L7Ae- and LSm-RNA interactomes of *Sulfolobus acidocaldarius*

Dissertation

zur

Erlangung des Doktorgrades

der Naturwissenschaften

(Dr. rer. nat.)

dem Fachbereich Biologie

der Philipps-Universität Marburg

vorgelegt von

Michael Daume

aus Großkrotzenburg

Marburg/Lahn, September 2017

Die Untersuchungen zur vorliegenden Arbeit wurden von November 2013 bis Juli 2017 unter der Betreuung von Herrn Dr. Lennart Randau am Max-Planck-Institut für terrestrische Mikrobiologie in Marburg durchgeführt.

Vom Fachbereich Biologie

der Philipps-Universität Marburg als Dissertation

angenommen am: 20.04.2018

Erstgutachter: Herr Dr. Lennart Randau
Zweitgutachter: Herr Prof. Dr. Martin Thanbichler

Tag der mündlichen Prüfung am: 03.05.2018

Teile dieser Arbeit wurden in folgendem Artikel veröffentlicht:

Daume M, Uhl M, Backofen R, Randau L. *RIP-Seq suggests translational regulation by L7Ae in Archaea.* mBio, 2017 Aug 1; 8 (4) pii: e00730-17

Dedicated to my family and friends

Table of Contents

Abbreviations..... I

Summary..... II

Zusammenfassung..... III

1. Introduction..... 1

1.1 Archaea – the third domain of life..... 1

1.2 Archaeal small RNAs..... 2

1.3 The kink-turn and its recognition by the L7Ae/L30 protein family..... 5

1.4 The Sm superfamily and the elusive function of the archaeal LSm proteins..... 8

1.5 The model organism *Sulfolobus acidocaldarius*..... 12

1.6 Aims of this work..... 13

2. Results..... 14

2.1 The small RNome of *S. acidocaldarius*..... 14

2.1.1 Identification of the small RNome of *S. acidocaldarius* by small RNA-Seq..... 14

2.2 The L7Ae-RNA interactome of *S. acidocaldarius*..... 17

2.2.1 Immunoprecipitation of genomically tagged L7Ae from *S. acidocaldarius*..... 17

2.2.2 Identification of the L7Ae-interacting RNAs using RIP-Seq analysis..... 20

2.2.3 Binding analysis of the *l7ae* 5' UTR by EMSA studies..... 23

2.2.4 L7Ae autoregulation analysis using β-galactosidase reporter assays..... 24

2.2.5 Establishment of a bacterial GFP reporter system for L7Ae/k-turn binding studies.. 25

2.2.6 Multiple sequence alignments of archaeal *l7ae* 5' UTRs..... 28

2.2.7 Analysis of archaeal *l7ae* 5' UTR binding by L7Ae via the GFP reporter system..... 29

2.2.8 Binding analysis to identified k-turn motifs in different mRNAs and SRP RNA..... 30

2.2.9 Analysis of the conservation of archaeal SRP RNA k-turns..... 32

2.2.10 Abolishment of L7Ae autoregulation in *S. acidocaldarius*..... 33

2.2.11 Competition of *l7ae* 5' UTR binding using natural substrates of L7Ae..... 33

2.3 The LSm-RNA interactome of *S. acidocaldarius*..... 34

2.3.1 Deletion of the three *lsm* genes of *S. acidocaldarius*..... 34

2.3.2 Growth analysis of the *lsm3* deletion strain..... 37

2.3.3 Genomic tagging of the *lsm* genes in *S. acidocaldarius*..... 37

2.3.4 Phenotypic analysis of the *lsm* mutant strains..... 38

2.3.5 Immunoprecipitation of His-tagged LSm proteins from *S. acidocaldarius*..... 40

2.3.6 Identification of the LSm-interacting RNAs using RIP-Seq analysis..... 42
2.3.7 Identification of two potential RNA binding motifs for LSm1 and LSm2..... 46
2.4 Subsequent findings of the LSm-RNA interactome study..... 49
2.4.1 LSm1 and LSm2 interaction study..... 49
2.4.2 Binding analysis of the U and UAG motifs by EMSAs..... 50
3. Discussion..... 52
3.1 RIP-Seq verifies RNA interaction partners of L7Ae in *S. acidocaldarius*..... 52
3.2 Archaeal SRP RNA – a novel RNA interactor of L7Ae?..... 52
3.3 L7Ae autoregulation as an adaption mechanism to changing C/D box sRNA pools..... 54
3.4 L7Ae regulates mRNAs encoding proteins involved in translation..... 57
3.5 L7Ae as a master regulator of translational processes – a theory..... 58
3.6 A bacterial reporter system that can be exploited by synthetic biology approaches..... 60
3.7 The LSm1 and LSm2 proteins of *S. acidocaldarius* form heteroheptameric rings..... 61
3.8 A putative role in mRNA degradation for *S. acidocaldarius* LSm1/2..... 61
3.9 LSm1/2 might be involved in several small RNA-regulated processes..... 64
3.10 Increased complexity for multiple archaeal LSm proteins?..... 66
3.11 Perspectives..... 69
4. Material and Methods..... 71
4.1 Materials, instruments and source of supplies..... 71
4.1.1 Chemicals, kits and enzymes..... 71
4.1.2 Instruments..... 72
4.1.3 Buffers and solutions..... 74
4.2 Strains and culture conditions..... 74
4.2.1 Strains..... 74
4.2.2 Culture conditions for *E. coli*..... 74
4.2.3 Culture conditions for *S. acidocaldarius*..... 75
4.3 Plasmids and oligonucleotides..... 75
4.3.1 Progenitor plasmids and constructed recombinant vectors..... 75
4.3.2 Oligonucleotides..... 79
4.4 Molecular working with DNA..... 88
4.4.1 DNA isolation..... 88
4.4.1.1 Phenol/Chloroform extraction of *S. acidocaldarius* genomic DNA..... 88
4.4.1.2 DNA precipitation..... 89
4.4.1.3 Plasmid DNA isolation from *E. coli*..... 89

4.4.2 Quantitation of DNA 89
4.4.2.1 Spectrophotometric quantitation and quality control 89
4.4.2.2 Fluorometric quantitation 89
4.4.3 Polymerase chain reaction (PCR) 89
4.4.3.1 Amplification of genomic DNA and plasmid DNA 90
4.4.3.2 Colony PCR 90
4.4.3.3 Overlap extension PCR 90
4.4.3.4 Site-directed mutagenesis PCR 91
4.4.4 Electrophoresis of DNA 91
4.4.4.1 Agarose gel electrophoresis 91
4.4.4.2 Non-denaturing polyacrylamide gel electrophoresis (native PAGE) 91
4.4.4.3 Denaturing polyacrylamide gel electrophoresis (denaturing PAGE) 92
4.4.5 Purification of DNA fragments 92
4.4.5.1 PCR purification 92
4.4.5.2 Gel extraction from agarose gels 92
4.4.5.3 Gel extraction from polyacrylamide gels 92
4.4.6 Hybridization of DNA oligonucleotides 93
4.4.7 Enzymatic modification of DNA 93
4.4.7.1 Restriction 93
4.4.7.2 Phosphorylation 93
4.4.7.3 Dephosphorylation 93
4.4.7.4 Ligation 93
4.4.7.5 Gibson Assembly 94
4.4.8 Transformation 94
4.4.8.1 Preparation of chemically competent *E. coli* cells 94
4.4.8.2 Transformation of *E. coli* 95
4.4.8.3 Preparation of electrocompetent *S. acidocaldarius* cells 95
4.4.8.4 Transformation of *S. acidocaldarius* 95
4.4.9 Sequencing 96
4.4.10 Radioactive labeling of DNA 5'-termini 96
4.5 Generation of genomic tags and mutations in *S. acidocaldarius* 96
4.6 Molecular working with RNA 96
4.6.1 Treatment of solutions, glassware and equipment 96
4.6.2 Isolation of *S. acidocaldarius* total and small RNAs 97

4.6.3 Co-immunoprecipitation of L7Ae-interacting RNAs..... 97
4.6.4 Co-immunoprecipitation of LSm-interacting RNAs..... 98
4.6.5 Quantitation of RNA..... 99
4.6.5.1 Spectrophotometric quantitation and quality control..... 99
4.6.5.2 Fluorometric quantitation..... 99
4.6.6 Denaturing polyacrylamide gel electrophoresis (denaturing PAGE)..... 99
4.6.7 Gel extraction from denaturing PAGE..... 99
4.6.8 RNA fragmentation by $ZnCl_2$ treatment and curation of RNA termini..... 99
4.6.9 Preparation of cDNA libraries for Illumina sequencing..... 100
4.6.10 Illumina HiSeq2500 sequencing..... 100
4.6.11 Run-off *in vitro* transcription of non-labeled and radioactively labeled RNA..... 101
4.6.12 Northern blot analysis..... 102
4.7 Biochemical methods..... 102
4.7.1 Heterologous production of *S. acidocaldarius* L7Ae in *E. coli*..... 102
4.7.2 Enrichment and purification of recombinant L7Ae..... 103
4.7.3 SDS-polyacrylamide gel electrophoresis (SDS-PAGE)..... 103
4.7.4 Western blot analysis..... 104
4.7.5 Bradford protein quantitation method..... 104
4.7.6 Electrophoretic mobility shift assay (EMSA)..... 105
4.7.7 Mass spectrometry..... 105
4.7.8 Reporter assays in *S. acidocaldarius* using β-galactosidase..... 105
4.8 Cell biological methods..... 106
4.8.1 Flow cytometry..... 106
4.8.1.1 GFP reporter system studies in *E. coli*..... 106
4.8.1.2 Analysis of *S. acidocaldarius lsm* mutant strains..... 106
4.8.2 Microscopy..... 107
4.9 Bioinformatic methods..... 107
4.9.1 Processing and mapping of Illumina sequencing data..... 107
4.9.2 Identification of the sRNome of *S. acidocaldarius*..... 108
4.9.3 Identification of the L7Ae and LSm interacting RNAs by the DESeq2 tool..... 108
4.9.4 Identification of potential LSm binding motifs using the MEME tool..... 109
4.9.5 Statistics..... 109
5. References..... 110
6. Appendix..... 125

Abbreviations

Δ	gene deletion
% (v/v)	percent by volume
% (w/v)	percent by weight
3' or 5' UTR	3' or 5' untranslated region of mRNA
5-FOA	5-fluoroorotic acid
A	Ampere
aKt	autoregulatory k-turn
Amp	ampicillin
arCOG	archaeal cluster of orthologous groups of proteins
APS	ammonium persulfate
ATP	adenosine triphosphate
bp	basepair(s)
C-terminal	carboxy-terminal
Cam	chloramphenicol
Cas	CRISPR-associated protein
cDNA	complementary DNA
cpm	counts per minute
CRISPR	Clustered Regularly Interspaced Short Palindromic Repeats
crRNA	CRISPR RNA
DEPC	diethylpyrocarbonate
DMSO	dimethyl sulfoxide
DNA	deoxyribonucleic acid
dNTP	deoxyribonucleoside triphosphate
dsDNA	double-stranded DNA
DTT	dithiothreitol
e.g.	for example ("exempli gratia")
EDTA	ethylene-diamine-tetraacetic acid
EMSA	electrophoretic mobility shift assay
et al.	and other ("et alteri")
FHA	Flag-HA tag
FPLC	Fast Protein Liquid Chromatography
frag RNA	RIP-Seq data of fragmented immunoprecipitated RNA
g	gram
x *g*	gravitational acceleration
GFP	green fluorescent protein
*gfp*Kt	*gfp*-regulatory k-turn
h	hour(s)
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
Hfq	bacterial Sm-like protein Host factor of phage Q
His	6x histidine tag
i.e.	that is ("id est")
IP	immunoprecipitation
IPTG	isopropyl β-$_D$-1-thiogalactopyranoside
Kan	kanamycin
kb	kilobases
kDa	kilo Dalton
KO	knock-out
Kt	kink-turn or k-turn
Kt-b	bulged strand of k-turn
Kt-n	non-bulged strand of k-turn
l	liter
LB	lysogeny broth
LSm	archaeal or eukaryotic Sm-like protein
M	molar (mol/l)
m	meter
min	minute(s)
MOPS	3-(*N*-morpholino)propanesulfonic acid
mRNA	messenger RNA
µ	micro (10^{-6})
n	nano (10^{-9})
N-terminal	amino-terminal
Ni-NTA	nickel-nitrilotriacetic acid
nt	nucleotides
NTP	nucleoside triphosphate
OD_{600}	optical density at 600 nm
ONPG	*ortho*-nitrophenyl-β-galactoside
PAGE	polyacrylamide gel electrophoresis
PCR	polymerase chain reaction
pH	potential of hydrogen
RIP-Seq	RNA-immunoprecipitation sequencing
RNA	ribonucleic acid
RNase	ribonuclease
RNA-Seq	high-throughput RNA sequencing
RNP	ribonucleoprotein
rpm	revolutions per minute
rRNA	ribosomal RNA
RrrR	RNase R-resistant RNA (sRNA)
RT	room temperature
s	second(s)
Sac-sR#	C/D box sRNA of *Sulfolobus acidocaldarius*
SD	Shine-Dalgarno sequence
SDS	sodium dodecyl sulfate
snRNA	eukaryotic small nuclear RNA
snoRNA	eukaryotic small nucleolar RNA
sRNA	small regulatory RNA
SRP RNA	signal recognition particle RNA
ssDNA	single-stranded DNA
shift	size difference of gel bands due to changes of mobility
tRNA	transfer-RNA
TAE	tris-acetate EDTA-buffer
TBE	tris-borate EDTA-buffer
TEMED	N,N,N',N'-tetramethylethylenediamine
Tris	tris-(hydroxymethyl)-aminomethane
U	uridine nucleoside
V	Volt
W	Watt
WT	wild-type

Summary

The archaeal L7Ae and Sm-like proteins (LSm) are universal RNA-binding proteins. L7Ae stabilizes non-coding RNA species, including ribosomal RNA, by recognizing a structural RNA motif, termed kink-turn (k-turn). Sm family proteins, like bacterial Hfq and eukaryotic Sm/LSm, are involved in multiple RNA-related processes including small RNA (sRNA)-based translational regulation, mRNA decay or splicing. However, the function of the archaeal members is elusive.

Using RNA-immunoprecipitation sequencing (RIP-Seq) methodology, this thesis aimed to identify the global RNA interaction partners (RNA interactome) of L7Ae and the three LSm proteins of the thermoacidophilic archaeon *Sulfolobus acidocaldarius*. Besides many known non-coding RNAs, the SRP RNA was identified as a novel binding partner of the L7Ae protein. Mobility shift assays demonstrated L7Ae binding to a k-turn motif that was found to be conserved among archaeal SRP RNAs. Interestingly, mRNAs, including the *l7ae* transcript, were enriched in the RIP-Seq analysis and found to comprise putative k-turns that facilitate L7Ae binding. *In vivo* studies showed that L7Ae autoregulates the translation of its mRNA by binding to a k-turn motif in the 5' untranslated region. A GFP reporter system was established in *Escherichia coli* that verified the conservation of L7Ae-mediated feedback regulation in archaea and provides a new tool for the modulation of synthetic gene circuits in bacteria. Mobility shift assays confirmed binding of L7Ae to a k-turn in the transcript of *nop5-fibrillarin*, suggesting that the synthesis of all C/D box sRNP core proteins (L7Ae, Nop5 and fibrillarin) is regulated by L7Ae. These studies revealed the regulation of mRNA translation as a novel function of the archaeal L7Ae protein.

The LSm RIP-Seq study found mRNAs and sRNAs as LSm1 and LSm2 interactors, including a recently reported sRNA that regulates biofilm formation in *S. acidocaldarius*. No RNA-binding capacity was observed for LSm3. A computational analysis of the interaction partners identified the U-rich 3' termination signal of RNAs and a motif composed of UAG triplets as potential LSm binding sites, which was verified by mobility shift assays. Knock-out studies revealed that only the *lsm3* gene is dispensable, whereas the genes for *lsm1* and *lsm2* seem to be essential. Mutant strains producing tagged versions of LSm1 and LSm2 displayed a pleiotropic phenotype. In resemblance to the roles of bacterial Hfq and eukaryotic LSm proteins, this study provides hints that archaeal LSm proteins may be involved in mRNA degradation, C/D box sRNA biogenesis and sRNA-regulated processes, like tRNA maturation and translational regulation of mRNAs.

Zusammenfassung

Die archaealen Proteine L7Ae und LSm (Sm-ähnliche Proteine) sind universelle RNA-Bindeproteine. L7Ae stabilisiert nicht-kodierende RNAs, einschließlich ribosomaler RNA, durch Erkennung eines RNA-Strukturmotivs, namens „kink-turn“ (k-turn). Proteine der Sm-Familie, wie die bakteriellen Hfq und die eukaryotischen Sm/LSm-Proteine, sind an multiplen RNA-basierten Prozessen beteiligt, wie der Regulation der mRNA-Translation durch kleine RNAs (sRNAs), der mRNA-Degradation oder dem Spleißen von prä-mRNAs. Die Funktion der archaealen Mitglieder dieser Familie ist hingegen noch nicht geklärt.

Das Ziel dieser Arbeit war die Identifikation aller RNA-Interaktionspartner (RNA-Interaktom) von L7Ae und der drei LSm-Proteine des thermoacidophilen Archaeons *Sulfolobus acidocaldarius* mittels RNA-Immunopräzipitation mit anschließender Hochdurchsatz-Sequenzierung (RIP-Seq). Neben vielen bereits bekannten nicht-kodierenden RNAs konnte die SRP RNA als ein neuartiger Bindepartner von L7Ae identifiziert werden. Mithilfe von Mobilitätsshift-Studien konnte die Bindung von L7Ae an ein konserviertes k-turn-Motiv der archaealen SRP RNA demonstriert werden. Interessanterweise zeigte die durchgeführte RIP-Seq-Analyse eine Anreicherung diverser mRNAs, welche potentielle k-turn-Motive enthielten und somit vermutlich die Bindung durch L7Ae ermöglichten. Unter diesen mRNAs befand sich auch das *l7ae*-Transkript. *In vivo*-Studien konnten nachweisen, dass L7Ae die Translation seiner eigenen mRNA durch Bindung an ein k-turn-Motiv in der 5'-untranslatierten Region reguliert. Ein in *Escherichia coli* etabliertes GFP-Reportersystem bestätigte die Konservierung der L7Ae-Autoregulation in Archaeen und kann zukünftig als Werkzeug für die Erstellung synthetischer Schaltkreise verwendet werden. Mobilitätsshift-Studien zeigten die Bindung von L7Ae an ein k-turn-Motiv im *nop5-fibrillarin*-Transkript. Dies lässt vermuten, dass L7Ae in der Lage ist die Synthese aller C/D box sRNP-Proteinkomponenten (L7Ae, Nop5, Fibrillarin) zu regulieren. Durch diese Studien wurde die translationale Regulation von mRNAs als neue Funktion von archaealem L7Ae aufgedeckt.

Die LSm-RIP-Seq-Analyse identifizierte mRNAs und sRNAs als LSm1- und LSm2-Bindepartner, einschließlich einer kürzlich entdeckten sRNA, die an der Biofilm-Herstellung in *S. acidocaldarius* beteiligt ist. Für das LSm3-Protein konnte keine RNA-Bindeaktivität beobachtet werden. Durch eine computergestützte Analyse der Interaktionspartner von LSm1 und LSm2 wurde das U-reiche 3'-Terminationssignal von RNAs, sowie ein aus UAG-Triplets bestehendes Motiv als potentielle LSm-Bindestellen identifiziert. Die Bindung der beiden Motive durch die LSm-Proteine wurde in Mobilitätsshift-Studien nachgewiesen. Während

Deletionsstudien zeigten, dass das *lsm3*-Gen entbehrlich ist, scheinen die Gene für *lsm1* und *lsm2* Gene essentiell für *S. acidocaldarius* zu sein. Darüber hinaus konnte für Mutanten, die getaggte LSm1- und LSm2-Proteine synthetisierten, ein pleiotropischer Phänotyp beobachtet werden. Vergleichbar mit den Funktionen, die für das bakterielle Hfq und für die eukaryotischen LSm-Proteine beschrieben wurden, deuten die Ergebnisse dieser Studie darauf hin, dass die archaealen LSm-Proteine an der Degradation von mRNA, der Herstellung von C/D box sRNAs und an sRNA-regulierten Prozessen wie der tRNA-Reifung und der translationalen Regulation von mRNAs beteiligt sind.

1. Introduction

1.1 Archaea – the third domain of life

About 100 years after their discovery in 1880, C. Woese and G. Fox introduced Archaea as a new domain of life [1-3]. The novel classification was based on an approach using 16S ribosomal RNA sequences to assess the phylogeny between species. Woese´s postulated three-domain (Bacteria, Archaea and Eukarya) model of cellular life became widely accepted in the scientific community (Fig. 1.1) [2, 4]. The first cultivated members of Archaea belonged to the two phyla Euryarchaeota and Crenarchaeota [4]. In the last decade, novel phyla, e.g. Korarchaeota, Thaumarchaeota, Aigarchaeota, Bathyarchaeota, Nanoarchaeota or Woesearchaeota, were uncovered by modern cultivation-independent sequencing approaches which include metagenomic studies or single-cell sequencing [5-10]. However, several phyla remain poorly characterized and their phylogenetic placement is still debated [11]. Recently, a new archaeal lineage, the Lokiarchaeota, was isolated from a hydrothermal vent site in the Arctic Ocean (so-called Loki´s Castle) [12, 13]. Similarities to archaea within the so-called TACK superphylum (Thaum-, Aig-, Cren- and Korarchaeota) were observed. Interestingly, the Lokiarchaeota genomes were found to harbor many eukaryotic features, which supports earlier theories that eukaryotes emerged from an archaeal ancestor related to the TACK superphylum [12, 14]. Current illustrations of the tree of life therefore often show Eukarya as a monophyletic group with the Lokiarchaeota (Fig. 1.1) [15].

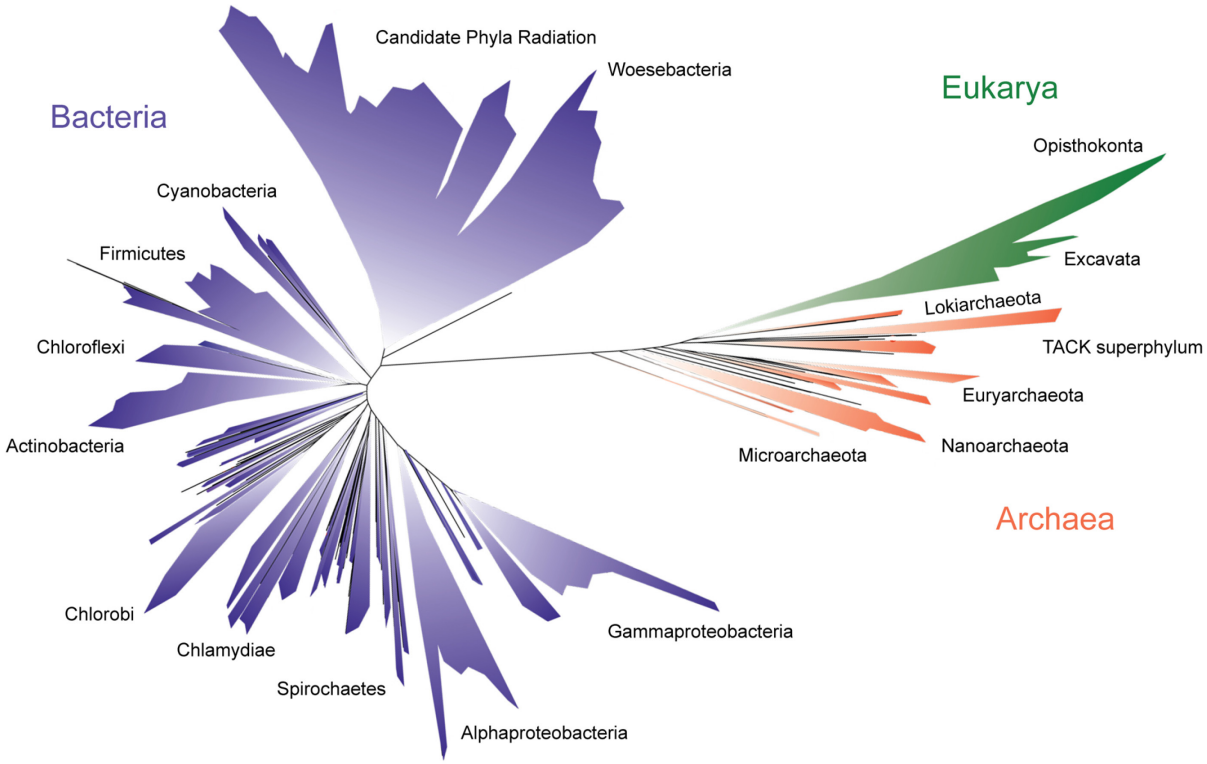


Figure 1.1 Current view of the tree of life. The relationship between species is illustrated (modified from [15]).

1.2 Archaeal small RNAs

Non-coding RNAs fulfill several essential cellular processes. Ribosomal RNAs (rRNAs) catalyze the translation of mRNAs into proteins. Transfer RNAs (tRNAs) decode the mRNA sequence and deliver amino acids to build peptide chains. The RNase P RNA cleaves 5' leader sequences of precursor tRNAs and the signal recognition particle RNA (SRP RNA) directs the secretion of proteins by guiding nascent peptide chains of the ribosome to the cell membrane/ER [16, 17]. With few exceptions, these four non-coding RNA classes are universally present in all organisms of the three domains of life [18-20]. Further classes of non-coding RNAs, which are also involved in key cellular processes, can be found within the individual domains. Due to their small size (< 300 nt), these non-coding RNA classes are commonly referred to as "small RNAs" (sRNAs). The following paragraphs will focus on the small RNAs that are present in archaea.

Archaea share small nucleolar-like (sno-like) RNAs with eukaryotes, which are divided into C/D box and H/ACA box sRNAs [21, 22]. These molecules guide the site-specific modification of target RNAs, mainly ribosomal RNAs and tRNAs [23]. C/D box sRNAs mediate the 2'-*O*-methylation of RNA [24, 25]. These 50-60 nt long RNAs contain conserved sequence motifs, termed C/C' (5'-RUGAUGA-3') and D/D' boxes (5'-CUGA-3') (Fig. 1.2a) [26, 27]. Paired C/D box elements are located at the RNA termini and form a secondary structure, termed kink-turn (see next chapter) [28, 29]. Another pair is present at the center of the RNA and forms a kink-loop structure [30]. These RNAs furthermore contain two unpaired sequences of 10-12 nt length upstream of the D/D' boxes, termed D/D' guides, which show complementarity to target RNA molecules [31]. The C/D box sRNAs form ribonucleoprotein (RNP) complexes with the archaeal proteins L7Ae, Nop5 and fibrillarin [25]. L7Ae proteins bind to the kink-turn/-loop structures of the RNA and recruit dimers of Nop5-fibrillarin [29, 32-34]. It was shown that archaeal C/D box sRNPs adopt a di-sRNP conformation, with two copies of C/D box sRNAs and four copies of each of the three protein subunits (Fig. 1.2b) [35, 36]. In these complexes, Nop5 serves as a bridge between the two RNAs [36]. The catalytic subunit of C/D box sRNPs is the methyltransferase fibrillarin [25, 37]. The RNA directs the associated fibrillarin to the target RNA via complementary binding of the D/D' guides. Afterwards, the matching nucleotide at the fifth position upstream of the D/D' boxes is positioned into the catalytic center of the fibrillarin protein [36, 38]. A methyl group donated by *S*-adenosyl methionine (SAM) is then transferred to the 2' moiety of the respective ribose [37, 39]. 2'-*O*-methylations favor the 3' endo-conformation of the ribose and block sugar-edge pairings, which increases the rigidity of RNA [40]. These modifications also prevent hydrolysis at the specific ribose [40]. 2'-*O*-

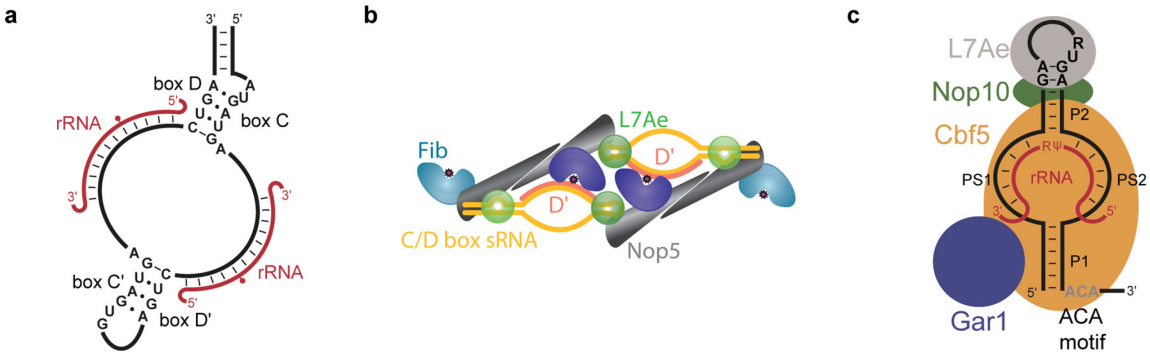


Figure 1.2 Archaeal C/D box and H/ACA box sRNPs. a) The schematic structure of an archaeal C/D box sRNA shows the conserved C/D and C'/D' boxes that form a terminal k-turn and internal k-loop, respectively. Ribosomal RNA (red) is targeted via complementary D/D' guide regions, which are located upstream of the D/D' boxes. The targeted nucleotide for 2'-*O*-methylation is indicated as a red dot (modified from [41]). b) An archaeal C/D box di-sRNP complex consisting of two C/D box sRNAs (yellow) and four copies of L7Ae (green), Nop5 (grey) and fibrillarin (light and dark blue) is illustrated (modified from [36]). c) The schematic structure of an archaeal H/ACA box sRNP is shown. L7Ae (grey) binds to an internal k-loop and Cbf5 (orange) attaches to the ACA motif and the pseudouridylation pocket of the RNA. Cbf5 catalyzes the pseudouridylation (ψ) of an unpaired uridine of the targeted rRNA (red). Nop10 (green) and Gar1 (blue) are attached to Cbf5 [23].

methylations are therefore presumed to increase the stability of rRNAs and tRNAs [42]. This is particularly important at high temperatures and catalytically active regions within the rRNA [40]. In agreement, archaea comprise hot spots for methylations in the ancient core of the ribosome and thermophilic archaea were found to contain increased numbers of C/D box sRNAs and methylation sites (>100) [40, 43].

H/ACA box sRNAs guide the pseudouridylation of RNA [44, 45]. In archaea, they usually consist of a single stem that is immediately followed by a single-stranded H (5'-ANANNA-3') or ACA box (5'-ACA-3') [42]. Within the stem, two unpaired strands form a bipartite loop (Fig. 1.2c) [42]. This loop represents the pseudouridylation pocket. Target RNA forms two 4-8 nt duplexes with the pocket, and a uridine (U) base, located between the two RNA duplexes, remains unpaired [44-46]. The proteins L7Ae, Nop10, Gar1 and Cbf5 are associated with the RNA to form H/ACA box sRNPs [46, 47]. The pseudouridine synthase Cbf5 binds the RNA via the H/ACA boxes [46, 47]. L7Ae recognizes a k-turn structure in the upper stem of the RNA, inducing a conformational change that places the target U into the active site of Cbf5, which catalyzes the pseudouridylation [33, 48, 49]. Nop10 organizes the interaction between L7Ae and Cbf5 and enhances the binding of the target RNA [42]. Gar1 supports the accurate placement of the target U and is necessary for the release of the RNA [49]. Pseudouridylations stabilize RNA structures and the loss of these modifications in rRNA was shown to impair translation [50, 51]. However, pseudouridylation sites are less abundant than 2'-*O*-methylation sites in archaeal ribosomal RNAs [42, 43]. In tRNAs, pseudouridylation can convert non-sense codons into sense codons [52].

The surprising presence of ~20 nt long sRNAs were reported in the thermophilic archaeon *Sulfolobus solfataricus* and suggested to comprise micro RNAs (miRNAs) or small interfering RNAs (siRNA) [53]. These 20-24 nt long sRNAs are components of a gene-silencing machinery, termed RNA interference (RNAi), which was so far only found in eukaryotes [54]. In this system, siRNAs and miRNAs are generated from exogenic (e.g. viral RNA) or endogenic dsRNA by the protein Dicer [54]. Subsequently, these small RNAs guide an interference complex to the invading RNA or mRNA, which prevents translation by cleavage of the targeted RNA. Archaea were found to comprise homologs for several RNAi key proteins, e.g. Dicer-like proteins and Argonaute-Piwi proteins [55]. However, so far no functional archaeal ancestor of eukaryotic RNAi systems was identified [55].

Instead, archaea and bacteria share a different class of small RNAs, the CRISPR RNAs (crRNAs), which show functional analogy to the eukaryotic RNAi machinery. The crRNAs are part of an immune system against mobile genetic elements, termed CRISPR-Cas. They guide a Cas protein interference complex to viral DNA/RNA or plasmid targets for their degradation [56]. In contrast to the RNAi mechanism, crRNAs are transcribed from a CRISPR locus (Clustered Regularly Interspaced Short Palindromic Repeats), which comprises DNA fragments (spacers) of the targeted mobile genetic elements that were integrated by a 'memory process' termed adaptation [56]. The prokaryotic CRISPR-Cas system therefore constitutes an adaptive, heritable immune system.

In the last years, many intergenic sRNAs were identified in archaea [57]. These sRNAs are well-studied in bacteria and play a crucial role in the regulation of gene expression [58]. They are usually produced upon specific environmental stimuli and target mRNAs in *trans* via short, partially complementary sequences [58, 59]. Due to this partial complementarity, a single *trans*-acting sRNA can often bind multiple mRNAs [59]. Targeting by the sRNA usually results in inhibition of mRNA translation, mostly by blocking the ribosomal binding site (RBS) within the 5' UTR of the mRNA (Fig. 1.3a) [60]. However, activation of translation was also found for bacterial sRNAs, e.g. in a process that allows for the unmasking of the RBS (Fig. 1.3b) [60]. So far, only few targets of archaeal sRNAs were identified [57]. The first identified target, the bicistronic MM2441-MM2442 mRNA of *Methanosarcina mazei*, was found to be regulated by blocking the RBS [61]. However, the mRNAs of many archaeal species lack 5' UTRs [62, 63]. It was shown that the 3' UTRs of archaeal mRNAs are important regulators of translation [64]. Translational regulation by a sRNA that targets the 3' UTRs of an mRNA was found in *S. solfataricus* [65]. In addition, regulation can involve interactions of the sRNA with the coding sequence of the mRNA, as recently suggested for a sRNA that is involved in biofilm regulation

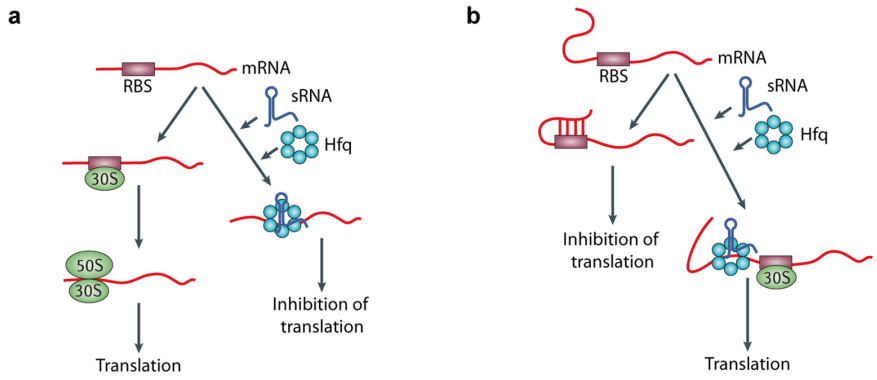


Figure 1.3 Translational regulation by bacterial small RNAs. a) In the translational repression pathway, a sRNA is targeting the ribosomal binding site (RBS) of an mRNA, which blocks the assembly of the ribosome (50S/30S) and inhibits translation. b) Translation of an mRNA is activated by a sRNA that unmasks the RBS and allows for the assembly of the ribosome. In both pathways (a and b), the sRNA-mRNA interactions are facilitated by the Hfq protein (modified from [60]).

in *Sulfolobus acidocaldarius* [66]. In bacteria, a common factor for sRNA-based mRNA regulation is the Hfq protein [60]. This ring-shaped protein acts as a chaperone facilitating interactions between the sRNA and the mRNA. Hfq homologs also exist in archaea, where they are termed LSm proteins (Sm-like proteins) [67]. However, in contrast to the well-studied Hfq protein, the function of the archaeal LSm proteins is largely unknown (see chapter 1.4) [67].

1.3 The kink-turn and its recognition by the L7Ae/L30 protein family

Non-coding RNAs usually comprise highly complex structures which often play crucial roles for the functionality of the RNAs [68]. For instance, the catalytic domain of the 23S/28S rRNA, the peptidyl transferase center, is formed by specific tertiary interactions within the RNA [69]. However, the complexity of RNA structures can be reduced to helical RNA duplexes containing internal loops and bulges (secondary structures) that are connected by different kind of junctions [70]. These junctions determine the trajectories of the RNA duplexes and define the architecture (tertiary structure) of RNA molecules [70]. The kink-turn (k-turn or Kt) represents one of these junctions.

A k-turn introduces a tight kink in the RNA, which bends the RNA helix axis by ~60° [71]. The element was first described as a structural motif in bacterial antiterminators and archaeal rRNA [28, 72]. It soon became clear that k-turns are widespread in many functional RNAs across all domains of life, i.e. rRNAs, C/D and H/ACA box s(no)RNAs, RNase P RNA, U4 snRNA and riboswitches [70]. In addition, SRP RNA was discussed to contain a k-turn fold involving potassium ions, which coined the term K^+-turn [73].

A standard k-turn is characterized by a short stem that is followed by an asymmetric three-nucleotide bulge and two conserved G•A pairs (Fig. 1.4a) [28, 70]. The G•A pairs (positions 1b•1n and 2b•2n) form so-called *trans* sugar edge (G) • Hoogsteen edge (A) interactions, in

which the bases are arranged “side by side” instead of “head by head” [74, 75]. This arrangement rotates the guanine base (G1b) of the first G•A pair out of plane and creates a strong kink in the RNA, which juxtaposes the two minor grooves of the canonical (C) helix (5' of the bulge) and the non-canonical (NC) helix (3' of the bulge) (Fig. 1.4b) [28, 74].

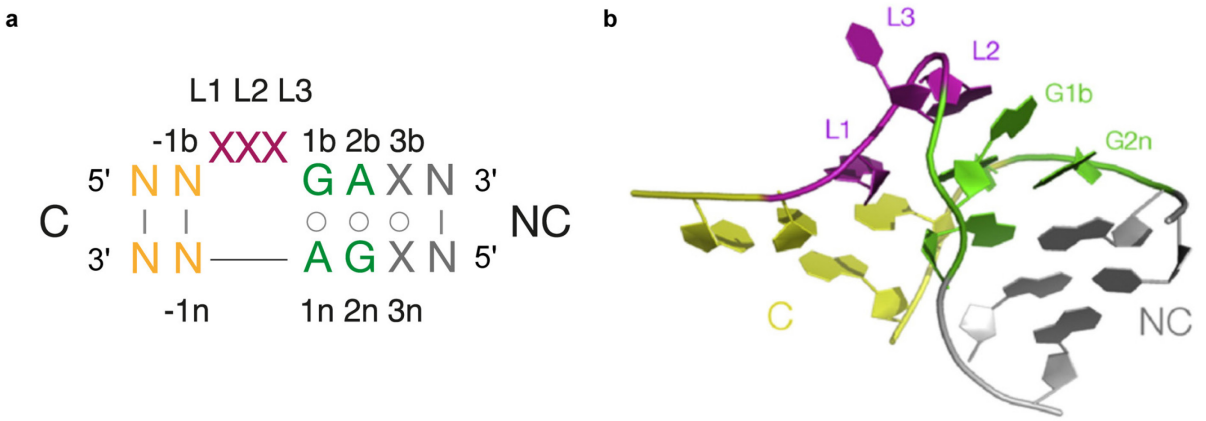


Figure 1.4 Structure of a standard k-turn. a) The consensus sequence of a standard k-turn is indicated. The bulged nucleotides (purple) are labeled by L1/L2/L3 and remaining nucleotides contain positive (3' of the bulge) or negative (5' of the bulge) numbers. The nucleotides of the bulged strand contain the suffix b and those on the non-bulged strand the suffix n. Watson-Crick base pairs of the canonical helix (C) are marked in yellow. The non-canonical helix (NC) comprises two conserved *trans* sugar edge (G) • Hoogsteen edge (A) pairs (green). The bases (3b•3n) involved in ion-induced folding can be paired or unpaired (grey) (modified from [76]). b) The 3D structure of the k-turn is illustrated [70]. The nucleotides are colored according to a).

Three different factors can promote the folding of k-turns into the kinked state [70]: i) tertiary interactions, ii) metal ions and iii) protein binding. Kink-turn folding via tertiary interactions was shown for the SAM-I riboswitch in which a long helix is bent so that its terminal loop can make long-range interactions with another helix to create the SAM binding pocket [76]. This interaction at the ligand binding pocket stabilizes the folded state of the k-turn. Metal-ion induced k-turn folding was shown for both divalent and monovalent metal ions [70]. However, not all k-turns (e.g. in C/D box sRNAs and U4 snRNA) undergo folding in response to metal ions. The base pair following the second G•A pair (position 3b•3n) was identified as a key determinant for ion-induced folding [77]. A Watson-Crick base pair (or G•U) at this position leads to an inability to fold upon metal ion addition, whereas 3n = G and 3b = C (but not both) permit folding. This rule has strong predictive value for k-turn formation in the presence of ions [77]. Finally, most k-turns serve as protein binding sites [71]. Diverse ribosomal proteins of highly varying structure and manner of binding were found to interact with k-turns [28]. However, the proteins of the L7Ae/L30 family are presumed to constitute the most ancient k-turn binding proteins [71].

Members of the L7Ae/L30 protein family are found in Bacteria, Archaea and Eukarya. Prominent examples are the archaeal and eukaryotic proteins rpL7A(e) and rpL30(e), which are

components of the large ribosomal subunit, the human/yeast proteins 15.5 kDa/Snu13p (C/D box snoRNPs/ U4 snRNPs), NHP2/Nhp2p (H/ACA box snoRNPs), Pop3/Pop3p and Rpp38/Rpp38p (RNase P and MRP RNPs), SBP2 (SECIS mRNA), as well as the bacterial proteins YbxF and YlxQ [78-85]. Crystal structures revealed that all members show structural homology and comprise a highly similar mode of k-turn interaction [84, 86-88]. L7Ae recognizes the k-turn via two RNA binding interfaces [88]: i) a highly basic β-strand:turn:α-helix and ii) a short hydrophobic loop. The α-helix of the first interface enters the widely accessible major groove of the NC helix and two conserved asparagine (N) and glutamic acid (E) residues form hydrogen bonds with the G1b and G2n bases (Fig. 1.5a). Additional residues of the α-helix interact with the phosphate backbone of the NC helix. The second interface, the hydrophobic loop, makes van-der-Waals contacts with the bulged nucleotides L1 and L2 and an E side chain forms hydrogen bonds with L1. These elements ensure a highly specific recognition of the k-turn structure [70]. A special feature of archaeal L7Ae is its capability of recognizing k-loop sequences, e.g. in C/D box sRNAs (see chapter 1.2), in which the C-helix is replaced by a terminal loop [30].

The main function of L7Ae proteins is the structuring of non-coding RNAs by stabilization of k-turn sequences. While eukaryotes comprise different L7Ae homologs for specific small RNA classes, archaeal L7Ae was found to be a multifunctional RNA-binding protein. It is a ribosomal protein that stabilizes domain I of the 23S rRNA by binding Kt-15 and interacting with the ribosomal protein L15e [28, 79]. Within C/D and H/ACA box sRNPs, archaeal L7Ae is not only crucial for the structuring of the RNA, but also for the recruitment of associated proteins (see chapter 1.2). L7Ae was furthermore identified as a fifth subunit of the RNase P RNP [89, 90]. It recognizes two k-turns in RNase P RNA and binding enhances the optimum temperature of the tRNA maturation complex [90, 91]. Finally, contacts with the SRP RNA from *S. solfataricus* were suggested [92]. In contrast to the many functions of archaeal L7Ae, the archaeal L30e was so far only identified as a subunit of the 50S ribosome [79].

Due to their high affinity, specificity, stability and defined structure, archaeal L7Ae-k-turn interactions were utilized as RNP modules in synthetic biology and biotechnological approaches. Saito and co-workers designed a synthetic ON/OFF switch to control the translation of an output protein in human cells using a C/D box k-turn and archaeal L7Ae (Fig. 1.5b) [93]. This tool was further developed to function in complex synthetic circuits for human cell fate control or feedback regulation of proteins in mammalian cells [94, 95].

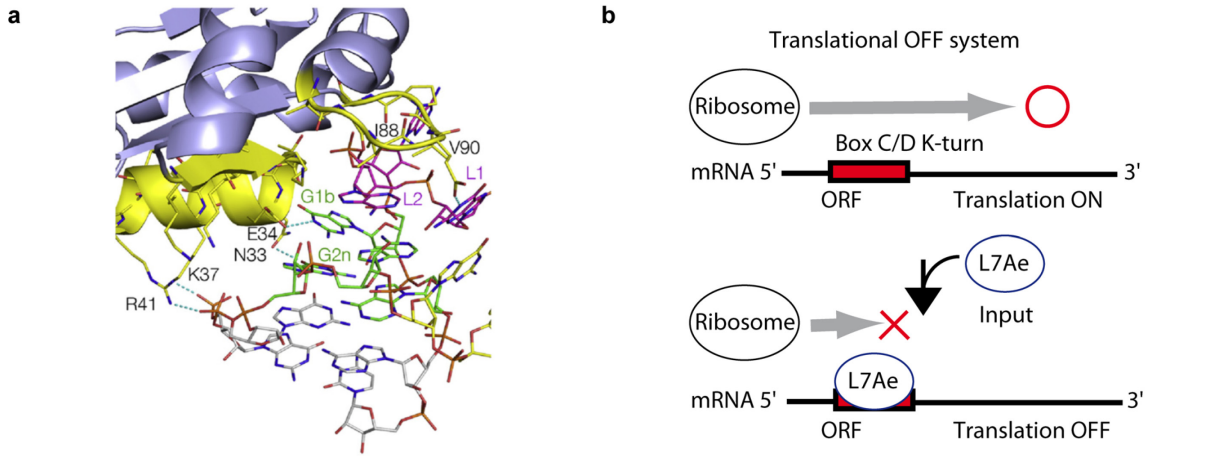


Figure 1.5 L7Ae-k-turn interaction and its utilization in synthetic translational regulation. a) The crystal structure of *A. fulgidus* L7Ae highlights interactions with ribosomal Kt-7 of *H. marismortui* [88]. The RNA binding interfaces of L7Ae are shown in yellow. Conserved amino acids (N33, E34, K37 and R41) within the α-helix of the β-strand:turn:α-helix interface make contacts with the G1b, G2n bases and the phosphate backbone of the NC helix. The amino acids of the hydrophobic loop (I88, E89 and V90) interact with the bulged nucleotides L1 and L2 [70]. b) L7Ae-k-turn interactions were utilized in a synthetic system for translational regulation in eukaryotes. In the OFF system, L7Ae targets a k-turn sequence that was integrated into the coding sequence of an mRNA. Binding of the k-turns blocks the ribosome and prevents translation [93].

1.4 The Sm superfamily and the elusive function of the archaeal LSm proteins

The archaeal LSm (Sm-like) proteins are members of the Sm superfamily, which include the eukaryotic Sm/LSm proteins and the bacterial Hfq protein [67]. These proteins are defined by their Sm fold, a tertiary structure that consists of five anti-parallel β-sheets and an N-terminal α-helix [96]. The Sm fold allows the formation of doughnut-shaped rings, a characteristic feature of the family [97]. Sm family proteins are involved in key processes of RNA metabolism and were shown to bind uridine-rich RNAs [97]. However, despite their structural homology and RNA-binding capacity, the assembly and functional roles of the Sm family members differ fundamentally between proteins of the three domains of life.

The bacterial Hfq protein was already discovered in the late 1960s as an essential host factor of bacteriophage Qβ in *Escherichia coli* [98]. Due to the lack of important sequence signatures that identified eukaryotic and archaeal family members, the realization that Hfq belongs to the Sm superfamily came relatively late (around 2002) when first structural analyses revealed the typical Sm fold and the formation of a homohexameric ring [99]. The Hfq protein functions as an RNA chaperone and plays a role in multiple RNA processes [60, 67]. First, it regulates mRNA translation by facilitating the interaction of sRNAs with target mRNAs (see chapter 1.2). Hfq changes the conformation of the bound sRNA by exposing "seed" sequences, which enables the targeting of the complementary target mRNA sequence [59, 100]. Second, Hfq can initiate degradation of the targeted mRNA by recruiting the ribonuclease RNase E, which degrades both the sRNA and mRNA [101]. Third, Hfq directly influences the stability of mRNAs by stimulating their polyadenylation, which triggers mRNA decay by exonucleases in

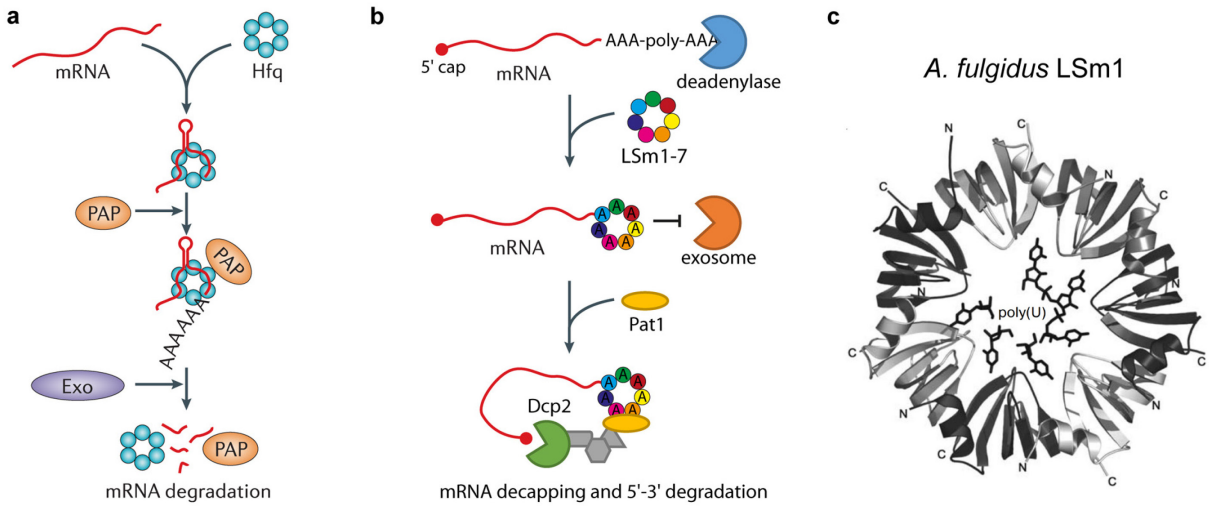


Figure 1.6 The role of Hfq and LSm1-7 in mRNA decay. a) In bacteria, Hfq stimulates the polyadenylation of mRNAs by interacting with the poly-A polymerase I (PAP). Polyadenylation activates exonucleases (Exo), which degrade the mRNA [60]. b) In eukaryotes, co-transcribed poly-A tails are shortened to oligo-A termini by a deadenylase. The oligo-A ends are bound by LSm1-7, which blocks 3'-5' degradation by the exosome. Pat1 interacts with LSm1-7 and recruits decapping factors that initiate RNA degradation in 5'-3' direction (modified from [60, 96]. c) The crystal structure of *A. fulgidus* LSm1 with a bound oligo-U RNA substrate is shown [102]. One uridine residue binds to one subunit of the complex. The RNA binds to the proximal face of the ring [103].

bacteria (Fig. 1.6a) [104, 105]. Hfq recognizes RNA via four different binding sites [106]. The proximal face is the predominant binding site for sRNAs and binds to uridine tracts at the 3' end of the RNA (3'-OH at the terminus) [107]. These RNA termini result from Rho-independent transcriptional termination observed for most sRNAs in bacteria [60, 108]. Each Hfq subunit binds to one uridine base [99, 107]. The distal face is the main binding site for mRNAs and recognizes A-rich sequences [106]. Gram-negative bacteria detect triplets ($[A\text{-}A\text{-}N]_n$ for *E. coli* Hfq), whereas gram-positive bacteria recognize every second adenine ($[A\text{-}N]_n$ for *Staphylococcus aureus* Hfq) [109]. The lateral face of the ring, the rim, is a further binding site for internal UA-rich sequences of the bound sRNA. A recent structure of full-length RydC sRNA bound to Hfq showed that the rim binds to the seed sequence of the sRNA, facilitating its interaction with the mRNA target [100]. Finally, the C-terminal domain, which shows the highest variability among Hfq proteins, was recently also found to contain a binding site for some sRNAs [106]. Due to the limiting amount of Hfq molecules in the cell and the high number of RNA substrates, the protein is thought to act like a hub, which performs active cycling of RNAs on its surface [110]. Once sRNA and mRNA are basepaired, Hfq is assumed to be released [60]. The broad role of Hfq in the cell is reflected by the pleiotropic effects that arise after deletion of the *hfq* gene [111]. However, Hfq is not present in all bacteria. Only around 50 % of all bacteria appear to contain the protein [60]. Proteins that comprise a FinO-domain, e.g. *E. coli* ProQ, were recently found to act similar as Hfq by mediating the interaction of small RNAs with target mRNAs [112].

Eukaryotes contain the most diverse Sm family with over 20 members, which are divided into the Sm and LSm subfamilies [96]. The Sm proteins were early discovered as core proteins of the spliceosome, which is a complex RNP-based machinery that performs intron excision in eukaryotic pre-mRNAs [113]. Seven Sm proteins assemble around internal U residues (5'-AUUUUUG-3') of a U-rich small nuclear RNA (U snRNA), building a heteroheptameric ring [114]. The ring serves as a scaffold for snRNP-specific proteins to generate mature U snRNPs [67]. Several U snRNPs (U1, U2, U4/U6·U5) assemble into the major spliceosome, which is involved in the maturation of most mRNAs [67]. The LSm subfamily contains up to 16 members and, similar to the Sm proteins, these proteins form heteroheptamers [97]. However, LSm rings differ from the Sm rings as they can form in the absence of RNA and, like Hfq, bind to the U-rich 3' end of RNAs using the proximal face of the ring [96]. Two LSm rings exist in eukaryotes: i) LSm2-8 and ii) LSm1-7. LSm2-8 is found in the nucleus and is also involved in mRNA splicing [96]. It is the core component of the U6 snRNP, which fulfills key reactions during splicing [115]. LSm2-8 binds to the processed end of the U6 snRNP and has a clear preference for a 3' terminal 2',3'-cyclic phosphate [116]. LSm1-7 is located in the cytoplasm and acts as a major regulator of mRNA decay (Fig. 1.6b) [117]. In eukaryotes, mRNAs are generated with long poly-A tails. Degradation of mRNA is initiated by a deadenylase, which trims the poly-A tail to a length of 10 adenines or less. LSm1-7 specifically binds these oligo-A tracts, presumably by the proximal face of the ring, and blocks 3'-5' decay by the exosome [96]. The protein serves as a scaffold for Pat1, which recruits decapping factors that decap the 5' terminus and induce mRNA degradation in 5'-3' direction [118]. The eukaryotic LSm proteins were furthermore found to interact with pre-RNase P RNA and snoRNAs and assumed to be involved in their biogenesis [119-121]. Due to this interaction, the proteins were also suggested to play a role in tRNA and rRNA maturation and processing [119, 120, 122].

The archaeal LSm proteins were identified due to their high sequence similarity with the described eukaryotic Sm/LSm proteins [115, 119]. With few exceptions, LSm proteins can be found in all archaea [67, 123]. Some euryarchaeal species, e.g. halophiles, contain only one LSm protein, many species contain two, whereas Crenarchaeota and Thaumarchaeota were found to comprise a third LSm member, which constitutes an extended version with a long C-terminal domain [67]. Archaeal LSm structure elucidation preceded their functional analysis. Similar to the eukaryotic Sm/LSm proteins, the archaeal LSm proteins form heptameric rings, and similar to Hfq, these rings are formed by single LSm subunits (homomeric) which assemble in the absence of RNA [102, 124-126]. LSm2 of *Archaeoglobus fulgidus* was found to form hexamers in the absence of RNA and at low pH (4.5), but was shown to assemble into stable

heptamers after the addition of oligo-U RNA [126, 127]. The crystal structure of the augmented LSm3 protein from *Pyrobaculum aerophilum* revealed the assembly of a 14-mer in the presence of divalent metal ions that consisted of two head-to-tail stacked LSm cores and an equatorial ring formed by the C-terminal domain [128]. The protein formed supraheptameric oligomers $[LSm3]_{n=multiples\ of\ 7}$ in solution. In general, archaeal LSm proteins (but also other members of the Sm superfamily) were found to form higher order assemblies including long fibers [129, 130]. However, the single-ring LSm proteins are generally considered to be the biologically relevant units [67]. Co-crystal structures of oligo-U RNA and LSm1 of *A. fulgidus* and *Pyrococcus abyssi* revealed that LSm1 binds uridine-rich RNA in a similar fashion to Hfq and eukaryotic LSm (Fig. 1.6c) [102, 131]. The RNA is bound within the inner pore of the proximal face and one uridine residue is complexed per subunit. However, despite the well-characterized structural features of the archaeal LSm proteins, the function of these RNA-binding proteins remains largely elusive. Similarities with eukaryotic and bacterial features raise the question if the archaeal LSm proteins act rather as chaperones (like Hfq) or as scaffolds (like eukaryotic Sm/LSm) [67]. Archaea do not comprise intron-containing mRNAs and therefore do not contain a spliceosome. Immunoprecipitation (IP) revealed that both LSm proteins from *A. fulgidus* interact with RNase P RNA (like eukaryotic LSm) [102]. In this study, LSm1 and LSm2 also co-immunoprecipitated the respective other protein, indicating that the two proteins interact *in vivo*. In a similar approach, the single LSm1 of *Haloferax volcanii* was found to interact with 20 sRNAs, including intergenic sRNAs, antisense sRNAs, one C/D box sRNA and SRP RNA [132]. EMSA studies showed that the protein also binds to tRNAs, suggesting a role in tRNA processing. A knock-out mutant of this protein was viable. Deletion of conserved amino acids of LSm1 showed differential gene expression, e.g. in motility genes. In congruence, the produced mutant exhibited a gain of function in swarming [133]. Recently, high-throughput sequencing of co-purified RNAs of LSm1 and LSm2 from *S. solfataricus* was performed. A high number and a large variety of archaeal small RNAs were identified. A role in tRNA/rRNA processing and s(no)RNA biogenesis was suggested [134]. Moreover, many interacting mRNAs were identified. Very recently, LSm1 and LSm2 of *S. solfataricus* were found to interact with proteins of the archaeal exosome, which degrades mRNA in 3'-5' direction [135]. The proteins were found to interact directly with DnaG, an accessory exosome subunit that binds poly-A tails. Overexpression of the two LSm proteins resulted in elevated levels of A-rich tails, indicating a function resembling Hfq activity. An Hfq-like function regarding small RNA-mediated mRNA regulation was, however, not yet shown for the archaeal LSm proteins. Both LSm proteins from *M. mazei* showed no significant binding affinity for $sRNA_{162}$, a sRNA

regulating the bicistronic transcript MM2441-MM2442 (see chapter 1.2) [61]. Interestingly, the archaeon *Methanocaldococcus jannaschii* was found to contain an Hfq-like protein instead of an LSm protein [123]. The protein forms an Hfq hexamer and was found to interact with different *E. coli* sRNAs. Furthermore, it could be shown to be functional in RhyB-mediated sodB mRNA degradation in *E. coli*. The protein also complemented the pleiotropic phenotype of an *E. coli* Δ*hfq* strain. Similar observations were made for Hfq from *M. jannaschii* in *S. aureus* [136].

1.5 The model organism *Sulfolobus acidocaldarius*

Sulfolobus acidocaldarius is a thermoacidophilic archaeon of the phylum Crenarchaeota. The strain DSM639, used in this work, was isolated from a hot acid spring at the Yellowstone National Park, USA [137]. This obligate aerobic archaeon grows at a temperature of 75-80°C and a pH optimum of 2-3. It can either grow chemolithotrophically by oxidization of elemental sulfur utilizing CO_2 as a carbon source, or chemoorganothropically using organic compounds as sugars or peptides. The ability for genetic manipulation established *S. acidocaldarius* as an archaeal model organism in the past 15 years. The possibility for positive selection (uracil auxotrophy) or counterselection (by 5-FOA) allowed the development of versatile genetic tools that can be used for markerless gene knockouts, genomic mutations (e.g. tags or point mutations), gene replacements, regulated gene expression, protein overproduction and reporter gene studies (using lacS) [138, 139].

S. acidocaldarius DSM639 contains a single *l7ae* gene (*saci_1520*) and the three *lsm* genes, *lsm1* (*saci_1224*), *lsm2* (*saci_0799*) and *lsm3* (*saci_0660*). The *l7ae* and *lsm1* genes exist as individual genes. The downstream transcription factor *lrs14* (*saci_1223*) overlaps *lsm1* in reverse direction. The *lsm2* and *lsm3* genes form dicistronic mRNAs with *matT* (*saci_0798*) and a tRNA ribosyltransferase gene (*saci_0659*), respectively.

1.6 Aims of this work

Kink-turn motifs are widespread in archaeal RNA molecules. L7Ae is known to be a multifunctional k-turn RNA binding protein, but a global analysis of its interaction partners has not yet been performed. This study therefore aims to identify novel RNA binding partners of archaeal L7Ae. Another aim of this work is to shed light on the elusive function of the archaeal LSm proteins. The identification of the RNA interaction partners, the RNA binding regions and the investigation of the essentiality of the LSm proteins is expected to provide clues about their role in archaea. To achieve these goals, three RNA sequencing (RNA-Seq) data sets will be generated. For the first data set, the sRNome, isolated small RNAs of *S. acidocaldarius* will be sequenced to obtain an overview of their abundance. For the second and third data sets, the L7Ae/LSm-RNA interactomes, the *l7ae* and *lsm* genes of *S. acidocaldarius* will be genomically tagged to allow for immunoprecipitation of the proteins and to co-isolate bound RNA. RNA immunoprecipitation sequencing (RIP-Seq) will be performed to identify the interaction partners of these proteins. Exemplary RNA-protein interactions of selected partners will be studied in more detail using EMSAs and GFP reporter assays. The potential RNA binding sites of the LSm proteins will be analyzed by computational analysis of the sequenced binding partners. Furthermore, deletion of the *lsm* genes will be attempted using a novel CRISPR-Cas genome editing approach [140].

2. Results

2.1 The small RNome of *S. acidocaldarius*

Previous studies revealed that L7Ae and LSm proteins bind to a large variety of small RNAs in archaea [92, 132, 134]. In order to obtain a global overview of the sRNA species that are present in *S. acidocaldarius* and to investigate the transcriptional level of potential binding partners of L7Ae and LSm proteins, the small RNome (complete set of sRNAs) of this organism was determined.

2.1.1 Identification of the small RNome of *S. acidocaldarius* by small RNA-Seq

Small RNA-Seq analysis was performed to identify the sRNome of *S. acidocaldarius* during optimal laboratory growth conditions. Total and small RNAs were isolated from logarithmically grown *S. acidocaldarius* MW001 cells and the sRNA fraction was used to prepare cDNA libraries (Fig. 2.1a). Typically, small RNAs range in size between 50 and 300 nt [57]. However, RNAs of approximately 20 nt length were reported in *Sulfolobus solfataricus* that are assumed to constitute RNAi-associated small interfering RNAs (siRNAs) and microRNAs (miRNAs), which so far have only been identified in eukaryotes [53]. To include these potential small RNA molecules, the prepared cDNA libraries were differentially selected by size: i) one duplicate contained RNA inserts of 50-150 nt (narrow range) and ii) one duplicate encompassed RNA inserts between 0-150 nt (wide range) (Fig. 2.1b). The second data set should therefore comprise the presumed RNAi-associated small RNAs, given that also *S. acidocaldarius* produces this type of RNAs.

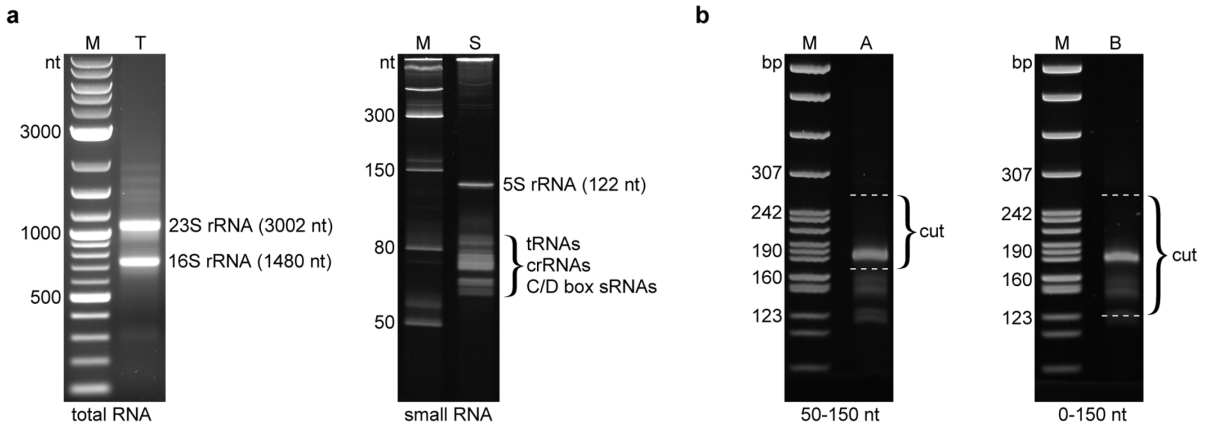


Figure 2.1 Isolation of *S. acidocaldarius* RNA and preparation of cDNA libraries. a) The left panel shows a native agarose gel of a total RNA (T) sample that is depleted of small RNAs. Two definite bands represent the 16S and 23S rRNAs. Their migration behavior reflects the compact folding of ribosomal RNA from hyperthermophilic organisms. The right panel shows a small RNA (S) isolate that was separated via denaturing PAGE. Bands of 60-120 nt represent different species of small RNAs from *S. acidocaldarius.* b) The PCR-amplified cDNA libraries were size-selected between 170-270 bp (A – 50-150 nt RNA inserts) and 120-270 bp (B – 0-150 nt RNA inserts) by gel extraction from native PAGE, respectively.

The cDNA libraries were sequenced by Illumina HiSeq2500 methodology. In total, 47 million sequencing reads were mapped onto the *S. acidocaldarius* genome (see Table 4.23 in chapter 4.9.1). First, the highest peaks of the obtained coverage plots were identified (Fig. 2.2). These peaks represent the most abundant RNAs in the data sets. Almost every peak constituted a C/D box sRNA, demonstrating their high abundance in *S. acidocaldarius.* Interestingly, a single C/D box sRNA, Sac-sR10, was found to be the most abundant RNA in both data sets. Every fifth sequenced read represented this RNA. One guide region of this RNA was proposed to target numerous tRNAs (e.g. Gly-CCC, Pro-CGG, Pro-GGG), which might explain the requirement for the observed large amounts of Sac-sR10 [21]. Other highly abundant RNAs are the 5S rRNA and a single CRISPR RNA (crRNA). Direct comparison of the two differently size-selected sRNome plots revealed that only a single highly abundant RNA smaller than 50 nt was present in the wide-range selected data set (Fig. 2.2; bottom track). This 16 nt long RNA constitutes a fragment of a 63 nt antisense RNA, which was identified in a previous RNA-Seq approach, where it showed resistance to RNase R treatment [141]. It was therefore termed RNase R resistant RNA (RrrR RNA). The RrrR RNA is a double-stranded non-coding RNA which is composed of a 125 nt long sense strand and the 63 nt antisense (as) strand. In both sRNome data sets, excessive amounts of the antisense strand were found to overlap with the sense strand (the detailed mapping pattern of the RrrR RNA of the sRNome analysis is illustrated in Fig. 2.27b).

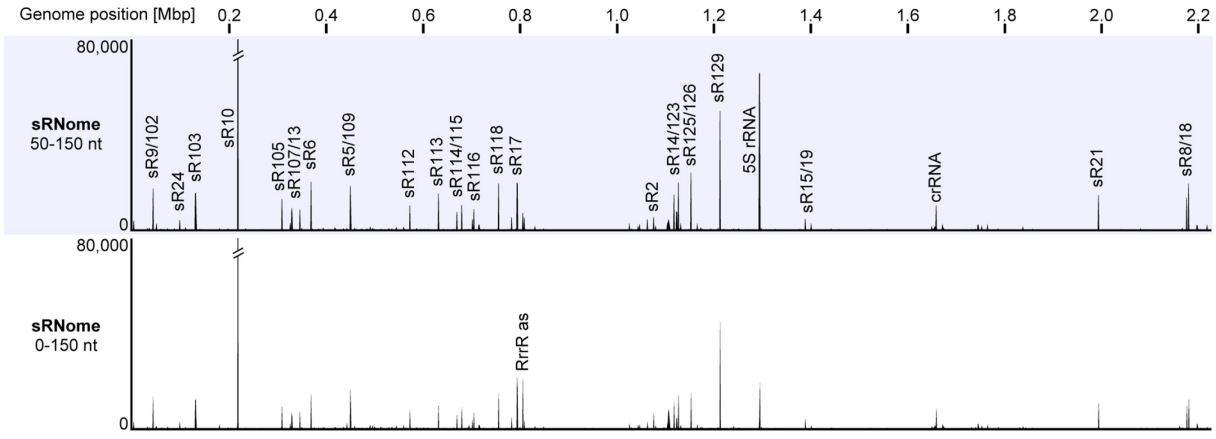


Figure 2.2 Overview of the most abundant small RNAs in *S. acidocaldarius*. The plots display the coverage (y-axis) of sequencing reads of the narrow- (top track) and wide-range (bottom track) selected sRNome data sets that were mapped onto the *S. acidocaldarius* genome (x-axis). Peaks represent the most abundant RNAs in the data sets and were labeled for the top track. C/D box sRNAs were labeled with (Sac-)sR# according to previous studies [21, 40]. The same RNA peaks are also present in the bottom track. For clearness, the only highly abundant RNA, which is found for the wide-range selected sRNome data set, was labeled in the bottom track (RrrR antisense (as)). The Sac-sR10 RNA constitutes the most abundant RNA in both data sets (covered by 200,000 reads). One million reads were mapped per data set.

The two RNA-Seq data sets were screened for sRNAs that contained at least 100 reads within one million mapped reads to identify the sRNome of *S. acidocaldarius*. This detailed analysis revealed 405 sRNAs that were produced during laboratory growth conditions (Appendix 1): 48 tRNAs, 1 SRP RNA, 1 RNase P RNA, 1 5S rRNA, 62 C/D box sRNAs, 2 H/ACA box sRNAs, 223 crRNAs, 46 antisense RNAs and 21 sRNAs of unknown function.

The coverage of the tRNAs was expectedly low in the RNA-Seq data sets, as their high modification rate effectively impairs the cDNA synthesis reaction of the reverse transcriptase [142]. In addition, the essential SRP and RNase P RNA molecules showed low coverage, which is likely caused by their exclusion during the cDNA size selection step as both RNAs have a length of around 300 nt. The 62 C/D box sRNAs showed a large variability in their relative abundance and their genes often overlapped with protein-coding genes or were found in close vicinity of adjacent genes. Recently, the co-transcription with these genes was shown, which implies a processing step of the transcript to produce mature C/D box sRNAs [143]. The two H/ACA box sRNAs were identified due to their homology with known H/ACA box sRNAs of other archaeal species. Additional H/ACA box sRNAs might be present in *S. acidocaldarius* but missed due to low RNA-Seq coverage. Except of a single crRNA, other crRNAs showed low abundance, which is due to the unusual RNA termini (5'-OH and 2',3'-cyclic phosphate) that occur during their maturation. Without a polynucleotide kinase (PNK) treatment step, these ends cannot be adapter ligated during cDNA library preparation. *S. acidocaldarius* comprises four CRISPR arrays which are associated with different reported CRISPR-Cas systems: type I-A, type I-D, type III-A, type III-B and type III-D [144-146].

Next, the mapped reads of the wide-range selected data set were filtered for read lengths of 19-25 nt to search for potential miRNA candidates, which typically contain a size of around 22 nt [147]. About 12 % of the mapped reads corresponded to this range. The highest peaks referred to reads of the antisense RrrR RNA and 5S/16S/23S rRNAs (Fig. 2.3a). Other high peaks constituted C/D box sRNA and tRNA fragments. The fragments were strongly biased towards the 5' or 3' termini of the RNAs. The peaks encompassed read lengths of the selected range and showed no overrepresentation of 22 nt fragments. Phased miRNA patterns were reported for several genes (*sso_0297; sso_1088; sso_0282; sso_0980*) in *S. solfataricus* [53]. Thus, the mappings of the homologous genes in *S. acidocaldarius* (*saci_0180; saci_1210; saci_0666*) were analyzed but did not reveal uncommon read patterns, which is exemplified by the coverage plot of the *sso_0282* homologue *saci_0666* (Fig. 2.3b). The genes displayed very similar distribution of mRNA fragment reads as other genes. Therefore, no evidence for the production of RNAi-associated small RNAs could be found in *S. acidocaldarius.*

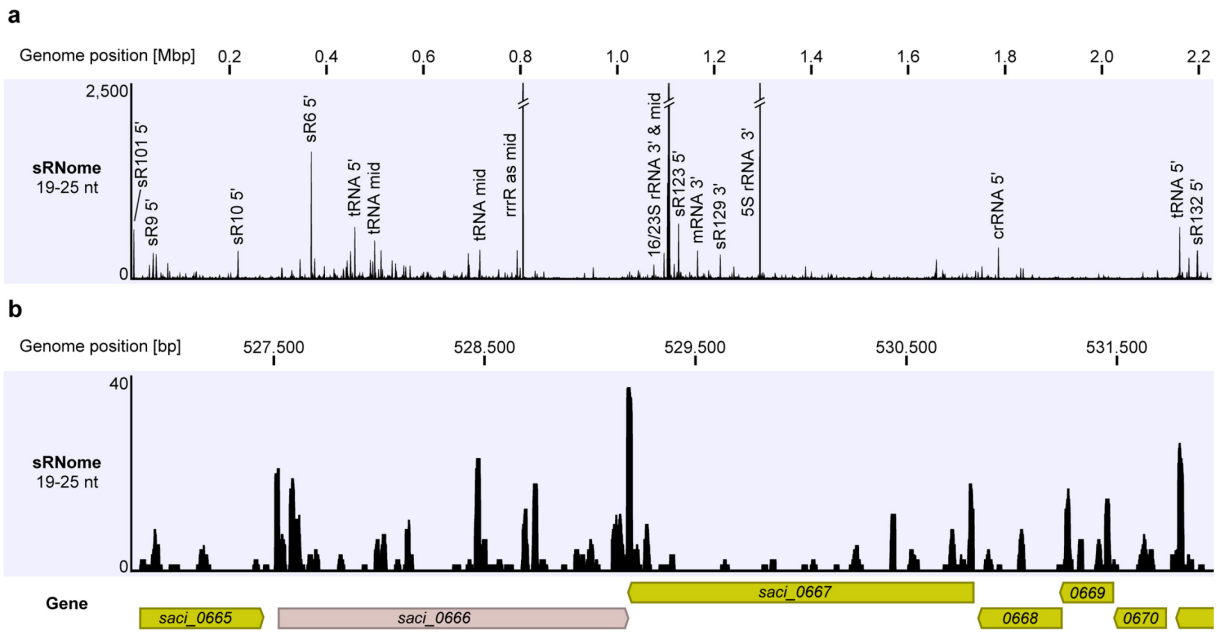


Figure 2.3 Mapping pattern of 19-25 nt sRNome reads. a) The coverage plot of the wide-range selected sRNome library is illustrated, which was filtered for read lengths of 19-25 nt. 115,827 reads were mapped onto the *S. acidocaldarius* genome. Labeled peaks show the identity of the most abundant reads. Peaks representing C/D box sRNAs were labeled with (Sac-)sR#. The location of the mapped reads within the identified RNAs is indicated (5': 5' terminus; mid: within RNA; 3': 3' terminus). b) The image shows the read mapping pattern of the *saci_0666* locus and the surrounding genes.

2.2 The L7Ae-RNA interactome of *S. acidocaldarius*

After obtaining an overview of the small RNAs that were produced in *S. acidocaldarius*, the aim was to identify RNAs that were bound by L7Ae. This goal required the isolation of the L7Ae protein from *S. acidocaldarius*. RNA-immunoprecipitation sequencing (RIP-Seq) was utilized for this purpose, in which the target protein is isolated from cell extracts via immunoprecipitation and the co-immunoprecipitated RNAs are analyzed by high-throughput RNA sequencing.

2.2.1 Immunoprecipitation of genomically tagged L7Ae from *S. acidocaldarius*

In order to purify the L7Ae protein and its RNA interaction partners, the single *l7ae* gene (*saci_1520*) of *S. acidocaldarius* was genomically fused to a C-terminal Flag-HA tag. This allowed for its purification via two consecutive immunoprecipitation steps. Immunoprecipitation (IP) yields high purity of the target protein, due to the specificity of the utilized antibodies for their epitope. The growth of the recombinant strain was monitored to investigate whether the presence of the tag influences the function of this essential protein. No differences in growth were observed for the L7Ae-CFHA strain in comparison to the reference strain MW001 (Fig. 2.4).

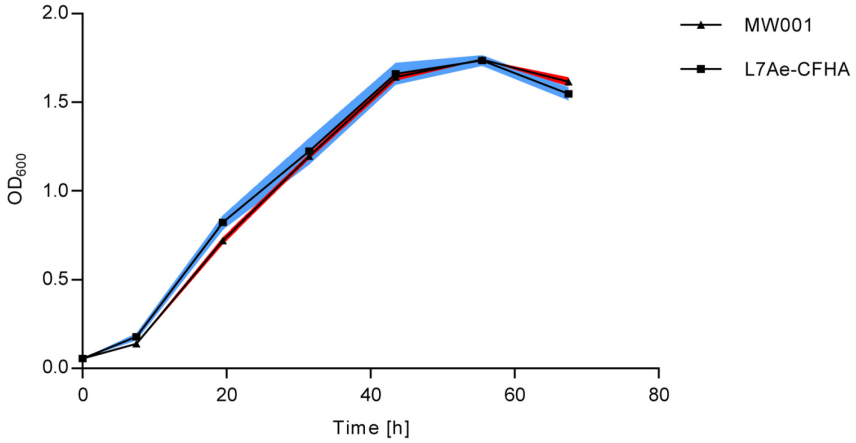


Figure 2.4 Growth study of the *S. acidocaldarius* strain with Flag-HA tagged L7Ae. The growth behavior of the *S. acidocaldarius* MW001 reference strain and the genomically Flag-HA tagged L7Ae strain (L7Ae-CFHA) is shown. Error bars (color filled area) demonstrate the standard deviation of three biological replicates.

L7Ae was purified from logarithmically grown cells via the Flag-HA tandem purification kit. The elution fractions displayed several protein bands. The three predominant bands constituted L7Ae and its C/D box sRNP interactors Nop5 and fibrillarin, verified by mass spectrometry (Fig. 2.5a). Other identified proteins included ribosomal proteins and Cbf5, the catalytic subunit of H/ACA box sRNP complexes. The specificity of the purification was determined by purifying a wild-type cell extract (MW001 with untagged *l7ae*) using identical protocols. This mock purification showed no isolated protein (Fig. 2.5a). The urea eluate of the L7Ae purification contained large amounts of co-immunoprecipitated RNA species that ranged in size from less than 50 nt to more than 1000 nt, whereas no visible amounts of RNA were detected for the mock purification (Fig. 2.5b). Northern blot analysis verified the presence of C/D box sRNAs in the eluate by using a probe against the highly abundant C/D box sRNA Sac-sR10, which was previously identified in the sRNome (Fig. 2.5c).

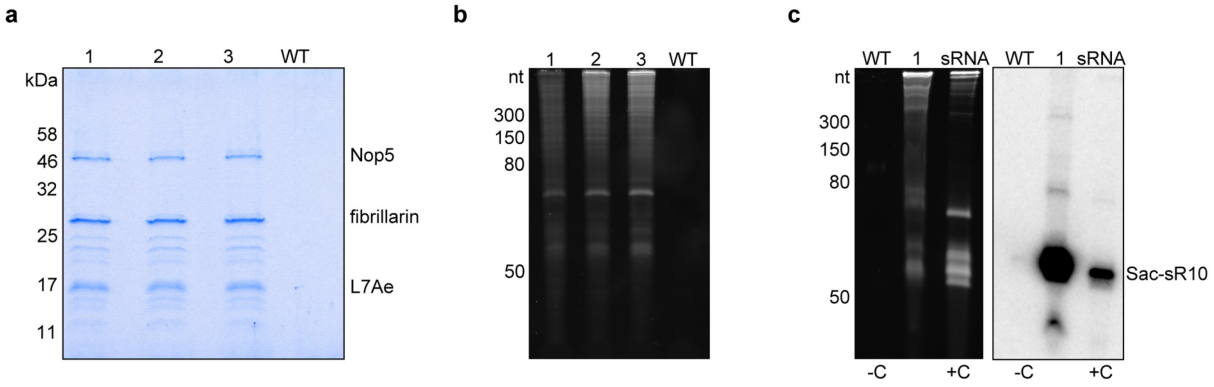


Figure 2.5 Co-immunoprecipitation of L7Ae-bound proteins and RNA. a) SDS-PAGE analysis illustrates the final purification fractions of L7Ae (16 kDa) after Flag- and HA-IP. Lanes 1-3 are three technical replicates and show the co-immunoprecipitation of the L7Ae-interacting proteins Nop5 (47 kDa), fibrillarin (25 kDa) and other proteins. The mock purification (WT) displays no protein bands. b) Denaturing PAGE shows the co-immunoprecipitated RNAs from a). L7Ae-interacting RNAs in varying sizes can be observed in lanes 1-3, whereas no nucleic acids are visible in the mock purification (WT). c) The left panel shows the denaturing PAGE of L7Ae co-immunoprecipitated RNA (1). The mock purification (WT) and small RNA isolation (sRNA) from *S. acidocaldarius* were applied as negative (-C) and positive (+C) controls, respectively. The gel was used for northern blot analysis (right panel). A radioactive probe against Sac-sR10 (58 nt) was utilized for C/D box sRNA detection. Signals are only visible in the L7Ae sample and the positive control.

The co-immunoprecipitated RNA of L7Ae was used to generate cDNA libraries. Illumina HiSeq2500 technology generates read length of up to 100 nt. In order to sequence L7Ae-bound RNA above 100 nt length, aliquots of the extracted RNAs were fragmented by $ZnCl_2$ treatment as described by Dominissini *et al*. (Fig. 2.6a) [148]. Bivalent metal ions can cleave RNA acting as Brønsted bases, which abstract a proton from the 2'-OH of the ribose [149]. This causes a nucleophilic attack of the phosphodiester bond, resulting in 2',3'-cyclic phosphate and 5'-OH RNA termini. After fragmentation, the RNA was treated with T4-PNK to cure the termini of the RNA, which allows proper adapter ligation during cDNA library preparation. Three kinds of cDNA libraries were prepared for Illumina sequencing (Fig. 2.6b): i) a duplicate prepared from un-fragmented RNA, containing L7Ae-interacting small RNAs (L7Ae sRNA), ii) a duplicate prepared from fragmented long RNAs (L7Ae frag RNA) and iii) one library prepared from a mock purification (WT control). Although no detectable amounts of RNA were isolated in the mock purification, PCR amplification of residual background RNA yielded sufficient amounts for Illumina sequencing during the cDNA library preparation (Fig. 2.6b).

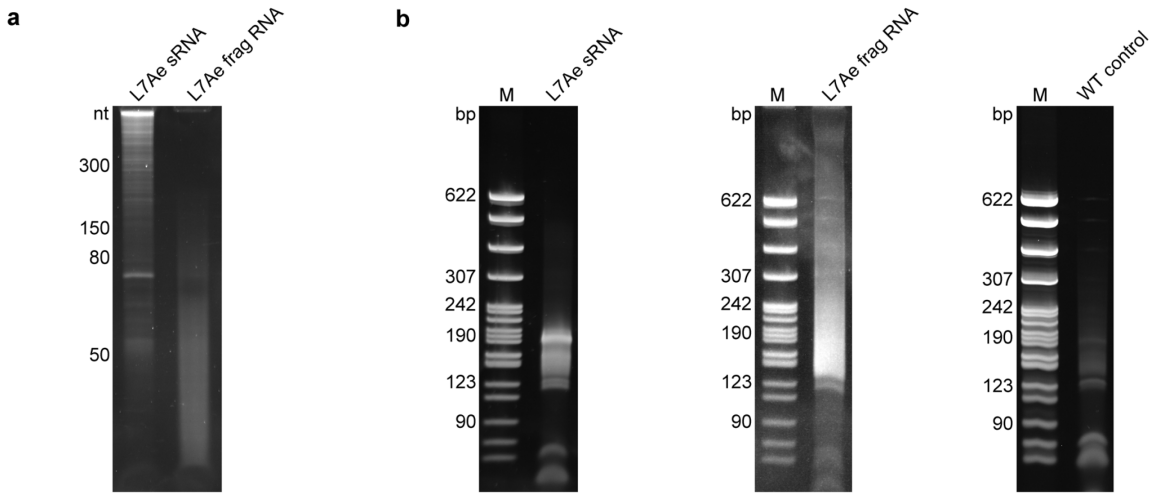


Figure 2.6 $ZnCl_2$ fragmentation and cDNA library preparation. a) Denaturing PAGE of L7Ae co-immunoprecipitated RNA (L7Ae sRNA) shows the fragmentation of the sample after $ZnCl_2$ treatment (L7Ae frag RNA). A smear of RNA bands is visible between 0 to 150 nt. b) The PCR-amplified cDNA libraries of L7Ae sRNA (left), frag RNA (middle) and WT control (right) were size-selected between 120 to 270 bp, which corresponds to RNA insert sizes of 0 to 150 nt.

2.2.2 Identification of the L7Ae-interacting RNAs using RIP-Seq analysis

The produced cDNA libraries of the co-immunoprecipitated RNAs were analyzed by Illumina sequencing to identify the RNA interaction partners of the L7Ae protein. In total, 32 million sequencing reads were mapped onto the *S. acidocaldarius* genome (see Table 4.23 in chapter 4.9.1). In agreement with the sRNome analysis (chapter 2.1.1), the two RIP-Seq plots of L7Ae displayed high read numbers for the C/D box sRNAs (Fig. 2.7a). Similar relative abundances of these RNAs were observed in the L7Ae sRNA data set and Sac-sR10 again represented the most abundant RNA. Ribosomal RNAs and RNase P RNA constituted other known interaction partners of L7Ae that were particularly overrepresented in the 'fragmented RNA' data set. In addition, abundant reads were found for the SRP RNA and for several mRNA fragments in both data sets. In contrast to the L7Ae sRNA and frag RNA cDNA libraries, where over 90 % of the sequenced reads mapped onto the *S. acidocaldarius* genome, only half of the reads of the WT control data set were specific for *S. acidocaldarius.* The reads of the other half represented unspecific sequences, e.g. from *E. coli* or human, which is caused by the extremely low amount of RNA that was obtained from the mock purification so that amplification of contaminating material is more likely during the cDNA library preparation. The obtained mapping pattern of the WT control resembled the sRNome plot. Small RNAs that were highly abundant in the sRNome, mainly C/D box sRNAs, were equally overrepresented in the WT control, which suggests that these RNAs were unspecifically isolated during the mock purification. This was problematic as many highly abundant RNAs, e.g. C/D box sRNAs, were known L7Ae interaction partners, so that normalization of the data sets was crucial.

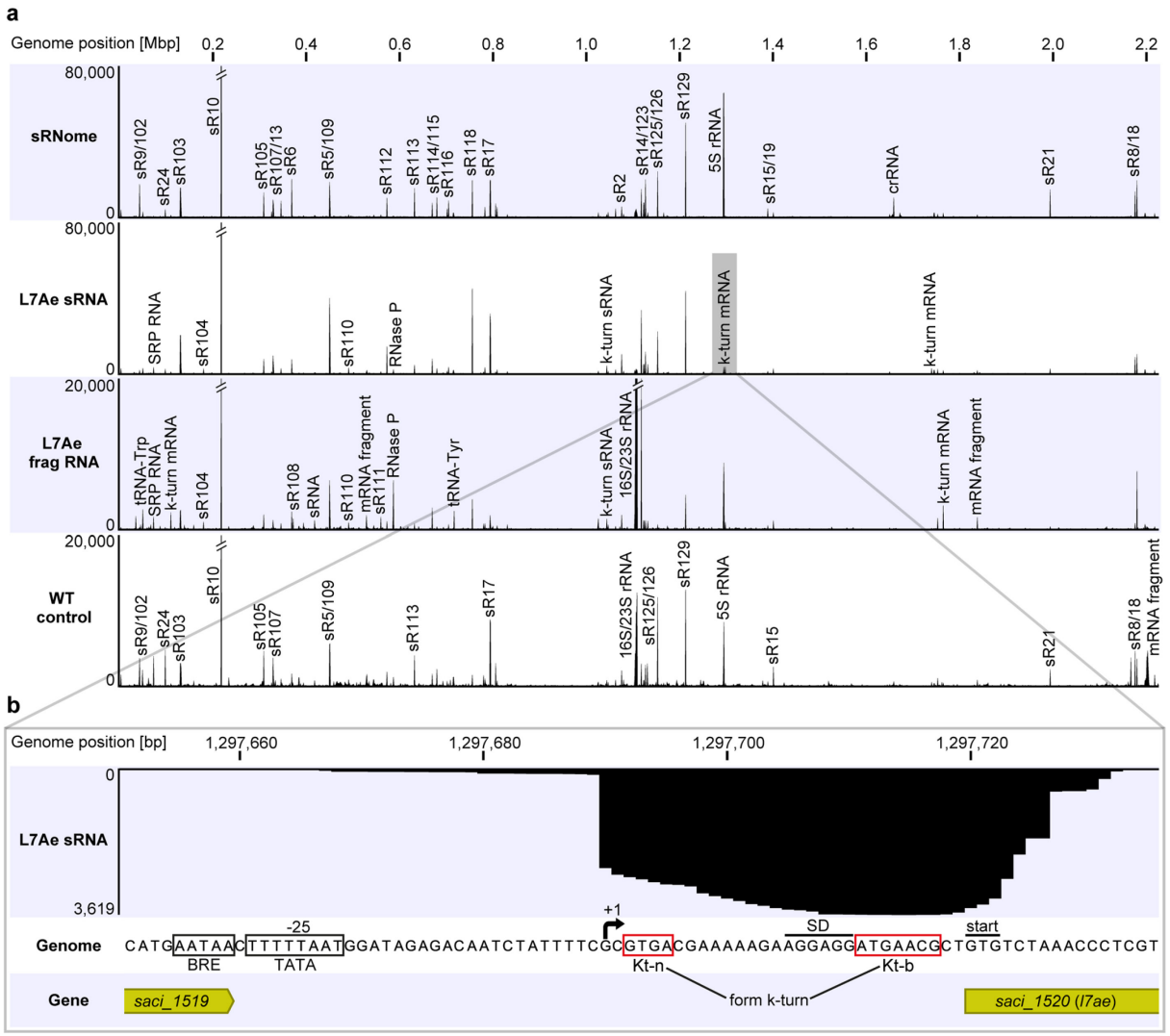


Figure 2.7 Global overview of the L7Ae-RNA interactome of *S. acidocaldarius*. a) The coverage plots of the sequenced L7Ae sRNA, L7Ae frag RNA and WT control cDNA libraries are illustrated. The sRNome was integrated to serve as a reference. Peaks representing C/D box sRNAs are labeled with (Sac-)sR#. Most C/D box sRNA peaks are also found in the two L7Ae data sets (unlabeled). For clearness, only the peaks which are unique for the two L7Ae plots were labeled. Each RNA profile contains one million mapped reads. The most abundant RNA of the sRNome, L7Ae sRNA and WT control data sets is Sac-sR10. The 16S and 23S rRNA represent the highest peak for the L7Ae frag RNA data set. b) The coverage plot of the *l7ae* promoter region of the L7Ae sRNA track is shown. A high number of reads is found downstream of position 1,297,690, which marks the transcriptional start site (+1) of the 5' UTR of the *l7ae* gene. Motifs for BRE (B recognition element) and TATA sites are boxed (black) upstream of the transcriptional start site. A Shine-Dalgarno (SD) motif is present 9 nt upstream of the GTG start codon and k-turn forming Kt-n and Kt-b strands are boxed (red).

Thus, a bioinformatic analysis of the RIP-Seq data sets was performed in collaboration with Michael Uhl (Group of Prof. Dr. Rolf Backofen, University of Freiburg) to identify bona fide interaction partners of L7Ae. The DESeq2 tool was used for a peak calling approach to find RNAs, which were significantly enriched (q-value < 0.1) over unspecific RNAs in the WT control data set [150]. This included a normalization of the read counts by a specific size factor, which was calculated for each library (for details see chapter 4.9.3). The identified RNAs were furthermore filtered by their normalized read count to remove false positive hits due to low-

abundance RNAs. This approach revealed 107 RNAs, which included 75 non-coding RNAs (59 C/D box sRNAs, 3 rRNAs (5S/16S/23S), 1 RNase P RNA, 1 SRP RNA, 1 tRNA, 1 H/ACA box sRNA, 3 antisense RNAs and 6 unknown sRNAs) and 32 fragments of mRNAs (Appendix 2). Most of the identified non-coding RNAs represented known interactors of L7Ae (C/D box sRNAs, rRNAs, RNase P RNA, H/ACA box sRNAs). In contrast, mRNAs were up to now not reported as binding partners of archaeal L7Ae. Interestingly, 20 of the 32 mRNA fragments comprised sequence motifs that matched or resembled the consensus sequence of C and D boxes (forming k-turns in C/D box sRNAs) (Appendix 2). To delineate the identified motifs from the terminology of the C/D box sRNAs, the C and D box-like motifs were termed Kt-b (bulged) and Kt-n (non-bulged) strand, respectively. The presence of the mRNAs might therefore be the result of k-turn motifs that are identified and bound by L7Ae. Putative k-turn motifs were found both in the coding sequences and in 5' untranslated regions (5' UTR) of mRNAs (Table 2.1).

The protein function of the majority (50 %) of the identified k-turn containing mRNAs is unknown (arCOG class S). Three mRNAs (*saci_0046*, *saci_0975* and *saci_1951*) encode proteins that categorize into arCOG class L, for proteins involved in replication, recombination and repair. Two k-turn containing mRNAs (*saci_1347* and *saci_1520*) categorize into arCOG class J (translation, ribosomal structure and biogenesis). The mRNAs of *saci_0139*, *saci_0669*, *saci_0886* and *saci_2027* encode proteins that belong to the arCOG classes N (cell motility), O (post-transcriptional modification), G (carbohydrate metabolism) and M (membrane biogenesis), respectively. Eight k-turn containing mRNAs are transcribed from operons and thus occur as polycistronic transcripts (*saci_0669*, *saci_0672*, *saci_0967*, *saci_0974*, *saci_0975*, *saci_1347*, *saci_1468* and *saci_1928*) (Table 2.1).

Interestingly, the k-turn containing mRNA of *saci_1520* encodes the L7Ae protein. The L7Ae sRNA data set displayed a high read abundance of the gene´s 5' UTR (Fig. 2.7b). The putative k-turn is formed by identified Kt-n and Kt-b sequences within this 5' UTR and the Kt-b sequence is preceded by a Shine-Dalgarno (SD) sequence. This finding led to the hypothesis that L7Ae regulates its own translation by binding to the k-turn formed within its 5' UTR, which masks the SD sequence and results in a reduced translation of the *l7ae* mRNA.

Table 2.1 List of k-turn containing mRNAs.

mRNA (k-turn locus)	arCOG	protein function	further mRNAs encoded within operon (arCOG/protein)
***saci_0046* (CDS)**	L	ATPase involved in DNA break repair	/
***saci_0139* (5' UTR)**	N	archaeal peptidase A24A	/
***saci_0237* (CDS)**	S	unknown	/
***saci_0333* (CDS)**	S	unknown	/
***saci_0630* (CDS)**	S	unknown	/
***saci_0669* (CDS)**	O	tRNA thio-modification	↑ *saci_0670* (J/30S rpS17e), ↓ *saci_0668* (S/unknown)
***saci_0672* (5' UTR)**	S	unknown	↓ *saci_0671* (G/FBPase)
***saci_0886* (CDS)**	G	sugar permease	/
***saci_0967* (CDS)**	S	unknown	↑ *saci_0968* (S/unknown)
***saci_0974* (CDS)**	S	unknown	↑ *saci_0975* (L/RPA)
***saci_0975* (CDS)**	L	replication protein A (RPA)	↓ *saci_0974* (S/unknown)
***saci_1004* (CDS)**	S	unknown	/
***saci_1347* (5' UTR)**	J	Nop5 (C/D box sRNA binding protein)	↓ *saci_1346* (J/fibrillarin), ↓ *saci_1345* (S/unknown), ↓ *saci_1344* (K/transcriptional regulator), ↓ *saci_1343* (J/N^2,N^2-dimethyl-guanosine tRNA methyltransferase TRM1)
***saci_1468* (5' UTR)**	S	unknown	↓ *saci_1467* (J/50S rpL39e), ↓ *saci_1466* (J/50S rpL31e), ↓ *saci_1465* (J/translation initiation IF-6), ↓ *saci_1464* (J/50S rpLX), ↓ *saci_1463* (O/prefoldin chaperone)
***saci_1468* (CDS)**	S	unknown	see above
***saci_1520* (5' UTR)**	J	50S rpL7Ae	/
***saci_1928* (CDS up)**	S	unknown	↓ *saci_1929* (S/unknown)
***saci_1928* (CDS down)**	S	unknown	↓ *saci_1929* (S/unknown)
***saci_1951* (CDS)**	L	transposase	/
***saci_2027* (CDS)**	M	glycosyltransferase	/

CDS: coding sequence; ↑ or ↓: upstream or downstream of the respective k-turn containing mRNA; L: replication, recombination, repair; N: cell motility; S: unknown function; O: post-transcriptional modification; G: carbohydrate metabolism; J: translation, ribosomal structure and biogenesis; M: cell membrane biogenesis

2.2.3 Binding analysis of the *l7ae* 5' UTR by EMSA studies

In order to test this autoregulation hypothesis, *in vitro* binding studies of the *l7ae* 5' UTR by the L7Ae protein were performed via electrophoretic mobility shift assays (EMSAs). The schematic structure of the *l7ae* 5' UTR reveals a standard k-turn: a Watson-Crick base pair is followed by a three-nucleotide bulge and the two characteristic G•A pairs (Fig. 2.8a). The SD sequence flanks the bulge at the 5' side. The formation of the k-turn was analyzed by mutating the nucleotides proposed to be involved in G•A pairing. Disruption of these non-Watson-Crick pairs was shown to be highly detrimental for proper k-turn folding and to prevent L7Ae binding [33, 151]. Recombinant His-Sumo tagged L7Ae was heterologously produced in *E. coli* and purified via immobilized metal affinity chromatography (IMAC) using Ni-NTA resin (Fig. 2.8b). The EMSAs showed that L7Ae can efficiently bind the native 5' UTR, whereas the mutation of a G•A pair resulted in binding deficiency (Fig. 2.8c).

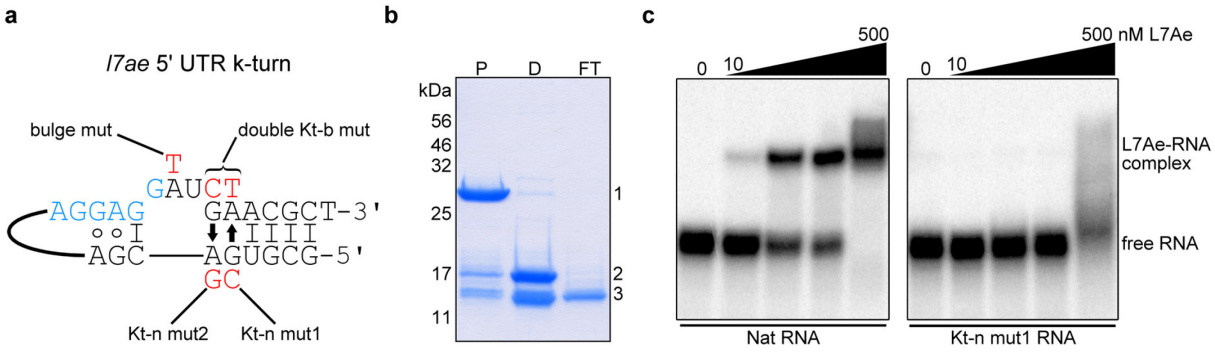


Figure 2.8 L7Ae binds its own 5' UTR. a) The schematic structure of the *l7ae* 5' UTR k-turn is illustrated. A three-nt bulge (GAU) is flanked at the 3' side by two *trans* sugar-Hoogsteen (arrows) G•A pairs and a four base pair stem. A single G-C base pair is present at its 5' side. The SD sequence is shown in blue. Mutants produced in this work are marked in red. b) SDS-PAGE analysis shows the purification steps of recombinant L7Ae: Lane P – pooled fractions of the first Ni-NTA purification, lane D – dialyzed protein after Sumo protease cleavage, lane FT – flow-through of the second Ni-NTA purification (1: His-Sumo-L7Ae, 2: cleaved His-Sumo tag, 3: pure L7Ae protein). c) EMSAs portray the binding of the *l7ae* 5' UTR by recombinant L7Ae. Binding of the native 5' UTR (Nat RNA) is observed with 10 nM L7Ae and increasing concentrations (10, 50, 100, 500 nM). The Kt-n strand mutated 5' UTR (Kt-n mut1) shows only unspecific binding at a concentration of 500 nM L7Ae.

2.2.4 L7Ae autoregulation analysis using β-galactosidase reporter assays

Next, β-galactosidase reporter assays were performed to test the autoregulation hypothesis *in vivo*. Variants of the *l7ae* promoter and the 5' UTR were fused to the *lacS* gene, integrated into a plasmid and transformed into *S. acidocaldarius*. The β-galactosidase enzyme activity was determined by measuring a yellow dye, which resulted from the hydrolysis of ONPG. *S. acidocaldarius* MW001 strains that were transformed with two independent mutants of the Kt-n strand (Kt-n mut1 and mut2) displayed a 1.6-fold increase in β-galactosidase activity in comparison to the native 5' UTR (Fig. 2.9). The tested mutations presumably impair k-turn formation and thereby abolish translational downregulation by L7Ae which results in a higher β-galactosidase activity. The effects were observed during logarithmic, early stationary and late stationary growth phases. Mutation of the BRE/TATA sites led to a loss of β-galactosidase activity suggesting that the *l7ae* promoter was disrupted. Next, augmentation of this negative feedback loop was attempted by using an *S. acidocaldarius* strain that lacked the most abundant L7Ae interactor Sac-sR10. Consequently, the amount of free L7Ae should be increased within this strain. A deletion mutant of this C/D box sRNA was successfully produced by Lydia Seelos (bachelor student under my supervision) [152]. In agreement, a ΔSac-sR10 strain that was transformed with the native *l7ae* 5' UTR displayed a 1.6-fold reduced β-galactosidase activity in comparison to MW001 wild-type strain transformants. The Kt-n sequence mutants showed a 2.1-fold higher enzyme activity in this background. These results demonstrate that the production of L7Ae is autoregulated in *S. acidocaldarius*.

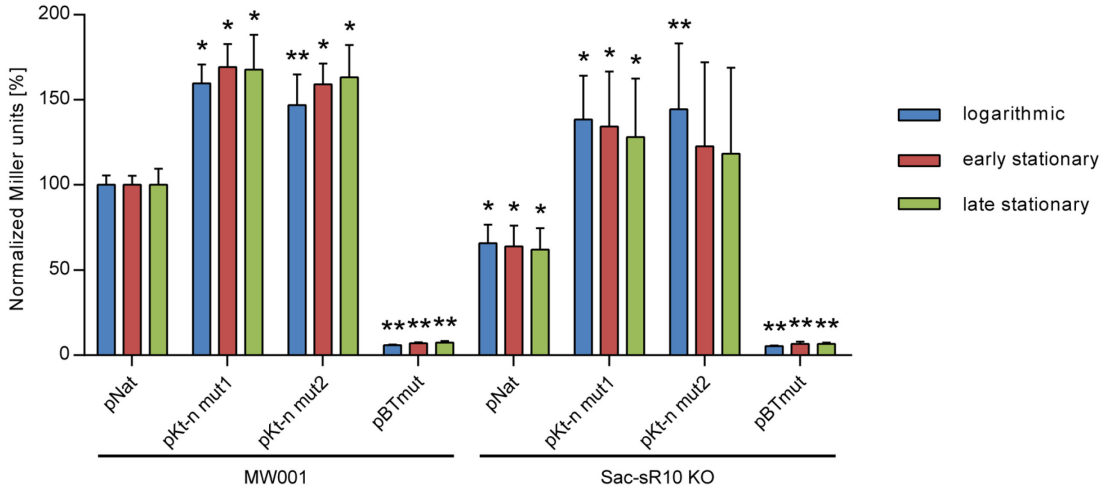


Figure 2.9 *In vivo* binding analysis of the *l7ae* 5' UTR by β-galactosidase reporter assays. Relative β-galactosidase activity (normalized Miller units) is shown for *S. acidocaldarius* MW001 and Sac-sR10 KO cells that were transformed with different plasmids: pNat (*l7ae* promoter + native 5' UTR), pKt-n mut1 and pKt-n mut2 (*l7ae* promoter + Kt-n mutant 1 or 2; Fig. 2.8a) and pBTmut (BRE/TATA site mutated *l7ae* promoter + native 5' UTR). The assay was performed with strains during logarithmic (blue), early stationary (red) and late stationary growth (green). The values are normalized to MW001 + pNat. Error bars demonstrate the standard deviation of five biological replicates. Asterisks (* = Student's t-test; ** = Welch's t-test) indicate the significance (p-value < 0.05) of the data with respect to the MW001 or Sac-sR10 KO + pNat strains, respectively.

2.2.5 Establishment of a bacterial GFP reporter system for L7Ae/k-turn binding studies

Due to the high levels of L7Ae substrates (most notably C/D box sRNAs), the β-galactosidase reporter set-up in *S. acidocaldarius* did not constitute an optimal system for testing further L7Ae/k-turn interactions. In bacteria, C/D box sRNAs are absent and only weak interactions of the endogenous L7Ae homologs YbxF and YlxQ with k-turns have been reported [84]. Therefore, a GFP reporter system for *E. coli* was designed to screen for motifs that are bound by L7Ae. *E. coli* Rosetta cells were transformed with a plasmid that contained IPTG-inducible *l7ae* from *S. acidocaldarius* and constitutively expressed *superfolding gfp (gfp)*. The *l7ae* 5' UTR was integrated upstream of the *gfp* gene to test for GFP downregulation by L7Ae and fluorescence was quantified by flow cytometry. These cells showed a GFP signal reduction of 85 % compared to a control strain without L7Ae (the *l7ae* gene of the control plasmid was destroyed by introducing a frameshift) (Fig. 2.10). The introduction of two mutations in the Kt-n strand of the *l7ae* 5' UTR resulted in a slightly higher fluorescence signal in comparison to the native UTR. However, an unexpected strong reduction of the overall fluorescence signal strength was noted for this strain. Mutation of the bulge, which should not affect k-turn formation, showed a GFP signal reduction similar to the native 5' UTR. Finally, a control UTR lacking a k-turn was tested. The *l7ae* 5' UTR was exchanged for the optimized LII-12 UTR of the P_m promoter of *E. coli,* described by Berg *et al.* [153]. This set-up was also found to result

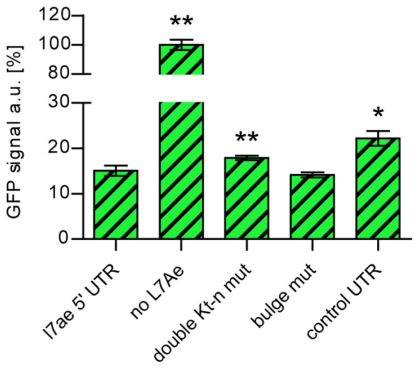


Figure 2.10 Toxicity effects of L7Ae overexpression in *E. coli*. The relative GFP signal of five *E. coli* transformants is shown. All transformants contain a plasmid with constitutively expressed *gfp* (preceded by the *l7ae* 5' UTR) and an IPTG-inducible *l7ae* gene. The plasmids contain the following variations: *l7ae* 5' UTR (no variation), no L7Ae (frameshifted *l7ae* gene), double Kt-n mut/bulge mut (double Kt-n/bulge mutated *l7ae* 5' UTR (Fig. 2.8a)) and control UTR (instead of *l7ae* 5' UTR). Error bars demonstrate the standard deviation of three biological replicates. Asterisks (* = Student's t-test; ** = Welch's t-test) indicate the significance (p-value < 0.05) of the data with respect to the strain containing the *l7ae* 5' UTR. The GFP fluorescence was recorded only from the GFP-positive population.

in less GFP signal (over 75 % compared to the respective control) and indicated that overproduction of L7Ae may lead to toxicity effects in *E. coli*, possibly due to the unspecific binding of essential RNA molecules.

The growth of the analyzed bacterial strains was monitored to investigate this possibility. In comparison to the control without L7Ae (no L7Ae), the *l7ae* expressing strain (no aKt) showed a decrease of the cell density three hours after induction, but later displayed recovery of the cell growth (Fig. 2.11a). After recovery of the cell growth, the culture was re-inoculated and again induced with IPTG. No decline of the cell density was observed for this re-inoculated strain (no aKt RI). The culture rather displayed the growth behavior of non-induced strains (Fig. 2.11b). These growth patterns confirm that overproduction of L7Ae is detrimental to *E. coli* cells and that recovery of the cell growth is due to the selection of escape mutants. In order to overcome this toxic effect, the negative feedback-loop strategy that *S. acidocaldarius* deploys to reduce L7Ae overproduction was introduced into *E. coli*. Therefore, an additional *l7ae* 5' UTR was integrated upstream of the *l7ae* gene to enable autoregulation of its expression. The resulting pMD-auto*l7ae-gfp* plasmid comprised an autoregulatory k-turn (aKt) and a *gfp*-regulatory k-turn (*gfp*Kt) sequence (Fig. 2.11c). Cells that were transformed with this plasmid (pMD) displayed normal growth behavior (Fig. 2.11a), demonstrating that autoregulation of L7Ae abolished its toxicity effects in *E. coli*. Flow cytometry also confirmed toxicity of the non-autoregulated L7Ae. Two cell populations and a high amount of debris could be observed for the strain without aKt (Fig. 2.11d). A significant portion of the cells was GFP-negative, likely

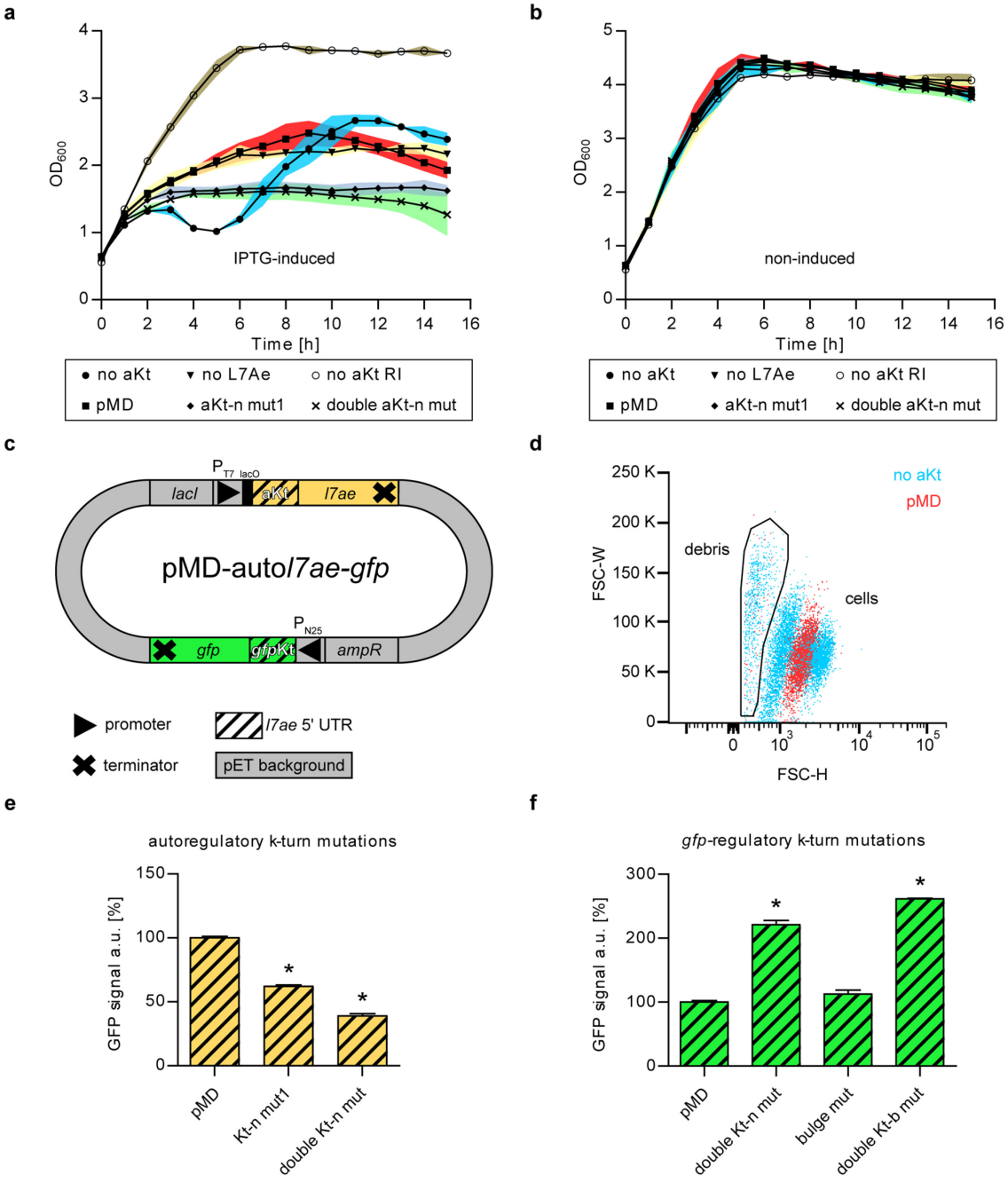


Figure 2.11 L7Ae toxicity can be cured by autoregulation. a) and b) Growth curves show IPTG-induced and non-induced *E. coli* strains containing different pMD-auto*l7ae-gfp* variants: no aKt (5' UTR of pET plasmid; RI: re-inoculated), no L7Ae (frameshifted *l7ae* gene), pMD (pMD-auto*l7ae-gfp*), aKt-n mut1 and double aKt-n mut (Fig. 2.8a). Error bars (color filled area) demonstrate the standard deviation of three biological replicates. c) The pMD-auto*l7ae-gfp* plasmid contains the *l7ae* gene under an IPTG-inducible T7 promoter (P_{T7}), which is followed by the *l7ae* 5' UTR (yellow streaked box). The k-turn formed by the 5' UTR leads to the negative autoregulation of L7Ae translation (autoregulatory k-turn or aKt). The *superfolding gfp* gene (*gfp*) is expressed by the constitutive N25 promoter (P_{N25}) from phage T5 [154]. A second *l7ae* 5' UTR (green streaked box) is upstream of the *gfp* gene and forms a GFP-regulatory Kt (*gfp*Kt). d) The dot plot illustrates the flow cytometry analysis of *E. coli* cell populations that carry the pMD-auto*l7ae-gfp* plasmid (pMD) or the aKt-absent variant (no aKt). The pMD strain shows one population, whereas the aKt-absent strain displays two cell populations and a higher amount of cell debris. e) The relative GFP signal of transformants that comprise mutations in the aKt is shown. The utilized pMD-auto*l7ae-gfp* plasmids contain a control UTR upstream of the *gfp* gene to show toxicity-caused GFP downregulation in the aKt mutants. f) The chart depicts the relative GFP signal of transformants that comprise mutations in the *gfp*Kt. In e) and f) the values were normalized to the pMD strain. Error bars demonstrate the standard deviation of three biological replicates. Asterisks (* = Student's t-test) indicate the significance (p-value < 0.05) of the data with respect to the pMD strain.

representing dead cells. The strain with autoregulated *l7ae* (pMD) displayed only a single population. Mutations of the aKt, introduced to abolish autoregulation, led to growth defects, which were less drastic than for the strain without aKt (Fig. 2.11a). This restored toxicity also manifested in the GFP signal. The fluorescence of aKt mutants was reduced in comparison to a functional aKt (pMD) (Fig. 2.11e). A double mutant of the aKt showed a more severe effect. Next, mutational analysis of the *gfp*Kt was performed. A 2.2-fold higher GFP signal was measured for the double Kt-n mutant (Fig. 2.11f), whereas the bulge mutant, which should not be affected in k-turn formation, showed a reduced GFP signal similar to the functional *gfp*Kt (pMD). A double mutant of the k-turn critical G•A nucleotides within the Kt-b strand led to 2.5-fold higher GFP fluorescence. These results demonstrate that the developed plasmid-borne autoregulatory GFP reporter system constitutes a reliable method to test L7Ae/k-turn interaction.

2.2.6 Multiple sequence alignments of archaeal *l7ae* 5' UTRs

The conservation of the *l7ae* 5' UTR among 121 archaeal species was investigated by multiple sequence alignments (Appendix 3). All investigated archaea, with the exception of Thermoproteales and *Nanoarchaeum equitans*, contained a SD sequence ~7-10 bp upstream of the *l7ae* coding sequence, indicating that almost all archaeal *l7ae* mRNAs comprise a 5' UTR. Potential transcriptional start sites could be assigned for most of the sequences due to the presence of conserved TATA boxes which revealed 5' UTR sizes from 10 nt up to 200 nt. High sequence conservation of the leader sequences was only identified among the Sulfolobales (Fig. 2.12). Bioinformatic prediction of k-turn formation within the 5' UTRs was attempted in collaboration with Prof. Michelle Meyer (Department of Biology, Boston College, MA, USA). In general, this is a challenging task due to the high sequence variability of k-turn/-loop structures. *In silico* k-turn folding was only possible for the Methanococcales, Methanocellales, Methanosarcinales and Haloferacales. However, conserved Kt-b and Kt-n strands within the 5' UTRs were identified by careful manual inspection of the alignments for almost all members. This indicates that k-turn formation is a conserved feature in archaeal *l7ae* 5' UTRs and suggests that autoregulation by the L7Ae protein is a widespread feature in archaea.

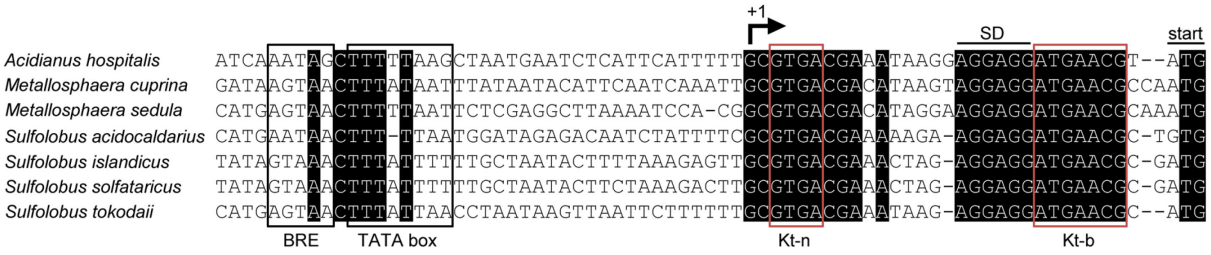


Figure 2.12 Alignment of Sulfolobales *l7ae* promoter sequences. The alignment shows the conservation of the *l7ae* 5' UTRs among the Sulfolobales. Regions of conservation are shaded in black. The BRE and TATA box promoter elements are found -25 bp upstream of the transcriptional start site (+1). The Shine-Dalgarno sequence (SD) precedes the Kt-n strand.

2.2.7 Analysis of archaeal *l7ae* 5' UTR binding by L7Ae via the GFP reporter system

The conservation of k-turn formation within other archaeal *l7ae* 5' UTRs was investigated by the developed GFP reporter system (chapter 2.2.5). The *gfp*Kt of pMD-auto*l7ae-gfp* was replaced by 50 bp of *l7ae* upstream regions of eight archaeal model organisms spread across the archaeal domain and tested for binding by *S. acidocaldarius* L7Ae (Fig. 2.13a). The *l7ae* leader sequences of *Staphylothermus marinus*, *Archaeoglobus fulgidus*, *Haloferax volcanii*, *Methanosarcina acetivorans*, *Thermococcus kodakarensis* and *Methanococcus maripaludis* comprised Kt-b and Kt-n strands, whereas only a Kt-n strand could be identified for *Aeropyrum pernix*. *Pyrobaculum aerophilum* belongs to the order Thermoproteales, which did not comprise an *l7ae* 5' UTR. The control UTR strain without *gfp*Kt showed around 80 % GFP fluorescence compared to strains with non-functional L7Ae (Fig. 2.13b). The reduction of the GFP signal by 20 % might account for residual toxicity of L7Ae. For comparison, strains with the *gfp*Kt (*S. acidocaldarius l7ae* 5' UTR) showed around 40 % GFP fluorescence. The *A. pernix* UTR showed no L7Ae downregulation, probably due to the absence of the Kt-b strand. However, downregulation was observed for all other tested archaeal 5' UTRs, particularly for the *A. fulgidus*, *M. acetivorans and M. maripaludis* UTRs. Strains with the *P. aerophilum* 5' UTR of *l7ae* displayed GFP signals close to background fluorescence and were therefore excluded. The results highlight that *S. acidocaldarius* L7Ae can bind to k-turns formed within the *l7ae* 5' UTR regions of various archaea, suggesting that the negative feedback loop of L7Ae is a conserved feature and operative in most archaeal organisms.

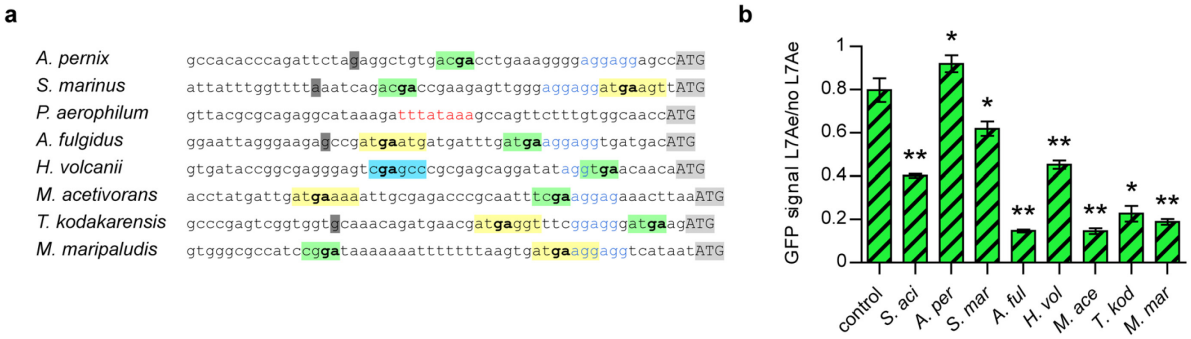


Figure 2.13 L7Ae binds the *l7ae* 5' UTRs of various archaea. a) The ~50 bp *l7ae* upstream sequences of eight taxonomically diverse archaea were extracted from the Clustal Omega alignment (Appendix 3). The proposed transcriptional start sites (dark grey box), potential Kt-b (yellow box) and Kt-n (green box) strands, Shine-Dalgarno sequences (blue letters) and start codons (light grey box) are marked. The GA nucleotides critical for k-turn formation are shown in bold. The *H. volcanii* sequence comprises a potential Kt-b strand (blue box) apart from a Kt-b strand that was identified further upstream in the alignment (Appendix 3). The *P. aerophilum l7ae* upstream sequence does not show any of the marked features, but comprises a TATA box (red letters) -25 bp upstream of the start codon. b) The relative GFP signal is shown for *E. coli* transformants that comprise a control UTR or the *l7ae* upstream sequences from a) in place of the *gfp*Kt (except *S. aci*) of the pMD-auto*l7ae-gfp* plasmid. The GFP signals represent the ratio of L7Ae/no L7Ae producing cells. Error bars demonstrate the standard deviation of three biological replicates. Asterisks (* = Student's t-test; ** = Welch's t-test) indicate the significance (p-value < 0.05) of the data with respect to the control strain.

2.2.8 Binding analysis to identified k-turn motifs in different mRNAs and SRP RNA

The RIP-Seq analyses of L7Ae revealed additional mRNAs that comprised potential k-turn motifs. This suggests that L7Ae might also be involved in regulating the translation of these transcripts. Three k-turn containing mRNAs were investigated for binding by the L7Ae protein: *saci_1468*, *saci_2027* and *saci_1347*. The proposed k-turn sequences of the *saci_1468* and *saci_2027* mRNAs were fused to the coding sequence of the *gfp* gene and tested in the established reporter system. The *saci_1468* and *saci_2027* genes encode for a DNA binding protein (hypothetical programmed cell death protein) and a glycosyltransferase (hypothetical protein), respectively. Both strains displayed downregulation of the GFP signal upon L7Ae induction (Fig. 2.14a). Unexpectedly, several attempts to clone the proposed k-turn sequence of the *saci_1347* mRNA into the GFP reporter plasmid failed. All obtained transformants (17) were mutated in the promoter sequence or in the coding sequence of the *saci_1347* Kt-*gfp* fusion. Two transformants displayed single mutations in the proposed Kt-b strand of the *saci_1347* Kt. Also site-directed mutagenesis to remove this single mutation repeatedly failed, indicating that the desired plasmid was somehow toxic for *E. coli.* The putative k-turn sequence of the *saci_1347* mRNA was therefore tested via EMSA. Interestingly, the *saci_1347* gene encodes the C/D box sRNP protein Nop5. In *S. acidocaldarius*, *saci_1347* is the first gene of an operon and the second, co-transcribed gene, *saci_1346,* encodes for the C/D box sRNP protein fibrillarin. L7Ae binding was observed with a similar affinity as for the *l7ae* 5' UTR (Fig. 2.14b).

A different potential L7Ae substrate is the SRP RNA as high read coverage was observed in the RIP-Seq data (Fig. 2.14c). Similar to *S. solfataricus*, a putative k-turn could be identified in close proximity to the 5e motif of helix 5 of the *S. acidocaldarius* SRP RNA (Fig. 2.14d) [92]. EMSA assays verified that this k-turn is also bound by L7Ae (Fig. 2.14e). This suggests that the SRP RNA is an authentic interaction partner of L7Ae.

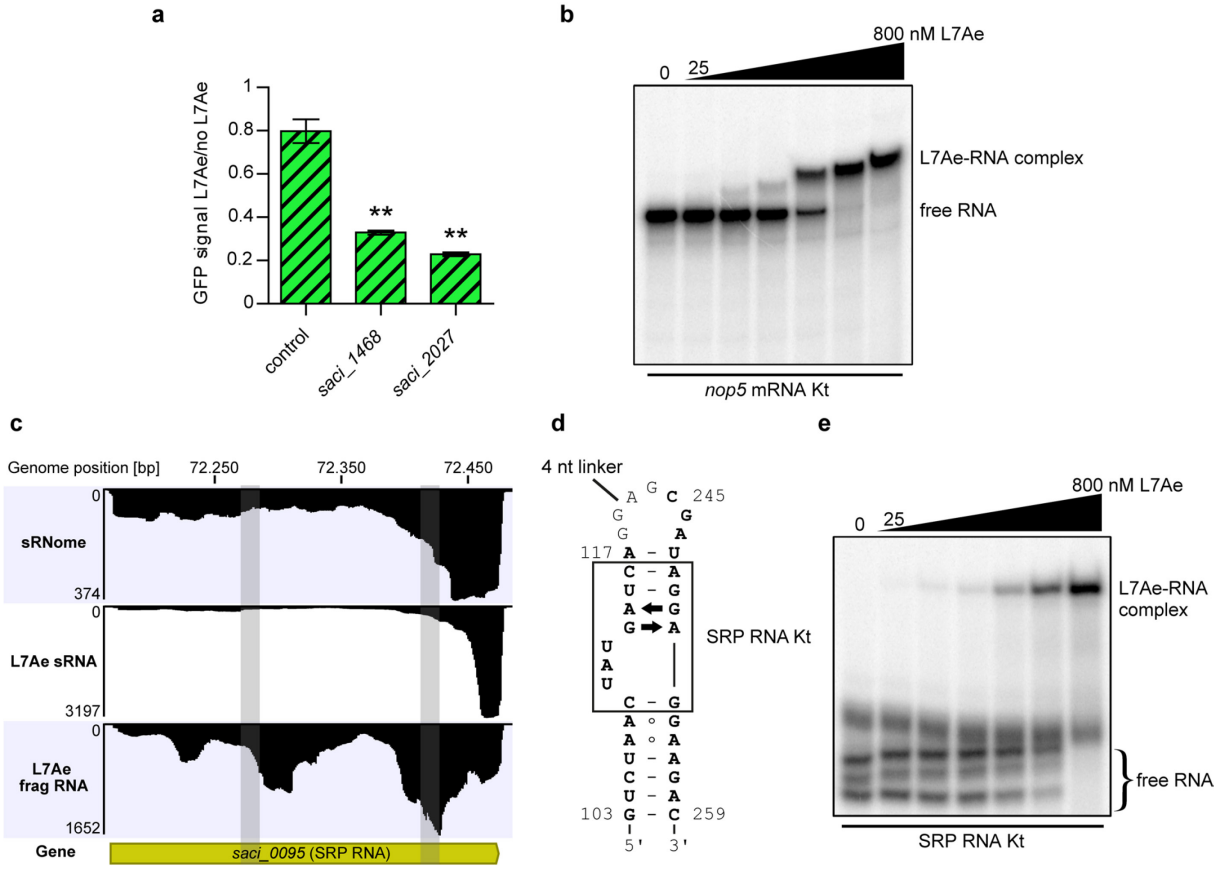


Figure 2.14 L7Ae binds to k-turn motifs identified in mRNA and SRP RNA. a) The relative GFP signal is shown for *E. coli* transformants which contain a control UTR in place of the *gfp*Kt. The *saci_1468* and *saci_2027* strains further contain the respective k-turn containing mRNA regions (Appendix 2) directly downstream of the start codon (GFP fusion). The GFP signals represent the ratio of L7Ae/no L7Ae producing cells. Error bars demonstrate the standard deviation of three biological replicates. Asterisks (** = Welch's t-test) indicate the significance (p-value < 0.05) of the data with respect to the control strain. b) The EMSA demonstrates full binding of the *nop5* mRNA k-turn at 400 nM L7Ae concentration (25, 50, 100, 200, 400, 800 nM gradient). The substrate comprises the first 125 nt of the mRNA, which was found to be enriched in the L7Ae RIP-Seq analysis (Appendix 2). c) The coverage plots for the SRP RNA locus of the sRNome, L7Ae sRNA and frag RNA data sets are illustrated. High read numbers are present for the two L7Ae RIP-Seq data sets. The two regions that form a potential k-turn in the SRP RNA are highlighted in grey. Each RNA profile contains one million mapped reads. d) The two RNA strands from nucleotides 103-117/245-259 (grey highlighted in c) of the SRP RNA form a potential k-turn (boxed). The shown sequence was used as substrate for the EMSA. The two RNA strands were joined by a four-nucleotide linker. e) L7Ae shows binding to the k-turn within the SRP RNA of *S. acidocaldarius*. Three secondary structures are formed by the free substrate that show full binding at an L7Ae concentration of 800 nM (25, 50, 100, 200, 400, 800 nM gradient).

2.2.9 Analysis of the conservation of archaeal SRP RNA k-turns

The finding of a putative k-turn in proximity to the 5e motif of the *S. acidocaldarius* and *S. solfataricus* SRP RNAs indicated that this structure could be a conserved feature in archaeal SRP RNAs. A multiple sequence alignment of 82 archaeal SRP RNA sequences listed in the SRP database (SRPDB: http://rth.dk/resources/rnp/SRPDB) was performed to investigate the conservation of this k-turn (Appendix 4). The alignment showed conserved GA bases within hinge 2 of the SRP RNAs. Hinge 2 is a loop within helix 5 that is found in close proximity to the 5e motif of all SRP RNAs (Fig. 2.15a) [155]. This loop comprised the k-turn sequence that was identified in *S. acidocaldarius* (Fig. 2.15b). The corresponding regions of the same archaeal model organisms that were investigated in chapter 2.2.7 (phylogenetically distant archaea) were manually folded (Fig. 2.15c). For all organisms, a putative k-turn sequence was identified at hinge 2 in a distance of 9-10 basepairs to the 5e motif. This suggests that k-turn formation at hinge 2 is a conserved feature of archaeal SRP RNAs.

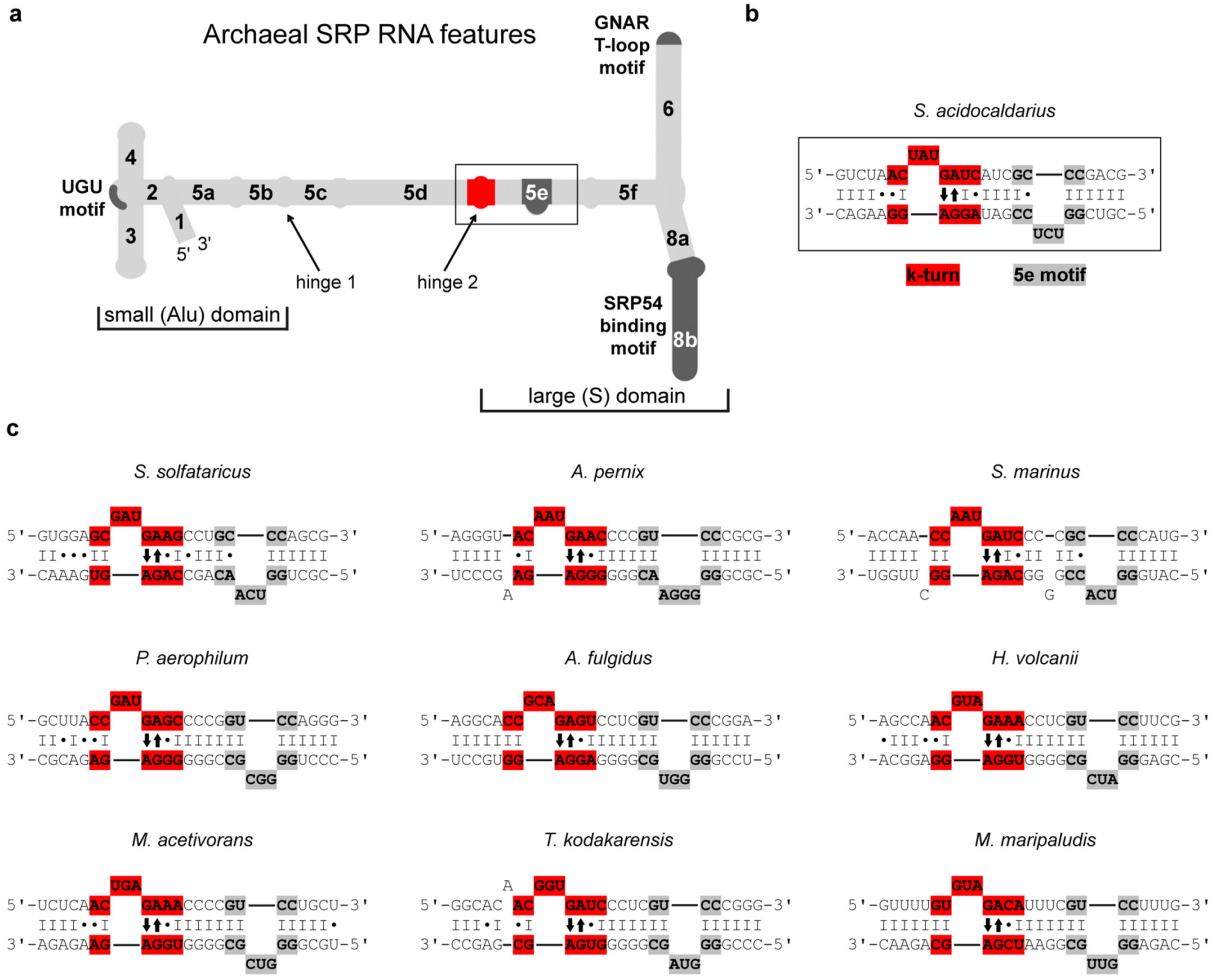


Figure 2.15 Putative k-turns within hinge 2 of archaeal SRP RNAs. a) The scheme of an archaeal SRP RNA is illustrated (modified from [156]). Important loci within the RNA are shown in grey. The locus of hinge 2 (red) and the 5e motif within helix 5 is boxed. b) The sequence shows the putative secondary structure of the *S. acidocaldarius* helix 5 region comprising hinge 2 and the 5e motif. A putative k-turn (red) is present in a distance of 9 basepairs to the 5e motif (grey). c) The putative secondary structures of hinge 2 and 5e motif of nine phylogenetically distant archaea shows that k-turn formation at hinge 2 is conserved among archaea.

2.2.10 Abolishment of L7Ae autoregulation in *S. acidocaldarius*

In *E. coli*, overproduction of the L7Ae protein led to toxic effects (chapter 2.2.5). The endogenous *l7ae* 5' UTR was mutated and the growth of the produced strains was monitored to screen for a similar phenotype in *S. acidocaldarius*. Strains with mutations of the Kt-n strand, which should abolish L7Ae autoregulation, showed only a minor growth delay under optimal laboratory growth conditions (Fig. 2.16). Thus, overproduction of L7Ae in *S. acidocaldarius* seems not to be as detrimental as observed for *E. coli*.

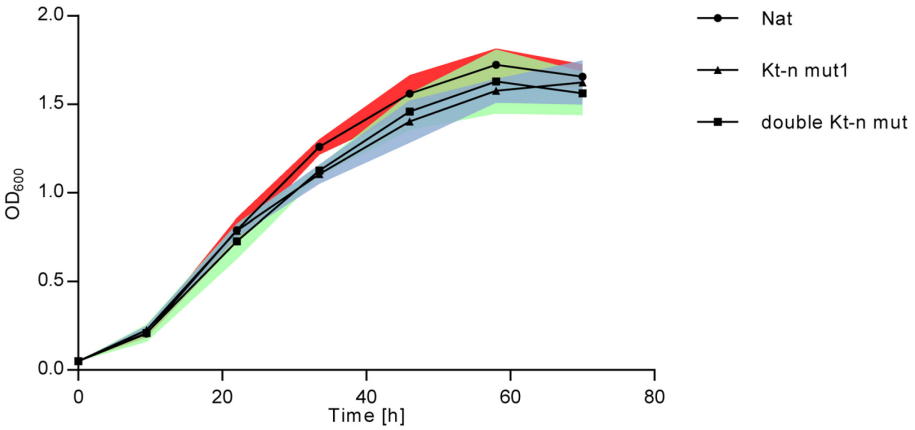


Figure 2.16 Growth curves of *l7ae* 5' UTR variants in the *S. acidocaldarius* genome. Depicted is the growth behavior of *S. acidocaldarius* strains (MW001 background) that comprise genomic mutations in the *l7ae* 5' UTR: Nat (no mutation), aKt-n mut1 and double aKt-n mut (Fig. 2.8a). The *l7ae* gene of all strains is genomically Flag-HA tagged, which simplified screening for the correct mutation. Error bars (color filled area) demonstrate the standard deviation of three biological replicates.

2.2.11 Competition of *l7ae* 5' UTR binding using natural substrates of L7Ae

Finally, competition analyses were performed to investigate the binding priority of the L7Ae protein for its substrates. Excess amounts of unlabeled total RNA and C/D box sRNA were added to radioactively labeled *l7ae* 5' UTR substrate. The utilized total RNA sample was depleted of small RNAs and ribosomal RNAs constitute most of the purified material, which represent L7Ae substrates (Fig. 2.1). Total RNA efficiently outcompeted the *l7ae* 5' UTR binding, whereas addition of the representative C/D box sRNA Sac-sR121 did not affect *l7ae* 5' UTR binding significantly (Fig. 2.17). The EMSA suggests that ribosomal RNA is the preferred L7Ae substrate and has a binding priority over the *l7ae* 5' UTR.

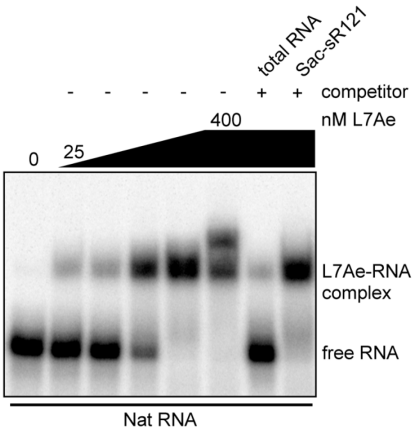


Figure 2.17 Significance of *l7ae* 5' UTR binding. a) The EMSA displays the competition analysis of *l7ae* 5' UTR (Nat RNA) binding. The Nat RNA is fully bound at a concentration of 400 nM L7Ae (25, 50, 100, 200, 400 nM L7Ae gradient). A ten-fold concentration of unlabeled total RNA of *S. acidocaldarius* shows effective competition of the binding as free Nat RNA is obtained. Almost no competition of the Nat RNA binding is observed for a hundred-fold excess of unlabeled C/D box sRNA Sac-sR121.

2.3 The LSm-RNA interactome of *S. acidocaldarius*

Previous studies identified several RNA species as binding partners of the archaeal LSm proteins, which indicates that these proteins are involved in various RNA-regulated processes in archaea [102, 132, 134]. However, the role of the proteins within these processes remains elusive. In order to shed more light on the function of the archaeal Sm homologs, the essentiality of the *lsm* genes from *S. acidocaldarius* was investigated and RNA interaction partners were identified via RIP-Seq.

2.3.1 Deletion of the three *lsm* genes of *S. acidocaldarius*

The *S. acidocaldarius* genome contains three *lsm* genes: *saci_1224* (*lsm1*), *saci_0799* (*lsm2*) and *saci_0660* (*lsm3*). In order to analyze the essentiality of the *S. acidocaldarius lsm* genes, two deletion approaches were conducted: i) deletion via the type I-A/III-B CRISPR-Cas systems and ii) deletion via the pyrEF-exchange method. The first technique was developed by Li and co-workers to generate marker-less mutations in *Sulfolobus islandicus* and resembles the well-established CRISPR-Cas9 genome editing approach [140]. It is based on the interference activity of the endogenous CRISPR-Cas type I-A and III-B systems of *S. islandicus*. In this method, a mutagenesis plasmid, which encodes a crRNA that targets the gene of interest, is transformed into *S. islandicus*. In addition, the plasmid comprises 'donor DNA', which encompasses the flanking regions of the gene of interest. Deletion of the gene of interest is achieved, when homologous recombination between the donor DNA and the locus of the gene of interest occurs (Fig. 2.18a). These cells survive, in contrast to transformants that contain the wild-type locus (no recombination), which are killed by the self-targeting interference

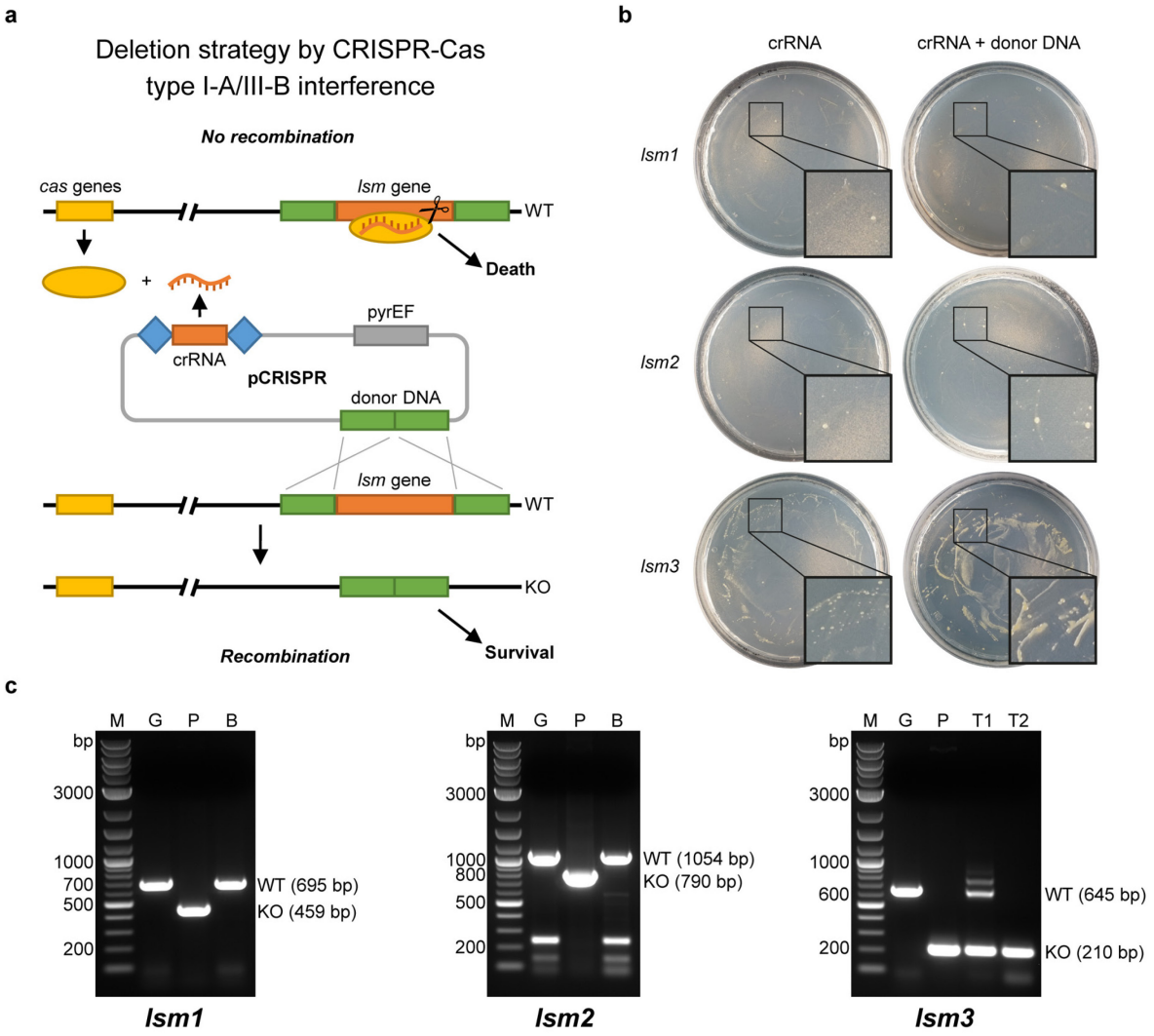


Figure 2.18 Deletion approach of the *lsm* genes via the endogenous type I-A/III-B CRISPR-Cas systems of *S. acidocaldarius*. a) The scheme illustrates the *lsm* deletion technique using the endogenous CRISPR-Cas systems of *S. acidocaldarius* (modified from [140]). The principle is described in the main text. b) The plates display the result of MW001 transformation with the produced pCRISPR plasmids lacking (crRNA) or containing the donor DNA (crRNA + donor DNA). The rectangles at the bottom right of the plates display a magnified image of the top left box. c) Colony PCRs of the *lsm* loci show bands representing the wild-type locus (WT) or the *lsm* deleted locus (KO). Genomic DNA (G) of *S. acidocaldarius* and the pCRISPR plasmids (P) were used as negative and positive controls, respectively. Representative background colonies (B) of the *lsm1* and *lsm2* deletion attempts display WT bands. Two types of transformants (T1/T2) were obtained for the *lsm3* deletion attempt: T1 shows the WT and KO bands as well as additional upper bands, whereas T2 displays a single band for the *lsm3* deleted locus.

activity of the endogenous type I-A and type III-B CRISPR-Cas systems. In the absence of donor DNA, all transformants are killed (negative control). The type I-A and type III-B CRISPR-Cas systems are also present in *S. acidocaldarius* [145, 146]. The group of Prof. Dr. Sonja-Verena Albers (University of Freiburg) modified the utilized mutagenesis plasmid to adapt it for *S. acidocaldarius*. In order to delete the *lsm* genes from *S. acidocaldarius*, suitable spacer sequences were cloned into the pCRISPR plasmid to generate *lsm* gene targeting crRNAs. As spacer sequences, the 36-38 nt regions adjacent to a TCN motif were chosen, which was reported as the recognized PAM (protospacer adjacent motif) of the type I-A system of *S.*

acidocaldarius [144, 145]. Furthermore, a donor DNA region was integrated into pCRISPR, which contained the flanking regions of the *lsm* genes (Fig. 2.18a).

S. acidocaldarius MW001 cells were transformed with the generated pCRISPR plasmids and streaked on Xylose supplemented plates to induce the overproduction of the crRNAs. Plasmids lacking the donor DNA were transformed in the same manner as negative control experiments to verify the interference activity of the utilized crRNAs. The plates of the negative control transformations for *lsm1* and *lsm2* (crRNA) displayed a low number of colonies (around ten) (Fig. 2.18b). A similar number of colonies was found for the donor DNA plasmid transformation (crRNA + donor DNA). Colony PCRs demonstrated that all picked clones of the *lsm1*/*lsm2* transformation did not contain the donor DNA plasmids and represented background colonies (Fig. 2.18c; left/middle panel). This demonstrates the activity of the CRISPR-Cas systems and indicates that *lsm1* and *lsm2* cannot be deleted. In contrast, over 90 colonies were obtained for the *lsm3* donor DNA plasmid transformation (Fig. 2.18b). Unexpectedly, also the negative control experiment showed many transformants (over 70), indicating that the chosen crRNA spacer sequence of the *lsm3* gene was not optimal for CRISPR-Cas interference. Two further spacer sequences for the *lsm3* gene were tested, which were chosen upstream and downstream of the initially tested spacer sequence. However, also the transformation of these two negative control plasmids displayed many colonies (data not shown). Colony PCRs demonstrated that 21 of 28 picked clones of the *lsm3* transformation comprised the donor DNA plasmid (Fig. 2.18c; right panel). Twenty of the transformants displayed bands for both the wild-type (genome) and the deleted locus (plasmid) of *lsm3*. These clones also showed additional bands above the wild-type band. However, one clone displayed a single band for the deleted *lsm3* locus, which was verified by sequencing. This shows that the *lsm3* gene was successfully deleted by the CRISPR method.

Knock-outs of the *lsm* genes were also attempted by the pyrEF-exchange method, which is a fast and straight-forward technique to generate marker-gene containing deletion strains. To this end, the pyrEF cassette, which encodes proteins that are essential for uracil synthesis, and the 50 bp flanking regions of the *lsm* genes were amplified by PCR and transformed into *S. acidocaldarius*. In a uracil-lacking medium, the uracil-auxotrophic MW001 strain can only survive upon homologous recombination with the pyrEF cassette, which generates *lsm* deletion strains. Again, only the *lsm3* gene could be successfully knocked-out by this method, whereas no deletions of the *lsm1* and *lsm2* genes could be obtained.

Taken together, these results indicate that the two *lsm* genes *lsm1* and *lsm2* are essential and non-redundant in *S. acidocaldarius*, whereas the *lsm3* gene is dispensable.

2.3.2 Growth analysis of the *lsm3* deletion strain

The single transformant which contained the CRISPR-deleted *lsm3* locus (chapter 2.3.1) was plated on second selection plates containing 5-FOA in order to lose the utilized donor DNA plasmid. This approach yielded markerless Δ*lsm*3 clones for further analyses. The growth of the produced strain (LSm3KO) was monitored. A similar growth behavior as for the MW001 reference strain could be observed (Fig. 2.19).

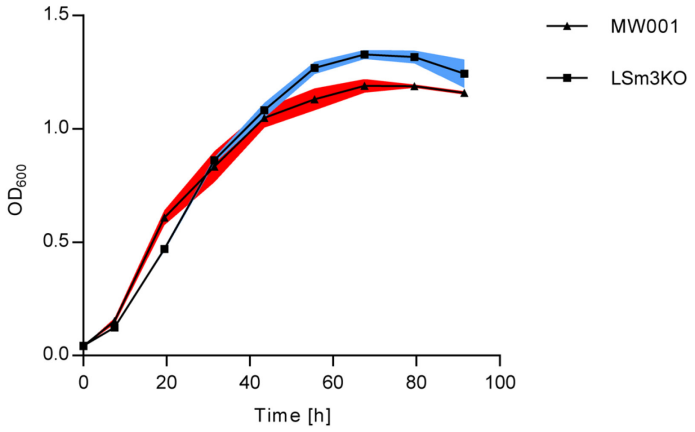


Figure 2.19 Growth curve of the *S. acidocaldarius* LSm3KO strain. The growth curves for the *S. acidocaldarius lsm3* deletion strain and MW001 reference strain are depicted. Error bars (color filled area) demonstrate the standard deviation of three biological replicates.

2.3.3 Genomic tagging of the *lsm* genes in *S. acidocaldarius*

A similar purification approach as described for the L7Ae RIP-seq studies (chapter 2.2.1) was pursued to isolate the LSm proteins and their interaction partners from *S. acidocaldarius*. Thus, genomic tagging by a Flag-HA tag (FHA) was attempted for the three *lsm* genes. N-terminal tagging was only successful for the *lsm2* and *lsm3* genes. The mutated *lsm2* strain displayed small colonies and delayed growth, which was not observed for the tagged *lsm3* strain (Fig. 2.20a). Therefore, C-terminal tagging of the *lsm1* and *lsm2* genes was conducted, making sure that the overlapping gene of *lsm1*, *saci_1223*, was not mutated. Only the *lsm1* gene could be tagged by a C-terminal Flag-HA tag. However, similar to N-terminally tagged *lsm2*, the LSm1-CFHA strain displayed delayed growth (Fig. 2.20a). Thus, the rather long Flag-HA tag (20 amino acids) might impair the ring formation of the LSm proteins. Therefore, a fusion with the substantially smaller His tag was chosen (6 amino acids). All three *lsm* genes could be C-terminally His-tagged. The growth of the produced strains was monitored and showed that the presence of the smaller tag did not negatively influence the growth behavior (Fig. 2.20b).

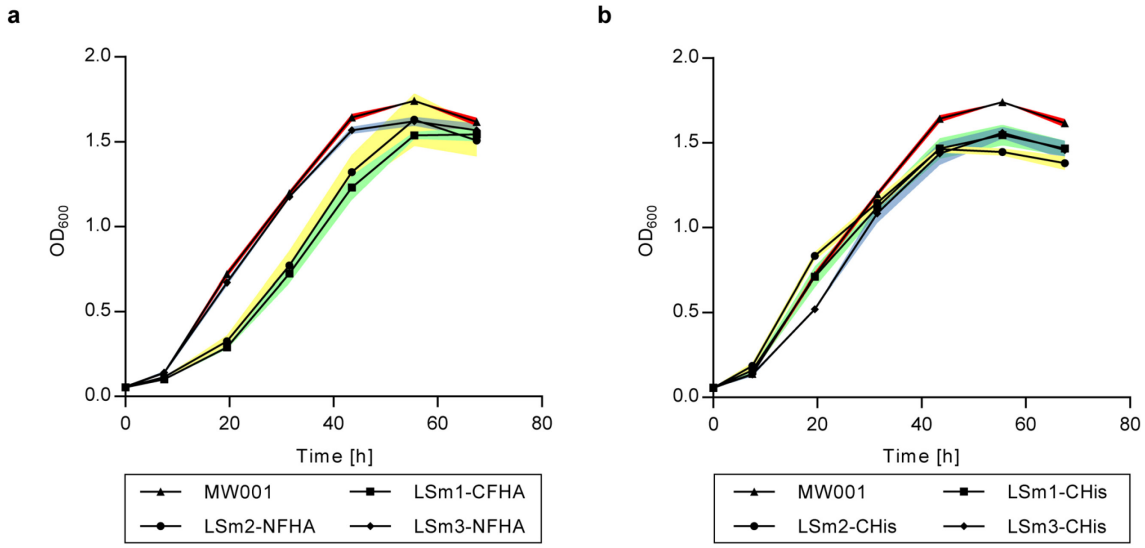
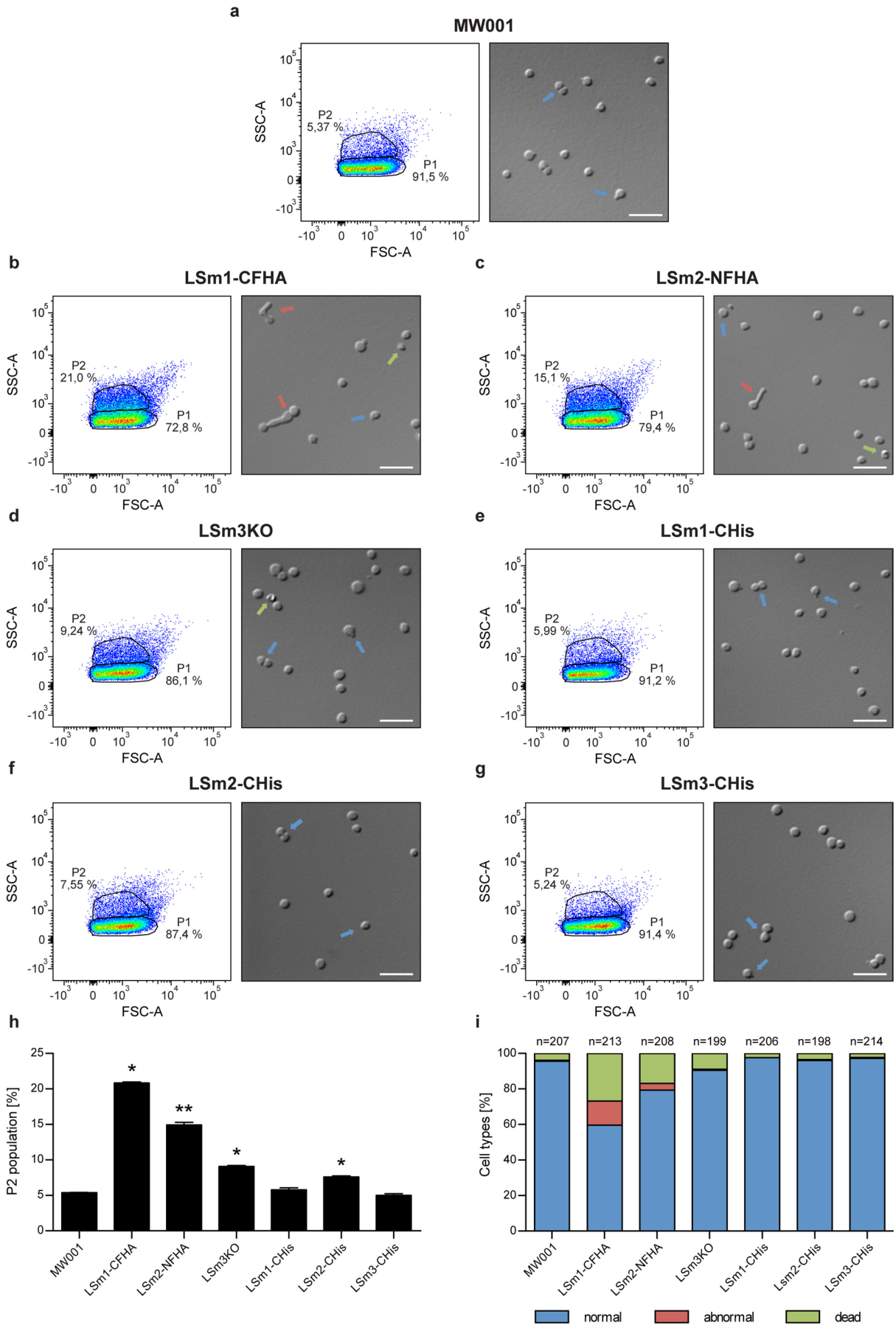


Figure 2.20 Growth curves of *S. acidocaldarius* strains with tagged *lsm* genes. a) and b) The growth behavior of genomically Flag-HA- or His-tagged *lsm* strains, respectively, was compared to the *S. acidocaldarius* MW001 reference strain. Error bars (color filled area) demonstrate the standard deviation of three technical replicates.

2.3.4 Phenotypic analysis of the *lsm* mutant strains

Flag-HA tagging of the *lsm1* and *lsm2* genes resulted in impaired growth of *S. acidocaldarius*, which highlights the important function of these genes (Fig. 2.20a). A flow cytometry analysis was performed to investigate whether the mutations also affect the morphology of the cells. In flow cytometry, the scattering of light is utilized to measure the volume and morphological complexity of cells. This is detected by forward scatter (FSC) and by side scatter (SSC), respectively. The FSC/SSC plots of logarithmically grown MW001 control and the Flag-HA tagged *lsm* strains showed two different populations (Fig. 2.21a-c). Population 1 (P1) constituted the main population, which made up over 90 % of the cells of the control strain. The second population (P2) represented cells of a higher SSC value, which accounted for 5.37 % of the cells of MW001. In contrast, P2 of LSm1-CFHA and LSm2-NFHA showed an increase to 21 and 15.1 % of the cells, respectively, resulting in a decrease of the P1 population. For comparison, the LSm3KO and the His-tagged strains of *lsm1/2/3* were also analyzed in the cytometer. The LSm3KO strain showed a slightly larger P2 population (9.24 %) than the control strain (Fig. 2.21d), whereas the three His-tagged *lsm* strains contained similar P2 sizes as MW001 (Fig. 2.21e-g). Statistical analysis of three technical replicates demonstrated a significant increase (p-value < 0.05) of the P2 populations for the LSm1-CFHA, LSm2-NFHA and LSm3KO strains (Fig. 2.21h). This analysis also revealed a significant increase of the P2 population of LSm2-CHis, although this strain showed P2 values similar to the control.



The legend for this figure is found on the next page.

Figure 2.21 Morphology of *lsm* mutant strains. a-g) The left panels show the flow cytometry analyses (SSC-A/FSC-A plots) of logarithmically grown *S. acidocaldarius* MW001 and *lsm* mutant strains. Two populations, P1 and P2, are marked. The right panels display visual analysis of the respective cultures using light microscopy. Three cell types are indicated by arrows: blue – normal cells (budding and dividing), red – abnormal cells and green – dead cells. Scale bars equal 5 µm. h) The size of the P2 populations is depicted in a bar chart. Error bars demonstrate the standard deviation of three technical replicates. Asterisks (* = Student's t-test; ** = Welch's t-test) indicate the significance (p-value < 0.05) of the data with respect to the MW001 strain. i) The chart illustrates the quantification of the three different cell types that were identified for each strain by microscopy. The number (n) of counted cells is indicated above of each bar.

Next, the utilized cell cultures were investigated by microscopy to visualize the morphological differences indicated by flow cytometry. In contrast to the control strain, which displayed typical dividing *S. acidocaldarius* cocci, the LSm1-CFHA and LSm2-NFHA strains frequently showed cells of abnormal shape (Fig. 2.21a-c) with variations that ranged from long and narrow cells to cells with multiple division sites. Moreover, the two strains contained many small cells of high granularity, which likely formed the P2 population in the flow cytometry analysis. These cells showed no movement and probably represent dead cells. Quantification of the microscopy images revealed that the LSm1-CFHA and the LSm2-NFHA cultures contained around 14 and 4 % abnormal shaped cells and 27 and 17 % dead cells, respectively (Fig. 2.21i). In contrast, the MW001 strain comprised less than 1 % abnormal cells and about 4 % of dead cells. The LSm3KO strain also showed less than 1 % of abnormal cells and around 9 % of dead cells (Fig. 2.21d,i). The three His-tagged LSm strains displayed similar microscopy results and cell shape patterns as the control strain (Fig. 2.21e-g,i).

These results highlight that the mutation of the *lsm1* and *lsm2* genes by the long Flag-HA tag leads to various morphological changes in *S. acidocaldarius*, whereas the mutants comprising the shorter His tag display a similar phenotype as the control strain. The His-tagged strains were therefore utilized for the subsequent purification of the LSm proteins.

2.3.5 Immunoprecipitation of His-tagged LSm proteins from *S. acidocaldarius*

In order to identify the RNA interaction partners of the LSm proteins during optimal growth conditions, the His-tagged *lsm* strains were grown to the logarithmic phase and then purified via Ni-NTA chromatography (Fig. 2.22a). The identity of all three proteins could be verified by western blot analysis using an anti-His antibody (α-His). Interestingly, the LSm2-CHis purification displayed two distinct bands. Western blot analysis demonstrated that the upper band constituted LSm2-CHis. A mass spectrometry approach of the Ni-NTA fraction was performed to determine the identity of the second band (Appendix 5). This analysis identified both, LSm1 and LSm2, among a large number of proteins, indicating that the lower band of the LSm2 purification might constitute untagged LSm1 and that the two proteins could form

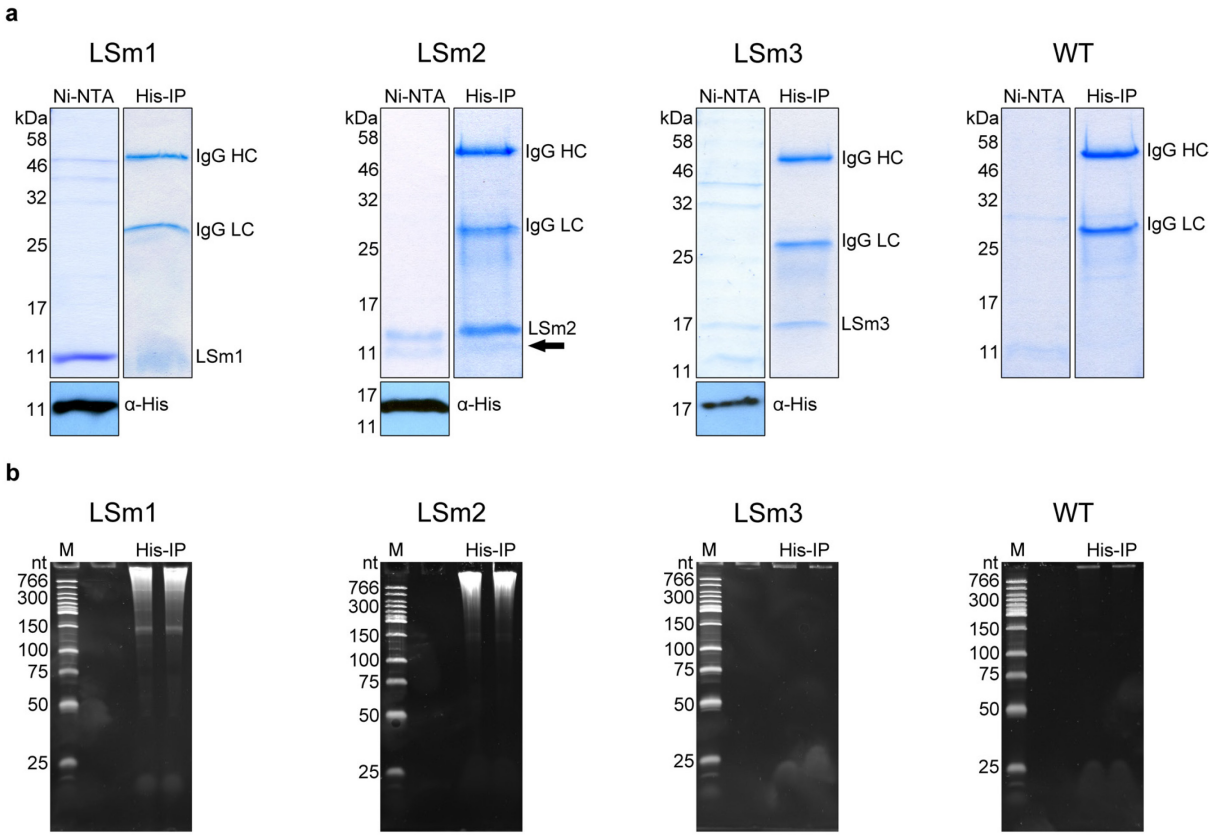


Figure 2.22 Isolation of the LSm proteins and interacting RNAs by immunoprecipitation. a) SDS-PAGE analyses show the purification fractions of the three LSm proteins and a mock purification (WT) after Ni-NTA chromatography (Ni-NTA) and immunoprecipitation using anti-His antibodies (His-IP). Western blot analysis (α-His) verifies the identity of the proteins in the Ni-NTA purification. Protein bands at the expected heights are visible (LSm1-CHis: 9.4 kDa; LSm2-CHis: 10.6 kDa; LSm3-CHis: 16.8 kDa). A second band in the LSm2 purification is marked by an arrow. Protein bands of the heavy (HC) and light chain (LC) of the utilized His-antibody appear in the SDS eluted His-IP. b) Denaturing PAGE analyses display LSm interacting RNAs that were co-immunoprecipitated with the proteins. The LSm1 and LSm2 purifications display RNA bands ranging from 150 to over 1000 nt. No visible amounts of RNA are present for LSm3 and the mock purification (WT).

heterocomplexes within the cell. Due to the moderate purity of the proteins after the Ni-NTA chromatography, a consecutive immunoprecipitation step was performed using anti-His antibodies that were coupled to protein G coated magnetic beads. This approach yielded high purity of the proteins and LSm2 again co-eluted with the smaller sized protein. The LSm-bound magnetic beads were analyzed by mass-spectrometry, however, the LSm2-interacting protein could not be identified. In addition, two attempts to identify the protein from gel extraction failed due to low protein amounts. All immunoprecipitated protein samples were analyzed by PAGE, which revealed large amounts of co-isolated RNA for purified LSm1 and LSm2 (Fig. 2.22b). Higher molecular sized RNAs of over 500 nt were predominantly present in the two purifications. The LSm3 purification, which generally yielded low protein amounts, displayed no visible amounts of nucleic acids, similar to the mock purification. Regardless, the isolated RNA of the three LSm and mock purifications was used to generate cDNA libraries. Fragmentation of the RNAs by $ZnCl_2$ (to include long RNAs like mRNAs) and size selection

of the produced cDNA libraries were performed as described in chapter 2.2.1. The following libraries were prepared for Illumina HiSeq2500 sequencing: i) a duplicate of LSm-interacting sRNAs (LSm1/2/3 sRNA), ii) a duplicate of fragmented long RNAs (LSm1/2/3 frag RNA) and iii) a triplicate prepared from mock purifications (WT control).

2.3.6 Identification of the LSm-interacting RNAs using RIP-Seq analysis

After sequencing, the RIP-Seq data sets were mapped onto the *S. acidocaldarius* genome. Over 90 % of the obtained sequencing reads of the LSm1 and LSm2 libraries could be mapped, whereas this was possible for only 10 % of the LSm3 and WT control reads (see Table 4.23 in chapter 4.9.1). Similar to the WT control data set for the L7Ae RIP-Seq analysis, the remaining 90 % of the reads of the two data sets originated from contaminating material caused by the extremely low RNA input in the cDNA library preparation protocol (see chapter 2.2.2).

In total, 48 million mapped reads were obtained. The coverage plots of the LSm1 and LSm2 libraries differed significantly from the one of the sRNome (Fig. 2.23). The identity of the highest peaks was investigated and revealed mainly fragments of mRNAs. Interestingly, many of these peaks were found for both the LSm1 and LSm2 data sets, demonstrating that the two proteins share RNA binding partners. In addition, several highly abundant non-coding RNAs were found, such as C/D box sRNAs, the RrrR RNA and a single crRNA, which were predominantly present in the LSm1 sRNA data set. As for the L7Ae WT control data set (see chapter 2.2.2), the LSm WT control data set mainly comprised peaks for RNAs that were equally overrepresented in the sRNome analysis, which shows that unspecific RNAs were isolated during the mock purification. The mapping pattern of the LSm3 sRNA data set strongly resembled the plot of the WT control and the fragmented data set only contained two peaks for the ribosomal RNAs of *S. acidocaldarius*. This demonstrates that no LSm3-specific RNAs were co-isolated during immunoprecipitation of His-tagged LSm3.

Using the DESeq2 tool, our collaboration partner Michael Uhl performed a peak calling approach to identify RNAs, which were significantly enriched over unspecific RNAs in the WT control data sets [150]. This analysis revealed 1216 LSm1 sRNA, 1003 LSm1 frag RNA, 996 LSm2 sRNA and 1051 LSm2 frag RNA peak calls (q-value < 0.05). Subsequently, the called peaks were investigated for overlaps with coding genes. In total, 873 and 562 enriched genes were found for the LSm1 and LSm2 data sets, which corresponds to 39 and 25 % of the coding genes (2224) of *S. acidocaldarius*, respectively. 502 of the enriched genes were found to be shared by both LSm proteins. In contrast, only 2 LSm3 sRNA and 21 LSm3 frag RNA peak calls were obtained, which mainly represented falsely mapped reads.

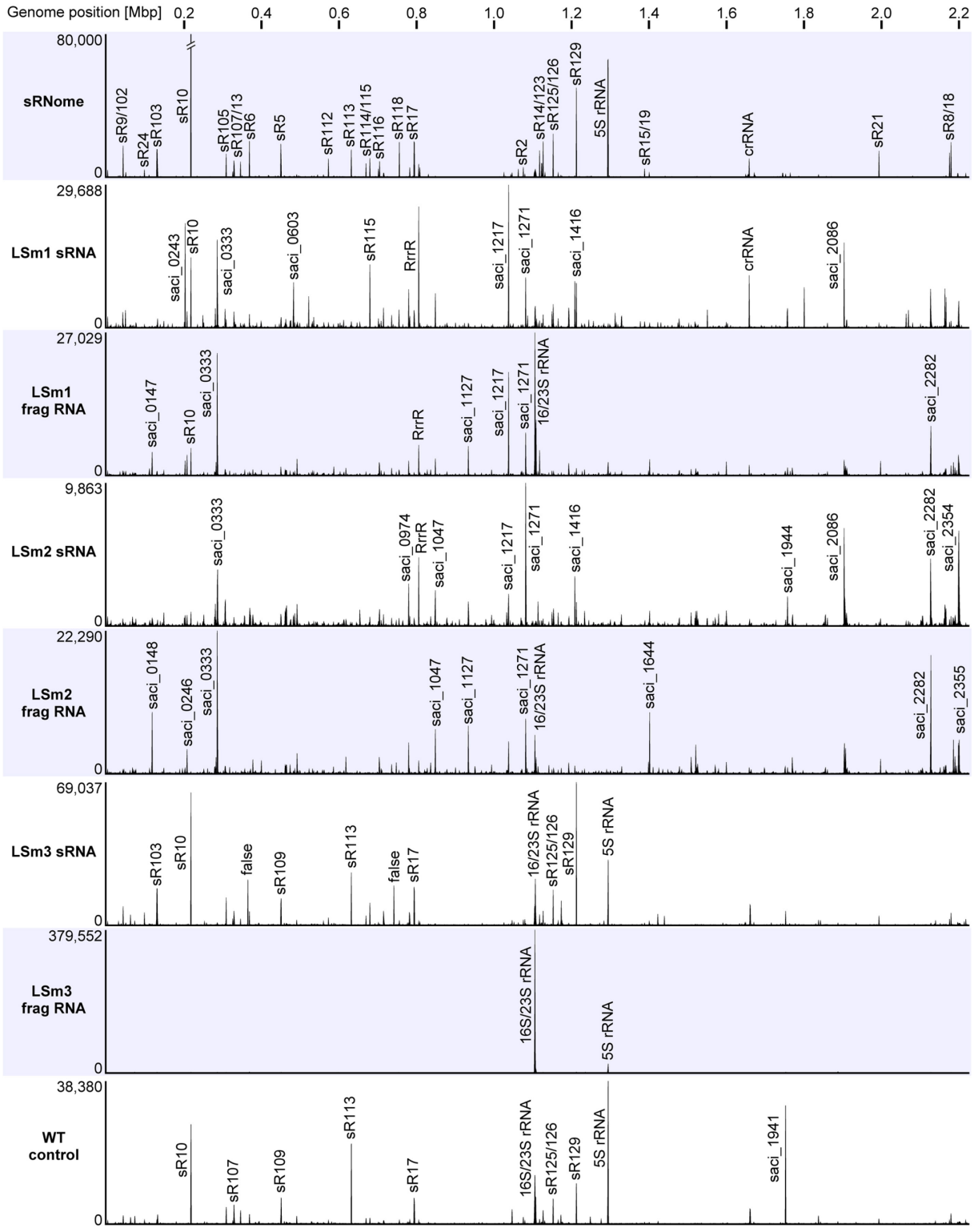


Figure 2.23 Global overview of the LSm-RNA interactome of *S. acidocaldarius.* The mapping patterns of the sequenced LSm sRNA, fragmented RNA and WT control cDNA libraries are depicted in coverage plots. Highly abundant reads are illustrated as distinct peaks. The sRNome was integrated as a reference. Peaks representing C/D box sRNAs are labeled with (Sac-)sR# and mRNAs by gene locus tags (saci_#). For clarity, only the 10-12 highest peaks of the LSm data sets were marked. Each RNA profile contains one million mapped reads. The LSm3 sRNA and WT control profiles contain 527,000 and 476,000 mapped reads, respectively.

In order to identify more relevant interaction partners of the two LSm proteins, a smaller subset was analyzed, which encompassed the 100 most significant peak calls of each LSm1 and LSm2 data set (in total 400 peaks). Manual inspection of these peaks revealed that the reads were often enriched in the 3' vicinity of genes. This was particularly true for the sRNA data sets of LSm1 (57 %) and LSm2 (28 %). Recently, Dar and co-workers performed a Term-Seq analysis in *S. acidocaldarius* to study the 3' termini of RNA transcripts and to identify signals of transcriptional termination [157]. Utilizing this data set (European Nucleotide Archive, accession no. PRJEB14292), it became apparent that the LSm-enriched regions mainly represented the 3' termini of RNAs (Fig. 2.24). Thus, the two LSm proteins seem to bind transcripts at their 3' ends. The LSm data sets even identified the 3' termini of RNAs, which were underrepresented in the Term-Seq approach, but were bound by the LSm proteins, e.g. the *saci_2218* mRNA (Fig. 2.24). The LSm1 and LSm2 frag RNA data sets displayed enriched 3' termini for only 7 and 11 of the 100 most significant peaks, respectively, due to the fragmentation of the bound transcripts. The reads of these data sets were more evenly distributed along the genes.

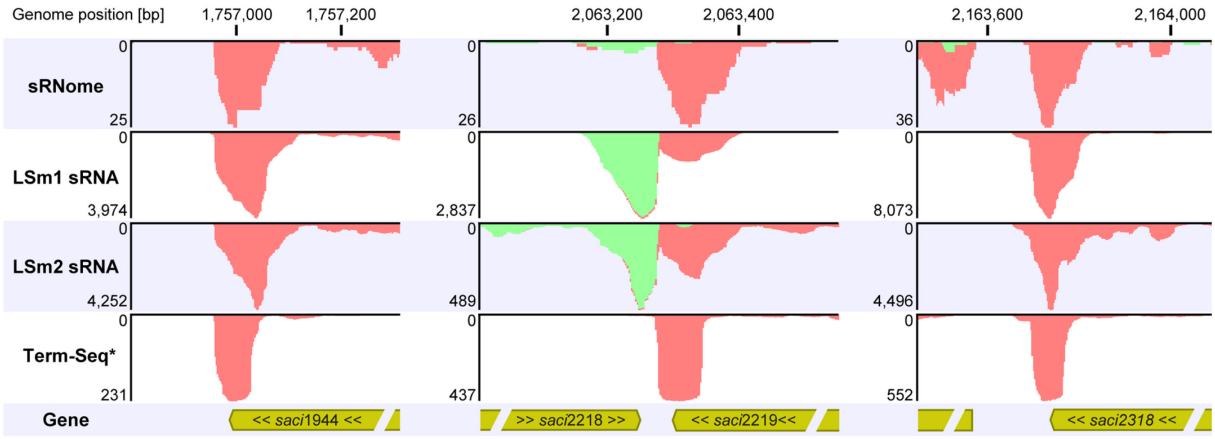


Figure 2.24 LSm1 and LSm2 bind to the 3' terminus of RNAs. The coverage plots at the 3' termini of the coding sequences (yellow arrows) of different genes are shown for the sRNome, the LSm1/LSm2 sRNA and the Term-Seq data sets. The mapping graphs of the Term-Seq data set represent the 3' termini of the respective mRNAs. The LSm1 and LSm2 data sets display high read numbers (y-axis) at the same region. Green (sense) and red (antisense) mapping graphs illustrate the orientation of the reads. One million reads were mapped per data set. The Term-Seq data set (*) was obtained from Dar *et al.* (ENA; accession no. PRJEB14292) [157].

Next, the identity of the most significant RNAs was investigated. 96 % of the LSm1 sRNA and frag RNA data set peaks belonged to mRNAs (191 fragments of 108 different mRNAs). Only 8 non-coding RNAs were present in the two subsets: the RrrR RNA, the single crRNA, Sac-sR115, RNase P RNA and antisense RNAs of *saci_0383/0887/1259/1747*. Additionally, the LSm2 sRNA and frag RNA data sets displayed a drastic overrepresentation (99 %) of mRNAs (198 fragments of 85 different mRNAs). The RrrR RNA (chapter 2.1.1) constituted the only non-coding RNA among the most significant peaks of LSm2. Noteworthy, this RNA belonged

to the top 10 of the most significant peaks for both sRNA data sets of LSm1 and LSm2. Both LSm1 and LSm2 displayed enrichment of exclusively the RrrR RNA sense strand, in contrast to the overrepresentation of the antisense RNA in the sRNome (the RNA is shown in Fig. 2.27b). Direct comparison of the subsets revealed that 60 mRNAs were shared between the proteins. 48 and 25 mRNAs were enriched for either LSm1 or LSm2 (Fig. 2.25).

The protein function of the identified mRNAs was investigated. The EggNOG database (version 4.5) was used to categorize the coding genes/encoded proteins into arCOGs (archaeal clusters of orthologous groups), which were established to classify archaeal proteins by their function [158-161]. Most mRNAs of the three different groups (LSm1-specific, LSm2-specific and LSm1&LSm2) belonged to arCOG class S (unknown protein function), which also represented the largest group in the *S. acidocaldarius* genome. Messenger RNAs that encoded proteins involved in energy conversion (arCOG class C) constituted the second largest fraction of the LSm2-specific and LSm1&LSm2-grouped mRNAs. Similarly, arCOG class C formed the second largest group in the *S. acidocaldarius* genome. In general, a corresponding representation of arCOG classes was observed for the three LSm groups and for the *S. acidocaldarius* genome. No specific functional class was significantly overrepresented in one of the three groups. However, the analysis is complicated by the high number of mRNAs encoding for proteins with unknown function.

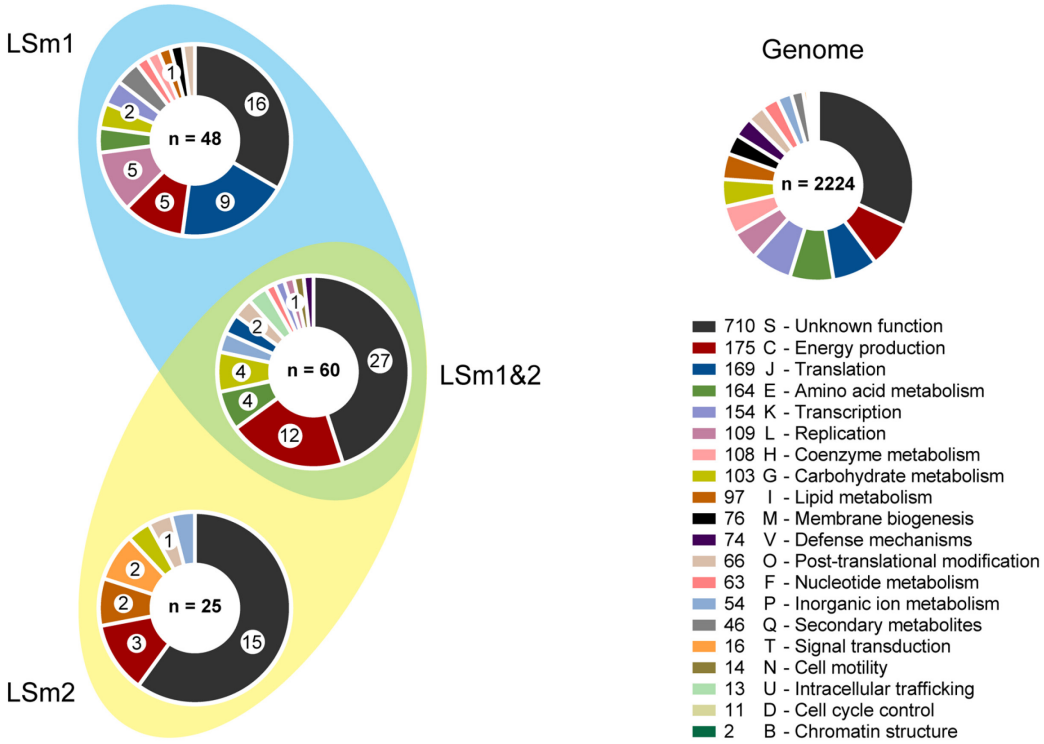


Figure 2.25 Functional classification of highly enriched LSm-interacting mRNAs. The Venn diagram (left) illustrates the number (n) of enriched mRNAs that were found to interact with either LSm1 (blue), LSm2 (yellow) or both proteins (green) and which were functionally categorized into arCOGs (pie charts). The pie chart on the right (genome) displays the classification of all identified *S. acidocaldarius* genes (2224 genes).

The RIP-Seq mapping patterns suggest that the LSm proteins bind RNAs at their 3' termini. Furthermore, the most significant interactors of the proteins were mainly mRNAs which belonged to arCOG classes that were well-represented in the genome. In order to determine whether the most significant peaks of the LSm data sets constituted merely highly transcribed RNAs, which were unspecifically bound at their 3' ends, the transcriptional level of the identified RNAs was investigated. To this end, an RNA-Seq study of Märtens and co-workers was utilized, in which the transcriptome of logarithmically grown *S. acidocaldarius* MW001 cells was examined [162]. The 108 and 85 identified interacting mRNAs of LSm1 and LSm2 displayed highly diverse expression levels, respectively (Fig. 2.26). No correlation between the significance (= enrichment of an RNA in the LSm data) of the called peaks and the transcriptional level could be observed. For comparison, the expression level of the 108 highest transcribed RNAs in *S. acidocaldarius* was investigated (Fig. 2.26; black line). Only few interactors of LSm1 and LSm2 belonged to these most expressed RNAs. The majority of the LSm interactors displayed expression levels that were 10- to 100-fold lower than the highest transcribed RNAs in *S. acidocaldarius*. This analysis concludes that mRNA interactors of the LSm proteins were not mainly enriched due to their strong transcription within the cell.

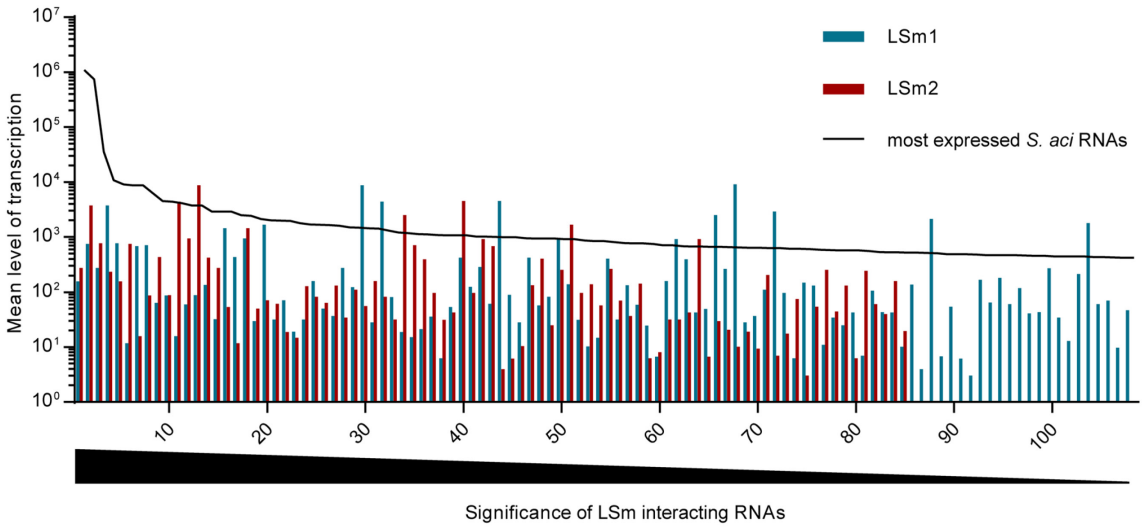


Figure 2.26 Transcription levels of the most significant LSm interacting RNAs. The bar chart shows the transcriptional level of the 108 and 85 identified interacting mRNAs of LSm1 (blue) and LSm2 (red), respectively. The RNAs are sorted by their significance (from left to right). A black line demonstrates the transcriptional level of the 108 most expressed RNAs of *S. acidocaldarius*. The data of *S. acidocaldarius* RNA transcription levels was obtained from Märtens *et al.* [162].

2.3.7 Identification of two potential RNA binding motifs for LSm1 and LSm2

The RIP-Seq analyses of LSm1 and LSm2 displayed defined regions of enriched reads, as illustrated by the 3' termini of RNAs (see Fig. 2.24). A motif-based sequence analysis was performed in order to identify potential LSm binding sites within these regions. Our collaboration partner Michael Uhl utilized the MEME tool (Multiple Expectation-Maximation

for Motif Elicitation) to investigate the called peaks that were identified during the DESeq2 analysis. MEME performs multiple sequence alignments to discover statistically significant motifs in the input sequences [163, 164]. This search revealed two significant sequence motifs that were found both for the LSm1 and LSm2 data sets: i) the U motif and ii) the UAG motif.

The consensus sequence of the first motif, the U motif, was composed of a 22 nt long stretch of highly conserved uracil residues (Fig. 2.27a). The U motif was found 253 (e-value: 6.9e-034) and 175 (e-value 3.9e-002) times in the called peaks of the LSm1 and LSm2 sRNA data sets, respectively. Localization of these motifs demonstrated that their presence was strongly biased towards the 3' termini of RNAs, as exemplified for the non-coding RNA RrrR (Fig. 2.27b). The Term-Seq analysis of Dar and co-workers revealed a termination consensus signal, which strongly resembled the identified U motif of the LSm proteins (Fig. 2.27a) [157]. A long stretch of U residues was present from position -21 towards the 3' terminus of RNAs. Similarly, a higher conservation of U bases existed between positions -11 and -21 (positions 1 to 7 for U motif). Remarkably, the second last residue of the U motif (position 21) showed a conserved cysteine which was also conserved at position -1 in the termination consensus sequence.

The second motif, the UAG motif, was identified in each of the four data sets (LSm1 sRNA/ frag RNA: 100/151 times, e-value: 1.1e-012/9.7e-043; LSm2 sRNA/frag RNA: 231/212 times, e-value: 3.4.e-057/3.4.e-055). It displayed a generally low conservation with the exception of two guanine residues at positions 3 and 15 (Fig. 2.27c). Two moderately conserved UA residues were found upstream of these positions. Inspection of the sequence revealed that five UAG triplets were present along the 25 nt consensus sequence (positions 1, 10, 13, 19 and 22) and partial UAG triplets existed for the nucleotide positions 6, 17 and 18. The motif was almost exclusively (94 %) found in the coding sequence of mRNAs. Within these mRNAs, the identified motifs were not clustered at start or stops of the coding sequences. The motif was identified in several highly enriched mRNAs, e.g. within the *saci_1217* mRNA, which represented the most enriched RNA of the LSm1 data set (Fig. 2.27d).

These results indicate that the LSm1 and LSm2 proteins might contain two binding sites for RNA molecules. One site presumably interacts with the U-rich termination signal of RNA transcripts, whereas the other site might recognize the coding sequence of mRNAs via the identified UAG motif.

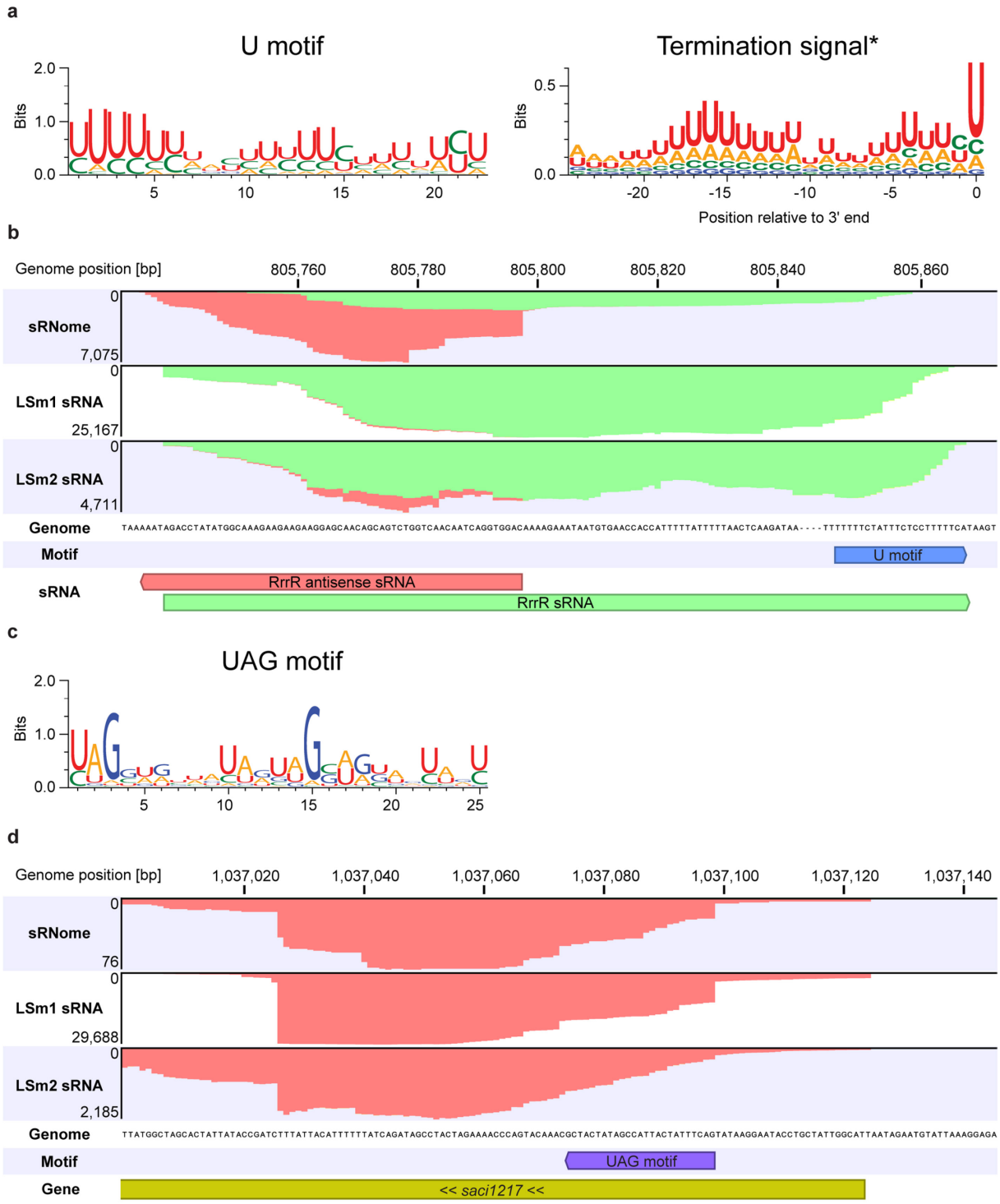


Figure 2.27 Two potential RNA binding motifs for LSm1 and LSm2. a) Sequence logos illustrate the consensus sequences of the identified U motif (left) of the LSm1 sRNA data set and the *S. acidocaldarius* transcription termination signal (right). The height of each base stack indicates the conservation at the respective position (measured in bits) and the height of the symbols (base residue) shows the relative frequency of the corresponding base at that position [165]. The sequence logo of the *S. acidocaldarius* termination signal (*) was obtained from Dar *et al.* [157]. b) The coverage plots for the non-coding RNA RrrR (green – forward reads) and its antisense RNA (red – reverse reads) are shown. The sRNome track displays an overrepresentation of the antisense RNA, whereas the two LSm1 and LSm2 sRNA plots only show enrichment of the RrrR RNA. The U motif is present at the 3' terminus of the RrrR RNA. c) The consensus sequence of the 25 nt long UAG motif of the LSm2 sRNA data set is depicted by a sequence logo. d) A UAG motif is present within a *saci_1217* mRNA fragment, which is highly enriched in the LSm1 sRNA data set.

2.4 Subsequent findings of the LSm-RNA interactome study

The experiments of the following paragraph were performed by Alexander Pastura (master student under my supervision) and are based on the findings presented in chapter 2.3. These results are briefly stated in this work to logically conclude the LSm-RNA interactome study.

2.4.1 LSm1 and LSm2 interaction study

Mass spectrometry indicated that the co-purified protein (second band) of the LSm2-CHis Ni-NTA purification constituted LSm1 (chapter 2.3.5; see Fig. 2.22a). Furthermore, the finding of many identical interaction partners in the RIP-Seq analysis provided hints that the two LSm proteins might interact in the cell (chapter 2.3.6). In order to test this hypothesis, the two *lsm* genes were recombinantly expressed in *E. coli* and analyzed for co-purification. First, both C-terminally His-tagged proteins were individually purified from *E. coli* via Ni-NTA chromatography (Fig. 2.28a; left panel). LSm2-CHis co-purified with large amounts of *E. coli* RNA, whereas no visible amounts of nucleic acids were observed for the LSm1-CHis purification (Fig. 2.28a; right panel). Next, C-terminally His-tagged *lsm1* and C-terminally Flag-tagged *lsm2* were provided on a single vector and co-expressed in *E. coli*. Purification of the LSm1-CHis protein via Ni-NTA chromatography co-isolated high amounts of LSm2-CFlag (Fig. 2.28a, left panel), demonstrating that these two proteins can form heterocomplexes. This purification also displayed high amounts of nucleic acid in an agarose gel (Fig. 2.28a, right panel). A control purification of LSm2-CFlag via Ni-NTA chromatography yielded no protein and no nucleic acids. In order to determine the oligomeric state of the LSm1-CHis/LSm2-CFlag heterocomplex, the purified proteins were analyzed by size exclusion chromatography. A high peak at 280 nm was obtained at an estimated size of 72 kDa, which corresponds to the expected size of a heptameric ring formed by LSm1-LSm2 (~70 kDa) (Fig. 2.28b). SDS-PAGE and western blot analysis verified that both proteins were present in this peak fraction.

These results suggest that LSm1 and LSm2 can form heterocomplexes in *S. acidocaldarius* and that LSm1 does not bind nucleic acid without the presence of LSm2. LSm2, however, is also capable of binding RNA in the absence of LSm1.

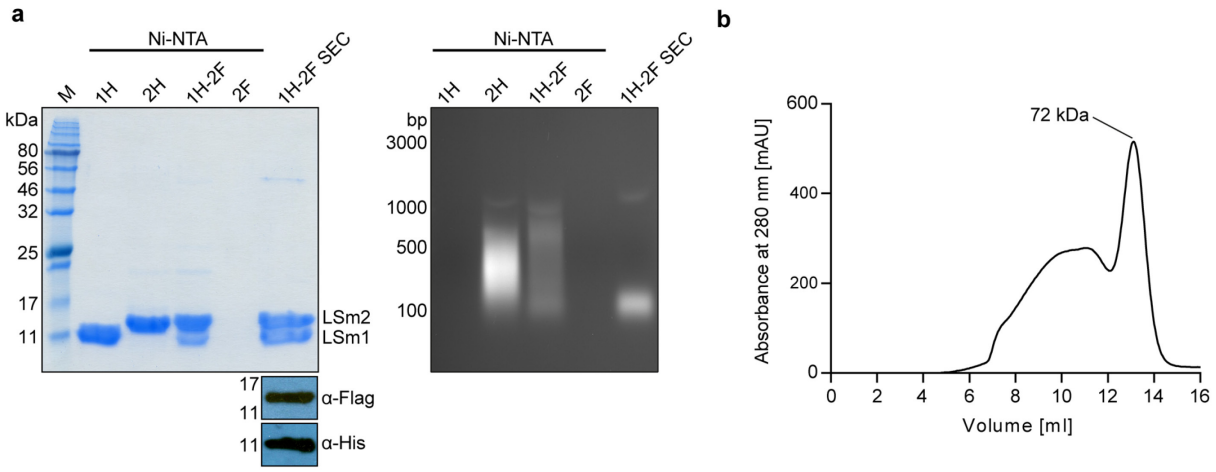


Figure 2.28 Heterocomplex formation of LSm1 and LSm2. a) SDS-PAGE analysis (left panel) shows the Ni-NTA purification of the following tagged LSm proteins: 1H – LSm1-CHis (9.3 kDa), 2H – LSm2-CHis (10.8 kDa), 1H-2F – LSm1-CHis&LSm2-CFlag (9.3 kDa & 10.7 kDa), 2F – LSm2-CFlag (10.7 kDa). The right lane (1H-2F SEC) represents the peak fraction (13 ml) of the size exclusion chromatography (SEC) analysis of the 1H-2F Ni-NTA purification shown in b). Flag-tagged LSm2 and His-tagged LSm1 were verified in this fraction by western blot analysis. The right panel displays agarose gel electrophoresis of the same purification fractions. *E. coli* RNA is visible in lanes 2H, 1H-2F and 1H-2F SEC. 5 µg of protein were loaded in each lane for both gels (except 2F). b) The graph illustrates the size exclusion chromatography analysis of the 1H-2F Ni-NTA purification. A high peak is obtained at 13 ml, which corresponds to an estimated molecular weight of 72 kDa. The experiments were performed by Alexander Pastura.

2.4.2 Binding analysis of the U and UAG motifs by EMSAs

Finally, the two identified RNA motifs of the RIP-Seq analysis were tested for binding by the LSm1-LSm2 heterocomplex. The assayed U motif substrate represented the termination signal (last 25 nt) of the RrrR RNA (see Fig. 2.27b), which contained 17 uracil residues. The EMSA displayed a high affinity of the LSm1-LSm2 heterocomplex for this substrate. Binding was observed at a low concentration of 25 nM LSm1-LSm2 and full binding of the substrate was present at around 400 nM (calculated for a heptameric ring) (Fig. 2.29). At higher protein concentrations (≥400 nM) the appearance of a second shift became visible, likely representing a higher molecular weight conformation of the protein-RNA complex. The 25 nt long motif of the *saci_1217* mRNA was utilized as the UAG motif substrate (see Fig. 2.27d). In contrast to the U motif, this RNA showed weak binding by the complex (Fig. 2.29). Full binding of this substrate could not be achieved, even with a protein concentration of 6000 nM. However, a variant of this substrate, which contained mutations of the two conserved G residues and alterations of the UAG triplets, displayed no binding. At last, a substrate that comprised eight triplets of UAG was tested in the EMSA. This polyUAG substrate displayed a similar binding affinity as for the native sequence from the *saci_1217* mRNA.

These binding studies demonstrate that the LSm1-LSm2 complex of *S. acidocaldarius* binds to the termination signal of RNAs. The general low affinity for the tested *saci_1217* fragment suggests that other factors might be required for efficient binding of the UAG motif.

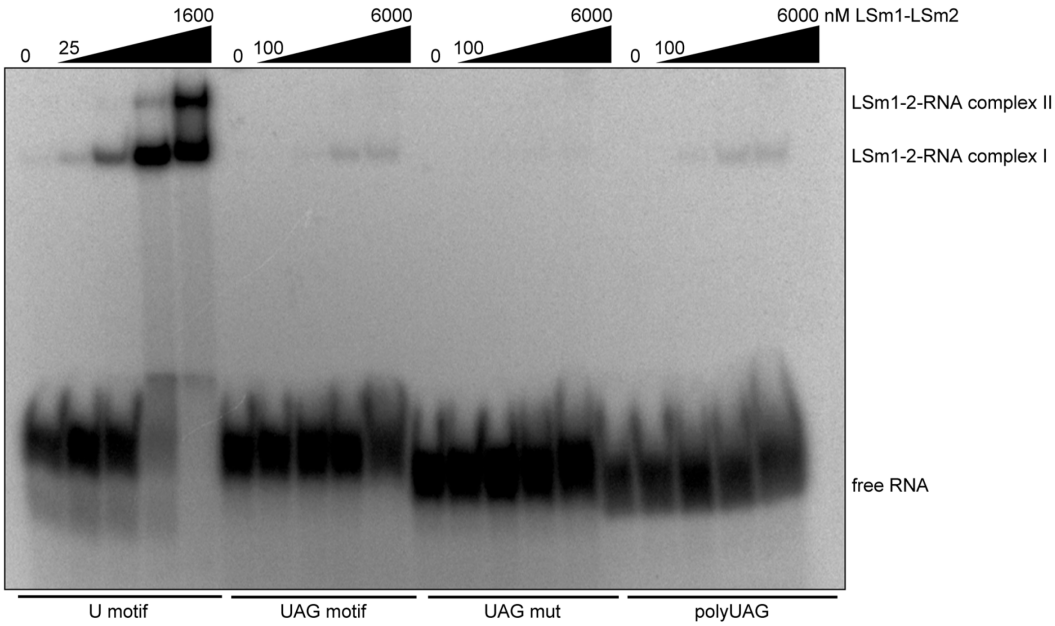


Figure 2.29 Binding study of the U and UAG motifs by the LSm1-LSm2 heterocomplex. EMSA studies highlight the binding of several RNA substrates by the LSm1-LSm2 heterocomplex. Binding of the U motif is observed with 25 nM LSm1-LSm2 and increasing protein concentrations (25, 100, 400, 1600 nM LSm1-LSm2). Protein concentrations of 100, 400, 1600 and 6000 nM were tested for the UAG motif, UAG mut (mutant) and polyUAG motif substrates. Binding of the UAG and polyUAG motif is observed with 400 nM LSm1-LSm2. The experiment was performed by Alexander Pastura.

3. Discussion

3.1 RIP-Seq verifies RNA interaction partners of L7Ae in *S. acidocaldarius*

The ribosomal protein L7Ae is an archaeal RNA-binding protein. It stabilizes the ribosomal RNA by recognizing a structural motif, termed k-turn [28]. The k-turn motifs are widespread and also found in other functional RNAs [28]. Previous studies demonstrated that L7Ae is an RNA structuring subunit of C/D box, H/ACA box and RNase P RNP complexes in archaea [33, 89, 92]. In this thesis work, RIP-Seq analysis of the L7Ae protein of *S. acidocaldarius* was utilized to identify its RNA interaction partners. All previously reported non-coding RNA interactors of archaeal L7Ae were found to be enriched in the obtained data sets. In addition, novel RNA interaction partners were identified by this technique. The potential roles of L7Ae interactions with these RNAs (i.e. SRP RNA and mRNAs) will be discussed below.

3.2 Archaeal SRP RNA – a novel RNA interactor of L7Ae?

The SRP RNA is a component of the signal recognition particle RNP (SRP RNP) which is involved in the trafficking of proteins in all domains of life. It recognizes the signal peptide of nascent proteins at the ribosome, halts translation and directs the ribosome to the cell membrane or ER to mediate the secretion of nascent proteins [166]. Accurate bending of the RNA is crucial for the function of the SRP RNP, which allows simultaneous interaction of the RNA with distant sites of the ribosome, i.e. the nascent chain exit site and the elongation factor binding site [167]. Bending occurs within helix 5, which links the two functional domains of the SRP RNA, the small domain (Alu) and large domain (S) (Fig. 2.15a) [167]. The 5e motif in helix 5, which is conserved in the SRP RNAs of all domains of life, was suggested to be the bending site within the RNA [168]. The motif consists of four base pairs that are interrupted by a three-nucleotide bulge (Fig. 2.15b) [169]. In human SRP RNA, the 5e motif was recently shown to adopt a k-turn-resembling structure, which, however, fundamentally deviates from classical k-turns due to the absence of the G•A pairs [73]. The observed structure is stabilized by a central potassium ion and was therefore termed K^+-turn [73]. The eukaryotic SRP72 protein specifically binds the K^+-turn and recognizes a conserved adenosine in the 5e bulge [170]. The RNA is kinked in the absence of SRP72, however SRP72 binding to the 5e motif was proposed to alter the bending angle of the RNA [171].

The archaeal SRP RNA was identified as a non-coding RNA binding partner of L7Ae by the utilized RIP-Seq approach. In line with this observation, this essential RNA was also found in a previous L7Ae immunoprecipitation study performed in *S. solfataricus* [92]. Using EMSA

studies and toe-printing assays, Zago and co-workers tested binding to full-length and truncated SRP RNA substrates and revealed L7Ae binding to a k-turn motif in close proximity to the 5e motif. A similar k-turn structure could also be found in *S. acidocaldarius* and EMSA analysis confirmed binding by L7Ae (Fig. 2.14d,e). Multiple sequence alignments (Appendix 4) and manual folding of helix 5 regions from phylogenetically distant archaea revealed that this putative k-turn is conserved among archaeal SRP RNAs and located in hinge 2 in a distance of 9-10 base pairs to the 5e motif (Fig. 2.15).

The archaeal SRP RNP is more basic than the protein-rich eukaryotic SRP RNP and is composed of the ~300 nt SRP RNA and the two proteins SRP19 and SRP54 [172]. A homolog of the eukaryotic SRP72 protein is absent in archaea [168]. It is therefore conceivable that the L7Ae protein substitutes for the role of SRP72 in archaeal SRP RNAs to mediate its bending. It appears likely that the 5e motif cannot serve as a substrate for L7Ae as it lacks structural hallmarks in the K^+-turn, e.g. the G•A pairs. Consequently, archaeal SRP RNAs would benefit from a classical k-turn in proximity to the 5e motif to allow L7Ae binding. However, this scenario raises the question, why the 5e motif is conserved in archaeal SRP RNAs. It is possible that L7Ae binding of the k-turn in hinge 2 is not sufficient for correct bending of the RNA. A k-turn in RNA introduces a kink with an angle of around 60° [71]. A cryo-EM analysis of full-length mammalian SRP RNP bound to the ribosome demonstrated that the bent RNA forms a larger angle of 90° [167]. The crystal structure of the *M. jannaschii* SRP RNA S domain revealed that the 5e motif bends the RNA in an angle of 30° [173]. The 5e motif could thus increase the bending of the RNA to the required angle. Presumably, both the L7Ae-bound k-turn at hinge 2 and the K^+-turn at the 5e motif are needed for the proper bending of the SRP RNA. Full-length crystal structures of archaeal SRP RNPs would be needed to verify this.

Could L7Ae be a core protein of archaeal SRP RNPs? Only few *in vivo* analyses were conducted on archaeal SRP RNPs in the past. Two studies performed a purification of SRP RNPs from archaeal cells by immunoprecipitation of SRP54 [174, 175]. Both studies co-isolated several proteins with SRP54, however, the identity of these proteins was not solved. The mass spectrometry analysis that was performed in the present work did not contain the SRP19 or SRP54 proteins. This analysis, however, aimed exclusively at the identification of the tagged L7Ae protein and utilized gel-extracted protein bands, which reduces the sensitivity of protein identification. Other known interactors of L7Ae were also absent in this approach, e.g. core proteins of the RNase P RNP. Repeating these analyses from soluble L7Ae isolates would be necessary to identify possible protein additions to the core SRP RNP.

Archaeal SRP research mainly focused on the analysis of crystal structures, which involved the heterologous production of SRP RNP components. The obtained structures revealed that SRP19 and SRP54 bind to the S domain, leaving an unprotected helix 5 and Alu domain, which could allow binding by L7Ae [173]. These observations, together with the findings of this work, provide a good basis for a future analysis of L7Ae-SRP RNP interactions.

3.3 L7Ae autoregulation as an adaption mechanism to changing C/D box sRNA pools

RIP-Seq analysis revealed 32 mRNA fragments that were enriched with L7Ae. The finding of putative k-turn sequences in 20 of the fragments led to the hypothesis that L7Ae is able to regulate the translation of these mRNAs. *In vivo* studies confirmed this hypothesis by demonstrating a downregulation of fluorescence intensity upon L7Ae induction for fusion constructs of three different putative mRNA k-turns to GFP. This suggests that archaeal L7Ae exhibits a second, previously unknown function by regulating mRNA translation.

One of the tested mRNA k-turns was found in the 5' UTR of L7Ae's own transcript, which suggests that L7Ae is an autoregulatory protein. Feedback regulation of ribosomal proteins is often found in nature and is thought to support a balanced and coordinated synthesis of proteins and ribosomal RNAs [176]. These autoregulatory ribosomal proteins usually recognize an RNA structure in their mRNA, which is structurally similar to their binding site in the ribosomal RNA [177]. However, L7Ae is not only a ribosomal protein, but fulfills many other important functions in the cell, which involve binding to a wide spectrum of non-coding RNAs. L7Ae autoregulation is therefore not only necessary to react to ribosomal RNA levels, but also to the high numbers of other non-coding RNAs in the cell, in particular the C/D box sRNAs. It is plausible that L7Ae is required to deal with changing C/D box sRNA pools in archaeal cells. This RNA family constitutes a large portion of L7Ae interactors, especially in thermophilic archaea, and the presence of highly variable C/D box sRNA genes in closely related archaeal organisms suggests their fast evolution [40]. Recently, our laboratory showed that C/D box sRNAs are often co-transcribed with adjacent genes whose promoters are “hijacked” for C/D box sRNA production [143]. This scenario can also explain the observed differential C/D box sRNA expression pattern in the *S. acidocaldarius* sRNome. The C/D box sRNA abundance is further coupled to the regulation of the “hijacked” promoters and environmental changes can have drastic effects on their production. Autoregulation of L7Ae allows to adjust its abundance in response to the availability of C/D box sRNA substrates. This theory is supported by the finding that the L7Ae feedback regulation could be augmented by deletion of the highly abundant C/D box sRNA Sac-sR10 (Fig. 2.9). Gene regulation by translational control also

constitutes a direct way to rapidly respond to environmental changes [178]. Remarkably, a k-turn that was verified to bind to L7Ae was also found in the 5' UTR of the *saci_1347* mRNA, which encodes the Nop5 protein. The *saci_1347* mRNA is part of a polycistronic transcript (see Table 2.1) and the downstream mRNA, *saci_1346*, encodes the fibrillarin protein. L7Ae can thus adjust the concentration of all three C/D box sRNP core protein components to the cellular C/D box sRNA levels. The identification of a k-turn in the bicistronic transcript of *nop5-fibrillarin* from *T. kodakarensis* provides support that this L7Ae regulation of Nop5 and fibrillarin may be conserved among archaea [179]. Jäger and co-workers also identified a k-turn in the transcript of *cbf5* in this organism, which encodes the catalytic subunit of H/ACA box sRNPs. This k-turn could not be identified in *S. acidocaldarius* and the gene´s transcript was not enriched in the RIP-Seq data. Nevertheless, L7Ae might also regulate the abundance of these sRNP complexes in other archaea.

L7Ae binds k-turn elements in three essential non-coding RNAs: ribosomal RNAs, RNase P RNA and presumably SRP RNA. These RNAs, in particular the rRNAs, are highly abundant within the cell. L7Ae is therefore needed in large amounts to allow proper ribosome and RNP assembly. This requires a strong promoter and an autoregulation mechanism that only represses L7Ae production after these three essential substrates are covered by L7Ae. The competition analysis (Fig. 2.17) indicates that these essential substrates are the preferred binding partners of L7Ae, as a sRNA-depleted total RNA sample that mostly contained 16S and 23S rRNAs efficiently outcompeted the *l7ae* 5' UTR binding. In contrast, an elevated amount of C/D box sRNA Sac-sR121 could not affect the binding significantly, which, at first glance, may seem contradictory to the above theory that L7Ae autoregulation is used to react to changing C/D box sRNA levels. The total loss of 2'-*O*-methylation in rRNAs was shown to be lethal [50, 180]. However, deletions of individual C/D box sRNAs were found to have no significant effect on the growth of *S. acidocaldarius* [152]. Full binding of the entire pool of C/D box sRNAs in the cell might not be necessary for efficient 2'-*O*-methylation and presumably not feasible considering that one *l7ae* gene is outnumbered by 62 C/D box sRNA genes. Accordingly, the L7Ae autoregulation mechanism may rather be used for a rough adjustment to changing C/D box sRNA levels than enabling L7Ae concentrations that allow the binding of each L7Ae substrate in the cell. This is supported by the finding that eliminating the autoregulation in *S. acidocaldarius* results in a rather moderate overproduction of L7Ae (~2-fold) as indicated by the Kt-n and Kt-b mutants of the β-galactosidase and GFP reporter assays. The L7Ae feedback regulation is therefore not a tight regulation in *S. acidocaldarius*, which is plausible considering the many substrates of L7Ae. This may be different in archaea containing only few C/D box

sRNAs. The presence of unbound C/D box sRNAs may also provide an explanation why preventing the L7Ae autoregulation in *S. acidocaldarius* did not cause the same strong toxicity phenotype that was observed in *E. coli*. In *E. coli*, no natural L7Ae substrates are available. Overproduction of L7Ae may therefore rapidly result in unspecific binding of essential RNA molecules causing the death of the cell. In contrast, unbound C/D box sRNAs in *S. acidocaldarius* may serve as a "buffer" to mitigate toxic effects of overproduced L7Ae.

The multiple sequence alignment of 121 archaeal *l7ae* loci revealed that the *l7ae* gene of all investigated archaea, except of Thermoproteales and *N. equitans*, comprised a possible 5' leader sequence. This observation is striking considering the general overrepresentation of leaderless transcripts in archaea [62, 181]. The identification of k-turn forming motifs in almost all 5' UTRs and the observed downregulation for 5' UTRs of phylogenetically distant archaea in the GFP reporter system revealed that the L7Ae autoregulation appears to be a conserved feature of archaea. The efficiency of GFP downregulation differed between the individual 5' UTRs. This differential pattern may give an indication for the stringency of L7Ae feedback regulation in the respective organisms. There is a tendency that organisms with few C/D box sRNA genes exhibit a more pronounced downregulation. The *l7ae* 5' UTRs of *A. fulgidus*, *M. acetivorans* and *M. maripaludis* showed the strongest downregulation. These archaea contain very few C/D box sRNA genes (<10) [40, 182, 183]. However, it should be noted that the environment, in which k-turn formation was tested, is artificial. Many of the investigated archaeal organisms thrive at high temperatures or in other extreme environments, e.g. high salt concentrations or anaerobic conditions. Hence, the k-turn formation may be influenced by these conditions, particularly by temperature. Moreover, each individual L7Ae protein from the different organisms might have a deviating affinity for its corresponding *l7ae* 5' UTR that may differ from the affinities observed for the utilized *S. acidocaldarius* L7Ae. The established GFP reporter system may therefore rather be used to identify downregulation due to k-turn formation than to quantify feedback regulation stringency.

N. equitans and the members of the Thermoproteales lack the *l7ae* 5' UTR required for autoregulation, but contain high numbers of C/D box sRNAs [40, 184]. It remains to be determined how L7Ae production is regulated in these archaea. Potentially, L7Ae binds to a k-turn within the coding sequence of the *l7ae* mRNA, as it has been found for the majority of mRNA binding partners of *S. acidocaldarius* (see Table 2.1).

The L7Ae homolog 15.5 kDa/Snu13p (human/yeast), which binds to C/D box snoRNAs in eukaryotes, was not reported to regulate its own production. However, eukaryotes can also comprise high C/D box snoRNA numbers and many of these RNAs are intron-encoded,

especially in higher eukaryotes. Thus, changing C/D box snoRNA numbers could also here be relevant [185]. It might be possible that autoregulation of 15.5 kDa has not yet been discovered. This may be supported by the finding that autoregulation of a different L7Ae/L30 family protein exists in eukaryotes: the L30e protein of *Saccharomyces cerevisiae* regulates the splicing and translation of its own mRNA [186-189]. The *l30e* transcript contains an intron that forms a k-turn at the splice site. L30e binding prevents the assembly of the spliceosome. Once the intron is spliced, the joint exons form a k-turn. Binding of L30e masks the start codon and thereby inhibits translation. The observation that L30e of *S. acidocaldarius* binds to the endogenous *l30e* mRNA of *S. cerevisiae* and regulates its splicing and translation *in vivo* underlines the preservation of the L7Ae/L30-k-turn interaction between Archaea and Eukarya [190]. In contrast to archaea and eukaryotes, the bacterial L7Ae homologs YbxF/YlxQ were shown to be dispensable and the k-turn interaction was shown to be weak [84, 191]. C/D box sRNAs are absent in bacteria. Taken together, these observations suggest that autoregulation of the L7Ae/L30 family originated in Archaea and coincided with the advent of C/D box sRNAs.

3.4 L7Ae regulates mRNAs encoding proteins involved in translation

In addition to the k-turn element that was identified in the 5' UTR of the *l7ae* transcript, the RIP-Seq approach revealed 19 other mRNA fragments that contained putative k-turn structures. These putative k-turns were either found in the 5' UTR (25 %) or coding sequence (75 %) of the identified mRNAs. As the performed GFP reporter studies demonstrated that L7Ae induction reduces the translation of GFP, it is likely that the translational regulation by L7Ae is repressive, presumably by blocking the progression of the ribosome.

Investigation of the predicted protein function of the k-turn containing mRNAs revealed no distinct category of L7Ae targets. The analysis was, however, complicated by a high number of proteins with unknown function (arCOG class S). The mRNAs located upstream or downstream of the identified k-turn containing mRNAs, which are encoded in a polycistronic transcript and are therefore also regulated by L7Ae, provide more information on protein function (see Table 2.1). Including these mRNAs, L7Ae may regulate the translation of 32 mRNAs. Although the majority (thirteen) of these RNAs still encodes proteins of unknown function, the products of many mRNAs (nine) categorize into arCOG class J. This functional category comprises proteins involved in translation, ribosomal structure and biogenesis. In contrast, other arCOG classes were represented by only one to three members. Most class J mRNAs were present in the two operons *saci_1347-1343* and *saci_1468-1463.* The k-turns in

their polycistronic transcript were both found in the 5' UTR of the first gene and each k-turn was verified for L7Ae binding by EMSA studies or GFP reporter assays.

The nine mRNAs encode for the ribosomal proteins S17e, L39e, L31e, LX and L7Ae, the C/D box sRNP proteins Nop5 and fibrillarin, the methyltransferase TRM1 and the translation initiation factor IF-6. Besides its own production, L7Ae can therefore regulate the production of four additional ribosomal proteins, which is advantageous for the balanced production of ribosome components [176]. Furthermore, it regulates the synthesis of all core protein of the C/D box sRNPs: Nop5, fibrillarin and itself. L7Ae is involved in the methylation of tRNAs via C/D box sRNPs. Interestingly, one of the nine mRNAs encodes the methyltransferase TRM1. TRM1 is an N^2,N^2-dimethyl-guanosine tRNA methyltransferase, which catalyzes the methylation of guanosine at position 26 in archaeal tRNAs [192, 193]. L7Ae also directly binds to a k-turn in a transcript encoding a tRNA-thio modification protein (*saci_0669*). As L7Ae could affect tRNA levels, due to its influence on the activity of RNase P RNPs (see also next chapter), these findings might suggest that, by their translational regulation, the amount of tRNA modifying enzymes is adjusted to the level of tRNAs [90].

Noteworthy, the mRNAs which did not comprise identifiable Kt-b and Kt-n strands may still contain k-turns, considering that k-turns allow high sequence variation (see Fig. 1.4a). These mRNAs include transcripts for three further ribosomal proteins (*saci_0080, saci_0088* and *saci_0593*). Remarkably, the respective genes are encoded in long operons that comprise genes for 26 further ribosomal proteins. In total, L7Ae could therefore potentially regulate the translation of 34 ribosomal proteins. Considering that ribosome biogenesis is a very energy-intensive process, placing k-turns in the transcripts of ribosomal proteins provides an elegant way to save energy by capping the production of excess amounts of ribosomal proteins.

3.5 L7Ae as a master regulator of translational processes – a theory

In this study, L7Ae of *S. acidocaldarius* was shown to bind to a large set of non-coding RNAs, including SRP RNA. Furthermore, the protein was demonstrated to autoregulate its production by binding to a k-turn structure formed within its mRNA leader. Additional k-turn substrates were identified in mRNAs. Archaeal L7Ae therefore comprises two major roles in the cell: i) structuring of non-coding RNAs and ii) translational regulation via k-turn elements in mRNA sequences (Fig. 3.1).

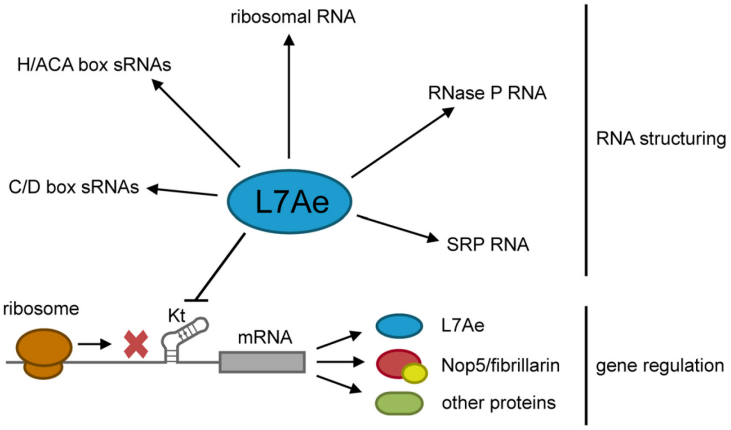


Figure 3.1 The dual function of archaeal L7Ae. The model highlights the two roles of archaeal L7Ae. L7Ae binds to k-turns in various non-coding RNAs and induces a conformational change which recruits additional proteins. In addition, L7Ae represses the translation of its own and of other mRNAs by binding to a k-turn structure in the leader or the coding sequence.

Previous studies already perceived that the functions of the non-coding RNAs, in which L7Ae fulfills its structuring role, are all related to translational processes in the cell [89, 194]. C/D and H/ACA box sRNAs were hypothesized to have evolved from a primitive translation apparatus due to the finding of the ribosomal L7Ae-k-turn RNP element [194]. Cho and co-workers found the same element in RNase P RNA and suggested a co-evolution of translation-related functions [89]. The authors argued that L7Ae´s influence on the architecture and activity of different RNPs could allow for their coordinated regulation [89, 90]. SRP RNA, which was identified as a putative binding partner of L7Ae in this work, constitutes another non-coding RNA that is linked to translation. The discovery of many putative L7Ae-regulated mRNAs that encode proteins involved in translation indicates that L7Ae can also regulate translation-related processes at the mRNA level. This could suggest that L7Ae is a master regulator of translation in archaea. The following scenario could be conceivable in archaeal cells:

L7Ae binds ribosomal RNA and directly participates in the translation of mRNA into proteins. By regulating the activity of RNase P RNP, it can adjust the level of mature tRNAs needed for mRNA decoding during the translation process. As a component of C/D box and H/ACA box sRNPs, it ensures the stability of the rRNA and tRNAs by their modification and facilitates correct folding of the rRNA. Within SRP RNPs, it can pause protein translation and guide the nascent peptide chain to the cell membrane. It further regulates the production of various translation-related proteins, such as ribosomal proteins and translation factors, to ensure their balanced availability. L7Ae autoregulation allows the adaption to changing substrate pools resulting from environmental changes, in particular varying C/D box sRNA levels, which in turn also affects the regulation of the mRNA interaction partners of L7Ae.

3.6 A bacterial reporter system that can be exploited by synthetic biology approaches

Saito and co-workers designed a synthetic switch to control the translation of proteins by utilizing the interaction of L7Ae with artificial k-turn elements in human cells [93]. In this work, a bacterial reporter system was designed in which natural archaeal L7Ae-k-turn interactions were used for the translational repression of GFP. The reporter system is provided on a single plasmid, is easily modifiable, and can be used to negatively control the translation of any gene of interest in *E. coli*. By fusing the gene of interest to the *gfp* gene or fusing the same k-turn that is used for regulation to a *trans*-encoded *gfp* gene, the system allows a direct read-out for the strength of regulation. Individual tuning can be achieved by adjustment of the inducer concentration (IPTG) or by utilizing different archaeal k-turns, which showed highly diverse efficiencies of repression. Strong repression can be established by placing several k-turns in a row. The system may also be used for spatiotemporal regulation of protein translation, e.g. by fusing L7Ae to a cell cycle-dependent protein. Similar to the applications developed for human cells, the archaeal L7Ae-k-turn interaction could be applied as an RNP module in synthetic biology approaches which utilize bacterial cells as host organisms [93-95]. For instance, feedback regulation of any protein of interest can be achieved by creating a fusion with L7Ae and providing a k-turn in the protein´s mRNA. This allows optimized and stable protein concentrations, which is often demanded in synthetic circuits.

In its current form, the system constitutes a translational repression system (OFF switch). To expand its potential for complex synthetic circuits, the system could be remodeled to allow translational activation upon L7Ae induction (ON switch). As demonstrated for human cells, this can be achieved by using a k-turn containing antisense RNA that represses the translation of an mRNA in the absence of L7Ae, whereas in the presence of L7Ae the antisense RNA is captured and the mRNA is translated [93]. The L7Ae-k-turn RNP module may also be utilized in CRISPR-Cas-based transcriptional regulation approaches. In these approaches, the crRNA-guided interference activity of modified Cas proteins, e.g. nuclease-dead Cas9 (dCas9), is exploited to repress or activate the transcription of genes [195]. An L7Ae-bound k-turn sequence in a single-guide RNA (sgRNA), which is used by dCas9 to target the gene of interest, could prevent the dCas9-mediated gene regulation by inhibiting the binding of dCas9 to the sgRNA. In the absence of L7Ae, gene regulation by dCas9 would be enabled.

It should be noted that a controlled production of L7Ae is necessary, when utilizing the L7Ae-k-turn RNP module in bacteria, as L7Ae overproduction proved to be toxic in *E. coli*. In the here designed system, this toxicity could be abolished by introducing a k-turn in the *l7ae* transcript to enable negative feedback regulation of the protein.

3.7 The LSm1 and LSm2 proteins of *S. acidocaldarius* form heteroheptameric rings

Various crystal structures revealed that the archaeal LSm proteins form homoheptameric rings [102, 124, 125, 131, 196]. These complexes were so far considered to be the functional RNA-binding units in the cell [67]. However, in this work, two observations indicated that the LSm1 and LSm2 proteins of *S. acidocaldarius* could form heteromers rather than homomers. First, a second prominent band at the estimated size of LSm1 was visible in the SDS-PAGE analysis of purified LSm2 (Fig. 2.22a). Mass spectrometry analysis of this purification fraction identified both LSm1 and LSm2 among other proteins (Appendix 5). In the LSm1 purification, a corresponding LSm2 band was not visible, potentially due to an overlap of similarly sized untagged LSm2 (9.8 kDa) with LSm1-CHis (9.4 kDa) (Fig. 2.22a). Noteworthy, the co-immunoprecipitation of LSm1 and LSm2 was also observed in two other studies [102, 135]. Second, the high number of identical binding partners of LSm1 and LSm2 identified in the RIP-Seq analysis indicated that the two proteins might interact *in vivo*. Finally, Alexander Pastura could proof the heteromerization of LSm1 and LSm2 in a heterologous system by co-purification of the two recombinant proteins from *E. coli* (Fig. 2.28a). A gel filtration analysis showed that the LSm1/2 heterocomplex forms a heptameric ring (Fig. 2.28b). During dialysis, this heteroheptamer displayed a significantly higher stability than the homomeric proteins that were purified individually, which demonstrates that the heteromer interactions are beneficial for stability [197]. Thus, it is likely that the LSm1/2 heterocomplex is the prevailing functional unit for RNA-binding in *S. acidocaldarius*. This notion is also supported by the finding that individually purified LSm1 from *E. coli* displayed very weak RNA-binding activity [197]. However, it should be noted that although most RNA binding partners of the proteins were identical, a certain degree of heterogeneity after RIP-Seq analysis of the two proteins was observed. This heterogeneity may result from the uneven representation of the two proteins within the heptameric complex, which could have an influence on the selection of the RNA binding partners (see also chapter 3.10).

3.8 A putative role in mRNA degradation for *S. acidocaldarius* LSm1/2

In this study, a bioinformatic analysis of the RNA binding partners of the LSm proteins from *S. acidocaldarius* revealed a U-rich motif as a potential LSm binding site. This U-rich motif corresponds to the 3' termination signal of RNAs that was recently identified by Dar and co-workers [157]. EMSA analysis confirmed a high binding affinity of the LSm1/2 heterocomplex for this 3' termination signal. In line with this finding, also the bacterial Hfq and eukaryotic LSm proteins bind uridine tracts at the 3' termini of RNAs [96]. However, for Hfq and

eukaryotic LSm, binding to U-rich termination sequences was primarily reported for small RNAs: Hfq binds the U-rich terminus of sRNAs that results from Rho-independent transcriptional termination and LSm2-8 binds to the processed 3' uridine tract of U6 snRNA [60, 96]. In contrast, most RNA binding partners of the *S. acidocaldarius* LSm proteins were mRNAs that contained the U-rich termination signal in a short 3' UTR (Fig. 2.24). Hfq also binds mRNAs, but recognizes these transcripts via A-rich sequences either within the mRNA (for sRNA regulatory interactions) or within short poly-A tails [60, 106, 198]. Hfq binding of the short poly-A tail stimulates excessive polyadenylation of the 3' terminus by poly-A polymerase I (PAP I), which triggers mRNA degradation by exonucleases (e.g. RNase R or PNPase) in bacteria (Fig. 1.6a) [60, 104]. Also eukaryotic LSm1-7 was shown to be involved in mRNA decay [117]. In contrast to bacterial mRNAs, eukaryotic mRNAs are generated with long poly-A tails that are deadenylated to induce their degradation [199]. The LSm1-7 complex binds the oligo-A tail of the deadenylated mRNAs, inhibits 3' to 5' degradation by the exosome and recruits decapping factors to initiate mRNA degradation from 5' to 3' direction (Fig. 1.6b) [96, 118]. LSm1-7 preferentially recognizes the oligo-A tail when the mRNAs contain a U-tract at their 3' terminus [118]. This raises the question if also the LSm proteins of *S. acidocaldarius* bind to the U-rich 3' terminus of mRNAs to induce their decay.

In contrast to eukaryotes, archaeal mRNAs are not generated with long poly-A tails. Similar to bacteria, the mRNAs of archaea are actively polyadenylated (or rather A-rich polynucleotidylated) to induce their degradation [200]. Archaea lack PAP I that generates homopolymeric poly-A tails and the PNPase which degrades these tails in bacteria [200]. Instead, they comprise a eukaryotic-like exosome complex, which is structurally and functionally similar to the bacterial PNPase [201, 202]. In nucleoside diphosphate (NDP) presence, the archaeal exosome complex adds heteropolymeric A-rich tails to mRNAs, which also contain the other three nucleotides in addition to adenosine, an ability that resembles bacterial PNPase activity [203-205]. In the presence of inorganic phosphate, the archaeal exosome then rapidly degrades the polynucleotidylated tails and the mRNA [203, 205].

A-rich polynucleotidylation of mRNAs was confirmed in *S. acidocaldarius* and other archaea [200, 206]. Therefore, the LSm RIP-Seq data sets were investigated for polynucleotidylated tails 3' of the termination signal. However, there was no evidence for such tails in the produced RIP-Seq data sets. Potentially, polynucleotidylated mRNAs are so rapidly degraded by the exosome, that they are underrepresented in the cell. A selection step would be required to enrich for these RNAs. For instance, the identification of tailed mRNAs in *S. acidocaldarius* mentioned above was performed by utilizing oligo-dT-primed RT-PCR [200].

Despite the lack of A-rich tailed mRNAs in the LSm RIP-Seq data sets, there is indication that the *Sulfolobus* LSm proteins stimulate mRNA polynucleotidylation to induce mRNA degradation. In a recent study, the LSm1 and LSm2 proteins from *S. solfataricus* were found to co-purify with components of the exosome which were also found in the mass spectrometry data of the LSm2-CHis Ni-NTA purification (Appendix 5) [135]. The authors demonstrated direct interaction of the LSm proteins with the poly-A binding and accessory exosomal subunit DnaG [135]. Overproduction of the LSm proteins stimulated A-rich tailing of mRNAs in *S. solfataricus* [135]. This could be demonstrated by RNA-Seq, which utilized RNA isolates of induced and non-induced cell cultures that were enriched for polynucleotidylated mRNAs by using oligo-dT affinity purification [135]. The LSm proteins of *Sulfolobus* thus seem to contain an Hfq-like (stimulation of polyadenylation) and a eukaryotic LSm-like function (3' terminal binding to uridine tracts) regarding their putative role in mRNA decay.

The degradation of mRNAs is a global process that affects the stability of potentially all produced mRNAs in the cell. This can explain the high number of mRNA binding partners of the LSm proteins and provide a rationale why a functional classification could not identify a specific group of mRNAs that was bound by the LSm proteins. However, it was shown that the enrichment of LSm-interacting mRNAs was not coupled to a strong transcription within the cell (Fig. 2.26). This suggests that the LSm proteins can stimulate the degradation of certain mRNAs more efficiently than others.

In the following model, the putative role of *Sulfolobus* LSm proteins in mRNA degradation is illustrated (Fig. 3.2).

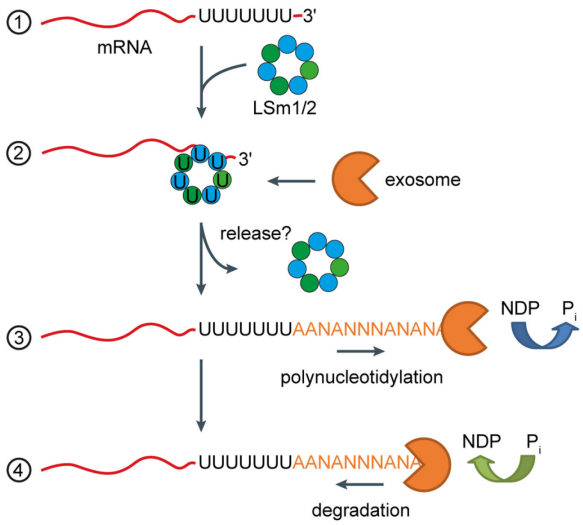


Figure 3.2 Model of *Sulfolobus* LSm function in mRNA degradation. In order to initiate the degradation of an mRNA, the LSm1/2 heterocomplex binds to the uridine tract of the 3' termination signal (1). LSm1/2 interacts with the archaeal exosome (2) and stimulates A-rich polynucleotidylation of the 3' terminus in the presence of NDP (3). Tailing of the mRNA might release the LSm complex from the termination signal, putatively due to its preference for terminal uridine residues. The exosome then rapidly degrades the polynucleotidylated mRNA in 3' to 5' direction in a phosphate-dependent manner (4).

3.9 LSm1/2 might be involved in several small RNA-regulated processes

Non-coding RNA binding partners were underrepresented in the analyzed subsets of the LSm RIP-Seq study (100 most significant peak calls). This was unexpected considering that the LSm proteins were implied to be part of small RNA-regulated processes [132].

A single crRNA was highly enriched for the LSm1 sRNA data set and also for the sRNome. This crRNA contained a 6 nt U tract in the spacer sequence. However, other crRNAs of *S. acidocaldarius* contain U-rich tracts in their spacers, but were not enriched in the LSm data sets. A possible reason for the specific enrichment of this crRNA is its fragmentation as the sRNome data set displayed an undefined 3' terminus for this RNA. Intact crRNAs comprise 5'-OH and 2',3'-cyclic phosphate termini resulting from Cas6 cleavage which cannot be adapter ligated during cDNA library preparation. This explains why all other crRNAs were underrepresented in the sRNome. In the LSm1 data set, the region of the 6 nt U tract showed the highest coverage, demonstrating that the LSm proteins preferentially bound to fragments that terminated with the U tract.

Sac-sR115 represents a highly enriched C/D box sRNA in the LSm1 data set. Several other C/D box sRNAs were enriched, although none of them were present within the 100 most significant peak calls. Previous studies already reported C/D box sRNAs and the associated proteins Nop5 and fibrillarin as binding partners of archaeal LSm proteins [132, 134, 135]. Nop5 and fibrillarin were also found in the mass spectrometry analysis of LSm2 (Appendix 5). Likewise, eukaryotic LSm proteins were found to be associated with C/D box snoRNAs [120, 121]. For instance, LSm2-8 were shown to be required for the 3' end processing of U3 C/D box snoRNA [121]. The archaeal LSm proteins were therefore also suggested to be involved in C/D box sRNA maturation [67]. In archaea, C/D box sRNAs are often co-transcribed with adjacent genes [143]. This requires a so far unknown processing step to generate mature C/D box sRNAs. The archaeal exosome was suggested as one candidate to maturate the 3' termini of these RNAs [143]. The Sac-sR115 sRNA is putatively co-transcribed with the overlapping gene *saci_0843* and located in the 3' UTR of the mRNA. The RNA terminates with a U-rich tract, which might be the binding site of LSm1/2. The role of LSm1/2 might be to recruit the archaeal exosome to the 3' end of the transcript, which allows the enhanced 3' maturation of the highly produced Sac-sR115.

Another enriched non-coding RNA was RNase P RNA. This RNA was also demonstrated to be an archaeal LSm interactor in previous studies [102, 134]. The RNase P RNP maturates tRNAs by cleaving the 5' leader sequences of precursor-tRNAs. The archaeal LSm proteins were therefore indicated to be involved in tRNA processing, similar to the eukaryotic LSm proteins

and Hfq [102, 119, 122, 134, 207]. A corresponding enrichment of tRNAs was not observed in the data sets. However, the inability to generate cDNA from these heavily modified RNAs, resulting in their absence in RNA-Seq data, prevents a validation of this association.

Only one genuine *trans*-acting sRNA, namely RrrR, was found to be highly enriched in the LSm data set. This RNA was recently shown to be a regulator of biofilm formation by our laboratory [66]. RrrR was predicted to target the *saci_0301* mRNA and a deletion mutant of this RNA resulted in an upregulation of *saci_0301* expression [66]. Furthermore, RrrR was shown to be negatively regulated by a short antisense RNA [66]. This antisense RNA outnumbered RrrR in the sRNome by more than 10-fold (Fig. 2.27b). In contrast, the LSm proteins bound almost exclusively to RrrR (Fig. 2.27b). Like Hfq, the *S. acidocaldarius* LSm proteins might therefore utilize RrrR as a *trans*-acting sRNA to regulate the translation of *saci_0301* mRNA. Further analyses are required to verify this theory. The *saci_0301* mRNA was not enriched in the LSm RIP-Seq data. This was not unexpected as the RIP-Seq analysis was performed from planktonically grown cells. In contrast, the *saci_0301* expression was shown to be strongly upregulated during biofilm growth [66].

In general, *trans*-acting sRNAs are often only expressed under specific growth or stress conditions [58]. This may explain their low abundance in the RIP-Seq data. In order to find more of these regulatory sRNAs, RIP-Seq analysis of the LSm proteins should be performed at varying conditions, e.g. biofilm growth, high pH and low temperature. This has the additional advantage of providing information on the function of these sRNAs.

Similar to Hfq, the *S. acidocaldarius* LSm proteins may bind *trans*-acting sRNAs via the U-rich termination signal. In line with this assumption, LSm1/2 showed binding to the U motif of RrrR in EMSA analyses. For sRNA-based mRNA regulation, Hfq recognizes A-rich sequences within mRNAs [106]. The *E. coli* Hfq was shown to bind $[A\text{-}A\text{-}N]_n$ triplets in the mRNA [109]. Each Hfq subunit binds one triplet. A motif search to identify LSm binding sites within the found RNA interaction partners revealed the UAG motif in addition to the U motif. The UAG motif was almost exclusively found in the coding sequence of mRNAs. This motif could constitute the archaeal equivalent of the $[A\text{-}A\text{-}N]_n$ binding site of Hfq. Several UAG triplets were identified in this motif, suggesting that each LSm subunit could bind one triplet. Only weak binding of the UAG motif of *saci_1217* mRNA could be observed during EMSA studies (Fig. 2.29). A mutated RNA lacking the respective motif, however, showed no binding. This mutant did not comprise a much lower uridine content than the original motif (five instead of seven uridine residues), to rule out that the motif was only bound due to the presence of uridines. Furthermore, it was shown that archaeal LSm proteins require at least five consecutive uridine

residues to efficiently bind RNA [196]. Due to the rather weak affinity of LSm1/2 to the UAG motif of the *saci_1217* mRNA, it should be considered that efficient binding to the UAG motif might only be possible, when a *trans*-acting sRNA is bound to the LSm ring and the sRNA interacted with the mRNA.

3.10 Increased complexity for multiple archaeal LSm proteins?

All proteins of the Sm superfamily are involved in RNA-related processes. Whereas bacteria contain the single Hfq protein, eukaryotes comprise over 20 members of this family. Hfq forms a homomeric ring and controls multiple processes in the cell, including translational regulation of mRNAs and mRNA turnover [60]. In contrast, eukaryotes harbor specialized heteromeric compositions of Sm/LSm rings that fulfill highly sophisticated roles, including mRNA splicing, mRNA decapping and decay, tRNA/rRNA maturation, 3' processing of histone mRNA, telomere maintenance and pre-RNase P RNA processing [96]. Archaea contain up to three LSm proteins. Halophiles contain only the LSm1 protein, other members of Euryarchaeota comprise one or two, and Crenarchaeota and Thaumarchaeota possess three LSm proteins [67]. Thus, phyla that are closer related to eukaryotes comprise a higher number of LSm proteins than phylogenetically distant archaea. The increase of functional complexity with the number of Sm proteins from Bacteria to Eukarya led Mura and co-workers to develop the hypothesis that also the archaeal LSm proteins adopted more specialized functions in essential cellular pathways as they increased in copy number (55).

The formation of heteromeric LSm complexes in archaea, which was so far known to be a eukaryotic feature, could be a first indication for this trend. LSm1 and LSm2 were shown to assemble into a heteroheptameric ring (chapter 3.7). As a heptamer consists of seven subunits, this implies that the two proteins cannot be equally represented within the ring. Potentially, different heptamer compositions of LSm1 and LSm2 fulfill different tasks, similar to the eukaryotic LSm proteins, where LSm2-8 is involved in mRNA splicing and LSm1-7 mediates mRNA decay [97]. Different specializations may explain the identification of slightly different RNA binding partners for the two proteins in the RIP-Seq analysis. For instance, non-coding RNAs were mostly enriched in the LSm1 than in the LSm2 RIP-Seq data sets (chapter 2.3.6). Potentially, LSm1/2 heterocomplexes with a higher proportion of LSm1 tend to bind non-coding RNAs. It may also be possible that an LSm1 homocomplex exists in *S. acidocaldarius* that preferably binds non-coding RNAs. However, this is contradicted by the observed RNA binding deficiency of recombinant LSm1 which was noted after purification from *E. coli* (Fig. 2.28a) and during EMSA assays performed by Alexander Pastura [197]. One novel function

that may result from the interaction of LSm1 and LSm2 is the putative involvement of the heterocomplex in mRNA degradation. Whereas the *Sulfolobus* LSm1 and LSm2 proteins are likely to be involved in this process, the single LSm1 proteins of halophilic archaea are certainly not: halophiles lack the archaeal exosome and thus polynucleotidylation proofed to be absent in these archaea [200]. Instead, RNA degradation is carried out by the exonuclease RNase R [206]. Besides halophiles, several methanogens lack exosome genes and polynucleotidylation, such as *M. maripaludis* and *M. jannaschii* [206]. Also *M. maripaludis* contains only the *lsm1* gene. *M. jannaschii* lacks all *lsm* genes, but comprises an Hfq protein [123]. In contrast, *M. mazei* or *Methanothermobacter thermautotrophicus* comprise the exosome and also two LSm proteins. This may indicate that an interaction of two different LSm proteins is a prerequisite for their involvement in the mRNA degradation process in archaea. In contradiction to this, Pyrococcus species only contain a single *lsm1* gene, but at the same time comprise exosomal genes. It should be noted that in bacteria, Hfq homocomplexes are also involved in mRNA degradation. However, the low sequence similarity between bacterial Hfq and archaeal/eukaryotic Sm/LSm suggests an early separation during evolution. Thus, Hfq´s involvement in mRNA degradation could have evolved separately from the archaeal/eukaryotic lineage. This scenario may be supported by different polyadenylated mRNA binding modes: Hfq recognizes poly-A tails via the distal face, while eukaryotic (and probably also the archaeal LSm) utilize the proximal face [96].

A further indication for an increased functional complexity and importance of LSm proteins along with their increasing copy number is that both LSm1 and LSm2 proteins of *S. acidocaldarius* seem to be essential, whereas the single *lsm1* gene of *H. volcanii* is dispensable [132]. This observation might be further connected to the LSm involvement in the mRNA degradation process in *S. acidocaldarius*. Many eukaryotic Sm and LSm proteins were reported to be essential [119]. In contrast, deletion mutants of *hfq* are viable, but often display pleiotropic defects [208]. Similar effects were reported for Δ*lsm1* from *H. volcanii*, but were invalidated when the authors found that the deletion of the *lsm1* gene resulted in the elimination of an internal promoter controlling a gene encoding a ribosomal protein [132, 133]. A newly produced defective mutant of *lsm1* was not affected in growth [133]. In contrast, in the current study the Flag-HA tagged *S. acidocaldarius* strains LSm1-CFHA and LSm2-NFHA both showed delayed growth rates. Considering that LSm1 and LSm2 are likely to be essential, the long tags seem to allow a limited functionality while not fully prohibiting complex formation. The respective strains displayed changes in morphology and an increased number of dead cells. These pleiotropic effects are likely to result from the involvement of LSm1 and LSm2 in the

global process of mRNA degradation. The observed effects were stronger for the LSm1-CFHA strain, which could indicate that the tag on the C-terminus of LSm1 affected the complex formation to a higher degree than the N-terminal tag on LSm2.

So far, no final proof for the essentiality of the *lsm1* and *lsm2* genes was obtained. However, as both conducted knock-out approaches failed to delete the *lsm1* and *lsm2* genes, whereas the *lsm3* gene could be knocked out by each strategy, essentiality of *lsm1* and *lsm2* is very likely. The successful deletion of *lsm3* as well as the absence of viable transformants in the negative control experiment for *lsm1* and *lsm2* of the CRISPR-based approach verified active targeting of the genome by the CRISPR-Cas interference complex. A potential reason for the presence of viable transformants in the negative control of the *lsm3* deletion is the emergence of escape mutants as indicated by the additional bands observed in the subsequent colony PCR. Complementation assays, in which an inducible extra copy of the genes is added and the original gene is deleted subsequently, constitute a way to demonstrate the essentiality of *lsm1* and *lsm2*. However, attempts to integrate a maltose inducible *lsm1* copy into the *S. acidocaldarius* genome failed (data not shown). Integration of a second *lsm2* gene was successful, however, the subsequent deletion of the original *lsm2* gene failed. These experiments may indicate that the LSm1 and LSm2 interaction strongly depends on a specific ratio between the two proteins which might be imbalanced by the integration of an additional copy or the expression of the respective genes from a foreign promoter.

The function of the third LSm protein, LSm3, remains elusive. The *lsm3* gene can be deleted from the *S. acidocaldarius* genome without observing growth defects, which demonstrates a less critical role for this protein than for LSm1 and LSm2. However, in comparison to the WT strain, the Δ*lsm3* strain displayed slightly increased numbers of dead cells in the flow cytometry analysis, indicating a certain impact of the gene on the fitness of *S. acidocaldarius.* Unlike LSm1 and LSm2, isolated LSm3 did not co-purify RNA. The possibility that the utilized His-tag destroyed the RNA-binding activity of LSm3 is unlikely, as this respective strain displayed the same phenotype as the WT in the flow cytometry analysis. Thus, LSm3 seems to lack RNA-binding capacity. A similar conclusion was drawn by Mura and co-workers as the protein was found to form supraheptameric oligomers in solution, in which the RNA-binding sites of the protein would be blocked [67, 128]. It should be considered that the long unknown C-terminal domain of LSm3 may confer a novel function to the protein that could be entirely different from what is known for other Sm family proteins.

3.11 Perspectives

In this study, L7Ae was found to interact with the archaeal SRP RNA. The putative binding site, a k-turn in close proximity to the 5e motif, was demonstrated to be conserved among the archaeal domain. It could therefore be possible that L7Ae is a core subunit of the archaeal SRP RNP. In order to investigate this, a mass spectrometry analysis of immunoprecipitated L7Ae and SRP54 should be performed to search for co-isolated SRP19/54 and L7Ae, respectively. In addition to the identified k-turn, the 5e motif is conserved within the SRP RNAs of archaea. Potentially, both the k-turn and the 5e motif are crucial for the accurate bending of the RNA. This may be demonstrated by Förster resonance energy transfer (FRET) experiments with an SRP RNA that contains fluorophores attached to the S and Alu domains. Mutation of the two motifs may change the bending of the RNA, which would result in altered fluorescence intensity. In addition, crystallization of the SRP RNA with L7Ae, SRP19 and SRP54 should be tested. This approach could be promising for obtaining the first full-length structure of an archaeal SRP RNP and could clarify how the bending is performed.

A key finding of this study was L7Ae´s ability to regulate the translation of proteins by binding to a k-turn motif within their mRNA transcripts. A high number of interacting mRNAs with putative k-turn motifs could be identified via L7Ae RIP-Seq analysis. L7Ae could thus prove to be a master regulator of protein translation in *S. acidocaldarius*. In order to investigate this putative role, a wild-type proteome could be compared with a proteome in which L7Ae autoregulation is abolished by utilizing a quantitative proteomics approach. This could allow for a global identification of the L7Ae regulated proteins in the cell.

The analysis of the LSm proteins from *S. acidocaldarius* revealed that LSm1 and LSm2 form stable heteroheptameric complexes. However, two proteins cannot be present in equal numbers in a heptamer. To analyze whether a specific subunit composition of the LSm1/2 heptamer complex is most prominent, native mass spectrometry analysis of recombinant LSm1/2 could be used, as previously performed for LSm1 of *H. volcanii* [132]. Non-coding RNA binding partners were primarily enriched in the RIP-Seq analysis of LSm1, which suggested that the subunit composition of the ring may have an influence on the selection of the RNA binding partners. By using a biotinylated non-coding RNA as a bait, e.g. RrrR, it could be screened if a certain LSm1/2 composition preferentially binds non-coding RNAs. Furthermore, crystallization of the LSm1/2 complex should be attempted to investigate how the ring is assembled by the two proteins.

Messenger RNAs were identified as the main binding partners of the LSm proteins. The identification of the 3' termination signal of mRNAs as a binding site and the previously

reported interaction of LSm proteins with the archaeal exosome suggested that the *Sulfolobus* LSm1 and LSm2 proteins are involved in the degradation of mRNAs by stimulating their tailing by the exosome [135]. To investigate if the *S. acidocaldarius* LSm proteins affect this tailing, the polynucleotidylation status of mRNAs should be compared between the wild-type strain and the growth affected LSm1-CFHA and LSm2-NFHA strains. This can be achieved by enrichment of tailed mRNAs via oligo-dT affinity purification followed by RNA-Seq as performed previously [135]. Furthermore, a RIP-Seq analysis of the Flag-HA tagged proteins could show if these proteins are affected in their ability to bind the 3' terminus of RNAs.

Halophilic archaea and several methanogens lack the archaeal exosome for mRNA degradation and only comprise a single LSm protein. It was hypothesized in this study that the involvement in mRNA degradation could be a unique function of interacting LSm1 and LSm2 proteins. Pyrococcus species contain genes for the archaeal exosome, but possess a single LSm1 protein. A RIP-Seq analysis of this protein could clarify whether LSm1 alone binds to the 3' terminus of mRNAs, which would indicate its involvement in mRNA decay.

Finally, it remains to be answered if archaeal LSm proteins fulfill a similar function as bacterial Hfq by mediating the interaction between *trans*-acting sRNAs and mRNAs. The binding of the *trans*-acting RNA RrrR by the LSm proteins suggests this. To investigate this putative function, EMSA studies with the non-coding RNA RrrR and the putative target mRNA *saci_0301* could be performed. Furthermore, it should be investigated if the identified UAG motif is the binding site of target mRNAs. A candidate motif, e.g. from *saci_1271*, could be mutated within the *S. acidocaldarius* genome. An LSm RIP-Seq analysis of the mutated strain could then show decreased levels of this mRNA. In addition, a CLIP-Seq (HITS-CLIP) analysis should be performed for the LSm proteins. This method allows identifying the exact RNA binding sites of the LSm proteins and could help to understand how the heteroheptameric ring recognizes RNA.

In conclusion, archaeal LSm proteins represent the missing link between the bacterial Hfq and eukaryotic Sm/LSm proteins. These proteins seem to be mosaics of Hfq function (sRNA-based translational regulation and stimulation of mRNA tailing) and Sm/LSm structure (assembly into heteroheptameric rings). Their investigation therefore offers a good opportunity to understand the evolution from the more generic role of the bacterial Hfq protein to the highly specialized functions of the Sm/LSm proteins of eukaryotes.

4. Material and Methods

4.1 Materials, instruments and source of supplies

4.1.1 Chemicals, kits and enzymes

The chemicals, kits and enzymes used in this work were obtained from the companies listed in Table 4.1. Remaining chemicals unlisted below were supplied by Carl Roth GmbH (Karlsruhe), Sigma-Aldrich (Taufkirchen) and BD (Heidelberg) in analytical grade.

Table 4.1 List of chemicals, kits, enzymes and consumables used in this study.

Chemicals	Company
5-Fluoroorotic acid (5-FOA)	Thermo Fisher Scientific Ltd. & Co. KG, Bonn
Acrylamide/Bisacrylamide (29:1 and 37.5:1)	Carl Roth GmbH, Karlsruhe
Ammonium Persulfate (APS)	Carl Roth GmbH, Karlsruhe
Antarctic Phosphatase	New England Biolabs GmbH, Frankfurt
Antibiotics (Ampicillin, Chloramphenicol, Kanamycin)	Carl Roth GmbH, Karlsruhe; Sigma-Aldrich, Taufkirchen
Anti-His IgG HRP-linked	Carl Roth GmbH, Karlsruhe
Anti-mouse IgG HRP-linked	Cell Signaling Technology, Leiden, NL
ATP [α-32P]	Hartmann Analytic GmbH, Braunschweig
ATP [γ-32P]	Hartmann Analytic GmbH, Braunschweig
Bacterial protease inhibitor	Carl Roth GmbH, Karlsruhe
β-Mercaptoethanol	Sigma-Aldrich, Taufkirchen
Bovine serum albumin (BSA)	Sigma-Aldrich, Taufkirchen
Bradford Reagent	BioRad Laboratories GmbH, Munich
ColorPlus Prestained Protein Ladder (10-230 kDa)	New England Biolabs GmbH, Frankfurt
Costar® Centrifuge Filter Tubes, pore size 0.45 µm	Sigma-Aldrich, Taufkirchen
Diethylpyrocarbonate (DEPC)	AppliChem GmbH, Darmstadt
Dimethyl Sulfoxide (DMSO)	Sigma-Aldrich, Taufkirchen
DNA Oligonucleotides	Eurofins MWG Operon, Ebersberg
dNTPs (dATP, dGTP, dCTP, dTTP)	New England Biolabs GmbH, Frankfurt
Dynabeads™ Protein G	Thermo Fisher Scientific Ltd. & Co. KG, Bonn
Ethidium Bromide	Carl Roth GmbH, Karlsruhe
Flag® HA Tandem Affinity Purification Kit	Sigma-Aldrich, Taufkirchen
Gel Breaker Tubes	IST Engineering, Milpitas, CA, USA
Gelpilot DNA Loading Dye (5x)	Qiagen GmbH, Hilden
Gelrite	Carl Roth GmbH, Karlsruhe
Gene Pulser® Cuvette, 0.1 cm gap	BioRad Laboratories GmbH, Munich
Glycogen	Roche Diagnostics GmbH, Mannheim
Instant Blue	Sigma-Aldrich, Taufkirchen (Expedeon)
Isopropyl-β-D-1-thiogalactopyranoside (IPTG)	Carl Roth GmbH, Karlsruhe
Low Molecular Weight DNA Ladder	New England Biolabs GmbH, Frankfurt
Low Range ssRNA Ladder	New England Biolabs GmbH, Frankfurt
Lysozyme	Sigma-Aldrich, Taufkirchen
***mir*Vana™ miRNA Isolation Kit**	Applied Biosystems, Darmstadt
Millex® AP20 Syringe Filter, pore size 20 µm	Merck KGaA, Darmstadt
Milk powder (skimmed)	Uelzena, Uelzen
NEBNext® Small RNA Library Prep Set	New England Biolabs GmbH, Frankfurt
NTPs (ATP/GTP/CTP/UTP)	Jena Bioscience GmbH, Jena
***ortho*-Nitrophenyl-β-galactoside (ONPG)**	Sigma-Aldrich, Taufkirchen

Phenylmethylsulfonylfluorid (PMSF)	Carl Roth GmbH, Karlsruhe
Phusion® High-Fidelity DNA Polymerase	Thermo Fisher Scientific Ltd. & Co. KG, Bonn
QIAGEN Plasmid Plus Maxi Kit	Qiagen GmbH, Hilden
QIAprep Spin Miniprep Kit	Qiagen GmbH, Hilden
QIAquick Gel Extraction Kit	Qiagen GmbH, Hilden
QIAquick PCR Purification Kit	Qiagen GmbH, Hilden
Qubit dsDNA HS Assay Kit	Thermo Fisher Scientific Ltd. & Co. KG, Bonn
Qubit RNA HS Assay Kit	Thermo Fisher Scientific Ltd. & Co. KG, Bonn
Quick-Load® 2-Log DNA Ladder (0.1-10.0 kb)	New England Biolabs GmbH, Frankfurt
Quick-Load® pBR322 DNA-MspI Digest	New England Biolabs GmbH, Frankfurt
Restriction endonucleases	New England Biolabs GmbH, Frankfurt
RNase Exitus Plus™	Applichem GmbH, Darmstadt
RNase Inhibitor (murine)	New England Biolabs GmbH, Frankfurt
Roti®-Lumin plus	Carl Roth GmbH, Karlsruhe
Roti®-Nylon plus, pore size 0.45 µm	Carl Roth GmbH, Karlsruhe
Roti®-Fluoro PVDF, pore size 0.2 µm	Carl Roth GmbH, Karlsruhe
SDS	Carl Roth GmbH, Karlsruhe
SigmaPrep™ Spin Column Kit	Sigma-Aldrich, Taufkirchen
Sumo Protease	Own production
SYBR Gold® Nucleic acid stain	Thermo Fisher Scientific Ltd. & Co. KG, Bonn
T4-DNA-Ligase	New England Biolabs GmbH, Frankfurt
T4-Polynucleotide-Kinase	New England Biolabs GmbH, Frankfurt
T4-Polynucleotide-Kinase (RNase-free)	Ambion, Darmstadt
T7-RNA-Polymerase	Own production
Tetramethylethylendiamine (TEMED)	Sigma-Aldrich, Taufkirchen
THE™ His-tag antibody (from mouse)	GenScript, Piscataway Township, NJ, USA
ULTRAhyb-Oligo hybridization buffer	Ambion, Darmstadt
Whatman GB004, 3MM	Schleicher & Schuell, Dassel
X-Ray CEA RP New film screen	CEA GmbH, Hamburg
ZelluTrans Dialysis tubes (6,000-8,000 MWCO)	Carl Roth GmbH, Karlsruhe

4.1.2 Instruments

Table 4.2 Instruments used in this study.

Instrument	Model and Company
Agarose gel electrophoresis	Chambers and Casting trays produced by company technician Philipps-University Marburg Power supply Consort E835; MS Laborgeräte, Dielheim
Aqua bidest, water system	PURELAB Plus, ELGA LabWater, Celle
Autoclave	5075 EL, Tuttnauer Europe B.V., Breda, NL
Bioanalyzer	Agilent 2100 Bioanalyzer, Agilent, Santa Clara, CA, USA
Centrifuges	Centrifuge 5424, Eppendorf AG, Hamburg; Sigma 3-30K, Sigma Laborzentrifugen GmbH, Osterode am Harz; Sorvall Lynx 4000, Thermo Fisher Scientific Ltd. & Co. KG, Bonn
Chromatography column	HisTrap HP 1 ml, GE Healthcare Europe GmbH, Freiburg
Electroporation device	Gene Pulser® Electroporation System, Pulse Controller Plus, Capacitance Extender Plus, BioRad Laboratories GmbH, Munich

Film developer	CP1000, Agfa, Mortsel, Belgium
Flow cytometer	BD LSRFortessa, BD Biosciences, Heidelberg
FPLC	Äkta-System: Pump P-900, Monitor UV-900, Monitor UPC-900, Valve INV-907, Mixer M-925; GE Healthcare Europe GmbH, Freiburg
French Press	FA-078 French Press, SLM Aminco, Urbana, IL, USA
Denaturing polyacrylamide gel electrophoresis	PROTEAN II Electrophoresis Chamber, BioRad Laboratories GmbH, Munich
Hybridization oven	Hybaid Shake 'n' Stack, Thermo Fisher Scientific Ltd. & Co. KG, Bonn
Incubators	KB53, Binder GmbH, Tuttlingen
Magnetic stirrer	IKA® RCT Standard, IKA®-Werke GmbH & Co. KG, Staufen
Magnetic Separation Rack	2-Tube Magnetic Separation Rack; New England Biolabs GmbH, Frankfurt
Microscope	Axioplan 2, Carl Zeiss Microscopy GmbH, Göttingen; CoolSnap HQ camera, Visitron Systems GmbH, Puchheim; FluoArc HBO Lamp, Carl Zeiss Microscopy GmbH, Göttingen
Nanodrop	NanoDrop® ND-1000 Spectrometer, Thermo Fisher Scientific Ltd. & Co. KG, Bonn
PCR-Cycler	C1000™ Thermal Cycler, BioRad Laboratories GmbH, Munich
Peristaltic pump	Peristaltic Pump P-1, GE Healthcare Europe GmbH, Freiburg
pH-meter	INOLAB pH level 1, WTW, Weilheim
Phosphorimager	Storm 840 phosphorimager, Molecular Dynamics, GE Healthcare Europe GmbH, Freiburg
Plate Reader	TECAN infinite M200Pro, Männedorf, Switzerland
Qubit Fluorometer	Qubit 2.0, Thermo Fisher Scientific Ltd. & Co. KG, Bonn
Rocker	Gyrorocker SSL3, Sigma-Aldrich, Taufkirchen
Scintillation counter	Beckmann LS 6500, Beckman Coulter GmbH, Krefeld
Semi-dry transfer cell	Trans-Blot® SD Semi-Dry Transfer Cell, BioRad Laboratories GmbH, Munich
SDS polyacrylamide gel electrophoresis	Mini-PROTEAN Tetra Cell, Bio-Rad Laboratories GmbH, Munich; Power supply PowerPac Basic, Bio-Rad Laboratories GmbH, Munich
Sonicator	Branson Sonifier 250, Branson Ultrasonicx, Danbury, CT, USA
Spectrophotometer	Ultrospec 3000 *pro*, GE Healthcare Europe GmbH, Freiburg
Thermomixer	Thermomixer Comfort 5350, Eppendorf AG, Hamburg
Thermoshaker	HT Thermotron, Infors AG, Bottmingen, Switzerland
UV-Crosslinker	UV Stratalinker® 1800, Stratagene, La Jolla, USA
UV-Transilluminator	BioDocd-IT system, UVP, Upland, CA, USA
Vortex Mixer	Vortex Genie 2, Scientific Industries, Bohemia, NY, USA

4.1.3 Buffers and solutions

Standard solutions and buffers were prepared according to Ausubel and Sambrook *et al.* [209, 210]. Special solutions and buffers are mentioned at the respective method. The media, solutions and buffers were, as necessary, autoclaved for 20 min at 121°C before use. Heat-sensitive solutions were filtered using a sterile filter unit (Steritop 0.2 µm pore size from Millipore, Billerica, MA, USA).

4.2 Strains and culture conditions

4.2.1 Strains

Table 4.3 Bacterial and archaeal strains used in this study.

Strain	Relevant genotype	Reference
***Escherichia coli* K12 DH5α**	F$^-$ Φ80*lac*ZΔM15 Δ(*lac*ZYA-*arg*F) U169 *rec*A1 *end*A1 *hsd*R17(r_K^-, m_K^+) *pho*A *sup*E44λ–*thi*-1 *gyr*A96 *rel*A1	**[211]**
***Escherichia coli* Top10**	F$^-$ *mcr*A Δ(*mrr-hsd*RMS-*mcr*BC) Φ80*lac*ZΔM15 Δ*lac*X74 *rec*A1 *ara*D139 Δ(*ara leu*) 7697 *gal*U *gal*K *rps*L (StrR) *end*A1 *nup*G	**Invitrogen™, Thermo Fisher Scientific Ltd. & Co. KG, Bonn**
***Escherichia coli* Rosetta2 (DE3) pLysS**	F$^-$ *ompT hsdS*$_B$(r_B^- m_B^-) *gal dcm* (DE3) pLysSRARE2 (Cam^R)	**Novagen, Darmstadt**
***Escherichia coli* K12 ER1821 bearing the pM.EsaBC4I plasmid**	F$^-$ *glnV44 e14$^-$(McrA$^-$) rfbD1? relA1? endA1 spoT1? thi-1 Δ(mcrC-mrr)114::IS10* (K12 ER1821)	**New England Biolabs GmbH, Frankfurt; [139]**
Sulfolobus acidocaldarius MW001	DSM639 Δ*pyrE* (*Saci_1597*; Δ91–412bp)	**[139]**

4.2.2 Culture conditions for *E. coli*

Cultivation of *E. coli* cell cultures (3-1500 ml) was carried out in Erlenmeyer flasks by shaking in a rotary shaker (200 rpm) at 37°C in LB (lysogeny broth) medium (1 % tryptone (w/v), 0.5 % yeast extract (w/v), 1 % NaCl (w/v), pH 7.2) or on solid medium plates (LB medium containing 1.5 % (w/v) agar-agar). Liquid LB medium supplemented with the appropriate plasmid-encoded antibiotics (ampicillin 100 µg/ml, chloramphenicol 34 µg/ml, kanamycin 50 µg/ml) was inoculated with a pre-culture (2 % (v/v)) and growth was monitored photometrically at OD_{600nm}. To ensure a linear dependence, cultures were diluted for values below OD_{600nm} 0.6. The OD_{600nm} of the corresponding culture medium was used as a reference value. Cultural stocks were obtained by the addition of 40 % (v/v) glycerol and were stored at -80°C. The *E. coli* K12 DH5a and Top10 strains were used for cloning, storage and preparation of plasmid DNA. The *E. coli* Rosetta2 (DE3) pLysS strain was used for the heterologous production of

recombinant L7Ae from *S. acidocaldarius*. Protein production was induced at OD_{600nm} 0.6-0.8 by the addition of 1 mM IPTG and incubation of the culture was continued for 3-4 hours at 37°C. Afterwards, the cells were harvested by centrifugation (12,000×*g*, 15 min, 4°C) and stored at -80°C. The *E. coli* K12 ER1821 pM.EsaBC4I strain was used to methylate plasmids before transformation into *S. acidocaldarius* to prevent restriction by the *SuaI* restriction enzyme.

4.2.3 Culture conditions for *S. acidocaldarius*

Sulfolobus acidocaldarius MW001, all mutant strains and all plasmid containing strains were aerobically grown at 180 rpm, 75°C and pH 3.5 in Brock media [137]. The media were supplemented with 0.1 % (w/v) tryptone or NZ-Amine, as well as 0.2 % (w/v) sucrose for normal growth or 0.3 % (w/v) xylose for induction of the *saci_1938* promoter (production of the crRNAs). 20 µg/ml uracil was added for strains without plasmid. The growth of the cells was monitored measuring the optical density at OD_{600nm}. Cell harvesting for RNA preparations and protein purifications was performed at OD_{600nm} of 0.4. For solid medium plates, two times concentrated Brock medium was prepared and supplemented with 6 mM $CaCl_2$ and 20 mM $MgCl_2$. 0.2 % NZ-Amine (w/v) and 0.4 % dextrin (w/v) were added for first selection plates. 0.2 % tryptone (w/v), 0.4 % dextrin (w/v) (or 0.3 % (w/v) xylose for crRNA production), 200 µg/ml 5-FOA and 20 µg/ml uracil were added for second selection plates. The prewarmed solution (75°C) was mixed with an equal volume of boiling 1.2 % gelrite solution (w/v). The mixture was adjusted to a pH of 3.5 using sulfuric acid and subsequently poured into petri dishes.

4.3 Plasmids and oligonucleotides

4.3.1 Progenitor plasmids and constructed recombinant vectors

Table 4.4 Progenitor plasmids used in this study.

Vector	Features	Application	Reference
pSVA406	Amp^r, pyrEF	Genomic mutation and tagging of *S. acidocaldarius* genes	Gift of Prof. Dr. Sonja-Verena Albers (Albert-Ludwigs-University Freiburg), **[139]**
pSVAxylFX-CRISPR	Amp^r, pyrEF, *lacZ*, *saci_1938* promoter, crRNA repeats	markerless deletions in *S. acidocaldarius* using CRISPR	Gift of Prof. Dr. Sonja-Verena Albers (Albert-Ludwigs-University Freiburg)
pSVA2003	Amp^r, pyrEF, Δ*lsm2* donor DNA	PCR template for Δ*lsm2* donor DNA	Gift of Prof. Dr. Sonja-Verena Albers (Albert-Ludwigs-University Freiburg)

pSVA431	Amp^r, pyrEF, *lacS* from *S. solfataricus*	PCR template for *lacS* gene	Gift of Prof. Dr. Sonja-Verena Albers (Albert-Ludwigs-University Freiburg), **[139]**
pSVA1431	Amp^r, pyrEF, mal promoter	β-galactosidase reporter assays in *S. acidocaldarius*	Gift of Prof. Dr. Sonja-Verena Albers (Albert-Ludwigs-University Freiburg), **[139]**
pEC-A-HiSumo	Amp^r, IPTG-inducible T7 promoter, His-Sumo tag fusion	Heterologous gene expression in *E. coli* Rosetta	Gift of Prof. Dr. Elena Conti (MPI Martinsried)
pASK-IBA3plus-*sfgfp*	Amp^r, Tet promoter, *superfolding gfp* gene, Strep tag fusion	PCR template for *superfolding gfp* gene	Gift of Prof. Dr. Regine Kahmann (MPI Marburg)
pDS43	Kan^r, IPTG-inducible T7 promoter, *pomZ-mcherry* fusion protein	PCR template for *mcherry* gene	Gift of Prof. Dr. Lotte Sogaard-Andersen (MPI Marburg)
pUC19	*lacZ*, Amp^r	RNA *in vitro* transcription	**[212]**

Table 4.5 Recombinant plasmids for the genomic tagging of *S. acidocaldarius* genes.

Plasmid + Insert	Description of the insert
pSVA406 + Saci_1520CFHA	C-terminally FLAG-HA tagged *l7ae* gene (*saci_1520*) with 1000 bp upstream and downstream regions from tag for homologous recombination; restriction sites: SacII, SalI; this work
pSVA406 + Saci_1224Flag-HA	N-terminally FLAG-HA tagged *lsm1* gene (*saci_1224*) with 500 bp upstream and downstream regions from tag for homologous recombination; restriction sites: NcoI, SalI; this work
pSVA406 + Saci_0799Flag-HA	N-terminally FLAG-HA tagged *lsm2* gene (*saci_0799*) with 500 bp upstream and downstream regions from tag for homologous recombination; restriction sites: NcoI, SalI; this work
pSVA406 + Saci_0660Flag-HA	N-terminally FLAG-HA tagged *lsm3* gene (*saci_0660)* with 500 bp upstream and downstream regions from tag for homologous recombination; restriction sites: NcoI, SalI; this work
pSVA406 + Saci_1224CFHA	C-terminally FLAG-HA tagged *lsm1* gene (*saci_1224*) with 1000 bp upstream and downstream regions from tag for homologous recombination; restriction sites: NcoI, SalI; this work
pSVA406 + Saci_0799CFHA	C-terminally FLAG-HA tagged *lsm2* gene (*saci_0799*) with 1000 bp upstream and downstream regions from tag for homologous recombination; restriction sites: NcoI, SalI; this work
pSVA406 + Saci_1224CHis	C-terminally His tagged *lsm1* gene (*saci_1224*) with 1000 bp upstream and downstream regions from tag for homologous recombination; restriction sites: NcoI, SalI; this work
pSVA406 + Saci_0799CHis	C-terminally His tagged *lsm2* gene (*saci_0799*) with 1000 bp upstream and downstream regions from tag for homologous recombination; restriction sites: NcoI, SalI; this work
pSVA406 + Saci_0660CHis	C-terminally His tagged *lsm3* gene (*saci_0660)* with 1000 bp upstream and downstream regions from tag for homologous recombination; restriction sites: NcoI, SalI; this work

Table 4.6 Recombinant plasmids for the genomic mutation of the *l7ae* 5' UTR.

Plasmid + Insert	Description of the insert
pSVA406 + Saci_1520CFHA g539c	C-terminally FLAG-HA tagged *l7ae* gene (*saci_1520*) with 1000 bp upstream and downstream regions from tag for homologous

	recombination; Kt-n mut1 mutation in *l7ae* 5' UTR ; restriction sites: SacII, SalI; this work
pSVA406 + Saci_1520CFHA g539c/a540t	C-terminally FLAG-HA tagged *l7ae* gene (*saci_1520*) with 1000 bp upstream and downstream regions from tag for homologous recombination; double Kt-n mut1 mutation in *l7ae* 5' UTR ; restriction sites: SacII, SalI; this work

Table 4.7 Recombinant plasmids for the genomic deletion of the *lsm* genes using CRISPR.

Plasmid + Insert	Description of the insert
pSVAxylFX-CRISPR + *lsm1* crRNA	Spacer against *lsm1* gene: position 110-145 of coding sequence; restriction site: SapI; this work
pSVAxylFX-CRISPR + *lsm1* crRNA-del-1224 donor DNA	Spacer against *lsm1* gene: position 110-145 of coding sequence; restriction site: SapI; donor DNA comprises deletion of *saci_1224* and 700-1000 bp upstream and downstream regions for homologous recombination; restriction sites: BamHI, EagI; this work
pSVAxylFX-CRISPR + *lsm2* crRNA	Spacer against *lsm2* gene: position 60-96 of coding sequence; restriction site: SapI; this work
pSVAxylFX-CRISPR + *lsm2* crRNA-del-0799 donor DNA	Spacer against *lsm2* gene: position 60-96 of coding sequence; restriction site: SapI; donor DNA comprises deletion of *saci_0799* and 700-1000 bp upstream and downstream regions for homologous recombination; restriction sites: BamHI, EagI; this work
pSVAxylFX-CRISPR + *lsm3* crRNA1	Spacer against *lsm3* gene: position 202-239 of coding sequence; restriction site: SapI; this work
pSVAxylFX-CRISPR + *lsm3* crRNA1-del-0660 donor DNA	Spacer against *lsm3* gene: position 202-239 of coding sequence; restriction site: SapI; donor DNA comprises deletion of *saci_0660* and 700-1000 bp upstream and downstream regions for homologous recombination; restriction sites: NheI, EagI; this work
pSVAxylFX-CRISPR + *lsm3* crRNA2	Spacer against *lsm3* gene: position 111-146 of coding sequence; restriction site: SapI; this work
pSVAxylFX-CRISPR + *lsm3* crRNA3	Spacer against *lsm3* gene: position 353-389 of coding sequence; restriction site: SapI; this work

Table 4.8 Recombinant plasmids for β-galactosidase reporter assays in *S. acidocaldarius*.

Plasmid	Description of the insert
pSVA1431 + L7PR-lacS Nat	pNat plasmid (Fig. 2.9); Native promoter and 5' UTR of *l7ae* cloned upstream of *lacS* gene of *S. solfataricus*; restriction sites: SacII, EagI; this work
pSVA1431 + L7PR-lacS Dmut	pKt-n mut1 plasmid (Fig. 2.9); Native promoter and Kt-n mut1 mutated 5' UTR of *l7ae* cloned upstream of *lacS* gene of *S. solfataricus*; restriction sites: SacII, EagI; this work
pSVA1431 + L7PR-lacS DmutGG	pKt-n mut2 plasmid (Fig. 2.9); Native promoter and Kt-n mut2 mutated 5' UTR of *l7ae* cloned upstream of *lacS* gene of *S. solfataricus*; restriction sites: SacII, EagI; this work
pSVA1431 + L7PR-lacS BTmut	pBTmut plasmid (Fig. 2.9); BRE/TATA site mutated promoter and native *l7ae* 5' UTR cloned upstream of *lacS* gene of *S. solfataricus*; restriction sites: SacII, EagI; this work

Table 4.9 Recombinant plasmid for the overproduction of L7Ae in *E. coli* Rosetta.

Plasmid + Insert	Description of the insert
pEC-A-HiSumo + Saci_1520	N-terminally His_{6x}-Sumo tagged *l7ae* gene; V. Tripp [141]

The pEC-A-HiSumo + Saci_1520 vector was used as a progenitor for the recombinant plasmids listed in Table 4.10.

Table 4.10 Recombinant plasmids for GFP reporter assays in *E. coli* Rosetta.

Plasmid	Description of the plasmid
pEC-A-*l7ae-sfgfp-mcherry* PN25	*l7ae* 5' UTR/no aKt plasmid (Fig. 2.10 and 2.11a,b,d); constitutive P_{N25} promoter [154] and native *l7ae* 5' UTR (*gfp*Kt) upstream of *superfolding gfp* gene and T1 terminator; constitutive chloramphenicol promoter P_{cat} upstream of *mcherry* gene and T1 terminator; *l7ae* gene under IPTG-inducible T7 promoter; pEC-A-HiSumo + Saci_1520 used as background plasmid; this work
pEC-A-*l7ae-sfgfp-mcherry* del112 pN25	no L7Ae plasmid (Fig. 2.10 and 2.11a,b); as pEC-A-*l7ae-sfgfp-mcherry* PN25; frameshifted (del112) *l7ae* gene; this work
pEC-A-*l7ae-sfgfp-mcherry* pN25 doublemut D	double Kt-n mut plasmid (Fig. 2.10); as pEC-A-*l7ae-sfgfp-mcherry* PN25; double Kt-n mutated *gfp*Kt; this work
pEC-A-*l7ae-sfgfp-mcherry* pN25 A4942T	bulge mut plasmid (Fig. 2.10); as pEC-A-*l7ae-sfgfp-mcherry* PN25; bulge mutated *gfp*Kt; this work
pEC-A-*l7ae-sfgfp-mcherry* pN25 control UTR	control UTR plasmid (Fig. 2.10); as pEC-A-*l7ae-sfgfp-mcherry* PN25; 5' UTR (control UTR) from P_m promoter [153] upstream of *superfolding gfp* gene; this work
pEC-A-*l7ae-sfgfp-mcherry* pN25 del112 control UTR	no L7Ae control UTR plasmid (Fig. 2.10); as pEC-A-*l7ae-sfgfp-mcherry* pN25 control UTR; frameshifted (del112) *l7ae* gene; used to define the percentage of downregulation for the pEC-A-*l7ae-sfgfp-mcherry* pN25 control UTR plasmid; this work
pMD2	pMD(-auto*l7ae-gfp*)/*S. aci* plasmid (Fig. 2.11a-d,f and 2.13b); as pEC-A-*l7ae-sfgfp-mcherry* PN25; additional native *l7ae* 5' UTR upstream of *l7ae* gene (aKt); this work
pMD20	aKt-n mut1 plasmid (Fig. 2.11a,b); as pMD2; Kt-n mut1 mutated aKt; this work
pMD21	double aKt-n mut plasmid (Fig. 2.11a,b); as pMD2; double Kt-n mutated aKt; this work
pMD1	pMD/control plasmid (Fig. 2.11e, Fig. 2.13b and Fig. 2.14a); as pEC-A-*l7ae-sfgfp-mcherry* pN25 control UTR; additional native *l7ae* 5' UTR upstream of *l7ae* gene (aKt); this work
pMD3	aKt-n mut1 plasmid (Fig. 2.11e); as pMD1; Kt-n mut1 mutated aKt; this work
pMD4	double aKt-n mut plasmid (Fig. 2.11e); as pMD1; double Kt-n mutated aKt; this work
pMD5	double *gfp*Kt-n mut plasmid (Fig. 2.11f); as pMD2; double Kt-n mutated *gfp*Kt; this work
pMD6	*gfp*Kt bulge mut plasmid (Fig. 2.11f); as pMD2; bulge mutated *gfp*Kt; this work
pMD22	double *gfp*Kt-b mut plasmid (Fig. 2.11f); as pMD2; double Kt-b mutated *gfp*Kt; this work
pMD9	*A. per* plasmid (Fig. 2.13b); as pMD2; 50 bp *A. pernix l7ae* upstream sequence (Fig. 2.13a) in place of *gfp*Kt; this work
pMD10	*S. mar* plasmid (Fig. 2.13b); as pMD2; 50 bp S. *marinus l7ae* upstream sequence (Fig. 2.13a) in place of *gfp*Kt; this work
pMD11	*P. aer* plasmid (data not shown); as pMD2; 50 bp P. *aerophilum l7ae* upstream sequence (Fig. 2.13a) in place of *gfp*Kt; this work
pMD12	*A. ful* plasmid (Fig. 2.13b); as pMD2; 50 bp *A. fulgidus l7ae* upstream sequence (Fig. 2.13a) in place of *gfp*Kt; this work

pMD13	*H. vol* plasmid (Fig. 2.13b); as pMD2; 50 bp *H. volcanii l7ae* upstream sequence (Fig. 2.13a) in place of *gfp*Kt; this work
pMD14	*M. ace* plasmid (Fig. 2.13b); as pMD2; 50 bp *M. acetivorans l7ae* upstream sequence (Fig. 2.13a) in place of *gfp*Kt; this work
pMD15	*T. kod* plasmid (Fig. 2.13b); as pMD2; 50 bp *T. kodakarensis l7ae* upstream sequence (Fig. 2.13a) in place of *gfp*Kt; this work
pMD16	*M. mar* plasmid (Fig. 2.13b); as pMD2; 50 bp *M. maripaludis l7ae* upstream sequence (Fig. 2.13a) in place of *gfp*Kt; this work
pMD1 del112	no L7Ae control plasmid (Fig. 2.13b and Fig. 2.14a); as pMD1; frameshifted (del112) *l7ae* gene; this work
pMD2 del112	no L7Ae *S. aci* plasmid (Fig. 2.13b); as pMD2; frameshifted (del112) *l7ae* gene; this work
pMD9 del112	no L7Ae *A. per* plasmid (Fig. 2.13b); as pMD9; frameshifted (del112) *l7ae* gene; this work
pMD10 del112	no L7Ae *S. mar* plasmid (Fig. 2.13b); as pMD10; frameshifted (del112) *l7ae* gene; this work
pMD11 del112	no L7Ae *P. aer* plasmid (data not shown); as pMD11; frameshifted (del112) *l7ae* gene; this work
pMD12 del112	no L7Ae *A. ful* plasmid (Fig. 2.13b); as pMD12; frameshifted (del112) *l7ae* gene; this work
pMD13 del112	no L7Ae *H. vol* plasmid (Fig. 2.13b); as pMD13; frameshifted (del112) *l7ae* gene; this work
pMD14 del112	no L7Ae *M. ace* plasmid (Fig. 2.13b); as pMD14; frameshifted (del112) *l7ae* gene; this work
pMD15 del112	no L7Ae *T. kod* plasmid (Fig. 2.13b); as pMD15; frameshifted (del112) *l7ae* gene; this work
pMD16 del112	no L7Ae M. *mar* plasmid (Fig. 2.13b); as pMD16; frameshifted (del112) *l7ae* gene; this work
pMD17	*saci_1468* plasmid (Fig. 2.14a); as pMD1; *saci_1468* Kt sequence (Appendix 2) is cloned in frame downstream of the *superfolding gfp* start codon; this work
pMD18	*saci_2027* plasmid (Fig. 2.14a); as pMD1; *saci_2027* Kt sequence (Appendix 2) is cloned in frame downstream of the *superfolding gfp* start codon; this work
pMD17 del112	no L7Ae *saci_1468* plasmid (Fig. 2.14a); as pMD17; frameshifted (del112) *l7ae* gene; this work
pMD18 del112	no L7Ae *saci_2027* plasmid (Fig. 2.14a); as pMD18; frameshifted (del112) *l7ae* gene; this work

Table 4.11 Recombinant plasmids used for RNA *in vitro* transcription

Plasmid + Insert	Description of the insert
pUC19 + L7P Nat	T7 promoter upstream of native *l7ae* 5' UTR and first 30 bp of *l7ae mRNA*; restriction sites: EcoRI, XbaI; this work
pUC19 + L7P Dmut	T7 promoter upstream of Kt-n mut1 mutated *l7ae* 5' UTR and first 30 bp of *l7ae mRNA*; restriction sites: EcoRI, XbaI; this work
pUC19 + C/D box sRNA 39	T7 promoter upstream of Sac-sR121 C/D box sRNA; restriction sites: BamHI, HindIII; V. Tripp [141]

4.3.2 Oligonucleotides

Genomic DNA of *S. acidocaldarius* MW001 was used as a PCR template for the oligonucleotides listed in Table 4.12.

Table 4.12 Oligonucleotides used for cloning of the plasmids listed in Table 4.5.

Name	Sequence 5' → 3'	Remarks
Saci_1520 CFHA LF for	TAGCGTCCGCGGCCAAAGGAGTTTATCATAGA CATCCCACAG	Left flank primer for *l7ae*; SacII
Saci_1520 CFHA LF rev	GGCGTAGTCTGGAACATCGTAAGGGTAGCTCT TAACCTTGTCATCGTCGTCCTTGTAGTCACTAC TTTTACCTTTTATTTCATTAACTCT	Left flank primer for *l7ae*; C-ter HA/Flag tag
Saci_1520 CFHA RF for	GACTACAAGGACGACGATGACAAGGTTAAGA GCTACCCTTACGATGTTCCAGACTACGCCTGA AAATTTTTGTTATAACTAATATTTTTA	Right flank primer for *l7ae*; C-ter Flag/HA tag
Saci_1520 CFHA RF rev	TTGGCTGTCGACTCCGCCTATTCCTAGACCTG	Right flank primer for *l7ae*; SalI
Saci_1224 LF for	ATTCTGCCATGGATTACTATACTCTGACACGAA AC	Left flank primer for *lsm1*; NcoI
Saci_1224 LF rev	ATGGACTACAAGGACGACGATGACAAGGTTAA GAGCTACCCTTACGATGTTCCAGACTACGCCA TAGTGGGTGTAATAGTGTC	Left flank primer for *lsm1;* N-ter Flag/HA tag
Saci_1224 RF for	GGCGTAGTCTGGAACATCGTAAGGGTAGCTCT TAACCTTGTCATCGTCGTCCTTGTAGTCCATAC ATGATCTAAGGATATTTAAC	Right flank primer for *lsm1;* N-ter HA/Flag tag
Saci_1224 RF rev	TCCAGTGTCGACTTAGTCTAGCTACACAGTCC	Right flank primer for *lsm1*; SalI
Saci_0799 LF for	ATTCTGCCATGGGTCCAAGTTGTGGTGTAA	Left flank primer for *lsm2*; NcoI
Saci_0799 LF rev	ATGGACTACAAGGACGACGATGACAAGGTTAA GAGCTACCCTTACGATGTTCCAGACTACGCCC AAGCCAAGATAGAAAATCC	Left flank primer for *lsm2;* N-ter Flag/HA tag
Saci_0799 RF for	GGCGTAGTCTGGAACATCGTAAGGGTAGCTCT TAACCTTGTCATCGTCGTCCTTGTAGTCCATAG TAAGTCTCCTCACTTTCTTTTTCTGA	Right flank primer for *lsm2;* N-ter HA/Flag tag
Saci_0799 RF rev	TGTAGTGTCGACGTAAGTGAAAATGGATTGGT TGA	Right flank primer for *lsm2*; SalI
Saci_0660 LF for	ATTCTGCCATGGCGTGCTTTGTCTCTAAGATTC	Left flank primer for *lsm3*; NcoI
Saci_0660 LF rev	ATGGACTACAAGGACGACGATGACAAGGTTAA GAGCTACCCTTACGATGTTCCAGACTACGCCA GTGTAAGTAGAAGAATTTCGGGGGAT	Left flank primer for *lsm3;* N-ter Flag/HA tag
Saci_0660 RF for	GGCGTAGTCTGGAACATCGTAAGGGTAGCTCT TAACCTTGTCATCGTCGTCCTTGTAGTCCATAC ATATTATGATATGACTAACATGGTTAAT	Right flank primer for *lsm3;* N-ter HA/Flag tag
Saci_0660 RF rev	TCCAGTGTCGACGTAATGTCTGGTGACGGTAT	Right flank primer for *lsm3*; SalI
Saci_1224 CFHA LF for	ATTCTGCCATGGGGTCAAAAGAGACTTGTCCT	Left flank primer for *lsm1*; NcoI
Saci_1224 CFHA LF rev	GACTACAAGGACGACGATGACAAGGTTAAGA GCTACCCTTACGATGTTCCAGACTACGCCTAA GACTTAAAAATATACTCTTTTAACTTCGCT	Left flank primer for *lsm1*; C-ter Flag/HA tag
Saci_1224 CFHA RF for	TTAGGCGTAGTCTGGAACATCGTAAGGGTAGC TCTTAACCTTGTCATCGTCGTCCTTGTAGTCAA CAGGAGACACTAAGATAACATTA	Right flank primer for *lsm1*; C-ter HA/Flag tag
Saci_1224 CFHA RF rev	TGTAGTGTCGACAGGTAACTACACCGGTAAGA	Right flank primer for *lsm1*; SalI

Saci_0799 CFHA LF for	ATTCTGCCATGGCCGCTGTGGCTTCAAGAGAC	Left flank primer for *lsm2*; NcoI
Saci_0799 CHis LF rev	GACTACAAGGACGACGATGACAAGGTTAAGA GCTACCCTTACGATGTTCCAGACTACGCCTAAT TTTATAATAAACTCATGAGAAGAGGGTGC	Left flank primer for *lsm2*; C-ter Flag/HA tag
Saci_0799 CHis RF for	TTAGGCGTAGTCTGGAACATCGTAAGGGTAGC TCTTAACCTTGTCATCGTCGTCCTTGTAGTCTTT TTCCTTACCCATTATTGATTCGTAGTC	Right flank primer for *lsm2*; C-ter HA/Flag tag
Saci_0799 CFHA RF rev	TGTAGTGTCGACCCGAAGTAAATCCAAATACG	Right flank primer for *lsm2*; SalI
Saci_1224 CHis LF rev	CACCACCACCACCACCACTAAGACTTAAAAAT ATACTCTTTTAACTTCGCT	Left flank primer for *lsm1*; C-ter His tag
Saci_1224 CHis RF for	TTAGTGGTGGTGGTGGTGGTGAACAGGAGACA CTAAGATAACATTA	Right flank primer for *lsm1*; C-ter His tag
Saci_0799 CHis LF rev	CACCACCACCACCACCACTAATTTTATAATAA ACTCATGAGAAGAGGGTGC	Left flank primer for *lsm2*; C-ter His tag
Saci_0799 CHis RF for	TTAGTGGTGGTGGTGGTGGTGTTTTTCCTTACC CATTATTGATTCGTAGTC	Right flank primer for *lsm2*; C-ter His tag
Saci_0660 CHis LF for	ATTCTGCCATGGATAATCGTCCTTCTTTTAGG	Left flank primer for *lsm3*; NcoI
Saci_0660 CHis LF rev	CACCACCACCACCACCACTGAAATGATAGGAG ATTTCGAAATTAAAG	Left flank primer for *lsm3*; C-ter His tag
Saci_0660 CHis RF for	TCAGTGGTGGTGGTGGTGGTGTGATCCTTTTTT CTTTTTATCAATA	Right flank primer for *lsm3*; C-ter His tag
Saci_0660 CHis RF rev	ATGGCTGTCGACCTTGAAGCCGAATACAATGA	Right flank primer for *lsm3*; SalI

The generated pSVA406 + Saci_1520CFHA plasmid was used as a PCR template for the oligonucleotides listed in Table 4.13.

Table 4.13 Oligonucleotides used for cloning of the plasmids listed in Table 4.6.

Name	Sequence 5' → 3'	Remarks
5UTR g539c for	GATAGAGACAATCTATTTTCGCGTCACGAAAA AGAAGGAGG	Mutagenesis primer for Kt-n mut1 mutation
5UTR g539c rev	CCTCCTTCTTTTTCGTGACGCGAAAATAGATTG TCTCTATC	Mutagenesis primer for Kt-n mut1 mutation
g539c_a540t for	GACAATCTATTTTCGCGTCTCGAAAAAGAAGG AGGATGA	Mutagenesis primer for double Kt-n mutation
g539c_a540t rev	TCATCCTCCTTCTTTTTCGAGACGCGAAAATAG ATTGTC	Mutagenesis primer for double Kt-n mutation

Genomic DNA of *S. acidocaldarius* MW001 was used as a PCR template for the oligonucleotides listed in Table 4.14. Exceptions are indicated.

Table 4.14 Oligonucleotides used for cloning of the plasmids listed in Table 4.7.

Name	Sequence 5' → 3'	Remarks
1224crRNA-hyb fw	TTAGCTCTTCAAAGCTAAGACTAAGTTCATATG CATATCATAACTTTTTAGATTGAAGAGCTTA	Hybridization primer for *lsm1* crRNA; SapI; Spacer

1224crRNA-hyb re	TAAGCTCTTCAATCTAAAAAGTTATGATATGC ATATGAACTTAGTCTTAGCTTTGAAGAGCTAA	Hybridization primer for *lsm1* crRNA; SapI; Spacer
del-1224 LF fw	AGTGGGATCCGGTCAAAAGAGACTTGTCCT	Left flank primer for *lsm1* donor DNA; BamHI
del-1224 LF re	CATTAAAGTTAAATATCCTTAGATCATGTTTAA GACTTAAAAATATACTCTTTTAACTTCGC	Left flank primer for *lsm1* donor DNA
del-1224 RF fw	GCGAAGTTAAAAGAGTATATTTTTAAGTCTTA AACATGATCTAAGGATATTTAACTTTAATG	Right flank primer for *lsm1* donor DNA
del-1224 RF re	GTAGACGGCCGAGGTAACTACACCGGTAAGA	Right flank primer for *lsm1* donor DNA; EagI
0799crRNA-hyb fw	TTAGCTCTTCAAAGTCCAACATATTCCTCACCA TTTTTGAGTTTAACTAGCGATTGAAGAGCTTA	Hybridization primer for *lsm2* crRNA; SapI; Spacer
0799crRNA-hyb re	TAAGCTCTTCAATCGCTAGTTAAACTCAAAAA TGGTGAGGAATATGTTGGACTTTGAAGAGCTA A	Hybridization primer for *lsm2* crRNA; SapI; Spacer
del-0799 LF fw[1]	AGTAGGATCCCCGCTGTGGCTTCAAGAGAC	Left flank primer for *lsm2* donor DNA, BamHI
del-0799 RF re[1]	TATTCGGCCGTTGGAGCCAGAGGCGTGGAA	Left flank primer for *lsm2* donor DNA, EagI
0660crRNA-hyb fw	TTAGCTCTTCAAAGCTGGGATCAAATAGGGGT GCAGATTTAACTAGAATTTCGATTGAAGAGCT TA	Hybridization primer for *lsm3* crRNA; SapI; Spacer
0660crRNA-hyb re	TAAGCTCTTCAATCGAAATTCTAGTTAAATCTG CACCCCTATTTGATCCCAGCTTTGAAGAGCTAA	Hybridization primer for *lsm3* crRNA; SapI; Spacer
del_0660 LF fw new	GATGGCTAGCATAATCGTCCTTCTTTTAGG	Left flank primer for *lsm3* donor DNA; NheI
del-0660 LF re	GAATAATATTATTAACCATGTTAGTCATATCAT AATATGTAATGATAGGAGATTTCGAAATTAAA GATGAAGAT	Left flank primer for *lsm1* donor DNA
del-0660 RF fw	ATCTTCATCTTTAATTTCGAAATCTCCTATCAT TACATATTATGATATGACTAACATGGTTAATA ATATTATTC	Right flank primer for *lsm1* donor DNA
del-0660 RF re	GATTCGGCCGGCAGGTGCAGGAATAGTAGG	Right flank primer for *lsm1* donor DNA; EagI

[1] PCR template: pSVA2003

The pSVA431 plasmid was used as a PCR template for the oligonucleotides listed in Table 4.15. Exceptions are indicated.

Table 4.15 Oligonucleotides used for cloning of the plasmids listed in Table 4.8.

Name	Sequence 5' → 3'	Remarks
L7PR Nat for	TAGCGTCCGCGGTCATATTCTCACATGAATAA CTTTTTAATGGATAGAGACAATCTATTTTCGCG TGACGAAAAAGAAGGAGGATGAACGCTGTGT ACTCATTTCCAAATAGCTT	Primer for *lacS* with native promoter and 5' UTR of *l7ae*; SacII
L7PR Dmut for	TAGCGTCCGCGGTCATATTCTCACATGAATAA CTTTTTAATGGATAGAGACAATCTATTTTCGCG	Primer for *lacS* with native promoter and Kt-

	TCACGAAAAAGAAGGAGGATGAACGCTGTGT ACTCATTTCCAAATAGCTT	n mut1 mutated 5' UTR of *l7ae*; SacII
L7PR BTmut for	TAGCGTCCGCGGTCATATTCTCACATGGCCGCC GGGCGCCGGGATAGAGACAATCTATTTTCGCG TGACGAAAAAGAAGGAGGATGAACGCTGTGT ACTCATTTCCAAATAGCTT	Primer for *lacS* with BRE/TATA mutated promoter of *l7ae*; SacII
L7PR HA new rev	AGCCAACGGCCGTTAGGCGTAGTCTGGAACAT CGTAAGGGTAGTGCCTTAATGGCTTTAC	Reverse primer for *lacS*; EagI
QC Dmut GG for[1]	GACAATCTATTTTCGCGTGGCGAAAAAGAAGG AGGATGA	Mutagenesis primer for Kt-n mut2 mutation
QC Dmut GG rev[1]	TCATCCTCCTTCTTTTTCGCCACGCGAAAATAG ATTGTC	Mutagenesis primer for Kt-n mut1 mutation

[1] PCR template: pSVA1431 + L7PR-lacS Nat

The utilized PCR templates for the oligonucleotides listed in Table 4.16 are indicated for each oligonucleotide.

Table 4.16 Oligonucleotides used for cloning of the plasmids listed in Table 4.9.

Name	Sequence 5' → 3'	Remarks
Inv PCR L7Ae for[1,6,7]	ATGTCTAAACCCTCGTATGTAAAAT	Inverse PCR primer for His-Sumo tag removal
Inv PCR L7Ae rev[1]	ATGTATATCTCCTTCTTAAAGTTAAAC	Inverse PCR primer for His-Sumo tag removal
pEC-A PCR for[2]	CGGGACCAGTGACGAAGGCT	Gibson Assembly primer for pEC-A-*l7ae* backbone part 1
pEC-A PCR rev[2]	AACTATGCGGCATCAGAGCA	Gibson Assembly primer for pEC-A-*l7ae* backbone part 1
ROP PCR for[2]	AAGCCAGTATACACTCCGCT	Gibson Assembly primer for pEC-A-*l7ae* backbone part 2
ROP PCR rev[2,5,6,8,9]	CCCACGGGTGCGCATGATCG	Gibson Assembly primer for pEC-A-*l7ae* backbone part 2; Inverse PCR primer
Pcat UTR GFP for[3]	CGATCATGCGCACCCGTGGGGGCACGTAAGAG GTTCCAACTTTCACCATAATGAAATAAGCGTG ACGAAAAAGAAGGAGGATGAACGCTATGAGC AAAGGAGAAGAAC	Gibson Assembly primer for *gfp* with native *l7ae* 5' UTR (*gfp*Kt)
T1 ter GFP rev[3]	AGCCTTCGTCACTGGTCCCGATAAAACGAAAG GCCCAGTCTTTCGACTGAGCCTTTCGTTTTATT TGATGCCTGGTTATTTTTCGAACTGCGGGT	Gibson Assembly primer for *gfp* with T1 terminator
Pcat mcherry fw1[4]	TCACTACCGGGCGTATTTTTTGAGTTATCGAGA TTTTCAGGAGCTAAGGAAGCTAAAATGGTGAG CAAGGGCGAGGA	Gibson Assembly primer 1 for *mcherry*
Pcat mcherry fw2[4]	TGCTCTGATGCCGCATAGTTGGCACGTAAGAG GTTCCAACTTTCACCATAATGAAATAAGATCA CTACCGGGCGTAT	Gibson Assembly primer 2 for *mcherry* with Pcat promoter
T1 ter mcherry rev[4]	AGCGGAGTGTATACTGGCTTATAAAACGAAAG GCCCAGTCTTTCGACTGAGCCTTTCGTTTTATT TGATGCCTGGTTACTTGTACAGCTCGTCCA	Gibson Assembly primer for *mcherry* with T1 terminator

PN25 5UTR gfp fw[5]	TATTTGCTTTCAGGAAAATTTTTCTGTATAATA GATTCGCGTGACGAAAAAGAAGGAG	Inverse PCR primer for insertion of the P_{N25} promoter [154]
del112[6,7]	AACTTTAAGAAGGAGATATACATATG−CTAAA CCCTCGTATGTAAAATTTG	Mutagenesis primer for frameshift of *l7ae*
del112-antisense[6,7]	CAAATTTTACATACGAGGGTTTAG−CATATGTA TATCTCCTTCTTAAAGTT	Mutagenesis primer for frameshift of *l7ae*
N25 doublmutD fw[6,8]	AAAATTTTTCTGTATAATAGATTCGCGTCTCGA AAAAGAAGGAGGATGAACG	Mutagenesis primer for double Kt-n mutated *gfp*Kt
N25 doublmutD re[6,8]	CGTTCATCCTCCTTCTTTTTCGAGACGCGAATC TATTATACAGAAAAATTTT	Mutagenesis primer for double Kt-n mutated *gfp*Kt
a4943t[6,8]	CTTTGCTCATAGCGTTCAACCTCCTTCTTTTTCG TCA	Mutagenesis primer for bulge mutated *gfp*Kt
a4943t_antisense[6,8]	TGACGAAAAAGAAGGAGGTTGAACGCTATGA GCAAAG	Mutagenesis primer for bulge mutated *gfp*Kt
control UTR for[6]	TATTTGCTTTCAGGAAAATTTTTCTGTATAATA GATTCAACATGTCCAATAATAATGGAGTATAA CATATGAGCAAAGGAGAAGAACTTTTCACTGG	Inverse PCR primer for insertion of the control (LII-12) UTR [153]
T7 L7AeUTR rev[6,7]	AGCGTTCATCCTCCTTCTTTTTCGTCACGCATT ATTTCTAGAGGGGAATTGTTATC	Inverse PCR primer for insertion of the native *l7ae* 5' UTR for autoregulation (aKt)
g83c[8,9]	AATTCCCCTCTAGAAATAATGCGTCACGAAAA AGAAGGAG	Mutagenesis primer for Kt-n mut1 mutated aKt
g83c_antisense[8,9]	CTCCTTCTTTTTCGTGACGCATTATTTCTAGAG GGGAATT	Mutagenesis primer for Kt-n mut1 mutated aKt
g83c_a84t[8,9]	CAATTCCCCTCTAGAAATAATGCGTCTCGAAA AAGAAGGAGGATG	Mutagenesis primer for double Kt-n mutated aKt
g83c_a84t_anti[8,9]	CATCCTCCTTCTTTTTCGAGACGCATTATTTCT AGAGGGGAATTG	Mutagenesis primer for double Kt-n mutated aKt
doubleCmut for[8]	TCGCGTGACGAAAAAGAAGGAGGATCTACGCT ATGAGCAA	Mutagenesis primer for double Kt-b mutated *gfp*Kt
doubleCmut rev[8]	TTGCTCATAGCGTAGATCCTCCTTCTTTTTCGT CACGCGA	Mutagenesis primer for double Kt-b mutated *gfp*Kt
APE_1818 UTR fw[8]	TATTTGCTTTCAGGAAAATTTTTCTGTATAATA GATTCGCCACACCCAGATTCTAGAGGCTGTGA CGACCTGAAAGGGGAGGAGGAGCCATGAGCA AAGGAGAAGAACTTTTCACTGG	Inverse PCR primer for insertion of the 50 bp *l7ae* upstream region of *A. pernix*
Smar_0825 UTR fw[8]	TATTTGCTTTCAGGAAAATTTTTCTGTATAATA GATTCATTATTTGGTTTTAAATCAGACGACCGA AGAGTTGGGAGGAGGATGAAGTTATGAGCAA AGGAGAAGAACTTTTCACTGG	Inverse PCR primer for insertion of the 50 bp *l7ae* upstream region of *S. marinus*
PAE3347 UTR fw[8]	TATTTGCTTTCAGGAAAATTTTTCTGTATAATA GATTCGTTACGCGCAGAGGCATAAAGATTTAT AAAGCCAGTTCTTTGTGGCAACCATGAGCAAA GGAGAAGAACTTTTCACTGG	Inverse PCR primer for insertion of the 50 bp *l7ae* upstream region of *P. aerophilum*

AF0764 UTR fw[8]	TATTTGCTTTCAGGAAAATTTTTCTGTATAATA GATTCGGAATTAGGGAAGAGCCGATGAATGAT GATTTGATGAAGGAGGTGATGACATGAGCAAA GGAGAAGAACTTTTCACTGG	Inverse PCR primer for insertion of the 50 bp *l7ae* upstream region of *A. fulgidus*
HVO_2737 UTR fw[8]	TATTTGCTTTCAGGAAAATTTTTCTGTATAATA GATTCGTGATACCGGCGAGGGAGTCGAGCCCG CGAGCAGGATATAGGTGAACAACAATGAGCA AAGGAGAAGAACTTTTCACTGG	Inverse PCR primer for insertion of the 50 bp *l7ae* upstream region of *H. volcanii*
MA1521 UTR fw[8]	TATTTGCTTTCAGGAAAATTTTTCTGTATAATA GATTCACCTATGATTGATGAAAAATTGCGAGA CCCGCAATTTCGAAGGAGAAACTTAAATGAGC AAAGGAGAAGAACTTTTCACTGG	Inverse PCR primer for insertion of the 50 bp *l7ae* upstream region of *M. acetivorans*
TK1311 UTR fw[8]	TATTTGCTTTCAGGAAAATTTTTCTGTATAATA GATTCGCCCGAGTCGGTGGTGCAAACAGATGA ACGATGAGGTTTCGGAGGGATGAAGATGAGCA AAGGAGAAGAACTTTTCACTGG	Inverse PCR primer for insertion of the 50 bp *l7ae* upstream region of *T. kodakarensis*
MmarC5 UTR fw[8]	TATTTGCTTTCAGGAAAATTTTTCTGTATAATA GATTCGTGGGCGCCATCCGGATAAAAAAATTT TTTTAAGTGATGAAGGAGGTCATAATATGAGC AAAGGAGAAGAACTTTTCACTGG	Inverse PCR primer for insertion of the 50 bp *l7ae* upstream region of *M. maripaludis*
del112+UTR for[8,9,10,11,12,13,14,15,16,17,18,19]	GGAGGATGAACGCTATG−CTAAACCCTCGTAT GTA	Mutagenesis primer for frameshift of *l7ae* in plasmids with aKt
del112+UTR rev[8,9,10,11,12,13,14,15,16,17,18,19]	TACATACGAGGGTTTAG−CATAGCGTTCATCCT CC	Mutagenesis primer for frameshift of *l7ae* in plasmids with aKt
Kt_1468 for[9]	CCAATAATAATGGAGTATAACATATGGAATAT GATGAAGAGTTAGATGAACTTCTGAAAAAAAG AGCAACAGAGAGCAAAGGAGAAGAACTTTTC ACTG	Inverse PCR primer for *saci_1468* Kt insertion after *gfp* start; part 1
Kt_2027 for[9]	CCAATAATAATGGAGTATAACATATGACTGAT GATGACGCTATACCCTCTGACACGTGGATCCA AAGCAAAGGAGAAGAACTTTTCACTG	Inverse PCR primer for *saci_2027* Kt insertion after *gfp* start; part 1
mRNA KT 2nd for[9]	TATTTGCTTTCAGGAAAATTTTTCTGTATAATA GATTCAACATGTCCAATAATAATGGAGTATAA CATATG	Inverse PCR Primer for mRNA Kt insertion + control UTR; part 2

[1] PCR template: pEC-A HiSumo + Saci_1520; [2] PCR template: pEC-A + Saci_1520 PCR product; [3] PCR template: pASK-IBA3plus-*sfgfp*; [4] PCR template: pDS43; [5] PCR template: pEC-A-*l7ae-sfgfp-mcherry*; [6] PCR template: pEC-A-*l7ae-sfgfp-mcherry* PN25; [7] PCR template: pEC-A-*l7ae-sfgfp-mcherry* PN25 control UTR; [8] PCR template: pMD2; [9] PCR template: pMD1; [10] PCR template: pMD9; [11] PCR template: pMD10; [12] PCR template: pMD11; [13] PCR template: pMD12; [14] PCR template: pMD13; [15] PCR template: pMD14; [16] PCR template: pMD15; [17] PCR template: pMD16; [18] PCR template: pMD17; [19] PCR template: pMD18

The hybridization oligonucleotides listed in Table 4.17 were directly cloned into pUC19.

Table 4.17 Oligonucleotides used for cloning of the plasmids listed in Table 4.11.

Name	Sequence 5' → 3'	Remarks
L7P Nat hyb for	GCTCGAATTCTAATACGACTCACTATAGCGTG ACGAAAAAGAAGGAGGATGAACGCTGTGTCT AAACCCTCGTATGTAAAATTTGAATCTAGAGC TC	Hybridization primer with T7 promoter and native *l7ae* 5' UTR; EcoRI, XbaI

L7P Nat hyb rev	GAGCTCTAGATTCAAATTTTACATACGAGGGT TTAGACACAGCGTTCATCCTCCTTCTTTTTCGT CACGCTATAGTGAGTCGTATTAGAATTCGAGC	Hybridization primer with T7 promoter and native *l7ae* 5' UTR; EcoRI, XbaI
L7P Dmut hyb for	GCTCGAATTCTAATACGACTCACTATAGCGTC ACGAAAAAGAAGGAGGATGAACGCTGTGTCT AAACCCTCGTATGTAAAATTTGAATCTAGAGC TC	Hybridization primer with T7 promoter and Kt-n mut1 mutation; EcoRI, XbaI
L7P Dmut hyb rev	GAGCTCTAGATTCAAATTTTACATACGAGGGT TTAGACACAGCGTTCATCCTCCTTCTTTTTCGT GACGCTATAGTGAGTCGTATTAGAATTCGAGC	Hybridization primer with T7 promoter and Kt-n mut1 mutation; EcoRI, XbaI

The PCR templates for the oligonucleotides listed in Table 4.18 are indicated for each oligonucleotide. The listed hybridization oligonucleotides were used as PCR templates to generate *in vitro* transcription templates.

Table 4.18 Oligonucleotides used to generate templates for *in vitro* transcription.

Name	Sequence 5' → 3'	Remarks
pUC19 PCR for[1,2]	GCTGCAAGGCGATTAAG	Left PCR primer for Nat and Kt-n mut1 RNA *in vitro* template
UTR+CDS PCR rev[1,2]	TTCAAATTTTACATACGAGGGTTT	Right PCR primer for Nat and Kt-n mut1 RNA *in vitro* template
pUC19IvFor[3]	TGTGCTGCAAGGCGATTAAG	Left PCR primer for Sac-sR121 RNA *in vitro* template
CD39PCRrev[3]	GAGATATCAGACATTAGAAGG	Right PCR primer for Sac-sR121 RNA *in vitro* template
Nop56L hyb for	GCTCGAATTCTAATACGACTCACTATAGAGAA TTGATGAAGATATATTTAGTGGAACATGCAAT AGGATCGTTTGGATATGACGAGAGCGGAAAAT TAAT	Hybridization primer with T7 promoter and Nop56 mRNA Kt
Nop56L hyb rev	GAGCTCTAGAAATTAAAGCTTCAGTAACTTTTC CAATATCTTTACTATTTGGTACAAAATCTATTA ATTTTCCGCTCTCGTCATATCCAAACGATCCTA T	Hybridization primer with T7 promoter and Nop56 mRNA Kt
SRP RNA hyb for	GCTCGAATTCTAATACGACTCACTATAGTCTAA CTATGATCAGGAGCGATAGGAGGAAGACTCTA GAGCTC	Hybridization primer with T7 promoter and SRP Kt
SRP RNA hyb rev	GAGCTCTAGAGTCTTCCTCCTATCGCTCCTGAT CATAGTTAGACTATAGTGAGTCGTATTAGAAT TCGAGC	Hybridization primer with T7 promoter and SRP Kt
T7 PCR for[4,5]	GCTCGAATTCTAATACGACTCACT	Left PCR primer for Nop56 mRNA Kt and SRP RNA Kt RNA *in vitro* template

Nop56L PCR rev[4]	AATTAAAGCTTCAGTAACTTTTCCAATATC	Right PCR primer for Nop56 mRNA Kt *in vitro* template
SRP RNA PCR rev[5]	GTCTTCCTCCTATCGCTCCT	Right PCR primer for SRP RNA Kt *in vitro* template

[1] PCR template: pUC19 + L7P Nat; [2] PCR template: pUC19 + L7P Dmut; [3] PCR template: pUC19 + C/D box sRNA 39; [4] PCR template: Nop56L hybrid; [5] PCR template: SRP RNA hybrid

Table 4.19 Oligonucleotides used for colony PCR.

Name	Sequence 5' → 3'	Remarks
1520cTag Seq for	TGTAGCTCATTTACC	Left primer to verify CFHA tag of *l7ae*
1520cTagcheckrev	CCTATCTTCAAACCC	Right primer to verify CFHA tag of *l7ae*
1224 Tag check for	TTGTACTTCCGTCATTCATT	Left primer to verify NFHA tag of *lsm1*
1224 Tag check rev	TCAGCAAAATACTCTTCGAG	Right primer to verify NFHA tag of *lsm1* or to check *lsm1* deletion
0799 Tag check for	TAAGTTCATTGTTCCATCGC	Left primer to verify NFHA tag of *lsm2*
0799 Tag check rev	ATGTCCTCTGGTTCAGTCAA	Right primer to verify NFHA tag of *lsm2*
0660 Tag check for	GCATTGAGTAGAGATACCAT	Left primer to verify NFHA tag of *lsm3*
0660 Tag check rev	TAATGCACCCTTTTCAATGG	Right primer to verify NFHA tag of *lsm3* or to check *lsm3* deletion
1224cTag Seq for	GGACTCAATGAAAAGG	Left primer to verify CFHA/CHis tag of *lsm1*
1224cTagcheckrev	TGACGGAAGTACAAG	Right primer to verify CFHA/CHis tag of *lsm1*
0799cTag Seq for	GAGTTGCTTGTCCTCCTAC	Left primer to verify CFHA/CHis tag of *lsm2*
0799cTagcheckrev	GAGTTGCTTGTCCTCCTAC	Right primer to verify CFHA/CHis tag of *lsm2*
Saci_0660 LF for	ATTCTGCCATGGCGTGCTTTGTCTCTAAGATTC	Left primer to verify CHis tag of *lsm3*
Saci_0660 N-His LF rev	ATGCACCACCACCACCACCACAGTGTAAGTAG AAGAATTTCGGGGGAT	Right primer to verify CHis tag of *lsm3*
Saci_1224 LF for	ATTCTGCCATGGATTACTATACTCTGACACGAA AC	Left primer to check *lsm1* deletion
Saci_0799 LF for	ATTCTGCCATGGGTCCAAGTTGTGGTGTAA	Left primer to check *lsm2* deletion
Saci_0799 RF rev	TGTAGTGTCGACGTAAGTGAAAATGGATTGGT TGA	Right primer to check *lsm2* deletion
Saci_0660 LF for	ATTCTGCCATGGCGTGCTTTGTCTCTAAGATTC	Left primer to check *lsm3* deletion

Genomic DNA of *S. acidocaldarius* MW001 was used as a PCR template for the oligonucleotides listed in Table 4.20.

Table 4.20 Oligonucleotides used for pyrEF-exchange PCR products.

Name	Sequence 5' → 3'	Remarks
1224 KO pyrEX for	TAGTACAAGCTTCGATAGCGAAGTTAAAAGAG TATATTTTTAAGTCTTAATTTGAGCAGTTCTAG	pyrEF-exchange PCR primer with 50 bp *lsm1* downstream region
1224 KO pyrEX rev	TGTTAATGAGGTTTGGCATATCATTAAAGTTAA ATATCCTTAGATCATGTGACCGGCTATTTTTTC AC	pyrEF-exchange PCR primer with 50 bp *lsm1* upstream region
0799 KO pyrEX for	ATTTATATTTTTTGCCATTAGCACCCTCTTCTCA TGAGTTTATTATAAAATTTGAGCAGTTCTAG	pyrEF-exchange PCR primer with 50 bp *lsm2* downstream region
0799 KO pyrEX rev	AACTCTTTTATATTAGACATTATCAGAAAAAG AAAGTGAGGAGACTTACTGACCGGCTATTTTT TCAC	pyrEF-exchange PCR primer with 50 bp *lsm2* upstream region
0660 KO pyrEX for	CTATTCTTCCTGCAAGATCTTCATCTTTAATTTC GAAATCTCCTATCATTTTTGAGCAGTTCTAG	pyrEF-exchange PCR primer with 50 bp *lsm3* downstream region
0660 KO pyrEX rev	CACATCCACTGAATAATATTATTAACCATGTTA GTCATATCATAATATGTGACCGGCTATTTTTTC AC	pyrEF-exchange PCR primer with 50 bp *lsm3* upstream region

Table 4.21 Oligonucleotide used for Northern blot analysis.

Name	Sequence 5' → 3'	Remarks
Saci11probe	CGCTTTTTGTCATCATTCTCAGATCCCGGATTC CACATC	DNA probe against Sac-sR10

4.4 Molecular working with DNA

4.4.1 DNA isolation

4.4.1.1 Phenol/Chloroform extraction of *S. acidocaldarius* genomic DNA

S. acidocaldarius cells were grown to the late logarithmic phase and 1.5 ml were harvested by centrifugation (14,600 rpm, 2 min, RT). The cell pellet was resuspended in 250 µl TEN buffer (10 mM Tris-HCl pH 8, 1 mM EDTA, 150 ml NaCl) and then mixed with 250 µl of TENST buffer (TEN supplemented with 0.12 % Triton X-100) by inverting the tube. After incubation for 30 min at RT, 500 µl of phenol/chloroform/isoamyl alcohol (25:24:1, v/v) were added and mixed using a vortex. After centrifugation (14,600 rpm, 2 min, RT), the intermediate phase was removed by a toothpick and 450 µl of the upper (aqueous) layer were transferred to a new reaction tube. The isolated DNA was subsequently concentrated by DNA precipitation (chapter 4.4.1.2).

4.4.1.2 DNA precipitation

DNA solutions were concentrated by precipitation using ethanol [210]. Two volumes of 100 % EtOH (v/v) and 0.3 M Na-acetate pH 5.8 were added to the solution and incubated for 30 min at -20°C. After centrifugation of the sample (14,600 rpm, 10 min, RT), the DNA pellet was washed two times with 70 % EtOH (v/v) (14,600 rpm, 2 min, RT), dried and resuspended in 10 mM Tris-HCl pH 8.5.

4.4.1.3 Plasmid DNA isolation from *E. coli*

Overnight cultures of *E. coli* (2-5 ml) were used for the preparation of plasmid DNA. Plasmid DNA was isolated using the QIAprep Spin Miniprep Kit or the QIAGEN Plasmid Plus Maxi Kit according to the manufacturer´s instructions.

4.4.2 Quantitation of DNA

4.4.2.1 Spectrophotometric quantitation and quality control

DNA concentrations in aqueous solutions were measured photometrically at a wavelength of λ = 260 nm using a spectrophotometer (Nanodrop). 50 µg of dsDNA/ml correlate to an extinction value of 1 [210]. In order to determine the quality of the DNA preparations, the spectrophotometer also measured the absorption at λ = 280 nm, which is the absorption wavelength of proteins. A260/A280 ratios of 1.8 to 2.0 corresponded to sufficiently pure DNA preparations, while lower values indicated high protein contamination.

4.4.2.2 Fluorometric quantitation

The Qubit fluorometer was used for high sensitivity quantification of low-yield DNA preparations, e.g. after cDNA library preparation (chapter 4.6.9). The Qubit utilizes fluorescent dyes (for ss/dsDNA, RNA or protein), which emit a signal only when bound to the specific target. By calibration with DNA/RNA/protein standards, the concentration of the utilized sample can be determined. DNA samples were quantified via the Qubit dsDNA HS Assay Kit according to the manufacturer´s instructions.

4.4.3 Polymerase chain reaction (PCR)

In vitro amplification of DNA was performed via the PCR technique [213, 214]. In this method, two sequence-specific DNA oligonucleotides (primers) hybridize to the 3' end of the sense and antisense strands of DNA, flanking the sequence of interest that has to be amplified. A heat-stable DNA polymerase enzyme then elongates the oligonucleotides in 5' to 3' direction. The

Phusion polymerase was used for all PCR reactions, which possesses a reduced mutation rate due to proofreading capability.

The PCR reaction includes the following three main steps: i) denaturing, ii) annealing and iii) elongation. During the denaturing step, the dsDNA that has to be amplified is melted into two ssDNA strands by incubation at over 95°C. In the annealing step, the primers hybridize to the melted ssDNA strands. During elongation, the annealed primers are extended via the 5' to 3' direction by incorporating complementary nucleotides by the DNA polymerase. Exponential amplification of the DNA is achieved by multiple rounds of the three steps. The annealing temperature of the primers is a crucial factor that affects the yield and purity of the amplified PCR product. The melting temperature of designed primers (chapter 4.3.2) was calculated using the following formula [215]:

$$Tm = 64.9 + 41 \times \frac{(nG + nC - 16.4)}{(nA + nT + nG + nC)}$$

4.4.3.1 Amplification of genomic DNA and plasmid DNA

PCR reactions were performed using 50 ng of template DNA (genomic or plasmid DNA), 0.5 µM of each primer, 200 µM of dNTPs, 1x Phusion GC reaction buffer, 3% (v/v) DMSO and 1 U of the Phusion polymerase. A thermal cycler was used for the PCR reaction. The standard cycling conditions are shown in the following table:

Table 4.22 PCR program for Phusion polymerase reactions.

Step	Cycle 1	Cycles 2-35	Cycle 36
Denaturation	98° C, 30s	98° C, 10s	
Annealing		42-65° C, 30s	
Elongation		72° C, 30s/kb	72° C, 5 min

4.4.3.2 Colony PCR

Colony PCRs of single colonies were performed to analyze the presence of plasmids, successful genomic integrations or recombination events after selection. For this, the picked colonies were lysed in 30 µl 0.2 M NaOH and afterwards neutralized with 70 µl of 0.2 M Tris pH 7.0. 1-2 µl of the lysate was used in 15 µl Phusion PCR reactions (Table 4.22) for amplification.

4.4.3.3 Overlap extension PCR

Overlap extension PCRs were used for the genomic tagging or deletion of the *l7ae* and *lsm* genes. First, left flank (LF) and right flank (LF) PCR products were generated. The utilized LF reverse and RF forward primers were designed in way to contain overlapping regions. The purified PCR products and the LF forward and RF reverse primers were then mixed. During

the reaction, the overlapping regions of the two PCR products anneal and are extended to form one joined PCR product. The added primers are then used to exponentially amplify this product. Phusion reactions with 25 ng of each PCR product were performed for the overlap extension PCRs.

4.4.3.4 Site-directed mutagenesis PCR

Site-directed mutagenesis PCR was used to introduce deletions, mutations or insertions into plasmid DNA sequences [216]. A primer set containing the desired mutations was designed using Agilent´s QuikChange primer design tool. Two 50 µl Phusion PCR reactions with one mutagenesis primer each was set up. Ten PCR cycles were performed for each reaction. Afterwards, 25 µl of the two reactions were mixed, 1 U of fresh Phusion polymerase was added and another 18 cycles were performed for the reaction. The annealing temperature was 55°C for all site-directed mutagenesis PCRs. The PCR reaction was afterwards restricted with the DpnI enzyme to degrade the used template plasmid and transformed into *E. coli* (see chapters 4.4.7.1 and 4.4.8.2)

4.4.4 Electrophoresis of DNA

4.4.4.1 Agarose gel electrophoresis

DNA mixtures of different lengths were separated in an electric field by agarose gel electrophoresis. DNA migrates towards the anode due to its negative charge. Agarose gels with 0.8 to 2 % (w/v) agarose in TAE buffer (40 mM Tris-acetate, 1 mM EDTA, 20 mM acetic acid pH 8.0) and 0.1 % (v/v) ethidium bromide were utilized. The DNA samples were mixed with 6x loading dye (40 % (v/v) sucrose, 0.25 % (w/v) bromophenol blue, 0.25 % (w/v) xylene cyanol) prior to loading on the gel. The electrophoresis was carried out at 100-120 V for 30-60 min in TAE buffer. The Quick-Load® 2-Log DNA Ladder and Low Molecular Weight DNA ladder were used as size standards. The DNA was visualized by UV irradiation at 254 nm using an UV transilluminator.

4.4.4.2 Non-denaturing polyacrylamide gel electrophoresis (native PAGE)

Small DNA fragments under 300 nt were separated under non-denaturing conditions (e.g. cDNA libraries) using polyacrylamide gel electrophoresis [210]. The concentration of the polyacrylamide (acrylamide/bisacrylamide, ratio 29:1) in the gels (90 mM Tris pH 8.0, 90 mM boric acid, 2 mM EDTA, 0.03% (v/v) APS, 0.005% (v/v) TEMED) varied between 6 to 12 % (v/v), depending on the size of the DNA. The DNA samples were mixed with 6x loading dye

(chapter 4.4.4.1) prior to loading on the gel. The electrophoresis was carried out at 8-12 W for 30-60 min in TBE buffer (90 mM Tris pH 8.0, 90 mM boric acid, 2 mM EDTA). The Quick-Load® pBR322 DNA-MspI Digest DNA ladder and Low Molecular Weight DNA ladder were used as size standards. After electrophoresis, the polyacrylamide gels were stained for 5 min in 1x SybrGold dissolved in TBE buffer. Visualization of the DNA was carried out at 254 nm using an UV transilluminator.

4.4.4.3 Denaturing polyacrylamide gel electrophoresis (denaturing PAGE)

Denaturing polyacrylamide gel electrophoresis was performed for DNA oligonucleotides (e.g. radiolabeled DNA oligonucleotides). The electrophoresis was performed as described for native PAGE (chapter 4.4.4.2), but using an 8 M urea concentration in the prepared polyacrylamide gels. The DNA samples were mixed with a 2x concentrated formamide loading dye (95 % (v/v) formamide, 0.025 % (w/v) bromophenol blue, 0.025 % (w/v) xylene cyanol, 5 mM EDTA pH 8.0) and incubated for 5 min at 95°C prior to loading the gel.

4.4.5 Purification of DNA fragments

4.4.5.1 PCR purification

PCR reactions that showed a single band after gel electrophoresis were directly purified using the QIAquick PCR Purification kit according to the manufacturer´s instructions.

4.4.5.2 Gel extraction from agarose gels

DNA fragments, e.g. after restriction or PCR, were excised from agarose gels using a scalpel. The DNA band was then purified via gel extraction using the QIAquick Gel Extraction kit following the instructions of the manufacturer.

4.4.5.3 Gel extraction from polyacrylamide gels

DNA was extracted from polyacrylamide gels by excising the respective bands from the gel using a scalpel. The gel piece was then transferred to a Gel Breaker tube (centrifuge tube with small holes) and centrifuged (14,600 rpm, 2 min, RT) into a 2 ml collection tube. The collected gel debris was subsequently covered with 500 µl gel elution buffer (20 mM Tris-HCl pH 7.5, 250 mM sodium acetate, 1 mM EDTA, 0.25 % SDS) and incubated overnight on ice while shaking (300 rpm). Afterwards, the DNA containing gel elution buffer was transferred to a Costar® centrifuge filter tube and centrifuged (14,600 rpm, 2 min, RT) to remove remaining gel debris. 1 % (v/v) glycogen was added to allow precipitation of small DNA molecules. The DNA

was subsequently precipitated using ethanol (chapter 4.4.1.2). The air-dried DNA pellet was then dissolved in nuclease-free 20 mM Tris-HCl pH 7.5.

4.4.6 Hybridization of DNA oligonucleotides

Complementary DNA oligonucleotides were hybridized by mixing 1 nmol of each oligonucleotide in a reaction volume of 10 µl ddH_2O. The mixture was incubated for 10 min at 95°C in a thermomixer and gradually cooled down to RT for 1-2 h. The hybridized oligonucleotides were then cloned into vectors or used as RNA *in vitro* transcription templates.

4.4.7 Enzymatic modification of DNA

4.4.7.1 Restriction

DNA was restricted with restriction endonucleases in the appropriate buffer according to the instructions of the manufacturer. A standard DNA restriction reaction contained 1 µg of DNA and 10-20 U restriction enzyme and was incubated at 37°C for 1-2 h or overnight. For site-directed mutagenesis PCR reactions, 200 U of DpnI and the recommended reaction buffer were added to the 50 µl reaction and incubated overnight.

4.4.7.2 Phosphorylation

Inverse PCR products were phosphorylated by T4 Polynucleotide Kinase (PNK) to allow self-ligation. To this end, 200 ng of inverse PCR product were mixed with 10 U T4 PNK in 1x DNA ligase reaction buffer containing ATP. The reaction was incubated for 1 h at 37°C.

4.4.7.3 Dephosphorylation

The 5' ends of restricted plasmid DNA were dephosphorylated using Antarctic Phosphatase to avoid self-ligation during the ligation reaction. A standard dephosphorylation reaction included 1 µg restricted plasmid DNA and 5-10 U Antarctic Phosphatase in the recommended reaction buffer of the manufacturer. The reaction was incubated at 37°C for 1-2 h and then heat inactivated at 65°C for 15 min.

4.4.7.4 Ligation

Restricted DNA fragments were ligated into vector DNA using the T4 DNA ligase. A standard ligation reaction contained 0.02 pmol vector DNA, 0.06 pmol insert DNA (ratio 1:3) and 4 U T4 DNA ligase in the recommended DNA ligase reaction buffer containing ATP. Phosphorylated inverse PCR products were self-ligated by addition of 10 U T4 DNA ligase to

the phosphorylation reaction (chapter 4.4.7.2). The reactions were incubated overnight at 16°C. Afterwards, the reactions were used for transformation.

4.4.7.5 Gibson Assembly

Gibson Assembly was performed to construct the pEC-A-*l7ae-sfgfp-mcherry* PN25 plasmid, which was used as a progenitor plasmid for the GFP reporter system developed in this work. This technique utilizes DNA fragments that contain overlapping regions (15-20 nt) to assemble these into one construct [217]. During isothermal conditions, a T5 exonuclease degrades dsDNA in 5' to 3' direction, resulting in long 3' overhangs which bind to the complementary overhangs of the neighboring DNA fragment. A DNA polymerase then fills up the single-stranded DNA by incorporating the complementary nucleotides. The resulting gaps are afterwards sealed by a DNA ligase. A self-made Gibson Assembly reaction mix was used, which did not contain the DNA ligase, as *E. coli* contains its own ligase. This Hot Fusion reaction mix was proved to contain a higher assembly efficiency than the original Gibson Assembly mix [218]. In total, four PCR products were assembled: i) pEC-A-*l7ae* backbone part 1, ii) pEC-A-*l7ae* backbone part 2, iii) *gfp*-T1 terminator PCR product with native *l7ae* 5' UTR and iv) *mcherry*-T1 terminator PCR product. The reaction contained 100 fmol of each PCR product, 1.5 U T5 exonuclease and 20 U Phusion DNA polymerase in pre-assembly buffer (100 mM Tris pH 8.0, 10 mM MgCl2, 200 µM dNTPs, 10 mM DTT and 5 % (v/v) PEG-8000) and was incubated for 1 h at 50°C. The reaction was afterwards transformed into *E. coli*.

4.4.8 Transformation

4.4.8.1 Preparation of chemically competent *E. coli* cells

All *E. coli* strains (Table 4.3) were treated with RbCl (rubidium chloride) and $CaCl_2$ to generate chemically competent cells [219]. 100 ml LB medium containing 10 mM $MgCl_2$ and 10 mM $MgSO_4$ were inoculated with 2 ml of an overnight culture and grown until OD_{600nm} 0.6. The culture was cooled on ice for 30 min and then harvested by centrifugation (3,000×*g*, 10 min, 4°C). Afterwards, the cell pellet was gently resuspended in 33 ml cold RF1 solution (30 mM potassium acetate pH 5.8, 100 mM RbCl, 50 mM $MnCl_2$, 10 mM $CaCl_2$ and 15 % glycerol) and incubated on ice for 30 min. Subsequently, the cells were centrifuged (3,000×*g*, 10 min, 4°C) and the pellet gently resuspended in 5 ml cold solution RF2 solution (10 mM RbCl, 10 mM MOPS pH 5.8, 75 mM $CaCl_2$, 15 % glycerol). The cells were again incubated for 30 min on ice and 100 µl aliquots were stored at -80°C.

4.4.8.2 Transformation of *E. coli*

Chemically competent *E. coli* cells were gently mixed with 1-10 µl of plasmid DNA and incubated for 15 min on ice. The cells were heat-shocked by incubation at 42°C for 45 s and then immediately placed on ice. After 5 min, 1 ml of LB medium was added to the cells and the mixture was incubated for 1 h at 37°C, gently shaking at 300 rpm. 100 µl of the transformed cells were plated on LB plates containing the appropriate antibiotics. The remaining cells were pelleted by centrifugation (8000 rpm, 30 s, RT) and resuspended in 100 µl LB medium prior to plating. The plates were incubated overnight at 37°C until visible colonies were formed. The colonies were afterwards screened for positive clones containing the correct recombinant plasmid.

4.4.8.3 Preparation of electrocompetent *S. acidocaldarius* cells

S. acidocaldarius MW001 and ΔSac-sR10 cells were grown in 50 ml of Brock medium pH 3.5 containing 0.1 % (w/v) tryptone (w/v), 0.2 % dextrin and 20 µg/ml uracil. The culture was grown to an OD_{600nm} of 0.3-0.7 (for 1-2 days) and 200 µl were transferred into 50 ml fresh medium. The cells were grown to an OD_{600nm} of 0.1-0.3. After 30 min of incubation on ice, the cell culture was harvested by centrifugation (2,000×*g*, 20 min, 4°C) and the pellet washed three times successively with 50 ml, 10 ml and 1 ml of ice-cold 20 mM sucrose. Finally, the pellet was resuspended in ice-cold 20 mM sucrose adjusted to an OD_{600nm} of 15 and 50 µl aliquots were stored at -80°C.

4.4.8.4 Transformation of *S. acidocaldarius*

Plasmids that were used for transformation into *S. acidocaldarius* (pSVA plasmids) were first methylated by transformation into the *E. coli* ER1821 strain, which contained the plasmid pM.EsaBC4I. This strain methylates the N4-position of the inner cytosine residues of GGCC, which is the recognition sequence of the *S. acidocaldarius* restriction endonuclease SuaI [220]. 200-500 ng of the methylated plasmids were afterwards electroporated into electrocompetent *S. acidocaldarius* MW001 or ΔSac-sR10 cells. Electroporation was performed in 1 mm cuvettes in a constant time protocol using the input parameter 1.5 kV, 25 µF and 600 Ω. 450 µl of Brock medium pH 4.5 containing 0.1 % (w/v) tryptone or NZ-Amine, as well as 0.2 % (w/v) sucrose for normal growth or 0.3 % (w/v) xylose for induction of the *saci_1938* promoter (production of the crRNAs) were used to resuspend the transformed cells. Regeneration was performed for 30 min at 75°C. Afterwards, 100 µl of the reaction was plated on first selection plates. The remaining cells were pelleted by centrifugation (8000 rpm, 30 s, RT) and resuspended in 100

µl regeneration medium prior to plating. The plates were incubated for seven days at 75°C and sealed in plastic bags to avoid drying-out.

4.4.9 Sequencing

All generated recombinant plasmids were verified via automated Sanger sequencing (Eurofins MWG Operon, Ebersberg) [221]. Sequencing always covered the inserts, the transition sites and parts of the backbone.

4.4.10 Radioactive labeling of DNA 5' termini

The single-stranded DNA oligonucleotide used for northern blot analysis (Saci11probe, Table 4.21) was 5'-labeled by radioactive P^{32}. For this, 10 pmol of Saci11probe DNA was mixed with 5 pmol [γ-32P]-ATP, 10 U T4 PNK and the corresponding reaction buffer and incubated for 1 h at 37°C. The radioactively labeled DNA was purified via denaturing PAGE and gel extraction (chapters 4.4.4.3 and 4.4.5.3). To visualize the radiolabeled DNA after denaturing PAGE, the gel was put into a plastic bag and exposed to a phosphor screen (15 min), which was subsequently analyzed by a phosphorimager. The radioactivity of the gel eluted DNA was determined using a scintillation counter.

4.5 Generation of genomic tags and mutations in *S. acidocaldarius*

Genomic Flag-HA/His tagging of the *l7ae* and *lsm* genes or mutation of the endogenous *l7ae* 5' UTR was performed using the pop-in/pop-out method described by Wagner *et al.* [139]. *S. acidocaldarius* MW001 cells were transformed with the respective plasmids (tables 4.5 and 4.6). Transformants were grown to the logarithmic phase in 50 ml Brock media supplemented with 0.1 % (w/v) tryptone and 0.2 % (w/v) sucrose. 10 µl of the cell cultures were diluted in 100 µl ddH_2O and plated on second selection plates bearing 5-FOA and uracil. The 5-FOA induced the loss of the plasmid, yielding a 50 % chance of the aimed mutation. The plates were incubated for 5 days and formed colonies were screened for correct mutations by colony PCR and sequencing (primers listed in Table 4.19).

4.6 Molecular working with RNA

4.6.1 Treatment of solutions, glassware and equipment

All solutions were treated with 0.1 % (v/v) DEPC, incubated overnight at RT and autoclaved to protect RNA from degradation by RNases. Disposable plastic ware (e.g. reaction tubes or pipette tips) was purchased in RNase-free conditions and autoclaved before use. Glassware was

sterilized by incubation at 200°C for at least 2 h. Non-heat resistant equipment (e.g. pipettes) was treated with RNase Exitus Plus™ before use.

4.6.2 Isolation of *S. acidocaldarius* total and small RNAs

Total and small RNAs (<200 nt) of *S. acidocaldarius* were isolated using the *mir*Vana™ miRNA isolation kit according to the manufacturer´s instructions. Briefly, 500 ng of *S. acidocaldarius* cells were resuspended in 1 ml lysis/binding buffer and lysed by a homogenizer for 3 min on ice. 100 µl miRNA Homogenate Additive™ was added to the cell lysate and incubated for 10 min on ice. Afterwards the RNA was isolated by acid-phenol/chloroform extraction. Small RNAs under 200 nt were separated from total RNAs by adjusting the ethanol concentration in the aqueous phase to 25 %. This immobilized larger RNAs on a glass-fiber filter tube, while small RNAs were collected in the filtrate. Afterwards, the EtOH concentration of the filtrate was increased to 55 % and the small RNAs were immobilized on a second glass-fiber filter tube. The two immobilized RNA fractions (small RNA-depleted total RNA and small RNAs) were washed and subsequently eluted in 100 µl nuclease-free water.

4.6.3 Co-immunoprecipitation of L7Ae-interacting RNAs

The L7Ae interacting RNAs were isolated using the Flag® HA Tandem Affinity Purification kit, which involves two subsequent immunoprecipitation steps. For this, 2-2.5 g *S. acidocaldarius* L7Ae-CFHA or MW001 (control) cells were resuspended in purification buffer (100 mM MOPS pH 6.5, 300 mM NaCl, 10 % (v/v) glycerol, 5 ml/g cells) and lysed by French Pressure (three times at 25,000 psi). After centrifugation of the cell lysate (30,000×*g*, 20 min, 4°C), 100 µl of washed 50 % Flag-Ab slurry were added to 10 ml of the filtered supernatant (Millex® syringe filter, pore size 20 µm). The mixture was incubated for 1 h at 4°C on a rotary shaker. After centrifugation (3,000×*g*, 2 min, 4°C), the supernatant was carefully removed without disturbing the pelleted resin. 1 ml of the supernatant was used to resuspend the pellet and to transfer the resin to a SigmaPrep™ spin column. The resin was washed three times (3,000×*g*, 2 min, 4°C) with 500 µl RIPA+PI buffer (50 mM Tris pH 8.0, 150 mM NaCl, 1 % (v/v) Nonidet P-40, 0.5 % (w/v) sodium deoxycholate, 0.1 % (w/v) SDS, 1 % (w/v) bacterial protease inhibitor). Afterwards, the L7Ae-CFHA protein was eluted from the resin by two subsequent elutions with 250 µl 3x Flag peptide in TBS (0.015 % (w/v) 3x Flag peptide, 50 mM Tris-HCl pH 7.6, 150 mM NaCl) and 15 min incubation at 4°C on a rotary shaker. 75 µl of washed 50 % HA slurry were added to the pooled eluates and the mixture was incubated for 30 min at 4°C on a rotary shaker. After three washing steps (3,000×*g*, 2 min, 4°C) with 500 µl

RIPA+PI buffer, the L7Ae protein was eluted from the HA resin by the addition of 75 µl 8 M urea and incubation for 15 min at RT (shaking). This eluate contained co-immunoprecipitated RNAs of L7Ae. All steps were performed on ice or at 4°C to avoid degradation of the RNA. The column was furthermore eluted with 36 µl SDS loading buffer (75 mM Tris-HCl pH 6.8, 0.6 % (w/v) SDS, 0.001 % (w/v) bromophenol blue, 15 % (v/v) glycerol, 1 M β-mercaptoethanol) for 15 min at RT (shaking) to elute the L7Ae protein and interacting proteins. The L7Ae-interacting RNAs of the urea eluate were afterwards separated by denaturing PAGE and purified via gel extraction (chapters 4.6.6 and 4.6.7). The SDS eluate was analyzed by SDS-PAGE (chapter 4.7.3)

4.6.4 Co-immunoprecipitation of LSm-interacting RNAs

Two subsequent purification steps were conducted to isolate the interacting RNAs of the LSm proteins: i) Ni-NTA purification and ii) immunoprecipitation. For this, 2-2.5 g *S. acidocaldarius* LSm1/2/3-CHis or MW001 (control) cells were resuspended in purification buffer (100 mM HEPES pH8 for LSm1, 100 mM MOPS pH 6.5 for LSm2/3, 100 mM NaCl, 10 % (v/v) glycerol, 10 mM imidazole, 5 ml/g cells) and lysed by French Pressure (three times at 25,000 psi). After centrifugation of the cell lysate (30,000×*g*, 20 min, 4°C), the filtered supernatant (Millex® syringe filter, pore size 20 µm) was cycled onto a 1 ml HisTrap HP Ni-NTA column for 45 min at 4°C. The loaded column was afterwards washed with wash buffer (purification buffer containing 25 mM imidazole) using an FPLC system. The LSm proteins were eluted from the column by a 500 mM imidazole gradient using elution buffer (purification buffer containing 500 mM imidazole). The four fractions that displayed the highest purity and protein amount in the SDS-PAGE analysis (chapter 4.7.3) were pooled, 30 µg of THE™ His-tag IgG antibody (from mouse) were added, the mixture was divided into two 2 ml tubes and incubated for 1 h at 4°C on a rotary shaker. 100 µl of washed Protein G Dynabeads™ were added to each of the tubes and the mixture was incubated for 20 min at RT. Afterwards, the magnetic beads, which were bound by the LSm-CHis-anti-His-Ab complex, were separated using a magnetic rack. The beads were washed three times with 1 ml His-IP wash buffer (purification buffer containing 250 mM imidazole). Afterwards, the beads of each tube were resuspended in 1 ml bead wash buffer and pooled into one tube. The beads were again separated using the magnetic stand, the supernatant was removed and the LSm bound RNAs were eluted with 100 µl 8 M urea by incubation for 15 min at RT (shaking at 1400 rpm in a thermo mixer). All steps were performed on ice or at 4°C to avoid degradation of the RNA, unless otherwise stated. In order to elute proteins, a second elution was performed with 64 µl SDS loading buffer (75 mM Tris-HCl pH

6.8, 0.6 % (w/v) SDS, 0.001 % (w/v) bromophenol blue, 15 % (v/v) glycerol, 1 M β-mercaptoethanol) by incubation for 5 min at 95°C (shaking at 1400 rpm in a thermo mixer). The LSm interacting RNAs of the urea eluate were afterwards separated by denaturing PAGE and purified via gel extraction (chapters 4.6.6 and 4.6.7). The SDS eluate was analyzed by SDS-PAGE (chapter 4.7.3)

4.6.5 Quantitation of RNA

4.6.5.1 Spectrophotometric quantitation and quality control

Quantitation and quality control of high-yield RNA preparations was performed using a spectrophotometer as described for DNA (chapter 4.4.2.1).

4.6.5.2 Fluorometric quantitation

Low-yield RNA preparations were quantified fluorometrically using the Qubit (see also chapter 4.4.2.2). Quantification was performed using the Qubit RNA HS Assay Kit according to the manufacturer´s instructions.

4.6.6 Denaturing polyacrylamide gel electrophoresis (denaturing PAGE)

RNA preparations were separated under denaturing conditions by polyacrylamide gel electrophoresis as described for DNA (chapter 4.4.4.3). The Low Range ssRNA ladder was used as a size standard. The gels were stained for 5 min using toluidine blue (0.1 % (w/v) toluidine blue O, 1 % (v/v) acetic acid, 40 % (v/v) methanol) for high-yield preparations or 1x SybrGold in TBE buffer for low-yield preparations. Toluidine blue stained gels were afterwards destained with ddH_2O until distinct RNA bands became visible. The RNA of SybrGold stained polyacrylamide gels was visualized at 254 nm using an UV transilluminator.

4.6.7 Gel extraction from denaturing PAGE

RNA was extracted from denaturing polyacrylamide gels as described for DNA (chapter 4.4.5.3).

4.6.8 RNA fragmentation by $ZnCl_2$ treatment and curation of RNA termini

Aliquots of L7Ae and LSm co-immunoprecipitated RNAs were fragmented by $ZnCl_2$ treatment (see also chapter 2.2.1) as described by Dominissini *et al.* to allow for Illumina sequencing of longer RNAs (>100 nt) [148]. To this end, 16 µl of the L7Ae and LSm immunoprecipitation urea eluates were mixed with 1.8 µl of 10x $ZnCl_2$ fragmentation buffer (100 mM Tris-HCl pH

6.8, 100 mM $ZnCl_2$), incubated for 150 s at 94°C and the fragmentation immediately stopped on ice by the addition of 2 µl 500 mM Na_2-EDTA pH 8.0. The fragmented RNA was subsequently separated by denaturing PAGE and gel extracted (chapter 4.6.6 and 4.6.7). Bivalent metal ion fragmentation is presumed to result in 2',3'-cyclic phosphate and 5'-OH RNA termini [149]. Therefore, the termini were cured by a dephosphorylation (2',3'-cyclic phosphate curation) and a phosphorylation (5'-OH curation) procedure using T4 PNK. For this, 15 µl of the fragmented RNA were mixed with 6 µl of 5x dephosphorylation buffer (500 mM Tris-HCl pH 6.5, 500 MgAc, 25 mM β-mercaptoethanol) and 1 µl of T4 PNK (Ambion) in a total volume of 30 µl and incubated for 6 h at 37°C. Subsequently, 1 mM ATP and 1 µl T4 PNK (Ambion) were added and incubated for 1 h at 37°C. The RNA was afterwards purified by EtOH precipitation using 1 % (v/v) glycogen.

4.6.9 Preparation of cDNA libraries for Illumina sequencing

The isolated small RNAs of *S. acidocaldarius* (small RNA-Seq) and the co-immunoprecipitated RNAs of L7Ae and the LSm proteins (RIP-Seq) were used to prepare cDNA libraries for Illumina sequencing. The library preparation was performed using the NEBNext® Multiplex Small RNA Library Prep Set for Illumina according to the instructions of the manufacturer. Briefly, 5' and 3' RNA adapters were ligated to the isolated RNA (100 ng RNA input material or the maximum volume of 6.5 µl for low-yield RNA preparations, e.g. mock purifications) and cDNA was produced by a reverse transcriptase, initiated from a primer, which hybridized to the ligated 3' RNA adapter. Next, the produced cDNA was amplified by PCR (15 cycles and 20 s elongation) utilizing bar-coded primers, which allowed to distinguish the different libraries after sequencing. The amplified cDNA libraries were afterwards separated by native PAGE and size selected by gel extraction (excision from 120-270 bp and 170-270 bp for RIP-Seq/narrow-range sRNome libraries and wide-range sRNome libraries, respectively). After gel elution, the concentration of the different libraries was measured fluorometrically (chapter 4.4.2.2) and the libraries were pooled equimolarly using a concentration of 2 nM cDNA in 10 mM Tris pH 8.5 and 0.1 % Tween-20. The produced libraries are listed in Table 4.23 on page 107.

4.6.10 Illumina HiSeq2500 sequencing

The produced cDNA libraries (chapter 4.6.9) were sequenced by the Illumina HiSeq2500 method at the Max Planck Genome Centre (Köln). In this method, the cDNA library mixture attaches to complementary oligonucleotides, which are anchored on the surface of an Illumina flow cell. A so-called bridge PCR is performed, which generates clusters of amplified DNA

clones (Fig. 4.1a). Next, primers, DNA polymerase and fluorescently labeled dNTPs are added to initiate the synthesis of the complementary strand. The competitive incorporation of the dNTPs results in the emission of a specific fluorescence signal at the first nucleotide to be sequenced (sequencing by synthesis), which is recorded by a detector (Fig. 4.1b). The labeled dNTPs are reversible terminators, which block the further sequencing reaction. After cleavage of the terminator component, fresh fluorescently labeled dNTPs are added and the process is repeated, until the whole cDNA is sequenced. The HiSeq2500 methodology allowed sequencing of 100 nt read lengths. The processing and mapping of the obtained sequencing reads is described in chapter 4.9.1.

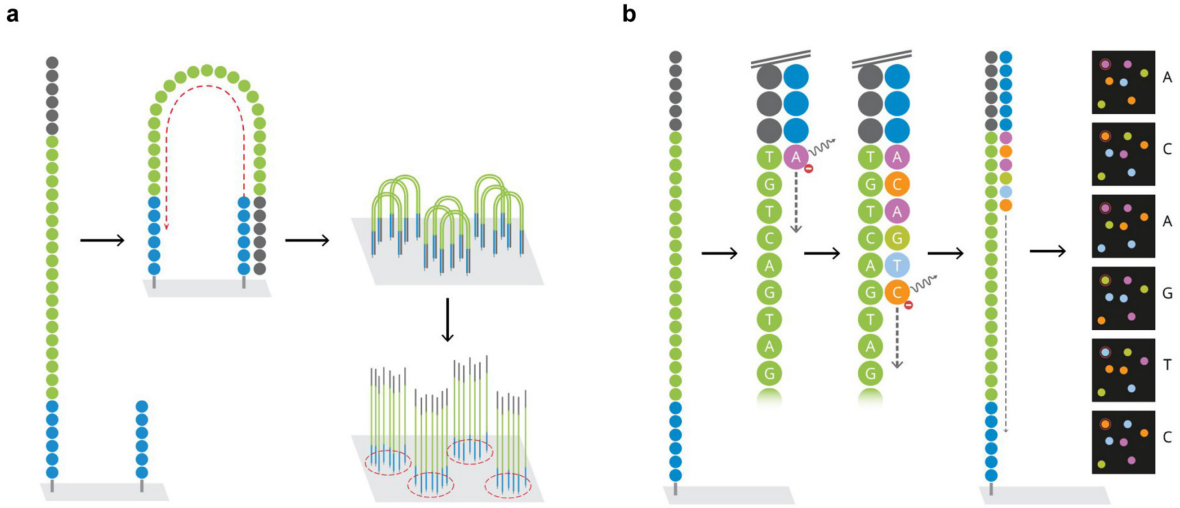


Figure 4.1 The Illumina sequencing method. a) The generation of clonal DNA clusters by bridge PCR is illustrated. b) The sequencing-by-synthesis reaction is shown, in which fluorescently labeled dNTPs are incorporated (modified from www.napoleome.ch).

4.6.11 Run-off *in vitro* transcription of non-labeled and radioactively labeled RNA

RNA transcripts were generated by run-off *in vitro* transcription using PCR products as templates (primers and templates are listed in tables 4.11/17/18). For non-labeled RNA *in vitro* transcriptions, a 50 µl Phusion PCR reaction was mixed with 40 mM HEPES/KOH pH 8.0, 22 mM $MgCl_2$, 5 mM dithiothreitol (DTT), 1 mM spermidine, 4 mM of each NTP, 400 U murine RNase inhibitor and 30 nM T7 RNA polymerase in a total volume of 1 ml. The reaction was incubated for 3 h at 37°C. For radioactively labeled RNA transcripts, 8 µl of a Phusion PCR reaction and 5 pmol [α-32P]-ATP (25 µCi) were mixed in a 20 µl *in vitro* transcription reaction. The reaction was incubated for 1 h at 37°C. The *in vitro* transcripts were afterwards purified via denaturing PAGE and gel extraction using 1 % (v/v) glycogen (chapters 4.6.6 and 4.6.7). To visualize the radiolabeled RNA after denaturing PAGE, the gels were put into plastic bags and exposed to phosphor screens (15 min), which were subsequently analyzed by a

phosphorimager. The radioactivity of gel eluted RNA transcripts was determined using a scintillation counter. The concentration of the purified RNAs was determined fluorometrically (chapter 4.6.5.2)

4.6.12 Northern blot analysis

RNA was separated using denaturing PAGE and transferred onto a positively charged nylon membrane using a semi-dry electrophoretic transfer system for 2 h at 20 V. As a preparation for the transfer, the gel, membrane and 12 Whatman papers (size of the gel) were equilibrated in TBE buffer for 5 min. The blot was assembled in the following order: 6 Whatman papers, nylon membrane, RNA-containing polyacrylamide gel and 6 Whatman papers. After the blotting, the RNA was immobilized on the membrane by UV crosslinking. The membrane was incubated for 30 min at 42°C in 8 ml (1 ml/10 cm^2 membrane) pre-hybridization buffer (Ultrahyb®-Oligo hybridization buffer). Afterwards, the 5'-terminal radiolabelled Saci11probe was heated to 95°C for 5 min and added to the hybridization buffer (10^6 cpm/ml hybridization buffer). Hybridization was performed overnight at 42°C. The membrane was washed with 15 ml low stringency buffer (2x SSC, 0.1 % SDS) and 15 ml high stringency buffer (1x SSC, 0.1 % SDS) for 30 min at 42°C to remove the unbound probe. Afterwards, the membrane was put into a plastic bag and exposed overnight to a phosphor screen. Visualization was performed using a phosphorimager.

4.7 Biochemical methods

4.7.1 Heterologous production of *S. acidocaldarius* L7Ae in *E. coli*

The recombinant pEC-A-HiSumo + Saci_1520 plasmid (Table 4.9), which is based on the pET vector (Novagen), was used for the heterologous production of *S. acidocaldarius* L7Ae in *E. coli*. The utilized *E. coli* Rosetta (DE3) expression strain harbors the DE3 prophage, which contains the T7 RNA polymerase gene under the control of the *lac* promoter. Due to the leakiness of the *lac* promoter, this strain also comprises the pLysS plasmid, which encodes the T7 lysozyme. This constitutively produced protein is a natural inhibitor of the T7 polymerase and is used to minimize the expression of the gene of interest in the non-induced state. The *l7ae* gene of the pEC-A-HiSumo + Saci_1520 plasmid is under the control of the T7 promoter. A *lac* operator sequence is present directly downstream of the T7 promoter, which is bound by the LacI repressor in the non-induced state, which additionally represses basal expression by the T7 promoter. Upon addition of IPTG, an artificial inducer of the *lac* operon that binds to the LacI repressor, the T7 polymerase is produced and binds to the T7 promoter sequence to

transcribe the *l7ae* gene. The pLysS plasmid furthermore encodes rare tRNAs (RARE2). This enhances the expression of eukaryotic and archaeal genes which contain codons that are rarely used by *E. coli*.

4.7.2 Enrichment and purification of recombinant L7Ae

2-3 g *E. coli* Rosetta (DE) cells with overproduced recombinant L7Ae were resuspended in lysis buffer (20 mM Tris-HCl pH 8.0, 10 % glycerol, 50 mM NaCl, 10 mM β-mercaptoethanol, 10 mM imidazole, 1.5 mg lysozyme/gram cells, 5 ml/g cells) and incubated for 30 min on ice. After sonication (6x 30 s, duty cycle: 40 %, output control: 4), the cell lysate was cleared by centrifugation (30,000×*g*, 20 min, 4°C) and the supernatant was filtered (Millex® syringe filter, pore size 20 µm). The supernatant was cycled onto a 1 ml HisTrap HP Ni-NTA column for 45 min at 4°C. The loaded column was afterwards washed with wash buffer (20 mM Tris-HCl pH 8.0, 10 % glycerol, 50 mM NaCl, 10 mM β-mercaptoethanol, 25 mM imidazole) using an FPLC system. The His-Sumo tagged L7Ae protein was eluted from the column by a 500 mM imidazole gradient using elution buffer (20 mM Tris-HCl pH 8.0, 10 % glycerol, 50 mM NaCl, 10 mM β-mercaptoethanol, 500 mM imidazole). His-Sumo-L7Ae protein containing fractions were identified via absorption at 280 nm wavelength and analyzed by SDS-PAGE (chapter 4.7.3). The respective fractions were pooled, Sumo protease (10 µg/ml) was added and the solution was dialyzed overnight in cleavage buffer (20 mM Tris-HCl pH 8.0, 10 % glycerol, 50 mM NaCl, 10 mM β-mercaptoethanol, 10 mM imidazole). A second Ni-NTA purification run was performed to separate the cleaved His-Sumo tag from the pure L7Ae protein. In the second run, the L7Ae protein eluted in the flow-through.

4.7.3 SDS-polyacrylamide gel electrophoresis (SDS-PAGE)

SDS-PAGE was used to separate protein samples by molecular weight and to determine their purity [222]. In this polyacrylamide gel electrophoresis, negatively charged sodium dodecyl sulfate (SDS) denatures proteins by attaching to the proteins. In an electric field, the denatured proteins then migrate towards the anode through the gel. The utilized polyacrylamide gels were composed of an upper stacking gel (125 mM Tris/HCl pH 6.8, 0.1 % (v/v) SDS, 0.1 % (v/v) APS und 0.001 % (v/v) TEMED) containing 4 % (v/v) polyacrylamide (acrylamide/bisacrylamide, ratio 37.5:1) and a lower separation gel (375 mM Tris/HCl pH 8.8, 0.1 % (v/v) SDS, 0.1 % (v/v) APS und 0.001 % (v/v) TEMED), which comprised 15 % (v/v) polyacrylamide. The gels were prepared by mixing the ingredients of the separation gel and pouring the mixture into the gel casting chamber until three quarters of the glass plates. The gel

was covered with ddH_2O until polymerization. Then, the stacking gel was poured on top and a 10-well comb was placed into the stacking gel. After polymerization, the gels were directly used in a Mini-PROTEAN Tetra Cell chamber filled with electrophoresis buffer (25 mM Tris/HCl pH 8.0, 0.1 % (w/v) SDS, 200 mM glycine) or stored for several days at 4°C. Protein samples were mixed with 2x SDS loading buffer (150 mM Tris-HCl pH 6.8, 1.2 % (w/v) SDS, 0.002 % (w/v) bromophenol blue, 30 % (v/v) glycerol, 2 M β-mercaptoethanol), boiled for 5 min at 95°C and loaded onto the gel. The gel run was carried out for 50 min at 200 V. After separation, the protein bands were visualized using Instant Blue.

4.7.4 Western blot analysis

Western blot analysis was performed to identify the His-tagged LSm proteins. In western blot analysis, proteins are transferred to a membrane and antibodies, which are specific for their epitope (often tags) are used to detect the proteins of interest [223]. The Ni-NTA fractions of the LSm purifications (chapter 4.6.4) were separated by SDS-PAGE and transferred to a PVDF membrane for 1 h at 20 V using a semi-dry electrophoretic transfer system. Prior to the transfer, the membrane was soaked for 20 s in 100 % methanol, then washed with ddH_2O and equilibrated in western transfer buffer (25 mM Tris pH 7.5, 200 mM glycine, 10 % methanol) for 5 min. Furthermore, 12 Whatman papers (size of the gel) and the protein containing polyacrylamide gel were equilibrated in western transfer buffer. The blot was assembled in the following order: 6 Whatman papers, PVDF membrane, SDS gel and 6 Whatman papers. After the transfer, the membrane was incubated overnight on a rocker at 4°C with blocking buffer (5 % (w/v) milk powder, 10 mM Tris pH 7.5, 150 mM NaCl, 0.1 % (v/v) Tween 20) to block unspecific binding by the antibodies. Anti-His-HRP antibody (1:10.000 dilution, HRP-linked antibody, from rabbit) was added to fresh blocking buffer and incubated with the membrane for 2 h at RT on a rocker. The membrane was washed five times with TBST (10 mM Tris pH 7.5, 150 mM NaCl, 0.1 % (v/v) Tween 20). Afterwards, the membrane was put into a plastic bag and the chemiluminescence reaction, catalyzed by the horseradish peroxidase (HRP) of the anti-His antibody, was started using the Roti®-Lumin plus reagent according to the manufacturer´s instructions. X-Ray films were exposed for 30 s to 60 min, depending on the signal strength. The films were developed in a film developer.

4.7.5 Bradford protein quantitation method

Protein concentrations were determined by the Bradford protein quantitation method using the Bio-Rad protein assay according to the instructions of the manufacturer [224]. The extinction

of the proteins was measured photometrically at a wavelength of 595 nm. Protein concentrations were measured by using a calibration curve of bovine serum albumin (BSA). BSA concentrations of 2, 4, 6, 8 and 10 mg/ml were used to create the curve.

4.7.6 Electrophoretic mobility shift assay (EMSA)

Electrophoretic mobility shift assays were performed to analyze the RNA binding activity of the L7Ae protein. In this method, native PAGE demonstrates a retardation of protein-bound nucleic acids compared to free nucleic acids. The assays were performed using around 30-50 nM radiolabeled *in vitro* RNA transcript (15,000 cpm) that were mixed with increasing concentrations of L7Ae in a volume of 10 µl EMSA binding buffer (20 mM Tris-HCl pH 8.0, 50 mM NaCl, 10 mM β-mercaptoethanol, 1 mM $MgCl_2$). Competition of the Nat RNA binding (Fig. 2.17) was performed by the addition of 10-fold unlabeled total RNA or 100-fold unlabeled C/D box sRNA Sac-sR121. The reactions were incubated for 10 min at 70°C and then separated by 10 % native PAGE as described for DNA (chapter 4.4.4.2). Afterwards, the gel was put into a plastic bag and exposed overnight to a phosphor screen. Visualization was performed using a phosphorimager.

4.7.7 Mass spectrometry

L7Ae/LSm proteins and their interaction partners that were obtained during co-immunoprecipitation experiments (chapters 4.6.3 and 4.6.4) were identified using mass spectrometry and were performed in collaboration with Jörg Kahnt at the Max-Planck Institute for terrestrial microbiology in Marburg.

4.7.8 Reporter assays in *S. acidocaldarius* using β-galactosidase

MW001 and Sac-sR10KO cultures, which were transformed with the β-galactosidase reporter plasmids (Table 4.8), were harvested at logarithmic (OD_{600nm} 0.5-0.6), early stationary (OD_{600nm} 1.4-1.6) and late stationary (OD_{600nm} 1.6) phase, respectively. The activity of the β-galactosidase enzyme was determined as described by Wagner *et al.* [139]. Two ml of each culture were pelleted and the cells resuspended in Z-buffer (10 mM KCl, 1 mM $MgSO_4$, 60 mM Na_2HPO_4, 40 mM NaH_2PO_4, pH 7.0), supplemented with 0.5 % (v/v) Triton-X and 1 mM PMSF, to an OD_{600nm} of 3.2. The cell suspension was lysed on a vortex and 20 µl were mixed with 170 µl Z-buffer and 10 µl ONPG solution (12 mg ONPG in 1 ml Z-buffer) and incubated at 42°C. The hydrolysis of ONPG by the β-galactosidase was measured at 410 nm for 3 h in 5 min intervals in a plate reader (Infinite M200 Pro, TECAN). Furthermore, the protein

concentration of the cell lysate was measured by the Bradford protein quantitation method. The activity of the β-galactosidase was quantified by the following equation [225, 226]:

$$Miller = \frac{60{,}000 \times \left(\text{A410 (t2} - \text{t1)} - \text{autolysis A410 (t2} - \text{t1)}\right) \times 7}{\text{time (s)} \times \text{volume of sample (ml)} \times \text{concentration of protein (mg/ml)}}$$

4.8 Cell biological methods

4.8.1 Flow cytometry

4.8.1.1 GFP reporter system studies in *E. coli*

E. coli Rosetta transformants that contained the GFP reporter system plasmids (Table 4.10) were grown in LB medium, adjusted to OD_{600nm} 0.6 and split into two cultures. One culture was induced with 1 mM IPTG. Two ml of both IPTG-induced and non-induced cultures were transferred to 24 well plates and incubated for 4 h at 37°C (shaking at 200 rpm). Afterwards, 1:300 dilutions of the cultures were prepared in 96 well plates in a total volume of 300 µl using PBS (10 mM Na_2HPO_4, 1.8 mM KH_2PO_4, 2.7 mM KCl, 137 mM NaCl, pH 7.4). 10,000 cells of each strain were recorded in the flow cytometer. The GFP fluorescence was excited by a laser at 488 nm. The data analysis was performed using FlowJo 10.1 (FlowJo LLC, Ashland, OR, USA). Data processing was performed by gating of cell populations i) to remove doublets, ii) to remove cell debris and iii) to exclude GFP-negative cells (iii, only for the toxic strains of Fig. 2.10, which do not contain the autoregulated L7Ae). The median value was used to calculate the GFP signal. All presented GFP ratios were calculated by dividing the GFP value of induced cells by the value of non-induced cells, which reduced variations between replicates. Furthermore, the GFP values included the ratio of IPTG-induced L7Ae/no L7Ae (frameshifted *l7ae* gene) strains. The strains containing the frameshifted *l7ae* gene constituted the proper control strains, as the non-induced strains displayed a different growth behavior than the induced strains (Fig. 2.11a,b). This allowed to measure L7Ae-induced signal reduction at similar growth conditions and to compare strains with different *gfp* expression strength, e.g. *H. volcanii* (GFP_{med}: 1133 a.u.) vs. *M. maripaludis* (GFP_{med}: 36354 a.u.) *l7ae* 5' UTR.

4.8.1.2 Analysis of *S. acidocaldarius lsm* mutant strains

Precultures of *S. acidocaldarius* mutant strains (Flag-HA/His tagged LSm strains and the LSm3KO strain) were grown for one day. The LSm1-CFHA and LSm2-NFHA strains were grown for two days, due to their delayed growth. Afterwards, the cultures were adjusted to

OD_{600nm} 0.1 (0.2 for LSm1-CFHA and LSm2-NFHA) and grown for one day until OD_{600nm} 0.8-1.0 (logarithmic phase). The cells were diluted 1:100 in PBS and analyzed by flow cytometry. The data analysis was performed using FlowJo 10.1 (FlowJo LLC, Ashland, OR, USA). Data processing was performed by gating of cell populations to remove doublets and to remove cell debris.

4.8.2 Microscopy

The same *lsm* mutant cultures which were analyzed by flow cytometry (chapter 4.8.1.2) were analyzed by light microscopy. 3 µl of the cultures were dripped on 2 % (w/v) agarose pads to immobilize the cells. Microscopy was performed using a 100x magnification (Plan-Apochromat 100x lens, Zeiss).

4.9 Bioinformatic methods

4.9.1 Processing and mapping of Illumina sequencing data

The CLC Genomics Workbench 9.5.3 software (Qiagen, Hilden) was used for processing and mapping of the obtained Illumina sequencing data sets. The data were processed by i) removal of sequences of low quality (quality score limit: 0.05; max. ambiguities: 2), ii) trimming of adapter sequences and iii) filtering by length (15 nt cut-off). The trimmed sequences were mapped to the *S. acidocaldarius* reference genome (accession no.: CP000077) by using default settings. The sRNome and L7Ae RIP-Seq data sets are available at Gene Expression Omnibus (GSE94748). Table 4.23 lists the mapping reports for the sequenced libraries.

Table 4.23 Mapping report of the sequenced cDNA libraries.

Name of library	Selected size	Number of sequenced reads	Number of mapped reads [%]
sRNome A	50-150 nt	12,114,316	12,070,084 [99.63]
sRNome B	50-150 nt	9,305,820	9,289,118 [99.82]
sRNome C	0-150 nt	11,909,827	11,470,982 [96.32]
sRNome D	0-150 nt	13,692,315	13,413,213 [97.96]
L7Ae sRNA1	0-150 nt	6,263,534	6,157,967 [98.31]
L7Ae sRNA2	0-150 nt	4,890,561	4,783,651 [97.81]
L7Ae frag RNA1	0-150 nt	7,470,228	6,632,643 [88.79]
L7Ae frag RNA2	0-150 nt	6,684,031	5,769,310 [86.31]
L7Ae WT control	0-150 nt	16,261,296	8,240,779 [50.68]
LSm1 sRNA1	0-150 nt	13,820,152	12,881,510 [93.21]
LSm1 sRNA2	0-150 nt	11,890,842	10,644,014 [89.51]
LSm1 frag RNA1	0-150 nt	4,086,444	3,614,701 [88.46]
LSm1 frag RNA2	0-150 nt	3,530,928	3,115,768 [88.24]
LSm1 WT control	0-150 nt	3,069,439	271,743 [8.85]
LSm2 sRNA1	0-150 nt	3,753,744	3,576,239 [95.27]
LSm2 sRNA2	0-150 nt	3,356,676	3,240,360 [96.53]

LSm2 frag RNA1	0-150 nt	3,770,105	3,443,990 [91.35]
LSm2 frag RNA2	0-150 nt	3,556,401	3,239,290 [90.83]
LSm2 WT control	0-150 nt	3,092,968	476,400 [15.40]
LSm3 sRNA1	0-150 nt	6,549,906	526,647 [8.04]
LSm3 sRNA2	0-150 nt	8,306,581	824,133 [9.92]
LSm3 frag RNA1	0-150 nt	2,417,059	1,175,677 [48.64]
LSm3 frag RNA2	0-150 nt	1,788,301	774,421 [43.30]
LSm3 WT control	0-150 nt	4,219,418	119,033 [2.82]

4.9.2 Identification of the sRNome of *S. acidocaldarius*

Small RNAs of *S. acidocaldarius* were identified by screening the four sRNome libraries (two biological replicates for narrow-range (50-150 nt) and wide-range (0-150 nt) selected data sets) for RNAs covered by at least 100 reads within one million mapped reads per library. The identified RNAs are listed in Appendix 1.

4.9.3 Identification of the L7Ae and LSm interacting RNAs by the DESeq2 tool

The identification of the L7Ae and LSm interacting RNAs was performed by Michael Uhl. A custom peak calling pipeline was applied to the mapped reads of the L7Ae and LSm RIP-Seq data sets as described by Holmqvist *et al.* [227]. First, the pipeline defined potential binding regions from the read data by utilizing the Blockbuster tool [228]. Next, the statistical significance of these regions in terms of their differential levels of abundance was assessed using the DESeq2 tool by comparing two replicates of each L7Ae and LSm RIP-Seq data set (sRNA and frag RNA data sets) with a control data set of untagged L7Ae/LSm (one WT control data set for each analyzed protein) [150]. The input to DESeq2 was composed of count tables that contained the number of reads counted for each identified binding region in the experiments as well as in the control libraries. On the basis of these counts, DESeq2 determined size factors for each library that were then used to calculate normalized read counts by dividing each read count by the corresponding library size factor. The determined normalized read counts were then used for comparisons between libraries, specifically for differential expression testing, and to calculate fold change and false-discovery-rate (q) values as described by Holmqvist *et al.* [227]. Identified L7Ae-interacting RNAs with a q-value of < 0.1 were further filtered by their normalized read count (at least 3,000 reads) to remove false positive hits due to low-abundance RNAs (the identified RNAs are listed in Appendix 2). The analyzed subsets of the identified LSm-interacting RNAs comprised the 100 most significant peak calls (q-values < 1.13e-009, 9.89e-011, 1.8e-006 and 2.64e-009 for LSm1 sRNA/frag and LSm2 sRNA/frag RNA, respectively).

4.9.4 Identification of potential LSm binding motifs using the MEME tool

Potential LSm binding motifs were identified by Michael Uhl using the MEME tool. MEME (Multiple Expectation-Maximation for Motif Elicitation) utilizes multiple sequence alignment to discover statistically significant motifs in the input sequences [163, 164]. As an input, the LSm1 and LSm2 peak calling regions that were identified by the DESeq2 analysis were used. In total, four MEME analyses were performed (for LSm1 sRNA/frag and LSm2 sRNA/frag RNA). Default MEME settings were utilized. The tool searched for 15 motifs containing lengths between 5-25 nt.

4.9.5 Statistics

Two-tailed, unpaired Student's t-tests were performed to calculate the significance of the data using a p-value of 0.05. Two-tailed, unpaired t-tests with Welch correction were applied in case of unequal variances.

5. References

1. **Farlow WG. 1880.** On The Nature Of The Peculiar Reddening Of Salted Codfish During The Summer Season. US Comm Fish and Fisheries, Report for 1878:969-973.
2. **Woese CR, Fox GE. 1977.** Phylogenetic structure of the prokaryotic domain: the primary kingdoms. Proc Natl Acad Sci U S A 74:5088-90.
3. **Cavicchioli R. 2011.** Archaea--timeline of the third domain. Nat Rev Microbiol 9:51-61.
4. **Woese CR, Kandler O, Wheelis ML. 1990.** Towards a natural system of organisms: proposal for the domains Archaea, Bacteria, and Eucarya. Proc Natl Acad Sci U S A 87:4576-9.
5. **Barns SM, Delwiche CF, Palmer JD, Pace NR. 1996.** Perspectives on archaeal diversity, thermophily and monophyly from environmental rRNA sequences. Proc Natl Acad Sci U S A 93:9188-93.
6. **Brochier-Armanet C, Boussau B, Gribaldo S, Forterre P. 2008.** Mesophilic Crenarchaeota: proposal for a third archaeal phylum, the Thaumarchaeota. Nat Rev Microbiol 6:245-52.
7. **Nunoura T, Takaki Y, Kakuta J, Nishi S, Sugahara J, Kazama H, Chee GJ, Hattori M, Kanai A, Atomi H, Takai K, Takami H. 2011.** Insights into the evolution of Archaea and eukaryotic protein modifier systems revealed by the genome of a novel archaeal group. Nucleic Acids Res 39:3204-23.
8. **Meng J, Xu J, Qin D, He Y, Xiao X, Wang F. 2014.** Genetic and functional properties of uncultivated MCG archaea assessed by metagenome and gene expression analyses. ISME J 8:650-9.
9. **Castelle CJ, Wrighton KC, Thomas BC, Hug LA, Brown CT, Wilkins MJ, Frischkorn KR, Tringe SG, Singh A, Markillie LM, Taylor RC, Williams KH, Banfield JF. 2015.** Genomic expansion of domain archaea highlights roles for organisms from new phyla in anaerobic carbon cycling. Curr Biol 25:690-701.
10. **Huber H, Hohn MJ, Rachel R, Fuchs T, Wimmer VC, Stetter KO. 2002.** A new phylum of Archaea represented by a nanosized hyperthermophilic symbiont. Nature 417:63-7.
11. **Eme L, Doolittle WF. 2015.** Archaea. Curr Biol 25:R851-5.
12. **Spang A, Saw JH, Jorgensen SL, Zaremba-Niedzwiedzka K, Martijn J, Lind AE, van Eijk R, Schleper C, Guy L, Ettema TJ. 2015.** Complex archaea that bridge the gap between prokaryotes and eukaryotes. Nature 521:173-9.
13. **Jorgensen SL, Hannisdal B, Lanzen A, Baumberger T, Flesland K, Fonseca R, Ovreas L, Steen IH, Thorseth IH, Pedersen RB, Schleper C. 2012.** Correlating microbial community profiles with geochemical data in highly stratified sediments from the Arctic Mid-Ocean Ridge. Proc Natl Acad Sci U S A 109:E2846-55.
14. **Guy L, Ettema TJ. 2011.** The archaeal 'TACK' superphylum and the origin of eukaryotes. Trends Microbiol 19:580-7.
15. **Hug LA, Baker BJ, Anantharaman K, Brown CT, Probst AJ, Castelle CJ, Butterfield CN, Hernsdorf AW, Amano Y, Ise K, Suzuki Y, Dudek N, Relman DA,**

Finstad KM, Amundson R, Thomas BC, Banfield JF. 2016. A new view of the tree of life. Nat Microbiol 1:16048.

16. **Altman S. 1990.** Nobel lecture. Enzymatic cleavage of RNA by RNA. Biosci Rep 10:317-37.

17. **Walter P, Blobel G. 1982.** Signal recognition particle contains a 7S RNA essential for protein translocation across the endoplasmic reticulum. Nature 299:691-8.

18. **Randau L, Schroder I, Soll D. 2008.** Life without RNase P. Nature 453:120-3.

19. **Lechner M, Rossmanith W, Hartmann RK, Tholken C, Gutmann B, Giege P, Gobert A. 2015.** Distribution of Ribonucleoprotein and Protein-Only RNase P in Eukarya. Mol Biol Evol 32:3186-93.

20. **Rosenblad MA, Samuelsson T. 2004.** Identification of chloroplast signal recognition particle RNA genes. Plant Cell Physiol 45:1633-9.

21. **Omer AD, Lowe TM, Russell AG, Ebhardt H, Eddy SR, Dennis PP. 2000.** Homologs of small nucleolar RNAs in Archaea. Science 288:517-22.

22. **Gaspin C, Cavaille J, Erauso G, Bachellerie JP. 2000.** Archaeal homologs of eukaryotic methylation guide small nucleolar RNAs: lessons from the Pyrococcus genomes. J Mol Biol 297:895-906.

23. **Watkins NJ, Bohnsack MT. 2012.** The box C/D and H/ACA snoRNPs: key players in the modification, processing and the dynamic folding of ribosomal RNA. Wiley Interdiscip Rev RNA 3:397-414.

24. **Kiss-Laszlo Z, Henry Y, Bachellerie JP, Caizergues-Ferrer M, Kiss T. 1996.** Site-specific ribose methylation of preribosomal RNA: a novel function for small nucleolar RNAs. Cell 85:1077-88.

25. **Omer AD, Ziesche S, Ebhardt H, Dennis PP. 2002.** In vitro reconstitution and activity of a C/D box methylation guide ribonucleoprotein complex. Proc Natl Acad Sci U S A 99:5289-94.

26. **Dennis PP, Omer A. 2005.** Small non-coding RNAs in Archaea. Curr Opin Microbiol 8:685-94.

27. **Kiss-Laszlo Z, Henry Y, Kiss T. 1998.** Sequence and structural elements of methylation guide snoRNAs essential for site-specific ribose methylation of pre-rRNA. EMBO J 17:797-807.

28. **Klein DJ, Schmeing TM, Moore PB, Steitz TA. 2001.** The kink-turn: a new RNA secondary structure motif. EMBO J 20:4214-21.

29. **Moore T, Zhang Y, Fenley MO, Li H. 2004.** Molecular basis of box C/D RNA-protein interactions; cocrystal structure of archaeal L7Ae and a box C/D RNA. Structure 12:807-18.

30. **Nolivos S, Carpousis AJ, Clouet-d'Orval B. 2005.** The K-loop, a general feature of the Pyrococcus C/D guide RNAs, is an RNA structural motif related to the K-turn. Nucleic Acids Res 33:6507-14.

31. **Tran E, Zhang X, Lackey L, Maxwell ES. 2005.** Conserved spacing between the box C/D and C'/D' RNPs of the archaeal box C/D sRNP complex is required for efficient 2'-O-methylation of target RNAs. RNA 11:285-93.

32. **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 30:931-41.

33. **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 31:869-77.

34. **Tran EJ, Zhang X, Maxwell ES. 2003.** Efficient RNA 2'-O-methylation requires juxtaposed and symmetrically assembled archaeal box C/D and C'/D' RNPs. EMBO J 22:3930-40.

35. **Bower-Phipps KR, Taylor DW, Wang HW, Baserga SJ. 2012.** The box C/D sRNP dimeric architecture is conserved across domain Archaea. RNA 18:1527-40.

36. **Lapinaite A, Simon B, Skjaerven L, Rakwalska-Bange M, Gabel F, Carlomagno T. 2013.** The structure of the box C/D enzyme reveals regulation of RNA methylation. Nature 502:519-23.

37. **Aittaleb M, Rashid R, Chen Q, Palmer JR, Daniels CJ, Li H. 2003.** Structure and function of archaeal box C/D sRNP core proteins. Nat Struct Biol 10:256-63.

38. **Reichow SL, Hamma T, Ferre-D'Amare AR, Varani G. 2007.** The structure and function of small nucleolar ribonucleoproteins. Nucleic Acids Res 35:1452-64.

39. **Wang H, Boisvert D, Kim KK, Kim R, Kim SH. 2000.** Crystal structure of a fibrillarin homologue from Methanococcus jannaschii, a hyperthermophile, at 1.6 A resolution. EMBO J 19:317-23.

40. **Dennis PP, Tripp V, Lui L, Lowe T, Randau L. 2015.** C/D box sRNA-guided 2'-O-methylation patterns of archaeal rRNA molecules. BMC Genomics 16:632.

41. **Ghalei H, Hsiao HH, Urlaub H, Wahl MC, Watkins NJ. 2010.** A novel Nop5-sRNA interaction that is required for efficient archaeal box C/D sRNP formation. RNA 16:2341-8.

42. **Yip WS, Vincent NG, Baserga SJ. 2013.** Ribonucleoproteins in archaeal pre-rRNA processing and modification. Archaea 2013:614735.

43. **Noon KR, Bruenger E, McCloskey JA. 1998.** Posttranscriptional modifications in 16S and 23S rRNAs of the archaeal hyperthermophile Sulfolobus solfataricus. J Bacteriol 180:2883-8.

44. **Ni J, Tien AL, Fournier MJ. 1997.** Small nucleolar RNAs direct site-specific synthesis of pseudouridine in ribosomal RNA. Cell 89:565-73.

45. **Ganot P, Bortolin ML, Kiss T. 1997.** Site-specific pseudouridine formation in preribosomal RNA is guided by small nucleolar RNAs. Cell 89:799-809.

46. **Baker DL, Youssef OA, Chastkofsky MI, Dy DA, Terns RM, Terns MP. 2005.** RNA-guided RNA modification: functional organization of the archaeal H/ACA RNP. Genes Dev 19:1238-48.

47. **Charpentier B, Muller S, Branlant C. 2005.** Reconstitution of archaeal H/ACA small ribonucleoprotein complexes active in pseudouridylation. Nucleic Acids Res 33:3133-44.

48. **Liang B, Xue S, Terns RM, Terns MP, Li H. 2007.** Substrate RNA positioning in the archaeal H/ACA ribonucleoprotein complex. Nat Struct Mol Biol 14:1189-95.

49. **Wang P, Yang L, Gao YQ, Zhao XS. 2015.** Accurate placement of substrate RNA by Gar1 in H/ACA RNA-guided pseudouridylation. Nucleic Acids Res 43:7207-16.

50. **Liang XH, Liu Q, Fournier MJ. 2009.** Loss of rRNA modifications in the decoding center of the ribosome impairs translation and strongly delays pre-rRNA processing. RNA 15:1716-28.

51. **Kierzek E, Malgowska M, Lisowiec J, Turner DH, Gdaniec Z, Kierzek R. 2014.** The contribution of pseudouridine to stabilities and structure of RNAs. Nucleic Acids Res 42:3492-501.

52. **Karijolich J, Yu YT. 2011.** Converting nonsense codons into sense codons by targeted pseudouridylation. Nature 474:395-8.

53. **Xu N, Li Y, Zhao YT, Guo L, Fang YY, Zhao JH, Wang XJ, Huang L, Guo HS. 2012.** Identification and characterization of small RNAs in the hyperthermophilic archaeon Sulfolobus solfataricus. PLoS One 7:e35306.

54. **Wilson RC, Doudna JA. 2013.** Molecular mechanisms of RNA interference. Annu Rev Biophys 42:217-39.

55. **Shabalina SA, Koonin EV. 2008.** Origins and evolution of eukaryotic RNA interference. Trends Ecol Evol 23:578-87.

56. **Vestergaard G, Garrett RA, Shah SA. 2014.** CRISPR adaptive immune systems of Archaea. RNA Biol 11:156-67.

57. **Babski J, Maier LK, Heyer R, Jaschinski K, Prasse D, Jager D, Randau L, Schmitz RA, Marchfelder A, Soppa J. 2014.** Small regulatory RNAs in Archaea. RNA Biol 11:484-93.

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

59. **Gorski SA, Vogel J, Doudna JA. 2017.** RNA-based recognition and targeting: sowing the seeds of specificity. Nat Rev Mol Cell Biol 18:215-228.

60. **Vogel J, Luisi BF. 2011.** Hfq and its constellation of RNA. Nat Rev Microbiol 9:578-89.

61. **Jager D, Pernitzsch SR, Richter AS, Backofen R, Sharma CM, Schmitz RA. 2012.** An archaeal sRNA targeting cis- and trans-encoded mRNAs via two distinct domains. Nucleic Acids Res 40:10964-79.

62. **Slupska MM, King AG, Fitz-Gibbon S, Besemer J, Borodovsky M, Miller JH. 2001.** Leaderless transcripts of the crenarchaeal hyperthermophile Pyrobaculum aerophilum. J Mol Biol 309:347-60.

63. **Wurtzel O, Sapra R, Chen F, Zhu Y, Simmons BA, Sorek R. 2010.** A single-base resolution map of an archaeal transcriptome. Genome Res 20:133-41.

64. **Brenneis M, Soppa J. 2009.** Regulation of translation in haloarchaea: 5'- and 3'-UTRs are essential and have to functionally interact in vivo. PLoS One 4:e4484.

65. **Martens B, Manoharadas S, Hasenohrl D, Manica A, Blasi U. 2013.** Antisense regulation by transposon-derived RNAs in the hyperthermophilic archaeon Sulfolobus solfataricus. EMBO Rep 14:527-33.

66. **Orell A, Tripp V, Aliaga-Tobar V, Albers SV, Maracaja-Coutinho V, Randau L. 2017.** A double-stranded RNA is involved in biofilm regulation in *Sulfolobus acidocaldarius*. Nucleic Acids Res [in press].

67. **Mura C, Randolph PS, Patterson J, Cozen AE. 2013.** Archaeal and eukaryotic homologs of Hfq: A structural and evolutionary perspective on Sm function. RNA Biol 10:636-51.

68. **Westhof E, Auffinger P. 2006.** Encyclopedia of Analytical Chemistry: RNA Tertiary Structure. John Wiley & Sons, Ltd.

69. **Polacek N, Mankin AS. 2005.** The ribosomal peptidyl transferase center: structure, function, evolution, inhibition. Crit Rev Biochem Mol Biol 40:285-311.

70. **Huang L, Lilley DM. 2016.** The Kink Turn, a Key Architectural Element in RNA Structure. J Mol Biol 428:790-801.

71. **Lilley DM. 2012.** The structure and folding of kink turns in RNA. Wiley Interdiscip Rev RNA 3:797-805.

72. **Winkler WC, Grundy FJ, Murphy BA, Henkin TM. 2001.** The GA motif: an RNA element common to bacterial antitermination systems, rRNA, and eukaryotic RNAs. RNA 7:1165-72.

73. **Becker MM, Lapouge K, Segnitz B, Wild K, Sinning I. 2017.** Structures of human SRP72 complexes provide insights into SRP RNA remodeling and ribosome interaction. Nucleic Acids Res 45:470-481.

74. **Lilley DM. 2014.** The K-turn motif in riboswitches and other RNA species. Biochim Biophys Acta 1839:995-1004.

75. **Katahira M, Sato H, Mishima K, Uesugi S, Fujii S. 1993.** NMR studies of G:A mismatches in oligodeoxyribonucleotide duplexes modelled after ribozymes. Nucleic Acids Res 21:5418-24.

76. **Schroeder KT, Daldrop P, Lilley DM. 2011.** RNA tertiary interactions in a riboswitch stabilize the structure of a kink turn. Structure 19:1233-40.

77. **McPhee SA, Huang L, Lilley DM. 2014.** A critical base pair in k-turns that confers folding characteristics and correlates with biological function. Nat Commun 5:5127.

78. **Koonin EV, Bork P, Sander C. 1994.** A novel RNA-binding motif in omnipotent suppressors of translation termination, ribosomal proteins and a ribosome modification enzyme? Nucleic Acids Res 22:2166-7.

79. **Ban N, Nissen P, Hansen J, Moore PB, Steitz TA. 2000.** The complete atomic structure of the large ribosomal subunit at 2.4 A resolution. Science 289:905-20.

80. **Watkins NJ, Segault V, Charpentier B, Nottrott S, Fabrizio P, Bachi A, Wilm M, Rosbash M, Branlant C, Luhrmann R. 2000.** A common core RNP structure shared between the small nucleoar box C/D RNPs and the spliceosomal U4 snRNP. Cell 103:457-66.

81. **Nottrott S, Hartmuth K, Fabrizio P, Urlaub H, Vidovic I, Ficner R, Luhrmann R. 1999.** Functional interaction of a novel 15.5kD [U4/U6.U5] tri-snRNP protein with the 5' stem-loop of U4 snRNA. EMBO J 18:6119-33.

82. **Henras A, Henry Y, Bousquet-Antonelli C, Noaillac-Depeyre J, Gelugne JP, Caizergues-Ferrer M. 1998.** Nhp2p and Nop10p are essential for the function of H/ACA snoRNPs. EMBO J 17:7078-90.

83. **Rosenblad MA, Lopez MD, Piccinelli P, Samuelsson T. 2006.** Inventory and analysis of the protein subunits of the ribonucleases P and MRP provides further evidence of homology between the yeast and human enzymes. Nucleic Acids Res 34:5145-56.

84. **Baird NJ, Zhang J, Hamma T, Ferre-D'Amare AR. 2012.** YbxF and YlxQ are bacterial homologs of L7Ae and bind K-turns but not K-loops. RNA 18:759-70.

85. **Caban K, Kinzy SA, Copeland PR. 2007.** The L7Ae RNA binding motif is a multifunctional domain required for the ribosome-dependent Sec incorporation activity of Sec insertion sequence binding protein 2. Mol Cell Biol 27:6350-60.

86. **Oruganti S, Zhang Y, Li H. 2005.** Structural comparison of yeast snoRNP and spliceosomal protein Snu13p with its homologs. Biochem Biophys Res Commun 333:550-4.

87. **Koo BK, Park CJ, Fernandez CF, Chim N, Ding Y, Chanfreau G, Feigon J. 2011.** Structure of H/ACA RNP protein Nhp2p reveals cis/trans isomerization of a conserved proline at the RNA and Nop10 binding interface. J Mol Biol 411:927-42.

88. **Huang L, Lilley DM. 2013.** The molecular recognition of kink-turn structure by the L7Ae class of proteins. RNA 19:1703-10.

89. **Cho IM, Lai LB, Susanti D, Mukhopadhyay B, Gopalan V. 2010.** Ribosomal protein L7Ae is a subunit of archaeal RNase P. Proc Natl Acad Sci U S A 107:14573-8.

90. **Fukuhara H, Kifusa M, Watanabe M, Terada A, Honda T, Numata T, Kakuta Y, Kimura M. 2006.** A fifth protein subunit Ph1496p elevates the optimum temperature for the ribonuclease P activity from Pyrococcus horikoshii OT3. Biochem Biophys Res Commun 343:956-64.

91. **Lai SM, Lai LB, Foster MP, Gopalan V. 2014.** The L7Ae protein binds to two kink-turns in the Pyrococcus furiosus RNase P RNA. Nucleic Acids Res 42:13328-38.

92. **Zago MA, Dennis PP, Omer AD. 2005.** The expanding world of small RNAs in the hyperthermophilic archaeon Sulfolobus solfataricus. Mol Microbiol 55:1812-28.

93. **Saito H, Kobayashi T, Hara T, Fujita Y, Hayashi K, Furushima R, Inoue T. 2010.** Synthetic translational regulation by an L7Ae-kink-turn RNP switch. Nat Chem Biol 6:71-8.

94. **Saito H, Fujita Y, Kashida S, Hayashi K, Inoue T. 2011.** Synthetic human cell fate regulation by protein-driven RNA switches. Nat Commun 2:160.

95. **Stapleton JA, Endo K, Fujita Y, Hayashi K, Takinoue M, Saito H, Inoue T. 2012.** Feedback control of protein expression in mammalian cells by tunable synthetic translational inhibition. ACS Synth Biol 1:83-8.

96. **Wilusz CJ, Wilusz J. 2013.** Lsm proteins and Hfq: Life at the 3' end. RNA Biol 10:592-601.

97. **Weichenrieder O. 2014.** RNA binding by Hfq and ring-forming (L)Sm proteins: a trade-off between optimal sequence readout and RNA backbone conformation. RNA Biol 11:537-49.

98. **Franze de Fernandez MT, Eoyang L, August JT. 1968.** Factor fraction required for the synthesis of bacteriophage Qbeta-RNA. Nature 219:588-90.

99. **Schumacher MA, Pearson RF, Moller T, Valentin-Hansen P, Brennan RG. 2002.** Structures of the pleiotropic translational regulator Hfq and an Hfq-RNA complex: a bacterial Sm-like protein. EMBO J 21:3546-56.

100. **Dimastrogiovanni D, Frohlich KS, Bandyra KJ, Bruce HA, Hohensee S, Vogel J, Luisi BF. 2014.** Recognition of the small regulatory RNA RydC by the bacterial Hfq protein. Elife 3.

101. **Masse E, Escorcia FE, Gottesman S. 2003.** Coupled degradation of a small regulatory RNA and its mRNA targets in Escherichia coli. Genes Dev 17:2374-83.

102. **Toro I, Thore S, Mayer C, Basquin J, Seraphin B, Suck D. 2001.** RNA binding in an Sm core domain: X-ray structure and functional analysis of an archaeal Sm protein complex. EMBO J 20:2293-303.

103. **Murina VN, Nikulin AD. 2011.** RNA-binding Sm-like proteins of bacteria and archaea. similarity and difference in structure and function. Biochemistry (Mosc) 76:1434-49.

104. **Mohanty BK, Maples VF, Kushner SR. 2004.** The Sm-like protein Hfq regulates polyadenylation dependent mRNA decay in Escherichia coli. Mol Microbiol 54:905-20.

105. **Blum E, Carpousis AJ, Higgins CF. 1999.** Polyadenylation promotes degradation of 3'-structured RNA by the Escherichia coli mRNA degradosome in vitro. J Biol Chem 274:4009-16.

106. **Updegrove TB, Zhang A, Storz G. 2016.** Hfq: the flexible RNA matchmaker. Curr Opin Microbiol 30:133-8.

107. **Sauer E, Weichenrieder O. 2011.** Structural basis for RNA 3'-end recognition by Hfq. Proc Natl Acad Sci U S A 108:13065-70.

108. **Otaka H, Ishikawa H, Morita T, Aiba H. 2011.** PolyU tail of rho-independent terminator of bacterial small RNAs is essential for Hfq action. Proc Natl Acad Sci U S A 108:13059-64.

109. **Robinson KE, Orans J, Kovach AR, Link TM, Brennan RG. 2014.** Mapping Hfq-RNA interaction surfaces using tryptophan fluorescence quenching. Nucleic Acids Res 42:2736-49.

110. **Fender A, Elf J, Hampel K, Zimmermann B, Wagner EG. 2010.** RNAs actively cycle on the Sm-like protein Hfq. Genes Dev 24:2621-6.

111. **Tsui HC, Leung HC, Winkler ME. 1994.** Characterization of broadly pleiotropic phenotypes caused by an hfq insertion mutation in Escherichia coli K-12. Mol Microbiol 13:35-49.

112. **Olejniczak M, Storz G. 2017.** ProQ/FinO-domain proteins: another ubiquitous family of RNA matchmakers? Mol Microbiol 104:905-915.

113. **Luhrmann R, Kastner B, Bach M. 1990.** Structure of spliceosomal snRNPs and their role in pre-mRNA splicing. Biochim Biophys Acta 1087:265-92.

114. **Leung AK, Nagai K, Li J. 2011.** Structure of the spliceosomal U4 snRNP core domain and its implication for snRNP biogenesis. Nature 473:536-9.

115. **Mayes AE, Verdone L, Legrain P, Beggs JD. 1999.** Characterization of Sm-like proteins in yeast and their association with U6 snRNA. EMBO J 18:4321-31.

116. **Licht K, Medenbach J, Luhrmann R, Kambach C, Bindereif A. 2008.** 3'-cyclic phosphorylation of U6 snRNA leads to recruitment of recycling factor p110 through LSm proteins. RNA 14:1532-8.

117. **Tharun S, He W, Mayes AE, Lennertz P, Beggs JD, Parker R. 2000.** Yeast Sm-like proteins function in mRNA decapping and decay. Nature 404:515-8.

118. **Chowdhury A, Mukhopadhyay J, Tharun S. 2007.** The decapping activator Lsm1p-7p-Pat1p complex has the intrinsic ability to distinguish between oligoadenylated and polyadenylated RNAs. RNA 13:998-1016.

119. **Salgado-Garrido J, Bragado-Nilsson E, Kandels-Lewis S, Seraphin B. 1999.** Sm and Sm-like proteins assemble in two related complexes of deep evolutionary origin. EMBO J 18:3451-62.

120. **Tomasevic N, Peculis BA. 2002.** Xenopus LSm proteins bind U8 snoRNA via an internal evolutionarily conserved octamer sequence. Mol Cell Biol 22:4101-12.

121. **Kufel J, Allmang C, Verdone L, Beggs J, Tollervey D. 2003.** A complex pathway for 3' processing of the yeast U3 snoRNA. Nucleic Acids Res 31:6788-97.

122. **Kufel J, Allmang C, Verdone L, Beggs JD, Tollervey D. 2002.** Lsm proteins are required for normal processing of pre-tRNAs and their efficient association with La-homologous protein Lhp1p. Mol Cell Biol 22:5248-56.

123. **Nielsen JS, Boggild A, Andersen CB, Nielsen G, Boysen A, Brodersen DE, Valentin-Hansen P. 2007.** An Hfq-like protein in archaea: crystal structure and functional characterization of the Sm protein from Methanococcus jannaschii. RNA 13:2213-23.

124. **Mura C, Cascio D, Sawaya MR, Eisenberg DS. 2001.** The crystal structure of a heptameric archaeal Sm protein: Implications for the eukaryotic snRNP core. Proc Natl Acad Sci U S A 98:5532-7.

125. **Collins BM, Harrop SJ, Kornfeld GD, Dawes IW, Curmi PM, Mabbutt BC. 2001.** Crystal structure of a heptameric Sm-like protein complex from archaea: implications for the structure and evolution of snRNPs. J Mol Biol 309:915-23.

126. **Toro I, Basquin J, Teo-Dreher H, Suck D. 2002.** Archaeal Sm proteins form heptameric and hexameric complexes: crystal structures of the Sm1 and Sm2 proteins from the hyperthermophile Archaeoglobus fulgidus. J Mol Biol 320:129-42.

127. **Kilic T, Sanglier S, Van Dorsselaer A, Suck D. 2006.** Oligomerization behavior of the archaeal Sm2-type protein from Archaeoglobus fulgidus. Protein Sci 15:2310-7.

128. **Mura C, Phillips M, Kozhukhovsky A, Eisenberg D. 2003.** Structure and assembly of an augmented Sm-like archaeal protein 14-mer. Proc Natl Acad Sci U S A 100:4539-44.

129. **Mura C, Kozhukhovsky A, Gingery M, Phillips M, Eisenberg D. 2003.** The oligomerization and ligand-binding properties of Sm-like archaeal proteins (SmAPs). Protein Sci 12:832-47.

130. **Arluison V, Mura C, Guzman MR, Liquier J, Pellegrini O, Gingery M, Regnier P, Marco S. 2006.** Three-dimensional structures of fibrillar Sm proteins: Hfq and other Sm-like proteins. J Mol Biol 356:86-96.

131. **Thore S, Mayer C, Sauter C, Weeks S, Suck D. 2003.** Crystal structures of the Pyrococcus abyssi Sm core and its complex with RNA. Common features of RNA binding in archaea and eukarya. J Biol Chem 278:1239-47.

132. **Fischer S, Benz J, Spath B, Maier LK, Straub J, Granzow M, Raabe M, Urlaub H, Hoffmann J, Brutschy B, Allers T, Soppa J, Marchfelder A. 2010.** The archaeal Lsm protein binds to small RNAs. J Biol Chem 285:34429-38.

133. **Maier LK, Benz J, Fischer S, Alstetter M, Jaschinski K, Hilker R, Becker A, Allers T, Soppa J, Marchfelder A. 2015.** Deletion of the Sm1 encoding motif in the lsm gene results in distinct changes in the transcriptome and enhanced swarming activity of Haloferax cells. Biochimie 117:129-37.

134. **Martens B, Bezerra GA, Kreuter MJ, Grishkovskaya I, Manica A, Arkhipova V, Djinovic-Carugo K, Blasi U. 2015.** The Heptameric SmAP1 and SmAP2 Proteins of the Crenarchaeon Sulfolobus Solfataricus Bind to Common and Distinct RNA Targets. Life (Basel) 5:1264-81.

135. **Martens B, Hou L, Amman F, Wolfinger MT, Evguenieva-Hackenberg E, Blasi U. 2017.** The SmAP1/2 proteins of the crenarchaeon Sulfolobus solfataricus interact with the exosome and stimulate A-rich tailing of transcripts. Nucleic Acids Res.

136. **Sittka A, Sharma CM, Rolle K, Vogel J. 2009.** Deep sequencing of Salmonella RNA associated with heterologous Hfq proteins in vivo reveals small RNAs as a major target class and identifies RNA processing phenotypes. RNA Biol 6:266-75.

137. **Brock TD, Brock KM, Belly RT, Weiss RL. 1972.** Sulfolobus: a new genus of sulfur-oxidizing bacteria living at low pH and high temperature. Arch Mikrobiol 84:54-68.

138. **Leigh JA, Albers SV, Atomi H, Allers T. 2011.** Model organisms for genetics in the domain Archaea: methanogens, halophiles, Thermococcales and Sulfolobales. FEMS Microbiol Rev 35:577-608.

139. **Wagner M, van Wolferen M, Wagner A, Lassak K, Meyer BH, Reimann J, Albers SV. 2012.** Versatile Genetic Tool Box for the Crenarchaeote Sulfolobus acidocaldarius. Front Microbiol 3:214.

140. **Li Y, Pan S, Zhang Y, Ren M, Feng M, Peng N, Chen L, Liang YX, She Q. 2016.** Harnessing Type I and Type III CRISPR-Cas systems for genome editing. Nucleic Acids Res 44:e34.

141. **Tripp V. 2016.** Pan-archaeal analysis of C/D box sRNA biogenesis and methylation targets. PhD thesis. Philipps University Marburg.

142. **Zheng G, Qin Y, Clark WC, Dai Q, Yi C, He C, Lambowitz AM, Pan T. 2015.** Efficient and quantitative high-throughput tRNA sequencing. Nat Methods 12:835-7.

143. **Tripp V, Martin R, Orell A, Alkhnbashi OS, Backofen R, Randau L. 2017.** Plasticity of archaeal C/D box sRNA biogenesis. Mol Microbiol 103:151-164.

144. **Lillestol RK, Shah SA, Brugger K, Redder P, Phan H, Christiansen J, Garrett RA. 2009.** CRISPR families of the crenarchaeal genus Sulfolobus: bidirectional transcription and dynamic properties. Mol Microbiol 72:259-72.

145. **Shah SA, Erdmann S, Mojica FJ, Garrett RA. 2013.** Protospacer recognition motifs: mixed identities and functional diversity. RNA Biol 10:891-9.

146. **Makarova KS, Wolf YI, Alkhnbashi OS, Costa F, Shah SA, Saunders SJ, Barrangou R, Brouns SJ, Charpentier E, Haft DH, Horvath P, Moineau S, Mojica FJ, Terns RM, Terns MP, White MF, Yakunin AF, Garrett RA, van der Oost J, Backofen R, Koonin EV. 2015.** An updated evolutionary classification of CRISPR-Cas systems. Nat Rev Microbiol 13:722-36.

147. **Kim VN. 2005.** MicroRNA biogenesis: coordinated cropping and dicing. Nat Rev Mol Cell Biol 6:376-85.

148. **Dominissini D, Moshitch-Moshkovitz S, Salmon-Divon M, Amariglio N, Rechavi G. 2013.** Transcriptome-wide mapping of N(6)-methyladenosine by m(6)A-seq based on immunocapturing and massively parallel sequencing. Nat Protoc 8:176-89.

149. **Forconi M, Herschlag D. 2009.** Metal Ion-Based Rna Cleavage as a Structural Probe. Methods in Enzymology, Vol 468: Biophysical, Chemical, and Functional Probes of Rna Structure, Interactions and Folding, Pt A 468:91-106.

150. **Love MI, Huber W, Anders S. 2014.** Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol 15:550.

151. **Goody TA, Melcher SE, Norman DG, Lilley DM. 2004.** The kink-turn motif in RNA is dimorphic, and metal ion-dependent. RNA 10:254-64.

152. **Seelos L. 2015.** Nachweis ribosomaler Methylierungsstellen in *Sulfolobus acidocaldarius*. Bachelor Thesis. Philipps University Marburg.

153. **Berg L, Kucharova V, Bakke I, Valla S, Brautaset T. 2012.** Exploring the 5'-UTR DNA region as a target for optimizing recombinant gene expression from the strong and inducible Pm promoter in Escherichia coli. J Biotechnol 158:224-30.

154. **Brunner M, Bujard H. 1987.** Promoter recognition and promoter strength in the Escherichia coli system. EMBO J 6:3139-44.

155. **Zwieb C, van Nues RW, Rosenblad MA, Brown JD, Samuelsson T. 2005.** A nomenclature for all signal recognition particle RNAs. RNA 11:7-13.

156. **Rosenblad MA, Larsen N, Samuelsson T, Zwieb C. 2009.** Kinship in the SRP RNA family. RNA Biol 6:508-16.

157. **Dar D, Prasse D, Schmitz RA, Sorek R. 2016.** Widespread formation of alternative 3' UTR isoforms via transcription termination in archaea. Nat Microbiol 1:16143.

158. **Tatusov RL, Galperin MY, Natale DA, Koonin EV. 2000.** The COG database: a tool for genome-scale analysis of protein functions and evolution. Nucleic Acids Res 28:33-6.

159. **Makarova KS, Sorokin AV, Novichkov PS, Wolf YI, Koonin EV. 2007.** Clusters of orthologous genes for 41 archaeal genomes and implications for evolutionary genomics of archaea. Biol Direct 2:33.

160. **Makarova KS, Wolf YI, Koonin EV. 2015.** Archaeal Clusters of Orthologous Genes (arCOGs): An Update and Application for Analysis of Shared Features between Thermococcales, Methanococcales, and Methanobacteriales. Life (Basel) 5:818-40.

161. **Huerta-Cepas J, Szklarczyk D, Forslund K, Cook H, Heller D, Walter MC, Rattei T, Mende DR, Sunagawa S, Kuhn M, Jensen LJ, von Mering C, Bork P. 2016.**

eggNOG 4.5: a hierarchical orthology framework with improved functional annotations for eukaryotic, prokaryotic and viral sequences. Nucleic Acids Res 44:D286-93.

162. **Martens B, Amman F, Manoharadas S, Zeichen L, Orell A, Albers SV, Hofacker I, Blasi U. 2013.** Alterations of the transcriptome of Sulfolobus acidocaldarius by exoribonuclease aCPSF2. PLoS One 8:e76569.

163. **Bailey TL, Boden M, Buske FA, Frith M, Grant CE, Clementi L, Ren J, Li WW, Noble WS. 2009.** MEME SUITE: tools for motif discovery and searching. Nucleic Acids Res 37:W202-8.

164. **Bailey TL, Williams N, Misleh C, Li WW. 2006.** MEME: discovering and analyzing DNA and protein sequence motifs. Nucleic Acids Res 34:W369-73.

165. **Crooks GE, Hon G, Chandonia JM, Brenner SE. 2004.** WebLogo: a sequence logo generator. Genome Res 14:1188-90.

166. **Akopian D, Shen K, Zhang X, Shan SO. 2013.** Signal recognition particle: an essential protein-targeting machine. Annu Rev Biochem 82:693-721.

167. **Halic M, Becker T, Pool MR, Spahn CM, Grassucci RA, Frank J, Beckmann R. 2004.** Structure of the signal recognition particle interacting with the elongation-arrested ribosome. Nature 427:808-14.

168. **Zwieb C, Bhuiyan S. 2010.** Archaea signal recognition particle shows the way. Archaea 2010:485051.

169. **Regalia M, Rosenblad MA, Samuelsson T. 2002.** Prediction of signal recognition particle RNA genes. Nucleic Acids Res 30:3368-77.

170. **Iakhiaeva E, Wower J, Wower IK, Zwieb C. 2008.** The 5e motif of eukaryotic signal recognition particle RNA contains a conserved adenosine for the binding of SRP72. RNA 14:1143-53.

171. **Iakhiaeva E, Iakhiaev A, Zwieb C. 2010.** Identification of amino acid residues in protein SRP72 required for binding to a kinked 5e motif of the human signal recognition particle RNA. BMC Mol Biol 11:83.

172. **Bhuiyan SH, Gowda K, Hotokezaka H, Zwieb C. 2000.** Assembly of archaeal signal recognition particle from recombinant components. Nucleic Acids Res 28:1365-73.

173. **Hainzl T, Huang S, Merilainen G, Brannstrom K, Sauer-Eriksson AE. 2011.** Structural basis of signal-sequence recognition by the signal recognition particle. Nat Struct Mol Biol 18:389-91.

174. **Moll RG. 2003.** Protein-protein, protein-RNA and protein-lipid interactions of signal-recognition particle components in the hyperthermoacidophilic archaeon Acidianus ambivalens. Biochem J 374:247-54.

175. **Rose RW, Pohlschroder M. 2002.** In vivo analysis of an essential archaeal signal recognition particle in its native host. J Bacteriol 184:3260-7.

176. **Nomura M, Gourse R, Baughman G. 1984.** Regulation of the synthesis of ribosomes and ribosomal components. Annu Rev Biochem 53:75-117.

177. **Nomura M, Yates JL, Dean D, Post LE. 1980.** Feedback regulation of ribosomal protein gene expression in Escherichia coli: structural homology of ribosomal RNA and ribosomal protein MRNA. Proc Natl Acad Sci U S A 77:7084-8.

178. **Holcik M, Sonenberg N. 2005.** Translational control in stress and apoptosis. Nat Rev Mol Cell Biol 6:318-27.

179. **Jager D, Forstner KU, Sharma CM, Santangelo TJ, Reeve JN. 2014.** Primary transcriptome map of the hyperthermophilic archaeon Thermococcus kodakarensis. BMC Genomics 15:684.

180. **Tollervey D, Lehtonen H, Carmo-Fonseca M, Hurt EC. 1991.** The small nucleolar RNP protein NOP1 (fibrillarin) is required for pre-rRNA processing in yeast. EMBO J 10:573-83.

181. **Torarinsson E, Klenk HP, Garrett RA. 2005.** Divergent transcriptional and translational signals in Archaea. Environ Microbiol 7:47-54.

182. **Richter H, Zoephel J, Schermuly J, Maticzka D, Backofen R, Randau L. 2012.** Characterization of CRISPR RNA processing in Clostridium thermocellum and Methanococcus maripaludis. Nucleic Acids Res 40:9887-96.

183. **Lui L. 2015.** Evolution of structure and function of kink-turn containing RNAs in the domain Archaea. PhD Thesis. University of California, Santa Cruz.

184. **Randau L. 2012.** RNA processing in the minimal organism Nanoarchaeum equitans. Genome Biol 13:R63.

185. **Lui L, Lowe T. 2013.** Small nucleolar RNAs and RNA-guided post-transcriptional modification. Essays Biochem 54:53-77.

186. **Dabeva MD, Post-Beittenmiller MA, Warner JR. 1986.** Autogenous regulation of splicing of the transcript of a yeast ribosomal protein gene. Proc Natl Acad Sci U S A 83:5854-7.

187. **Eng FJ, Warner JR. 1991.** Structural basis for the regulation of splicing of a yeast messenger RNA. Cell 65:797-804.

188. **Dabeva MD, Warner JR. 1993.** Ribosomal protein L32 of Saccharomyces cerevisiae regulates both splicing and translation of its own transcript. J Biol Chem 268:19669-74.

189. **Li B, Vilardell J, Warner JR. 1996.** An RNA structure involved in feedback regulation of splicing and of translation is critical for biological fitness. Proc Natl Acad Sci U S A 93:1596-600.

190. **Vilardell J, Yu SJ, Warner JR. 2000.** Multiple functions of an evolutionarily conserved RNA binding domain. Mol Cell 5:761-6.

191. **Sojka L, Fucik V, Krasny L, Barvik I, Jonak J. 2007.** YbxF, a protein associated with exponential-phase ribosomes in Bacillus subtilis. J Bacteriol 189:4809-14.

192. **Ihsanawati, Nishimoto M, Higashijima K, Shirouzu M, Grosjean H, Bessho Y, Yokoyama S. 2008.** Crystal structure of tRNA N2,N2-guanosine dimethyltransferase Trm1 from Pyrococcus horikoshii. J Mol Biol 383:871-84.

193. **Constantinesco F, Benachenhou N, Motorin Y, Grosjean H. 1998.** The tRNA(guanine-26,N2-N2) methyltransferase (Trm1) from the hyperthermophilic archaeon Pyrococcus furiosus: cloning, sequencing of the gene and its expression in Escherichia coli. Nucleic Acids Res 26:3753-61.

194. **Tran E, Brown J, Maxwell ES. 2004.** Evolutionary origins of the RNA-guided nucleotide-modification complexes: from the primitive translation apparatus? Trends Biochem Sci 29:343-50.

195. **Qi LS, Arkin AP. 2014.** A versatile framework for microbial engineering using synthetic non-coding RNAs. Nat Rev Microbiol 12:341-54.

196. **Achsel T, Stark H, Luhrmann R. 2001.** The Sm domain is an ancient RNA-binding motif with oligo(U) specificity. Proc Natl Acad Sci U S A 98:3685-9.

197. **Pastura A. 2017.** RNA substrate binding of LSm1/LSm2 complexes from Sulfolobus acidocaldarius. Master Thesis. Philipps University Marburg.

198. **Link TM, Valentin-Hansen P, Brennan RG. 2009.** Structure of Escherichia coli Hfq bound to polyriboadenylate RNA. Proc Natl Acad Sci U S A 106:19292-7.

199. **Chen CY, Shyu AB. 2011.** Mechanisms of deadenylation-dependent decay. Wiley Interdiscip Rev RNA 2:167-83.

200. **Portnoy V, Evguenieva-Hackenberg E, Klein F, Walter P, Lorentzen E, Klug G, Schuster G. 2005.** RNA polyadenylation in Archaea: not observed in Haloferax while the exosome polynucleotidylates RNA in Sulfolobus. EMBO Rep 6:1188-93.

201. **Evguenieva-Hackenberg E, Walter P, Hochleitner E, Lottspeich F, Klug G. 2003.** An exosome-like complex in Sulfolobus solfataricus. EMBO Rep 4:889-93.

202. **Lorentzen E, Walter P, Fribourg S, Evguenieva-Hackenberg E, Klug G, Conti E. 2005.** The archaeal exosome core is a hexameric ring structure with three catalytic subunits. Nat Struct Mol Biol 12:575-81.

203. **Evguenieva-Hackenberg E. 2011.** The archaeal exosome. Adv Exp Med Biol 702:29-38.

204. **Mohanty BK, Kushner SR. 2000.** Polynucleotide phosphorylase functions both as a 3' right-arrow 5' exonuclease and a poly(A) polymerase in Escherichia coli. Proc Natl Acad Sci U S A 97:11966-71.

205. **Evguenieva-Hackenberg E, Roppelt V, Finsterseifer P, Klug G. 2008.** Rrp4 and Csl4 are needed for efficient degradation but not for polyadenylation of synthetic and natural RNA by the archaeal exosome. Biochemistry 47:13158-68.

206. **Portnoy V, Schuster G. 2006.** RNA polyadenylation and degradation in different Archaea; roles of the exosome and RNase R. Nucleic Acids Res 34:5923-31.

207. **Lee T, Feig AL. 2008.** The RNA binding protein Hfq interacts specifically with tRNAs. RNA 14:514-23.

208. **Sobrero P, Valverde C. 2012.** The bacterial protein Hfq: much more than a mere RNA-binding factor. Crit Rev Microbiol 38:276-99.

209. **Ausubel FM. 1987.** Current Protocols in Molecular Biology. Wiley.

210. **Sambrook JF, Fritsch EF, Maniatis, T. 1989.** Molecular Cloning: A laboratory manual. Cold Spring Harbor Laboratory Press.

211. **Hanahan D. 1983.** Studies on transformation of Escherichia coli with plasmids. J Mol Biol 166:557-80.

212. **Vieira J, Messing J. 1982.** The pUC plasmids, an M13mp7-derived system for insertion mutagenesis and sequencing with synthetic universal primers. Gene 19:259-68.

213. **Mullis K, Faloona F, Scharf S, Saiki R, Horn G, Erlich H. 1986.** Specific enzymatic amplification of DNA in vitro: the polymerase chain reaction. Cold Spring Harb Symp Quant Biol 51 Pt 1:263-73.

214. **Saiki RK, Gelfand DH, Stoffel S, Scharf SJ, Higuchi R, Horn GT, Mullis KB, Erlich HA. 1988.** Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase. Science 239:487-91.

215. **Wallace RB, Shaffer J, Murphy RF, Bonner J, Hirose T, Itakura K. 1979.** Hybridization of synthetic oligodeoxyribonucleotides to phi chi 174 DNA: the effect of single base pair mismatch. Nucleic Acids Res 6:3543-57.

216. **Ho SN, Hunt HD, Horton RM, Pullen JK, Pease LR. 1989.** Site-directed mutagenesis by overlap extension using the polymerase chain reaction. Gene 77:51-9.

217. **Gibson DG, Young L, Chuang RY, Venter JC, Hutchison CA, 3rd, Smith HO. 2009.** Enzymatic assembly of DNA molecules up to several hundred kilobases. Nat Methods 6:343-5.

218. **Fu C, Donovan WP, Shikapwashya-Hasser O, Ye X, Cole RH. 2014.** Hot Fusion: an efficient method to clone multiple DNA fragments as well as inverted repeats without ligase. PLoS One 9:e115318.

219. **Inoue H, Nojima H, Okayama H. 1990.** High efficiency transformation of Escherichia coli with plasmids. Gene 96:23-8.

220. **Berkner S, Grogan D, Albers SV, Lipps G. 2007.** Small multicopy, non-integrative shuttle vectors based on the plasmid pRN1 for Sulfolobus acidocaldarius and Sulfolobus solfataricus, model organisms of the (cren-)archaea. Nucleic Acids Res 35:e88.

221. **Sanger F, Nicklen S, Coulson AR. 1992.** DNA sequencing with chain-terminating inhibitors. 1977. Biotechnology 24:104-8.

222. **Laemmli UK. 1970.** Cleavage of structural proteins during the assembly of the head of bacteriophage T4. Nature 227:680-5.

223. **Burnette WN. 1981.** "Western blotting": electrophoretic transfer of proteins from sodium dodecyl sulfate--polyacrylamide gels to unmodified nitrocellulose and radiographic detection with antibody and radioiodinated protein A. Anal Biochem 112:195-203.

224. **Bradford MM. 1976.** A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. Anal Biochem 72:248-54.

225. **Miller JH. 1972.** Experiments in Molecular Genetics. Cold Spring Harbor Laboratory Press, New York.

226. **Wagner M, Wagner A, Ma X, Kort JC, Ghosh A, Rauch B, Siebers B, Albers SV. 2014.** Investigation of the malE promoter and MalR, a positive regulator of the maltose regulon, for an improved expression system in Sulfolobus acidocaldarius. Appl Environ Microbiol 80:1072-81.

227. **Holmqvist E, Wright PR, Li L, Bischler T, Barquist L, Reinhardt R, Backofen R, Vogel J. 2016.** Global RNA recognition patterns of post-transcriptional regulators Hfq and CsrA revealed by UV crosslinking in vivo. EMBO J 35:991-1011.

228. **Langenberger D, Bermudez-Santana C, Hertel J, Hoffmann S, Khaitovich P, Stadler PF. 2009.** Evidence for human microRNA-offset RNAs in small RNA sequencing data. Bioinformatics 25:2298-301.

6. Appendix

Appendix 1. List of small RNAs from *S. acidocaldarius.* The table summarizes the small RNAs of *S. acidocaldarius* that were identified by the sRNome analysis by screening for RNAs that are covered by at least 100 reads in one of the four libraries (A, B, C and D). For clarity, rRNAs, tRNAs, C/D box sRNAs and crRNAs were excluded from the list.

Appendix 2. List of L7Ae-interacting RNAs. 107 L7Ae-interacting RNAs are listed for the RIP-Seq analysis that were identified by a custom peak calling approach (q-value < 0.1) and by filtering for the mean normalized read count of the L7Ae sRNA data sets (at least 3000 reads). Identified interacting C/D box sRNAs are named by Sac-sR#. The sequence of the region that was enriched for the k-turn containing mRNAs is shown at the right side of the table. Putative Kt-b and Kt-n strands are marked in red and green, respectively. The k-turn containing mRNAs that were analyzed in this study are marked in yellow.

Appendix 3. Multiple sequence alignment of 121 archaeal *l7ae* promoter sequences. The alignment of the 121 (Archaeal Genome Browser list) archaeal *l7ae* promoter sequences was separately performed for the phyla Crenarchaeota, Thaumarchaeota, Euryarchaeota and for other archaea by using the alignment tool Clustal Omega. Putative TATA box elements, transcription start sites (TSS), Shine-Dalgarno (SD) sequences and the Kt-n and Kt-b sequences are indicated. For clarity, an extract of the alignment is shown. The full alignment can be downloaded at: http://mbio.asm.org/content/8/4/e00730-17/DC6/embed/inline-supplementary-material-6.docx

Appendix 4. Multiple sequence alignment of 82 archaeal SRP RNAs. 82 (list of archaea from the SRP database) archaeal SRP RNAs were aligned using the tool Clustal Omega. Putative Kt-b and Kt-n strands are indicated. Conservation of the GA nucleotides of the putative k-turn sequences is found for the SRP RNAs. The bulged (5e-b) and non-bulged (5e-n) strands of the 5e motif are indicated. For clarity, an extract of the alignment is shown.

Appendix 5. Mass spectrometry analysis of Ni-NTA-purified LSm2-CHis. The table lists the co-isolated proteins of LSm2-CHis that were identified by mass spectrometry analysis. Identified proteins which are referred to in this study are marked in yellow. For clarity, cells were hidden, which is indicated by black lines. The analysis was performed by Jörg Kahnt.

Appendix 1. List of small RNAs from *S. acidocaldarius.*

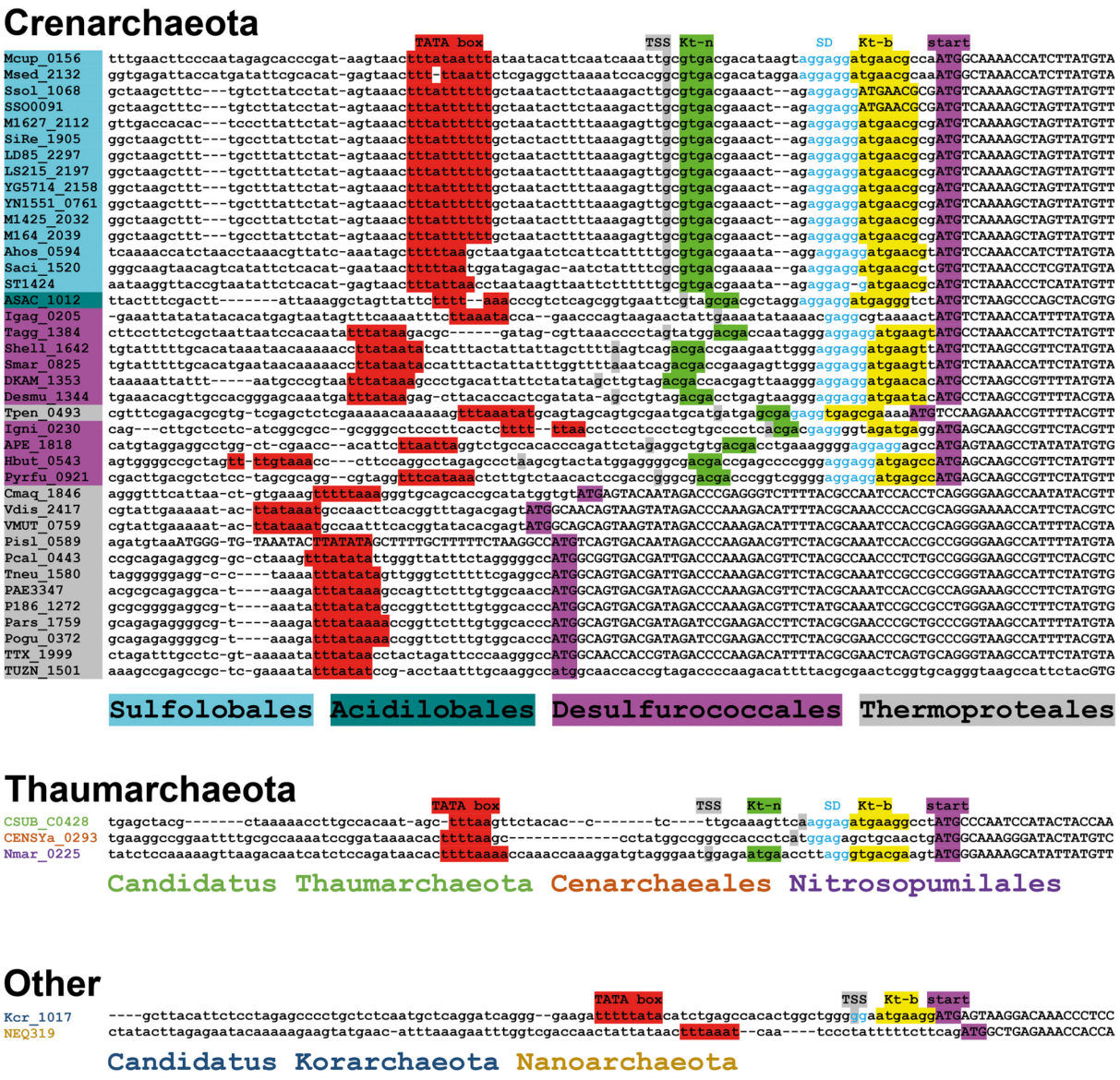
List of sRNome RNAs (at least 100 reads for one of the four sRNome libraries; exluded are mRNA, tRNA and rRNA fragments in sense direction, C/D box sRNAs and crRNAs)								
					sRNome (1.000.000 mapped reads per library)			
Number	Region of enriched RNA	Length	Name	Remarks	# A (50-150 nt)	# B (50-150 nt)	# C (0-150 nt)	# D (0-150 nt)
1	complement(30463..30503)	41	K-turn mRNA	coding sequence of Saci_0046; overlaps with Saci_0045	467	317	743	663
2	complement(34704..34761)	44	antisense RNA	antisense RNA of Saci_0049	585	106	474	569
3	49287..49313	27	antisense tRNA	antisense tRNA of Saci_0066	278	71	509	588
4	72171..72477	307	SRP RNA		374	182	684	316
5	72645..72668	24	antisense tRNA	antisense tRNA of Saci_0096	17	4	119	192
6	complement(92843..92875)	33	antisense RNA	antisense RNA of Saci_0123	11	7	120	89
7	109661..109730	70	K-turn mRNA	5' UTR of Saci_0139; correct gene length: WP_015385366.1	820	264	499	343
8	complement(110496..110526)	31	antisense RNA	antisense RNA of Saci_0140	14	5	138	110
9	complement(124025..124050)	20	antisense RNA	antisense RNA of Saci_0154	49	29	189	156
10	126801..126848	48	antisense RNA	antisense RNA of Saci_0156	46	30	116	122
11	complement(130023..130091)	56	H/ACA box sRNA	according to Sebastien Muller 2009 NAR	125	47	85	79
12	196988..197059	72	K-turn mRNA	coding sequence of Saci_0237	247	111	280	142
13	complement(202601..202647)	47	antisense RNA	antisense RNA of Saci_0242	71	24	211	94
14	301602..301627	26	antisense RNA	antisense RNA of Saci_0358	43	24	110	119
15	complement(345774..345831)	58	K-turn sRNA	probably C/D box sRNA	352	97	109	54
16	complement(349691..349729)	39	sRNA		501	322	546	496
17	complement(395118..395164)	47	antisense tRNA	antisense tRNA of Saci_0470	95	26	135	60
18	complement(395666..395699)	34	antisense RNA	antisense RNA of Saci_0472	47	10	417	194
19	417453..417526	74	sRNA	antisense of Sac-sR39	774	134	352	230
20	complement(418864..419030)	167	antisense RNA	antisense RNA of Saci_0504	490	224	488	168
21	440237..440283	47	sRNA		122	52	196	197
22	complement(453328..453351)	24	antisense RNA	antisense RNA of Saci_0556	36	28	125	270
23	458698..458724	27	antisense tRNA	antisense tRNA of Saci_0561	167	362	1068	1123
24	458981..459059	79	H/ACA box sRNA	also found in *S. solfataricus* -> Zago 2005 Mol Mic	132	62	137	77
25	476049..476076	28	antisense RNA	antisense RNA of Saci_0597	23	3	167	142
26	489282..489345	64	antisense RNA	antisense RNA of Saci_0611 and Saci_0612	125	54	87	62
27	499628..499664	37	sRNA		183	178	412	513
28	complement(501154..501185)	32	K-turn mRNA	coding sequence of Saci_0630	27	15	235	138
29	519616..519665	50	antisense RNA	antisense RNA of Saci_0655	23	9	110	100
30	529073..529257	185	antisense RNA	antisense RNA of Saci_0667	370	457	368	218
31	complement(535192..535236)	45	antisense tRNA	antisense tRNA of Saci_0674	357	76	662	540
32	563570..563598	29	antisense tRNA	antisense tRNA of Saci_0705	63	15	256	99
33	584601..584661	61	sRNA		124	61	138	109
34	585936..586248	313	RNase P RNA		470	445	455	397
35	610562..610585	24	antisense tRNA	antisense tRNA of Saci_0764	52	18	140	156
36	complement(611192..611275)	84	antisense RNA	antisense RNA of Saci_0765	146	72	140	185
37	complement(686936..687007)	72	antisense RNA	antisense RNA of Saci_0856	94	47	110	48
38	complement(691848..691906)	59	antisense RNA	antisense RNA of Saci_0863	54	45	114	131
39	complement(692154..692177)	24	antisense RNA	antisense RNA of Saci_0864	70	66	450	1340
40	complement(694089..694120)	32	antisense tRNA	antisense tRNA of Saci_0868	199	23	1433	356
41	complement(694256..694294)	39	sRNA		54	16	94	107
42	complement(754745..754774)	30	antisense RNA	antisense RNA of Saci_0941	132	103	230	378
43	complement(773358..773386)	29	K-turn mRNA	coding sequence of Saci_0967	154	129	230	153
44	complement(779195..779258)	64	K-turn mRNA	coding sequence of Saci_0974	133	233	174	163
45	complement(782506..782565)	60	antisense RNA	antisense RNA of Saci_0977	202	72	205	209
46	805391..805427	37	K-turn mRNA	coding sequence of Saci_1004	84	74	194	152
47	805735..805858	124	sRNA + antisense RNA	RrrR RNA	7075	1775	20734	27919
48	834024..834052	29	antisense RNA	antisense RNA of Saci_1035	9	3	101	131
49	842930..842963	34	antisense RNA	antisense RNA of Saci_1041 and Saci_1042	88	33	203	160
50	1030957..1030991	35	antisense RNA	antisense RNA of Saci_1212	32	31	70	136
51	1043124..1043206	83	K-turn sRNA	could be H/ACA box sRNA	1212	643	1474	910
52	complement(1103840..1103873)	34	sRNA	close to Saci_1299 (23S rRNA)	239	97	1525	1146
53	complement(1108682..1108726)	45	sRNA	close to Saci_1300 (16S rRNA)	121	67	94	49
54	1129862..1129909	48	antisense RNA	antisense RNA of SACI_RS06310	278	94	338	218
55	complement(1130243..1130279)	37	antisense RNA	antisense RNA of Saci_1322	82	57	142	93
56	complement(1148823..1148947)	125	K-turn mRNA	5' UTR of Saci_1347 (nop5)	108	70	98	72
57	1160136..1160197	62	antisense RNA	antisense RNA of Saci_1358	112	62	137	92
58	1168227..1168374	148	sRNA		133	48	124	70
59	complement(1236240..1236309)	70	sRNA		123	45	290	160
60	complement(1250761..1250815)	33	K-turn mRNA	coding sequence of Saci_1468	101	64	211	152
61	complement(1250990..1251036)	47	K-turn mRNA	5' UTR of Saci_1468	317	194	288	123
62	1251499..1251531	33	antisense RNA	antisense RNA of Saci_1469	6	2	120	82
63	1268345..1268396	52	antisense RNA	antisense RNA of Saci_1491	101	9	100	30
64	complement(1292527..1292555)	29	antisense RNA	antisense RNA of Saci_1514	87	48	383	319
65	1293914..1294038	125	5S rRNA		65728	7431	19169	51455
66	1297690..1297732	43	K-turn mRNA	5' UTR of Saci_1520 (l7ae)	472	317	627	435
67	1338266..1338294	29	antisense RNA	antisense RNA of Saci_1565	29	16	143	85
68	1387630..1387703	74	antisense RNA	antisense RNA of Saci_1624	113	63	151	147
69	1400450..1400505	54	K-turn sRNA	transcribed with Sac-sR15; could be C/D box sRNA	1464	577	787	545
70	1458935..1458989	55	sRNA		130	97	82	88
71	complement(1507703..1507796)	94	antisense RNA	antisense RNA of Saci_1747	76	117	75	70
72	1522346..1522379	34	antisense RNA	antisense RNA of Saci_1762	88	33	186	187
73	complement(1523500..1523548)	49	antisense RNA	antisense RNA of Saci_1763	188	175	135	280
74	complement(1523945..1523995)	51	antisense RNA	antisense RNA of Saci_1763	110	65	171	148
75	complement(1524036..1524082)	47	antisense RNA	antisense RNA of Saci_1763	61	65	171	148
76	complement(1650666..1658600)		CRISPR Array 1	132 spacer sequences				
77	1670945..1675658		CRISPR Array 2	77 spacer sequences				
78	complement(1676267..1676310)	44	sRNA		35	6	129	36
79	complement(1736546..1736646)	101	sRNA		159	52	127	54
80	1739282..1739381	100	K-turn mRNA	coding sequence of Saci_1928	93	598	301	304
81	1739540..1739639	100	K-turn mRNA	coding sequence of Saci_1928	217	663	452	505
82	1763226..1763256	31	sRNA		15	2	139	101
83	1764722..1764793	72	K-turn mRNA	coding sequence of Saci_1951	1774	1254	1487	846
84	complement(1786368..1786621)		CRISPR Array 3	4 spacer sequences				
85	complement(1825968..1826650)		CRISPR Array 4	10 spacer sequences				
86	1830280..1830313	34	K-turn sRNA	might be antisense to 3' UTR of Saci_2016	48	71	458	487
87	1842184..1842221	38	K-turn mRNA	coding sequence of Saci_2027	46	73	328	385
88	complement(1870638..1870717)	80	sRNA		117	61	88	60
89	complement(2157472..2157500)	29	K-turn mRNA	coding sequence of Saci_2311	74	27	226	105
90	2161778..2161815	38	antisense RNA	antisense RNA of Saci_2316	115	32	266	217
91	2196939..2196987	49	K-turn sRNA	transcribed with Sac-sR16 and Sac-sR132; could be C/D box sRNA	628	433	589	465

Appendix 2. List of L7Ae-interacting RNAs.

List of L7Ae-interacting RNAs (significance below q-value 0.1 and filtered for the mean normalized read count of L7Ae sRNA1&2 (at least 3000 reads))

Number	Exact Locus of Called RNA	Length	Name	Remarks	Peak Call	Peak Start	Peak End	Strand	Normalized Read Count				Log2 Fold Change	q-value	K-turn mRNA/sRNA sequences with marked Kt-b (red) and Kt-n (green) strands
									Control	L7Ae sRNA1	L7Ae sRNA2	Mean sRNA1&2			
1	2497..2555	59	Sac-sR101		clip_peak_11	2439	2589	+	3160	32382	42235	37308	3.45	3.10E-014	
2	complement(30463..30503)	41	K-turn mRNA	coding sequence of Saci_0046; overlaps with Saci_0045	clip_peak_2028	30426	30557	-	499	8553	6644	7599	3.81	4.32E-017	AAGAGTGATGAGACGTTACAAGTCAAGACTGAACGACTTTT
3	complement(41317..41392)	76	mRNA fragment	coding sequence of Saci_0054	clip_peak_2060	41281	41435	-	1473	1995	4647	3321	1.12	0.096292208	
4	42135..42193	59	Sac-sR9		clip_peak_27	42123	42251	+	7545	22776	21510	22143	1.51	0.002003022	
5	49051..49109	59	Sac-sR102		clip_peak_40	49009	49157	+	1059	35392	24952	30172	4.67	1.49E-024	
6	complement(49180..49311)	132	tRNA fragment	tRNA fragment of Saci_0066; mandy reads of intron	clip_peak_2068	49140	49317	-	652	6143	5187	5665	3.02	2.60E-011	
7	62922..62965	44	mRNA fragment	coding sequence of Saci_0080	clip_peak_51	62833	63033	+	1402	3633	9748	6691	2.14	0.000635448	
8	66860..66901	42	mRNA fragment	coding sequence of Saci_0088 and Saci_0089	clip_peak_54	66783	66969	+	476	3835	2656	3245	2.68	3.45E-008	
9	72171..72478	308	SRP RNA	high enrichment at locus of the two K-turns within SRP RNA	clip_peak_76	72352	72501	+	9003	40185	16124	28154	1.57	0.014786654	
10	97321..97378	58	Sac-sR24		clip_peak_90	97295	97416	+	9725	26126	20020	23073	1.21	0.022468405	
11	109661..109730	70	K-turn mRNA	5' UTR of SACI_RS00660 (WP_015385366.1)	clip_peak_101	109648	109760	+	581	10243	5479	7861	3.61	1.66E-012	ATAAGTAGGGGTATCTAGCCCTCTGACTGCCTCGCAGAGGGCTGTGTGAGGAGGGCTTCCTGCCGTGAAT
12	complement(129966..130023)	58	Sac-sR103	preceded by an H/ACA box sRNA	clip_peak_2143	129924	130080	-	2435	55807	34562	45184	4.06	2.30E-017	
13	complement(130023..130091)	56	H/ACA box sRNA	according to Sebastien Muller 2009 NAR	clip_peak_2143	129924	130080	-	2435	55807	34562	45184	4.06	2.30E-017	
14	131786..131845	60	Sac-sR11		clip_peak_116	131741	131879	+	4968	42170	32103	37137	2.81	1.24E-009	
15	158261..158501	241	mRNA fragment	coding sequence of Saci_0195	clip_peak_162	158248	158518	+	1589	4324	3808	4066	1.31	0.009974763	
16	complement(179538..179593)	56	Sac-sR104	in L7Ae data sets only 1st half enriched	clip_peak_2211	179517	179597	-	376	28859	20934	24897	5.85	1.91E-038	
17	196988..197059	72	K-turn mRNA	coding sequence of Saci_0237	clip_peak_203	196973	197109	+	64	4041	3442	3741	5.69	4.09E-037	AATGATGATGAAGCTACAATTGTCATAAAGAAAGTTCGCAGACAAATAAGGGATGATGAACACGTTTGTGTT
18	217035..217093	59	Sac-sR10	most abundant sRNA in sRNome and L7Ae sRNA data	clip_peak_221	216975	217141	+	87257	4843146	4182513	4512830	5.52	2.94E-037	
19	285911..285956	46	K-turn mRNA	coding sequence of Saci_0333	clip_peak_297	285844	286114	+	847	3709	5387	4548	2.34	1.95E-006	TCCAACCTGTTAGAGATTTTGAGGGAAATGGATGATAACTCAGTAA
20	308044..308101	58	Sac-sR105		clip_peak_327	308022	308148	+	9363	96362	80278	88320	3.14	2.48E-012	
21	325553..325612	60	Sac-sR106		clip_peak_347	325506	325652	+	870	11246	7227	9237	3.29	9.22E-012	
22	328393..328450	58	Sac-sR107		clip_peak_351	328365	328502	+	7354	113927	74625	94276	3.55	6.35E-014	
23	329195..329252	58	Sac-sR36		clip_peak_352	329178	329271	+	2563	56143	57114	56629	4.34	1.18E-023	
24	345686..345746	61	Sac-sR13		clip_peak_359	345675	345800	+	1843	31592	28842	30217	3.92	3.41E-019	
25	368184..368239	56	Sac-sR6		clip_peak_384	368129	368277	+	3342	90427	72846	81636	4.47	4.81E-024	
26	371690..371743	54	Sac-sR108		clip_peak_394	371672	371762	+	151	4354	3119	3736	4.48	4.70E-022	
27	complement(393868..393908)	41	Sac-sR7		clip_peak_2468	393792	393938	-	852	3914	2938	3426	1.94	8.82E-005	
28	417453..417526	74	sRNA	antisense of Sac-sR39 (annotated in archaeal genome browser)	clip_peak_424	417425	417540	+	1611	7736	5957	6846	2.02	3.10E-005	
29	complement(418940..419000)	61	antisense RNA	antisense RNA of Saci_0504	clip_peak_2496	418886	419097	-	286	5086	3112	4099	3.70	4.07E-014	
30	complement(449458..449515)	58	Sac-sR5		clip_peak_2521	449416	449559	-	11033	497360	344006	420683	5.08	8.64E-029	
31	450909..450965	57	Sac-sR109		clip_peak_444	450892	451013	+	11435	295721	185332	240526	4.24	5.07E-019	
32	458753..458931	179	Sac-sR20	only 2nd half of C/D box sRNA is enriched	clip_peak_450	458729	458930	+	422	3410	2647	3028	2.75	4.74E-009	
33	complement(472892..472946)	55	mRNA fragment	coding sequence of Saci_0593	clip_peak_2568	472818	473020	-	1580	4658	7732	6195	1.90	0.000345196	
34	complement(490815..490870)	56	Sac-sR110	in L7Ae data sets only 2nd half enriched	clip_peak_2622	490749	490899	-	1860	36081	52694	34387	4.08	7.84E-021	
35	complement(501158..501185)	28	K-turn mRNA	coding sequence of Saci_0630	clip_peak_2648	501060	501218	-	472	6533	4263	5398	3.39	1.66E-012	ACGGATGATGAAAAGAAAAAGGAGCTTG
36	528451..528478	28	mRNA fragment	coding sequence of Saci_0666	clip_peak_493	528346	528578	+	1587	2561	4135	3348	1.04	0.078888327	
37	528670..528750	81	mRNA fragment	coding sequence of Saci_0666	clip_peak_492	528606	528874	+	5679	21023	27754	24388	2.04	2.57E-005	
38	528773..528829	57	mRNA fragment	coding sequence of Saci_0666	clip_peak_492	528606	528874	+	5679	21023	27754	24388	2.04	2.57E-005	
39	529107..529208	102	antisense RNA	antisense RNA of Saci_0667	clip_peak_491	529045	529252	+	4111	22600	16537	19568	2.18	6.80E-006	
40	complement(531406..531486)	81	K-turn mRNA	coding sequence of Saci_0669	clip_peak_2689	531341	531574	-	1373	3451	3550	3501	1.31	0.009762397	AGAATGAGTGTGAAGATAAGGCTAAAAGGTCCTCTGGCGACAAGGTTAGGAAGAGACGAATTTGTTATTTCTCTGAAAGCT
41	complement(534770..534799)	30	K-turn mRNA	5' UTR of Saci_0672	clip_peak_2702	534709	534910	-	2422	5933	4938	5435	1.13	0.032469591	AAGCGTGATGAACGTCTGAGAGTTGCTGTA
42	complement(559319..559377)	59	Sac-sR111		clip_peak_2745	559278	559386	-	637	11551	8216	9884	3.83	1.38E-016	
43	complement(572676..572735)	60	Sac-sR112		clip_peak_2765	572630	572755	-	3667	175287	132095	153691	5.22	9.68E-032	
44	585936..586248	313	RNaseP	high enrichment at locus of the two K-turns within RNase P RNA	clip_peak_544	585944	586148	+	451	8947	5412	7179	3.85	3.23E-015	
45	631932..631987	56	Sac-sR113		clip_peak_617	631880	632020	+	8021	55759	49313	52536	2.63	5.93E-009	
46	complement(669555..669612)	58	Sac-sR114		clip_peak_2851	669501	669674	-	3419	96049	82401	89225	4.57	1.28E-025	
47	679119..679177	59	Sac-sR115		clip_peak_670	679048	679192	+	4956	38641	31900	35271	2.75	1.63E-009	
48	complement(693495..693541)	47	K-turn mRNA	coding sequence of Saci_0886	clip_peak_2896	693387	693616	-	473	5258	4297	4777	3.23	1.17E-012	ATATTTGATGAAGAACGTGGAGAATATATTTGTACTGAGACAGGTGA
49	701895..701953	59	Sac-sR22		clip_peak_688	701867	701987	+	1075	14419	10735	12577	3.43	7.17E-014	
50	704342..704395	54	Sac-sR116		clip_peak_692	704319	704441	+	511	40035	43226	41630	6.16	8.71E-047	
51	714983..715035	53	Sac-sR117		clip_peak_701	714942	715086	+	2373	21486	13561	17523	2.78	1.54E-008	
52	755543..755609	67	Sac-sR118		clip_peak_755	755509	755651	+	2978	595404	432013	513709	7.19	9.44E-059	
53	complement(773358..773386)	29	K-turn mRNA	coding sequence of Saci_0967	clip_peak_2972	773245	773451	-	514	4560	4246	4403	3.01	2.30E-011	ACTGATGATGAAATTGAGTACGCGATTGA
54	complement(779094..779178)	30	mRNA fragment	coding sequence and 3' UTR of Saci_0974	clip_peak_2982	779099	779330	-	5702	14223	17189	15706	1.42	0.004941735	
55	complement(779195..779258)	64	K-turn mRNA	coding sequence of Saci_0974	clip_peak_2982	779099	779330	-	5702	14223	17189	15706	1.42	0.004941735	GTAATCGAACTGTGACCGAGATTGTTGACAGAGCGAGTAATGATACAAATCTTCCGTCTGATCA
56	complement(779398..779439)	42	K-turn mRNA	coding sequence of Saci_0974 and Saci_0975	clip_peak_2983	779400	779598	-	3993	7769	10542	9156	1.16	0.032067084	GAGCGAAGGAGAGGAAGATGAATGATTTTGATGAAGTAAAA
57	782598..782662	65	Sac-sR27		clip_peak_786	782570	782715	+	1732	40858	31984	36421	4.26	1.38E-021	
58	complement(793953..794012)	60	Sac-sR17		clip_peak_3015	793936	794056	-	17772	397306	278091	337698	4.11	4.25E-019	
59	805391..805427	37	K-turn mRNA	coding sequence of Saci_1004	clip_peak_797	805367	805454	+	242	9137	5479	7308	4.74	1.38E-022	AAGGATGATGAAGAAGGAATATGGATTCTGAGGGTAA
60	808273..808331	59	Sac-sR28		clip_peak_802	808234	808357	+	1412	22944	16644	19794	3.69	1.07E-015	
61	complement(830368..830430)	63	Sac-sR119		clip_peak_3040	830332	830481	-	639	13950	10796	12373	4.14	3.27E-020	
62	1025710..1025737	27	Sac-sR23	in L7Ae data sets only 1st half enriched	clip_peak_984	1025701	1025817	+	1023	15046	10280	12663	3.51	7.98E-014	
63	1043124..1043202	45	K-turn mRNA	could be H/ACA box sRNA; in L7Ae data sets only 2nd half enriched	clip_peak_1002	1043108	1043243	+	1488	50260	60750	55505	5.06	4.94E-031	ATTCAAATAATCAGTAAACGCTGGGATGCTGATAAGTGAAGAGAGACCACCGGTCTGAATGTATGATGTTTAGAACAA
64	complement(1046479..1046529)	51	Sac-sR120		clip_peak_3227	1046429	1046547	-	878	12988	6510	9749	3.33	2.57E-010	
65	complement(1062256..1062318)	63	Sac-sR121		clip_peak_3242	1062217	1062354	-	1658	23934	20027	21980	3.62	4.55E-016	
66	1075502..1075560	59	Sac-sR2		clip_peak_1020	1075494	1075609	+	4607	135328	112916	124122	4.61	7.60E-026	
67	complement(1079251..1079312)	62	Sac-sR122		clip_peak_3260	1079214	1079344	-	2037	16200	11060	13630	2.65	3.81E-008	
68	complement(1103840..1103873)	34	sRNA	close to Saci_1299 (23S rRNA)	clip_peak_3291	1103805	1104002	-	3355	25471	19360	22416	2.65	1.25E-008	
69	complement(1103901..1103931)	31	rRNA fragment	rRNA fragment of Saci_1299 (23S rRNA)	clip_peak_3291	1103805	1104002	-	3355	25471	19360	22416	2.65	1.25E-008	
70	complement(1105472..1105523)	52	rRNA fragment	rRNA fragment of Saci_1299 (23S rRNA)	clip_peak_3293	1105419	1105631	-	10208	21533	20911	21222	1.03	0.049987661	
71	complement(1105563..1105601)	39	rRNA fragment	rRNA fragment of Saci_1299 (23S rRNA)	clip_peak_3293	1105419	1105631	-	10208	21533	20911	21222	1.03	0.049987661	
72	complement(1117450..1117506)	57	Sac-sR14		clip_peak_3318	1117388	1117560	-	6073	448003	290321	369162	5.72	6.08E-035	
73	1117684..1117740	57	Sac-sR3		clip_peak_1046	1117652	1117759	+	1315	34151	20377	27264	4.21	3.52E-018	
74	1122973..1123025	53	Sac-sR25		clip_peak_1047	1122963	1123078	+	1729	62405	44029	53217	4.78	9.92E-026	
75	1126135..1126192	58	Sac-sR123		clip_peak_1054	1126093	1126212	+	6300	143105	117640	130372	4.24	7.44E-022	
76	complement(1131214..1131269)	56	Sac-sR124		clip_peak_3349	1131152	1131295	-	1123	61308	42669	51989	5.35	7.32E-032	
77	complement(1148823..1148947)	125	K-turn mRNA	5' UTR of Saci_1347 (nop5)	clip_peak_3382	1148763	1148962	-	628	3616	3877	3747	2.50	5.10E-008	GAGAATTGATGAAGATATATTTAGTGGAACATGCAATAGGATCGTTTGGATATGACGAGAGCGGAAAATTAATAGATTTTGTACCAAATAGTAAAGATATTGGAAAAGTTACTGAAGCTTTAATT
78	1152391..1152447	57	Sac-sR125	upstream of Sac-sR126	clip_peak_1096	1152339	1152554	+	31256	356521	364095	360308	3.42	5.22E-015	
79	1152452..1152513	62	Sac-sR126	preceded by Sac-sR125	clip_peak_1096	1152339	1152554	+	31256	356521	364095	360308	3.42	5.22E-015	
80	1165288..1165344	57	Sac-sR127		clip_peak_1108	1165280	1165383	+	634	8934	4642	5788	3.08	1.36E-010	
81	1172734..1172794	61	Sac-sR128		clip_peak_1127	1172728	1172857	+	906	6079	5552	5816	2.60	1.07E-008	
82	complement(1175493..1175520)	28	Sac-sR26	in L7Ae data sets only 2nd half enriched	clip_peak_3415	1175446	1175580	-	577	8733	6483	7608	3.60	4.57E-015	
83	1212722..1212778	57	Sac-sR129		clip_peak_1176	1212675	1212828	+	26109	545546	395572	470559	4.04	8.12E-019	
84	complement(1250761..1250815)	53	K-turn mRNA	coding sequence of Saci_1468	clip_peak_3539	1250696	1250894	-	2018	7039	7707	7373	1.81	1.44E-004	ATCCGGCAGAATTCAAGCACAAATAACTGATGATGAACTAAAACAAATATTAGCT
85	complement(1250990..1251036)	47	K-turn mRNA	5' UTR of Saci_1468	clip_peak_3538	1250928	1251091	-	1943	9139	15407	12273	2.56	4.86E-007	GAATATGATGAAGAGTTAGATGAACTTCTGAAAAAAAGAGCAACAGA
86	1293914..1294038	125	5S rRNA		clip_peak_1269	1293931	1294044	+	18756	46272	34702	40487	1.08	0.047326759	
87	1297690..1297726	37	K-turn mRNA	5' UTR of Saci_1520 (l7ae)	clip_peak_1279	1297643	1297782	+	770	43638	22760	33199	5.21	3.65E-025	GCGTGACGAAAAAGAAGGAGGATGAACGCTGTGTGTCTA
88	1308814..1308897	84	mRNA fragment	coding sequence of Saci_1534	clip_peak_1290	1308778	1308925	+	55	3493	2674	3084	5.62	1.16E-034	
89	1388930..1388988	59	Sac-sR19		clip_peak_1447	1388887	1389011	+	1111	14361	10626	12494	3.38	2.03E-013	
90	1400391..1400449	59	Sac-sR15	RNA contains an extended 3' end	clip_peak_1459	1400368	1400560	+	5900	42043	33055	37549	2.59	2.30E-008	
91	1400450..1400503	54	K-turn sRNA	transcribed with Sac-sR15; could be C/D box sRNA	clip_peak_1459	1400368	1400560	+	5900	42043	33055	37549	2.59	2.30E-008	TGAAAGGTGATTAGTGAGGTGATCGCTGATAAATGATGAAAGTGTTAAAACTCT
92	1649007..1649063	57	Sac-sR29		clip_peak_1616	1649001	1649096	+	395	10722	7631	9177	4.39	1.38E-021	
93	1739282..1739381	100	K-turn mRNA	coding sequence of Saci_1928	clip_peak_1644	1739278	1739410	+	316	5157	5081	5119	3.90	8.89E-019	AAGGATGATGAGACTGTTAGGGAGATAGCTAGGACAGCACTGGAGGTGAACGCAAAGTATAAATGCTCAATAAACTTCCACGTTGCAAGTGAAAACGATA
94	1739540..1739639	100	K-turn mRNA	coding sequence of Saci_1928	clip_peak_1643	1739528	1739668	+	851	27484	31202	29343	4.96	2.92E-030	AAGGATGATGAGACTGTTAGGGAGATAGCTAGGACAGCACTGGAGGTGAACGCAAAGTATAAATGCTCAATAAACTTCCACGTTGTCGAAGAAATGAAC
95	complement(1744951..1745018)	68	Sac-sR130		clip_peak_3889	1744940	1745025	-	679	22279	14567	18423	4.60	1.11E-022	
96	complement(1752050..1752144)	95	mRNA fragment	coding sequence of Saci_1941 and Saci_1940	clip_peak_3902	1752007	1752163	-	2223	26257	13420	19839	3.03	9.00E-009	
97	1764690..1764803	114	K-turn mRNA	coding sequence of Saci_1951	clip_peak_1656	1764670	1764863	+	1214	26395	13345	19870	3.87	7.73E-014	TAAAGTGGGCTAAGACCAGCCCAGTGGTCTACCGCTGGACGAGTGGAGCGGGGTGGGTGCTTAATGCATCCACTAGCTATGAAGCGATGAGAGTGAAGGCGGTAAACCATAAAC
98	1830280..1830314	35	K-turn sRNA	might be antisense to 3' UTR of Saci_2016	clip_peak_1695	1830242	1830369	+	271	6665	4691	5678	4.24	5.99E-020	AAACATGATGAGAAGTGTGGGGCTCCGCAATACAA
99	1837726..1837818	93	mRNA fragment	coding sequence and 3' UTR of Saci_2023	clip_peak_1717	1837691	1837854	+	1850	19441	9694	14568	2.85	1.00E-007	
100	1842184..1842221	38	K-turn mRNA	coding sequence of Saci_2027	clip_peak_1723	1842171	1842235	+	409	6901	5581	6241	3.81	2.38E-017	ACTGATGATGACGCTATACCCTCTGACACGTGGATCCA
101	complement(1993082..1993121)	40	Sac-sR21		clip_peak_4018	1993018	1993159	-	4491	32505	29737	31121	2.71	1.57E-009	
102	complement(2175743..2175802)	60	Sac-sR8		clip_peak_4153	2175689	2175848	-	9341	106045	71496	88771	3.14	3.61E-011	
103	2179505..2179563	59	Sac-sR18		clip_peak_1988	2179483	2179591	+	7434	137606	88970	113288	3.79	1.21E-015	
104	2196893..2196949	57	Sac-sR16	Sac-sR16 and Sac-sR132 are linked by a 39 nt RNA	clip_peak_1991	2196873	2197073	+	2423	43537	29077	36307	3.77	9.84E-016	
105	2196939..2196987	49	K-turn sRNA	transcribed with Sac-sR16 and Sac-sR132; could be C/D box sRNA	clip_peak_1991	2196873	2197073	+	2423	43537	29077	36307	3.77	9.84E-016	ACTGATGATGATGTGGGAAGAGCCAGATTAAGATGAAGACCTCTCCTAT
106	2196988..2197045	58	Sac-sR132	Sac-sR16 and Sac-sR132 are linked by a 39 nt RNA	clip_peak_1991	2196873	2197073	+	2423	43537	29077	36307	3.77	9.84E-016	
107	complement(2217710..2217166)	57	Sac-sR4		clip_peak_4221	2217058	2217170	-	2909	18602	15613	17107	2.48	7.08E-008	

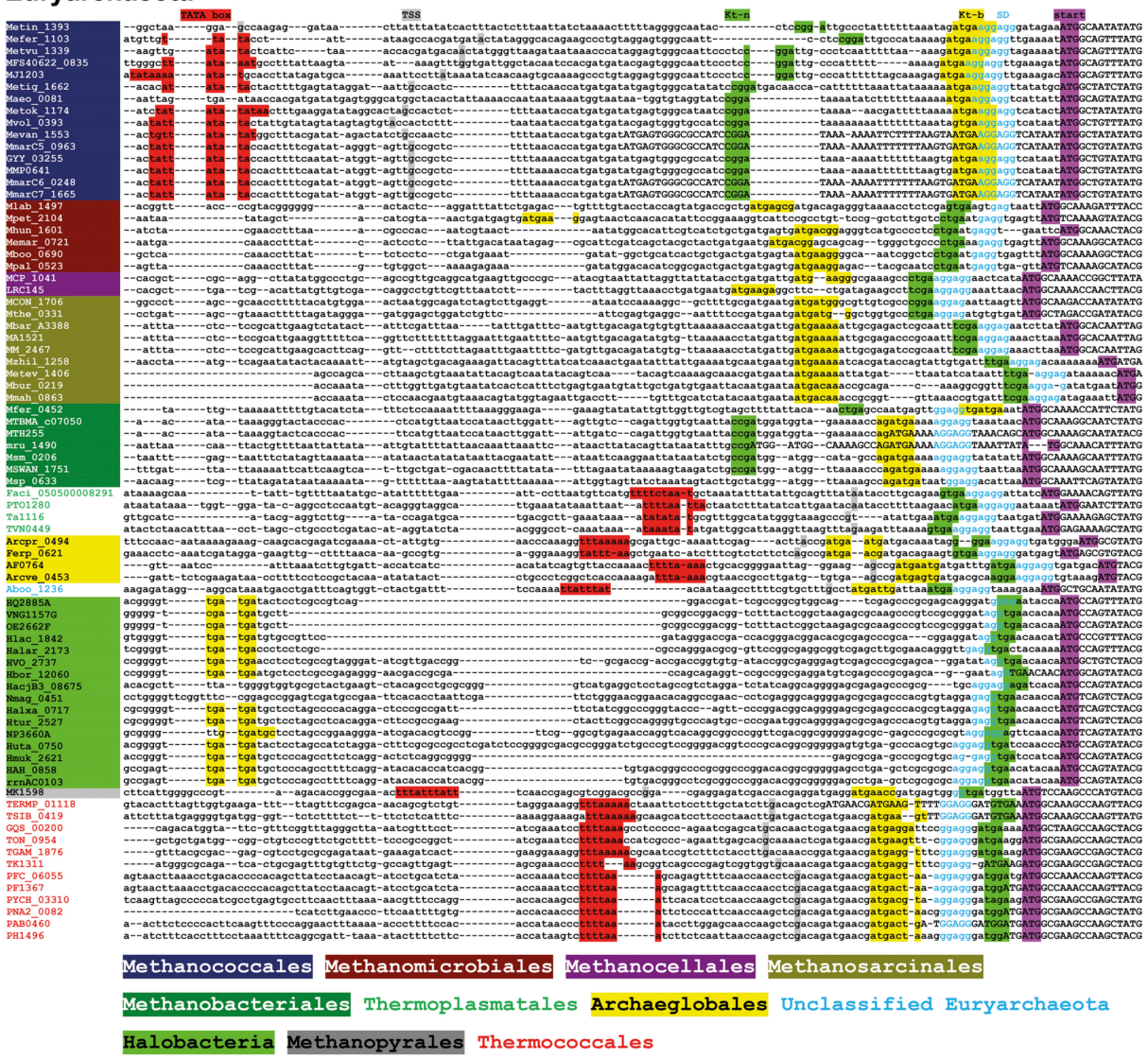
Appendix 3. Multiple sequence alignment of 121 archaeal *l7ae* promoter sequences.



The alignment for the phylum Euryarchaeota is found on the next page.

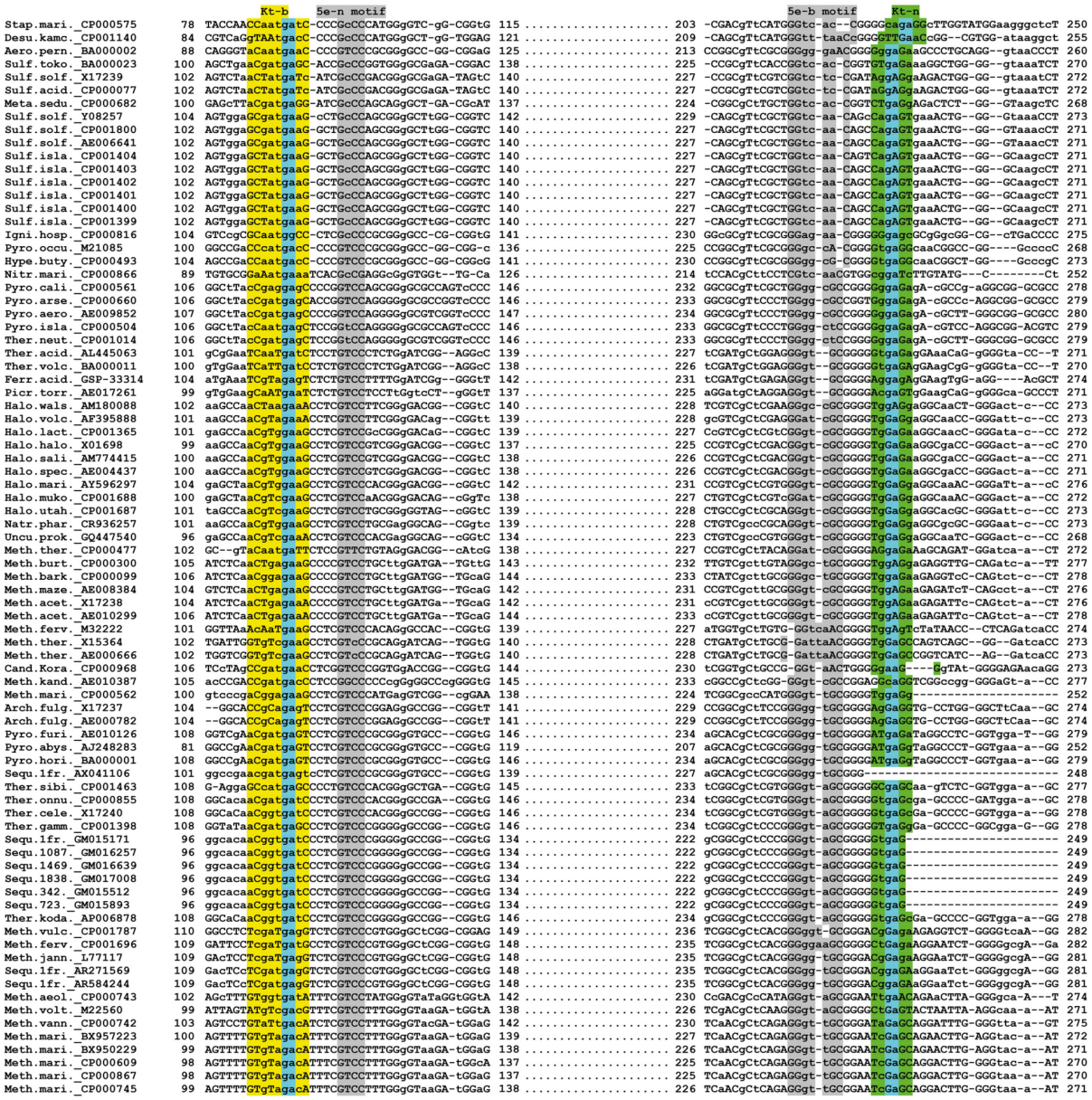
Appendix 3. Multiple sequence alignment of 121 archaeal *l7ae* promoter sequences.

Euryarchaeota



The alignments for the phyla Crenarchaeota, Thaumarchaeota and of other archaea are found on the previous page.

Appendix 4. Multiple sequence alignment of 82 archaeal SRP RNAs.



Appendix 5. Mass spectrometry analysis of Ni-NTA-purified LSm2-CHis.

Number	Accession	Description	Score	Coverage	# Proteins	# Unique Peptides	# Peptides	# PSMs	# AAs	MW [kDa]	Remarks
1	68567724	universally conserved protein [Sulfolobus acidocaldarius DSM 639]	822,6	77,80%	2	36	89	250	482	53,4	
2	68566818	pyridine nucleotide-disulphide oxidoreductase [Sulfolobus acidocaldarius DSM 639]	224,81	63,90%	2	13	29	66	385	43,2	
3	68567947	GTP-binding protein [Sulfolobus acidocaldarius DSM 639]	218,4	85,33%	1	14	40	68	259	29,9	
4	K2C1_HUMAN-Keratin	cytoskeletal 1 Homo sapiens	196,92	44,10%	1	15	39	64	644	66	
5	13813782	Orotate phosphoribosyltransferase (OPRT) (pyrE) [Sulfolobus solfataricus P2]	194,23	79,49%	1	11	27	61	195	21,6	
6	68567980	aspartate carbamoyltransferase [Sulfolobus acidocaldarius DSM 639]	179,62	48,49%	1	12	28	60	299	34,1	
7	68568681	(R)-citramalate synthase [Sulfolobus acidocaldarius DSM 639]	179,4	55,07%	1	13	34	58	523	58,1	
8	68567503	conserved Archaeal protein [Sulfolobus acidocaldarius DSM 639]	163,85	74,07%	2	9	24	49	135	15,8	
9	68567154	conserved Archaeal radical SAM superfamily protein [Sulfolobus acidocaldarius DSM 639]	163,81	48,29%	3	8	40	68	584	67,9	
10	68566608	iron dependent repressor [Sulfolobus acidocaldarius DSM 639]	142,62	62,15%	2	12	23	38	214	24,5	
11	68567121	thermosome beta subunit [Sulfolobus acidocaldarius DSM 639]	141,15	43,94%	2	6	28	38	553	60,1	
12	68567144	DNA directed RNA polymerase subunit B [Sulfolobus acidocaldarius DSM 639]	138,42	28,15%	4	10	32	42	1126	126,4	
13	68567802	thermosome alpha subunit [Sulfolobus acidocaldarius DSM 639]	138,21	54,30%	3	11	33	43	558	59,8	
14	68567296	tRNA-intron endonuclease [Sulfolobus acidocaldarius DSM 639]	136,23	59,12%	1	11	23	44	181	20,9	
15	68566887	precorrin-8X methylmutase [Sulfolobus acidocaldarius DSM 639]	123,57	76,78%	1	7	22	40	211	23,9	
16	68568521	conserved protein [Sulfolobus acidocaldarius DSM 639]	116,79	50,00%	1	9	18	35	248	29,1	
17	68568712	conserved Archaeal protein [Sulfolobus acidocaldarius DSM 639]	111,34	61,14%	2	10	25	32	422	47,3	
18	68566927	universally conserved ATP-binding protein [Sulfolobus acidocaldarius DSM 639]	110,61	49,49%	3	12	19	32	295	32,1	
19	68567137	30S ribosomal protein S7P [Sulfolobus acidocaldarius DSM 639]	99,27	78,97%	3	6	21	33	195	22	
20	keratin_10_epidermolytic	hyperkeratosis Homo sapiens	93,97	32,05%	1	6	25	34	624	63,3	
21	68567793	universally conserved protein [Sulfolobus acidocaldarius DSM 639]	92,73	39,04%	2	7	20	30	479	54,3	
22	68567240	Sm-like protein [Sulfolobus acidocaldarius DSM 639]	89,92	93,10%	1	7	17	31	87	9,8	LSm2 (Saci_0799)
23	68567136	elongation factor 1-alpha [Sulfolobus acidocaldarius DSM 639]	89,05	47,59%	3	8	23	33	435	48,2	
24	68567579	HTH domain protein [Sulfolobus acidocaldarius DSM 639]	87,86	42,41%	1	7	20	36	349	40,3	
25	68567170	cytidylyltransferase [Sulfolobus acidocaldarius DSM 639]	87,48	59,65%	1	8	15	26	171	19,9	
26	68568347	hypothetical protein Saci_1974 [Sulfolobus acidocaldarius DSM 639]	87,15	48,17%	1	6	14	26	301	35,1	
27	68567641	hypothetical protein Saci_1224 [Sulfolobus acidocaldarius DSM 639]	81,6	73,42%	2	7	12	25	79	8,5	LSm1 (Saci_1224)
28	68567018	TBP-interacting proten [Sulfolobus acidocaldarius DSM 639]	80,18	32,96%	1	4	15	30	452	50,4	
29	68567098	hypothetical metallo-beta-lactamase superfamily protein [Sulfolobus acidocaldarius DSM 63	76,48	31,85%	2	6	16	26	631	70,8	
30	68567764	serine hydroxymethyltransferase [Sulfolobus acidocaldarius DSM 639]	74,5	44,57%	1	3	17	24	433	48,3	
31	68566852	conserved protein [Sulfolobus acidocaldarius DSM 639]	73,88	42,19%	1	5	19	26	320	37,5	
32	68566579	30S ribosomal protein S4P [Sulfolobus acidocaldarius DSM 639]	72,57	60,23%	1	3	14	29	176	20,4	
33	68567849	50S ribosomal protein L10P [Sulfolobus acidocaldarius DSM 639]	70,08	34,03%	2	8	12	18	335	36,5	
34	68566886	precorrin-8X methylmutase [Sulfolobus acidocaldarius DSM 639]	69,74	22,06%	1	7	11	19	340	38,4	
35	68566820	conserved protein [Sulfolobus acidocaldarius DSM 639]	69,2	46,73%	1	7	16	22	306	34,2	
36	68567969	conserved Archaeal protein [Sulfolobus acidocaldarius DSM 639]	67,47	39,67%	2	4	19	25	489	56,3	
37	68567228	conserved Archaeal protein [Sulfolobus acidocaldarius DSM 639]	62,72	40,80%	1	7	13	19	326	37,3	
38	68567160	conserved Prokaryal protein [Sulfolobus acidocaldarius DSM 639]	60,53	33,83%	2	6	17	20	467	51,9	
39	68567663	conserved protein [Sulfolobus acidocaldarius DSM 639]	59,19	33,67%	1	3	16	20	398	46,1	
40	68568092	dihydroxyacid dehydratase/phosphogluconate dehydratase [Sulfolobus acidocaldarius DSM 6	58,71	27,99%	1	5	14	18	561	59,8	
41	68566816	conserved protein [Sulfolobus acidocaldarius DSM 639]	58,21	29,68%	1	3	17	25	438	50,9	
42	68567054	30S ribosomal protein S3P [Sulfolobus acidocaldarius DSM 639]	57,3	61,04%	1	3	14	18	231	25,3	
43	68566872	3-hydroxyisobutyrate dehydrogenase [Sulfolobus acidocaldarius DSM 639]	56,44	54,36%	1	5	12	17	287	31,3	
44	68566807	formate dehydrogenase alpha chain [Sulfolobus acidocaldarius DSM 639]	55	19,57%	1	4	14	17	976	109	
45	68567741	conserved Archaeal protein [Sulfolobus acidocaldarius DSM 639]	54,92	53,15%	1	4	12	17	254	29,3	
46	68567792	ABC transporter [Sulfolobus acidocaldarius DSM 639]	53,27	60,89%	2	3	14	19	248	27,4	
47	68566553	DNA double-strand break repair Mre11 nuclease [Sulfolobus acidocaldarius DSM 639]	53,01	23,82%	1	4	9	18	382	44,1	
48	68566764	conserved Archaeal protein [Sulfolobus acidocaldarius DSM 639]	52,64	32,69%	2	3	10	15	309	34,8	
49	68568505	glutamine synthetase [Sulfolobus acidocaldarius DSM 639]	51,69	35,27%	1	3	13	18	431	49,3	
50	68568069	alcohol dehydrogenase [Sulfolobus acidocaldarius DSM 639]	51,03	47,13%	1	3	14	18	331	36,5	
51	68568370	ABC transporter, ATPase component [Sulfolobus acidocaldarius DSM 639]	49,94	55,19%	1	3	15	18	270	30,7	
52	68567044	50S ribosomal protein L6P [Sulfolobus acidocaldarius DSM 639]	49,84	55,91%	1	4	13	16	186	20,9	
53	68567350	carbon monoxide dehydrogenase medium chain [Sulfolobus acidocaldarius DSM 639]	49,23	44,64%	2	4	14	17	280	30,9	
54	13813422	Thermosome beta subunit(thermophilic factor 55) (ring complex beta subunit)(chaperonin be	49,1	22,62%	1	1	10	14	557	60,3	
55	68566873	conserved protein [Sulfolobus acidocaldarius DSM 639]	49,01	61,72%	1	6	12	17	128	14,8	
56	68566755	conserved protein [Sulfolobus acidocaldarius DSM 639]	48,99	42,27%	1	2	14	20	220	25,3	
57	68567617	glucose-1-phosphate adenylyltransferase [Sulfolobus acidocaldarius DSM 639]	48,64	28,06%	1	2	12	15	417	48,1	
58	68566702	CBS domain protein [Sulfolobus acidocaldarius DSM 639]	48,5	31,56%	1	2	13	18	377	42,9	
59	68568253	hypothetical protein Saci_1877 [Sulfolobus acidocaldarius DSM 639]	48,03	36,93%	1	4	11	16	287	33,4	
60	68567033	conserved Archaeal protein [Sulfolobus acidocaldarius DSM 639]	47,81	32,42%	1	5	12	14	182	20,7	
61	68567208	thiamine pyrophosphate-requiring enzyme [Sulfolobus acidocaldarius DSM 639]	47,6	34,63%	1	5	14	15	540	59,9	
62	68566685	[Fe] superoxide dismutase [Sulfolobus acidocaldarius DSM 639]	47,32	56,40%	2	4	14	17	211	24,3	
63	68566756	putative inactivated transposase [Sulfolobus acidocaldarius DSM 639]	46,81	48,72%	1	3	12	15	273	31,1	
64	68567848	50S ribosomal protein L12P [Sulfolobus acidocaldarius DSM 639]	46,73	71,43%	2	4	7	12	105	11,1	
65	68566621	conserved protein [Sulfolobus acidocaldarius DSM 639]	45,58	23,98%	1	3	15	15	738	85,2	
66	68568626	carbon monoxide dehydrogenase medium chain [Sulfolobus acidocaldarius DSM 639]	45,31	44,84%	1	3	11	16	281	30,8	
67	68566666	50S ribosomal protein L10E [Sulfolobus acidocaldarius DSM 639]	44,63	53,41%	1	4	12	17	176	19,9	
68	68567938	30S ribosomal protein S26E [Sulfolobus acidocaldarius DSM 639]	44,55	48,42%	1	3	13	17	95	10,8	
69	68567621	thermosome gamma subunit [Sulfolobus acidocaldarius DSM 639]	43,83	35,81%	1	2	16	17	539	58,5	
70	68566580	30S ribosomal protein S11P [Sulfolobus acidocaldarius DSM 639]	43,47	56,06%	2	3	7	12	132	14,1	
71	68567059	50S ribosomal protein L4E [Sulfolobus acidocaldarius DSM 639]	43,22	51,88%	2	2	19	26	266	29,1	
72	68567462	ABC transporter [Sulfolobus acidocaldarius DSM 639]	42,67	35,56%	1	5	12	14	419	47	
73	68566939	DNA/pantothenate metabolism flavoprotein [Sulfolobus acidocaldarius DSM 639]	42,37	20,29%	1	3	8	13	409	45,5	
74	68567373	2-isopropylmalate synthase [Sulfolobus acidocaldarius DSM 639]	42,16	24,09%	2	3	10	14	386	42,2	
75	68566741	3-isopropylmalate dehydratase large subunit [Sulfolobus acidocaldarius DSM 639]	42,07	26,81%	1	2	10	13	414	44,6	
76	68567239	S-adenosylmethionine synthetase [Sulfolobus acidocaldarius DSM 639]	41,94	36,97%	2	4	11	14	403	44,1	
77	H3SSU7	60 kDa chaperonin OS=Pseudomonas aeruginosa MPAO1/P1 GN=groL PE=3 SV=1 - [H3S	41,15	31,26%	2	5	12	15	547	57	
78	68567354	GTP-binding protein 1 [Sulfolobus acidocaldarius DSM 639]	40,83	21,48%	1	2	9	13	526	58,9	
79	68567126	conserved Archaeal protein [Sulfolobus acidocaldarius DSM 639]	40,52	34,81%	2	4	10	13	385	42,6	
80	68567753	NOP56/58-like protein [Sulfolobus acidocaldarius DSM 639]	40,47	35,44%	3	4	12	14	412	46,6	Nop5 (Saci_1347) - C/D box sRNP subunit
81	68568208	conserved Archaeal protein [Sulfolobus acidocaldarius DSM 639]	40,37	41,55%	1	4	7	12	207	24,5	
82	68567269	30S ribosomal protein S6E [Sulfolobus acidocaldarius DSM 639]	39,52	36,62%	1	4	9	14	213	23,7	
83	68567728	DNA binding protein alba-2 [Sulfolobus acidocaldarius DSM 639]	39,09	70,10%	1	1	15	18	97	10,5	
84	68568611	aldose reductase [Sulfolobus acidocaldarius DSM 639]	38,89	25,86%	1	4	7	12	290	32,9	
85	68567972	transcriptional regulator [Sulfolobus acidocaldarius DSM 639]	38,45	41,29%	3	4	8	18	155	17,8	
86	68567979	aspartate carbamoyltransferase regulatory chain [Sulfolobus acidocaldarius DSM 639]	37,74	50,00%	1	1	11	19	164	18,3	
87	68567181	conserved Archaeal protein [Sulfolobus acidocaldarius DSM 639]	37,27	25,45%	1	2	11	15	393	45,3	
88	68567932	V-type ATP synthase alpha chain [Sulfolobus acidocaldarius DSM 639]	37,21	26,35%	3	4	12	13	592	66,1	
89	68566573	conserved Archaeal protein [Sulfolobus acidocaldarius DSM 639]	37,09	32,77%	2	1	10	14	412	45,7	DnaG (Saci_0075) - associates with exosome
90	68567050	50S ribosomal protein L14P [Sulfolobus acidocaldarius DSM 639]	36,6	44,20%	1	5	6	9	138	15,2	
91	68567073	ribonuclease PH [Sulfolobus acidocaldarius DSM 639]	35,09	41,15%	3	1	11	16	243	27	Rrp41 (Saci_0610) - archaeal exosome subunit
92	68567762	glyceraldehyde-3-phosphate dehydrogenase [Sulfolobus acidocaldarius DSM 639]	34,75	26,24%	1	3	11	13	343	37,9	
130	68567072	exosome complex exonuclease RRP43-like protein [Sulfolobus acidocaldarius DSM 639]	24,32	28,26%	1	2	6	8	276	30,5	Rrp42 (Saci_0609) - archaeal exosome subunit
131	68567410	conserved Archaeal protein [Sulfolobus acidocaldarius DSM 639]	24,17	14,63%	1	2	6	6	451	51,3	
135	68567752	fribrillarin-like pre-rRNA processing protein [Sulfolobus acidocaldarius DSM 639]	23,19	32,90%	1	2	6	7	231	26,4	fibrillarin (Saci_1346) - C/D box sRNP subunit
136	68567045	30S ribosomal protein S8P [Sulfolobus acidocaldarius DSM 639]	22,77	61,65%	1	1	9	11	133	14,9	
150	68567074	conserved Archaeal KH domain protein [Sulfolobus acidocaldarius DSM 639]	18,91	16,06%	1	2	4	5	249	27,7	Rrp4 (Saci_0611) - archaeal exosome subunit
151	68568267	hypothetical protein Saci_1893 [Sulfolobus acidocaldarius DSM 639]	18,83	9,64%	1	2	4	5	394	44,5	
366	68566594	conserved protein [Sulfolobus acidocaldarius DSM 639]	3,25	3,88%	1	1	1	1	361	40,2	

Abgrenzung der Eigenleistung

Die in dieser Arbeit präsentierten Ergebnisse wurden von mir selbstständig ohne andere als die hier aufgeführte Hilfe durchgeführt. Im Folgenden werden weitere an dieser Arbeit beteiligten Personen sowie deren experimentellen Beiträge genannt:

Jörg Kahnt (Massenspektrometrie-Einrichtung des Max-Planck-Institutes in Marburg)
Hat die massenspektrometrische Analyse der L7Ae- und LSm-Aufreinigungen durchgeführt und die Tabelle aus Appendix 5 bereitgestellt.

Michael Uhl (AG Backofen, Albert-Ludwigs-Universität Freiburg)
Hat als Kollaborationspartner die bioinformatische Auswertung (Peak-Calling-Analyse und MEME-Analyse) der L7Ae- und LSm-RIP-Seq-Studien durchgeführt, um die Bindepartner der L7Ae- und LSm-Proteine zu ermitteln und die potentiellen Bindemotive der LSm-Proteine zu identifizieren. Michael Uhl hat die Sequenzlogos in Abbildung 2.27a und 2.27b bereitgestellt.

Lydia Seelos (Bachelor-Studentin)
Hat im Rahmen ihrer Bachelorarbeit eine *Sulfolobus acidocaldarius* Deletionsmutante für die C/D box sRNA Sac-sR10 hergestellt (ΔSac-sR10).

Prof. Michelle Meyer (Department of Biology, Boston College, MA, USA)
Hat im Rahmen einer Kollaboration eine *in silico*-Faltung der putativen k-turn-Motive in den 121 archaealen *l7ae*-Promotorregionen durchgeführt.

Alexander Pastura (Master-Student)
Hat im Rahmen seiner Masterarbeit eine Interaktionsstudie der Proteine LSm1 und LSm2 durchgeführt und die Bindung des LSm1/2-Heterokomplexes an die von Michael Uhl identifizierten potentiellen Bindemotive der LSm-Proteine per EMSA analysiert. Die Abbildungen 2.28 und 2.29 wurden von Alexander Pastura bereitgestellt.

Erklärung

Hiermit erkläre ich, dass ich meine Dissertation mit dem Titel „L7Ae- and LSm-RNA interactomes of *Sulfolobus acidocaldarius*“ selbstständig und ohne unerlaubte Hilfe angefertigt und mich dabei keiner anderen als der von mir ausdrücklich bezeichneten Quellen und Hilfsmittel bedient habe.

Diese Dissertation wurde in der jetzigen oder einer ähnlichen Form noch bei keiner anderen Hochschule eingereicht und hat noch keinen sonstigen Prüfungszwecken gedient.

_____ _____

Ort, Datum Michael Daume