

Chapter 10

Diverse Functions of Small RNAs (sRNAs) in Halophilic Archaea: From Non-coding Regulatory sRNAs to Microprotein-Encoding sRNAs

Jana Kliemt and Jörg Soppa

Abstract The number of experimentally-verified small RNAs (sRNAs) in haloarchaea has increased tremendously since their discovery in 2009. A recent dRNA-Seq study led to the identification of about 2900 non-coding primary transcripts, a number that was much higher than the less than 1900 protein-coding mRNAs that were detected. Intergenic sRNAs have been studied intensively, and it was revealed that they have important regulatory functions under many different conditions. However, the numbers of cis-antisense RNAs and of cis-sense RNAs are much higher, and thus future research will shift to concentrate on their characterization. Other classes of haloarchaeal sRNAs include tRNA-derived fragments, which have been shown to be regulators of translation, CRISPR/CAS defensive sRNAs, C/D box sRNAs that guide methylation, and 7S RNA involved in membrane protein biogenesis. Last, but not least haloarchaea contain hundreds of small mRNAs that encode microproteins, which represent an emerging field of research. Taken together, the network of small RNAs in haloarchaea is much larger than anticipated only a few years ago, and the multitude of sRNAs has changed the view of the function of haloarchaeal genomes.

10.1 Introduction

Small non-coding regulatory RNAs (sRNAs) exist in all three domains of life, archaea, bacteria, and eukaryotes. In recent years an ever increasing number of sRNAs have been discovered, and they were found to be involved in and important for many different biological processes. Several recent reviews summarize various aspects of sRNAs in prokaryotes (Wagner and Romby 2015; Kopf and Hess 2015; Murina and Nikulin 2015; van Puyvelde et al. 2015; Georg and Hess 2011; Waters

J. Kliemt • J. Soppa (✉)

Biocentre, Institute for Molecular Biosciences, Goethe-University, Max-von-Laue-Str. 9, 60438 Frankfurt, Germany

e-mail: kliemt@bio.uni-frankfurt.de; soppa@bio.uni-frankfurt.de

and Storz 2009) and in eukaryotes (Catalanotto et al. 2016; Yang et al. 2016; Borges and Marienssen 2015; Huang et al. 2013). In addition, very recently it has been acknowledged that small RNAs can also contain open reading frames and that the encoded microproteins can have very important functions (reviews: Ramamurthi and Storz 2014; Storz et al. 2014). This chapter will concentrate on the various groups of non-coding and coding sRNAs from halophilic archaea, including e.g. cis antisense RNAs, snoRNAs, and tRNA-derived fragments. On the one hand it updates two earlier reviews (Babski et al. 2014; Schmitz-Streit et al. 2011), on the other hand it gives a broader view and includes additional classes like microprotein-encoding sRNAs. sRNAs from other phylogenetic groups of archaea are discussed in other chapters of this book.

10.2 Identification of Small RNAs and the Changing View of the Haloarchaeal Transcriptome

The first archaeal sRNAs were detected in the euryarchaeon *Archaeoglobus fulgidus* at the beginning of this century (Tang et al. 2002). Shortly thereafter, it was unraveled that sRNAs occur also in crenarchaeota (Tang et al. 2005). The first study with a halophilic archaeon was published a few years later (Straub et al. 2009). A small scale RNomics study with *Haloferax volcanii* led to the identification of 21 intergenic sRNAs and 18 antisense sRNAs (asRNAs). Northern blot analyses revealed that many of the sRNA genes were differentially expressed, indicating that the regulatory roles of the respective sRNAs are confined to specific environmental conditions. The next approach was the bioinformatics comparison of the intergenic regions of *Hfx. volcanii* with those of four other haloarchaea, one crenarchaeon, and one halophilic bacterium (Babski et al. 2011). More than 120 conserved regions that might represent conserved sRNA genes were found. The expression of 61 of these putative sRNA genes was analyzed using a dedicated DNA microarray, and 37 genes were found to be expressed under at least one of the three conditions tested, verifying that the bioinformatics predictions could successfully identify sRNA genes.

High Throughput Sequencing (HTS) of cDNA libraries, which was relatively new at that time, was used to characterize the small transcriptome of sRNAs with lengths between 17 and 500 nt (Heyer et al. 2012). Thereby, the number of sRNAs was increased to 145 intergenic sRNAs and 45 asRNAs. RNAs from cultures grown at six different conditions were used for cDNA library generation, and multiplexing was used to sort the sequences bioinformatically after a single HTS run. Again, it was found that many sRNA genes were differentially expressed. Notably, many sRNAs could only be detected in cultures that were grown under low salt, a condition that represents considerable stress for haloarchaea. Haloarchaea use the so-called “salt in” strategy for osmoadaptation. The salt concentration in the

cytoplasm is as high as in the environment. The consequence of this strategy is that all biological processes have to be evolutionary adapted to function in the presence of molar concentrations of salt. Haloarchaeal proteins contain 20% aspartic and glutamic acid residues and have a high negative charge density at their surface. This makes them soluble at high salt concentrations, but on the other hand, this makes them very sensitive to low salt concentrations. Typical haloarchaeal proteins denature at salt concentrations below 1 M NaCl. Most halophilic bacteria apply the so called “salt out” strategy, they have a low salt concentration in the cytoplasm and use organic compatible solutes for osmoadaptation.

Very recently a state of the art differential RNA-Seq (dRNA-Seq) approach was used to characterize the primary transcriptome of *Hfx. volcanii* (Babski et al. 2016). dRNA-Seq makes use of an enzyme that degrades all transcripts without a triphosphate at their 5'-end, while transcripts with a triphosphate remain untouched. Comparison of treated and untreated samples allows the differentiation between primary transcripts and transcripts that were generated by processing or are degradation intermediates. The highly increased sequencing depth led to the identification of nearly 2800 novel non-coding transcripts. Remarkably, the total number of non-coding RNAs was with 2900 considerably higher than the total number of protein-coding RNAs with less than 1900. Taken together, the view of the transcriptome and of the genome function of *Hfx. volcanii* has changed dramatically within the last 6 years. Figure 10.1a schematically shows that according to the original annotation the genome contained nearly exclusively protein-coding genes (Hartman et al. 2010). Figure 10.1b illustrates that small-scale RNomics and the HTS approaches led to the identification of about 200 sRNA genes (Babski et al.

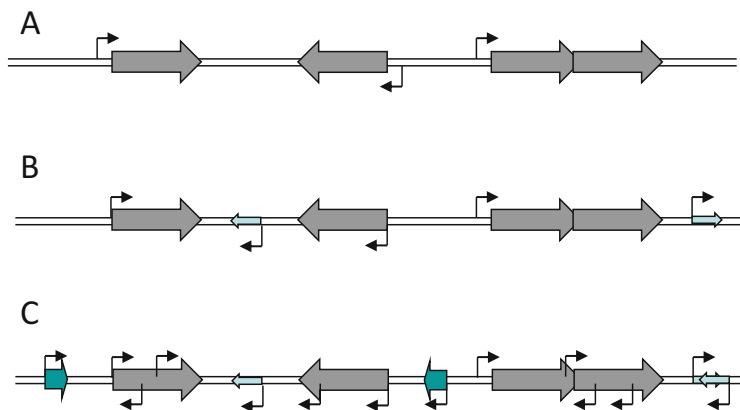


Fig. 10.1 Changing view of the transcriptome and genome function of *Hfx. volcanii* during recent years. (a) View based on genome sequencing and annotation in 2010 (Hartmann et al. 2010). (b) View based on the identification of sRNAs genes in intergenic regions (Heyer et al. 2012). (c) View based on the very recent dRNA-Seq study (Babski et al. 2016). The arrows indicate presumed transcription start sites (a, b) and experimentally verified transcription start sites (c)

2011; Heyer et al. 2012). In stark contrast, the dRNA-Seq study uncovered that the number of non-coding RNAs is in fact higher than the number of protein-coding genes (Fig. 10.1c) (Babski et al. 2016). Various different classes of non-coding RNAs were found, which are discussed below. It can be expected that not all sRNAs of *Hfx. volcanii* have been identified yet, because the dRNA-Seq study was performed using cultures grown under optimal conditions. Because it has been shown that sRNA genes can be differentially expressed and can be silent under optimal conditions (see above), it can safely be predicted that further studies with cultures grown under non-optimal conditions will further enhance the number of sRNAs of *Hfx. volcanii*.

dRNA-Seq studies have been performed only for three additional archaeal species, i.e. *Methanobolus psychrophilus* (Li et al. 2015), *Thermococcus kodakarensis* (Jäger et al. 2014), and *Methanosarcina mazei* Gö1 (Jäger et al. 2009). For all three species the number of non-coding RNAs was much smaller than the number of protein-coding RNAs, e.g. only 195 of 2056 transcripts from *M. psychrophilus* were non-coding sRNAs. Therefore, the fraction of non-coding RNAs is not uniformly high in all archaea, and the situation in haloarchaea resemble the situation in higher eukaryotes, which also contain a higher number of non-coding transcripts than protein-coding transcripts (Wan et al. 2014).

Table 10.1 gives an overview of the number of three classes of sRNAs and the number of annotated protein-coding genes for six archaeal species. Only RNA-Seq studies and dRNA-Seq studies since 2009 have been included, because earlier small-scale RNomics studies led to much smaller numbers of identified sRNAs. It should be noted that the numbers should be handled with care, because the number of different culturing conditions, the sequencing depth, and the bioinformatics analysis pipeline can tremendously influence the results. Nevertheless, it can be seen that the numbers especially of intergenic sRNAs and asRNAs (and their ratios) differ considerably in the six investigated species.

Table 10.1 Numbers of protein coding genes (genome annotation) and of three classes of non-coding sRNAs (RNA-Seq or dRNA-Seq) in selected archaeal species

| Species | No. of ORFs | Interg. sRNAs | asRNAs | iRNAs | Reference |
|-----------------------------|-------------|---------------|--------|-------|------------------------------|
| <i>Haloflex volcanii</i> | 4040 | 395 | 1244 | 1153 | Babski et al. (2016) |
| <i>Methanobolus psychr.</i> | 3167 | 195 | 1110 | 1440 | Li et al. (2015) |
| <i>Thermococcus kodak.</i> | 2306 | 69 | 1018 | 644 | Jäger et al. (2014) |
| <i>Pyrococcus abyssi</i> | 1784 | 107 | 215 | n.r. | Toffano-Nioche et al. (2013) |
| <i>Haloflex volcanii</i> | 4040 | 145 | 45 | n.r. | Heyer et al. (2012) |
| <i>Sulfolobus solft.</i> | 2994 | 125 | 185 | n.r. | Wurtzel et al. (2010) |
| <i>Methanosarcina m.</i> | 3371 | 199 | 43 | n.r. | Jäger et al. (2009) |

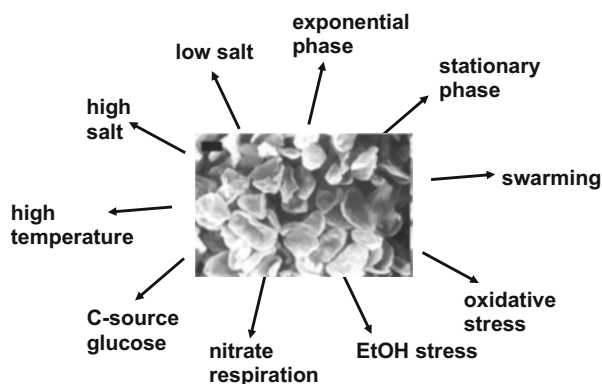
n.r. not reported

10.3 Various Classes of Small Non-coding Regulatory Haloarchaeal RNAs

10.3.1 Intergenic sRNAs

The intergenic sRNAs were the first sRNAs that have been systematically characterized. Already in the first study two gene deletion mutants have been constructed and phenotypically analyzed (Straub et al. 2009). One of the mutants could not grow at the elevated temperature of 51 °C, the other mutant had a severe growth defect at the low salt concentration of 0.9 M NaCl. Both phenotypes underscored the high importance of sRNAs for the physiology of *H. volcanii*. In a subsequent study 27 sRNA gene deletion mutants were generated and characterized (Jaschinski et al. 2014). For 24 of the 27 mutants a phenotypic difference from the wild-type could be detected under at least 1 of the 12 tested conditions. In addition, differential expression of sRNA genes was studied using a variety of different experimental approaches (Northern blot analyses, reporter gene assays, DNA microarray analyses). The results of all approaches revealed that sRNAs are important for the regulation of many biological functions in haloarchaea, which is schematically illustrated in Fig. 10.2. The biological functions include stress adaptation (which is proposed to be the major function of sRNAs in bacteria), but also metabolic regulation, adaptation to the extremes of growth conditions, and, last but not least, regulation of behavior. Remarkably, more than 10 of the 27 deletion mutants exhibited a gain-of-function phenotype. As yet this is unprecedented for any sRNA gene deletion mutant in bacteria. However, gain-of-function phenotypes have also been described for deletion or depletion of miRNAs in higher eukaryotes (Daniel et al. 2014). These results illustrate that regulatory circuits did not evolve to ensure the highest growth rate under one specific (laboratory) condition, but that regulatory networks were favored that had the highest stability and flexibility under the ever changing conditions of natural environments.

Fig. 10.2 Schematic overview of the diverse biological functions of sRNAs in *Hfx. volcanii*. The functions have been deduced from the phenotypes of sRNA gene deletion mutants and from elevated sRNA levels under specific conditions. The EM picture of *Hfx. volcanii* was supplied by J. Babski, K. Jaschinski, and J. Soppa (unpublished data)



The recent dRNA-Seq study increased the number of intergenic sRNAs to more than 400. Only a small fraction of them have been studied until now, therefore, the already uncovered manifold functions of sRNAs in haloarchaea (Fig. 10.2) represent only the tip of the iceberg. Further work is also needed to identify the molecular targets of sRNAs, which are presumably primarily protein-coding mRNAs, as well as the molecular details of sRNA-target RNA interactions and the molecular mechanisms of regulation. The bioinformatics target prediction algorithms that have been successfully used with bacteria and methanogenic archaea have as yet not been successful with haloarchaea, possibly because the conditions in the high salt cytoplasm are so different from the conditions in mesohalic species. However, comparisons of the transcriptomes of sRNA deletion mutants and the wild-type have already led to the discovery of the target mRNAs for several intergenic sRNAs, and thus it can be expected that experimental approaches will soon shed light on the details of the regulatory functions of intergenic haloarchaeal sRNA. Because 72% of all haloarchaeal protein-coding transcripts are leaderless (Babski et al. 2016), it has been predicted that many sRNAs might bind to the 3'-UTRs of their target mRNAs. This would be analogous to the eukaryotic miRNAs, which also bind to 3'UTRs, and in contrast to bacterial sRNAs, which typically bind to the 5'-region. The sRNA-target mRNA interaction does not seem to be uniform in archaea, first examples include the binding to the 5'-region in *M. mazei* (Prasse et al. 2013; Jäger et al. 2012) as well as the binding to the 3'UTR in *Hfx. volcanii* (Kliemt, Jaschinski, and Soppa, unpublished data) and in *Sulfolobus solfataricus* (Martens et al. 2013).

10.3.2 *Cis Sense sRNAs*

The analysis of the primary transcriptome of *Hfx. volcanii* led to the identification of more than 1100 sRNAs that were encoded in the same direction and within ORFs of protein-encoding genes (cis sense sRNAs) (Babski et al. 2016). This class of sRNAs had also been found in previous studies, but had not been further discussed, because these RNAs might be meta-stable degradation intermediates of the mRNAs. However, this possibility could be excluded by the experimental design of dRNA-Seq, which enriches for primary transcripts with a triphosphate at the 5'-end. In addition, a high fraction of these internal sRNA genes were preceded by promoter motifs with a high promoter score, also indicating that a high number of ORF-internal promoters exist in *Hfx. volcanii*. Also for *Hbt. salinarum* a large number of internal transcripts have been described, that were preceded by transcription factor binding sites. Therefore, it was concluded that a high number of ORF-internal promoters and cis sense sRNAs exists in *Hbt. salinarum* (Koide et al. 2009). ORF-internal promoters at the 3'-end of the first of two overlapping genes can drive the expression of the down-stream protein-coding gene, and in these cases the transcripts would not be bona fide sRNAs. Such an example has been characterized for the *HVO_2723/HVO_2722* gene pair of *Hfx. volcanii*

(Maier et al. 2015a). However, these cases will be only a very minor fraction, because the ORF-internal TSS were not enriched at the 3'-ends of genes, but distributed throughout the ORFs (Babski et al. 2016). The functionality of stand-alone internal sRNA genes was proven by overexpression of two examples, which led to a clear phenotypic difference between the overexpression mutants and the wild-type (Gomez-Filho et al. 2015). Overexpression of two internal sRNAs (VNG_aot0042 and VNG_R0052) resulted in a slight increase in growth rate and an about 50% increase in growth yield, compared with control cultures containing the empty expression vector. The molecular mechanism of action of the cis sense sRNAs is not clear and has to be clarified in the future. One obvious possibility is that the internal sRNAs regulate the protein-coding mRNAs via competition for RNA-binding proteins, in particular RNases. However, as yet there is no experimental evidence that this is the mode of operation of internal sRNAs. The very high number of more than 1100 internal sRNAs indicates that they will have an important influence on the physiology of haloarchaea, and that the understanding of regulatory networks in haloarchaea will remain incomplete without the analysis of internal sRNAs.

A special class of internal sRNAs are Transcription Start Site associated RNAs (TSSaRNAs). They represent transcripts that are initiated at the promoters of protein-coding genes, but are terminated soon after initiation (Zaramela et al. 2014). They have been found to occur and to be ubiquitous in all three domains of life. Most probably they are products of regulatory mechanisms that involve pausing of RNA polymerase to differentially decide about further elongation or termination of transcription. Therefore, these sRNAs do not have a regulatory function themselves, but they are the products of a co-transcriptional regulatory mechanism.

10.3.3 *Cis Antisense sRNAs (asRNAs)*

The dRNA-Seq study has revealed that *Hfx. volcanii* contains more than 1200 cis antisense RNAs, and thus asRNAs form the largest group of non-coding RNAs (Babski et al. 2016). During exponential growth under optimal conditions asRNAs were present to 30% of all protein-coding genes. Figure 10.3 shows that the levels of the asRNAs and the levels of the cognate sense mRNAs exhibited a very strong negative correlation, i.e. when the levels of the asRNAs were high, the levels of the cognate mRNAs were very low, and vice versa. This is a strong indication that the antisense RNAs are negative regulators of gene expression, and that duplex formation between mRNA and asRNA leads to degradation. This would require the presence of a double-strand specific RNase, which still needs to be identified.

The presence of a high fraction of asRNAs is not confined to haloarchaea, but seems to be widespread in various archaea and bacteria. For example, asRNAs to 26% of all genes have been found in *Methanococcus psychrophilus* (Li et al. 2015) and in *Pyrococcus abyssi* (Toffano-Nioche et al. 2013). Examples for

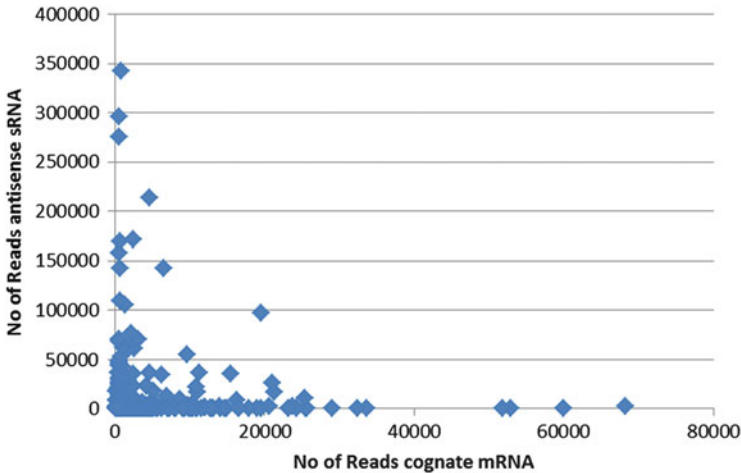


Fig. 10.3 Scatter plot of the levels of antisense sRNAs and the corresponding mRNAs. The scatter plot shows the strong negative correlation between the levels of the asRNAs and the cognate target mRNAs (taken from Babski et al. 2016)

bacteria with a high fraction of asRNAs are *Staphylococcus aureus* with 50% and *Prochlorococcus* strains with up to 75% (Georg and Hess 2011).

One specific function of asRNAs in haloarchaea seems to be the regulation of transposition, which has also been described to be true for bacteria (Brantl 2007). 134 of the 1244 asRNAs of *Hfx. volcanii* were antisense to the genes of transposases, underscoring the model of antisense regulation of transposition. This is not confined to haloarchaea, e.g. asRNAs to transposons have also been described to occur in *T. kodakarensis* (Jäger et al. 2014), *Sulfolobus solfataricus* (Wurtzel et al. 2010), and *M. mazei* (Jäger et al. 2009). Regulation of transposition by asRNAs has also been described to operate in bacteria (review: Ellis and Haniford 2016). While the regulation of transposition by asRNAs seems to be wide-spread in prokaryotes, only a minor fraction of asRNAs target transposons, most are directed against mRNAs encoding proteins of the cellular metabolism. Future work is needed to unravel the molecular regulatory mechanism of asRNAs in *H. volcanii*, irrespective of the identity of their target mRNAs.

10.4 tRNA-Derived Fragments

Recently it was discovered that tRNAs can be cleaved into “tRNA-derived fragments” (tRFs) in all three domains of life. The processing is induced by specific conditions, e.g. stress conditions, and the resulting fragments can have very different half lives and functions (review: Gebetsberger and Polacek 2013). The existence of tRFs in haloarchaea was discovered in the course of a transcriptome

analysis via High Throughput Sequencing (Heyer et al. 2012). tRFs were found for 11 of the 51 tRNAs of *Hfx. volcanii*. The tRFs were typically detected under one or two of the six tested conditions, underscoring the differential generation of tRFs. Northern blot analysis was performed to determine the lengths of the tRFs from tRNA^{Gln} (about 40 nt) and from tRNA^{His} (about 65 nt) and to show differential levels under different conditions.

In an independent approach all sRNAs were identified that could be co-purified with ribosomes using density gradients (Gebetsberger et al. 2012). The ribosomes were isolated from cultures that had been exposed to 1 of 11 different stress conditions, respectively, and the co-isolated sRNAs in the range from 20 to 500 nt were identified by HTS. In total, tRFs from 12 tRNAs could be identified, which had a length distribution from 10 to 49 nt. However, 1 tRF of 26 nt dominated the library and generated more than 85% of all reads. It was derived from two paralogous valine tRNAs (GAC) that are encoded adjacently in the genome of *Hfx. volcanii*. Processing of the tRNA^{Val} into the tRF was condition-dependent and occurred nearly exclusively under alkaline stress at a pH of 8.5. In contrast, the tRF was absent under optimal conditions and under various other stress conditions, e.g. a hypoosmotic shock or UV irradiation. The tRF^{Val} was shown to bind to the small subunit of the ribosome, and it could severely inhibit translation in an *in vitro* translation system (Gebetsberger et al. 2012). Furthermore, it could be shown that binding of tRF^{Val} to the ribosome can displace the mRNA, which results in a stress-induced global attenuation of translation *in vitro* and *in vivo* (Gebetsberger et al. 2016). The processing of tRNAs into tRFs that bind to the ribosome and inhibit translation represent an extremely fast response to the onset of stress conditions. In addition, tRNAs are extremely old, and thus it can be speculated that their usage in stress response circuits started early in evolution, in agreement with the occurrence of tRFs in all three domains of life.

10.5 CRISPR/Cas Defence Systems in Haloarchaea

It was not less than a sensation as it was discovered about a decade ago that prokaryotes contain adaptive immune systems that are directed against invading nucleic acids like phages or plasmids (reviews: van der Oost et al. 2014; Westra et al. 2014). The systems are comprised of “Clustered Regularly Interspaced Palindromic Repeats” (CRISPR) and “CRISPR Associated” (Cas) protein genes. In short, when cells survive the attack of a virus or a plasmid, short sequences of the attacking nucleic acid are integrated into CRISPR locus as spacer sequences between repeated motifs. Transcription of the CRISPR locus results in long transcripts, which are processed into small crRNAs that each contain the recognition motif for one invader. Upon a new infection, the crRNAs direct the Cas proteins to the foreign DNA (or RNA) and enables its destruction. About half of all bacteria and nearly all archaea contain such CRISPR/Cas systems. The CRISPR/Cas

systems are not identical, based on the inventory of the Cas proteins they have been classified into several groups (Makarova et al. 2011).

The CRISPR/Cas system of *H. volcanii* has been intensely studied in recent years (Maier et al. 2012, 2013, 2015b, c; Marchfelder et al. 2012). *H. volcanii* contains three CRISPR loci and eight Cas genes. All three CRISPR loci are transcribed constitutively and thus the system is active in the absence of any invader. Genetic approaches have been established that allowed the characterization of the importance of the repeat sequences, the spacers, and the Cas proteins. The system has also been modified as a molecular genetic tool to down-regulate the expression of any gene of interest. A whole chapter of this book is devoted to the haloarchaeal CRISPR/Cas system, therefore, it will not be discussed any further in this chapter.

10.6 sRNAs That Are Not Well-Studied in Haloarchaea

In eukaryotes small nucleolar RNAs (snoRNAs) form a large and important class of sRNAs with a variety of functions (Lui and Lowe 2013). Their canonical functions are to be part of RNP complexes and guide enzymes to target sites on ribosomal RNAs, leading either to 2'-O-methylation of ribose (C/D box snoRNAs) or to the formation of pseudouridine (H/ACA snoRNAs). Because archaeal sRNAs fulfill the same functions and interact with archaeal proteins that are homologous to eukaryotic proteins, they were also called “snoRNAs” in spite of the lack of a nucleolus or nucleus in archaea. Recently, it has been proposed to rename them to C/D box sRNAs and H/ACA guide sRNAs. This terminology will be used when one class of these sRNAs is discussed, the term “snoRNAs” will still be used when both classes are summarized. Two recent reviews summarize the knowledge about these classes of archaeal sRNAs (Tripp et al. 2017; Lui and Lowe 2013).

The number of snoRNAs is especially high in thermophilic archaea, e.g. more than 80 C/D box sRNAs have been identified in several species of *Pyrococcus* (Bernick et al. 2012). In contrast, the number of snoRNAs is very low in haloarchaea, and only a single C/D box sRNA is present in the genome annotation of *H. volcanii* (Hartman et al. 2010). In addition, a second C/D box sRNA has been characterized that is encoded in an intron of the tRNA^{Trp}, and which was shown to be essential for methylation of the pre-tRNA^{Trp} at positions 34 and 39 (Clouet d'Orval et al. 2001). In any case, the number of snoRNAs in haloarchaea is very low. Therefore, it is not surprising that the role of archaeal snoRNAs has not been analyzed in haloarchaea, but in other archaeal groups. A chapter of this book is devoted to the characterization of archaeal snoRNAs.

In eukaryotes, the 7S RNA is part of the signal recognition particle (SRP), which is important for the direction of membrane proteins to the cytoplasmic membrane and their faithful integration. The current model is that the SRP stops translation of mRNAs for membrane proteins after the signal sequence has been translated, and translation is restarted after the interaction between SRP and the SRP receptor in the

membrane have ensured the correct localization of the translating ribosome. One very early study showed that the 7S RNA is important for the expression of the gene for the major membrane protein of *Hbt. salinarum*, Bacterioopsin (Gropp et al. 1992). It was concluded that the 7S RNA is probably essential for the expression of membrane protein genes in general. Unfortunately, no study followed to verify this claim for *Hbt. salinarum* or any other haloarchaeal species. However, it is very likely that the 7S RNA is indeed important for membrane protein biosynthesis in haloarchaea in general, because it is conserved also in species that are devoid of bacterioopsin.

In eukaryotes, many circular RNAs (circRNAs) have been described, and they are thought to play important regulatory roles in physiological as well as pathological processes. Recently, the database “circRNADb” was generated, which contains more than 30,000 human exonic circRNAs (Chen et al. 2016). In archaea a single study exists that has used circRNA-seq to systematically identify circRNAs in *Sulfolobus solfataricus* (Danan et al. 2012). A large number of circRNAs have been found, including expected circRNAs like tRNA introns, but the majority were novel circRNAs of unknown function. Also circular forms of C/D box sRNAs and of RNase P were found. Also for *Pyrococcus furiosus* it has been reported that most, if not all C/D box sRNAs exist not only in linear, but also in circular form (Starostina et al. 2004). In haloarchaea, it is known that the splicing of introns from tRNAs and the processing of pre-rRNA leads to circular RNAs (Salgia et al. 2003). However, these circles are thought to be processing intermediates without further biological function, which are degraded soon after their generation. Based on the wide distribution in other phylogenetic groups it is tempting to speculate that also haloarchaea contain circular RNAs with biological (regulatory) functions, however, no experimental evidence has been presented as yet.

In eukaryotes, studies of non-coding regulatory RNAs have initially focused on very short RNAs of only about 20 nt, e.g. miRNAs, siRNAs, and piRNAs. In recent years the so-called “long non-coding RNAs” (lncRNAs) came into focus, and it was discovered that eukaryotes contain thousands of lncRNAs. However, per definition lncRNAs are longer than 200 nt. Because the definition of sRNAs in archaea covers non-coding RNAs from 20 to 500 nt, non-coding RNAs of 200–500 nt are termed lncRNAs (=long) in eukaryotes and sRNAs (=short) in archaea. The majority of the eukaryotic lncRNAs is longer than 500 nt and lncRNAs can be up to several thousand nt long. Non-coding RNAs of such lengths have not been discovered yet for haloarchaea or any other prokaryote.

10.7 Small RNAs Encoding Microproteins

The first genome sequence was published in 1995, it was the genome sequence of the bacterium *Mycoplasma genitalium*, an intracellular pathogen with a reduced genome size. The first genome sequence of the first archaeon, *Methanococcus jannaschii*, followed soon after in 1996. Today, only 20 years later, more than

60,000 prokaryotic genome sequences are available. For the annotation of open reading frames (ORFs) typically a minimal cutoff of 100 codons was used to avoid the massive annotation of false positive small ORFs, which are not real genes. However, this meant that also real small genes that encode microproteins of less than 100 amino acids escaped annotation. In recent years it became evident that microproteins are prevalent and have very important functions in all three domains of life (Eguen et al. 2015; Ramamurthi and Storz 2014; Storz et al. 2014; Cheng et al. 2011).

Already about 10 years ago a study had focused on the characterization of the “low molecular weight proteome” of *Hbt. salinarum* (Klein et al. 2007). The optimization of several techniques was necessary, because standard experimental approaches usually work well with medium-sized proteins, but do not perform well for microproteins. In total, 380 microproteins of less than 100 amino acids could be identified, which are equivalent to 14% of the annotated proteome. Thus the microproteins make up a non-negligible part of the total proteome. It was noted that 20 of these microproteins contain two CPXCG double cysteine motifs and they were proposed to be one-domain zinc finger proteins. As a proof-of-principle that these putative zinc finger microproteins can have important regulatory functions the gene for one of these proteins was deleted (Tarasov et al. 2008). The mutant was defective in the expression of the bacterioopsin (*bop*) gene and consequently could no longer grow phototrophically. Also the replacement of one of the cysteines by a serine led to a loss *bop* gene expression and the ability to use light to drive the energy metabolism. Furthermore, the mutants were unable to synthesize carotenoids because the transcript level of the phytoene synthase was decreased. These results underscore the importance of one 60 amino acid microprotein for the physiology of *Hbt. salinarum*. Subsequently it was discovered that the transcript was in fact bicistronic and downstream of the zinc finger microprotein another microprotein of 55 amino acids was encoded, which also is involved in regulation of *bop* gene expression (Tarasov et al. 2011). This further enlarged the regulatory network of phototrophy of *H. salinarum*, which was known before to contain several normal-sized proteins.

The experimental analysis of the low molecular weight proteome has aided the annotation of small protein genes in other haloarchaea. For example, the annotation of the genome of *Hfx. volcanii* currently contains 575 genes for microproteins of less than 100 amino acids, 69 of which are putative one-domain zinc finger proteins with two CPXCG motifs. This is equivalent to 14% of all proteins, like in *Hbt. salinarum*, and the fraction is higher than the average fraction in prokaryotes, which is around 11% (Cheng et al. 2011). The vast majority of these microproteins (72%) are annotated as “hypothetical proteins” and do not have known functions. Some examples of microproteins with known functions are several ribosomal proteins, the cold shock proteins, and the Lsm protein. The Lsm (Like Sm) protein belongs to a large family of RNA-binding proteins. In eukaryotes, Sm and Lsm proteins have many different functions, for example, they are components of the spliceosome (review: Wilusz and Wilusz 2013). The Hfq protein also belongs to this protein family, which is important for the function of intergenic sRNAs in bacteria (review:

negative (D, E) and positive (K, R) charges / S, T, Y / N, Q

HVO_

0197A MPVCDHCGSHVSERFARVFDKNGQVLACPNCSSANAGIAEVARQQRARTA

0241 MPAVQCRECGRDVAVHHEIETTCKTTPDGFDTTRYRCPYCKSEMARDVKTRIV

0325 MSQCNCDAFVSNFVRVFGDEEDGNVYACPSCSANAGISQVSSERRASSL

0416 MASAPSDDLFDQFLTDRGHETE PARWDRSYNKLQCPDCGALHDMGAATCSVCGWVPEA

0489 MSATTTDDYEFTECSQCEGFVAVNGGMRAAVLERGCPICGSPVSADAFDPCPA

0490 MDSADYTFVCPCEAESMLVNDMSRDAPLENGCVVCSAALTTDAFSAT

0546 MRDAPDPAGPPPVCDDCGARHSFRRRFVTTGGWRVTVYRCSECGSRASSDGDGDE

0695 MHTCGNCGEFVSRDFVRFVGNMDEVVGCPCACMNMREVMQDGAAGQTSGRVWRTRA

0758 MKTTRKGLRDGELEKDTYGRLLTCECGESLKKKNDPEVFSVRICADCGREWKELR

0919 MWCECRGRDTTVRKHAVDEFTRFLCNDCRVAVWDRFVSA

0993 MSFEWVSYFTCEECGLERPSPEVVPYDRLGAVCPGCGAETRPAAAAMADASTPVDG

1118 MREFDVTCPCEGGERYRVNEPMMRTLRETTGCVLCTAPLNDAAKSA

1533 MVLHNSVIDDYHPTEGGYECRSCRTRTVSASHLSECPDCGGSVRNIAVARE

1677 MVEAFVRLLCPECGKDWETNPTELPARHNDNYSQSCGATRRTAEFMRTERDLQTLKQFE

1848A MPTCONCNSFVTEGYVRVFAPEGMDAPRVCPHCEDLVRDGSQVREARTRH

2400 MSDLEIERECPACGNDTFYLAASMEIHLGTTKTKWHCTECDYGYIHIITDDIETYAKAEA

2753 MSESEQRHAHQCVSCGINIAGMSAATFKCPDCGQEISRCSKCRKQSNLYECPDCGFMGP

2805A MPECONCGSFVTPAYARVFTPDGMENPRVCPNCEDMVVDGAQVRAARS PRNH

2901 MAGLQQQRARGRDMLECRGGCAVFPPEGRATNDGWTVYVCPCEQVEGIGEGLRRL

Fig. 10.4 Sequences of 19 arbitrarily chosen one-domain zinc finger microproteins of *H. volcanii*. The two CPXCG motifs are *underlined*. Charged and hydrophilic amino acids are color-coded, as indicated on *top*. The HVO numbers (*left*) are the gene designations in the genome annotation of *H. volcanii* (www.halolex.mpg.de)

Wilusz and Wilusz 2013). The haloarchaeal Lsm protein has also been shown to bind sRNAs and to have important regulatory functions, because a deletion mutant has a very severe growth defect (Fischer et al. 2010).

Also only very few of the 69 *Hfx. volcanii* putative one-domain zinc finger microproteins with CPXCG motifs have annotated functions, e.g. as ribosomal proteins or a small subunit of RNA polymerase. Figure 10.4 shows that proteins of this family have a very high fraction of charged and hydrophilic amino acids, which are indicative of binding to many interaction partners and of posttranslational modifications. It will be interesting to unravel functions of more examples of this interesting family of microproteins.

The characterization of microproteins has also been started in methanogenic archaea. Three microproteins of 23–61 amino acid lengths have been identified by LC-MSMS in cell extracts of *Methanosarcina mazei* (Prasse et al. 2015). Two of them had increased levels during mid-exponential growth phase under nitrogen limitation. Overproduction of the three microproteins resulted in transcript level changes of 40–159 transcripts. However, phenotypic changes between the wild-type and the three microprotein overproducers could not be observed (Prasse et al. 2015). Optimization of experimental approaches, e.g. including gel-free LC-MS, increased the number of experimentally verified microproteins of *M. mazei* to 28 (Cassidy et al. 2016), and it is easy to predict that the number will further increase in the future.

The characterization of microproteins and their biological roles is an emerging field in molecular biology. The German Research Council (DFG) has reacted to this challenge and is currently setting up a Priority Program, which is devoted to the analysis of microproteins in prokaryotes and will operate from August 2017 to July 2023.

10.8 Conclusions and Outlook

The recent improvements of RNA-Seq and derivatives thereof have led to the identification of thousands of non-coding sRNAs, not only in haloarchaea, but also in other phylogenetic groups of archaea and bacteria. However, the prevalence of non-coding sRNAs over protein-coding mRNAs is not universally conserved, but specific for certain species or groups. Most probably not all sRNAs have been discovered yet, because their levels vary substantially in different environmental conditions, and thus further studies under additional conditions will most probably further increase the numbers of haloarchaeal sRNAs.

The most intensely studied group of sRNAs are intergenic sRNAs, which have been shown to be important for many biological functions. Future studies will concentrate on the identification of their target mRNAs, either by experimental approaches or using optimized bioinformatics approaches, and on the analysis of their molecular mechanisms of action. The high fraction of leaderless mRNAs in haloarchaea makes it likely that many sRNAs will be found to interact with the 3'-UTRs of their targets.

Haloarchaea contain a much higher number of cis sense sRNAs than intergenic sRNAs. These cis sense sRNAs have hardly been studied and were long thought to be degradation intermediates. However, the dRNA-Seq approach ensured that all of the listed cis sense sRNAs are primary transcripts and no processing intermediates, and two characterized examples verified that they have a regulatory function *in vivo*. The largest group of haloarchaeal sRNAs are asRNAs. The high negative correlation between their levels and the levels of the cognate mRNAs led to the prediction that most of them will turn out to be negative regulators of gene expression. Additional classes of haloarchaeal sRNAs include the crRNAs from the CRISPR/Cas systems and the tRNA-derived fragments, both of which are being studied intensively, and snoRNAs and circular RNAs, which have not been studied until now.

The last group of sRNAs is formed by small mRNAs that encode microproteins of less than 100 amino acids. The analysis of the roles and mechanisms of microproteins is an emerging field of research, which has been initiated not only with haloarchaea, but also with methanogenic archaea and many groups of bacteria.

In summary, haloarchaea contain a zoo of different small RNAs, most of which have only been identified very recently. It is easy to predict that future work will lead to unprecedented insight into the RNA regulatory networks in haloarchaea—and will yield many surprises. The change of concept has only been started, that will

change of the view on haloarchaeal genomes from protein-encoding entities with a few RNA genes to DNA molecules that encode mostly RNA—and additionally contains a minor fraction of protein-encoding genes.

References

- Babski J, Tjaden B, Voss B, Jellen-Ritter A, Marchfelder A, Hess WR, Soppa J (2011) Bioinformatic prediction and experimental verification of sRNAs in the haloarchaeon *Haloflexax volcanii*. *RNA Biol* 8:806–816
- Babski J, Maier LK, Heyer R, Jaschinski K, Prasse D, Jäger D, Randau L, Schmitz RA, Marchfelder A, Soppa J (2014) Small regulatory RNAs in Archaea. *RNA Biol* 11:489–493
- Babski J, Haas KA, Nather-Schindler D, Pfeiffer F, Forstner KU, Hammelmann M, Hilker R, Becker A, Sharma CM, Marchfelder A, Soppa J (2016) Genome-wide identification of transcriptional start sites in the haloarchaeon *Haloflexax volcanii* based on differential RNA-Seq (dRNA-Seq). *BMC Genomics* 17:629
- Bernick DL, Dennis PP, Lui LM, Lowe TM (2012) Diversity of antisense and other non-coding RNAs in archaea revealed by comparative small RNA sequencing in four *Pyrobaculum* species. *Front Microbiol* 3:231
- Borges F, Martienssen RA (2015) The expanding world of small RNAs in plants. *Nat Rev Mol Cell Biol* 16:727–741
- Brantl S (2007) Regulatory mechanisms employed by cis-encoded antisense RNAs. *Curr Opin Microbiol* 10:102–109
- Cassidy L, Prasse D, Linke D, Schmitz RA, Tholey A (2016) Combination of bottom-up sD-LC-MS and semi-top-down gelFree-LC-MS enhances coverage of proteome and low molecular weight short open reading frames encoded peptides of the archaeon *Methanosarcina mazei*. *J Proteome R* 15:3773–3783
- Catalanotto C, Cogoni C, Zardo G (2016) MicroRNA in control of gene expression: an overview of nuclear functions. *Int J Mol Sci* 17:E1712
- Chen X, Han P, Zhou T, Guo X, Song X, Li Y (2016) circRNADb: a comprehensive database for human circular RNAs with protein-coding annotations. *Sci Rep* 6:34985
- Cheng H, Chan WS, Li Z, Wang D, Liu S, Zhou Y (2011) Small open reading frames: current prediction techniques and future prospect. *Curr Protein Pept Sci* 12:503–507
- Clouet d’Orval B, Bortolin ML, Gaspin C, Bachellerie JP (2001) Box C/D RNA guides for the ribose methylation of archaeal tRNAs. The tRNATrp intron guides the formation of two ribose-methylated nucleosides in the mature tRNATrp. *Nucleic Acids Res* 29:4518–4529
- Danan M, Schwartz S, Edelheit S, Sorek R (2012) Transcriptome-wide discovery of circular RNAs in Archaea. *Nucleic Acids Res* 40:3131–3142
- Daniel JM, Penzkofer D, Teske R, Dutzmann J, Koch A, Bielenberg W, Bonauer A, Boon RA, Fischer A, Bauersachs J, van Rooij E, Dimmeler S, Sedding DG (2014) Inhibition of miR-92a improves re-endothelialization and prevents neointima formation following vascular injury. *Cardiovasc Res* 103:564–572
- Eguen T, Straub D, Graeff M, Wenkel S (2015) MicroProteins: small size-big impact. *Trends Plant Sci* 20:477–482
- Ellis MJ, Haniford DB (2016) Riboregulation of bacterial and archaeal transposition. *WIREs RNA* 7:382–398
- Fischer S, Benz J, Spath B, Maier LK, Straub J, Granzow M, Raabe M, Urlaub H, Hoffmann J, Brutschy B, Allers T, Soppa J, Marchfelder A (2010) The archaeal Lsm protein binds to small RNAs. *J Biol Chem* 285:34429–34438
- Gebetsberger J, Polacek N (2013) Slicing tRNAs to boost functional ncRNA diversity. *RNA Biol* 10:1798–1806

- Gebetsberger J, Zywicki M, Kunzi A, Polacek N (2012) tRNA-derived fragments target the ribosome and function as regulatory non-coding RNA in *Haloferax volcanii*. *Archaea* 2012:260909
- Gebetsberger J, Wyss L, Mleczko AM, Reuther J, Polacek N (2016) A tRNA-derived fragment competes with mRNA for ribosome binding and regulates translation during stress. *RNA Biol*. doi:10.1080/15476286.2016.1257470
- Georg J, Hess WR (2011) *cis*-antisense RNA, another level of gene regulation in bacteria. *Microbiol Mol Biol Rev* 75:286–300
- Gomes-Filho JV, Zaramela LS, Italiani VC, Baliga NS, Vencio RZ, Koide T (2015) Sense overlapping transcripts in *ISJ341*-type transposase genes are functional non-coding RNAs in archaea. *RNA Biol* 12:490–500
- Gropp R, Gropp F, Betlach MC (1992) Association of the haloarchaeal 7S RNA to the polysome correlates with expression of the membrane protein bacterioopsin. *Proc Natl Acad Sci U S A* 89:1204–1208
- Hartman AL, Norais C, Badger JH, Delmas S, Haldenby S et al (2010) The complete genome sequence of *Haloferax volcanii* DS2, a model archaeon. *PLoS One* 5:e9605
- Heyer R, Dorr M, Jellen-Ritter A, Spath B, Babski J, Jaschinski K, Soppa J, Marchfelder A (2012) High throughput sequencing reveals a plethora of small RNAs including tRNA derived fragments in *Haloferax volcanii*. *RNA Biol* 9:1011–1018
- Huang Y, Zhang JL, Yu XL, Xu TS, Wang ZB, Cheng XC (2013) Molecular functions of small regulatory noncoding RNA. *Biochemistry (Moscow)* 78:221–230
- Jäger D, Sharma CM, Thomsen J, Ehlers C, Vogel J, Schmitz RA (2009) Deep sequencing analysis of the *Methanosarcina mazei* Go1 transcriptome in response to nitrogen availability. *Proc Natl Acad Sci U S A* 106:21878–21882
- Jäger D, Pernitzsch SR, Richter AS, Backofen R, Sharma CM, Schmitz RA (2012) An archaeal sRNA targeting *cis*- and *trans*-encoded mRNAs via two distinct domains. *Nucleic Acids Res* 40:10964–10979
- Jäger D, Förstner KU, Sharma CM, Santangelo TJ, Reeve JN (2014) Primary transcriptome map of the hyperthermophilic archaeon *Thermococcus kodakarensis*. *BMC Genomics* 15:568
- Jaschinski K, Babski J, Lehr M, Burmester A, Benz J, Heyer R, Dorr M, Marchfelder A, Soppa J (2014) Generation and phenotyping of a collection of sRNA gene deletion mutants of the haloarchaeon *Haloferax volcanii*. *PLoS One* 9:e90763
- Klein C, Aivaliotis M, Olsen JV, Falb M, Besir H, Scheffer B, Bisle B, Tebbe A, Konstantinidis K, Siedler F, Pfeiffer F, Mann M, Oesterhelt D (2007) The low molecular weight proteome of *Halobacterium salinarum*. *J Proteome Res* 6:1510–1518
- Koide T, Reiss DJ, Bare JC, Pang WL, Facciotti MT, Schmid AK, Pan M, Marzolf B, Van PT, Lo FY, Pratap A, Deutsch EW, Peterson A, Martin D, Baliga NS (2009) Prevalence of transcription promoters within archaeal operons and coding sequences. *Mol Syst Biol* 5:285
- Kopf M, Hess WR (2015) Regulatory RNAs in photosynthetic cyanobacteria. *FEMS Microbiol Rev* 39:301–315
- Li J, Guo Y, Yue L, Li Y, Ge W et al (2015) Global mapping of transcriptional start sites revealed both transcriptional and post-transcriptional regulation of cold adaptation in the methanogenic archaeon *Methanoblobus psychrophilus*. *Sci Rep* 5:9209
- Lui L, Lowe T (2013) Small nucleolar RNAs and RNA-guided post-transcriptional modification. *Biochem Soc Essays Biochem* 54:53–77
- Maier LK, Fischer S, Stoll B, Brendel J, Pfeiffer F, Dyall-Smith M, Marchfelder A (2012) The immune system of halophilic archaea. *Mob Genet Elements* 2:228–232
- Maier LK, Lange SJ, Stoll B, Haas KA, Fischer S, Fischer E, Duchardt-Ferner E, Wohnert J, Backofen R, Marchfelder A (2013) Essential requirements for the detection and degradation of invaders by the *Haloferax volcanii* CRISPR/Cas system I-B. *RNA Biol* 10:865–874
- Maier LK, Benz J, Fischer S, Alstetter M, Jaschinski K, Hilker R, Becker A, Allers T, Soppa J, Marchfelder A (2015a) Deletion of the SM1 encoding motif in the *lsm* gene results in distinct

- changes in the transcriptome and enhanced swarming activity of *Haloferax* cells. *Biochimie* 117:129–137
- Maier LK, Dyall-Smith M, Marchfelder A (2015b) The adaptive immune system of *Haloferax volcanii*. *Life (Basel)* 5:521–537
- Maier LK, Stachler AE, Saunders SJ, Backofen R, Marchfelder A (2015c) An active immune defense with a minimal CRISPR (clustered regularly interspaced short palindromic repeats) RNA and without the Cas6 protein. *J Biol Chem* 290:4192–4201
- Makarova KS, Haft DH, Barrangou R, Brouns SJ, Charpentier E, Horvath P, Moineau S, Mojica FJ, Wolf YI, Yakunin AF, van der Oost J, Koonin EV (2011) Evolution and classification of the CRISPR-Cas systems. *Nat Rev Microbiol* 9:467–477
- Marchfelder A, Fischer S, Brendel J, Stoll B, Maier LK, Jäger D, Prasse D, Plagens A, Schmitz RA, Randau L (2012) Small RNAs for defence and regulation in archaea. *Extremophiles* 16:685–696
- Martens B, Manoharadas S, Hasenohrl D, Manica A, Blasi U (2013) Antisense regulation by transposon-derived RNAs in the hyperthermophilic archaeon *Sulfolobus solfataricus*. *EMBO Rep* 14:527–533
- Murina VN, Nikulin AD (2015) Bacterial small regulatory RNAs and Hfq protein. *Biochemistry (Moscow)* 80:1647–1654
- Prasse D, Ehlers C, Backofen R, Schmitz RA (2013) Regulatory RNAs in archaea: first target identification in methanoarchaea. *Biochem Soc Trans* 41:344–349
- Prasse D, Thomsen J, De Santis R, Muntel J, Becher D, Schmitz RA (2015) First description of small proteins encoded by spRNAs in *Methanosarcina mazei* strain Go1. *Biochimie* 117:138–148
- Ramamurthi KS, Storz G (2014) The small protein floodgates are opening; now the functional analysis begins. *BMC Biol* 12:96
- Salgia SR, Singh SK, Gurha P, Gupta R (2003) Two reactions of *Haloferax volcanii* RNA splicing enzymes: joining of exons and circularization of introns. *RNA* 9:319–330
- Schmitz-Streit R, Jäger D, Jellen-Ritter A, Babski J, Soppa J, Marchfelder A (2011) Archaea employ small RNAs as regulators. In: Hess M (ed) *Regulatory RNAs in prokaryotes*. Springer, Wien, pp 131–145
- Starostina NG, Marshburn S, Johnson LS, Eddy SR, Terns RM, Terns MP (2004) Circular box C/D RNAs in *Pyrococcus furiosus*. *Proc Natl Acad Sci U S A* 101:14097–14101
- Storz G, Wolf YI, Ramamurthi KS (2014) Small proteins can no longer be ignored. *Annu Rev Biochem* 83:753–777
- Straub J, Brenneis M, Jellen-Ritter A, Heyer R, Soppa J, Marchfelder A (2009) Small RNAs in haloarchaea: identification, differential expression and biological function. *RNA Biol* 6 (3):281–292
- Tang TH, Bachelier JP, Rozhdestvensky T, Bortolin ML, Huber H, Drungowski M, Elge T, Brosius J, Huttenhofer A (2002) Identification of 86 candidates for small non-messenger RNAs from the archaeon *Archaeoglobus fulgidus*. *Proc Natl Acad Sci U S A* 99:7536–7541
- Tang TH, Polacek N, Zywicki M, Huber H, Brugger K, Garrett R, Bachelier JP, Huttenhofer A (2005) Identification of novel non-coding RNAs as potential antisense regulators in the archaeon *Sulfolobus solfataricus*. *Mol Microbiol* 55:469–481
- Tarasov VY, Besir H, Schwaiger R, Klee K, Furtwängler K, Pfeiffer F, Oesterhelt D (2008) A small protein from the *bop-brp* intergenic region of *Halobacterium salinarum* contains a zinc finger motif and regulates *bop* and *crtB1* transcription. *Mol Microbiol* 67:772–780
- Tarasov V, Schwaiger R, Furtwängler K, Dyall-Smith M, Oesterhelt D (2011) A small basic protein from the *brz-brp* operon is involved in the regulation of *bop* transcription in *Halobacterium salinarum*. *BMC Mol Biol* 12:42
- Toffano-Nioche C, Ott A, Crozat E, Nguyen AN, Zytynic M, Leclerc F, Forterre P, Boulloc P, Gautheret D (2013) RNA at 92°C: the non-coding transcriptome of the hyperthermophilic archaeon *Pyrococcus abyssi*. *RNA Biol* 10:1211–1220

- Tripp V, Martin R, Orell A, Alkhnbashi OS, Backofen R, Randau L (2017) Plasticity of archaeal C/D box sRNA biogenesis. *Mol Microbiol* 103:151–164
- Van der Oost J, Westra ER, Jackson RN, Wiedenheft B (2014) Unravelling the structural and mechanistic basis of CRIPR-Cas systems. *Nat Rev Microbiol* 12:479–492
- Van Puyvelde S, Vanderleyden J, De Keersmaecker SC (2015) Experimental approaches to identify small RNAs and their diverse roles in bacteria—what we have learnt in one decade of MicA research. *Microbiologyopen* 4:699–711
- Wagner EG, Romby P (2015) Small RNAs in bacteria and archaea: who they are, what they do, and how they do it. *Adv Genet* 90:133–208
- Wan G, Liu Y, Han C, Zhang X, Lu X (2014) Noncoding RNAs in DNA repair and genome integrity. *Antioxid Redox Signal* 20:655–677
- Waters LS, Storz G (2009) Regulatory RNAs in bacteria. *Cell* 136:615–628
- Westra ER, Buckling A, Fineran PC (2014) CRISPR-Cas systems: beyond adaptive immunity. *Nat Rev Microbiol* 12:317–326
- Wilusz CJ, Wilusz J (2013) Lsm proteins and Hfq: life at the 3'-end. *RNA Biol* 10:592–601
- Wurtzel O, Sapra R, Chen F, Zhu Y, Simmons BA, Sorek R (2010) A single-base resolution map of an archaeal transcriptome. *Genome Res* 20:133–141
- Yang JX, Rastetter RH, Wilhelm D (2016) Non-coding RNAs: an introduction. *Adv Exp Med Biol* 886:13–32
- Zaramela LS, Vencio RZN, ten Caten F, Baligy N, Koide T (2014) Transcription start site associated RNAs (TSSaRNAs) are ubiquitous in all domains of life. *PLoS One* 9:e107680