Peer-reviewed version available at *Genes* **2018**, *9*, 141; doi:10.3390/genes9030141

The non-coding regulatory RNA revolution in Archaea

Diego R. Gelsinger[a] and Jocelyne DiRuggiero[a]#

Department of Biology, The Johns Hopkins University, Baltimore, Maryland, USA[a]

#Address correspondence to Jocelyne DiRuggiero, jdiruggiero@jhu.edu

Abstract

Small non-coding RNAs (sRNAs) are ubiquitously found in the three domains of life playing large-scale roles in gene regulation, transposable element silencing, and defense against foreign elements. While a substantial body of experimental work has been done to uncover function of sRNAs in Bacteria and Eukarya, the functional roles of sRNAs in Archaea are still poorly understood. Recently, high throughput studies using RNA-sequencing revealed that sRNAs are broadly expressed in the Archaea, comprising thousands of transcripts within the transcriptome during non-challenged and stressed conditions. Antisense sRNAs, which overlap a portion of a gene on the opposite strand (*cis*-acting), are the most abundantly expressed non-coding RNAs and they can be classified based on their binding patterns to mRNAs (3' UTR, 5' UTR, CDS-binding). These antisense sRNAs target many genes and pathways, suggesting extensive roles in gene regulation. Intergenic sRNAs are less abundantly expressed and their targets are difficult to find because of a lack of complete overlap between sRNAs and target mRNAs (*trans*-acting). While many sRNAs have been validated experimentally, a regulatory role has only been reported for very few of them. Further work is needed to elucidate sRNA-RNA binding mechanisms, the molecular determinants of sRNA-mediated regulation, whether protein components are involved, and how sRNAs integrate with complex regulatory networks.

Introduction

Small RNAs (sRNAs) are important regulators for multiple cellular functions and they are ubiquitous in all domains of life. sRNAs, also called non-coding RNAs (ncRNAs), are RNAs that do not function as messenger RNAs (mRNAs), ribosomal RNAs, or transfer RNAs in the cell. sRNAs in bacteria and eukarya play essential roles in transcriptional regulation, chromosome replication, RNA processing and modification, mRNA stability and translation, and even protein degradation and translocation [3-5]. Recently, it was discovered that archaeal genomes encode for large numbers of sRNAs and that many of them are responsive to environmental stresses [1, 2, 6-14].

While much remains to be elucidated about sRNAs in *Archaea*, decades of research in Eukarya and Bacteria have built a body of knowledge on their functional roles and their mechanisms of action. In *Eukarya,* several types of sRNAs have been identified, including microRNAs (miRNAs), PIWI-associated RNAs (piRNAs), and endogenous small interfering RNAs (siRNAs) [3]. The most studied of these, miRNAs, regulate protein expression in key cellular processes. miRNAs are typically 20-30 nucleotides (nt) long, target the 3' end of their cognate mRNAs, and form complexes with Argonaute (Ago) proteins [15]. While Ago homologs have also been found in archaeal genomes, there is no evidence for eukaryotic-like RNA interference in these organisms [16]. Rather, a defensive role against foreign genetic material was recently proposed whereby archaeal Ago proteins direct guide-dependent cleavage of foreign DNA [17].

In bacteria, sRNAs are typically 50 to 300 nt in length and act by targeting mRNA stability, translation, or by binding to proteins [18]. Base-pairing with their mRNAs targets are of two types. *Cis*-encoded antisense RNAs (asRNAs) are encoded on the DNA strand opposite their

target gene and thus can act via extensive base pairing (**Fig. 1A**). asRNAs have been found to repress transposons and toxic protein synthesis and to modulate the expression of transcriptional regulators [4, 18]. In contrast, *trans*-encoded sRNAs are encoded at genomic location distinct from their target mRNAs, such as intergenic regions, and act via limited base pairing [19] (**Fig. 1B**). These sRNAs bind at the 5' end or 3' end of their target, either blocking ribosome binding and/or triggering degradation of target mRNAs via the endoribonuclease RNaseE [18]. sRNAs can also activate translation by preventing the formation of inhibitory secondary structures and therefore increasing ribosome binding [4]. *Trans*-encoded sRNAs can target multiple genes, including key transcription factors and regulators and, as a consequence, a single sRNA can modulate the expression of very large regulons [15, 18]. A typical example of that is the sRNA OxyS involved in the oxidative stress response in *E. coli* [20]. In most gram-negative bacteria, the RNA binding protein Hfq is required for function and/or stability of the sRNAs [4]. Hfq is structurally related to the Sm/Lsm family of proteins and acts by stabilizing RNAs or by promoting rapid mRNA-sRNA base-pairing and recruiting of RNAseE for degradation [4]. However, other bacteria do not require Hfq for sRNA-mediated regulation even when the protein is encoded in their genome. Recently, novel RNA-binding proteins such as CsrA and ProQ have also been proposed to function as sRNAs chaperones in bacteria [21].

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 25 January 2018 doi:10.20944/preprints201801.0232.v1

Peer-reviewed version available at Genes 2018, 9, 141; doi:10.3390/genes9030141

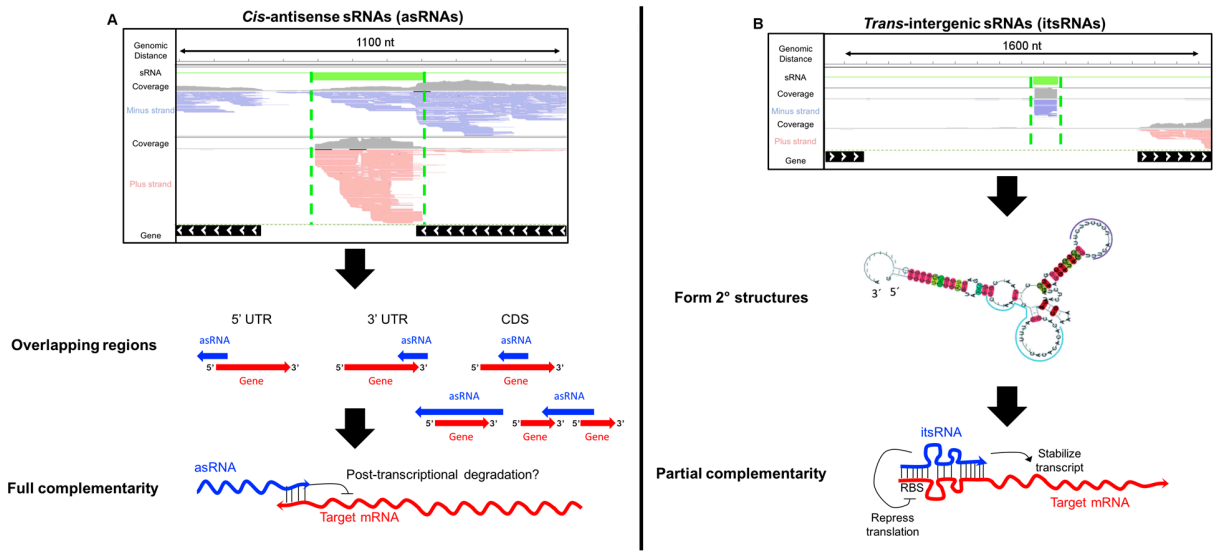

Figure 1: Classes of sRNAs discovered in *Archaea*. Genome viewer of (A) antisense sRNAs (*cis*-acting) and (B) intergenic sRNAs (*trans*-acting). Paired-end reads (100 bases) were mapped to the *H. volcanii* NCBI reference genome. Reference genes are marked as black lines with white arrows indicating their location on the plus strand (>) or minus strand (<). Reads marked in red are transcribed from the minus strand while blue reads are transcribed from the plus strand. Green lines indicated sRNAs. Coverage plots are in gray. Diagrams of classes of asRNAs are shown based on binding attributes: overlapping the 3' UTR, the 5' UTR, within the coding sequence (CDS), extending past the CDS, and overlapping multiple mRNAs. An example of an intergenic sRNA secondary structure is shown in (B) [2]. Reported regulatory roles of archaeal sRNAs are shown at the bottom.

In archaea, the functional and mechanistic characterization of sRNAs is still in its infancy. The size range of archaeal sRNAs is 50 to 500 nt in length. Both *cis* asRNAs and *trans*-encoded sRNAs (thereafter called itsRNAs for intergenic sRNAs) have been reported in a number of archaeal species (**Fig. 1**), as well as *cis* sense sRNAs that are transcribed within the open reading frame (ORF) of genes [1, 14, 22-27]. In addition to these ncRNAs, small

Peer-reviewed version available at Genes 2018, 9, 141; doi:10.3390/genes9030141

nucleolar RNAs (snoRNAs), tRNA-derived fragments, and CRISPR RNAs (crRNAs) have been found in Archaea. This review will focus on the *cis*- and *trans*-encoded archaeal sRNAs as it becomes more evident that they play essential roles in gene regulation and because there have been exciting new developments in the last few years (since the last sRNA review) in unraveling the functional roles of these sRNAs. We will first document the sRNAs identified so far in the Archaea and discuss the state-of-the art methods for sRNA detection. We will then describe the molecular mechanisms that have been elucidated for a small number of archaeal sRNAs, give an overview of sRNA-interacting partners, and as such provide insights into the biological roles of these sRNAs. Lastly we will discuss the future prospects for archaeal sRNA investigations.

Identification of sRNAs: what has been found so far?

In contrast to the wealth of knowledge on bacterial and eukaryal RNA regulators, our knowledge of sRNAs in Archaea is limited to a handful of studies for hyperthermophiles, methanogens, and the haloarchaeon *Haloferax volcanii* [9, 11, 28]. Both classes of trans- and cis-encoded regulatory sRNAs have been found in the Archaea. Initial identification of sRNAs relied on bioinformatic (RNomics) prediction using archaeal whole genome sequences. Later, microarray and 454-pyrosequencing technologies provided a mean to validate these predicted sRNAs and further identified novel sRNAs by the hundreds. However, it is the unprecedented discovery of more than 2,900 sRNAs in *H. volcanii* by two recent high-throughput sequencing (HTS) studies - which is quite remarkable considering that the genome of this organism encode for just over 4,000 proteins [1, 14] - that has permanently altered our view of the archaeal transcriptome. From these studies it is now clear that a large proportion of RNAs are non-coding, nearly rivaling the number of RNAs encoding for proteins [1, 14]. Additionally, non-coding RNAs up

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 25 January 2018 doi:10.20944/preprints201801.0232.v1

Peer-reviewed version available at Genes 2018, 9, 141; doi:10.3390/genes9030141

to 1,000 nt in length were also reported [1], thus increasing the size range of identified archaeal sRNAs. In the model haloarchaeon, *H. volcanii*, as many as 1,500, asRNAs and 400 intergenic sRNAs have been identified, indicating that most sRNAs in this organism are antisense to coding regions. Furthermore, as much as 30% of the sRNAs discovered in *H. volcanii* contained stringent basal transcriptional promoters, such as a TATA-box, and exhibited expression levels comparable to mRNAs, underling their relevance in the global regulation of gene networks [1, 14].

Table 1: Summary of sRNA discovered in the *Archaea*

	Number of Genes	Total Number of sRNAs	Number of itsRNAs	Number of asRNAs	Number of iRNAs	Reference
Euryarchaeota						
Haloferax volcanii	4023	1557	77	1480	N/A	Gelsinger, *et al.* (in review)
Haloferax volcanii	4023	2792	395	1244	1153	Babski, *et al.* (2016)
Haloferax volcanii	4023	190	145	45	N/A	Heyer, *et al.* (2012)
Methanolobus psychrophilus	2974	2745	195	1110	1440	Li, *et al.* (2015)
Methanosarcina mazei	3551	242	199	43	N/A	Jäger, *et al.* (2009)
Thermococcus kodakarensis	2328	1731	69	1018	644	Jäger, *et al.* (2014)
Pyrococcus abyssi	1969	322	107	215	N/A	Toffano-Nioche, *et al.* (2013)
Archaeoglobus fulgidus	248	45	9	33	3	Tang, *et al.* (2002)
Crenarchaeaota						
Sulfolobus solfataricus	3254	310	125	185	N/A	Wurtzel, *et al.* (2010)
Pyrobaculum aerophilum, arsenaticum, calidifontis, & islandicum	2706, 2407, 2200, 2075	Number Not Reported	Number Not Reported	3	N/A	Lowe, *et al.* (2012)
Nanoarchaeaota						
Nanoarchaeum equitans	553	Number Not Reported	Number Not Reported	Number Not Reported	N/A	Randau. (2012)
asRNAs: antisense sRNAs; itsRNAs: intergenic sRNAs; iRNAs: internally transcribed sRNAs; N/A: these types of sRNAs were not reported.						

While sRNAs are particularly numerous in haloarchaea genomes [14], they have also been found in a number of other archaea, including *Sulfolobus* [22], *Methanosarcina* [23],

Pyrobaculum [24], *Pyrococcus* [29], *Thermococcus* [30], and *Methanolobus* [31] (**Table 1**). In *Sulfolobus solfataricus*, 125 trans-encoded sRNAs and 185 cis-encoded asRNAs were identified using HTS, suggesting that 6.1% of all genes in *S. solfataricus* are associated with sRNAs [22]. A comparative genome analysis of *Methanosarcina mazei*, *M. bakeri*, and *M. acetivorans* revealed that 30% of the antisense and 21% of the intergenic sRNAs identified were conserved across the 3 species [23]. Similarly to bacteria, the number of antisense sRNAs reported in the archaea numbers in the 100s and further work is greatly needed to validate and characterize their functional roles [19].

Best methods for sRNA discovery

Methods at the forefront of sRNA discovery in Archaea are all RNA sequencing methods that take advantage of the sequencing depth and high throughput of Illumina technologies. These methods are (1) differential RNA-sequencing (dRNA-seq) and (2) size-selected, strand-specific sRNA-sequencing (sRNA-seq) [14, 32]. Both methods preserve the strand-specificity that has been key to the identification of antisense sRNAs in *Archaea*.

Differential RNA-seq was used to identify hundreds to thousands of sRNAs in *H. volcanii*, *M. psychrophilus*, *T. kodakerensis*, and *P. abyssi* [14, 29-31]. The dRNA-seq method is based on the selective enrichment of primary transcripts, which in prokaryotes carry a triphosphate at their 5' ends. In contrast, processed transcripts characteristically carry a monophosphate at their 5′ ends. To construct dRNA-seq libraries, the RNA pool is aliquoted into two fractions; one fraction is treated with a terminator exonuclease (TEX) that specifically degrades 5′ monophosphate RNAs, thus selectively enriching for primary transcripts. The other fraction is left untreated (no TEX) to preserve all RNA species, both primary and processed. To

preserve strand specificity, the 5' ends of both RNA fractions are treated with tobacco acid pyrophosphatase (TAP) to convert 5′ triphosphate ends into 5′ monophosphate ends. A RNA linker is then added to the 5' end of the RNA molecules, followed by cDNA synthesis and sequencing library construction. TSS mapping is carried out by comparing the TEX treated and untreated libraries to find relative coverage differences at the 5' end of primary transcripts (i.e. primary transcripts will have more coverage at the 5' ends in the TEX treated libraries) [33]. This provides a global approach to identify all primary RNAs and the exact position at which they are transcribed, under any condition [33]. However, a significant drawback to this method is that it does not provide information on the length of the sRNAs because it is restricted to the 5′-ends of the transcripts; it is also biased against processed sRNAs [14].

RNA-seq in itself is inherently not strand-specific, motivating the development of multiple methods to ensure strand-specificity. These methods include the ligation of a RNA linker either at the 5' or 3' end of a RNA molecule and the use of the linker for priming cDNA synthesis. Alternatively, dUTP is incorporated into the second strand synthesis, which is then degraded, leaving the first strand for sequencing or template switching PCR [33]. In Gelsinger *et al.* (2018) a modified sRNA-seq protocol was adapted from eukaryotic piRNA-seq methods to construct a sRNA-seq library that enabled strand-specific deep sequencing and identification of thousands of sRNAs in *H. volcanii*. In this method, sRNA in the 50-500 nt range were size-selected using polyacrylamide gel electrophoresis. The sRNAs were eluted and enriched and then strand-specific libraries were prepared using either dUTP incorporation into second strand synthesis or template switching PCR [as used in the SMART-seq kit (Takara)]. Paired-end sequencing (2 x 100 bp) was then carried out on the Illumina HiSeq 2500 platform with an average of 10 million reads per library [1]. By significantly enriching for sRNAs, this method

provides better detection of full length of sRNAs. and its strand specificity allowed for the clear identification of antisense and intergenic sRNAs. However, the detection of internal sense sRNAs appeared to be difficult because of their masking by mRNA reads.

Besides library preparation and sequencing, another major difficulty in sRNA identification is in the bioinformatic analysis of the RNA-seq data. General principles of sRNA identification are to map short sequencing reads against a reference genome (if available, if not *de novo* assembly of transcripts), to assemble these aligned reads into transcripts, to quantify transcript expression levels, and to predict ORFs (or lack thereof) [1]. One challenge in this process is the read mapping to archaeal genomes using software only designed for Eukaryotes (splice-aware) or that require a previously known transcriptome [1]. Another challenge is the assembly of these mapped reads into transcripts for archaeal genomes where transcript boundaries are not well defined and neighboring transcripts frequently overlap [1]. While, no single pipeline has been published to specifically identify sRNA in Archaea, the computational strategy used in Gelsinger et al. (2018) presents a significant step forward in designing analytical pipeline specific for archaeal and bacterial sRNA discovery.

Molecular mechanisms of sRNA regulation in Archaea

Despite the discovery of thousands of sRNAs in archaeal transcriptomes, functional and mechanistic characterizations of sRNAs in the *Archaea* is in its infancy. Initial insight into antisense sRNA mechanisms comes from recent work in *H. volcanii,* showing that an overwhelming majority of all sRNAs expressed in this organism are antisense [1, 14] (**Fig. 2**). Of these, only a minority (7%) overlapped the 5’ UTR of mRNAs, which is in concurrence with findings that most mRNAs in *H. volcanii* are leaderless (lacking a 5' UTR) [1], while most (67%)

Peer-reviewed version available at Genes 2018, 9, 141; doi:10.3390/genes9030141

overlapped within the coding sequence (CDS) of mRNAs (**Fig. 2**). In bacterial itsRNAs and eukaryal sRNAs, the region of interaction (hybridization) between a sRNA and its target mRNA has been termed a "seed" region [1]. In *H. volcanii* no "seed" binding region for CDS-binding sRNAs could be found, indicating that they likely have full occupancy upon the mRNA [1]. Lastly, a smaller fraction (26%) of asRNAs overlapped the 3' UTR of mRNAs [1] **(Fig. 2).**

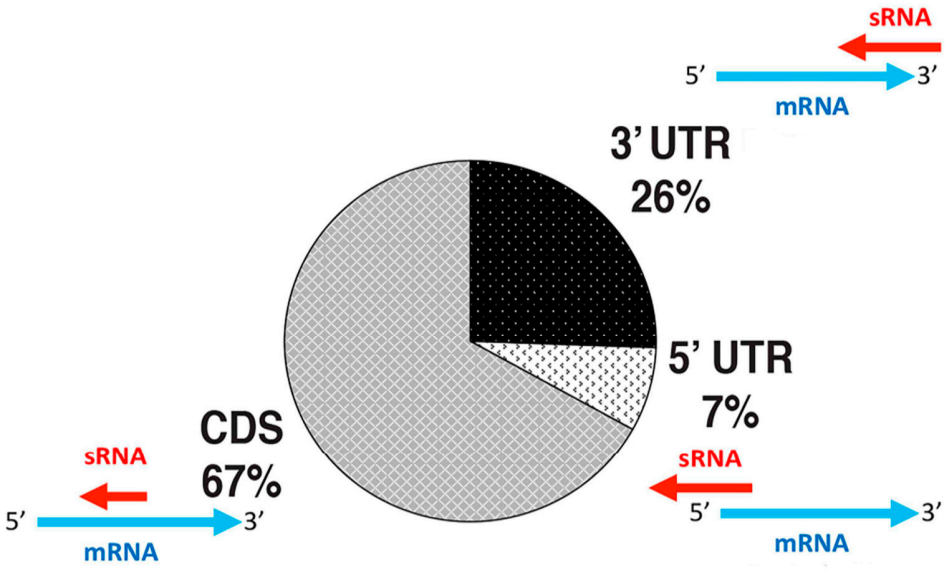

Figure 2: Distribution of binding regions for antisense sRNAs. UTR, untranslated region; CDS, coding sequence [1].

In a recent study of sRNAs in *H. volcanii,* Gelsinger, *et al*. [1] found that a large number of asRNAs were either up-regulated and down-regulated during oxidative stress, revealing two types of antisense sRNA populations. An anti-correlation was observed for a group of up-regulated antisense sRNAs and their down-regulated *cis*-encoded putative targets, indicating a potential mechanism of negative regulation [1]. In contrast, many *cis*-encoded putative mRNA targets of differentially regulated asRNAs exhibited a positive correlation in their expression patterns to oxidative stress (*cis*-pairs differentially expressed in the same orientation), suggesting

a positive regulatory effect between asRNA-mRNA *cis*-pairs [1]. Although negative regulatory effect of asRNAs on their target mRNAs was also suggested by results from another study, also in *H. volcanii* [14], experimental evidence are still lacking because of the inherent difficulty at manipulating such overlapping sRNA-mRNAs pairs.

While the regulatory effects of asRNAs can be readily inferred because of the overlap with their mRNAs targets, it is rather different with intergenic sRNAs where finding targets is a particularly difficult task. As a consequence, mechanistic insights into the regulation of intergenic sRNAs have only been provided for very few specific sRNAs. Prasse et al. [2] determined that, in *M. mazei,* a nitrogen starvation-specific sRNA, $sRNA_{154}$, stabilized some mRNAs while inhibiting translation initiation for other mRNAs, thus playing a dual regulatory role. $sRNA_{154}$ is highly conserved in the *Methanosarcina* and it was predicted to form a stable secondary structure with two loops required for stabilization of mRNA targets. The authors of the study proposed that the mechanistic role of the two loops was to mask endonucleolytic cleaveage sites of RNases by hybridizing to the mRNAs target and, thus, preserving the mRNA for translation [2]. In contrast, they also showed that loop 2 of $sRNA_{154}$ contains anti-ribosome binding site (RBS) sequences that masked the RBS of the $glnA_2$-mRNA target, repressing translation initiation [2]. Another sRNA in *M. mazei*, $sRNA_{41}$, was also found to repressed its targets at the translation level by masking ribosome binding sites within polycistronic mRNAs [13]. Other molecular mechanisms have been identified in the archaea such as the binding of itsRNAs to the 3'UTR of mRNA targets in *S. solfataricus* [34] and, more recently, in *H. volcanii* (Kliemt, Jaschinski, and Soppa, unpublished data). This is of particular interest in the haloarchaea because 72% of their transcripts are leaderless [14]. Taken together these studies are building a narrative of sRNA (both antisense and intergenic) mechanisms in the archaea,

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 25 January 2018 doi:10.20944/preprints201801.0232.v1

Peer-reviewed version available at Genes 2018, 9, 141; doi:10.3390/genes9030141

combining global approaches with individual targeted sRNA studies, demonstrating that sRNAs are also essential partners in gene regulation in the third domain of life **(Fig. 3)**.

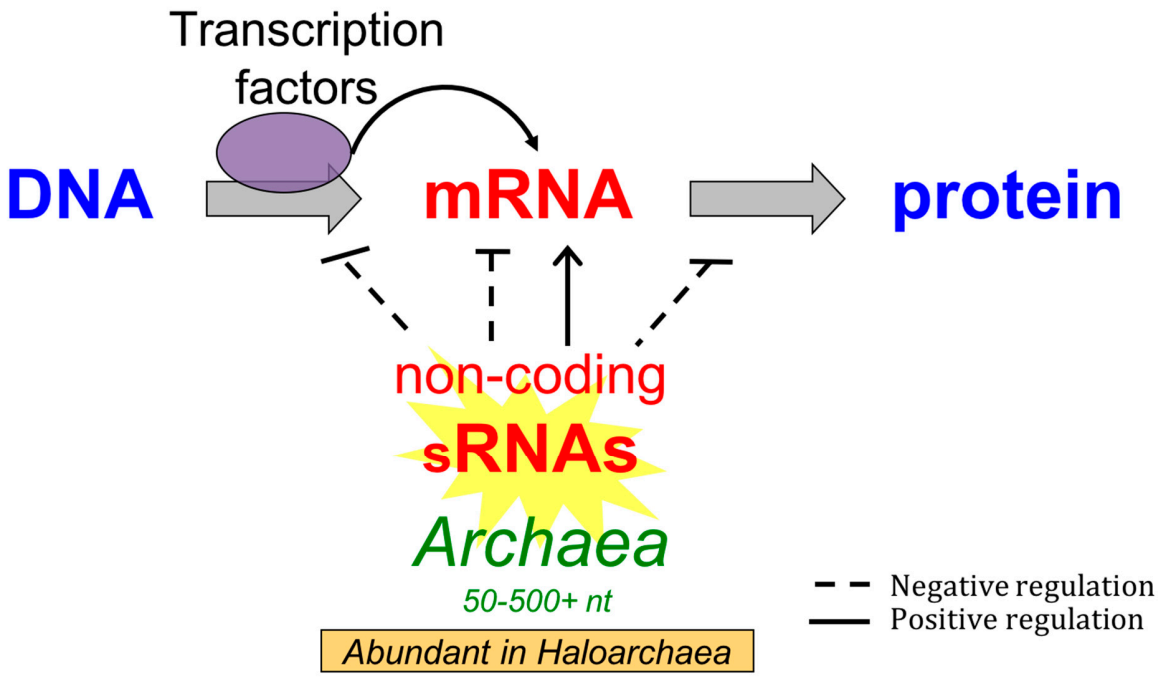

Figure 3: Correlation- and experimental-based regulatory mechanisms for sRNAs reported in the *Archaea*.

Gene regulatory networks and sRNAs

There are many known advantages of sRNA regulators including reduce metabolic cost, additional levels of regulation, unique regulatory properties, and faster response to stresses. Indeed the regulatory effects of sRNAs are often observed within minutes in bacterial systems [4]. Furthermore, sRNAs in bacteria can regulate very large gene networks as well as key transcription factors [4]. Examples of these include OxyS and SgrS in *E. coli,* involved in oxidative and glucose-phosphate stress, respectively [20].

In *H. volcanii*, many intergenic sRNAs are differentially expressed in response to varying environmental conditions, including elevated temperature, osmotic stress, nutrient limitation, and

oxidative stress, [1, 8, 12, 35, 36]. While phenotypic characterization of sRNA deletion mutants, including 10 gain-of-function phenotypes out of 27 mutants tested, confirmed their roles in metabolic regulation, stress adaptation and complex behavior [12, 36], their targets are still unknown with a few exceptions. Some of these exceptions comes from the study of *M. mazei* cultures grown under nitrogen starvation conditions where RNA-seq experiments revealed the differential expression of a number of sRNAs in response to nitrogen availability [23, 37]. This then resulted in the identification of the first *in vivo* targets for archaeal intergenic sRNAs [23, 37]. A potential target for one of these sRNAs, $sRNA_{162}$, was a bicistronic mRNA encoding for a transcription factor involved in regulating the switch between carbon sources and for a protein of unknown function [37]. Another sRNA, $sRNA_{154}$, was also exclusively expressed during nitrogen starvation conditions and the multiple targets for $sRNA_{154}$ included mRNAs for the α subunit of nitrogenase (*nif*H), the transcriptional activator of the *nif* operon (*nrp*A), and glutamine synthase1/2 (*gln*A_1/*gln*A_2). $sRNA_{154}$ was found to stabilize nifH-, nrpA-, and $glnA_1$-mRNAs but to block the translation of $glnA_2$-mRNA. Thus the proposed functional role of $sRNA_{154}$ was to regulate N_2-fixation under nitrogen limiting conditions by stabilizing transcripts involved in nitrogenase production (both regulators of and the nitrogenase itself), leading to a feed forward regulatory system [2]. Most recently, another *M. mazei* itsRNA, $sRNA_{41}$, was found to be down-regulated during nitrogen limiting conditions. Targets of $sRNA_{41}$ were involved in acetyl-CoA- decarbonylase/synthase complexes (ACDS) and were repressed at the translational level [13]. Thus $sRNA_{41}$ was predicted to play a role in repressing ACDS protein levels, however, under nitrogen limiting conditions, $sRNA_{41}$ was down-regulated, thus allowing ACDS levels to increase, which in turn provided sufficient amino acids for nitrogenase synthesis and energy for N_2-fixation [13]. While these studies provide great examples of gene network

regulated by sRNAs in the archaea, additional work is needed to identify many more molecular targets of archaeal itsRNAs and the diverse mechanisms of their sRNA-mRNA interactions.

In contrast, asRNAs, which are by far the largest group of sRNAs found in the *Archaea*, are encoded in the opposite strand of their putative target. In the hyperthermophile *Pyrobaculum*, 3 antisense sRNAs were found opposite a ferric uptake regulator, a triose-phosphate isomerase, and transcription factor B, supporting a potential role in the regulation of iron, transcription, and core metabolism [24]. Target enrichment of asRNAs differentially regulated by oxidative stress in *H. volcanii* included mRNAs involved in transposon mobility, chemotaxis signaling, peptidase activity, and transcription factors [1]. The functional enrichment of transposon targeted by asRNAs suggests that during oxidative stress transposon activity is tightly regulated in *H. volcanii*, potentially explaining its increased resistance to oxidative stress conditions [1]. Indeed, transposons are genetic elements that hop around in the genome causing double strand breaks. This added stress would likely be detrimental to a cell under oxidative stress, hence a need to be silenced [38-40]. sRNAs antisense to transposons were also reported for *Thermococcus kodakarensis* [30], *S. solfataricus* [22], and *M. mazei* [23] suggesting that, similarly to bacteria, regulation of transposition is mediated by asRNAs in archaea [41].

Future Prospects

To date, very few of the newly reported candidate sRNAs in the archaea have been functionally characterized [2, 13, 37] and many questions remain: *what are the targets of the multitude of sRNAs discovered in the Archaea? What are the regulatory effects of these sRNAs? And more importantly, what type of molecular mechanisms can we expect in a domain of life where information processing systems are a mosaic of bacterial and eukaryal systems* [42]?

Co-immuno-precipitation with the Lsm protein, the archaeal homolog of Hfq, was used to "capture" sRNAs in vitro [6]. However, the role of Lsm – or any other RNA-binding protein - remains to be elucidated in the archaea. Homologs of eukaryotic miRNA interacting proteins (Argonautes) have been found in the archaea, but rather than RNA interference, a defensive role again foreign genetic elements have been proposed [43]. Questions about what ribonucleoprotein complexes are involved in archaeal sRNA regulation and their mechanistic roles remain unanswered. Furthermore, negative regulatory effects have been reported for archaea sRNAs but we still have no information if there are RNases involved in degrading target mRNAs, as observed in *Bacteria*.

Outstanding questions also remain regarding the role of more than 1100 cis-sense sRNAs recently discovered in *H. volcanii* [14], the prevalence of regulatory tRNA-derived fragments in the archaea [8, 44] and the potential for sRNAs to encode small peptides such as in bacteria and eukarya [45]. Finally, *in vivo* quantitative measurements of sRNA-mediated regulation, such as those currently being made in the bacteria [46], are necessary to understand, at a system-level, how sRNA-based regulation is integrated within a cell's regulatory networks.

Acknowledgements: This work was supported by grant FA9950-14-1-0118 from the AFOSR.

References:

1. Gelsinger, D.R. and J. DiRuggiero, *Transcriptional landscape and regulatory roles of small non-coding RNAs in the oxidative stress response of the haloarchaeon Haloferax volcanii.* bioRxiv, 2017.

2. Prasse, D., et al., *sRNA154 a newly identified regulator of nitrogen fixation in Methanosarcina mazei strain Gö1.* RNA Biology, 2017. **14**(11): p. 1544-1558.

3. Cech, T.R., Steitz, J.A, *The noncoding RNA revolution-trashing old rules to forge new ones.* . Cell 2014. **157**(1): p. 77-94.

4. Storz, G., J. Vogel, and K.M. Wassarman, *Regulation by small RNAs in bacteria: expanding frontiers.* Mol Cell, 2011. **43**(6): p. 880-91.

5. Wagner, E.G.H. and P. Romby, *Chapter Three - Small RNAs in Bacteria and Archaea: Who They Are, What They Do, and How They Do It*, in *Advances in Genetics*, T. Friedmann, J.C. Dunlap, and S.F. Goodwin, Editors. 2015, Academic Press. p. 133-208.

6. Fischer, S., et al., *The archaeal Lsm protein binds to small RNAs.* J Biol Chem, 2010. **285**(45): p. 34429-38.

7. Fischer, S., et al., *Regulatory RNAs in Haloferax volcanii.* Biochem Soc Trans, 2011. **39**(1): p. 159-62.

8. Heyer, R., et al., *High throughput sequencing reveals a plethora of small RNAs including tRNA derived fragments in Haloferax volcanii.* RNA Biol, 2012. **9**(7): p. 1011-8.

9. Marchfelder, A., et al., *Small RNAs for defence and regulation in archaea.* Extremophiles, 2012. **16**(5): p. 685-96.

10. Schmitz-Streit, R., et al., *Archaea employ small RNAs as regulators*, in *Regulatory RNAs in Prokaryotes*, W.R. Hess and A. Marchfelder, Editors. 2011, Springer Verlag: Wien. p. 131-145.

11. Soppa, J., et al., *Small RNAs of the halophilic archaeon Haloferax volcanii.* Biochem Soc Trans, 2009. **37**(Pt 1): p. 133-6.

12. Straub, J., et al., *Small RNAs in haloarchaea: identification, differential expression and biological function.* RNA Biol, 2009. **6**(3): p. 281-92.

13. Buddeweg, A., et al., *sRNA41 affects ribosome binding sites within polycistronic mRNAs in Methanosarcina mazei Gö1.* Molecular Microbiology, 2017: p. n/a-n/a.

14. Babski, J., et al., *Genome-wide identification of transcriptional start sites in the haloarchaeon Haloferax volcanii based on differential RNA-Seq (dRNA-Seq).* BMC Genomics, 2016. **17**(1): p. 629.

15. Morris KV, M.J., *The rise of regulatory RNA.* Nature Reviews Genetics, 2014. **15**(6): p. 423–437.

16. Li, Y., Liu, X., Huang, L., Guo, H., Wang, X.J. , *Potential coexistence of both bacterial and eukaryotic small RNA biogenesis and functional related protein homologs in Archaea.* . J Genet Genomics, 2010(8): p. 493-503.

17. Zander, A., et al., *Guide-independent DNA cleavage by archaeal Argonaute from Methanocaldococcus jannaschii.* Nature Microbiology, 2017. **2**: p. 17034.

18. Gottesman, S. and G. Storz, *Bacterial small RNA regulators: versatile roles and rapidly evolving variations.* Cold Spring Harb Perspect Biol, 2011. **3**(12).

19. Thomason, M.K. and G. Storz, *Bacterial antisense RNAs: How many are there and what are they doing?* Annual review of genetics, 2010. **44**: p. 167-188.

20. Altuvia, S., et al., *A small, stable RNA induced by oxidative stress: role as a pleiotropic regulator and antimutator.* Cell, 1997. **90**(1): p. 43-53.

21. Smirnov, A., et al., *Grad-seq guides the discovery of ProQ as a major small RNA-binding protein.* Proceedings of the National Academy of Sciences, 2016. **113**(41): p. 11591-11596.

22. Wurtzel, O., et al., *A single-base resolution map of an archaeal transcriptome.* Genome Res, 2010. **20**(1): p. 133-41.

23. Jager, D., et al., *Deep sequencing analysis of the Methanosarcina mazei Go1 transcriptome in response to nitrogen availability.* Proc Natl Acad Sci U S A, 2009. **106**(51): p. 21878-82.

24. Bernick, D.L., et al., *Diversity of Antisense and Other Non-Coding RNAs in Archaea Revealed by Comparative Small RNA Sequencing in Four Pyrobaculum Species.* Front Microbiol, 2012. **3**: p. 231.

25. Toffano-Nioche, C., et al., *RNA at 92 °C: the non-coding transcriptome of the hyperthermophilic archaeon Pyrococcus abyssi.* RNA Biol, 2013. **10**.

26. Jäger, D., et al., *Primary transcriptome map of the hyperthermophilic archaeon Thermococcus kodakarensis.* BMC Genomics, 2014. **15**.

27. Li, J., et al., *Global mapping transcriptional start sites revealed both transcriptional and post-transcriptional regulation of cold adaptation in the methanogenic archaeon Methanolobus psychrophilus.* Sci Rep, 2015. **5**.

28. Prasse, D., et al., *Regulatory RNAs in archaea: first target identification in Methanoarchaea.* Biochem Soc Trans, 2013. **41**(1): p. 344-9.

29. Toffano-Nioche, C., et al., *RNA at 92°C: The non-coding transcriptome of the hyperthermophilic archaeon Pyrococcus abyssi.* RNA Biology, 2013. **10**(7): p. 1211-1220.

30. Jäger, D., et al., *Primary transcriptome map of the hyperthermophilic archaeon Thermococcus kodakarensis.* BMC Genomics, 2014. **15**(1): p. 684.

31. Li, J., et al., *Global mapping transcriptional start sites revealed both transcriptional and post-transcriptional regulation of cold adaptation in the methanogenic archaeon Methanolobus psychrophilus.* Scientific Reports, 2015. **5**: p. 9209.

32. Crits-Christoph, A., et al., *Functional analysis of the archaea, bacteria, and viruses from a halite endolithic microbial community.* Env. Microbiol. , 2016. **18**: p. 2064-2077. doi: 10.1111/1462-2920.13259.

33. Sharma, C.M. and J. Vogel, *Differential RNA-seq: the approach behind and the biological insight gained.* Current Opinion in Microbiology, 2014. **19**: p. 97-105.

34. Märtens, B., et al., *Antisense regulation by transposon‐derived RNAs in the hyperthermophilic archaeon <em>Sulfolobus solfataricus</em&gt.* EMBO reports, 2013. **14**(6): p. 527.

35. Babski, J., et al., *Bioinformatic prediction and experimental verification of sRNAs in the haloarchaeon Haloferax volcanii.* RNA Biol, 2011. **8**(5): p. 806-16.

36. Jaschinski, K., et al., *Generation and Phenotyping of a Collection of sRNA Gene Deletion Mutants of the Haloarchaeon Haloferax volcanii.* PLOS ONE, 2014. **9**(3): p. e90763.

37. Jager, D., et al., *An archaeal sRNA targeting cis- and trans-encoded mRNAs via two distinct domains.* Nucleic Acids Res, 2012. **40**(21): p. 10964-79.

38. Whitehead, K., et al., *An integrated systems approach for understanding cellular responses to gamma radiation.* Mol Syst Biol 2006. **47**: p. doi:10.1038/msb4100091.

39. Wheeler, B.S., *Small RNAs, big impact: small RNA pathways in transposon control and their effect on the host stress response.* Chromosome Research, 2013. **21**(6): p. 587-600.

40. Capy, P., et al., *Stress and transposable elements: co-evolution or useful parasites?* Heredity, 2000. **85**(2): p. 101-106.

41. Balasubramanian, D., et al., *A Prophage-encoded small RNA controls metabolism and cell division in Escherichia coli.* mSystems, 2016. **1**(1).

42. Bell, S.D. and S.P. Jackson, *Transcription and translation in Archaea: a mosaic of eukaryal and bacterial features.* Trends in Microbiology, 1998. **6**(6): p. 222-228.

43. Willkomm, S., et al., *Structural and mechanistic insights into an archaeal DNA-guided Argonaute protein.* Nature Microbiol, 2017. **2**: p. 17035.

44. Gebetsberger, J., et al., *A tRNA-derived fragment competes with mRNA for ribosome binding and regulates translation during stress.* RNA Biology, 2017. **14**(10): p. 1364-1373.

45. Bronsard, J., et al., *sRNA and cis-antisense sRNA identification in Staphylococcus aureus highlights an unusual sRNA gene cluster with one encoding a secreted peptide.* Scientific Reports, 2017. **7**(1): p. 4565.

46. Bobrovskyy, M., et al., *Determinants of target prioritization and regulatory hierarchy for the bacterial small RNA SgrS.* bioRxiv, 2017.