

## Small RNAs for defence and regulation in archaea

Anita Marchfelder · Susan Fischer · Jutta Brendel · Britta Stoll ·  
Lisa-Katharina Maier · Dominik Jäger · Daniela Prasse ·  
André Plagens · Ruth A. Schmitz · Lennart Randau

Received: 26 March 2012 / Accepted: 14 June 2012 / Published online: 5 July 2012  
© The Author(s) 2012. This article is published with open access at Springerlink.com

**Abstract** Non-coding RNAs are key players in many cellular processes within organisms from all three domains of life. The range and diversity of small RNA functions beyond their involvement in translation and RNA processing was first recognized for eukaryotes and bacteria. Since then, small RNAs were also found to be abundant in archaea. Their functions include the regulation of gene expression and the establishment of immunity against invading mobile genetic elements. This review summarizes our current knowledge about small RNAs used for regulation and defence in archaea.

**Keywords** sRNA · Lsm · Hfq · Archaea · CRISPR · crRNA

### Abbreviations

sRNA Small RNA  
HTS High throughput sequencing  
crRNA CRISPR RNA

CRISPR Clustered regularly interspaced short  
palindromic repeats

Cas CRISPR associated

### Small RNAs regulate gene expression

Bacteria and eukaryotes use a plethora of non-coding RNAs to regulate gene expression (Hüttenhofer et al. 2005; Brantl 2009; Waters and Storz 2009; Krol 2010). The mechanisms by which bacterial as well as eukaryotic small RNAs act have been studied in detail revealing common characteristics as well as differences. Bacterial small regulatory RNAs are often required for regulation of metabolic pathways (Gottesman 2004a, b). Bacterial *trans*-encoded sRNAs can act by masking the ribosome binding site or the start codon or binding to sequences close to these sites (Gottesman 2004a, b; Waters and Storz 2009). *Trans*-encoded sRNAs are only partially complementary to their target and often require the help of a protein (i.e. Hfq) for activity. In the last years more and more bacterial RNA populations have been analyzed with deep-sequencing methodologies revealing also a high amount of *cis*-encoded antisense RNAs (Rasmussen et al. 2009; Thomason and Storz 2010; Georg and Hess 2011; Brantl 2012). *Cis*-antisense RNAs are encoded on the opposite strand of a gene and are therefore completely complementary to their target. It is estimated that a bacterial genome encodes about 200–300 sRNAs and in *Escherichia coli* for instance about 140 sRNAs are known but a biological role has been defined for only 25 of these, showing how difficult it is to unravel their in vivo function (Brantl 2012). In addition, sRNAs like the 6S RNA (Gildehaus et al. 2007) have been shown to regulate gene expression by specific binding to proteins (Brantl 2009).

Communicated by S. Albers.

A. Marchfelder · S. Fischer · J. Brendel · B. Stoll · L.-K. Maier  
Biologie II, Ulm University, Albert-Einstein-Allee 11,  
89069 Ulm, Germany

D. Jäger · D. Prasse · R. A. Schmitz  
Institute for General Microbiology,  
Christian-Albrechts-University Kiel,  
Am Botanischen Garten 1-9, 24118 Kiel, Germany

A. Plagens · L. Randau (✉)  
Prokaryotic Small RNA Biology,  
Max Planck Institute for Terrestrial Microbiology,  
Karl-von-Frisch Strasse 10, 35043 Marburg, Germany  
e-mail: lennart.randau@mpi-marburg.mpg.de

In eukaryotes, the majority of the non-coding miRNAs binds to the 3' end of the target mRNAs, which triggers degradation or leads to inhibition of translation (Meister 2007; Guo et al. 2010; Krol 2010). It has been predicted that 30–50 % of all human genes are regulated by miRNAs, emphasizing the importance of small RNAs in regulation of gene expression (Lewis et al. 2005; Krol 2010).

It has been suggested that archaea also use small RNAs to regulate gene expression, but at the moment we do not know details about the molecular mechanism involved in archaeal sRNA regulation. Up to date in only six archaeal organisms the small RNA population has been investigated: *Archaeoglobus fulgidus*, *Sulfolobus solfataricus*, *Methanocaldococcus jannaschii*, *Methanosarcina mazei*, *Pyrococcus furiosus*, and *Haloferax volcanii* (Klein et al. 2002; Schattner 2002; Tang et al. 2002, 2005; Zago et al. 2005; Jäger et al. 2009; Soppa et al. 2009; Straub et al. 2009; Wurtzel et al. 2009; Fischer et al. 2010; Babski et al. 2011). Details about the site and the mode of the interactions between the sRNA and the target are not known and it is therefore not clear whether additional factors are required.

Here, we present the current state of the art about sRNAs in Archaea, their diversity and potential biological functions.

### Elucidation and analysis of the archaeal sRNA population

The first approaches to identify small RNA populations in archaeal organisms were carried out 10 years ago using experimental and computational methods.

### Prediction of sRNA genes

The prediction of non-coding RNA genes is not as straightforward as the prediction of protein coding genes. In bacterial organisms where promoter and terminator elements have been clearly defined those can be used as a tool to predict sRNA genes. This approach has been used successfully together with comparative genome analysis to predict sRNAs in *E. coli* (Argaman et al. 2001; Wassarman et al. 2001). Although promoter and terminator elements have also been defined for archaeal genes, in silico prediction of sRNA genes using these parameters was not successful to date (Soppa et al. 2009; Jäger, Schmitz, Liesegang, unpublished). In addition, for only 38 % of the identified sRNA genes from *Haloferax* and 44 % of the sRNA genes from *Methanosarcina* basal promoter elements were found in an appropriate distance to the transcriptional start site, suggesting that either the sRNAs are processed from precursors or that the sRNA genes have unusual promoter elements. Therefore, other approaches for sRNA prediction are used one being the analysis of the GC content. For example, in hyperthermophiles non-coding RNAs have a higher GC content and thus non-coding RNAs are predicted by the identification of GC rich regions. Another approach employed for sRNA gene prediction is comparative genome analysis, here non-coding RNAs are identified as intergenic regions conserved between at least two organisms.

The first bioinformatics approach to identify archaeal small RNAs was applied in *M. jannaschii* and *P. furiosus* (Klein et al. 2002; Schattner 2002) (Table 1). Since both organisms have a high A/T content, the screen for novel small RNAs used a GC content bias as well as the

**Table 1** sRNAs identified in Archaea

Type of sRNA	<i>M. jannaschii</i>		<i>P. fu</i>	<i>H. volcanii</i>			
(a) In silico identification							
Prediction	(1)	(2)		(1)	(2)		
Intergenic sRNA	5	18	5	31	94		
Type of sRNA	<i>A. fu</i>		<i>S. solfataricus</i>		<i>H. volcanii</i>		<i>M. ma</i>
Method of identification	RNomics	RNomics	CoIP	HTS	RNomics	HTS	HTS
(b) Experimental identification							
Intergenic sRNA	9	11	3	125	21	145	199
<i>cis</i> -Antisense	33	19	8	185	18	45	43

(a) In silico approaches were used to predict sRNAs in *M. jannaschii* (Klein et al. 2002; Schattner 2002); *P. furiosus* (*P. fu*) (Klein et al. 2002) and *H. volcanii* (Babski et al. 2011). The number of sRNAs identified with these approaches is shown. For *M. jannaschii* and *H. volcanii* two different approaches were used

(b) The sRNA populations from *A. fulgidus* (*A. fu*) (Tang et al. 2002); *S. solfataricus* (Tang et al. 2002; Hüttenhofer et al. 2005; Zago et al. 2005; Wurtzel et al. 2009), *H. volcanii* (Straub et al. 2009; Heyer et al. 2012) and *M. mazei* (*M. ma*) (Jäger et al. 2009). The number of sRNAs identified with experimental RNomics, co-immuno precipitation (CoIP) or HTS is shown

programme QRNA finder, which uses a comparative sequence analysis algorithm to detect conserved structural RNAs (Rivas and Eddy 2001). Using this method five new small RNAs were identified in both organisms. The second approach used local base composition statistics to identify small RNAs in *M. jannaschii*. This search resulted in the identification of 18 putative new small RNAs (Schattner 2002). It took another couple of years until sRNA predictions for a third archaeon, *H. volcanii*, were published. Here, two separate bioinformatics approaches were used to elucidate the small RNA population (Babski et al. 2011). Comparison of intergenic regions of *H. volcanii* with one halophilic bacterium, one crenarchaeal species and three haloarchaeal species was used in the first approach. Since genomes from different phylogenetic groups were used for comparison only highly conserved sRNAs could be identified, resulting in the prediction of 31 sRNAs. A comparative analysis restricted to haloarchaeal organisms was used in the second bioinformatics approach; here, 94 putative sRNA genes were identified, which were conserved in at least two or three haloarchaea.

The prediction of sRNAs is an important tool for the analysis of regulatory RNAs but for the predicted sRNA candidates an experimental verification is essential, since the identified conserved structures might be conserved riboswitch-like regulatory elements which are part of a 5' or 3' UTR and do not represent sRNA genes.

#### Experimental identification of small RNAs

The first experimental identification of an archaeal sRNA population was performed with the euryarchaeon *Archaeoglobus fulgidus* (Tang et al. 2002). Tang et al. generated a cDNA library from a size selected RNA fraction (50–500 nt). Sequencing revealed nine potential *trans*-encoded sRNAs and 33 *cis*-antisense sRNAs, in addition 22 short RNA involved in the CRISPR/Cas prokaryotic immune system (crRNAs) were found.

The next set of archaeal sRNAs was analyzed in the crenarchaeon *Sulfolobus solfataricus*. Here, two different approaches were used: experimental RNomics (Tang et al. 2005) and co-immuno-precipitation with the protein L7Ae (Zago et al. 2005). Later a high through sequencing approach was also applied (Wurtzel et al. 2009). In the experimental RNomics approach a cDNA library representing the RNA population from 50 to 500 nucleotides was generated. To avoid identification of low level unspecific transcripts novel small RNA candidates were only further investigated if their expression was confirmed by northern analyses. This approach identified 19 *cis*-antisense sRNAs and 11 *trans*-encoded sRNAs and one crRNA. The majority of the *cis*-antisense sRNAs was encoded opposite to transposase genes, suggesting that the

sRNAs are involved in regulation of transposons (Tang et al. 2005). For the co-immunoprecipitation approach, the *Sulfolobus* L7Ae protein was used to isolate sRNAs (Zago et al. 2005). The archaeal L7Ae protein is a component of the large subunit of the ribosome and part of ribonucleo-protein complexes (RNPs) which are responsible for ribose methylation and pseudouridylation (Zago et al. 2005). The L7Ae protein has been shown to bind to archaeal snoRNAs (Kuhn et al. 2002; Rozhdetsvensky et al. 2003). The co-immuno-precipitation with L7Ae resulted in the identification of three *trans* sRNAs and five *cis*-antisense RNAs (Zago et al. 2005). The HTS approach detected 310 sRNAs including all formerly identified sRNAs (Wurtzel et al. 2009). Interestingly, more *cis*-antisense sRNAs (185) than *trans*-encoded sRNA (125) were detected. Although an increasing number of *cis*-antisense sRNAs are being identified, the number of *trans*-encoded sRNAs has so far been higher than the number of *cis*-antisense sRNAs found. This is the first case, where *cis*-antisense RNAs outnumber the *trans*-encoded sRNAs (Wurtzel et al. 2009).

In the same year the sRNA populations of two euryarchaeota were published: *M. mazei* and *H. volcanii*. RNAseq analysis of *M. mazei* growing under nitrogen starvation and nitrogen saturation identified 199 *trans*-encoded sRNAs as well as 43 *cis*-antisense sRNAs in this organism (Jäger et al. 2009). Comparative genome analysis further revealed that a significant number of the identified *cis*-antisense sRNAs (30 %) and the *trans*-encoded sRNAs (21 %) were conserved in *M. mazei* and *Methanosarcina barkeri* and *M. acetivorans*. Several of the *cis*-antisense sRNA candidates are encoded on the opposite strand to transposase genes indicating a post-transcriptional regulation of transposon mobility by antisense RNAs, a mechanism previously observed for insertion elements in *S. solfataricus* (Tang et al. 2005) and the bacterial transposons Tn10 and Tn30 in *Escherichia coli* (Ma and Simons 1990; Arini et al. 1997). The finding that in *M. mazei* several of those *cis*-antisense sRNAs show different expression levels depending on the nitrogen source (Jäger et al. 2009), strongly suggests that transposition events in *M. mazei* are regulated in response to the nitrogen availability.

40 *trans*-encoded sRNA candidates have the potential to encode peptides smaller than 30 amino acids. As the majority of those small potential ORFs as well as the flanking non-coding RNA region show high conservation in *Methanosarcina barkeri* and *M. acetivorans* it is tempting to speculate that those *trans*-encoded sRNAs might have a dual function as regulatory sRNA and mRNA. Very recently three of those small peptides have been shown to be expressed (Prasse, Jäger, Thomsen, Becher, Hecker and Schmitz, unpublished). Up to now the expression of 88 out of 130 randomly selected sRNA candidates was confirmed

using northern blot analyses. Differential expression in response to the nitrogen source was confirmed for 18 small RNAs, representing the first prokaryotic regulatory RNAs potentially involved in nitrogen stress response. Based on these findings and the strong conservation a crucial role of sRNAs in the nitrogen or general stress response has been proposed for *M. mazei* (Jäger et al. 2009).

For the identification of sRNAs in the halophilic archaeon *H. volcanii* three different experimental approaches were employed, i.e. experimental RNomics (Straub et al. 2009), HTS (Heyer et al. 2012) and co-purification with the Lsm protein (Fischer et al. 2010). Altogether 145 intergenic sRNAs were identified and 45 *cis*-antisense RNAs (Table 1).

For the experimental RNomics approach, a cDNA library was generated from a *Haloferax* RNA fraction containing RNA molecules ranging in size from 130 to 460 nts. Sequencing of this library led to the identification of 18 antisense sRNAs and 21 intergenic sRNAs (Straub et al. 2009). Northern analyses of the candidate sRNAs showed differential expression of several sRNAs. One sRNA has the potential to code for a peptide of 34 amino acids length, which is conserved between haloarchaeal organisms.

In a second approach, a tagged version of the *Haloferax* Lsm protein was used to co-immuno-precipitate sRNAs (Fischer et al. 2010). The archaeal Lsm protein is like the bacterial Hfq protein a member of the Sm/Lsm protein family. Proteins of this family have been shown to be important players in cellular RNA pathways and in many bacteria the Hfq protein is required for interaction of some *trans*-encoded sRNAs with their target mRNAs. Expression of a FLAG-tagged Lsm protein in *Haloferax* cells confirmed that the archaeal protein binds to small RNAs. Sequencing of the co-immuno-precipitated RNAs revealed that 10 intergenic sRNAs and 7 of the bioinformatically predicted sRNAs were bound to the Lsm protein (Fischer et al. 2010). These data suggest that in archaea the Lsm protein might be involved in the sRNA regulation pathway.

In the third experimental approach cDNA libraries generated from size selected RNAs (17–500 nts) were sequenced using HTS. Six different cDNA libraries were constructed from RNA isolated from cultures grown at standard growth conditions, low salt conditions and high temperature, each at exponential and stationary phase. Altogether 145 *trans*-encoded sRNAs and 45 *cis*-antisense RNAs were identified by this approach (Heyer et al. 2012). All sRNAs identified with the experimental RNomics approach were also detected in the HTS analysis. The HTS approach identified also a new class of sRNAs, which were recently discovered in eukaryotes: tRNA-derived fragments (tRFs) (Cole et al. 2009; Lee et al. 2009). Several tRNA 5' fragments, tRNA 3' fragments and tRNA 3' trailer

sequences have been identified as stable molecules in eukaryotic cells. In human cells these fragments are generated from mature tRNAs or precursor-tRNAs and represent an abundant class of small RNAs. The tRNA 3' processing endonuclease tRNase Z (Cole et al. 2009; Hartmann et al. 2009; Lee et al. 2009) as well as Dicer (Jinek and Doudna 2009) seem to be involved in generation of these fragments. Several tRFs have been shown to be associated with Argonaute proteins (Haussecker et al. 2010). Recent studies suggest that these tRNA fragments are not random by-products of tRNA biogenesis and degradation but are an abundant and novel class of short regulatory RNAs, which show specific expression patterns (summarized in Sobala and Hutvagner 2011). Lee et al. (2009) could show that overexpression of a tRF in human cell lines results in higher proliferation rates.

The HTS approach used with *H. volcanii* revealed the presence of 11 tRFs, suggesting that these types of small RNAs are also active in archaea. Northern blot analyses showed the stable presence of these fragments in the cell (Heyer et al. 2012).

The detection of tRFs in *Haloferax* suggests that these are also used in Archaea to regulate gene expression. One potential biological function of these molecules might be the inhibition of translation. In addition, archaeal tRFs might be associated with RNA interacting proteins as seen in eukaryotes. It is apparent that more detailed analyses are required to confirm these speculations and to elucidate their specific role in the cell.

#### Functional analysis

After the identification of the pool of potential sRNAs the next and crucial step is to select the best candidates for regulatory RNAs for further in depth studies. One approach to do that is the construction of sRNA gene deletion mutants and the analysis of their phenotypes in comparison to wild type strains. For *H. volcanii* more than ten deletion strains were successfully generated and with two exceptions all were viable. Phenotypic differences between wild type and sRNA gene deletion mutants could be detected under specific conditions such as high temperature, low salt concentrations and different carbon sources (Straub et al. 2009; Fischer et al. 2011; Heyer et al. 2012). For instance, two mutants showed severe growth defects at high temperature and low salt concentrations, respectively (Straub et al. 2009). These data suggest that in *H. volcanii* sRNAs may be required for metabolic regulation.

Very recently, target analysis based on whole transcriptome changes in mutant strains combined with genome-wide in silico target predictions using the tool IntaRNA (Busch et al. 2008) facilitated the identification of the first *trans*-encoded target mRNA for an archaeal sRNA

in *M. mazei* (Jäger, Pernitzsch, Backofen, Richter, Sharma and Schmitz, submitted). Demonstration of sRNA sequestering the predicted ribosome binding site as well as the translation start codon of its target mRNA in vitro strongly argues that the interaction is likely to inhibit translation initiation, a mechanisms up to now exclusively described for bacterial sRNAs. Additional evidence was obtained by a genetic approach that this sRNA is involved in regulating the metabolic switch from methanol to methylamines fueled methanogenesis, most likely by the above-mentioned post-transcriptional regulation (Jäger, Pernitzsch, Backofen, Richter, Sharma and Schmitz, submitted).

### Outlook for archaeal sRNA research

The six archaeal organisms studied up until now in regard to their sRNA population represent only the tip of the iceberg for our knowledge about sRNAs in archaea. Furthermore, from the phylum of crenarchaeota only one organism was analyzed and none from the phylum thaumarchaeota. One might also expect a certain diversity of sRNA regulation mechanisms within the archaeal domain. Therefore, much data are required to find out how archaea regulate gene expression with small RNAs.

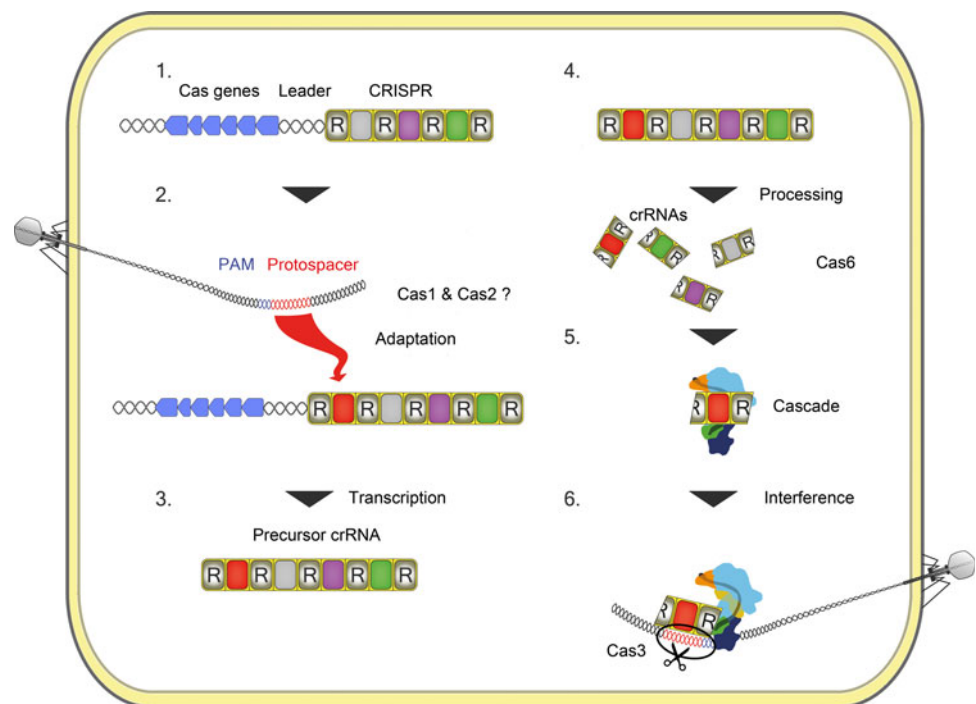
Nevertheless the data acquired up until now show that archaea encode sRNAs including *trans*-encoded sRNA and *cis*-antisense sRNAs. But the details of the interaction between sRNAs and their targets have not been unraveled yet. Further, it is not known what kind of effect an sRNAs

has on its target, whether it, for instance, triggers the degradation of the target mRNA. Also, we do not know yet whether archaeal sRNAs can also act on proteins to regulate gene expression like their bacterial counterparts. In light of the extreme growth conditions found in the archaeal domain the investigation of the sRNA mechanism promises to reveal interesting discoveries about RNA–RNA interactions and RNA–protein interactions. Considering the observation that in some archaea mRNAs contain very short or no 5' UTRs, an interaction of an sRNA with the 3' UTR might also be likely. Altogether the analysis of sRNA regulation in archaea will remain fascinating and promises exciting discoveries in the near future.

### CRISPR/Cas, the adaptive immune system of prokaryotes

In the following sections, we would like to shift the focus to a different family of regulatory RNAs, namely the crRNAs that fulfill executive roles in the antiviral defense mediated by CRISPR/Cas systems. CRISPR sequences were initially discovered in *E. coli* in 1987 (Ishino et al. 1987), but only named and classified as a common family of repetitive DNA sequences in the genomes of many bacteria and archaea in 2002 (Jansen et al. 2002). CRISPR is short for clustered regularly interspaced short palindromic repeats, which describes some of the structural features of these clusters (Fig. 1). The nearly identical repeat sequences of a standard CRISPR array vary in

**Fig. 1** Schematic overview of the CRISPR/Cas type I mechanism. A viral sequence (protospacer, red) is incorporated into the host CRISPR cluster (*R* indicates repeat sequences). The universal Cas1 and Cas2 proteins are proposed to mediate this adaptation step. The CRISPR cluster is transcribed into a pre-crRNA that is processed into mature crRNAs by Cas6. The crRNAs are loaded onto the Cascade complex that delivers crRNA to the viral DNA during a repeat attack. The virus is identified via base complementarity between crRNA and protospacer and is subsequently degraded by Cas3 (color figure online)



length between 23 and 47 nucleotides and are interspaced with so-called spacer sequences of similar size. It was first observed in 2005 that these spacer sequences can represent fragments of the genomes of viruses and other foreign genetic elements (Mojica et al. 2005). Finally in 2007, one major function of the ubiquitous CRISPR arrays was discovered. It was shown that *Streptococcus thermophilus* can acquire resistance against a virus by integrating a genome fragment of this specific virus into its CRISPR cluster (Barrangou et al. 2007). Therefore, CRISPR is described as the adaptive immune system of prokaryotes which contains a set of spacer sequences that dictate specificity. The co-evolution of viruses and their hosts guarantees a constant potential of adding new spacer sequences (and therefore new resistances) to a growing CRISPR locus. Indeed, some CRISPR loci contain several hundred spacer sequences. However, recombination at repeat sequences is a potential mechanism for reducing the total size of repeat-spacer units which is most often found to be averaged between 20 and 30 units. The spacer sequences are inserted at one specific end of the CRISPR loci that is close to a so-called leader region. This leader region was shown to contain a promoter for the transcription of the CRISPR locus (Pul et al. 2010) and proposed to contain elements that direct spacer integration. Several CRISPR-associated (Cas) genes are often found in the direct vicinity of the leader region and the Cas proteins fulfill many critical roles within this prokaryotic immune system which will be detailed below.

Recent computational studies showcased the divergence of CRISPR/Cas systems in Bacteria and Archaea (Haft et al. 2005; Makarova et al. 2011a, b). CRISPR/Cas types and subtypes differ in length and structure of the repeats but mostly in the set of utilized Cas proteins. CRISPR/Cas systems were classified into three major types and ten subtypes (Makarova et al. 2011a, b). Type I CRISPR/Cas systems are proposed to target only DNA and are distributed in a non-uniform fashion among archaeal and bacterial lineages. Type II CRISPR/Cas systems appear to be found exclusively in Bacteria and utilize a *trans*-encoded guide RNA and host RNase III for crRNA maturation (Deltcheva et al. 2011). Archaea most likely do not possess RNase III-like enzymes (Condon and Putzer 2002) which would explain the observed difference in type II distribution. Type III CRISPR/Cas systems are more commonly found in Archaea and target either RNA or DNA. It is intriguing to see archaeal CRISPR/Cas systems that target RNA even though archaeal RNA viruses have not been described yet. First indications for the existence of archaeal RNA viruses were only recently obtained (Bolduc et al. 2012). It might be that the mRNA presents a target for destruction rather than the DNA genome of an archaeal virus. The analysis of the occurrence of CRISPR clusters within prokaryotic genomes reveals two general trends. First, CRISPR clusters

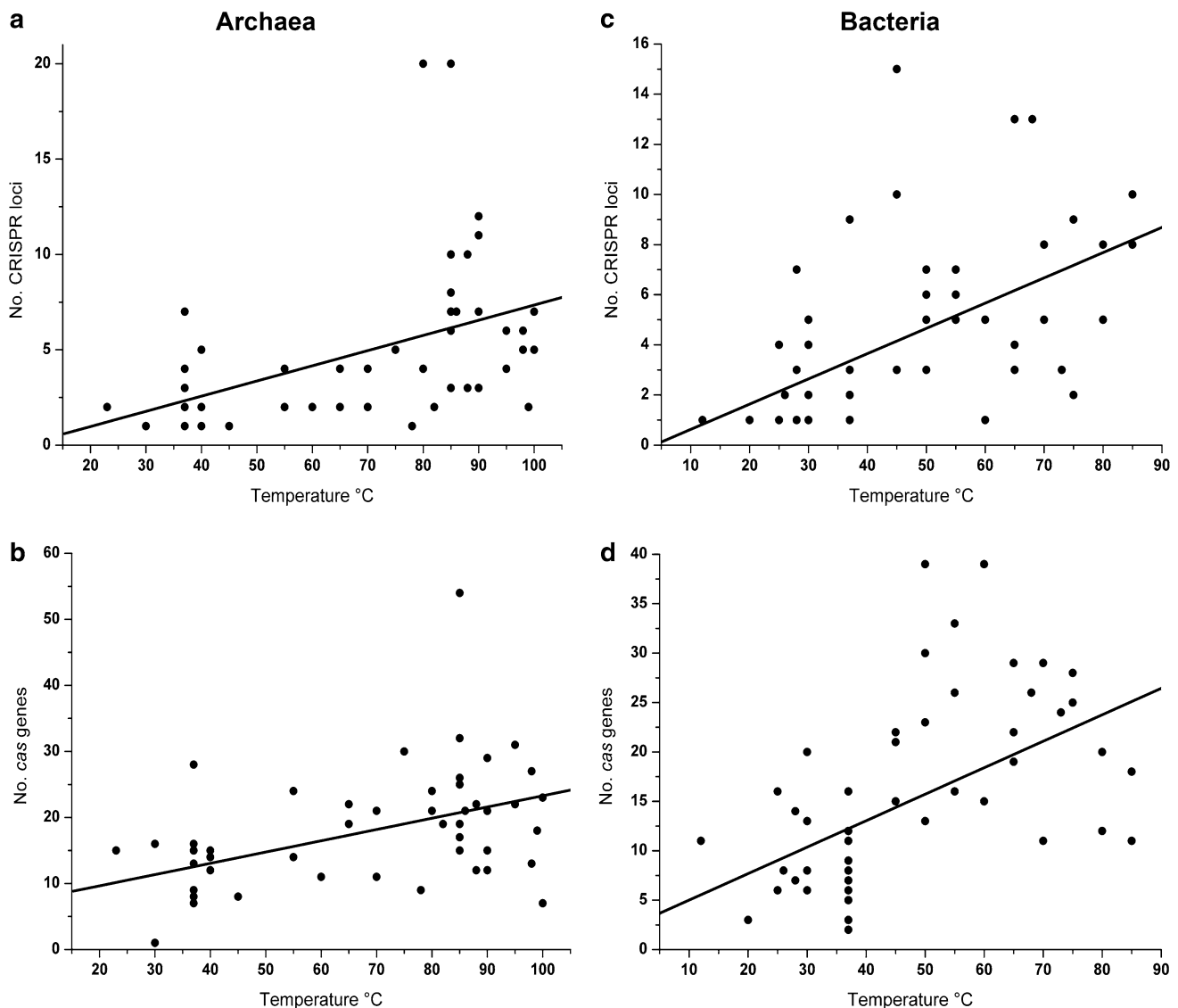
are more often found in archaeal genomes (~90 %) than in bacterial genomes (~40 %) (Grissa et al. 2007) and second, multiple CRISPR clusters are more often occurring in thermophilic organisms than in meso- or psychrophilic strains (Anderson et al. 2011) (Fig. 2). One striking example is the number of CRISPR clusters found within methanogenic archaea. Mesophilic *Methanococcus* strains have no or very few CRISPR clusters, while some hyperthermophilic *Methanocaldococcus* genomes contain over 20 clusters that can make up over 1 % of their total genome (Lillestøl et al. 2006).

### Mechanism of CRISPR-mediated immunity

The common principles of CRISPR/Cas functions, the role of individual Cas proteins and the functional differences among the CRISPR subtypes are studied intensively. In several bacteria, *Pyrococcus furiosus* and *Sulfolobus*, some aspects have already been elucidated in substantial detail while other parts remain mostly obscure. The main features of CRISPR-mediated immunity are likely shared between all subtypes and are commonly classified into three main stages that we will follow with a specific viral sequence (in our example spacerX, indicated in red in Fig. 1).

In the first stage, termed adaptation, the viral DNA is injected into a host cell and protospacerX is recognized and targeted for insertion into the host CRISPR. It is not clear how the acquisition of a new spacer is mediated but, based on gene deletion analyses, the universal proteins Cas1 and Cas2 are proposed to play key roles in this process (Brouns et al. 2008). Short conserved sequences (usually 2 or 3 nucleotides) called protospacer adjacent motifs (PAMs) flank the spacer sequence in the viral genome (termed protospacer) and thus determine the targets of most CRISPR/Cas systems (Mojica et al. 2009). CRISPR clusters start and end with a repeat sequence. Therefore, a new repeat has to be generated via an unknown mechanism at the region flanking the leader sequence and spacerX is inserted between this new repeat and the former first repeat. The matrix for the new repeat generation or potential mechanisms for repeat duplication are still to be identified.

The second stage encompasses the transcription of the CRISPR cluster via a promoter sequence within the leader into a long precursor-crRNA (pre-crRNA) which is subsequently processed by Cas6 or related enzymes into short, regulatory crRNAs (Carte et al. 2008; Haurwitz et al. 2010). Cas6 was shown to mature crRNA into a sequence that contains the complete spacer sequence flanked by repeat fragments that define a specific 8-nucleotide 5'-hydroxyl terminus and 2'-3' cyclic phosphate ends of varying length (Gesner et al. 2011; Jore et al. 2011; Sashital et al. 2011). Thus, one crRNA contains the entire



**Fig. 2** Correlation of archaeal and bacterial growth temperatures with CRISPR/Cas content. Over 50 archaeal (**a**, **b**) and bacterial (**c**, **d**) genomes were analyzed for number of CRISPR and *cas* genes, respectively, and compared to the respective optimal growth

temperatures. Indicated is the best-fit line of raw data. Significance of correlation was confirmed by Spearman's rank correlation test (**a**  $\rho = 0.459$ , **b**  $\rho = 0.602$ , **c**  $\rho = 0.609$ , **d**  $\rho = 0.602$ ; **a-d**  $p < 0.001$ )

spacerX sequence. Type III CRISPR/Cas systems utilize further Cas proteins for the 3'-terminal maturation of crRNAs after primary processing events (Hatoum-Aslan et al. 2011).

The final stage is termed interference and describes the targeted destruction of viral DNA. A repeated attack by the virus that initially provided the spacer sequence can now be interfered with the use of the now present crRNA containing spacerX. The crRNAs are inserted into a multi-subunit complex called Cascade (CRISPR-associated complex for antiviral defence) (Brouns et al. 2008; Wiedenheft et al. 2011). The Cascade complexes use the crRNA to identify the spacerX sequence in invading viral DNA by base-pairing starting at the 5' end of the crRNA.

The absence of base-pairing between the 5' and 3' terminal tags of the crRNA and the viral protospacer and PAM sequence ensure that only viral DNA is cleaved and not host DNA that would otherwise present a target at the CRISPR cluster that generated the crRNA (Marraffini and Sontheimer 2010). The base-pairing between crRNA and protospacer induces conformational changes that might trigger the recruitment of Cas3 in type I CRISPR/Cas systems (Wiedenheft et al. 2011). Cas3 often consists of a helicase domain that mediates the unwinding of RNA/DNA and DNA/DNA duplexes and a HD nuclease domain that finally cleaves the invading single-stranded DNA strand (Beloglazova et al. 2011; Mulepati and Bailey 2011; Sinkunas et al. 2011).

The current general understanding of the impact and activity of CRISPR has been described in detail (Bhaya et al. 2011; Terns and Terns 2011) and we will therefore focus on the description of recent findings for selected archaeal CRISPR/Cas systems.

#### CRISPR/Cas systems of Archaea

##### *Pyrococcus furiosus*

Detailed studies of archaeal CRISPR/Cas systems were conducted for *Pyrococcus furiosus* (Hale et al. 2008, 2009). Here, long CRISPR transcripts are processed by Cas6 into crRNA via a single endonucleolytic cut within the repeat that generates the characteristic 5' terminal 8 nucleotide tags. Cas6 cleavage products are further processed at the 3' end by a currently unknown mechanism that is thought to involve an exonuclease (Carte et al. 2008). The crystal structure of *P. furiosus* Cas6 is available in complex with a repeat RNA and provides insights into pre-crRNA recognition and cleavage (Wang et al. 2011). Cas6 is a metal-independent endoribonuclease that is proposed to cleave pre-crRNAs via a general acid–base chemistry transesterification reaction employing an active-site catalytic triad that is also commonly found in archaeal tRNA splicing endonucleases (Calvin and Li 2008). RNA substrate recognition differs between Cas6 homologs and *P. furiosus* Cas6 was shown to bind to the 5' terminus of unstructured pre-crRNA substrates, while other Cas6-like homologs (Cse3 and Csy4) recognize a stable hairpin that lies immediately upstream of their cleavage sites (Gesner et al. 2011; Sashital et al. 2011; Wang 2012). Structural studies on the Cas6 protein of *Pyrococcus horikoshii* suggest that Cas6 can form dimers in an RNA sequence-dependent manner, which were proposed to benefit the efficiency and specificity of endonuclease activity (Wang et al. 2012).

Other studies on the *P. furiosus* CRISPR focused on the effector complex of CRISPR/Cas subtype III-B. This complex consists of Cas proteins that are often found in a separate module. The common occurrence of a conserved structural motif termed RAMP (repeat associated mysterious protein) within these proteins coined the classification as Cmr (Cas module RAMP) proteins. Isolated *P. furiosus* Cmr module complexes contained six to seven Cas proteins (Cmr 1–1, Cmr 1–2, Cmr 2–6) and mature crRNAs (Hale et al. 2009, 2012). In contrast to the Cascade complex of *E. coli* and probably other type I and II CRISPR/Cas systems, the Cas RAMP module (CMR) complex targets and cleaves RNA instead of DNA (Hale et al. 2009). Cleavage activity depends on divalent cations and is specific for single-stranded RNA substrates that are complementary to the crRNA that is bound to the CMR complex.

The omission of any of the six Cmr proteins but Cmr5 in the effector complex assembly reduced cleavage activity significantly (Hale et al. 2009). Finally, it could be shown that synthetic crRNAs could be designed to direct the targeting and cleavage activity towards an mRNA of choice (Hale et al. 2012). Thus, the Cmr complex can use such crRNAs with a defined 8 nt 5' tag and a guide sequence that is complementary to a target mRNA to mediate the cleavage of this molecule. These experiments suggest that Cmr complexes have the potential to play a role in the regulation of endogenous mRNAs.

##### *Sulfolobus*

Processed archaeal crRNAs were first detected experimentally in *Archaeoglobus fulgidus*, *Sulfolobus solfataricus* and *Sulfolobus acidocaldarius* as part of RNomics approaches to identify small RNAs (Tang et al. 2002, 2005; Lillestøl et al. 2006). Effector complexes of both a DNA-targeting Cascade-like I-A subtype complex and an RNA-targeting CMR subtype III-B complex have been identified in *Sulfolobus solfataricus* (Lintner et al. 2011; Zhang et al. 2012). The archaeal Cascade-like complex (aCascade) was shown to contain Cas7, Cas5, Csa5, the crRNA maturation enzyme Cas6 and crRNA. Other Cas proteins might be associated in vivo. Nevertheless, recombinant Cas7 and Cas5 proteins are sufficient to form a stable complex that binds crRNAs and complementary ssDNA. The production of recombinant Cas7 yields extended right-handed helical assemblies. These assemblies are thought to support and protect the crRNA along their entire length (Lintner et al. 2011). The model is consistent with the available structures for the *E. coli* Cascade complex (Wiedenheft et al. 2011). However, it is unclear if and how the number of aggregated Cas7 subunits is limited in aCascade as the *E. coli* Cascade complex contains a defined hexameric CasC core.

The CMR complex of *S. solfataricus* consists of seven subunits (Cmr 1–7) and a crRNA (Zhang et al. 2012). Cmr7 is an additional subunit that is not present in the analyzed *P. furiosus* CMR complex and homologs are only identifiable in the *Sulfolobales*. It was demonstrated that the CMR complex cleaves the annealed crRNA and target RNA at UA dinucleotides in the presence of manganese ions. This activity was initially proposed to be mediated by the HD nuclease domain of Cmr2 but recent studies indicated that this component is not the catalytic domain (Cocozaki et al. 2012). One notable difference between the cleavage reactions of CMR complexes from *P. furiosus* and *S. solfataricus* is the generation of different termini structures. *P. furiosus* CMR complexes generate 3'-cyclic phosphate while the *S. solfataricus* CMR complex produces 3'-OH and 5'-phosphate ends.



*S. solfataricus* proves also to be an ideal system to study CRISPR/Cas activity in vivo due to the availability of several viruses and tools for its genetic manipulation (Prangishvili and Garrett 2004; Wagner et al. 2009). An in vivo CRISPR/Cas system is available that follows the targeting of DNA in SSV1 transfection assays (Manica et al. 2011). Furthermore, the strains *S. solfataricus* P2 and *S. islandicus* REY15A were used to follow the changes of the CRISPR loci when challenged with viral and plasmid protospacers on a positively selected plasmid (Gudbergsdottir et al. 2011). These studies revealed that transformation success depended on the deletion of (a) CRISPR loci of different sizes, (b) entire CRISPR/Cas modules but also (c) the precise excision of single spacers that matched the protospacer sequence. The mechanism of this excision event remains to be identified.

#### Other archaea

Since there is only limited sequence information available concerning archaeal viruses it is often not possible to identify PAM sequences for archaeal CRISPR/Cas systems by matching spacer sequences to viral sequences. Therefore, plasmid-based invader sequences have been established and successfully been used (Gudbergsdottir et al. 2011; Fischer, Maier, Stoll, Brendel, Fischer, Pfeiffer, Dyall-Smith, Marchfelder, in revision). A systematic approach using a plasmid-based set-up was used to identify PAM sequences for *Haloferax volcanii* (Fischer, Maier, Stoll, Brendel, Fischer, Pfeiffer, Dyall-Smith, Marchfelder, submitted). Here, a spacer from one of the genomic CRISPR loci was cloned into a *Haloferax* plasmid and potential PAM motifs were added adjacent to this spacer sequence. Using this method, six different PAM sequences were identified which are required for target recognition during the defence reaction in *H. volcanii*. This is the first CRISPR group for which more than two PAM sequences have been identified. Using the plasmid-based invader set-up it was shown that the CRISPR/Cas system of *H. volcanii* is targeting DNA. Cells which could escape the defence reaction did so by deleting the complete *cas* gene cassette or the protospacer on the plasmid or the corresponding spacer in the chromosomal CRISPR locus (Fischer, Maier, Stoll, Brendel, Fischer, Pfeiffer, Dyall-Smith, Marchfelder, in revision).

Analysis of the CRISPR/Cas I-A system of the crenarchaeon *Thermoproteus tenax* revealed two discrete polycistronic transcripts that contained (a) the genes for the aCascade unit (*csa5*, *cas7*, *cas5a*, *cas3*, *cas3'*, *cas8a2*) and (b) a proposed Cascis unit (CRISPR-associated complex for the integration of spacers) (*cas4*, *cas1/2*, *csa1*). The transcription of the *cascade* genes is modulated by changes of environmental parameters, like UV-light exposure or

high ionic strength, indicating a tight regulation of *cas* genes in vivo. The two multi-subunit complexes aCascade and Cascis could be generated by the reconstitution of all individually insoluble recombinant proteins encoded within the operons, which suggests interaction and coordination of these Cas proteins (Plagens et al. 2012).

#### Archaeal viruses

CRISPR/Cas systems are more often found in Archaea than in Bacteria and extremophilic archaea contain the highest numbers of CRISPR clusters and available spacer sequences (Anderson et al. 2011) (Fig. 1). This is, however, in stark contrast to the very limited available amount of information for potential viruses that could be targeted by the spacer sequences within the crRNAs of extremophilic archaea. The correlation of spacer sequence and matching viral targets is most successful for crenarchaeal acidothermophiles where up to 30 % of the spacer sequences could be assigned to viral or plasmid genomes (Shah et al. 2009). In most other cases, spacer sequences can be used to estimate the number of archaeal viruses and conjugative plasmids that have not been cultured and/or sequenced. With one exception for a single-stranded DNA virus (Pietila et al. 2009), isolated archaeal viruses contain a dsDNA genome and exhibit a fascinating range of virion morphotypes (Prangishvili et al. 2006). Archaeal viruses lyse their host cells less often than bacterial viruses (Bize et al. 2009) and hyperthermophilic viruses tend to integrate into the host genome.

The short spacer sequences directly provide a tiny fragment of a viral genome from a previous infection that can be used to hunt for unknown archaeal viruses in the environment (Snyder et al. 2010). CRISPR spacer sequences were extracted from metagenomic data obtained from acidic hot spring environments in Yellowstone National Park. Viruses were detectable by hybridization of virus-enriched environmental samples to a microarray platform that contained spacer sequence probes (Snyder et al. 2010). The advent of highly sensitive deep-sequencing technologies should enable the sequencing of viral genomes that have been captured by the short CRISPR spacer sequences.

#### CRISPR targets: DNA or RNA?

What are the consequences for the CRISPR/Cas system that targets either DNA (e.g. Cascade and Cas3) or RNA (e.g. CMR complex)? The targeting of DNA allows for the usage of spacers that derive from either the sense or the antisense strand of a viral genome. Targeting RNA of mobile genetic elements would limit the choice of protospacers to sequences that are complementary to the given

mRNA. On the other hand, DNA targeting poses the problem of discriminating viral DNA from host DNA as any genomic CRISPR sequence would provide a crRNA target. This discrimination is achieved by the use of PAM sequences as described above. PAM sequences were not observed to be required in the targeting of the CMR complex. Therefore, this mechanism might have evolved to free the CRISPR/Cas system from the requirement of certain protospacer flanking sequences and would provide an antiviral defense system that can complement DNA-targeting CRISPR/Cas subtypes.

## Conclusion

The described identification of diverse small non-coding RNA families in archaea showcases the growing realization of the importance of these RNA species in this third domain of life. Modern high-throughput sequencing methods will reveal the RNome composition of further archaeal species. This information can then be combined with necessary functional studies on individual sRNA and crRNA molecules together with the identification of their target sites and should help us to validate the full scope of their influence on cellular processes.

**Open Access** This article is distributed under the terms of the Creative Commons Attribution License which permits any use, distribution, and reproduction in any medium, provided the original author(s) and the source are credited.

## References

- Anderson RE, Brazelton WJ, Baross JA (2011) Using CRISPRs as a metagenomic tool to identify microbial hosts of a diffuse flow hydrothermal vent viral assemblage. *FEMS Microbiol Ecol* 77(1):120–133
- Argaman L, Hershberg R, Vogel J, Bejerano G, Wagner EG, Margalit H, Altuvia S (2001) Novel small RNA-encoding genes in the intergenic regions of *Escherichia coli*. *Curr Biol* 11(12):941–950
- Arini A, Keller MP, Arber W (1997) An antisense RNA in IS30 regulates the translational expression of the transposase. *Biol Chem* 378(12):1421–1431
- 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 *Haloflex volcanii*. *RNA Biol* 8:5
- Barrangou R, Fremaux C, Deveau H, Richards M, Boyaval P, Moineau S, Romero DA, Horvath P (2007) CRISPR provides acquired resistance against viruses in prokaryotes. *Science* 315(5819):1709–1712
- Beloglazova N, Petit P, Flick R, Brown G, Savchenko A, Yakunin AF (2011) Structure and activity of the Cas3 HD nuclease MJ0384, an effector enzyme of the CRISPR interference. *EMBO J* 30(22):4616–4627
- Bhaya D, Davison M, Barrangou R (2011) CRISPR-Cas systems in bacteria and archaea: versatile small RNAs for adaptive defense and regulation. *Annu Rev Genet* 45:273–297
- Bize A, Karlsson EA, Ekefjard K, Quax TE, Pina M, Prevost MC, Forterre P, Tenaillon O, Bernander R, Prangishvili D (2009) A unique virus release mechanism in the Archaea. *Proc Natl Acad Sci USA* 106(27):11306–11311
- Bolduc B, Shaughnessy DP, Wolf YI, Koonin E, Roberto FF, Young M (2012) Identification of novel positive-strand RNA viruses by metagenomic analysis of archaea-dominated Yellowstone Hot Springs. *J Virol* 86(10):5562–5573
- Brantl S (2009) Bacterial chromosome-encoded small regulatory RNAs. *Future Microbiol* 4(1):85–103
- Brantl S (2012) Small regulatory RNAs (sRNAs): key players in prokaryotic metabolism, stress response, and virulence. In: Mallick B (ed) *Regulatory RNAs*. Springer, Berlin, pp 73–109
- Brouns SJ, Jore MM, Lundgren M, Westra ER, Slijkhuys RJ, Snijders AP, Dickman MJ, Makarova KS, Koonin EV, van der Oost J (2008) Small CRISPR RNAs guide antiviral defense in prokaryotes. *Science* 321(5891):960–964
- Busch A, Richter AS, Backofen R (2008) IntaRNA: efficient prediction of bacterial sRNA targets incorporating target site accessibility and seed regions. *Bioinformatics* 24(24):2849–2856
- Calvin K, Li H (2008) RNA-splicing endonuclease structure and function. *Cell Mol Life Sci* 65(7–8):1176–1185
- Carte J, Wang R, Li H, Terns RM, Terns MP (2008) Cas6 is an endoribonuclease that generates guide RNAs for invader defense in prokaryotes. *Genes Dev* 22(24):3489–3496
- Cocozaki AI, Ramia NF, Shao Y, Hale CR, Terns RM, Terns MP, Li H (2012) Structure of the Cmr2 subunit of the CRISPR-Cas RNA silencing complex. *Structure* 20(3):545–553
- Cole C, Sobala A, Lu C, Thatcher SR, Bowman A, Brown JW, Green PJ, Barton GJ, Hutvagner G (2009) Filtering of deep sequencing data reveals the existence of abundant Dicer-dependent small RNAs derived from tRNAs. *RNA* 15(12):2147–2160 (Epub 2009 Oct 22)
- Condon C, Putzer H (2002) The phylogenetic distribution of bacterial ribonucleases. *Nucleic Acids Res* 30(24):5339–5346
- Deltcheva E, Chylinski K, Sharma CM, Gonzales K, Chao Y, Pirzada ZA, Eckert MR, Vogel J, Charpentier E (2011) CRISPR RNA maturation by trans-encoded small RNA and host factor RNase III. *Nature* 471(7340):602–607
- Fischer S, Benz J, Späth B, Maier L-K, 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(45):34429–34438
- Fischer S, Benz J, Späth B, Jellen-Ritter A, Heyer R, Dörr M, Maier LK, Menzel-Hobeck C, Lehr M, Jantzer K, Babski J, Soppa J, Marchfelder A (2011) Regulatory RNAs in *Haloflex volcanii*. *Biochem Soc Trans* 39(1):159–162
- Georg J, Hess WR (2011) Natural antisense transcripts in bacteria. In: Hess WR, Marchfelder A (eds) *Regulatory RNAs in prokaryotes*. Springer, Wien, pp 95–108
- Gesner EM, Schellenberg MJ, Garside EL, George MM, Macmillan AM (2011) Recognition and maturation of effector RNAs in a CRISPR interference pathway. *Nat Struct Mol Biol* 18(6):688–692
- Gildehaus N, Neusser T, Wurm R, Wagner R (2007) Studies on the function of the riboregulator 6S RNA from *E. coli*: rRNA polymerase binding, inhibition of in vitro transcription and synthesis of RNA-directed de novo transcripts. *Nucleic Acids Res* 35(6):1885–1896 (Epub 2007 Mar 1881)
- Gottesman S (2004a) The small RNA regulators of *Escherichia coli*: roles and mechanisms. *Annu Rev Microbiol* 58:303–328
- Gottesman S (2004b) Small RNAs shed some light. *Cell* 118(1):1–2
- Grissa I, Vergnaud G, Pourcel C (2007) The CRISPRdb database and tools to display CRISPRs and to generate dictionaries of spacers and repeats. *BMC Bioinformatics* 8:172
- Gudbergsdottir S, Deng L, Chen Z, Jensen JV, Jensen LR, She Q, Garrett RA (2011) Dynamic properties of the Sulfolobus

- CRISPR/Cas and CRISPR/Cmr systems when challenged with vector-borne viral and plasmid genes and protospacers. *Mol Microbiol* 79(1):35–49
- Guo H, Ingolia NT, Weissman JS, Bartel DP (2010) Mammalian microRNAs predominantly act to decrease target mRNA levels. *Nature* 466(7308):835–840
- Haft DH, Selengut J, Mongodin EF, Nelson KE (2005) A guild of 45 CRISPR-associated (Cas) protein families and multiple CRISPR/Cas subtypes exist in prokaryotic genomes. *PLoS Comput Biol* 1(6):e60
- Hale C, Kleppe K, Terns RM, Terns MP (2008) Prokaryotic silencing (psi)RNAs in *Pyrococcus furiosus*. *RNA* 14(12):2572–2579
- Hale CR, Zhao P, Olson S, Duff MO, Graveley BR, Wells L, Terns RM, Terns MP (2009) RNA-guided RNA cleavage by a CRISPR RNA-Cas protein complex. *Cell* 139(5):945–956
- Hale CR, Majumdar S, Elmore J, Pfister N, Compton M, Olson S, Resch AM, Glover CV III, Graveley BR, Terns RM, Terns MP (2012) Essential features and rational design of CRISPR RNAs that function with the Cas RAMP module complex to cleave RNAs. *Mol Cell* 45:292–302
- Hartmann RK, Gossringer M, Spath B, Fischer S, Marchfelder A (2009) The making of tRNAs and more—RNase P and tRNase Z. *Prog Mol Biol Transl Sci* 85:319–368
- Hatoum-Aslan A, Maniv I, Marraffini LA (2011) Mature clustered, regularly interspaced, short palindromic repeats RNA (crRNA) length is measured by a ruler mechanism anchored at the precursor processing site. *Proc Natl Acad Sci USA* 108(52):21218–21222
- Haurwitz RE, Jinek M, Wiedenheft B, Zhou K, Doudna JA (2010) Sequence- and structure-specific RNA processing by a CRISPR endonuclease. *Science* 329(5997):1355–1358
- Haussecker D, Huang Y, Lau A, Parameswaran P, Fire AZ, Kay MA (2010) Human tRNA-derived small RNAs in the global regulation of RNA silencing. *RNA* 16(4):673–695
- Heyer R, Dörr M, Jellen-Ritter A, Späth B, Babski J, Jaschinski K, Soppa J, Marchfelder A (2012) High throughput sequencing reveals a plethora of small RNAs including tRNA derived fragments in *Haloflex volcanii*. *RNA Biol* (in press)
- Hüttenhofer A, Schattner P, Polacek N (2005) Non-coding RNAs: hope or hype? *Trends Genet* 21(5):289–297
- Ishino Y, Shinagawa H, Makino K, Amemura M, Nakata A (1987) Nucleotide sequence of the *iap* gene, responsible for alkaline phosphatase isozyme conversion in *Escherichia coli*, and identification of the gene product. *J Bacteriol* 169(12):5429–5433
- 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 USA* 106(51):21878–21882
- Jansen R, Embden JD, Gaastra W, Schouls LM (2002) Identification of genes that are associated with DNA repeats in prokaryotes. *Mol Microbiol* 43(6):1565–1575
- Jinek M, Doudna JA (2009) A three-dimensional view of the molecular machinery of RNA interference. *Nature* 457(7228):405–412
- Jore MM, Lundgren M, van Duijn E, Bultema JB, Westra ER, Waghmare SP, Wiedenheft B, Pul U, Wurm R, Wagner R, Beijer MR, Barendregt A, Zhou K, Snijders AP, Dickman MJ, Doudna JA, Boekema EJ, Heck AJ, van der Oost J, Brouns SJ (2011) Structural basis for CRISPR RNA-guided DNA recognition by Cascade. *Nat Struct Mol Biol* 18(5):529–536
- Klein RJ, Misulovin Z, Eddy SR (2002) Noncoding RNA genes identified in AT-rich hyperthermophiles. *Proc Natl Acad Sci USA* 99(11):7542–7547
- Krol J, Loedige I, Filipowicz W (2010) The widespread regulation of microRNA biogenesis, function and decay. *Nat Rev Genet* 11(9):597–610 (Epub 2010 Jul 27)
- Kuhn JF, Tran EJ, Maxwell ES (2002) Archaeal ribosomal protein L7 is a functional homolog of the eukaryotic 15.5kD/Snu13p snoRNP core protein. *Nucleic Acids Res* 30(4):931–941
- Lee YS, Shibata Y, Malhotra A, Dutta A (2009) A novel class of small RNAs: tRNA-derived RNA fragments (tRFs). *Genes Dev* 23(22):2639–2649
- Lewis BP, Burge CB, Bartel DP (2005) Conserved seed pairing, often flanked by adenosines, indicates that thousands of human genes are microRNA targets. *Cell* 120(1):15–20
- Lillestøl RK, Redder P, Garrett RA, Brügger K (2006) A putative viral defence mechanism in archaeal cells. *Archaea* 2(1):59–72
- Lintner NG, Kerou M, Brumfield SK, Graham S, Liu H, Naismith JH, Sdano M, Peng N, She Q, Copie V, Young MJ, White MF, Lawrence CM (2011) Structural and functional characterization of an archaeal clustered regularly interspaced short palindromic repeat (CRISPR)-associated complex for antiviral defense (CASCADE). *J Biol Chem* 286(24):21643–21656
- Ma C, Simons RW (1990) The *IS10* antisense RNA blocks ribosome binding at the transposase translation initiation site. *EMBO J* 9(4):1267–1274
- Makarova KS, Aravind L, Wolf YI, Koonin EV (2011a) Unification of Cas protein families and a simple scenario for the origin and evolution of CRISPR-Cas systems. *Biol Direct* 6:38
- 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 (2011b) Evolution and classification of the CRISPR-Cas systems. *Nat Rev Microbiol* 9(6):467–477
- Manica A, Zebec Z, Teichmann D, Schleper C (2011) In vivo activity of CRISPR-mediated virus defence in a hyperthermophilic archaeon. *Mol Microbiol* 80(2):481–491
- Marraffini LA, Sontheimer EJ (2010) Self versus non-self discrimination during CRISPR RNA-directed immunity. *Nature* 463(7280):568–571
- Meister G (2007) miRNAs get an early start on translational silencing. *Cell* 131(1):25–28
- Mojica FJ, Diez-Villasenor C, Garcia-Martinez J, Soria E (2005) Intervening sequences of regularly spaced prokaryotic repeats derive from foreign genetic elements. *J Mol Evol* 60(2):174–182
- Mojica FJ, Diez-Villasenor C, Garcia-Martinez J, Almendros C (2009) Short motif sequences determine the targets of the prokaryotic CRISPR defence system. *Microbiology* 155(Pt 3):733–740
- Mulepati S, Bailey S (2011) Structural and biochemical analysis of nuclease domain of clustered regularly interspaced short palindromic repeat (CRISPR)-associated protein 3 (Cas3). *J Biol Chem* 286(36):31896–31903
- Pietila MK, Roine E, Paulin L, Kalkkinen N, Bamford DH (2009) An ssDNA virus infecting archaea: a new lineage of viruses with a membrane envelope. *Mol Microbiol* 72(2):307–319
- Plagens A, Tjaden B, Hagemann A, Randau L, Hensel R (2012) Characterization of the CRISPR/Cas I-A system of the hyperthermophilic Crenarchaeon *Thermoproteus tenax*. *J Bacteriol* 194(10):2491–2500
- Prangishvili D, Garrett RA (2004) Exceptionally diverse morphotypes and genomes of crenarchaeal hyperthermophilic viruses. *Biochem Soc Trans* 32(Pt 2):204–208
- Prangishvili D, Forterre P, Garrett RA (2006) Viruses of the Archaea: a unifying view. *Nat Rev Microbiol* 4(11):837–848
- Pul U, Wurm R, Arslan Z, Geissen R, Hofmann N, Wagner R (2010) Identification and characterization of *E. coli* CRISPR-cas promoters and their silencing by H-NS. *Mol Microbiol* 75(6):1495–1512
- Rasmussen S, Nielsen HB, Jarmer H (2009) The transcriptionally active regions in the genome of *Bacillus subtilis*. *Mol Microbiol* 73(6):1043–1057 (Epub 2009 Aug 4)

- Rivas E, Eddy SR (2001) Noncoding RNA gene detection using comparative sequence analysis. *BMC Bioinformatics* 2:8 (Epub 2001 Oct 2010)
- Rozhdestvensky TS, Tang TH, Tchirkova IV, Brosius J, Bachellerie JP, Huttenhofer A (2003) Binding of L7Ae protein to the K-turn of archaeal snoRNAs: a shared RNA binding motif for C/D and H/ACA box snoRNAs in Archaea. *Nucleic Acids Res* 31(3): 869–877
- Sashital DG, Jinek M, Doudna JA (2011) An RNA-induced conformational change required for CRISPR RNA cleavage by the endoribonuclease Cse3. *Nat Struct Mol Biol* 18(6):680–687
- Schattner P (2002) Searching for RNA genes using base-composition statistics. *Nucleic Acids Res* 30(9):2076–2082
- Shah SA, Hansen NR, Garrett RA (2009) Distribution of CRISPR spacer matches in viruses and plasmids of crenarchaeal acidothermophiles and implications for their inhibitory mechanism. *Biochem Soc Trans* 37(Pt 1):23–28
- Sinkunas T, Gasiunas G, Fremaux C, Barrangou R, Horvath P, Siksnys V (2011) Cas3 is a single-stranded DNA nuclease and ATP-dependent helicase in the CRISPR/Cas immune system. *EMBO J* 30(7):1335–1342
- Snyder JC, Bateson MM, Lavin M, Young MJ (2010) Use of cellular CRISPR (clusters of regularly interspaced short palindromic repeats) spacer-based microarrays for detection of viruses in environmental samples. *Appl Environ Microbiol* 76(21):7251–7258
- Sobala A, Hutvagner G (2011) Transfer RNA-derived fragments: origins, processing, and functions. *Wiley Interdiscip Rev RNA* 2(6):853–862
- Soppa J, Straub J, Brenneis M, Jellen-Ritter A, Heyer R, Fischer S, Granzow M, Voss B, Hess WR, Tjaden B, Marchfelder A (2009) Small RNAs of the halophilic archaeon *Haloferax volcanii*. *Biochem Soc Trans* 37(Pt 1):133–136
- 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* 6:281–292
- Tang TH, Bachellerie 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 USA* 99(11):7536–7541
- Tang TH, Polacek N, Zywicki M, Huber H, Brugger K, Garrett R, Bachellerie JP, Huttenhofer A (2005) Identification of novel non-coding RNAs as potential antisense regulators in the archaeon *Sulfolobus solfataricus*. *Mol Microbiol* 55(2):469–481
- Terns MP, Terns RM (2011) CRISPR-based adaptive immune systems. *Curr Opin Microbiol* 14(3):321–327
- Thomason MK, Storz G (2010) Bacterial antisense RNAs: how many are there, and what are they doing? *Annu Rev Microbiol* 44:167–188
- Wagner M, Berkner S, Ajon M, Driessen AJ, Lipps G, Albers SV (2009) Expanding and understanding the genetic toolbox of the hyperthermophilic genus *Sulfolobus*. *Biochem Soc Trans* 37(Pt 1):97–101
- Wang R, Preamplume G, Terns MP, Terns RM, Li H (2011) Interaction of the Cas6 ribonuclease with CRISPR RNAs: recognition and cleavage. *Structure* 19(2):257–264
- Wang R, Zheng H, Preamplume G, Shao Y, Li H (2012) The impact of CRISPR repeat sequence on structures of a Cas6 protein-RNA complex. *Protein Sci* 21(3):405–417
- Wassarman KM, Repoila F, Rosenow C, Storz G, Gottesman S (2001) Identification of novel small RNAs using comparative genomics and microarrays. *Genes Dev* 15(13):1637–1651
- Waters LS, Storz G (2009) Regulatory RNAs in bacteria. *Cell* 136(4):615–628
- Wiedenheft B, Lander GC, Zhou K, Jore MM, Brouns SJ, van der Oost J, Doudna JA, Nogales E (2011) Structures of the RNA-guided surveillance complex from a bacterial immune system. *Nature* 477(7365):486–489
- Wurtzel O, Sapra R, Chen F, Zhu Y, Simmons BA, Sorek R (2009) A single-base resolution map of an archaeal transcriptome. *Genome Res* 20(1):133–141
- Zago MA, Dennis PP, Omer AD (2005) The expanding world of small RNAs in the hyperthermophilic archaeon *Sulfolobus solfataricus*. *Mol Microbiol* 55(6):1812–1828
- Zhang J, Rouillon C, Kerou M, Reeks J, Brugger K, Graham S, Reimann J, Cannone G, Liu H, Albers SV, Naismith JH, Spagnolo L, White MF (2012) Structure and mechanism of the CMR complex for CRISPR-mediated antiviral immunity. *Mol Cell* 45(3):303–313