus. Based on the conserved function, these snoRNAs are also called small “nucleolar” RNAs in archaea. C/D box snoRNAs guide ribose methylation at specific sites of the rRNAs. They were first predicted to exist in three archaeal species, and subsequently, experimentally verified in Sulfolobus acidocaldarius (reviewed in refs. 23 and 26). Furthermore, H/ACA box snoRNAs have been identified in archaea. H/ACA box snoRNAs guide the conversion of uridine to pseudouridine at specific sites of the rRNAs. To date, the first RNA-Seq study with an archaeal species was performed for Methanococcus mazei GoI under different nitrogen availabilities, leading to the identification of 242 intergenic and antisense sRNAs, including six cis-antisense sRNAs overlapping with transposase genes and 40 tRNA candidates containing very short ORFs potentially encoding peptides smaller than 30 amino acids.10 The transcriptome of Halobacterium volcanii was characterized in exponentially growing and in stationary phase cultures grown under optimal conditions as well as, respectively, under reduced salt concentration and elevated temperature using a multiplexing RNA-Seq approach.11 One hundred and forty-five intergenic sRNAs and 45 antisense sRNAs were identified and it was revealed that the levels of many sRNAs differed depending on growth phase and/or external conditions. Sense sRNAs were also observed, but not further characterized, because they might correspond to degradation intermediates of full-length transcripts. Notably, also tRNA-derived fragments were observed, an indication that they might play regulatory roles, similar to their function in eukaryotes. The 190 intergenic and antisense sRNAs included all sRNAs identified in a previous small scale RNomics study,10 but not all sRNAs that had been predicted using genome comparisons.11 Thus, it is very likely that not all sRNA genes were expressed under the three tested conditions and the actual number is higher than the 190 sRNAs observed in this RNA-Seq study.11 A RNA-Seq study was also performed with four Pyrobaculum species, i.e., P. aerophilum, P. arsenaticum, P. caldifontis, and P. islandicum.14 However, this study focused on the characterization of smRNAs and CRISPR-derived crRNAs. In all four species approx. 85 C/D box snoRNAs and 10 H/ACA snoRNAs were identified.
nucleotides from an associated archaeon, compacted genome and its RNA production relies on the import of very small sRNAs in another RNA-Seq study concentrated on the identification that in the euryarchaeal species characterized thus far. Recently, genes, the fraction of sRNA genes is about 10% and higher than is 2.99 Mbp and encodes less than 3000 protein-encoding using small scale RNomics.\textsuperscript{36,49} As the genome of approach included all sRNAs that had been identified earlier discussed above. The sRNAs identified with the RNA-Seq (125 members) by far, in contrast to the euryarchaeal species 185 members, which exceed the fraction of intergenic sRNAs the highest fraction was comprised of antisense sRNAs with antisense sRNAs overlapping with transposase genes.\textsuperscript{48} Notably, The first study led to the identification of more than 300 sRNAs, including 13 snorRNAs, 18 cRNAs, and 28 cis-antisense sRNAs overlapping with transposase genes.\textsuperscript{46} Notably, the highest fraction was comprised of antisense sRNAs with 185 members, which exceed the fraction of intergenic sRNAs (125 members) by far, in contrast to the euryarchaeal species discussed above. The sRNAs identified with the RNA-Seq approach included all sRNAs that had been identified earlier using small scale RNomics.\textsuperscript{45} As the genome of S. solfataricus is 2.99 Mbp and encodes less than 3000 protein-encoding genes, the fraction of sRNA genes is about 10% and higher than that in the euryarchaeal species characterized thus far. Recently, another RNA-Seq study concentrated on the identification of very small sRNAs in S. solfataricus.\textsuperscript{48} Total RNA was size-fractionated and RNA from about 18–30 nt was analyzed. A large number of sRNAs around 20 nt was found, verifying that very small sRNAs exist in Archaea. Only few members of the phylum Nanoarchaeota have been identified, with Nanoarchaeum equitans being the single cultured organism.\textsuperscript{51,52} N. equitans contains a minimal and highly compacted genome and its RNA production relies on the import of nucleotides from an associated archaeon, Ignicoccus hospitalis. Small

| Number of genes for | Intergenic sRNAs | allsRNAs | tRFs | Reference |
|---------------------|------------------|----------|------|-----------|
| **Euryarchaeota**   |                  |          |      |           |
| M. mazei            | 199              | 43       | n.d. | 41        |
| M. vulcani          | 145              | 45       | 11   | 42        |
| *P. abyssi*         | 107              | 215      | n.d. | 46        |
| **Crenarchaeota**   |                  |          |      |           |
| S. solfataricus     | 43               | 185      | n.d. | 48        |
| Pyrobaculum (4 species) | present\textsuperscript{42} | present\textsuperscript{42} | present\textsuperscript{42} | 44        |
| **Nanoarchaeota**   |                  |          |      |           |
| *N. equitans*       | present\textsuperscript{42} | present\textsuperscript{42} | present\textsuperscript{42} | 53        |

\*Not detectable due to the method of library construction. \*Number of genes not included in the publication.

Using a combination of experimental and in silico approaches, small RNAs were also detected in *Pyrococcus abyssi*.\textsuperscript{46} Recently, the non-coding transcriptome of this hyperthermophilic archaeon was additionally analyzed by RNA-Seq, surprisingly demonstrating that several highly expressed or highly conserved sRNAs are AU-rich, suggesting RNA functions that do not require extensive secondary structure in the high-temperature environment of *P. abyssi*.\textsuperscript{46}

Record numbers of 126 C/D box snoRNAs could be identified in the sRNA profile of the hyperthermophilic *Methanopyrus kandleri*.\textsuperscript{14} The high amount of these RNA species appears to correlate with the increased growth temperature of up to 110 °C, which would necessitate 2-O-methylation for the stabilization of tRNA folding.

In the subdomain of Crenarchaeota, two RNA-Seq studies have been performed with the species *Sulfolobus solfataricus*. The first study led to the identification of more than 300 sRNAs, including 13 snoRNAs, 18 cRNAs, and 28 cis-antisense sRNAs overlapping with transposase genes.\textsuperscript{46} Notably, the highest fraction was comprised of antisense sRNAs with 185 members, which exceed the fraction of intergenic sRNAs (125 members) by far, in contrast to the euryarchaeal species discussed above. The sRNAs identified with the RNA-Seq approach included all sRNAs that had been identified earlier using small scale RNomics.\textsuperscript{45} As the genome of *S. solfataricus* is 2.99 Mbp and encodes less than 3000 protein-encoding genes, the fraction of sRNA genes is about 10% and higher than that in the euryarchaeal species characterized thus far. Recently, another RNA-Seq study concentrated on the identification of very small sRNAs in *S. solfataricus*.\textsuperscript{48} Total RNA was size-fractionated and RNA from about 18–30 nt was analyzed. A large number of sRNAs around 20 nt was found, verifying that very small sRNAs exist in Archaea. Only few members of the phylum Nanoarchaeota have been identified, with *Nanoarchaeum equitans* being the single cultured organism.\textsuperscript{51,52} *N. equitans* contains a minimal and highly compacted genome and its RNA production relies on the import of nucleotides from an associated archaeon, *Ignicoccus hospitalis*. Small RNA profiling via RNA-Seq methodology was applied to identify sRNA molecules that are produced under these constraints.\textsuperscript{48} It was shown that C/D box snoRNAs and CRISPR RNAs are abundant in the cell, which underlines the importance of small RNA-guided RNA modification and viral defense mechanisms. In addition, several novel sense and antisense sRNAs were identified that could fulfill regulatory functions in the cell.

Until now, no sRNA investigation studies have been performed in species belonging to the kingdoms of Korarchaeota and Thaumarchaeota. Nevertheless, the occurrence of sRNAs in all species from the kingdoms Euryarchaeota, Crenarchaeota, and Nanoarchaeota that have been studied thus far underscores the wide-spread or even universal occurrence of sRNAs in archaea. Table 1 gives an overview of the results discussed above and lists several subclasses of sRNAs that have been experimentally verified by RNomics or RNA-Seq approaches. It should be noted that most archaeal sRNAs are not well conserved; many sRNA genes are not even shared by species of the same genus. For example, only very few sRNAs that have been found in *H. volcanii* are also present in *Haloferax mediterranius*.\textsuperscript{49} In *M. mazei*, the majority of sRNAs genes are conserved in other species of methanogenic archaea, but are not found outside of the methanogens.\textsuperscript{46} However, this observation is not confined to archaea, but similar findings have been reported for bacterial sRNA genes. For example, more than 90% of the 500 sRNAs identified in *Pseudomonas aeruginosa* have no homolog in any other bacterial species.\textsuperscript{53} In *Salmonella*, sRNAs were identified, which are not existent in *E. coli*.\textsuperscript{56} In addition, “highly conserved” sRNAs are typically confined to one taxonomic group, e.g., highly conserved cyanobacterial sRNAs are not found outside of the cyanobacteria.\textsuperscript{57} Therefore, it seems that in prokaryotes, both archaea and bacteria, the pace of evolution is much higher for sRNA genes than for protein-encoding genes. An interesting but as-yet-unresolved question is whether the species-specific sRNA genes were generated de novo during species evolution or whether the mutation rates of preexisting sRNA genes were so high that the similarity of homologous genes in different species has dropped below recognizable limits.

Table 1. Trans-encoded intergenic sRNAs, cis-encoded antisense sRNAs (allsRNAs), and tRNA-derived fragments (tRFs) experimentally identified in various archaea using RNA-Seq approaches

| Reference |
|-----------|
| 41        |
| 42        |
| 46        |
| 48        |
| 44        |
| 53        |
In Vivo Functions of sRNAs in Archaea

Although sRNAs have been identified in a variety of archael species, detailed analyses of the biological functions in vivo have until now only been performed with two species, *M. mazei* G101 and *H. volcanii*. In both species differential sRNA levels have been quantified using northern blot, DNA microarray, and RNA-Seq analyses, and mutants have been generated and characterized.

The methanogenic archaeon *M. mazei* is able to fix molecular nitrogen under conditions of nitrogen limitation and is genetically tractable, thus has been used for many years as a model to study nitrogen-dependent gene regulation in archaea. The recent discovery of regulatory sRNAs and their predicted role in nitrogen-dependent differential gene expression add another layer of regulation in addition to the well-studied transcriptional regulation. RNA-Seq analysis revealed that *M. mazei* contains nearly 250 sRNAs, 40 of which with coding capacity that might be oligopeptide-encoding mRNAs or dual function sRNAs. One hundred and thirty-five sRNAs were found to have differential levels in response to nitrogen availability, 36 were solely detected during nitrogen abundance, and 99 under nitrogen limitation. For several sRNAs, differential levels in response to the nitrogen availability were verified by northern blot analyses. Notably, in several cases, conserved motifs could be identified in the promoters of nitrogen-regulated sRNAs, indicating a network of coordinated transcriptional regulation.

One example, sRNA154, was further functionally characterized by a genetic approach. sRNA154 was found to be exclusively present under nitrogen-limiting conditions and has been shown to be under strict transcriptional control of the general nitrogen regulatory protein NrpR, indicating that it is relevant under nitrogen-fixating conditions. A deletion mutant of the sRNA154 gene had a severe growth defect under nitrogen-limiting conditions. However, under nitrogen sufficiency, it grew indistinguishable from the parent strain, emphasizing a crucial role of sRNA154 for nitrogen fixation. An in-depth characterization of a further sRNA, which led to the discovery of a first archaeal target mRNA, is discussed in the next chapter.

Differential sRNA levels under various conditions have also been analyzed in the halophilic archaeon *H. volcanii*, which has an optimal salt concentration of 2.1 M NaCl. For example, northern blot analyses revealed that sRNA194 is highly expressed during exponential phase, but absent or barely detectable in stationary phase cells. A 6-fold multiplexed RNA-Seq analysis...
yielded a genome-wide overview of sRNA levels in cultures grown under three different conditions to exponential phase and to stationary phase, respectively. The largest difference was found between optimal salt concentration and reduced salt concentration, 24 sRNAs had higher levels at the optimal salt concentration, while 19 sRNAs had higher levels at the reduced salt concentration. The six highest sRNA levels were found in cells grown at the reduced salt concentration, indicating that sRNAs have important regulatory functions for osmotic adaptation in *H. volcanii*. For *H. volcanii*, a very efficient genetic system has been established and optimized. Generation of two deletion mutants revealed that phenotypic comparison of mutants and parent strain is an effective approach to unravel the biological roles of sRNAs in haloarchaea. deletion mutant of sRNA63 could not grow at 51 °C in contrast to the parent strain, while deletion mutant of sRNA63 had a severe growth defect at low salt concentrations. Recently, growth of *H. volcanii* in microtiter plates was established, enabling the parallel comparison of growth parameters of many cultures. The efficient genetic system and the possibility of massive parallel cultivation together enabled the generation of a set of 27 deletion mutants of sRNA genes and their phenotyping under 11 different growth or stress conditions. In addition, cell morphology was checked and swarming was analyzed. Interestingly, 24 of the 27 sRNA gene deletion mutants exhibited a phenotype under at least one of the tested conditions, unraveling the broad importance of sRNAs in haloarchaea for many biological functions. They are involved in metabolic regulation, adaptation to extremes of temperature and osmolarity, stress response, growth phase adaptation, and even regulation of cellular morphology and behavior. Notably, seven of the 27 deletion mutants showed a gain-of-function phenotype, e.g., shorter lag phase, faster growth rate, faster swarming than the parent strain. It is known that depletion of mRNAs in higher eukaryotes can result in a gain-of-function phenotype. Thus, it appears that in archaea and eukaryotes evolution of sRNAs did not result in the highest possible function of a single biological process, but in optimized flexibility and stability of the regulatory network. Loss of sRNA function can also lead to a severe loss of function, both in archaea (see above) and in eukaryotes. For example, mutations in miRNA genes in *C. elegans* can impair developmental timing in larvae. In humans, miRNA dysfunction or deficiency induces strong disease patterns, cancer (reviewed in ref. 73), multiple sclerosis, Parkinson disease, or Alzheimer disease. In contrast, sRNA gene deletion mutants in bacteria often lack a detectable phenotypic difference to the parent strain or have only a very mild phenotype; therefore, it has sometimes been proposed that the major role of bacterial sRNAs is the “fine-tuning of gene expression.”

**Targets of Archaeal sRNAs**

In bacteria, sRNAs typically interact with their target mRNAs near the 5′-end, often masking the ribosome binding site (RBS) or leading to an unfolding of the mRNA that makes the RBS accessible for ribosomes. In eukaryotes, miRNAs typically bind to the 3′-UTRs of their target mRNAs, leading to translational inhibition and/or degradation of the mRNAs (reviewed in ref. 80). In both domains, a number of computational approaches have been established that allow a genome-wide in silico search for putative targets of a specific sRNA. However, these programs designed for and trained with bacterial and eukaryotic sRNAs, respectively, cannot necessarily be used for a genome-wide bioinformatic prediction of targets of archaeal sRNAs without specific adaptations. Early attempts to apply prediction programs for bacterial sRNA targets failed for archaeal sRNAs, indicating that the principles of interactions between sRNAs and target mRNAs might be different in archaea and bacteria. Thus, experimental approaches appear to be crucial to identify target mRNAs for archaeal sRNAs, subsequently verified targets could be used to guide bioinformatic predictions. Very recently, the group of Rolf Backofen optimized the target prediction program IntaRNA and integrated comparative genomics into the prediction pipeline (see also contribution in this volume by Backofen, et al.), which was successfully used in combination with an experimental approach for the identification of the first target of an archaeal sRNA. This first target for an archaeal sRNA was identified in *M. mazei* combining genetic approaches, genome-wide transcriptome analysis of sRNA mutants, and computational target predictions using the tool IntaRNA. The bicistronic mRNA (MM2441-MM2442) was identified as the trans-encoded target mRNA of the respective archaeal sRNA (sRNA30). The predicted target interaction between the non-structured single-stranded linker region of sRNA30 and its target results in masking the RBS of MM2441 by base-pairing as well as the translational start codon, leading to a dis-coordinated expression of the bicistronic mRNA, and thus, to a depletion of the transcriptional regulator MM2441. The predicted interaction between sRNA30 and the RBS of MM2441 was verified in vitro (by EMSA) and in vivo (by ectopic expression of sRNA30 and various mutants). Further studies strongly indicated the involvement of sRNA30 in the expression regulation of soluble methyltransferases—most likely due to translation inhibition of the transcriptional regulator MM2441—following the metabolic switch between growth on methanol vs. methylvamine. Notably, it was also observed that sRNA162 overlaps the 5UTR of the MM2442 transcript, and thus, is also acting as a cis-encoded antisense RNA. In addition to regulating MM2442 expression as a trans-encoded sRNA (compare Fig. 1). Besides the recent report on an sRNA acting in trans in *S. elongatus* (reviewed in ref. 80), an sRNA acting as an antisense RNA on cis-encoded as well as on trans-encoded mRNAs most likely via two distinct domains has not been observed in prokaryotes so far. Thus, the methanoarchaeal sRNA162 is the prototype of a novel class of “dual-mechanistic” sRNAs, which has the ability to act simultaneously in cis and in trans. The interaction of sRNA162 with the 5′-region of at least two genes, overlapping the RBS, resembles the molecular mechanisms of typical bacterial sRNAs. Transcripts of methanoarchaea usually have long 5′-UTRs, thus interactions between sRNAs and 5′-UTRs are very likely. Further studies are required to unravel
whether it is indeed typical for the action of methanocalarchal sRNAs.

In contrast to Methanorarchaea, various archaeal species exist in which most or all transcripts do not contain 5′-UTRs, e.g., Pyrobaculum aerophilum,76 Sulfolobus solfataricus,77 Halobacterium salinarum, and H. volcanii.86 If differential regulation of translational efficiencies and/or stabilities of these leaderless transcripts would occur, respective motifs must be localized either in within the ORF or in the 3′-UTR. Genome-wide studies of translational regulation (translome analyses) in H. salinarum and H. volcanii revealed that translational regulation is not seldom in these two species. In fact 20% and 10%, respectively, of all transcripts have growth phase-specific differential translational efficiencies.87 Further studies showed that 5′-UTRs and 3′-UTRs are sufficient to transfer translational regulation to a reporter transcript, and that the direction of translational regulation (up- or downregulation) is encoded in the 3′-UTR.88 To our knowledge, this was the first report that 3′-UTRs are involved in translational regulation in a prokaryotic species. However, it remained unclear whether sRNAs were involved in these examples. Recently, the transcriptomes of seven sRNA gene deletion mutants were compared with that of the parent strain with the aim to identify putative sRNA targets experimentally (Jaschinski, Babki, and Soppa, unpublished results). In each case, several or many genes differed in their transcript levels between mutant and parent strain. The numbers of transcripts with differential levels ranged from less than 10 to about 100, indicating that the regulatory role of haloarchaeal sRNAs can be very specific or rather broad, depending on the identity of the sRNA. Bioinformatic comparisons of the sequences of sRNAs and putative target mRNAs led to the discovery of several examples of putative interactions between sRNAs and the 3′-UTRs of mRNAs. Currently, experimental approaches to analyze whether these interactions occur are relevant are in vivo are under way, e.g., several UTRs of several transcripts have been transferred to a reporter transcript, and quantification of translational efficiencies in sRNA deletion mutants vs. parent strain indicate that the interaction between sRNA and the 3′-UTR can indeed guide translational efficiency (Jaschinski, Höfle, and Soppa, unpublished results).

Additional implication for an archaeal sRNA targeting a 3′-UTR has been obtained from a recent comparative RNA sequencing approach studying four Pyrobaculum strains.89 Here, a 65-nucleotide-long antisense transcript, a53, has been identified that is conserved in Pyrobaculum and binds to the 3′-end of the tpi gene (encoding trisac-phosphate-isomerase), either overlapping the stop codon or binding to the 3′-UTR. Furthermore, a conserved structural element located close to the stop codon has been recognized in the 3′-region of the tpi gene (designated tpi element). Binding of a53 to the tpi mRNA might be able to compete against the formation of the intramolecular tpi element structure, and consequently, might modulate the function of the highly conserved tpi element.

Recently, an interaction between a sRNA and the 3′-UTR of a target mRNA was reported for S. solfataricus.77 The sRNA was regulated in a phosphate-dependent manner and had a high level during phosphate sufficiency and a low level during phosphate starvation. The transcript of the gene Sso1183 was found to be inversely regulated in response to changing phosphate concentrations. A sequence comparison in silico revealed a very long stretch of high complementarity of about 60 nt between the sRNA and the 3′-UTR of the mRNA. The 3′-UTR and a mutated version of the 3′-UTR (with highly reduced complementarity) were fused to the reporter transcript lacS, and the transcript levels of the sRNA and the 3′-UTR were quantified in cultures grown in the presence or absence of phosphate. It could indeed be shown that the level of lacS containing the 3′-UTR of Sso1183 was inversely regulated to the sRNA level, in contrast to two lacS controls lacking a 3′-UTR or containing a 3′-UTR of an unrelated gene. Thus, first examples are accumulating that archaeal sRNAs can execute their regulatory role via interaction with the 3′-UTRs of target genes.

In summary, experimental approaches have led to the identification of first target mRNAs in three species of archaea. Already, these first examples reveal that the molecular mode of action of archaeal sRNAs is not universal in this domain of life and that the characterization of archaeal sRNAs leads to the identification of molecular mechanisms of sRNAs that have not been found for any bacterial sRNA. The modes of sRNA target mRNA interactions are schematically summarized in Figure 1.

tRNA-Derived Fragments

Recently, it has been discovered that eukaryotes contain small RNAs that are derived from tRNAs (tRNA-derived fragments, tRF). Initially, it was thought that tRFs are simple degradation products with increased stability, but now it is clear that they fulfill specific regulatory functions in many biological processes (reviewed in ref. 22), e.g., cell proliferation,77 in protein biosynthesis, and in the siRNA and miRNA pathway.95 In accordance with their regulatory roles, the processing of tRNAs to tRFs is itself regulated and is induced under specific conditions, e.g., stress application.96 Subsequently, tRFs were also identified in bacteria and archaea. In archaea, the existence of tRFs has been reported for several species of Pyrobaculum76 and H. volcanii.86 In H. volcanii, tRFs could be observed from 11 of the 51 tRNAs using RNA-Seq, and their existence could be verified using northern blot analyses.97 In another study, all small RNAs that could be co-isolated with ribosomes were identified by RNA-Seq.42 Twenty-six percent of the co-isolated sRNAs were tRFs, indicating that tRFs can directly or indirectly be associated with the ribosome. The tRFs of the tRNA-Valine (tRF[Val]) were the most prominent co-isolated small RNAs. It could be verified by density gradient centrifugation and in vitro binding studies that tRF[Val] directly binds to the small ribosomal subunit. Furthermore, using an in vitro translation system, it was shown that binding of tRF[Val] to the ribosome inhibits translation. In addition, it was revealed that the generation of tRF[Val] from the tRNA[Val] was highly induced under specific stress conditions, e.g., alkaline stress.42 Taken together, these results
show that rRFs can potentially fulfill regulatory functions not only in eukaryotes, but also in archaea.

CRISPR RNAs in Archaea

In prokaryotes, one family of sRNAs, the CRISPR RNAs (crRNAs), play a key role in the defense against foreign DNA, e.g., viruses or conjugative plasmids. Genomic CRISPR arrays were first identified in E. coli. The term CRISPR is an acronym for "clustered regularly interspaced short palindromic repeats."106,107 and as such, CRISPR loci are characterized by series of short repeated DNA sequences interspaced by unique spacer sequences. These spacers can represent foreign DNA elements that have been acquired at the promoter-proximal position of a CRISPR array. The CRISPR elements are transcribed and processed into small CRISPR RNAs, which contain a single spacer element flanked by repeat RNA fragments. These spacer sequences can guide a complex of CRISPR-associated (Cas) proteins to cDNA targets. This allows the detection of a repeated exposure to foreign DNA, which can subsequently be degraded by this interference ribonucleoprotein complex. Thus, CRISPR and Cas proteins represent an adaptive immune system, which follows rules of Lamarckian evolution.103 Different CRISPR-Cas systems exist with a plethora of Cas protein families that highlight functional differences (e.g., recognition of DNA or RNA) and the co-evolution of CRISPR and anti-CRISPR systems in prokaryotes and viruses.104,105 The diversification of Cas proteins is most obvious for cRNA maturation endonucleases and large subunits of the interference complex.106–108 The biogenesis of CRISPR RNA molecules and the mechanisms of CRISPR-Cas interference have been reviewed in detail before (e.g., refs. 108–110). A comparison of bacterial and archaeal CRISPR systems reveals that the latter contain (1) more and longer CRISPR systems, (2) a more diverse Cas protein landscape, and (3) more often a constitutive production of crRNAs. However, these observations also correlate with an increasingly hot environment, which is most often reserved for thermophilic or hyperthermophilic archaeal organisms. It remains to be analyzed if extreme growth conditions limit the diversity of host/virus interactions, which results in an accelerated evolution of defense and counter-defense measures.

Future Aspects

As summarized above, in recent years, a plethora of sRNAs have been found in any archaeal species that has been studied, and their importance for many different biological processes has been elucidated. Notably, first identifications of mRNA targets of sRNAs has led to the discovery of mechanisms of action unprecedented in prokaryotes, i.e., a dual-mechanistic sRNA that is a cis-antisense sRNA and a trans-acting sRNA at the same time, and a few sRNAs that act via binding to the 3’-UTR. Nevertheless, knowledge about archaeal sRNAs is lacking behind knowledge obtained with bacteria and eukaryotic sRNAs, and the following developments seem to be important for the near future.

Until now, the genome-wide prediction of sRNA targets has been successful for M. mazei in combination with genetic approaches, but not for other archaea. The experimentally identified targets should be used to train bioinformatic prediction programs with the aim to develop reliable high-quality predictions, possibly in an iterative exchange between experimental results and bioinformatic improvements. The results with M. mazei systems show that a domain-wide “archaeal” prediction program will not be possible, but several attempts for different groups of archaea are necessary. Molecular analysis of a significant number of sRNA-target mRNA pairs will be necessary to understand the modes of action and to enable generalization of principles. In E. coli and Salmonella, a dual plasmid system with a reporter gene (carrying UTRs and modified versions thereof) on one plasmid and the cognate sRNA gene on another plasmid has been proven to be extremely useful (e.g., refs. 111–114). Equivalent systems should be developed for several species of archaea to allow detailed analyses of sRNAs and motifs within UTRs of target mRNAs in vivo.

It is currently unknown whether and which proteins are required for the regulatory functions of sRNAs. In H. volcanii, the Lsm protein has been shown to bind a small subset of the existing sRNA pool, but it is unclear whether this is important for function or whether other sRNAs need different proteins or regulate solely at the RNA level. Methods for the affinity isolation of sRNAs and/or regulated mRNAs would be desirable to enable the identification of interaction partners. It has been shown for many sRNAs that their levels are very different under various growth conditions. However, the regulation of sRNA generation and decay has not been studied yet, neither for sRNA genes with an own promoter, nor for sRNAs like rRFs that are processed from a longer precursor.

In summary, significant progress has been made in recent years concerning the presence and distribution of sRNAs in archaea, their biological functions, and initial successes have been made in target identification and characterization of sRNA-target RNA interactions, but it is also clear that considerable further work is needed to deepen our understanding of archaeal sRNAs and their regulatory networks.

Disclosure of Potential Conflicts of Interest

No potential conflicts of interest were disclosed.

Acknowledgments

The work in the groups of Marchfelder A, Schmitz RA, and Soppa J were funded by the German Research Council (Deutsche Forschungsgemeinschaft, DFG) through the Priority Program SPP 1258 "Sensory and Regulatory RNAs in Prokaryotes" (grants DFG Ma1538/11, Schm 1052/9, So264/14). The group of Randau L was associated to the Priority Program and was funded by the Max-Planck-Society. Currently, the collaborative project of the groups of Marchfelder A and Soppa J are funded through the DFG grants Ma5358/18 and So264/21.
60 Gunsalus RP, Deppenmeier U, Schmitz RA. Global transcriptional analysis of Koonin EV, Reysenbach AL. Insights into archaeal archaeon world of small RNAs in the hyperthermophilic JB.187.17.6147-6154.2005

50 Direct 2013; 8:9; PMID:23607440; http://dx.doi.org/10.1186/1745-6150-8-9

40 Randau L. RNA processing in the minimal organism Nanoarchaeum equitans. Genome Biol 2012; 13:R63; PMID:22618407; http://dx.doi.org/10.1186/1475-2867-13-6

30 Schmitz RA. Insights into the NrpR regulon in MJansson MD, Lund AH. MicroRNA and cancer. Mol Cell 2005; 273:290-8; PMID:15824966; http://dx.doi.org/10.1016/j.mcb.2005.07.029

20 Ehlers C, Veit K, Greimisch G, Schmidt RA. Functional organization of a single riboswitch cluster in the extremophile archaeon Methanothermus marisnigri. Genet. Res. 2002; 79:143-50; PMID:12138602; http://dx.doi.org/10.1017/S0016672302007135

10 Ehlers C, Jager D, Schmidt RA. Establishing a comprehensive gene expression “card” for Methanobrevibacter smithii strain GB for constructing a complete set of annotated genes for the strain. Genet. Res. 2011; 98:3600-4; PMID:21939464; http://dx.doi.org/10.1017/S0016672311002761

...
regulation in halophilic archaea. BMC Genomics 2012; 13:157. PMID:22552946; http://dx.doi.org/10.1186/1471-2164-13-157.

Brenneis M, Soppa J. Regulation of translation in haloarchaea: 5′ and 3′ UTRs are essential and have to functionally interact in vivo. PLoS One 2009; 4:e4484. PMID:19312437; http://dx.doi.org/10.1371/journal.pone.0004484.

Boscoletti J, Zvonek M, Kázmár A, Balázs N. RNA-derived fragments target the ribosome and function as regulatory non-coding RNA in Haloferax volcanii. Archen 2012; 8:415. PMID:22652609; http://dx.doi.org/10.1186/1471-2164-8-415.

Grobinger J, Zvonek M, Kázmár A, Balázs N. RNA-derived fragments target the ribosome and function as regulatory non-coding RNA in Haloferax volcanii. Archen 2012; 8:415. PMID:22652609; http://dx.doi.org/10.1186/1471-2164-8-415.

Meriaux B, Massenault S, Hausenedl D, Mestres A, Blisi U. Antisense regulation by transposon-derived RNAs in the hyperthermophilic archaeon Sulfolobus solfataricus. EMBO Rep 2013; 14:527-35. PMID:23797902; http://dx.doi.org/10.1038/embobiol.2013.47.

Lee YS, Shibuya Y, Milkova A, Davis H. A novel class of small RNAs: RNA-derived RNA fragments (tRFs). Genes Dev 2009; 23:2639-49. PMID:19933153; http://dx.doi.org/10.1101/gad.1837609.

Hanneschel D, Haengg Y, Lutz A, Zimmermann P, Fri AZ, Kay MA. Human RNA-derived small RNAs in the global regulation of RNA silencing. RNA 2010; 16:473-85. PMID:20618376; http://dx.doi.org/10.1261/rna.204810.

Fu H, Feng J, Lai Q, Sun F, Tu Y, Zhu J, Xing B, Yan Z, Zheng X. Stress induces RNA cleavage by angiogenin in mammalian cells. FEBS Lett 2009; 583:437-42. PMID:19104460; http://dx.doi.org/10.1016/j.febslet.2008.12.043.

Thompson DM, Lu C, Gepstein P, Barker P. RNA cleavage is a conserved response to oxidative stress: in eukaryotes. RNA 2008; 14:2959-60. PMID:18759243; http://dx.doi.org/10.1261/rna.152589.

Boscoletti DL, Denisso PP, Lui LM, Lowe TM. Diversity of Antisense and Other Non-Coding KSs in Archaea Revealed by Comparative Small RNA Sequencing in Four Pyrolobus Fumarii Species. Front Microbiol 2012; 3:255. PMID:22752866; http://dx.doi.org/10.3389/fmicb.2012.00251.

Grobinger J, Zvonek M, Kázmár A, Balázs N. RNA-derived fragments target the ribosome and function as regulatory non-coding RNA in Haloferax volcanii. Archen 2012; 8:415. PMID:22652609; http://dx.doi.org/10.1186/1471-2164-8-415.

Makarewa KS, Haft DH, Baojun B, Blattner FJ, Wolf YI, Yakunin AF, et al. Evolution and function of CRISPR-Cas bacterial immune system. Nature 2013; 493:429-35. PMID:23242138; http://dx.doi.org/10.1038/nature11723.

Papenfort K, Makarova KS, Haft DH, Barrangou R, Brouns FJ, Wolf YI, et al. Type I-A Cascade interference complex. Nucleic Acids Res 2011; 39:4289-32. PMID:22342138; http://dx.doi.org/10.1093/nar/gkq1372.

Makarewa KS, Blattner FJ, Baojun B, Barrangou R, Makarova KS, et al. The type I-A CRISPR-Cas system disperses antibiotic resistance genes in bacteria. Proc Natl Acad Sci U S A 2012; 109:7923-8. PMID:22858617; http://dx.doi.org/10.1073/pnas.1111673109.

Winkler M, Schirmer C, Bauhaus CM, Mika F, Nierhaus KH, Vogel J. RNA binding to a 5′ terminal region triggers translational initiation. Mol Cell 2004; 12:627-37. PMID:15161622; http://dx.doi.org/10.1016/j.molcel.2004.07.027.