Pan-archaeal analysis of C/D box sRNA biogenesis and methylation targets

Dissertation

zur

Erlangung des Doktorgrades

der Naturwissenschaften

(Dr. rer. nat.)

Dem Fachbereich Biologie

der Philipps-Universität Marburg

vorgelegt von

Vanessa Tripp

aus Frankfurt

Marburg/Lahn

April 2016

Die Untersuchungen der vorliegenden Arbeit wurden von November 2012 bis März 2016 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: 04.05.2016

Erstgutachter: Dr. Lennart Randau
Zweitgutachter: Prof. Dr. Torsten Waldminghaus

Tag der mündlichen Prüfung: 02.06.2016

Teile dieser Arbeit wurden in folgendem Artikel veröffentlicht:

Dennis, P.*, Tripp, V.*, Lui, L, Lowe, T. and Randau, L. (2015) *C/D box sRNA-guided 2'-O-methylation patterns of archaeal rRNA molecules.* BMC Genomics, 16:632 (* shared first authorship)

Weitere Publikationen:

Plagens, A., Tripp, V., Daume, M., Sharma, K., Klingl, A., Hrle, A.,Conti, E., Urlaub, H. and Randau, L. (2014) *In vitro assembly and activity of an archaeal CRISPR-Cas type I-A Cascade interference complex.* Nucleic Acids Res, 42, 5125-5138.

Su, A.A.H, Tripp, V. and Randau, L. (2013) *RNA-Seq analyses reveal the order of tRNA processing events and the maturation of C/D box and CRISPR RNAs in the hyperthermophile Methanopyrus kandleri* Nucleic Acids Res, 41, 6250-6258.

Eidesstattliche Erklärung

Hiermit erkläre ich, dass ich meine Dissertation mit dem Titel “Pan-archaeal analysis of C/D box sRNA biogenesis and methylation targets“ selbstständig und ohne unerlaubte Hilfsmittel angefertigt und ich keine als die von mir ausdrücklich angegebenen Quellen verwendet habe.

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

Marburg, den 01.04.2016

Vanessa Tripp

I Summary

Post-transcriptional modifications of RNA molecules occur in all three domains of life and influence RNA stability and functionality. The most numerous modifications are 2'-*O*-methylations at the ribose moiety and pseudouridylations. In archaea, modified bases are abundant in ribosomal RNAs (rRNAs) and transfer RNAs (tRNAs). The introduction of both modifications is guided by small RNAs that are incorporated into ribonucleoprotein complexes (RNPs). 2'-*O*-methylations are guided by C/D box sRNAs in archaea. C/D box sRNAs are characterized by the conserved sequence elements boxC/C' (consensus sequence: RUGAUGA) and boxD/D' (consensus sequence: CUGA). Upon C/D box sRNA folding, both sequence elements base-pair (boxC with boxD and boxC' with boxD'), which results in the formation of two kink-turn motifs that are stabilized by binding of the protein L7Ae. The sequences between the two kink-turn elements show complementarity to the sequences of the target RNA and thereby serve as guide sequences that determine the sites of 2'-*O*-methylation. The modifications are introduced site-specifically at the nucleotide of the target RNA that is complementary to the fifth nucleotide upstream of the boxD/D' motif by the methyltransferase fibrillarin.

Based on the guide sequences, C/D box sRNA targets of seven archaea were predicted and mapped onto the consensus structure of the 16S and 23S rRNA. Conserved methylation hotspots were observed in ancient core regions of the rRNAs that are important for ribosome integrity and functionality and that are not protected by ribosomal proteins. Therefore, the modifications might contribute to the folding, structural stabilization and function of the rRNAs.

The biogenesis of archaeal C/D box sRNAs is largely unknown as independent promoters cannot be identified for the majority of the C/D box sRNA genes. The analysis of C/D box sRNA genes in six archaeal model organisms revealed diverse genetic contexts, providing opportunities for transcription without the necessity of an independent promoter. C/D box sRNA genes localize e.g. in the 5' or 3'-UTR of flanking protein-coding regions and polycistronic C/D box sRNA transcripts exist. Plasmid-based C/D box sRNA *in vivo* analyses were performed in *Sulfolobus acidocaldarius* in which C/D box sRNA genes variants with their native or random upstream and downstream sequences were used to identify C/D box sRNA stabilization and maturation requirements. The analyses revealed that the maturation of C/D box sRNAs occurs independently of the upstream and downstream sequences. The integrity of the k-turn is important for C/D box sRNA stability. Archaeal C/D box sRNAs exhibit a transcriptional plasticity and their maturation is suggested to include the action of unspecific exoribonucleases. Complete degradation might be prevented by co-transcriptional L7Ae binding or complete C/D box sRNP assembly.

Circular forms of the C/D box sRNAs were identified in several hyperthermophilic archaea and the circularization reaction should protect RNAs from degradation. C/D box sRNA gene upstream and downstream sequences were shown not to be required for circularization but the responsible RNA ligase remains to be identified.

Thus, this thesis provides insights into the transcription and maturation of archaeal C/D box sRNAs and highlights conserved 2'-*O*-methylation pattern in archaeal rRNAs.

II Zusammenfassung

Posttranskriptionale Modifikationen sind in RNA Molekülen von allen drei Domänen des Lebens zu finden und sie beeinflussen die Stabilität und Funktionalität der RNAs. Die häufigsten Modifikationen sind 2'-*O*-Methylierungen der Ribose oder Pseudouridylierungen. In Archaeen kommen sie am häufigsten in ribosomalen RNAs (rRNAs) und Transfer-RNAs (tRNAs) vor. Die Position in der diese Modifikationen eingefügt werden, wird durch kleine RNAs bestimmt die an einen Ribonukleoproteinkomplex (RNP) gebunden sind. Für die Positionsbestimmung der 2'-*O*-Methylierungen in den Ziel-RNAs sind in Archaeen C/D box sRNAs zuständig. Charakteristische Sequenzmotive der C/D box sRNAs sind die BoxC/C' (Konsensussequenz: RUGAUGA) und BoxD/D' (Konsensussequenz: CUGA). Während der C/D box sRNA Faltung bilden diese Motive Basenpaare (BoxC mit BoxD sowie BoxC' mit BoxD') und dadurch entstehen zwei Kink-turn Motive, die durch die Bindung von L7Ae stabilisiert werden. Die Sequenzen zwischen den BoxC und BoxD' bzw. BoxC' und BoxD Motiven weisen Komplementarität zu den Sequenzen der Ziel-RNA auf und dadurch bestimmen sie das Ribonukleotid an dem die Ribose 2'-*O*-methyliert wird. Das Ribonukleotid der Ziel-RNA, das komplementär zum fünften Ribonukleotid stromaufwärts der BoxD/D' Motive ist, wird durch die Methyltransferase Fibrillarin modifiziert.

Basierend auf den zur Ziel-RNA komplementären Bereichen, wurden die C/D box sRNA Ziele von sieben Archaeen vorhergesagt und in den Konsensusstrukturen der 16S und 23S rRNA gekennzeichnet. Konservierte Methylierungshotspots wurden in evolutionär ursprünglichen Kernregionen der rRNAs gefunden, die wichtig für den Ribosomenaufbau und die Ribosomenfunktionalität sind und nicht von ribosomalen Proteinen geschützt werden. Deshalb könnten die Modifikationen zur Faltung, Stabilisierung von Sekundärstrukturen sowie der Funktion der rRNA beitragen.

In Archaeen ist die Biogenese der C/D box sRNAs weitgehend unbekannt, da für die Mehrheit der C/D box sRNA Gene keine eigenen Promotoren identifiziert werden konnten. Die Analyse des genomischen Kontexts von C/D box sRNA Genen in diversen Archaeen hat gezeigt, dass dieser sehr unterschiedlich sein kann. Es wurden verschiedene C/D box sRNA Gen Lokalisierungen gefunden, bei denen die Transkription ohne das Vorhandensein eines eigenen Promotors möglich ist. Es konnten z.B. C/D box sRNA Gene identifiziert werden die in 5' oder 3'-UTRs von flankierenden protein-codierenden Bereichen liegen sowie Cluster von mehreren C/D box sRNA Genen, die polycistronisch transkribiert werden. Um die C/D box sRNA Stabilisierungs- und Reifungs-Voraussetzungen zu identifizieren, wurden Plasmid-basierte *in vivo* Analysen in *Sulfolobus acidocaldarius* durchgeführt, wobei C/D box sRNA Gen Varianten mit den nativen oder beliebigen stromaufwärts und stromabwärts gelegenen Sequenzen genutzt wurden. Die Analysen haben gezeigt, dass die Reifung der C/D box sRNAs unabhängig von den stromaufwärts oder stromabwärts gelegenen Sequenzen ist. Die

Integrität des k-turns ist wichtig für die C/D box sRNA Stabilität. C/D box sRNAs weisen eine transkriptionale Plastizität auf und die Reifung der RNAs scheint die Aktivität von unspezifischen Exoribonukleasen einzuschließen. Die vollständige Degradation der C/D box sRNAs könnte durch die co-transkriptionale Assemblierung von L7Ae oder des kompletten C/D box sRNPs verhindert werden.

In verschiedenen hyperthermophilen Archaeen kommen zudem zirkuläre Formen der C/D box sRNAs vor, die wahrscheinlich einen Schutz vor Degradation darstellen. Es konnte gezeigt werden, dass die stromaufwärts und stromabwärts gelegenen Sequenzen der C/D box sRNA Gene für die Zirkularisierung nicht relevant sind, allerdings muss die für die Zirkularisierung zuständige Ligase noch identifiziert werden.

Somit liefert diese Arbeit Einblicke in die Transkription und Reifung von C/D box sRNAs in Archaeen und zeigt, dass konservierte 2'-*O*-Methylierungsmuster in rRNAs von Archaeen existieren.

III Abbreviations

ad	add to the volume
APS	ammonium persulfate
ATP	adenosine triphosphate
β-Me	β-mercaptoethanol
BHB	bulge-helix-bulge (motif)
bp	basepair
BSA	bovine serum albumin
C/D box sRNA	C/D box sno-like RNA
cDNA	complementary DNA
C-terminal	carboxy-terminal
cpm	counts per minute
CV	column volume
Da	Dalton
ddH2O	two times destilled water
DEPC	diethylpyrocarbonate
DMSO	dimethyl sulfoxide
DNA	deoxyribonucleic acid
DNase	desoxyribonuclease
dNTP	deoxyribonucleotide triphosphate
ds	downstream
dsDNA	double-stranded DNA
DTT	dithiothreitol
e.g.	for example
EDTA	ethylene-diamine-tetraacetic acid
EM	electron microscopy
EMSA	electrophoretic mobility shift assay
et al.	*et alteri* = and others
FPLC	Fast Protein Liquid Chromatography
g	gram
x g	gravitational acceleration
h	hour
H/ACA box sRNA	H/ACA box sno-like RNAs
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
His-tag	histidine tag
IPTG	isopropyl β-D-1-thiogalactopyranoside
kb	kilobase
l	liter
LB	lysogeny broth
M	molar (mol/l)
min	minutes
MW	molecular weight
µ	micro (10^{-6})
n	nano (10^{-9})
Ni^{2+}-NTA	nickel nitriloacetic acid
NMR spectroscopy	nuclear magnetic resonance spectroscopy
nt	nucleotides
N-terminal	amino-terminal
NTP	nucleoside triphosphate
OD_{600nm}	optical density at 600 nm
ORF	open reading frame
PAGE	polyacrylamide gel electrophoresis
PCR	polymerase chain reaction
pH	negative logarithm of the hydrogen ion (H+) concentration
RNA	ribonucleic acid
RNase	ribonuclease
RNP	ribonucleoprotein complex
rpm	rounds per minute
rRNA	ribosomal RNA
RT-PCR	reverse transcription polymerase chain reaction
RT	room temperature
s	seconds
SANS	small-angle neutron scattering
SAXS	small-angle X-ray scattering
SDS	sodium dodecyl sulfate
snRNA	small nuclear RNAs
ssDNA	single-stranded DNA
ssRNA	single stranded RNA
tRNA	transfer RNA
TAE	tris-acetate-EDTA-buffer
TBE	tris-borate-EDTA-buffer
TEMED	N,N,N';N'-tetramethylethylenediamide
Tris	tris-(hydroxymethyl)-aminomethane
U	unit (enzyme activity)
us	upstream

UTR	untranslated region
UV	ultraviolet
W	watt
w/o	without
% (v/v)	percent by volume
% (w/v)	percent by weight
>	higher than
<	lower than
Δ	deletion

Table of contents

1. Introduction

1.1 Small RNA-guided posttranscriptional modifications

Post-transcriptional modifications occur in RNA molecules of all three domains of life. In eukaryotes and archaea, modified nucleotides are most commonly 2'-*O*-methylated at the ribose moiety or pseudouridylated. Target RNAs comprise ribosomal RNAs (rRNA) and transfer RNAs (tRNA) in archaea and small nuclear RNAs (snRNA) are additionally targeted in eukaryotes [1-4]. In eukaryotes, about 50-100 nucleotides are modified in rRNAs [5]. In archaea, a correlation between the growth temperature and the number of modifications was suggested to exist. Mesophilic archaea possess only few modifications and thermophilic archaea contain a large number of 2'-*O*-methylations [1, 6-9]. In bacteria, 2'-*O*-methylations and pseudouridylations are rare RNA modifications. *Escherichia coli* rRNAs possess only four 2'-*O*-methylated nucleotides. Base methylations are more common in bacteria and these modifications are introduced in RNA-independent fashion by site or region specific enzymes [5]. These enzymes do not exist in archaea and eukaryotes. Instead, site-specific 2'-*O*-methylations and pseudouridylations are introduced by RNA-guided mechanisms. RNAs that exhibit complementary sequences to the target RNAs serve as guide molecules in RNA-protein complexes and direct their modification activity towards their targets [1, 10, 11].

Two major families of guide RNAs exist, which are termed C/D box snoRNAs and H/ACA box snoRNAs in eukaryotes. Their archaeal homologues are termed C/D box sno-like RNAs (C/D box sRNA) and H/ACA box sno-like RNAs (H/ACA box sRNA). C/D box s(no)RNAs guide the 2'-*O*-methylation of target nucleotides and H/ACA box s(no)RNAs guide the conversion of uridine to pseudouridine [1, 3, 12, 13]. Both modifications influence the stability of the RNA targets and they protect RNA from ribonucleolytic cleavage. Additionally, rRNA folding may be influenced as several C/D box sRNAs exist that target nucleotides that are apart in the rRNA sequence but close in the rRNA secondary structure. Thus, C/D box s(no)RNAs might act as RNA chaperones [14-19].

Analyses of all modified nucleotides in the rRNAs of yeast and *E. coli* showed that most modifications occur in functional important regions, including the peptidyl transferase center, the subunit-subunit interaction sites or interaction sites of tRNA and mRNA [5]. Modifications of most specific nucleotides do not seem to be conserved. Only loss of the single modification at position U2552 close to the peptidyl transferase center could be shown to lead to growth and translational defects in *E. coli* [20]. In contrast, clusters of modifications seem to be important as yeast cells lacking modification clusters are impaired in growth. Yeast cells lacking all modifications are not viable [7, 21-23]. Additionally, partial modification of individual rRNA nucleotides could be observed in organisms of all three domains of life. Therefore, the modifications add to the heterogeneity of rRNAs [24-26].

In eukaryotes, rRNAs are often transcribed as precursor molecules and few snoRNAs were identified that do not guide modification but are required for pre-rRNA processing. They assist in cleavage site determination as they support rRNA folding [27, 28]. Additionally, orphan snoRNAs that do not show complementarity to known target RNAs were identified in archaea and eukaryotes. Their function remains to be investigated [1, 29, 30].

1.2 Features of C/D box s(no)RNAs

C/D box s(no)RNAs have a length of approximately 50-100 nt in eukaryotes and 50-70 nt in archaea. Eukaryotic C/D box snoRNAs that are involved in rRNA maturation and folding show extended complementarity to rRNA sequences and can be up to 300 nt long [31-33].

C/D box s(no)RNAs are characterized by conserved sequence elements and a characteristic secondary structure. The conserved sequence elements are boxC (consensus sequence: RUGAUGA) and boxD (consensus sequence: CUGA) which are situated close to the 5' and 3' terminus, respectively. Both elements can be duplicated in the central region (boxC', boxD') but, in eukaryotes, these two boxes are less conserved [34, 35]. Upon C/D box s(no)RNA folding, the boxC and boxD sequences base-pair resulting in a helix-internal loop-helix structure termed kink-turn (k-turn) [36, 37]. A terminal helix of the k-turn motif that is formed by the sequences upstream of boxC and downstream of boxD, respectively (helixI), is common for eukaryotic C/D box snoRNAs, but is absent in most archaeal C/D box sRNAs [1, 8, 9, 17]. The sequences of the boxC and boxD are part of the asymmetric internal loop with three unpaired nucleotides at one side and the internal helix (helixII). The first uridine of boxC is part of the internal loop. The following guanosine and adenosine form sheared GA base pairs with the corresponding nucleotides of boxD. The helixII also contains a U-U mismatch and a Watson-Crick base pair formed by nucleotides from boxC and D (figure 1.1) [37]. The k-turn motif can be also found in different other RNAs e.g. rRNAs, H/ACA box s(no)RNAs, human ribonuclease P and untranslated regions of mRNA including several riboswitches [37-43]. Typically, a second version of the k-turn motif can be found in C/D box s(no)RNAs. It results of base pairing between the boxC' and boxD' sequences. The terminal helix is replaced by a loop which coined the term k-loop (figure 1.1) [44].

The sequences that are located between the boxC and boxD' motif, as well as between the boxC' and boxD motif, respectively, show complementarity to the sequences of the target RNA and serve as guide sequences that determine the sites for the 2'-*O*-methylation reaction. The guide sequences usually have a length of 10-12 nt in archaea and 10-21 nt in eukaryotes. The length of the guide sequences is crucial for the 2'-*O*-methylation reaction and the modification is introduced site-specifically at the nucleotide of the target RNA that is complementary to the fifth nucleotide upstream of the boxD/D' motif [3, 45]. A Watson-Crick base pair is required at this position and extended Watson-Crick base pairs along the guide

RNA-target RNA duplex are crucial for efficient methylation [46-49]. In yeast, the region between the boxC' and boxD' sequences has additional complementarity with the target RNA, which stimulates methylation efficiency [50].

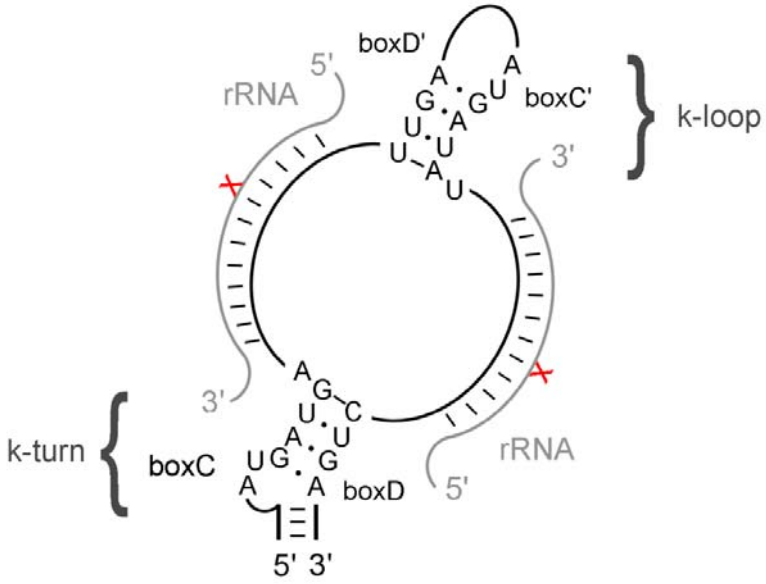

Figure 1.1: Schematic representation of a C/D box s(no)RNA. The conserved boxC/C' and boxD/D' sequences base-pair and k-turn and k-loop structures are formed. The guide regions show complementarity to the target RNA and the nucleotide that is complementary to the fifth nucleotide upstream of the boxD/D' motif becomes 2'-*O*-methylated (red).

1.3 Structure of C/D box s(no)RNPs

All C/D box s(no)RNAs form ribonucleoprotein complexes (RNPs) with three highly conserved core proteins in archaea and four core proteins in eukaryotes. The proteins that form the C/D box s(no)RNP complex were identified for both domains but the structure of the eukaryotic complex is not fully understood. The heterologous production and reconstitution of archaeal C/D box sRNP proteins proved to be more successful. Crystal structures for individual archaeal C/D box sRNP proteins, subcomplexes and the complete C/D box sRNP exist and provide insights into their composition, assembly and function.

In archaea, L7Ae is the primary RNA binding protein that binds and stabilizes the k-turn and k-loop motifs [40, 51]. Additionally, L7Ae has a function as a ribosomal protein subunit [52]. Essential features for L7Ae binding to the C/D box sRNA are (i) the terminal stem at the C/D box sRNA 5' and 3' ends, which juxtaposes the boxC and boxD sequences, (ii) the two sheared GA base pairs formed by pairing of the boxC and boxD sequences and (iii) the uridine of boxC that is part of the internal loop [51]. Direct interaction of L7Ae with six nucleotides of the k-turn could be observed in the crystal structure of an *in vitro* assembled L7Ae-C/D box sRNA complex of *Archaeoglobus fulgidus* [53]. These six nucleotides are the U and G of the boxD sequence, the conserved purine and the adjacent U and G nucleotides

of the boxC motif as well as the nucleotide upstream of the purine in the boxC motif (figure 1.1).

Following initial binding of L7Ae to the RNA, the archaeal C/D box sRNP protein Nop5 recognizes the sRNA-L7Ae complex. Nop5 possesses several important domains for the assembly of the complex. The carboxy-terminal domain exhibits RNA-binding sites and interacts with the L7Ae-C/D box sRNA complex [54]. Additionally, electrostatic interactions exist between Nop5 and the C/D box sRNA guide sequences [55, 56]. The coiled-coil domains of two Nop5 molecules interact, leading to the dimerization of the proteins. The amino-terminal domain of Nop5 interacts with fibrillarin, the third core protein of archaeal C/D box sRNPs [54, 57]. Fibrillarin uses S-adenosyl-L-methionine as methyl group donor for the 2'-*O*-methylation reaction. Independent methyl transfer activity of fibrillarin could not be observed indicating that the activity is dependent on C/D box sRNP formation [46, 58]. In the absence of L7Ae and the C/D box sRNA, fibrillarin and Nop5 build stable complexes with two subunits of both proteins [54, 59]. L7Ae can only bind to this subcomplex in the presence of a C/D box sRNA. It is not clear whether a step-wise assembly of all proteins occurs *in vivo* or whether a Nop5-fibrillarin subcomplex is preformed that binds to the L7Ae-C/D box sRNA subcomplex [60].

Conflicting biochemical and structural studies of the archaeal C/D box sRNP described the native complex either as a monomer with one C/D box sRNA and two copies of each protein or as a dimer of two C/D box sRNAs and four copies of each protein. The first electron microscopy (EM) structure of a C/D box sRNP was obtained for an *in vitro* reconstituted complex with recombinant *Methanocaldococcus jannaschii* proteins. This complex revealed a di-sRNP architecture [60, 61]. The crystal structure of a reconstituted complex with *Pyrococcus furiosus* proteins supported the existence of this di-sRNP architecture [62]. However, a crystal structure of an *in vitro* reconstituted C/D box sRNP with recombinant *Sulfolobus solfataricus* proteins revealed a mono-sRNP structure [63]. The difference of the assembly approaches resided in the structure of the utilized C/D box sRNA. The reconstitution of the *M. jannaschii* and *P. furiosus* complexes was performed with *in vitro* transcribed natural C/D box sRNA [61]. Reconstitution of the *S. solfataricus* complex was achieved with an artificial two-stranded RNA without k-loop motif [63]. Therefore, it is plausible that differences in the C/D box sRNP reconstitutions occurred due to the usage of different RNAs as it was shown that C/D box sRNP assembly strategies with non-natural C/D box sRNAs can lead to monomeric complex assemblies [60, 64].

Later NMR spectroscopy (nuclear magnetic resonance spectroscopy) and small-angle X-ray and neutron scattering (SAXS and SANS) experiments of reconstituted *P. furiosus* C/D box sRNPs and a physiological C/D box sRNA substrate indicated a dimeric structure and the analyses revealed insights into the working mechanism. It was found that the Nop5 subunits

interact with each other via their coiled-coil domains but the proteins are no strictly assigned to one guide sequence as thought before (figure 1.2A) [55]. The structure also revealed that the target RNAs are likely 2'-*O*-methylated in a sequential manner. Sequential 2'-*O*-methylation was observed previously and the structure unravels the mechanism [65]. Here, all guide regions form base pairs with the target RNA, but only the two diagonally opposing fibrillarin subunits are in contact with the target RNA at the same time and therefore only two modifications can be introduced at the same time (figure 1.2B). The modifications that are guided by the C/D box sRNA regions located between the boxC and boxD' motifs are targeted first and more efficiently [55, 66]. It remains to be elucidated whether sequential methylation is used to regulate rRNA folding.

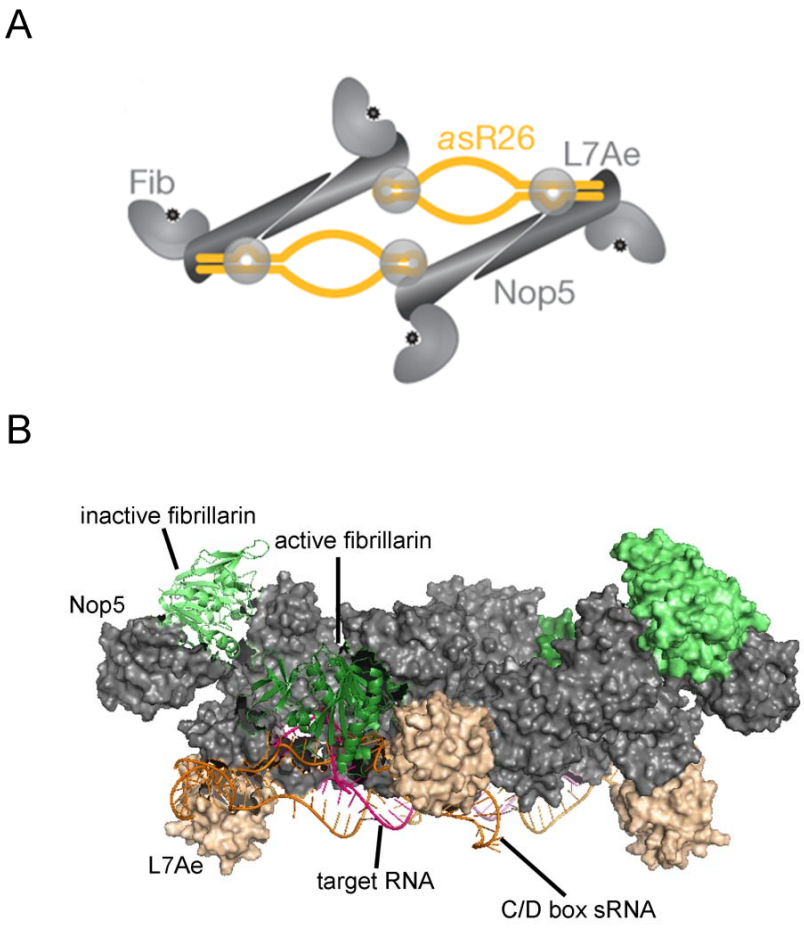

Figure 1.2: Schematic representation (A) and structure (B) of a dimeric archaeal C/D box sRNP. (A) The model includes four copies of each of the three core proteins fibrillarin (fib), Nop5 and L7Ae. The proteins are not assigned to one C/D box sRNA (yellow) but the Nop5 proteins interact to build the platform of the complex with two incorporated C/D box sRNAs (modified from [55]). (B) The complex reveals that only two fibrillarin subunits are in contact with the target RNA (active) (pdb: 4BY9).

In eukaryotes, orthologues of the three archaeal C/D box sRNP forming proteins exist. The protein that stabilizes the k-turn motif is the 15.5 kDa protein (Snu13 in yeast). K-loop motif binding could not be shown *in vitro* for the eukaryotic protein. This indicates that the C/D box

s(no)RNP assembly differs between archaea and eukaryotes [35, 36, 67-69]. The structural subunits of the eukaryotic C/D box snoRNP are Nop56 and Nop58, which exhibit highly similar sequences. They are predicted to form a heterodimer and replace the archaeal Nop5 as the coiled-coil domains of the two proteins show a similar sequence to the coiled-coil domain in Nop5 [54, 68, 70, 71]. Nop56 interacts with the guide regions of the snoRNA suggesting its involvement in substrate recognition. Nop58 is required for snoRNA stability [50, 68, 72, 73]. The catalytic subunit of the eukaryotic C/D box snoRNPs is fibrillarin (Nop1 in yeast) and S-adenosyl-L-methionine is used as methyl group donor for the 2'-*O*-methylation reaction [21, 74-76]. Moreover, several snoRNP-specific proteins exist in eukaryotes that have functions in snoRNP assembly and snoRNA protection during the maturation process [16].

In eukaryotes, the specific binding of archaeal L7Ae to k-turns was used to build synthetic translational regulators. Translation can be blocked by binding of L7Ae to artificial k-turns in 5' UTRs [77].

1.4 Biogenesis of C/D box s(no)RNAs

Archaeal and eukaryotic C/D box s(no)RNAs fulfil the same function but the genomic organization of their genes differs. This implies the presence of differing biogenesis pathways in both domains. In eukaryotes, the biogenesis pathway is well understood. The genomic organization of C/D box snoRNA genes and the transcription and maturation are highly variable. In general, a trend exists from lower to higher eukaryotes towards the reduction of the number of independent promoters by arranging C/D box snoRNA genes in polycistrons or within introns [78]. Furthermore, mobility of snoRNA genes was hypothesized for eukaryotes. Computational analyses of human snoRNA orthologues in other mammalian genomes revealed several paralogues that display retroposon characteristics: an A-rich tail and an approximately 14 bp long target site duplication [79].

In yeast, as well as in plants, most snoRNA genes are transcribed from independent RNA polymerase II (or less frequent RNA polymerase III) promoters either as mono- or polycistronic transcripts. In plants, C/D box snoRNA genes exist almost exclusively as polycistonic clusters [78, 80-83]. Additionally, dicistronic tRNA-C/D box snoRNA genes are reported for plants [84, 85]. In vertebrates, only few genes are transcribed from independent promoters. In these organisms, most of the genes are located within introns of protein-coding and non-protein-coding genes. The intronic C/D box snoRNA genes can either exist individually or as polycistrons, but individual localizations are more common [79, 86-88]. In general, clusters of several C/D box snoRNAs within an intron can be composed of homologous or heterologous snoRNA genes [30]. Interestingly, a decent amount of snoRNA genes are located in introns of housekeeping genes that have functions in ribosome

biogenesis. This implies that a correlation between snoRNA gene expression and the expression of proteins that are involved in the same processes exists [10, 78, 89].

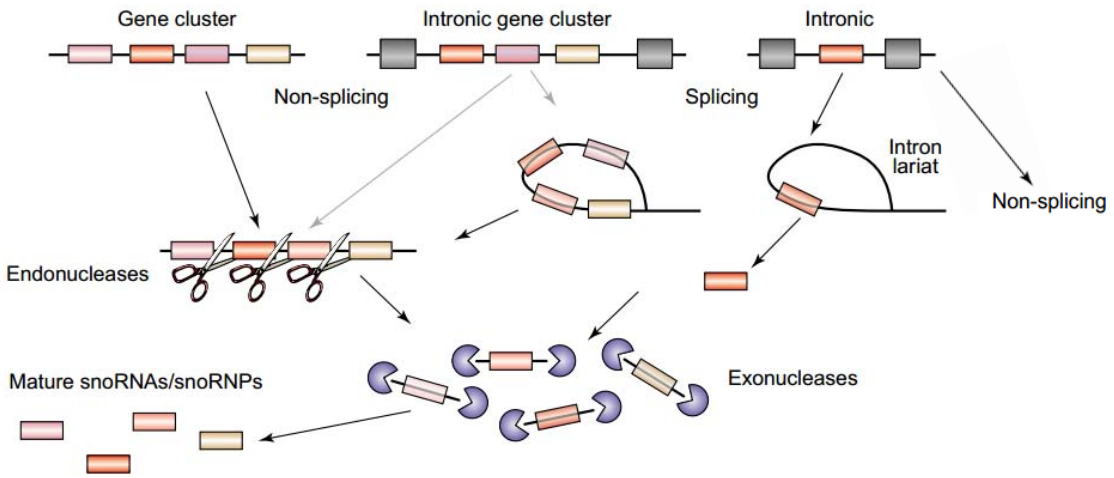

Figure 1.3: Processing of eukaryotic snoRNA genes. Polycistronic transcripts are processed by endoribonucleolytic cleavage and further trimming by exoribonucleases. A splicing-dependent and a splicing-independent pathway exists for intron-encoded snoRNAs (modified from [82]).

C/D box snoRNA genes in polycistronic clusters or in introns require maturation into functional C/D box snoRNAs before they can guide the 2'-*O*-methylation of their targets. Endo- and exoribonucleases are responsible for the processing of polycistronic C/D box snoRNA transcripts (figure 1.3). In yeast, polycistronic snoRNAs are flanked by sequences that form short hairpin structures and that are recognized by an RNaseIII-like endoribonuclease [90, 91]. Further trimming of the 5' and 3' ends occurs by 5'-3' and 3'-5' exoribonucleases that are not specific for the maturation of snoRNAs (Rat1, Xrn1; exosome) [92-94]. The dicistronic tRNA-C/D box snoRNA transcripts reported in plants are processed by RNaseZ or an alternative pathway which involves cleavage by unknown RNases [84, 85].

Intron-encoded C/D box snoRNAs are matured by two different pathways of which one is splicing-dependent and the other is splicing-independent [95]. Predominantly, the introns are spliced out of the pre-mRNA resulting in lariat structures that are debranched. Subsequently, mature C/D box snoRNAs are obtained by exoribonuclease activities [96, 97]. In the second pathway, introns are not spliced out of the pre-mRNA, but mature C/D box snoRNAs are revealed by the action of endoribonucleases that cleave up- and downstream of the C/D box sRNA and further trimming occurs by exoribonucleases. The second pathway seems to be important for snoRNA processing from introns in plants but also exists in yeast [96, 98].

The biogenesis pathway of archaeal C/D box sRNAs is largely unexplored but the observation of different genomic contexts of archaeal and eukaryotic C/D box s(no)RNA genes indicates that alternative mechanism must exist in archaea. Further evidence for a unique biogenesis pathway can be derived from circular archaeal C/D box sRNAs that are

present in addition to linear molecules [8, 9, 99, 100]. Circular RNAs are rare and, besides C/D box sRNAs, they are described for tRNA and rRNA introns and processing intermediates and other non-coding RNAs as H/ACA box sRNAs and RNaseP in archaea [99]. In *Thermoproteus* species, circular SRP RNA molecules were observed in which the circularization restores the functionality of these RNAs [101]. In the case of tRNA and rRNA introns, the splicing endonuclease recognizes and cleaves bulge-helix-bulge (BHB) structures that are formed by pairing of the regions at the junctions of the introns and the released introns are circularized by a dedicated tRNA ligase [102-105]. For circular SRP RNAs a similar mechanism is suggested [101]. However, C/D box sRNAs do not appear to form canonical BHB motifs that could explain the circularization of these RNAs [100]. Relaxed BHB structures with extended helices or lacking bulges were identified at the borders of C/D box sRNAs of several archaea but they are not likely to be cleaved by known splicing endonucleases [106]. Since circularized C/D box sRNAs were only detected in thermophilic archaea, it is suggested that the circularization stabilizes the C/D box sRNAs at elevated temperatures [99, 100].

Most C/D box snoRNA genes in eukaryotes contain independent promoters but the archaeal C/D box sRNA genes are predominantly located in intergenic regions or overlap with the 3' or 5' end of flanking ORFs [7, 9, 17, 107]. The archaeon *Nanoarchaeum equitans*, which contains a minimal and highly compact genome with several split protein-encoding genes, possesses C/D box sRNA genes directly adjacent to the split genes [9, 108, 109]. The lack of independent promoters raises the question about the transcription mechanism of archaeal C/D box sRNA genes. For several C/D box sRNA genes the genomic localization ensures their transcription and maturation. Similar to eukaryotes, few polycistronic C/D box sRNA transcripts are described. In *P. furiosus*, two clusters with two C/D box sRNA genes can be observed and in *S. solfataricus*, one cluster of two C/D box sRNA genes exists [7, 17]. However, the endoribonuclease RNaseIII that is important for processing of polycistronic transcripts in eukaryotes has not been identified in archaea which suggests that different enzymes have to be involved in the maturation process [110].

Different C/D box sRNA gene locations that ensure the C/D box sRNA transcription are downstream regions of tRNA genes, in which the C/D box sRNA is transcribed together with the tRNA. A dicistronic tRNA-C/D box sRNA is reported for the most abundant C/D box sRNA of *N. equitans*. Its gene is located downstream of the gene for $tRNA^{Val}$ and the tRNA maturation at the 3' terminus by RNaseZ is suggested to generate the 5' terminus of the C/D box sRNA [9]. C/D box sRNA genes can also exist in tRNA introns. The pre-$tRNA^{Trp}$ of *Haloferax volcanii* and *Halobacterium salinarum* possess C/D box sRNA genes in their introns. Interestingly, the *H. volcanii* C/D box sRNA guides modification of pre-$tRNA^{Trp}$ *in cis* and *in trans* [9, 65, 111]. Potential homologs of these C/D box sRNAs encoded in pre-

$tRNA^{Trp}$ introns were identified in additional euryarchaeal genomes [1, 65, 107, 111]. The splicing endonuclease recognizes the BHB motifs of the introns which effectively releases the C/D box sRNA. The intron of pre-$tRNA^{Trp}$ in *S. solfataricus* is also highly stable but it does not contain the characteristic boxC and boxD motifs. As it shows sequence similarity to the cognate $tRNA^{Trp}$ it could also play a role in modifying this RNA [112]. The pre-$tRNA^{Met}$ of *N. equitans* exhibits a stable intron with boxC and boxD motifs that would be able to form into a k-turn structure but a guide function of this RNA is unlikely [9].

1.5 C/D box sRNAs in *Sulfolobus acidocaldarius*

Sulfolobus acidocaldarius strain DSM639 is a hyperthermoacidophilic archaeon that was isolated from a solfatara [113]. It belongs to the phylum Crenarchaeota, metabolizes sulfur and grows under aerobic conditions at temperatures between 75°C and 80°C and a pH range between 2 and 3. Its genome was sequenced in 2005 and genetic systems have been developed [114-117].

S. acidocaldarius is the first archaeon in which C/D box sRNAs were identified and analyzed. Eighteen C/D box sRNAs could be detected in coimmunoprecipitation studies with fibrillarin and Nop5 [1]. L7Ae could be identified as the third core protein in purified C/D box sRNPs from the related species *S. solfataricus*. Reconstituted recombinant *S. solfataricus* C/D box sRNPs revealed information about the archaeal C/D box sRNP structure and about important regions in the C/D box sRNAs. The assembly of an active complex onto the RNA could only be reached in the order L7Ae, Nop5, fibrillarin and activity was shown to be dependent on intact k-turn and k-loop motifs. EM analyses revealed a di-sRNP structure of the reconstituted complex [40, 46, 60, 118]. Circular C/D box sRNAs were detected in *S. solfataricus* [99].

1.6 Aim of this study

The biogenesis pathway of archaeal C/D box sRNAs is largely unknown. One aim of this project is to obtain insights into the transcription mechanism and maturation pathway of members of this abundant and important RNA family. The genomic organization of C/D box sRNAs of six archaeal model organisms is analyzed to provide an overview of possible transcription mechanisms. Illumina RNA-Seq methodology is performed to identify C/D box sRNAs in the genetically tractable archaeon *S. acidocaldarius* and to analyze C/D box sRNA circularization in this organism. Plasmid-based C/D box sRNA *in vivo* analyses are performed in *Sulfolobus acidocaldarius* in which C/D box sRNA genes variants are used to identify C/D box sRNA maturation, stabilization and circularization requirements. The influence of the C/D box sRNA gene's upstream and downstream sequences and the

integrity of the boxC/C' and boxD/D' motifs and the k-turn/k-loop structures for these processes is tested.

Guide sequences of identified C/D box sRNAs are extracted to identify conserved 2'-*O*-methylation sites in archaeal target rRNAs. The C/D box sRNA targets of seven archaea are mapped onto the consensus structure of the 16S and 23S rRNA to identify methylation hotspots.

2. Results

2.1 Genetic context of archaeal C/D box sRNA genes

Previous studies revealed that archaeal C/D box sRNA genes often do not possess an external promoter that is characterized by a TATA-box sequence in a distance of -26 +/- 3 bp upstream of the transcription start site [9, 100, 119]. Therefore, the genomic 50 nt upstream region of 343 C/D box sRNA genes from six archaea (table 2.1) was computationally screened for putative promoter sequences or other conserved motifs with a length of 6-15 nt (in cooperation with Omer Alkhnbashi, group of Rolf Backofen, Albert-Ludwigs-University Freiburg). 58 C/D box sRNA genes of *Ignicoccus hospitalis* are not included in the analysis because they were identified at a later date.

Table 2.1: List of the organisms of which the genetic context of C/D box sRNA genes was analyzed. The archaeal organisms, their phylogenetic classification, the growth temperature, the GC content and the number of C/D box sRNA genes are indicated.

Organism	Phylum/Order	Growth temperature	GC content (%)	No. of C/D box sRNA genes	Reference
Methanococcus maripaludis C5 (Mma)	Euryarchaeota/ Methanococcales	35 – 40 °C	33.0	7	[120]
Nanoarchaeum equitans (Neq)	Nanoarchaeota	80 – 100 °C	31.6	26	[9]
Sulfolobus acidocaldarius (Sac)	Crenarchaeota/ Sulfolobales	67 – 80 °C	36.7	61	[1, 7, 121], this thesis
Thermoproteus tenax (Tte)	Crenarchaeota/ Thermoproteales	70 – 97 °C	55.1	52	[122]
Methanopyrus kandleri (Mka)	Euryarchaeota/ Methanopyrales	84 – 110 °C	61.2	127	[8]
Ignicoccus hospitalis (Iho)	Crenarchaeota/ Desulfurococcales	80–100 °C	56.5	128	[9]

The only conserved sequences that can be identified are TATA-box elements and just about 20 % of the C/D box sRNA genes possess these elements within the 50 nt upstream region (figure 2.1A). Three different AT-rich motifs can be found and the motif 2 is the most prevalent motif. It can be observed for 49 % of the C/D box sRNAs genes where one of the three motifs can be identified in the upstream region but the motif 2 is also the most variable one of the three identified motifs. In *Methanococcus maripaludis* the motif 2 does not exist upstream of C/D box sRNA genes (figure 2.1C). In general, in *M. maripaludis*, *Nanoarchaeum equitans*, *Thermoproteus tenax* and *Sulfolobus acidocaldarius* about half of the C/D box sRNA genes possess one of the identified TATA box elements. In *I. hospitalis* and *Methanopyrus kandleri*, the two analyzed organisms with the highest GC content, only 6 out of 70 and 3 out of 127 C/D box sRNA genes, respectively, possess the indicated TATA-

box motifs within the 50 nt upstream region and only the motifs 1 and 2 can be identified (table 2.1; figure 2.1B, C). For the C/D box sRNA genes where a TATA box motifs can be identified, only about 40 % possess the promoter sequences at the typical position -26 +/- 3 bp (figure 2.1B). This correlates to 8.8 % of all analyzed C/D box sRNAs genes. Other AT-rich motifs at this position can be identified for a minor fraction of C/D box sRNA genes but the majority of C/D box sRNA genes does not appear to utilize promoters that would allow primary C/D box sRNA transcript production.

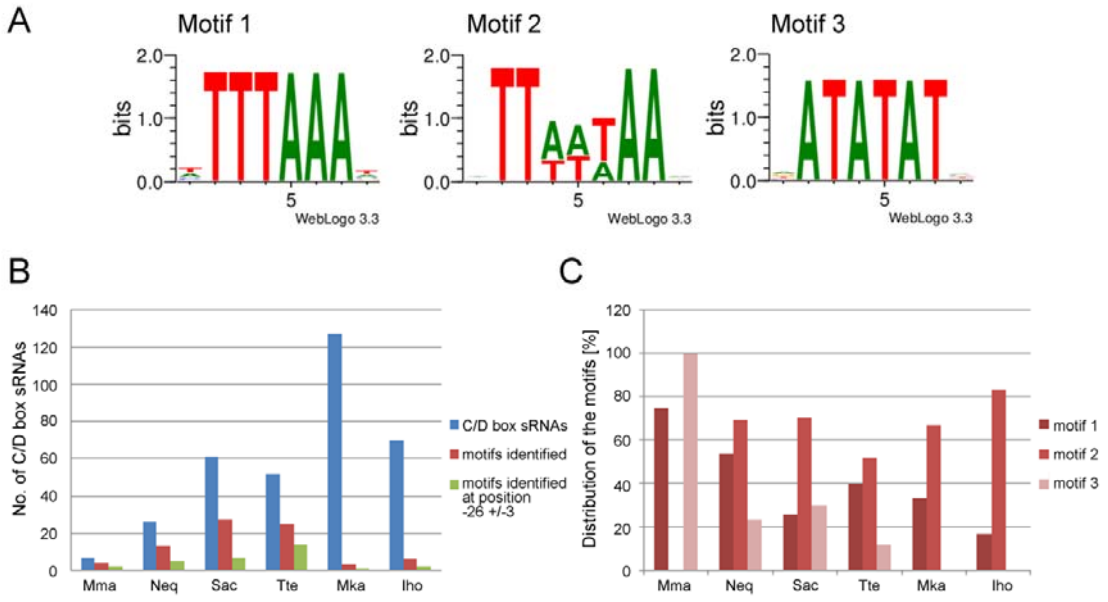

Figure 2.1: Conserved sequence motifs within the genomic 50 nt upstream region of archaeal C/D box sRNA genes. (A) The genomic upstream regions of 343 archaeal C/D box sRNA genes were analyzed for conserved sequences and the displayed sequences were identified. (B) The diagram shows the total amount of C/D box sRNA genes in the individual analyzed organisms and the amount of genes were the conserved motifs could be identified. Additionally, the minor fraction of C/D box sRNA genes that possess the promoter motifs at the typical position -26 +/- 3 bp is displayed. (C) The diagram shows the distribution of the three identified promoter motifs upstream of C/D box sRNA genes in the six analyzed organisms.

Previous analyses of the genetic context of C/D box sRNA genes in several archaea revealed that they are located at positions that enable transcription without the necessity of an own promoter. They can e.g. be found within tRNA introns or downstream of tRNA genes [1, 9, 17, 123]. To identify additional potential transcription pathways, the genetic context of the archaeal C/D box sRNA genes summarized in table 2.1, with the exception of *N. equitans*, was analyzed. *N. equitans* C/D box sRNAs were excluded from the analyses as their genetic context was analyzed previously [9]. More than one fourth of the C/D box sRNA genes overlaps with the 3' end of flanking ORFs (figure 2.2A). In *I. hospitalis* alone, 40 % of all C/D box sRNA genes are found in this context. The position of C/D box sRNA genes at the 3' end of flanking ORFs implies that the transcripts of the neighbouring genes possess a

3'-UTR. In addition to C/D box sRNA genes that overlap with the 3' end of flanking ORFs, several C/D box sRNA genes can be identified that are located downstream of ORFs and only few nucleotides separate the C/D box sRNA gene and the flanking protein-coding gene. In these cases, an extended transcript may exist in which the C/D box sRNAs are part of the 3'-UTR of the flanking protein-coding region.

C/D box sRNAs that overlap with the 3' end of flanking protein-coding regions contain the stop codon at the end of the coding region within their sequence. In nearly half of the C/D box sRNAs this stop codon is situated within the boxC motif, but it can be also situated in the boxD', boxC' or boxD motifs (figure 2.2B). Additionally, stop codons can be found within guide sequences, in the sequence between the boxC' and boxD' motif or in the sequence upstream of the boxC motif of the C/D box sRNAs. Some of these stop codon positions result in the occurrence of one or both guide regions within the ORF.

In addition to C/D box sRNA genes that overlap with the 3' end of flanking ORFs, C/D box sRNA genes can also be found that overlap with the 5' end of flanking ORFs. However, only 7 % of the C/D box sRNA genes of the six analyzed archaea are located at this position. The position of C/D box sRNA genes at the 5' end of flanking ORFs implies that the transcripts of the neighbouring genes possess a 5'-UTR. C/D box sRNAs that overlap with the 5' end of flanking protein-coding regions contain the start codon of the coding region within their sequence. Most start codons are situated in the guide sequence or in the sequence downstream of the boxD motif (figure 2.2C). Additionally, the start codon can be found within all box motifs. When it is situated in the boxC, boxC' or boxD' motif, one or both guide sequences of the C/D box sRNAs are part of the neighbouring coding region.

In addition to C/D box sRNA genes that overlap with the 5' end of flanking ORFs, 7 % of C/D box sRNA genes are located upstream of the neighbouring coding gene and the distance between both genes is less than 26 nt. Consequently, the promoter sequence for the protein-coding genes are located within the C/D box sRNA gene sequences or upstream of the C/D box sRNA genes. In the latter case, the C/D box sRNAs might be on a shared transcript with the flanking protein-coding regions in which the C/D box sRNAs are part of their 5'-UTR sequences.

Archaeal transcriptome data from *Sulfolobus solfataricus* and *Pyrobaculum aerophilum* revealed that most transcripts do not possess a 5'-UTR [124, 125]. However, careful analysis of the *S. solfataricus* transcripts that possess a 5'-UTR shows that two 5'-UTRs exhibit C/D box sRNA characteristics. The SSO0495 transcript exhibits a 72 nt long 5'-UTR and in this UTR a complete C/D box sRNA can be found (Sso-sR6). The SSO0887 transcript has a 25 nt long 5'-UTR and this UTR contains a C/D box sRNA half (Sso-sR2). The remaining C/D box sRNA gene half overlaps with *SSO0887* and the start codon for the synthesis of SSO0887 is situated in the boxC' motif.

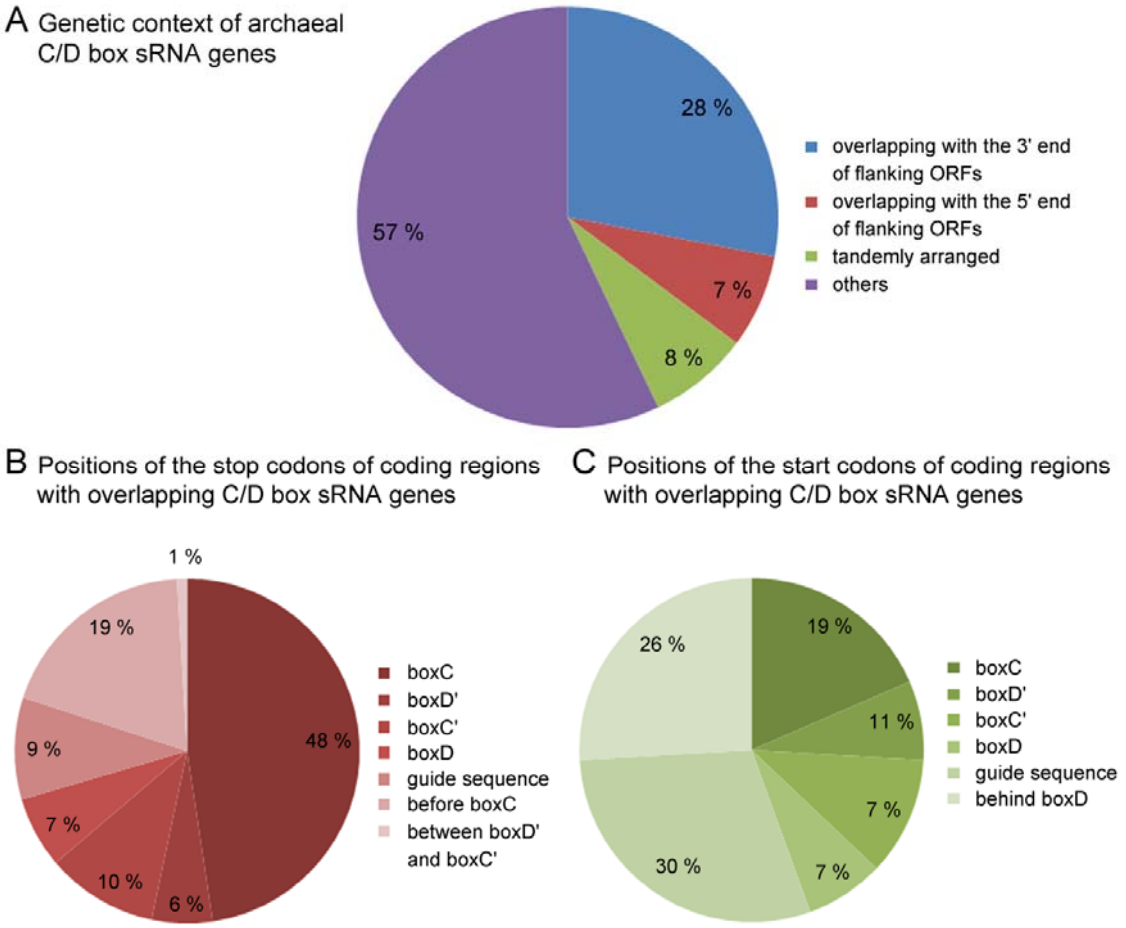

Figure 2.2: Genetic context of archaeal C/D box sRNA genes and positions of the start and stop codon within C/D box sRNAs that overlap with the 5' or 3' end of flanking ORFs. (A) The genetic context of 375 C/D box sRNA genes was analyzed. C/D box sRNA genes that overlap with the flanking ORFs imply that the transcripts possess UTRs. C/D box sRNA genes that are located upstream or downstream of the neighbouring gene, separated by only few nt may also exist as shared transcripts but are included in the diagram as 'others'. In total, 15 % of the C/D box sRNA genes are separated from the upstream or downstream gene by less than 25 nt (B) Positions of the stop codons of coding regions with overlapping C/D box sRNA genes. (C) Positions of the start codons of coding regions with overlapping C/D box sRNA genes.

Another frequent organisation of C/D box sRNA genes are clusters of two or three C/D box sRNA genes suggesting polycistronic transcription. This arrangement can be found multiple times in *S. acidocaldarius*, *M. kandleri* and *I. hospitalis.*

In *T. tenax* and *I. hospitalis*, two putative dicistronic tRNA-C/D box sRNA pairs exist. A putative dicistronic tRNA-C/D box sRNA was already identified in *N. equitans* and the 3' processing activity of RNaseZ was proposed to generate the 5' terminus of the C/D box sRNA [9]. In *T. tenax*, the C/D box sRNA *TtesR134* gene is located directly downstream of a $tRNA^{Pro}$ gene and in *I. hospitalis* the C/D box sRNA *IhosR11* gene is located downstream of a $tRNA^{Ser}$ gene. If a dicistronic transcript exists, the tRNA 3' processing activity of RNaseZ would be not sufficient to create a mature C/D box sRNA 5' terminus in *I. hospitalis* as a 14 nt spacer separates the tRNA and the C/D box sRNA.

The majority of the remaining C/D box sRNA genes is located in intergenic regions. The gene of C/D box sRNA MKAsR55 of *M. kandleri* exhibits an interesting localization as it is encoded on the reverse strand between *purS* and *purL* and thereby disrupts an operon of three purine biosynthesis genes [8]. Eight other purine biosynthesis genes exist in the genome of *M. kandleri* and two of them are also associated with C/D box sRNA genes but in these cases the C/D box sRNA genes may be transcribed together with the protein-coding genes and create putative 5'-UTRs of the protein-coding regions.

All these observations concerning the genetic context of C/D box sRNA genes indicate that C/D box sRNA transcription is highly variable and that C/D box sRNA genes do not necessarily need to possess dedicated promoters to ensure their transcription but that they can rather be part of longer precursors with adjacent transcription products. However, the transcription of longer precursors implies that processing is required for the maturation of individual C/D box sRNAs.

2.2 Small RNA sequencing and identification of C/D box sRNAs of *S. acidocaldarius*

Illumina RNA-Seq methodology was utilized to sequence small RNA species (<200 nt) and to identify C/D box sRNAs in *S. acidocaldarius* MW001, a uracil auxotrophic strain that is used for genetic work with this archaeon [114]. The obtained RNA-Seq reads were mapped to the genome of wild type *S. acidocaldarius* DSM639.

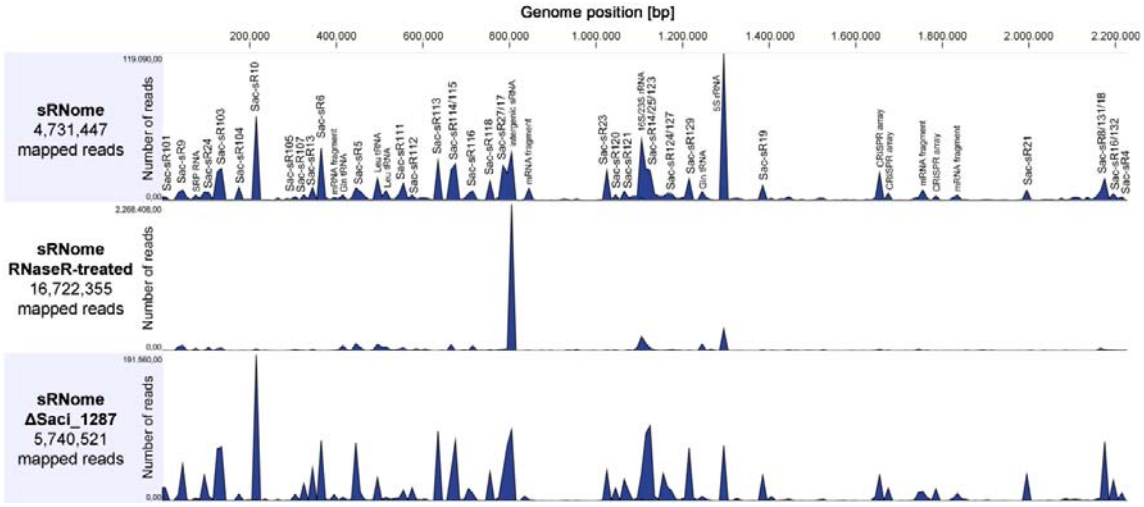

Figure 2.3: The sRNome of *S. acidocaldarius*. The graphs represent the genome-wide coverage of Illumina HiSeq2000 reads that were mapped to the genome of *S. acidocaldarius* DSM639 (NC_007181.1, 2225959 bp). Small RNAs corresponding to the prominent peaks were identified. Three datasets are visualized: the sRNome of *S. acidocaldarius* MW001 (upper section), the RNaseR-treated sRNome of *S. acidocaldarius* MW001 (middle section) and the sRNome of the mutant strain *S. acidocaldarius* ΔSaci_1287 (lower section).

In total, this mapping contains 4,731,447 reads with an average length of 23 nt. The genome-wide coverage of sequencing reads indicates that the largest amount of sequencing reads

were obtained for rRNA fragments and C/D box sRNAs (figure 2.3 upper section). In total, 61 C/D box sRNAs were identified.

As described in previous studies, the abundance of individual C/D box sRNAs differs significantly [8]. Some RNAs are just covered by a few hundreds of sequencing reads and others are covered by ten thousands of reads. The length of the 61 C/D box sRNAs varies between 53 and 76 nt but the length of individual C/D box sRNAs is not always absolute. Variations of up to seven nucleotides on the 5' and 3' termini of the RNAs can be observed (figure 2.4). Interestingly, variable 5' termini can be observed even if the C/D box sRNA genes are likely transcribed from their own dedicated promoter.

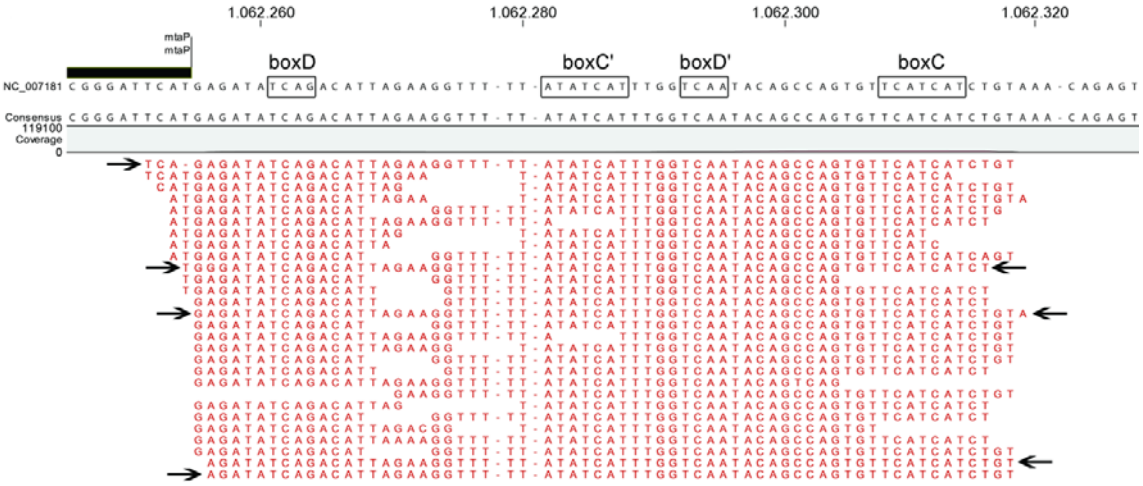

Figure 2.4: RNA-Seq reads of C/D box sRNA Sac-sR121 mapped to the reference genome (*S. acidocaldarius* DSM639; NC_007181.1). The C/D box sRNA is encoded on the negative strand and the characteristic boxC/C' and boxD/D' motifs are marked on the reference genome. Arrows indicate sequencing reads with different 5' and 3' termini of the RNA.

2.3 Circular C/D box sRNAs

Previous studies showed that archaeal C/D box sRNAs can exist in a circular form in addition to the linear form. Permuted RNA-Seq reads in which the order of the four conserved boxes is not C-D'-C'-D but for example C'-D-C-D' were identified in *M. kandleri, S. solfataricus* and *N. equitans* which hints at the existence of circular C/D box sRNA molecules [8, 99, 100]. In the RNA-Seq data of *S. acidocaldarius*, permuted sequencing reads cannot be observed. On the one hand, this could be due to the RNA isolation and library preparation procedure as circular RNA molecules need internal cleavage for effective adapter ligation. On the other hand, the read length is much shorter in the *S. acidocaldarius* dataset in comparison to the *M. kandleri* dataset and short permuted reads might be not properly represented in the read mapping.

To enrich for circular RNA molecules before RNA-Seq library preparation, the isolated small RNAs were treated with RNaseR, a 3'-5' exoribonuclease that digests linear RNA and leaves lariat structures and circular RNAs intact. The RNA-Seq run revealed 16,722,355 reads with

an average length of 20 nt that map to the genome of *S. acidocaldarius* DSM639 (figure 2.3 middle section). All 61 C/D box sRNAs can be identified in the RNA-Seq data of the RNaseR-treated small RNA sample indicating that circular C/D box sRNAs exist in *S. acidocaldarius*. However, portions of the rRNAs are also present and this might be a result of incomplete RNaseR degradation. Strikingly, reads for a small RNA in an intergenic region at genome position 805,747 are highly enriched and represent the majority of RNA-Seq reads. Recent data from our laboratory suggest the discovered RNA to be a trans-encoded sRNA with a vital regulatory function in biofilm formation (Dr. Alvaro Orell, unpublished data). Northern blot analyses and inverse reverse transcriptase-polymerase chain reaction (iRT-PCR) studies of specific C/D box sRNAs provided further evidences for C/D box sRNA circularization in *S. acidocaldarius*. Northern blot analyses with specific probes for a set of C/D box sRNAs revealed the existence of two bands (figure 2.5).

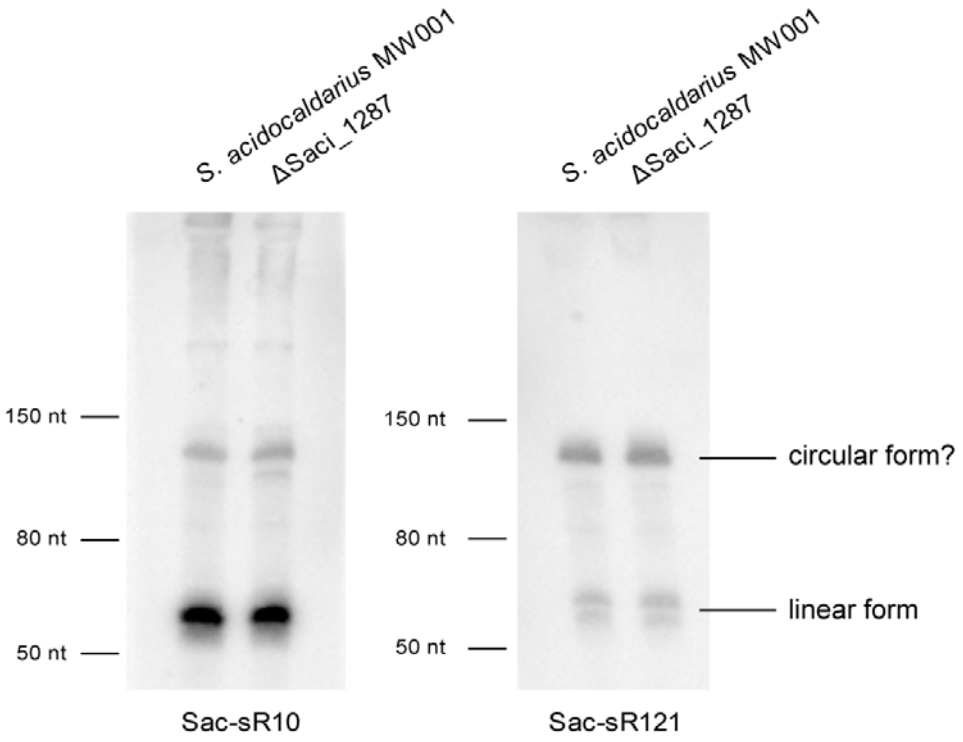

Figure 2.5: Northern blot analyses of C/D box sRNA Sac-sR10 and Sac-sR121. The analyses were performed with 10 µg extracted total RNA from *S. acidocaldarius* MW001 and the mutant strain *S. acidocaldarius* ΔSaci_1287 using a 5'-radiolabeled probe complementary to the C/D box sRNA sequence. Two major RNA forms are visible. The size of approx. 60-70 nt corresponds to the linear form of these RNAs. For *Pyrococcus furiosus* C/D box sRNAs a slower migrating form was observed in denaturing polyacrylamide gels and shown to be a circular variant of the C/D box sRNAs [100].

Previous studies observed slower migration of circular *Pyrococcus furiosus* C/D box sRNAs in denaturing gels [100]. However, RNA extraction of both *S. acidocaldarius* forms from a denaturing polyacrylamide gel and subsequent inverse RT-PCR revealed circular C/D box sRNAs in both samples. This result could be due to insufficient separation of both variants on the gel. Northern blot probes that cross the circularization junction were difficult to design as

they have to fulfill two criteria. First, they have to be short enough to not also bind to linear C/D box sRNAs and, second, they have to be specific for the recognition of individual C/D box sRNAs. As the boxC and boxD motifs are close to the circularization junction it is difficult to reach sufficient specificity. Furthermore, as shown in section 2.2, C/D box sRNAs display length heterogeneity at their 5' and 3' termini. Thus, different circularization junctions exist that cannot be detected in Northern blot analyses with a single probe that crosses the circularization junction.

Inverse RT-PCR analyses with specific outward-facing oligonucleotides for the amplification of different circular C/D box sRNAs (Sac-sR24, Sac-sR121, Sac-sR10) were successful and sequencing of the amplification products confirmed the presence of circularized C/D box sRNA molecules. The observed ladder-like pattern of RT-PCR products on an agarose gel is likely caused by multiple rounds of reverse transcription around a circular molecule (figure 2.6A).

Analyses of the amplified sequences provided information about the circularization junction. Amplification products from iRT-PCR analyses with specific oligonucleotides for C/D box sRNA Sac-sR121 reveal the existence of different circularization junctions within one amplification product (figure 2.6B). The occurrence of different circularization junctions in single amplification products might be PCR artefacts. The products of the first amplification cycle can serve as primers for the amplification in the next cycle. When the circularization junctions of the molecules that serve as primer and of the molecules that become amplified differ, amplification products with different circularization junctions within a single molecule are generated. Nevertheless, all observed circularization junctions represent real junctions of C/D box sRNA Sac-sR121 in *S. acidocaldarius*. The fact that different circularization junctions can be observed, fits to the RNA-Seq-derived observation that C/D box sRNA variants with diverse 5' and 3' termini exist. Different circularization junctions were revealed by sequencing of three different PCR products from iRT-PCR analyses with specific oligonucleotides for C/D box sRNA Sac-sR121 (Fig. 2.6B). Interestingly, in one case the sequence upstream of the boxC motif is 24 nt long although the predicted (and in section 2.4.2 confirmed) TATA-box is located at position -25 upstream of the boxC motif. This indicates that this promoter is not the only one to initiate C/D box sRNA Sac-sR121 transcription.

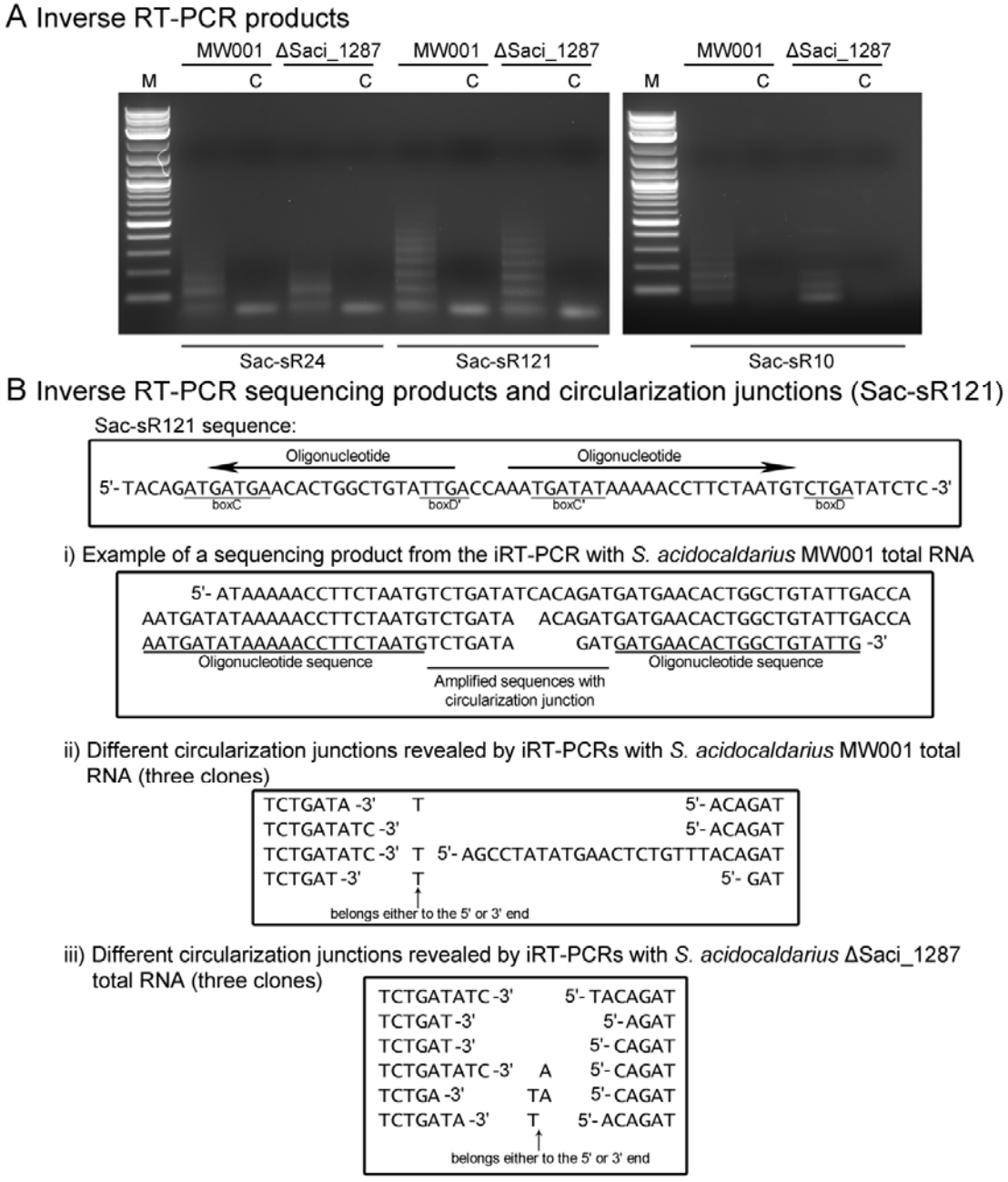

Figure 2.6: Inverse RT-PCR analyses for the verification of circular C/D box sRNAs and determination of the circularization junction. (A) Agarose gels of iRT-PCR products of C/D box sRNAs Sac-sR24, Sac-sR121 and Sac-sR10. For the reverse transcription reaction extracted total RNA from *S. acidocaldarius* MW001 and the mutant strain *S. acidocaldarius* ΔSaci_1287 as well as specific oligonucleotides for the indicated C/D box sRNAs were used. In the control reaction no reverse transcriptase was added. A ladder-like pattern of the RT-PCR products that results from multiple rounds of reverse transcription around a circular RNA template was observed. (B) The C/D box sRNA Sac-sR121 sequence with the positions of the oligonucleotides that were used for amplification is indicated. The sequencing of the amplification products revealed the circularization junction but varying junctions exist in the same molecule (i). Sequencing of several iRT-PCR products of *S. acidocaldarius* MW001 (ii) or the mutant strain *S. acidocaldarius* ΔSaci_1287 (iii) show diverse circularization junctions. For some nucleotides it cannot be determined whether they originally belong to the 5' or 3' terminus of the linear C/D box sRNA molecule.

RNA ligases are obvious candidates to perform the circularization of C/D box sRNAs. From the three known archaeal RNA ligases, two are annotated in *S. acidocaldarius*: the 2'-5' RNA ligase (Saci_1287) and the tRNA ligase RtcB (Saci_1317). A potential 5'-3' RNA ligase could not be identified in *S. acidocaldarius*. The RNA ligase RtcB ligates tRNA molecules after intron removal by the tRNA splicing endonuclease [126-128]. The 2'-5' RNA ligase is highly conserved in archaea and bacteria but its physiological role is unknown [129, 130]. Therefore, this RNA ligase was a good candidate enzyme for the circularization of C/D box sRNAs. To test the influence of both RNA ligases for the circularization of C/D box sRNAs, markerless deletion of both ligase genes was attempted. A markerless deletion mutant of *saci_1287* could be obtained, however the deletion of Saci_1317 was not possible. It is likely that this tRNA ligase is essential.

Growth studies of the mutant strain *S. acidocaldarius* ΔSaci_1287 revealed a slightly improved growth rate compared to the strain *S. acidocaldarius* MW001 (figure 2.7).

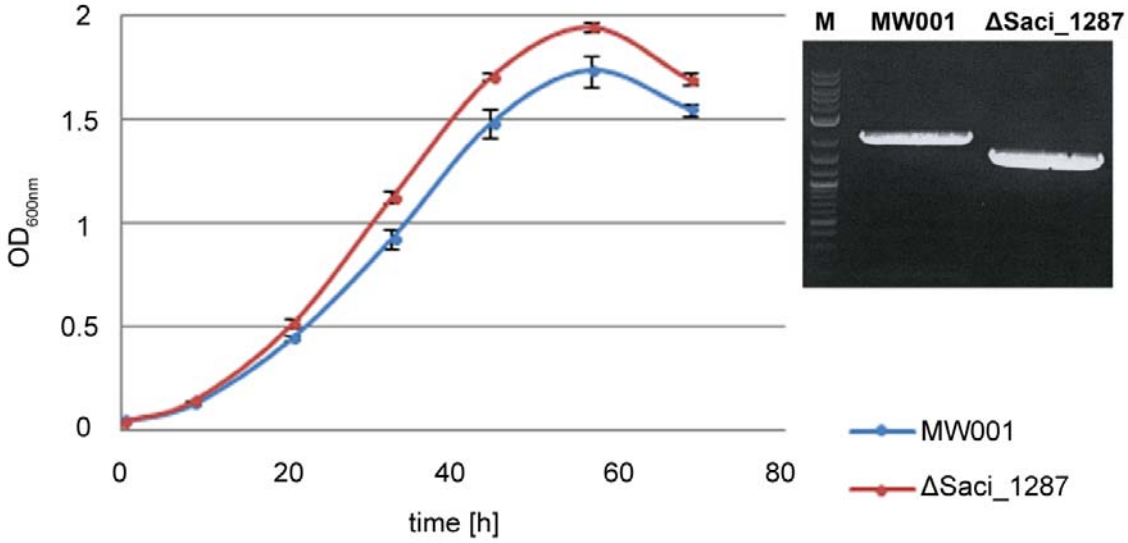

Figure 2.7: Growth curve of *S. acidocaldarius* MW001 and the mutant strain *S. acidocaldarius* ΔSaci_1287 and verification of the deletion of *saci_1287* via PCR analysis.

Illumina RNA-Seq methodology was utilized to sequence small RNA species of *S. acidocaldarius* ΔSaci_1287. 5,740,521 reads with an average length of 34 nt mapped to the genome of *S. acidocaldarius* DSM639 (figure 2.3 lower section). The overall pattern looks similar to the sRNome data of *S. acidocaldarius* MW001. The RNA-Seq coverage plots of *S. acidocaldarius* ΔSaci_1287 and *S. acidocaldarius* MW001 were compared to identify the effect of the 2'-5' RNA ligase deletion. However, discernable differences were not observed and missing ligation products were not identified. Interestingly, permuted C/D box sRNA reads can be observed for few C/D box sRNAs. Additionally, Northern blot and inverse RT-PCR analyses indicate that C/D box sRNA circularization occurs in *S. acidocaldarius*

ΔSaci_1287 (figure 2.5 and 2.6B). The circularization junctions are similar to the junctions that can be observed in *S. acidocaldarius* MW001. Therefore, the 2'-5' RNA ligase does not seem to be involved in the circularization of C/D box sRNAs. The physiological function of this universal RNA ligase remains enigmatic. It is possible that RNA circularization is mediated by RtcB as it was shown to fulfil other functions apart from tRNA maturation [101, 131].

RtcB ligates tRNA molecules after intron excision by the tRNA splicing endonuclease. The tRNA splicing endonuclease recognizes and cleaves bulge-helix-bulge (BHB) motifs that are formed at the exon-intron junctions [105]. In previous studies, canonical BHB motifs could not be identified at the borders of C/D box sRNAs [100]. Only relaxed BHB structures with extended helices or lacking helices were identified at the borders of C/D box sRNAs of several archaea but they are not likely to be cleaved by known splicing endonucleases [106].

To identify putative structures that could be recognized by specific endoribonucleases, a bioinformatic screen for conserved structures in the 50 nucleotide upstream and downstream region of 343 C/D box sRNA genes (see section 2.1) was performed (in cooperation with Omer Alkhnbashi, group of Rolf Backofen, Albert-Ludwigs-University Freiburg). However, convincing structures were not identified. In conclusion, C/D box sRNA circularization occurs in *S. acidocaldarius* but the RNA ligase that is responsible for the ligation as well as potential structures in the upstream and downstream regions of the C/D box sRNAs that could be recognized by the ligase were not identified.

2.4 Maturation of C/D box sRNAs

2.4.1 Verification of C/D box sRNAs in 5' and 3'-UTRs of flanking coding regions and verification of dicistronic C/D box sRNA transcripts

The genetic context of several C/D box sRNA genes suggests their transcription as a precursor, e.g. together with the flanking gene's mRNA when the distance between C/D box sRNA gene and the flanking coding region is small. In addition, polycistronic transcripts of several proximal C/D box sRNA genes are possible. Northern blot analyses with probes hybridizing with both individual RNAs of the C/D box sRNA Sac-sR125 and Sac-sR126 tandem show two major bands. The size of the lower band corresponds to the size of a single C/D box sRNA and the size of the upper band corresponds to the size of the C/D box sRNA dicistron (figure 2.8A). RNaseR treatment prior to the analysis confirmed that the transcript with a size of around 140 nt was the dicistronic transcript and not the circular form of the C/D box sRNA because both bands disappeared after the treatment. RT-PCR analyses with an oligonucleotide pair in which each oligonucleotide binds within one of the C/D box sRNAs of the tandem C/D box sRNAs reveal a product and confirm that dicistronic precursor transcript exist (figure 2.8B).

Further RT-PCR analyses were performed to test whether C/D box sRNA genes that are located in close proximity to ORFs, are transcribed together with the flanking gene and therefore are part of the 5' or 3'-UTR of the flanking mRNA region. RT-PCR analyses were performed with oligonucleotide pairs and one oligonucleotide bound within the C/D box sRNA and the other one within the flanking mRNA portion.

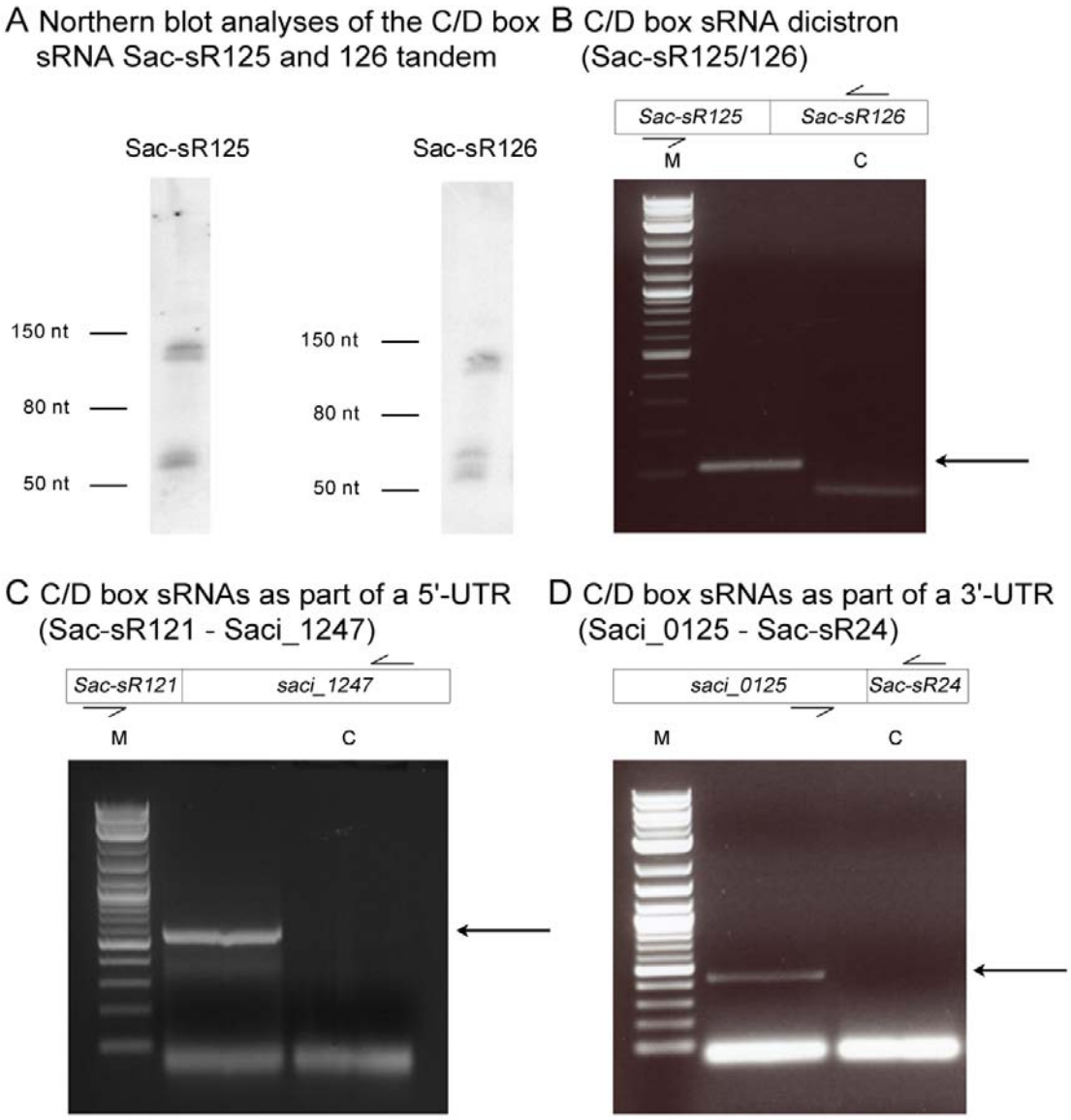

Figure 2.8: Verification of dicistronic C/D box sRNA transcripts and C/D box sRNAs in 5' and 3'-UTRs of flanking coding regions via Northern blot and RT-PCR analyses. RT-PCR analyses were performed with oligonucleotide pairs as indicated. In the control reaction, no reverse transcriptase was added. Arrows indicate amplification products of the expected size. (A) Northern blot analyses were performed with 10 µg extracted total RNA from *S. acidocaldarius* MW001 and 5'-labeled probes complementary to C/D box sRNA Sac-sR125 and Sac-sR126, respectively. Both analyses reveal two major bands that correspond to the size of a mature C/D box sRNA (~60/65 nt) and the C/D box sRNA dicistron (~125 nt). (B) RT-PCR analyses of the C/D box sRNA tandem Sac-sR125/126 reveal the existence of a dicistronic transcript. (C) RT-PCR analyses to test for the existence of C/D box sRNA Sac-sR121 in the 5'-UTR of Saci_1247 reveal a faint product of the estimated size. (D) RT-PCR analyses to test for the existence of C/D box sRNA Sac-sR24 in the 3'-UTR of Saci_0125 reveal a faint product of the estimated size.

The C/D box sRNA Sac-sR121 gene is located directly upstream of *saci_1247* and the C/D box sRNA Sac-sR24 gene is located directly downstream of *saci_0125*. RT-PCR analyses with the indicated oligonucleotides reveal products of the estimated sizes indicating that shared precursor transcripts exist (figure 2.8C und D). However, the products of the RT-PCR analyses are very faint and large amounts of unbound oligonucleotides are visible potentially indicating low amounts of the precursor transcripts. This observation is supported by the absence of signals for the C/D box sRNA-mRNA precursor transcripts in Northern blot analyses with probes hybridizing to the C/D box sRNAs (figure 2.5). However, it is also difficult to determine the expected height as little is known about transcription termination in archaea and operon arrangements of the investigated genes are also possible.

2.4.2 C/D box sRNA maturation - *in vivo* assays with C/D box sRNA Sac-sR121

The existence of C/D box sRNA precursors necessitates the presence of a maturation mechanism. The maturation could be achieved via cleavage by unspecific exoribonucleases and/or site-specific endoribonucleases. Variable 5' and 3' termini exist (figure 2.4), which provide a hint for unspecific degradation but it cannot be excluded that endoribonucleases act prior to the unspecific trimming by exoribonucleases. However, the computational screen for conserved structures in the 50 nt upstream and downstream sequences of the C/D box sRNAs that could potentially be recognized by endoribonucleases did not reveal any conserved elements. Moreover, *S. acidocaldarius* cell extract fractionation with liquid chromatography either on a heparin or a MonoS column and treatment of 5' radiolabeled C/D box sRNA precursor *in vitro* transcripts, with a 156 nt extension on the 3' end, with the chromatographic fractions could not identify an endoribonuclease activity in any of the fractions. Except for thermal degradation that also occurred in the control where no chromatographic fraction was added, the C/D box sRNA precursor remained unaltered.

To screen C/D box sRNA maturation requirements, *in vivo* studies were performed in *S. acidocaldarius* MW001. The plasmid pSVA1431 containing an inducible maltodextrin promoter was used as backbone in the experiments. This vector is described as pCmalLacS in Berkner *et al.* 2010 [116] but the distance between the TATA box and the transcription start site is 2 bp shorter in pSVA1431 (Michaela Wagner, group of Sonja-Verena Albers, Albert-Ludwigs-University Freiburg, personal communication). The C/D box sRNA *Sac-sR121* gene, found directly upstream of *saci_1247*, was cloned with 50 bp of the native upstream and downstream sequence behind the inducible maltodextrin promoter into this plasmid. The D' guide was exchanged partially against a unique 12 nt sequence (C/D box sRNA Sac-sR121*) to enable detection of the transcripts and to distinguish them from native C/D box sRNA Sac-sR121 transcripts. The general C/D box sRNA Sac-sR121* cloning scheme is shown in figure 2.9. This construct was used to introduce changes in the

promoter, C/D box sRNA up- and downstream region or within the C/D box sRNA gene as detailed in the respective sections.

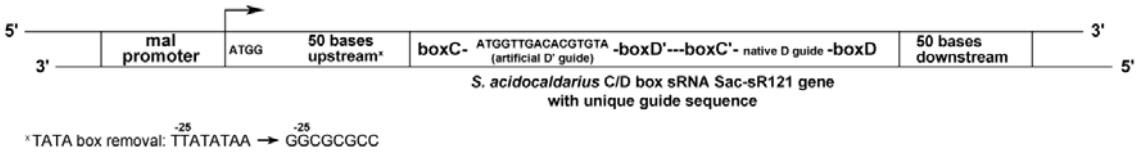

Figure 2.9: Cloning scheme of C/D box sRNA Sac-sR121* *in vivo* constructs. The constructs contain C/D box sRNA Sac-sR121 with an artificial D' guide and 50 bp of the native upstream and downstream sequence of the C/D box sRNA gene. The constructs are cloned into the plasmid pSVA1431 containing an inducible maltodextrin (mal) promoter as indicated in the scheme. When the inducible promoter was used, the TATA box in the 50 bp upstream region was removed as specified. Further changes of the C/D box sRNA gene or the up- and downstream region are detailed in the respective sections.

Northern blot analyses with total RNA from *S. acidocaldarius* cells harboring the C/D box sRNA Sac-sR121* *in vivo* constructs and a 16 nt long LNA probe hybridizing with the boxC sequence as well as the artificial guide sequence reveal that TATA-boxes in the upstream region are recognized during transcription initiation (figure 2.10A lane 2 and 3). For this reason, the TATA-box sequence was exchanged against a GC-containing sequence in constructs in which the inducible promoter was used. This design ensures the presence of a single transcription start site (figure 2.9; figure 2.10A lane 4).

The transcripts that can be detected in the Northern blot experiments reveal a size between 50 and 80 nt. This corresponds to the size of a mature C/D box sRNA. Inverse RT-PCR analyses and sequencing of the amplified products verified the existence of a mature C/D box sRNA transcript. Consequently, the upstream region of the C/D box sRNA (that results from the transcription start as determined by the external promoter) was processed and transcription was either terminated at the C/D box sRNA 3' terminus or processing of a longer transcript occurred. As a consequence of this observation, either a transcription termination or a processing 'signal' has to exist. An indication that the transcription was not terminated (or not completely terminated) was revealed by Northern blot experiments with total RNA from *S. acidocaldarius* cells harbouring C/D box sRNA Sac-sR121* in which the downstream region was exchanged against a sequence that is able to form a stable RNA hairpin. In this case a longer transcript is visible in the Northern blot analysis (figure 2.10A lane 5).

Assuming that the transcription is not terminated at the C/D box sRNA 3' terminus, a longer transcript has to be processed and processing of the 5' terminus definitely has to occur. The sequences of the upstream and downstream region of the C/D box sRNA do not seem to be important for processing. Mature C/D box sRNA Sac-sR121* transcripts are visible in Northern blot analyses with total RNA from cells harboring *in vivo* constructs in which the

upstream and downstream sequences are exchanged to random sequences (figure 2.10A lanes 6-9).

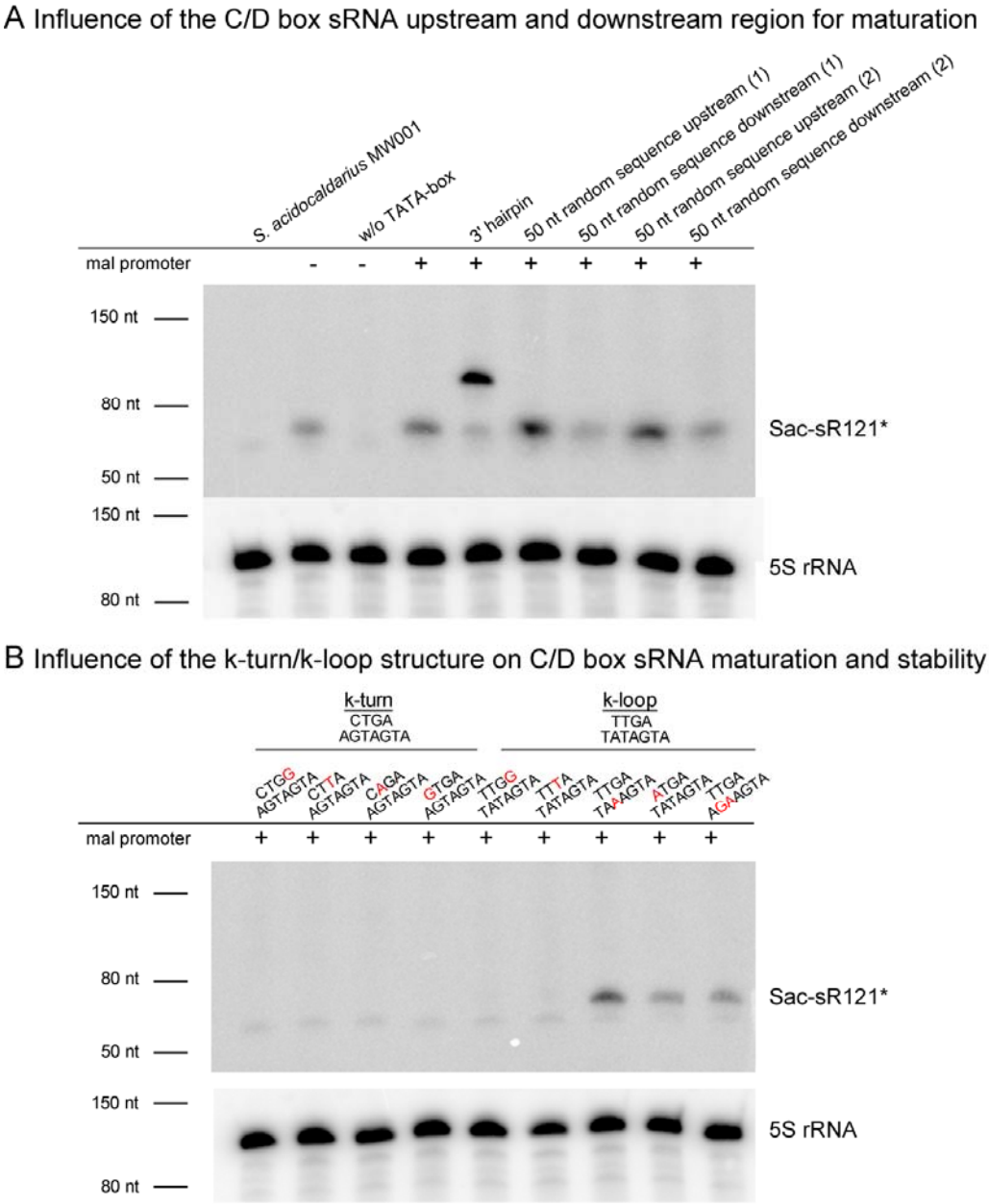

Figure 2.10: Northern blot analyses of C/D box sRNA Sac-sR121* *in vivo* constructs. The analyses were performed with 10 µg of extracted total RNA from *S. acidocaldarius* MW001 harboring plasmids with different versions of C/D box sRNA Sac-sR121* that exhibit the indicated differences in the promoter, in the C/D box sRNA up- and downstream sequence or within the C/D box sRNA sequence. A 16 nt 5'-radiolabeled LNA probe was used which hybridizes to the 12 nt artificial guide and 4 nt of the boxC motif. A probe complementary to 5S rRNA transcripts was used as loading control. (A) The existence of a TATA box in the 50 bp upstream sequence was determined. Exchange of the 50 bp upstream and downstream regions of C/D box sRNA genes to random sequences showed that these sequences are not important for C/D box sRNA maturation. Unspecific binding of the probe to *S. acidocaldarius* MW001 RNAs is visualized in lane 1. (B) K-turn and k-loop mutants were analyzed for their impact on C/D box sRNA maturation and stability. Mutations to the native boxC/C' and boxD/D' motifs that are indicated in the upper part of the figure are marked in red. Except of mutations in the conserved GA bp, mutations in the k-loop are tolerated whereas mutations in the k-turn are not tolerated at all.

If a processing signal (sequence or structure) existed within in the C/D box sRNA, it would be most likely located within the boxC/C' or boxD/D' motifs or the k-turn/k-loop structures as these are the only sequences and structures that are conserved within all C/D box sRNAs. C/D box sRNA Sac-sR121* transcripts harboring single mutations of the nucleotides of the boxD motif as well as of the two GA base pairs involved in k-loop formation do not show any signal in the Northern blot analyses (figure 2.10B lanes 1-6). Only single or double mutations disrupting the Watson-Crick base pair or the mismatch base pair involved in k-loop formation show mature C/D box sRNA transcripts (figure 2.10B lanes 7-9). The observation that changes are tolerated in the boxC' and boxD' motifs that are involved in k-loop formation also fits to sequence data of the boxC/C' and boxD/D' motifs of all 61 *S. acidocaldarius* C/D box sRNAs (figure 2.11). The sequence logos show that the boxC and boxD sequences are more conserved than the boxC' and boxD' sequences. However, the nucleotides of the boxC' and boxD' motifs that form the two GA base pairs in the k-loop are also highly conserved.

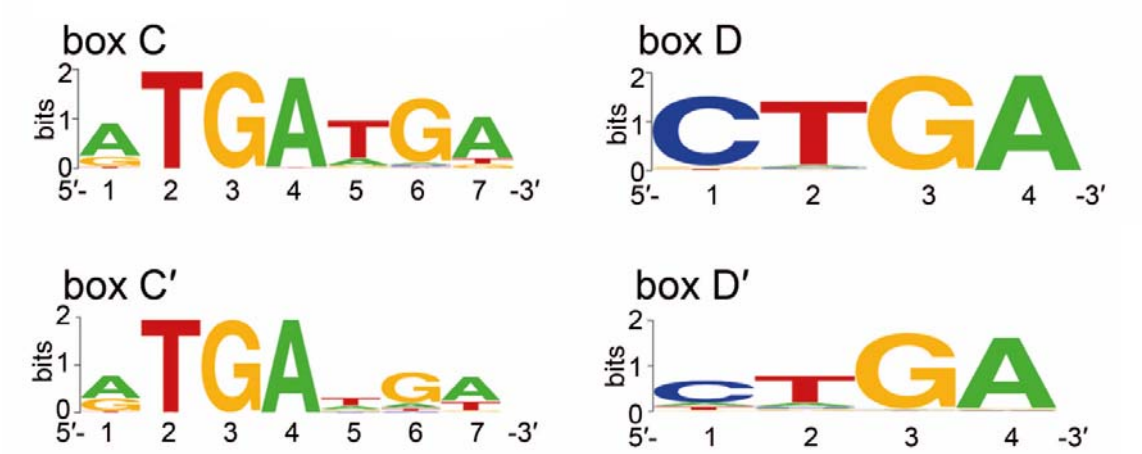

Figure 2.11: Sequence logos of *S. acidocaldarius* boxC/C' and boxD/D' motifs.

Northern Blot signals cannot be observed for C/D box sRNA Sac-sR121* k-turn mutants and GA base pair mutants of the k-loop, which might indicate that k-turn and k-loop formation is crucial for C/D box sRNA stability. When the sequences of the box motifs are not intact, L7Ae binding is abolished and the other C/D box sRNP proteins cannot assemble and stabilize the RNA. Nevertheless, the Watson-Crick base pair and the mismatch base pair seem to be only essential for L7Ae binding to the k-turn and not for binding to the k-loop. It is possible that the k-loop is formed faster due to the closer proximity of the boxC' and boxD' motifs in the primary sequence. Another possibility is, that L7Ae binding to the k-turn is more important in order to protect the 5' and 3' ends of the RNAs. Interestingly, few native C/D box sRNAs exist that do not possess a mismatch pair but a Watson-Crick bp in the k-turn or no GA-bp in the k-loop.

In general, it was observed that C/D box sRNA maturation is possible without the necessity of a specific sequence in the upstream and downstream region. This substantiates the

assumption that a specific endoribonuclease does not have to exist for maturation. Additionally, no specific sequence seems to be important for C/D box sRNA circularization. The performed iRT-PCR analyses revealed that all stable C/D box sRNA Sac-sR121* constructs are circular. Interestingly, in the Northern blot analyses a signal for a slower migrating form cannot be observed. This questions that the signal for the slower migrating RNA that could be observed in the Northern blot analyses presented in figure 2.5 is the circular C/D box sRNA variant as it was reported for *P. furiosus* C/D box sRNAs [100].

2.5 Production of C/D box sRNP proteins and assembly of the C/D box sRNP complex

To show that the binding of L7Ae to the k-turn and the k-loop is necessary for complete C/D box sRNP assembly, Electrophoretic Mobility Shift Assays (EMSAs) with C/D box sRNA *in vitro* transcripts containing mutated boxD/D' motifs were performed. In these transcripts k-turn and k-loop formation and thereby L7Ae binding should be inhibited and the addition of the remaining C/D box sRNP proteins Nop5 and fibrillarin should not lead to C/D box sRNP assembly. Therefore, *S. acidocaldarius* C/D box sRNP proteins were produced heterologously in *E. coli* and purified. L7Ae was overproduced with a N-terminal Sumo protein fusion and a 6x His-tag. Nop5 and N-terminal 6x His-tagged fibrillarin were overproduced together in a Duet vector as both proteins were pre-incubated in previous C/D box sRNP assemblies [55].

L7Ae could be purified via affinity chromatography using a nickel-NTA column and size-exclusion chromatography (figure 2.12). The Sumo/6x His-tag could be cleaved off using a sumo-protease which resulted in tag-free L7Ae that elutes at an elution volume of 13.9 to 18.0 ml with the majority of protein eluting between 15.0-18.0 ml (fractions B9-B7) from the size-exclusion column. According to a calibration curve, the peak corresponds to a molecular weight of approx. 30 kDa. The molecular weight of L7Ae is 13.6 kDa but the observed size shift could e.g. be caused by bound nucleic acid (figure 2.12B). EMSA studies with L7Ae and a k-turn possessing *in vitro* transcript that should allow binding of a single L7Ae molecule show a single mobility shift as expected (figure 2.13A and B). However, EMSA studies with L7Ae and C/D box sRNA *in vitro* transcripts show multiple shifts. Mutations of the boxD and boxD' motif nucleotides that are important for the formation of the GA base pair in the k-turn and k-loop did not abolish L7Ae binding (figure 2.13A, C and D).

Hence, the necessity of L7Ae binding to the C/D box sRNA for complete C/D box sRNP assembly could not be shown. An explanation for the binding of multiple L7Ae proteins to the C/D box sRNA *in vitro* transcript could be that unspecific L7Ae binding might be promoted by nucleic acid contaminations that could not be removed during the purification steps (figure

2.12B). Anion exchange chromatography with a MonoQ column yielded a small fraction of nucleic-acid free protein but it was very unstable and precipitated quickly.

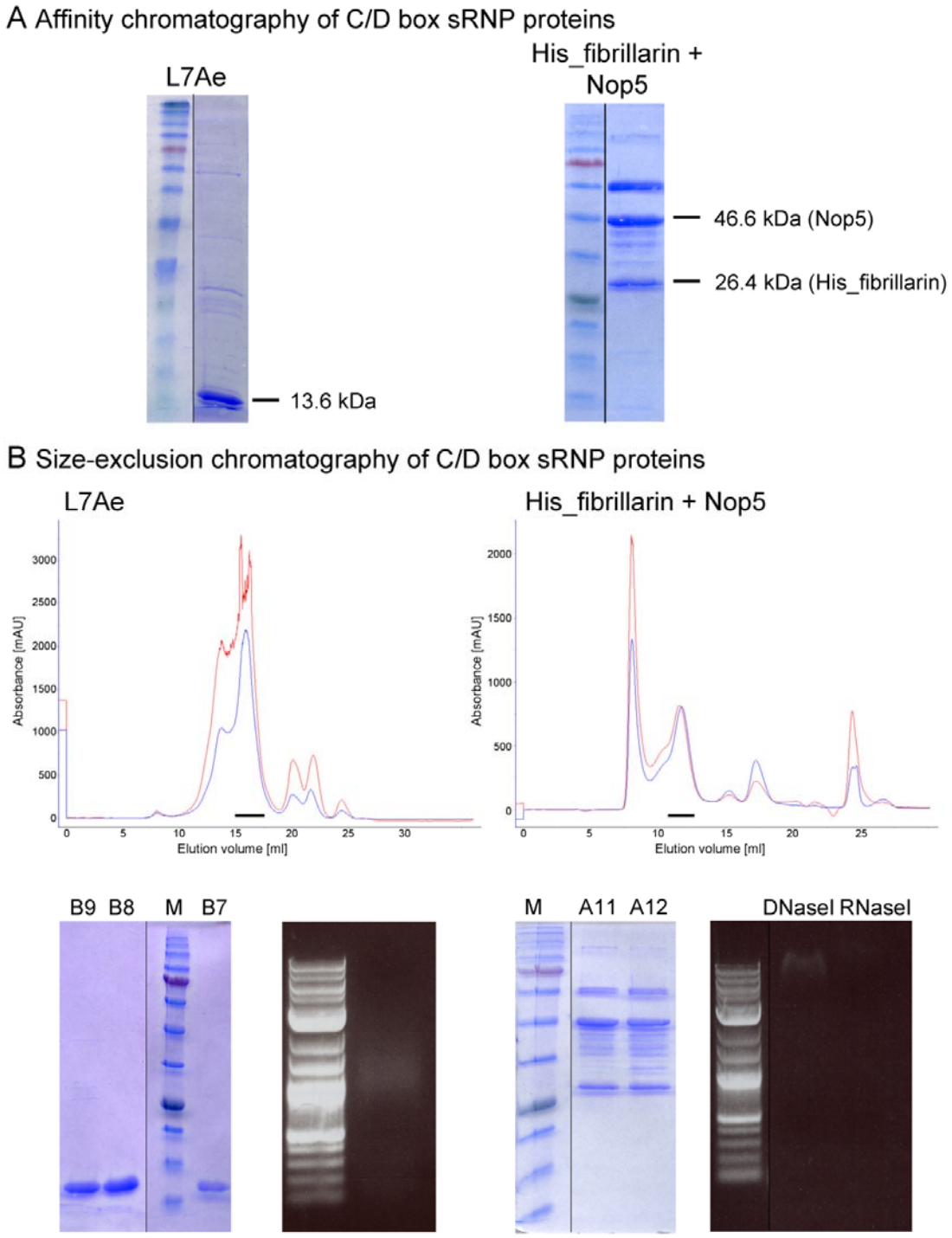

Figure 2.12: Purification of C/D box sRNP components. (A) SDS-PAGE of L7Ae and the fibrillarin-Nop5 subcomplex after Ni^{2+}-NTA-affinity chromatography. (B) Size-exclusion chromatography of L7Ae and the fibrillarin-Nop5 subcomplex. The marked peak fractions were analyzed via SDS-PAGE. Agarose gel electrophoresis of the pooled peak fractions revealed bound nucleic acids and DNaseI and RNaseI treatment of the fibrillarin-Nop5 subcomplex revealed that RNA is bound.

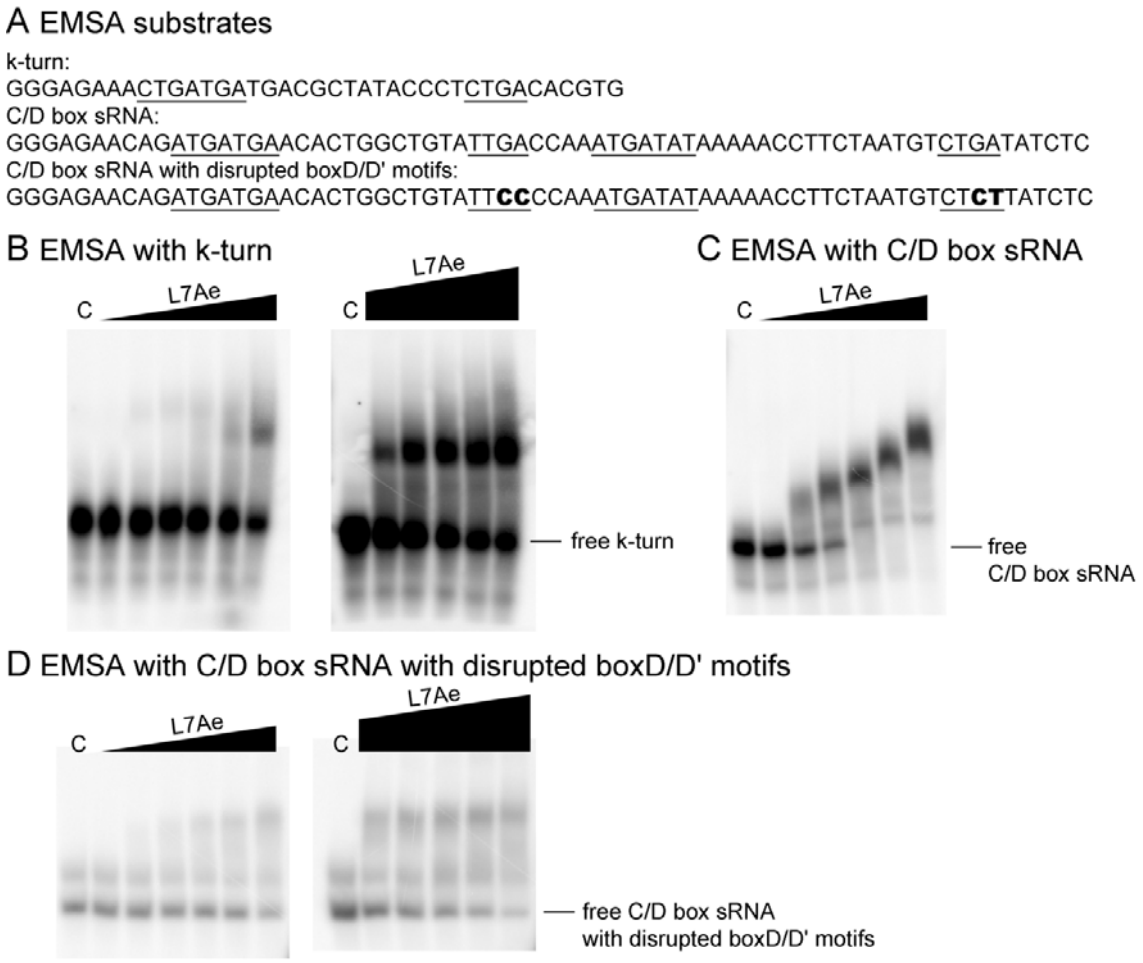

Figure 2.13: K-turn binding ability of L7Ae. Increasing concentrations of L7Ae were used for EMSA studies with different 5'-terminal radiolabeled substrates (16 nM). For the EMSAs in part B and D 0.035, 0.136, 0.272, 0.54, 1.0, 3.5 µM and 1.75, 3.5, 5.25, 7.0, 12.25 µM L7Ae were used, respectively. For the EMSA in part C 0.035, 0.136, 0.272, 0.54, 1.0 and 3.5 µM L7Ae was used. Bands were separated by 6 % native PAGE. (A) Used substrates comprise a single k-turn, a native C/D box sRNA and a C/D box sRNA with disrupted GA base pairs that should inhibit k-turn/k-loop motif formation. (B) A mobility shift with the k-turn as substrate can be observed starting at 1 µM L7Ae concentration. (C) Mobility shifts with the C/D box sRNA substrate can be observed starting at 0.136 µM L7Ae concentration and increase with increasing L7Ae concentrations. (D) A mobility shift with the C/D box sRNA with disrupted k-turn/k-loop motifs as substrate can be observed starting at 0.136 µM L7Ae concentration.

The fibrillarin/Nop5 subcomplex was purified via affinity chromatography using a nickel-NTA column and both proteins elute in the same fraction. This already indicates that both proteins form a complex (figure 2.12A). Subsequent size-exclusion chromatography reveals a peak at an elution volume of 8.8 ml to 13.0 ml with the majority of both proteins eluting between 10.0-12.0 ml (fractions A11/A12). According to a calibration curve, the peak corresponds to a molecular weight of approx. 150 kDa (figure 2.12B). Nop5 possesses a molecular weight of 46.6 kDa and fibrillarin a molecular weight of 26.4 kDa. Therefore, the peak fractions seem to contain a His_fibrillarin/Nop5 subcomplex with two copies of each protein. However, other proteins exist in the same elution fractions. Additionally, RNA is bound to the

His_fibrillarin/Nop5 subcomplex that could not be removed via anion exchange chromatography using a MonoQ column (figure 2.12B). C/D box sRNP assembly with a C/D box sRNA *in vitro* transcript, purified L7Ae and the His_fibrillarin/Nop5 subcomplex according to previous studies and subsequent size-exclusion chromatography did not reveal the formation of a complete C/D box sRNP [55]. Therefore, it could not be proven experimentally that L7Ae binding to the k-turn and the k-loop is necessary for complete C/D box sRNP assembly.

2.6 Maturation of dicistronic C/D box sRNA transcripts - *in vivo* assays with C/D box sRNA tandem Sac-sR125/126

Gene clusters of two or three C/D box sRNA genes were observed in multiple archaea and Northern blot and iRT-PCR analyses confirmed the transcription as dicistron for one of these arrangements in *S. acidocaldarius* (section 2.1 and 2.4.1). This opens the question how boxC and boxD motifs are selected for the specific maturation of two C/D box sRNAs because the formation of an alternative C/D box sRNA can be also imagined (figure 2.14A). An alternative transcript could not be observed in the RNA-Seq data. To analyze C/D box sRNA maturation specificity in a tandem arrangement, plasmid-based *in vivo* assays were performed similar to the *in vivo* assays described in section 2.4.2. The plasmid pSVA1431 was used as a backbone and the sequence of the tandem C/D box sRNA genes *Sac-sR125* and *Sac-sR126* was cloned behind the inducible maltodextrin promoter [114]. To easier follow the production of the alternative C/D box sRNA, two versions of the C/D box sRNA tandem were cloned. These constructs possess one complete C/D box sRNA gene and one C/D box sRNA gene half and are termed *in vivo* construct tandem version 1 and 2 (V1 and V2) (figure 2.14B). To enable detection of the transcripts via Northern blot analyses, one guide sequence was exchanged against an artificial 12 nt sequence. A 16 nt LNA probe hybridizing to the boxC motif as well as the artificial guide sequence was used for detection of the transcripts in Northern blot analyses. Further changes in the *in vivo* constructs are detailed in the respective section.

First, a sequence that is able to form a stable RNA hairpin was inserted into the downstream region of the V1 *in vivo* construct, to clarify that V1 *in vivo* constructs were transcribed as precursors and processing of the transcripts occurred (figure 2.16A; construct 5). Northern blot analyses with total RNA from cells harboring this construct revealed a longer transcript indicating that transcription is not terminated at the end of the C/D box sRNA tandem V1 (figure 2.16B; construct 5).

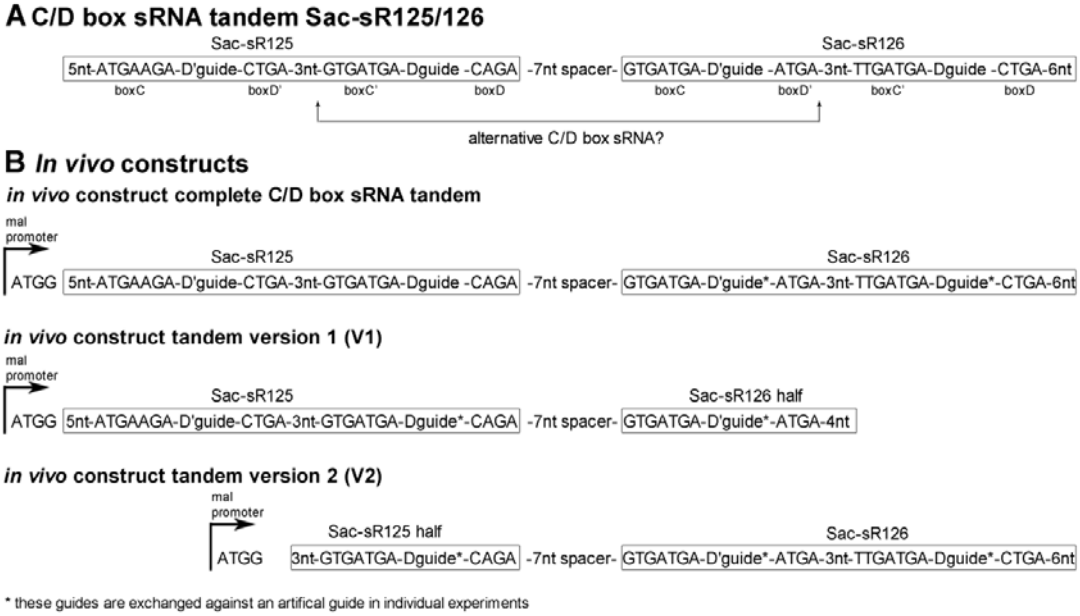

Figure 2.14: Scheme of the C/D box sRNA tandem Sac-sR125/126 and different *in vivo* constructs that were used for the analysis of the maturation of dicistronic C/D box sRNA transcripts. (A) The scheme shows the structure of the tandem C/D box sRNAs Sac-sR125/126. The alternative C/D box sRNA whose existence should be analyzed with the *in vivo* assays is indicated. (B) The *in vivo* assay constructs contain either (i) the complete C/D box sRNA tandem, (ii) a version without the second half of *Sac-sR126* beginning with the boxC' motif (V1) or (iii) a version without the first half of *Sac-sR125* containing the boxC and boxD' motifs (V2). Individual guide sequences are exchanged against the artificial guide in certain experiments (*). The constructs are cloned into the plasmid pSVA1431 that contains the inducible maltodextrin (mal) promoter as indicated in the scheme. Further changes of the C/D box sRNA genes are detailed in the respective section.

Previous analyses of archaeal C/D box sRNAs showed that the majority of guide sequences has a length of 12 nt and that the distance between the boxD' and boxC' motifs is almost exclusively 2 nt [17]. Analyses of the length of the guide sequences and the distance between the boxD' and boxC' motifs of all 61 *S. acidocaldarius* C/D box sRNAs revealed that most of the guide sequences have a length of 12 nt and the distance between the boxD' and boxC' motifs is 3 nt or 4 nt in most cases (figure 2.15). Both guide sequences of the alternative C/D box sRNA would be 12 nt long but the spacer between the boxD' and boxC' motifs would be 7 nt in the alternative C/D box sRNA. *In vivo* tandem V1 constructs were designed in which the spacer length between both C/D box sRNA genes was varied from native 7 nt to 3 nt and 2 nt, to test whether the length of the spacer sequence influences the possible creation of an alternative C/D box sRNA (figure 2.16A; constructs 1, 2 and 3). Northern blot analyses with total RNA from cell harboring these *in vivo* constructs revealed a transcript with a size of around 90 nt (figure 2.16B; constructs 1, 2 and 3). Besides the commercially available size standard, two additional size markers were used. On the one hand, a 89 nt long *in vitro* transcript of a C/D box sRNA tandem V1 with additional 10 nt facilitating the start of the transcription was used. On the other hand, a C/D box sRNA Sac-sR126 *in vivo* construct with the artificial guide sequence served as size marker (figure 2.16

and figure 2.17A and B; construct C2). The different size markers indicate that visible transcripts with the variable spacer length correspond to the C/D box sRNA tandem V1 and the alternative C/D box sRNA cannot be detected. Interestingly, a C/D box sRNA Sac-sR125 control *in vivo* construct did not reveal transcripts in Northern blot analyses with total RNA from cells harboring this construct (figure 2.17A and B; construct C1). Cells harboring this *in vivo* construct are additionally impaired in growth (figure 2.17C and D; construct C1).

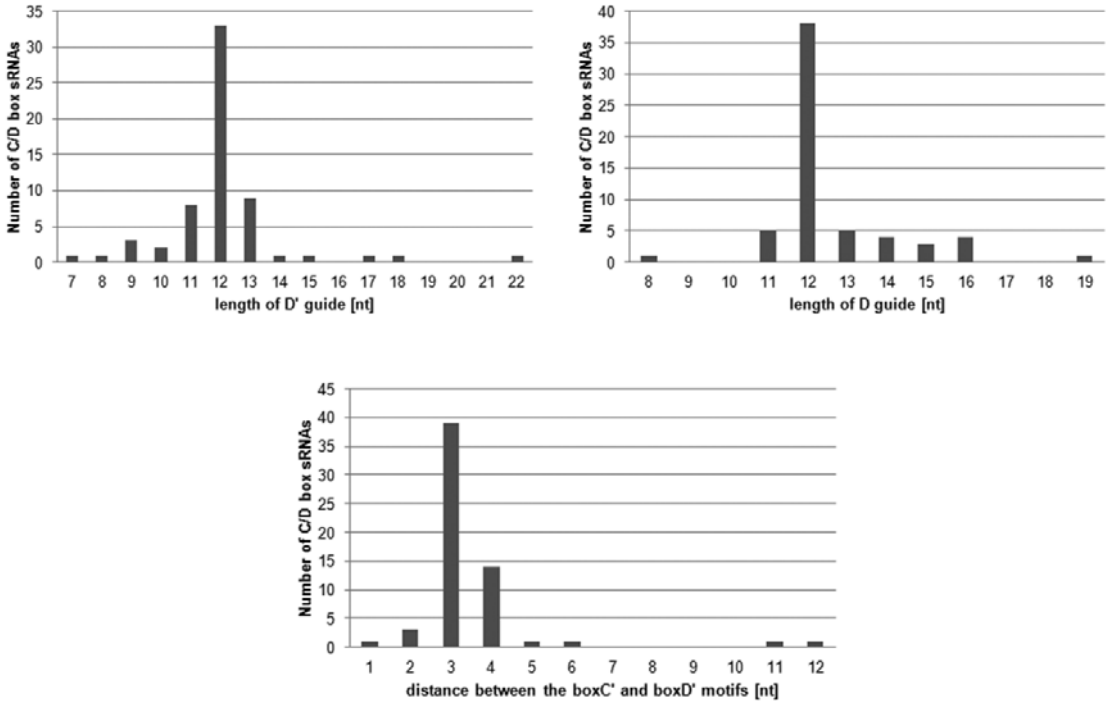

Figure 2.15: Analyses of length distributions of C/D box sRNA guide sequenes and the spacer region between the boxC' and boxD' motifs. The analysis includes all 61 *S. acidocaldarius* C/D box sRNAs.

The positions of the artificial guide in the analyzed V1 *in vivo* constructs only allowed for the detection of the alternative C/D box sRNA but the existence of synthetic Sac-sR125 could not be examined. For this reason the position of the artificial guide region was changed (figure 2.16A; construct 4). Surprisingly, this construct cannot be transformed or the cells might exhibit a significant growth defect that counteracts transformation. Transformation assays in which the same amount of plasmid was transformed into competent *S. acidocaldarius* MW001 cells in triplicates illustrate this phenomenon (figure 2.16C; construct 4).

Next, complete tandem C/D box sRNA *in vivo* constructs were analyzed, to check whether the problems only occur with V1 *in vivo* constructs as these are unnatural arrangements of a complete C/D box sRNA and a C/D box sRNA half. The *in vivo* constructs have the native arrangement of the C/D box sRNA *Sac-sR125* and *Sac-sR126* genes but the artificial guide sequence is inserted at different positions. Additionally, the space between both C/D box

sRNA genes was exchanged to 2 nt in one *in vivo* construct (figure 2.16A; constructs 6, 7 and 8). Northern blot analyses with total RNA from cells harboring the different complete tandem *in vivo* constructs revealed a transcript with a size of approximately 140 nt and this size corresponds to the size of a C/D box sRNA dicistron. For the construct with the shortened spacer sequence a second band with a size of approximately 90 nt is visible but a mature C/D box sRNA could not be observed in any of the experiments (figure 2.16B; constructs 6, 7 and 8). Surprisingly, cells harbouring the complete tandem C/D box sRNA *in vivo* construct. in which the D' guide sequence of *Sac-sR126* was exchanged against the artificial guide sequence, show significant growth defects (figure 2.16C and D; construct 6).

Mutations were introduced in the conserved GA base pairs of the boxC and boxD' motifs and these boxes should not be able to participate in k-turn formation, to elucidate why the tandem V1 *in vivo* constructs were stabilized and a mature C/D box sRNA was not produced (figure 2.17A; constructs 1-4). The V1 tandem construct with a mutated boxD' motif in the Sac-sR126 half and an artificial Sac-sR125 D guide cannot be transformed or the cells might exhibit a significant growth defect counteracting observation of transformants (figure 2.17C; construct 4). Northern blot analyses with total RNA of cells harboring the remaining constructs, reveal transcripts with a size of approximately 90 nt (figure 2.17B; constructs 1-3). These results show that the tandem V1 *in vivo* constructs are stably maintained despite of the introduced mutations.

In addition, it should be tested whether the non-consensus sequences of several box motifs prohibited the formation of the alternative C/D box sRNA. Therefore tandem V1 *in vivo* constructs were designed in which the non-consensus sequences of the boxC, boxD and boxD' motifs were changed to the consensus sequence, either individually or combined that all boxes had the consensus sequence (figure 2.17A; constructs 5, 6 and 7). Individual mutations in the box sequences lead to the formation and stabilization of a transcript with approximately 90 nt that was also observed for other tandem V1 *in vivo* constructs and mature C/D box sRNA transcripts are not visible (figure 2.17B; constructs 5 and 6). When all box motifs have the perfect consensus sequence, the *in vivo* construct is not transformable or the cells might have a serious growth defect (figure 2.17C; construct 7).

As the Northern blot analyses with the tandem V1 *in vivo* constructs did neither show the alternative C/D box sRNA transcript nor a mature synthetic C/D box sRNA Sac-sR125 transcript, further experiments were performed with tandem V2 *in vivo* constructs that possess a C/D box sRNA *Sac-sR125* half and the complete C/D box sRNA *Sac-sR126* (figure 2.14). To identify the guide sequence with the best-tolerated artificial guide, three constructs were created in which each of the three guide sequences was individually exchanged with the artificial guide sequence (figure 2.18A; constructs 1-3).

Northern blot analyses with total RNA from cells harboring these three constructs show a transcript with a size of approximately 100 nt for all constructs. The increased size compared to tandem V1 *in vivo* constructs is due to a size difference of 8 nt between C/D box sRNA Sac-sR125 and C/D box sRNA Sac-sR126. In none of the cases a mature C/D box sRNA or the alternative C/D box sRNA transcript are visible. Cells that harbor the tandem V2 *in vivo* construct in which the sequence of the D' guide of C/D box sRNA *Sac-sR126* was exchanged, show a slightly impaired growth phenotype (figure 2.18C and D; construct 2). The transcript of the *in vivo* construct with the artificial guide sequence instead of the D guide of C/D box sRNA Sac-sR126 shows the strongest signal. Therefore, additional tests to elucidate how the visible tandem V2 *in vivo* transcripts are stabilized were performed with the unique guide at this position.

Tandem V2 *in vivo* constructs with poly-A sequences instead of boxC' or boxD motifs that should inhibit k-turn formation still reveal stable transcripts with a size of approximately 100 nt in Northern blot analyses but only with total RNA from cells that harbor the construct with the mutated boxD motif in *Sac-sR125*. The remaining two mutants do not show a transcript in Northern blot analyses (figure 2.18A and B; constructs 4-6). This indicates that L7Ae binding is responsible for the maintenance of the transcript with a size of approximately 100 nt. However, a mature C/D box sRNA cannot be stably maintained.

Additionally, V2 *in vivo* constructs were designed in which non-consensus sequences of several box motifs were changed to consensus sequence, to check whether these changes could induce the formation of the alternative C/D box sRNA (figure 2.18A; constructs 7-9). Northern blot analyses with total RNA from cells harboring *in vivo* constructs in which mutations in individual box motifs are introduced, show the approximately 100 nt transcript. Additionally, a smaller transcript is visible but the alternative C/D box sRNA is not stabilized (figure 2.18B; constructs 7,8). The *in vivo* construct in which all box motifs possess consensus sequences is not transformable or the cells harboring this *in vivo* construct show a significant growth defect (figure 2.18C; construct 9).

In general, it seems that the processing of C/D box sRNA tandem transcripts is problematic, which is possibly due to multiple binding options for the protein L7Ae. However, it could not be clarified why the tandem V1 and V2 transcripts were stably maintained. It might be that the exchange of the guide sequences to the artificial guide sequence was not tolerated for the Sac-sR125/126 tandem of *S. acidocaldarius*.

rRNA targets of C/D box sRNA Sac-sR126 could not be predicted but the D guide region of C/D box sRNA Sac-sR125 targets a nucleotide in helix 69 and the D' guide targets a nucleotide in helix 70 of the 23S rRNA (figure 2.19). Helix 69 is a 2'-*O*-methylation hotspot in archaeal C/D box sRNAs (see section 2.7.1). To test whether the methylation by the D guide of C/D box sRNA Sac-sR125 is crucial, the guide should be exchanged in the genome of

S. acidocaldarius MW001. This was not successful even if a second copy of C/D box sRNA *Sac-sR125* was introduced in the genome (in the α-amylase locus) [132]. However, the C/D box sRNA *Sac-sR125* guide was located directly downstream of *saci_1350*, encoding for the essential subunit E of the glutamyl-tRNA(Gln) amidotransferase and it cannot be excluded that the D guide exchange disrupts regulatory elements of *saci_1350*.

In conclusion, the plasmid-based *in vivo* studies could show that a dicistronic C/D box sRNA Sac-sR125/126 transcript is made and stabilized. However, either the complete dicistron or the artificial tandem V1 and V2 transcripts are stably maintained and mature C/D box sRNAs are not generated. Surprisingly, several constructs cannot be transformed or the cells exhibit significant growth defects. This was not expected as the C/D box sRNA tandem Sac-sR125/126 additionally exists in its native locus. It remains to be seen whether these effects arise because of the overexpression of the *in vivo* constructs with the altered guide region whereby the native RNAs are outcompeted from efficient modification of their RNA targets.

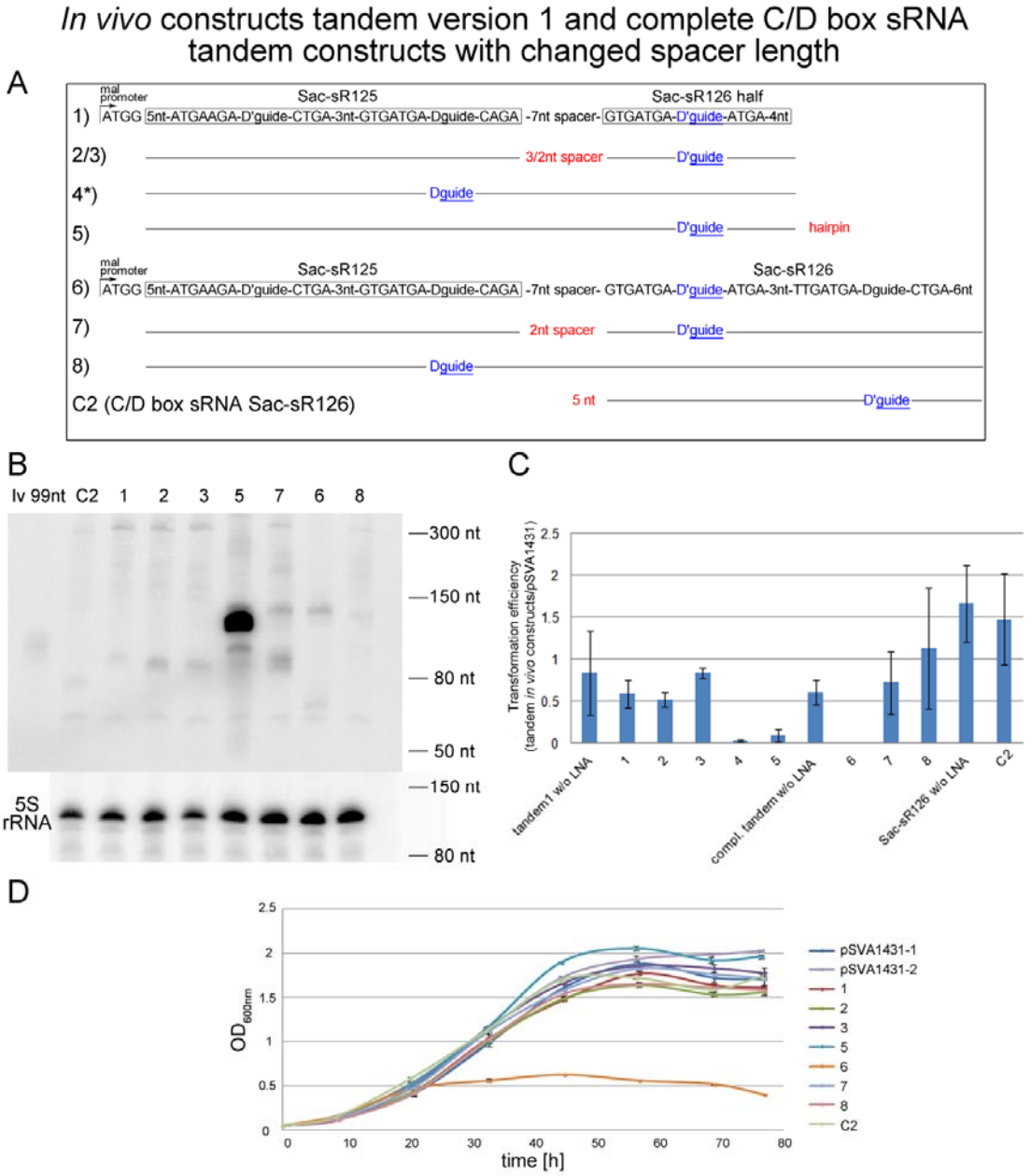

Figure 2.16: *In vivo* assays with complete tandem constructs and tandem V1 constructs. (A) In the detailed scheme of the tested *in vivo* constructs differences within the constructs are indicated in red. The position of the artificial guide is indicated in blue. Numbering of the constructs is consistent in the figure. Cells harboring the construct marked with the asterisk (*) could not be transformed or showed a strong growth deficiency. (B) Northern blot analyses of C/D box sRNA tandem *in vivo* constructs were performed with 10 µg extracted total RNA from *S. acidocaldarius* MW001 harboring plasmids with different complete tandem or tandem V1 constructs. As additional control a 99 nt *in vitro* transcript is used. A 16 nt 5'-γ-[^{32}P]-ATP LNA probe hybridizing to the 12 nt artificial guide and 4 nt of the boxC motif was used for detection of the transcripts. A probe complementary to 5S rRNA transcripts was used as loading control. (C) In a transformation assay, 100 ng plasmid DNA with the different *in vivo* constructs was transformed in triplicates and the colonies were counted after 5 days of incubation. The transformation efficiency was calculated by comparing the amount of colonies that were obtained for the transformation of the *in vivo* constructs versus the amount of colonies that were obtained for the transformation of the empty plasmid pSVA1431. *In vivo* constructs without the artificial guide (w/o LNA) serve as controls. (D) In the growth curve of *S. acidocaldarius* MW001 harboring plasmids with the different constructs, the control pSVA1431-2 is valid for construct 5 and the control pSVA1431-1 for the remaining constructs. The error bars represent the standard deviation of three cultures that were grown simultaneously.

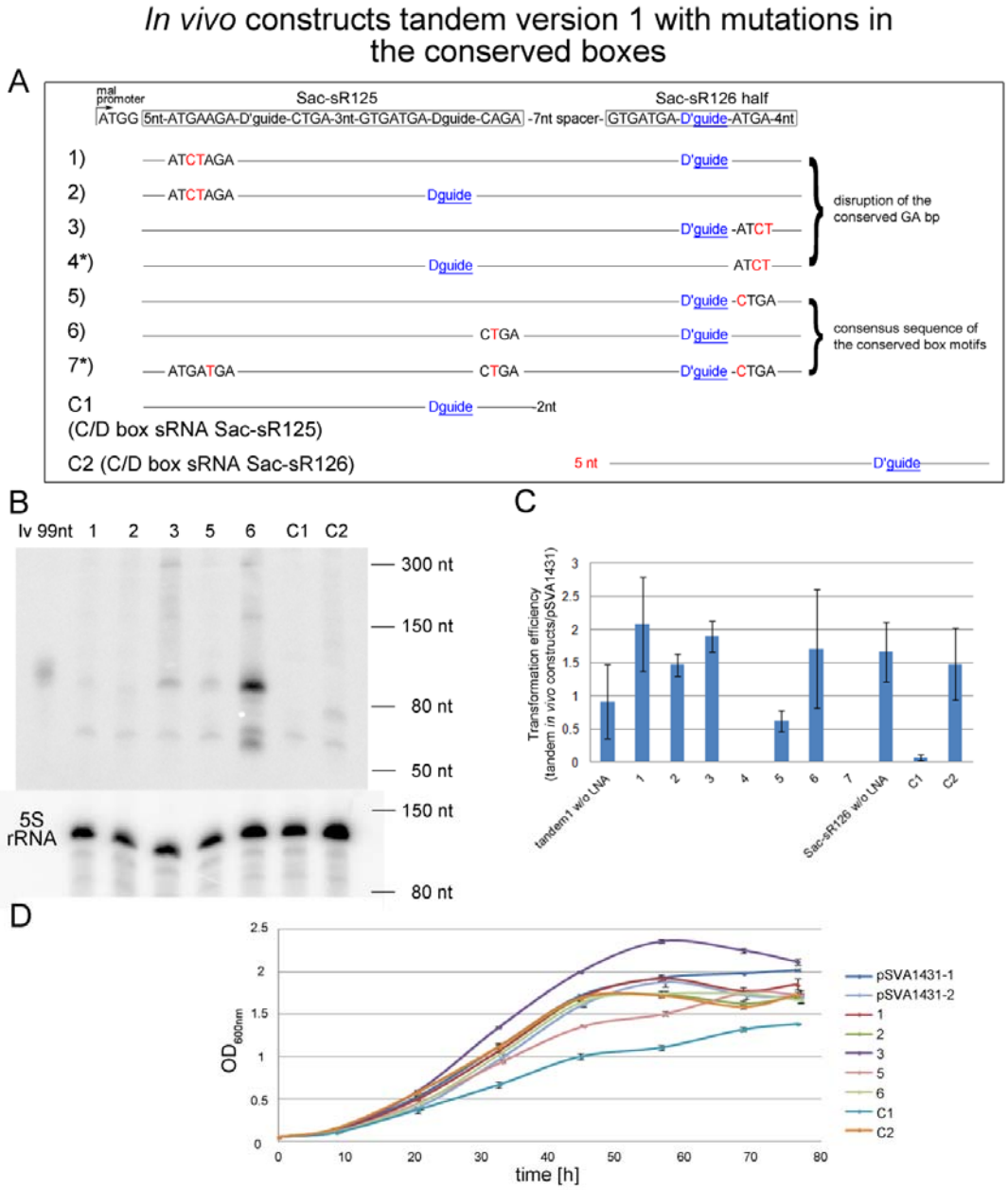

Figure 2.17: *In vivo* assays with tandem V1 constructs with mutations in the conserved boxes. (A) The detailed scheme of the tested *in vivo* constructs indicates different nucleotides of the boxC and boxD motifs that were mutated (red) and the position of the artificial guide (blue) that differs. Numbering of the constructs is consistent in the figure. Cells harboring the constructs marked with the asterisk (*) could not be transformed or showed a strong growth deficiency. (B) Northern blot analyses of C/D box sRNA tandem *in vivo* constructs were performed with 10 µg extracted total RNA from *S. acidocaldarius* MW001 harbouring plasmids with different tandem V1 constructs. As additional control a 99 nt *in vitro* transcript is used. A 16 nt 5'-γ-[^{32}P]-ATP LNA probe hybridizing to the 12 nt artificial guide and 4 nt of the boxC motif was used for detection of the transcripts. A probe complementary to 5S rRNA transcripts was used as loading control. (C) In a transformation assay, 100 ng plasmid DNA with the different *in vivo* constructs was transformed in triplicates and the colonies were counted after 5 days of incubation. The transformation efficiency was calculated by comparing the amount of colonies that were obtained for the transformation of the *in vivo* constructs versus the amount of colonies that were obtained for the transformation of the empty plasmid pSVA1431. *In vivo* constructs without the artificial guide (w/o LNA) serve as controls. (D) In the growth curve of *S. acidocaldarius* MW001 harboring plasmids with the different constructs, the control pSVA1431-2 is valid for constructs 5 and 6 and the control pSVA1431-1 for the remaining constructs. The error bars represent the standard deviation of three cultures that were grown simultaneously.

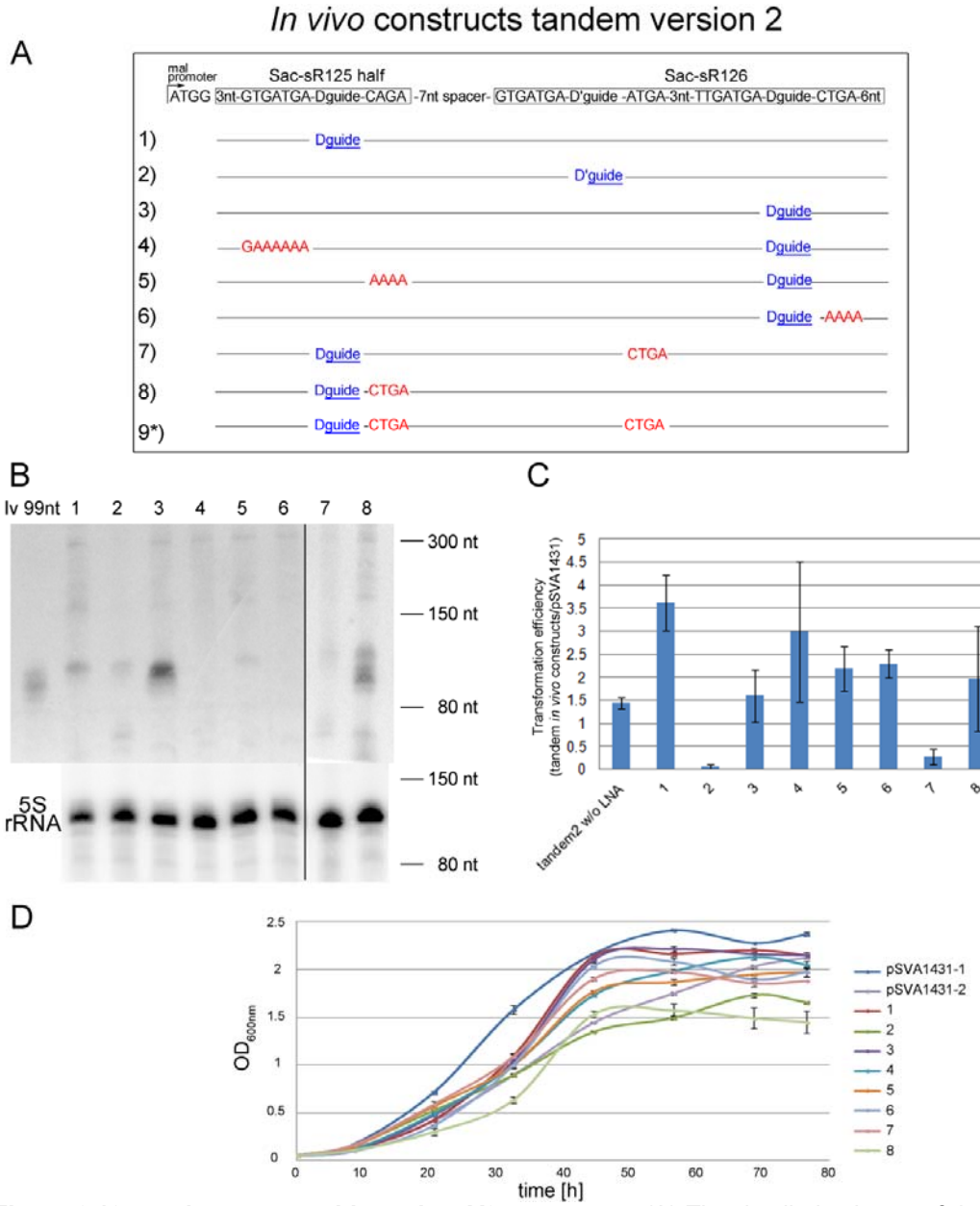

Figure 2.18: *In vivo* assays with tandem V2 constructs. (A) The detailed scheme of the tested *in vivo* constructs indicates different nucleotides of the boxC and boxD motifs that were mutated (red) and the position of the artificial guide (blue) that differs. Numbering of the constructs is consistent in the figure. Cells harboring the construct marked with the asterisk (*) could not be transformed or showed a strong growth deficiency. (B) Northern blot analyses of C/D box sRNA tandem *in vivo* constructs were performed with 10 µg extracted total RNA from *S. acidocaldarius* MW001 harboring plasmids with different tandem V2 constructs. As additional control a 99 nt *in vitro* transcript is used. A 16 nt 5'-γ-[^{32}P]-ATP LNA probe hybridizing to the 12 nt artificial guide and 4 nt of the boxC motif was used for detection of the transcripts. A probe complementary to 5S rRNA transcripts was used as loading control. (C) In a transformation assay, 100 ng plasmid DNA with the different *in vivo* constructs was transformed in triplicates and the colonies were counted after 5 days of incubation. The transformation efficiency was calculated by comparing the amount of colonies that were obtained for the transformation of the *in vivo* constructs versus the amount of colonies that were obtained for the transformation of the empty plasmid pSVA1431. *In vivo* constructs without the artificial guide (w/o LNA) serve as control. (D) In the growth curve of *S. acidocaldarius* MW001 harboring plasmids with the different constructs, the control pSVA1431-2 is valid for construct 2 and the control pSVA1431-1 for the remaining constructs. The error bars represent the standard deviation of three cultures that were grown simultaneously.

2.7 C/D box sRNA guided 2'-O-methylation patterns of archaeal rRNA molecules

For several bacteria and eukaryotes maps of post-transcriptional modifications exist and several hotspots could be identified in which modifications cluster e.g. near the peptidyl transferase center, the A, P and E sites of tRNA and mRNA binding and the interacting region of small and large ribosomal subunits [5]. No complete overview of the 2'-*O*-methylation target sites exists for archaea. A previous study with the aim to map the complete methylation targets of archaeal C/D box sRNAs used C/D box sRNAs of *S. acidocaldarius* that were coimmunoprecipitated with fibrillarin and Nop56. In this study only 18 different C/D box sRNAs were identified and the predicted methylation targets were verified with primer extension analysis with low dNTP concentration [1]. Another study that also included *S. acidocaldarius*, but additionally *Haloarcula marismortui*, only focused on the postranscriptional modifications of helices 90-92 in domain V of the 23S rRNA [26].

Therefore, the 2'-*O*-methylation target sites in archaeal rRNAs were predicted and mapped onto the consensus structures of these RNAs. Included in this study are the C/D box sRNA guide sequences of the six archaea shown in table 2.1 and the guide sequences of the crenarchaeon *Pyrobaculum calidifontis* are included additionally [133]. In total, a list of 489 C/D box sRNAs was compiled and shared with our collaborators Lauren Lui (group of Todd Lowe, University of California) and Prof. Patrick Dennis (Janelia Research Campus, Howard Hughes Medical Institute) who performed target predictions. In total, 719 guide interactions with 16S and 23S rRNA molecules could be identified. The modifications are introduced site-specifically at the nucleotide of the target RNA that is complementary to the fifth nucleotide upstream of the box D/D' motif. Interestingly, for one third of the guide regions, rRNA targets could not be identified.

The data were also used to by our collaborators to analyze the evolution of C/D box sRNA guide sequences and a function of C/D box sRNAs acting as RNA chaperons facilitating rRNA folding is proposed as the two guides can target rRNA positions that are not in proximity in the primary sequence but are close in the secondary structure.

453 2'-*O*-methylation sites are predicted in the 23S rRNA. They are illustrated in the secondary structure of the 23S rRNA of *H. marismortui* which is based on the crystal structure of the large ribosomal subunit [52] (figure 2.19). To visualize the modifications in the 23S rRNA secondary structure, a Java based algorithm was developed (Dr. Lennart Randau). The input are the relative x and y-pixel values of individual rRNA nucleotides to represent the overall rRNA structure and individual parameters as the predicted 2'-*O*-methylation sites were included additionally. In total, 334 different nucleotides become modified in the 23S rRNA and 2'-*O*-methylations by multiple archaea at a given nucleotide are displayed by increasing dot size. The color change from blue to red indicates the

methylation conservation within a nine nucleotide window i.e. also the methylation status of the nucleotides in the vicinity is included.

The primary sequence and secondary structure of the 23S rRNA are highly conserved [134]. The secondary structure shows a central core and six domains. 2'-*O*-methylation hotspots can be observed within hairpins 35 and 35a in domain II, helices 61 and 68 to 71 in domain IV and hairpins 90 to 93 in domain V.

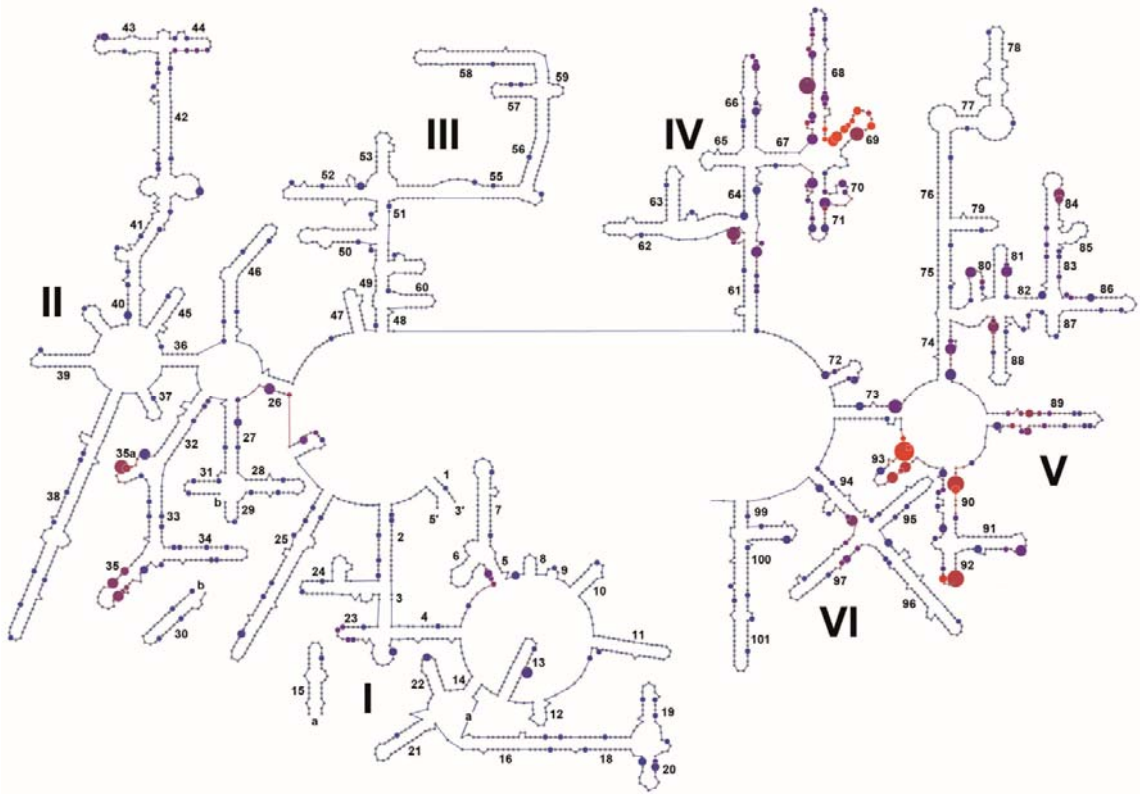

Figure 2.19: Distribution of archaeal 2'-O-methylation sites in 23S rRNA. The methylation site targets for C/D box sRNA guides from seven archaeal species are mapped onto the 23S rRNA structure that is based on the crystal structure from *Haloarcula marismortui* [134]. Increasing dot size indicates occurrence of methylation by several organisms at a given nucleotide. The color change of the dots from blue to red indicates methylation conservation within a nine nucleotide window.

266 2'-*O*-methylation sites were predicted for the 16S rRNA that occur at 195 different positions. For the visualization of the 16S rRNA secondary structure of *S. acidocaldarius* and the distribution of the 2'-*O*-methylated nucleotides, the Java based visualization algorithm was used (figure 2.20). The 16S rRNA consists of a central core that is close to the decoding center of the small ribosomal subunit where codon/anticodon complementarity is monitored. The central core connects the four domains of the RNA [135]. Most conserved 2'-*O*-methylations can be identified in helices 3,18 and 27 in the central core.

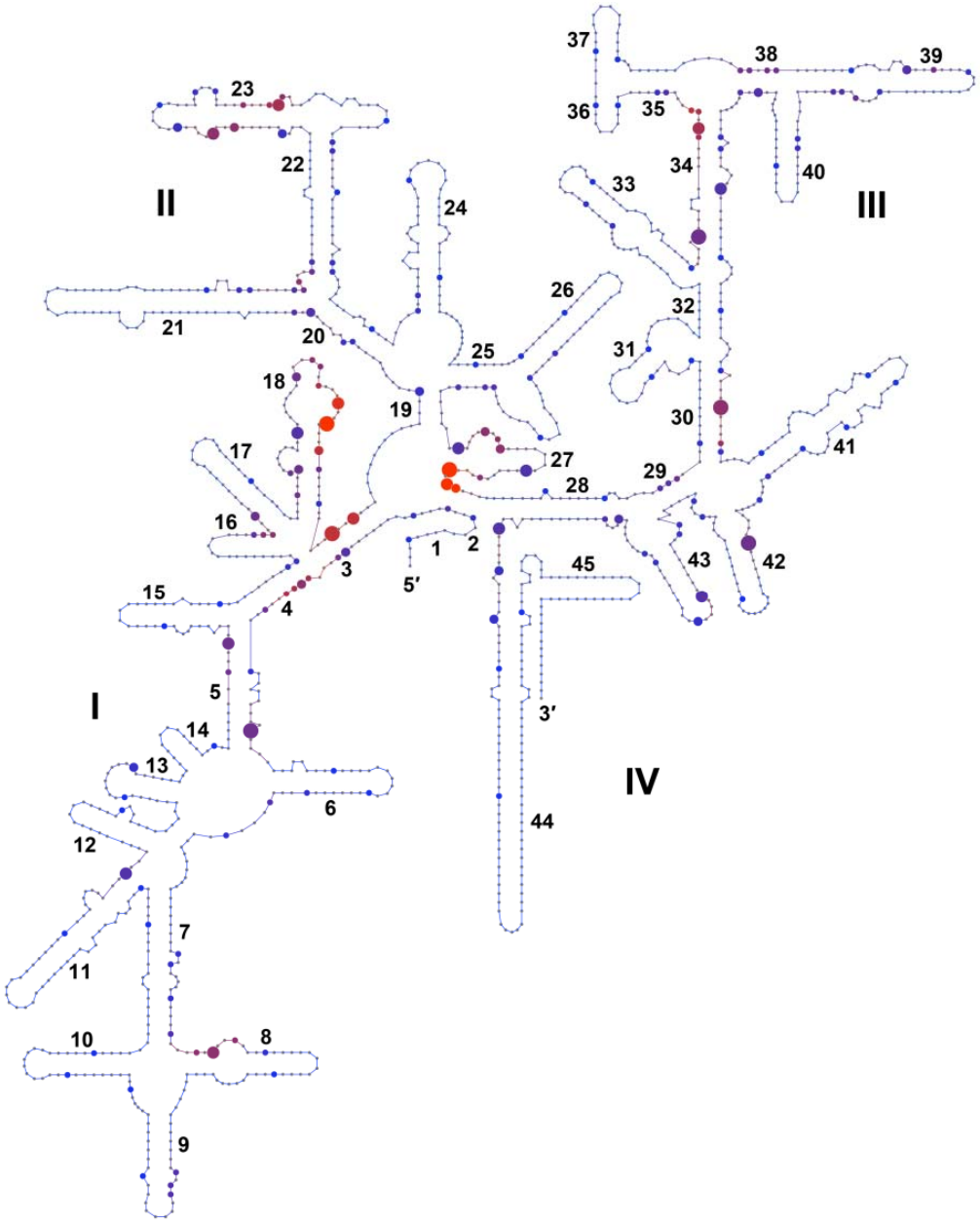

Figure 2.20: Distribution of archaeal 2'-O-methylation sites in 16S rRNA. The methylation site targets for C/D box sRNA guides from seven archaeal species are mapped onto the consensus 16S rRNA structure [134]. Increasing dot size indicates occurrence of methylation by several organisms at a given nucleotide. The color change of the dots from blue to red indicates methylation conservation within a nine nucleotide window.

3. Discussion

3.1 Genetic context of C/D box sRNA genes

Previous studies revealed that a large variety of C/D box snoRNA gene arrangements exists in eukaryotes with a trend towards the reduction of the number independent promoters from lower to higher eukaryotes [83, 136, 137]. The absence of independent promoters is possible by polycistronic arrangements of C/D box snoRNA genes or their localization in introns of protein-coding and non-protein-coding genes [78]. Archaeal C/D box sRNA genes seem to exhibit similar characteristics. Already in the 1990s, mutational studies indicated that archaeal promoters share similarities with eukaryotic polymerase II promoters [138, 139]. The data revealed similarities (i) to the TATA box sequences approximately 26 nt upstream of the transcription start site, (ii) to the transcription factor B recognition element (BRE) upstream of the TATA box and (iii) to the initiator element near the transcription start site [119]. Transcriptome data allowed for a global analysis of archaeal TATA box promoter elements [124, 140]. A computational search for conserved sequences within the 50 nt upstream region of 343 C/D box sRNA genes from six archaea (Table 2.1) revealed AT-rich sequences that resemble the archaeal consensus TATA box promoter elements. Different conserved features were not detected. However, TATA box sequences that suggest C/D box sRNA transcription from independent promoters could only be observed for a small amount of C/D box sRNA genes in this pan-archaeal analysis. The majority of C/D box sRNA genes did not exhibit promoters for primary transcript production. A detailed analysis of archaeal C/D box sRNA genes revealed a diverse genomic context providing opportunities for C/D box sRNA transcription in the absence of independent promoters.

A significant amount of C/D box sRNA genes overlaps with flanking ORFs or is located directly at their 5' or 3' ends [7, 9, 17, 107]. These C/D box sRNA gene localizations imply that the transcripts of the flanking genes possess a 5' and a 3'-UTR, respectively. The existence of 3'-UTRs was observed in archaea but the existence of C/D box sRNAs in 5'-UTRs is interesting as most archaeal transcripts lack a 5'-UTR, a phenomenon termed leaderless transcripts [125, 141]. One example is *S. solfataricus*, in which the transcription of 69 % of protein-coding transcripts starts at the 'A' of the ATG translation start codon or 1-3 bases upstream. However, several longer 5'-UTRs exist and approximately 6 % of the transcripts possess 5'-UTRs that are larger than 20 nucleotides [124]. Analysis of the longer 5'-UTRs in *S. solfataricus* showed that two UTRs exhibit C/D box sRNA characteristics and this demonstrates that C/D box sRNAs can indeed be part of 5'-UTRs of the flanking protein-coding sequences even though the majority of archaeal transcripts is leaderless.

The existence of C/D box sRNAs in 5' and a 3'-UTRs implies that the genes do not need to possess an own promoter. The observation that the amount of individual C/D box sRNAs differs significantly may be related to the strength and regulation of the promoters that were

'hijacked'. It can be excluded that the differences are RNA folding artefacts as all C/D box sRNAs possess a similar size and structure and the structure is mainly stabilized by the C/D box sRNP proteins but these are removed during the RNA isolation procedure. Additionally, the RNA-Seq library preparation protocol includes a denaturation step were RNA secondary structures are resolved.

In addition to C/D box sRNA genes that overlap with flanking ORFs, kink-turn (k-turn) motifs that overlap with the 5' or 3' end of coding sequences were identified in *S. solfataricus* transcripts by immunoprecipation studies with the k-turn binding protein L7Ae and its binding partners from *S. solfataricus* cell extracts [40]. The occurence of k-turn motifs or C/D box sRNAs at the 5' end of flanking protein-coding sequences is curious due to the predominant existence of leaderless transcripts. *In vitro* and *in vivo* studies with synthetic k-turn motifs in 5'-UTRs showed that L7Ae binding to the k-turn motifs acts as efficient translational block [77]. Additionally, C/D box sRNA-reporter gene fusions revealed that C/D box sRNA genes that were fused to the 5' end of the reporter gene negatively influence the enzyme activity. However, the effect could not clearly be shown to be dependent on L7Ae binding as an effect was already observed on the transcriptional level. C/D box sRNA gene fusions to the 3' end of the reporter gene generated either neutral or negative effects (master thesis of Roman Martin). The mRNA stability does not seem to be influenced when C/D box sRNAs exist on the same transcripts. Studies of the mRNA half-life in *S. solfataricus* showed that the two transcripts that possess C/D box sRNA features in their 5'-UTRs match the median mRNA half-life [124, 142].

The negative effect of C/D box sRNAs in 5'-UTRs on the expression of protein-coding genes raises the question how this localization evolved. It was suggested for eukaryotes that snoRNA genes evolve faster than protein-coding genes as their sequence is more variable. An open reading frame does not have to be maintained and the function of individual C/D box snoRNAs does not seem to be essential. The only sequences that are required for stability and functionality are the boxC/C' and boxD/D' motifs and to a certain extent the guide regions [82]. Interestingly, our dataset shows that many of the C/D box sRNA genes that overlap with the neighboring gene, possess the sequence of the start or stop codon within their boxC/C' or boxD/D' motif. Therefore it is an intriguing possibility that gene start and stop codons serve as origin for the formation of k-turn motifs within C/D box sRNA genes.

3.2 Maturation of archaeal C/D box sRNAs

The biogenesis pathway of archaeal C/D box sRNAs is, in comparison to their eukaryotic homologues, largely unexplored. The *in vivo* analyses of artificial C/D box sRNAs in this study reveal first details about C/D box sRNA maturation in archaea. The existence of

archaeal C/D box sRNA genes in diverse genetic contexts implies that a universal maturation mechanism must exist which tolerates C/D box sRNA precursors that result from transcription initiated by the different and often distant promoters. The significant amount of C/D box sRNA genes that overlap with flanking genes suggests that signals that are required for the processing of the C/D box sRNAs are not necessarily to be found in the C/D box sRNA adjacent regions. Interestingly, the C/D box sRNA *in vivo* production studies in *S. acidocaldarius* showed that the C/D box sRNA gene upstream or downstream sequences are not required for efficient C/D box sRNA maturation. Consequently, sequence- or structure-specific endoribonucleases do not seem to be involved in the C/D box sRNA maturation process unless these nucleases recognize the C/D box sRNA structure or sequences within the C/D box sRNAs. Therefore, precursor-processing activity of unspecific exoribonucleases is plausible.

The C/D box snoRNA gene arrangements are diverse in eukaryotes but the maturation and especially the final trimming of eukaryotic snoRNAs always includes the action of exoribonucleases. In yeast, the 5' processing of snoRNAs that were processed from introns or were released by endoribonucleolytic cleavage from polycistronic transcripts, is carried out by the homologous exoribonucleases Rat1p and Xrn1p which have additional functions in pre-rRNA processing [93, 94]. The 3' terminus of eukaryotic snoRNAs is processed by several subunits of the exosome of which the only non-essential exosome subunit Rrp6p executes the final trimming reaction [92, 143, 144].

First insights into RNA processing in archaea were revealed by analyses of mRNA degradation but this process is not completely understood yet. However, enzymes for the processing of the 3' and 5' end of mRNA transcripts have been identified and their potential role in C/D box sRNA maturation cannot be excluded. Homologues of the eukaryotic snoRNA 5' end processing enzymes Rat1p and Xrn1p do not exist in archaea [145]. Archaeal homologues of the bacterial RNaseJ family, exhibiting 5'-3' exoribonucleolytic activity, have been identified. This group comprises aRNaseJ, aCPSF2 and aCPSF1. These RNases have a characteristic β-CASP metallo-β-lactamase fold and degrade transcripts with 5'-monophosphate or 5'-hydroxyl ends. The enzyme aCPSF1 possesses also an endoribonuclease activity and single-stranded CA dinucleotides become preferentially cleaved. Thermoprotei, e.g. *Sulfolobus* species, do contain the proteins aCPSF1 and aCPSF2. A deletion of aCPSF2 in *S. acidocaldarius* revealed differential transcript abundance for a large amount of genes. Nevertheless, the enzyme is not essential and cells lacking aCPSF2 do not show growth defects indicating that essential exoribonucleolytic activities can be substituted for by other enzymes [145-148].

Exosomes that mediate the 3' end processing of mRNA transcripts were identified in most archaea with the exception of the *Halophiles* and *Methanococcales*. The archaeal exosome

is a nine subunit complex of which a hexameric ring of Rrp41-Rrp42 dimers is the catalytic core. Entry to this core is restricted to single-stranded RNA due to size constraints. The catalytic core shows more structural and functional similarity to the bacterial polynucleotide phosphorylase (PNPase) than to the non-catalytic core of the exosome in eukaryotes. The proteins Rrp4 and Csl4 build a flexible structure at the top of the catalytic hexamer and are involved in RNA recruitment to the catalytic chamber [149-152].

The observation that archaeal C/D box sRNA maturation occurs independent of a specific genomic upstream and downstream context implies that exoribonucleases are involved in the process. The described aCPSF1 and aCPSF2 enzymes, as well as the exosome, which is also involved in eukaryotic C/D box snoRNA 3' end processing, could be involved in the C/D box sRNA maturation. However, it cannot be excluded that other unknown exoribonucleases are necessary for archaeal C/D box sRNA maturation. The eukaryotic enzymes involved in snoRNA maturation were identified by knockout studies of known ribonucleases. A description of a *S. acidocaldarius* aCPSF2 deletion strain does not include an analysis of small non-coding RNAs. An aCPSF1 deletion is not available [148]. Analyses of the sRNome of aCPSF knockout mutants would be a feasible approach to check whether these exoribonucleases are involved in C/D box sRNA maturation.

A requirement for C/D box sRNA processing by exoribonucleases is the protection of the C/D box sRNA from degradation. Different protection mechanisms exist for mRNAs. In archaea, binding of the translation initiation factor a/eIF2(γ) to tri-phosphorylated mRNA provides protection from the activity of RNaseJ [153]. Bacterial and eukaryotic mRNAs are protected from the activity of ribonucleases by stem-loop structures, the cap structure or poly-A tails. The mRNA molecules in bacteria can be also stabilized by high-affinity ribosome binding [154]. All these factors prevent exoribonucleolytic degradation of the mRNA ends but a different protection is necessary for C/D box sRNA maturation as the RNA ends need to be degraded but the mature C/D box sRNA should be maintained. In eukaryotes, it could be shown that the snoRNP proteins associate with the snoRNA before processing occurs [155, 156]. The terminal stem that is formed by base pairing of the sequences at the 5' and 3' end of the C/D box snoRNA and the boxC and boxD motifs, is required for the assembly of the snoRNP [157, 158]. The C/D box sRNA *in vivo* assay data with boxC' and boxD/D' mutants (section 2.4.2) showed that mutations that are involved in k-turn formation result in instable C/D box sRNA transcripts, which is likely a result of prevented L7Ae binding. Therefore, it is plausible that a C/D box sRNA protection mechanisms exists in archaea in which co-transcriptional L7Ae or complete C/D box sRNP binding protects the C/D box sRNAs from complete degradation. A good indication for L7Ae or complete C/D box sRNP binding prior to C/D box sRNA processing stems from RNA immunoprecipitation studies with L7Ae and subsequent RNA-Seq analyses of the associated RNAs (RIP-Seq). These data show that

longer, not fully processed reads, e.g. with nucleotides of the neighboring mRNA, exist for C/D box sRNAs with flanking ORFs (Michael Daume, unpublished data). However, with these data it cannot be determined whether only L7Ae or the complete C/D box sRNP assembles cotranscriptionally. The dimeric structure of the C/D box sRNP with two bound sRNAs makes it hard to imagine that the whole complex assembles cotranscriptionally.

It was observed that k-turn integrity is crucial for the maintenance of mature cellular C/D box sRNAs. In contrast, single or double mutations in the boxC' and boxD' sequences that disrupt the Watson-Crick base pair or the mismatch base pair in the k-loop are allowed. The importance of the two sheared GA base pairs in the k-loop for L7Ae binding was observed previously [44]. Differences in L7Ae binding to varied k-loop motifs were not analyzed but the loop that is formed by the nucleotides between the boxC' and boxD' motifs comprises only three or four nucleotides in the majority of *S. acidocaldarius* C/D box sRNAs. Therefore, the two boxes are already in close proximity and k-loop formation is facilitated in comparison to k-turn formation. This may be the reason why several mutations are allowed in the boxC' and boxD' motifs. Furthermore, a k-loop does not necessarily always exist in eukaryotic snoRNAs. The architecture of eukaryotic C/D box snoRNPs is not completely understood, but assembly seems to be possible without a k-loop motif [50]. This suggests that the weaker conservation of the archaeal k-loop in comparison to the k-turn will not interfere with complete C/D box sRNP assembly.

It is concluded that co-transcriptional L7Ae binding or complete C/D box sRNP assembly blocks degradation of archaeal C/D box sRNAs and the 5' and 3' ends are defined by steric hindrance of L7Ae or the C/D box sRNP prohibiting further degradation (figure 3.1).

Variable ends with variations of up to seven nucleotides were observed in the RNA-Seq data of *S. acidocaldarius*. A length heterogeneity of several nucleotides was also observed in plant snoRNAs and at the circularization junctions of *S. solfataricus* and *P. furiosus* C/D box sRNAs [83, 99, 100, 159]. Additionally, vertebrate snoRNAs that were expressed in plant cells were slightly larger than snoRNAs expressed from their native locus in vertebrates and it was concluded that these differences result from different protection of the proteins and therefore different accessibility of the exoribonuclease [160]. The length heterogeneity of several nucleotides might be a result of complexes that exhibit a certain degree of flexibility. In agreement, large conformational rearrangements were shown to occur in C/D box sRNP of *P. furiosus* upon substrate binding [55].

Based on crystal structures of C/D box sRNA bound L7Ae, direct interactions of L7Ae could be detected for nucleotides of the boxC and boxD motifs, as well as for the nucleotide directly upstream of the boxC motif [53]. Footprinting studies of L7Ae on C/D box sRNAs using RNaseT1 showed that nucleotides in the boxC motif and in the flanking upstream and downstream region as well as nucleotides of the boxC' and boxD' motifs are protected from

RNaseT1 digestion [44]. In general, the experiments show that in addition to nucleotides of the conserved box motifs, nucleotides upstream of the boxC motif are also protected by L7Ae binding. However, RNaseT1 digestion is restricted to cleavage at guanosine residues in single-stranded RNA and therefore it could not be determined exactly with the used substrate how many nucleotides upstream of the boxC and downstream of the boxD motif are protected by L7Ae binding. Furthermore, it is possible that the Nop5 subunits additionally protect the RNA ends in the complete C/D box sRNP, but footprinting studies with a complete archaeal C/D box sRNP are not available.

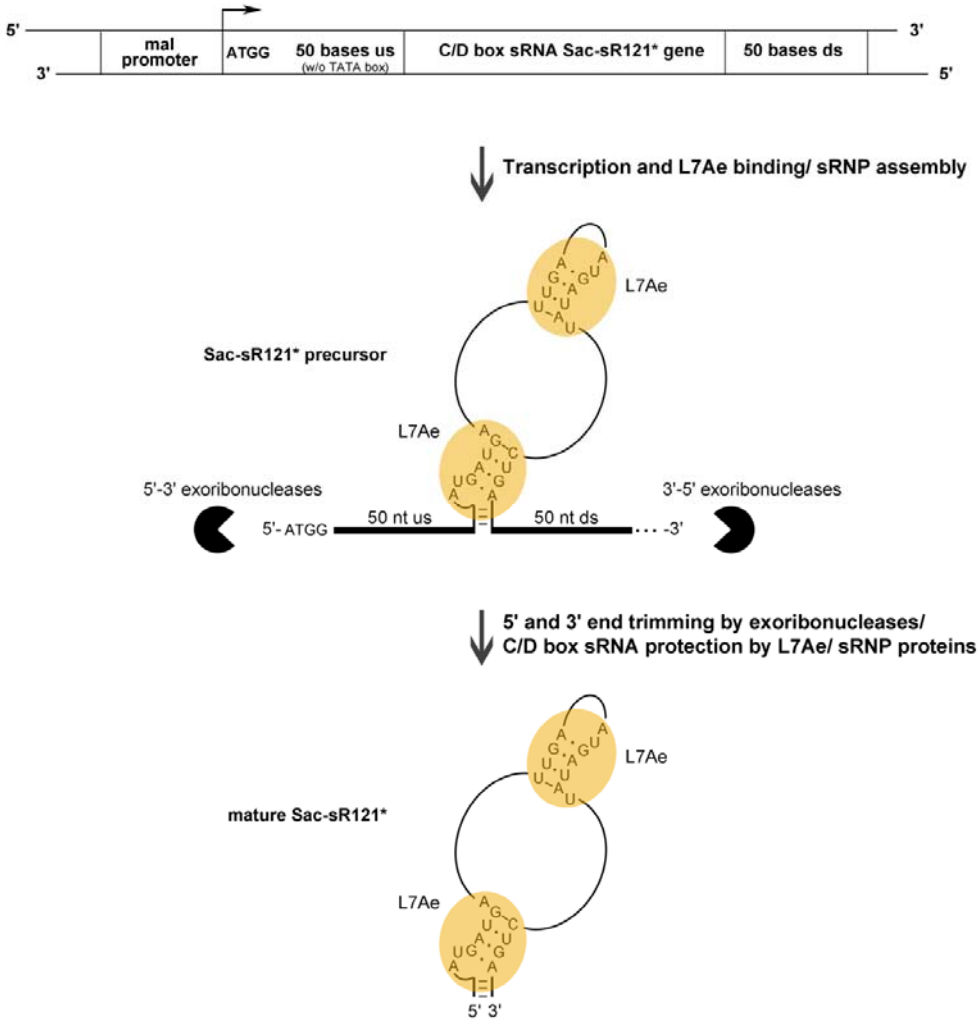

Figure 3.1: Model for maturation of C/D box sRNA Sac-sR121*. Transcription from the external maltodextrin (mal) promoter produces a primary C/D box sRNA transcript that also exhibits C/D box sRNA upstream and downstream sequences. The 5' and 3' ends become trimmed by the action of exoribonucleases. Co-transcriptional binding of L7Ae or complete C/D box sRNP assembly protects the C/D box sRNA from degradation by exoribonucleases and the 5' and 3' ends are defined by steric hindrance of the proteins.

Assembly of the C/D box sRNP with recombinant *S. acidocaldarius* proteins was not successful. In general, the *in vitro* C/D box sRNP assembly seems to be problematic and

requires proper substrate RNAs. The mono- or di-sRNP architecture of the archaeal complex was controversially discussed for several years in reference to C/D box sRNP structures of several archaea. This controversy could be solved just recently by the determination of a structure that also showed a mechanism for sequential C/D box sRNA methylation [55, 60-64].

One explanation for the unsuccessful assembly of the *S. acidocaldarius* C/D box sRNP could be the binding specificity of L7Ae. L7Ae binding to a single k-turn resulted in a single shift in EMSA studies, however, usage of a native C/D box sRNA substrate revealed multiple shifts with increasing L7Ae concentrations. The multiple shifts could be also observed with a substrate with disrupted boxD and boxD' motifs that should be not able to form a stable k-turn motif and therefore should prevent L7Ae binding. Binding of L7Ae to this substrate, as well as the presence of multiple shifts are indications for unspecific binding of L7Ae to the C/D box sRNA substrate. They could be caused by unspecifically bound *E. coli* RNA from the recombinant production of the protein. Multiple shifts were already observed in previous EMSA studies with a S*. acidocaldarius* C/D box sRNA and recombinant *S. solfataricus* L7Ae that were performed under similar conditions. The authors of this study suggested that multiple conformational isomers of the C/D box sRNA-L7Ae complex exist [46]. Follow-up EMSA studies with a C/D box sRNA substrate with a disrupted k-loop showed rather a smear than a distinct shift upon L7Ae addition and this is also an indication for unspecific binding. However, the addition of Nop5 and fibrillarin revealed the formation of non-functional complexes that were not able to methylate target RNA [118]. In addition, recombinant *M. jannaschii* L7Ae showed, apart from two shifts for RNP complexes, a further shift in EMSA studies with excess concentrations of L7Ae [161]. All these experiments indicate that L7Ae can also bind unspecific to C/D box sRNAs without the recognition of k-turn or k-loop motifs but via charge-charge interactions. These interactions do not seem to be restricted to C/D box sRNAs as low affinity binding of the *M. jannaschii* L7Ae could be also shown with rRNA molecules [51].

Taken together, the current archaeal C/D box sRNA maturation model highlights a necessity for unspecific ribonuclease activities and co-transcriptional L7Ae or entire C/D box sRNP assembly is suggested to prevent complete sRNA degradation.

A minor fraction of archaeal C/D box sRNAs exists in tandem arrangements and Northern blot analyses revealed the existence of a dicistronic transcript, but also showed that both C/D box sRNAs exist individually. The *in vivo* studies on the maturation of tandem C/D box sRNAs (section 2.6) were aimed to address the question how these RNAs are processed in archaea. It was difficult to draw a final conclusion. None of the *in vivo* constructs revealed the production of a mature C/D box sRNA transcript even when a complete tandem arrangement was analyzed. Instead, the dicistronic transcript or the artificially designed 'tandem' consisting

of a complete C/D box sRNA and a C/D box sRNA half were detected even if the boxC/C' or boxD/D' sequences were disrupted. Additionally, transformation of several constructs was not possible or the cells harboring the constructs showed strong growth defects. Northern blot analyses with total RNA from cells that only harbored the control construct C/D box sRNA Sac-sR125 with one artificial guide did not show a stable transcript of this C/D box sRNA and the cells that harbored the plasmid with this *in vivo* construct were also impaired in growth. As the C/D box sRNA tandem exists in its native locus in addition to the copy in the *in vivo* constructs, such effects were not expected. The facts indicate a potential general problem with the plasmid-based *in vivo* C/D box sRNA assays for the analysis of the C/D box sRNA tandem arrangement Sac-sR125/126. The problem might arise from the guide exchange to the artificial guide sequence but that could not be fully clarified. Both putative modification target sites are located in close vicinity in the 23S rRNA. This could indicate that both guide regions or even the presence of both guide regions in one C/D box sRNA molecule are essential for rRNA modification at both target sites or for rRNA folding (RNA chaperone activity). It is possible that the synthetic C/D box sRNA Sac-sR125 is degraded as the binding of the native guide of this RNA to the complementary region in the 23S rRNA could block the binding of the native C/D box sRNA Sac-sR125 which exhibits two intact guide regions. In general, it could be that the *in vivo* constructs that were not transformable or showed strong growth defects would lead to the formation of mature C/D box sRNAs but, due to the guide exchange to the artificial guide, these RNAs can only bind with one guide to the target rRNA and thereby block the binding of the native C/D box sRNA.

In eukaryotes, polycistronic snoRNAs can be often found, especially in higher eukaryotes [78]. In yeast, the initial processing of polycistronic snoRNAs is performed by the endoribonuclease Rnt1p, the homologue of *E. coli* RNaseIII. This endoribonuclease cleaves in intergenic double-stranded RNA regions that are closed with AGNN tetraloops. This step separates individual snoRNAs of the polycistronic precursor and further trimming of the snoRNA ends occurs by exoribonucleases [91, 94]. In plants, the enzymes that are responsible for the processing of polycistronic snoRNAs are not yet identified, but the processing is splicing-independent and an activity of endoribonucleases, is suggested [82, 83]. In maize, the distances between individual C/D box snoRNA genes within a polycistron are, with one exception that possesses an 749 bp long intergenic region, 31-136 bp long, leaving enough space for sequence information that contains endoribonuclease recognition signals [83]. In the tandem arrangements that were identified in archaea, only few nucleotides separate the boxD motif of the first C/D box sRNA and the boxC motif of the second C/D box sRNA within the polycistron. This reduces the possibility for specific endoribonuclease interactions that were identified in yeast and postulated for plant snoRNAs. The binding of L7Ae to the k-turn motifs blocks exoribonuclease activity which makes it

difficult to imagine how processing without an endoribonuclease activity is possible for the archaeal C/D box sRNA polycistron. One possibility is that the C/D box sRNA structure is recognized or that the signals for processing are encoded within the guide regions. In the second scenario, the guide region exchange in the *in vivo* constructs could cause problems. In most *in vivo* experiments, the D guide of C/D box sRNA Sac-sR125 or the D' guide of C/D box sRNA of Sac-sR126 were exchanged against the artificial guide i.e. the guide regions next to the potential endoribonuclease cleavage site. Interestingly, the V2 *in vivo* construct, in which the D guide of C/D box sRNA Sac-sR126 was exchanged which is not adjacent to the potential endoribonuclease cleavage site, showed the strongest signal in the Northern blot analyses in comparison to *in vivo* constructs that possess the artificial guide at a different position. However, a signal for a mature C/D box sRNA transcript was absent.

Identified endoribonucleases in archaea include (i) the splicing endonuclease that is involved in tRNA and rRNA processing, (ii) RNaseZ, responsible for the 3' end processing of tRNAs and (iii) Cas6, which has a function in CRISPR RNA processing [104, 162, 163]. Additionally, aCPSF1 was identified to play a key role in mRNA degradation and to cleave single-stranded CA dinucleotides [164]. In the case of the analyzed C/D box sRNA polycistron from *S. acidocaldarius*, a CA dinucleotide could not be observed between both C/D box sRNAs and the other endoribonucleases are not likely to be involved in processing of the dicistronic C/D box sRNA precursor as they recognize specific structures that cannot be identified. It is also unlikely that an endoribonuclease exists that is specifically responsible for the maturation of polycistronic C/D box sRNAs as this arrangement is not predominant in archaea. However, it cannot be excluded that a yet unknown endoribonuclease is involved in the processing of dicistronic C/D box sRNAs in archaea.

The *in vivo* studies did not reveal a crucial factor for C/D box sRNA tandem processing. The spacer length between both C/D box sRNAs or the non-consensus sequence of several boxC/C' and boxD/D' motifs did not facilitate correct C/D box sRNA tandem processing. However, it also cannot be excluded that one of these factors is crucial for correct processing as some plasmids were not transformable or resulted in strong growth deficiencies and in most *in vivo* experiments an unnatural transcript was stabilized. Intact boxC and boxD motifs do not seem to be essential for the stabilization of these transcripts as disruption of the conserved GA base pairs or complete disruption of the box sequences still revealed the stable non-mature C/D box sRNA transcript in most cases. It could be shown previously that complexes of the C/D box sRNP proteins can form *in vitro* e.g. with only a minimal k-turn substrate or with C/D box sRNAs that have a disrupted k-loop motif, but these complexes were not functional [118]. It can therefore not be excluded that artificial complexes are built that stabilize the C/D box sRNA tandem *in vivo* constructs. Future experiments on tandem C/D box sRNA maturation could be performed with (i) complete tandem arrangements with

two artificial guide regions or with (ii) the second tandem C/D box sRNA in *S. acidocaldarius* to identify whether similar problems in the *in vivo* experiments occur or whether they are specific for the analyzed tandem.

3.3 C/D box sRNA circularization

The RNA-Seq data obtained for *S. acidocaldarius* C/D box sRNAs did not identify circular C/D box sRNAs in this organism probably due to methodical problems. However, Northern blot and inverse RT-PCR analyses do indicate C/D box sRNA circularization. Archaeal C/D box sRNA were shown to be circular in *S. solfataricus* and *P. furiosus* and C/D box sRNA circularization is assumed for *N. equitans* [9, 99, 100]. These organisms live at temperatures above 65 °C and it is possible that circular C/D box sRNAs are unique for organisms with a hyperthermophilic lifestyle. Circular C/D box sRNAs seem to be a functional RNA form as they were identified in coimmunoprecipitation experiments with C/D box sRNP proteins [100].
In general, circular RNAs can be found in all domains of life but they are rare in comparison to linear RNAs [165]. In archaea, they mainly comprise tRNA and rRNA introns that are removed by the splicing endonuclease. This enzyme cleaves within BHB structures that are formed at the exon-intron junctions [103-105]. The RNA ligase RtcB ligates the tRNA (and rRNA) molecules after the introns were spliced out by the splicing endonuclease and is also responsible for the circularization of the introns as they have the same termini (2',3'-cyclic phosphate and 5'-OH) [126-128]. Two other RNA ligases are known for archaea that could be involved in C/D box sRNA ligation, the 5'-3' RNA ligase and the 2'-5' RNA ligase. The 5'-3' RNA ligase, which is suggested to have a function in RNA repair is not encoded in the *S. acidocaldarius* genome [166, 167]. The 2'-5' RNA ligase is highly conserved in archaea and bacteria but its physiological role is unknown [129, 130]. A function in C/D box sRNA circularization could be excluded as the created deletion mutant still harbored circular C/D box sRNAs. The RNA-Seq data of the deletion mutant also did not give a hint about the actual function of the 2'-5' RNA ligase. The enzymatic activity that is responsible for C/D box sRNA circularization is not known. It might be a moonlighting activity of RtcB or a yet unidentified RNA ligase. Circular C/D box sRNAs were identified in C/D box sRNP coimmunoprecipitation studies which implies that the RNA ligase should be transiently associated, especially when the C/D box sRNP proteins assemble co-transcriptionally as described in eukaryotes. Therefore, similar studies could be performed with additional RNA-protein cross-linking (CLIP) to identify the RNA ligase [100, 168].
The *S. acidocaldarius in vivo* experiments revealed circularization of the C/D box sRNAs when the upstream and downstream sequences were exchanged against random sequences. This indicates that 'signals' for the circularization reaction do not exist within these sequences. The proposed C/D box sRNA maturation by exoribonucleolytic degradation

results in short sequences upstream of the boxC and downstream of the boxD motif which stick out of the assembled C/D box sRNP and remain in close proximity. It is therefore plausible that these ends are ligated without the recognition of additional factors by a RNA ligase. Computational analyses did not identify conserved sequences or structures in the upstream or downstream regions which is also a hint that circular C/D box sRNAs are not the result of the C/D box sRNA processing mechanism, as it is the case for circular rRNA and tRNA introns.

Relaxed BHB structures with extended helices or lacking bulges that could likely not be cleaved by known splicing endonucleases were identified at the borders of C/D box sRNAs of several archaea including C/D box sRNAs from *S. acidocaldarius*. Of 24 analyzed *S. acidocaldarius* C/D box sRNAs only two did not show the identified structures at the C/D box sRNA borders. However, the identification of new C/D box sRNAs based on the designed algorithm revealed only five more C/D box sRNAs in *S. acidocaldarius* [106]. Therefore, the majority of C/D box sRNAs that were identified with the small RNA sequencing approach in this study, is without these motifs. Additionally, our data revealed the existence of varying circularization junctions and this phenomenon was also observed in *P. furiosus* and *S. solfataricus* [99, 100]. However, this is not in agreement with the action of the splicing endonuclease that recognizes and cleaves canonical BHB structures at the borders of tRNA and rRNA introns and the subsequent action of RtcB. For tRNA and rRNA introns a heterogeneity of the circularization junction could not be observed [99]. These facts question the recognition of relaxed BHB motifs at the borders of archaeal C/D box sRNAs.

One exception, in which C/D box sRNA circularization is the result of a processing mechanism, exists for C/D box sRNAs that are encoded in tRNA introns. These sRNAs are released and circularized after intron splicing. C/D box sRNA genes in tRNA introns are reported for the pre-$tRNA^{Trp}$ in most euryarchaea. Based on their existence in most euryarchaea, it was suggested that this RNA was previously a trans-acting RNA and was inserted in the intron in a common ancestor of euryarchaea [111]. It is possible that this position was additionally favored and maintained because of the efficient transcription, C/D box sRNA maturation and stabilization through circularization. Previous studies revealed that RNA circularization increases the RNA stability and nuclease degradation is prevented [169, 170]. These factors are both important for C/D box sRNAs. The secondary structure, as well as protection against degradation, have to be maintained to ensure efficient 2'-*O*-methylation of the target RNAs. In eukaryotic C/D box snoRNAs, the RNA ends form the double stranded terminal stem to ensure these requirements, but archaeal C/D box sRNAs often do not have this stabilization as terminal stems are absent or too short [1, 8, 9, 17, 171]. The high abundance of C/D box sRNAs in archaea indicates that the RNAs are extremely stable. Therefore, C/D box sRNA circularization might be performed to stabilize the RNA structure

and to provide protection against degradation in hyperthermophilic organisms in addition to L7Ae or C/D box sRNP binding to the C/D box sRNA.

3.4 2'-O-methylation patterns of archaeal rRNA molecules

In contrast to bacteria and eukaryotes in which maps of post-transcriptional modifications exist, no complete overview of archaeal 2'-*O*-methylation targets in rRNAs exists up to now. Previous mappings either focused on specific helices in the rRNA structure or included only a limited amount of C/D box sRNAs due to small data availability [1, 7, 26]. The mapping of the predicted 2'-*O*-methylation target sites of seven archaea shows interesting hotspots of methylation conservation. All these regions belong to the ancient core of the rRNAs [172]. Additionally, the regions where methylation hotspots can be identified are important for the functionality of the ribosome. The helices 90-93 in domain V of the 23S rRNA are close to the catalytic peptidyl transferase center where peptide bond formation and release occurs [173]. The helices 61 and 68-71, as well as the helices 35 and 35a, are part of the interface between the large and the small ribosomal subunit and the helix 69 is known to directly interact with the 16S rRNA [174]. In the 16S rRNA, most predicted 2'-*O*-methylation target sites are located in the central core connecting the four domains. It is formed in the early stage of the assembly and is close to the decoding center where codon-anticodon base pairing is monitored [135]. In summary, conserved 2'-*O*-methylation target sites seem to be located in ancient core regions of the rRNAs that are important for ribosome integrity and functionality and the modifications might contribute to the folding, structural stabilization and function of the rRNAs. As these regions include RNA-RNA interaction sites between the 23S rRNA and the 16S rRNA, as well as between the 23S rRNA and the tRNAs in the peptidyl transferase center, it could be that these interactions are stabilized by these modifications [174, 175]. Additionally, the regions in which methylation hotspots can be observed are not protected by ribosomal proteins and the modifications may help in the stabilization of the rRNA structure in these unprotected regions [52].

In bacteria and eukaryotes, where post-transcriptional modification maps exist for several organisms, similar hotspots of modifications could be identified in the 23S rRNA [5, 19]. Hence, a conserved modification pattern exists in all three domains of life although different kinds of machineries are responsible for the formation of the modifications. In bacteria, the mechanism is RNA-independent and site-specific enzymes exist in comparison to eukaryotes and archaea where the RNA-dependent modification is predominant [1, 5, 11].

Interestingly, not a single nucleotide in 16S or 23S rRNA is modified by all seven archaea that were analyzed. One nucleotide of the 23S rRNA is modified by six analyzed archaea and three nucleotides by five archaea. The low conservation of modifications of most specific nucleotides was also observed in bacteria and eukaryotes and mutational studies also

revealed that modification of specific nucleotides is not crucial, but the deletion of clusters of modifications has negative effects on cell growth [22, 23]. This indicates that the positive effect of the modifications is only provided by the accumulation of modifications in specific regions. Furthermore, the low conservation of modifications at specific nucleotides led to the hypothesis that the selection for modifications at definite nucleotides is weak and therefore the guide regions are dynamic and the evolution of guide regions is faster than the evolution of changes in the target rRNA sequences. Nucleotide substitutions occur in the guide region that lead to mismatches between the guide and the target RNA and the accumulation of mutations leads to orphan guides that are free to evolve guides for new methylation targets [7]. This scenario would also explain the huge amount of orphan guides with no identified targets.

The 2'-*O*-methylation target sites presented in this study are computational predictions. Two 2'-*O*-methylated nucleotides in the decoding center of the 16S rRNA and five modifications in the helices 20, 90 and 93 of the 23S rRNA of *S. acidocaldarius* could be validated with RTL-P studies (Reverse Transcription at low dNTP concentrations followed by PCR) (bachelor thesis of Lydia Seelos). This method was established to detect 2'-*O*-methylated nucleotides in human and yeast rRNAs and it is based on the ability of the reverse transcriptase to arrest at modified sites at low dNTP concentrations [176]. Additional modifications in helix 92 of *H. marismortui* and *S. acidocaldarius* were previously identified in MALDI-MS studies [26]. Observation of conserved prediction of nucleotides in functionally important regions and a general agreement with limited available data on rRNA modifications provide support for the reliability of the predictions. Future global and differential RNA-seq will be utilized to verify the existence of these modifications biochemically.

4. Material and Methods

4.1 Material and sources of supply

4.1.1 Chemicals, enzymes and kits

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

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

Chemicals	Company
Acrylamid/Bisacrylamid (29:1 und 37.5:1)	Carl Roth GmbH, Karlsruhe
Amicon ® Ultra Centrifugal Filters	Merck Millipore KGaA, Darmstadt
Antarctic Phosphatase	New England Biolabs GmbH, Frankfurt
Antibiotics (Ampicillin, Chloramphenicol, Kanamycin)	Carl Roth GmbH, Karlsruhe; Sigma- Aldrich, Taufkirchen
Bovine serum albumin (BSA)	Sigma-Aldrich, Taufkirchen
Bradford Reagent	BioRad Laboratories GmbH, Munich
ColorPlus Prestained Protein Ladder, Broad Range (10-230 kDa)	New England Biolabs GmbH, Frankfurt
Diethylpyrocarbonate (DEPC)	AppliChem GmbH, Darmstadt
DNaseI	Thermo Fisher Scientific Germany Ltd. & Co. KG, Bonn
dNTPs	New England Biolabs GmbH, Frankfurt
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-thiogalactopyranosid (IPTG)	Carl Roth GmbH, Karlsruhe
Low Molecular Weight Ladder	New England Biolabs GmbH, Frankfurt
Low Range ssRNA Ladder	New England Biolabs GmbH, Frankfurt
Lysozym	Sigma-Aldrich, Taufkirchen
mirVana™ miRNA Isolation Kit	Applied Biosystems, Darmstadt
NTPs	Jena Bioscience GmbH, Jena
Phenol (Roti-Phenol & Roti-Aqua-Phenol)	Carl Roth GmbH, Karlsruhe
Phusion ® High-Fidelity DNA Polymerase	Thermo Fisher Scientific Germany 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
Quick-Load ® 2-Log DNA Ladder (0.1-10.0 kb)	New England Biolabs GmbH, Frankfurt
Restriction endonucleases	New England Biolabs GmbH, Frankfurt
RNase Exitus Plus™	Applichem GmbH, Darmstadt
RNase Inhibitor	New England Biolabs GmbH, Frankfurt
RNaseR	Epicentre, Madison, USA
Roti ®-Nylon plus, pore size 0.45 µm	Carl Roth GmbH, Karlsruhe
SDS	Carl Roth GmbH, Karlsruhe
SuperScript ® III RT	Thermo Fisher Scientific Germany Ltd. & Co. KG, Bonn
SYBR Gold ® Nucleic acid stain	Thermo Fisher Scientific Germany Ltd. & Co. KG, Bonn

T4-DNA-Ligase	New England Biolabs GmbH, Frankfurt
T4-Polynucleotide-Kinase	New England Biolabs GmbH, Frankfurt
T7-RNA-Polymerase	own production
Taq DNA Polymerase	New England Biolabs GmbH, Frankfurt
Topo ® TA cloning ®	Thermo Fisher Scientific Germany Ltd. & Co. KG, Bonn
TRIzol Reagent	Ambion, Darmstadt
ULTRAhyb-Oligo hybridization buffer	Ambion, Darmstadt
Whatman GB004, 3MM	Schleicher & Schuell, Dassel
ZelluTrans Dialysis tubes	Carl Roth GmbH, Karlsruhe
5-Fluoroorotic acid (5-FOA)	Thermo Fisher Scientific Germany Ltd. & Co. KG, Bonn

4.1.2. Instruments

Table 4.2: Instruments used in this study.

Agarose gel electrophoreses	Chambers and Casting tray: 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
Semi-dry transfer cell	Trans-Blot® SD Semi-Dry Transfer Cell, BioRad Laboratories GmbH, Munich
Centrifuges	Centrifuge 5424, Eppendorf AG, Hamburg; Sorvall RC5B Plus, Thermo Fisher Scientific Germany Ltd. & Co. KG, Bonn
Chromatography columns	HisTrap HP 1 ml/ MonoQ 5/50 GL/ Superdex 200 10/300 GL/HiTrap Heparin HP 1 ml/ MonoS 5/50 GL; GE Healthcare Europe GmbH, Freiburg
FPLC	Äkta-System: Pump P-900, Monitor UV-900, Monitor UPC-900, Valve INV-907, Mixer M-925; GE Healthcare Europe GmbH, Freiburg
Denaturing polyacrylamide gel chambers	PROTEAN II Electrophoresis Chamber, BioRad Laboratories GmbH, Munich
Gene Pulser	Gene Pulser ® Electroporation System, Pulse Controller Plus, Capacitance Extender Plus; BioRad Laboratories GmbH, Munich
Hybridisation oven	Hybaid Shake 'n' Stack, Thermo Fisher Scientific Germany Ltd. & Co. KG, Bonn
Incubators	Thermotron, Infors AG, Bottmingen, Switzerland
Magnetic stirrer	IKA® RCT Standard, IKA®-Werke GmbH & Co. KG, Staufen
Nanodrop	NanoDrop® ND-1000 Spectrometer, Thermo Fisher Scientific Germany Ltd. & Co. KG, Bonn
PCR-Cycler	C1000™ Thermal Cycler, BioRad Laboratories GmbH, Munich
Phosphorimager	Storm 840 phosphorimager, Molecular Dynamics
Polyacrylamide gel electrophoresis	Mini-PROTEAN Tetra Cell, Bio-Rad Laboratories GmbH, Munich; Power supply PowerPac Basic, Bio-Rad Laboratories GmbH, Munich

Scintillation counter	Beckmann LS 6500, Beckman Coulter, Krefeld
Spectrophotometer	Ultrospec 3000 *pro*, GE Healthcare Europe GmbH, Freiburg
Thermomixer	Thermomixer Comfort 5350, Eppendorf AG, Hamburg
UV-Crosslinker	UV Stratalinker ® 1800, Stratagene, La Jolla, USA
UV-Transilluminator	BioDocd-IT system, UVP, Upland, 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*ZΔM15Δ(*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	Hanahan D [177]
***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 plasmid pM.EsaBC4I**	F⁻ *glnV44 e14⁻(McrA⁻) rfbD1? relA1? endA1 spoT1? thi-1 Δ(mcrC-mrr)114::IS10 (K12 ER1821)*	New England Biolabs GmbH, Frankfurt; [114]
Sulfolobus acidocaldarius MW001	DSM639 Δ*pyrE* (*Saci_1597*; Δ91–412bp)	[114]

4.2.2. Culture conditions for *Escherichia coli*

Aerobic cultivation of *E. coli* batch cultures (5-1000 ml) was carried out in Erlenmeyer flasks by shaking in a rotatory shaker at 200 rpm at 37°C in lysogeny broth (LB) medium (1 % tryptone (w/v), 0.5 % yeast extract, 1 % NaCl (w/v), pH 7.2) or on solid medium plates (LB medium containing 1.5 % (w/v) agar-agar). Liquid LB medium containing the appropriate plasmid-encoded antibiotics (chloramphenicol 34 µg/ml, kanamycin 50 µg/ml and ampicillin 100 µg/ml) was inoculated with a pre-culture (2 % (v/v)) and growth was monitored at 600 nm. Heterologous production of recombinant proteins was induced at an OD_{600nm} of 0.6-0.8 by the addition of 1 mM IPTG. After incubation at 37°C for 3-4 hours, cells were harvested via centrifugation (6,000 x g, 10 min, 4°C) and stored at -80°C. The *E. coli* strain K12 DH5α was used for cloning, storage and preparation of plasmid DNA. The strain *E. coli* Rosetta2 (DE3) pLysS was used for the heterologous expression of recombinant *S. acidocaldarius* proteins. The strain *E. coli* ER1821 pM.EsaBC4I was used to methylate plasmids prior to transformation into *S. acidocaldarius* to prevent restriction by the *SuaI* restriction enzyme.

4.2.3. Culture conditions for *Sulfolobus acidocaldarius*

Sulfolobus acidocaldarius MW001, all mutant strains and all plasmid containing strains were aerobically grown in Brock media [113] with a pH of 3.5 at 75°C. The media were supplemented with 0.1 % (w/v) tryptone as well as 0.2 % (w/v) sucrose for normal growth or 0.2 % (w/v) dextrin for induction of the maltose-inducible (mal) promoter. The growth of the cells was monitored by measurement of the optical density at 600 nm. Cell harvesting for total RNA preparations was performed at OD_{600nm} of 0.4-0.6. For pouring plates a two times concentrated Brock medium was prepared and supplemented with 6 mM $CaCl_2$ and 20 mM $MgCl_2$. For first selection plates 0.2 % NZ-Amine (w/v) and 0.4 % dextrin (w/v) were added. For second selection plates 0.2 % tryptone (w/v), 0.4 % dextrin (w/v), 200 µg/ml 5-FOA and 20 µg/ml uracil were added. The solution was prewarmed at 75°C and mixed with an equal volume of boiling 1.2 % gelrite solution (w/v). The pH of the mixture was adjusted to 3.5 with sulfuric acid and the mixture subsequently poured in petri dishes.

4.3 Oligonucleotides, plasmids and constructed recombinant vectors

4.3.1 Plamids and constructed recombinant vectors

Table 4.4: Plasmids used in this study.

Vector	Features	Application	Source
pEC-A-HiSumo	Amp^r, Sumo protein fusion, His-tag	heterologous gene expression	Gift of Prof. Dr. Elena Conti (MPI Martinsried)
pUC19	*lacZ*, Amp^r	*In vitro* run-off transcription	
pET-Duet1	Amp^r	heterologous gene expression	
pSVA1431	Amp^r, pyrEF, mal promoter	gene expression in *S. acidocaldarius*	Gift of Prof. Dr. Sonja-Verena Albers (Albert-Ludwigs-University Freiburg), [114]
pSVA406	Amp^r, pyrEF	markerless deletions or insertions in *S. acidocaldarius*	Gift of Prof. Dr. Sonja-Verena Albers (Albert-Ludwigs-University Freiburg), [114]
pSVA431	lacS, Amp^r, pyrEF	markerless deletions in *S. acidocaldarius*	Gift of Prof. Dr. Sonja-Verena Albers (Albert-Ludwigs-University Freiburg), [114]
pSVAMCS	Amp^r, pyrEF	control plasmid of pSVA1431 w/o mal promoter	

Table 4.5: Constructed recombinant plasmids for markerless deletions in *S. acidocaldarius*.

Plasmid + Insert	Description of the insert
pSVA431 + Saci1287KO	knockout of Saci_1287; contains part of Saci_1287 (NcoI, KpnI) and upstream and downstream regions for homologous recombination (BamHI, SalI)

pSVA431 + Saci1317KO	knockout of Saci_1317; contains part of Saci_1317 (NcoI, KpnI) and upstream and downstream regions for homologous recombination (BamHI, SalI)

Table 4.6: Constructed recombinant plasmids for protein production.

Plasmid + Insert	Description of the insert
pEC-A-Hi-Sumo + *Saci1520*	Saci_1520 with Sumo protein fusion and 6x His-tag (N-terminal)
pET-Duet 1 + *fibrillarin* + *Nop5*	Saci_1346 with a 6x His-tag (N-terminal) (BamHI, SalI) and Saci_1347 (NdeI)

Table 4.7: Constructed recombinant plasmids for *in vitro* run-off transcription.

Plasmid + Insert	Description of the insert
pUC19 + tandem version 1	Sac-sR125 & Sac-sR126 half with T7 promoter, restriction sites BamHI/HindIII
pUC19 + C/D 47/48	Sac-sR125 & Sac-sR126 with T7 promoter, restriction sites BamHI/HindIII
pUC19 + C/D box sRNA 39 + 156 nt downstream region	Sac-sR121 with T7 promoter and 156 nt native downstream region, restriction sites BamHI/HindIII
pUC19 + k-turn	Sac-sR121 half with T7 promoter, restriction sites BamHI/HindIII
pUC19 + C/D39	Sac-sR121 with T7 promoter, restriction sites BamHI/HindIII
pUC19 + C/D39KKO	Sac-sR121 with boxD motif CTCT and boxD' motif TTCC with T7 promoter, restriction sites BamHI/HindIII

Table 4.8: Constructed recombinant plasmids for introduction of a 2nd copy of Sac-sR125 in the *S. acidocaldarius* genome and exchange of the native Sac-sR125 D guide.

Plasmid + Insert	Description of the insert
pSVA406 + 2nd copy Saci 47	introduction of a 2nd copy of Sac-sR125 in the α-amylase locus; Sac-sR125 with the upstream and downstream region of Saci_1162 for homologous recombination (ApaI, BamHI)
pSVA406 + exchange Saci 47 2nd guide sequence against LNA guide	exchange of the D guide in Sac-sR125 to the artificial guide; Sac-sR125 with the genomic upstream and downstream region for homologous recombination (SacII, SalI)

Table 4.9: Constructed recombinant plasmids for C/D box sRNA *in vivo* analyses in *S. acidocaldarius*. All inserts were cloned with the NcoI and EagI restriction sites in the plasmids. The random sequences were obtained with a random DNA sequence generator. (us=upstream, ds=downstream, w/o =without)

Plasmid + Insert	Description of the insert
pSVAMCS + C/D39	Sac-sR121 with 50 us and ds region, artificial D' guide
pSVAMCS + C/D39 with 50 us and ds w/o TATA	Sac-sR121 with 50 us and ds w/o TATA box in us region, artificial D' guide
pSVA1431 + C/D39 with 50 us and ds w/o TATA	Sac-sR121 with 50 us and ds w/o TATA box in us region, artificial D' guide
pSVA1431 + C/D39 + 50 us; ds region GC stretch (long)	Sac-sR121 with 50 us w/o TATA box and 50 ds + GC stretch, artificial D' guide
pSVA1431 + C/D39 + 50 nt ds region + 50 nt random us region (Version1)	Sac-sR121 with 50 ds and 50 us random (Version1), artificial D' guide
pSVA1431 + C/D39 + 50 nt us region + 50 nt random ds	Sac-sR121 with 50 us w/o TATA box and + 50 ds random (Version1), artificial D' guide

region (Version1)	
pSVA1431 + C/D39 + 50 nt ds region + 50 nt random us region (Version2)	Sac-sR121 with 50 ds and 50 us random (Version2), artificial D' guide
pSVA1431 + C/D39 + 50 nt us region + 50 nt random ds region (Version2)	Sac-sR121 with 50 us w/o TATA box and + 50 ds random (Version2), artificial D 'guide
pSVA1431 + Saci 39 boxD CTGG	Sac-sR121 with 50 us w/o TATA box and 50 ds and boxD CTGG, artificial D' guide
pSVA1431 + Saci 39 boxD CTTA	Sac-sR121 with 50 us w/o TATA box and 50 ds and boxD CTTA, artificial D' guide
pSVA1431 + C/D39 boxD CAGA, boxC TGATGA	Sac-sR121 with 50 us w/o TATA box and 50 ds and boxD CAGA, artificial D' guide
pSVA1431 + C/D39 + 50 us and ds region; boxD GTGA	Sac-sR121 with 50 us w/o TATA box and 50 ds and boxD GTGA, artificial D' guide
pSVA1431 + Saci 39 boxD' TTGG	Sac-sR121 with 50 us w/o TATA box and 50 ds and boxD' TTGG, artificial D' guide
pSVA1431 + Saci 39 boxD' TTTA	Sac-sR121 with 50 us w/o TATA box and 50 ds and boxD' TTTA, artificial D' guide
pSVA1431 + C/D39 boxD' TTGA, boxC' TGAAAT	Sac-sR121 with 50 us w/o TATA box and 50 ds and boxC' ATGAAAT, artificial D' guide
pSVA1431 + C/D39 boxD' ATGA, boxC' TGATAT	Sac-sR121 with 50 us w/o TATA box and 50 ds and boxD' ATGA, artificial D' guide
pSVA1431 + C/D39 boxD' TTGA, boxC' TGAAGA	Sac-sR121 with 50 us w/o TATA box and 50 ds and boxC' ATGAAGA, artificial D' guide
pSVA1431 + tandem 4748 (8nt) guide 2	Sac-sR125 and Sac-sR126 half with 7 nt spacer and artificial Sac-sR126 D' guide
pSVA1431 + tandem 4748 (4nt)	Sac-sR125 and Sac-sR126 half with 3 nt spacer and artificial Sac-sR126 D' guide
pSVA1431 + tandem 4748 (3nt)	Sac-sR125 and Sac-sR126 half with 2 nt spacer and artificial Sac-sR126 D' guide
pSVA1431 + tandem 4748, LNA 47C'D	Sac-sR125 and Sac-sR126 half with artificial Sac-sR125 D guide
pSVA1431 + tandem 4748, GC stretch ds	Sac-sR125 and Sac-sR126 half with artificial Sac-sR126 D' guide and GC stretch in the ds region
pSVA1431 + tandem 4748 komplett (8nt)	Sac-sR125 and Sac-sR126 with artificial Sac-sR126 D' guide
pSVA1431 + tandem 4748 komplett (3nt	Sac-sR125 and Sac-sR126 with 2 nt spacer and artificial Sac-sR126 D' guide
pSVA1431 + tandem4748 complete, LNA48C'D	Sac-sR125 and Sac-sR126 with artificial Sac-sR125 D guide
pSVA1431 + C/D 48 with LNA guide	Sac-sR126 with artificial D' guide
pSVA1431 + tandem 4748, LNA 48, boxC 47 destroyed	Sac-sR125 and Sac-sR126 half with artificial Sac-sR126 D' guide and Sac-sR125 boxC ATCTAGA
pSVA1431 + tandem 4748, LNA 47, boxC 47 destroyed	Sac-sR125 and Sac-sR126 half with artificial Sac-sR125 D guide and Sac-sR125 boxC ATCTAGA
pSVA1431 + tandem 4748, LNA 48, boxD' 48 destroyed	Sac-sR125 and Sac-sR126 half with artificial Sac-sR126 D' guide and Sac-sR126 boxD' ATCT
pSVA1431 + tandem 4748, LNA 47, boxD' 48 destroyed	Sac-sR125 and Sac-sR126 half with artificial Sac-sR125 D guide and Sac-sR126 boxD' ATCT
pSVA1431 + tandem 4748, LNA48, 48boxD'-CTGA statt	Sac-sR125 and Sac-sR126 half with artificial Sac-sR126 D' guide and Sac-sR126 box D‘ CTGA

ATGA	
pSVA1431 + tandem 4748, LNA48, 47boxD-CTGA statt CAGA	Sac-sR125 and Sac-sR126 half with artificial Sac-sR126 D' guide and Sac-sR125 box D CTGA
pSVA1431 + tandem 4748, LNA48, 47boxC TGATGA-boxD CTGA-48boxD' CTGA	Sac-sR125 and Sac-sR126 half with artificial Sac-sR126 D' guide and Sac-sR125 boxC ATGATGA, Sac-sR125 boxD CTGA, Sac-sR126 boxD' CTGA
pSVA1431 + C/D 47	Sac-sR125 with artificial D guide
pSVA1431 + tandem4748-2 LNA guide 47C'D	Sac-sR125 half and Sac-sR126 with artificial Sac-sR125 D guide
pSVA1431 + tandem4748-2 LNA guide 48CD'	Sac-sR125 half and Sac-sR126 with artificial Sac-sR126 D' guide
pSVA1431 + tandem4748-2 LNA guide 48C'D	Sac-sR125 half and Sac-sR126 with artificial Sac-sR126 D guide
pSVA1431 + tandem4748-2, LNA 48C'D, 47boxC' poly-A	Sac-sR125 half and Sac-sR126 with artificial Sac-sR126 D guide and Sac-sR125 boxC' poly-A
pSVA1431 + tandem4748-2, LNA 48C'D, 47boxD poly-A	Sac-sR125 half and Sac-sR126 with artificial Sac-sR126 D guide and Sac-sR125 boxD poly-A
pSVA1431 + tandem4748-2, LNA 48C'D, 48boxD poly-A	Sac-sR125 half and Sac-sR126 with artificial Sac-sR126 D guide and Sac-sR126 boxD poly-A
pSVA1431 + tandem4748-2 LNA 47C'D, boxD' CTGA	Sac-sR125 half and Sac-sR126 with artificial Sac-sR125 D guide and Sac-sR126 boxD' CTGA
pSVA1431 + tandem4748-2 LNA 47C'D, boxD CTGA	Sac-sR125 half and Sac-sR126 with artificial Sac-sR125 D guide and Sac-sR125 boxD CTGA
pSVA1431 + tandem4748-2, all boxes perfect (CTGA/TGATGA)	Sac-sR125 half and Sac-sR126 with artificial Sac-sR125 D guide and Sac-sR125 boxD CTGA and Sac-sR126 boxD' CTGA

4.3.2 Oligonucleotides

All oligonucleotides were synthesized from Eurofins MWG Operon (Ebersberg). The LNA oligonucleotide was synthesized by Exiqon (Vedbæk, Denmark).

Table 4.10: Oligonucleotides used for the amplification of genomic DNA from *S. acidocaldarius* for the creation of in-frame deletion mutants and for the creation of plasmids for the heterologous production of proteins.

Name	Sequence 5' → 3'
In frame deletion of Saci_1287 (pSVA431)	
Saci1287fwNcoI	GAT CCA TGG CAG GTA TTA ATG TGC CAA AC
Saci1287reKpnI	GAT GGT ACC TCA TCT AGG TCC TCT AAC TCT TC
Saci1287KOUp	CTC CAC TGA CAA CTT TAA ATT CCC CTA TTC TG
Saci1287KODown	GGG AAT TTA AAG TTG TCA GTG GAG GAG AAA GTT TTA G
Saci1287KOUpBamHI	GAT GGA TCC TAA TCT GCC AGA GCA ATG
Saci1287KODownSalI	GAT GTC GAC CTA AGT GAG TGG TCA CAA AC
In frame deletion of Saci_1317 (pSVA431)	
Saci1317fwNcoI	GAT CCA TGG GAT AAC ACA TTG GGC CAG AG
Saci1317reKpnI	GAT GGT ACC TAA CTT CTG ACC GCG GCT TC
Saci1317KOUp	CCT CTC TCT TCA TAA GGT TTC CTA CTT ACA TCA TAC
Saci1317KODown	GTA GGA AAC CTT ATG AAG AGA GAG GAT TTA TTA AC
Saci1317KOUpBamHI	GAT GGA TCC TAC CAC CGG ATA TGT TCA ATG

Saci1317KODownSalI	GAT GTC GAC CTT TTC TCA CTG AGA CTA TTT C
Cloning of Saci_1520 into pEC-A-HiSumo	
Saci1520rev	GCA AAG CAC CGG CCT CGT TAA CTA CTT TTA CCT TTT ATT TCA TTA ACT C
Saci1520fw	ACC AGG AAC AAA CCG GCG GCC GCT CGA TGT CTA AAC CCT CGT ATG
Cloning of Saci_1346 & 1347 into pET-Duet	
N-Fibr-rev1	GTG GTC GAC TCA CTT ATA CTT TAC TAC TAT C
N-Fibr-for1	TGA GGA TCC GAT GTC TGA AAT CGA AAA AGT TAC
Nop5-for1	TGA CAT ATG AAG ATA TAT TTA GTG GAA C
Nop5-rev1	ACA CAT ATG TTA TCT CCT TCC CTC TTT TTT AC

Table 4.11: Oligonucleotides used for amplification of C/D box sRNA tandem version 1 as control for Northern blot analyses. The T7 promoter sequence is underlined.

Name	Sequence 5' → 3'
tandem version1	
invitroLNA48For	TAA TAC GAC TCA CTA TAG GGA GAA TGG TAA AGA TGA AGA ATT AC
invitroLNA48Rev	ATT TTC ATC GTG TCA ACC ATT C

Table 4.12: Oligonucleotides used for hybridization and subsequent ligation into pUC19 for *in vitro* run-off transcription. The T7 promoter sequences are underlined.

Name	Sequence 5' → 3'
Sac-sR125/126 tandem	
S4748BamHind1for	GAT CCT AAT ACG ACT CAC TAT AGG GAG ATA AAG ATG AAG AAT TAC CCC CGA ACC TGA AAA GTG ATG AAA AGG
S4748EcoHind1rev	TCT GAC AAA TGC CTT TTC ATC ACT TTT CAG GTT CGG GGG TAA TTC TTC ATC TTT ATC TCC CTA TAG TGA GTC GTA TTA G
S4748EcoHind2for	CAT TTG TCA GAT TAT TTA GTG ATG AGC TTG ACC GCA TAT GAA AAT TGA TGA TTT AAT AAA GGT AGC CTC TGA TTT TCC A
S4748EcoHind2rev	AGC TTG GAA AAT CAG AGG CTA CCT TTA TTA AAT CAT CAA TTT TCA TAT GCG GTC AAG CTC ATC ACT AAA TAA
Sac-sR121 with 156 nt downstream region	
nS39UTRBamHind1for	GAT CCT AAT ACG ACT CAC TAT AGG GAG ATA ACA GAT GAT GAA CAC TGG CTG TAT TGA CCA AAT GAT ATA AAA ACC TTC TAA TGT CTG A
nS39UTRBamHInd1rev	TTC ATG AGA TAT CAG ACA TTA GAA GGT TTT TAT ATC ATT TGG TCA ATA CAG CCA GTG TTC ATC ATC TGT TCT CCC TAT AGT GAG TCG TAT TAG
nS39UTRBamHind2for	TAT CTC ATG AAT CCC GTA CAT ATT CTC GCA AAG AAA GGA GAT GTG GCA GAA AGA GTA ATA ATA GCA GGT GAT CCC GGC AGG GTA AGG TAT ATA
nS39UTRBamHind2rev	AGC TTA TAT ACC TTA CCC TGC CGG GAT CAC CTG CTA TTA TTA CTC TTT CTG CCA CAT CTC CTT TCT TTG CGA GAA TAT GTA CGG GA
k-turn	
Kturn-For	GAT CCT AAT ACG ACT CAC TAT AGG GAG AAA CTG ATG ATG ACG CTA TAC CCT CTG ACA CGT GA
Kturn-Rev	AGC TTC ACG TGT CAG AGG GTA TAG CGT CAT CAT CAG TTT CTC CCT ATA GTG AGT CGT ATT AG
C/D box sRNA	
CD39IvFor	GAT CCT AAT ACG ACT CAC TAT AGG GAG AAC AGA TGA TGA ACA CTG GCT GTA TTG ACC AAA TGA TAT AAA AAC CTT CTA ATG TCT GAT ATC TCA

CD39IvRev	AGC TTG AGA TAT CAG ACA TTA GAA GGT TTT TAT ATC ATT TGG TCA ATA CAG CCA GTG TTC ATC ATC TGT TCT CCC TAT AGT GAG TCG TAT TAG
C/D box sRNA with disrupted boxD/D' motifs	
CD39KKOFor	GAT CCT AAT ACG ACT CAC TAT AGG GAG AAC AGA TGA TGA ACA CTG GCT GTA TTC CCC AAA TGA TAT AAA AAC CTT CTA ATG TCT CTT ATC TCA
CD39KKORev	AGC TTG AGA TAA GAG ACA TTA GAA GGT TTT TAT ATC ATT TGG GGA ATA CAG CCA GTG TTC ATC ATC TGT TCT CCC TAT AGT GAG TCG TAT TAG

Table 4.13: Oligonucleotides used for inverse RT-PCR analyses.

Name	Sequence 5' → 3'
Sac-sR121*	
S39RTFor	GAT ATA AAA ACC TTC TAA TGT CTG
39RT-PCR Rev (zi	TGG TCA ATA CAC GTG TCA AC
tandem version 1 spacer 3 nt	
tandemFor (RT)	TTC AGG TTC GGG GGT AAT TCT TC
tandem4ntRev	GAT TAG TGA TGA ATG GTT GAC
complete tandem spacer 2 nt	
tandemFor (RT)	TTC AGG TTC GGG GGT AAT TCT TC
Tandem3ntRev	CAG AAT TTG ATG AAT GGT TGA C
Sac-sR24	
Saci6IRTfor	ATG TCC TGG ATC CTT GCT G
Saci6IRTrev	TAG CTC AGC CCT TGT AAG
Sac-sR121	
Saci39IRTfor	AAT GAT ATA AAA ACC TTC TAA TG
Saci39IRTrev	CAA TAC AGC CAG TGT TCA TC
Sac-sR10	
Saci11IRTfor	GAT GAC AAA AAG CGC GAG CG
Saci11IRTrevRT	TTC TCA GAT CCC GGA TTC CAC

Table 4.14: Oligonucleotides used for RT-PCR analyses.

Name	Sequence 5' → 3'
Sac-R125/126	
Saci4748forRT	AGA GGC TAC CTT TAT TAA ATC ATC
Saci4748revRT	GAT GAA GAA TTA CCC CCG AAC CTG
Sac-sR121 - Saci_1247	
Saci39forRT-2	CTC AGT ATG AAC AGT GTA GAG C
Saci39revRT	GAA CAC TGG CTG TAT TGA CCA AAT G
Saci_0125 - Sac-sR24	
Saci6revRT-2	ATT TGG TGT TGC TCT ACT CC
Saci6forRT	AGC AAG GAT CCA GGA CAT CAT AGC

Table 4.15: Oligonucleotides used for amplification of genomic DNA for the introduction of a 2nd copy of Sac-sR125 in the *S. acidocaldarius* genome and exchange of the native D guide.

Name	Sequence 5' → 3'
2 nd copy Sac-sR125 (pSVA406)	
Saci_1162_KI_Fw_up_ApaI (Alvaro Orell)	CTC ACT GGG CCC GAC CAA GAG CAG ACA AAG AG
Saci_1162_KI_Rv_dwn_BamHI (Alvaro Orell)	CGC CGA GGA TCC GTC ACT GAA CAA TAT CCC CTC
47Uprev	CCC GAA CCT GAA AAG TGA TGA AAA GGC ATT TGT CAG ATT CAT AAT TTA TAA CTA AAA TAT AAG G
47Downfor	TCA TCA CTT TTC AGG TTC GGG GGT AAT TCT TCA TCT TTA AAA TAA AAA TAG ATA TTC AAG
Sac-sR125 exchange D guide (pSVA406)	
CD47KORev	GTG GTC GAC TGT CCA GTT CTA ACG AGA AG
CD47KOFor	GTG CCG CGG AGG CTC GGC AAG AAT GTA TC
CD47g2LNArev	CTA AAT AAT CTG CGT GTC AAC CAT TCA TCA CTT TTC AGG TTC GGG GGT AAT TCT TC
CD47g2LNAfor	AAA GTG ATG AAT GGT TGA CAC GCA GAT TAT TTA GTG ATG AGC TTG ACC G

Table 4.16: Oligonucleotides used for hybridization and subsequent ligation into pSVA1431/pSVAMCS for C/D box sRNA Sac-sR121 *in vivo* assays in *S. acidocaldarius.*

Name	Sequence 5' → 3'
Sac-sR121*	
S39NcoEag 1for	CAT GGT ATT ATC CTA ATT TAT TCT AAT GCT TAT ATA AGC CTA TAT GAA CTC TGT TTA CAG ATG ATG AAT GGT TGA CAC GTG TA
S39NcoEag1rev	ATC ATT TGG TCA ATA CAC GTG TCA ACC ATT CAT CAT CTG TAA ACA GAG TTC ATA TAG GCT TAT ATA AGC ATT AGA ATA AAT TAG GAT AAT AC
S39NcoEag2for	TTG ACC AAA TGA TAT AAA AAC CTT CTA ATG TCT GAT ATC TCA TGA ATC CCG TAC ATA TTC TCG CAA AGA AAG GAG ATG TGG CAG AAA GAG TC
S39NcoEag2rev	GGC CGA CTC TTT CTG CCA CAT CTC CTT TCT TTG CGA GAA TAT GTA CGG GAT TCA TGA GAT ATC AGA CAT TAG AAG GTT TTT AT
w/o TATA box	
S39NcoEag1rev_2	ATC ATT TGG TCA ATA CAC GTG TCA ACC ATT CAT CAT CTG TAA ACA GAG TTC ATA TAG GCG GCG CGC CGC ATT AGA ATA AAT TAG GAT AAT AC
S39NcoEag1for_2	CAT GGT ATT ATC CTA ATT TAT TCT AAT GCG GCG CGC CGC CTA TAT GAA CTC TGT TTA CAG ATG ATG AAT GGT TGA CAC GTG TA
S39NcoEag2for	TTG ACC AAA TGA TAT AAA AAC CTT CTA ATG TCT GAT ATC TCA TGA ATC CCG TAC ATA TTC TCG CAA AGA AAG GAG ATG TGG CAG AAA GAG TC
S39NcoEag2rev	GGC CGA CTC TTT CTG CCA CAT CTC CTT TCT TTG CGA GAA TAT GTA CGG GAT TCA TGA GAT ATC AGA CAT TAG AAG GTT TTT AT
3' hairpin	
S39NcoEag1rev_2	ATC ATT TGG TCA ATA CAC GTG TCA ACC ATT CAT CAT CTG TAA ACA GAG TTC ATA TAG GCG GCG CGC CGC ATT AGA ATA AAT TAG GAT AAT AC
S39NcoEag1for_2	CAT GGT ATT ATC CTA ATT TAT TCT AAT GCG GCG CGC CGC CTA TAT GAA CTC TGT TTA CAG ATG ATG AAT GGT TGA CAC GTG TA
S39longHP_for	TTG ACC AAA TGA TAT AAA AAC CTT CTA ATG TCT GAT ATC TCC GGC CGC CCC CGG GCC CCC GGC CGG GGG CCC GGG GGC
S39longHP_rev	GGC CGC CCC CGG GCC CCC GGC CGG GGG CCC GGG GGC GGC CGG AGA TAT CAG ACA TTA GAA GGT TTT TAT
50 nt random us 1	

S39usrandomFor	CAT GGA CTC GAA TCG TGC TAG AGT TCT CCC AGA TCG TAG TGG GTG TTC ACG AAC CTA CAG ATG ATG AAT GGT TGA CAC GTG TA
S39usrandomRev	CTG TAG GTT CGT GAA CAC CCA CTA CGA TCT GGG AGA ACT CTA GCA CGA TTC GAG TC
S39NcoEag2for	TTG ACC AAA TGA TAT AAA AAC CTT CTA ATG TCT GAT ATC TCA TGA ATC CCG TAC ATA TTC TCG CAA AGA AAG GAG ATG TGG CAG AAA GAG TC
S39NcoEag2rev	GGC CGA CTC TTT CTG CCA CAT CTC CTT TCT TTG CGA GAA TAT GTA CGG GAT TCA TGA GAT ATC AGA CAT TAG AAG GTT TTT AT
50 nt random ds 1	
S39NcoEag1rev_2	ATC ATT TGG TCA ATA CAC GTG TCA ACC ATT CAT CAT CTG TAA ACA GAG TTC ATA TAG GCG GCG CGC CGC ATT AGA ATA AAT TAG GAT AAT AC
S39NcoEag1for_2	CAT GGT ATT ATC CTA ATT TAT TCT AAT GCG GCG CGC CGC CTA TAT GAA CTC TGT TTA CAG ATG ATG AAT GGT TGA CAC GTG TA
S39dsrandomFor	TTG ACC AAA TGA TAT AAA AAC CTT CTA ATG TCT GAT ATC TCA CTC GAA TCG TGC TAG AGT TCT CCC AGA TCG TAG TGG GTG TTC ACG AAC CC
S39dsrandomRev	GGC CGG GTT CGT GAA CAC CCA CTA CGA TCT GGG AGA ACT CTA GCA CGA TTC GAG TGA GAT ATC AGA CAT TAG AAG GTT TTT AT
50 nt random us 2	
S39usrandomFor2	CAT GGA GGG ATG ACA TAT TAT TTT AAG AAA CGG ACC GTT AGA CTA CTC AAT TCT CTA CAG ATG ATG AAT GGT TGA CAC GTG TA
S39usrandomRev2	ATC ATT TGG TCA ATA CAC GTG TCA ACC ATT CAT CAT CTG TAG AGA ATT GAG TAG TCT AAC GGT CCG TTT CTT AAA ATA ATA TGT CAT CCC TC
S39NcoEag2for	TTG ACC AAA TGA TAT AAA AAC CTT CTA ATG TCT GAT ATC TCA TGA ATC CCG TAC ATA TTC TCG CAA AGA AAG GAG ATG TGG CAG AAA GAG TC
S39NcoEag2rev	GGC CGA CTC TTT CTG CCA CAT CTC CTT TCT TTG CGA GAA TAT GTA CGG GAT TCA TGA GAT ATC AGA CAT TAG AAG GTT TTT AT
50 nt random ds 2	
S39NcoEag1rev_2	ATC ATT TGG TCA ATA CAC GTG TCA ACC ATT CAT CAT CTG TAA ACA GAG TTC TA TAG GCG GCG CGC CGC ATT AGA ATA AAT TAG GAT AAT AC
S39NcoEag1for_2	CAT GGT ATT ATC CTA ATT TAT TCT AAT GCG GCG CGC CGC CTA TAT GAA CTC TGT TTA CAG ATG ATG AAT GGT TGA CAC GTG TA
S39dsrandomFor2	TTG ACC AAA TGA TAT AAA AAC CTT CTA ATG TCT GAT ATC TCA GGG ATG ACA TAT TAT TTT AAG AAA CGG ACC GTT AGA CTA CTC AAT TCT CC
S39dsrandomRev2	GGC CGG AGA ATT GAG TAG TCT AAC GGT CCG TTT CTT AAA ATA ATA TGT CAT CCC TGA GAT ATC AGA CAT TAG AAG GTT TTT AT
boxD CTGG	
S39NcoEag1rev_2	ATC ATT TGG TCA ATA CAC GTG TCA ACC ATT CAT CAT CTG TAA ACA GAG TTC ATA TAG GCG GCG CGC CGC ATT AGA ATA AAT TAG GAT AAT AC
S39NcoEag1for_2	CAT GGT ATT ATC CTA ATT TAT TCT AAT GCG GCG CGC CGC CTA TAT GAA CTC TGT TTA CAG ATG ATG AAT GGT TGA CAC GTG TA
S39DCTGG 2for	TTG ACC AAA TGA TAT AAA AAC CTT CTA ATG TCT GGT ATC TCA TGA ATC CCG TAC ATA TTC TCG CAA AGA AAG GAG ATG TGG CAG AAA GAG TC
S39DCTGG 2rev	GGC CGA CTC TTT CTG CCA CAT CTC CTT TCT TTG CGA GAA TAT GTA CGG GAT TCA TGA GAT ACC AGA CAT TAG AAG GTT TTT AT
boxD CTTA	
S39NcoEag1rev_2	ATC ATT TGG TCA ATA CAC GTG TCA ACC ATT CAT CAT CTG TAA ACA GAG TTC ATA TAG GCG GCG CGC CGC ATT AGA ATA AAT TAG GAT AAT AC
S39NcoEag1for_2	CAT GGT ATT ATC CTA ATT TAT TCT AAT GCG GCG CGC CGC CTA TAT GAA CTC TGT TTA CAG ATG ATG AAT GGT TGA CAC GTG TA

S39DCTTA2for	TTG ACC AAA TGA TAT AAA AAC CTT CTA ATG TCT TAT ATC TCA TGA ATC CCG TAC ATA TTC TCG CAA AGA AAG GAG ATG TGG CAG AAA GAG TC
S39DCTTA 2rev	GGC CGA CTC TTT CTG CCA CAT CTC CTT TCT TTG CGA GAA TAT GTA CGG GAT TCA TGA GAT ATA AGA CAT TAG AAG GTT TTT AT
boxD CAGA	
S39NcoEag1rev_2	ATC ATT TGG TCA ATA CAC GTG TCA ACC ATT CAT CAT CTG TAA ACA GAG TTC ATA TAG GCG GCG CGC CGC ATT AGA ATA AAT TAG GAT AAT AC
S39NcoEag1for_2	CAT GGT ATT ATC CTA ATT TAT TCT AAT GCG GCG CGC CGC CTA TAT GAA CTC TGT TTA CAG ATG ATG AAT GGT TGA CAC GTG TA
CD CAGA/TGATGAfo	TTG ACC AAA TGA TAT AAA AAC CTT CTA ATG TCA GAT ATC TCA TGA ATC CCG TAC ATA TTC TCG CAA AGA AAG GAG ATG TGG CAG AAA GAG TC
CDCAGA/TGATGA-r	GGC CGA CTC TTT CTG CCA CAT CTC CTT TCT TTG CGA GAA TAT GTA CGG GAT TCA TGA GAT ATC TGA CAT TAG AAG GTT TTT AT
boxD GTGA	
S39NcoEag1rev_2	ATC ATT TGG TCA ATA CAC GTG TCA ACC ATT CAT CAT CTG TAA ACA GAG TTC ATA TAG GCG GCG CGC CGC ATT AGA ATA AAT TAG GAT AAT AC
S39NcoEag1for_2	CAT GGT ATT ATC CTA ATT TAT TCT AAT GCG GCG CGC CGC CTA TAT GAA CTC TGT TTA CAG ATG ATG AAT GGT TGA CAC GTG TA
S39boxD-GTGA_for	TTG ACC AAA TGA TAT AAA AAC CTT CTA ATG TGT GAT ATC TCA TGA ATC CCG TAC ATA TTC TCG CAA AGA AAG GAG ATG TGG CAG AAA GAG TC
S39boxD-GTGA_rev	GGC CGA CTC TTT CTG CCA CAT CTC CTT TCT TTG CGA GAA TAT GTA CGG GAT TCA TGA GAT ATC ACA CAT TAG AAG GTT TTT AT
boxD' TTGG	
S39D?TTGG1Rev-2	ATC ATT TGG CCA ATA CAC GTG TCA ACC ATT CAT CAT CTG TAA ACA GAG TTCATA TAG GCG GCG CGC CGC ATT AGA ATA AAT TAG GAT AAT AC
S39NcoEag1for_2	CAT GGT ATT ATC CTA ATT TAT TCT AAT GCG GCG CGC CGC CTA TAT GAA CTCTGT TTA CAG ATG ATG AAT GGT TGA CAC GTG TA
S39D?TTGG2for	TTG GCC AAA TGA TAT AAA AAC CTT CTA ATG TCT GAT ATC TCA TGA ATC CCG TAC ATA TTC TCG CAA AGA AAG GAG ATG TGG CAG AAA GAG TC
S39NcoEag2rev	GGC CGA CTC TTT CTG CCA CAT CTC CTT TCT TTG CGA GAA TAT GTA CGG GAT TCA TGA GAT ATC AGA CAT TAG AAG GTT TTT AT
boxD' TTTA	
S39D?TTTA1Rev-2	ATC ATT TGG TAA ATA CAC GTG TCA ACC ATT CAT CAT CTG TAA ACA GAG TTC ATA TAG GCG GCG CGC CGC ATT AGA ATA AAT TAG GAT AAT AC
S39NcoEag1for_2	CAT GGT ATT ATC CTA ATT TAT TCT AAT GCG GCG CGC CGC CTA TAT GAA CTCTGT TTA CAG ATG ATG AAT GGT TGA CAC GTG TA
S39D?TTTA2for	TTT ACC AAA TGA TAT AAA AAC CTT CTA ATG TCT GAT ATC TCA TGA ATC CCG TAC ATA TTC TCG CAA AGA AAG GAG ATG TGG CAG AAA GAG TC
S39NcoEag2rev	GGC CGA CTC TTT CTG CCA CAT CTC CTT TCT TTG CGA GAA TAT GTA CGG GAT TCA TGA GAT ATC AGA CAT TAG AAG GTT TTT AT
boxC' TGAAAT	
C'TTGA/TGAAGAr1	TTC ATT TGG TCA ATA CAC GTG TCA ACC ATT CAT CAT CTG TAA ACA GAG TTC ATA TAG GCG GCG CGC CGC ATT AGA ATA AAT TAG GAT AAT AC
S39NcoEag1for_2	CAT GGT ATT ATC CTA ATT TAT TCT AAT GCG GCG CGC CGC CTA TAT GAA CTCTGT TTA CAG ATG ATG AAT GGT TGA CAC GTG TA
C'TTGA/TGAAATf	TTG ACC AAA TGA AAT AAA AAC CTT CTA ATG TCT GAT ATC TCA TGA ATC CCG TAC ATA TTC TCG CAA AGA AAG GAG ATG TGG CAG AAA GAG TC
S39NcoEag2rev	GGC CGA CTC TTT CTG CCA CAT CTC CTT TCT TTG CGA GAA TAT

	GTA CGG GAT TCA TGA GAT ATC AGA CAT TAG AAG GTT TTT AT
boxD' ATGA	
C'ATGA/TGATATr1	ATC ATT TGG TCA TTA CAC GTG TCA ACC ATT CAT CAT CTG TAA ACA GAG TTC ATA TAG GCG GCG CGC CGC ATT AGA ATA AAT TAG GAT AAT AC
S39NcoEag1for_2	CAT GGT ATT ATC CTA ATT TAT TCT AAT GCG GCG CGC CGC CTA TAT GAA CTC TGT TTA CAG ATG ATG AAT GGT TGA CAC GTG TA
C'ATGA/TGATATf2	ATG ACC AAA TGA TAT AAA AAC CTT CTA ATG TCT GAT ATC TCA TGA ATC CCG TAC ATA TTC TCG CAA AGA AAG GAG ATG TGG CAG AAA GAG TC
S39NcoEag2rev	GGC CGA CTC TTT CTG CCA CAT CTC CTT TCT TTG CGA GAA TAT GTA CGG GAT TCA TGA GAT ATC AGA CAT TAG AAG GTT TTT AT
boxC' TGAAGA	
C'TTGA/TGAAGAr1	TTC ATT TGG TCA ATA CAC GTG TCA ACC ATT CAT CAT CTG TAA ACA GAG TTC ATA TAG GCG GCG CGC CGC ATT AGA ATA AAT TAG GAT AAT AC
S39NcoEag1for_2	CAT GGT ATT ATC CTA ATT TAT TCT AAT GCG GCG CGC CGC CTA TAT GAA CTC TGT TTA CAG ATG ATG AAT GGT TGA CAC GTG TA
C'TTGA/TGAAGAf2	TTG ACC AAA TGA AGA AAA AAC CTT CTA ATG TCT GAT ATC TCA TGA ATC CCG TAC ATA TTC TCG CAA AGA AAG GAG ATG TGG CAG AAA GAG TC
C'ATGA/TGATGAr	GGC CGA CTC TTT CTG CCA CAT CTC CTT TCT TTG CGA GAA TAT GTA CGG GAT TCA TGA GAT ATC AGA CAT TAG AAG GTT TTT TC

Table 4.17: Oligonucleotides used for hybridization and subsequent ligation into pSVA1431/pSVAMCS for C/D box sRNA Sac-sR121 *in vivo* assays in *S. acidocaldarius.*

Name	**Sequence 5' → 3'**
7 nt spacer	
4748-8nt-2for	CAT GGT AAA GAT GAA GAA TTA CCC CCG AAC CTG AAA AGT GAT GAA AAG GCA TTT GTC AGA TTA TTT AGT GAT GAA TGG TTG ACA CGA TGA AAA TC
4748-8nt-2rev	GGC CGA TTT TCA TCG TGT CAA CCA TTC ATC ACT AAA TAA TCT GAC AAA TGC CTT TTC ATC ACT TTT CAG GTT CGG GGG TAA TTC TTC ATC TTT AC
3 nt spacer	
4748-4nt-1for	CAT GGT AAA GAT GAA GAA TTA CCC CCG AAC CTG AAA AGT GAT GAA AAG GCA TTT GTC AGA TTA GTG ATG AAT GGT TGA CAC GTG TAA TGA AAA TC
4748-4nt-1rev	GGC CGA TTT TCA TTA CAC GTG TCA ACC ATT CAT CAC TAA TCT GAC AAA TGC CTT TTC ATC ACT TTT CAG GTT CGG GGG TAA TTC TTC ATC TTT AC
2 nt spacer	
4748-3nt-1for	CAT GGT AAA GAT GAA GAA TTA CCC CCG AAC CTG AAA AGT GAT GAA AAG GCA TTT GTC AGA ATT TGA TGA ATG GTT GAC ACG TGT AAT GAA AAT C
4748-3nt-1rev	GGC CGA TTT TCA TTA CAC GTG TCA ACC ATT CAT CAA ATT CTG ACA AAT GCC TTT TCA TCA CTT TTC AGG TTC GGG GGT AAT TCT TCA TCT TTA C
sR125-DguideLNA	
4748LNAC?D8nt-a	CAT GGT AAA GAT GAA GAA TTA CCC CCG AAC CTG AAA AGT GAT GAA TGG TTG ACA CGC AGA TTA TTT AGT GAT GAG CTT GAC CGC ATA TGA AAA TC
4748LNAC?D8nt-b	GGC CGA TTT TCA TAT GCG GTC AAG CTC ATC ACT AAA TAA TCT GCG TGT CAA CCA TTC ATC ACT TTT CAG GTT CGG GGG TAA TTC TTC ATC TTT AC
hairpin	
4748longGC-a	CAT GGT AAA GAT GAA GAA TTA CCC CCG AAC CTG AAA AGT GAT

	GAA AAG GCA TTT GTC AGA TTA TTT AGT GAT GA
4748longGC-b	GCC TTT TCA TCA CTT TTC AGG TTC GGG GGT AAT TCT TCA TCT TTA C
4748longGC-c	ATG GTT GAC ACG ATG AAA ATC CGG CCG CCC CCG GGC CCC CGG CCG GGG GCC CGG GGG C
4748longGC-d	GGC CGC CCC CGG GCC CCC GGC CGG GGG CCC GGG GGC GGC CGG ATT TTC ATC GTG TCA ACC ATT CAT CAC TAA ATA ATC TGA CAA AT
complete 7 nt sR126-D'guideLNA	
4748komp8nt-a	CAT GGT AAA GAT GAA GAA TTA CCC CCG AAC CTG AAA AGT GAT GAA AAG GCA TTT GTC AGA TTA TTT AGT GAT GAA TGG TTG ACA CG
4748komp8nt-b	CCT TTT CAT CAC TTT TCA GGT TCG GGG GTA ATT CTT CAT CTT TAC
4748komp8nt-c	ATG AAA ATT GAT GAT TTA ATA AAG GTA GCC TCT GAT TTT CCC
4748komp8nt-d	GGC CGG GAA AAT CAG AGG CTA CCT TTA TTA AAT CAT CAA TTT TCA TCG TGT CAA CCA TTC ATC ACT AAA TAA TCT GAC AAA TG
complete 2 nt	
4748komp3nt-a	CAT GGT AAA GAT GAA GAA TTA CCC CCG AAC CTG AAA AGT GAT GAA AAG GCA TTT GTC AGA ATT TGA TGA ATG GTT GAC ACG
4748komp8nt-b	CCT TTT CAT CAC TTT TCA GGT TCG GGG GTA ATT CTT CAT CTT TAC
4748komp8nt-c	ATG AAA ATT GAT GAT TTA ATA AAG GTA GCC TCT GAT TTT CCC
4748komp3nt-d	GGC CGG GAA AAT CAG AGG CTA CCT TTA TTA AAT CAT CAA TTT TCA TCG TGT CAA CCA TTC ATC AAA TTC TGA CAA ATG
complete 7 nt sR125-DguideLNA	
7ntcompLNA48C1f	CAT GGT AAA GAT GAA GAA TTA CCC CCG AAC CTG AAA AGT GAT GAA AAG GCA TTT GTC AGA TTA TTT AGT GAT GAG CTT GAC CGC AT
4748komp8nt-b	CCT TTT CAT CAC TTT TCA GGT TCG GGG GTA ATT CTT CAT CTT TAC
7ntcompLNA48C2f	ATG AAA ATT GAT GAA TGG TTG ACA CGC TGA TTT TCC C
7ntcompLNA48C2r	GGC CGG GAA AAT CAG CGT GTC AAC CAT TCA TCA ATT TTC ATA TGC GGT CAA GCT CAT CAC TAA ATA ATC TGA CAA ATG
sR126alone	
Saci48all-2for	CAT GGA TTT AGT GAT GAG CTT GAC CGC ATA TGA AAA TTG ATG AAT GGT TGA CAC GCT GAT TTT CCC
Saci48all-2rev	GGC CGG GAA AAT CAG CGT GTC AAC CAT TCA TCA ATT TTC ATA TGC GGT CAA GCT CAT CAC TAA ATC
sR125boxC, sR126D'guideLNA	
4748boxCLNA48a	CAT GGT AAA GAT CTA GAA TTA CCC CCG AAC CTG AAA AGT GAT GAA AAG GCA TTT GTC AGA TTA TTT AGT GAT GAA TGG TTG ACA CGA TGA AAA TC
4748boxCLNA48b	GGC CGA TTT TCA TCG TGT CAA CCA TTC ATC ACT AAA TAA TCT GAC AAA TGC CTT TTC ATC ACT TTT CAG GTT CGG GGG TAA TTC TAG ATC TTT AC
sR125boxC, DguideLNA	
4748boxCLNA47a	CAT GGT AAA GAT CTA GAA TTA CCC CCG AAC CTG AAA AGT GAT GAA TGG TTG ACA CGC AGA TTA TTT AGT GAT GAG CTT GAC CGC ATA TGA AAA TC
4748boxCLNA47b	GGC CGA TTT TCA TAT GCG GTC AAG CTC ATC ACT AAA TAA TCT GCG TGT CAA CCA TTC ATC ACT TTT CAG GTT CGG GGG TAA TTC TAG ATC TTT AC
sR126boxD', D'guide LNA	
4748boxD'LNA48a	CAT GGT AAA GAT GAA GAA TTA CCC CCG AAC CTG AAA AGT GAT GAA AAG GCA TTT GTC AGA TTA TTT AGT GAT GAA TGG TTG ACA CGA TCT AAA TC

4748boxD'LNA48b	GGC CGA TTT AGA TCG TGT CAA CCA TTC ATC ACT AAA TAA TCT GAC AAA TGC CTT TTC ATC ACT TTT CAG GTT CGG GGG TAA TTC TTC ATC TTT AC
sR126boxD', sR125Dguide LNA	
4748boxD'LNA47a	CAT GGT AAA GAT GAA GAA TTA CCC CCG AAC CTG AAA AGT GAT GAA TGG TTG ACA CGC AGA TTA TTT AGT GAT GAG CTT GAC CGC ATA TCT AAA TC
4748boxD'LNA47b	GGC CGA TTT AGA TAT GCG GTC AAG CTC ATC ACT AAA TAA TCT GCG TGT CAA CCA TTC ATC ACT TTT CAG GTT CGG GGG TAA TTC TTC ATC TTT AC
sR126boxD'CTGA	
48boxD'-CTGA-for	CAT GGT AAA GAT GAA GAA TTA CCC CCG AAC CTG AAA AGT GAT GAA AAG GCA TTT GTC AGA TTA TTT AGT GAT GAA TGG TTG ACA CGC TGA AAA TC
48boxD'-CTGA-rev	GGC CGA TTT TCA GCG TGT CAA CCA TTC ATC ACT AAA TAA TCT GAC AAA TGC CTT TTC ATC ACT TTT CAG GTT CGG GGG TAA TTC TTC ATC TTT AC
sR125boxD CTGA	
47boxD-CTGA-for	CAT GGT AAA GAT GAA GAA TTA CCC CCG AAC CTG AAA AGT GAT GAA AAG GCA TTT GTC TGA TTA TTT AGT GAT GAA TGG TTG ACA CGA TGA AAA TC
47boxD-CTGA-rev	GGC CGA TTT TCA TCG TGT CAA CCA TTC ATC ACT AAA TAA TCA GAC AAA TGC CTT TTC ATC ACT TTT CAG GTT CGG GGG TAA TTC TTC ATC TTT AC
all boxes consensus	
47CD-48D'-for	CAT GGT AAA GAT GAT GAA TTA CCC CCG AAC CTG AAA AGT GAT GAA AAG GCA TTT GTC TGA TTA TTT AGT GAT GAA TGG TTG ACA CGC TGA AAA TC
47CD-48D'-rev	GGC CGA TTT TCA GCG TGT CAA CCA TTC ATC ACT AAA TAA TCA GAC AAA TGC CTT TTC ATC ACT TTT CAG GTT CGG GGG TAA TTC ATC ATC TTT AC
sR125alone	
Saci47For	CAT GGT AAA GAT GAA GAA TTA CCC CCG AAC CTG AAA AGT GAT GAA TGG TTG ACA CGC AGA TTC
Saci47Rev	GGC CGA ATC TGC GTG TCA ACC ATT CAT CAC TTT TCA GGT TCG GGG GTA ATT CTT CAT CTT TAC
T2sR125DguideLNA	
Tandem2-47C'Dfo	CAT GGA AAG TGA TGA ATG GTT GAC ACG CAG ATT ATT TAG TGA TGA GCT TGA CCG CAT ATG AAA ATT GAT GAT TTA ATA AAG GTA GCC TCT GAT TTT CCC
Tandem2-47C'Dre	GGC CGG GAA AAT CAG AGG CTA CCT TTA TTA AAT CAT CAA TTT TCA TAT GCG GTC AAG CTC ATC ACT AAA TAA TCT GCG TGT CAA CCA TTC ATC ACT TTC
T2sR126D'guideLNA	
Tandem2-48CD'fo	CAT GGA AAG TGA TGA AAA GGC ATT TGT CAG ATT ATT TAG TGA TGA ATG GTT GAC ACG ATG AAA ATT GAT GAT TTA ATA AAG GTA GCC TCT GAT TTT CCC
Tandem2-48CD're	AGG CTA CCT TTA TTA AAT CAT CAA TTT TCA TCG TGT CAA CCA TTC ATC ACT AAA TAA TCT GAC AAA TGC CTT TTC ATC ACT TTC
T2sR126DguideLNA	
Tandem2-48C'Dfo	CAT GGA AAG TGA TGA AAA GGC ATT TGT CAG ATT ATT TAG TGA TGA GCT TGA CCG CAT ATG AAA ATT GAT GAA TGG TTG ACA CGC TGA TTT TCC C
Tandem2-48C'Dre	GGC CGG GAA AAT CAG CGT GTC AAC CAT TCA TCA ATT TTC ATA TGC GGT CAA GCT CAT CAC TAA ATA ATC TGA CAA ATG CCT TTT CAT CAC TTT C
T2sR125boxC'polyA	

T2-48C'D47C'pAf	CAT GGA AAG AAA AAA AAA GGC ATT TGT CAG ATT ATT TAG TGA TGA GCT TGA CCG CAT ATG AAA ATT GAT GAA TGG TTG ACA CGC TGA TTT TCC C
T2-48C'D47C'pAr	GGC CGG GAA AAT CAG CGT GTC AAC CAT TCA TCA ATT TTC ATA TGC GGT CAA GCT CAT CAC TAA ATA ATC TGA CAA ATG CCT TTT TTT TTC TTT C
T2sR125boxDpolyA	
T2-48C'D-47DpAf	CAT GGA AAG TGA TGA AAA GGC ATT TGT AAA ATT ATT TAG TGA TGA GCT TGA CCG CAT ATG AAA ATT GAT GAA TGG TTG ACA CGC TGA TTT TCC C
T2-48C'D-47DpAr	GGC CGG GAA AAT CAG CGT GTC AAC CAT TCA TCA ATT TTC ATA TGC GGT CAA GCT CAT CAC TAA ATA ATT TTA CAA ATG CCT TTT CAT CAC TTT C
T2sR126boxDpolyA	
T2-48C'D-48DpAf	CAT GGA AAG TGA TGA AAA GGC ATT TGT CAG ATT ATT TAG TGA TGA GCT TGA CCG CAT ATG AAA ATT GAT GAA TGG TTG ACA CGA AAA TTT TCC C
T2-48C'D-48Dpr	GGC CGG GAA AAT TTT CGT GTC AAC CAT TCA TCA ATT TTC ATA TGC GGT CAA GCT CAT CAC TAA ATA ATC TGA CAA ATG CCT TTT CAT CAC TTT C
T2sR126boxD'CTGA	
T2-47C'DbDCTGAf	CAT GGA AAG TGA TGA ATG GTT GAC ACG CAG ATT ATT TAG TGA TGA GCT TGA CCG CAT CTG AAA ATT GAT GAT TTA ATA AAG GTA GCC TCT GAT TTT CCC
T2-47C'DbDCTGAr	GGC CGG GAA AAT CAG AGG CTA CCT TTA TTA AAT CAT CAA TTT TCA GAT GCG GTC AAG CTC ATC ACT AAA TAA TCT GCG TGT CAA CCA TTC ATC ACT TTC
T2sR125boxDCTGA	
T2-47C'DbD'CTGAf	CAT GGA AAG TGA TGA ATG GTT GAC ACG CTG ATT ATT TAG TGA TGA GCT TGA CCG CAT ATG AAA ATT GAT GAT TTA ATA AAG GTA GCC TCT GAT TTT CCC
T2-47C'DbD'CTGAr	GGC CGG GAA AAT CAG AGG CTA CCT TTA TTA AAT CAT CAA TTT TCA TAT GCG GTC AAG CTC ATC ACT AAA TAA TCA GCG TGT CAA CCA TTC ATC ACT TTC
T2allboxesconsensus	
T247C'Dperf-for	CAT GGA AAG TGA TGA ATG GTT GAC ACG CTG ATT ATT TAG TGA TGA GCT TGA CCG CAT CTG AAA ATT GAT GAT TTA ATA AAG GTA GCC TCT GAT TTT CCC
T247C'Dperf-rev	GGC CGG GAA AAT CAG AGG CTA CCT TTA TTA AAT CAT CAA TTT TCA GAT GCG GTC AAG CTC ATC ACT AAA TAA TCA GCG TGT CAA CCA TTC ATC ACT TTC

Table 4.18: Oligonucleotides used as probes in Northern blot analyses. The probe which hybridizes to the 5S rRNA was already used in previous studies [148]. (+= LNA bases)

Target	Name	Sequence 5' → 3'
Artificial guide region	GuideSaci	CGTGT+CAA+CCAT+TCAT
Sac-sR10	Saci11probe	CGC TTT TTG TCA TCA TTC TCA GAT CCC GGA TTC CAC ATC
Sac-sR121	ProbeSaci39	TCA GAC ATT AGA AGG TTT TTA TAT CAT TTG GTC AAT ACA GCC AGT GTT C
Sac-sR125	NbSaci47	GAC AAA TGC CTT TTC ATC ACT TTT CAG GTT CGG GGG TAA TTC TTC
Sac-sR126	NbSaci48	GAG GCT ACC TTT ATT AAA TCA TCA ATT TTC ATA TGC GGT CAA GCT CAT C
5S rRNA	Nb5SrRNA	GAG CGG CTT AAC TTC CGG GT

Table 4.19: Oligonucleotides used for sequencing of constructs cloned into pSVA1431 and pSVAMCS

Name	Sequence 5' → 3'
Seqpr2_Saci1431	CGT ATT ACC GCC TTT GAG TGA GC
S1431KOPrrev-2	CTT TCT ATT TCG TCG GGT TC

4.4 Working with DNA

4.4.1 Preparation of plasmid DNA from *E. coli*

Plasmid DNA was prepared from *E. coli* overnight cell cultures using the QIAprep Spin Miniprep Kit or the QIAGEN Plasmid Plus Maxi Kit according to the manufacturer's instructions.

4.4.2 Nucleic acid precipitation

Nucleic acids were concentrated via ethanol precipitation [178]. 0.3 M Na-acetate and two volumes of 100 % ethanol (v/v) were added to the sample and incubated for 30 min at -20°C. The sample was centrifuged at 12,000 x g for 10 min at RT. Afterwards, the supernatant was removed and the DNA pellet was washed with 70 % ethanol (v/v) (12,000 x g, 2 min, RT). The supernatant was discarded and the DNA pellet was dried and resuspended in ddH_2O.

4.4.3 Phenol/chloroform extraction of DNA

To extract DNA from solutions, the sample was mixed with 1 volume of phenol/chloroform (1:1) pH 8.0. After centrifugation (15,000 x g, 1 min, RT), the upper aqueous phase was transferred into a fresh tube and 1 volume of chloroform was added and mixed with the sample. After centrifugation (15,000 x g, 1 min, RT), the upper aqueous phase was transferred into a fresh tube and DNA precipitation was performed subsequently (section 4.4.2).

4.4.4 Quantitative and qualitative analysis of DNA

The determination of the concentration and the purity control of DNA was performed with a spectrophotometer. The concentration of double-stranded DNA can be determined by measuring the extinction at a wavelength of λ=260 nm. An extinction value of 1 correlates to a concentration of 50 µg dsDNA/ml [178]. Proteins absorb at a wavelength of 280 nm. Therefore the A260/A280 ratio provides information on the purity of the DNA. Sufficiently pure DNA preparations show a ratio between 1.8 and 2.0.

4.4.5 Agarose gel electrophoresis of DNA

Agarose gel electrophoresis of DNA fragments (plasmid DNA and PCR products) was carried out to determine their size and amount. Depending on the size of the analyzed DNA

fragments, agarose gels with 1-2 % (w/v) agarose in TAE buffer (40 mM Tris-acetate, 1 mM EDTA, 20 mM acetic acid pH 8.0) and 0.1 µl/ml ethidium bromide (10 mg/ml) were used. Before loading onto the gel the samples were mixed with 6x loading dye (40 % (v/v) sucrose, 0.25 % (w/v) bromphenol blue, 0.25 % (w/v) xylene cyanol). Electrophoresis was performed at 90-120 V in TAE buffer. As size marker the 2-log DNA Ladder was used. After the electrophoresis the DNA was visualized by UV irradiation at 254 nm.

4.4.6 Gel extraction of agarose gels

For the extraction of DNA fragments from agarose gels, the appropriate bands were sliced out of the gel on a UV table. The extraction was performed using the QIAquick Gel Extraction Kit according to the manufacturer's instructions. The DNA was eluted in 20-30 µl ddH_2O.

4.4.7 Polymerase chain reactions (PCR)

The PCR technique was utilized for the exponential amplification of DNA *in vitro* [179, 180]. Two sequence-specific oligonucleotides hybridize to the 5' end of the coding and the non-coding strand, flanking the region that has to be amplified. They are the starting point for the elongation that is catalyzed by a heat-stable DNA polymerase.

The main steps are denaturation of dsDNA into ssDNA, annealing of oligonucleotides to their complementary DNA and elongation of the oligonucleotides in 5' to 3' direction by addition of complementary nucleotides. Multiple rounds of the three steps result in multiple copies of the target sequence. Every oligonucleotide has a certain annealing temperature that can be estimated by the formula Tm= 64.9 + 41 x (nG+nC-16.4)/(nA+nT+nG+nC) [181].

4.4.7.1 Amplification of plasmid DNA, genomic DNA or cDNA

The PCR amplifications were performed using 50 ng of template DNA (genomic, plasmid or cDNA) 0.5 µM of each oligonucleotide, 200 µM dNTPs, 3 % (v/v) dimethyl sulfoxide (DMSO) and the 10x reaction buffer of the suitable polymerase. 1 U of the Phusion polymerase or 2 U of the Taq polymerase were used. The PCR reactions were carried out in a thermal cycler. Standard cycling conditions are shown in table 4.20 and 4.21 and optimal cycling conditions were determined empirically for each DNA fragment.

Table 4.20: Standard-PCR using Phusion polymerase

	Step 1	**Step 2-35**		**Step 36**
Denaturation	98°C, 30 s	98°C	10 s	
Annealing		45-72°C	30 s	
Elongation		72°C	30 s/kb	72°C, 10 min

Table 4.21: Standard-PCR using Taq polymerase

	Step 1	Step 2-35		Step 36
Denaturation	95°C, 30 s	95°C	15 s	
Annealing		45-68°C	30 s	
Elongation		68°C	1 min/kb	68°C, 5 min

4.4.7.2 Overlap extension PCR

In a overlap extension PCR two PCR products with overlapping sequences are fused together to obtain the complete product. In total, four oligonucleotides were used. For the first PCR reactions that were performed according to section 4.4.7.1 two products with overlapping sequences were created. After purification of the PCR products both products were mixed and two external oligonucleotides were used that generate the long fusion product.

4.4.7.3 Colony-PCR of *S. acidocaldarius* transformants

Single colonies appearing on plates were analyzed for the acceptance of plasmids, successful genomic integrations or recombination events after selection. Single-colonies were picked and lysed in 30 µl 0.2 M NaOH and previous to the PCR reaction neutralized with 70 µl 0.2 M Tris pH 7.8. For the PCR reactions 0.5 µl lysate were used as template in 30 µl reactions using Phusion polymerase for amplification.

4.4.8 Enzymatic modification of DNA

4.4.8.1 Restriction of DNA

The restriction of DNA was performed with restriction endonucleases according to the manufacturer's instructions in the recommended buffer. The DNA was mixed with 10-20 U restriction enzyme/µg DNA and incubated at 37°C for 2 h.

4.4.8.2 5'-dephosphorylation of linearized vector-DNA

In order to avoid re-ligation of the digested plasmid DNA, Antarctic Phosphatase treatment was used to remove the 5'-end phosphoryl groups. 0.5 U/ pmol DNA of Antarctic phosphatase and the corresponding buffer were added to the restriction mixture and then further incubated at 37°C for 1 h.

4.4.9 Ligation

T4 DNA ligase catalyzes the formation of a phosphodiester bond between 3' hydroxyl and 5' phosphate groups of duplex DNA using ATP as energy source and coenzyme. A molar ration of PCR fragment and restricted, dephosphorylated plasmid DNA of 4:1 was used for

the ligation reaction. A reaction mixture of 8 µl was prepared and incubated at 45°C for 5 min to dissolve secondary structures. After the addition of 1 µl T4 DNA ligase and the appropriate ligation buffer (10x) the mixture was incubated overnight at 16°C. After inactivation of the T4 DNA ligase for 10 min at 70°C, the recombinant plasmids were used for transformation.

4.4.10 Phosphorylation of oligonucleotides

Prior to hybridization 1 nmol of each oligonucleotide was 5'-phosphorylated in separate reactions that contained 50 U of T4-Polynucleotide Kinase (PNK), 1x of the corresponding reaction buffer, and 20 mM ATP in a 20 µl reaction volume. The reaction mixture was incubated at 37°C for 1 h.

4.4.11 Hybridization of oligonucleotides

For the hybridization of oligonucleotides, 1 µl of each phosphorylated oligonucleotide was mixed with T4 DNA ligase buffer (10x) and the reaction volume was filled with ddH_2O to 10 µl. The hybridization occured by heating the reaction for 5 min at 95°C and gradual cool down to RT for 1-2 h. After the hybridisation, the mixture was ligated into appropriate plasmids (section 4.4.9).

4.4.12 Transformation

4.4.12.1 Preparation of chemical competent *E. coli* cells

For the preparation of $CaCl_2$ competent *E. coli* cells of all strains 100 ml LB medium was inoculated with 2 ml of an overnight culture. Additionally, 10 mM $MgCl_2$ und 10 mM $MgSO_4$ were supplemented. The cultures were incubated at 37°C and 200 rpm up to an OD_{600nm} of 0.6. Then the culture was incubated on ice for 30 min and centrifuged (8000 rpm, 8 min, 4°C). The cell pellet was resuspended in 33 ml cold RF1 solution (100 mM RbCl, 50 mM $MnCl_2$ x 4 H_2O, 30 mM potassium acetate, 10 mM $CaCl_2$ x 2 H_2O, 15 % (v/v) glycerol pH 5.8). After an incubation on ice for 30 min, the cells were centrifuged (8000 rpm, 8 min, 4°C). The cell pellet was resuspended in 5 ml cold RF2 solution (10 mM RbCl, 10 mM MOPS (0.5 M stock solution pH 6.8), 75 mM $CaCl_2$ x H_2O, 15 % (v/v) glycerol pH 5.8) and incubated on ice for 30 min. Aliquots of 100 µl were stored at -80°C.

4.4.12.2 Transformation of competent *E. coli* cells

For the transformation of plasmid DNA into *E. coli* the complete ligation reaction or 2 µl of isolated plasmid DNA were mixed with 100 µl of competent *E. coli* cells. After incubation on ice for 30 min the transformation occured by an heat-shock at 42°C for 1 min and subsequent incubation on ice for 1 min. After addition of 1 ml LB medium, the cells were incubated at 37°C for 1 h. 100 µl of the cells were plated on LB agar plates that contain the

respective antibiotics. The remaining cells were centrifuged at 400 rpm for 4 min and resuspended in 100 µl of LB medium prior to plating on a second antibiotics-containing LB agar plate. The LB agar plates were incubated at 37°C overnight and colonies were screened for positive clones that contain the recombinant plasmid by restriction digest (section 4.4.8.1) of the isolated plasmid DNA (section 4.4.1).

4.4.12.3 Preparation of electrocompetent *S. acidocaldarius* cells

S. acidocaldarius MW001 cells were grown in 50 ml Brock medium supplemented with 0.1 % (w/v) tryptone, 0.2 % (w/v) dextrin and 20 µg/ml uracil and adjusted to pH 3.5 with sulfuric acid. When the culture reached an OD_{600} of 0.3-0.7 after 1-2 days growth, an aliquot (200-500 µl) was transferred into 50 ml fresh medium and harvested at an OD_{600} of 0.1-0.3. The culture was cooled down on ice, then centrifuged for 20 min at 2000 x g and washed three times successively with 50 ml, 10 ml and 1 ml ice cold 20 mM sucrose. The pellet was resuspended in ice cold 20 mM sucrose to a theoretical final OD_{600nm} of 15 and aliquotted in 50 µl portions. The aliquots were directly used for transformation or frozen at -80°C without using liquid nitrogen for storage.

4.4.12.4 Transformation of competent *S. acidocaldarius* cells

Prior to transformation into *S. acidocaldarius* the plasmids were methylated to prevent restriction by the SuaI restriction enzyme. For this purpose the plasmids were transformed into *E. coli* ER1821 bearing the additional plasmid pM.EsaBC4I. Methylated plasmids were electroporated in electrocompetent *S. acidocaldarius* MW001 cells using a GenePulser II with a constant time protocol with the input parameters 1.5 kV, 25 µF, 600 Ω in 1 mm cuvettes. Cells were regenerated in 450 µl Brock medium supplemented with 0.1 % (w/v) tryptone and 0.2 % (w/v) dextrin and adjusted to pH 4.5 and incubated for 30 min at 75°C and 300 rpm in a heating block. 100 µl of the cells were plated on first selection plates. The remaining cells were centrifuged at full speed for 15 s and resuspended in 100 µl Brock medium prior to plating onto a second first selection plate. The plates were sealed in plastic bags to avoid drying-out and incubated for seven days at 75°C. Transformants were analyzed by colony-PCR.

For the determination of the transformation efficiency 100 ng plasmid DNA was transformed in triplicates and the colonies were counted after 5 days of incubation. The transformation efficiency was calculated by the amount of colonies that were obtained for the transformation of the *in vivo* constructs versus the amount of colonies that were obtained for the transformation of the empty plasmid pSVA1431.

4.4.13 Sequencing

Automated sequencing of DNA was performed by Eurofins MWG Operon (Ebersberg). For the sequencing reaction standard primer were used except for constructs cloned into the vectors pSVA1431 and pSVAMCS. To sequence these constructs, the primers indicated in table 4.19 were used.

4.4.14 5'- terminal radioactive labeling of DNA

Single-stranded DNA oligonucleotides, synthesized by MWG Operon were used for radioactive labeling. 100 pmol of the oligonucleotides were mixed with 5 pmol of γ[^{32}P]-ATP (Hartmann Analytic), 10 U of T4 PNK and the corresponding reaction buffer (10x) and incubated at 37°C for 1 h.

4.4.15 Denaturing polyacrylamide gel electrophoresis of radiolabeled DNA

Electrophoretic separation of the radiolabeled DNA was performed with 8 to 12 % of the polyacrylamide stock solution containing acrylamide and bisacrylamide in a ratio of 29:1, depending on the size of the oligonucleotide. Additionally the gels contained 8 M urea, TBE buffer (90 mM Tris, 2 mM EDTA pH 8.0, 90 mM boric acid), 0.1 % (v/v) APS und 0.1 % (v/v) TEMED. Prior to loading on the gel the samples were mixed with 2x formamide loading buffer (95 % (v/v) formamide, 0.025 % (w/v) bromphenol blue, 0.025 % (w/v) xylene cyanol, 5 mM EDTA pH 8.0) and incubated at 95°C for 5 min. The separation was carried out at 12 W.

4.4.16 Detection of radiolabeled DNA by phosphorimaging

Gels or blots were put into plastic bags and subsequently exposed to phosphor screens overnight. The bands on the phosphor screen were visualized with a phosphoimager.

4.4.17 Extraction of radiolabeled DNA from urea-polyacrylamide gels

The radiolabeled DNA visualized by phosphorimaging was cut out from the gel and dissolved in 500 µl of gel elution buffer (20 mM Tris/HCl pH 7.5, 250 mM sodium acetate, 1 mM EDTA pH 8.0). After incubation at -20°C for 30 min the samples were placed on ice and shaked overnight. The supernatant was ethanol precipitated (section 4.4.2) but additionally glycogen was added in a 1:100 ratio. The radioactivity was measured in a scintillation counter.

4.5 Working with RNA

4.5.1 Treatment of solutions, glassware and equipment

To ensure protection against RNases all buffers and solutions were treated with with 0.1 % (v/v) diethylpyrocarbonate (DEPC), incubated overnight at RT and autoclaved. Reusable plastic ware was treated with RNase Exitus Plus™.

4.5.2 Isolation of small and total RNA from *S. acidocaldarius* and treatment with RNaseR

4.5.2.1 mirVana™ miRNA Isolation Kit (Ambion)

For the isolation of small RNAs (< 200 nt), the mirVana™ Isolation Kit was used. The isolation was performed according to the manufacturer's instructions. The homogenization of 0.1 g cells of *S. acidocaldarius* was carried out in 1 ml lysis/binding buffer using a glass Teflon homogenizer for 3 min on ice. After the addition of 100 µl miRNA Homogenate Additive™ and 10 min incubation on ice, the RNAs were phenol/chloroform extracted. To enrich for small RNAs the sample was brought to an ethanol concentration of 25 % to immobilize large RNAs on a glass-fiber filter. Small RNAs were collected in the filtrate. The ethanol concentration in the filtrate was increased to 55 % and the filtrate was applied to a second glass-fiber filter to immobilize small RNAs. Both RNA fractions were washed and eluted in 100 µl Elution Solution.

4.5.2.2 TRIzol Reagent (Ambion)

For the isolation of total RNA from *S. acidocaldarius* 14 ml from an exponentially grown culture (OD_{600nm} 0.4-0.6) were harvested. The cell pellet was resuspended in 1 ml TRIzol Reagent and lysed for 5 min at room temperature. After addition of 200 µl chloroform, mixing and incubation at room temperature for 10 min, the solution was centrifuged (16,000 x g, 5 min, 4°C). The upper aqueous phase was transferred to a fresh tube and the RNA precipitated by the addition of 500 µl isopropanol and incubation at -20°C for 10 min. The RNA was pelleted by centrifugation (16,000 x g, 10 min, 4°C), washed two times with 1 ml 70 % ethanol, dried at room temperature and resuspended in 50 µl of H_2O_{DEPC}.

4.5.2.3 RNaseR treatment of *S. acidocaldarius* total RNA

The RNaseR treatment was performed as described in Danan *et al.* 2012 [99]. 60 U RNaseR were added to 10 µg small RNA of *S. acidocaldarius* MW001 and the reaction was carried out in 10 x reaction buffer at 37°C for 45 min. To remove the enzyme and salts from the reaction, ethanol precipitation was performed (section 4.4.2).

4.5.2.4 DNaseI treatment of total RNA

1 µg total RNA was incubated with 1 U DNaseI in the corresponding 10x reaction buffer supplemented with 2.5 mM $MnCl_2$ in a total reaction volume of 10 µl at 37°C for 1 h. After this incubation step, additional 1 U DNaseI was added to the reaction and incubation proceeded for 1 h at 37°C. The DNaseI was inactivated by the addition of 5 mM EDTA and incubation at 65°C for 10 min.

4.5.3 Phenol/chloroform extraction of RNA

To extract RNA from solutions, the samples were mixed with 1 volume of phenol/chloroform (1:1) pH 5.0. After centrifugation (15,000 x g, 1 min, RT), the upper aqueous phase was transferred into a fresh tube and 1 volume of chloroform was added and mixed with the sample. After centrifugation (15,000 x g, 1 min, RT), the upper aqueous phase was transferred into a fresh tube and RNA precipitation was performed subsequently (section 4.4.2).

4.5.4 Quantitative and qualitative analysis of RNA

The determination of the concentration and the purity control of RNA was performed with a spectrophotometer. The concentration of single-stranded RNA can be determined by measuring the extinction at a wavelength of λ=260 nm. An extinction value of 1 correlates to a concentration of 40 µg ssRNA/ml for RNAs that are smaller than 200 nt [178]. Proteins absorb at a wavelength of 280 nm. Therefore the A260/A280 ratio provides information on the purity of the RNA. Sufficiently pure RNA preparations show a ratio that is higher than 1.8.

4.5.5 *In vitro* run-off transcription

To generate RNA transcripts *in vitro*, run-off transcription was carried out. PCR products containing a T7 RNA promoter sequence were used as DNA templates. A 50 µl PCR reaction was mixed with 40 mM HEPES/KOH pH 8.0, 22 mM $MgCl_2$, 5 mM Dithiothreitol, 1 mM Spermidin, 4 mM of each NTP and 30 nM T7 RNA polymerase in a reaction volume of 1 ml. The reaction mixture was incubated at 37°C for 3 h. Afterwards the RNA was purified using denaturing polyacrylamide gel electrophoresis, gel-extraction and ethanol precipitation (section 4.4.2, 4.5.6, 4.5.7).

4.5.6 Denaturing polyacrylamide gel electrophoresis

Electrophoretic separation of RNA fragments was performed as described in section 4.4.15. As size standard the Low Range ssRNA Ladder was used. The separation was carried out at 150-200 V. The gels were stained with SYBR Gold ® Nucleic acid stain and the RNA was visualized by UV irradiation at 254 nm.

4.5.7 Gelextraction of RNA

The RNA with the appropriate size was cut out from the gel and dissolved in 500 µl of gel elution buffer (20 mM Tris/HCl pH 7.5, 250 mM sodium acetate, 1 mM EDTA pH 8.0). After incubation at -20°C for 30 min the samples were placed on ice and shaked overnight. The supernatant was ethanol precipitated (section 4.4.2) but additionally glycogen was added in a 1:100 ratio.

4.5.8 5'-terminal radioactive labeling of RNA

RNA transcripts generated via *in vitro* run-off transcription, were used for 5'-terminal radioactive labeling with γ[^{32}P] in a T4 PNK reaction as described for DNA in section 4.4.14. Prior to the labeling, RNA transcripts had to be 5'-dephosphorylated (section 4.4.8.2).

4.5.9 Northern blot analyses

A semi dry electrophoretic transfer system was used to transfer the RNA that was separated on denaturing polyacrylamide gels (section 4.4.15) onto a positively charged nylon membrane (Roti®-Nylon plus, pore size 0.45 µm). Prior to the transfer, the membrane, the gels as well as Whatman GB004, 3MM Paper were equilibrated in 1x TBE buffer for 5 min. The blot was assembled in the order 6x Whatman paper, nylon membrane, polyacrylamide gel, 6x Whatman paper and the transfer was performed for 2 h at 20 V. Subsequently, the RNA was UV-crosslinked to the membrane.

The membrane was pre-hybridized for 30 min at 42°C in ULTRAhyb-Oligo Hybridization Buffer (1 ml/10 cm^2 membrane) to block non-specific binding sites. The 5'-terminal radiolabeled probes (10^6 cpm/ml hybridization buffer) were applied to the hybridization buffer after incubation at 95°C for 5 min. The hybridization was performed over night at 42°C. The blot was washed twice, with 15 ml low stringency buffer (2x SSC, 0.1 % SDS) and with 15 ml high stringency buffer (1x SSC, 0.1 % SDS) for 30 min at 42°C each to remove unbound probe. The membranes were exposed to phosphor screens overnight and the bands on the phosphor screen were visualized with a phosphorimager.

4.5.10 Electrophoretic Mobility Shift Assays

The RNA binding activity of L7Ae was assessed with Electrophoretic Mobility Shift Assays and 5'-terminal radiolabeled *in vitro* transcribed substrates (section 4.5.5 and 4.5.8). 16 nM of labeled substrate and 0.035 to 12.25 µM L7Ae were mixed in a volume of 15 µl buffer (20 mM Tris pH 8.0, 200 mM NaCl, 10 mM β-Me, 10 % glycerol). The reactions were incubated at 70°C for 10 min and separated on 6 % native TBE polyacrylamide gels (section 4.4.15 without the addition of urea). Detection of radioactivity was carried out by phosphorimaging (section 4.4.16).

4.5.11 (Inverse) Reverse Transcription-PCR

cDNA synthesis with the SuperScriptIII Reverse Transcriptase was performed according to the manufacturer's instructions. 1 µg total RNA, 2 pmol of the gene-specific primer and 1 µl 10 mM dNTP mix were incubated in a total reaction volume of 13 µl at 65°C for 5 min to resolve secondary structures. After an incubation on ice for at least 1 min, 4 µl 5x First-Strand Buffer, 1 µl 0.1 M DTT, 1 µl RNase-Inhibitor and 1 µl SuperScript III Reverse Transcriptase were added to the reaction. The cDNA synthesis was performed at 55°C for 45 min. The enzyme was inactivated by heating the reaction to 70°C for 15 min. 2 µl of the cDNA was used for amplification with the Taq polymerase in a 50 µl reaction (section 4.4.7.1). For sequence determination of the iRT-PCR products, Topo cloning with the Topo ® TA cloning kit was performed according to the manufacturer's instructions.

4.5.12 RNA-sequencing, identification of C/D box sRNAs and prediction of their targets

Small RNA that was extracted with the mirVana miRNA isolation kit (section 4.5.2.1) was used. Preparation of the RNA-Seq libraries and Illumina HiSeq2000 sequencing were performed at the Max-Planck Genome Centre, Cologne. Sequencing reads were trimmed by the removal of linker sequences and other sequences using a quality score limit of 0.05. The trimmed reads were mapped to the reference genome of *S. acidocaldarius* DSM639 (GenBank: NC_007181) using CLC Genomics Workbench 5.0 (CLC Bio, Aarhus, Denmark). The following mapping parameters were used: mismatch cost: 2, insertion cost: 3, deletion cost: 3, length fraction: 0.5, similarity: 0.8. RNA molecules with a read coverage > 500 were manually checked for the existence of the conserved boxC/C' and boxD/D' motifs. 2'-*O*-methylations occur in the nucleotide that is complementary to the nucleotide in the guide region that is located five positions upstream of the boxD/D' motifs. These sites were computationally predicted by scanning for extended complementarity between the guide regions and the rRNAs. One mismatch, two GU base pairs and no bulges were allowed. The target sites were then mapped manually onto the rRNA alignment and the rRNA structure.

4.5.13 Bioinformatic analyses of the C/D box sRNA upstream and downstream regions

The 50 nt upstream and downstream sequences of C/D box sRNA genes were analyzed for conserved sequences and structures. To check for sequence conservation, sequences were clustered into related families based on a similarity matrix using Markov clustering with reasonable cutoff (80%) [182]. Afterwards, multiple sequence alignments were generated using MAFFT [183] and regions with at least 85 % sequence conservation and a length of 6 to 15 nucleotides were extracted. Finally, sequence logos were generated using WebLogo

[184]. For the identification of conserved structure motifs hierarchical cluster trees were generated using RNAclust which reflected sequence and structure similarity which was given by LocARNA [185]. For each node in the cluster tree consensus structures were searched.

4.6 Biochemical methods

4.6.1 Heterologous production of *S. acidocaldarius* proteins in *E. coli*

For the heterologous production of *S. acidocaldarius* proteins the plasmid constructs were used that are described in table 4.6. As vector the pET vector system (Novagen) and the pec-A-HI-SUMO vector were used. The heterologous expression was performed in *E. coli* Rosetta2 (DE3) pLysS described in table 4.3. This *E. coli* strain contains the RNA polymerase of the T7 bacteriophage and the genes were cloned that they are under the control of a T7 lac promoter. Through the addition of IPTG, an inductor of the lac operon, the T7 RNA polymerase gene is transcribed and the T7 RNA polymerase produced. The T7 RNA polymerase binds to the promoter sequence allowing for gene expression and subsequent protein production of the desired protein. The pLys plasmid encodes the T7 lysozyme which is a natural inhibitor of the T7 RNA polymerase. This ensures minimal expression of the target genes before induction with IPTG. The culture conditions are described in section 4.2.2.

4.6.2 Cell disruption, enrichment and purification of recombinant enzymes

Recombinant *E. coli* Rosetta2(DE) pLysS cells were resuspended in 5 ml buffer (20 mM Tris pH 8.0, 300 mM NaCl, 10 mM β-Me, 10 % glycerol, 10 % imidazole) per g cells and incubated with 1.5 mg lysozyme at 4°C for 30 min. Cells were sonicated 6 x 30 sec (40 % duty cycle, output control 4) while kept on ice. Unbroken cells and cell debris were removed via centrifugation (45,000 x g, 4°C, 45 min). Ni^{2+}-NTA affinity chromatography was performed on a FPLC. The soluble proteins were loaded onto an equilibrated Ni sepharose column and washed with buffer. For the 1 ml elution fractions a linear imidazole gradient was used to a final concentration of 1 M imidazole. Fractions that contained protein were identified via absorption at 280 nm wavelength and analyzed via SDS-PAGE (section 4.6.3). Fractions containing the heterologous produced protein were pooled and dialyzed over night in ZelluTrans Dialysis tubes (Carl Roth GmbH) with appropriate molecular weight cut-off (MWCO) to remove the imidazole. For further purification of L7Ae the sumoprotease was added during dialysis to remove the Sumo/His-tag. The Sumo/His-tag was removed from the solution by applying the dialyzed sample onto another Ni sepharose column and L7Ae was collected in the flow-through. Further purification and size determination of L7Ae as well as the Nop5/fibrillarin subcomplex were performed via gel-filtration chromatography using an equilibrated Superdex 200 column. The protein complexes were separated at a flow rate of

0.4 ml/min and 1 ml fractions were collected and analyzed. For the attempt to remove the bound nucleic acid, dialysis was performed to remove the salt from the buffer and the samples were applied to anion-exchange chromatography with a MonoQ column. The solution was loaded onto the equilibrated column, washed with buffer and eluted in 1 ml fractions with a linear salt gradient to a final concentration of 1 M NaCl.

4.6.3 Cell extract fractionation and test for nuclease activity

S. acidocaldarius cells were resuspended in 5 ml buffer (50 mM MES pH 6.0, 20 mM $MgCl_2$) per g cells. Cells were sonicated 4 x 30 sec (40 % duty cycle, output control 4) while kept on ice. Unbroken cells and cell debris were removed via centrifugation (25,000 x g, 4°C, 1 h). The soluble proteins were loaded either on a MonoS or a heparin column. For the 1 ml elution fractions a linear NaCl gradient was used to a final concentration of 1 M NaCl. The protein amount in the fractions was determined and 10 µg of the fractions were used in activity assays with 100.000 cpm 5'-labeled C/D box sRNA precursor *in vitro* transcripts. Additionally, 1 µg L7Ae and fibrillarin were added and the reaction volume was filled with buffer (50 mM MES pH 6.0, 20 mM $MgCl_2$) to 40 µl. The reaction as incubated for 10 min at 50°C and the products were separated on denaturing polyacrylamide gels (section 4.4.15).

4.6.4 SDS-polyacrylamide gel electrophoresis (SDS-PAGE)

SDS gel electrophoreses was used to separate denatured proteins in polyacrylamide gels [186]. The gels consist of a 4 % polyacrylamide (v/v) 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) and a 10-15 % separating gel (375 mM Tris/HCl pH 8.8, 0.1 % (v/v) SDS, 0.1 % (v/v) APS und 0.001 % (v/v) TEMED). The polyacrylamide concentration of the separating gel depends on the molecular weight of the analyzed protein. For the preparation of the gels an acrylamide stock solution was used that contains acrylamide und bisacrylamide in a ratio of 37.5:1. The ingredients of the separating gels were mixed and subsequently poured into the gel casting chamber. The gel was covered with ddH_2O and polymerized. Afterwards, the stacking gel solution was poured on top and a 10-sample well was placed within the stacking gel. After polymerization the gels were either used directly for electrophoresis or stored at 4°C.

Before electrophoresis the protein samples were mixed with 2x denaturing buffer (150 mM Tris/HCl pH 6.8, 1.2 % (w/v) SDS, 30 % (v/v) glycerol, 6.7 % (v/v) β-Mercaptoethanol, 0.0018 % (w/v) bromphenolblue) and incubated at 95°C for 5 min. SDS attaches in a ratio of 1.4 g/g protein and covers the charge of the proteins that they have a constant negative charge. Separation was performed at a constant voltage of 200 V in a Mini-PROTEAN Tetra Cell running chamber filled with electrophoresis buffer (0.1 % (w/v) SDS, 25 mM Tris/HCl pH

8.0, 200 mM glycine) for 45 min. As size standard the ColorPlus Prestained Protein Ladder, Broad Range was used. Proteins were visualized by gel staining with Instant Blue.

4.6.5 Protein quantitation

The Bio-Rad protein assay based on Bradford protein quantitation method [187] was used according to the manufacturer's instructions. The extinction of the samples was measured photometrically at a wavelength of 595 nm (Ultrospec 3000 *pro;* GE Healthcare Europe GmbH). A calibration curve was created with bovine serum albumine (BSA) using concentrations of 2, 4, 6, 8 und 10 µg/ml.

5. References

1. Omer AD, Lowe TM, Russell AG, Ebhardt H, Eddy SR, Dennis PP: Homologs of small nucleolar RNAs in Archaea. *Science* 2000, 288:517-522.
2. Maden BE, Corbett ME, Heeney PA, Pugh K, Ajuh PM: Classical and novel approaches to the detection and localization of the numerous modified nucleotides in eukaryotic ribosomal RNA. *Biochimie* 1995, 77:22-29.
3. Kiss-Laszlo Z, Henry Y, Bachellerie JP, Caizergues-Ferrer M, Kiss T: Site-specific ribose methylation of preribosomal RNA: a novel function for small nucleolar RNAs. *Cell* 1996, 85:1077-1088.
4. Tycowski KT, You ZH, Graham PJ, Steitz JA: Modification of U6 spliceosomal RNA is guided by other small RNAs. *Mol Cell* 1998, 2:629-638.
5. Decatur WA, Fournier MJ: rRNA modifications and ribosome function. *Trends Biochem Sci* 2002, 27:344-351.
6. Noon KR, Bruenger E, McCloskey JA: Posttranscriptional modifications in 16S and 23S rRNAs of the archaeal hyperthermophile *Sulfolobus solfataricus*. *J Bacteriol* 1998, 180:2883-2888.
7. Dennis PP, Omer A, Lowe T: A guided tour: small RNA function in Archaea. *Mol Microbiol* 2001, 40:509-519.
8. Su AA, Tripp V, Randau L: RNA-Seq analyses reveal the order of tRNA processing events and the maturation of C/D box and CRISPR RNAs in the hyperthermophile *Methanopyrus kandleri*. *Nucleic Acids Res* 2013, 41:6250-6258.
9. Randau L: RNA processing in the minimal organism *Nanoarchaeum equitans*. *Genome Biol* 2012, 13:R63.
10. Maxwell ES, Fournier MJ: The small nucleolar RNAs. *Annu Rev Biochem* 1995, 64:897-934.
11. Balakin AG, Smith L, Fournier MJ: The RNA world of the nucleolus: two major families of small RNAs defined by different box elements with related functions. *Cell* 1996, 86:823-834.
12. Ganot P, Bortolin ML, Kiss T: Site-specific pseudouridine formation in preribosomal RNA is guided by small nucleolar RNAs. *Cell* 1997, 89:799-809.
13. Ni J, Tien AL, Fournier MJ: Small nucleolar RNAs direct site-specific synthesis of pseudouridine in ribosomal RNA. *Cell* 1997, 89:565-573.
14. Helm M: Post-transcriptional nucleotide modification and alternative folding of RNA. *Nucleic Acids Res* 2006, 34:721-733.
15. Herschlag D, Eckstein F, Cech TR: The importance of being ribose at the cleavage site in the *Tetrahymena* ribozyme reaction. *Biochemistry* 1993, 32:8312-8321.
16. Watkins NJ, Bohnsack MT: 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* 2012, 3:397-414.
17. Gaspin C, Cavaille J, Erauso G, Bachellerie JP: Archaeal homologs of eukaryotic methylation guide small nucleolar RNAs: lessons from the *Pyrococcus* genomes. *J Mol Biol* 2000, 297:895-906.
18. Steitz JA, Tycowski KT: Small RNA chaperones for ribosome biogenesis. *Science* 1995, 270:1626-1627.
19. Polikanov YS, Melnikov SV, Soll D, Steitz TA: Structural insights into the role of rRNA modifications in protein synthesis and ribosome assembly. *Nat Struct Mol Biol* 2015, 22:342-344.

20. Caldas T, Binet E, Bouloc P, Richarme G: Translational defects of *Escherichia coli* mutants deficient in the Um(2552) 23S ribosomal RNA methyltransferase RrmJ/FTSJ. *Biochem Biophys Res Commun* 2000, 271:714-718.
21. Schimmang T, Tollervey D, Kern H, Frank R, Hurt EC: A yeast nucleolar protein related to mammalian fibrillarin is associated with small nucleolar RNA and is essential for viability. *EMBO J* 1989, 8:4015-4024.
22. King TH, Liu B, McCully RR, Fournier MJ: Ribosome structure and activity are altered in cells lacking snoRNPs that form pseudouridines in the peptidyl transferase center. *Mol Cell* 2003, 11:425-435.
23. Liang XH, Liu Q, Fournier MJ: Loss of rRNA modifications in the decoding center of the ribosome impairs translation and strongly delays pre-rRNA processing. *RNA* 2009, 15:1716-1728.
24. Buchhaupt M, Sharma S, Kellner S, Oswald S, Paetzold M, Peifer C, Watzinger P, Schrader J, Helm M, Entian KD: Partial methylation at Am100 in 18S rRNA of baker's yeast reveals ribosome heterogeneity on the level of eukaryotic rRNA modification. *PLoS One* 2014, 9:e89640.
25. Andersen TE, Porse BT, Kirpekar F: A novel partial modification at C2501 in *Escherichia coli* 23S ribosomal RNA. *RNA* 2004, 10:907-913.
26. Hansen MA, Kirpekar F, Ritterbusch W, Vester B: Posttranscriptional modifications in the A-loop of 23S rRNAs from selected archaea and eubacteria. *RNA* 2002, 8:202-213.
27. Henras AK, Soudet J, Gerus M, Lebaron S, Caizergues-Ferrer M, Mougin A, Henry Y: The post-transcriptional steps of eukaryotic ribosome biogenesis. *Cell Mol Life Sci* 2008, 65:2334-2359.
28. Tycowski KT, Shu MD, Steitz JA: Requirement for intron-encoded U22 small nucleolar RNA in 18S ribosomal RNA maturation. *Science* 1994, 266:1558-1561.
29. Huttenhofer A, Kiefmann M, Meier-Ewert S, O'Brien J, Lehrach H, Bachellerie JP, Brosius J: RNomics: an experimental approach that identifies 201 candidates for novel, small, non-messenger RNAs in mouse. *EMBO J* 2001, 20:2943-2953.
30. Chen CL, Chen CJ, Vallon O, Huang ZP, Zhou H, Qu LH: Genomewide analysis of box C/D and box H/ACA snoRNAs in *Chlamydomonas reinhardtii* reveals an extensive organization into intronic gene clusters. *Genetics* 2008, 179:21-30.
31. Lui L, Lowe T: Small nucleolar RNAs and RNA-guided post-transcriptional modification. *Essays Biochem* 2013, 54:53-77.
32. Lestrade L, Weber MJ: snoRNA-LBME-db, a comprehensive database of human H/ACA and C/D box snoRNAs. *Nucleic Acids Res* 2006, 34:D158-162.
33. Dutca LM, Gallagher JE, Baserga SJ: The initial U3 snoRNA:pre-rRNA base pairing interaction required for pre-18S rRNA folding revealed by *in vivo* chemical probing. *Nucleic Acids Res* 2011, 39:5164-5180.
34. Kiss-Laszlo Z, Henry Y, Kiss T: Sequence and structural elements of methylation guide snoRNAs essential for site-specific ribose methylation of pre-rRNA. *EMBO J* 1998, 17:797-807.
35. Szewczak LB, DeGregorio SJ, Strobel SA, Steitz JA: Exclusive interaction of the 15.5 kD protein with the terminal box C/D motif of a methylation guide snoRNP. *Chem Biol* 2002, 9:1095-1107.
36. Watkins NJ, Segault V, Charpentier B, Nottrott S, Fabrizio P, Bachi A, Wilm M, Rosbash M, Branlant C, Luhrmann R: A common core RNP structure shared between

the small nucleoar box C/D RNPs and the spliceosomal U4 snRNP. *Cell* 2000, 103:457-466.

37. Klein DJ, Schmeing TM, Moore PB, Steitz TA: The kink-turn: a new RNA secondary structure motif. *EMBO J* 2001, 20:4214-4221.
38. Rozhdestvensky TS, Tang TH, Tchirkova IV, Brosius J, Bachellerie JP, Huttenhofer A: 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* 2003, 31:869-877.
39. Reiter NJ, Osterman A, Torres-Larios A, Swinger KK, Pan T, Mondragon A: Structure of a bacterial ribonuclease P holoenzyme in complex with tRNA. *Nature* 2010, 468:784-789.
40. Zago MA, Dennis PP, Omer AD: The expanding world of small RNAs in the hyperthermophilic archaeon *Sulfolobus solfataricus*. *Mol Microbiol* 2005, 55:1812-1828.
41. Blouin S, Lafontaine DA: A loop loop interaction and a K-turn motif located in the lysine aptamer domain are important for the riboswitch gene regulation control. *RNA* 2007, 13:1256-1267.
42. Heppell B, Lafontaine DA: Folding of the SAM aptamer is determined by the formation of a K-turn-dependent pseudoknot. *Biochemistry* 2008, 47:1490-1499.
43. Lilley DM: The K-turn motif in riboswitches and other RNA species. *Biochim Biophys Acta* 2014, 1839:995-1004.
44. Nolivos S, Carpousis AJ, Clouet-d'Orval B: 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* 2005, 33:6507-6514.
45. Tran E, Zhang X, Lackey L, Maxwell ES: 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* 2005, 11:285-293.
46. Omer AD, Ziesche S, Ebhardt H, Dennis PP: *In vitro* reconstitution and activity of a C/D box methylation guide ribonucleoprotein complex. *PNAS* 2002, 99:5289-5294.
47. Appel CD, Maxwell ES: Structural features of the guide:target RNA duplex required for archaeal box C/D sRNA-guided nucleotide 2'-O-methylation. *RNA* 2007, 13:899-911.
48. Cavaille J, Nicoloso M, Bachellerie JP: Targeted ribose methylation of RNA *in vivo* directed by tailored antisense RNA guides. *Nature* 1996, 383:732-735.
49. Ziesche SM, Omer AD, Dennis PP: RNA-guided nucleotide modification of ribosomal and non-ribosomal RNAs in Archaea. *Mol Microbiol* 2004, 54:980-993.
50. van Nues RW, Granneman S, Kudla G, Sloan KE, Chicken M, Tollervey D, Watkins NJ: Box C/D snoRNP catalysed methylation is aided by additional pre-rRNA base-pairing. *EMBO J* 2011, 30:2420-2430.
51. Kuhn JF, Tran EJ, Maxwell ES: Archaeal ribosomal protein L7 is a functional homolog of the eukaryotic 15.5kD/Snu13p snoRNP core protein. *Nucleic Acids Res* 2002, 30:931-941.
52. Ban N, Nissen P, Hansen J, Moore PB, Steitz TA: The complete atomic structure of the large ribosomal subunit at 2.4 A resolution. *Science* 2000, 289:905-920.
53. Moore T, Zhang Y, Fenley MO, Li H: Molecular basis of box C/D RNA-protein interactions; cocrystal structure of archaeal L7Ae and a box C/D RNA. *Structure* 2004, 12:807-818.

54. Aittaleb M, Rashid R, Chen Q, Palmer JR, Daniels CJ, Li H: Structure and function of archaeal box C/D sRNP core proteins. *Nat Struct Biol* 2003, 10:256-263.
55. Lapinaite A, Simon B, Skjaerven L, Rakwalska-Bange M, Gabel F, Carlomagno T: The structure of the box C/D enzyme reveals regulation of RNA methylation. *Nature* 2013, 502:519-523.
56. Ghalei H, Hsiao HH, Urlaub H, Wahl MC, Watkins NJ: A novel Nop5-sRNA interaction that is required for efficient archaeal box C/D sRNP formation. *RNA* 2010, 16:2341-2348.
57. Gagnon KT, Biswas S, Zhang X, Brown BA, 2nd, Wollenzien P, Mattos C, Maxwell ES: Structurally conserved Nop56/58 N-terminal domain facilitates archaeal box C/D ribonucleoprotein-guided methyltransferase activity. *J Biol Chem* 2012, 287:19418-19428.
58. Wang H, Boisvert D, Kim KK, Kim R, Kim SH: Crystal structure of a fibrillarin homologue from *Methanococcus jannaschii*, a hyperthermophile, at 1.6 A resolution. *EMBO J* 2000, 19:317-323.
59. Zhang X, Champion EA, Tran EJ, Brown BA, 2nd, Baserga SJ, Maxwell ES: The coiled-coil domain of the Nop56/58 core protein is dispensable for sRNP assembly but is critical for archaeal box C/D sRNP-guided nucleotide methylation. *RNA* 2006, 12:1092-1103.
60. Bower-Phipps KR, Taylor DW, Wang HW, Baserga SJ: The box C/D sRNP dimeric architecture is conserved across domain Archaea. *RNA* 2012, 18:1527-1540.
61. Bleichert F, Gagnon KT, Brown BA, 2nd, Maxwell ES, Leschziner AE, Unger VM, Baserga SJ: A dimeric structure for archaeal box C/D small ribonucleoproteins. *Science* 2009, 325:1384-1387.
62. Xue S, Wang R, Yang F, Terns RM, Terns MP, Zhang X, Maxwell ES, Li H: Structural basis for substrate placement by an archaeal box C/D ribonucleoprotein particle. *Mol Cell* 2010, 39:939-949.
63. Lin J, Lai S, Jia R, Xu A, Zhang L, Lu J, Ye K: Structural basis for site-specific ribose methylation by box C/D RNA protein complexes. *Nature* 2011, 469:559-563.
64. Bleichert F, Baserga SJ: Dissecting the role of conserved box C/D sRNA sequences in di-sRNP assembly and function. *Nucleic Acids Res* 2010, 38:8295-8305.
65. Singh SK, Gurha P, Tran EJ, Maxwell ES, Gupta R: Sequential 2'-O-methylation of archaeal pre-tRNATrp nucleotides is guided by the intron-encoded but trans-acting box C/D ribonucleoprotein of pre-tRNA. *J Biol Chem* 2004, 279:47661-47671.
66. Graziadei A, Masiewicz P, Lapinaite A, Carlomagno T: Archaea box C/D enzymes methylate two distinct substrate rRNA sequences with different efficiency. *RNA* 2016.
67. Gagnon KT, Zhang X, Qu G, Biswas S, Suryadi J, Brown BA, 2nd, Maxwell ES: Signature amino acids enable the archaeal L7Ae box C/D RNP core protein to recognize and bind the K-loop RNA motif. *RNA* 2010, 16:79-90.
68. Granneman S, Kudla G, Petfalski E, Tollervey D: Identification of protein binding sites on U3 snoRNA and pre-rRNA by UV cross-linking and high-throughput analysis of cDNAs. *PNAS* 2009, 106:9613-9618.
69. Watkins NJ, Dickmanns A, Luhrmann R: Conserved stem II of the box C/D motif is essential for nucleolar localization and is required, along with the 15.5K protein, for the hierarchical assembly of the box C/D snoRNP. *Mol Cell Biol* 2002, 22:8342-8352.
70. Newman DR, Kuhn JF, Shanab GM, Maxwell ES: Box C/D snoRNA-associated proteins: two pairs of evolutionarily ancient proteins and possible links to replication and transcription. *RNA* 2000, 6:861-879.

71. Caffarelli E, Losito M, Giorgi C, Fatica A, Bozzoni I: *In vivo* identification of nuclear factors interacting with the conserved elements of box C/D small nucleolar RNAs. *Mol Cell Biol* 1998, 18:1023-1028.
72. Lafontaine DL, Tollervey D: Synthesis and assembly of the box C+D small nucleolar RNPs. *Mol Cell Biol* 2000, 20:2650-2659.
73. Lafontaine DL, Tollervey D: Nop58p is a common component of the box C+D snoRNPs that is required for snoRNA stability. *RNA* 1999, 5:455-467.
74. Tyc K, Steitz JA: U3, U8 and U13 comprise a new class of mammalian snRNPs localized in the cell nucleolus. *EMBO J* 1989, 8:3113-3119.
75. Ochs RL, Lischwe MA, Spohn WH, Busch H: Fibrillarin: a new protein of the nucleolus identified by autoimmune sera. *Biol Cell* 1985, 54:123-133.
76. Galardi S, Fatica A, Bachi A, Scaloni A, Presutti C, Bozzoni I: Purified box C/D snoRNPs are able to reproduce site-specific 2'-O-methylation of target RNA *in vitro*. *Mol Cell Biol* 2002, 22:6663-6668.
77. Saito H, Kobayashi T, Hara T, Fujita Y, Hayashi K, Furushima R, Inoue T: Synthetic translational regulation by an L7Ae-kink-turn RNP switch. *Nat Chem Biol* 2010, 6:71-78.
78. Dieci G, Preti M, Montanini B: Eukaryotic snoRNAs: a paradigm for gene expression flexibility. *Genomics* 2009, 94:83-88.
79. Weber MJ: Mammalian small nucleolar RNAs are mobile genetic elements. *PLoS Genet* 2006, 2:e205.
80. Li SG, Zhou H, Luo YP, Zhang P, Qu LH: Identification and functional analysis of 20 Box H/ACA small nucleolar RNAs (snoRNAs) from *Schizosaccharomyces pombe*. *J Biol Chem* 2005, 280:16446-16455.
81. Liang D, Zhou H, Zhang P, Chen YQ, Chen X, Chen CL, Qu LH: A novel gene organization: intronic snoRNA gene clusters from *Oryza sativa*. *Nucleic Acids Res* 2002, 30:3262-3272.
82. Brown JW, Echeverria M, Qu LH: Plant snoRNAs: functional evolution and new modes of gene expression. *Trends Plant Sci* 2003, 8:42-49.
83. Leader DJ, Clark GP, Watters J, Beven AF, Shaw PJ, Brown JW: Clusters of multiple different small nucleolar RNA genes in plants are expressed as and processed from polycistronic pre-snoRNAs. *EMBO J* 1997, 16:5742-5751.
84. Kruszka K, Barneche F, Guyot R, Ailhas J, Meneau I, Schiffer S, Marchfelder A, Echeverria M: Plant dicistronic tRNA-snoRNA genes: a new mode of expression of the small nucleolar RNAs processed by RNase Z. *EMBO J* 2003, 22:621-632.
85. Barbezier N, Canino G, Rodor J, Jobet E, Saez-Vasquez J, Marchfelder A, Echeverria M: Processing of a dicistronic tRNA-snoRNA precursor: combined analysis *in vitro* and *in vivo* reveals alternate pathways and coupling to assembly of snoRNP. *Plant Physiol* 2009, 150:1598-1610.
86. Tycowski KT, Aab A, Steitz JA: Guide RNAs with 5' caps and novel box C/D snoRNA-like domains for modification of snRNAs in metazoa. *Curr Biol* 2004, 14:1985-1995.
87. Leader DJ, Sanders JF, Waugh R, Shaw P, Brown JW: Molecular characterisation of plant U14 small nucleolar RNA genes: closely linked genes are transcribed as polycistronic U14 transcripts. *Nucleic Acids Res* 1994, 22:5196-5203.
88. Pelczar P, Filipowicz W: The host gene for intronic U17 small nucleolar RNAs in mammals has no protein-coding potential and is a member of the 5'-terminal oligopyrimidine gene family. *Mol Cell Biol* 1998, 18:4509-4518.

89. Bratkovic T, Rogelj B: Biology and applications of small nucleolar RNAs. *Cell Mol Life Sci* 2011, 68:3843-3851.
90. Chanfreau G, Rotondo G, Legrain P, Jacquier A: Processing of a dicistronic small nucleolar RNA precursor by the RNA endonuclease Rnt1. *EMBO J* 1998, 17:3726-3737.
91. Chanfreau G, Legrain P, Jacquier A: Yeast RNase III as a key processing enzyme in small nucleolar RNAs metabolism. *J Mol Biol* 1998, 284:975-988.
92. Allmang C, Kufel J, Chanfreau G, Mitchell P, Petfalski E, Tollervey D: Functions of the exosome in rRNA, snoRNA and snRNA synthesis. *EMBO J* 1999, 18:5399-5410.
93. Petfalski E, Dandekar T, Henry Y, Tollervey D: Processing of the precursors to small nucleolar RNAs and rRNAs requires common components. *Mol Cell Biol* 1998, 18:1181-1189.
94. Qu LH, Henras A, Lu YJ, Zhou H, Zhou WX, Zhu YQ, Zhao J, Henry Y, Caizergues-Ferrer M, Bachellerie JP: Seven novel methylation guide small nucleolar RNAs are processed from a common polycistronic transcript by Rat1p and RNase III in yeast. *Mol Cell Biol* 1999, 19:1144-1158.
95. Brown JW, Marshall DF, Echeverria M: Intronic noncoding RNAs and splicing. *Trends Plant Sci* 2008, 13:335-342.
96. Villa T, Ceradini F, Presutti C, Bozzoni I: Processing of the intron-encoded U18 small nucleolar RNA in the yeast *Saccharomyces cerevisiae* relies on both exo- and endonucleolytic activities. *Mol Cell Biol* 1998, 18:3376-3383.
97. Kiss T, Filipowicz W: Exonucleolytic processing of small nucleolar RNAs from pre-mRNA introns. *Genes Dev* 1995, 9:1411-1424.
98. Leader DJ, Clark GP, Watters J, Beven AF, Shaw PJ, Brown JW: Splicing-independent processing of plant box C/D and box H/ACA small nucleolar RNAs. *Plant Mol Biol* 1999, 39:1091-1100.
99. Danan M, Schwartz S, Edelheit S, Sorek R: Transcriptome-wide discovery of circular RNAs in Archaea. *Nucleic Acids Res* 2012, 40:3131-3142.
100. Starostina NG, Marshburn S, Johnson LS, Eddy SR, Terns RM, Terns MP: Circular box C/D RNAs in *Pyrococcus furiosus*. *PNAS* 2004, 101:14097-14101.
101. Plagens A, Daume M, Wiegel J, Randau L: Circularization restores signal recognition particle RNA functionality in *Thermoproteus*. *Elife* 2015, 4.
102. Chant J, Dennis P: Archaebacteria: transcription and processing of ribosomal RNA sequences in *Halobacterium cutirubrum*. *EMBO J* 1986, 5:1091-1097.
103. Tang TH, Rozhdestvensky TS, d'Orval BC, Bortolin ML, Huber H, Charpentier B, Branlant C, Bachellerie JP, Brosius J, Huttenhofer A: RNomics in Archaea reveals a further link between splicing of archaeal introns and rRNA processing. *Nucleic Acids Res* 2002, 30:921-930.
104. Kleman-Leyer K, Armbruster DW, Daniels CJ: Properties of *H. volcanii* tRNA intron endonuclease reveal a relationship between the archaeal and eucaryal tRNA intron processing systems. *Cell* 1997, 89:839-847.
105. Diener JL, Moore PB: Solution structure of a substrate for the archaeal pre-tRNA splicing endonucleases: the bulge-helix-bulge motif. *Mol Cell* 1998, 1:883-894.
106. Berkemer SJ, Höner zu Siederdissen C, Amman F, Wintsche A, Will S, Hofacker I, Prohaska SJ, Stadler PF: Processed Small RNAs in Archaea and BHB Elements. *GCB* 2015, 1:e18.

107. Weisel J, Wagner S, Klug G: The Nop5-L7A-fibrillarin RNP complex and a novel box C/D containing sRNA of *Halobacterium salinarum* NRC-1. *Biochem Biophys Res Commun* 2010, 394:542-547.
108. Waters E, Hohn MJ, Ahel I, Graham DE, Adams MD, Barnstead M, Beeson KY, Bibbs L, Bolanos R, Keller M, et al: The genome of *Nanoarchaeum equitans*: insights into early archaeal evolution and derived parasitism. *PNAS* 2003, 100:12984-12988.
109. Randau L, Munch R, Hohn MJ, Jahn D, Soll D: *Nanoarchaeum equitans* creates functional tRNAs from separate genes for their 5'- and 3'-halves. *Nature* 2005, 433:537-541.
110. Lamontagne B, Larose S, Boulanger J, Elela SA: The RNase III family: a conserved structure and expanding functions in eukaryotic dsRNA metabolism. *Curr Issues Mol Biol* 2001, 3:71-78.
111. Clouet d'Orval B, Bortolin ML, Gaspin C, Bachellerie JP: Box C/D RNA guides for the ribose methylation of archaeal tRNAs. The tRNATrp intron guides the formation of two ribose-methylated nucleosides in the mature tRNATrp. *Nucleic Acids Res* 2001, 29:4518-4529.
112. Danan M, Schwartz S, Edelheit S, Sorek R: Transcriptome-wide discovery of circular RNAs in Archaea. *Nucleic Acids Res* 2012, 40:3131-3142.
113. Brock TD, Brock KM, Belly RT, Weiss RL: Sulfolobus: a new genus of sulfur-oxidizing bacteria living at low pH and high temperature. *Arch Mikrobiol* 1972, 84:54-68.
114. Wagner M, van Wolferen M, Wagner A, Lassak K, Meyer BH, Reimann J, Albers SV: Versatile Genetic Tool Box for the Crenarchaeote *Sulfolobus acidocaldarius*. *Front Microbiol* 2012, 3:214.
115. Berkner S, Grogan D, Albers SV, Lipps G: 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* 2007, 35:e88.
116. Berkner S, Wlodkowski A, Albers SV, Lipps G: Inducible and constitutive promoters for genetic systems in *Sulfolobus acidocaldarius*. *Extremophiles* 2010, 14:249-259.
117. Chen L, Brugger K, Skovgaard M, Redder P, She Q, Torarinsson E, Greve B, Awayez M, Zibat A, Klenk HP, Garrett RA: The genome of *Sulfolobus acidocaldarius*, a model organism of the Crenarchaeota. *J Bacteriol* 2005, 187:4992-4999.
118. Omer AD, Zago M, Chang A, Dennis PP: Probing the structure and function of an archaeal C/D-box methylation guide sRNA. *RNA* 2006, 12:1708-1720.
119. Soppa J: Transcription initiation in Archaea: facts, factors and future aspects. *Mol Microbiol* 1999, 31:1295-1305.
120. Richter H, Zoephel J, Schermuly J, Maticzka D, Backofen R, Randau L: Characterization of CRISPR RNA processing in *Clostridium thermocellum* and *Methanococcus maripaludis*. *Nucleic Acids Res* 2012, 40:9887-9896.
121. Dennis PP, Tripp V, Lui L, Lowe T, Randau L: C/D box sRNA-guided 2'-O-methylation patterns of archaeal rRNA molecules. *BMC Genomics* 2015, 16:632.
122. Plagens A, Tripp V, Daume M, Sharma K, Klingl A, Hrle A, Conti E, Urlaub H, Randau L: *In vitro* assembly and activity of an archaeal CRISPR-Cas type I-A Cascade interference complex. *Nucleic Acids Res* 2014, 42:5125-5138.
123. Salgia SR, Singh SK, Gurha P, Gupta R: Two reactions of *Haloferax volcanii* RNA splicing enzymes: joining of exons and circularization of introns. *RNA* 2003, 9:319-330.
124. Wurtzel O, Sapra R, Chen F, Zhu Y, Simmons BA, Sorek R: A single-base resolution map of an archaeal transcriptome. *Genome Res* 2010, 20:133-141.

125. Slupska MM, King AG, Fitz-Gibbon S, Besemer J, Borodovsky M, Miller JH: Leaderless transcripts of the crenarchaeal hyperthermophile *Pyrobaculum aerophilum*. *J Mol Biol* 2001, 309:347-360.
126. Calvin K, Li H: RNA-splicing endonuclease structure and function. *Cell Mol Life Sci* 2008, 65:1176-1185.
127. Englert M, Sheppard K, Aslanian A, Yates JR, 3rd, Soll D: Archaeal 3'-phosphate RNA splicing ligase characterization identifies the missing component in tRNA maturation. *PNAS* 2011, 108:1290-1295.
128. Tanaka N, Meineke B, Shuman S: RtcB, a novel RNA ligase, can catalyze tRNA splicing and HAC1 mRNA splicing *in vivo*. *J Biol Chem* 2011, 286:30253-30257.
129. Greer CL, Javor B, Abelson J: RNA ligase in bacteria: formation of a 2',5' linkage by an *E. coli* extract. *Cell* 1983, 33:899-906.
130. Kato M, Shirouzu M, Terada T, Yamaguchi H, Murayama K, Sakai H, Kuramitsu S, Yokoyama S: Crystal structure of the 2'-5' RNA ligase from *Thermus thermophilus* HB8. *J Mol Biol* 2003, 329:903-911.
131. Kosmaczewski SG, Edwards TJ, Han SM, Eckwahl MJ, Meyer BI, Peach S, Hesselberth JR, Wolin SL, Hammarlund M: The RtcB RNA ligase is an essential component of the metazoan unfolded protein response. *EMBO Rep* 2014, 15:1278-1285.
132. Meyer BH, Albers SV: AglB, catalyzing the oligosaccharyl transferase step of the archaeal N-glycosylation process, is essential in the thermoacidophilic crenarchaeon *Sulfolobus acidocaldarius*. *Microbiologyopen* 2014, 3:531-543.
133. Bernick DL, Dennis PP, Lui LM, Lowe TM: Diversity of Antisense and Other Non-Coding RNAs in Archaea Revealed by Comparative Small RNA Sequencing in Four *Pyrobaculum* Species. *Front Microbiol* 2012, 3:231.
134. Cannone JJ, Subramanian S, Schnare MN, Collett JR, D'Souza LM, Du Y, Feng B, Lin N, Madabusi LV, Muller KM, et al: The comparative RNA web (CRW) site: an online database of comparative sequence and structure information for ribosomal, intron, and other RNAs. *BMC Bioinformatics* 2002, 3:2.
135. Wimberly BT, Brodersen DE, Clemons WM, Jr., Morgan-Warren RJ, Carter AP, Vonrhein C, Hartsch T, Ramakrishnan V: Structure of the 30S ribosomal subunit. *Nature* 2000, 407:327-339.
136. Lee CY, Lee A, Chanfreau G: The roles of endonucleolytic cleavage and exonucleolytic digestion in the 5'-end processing of *S. cerevisiae* box C/D snoRNAs. *RNA* 2003, 9:1362-1370.
137. Tycowski KT, Steitz JA: Non-coding snoRNA host genes in *Drosophila*: expression strategies for modification guide snoRNAs. *Eur J Cell Biol* 2001, 80:119-125.
138. Reiter WD, Hudepohl U, Zillig W: Mutational analysis of an archaebacterial promoter: essential role of a TATA box for transcription efficiency and start-site selection *in vitro*. *PNAS* 1990, 87:9509-9513.
139. Palmer JR, Daniels CJ: *In vivo* definition of an archaeal promoter. *J Bacteriol* 1995, 177:1844-1849.
140. Jager D, Forstner KU, Sharma CM, Santangelo TJ, Reeve JN: Primary transcriptome map of the hyperthermophilic archaeon *Thermococcus kodakarensis*. *BMC Genomics* 2014, 15:684.
141. Brenneis M, Hering O, Lange C, Soppa J: Experimental characterization of Cis-acting elements important for translation and transcription in halophilic archaea. *PLoS Genet* 2007, 3:e229.

142. Andersson AF, Lundgren M, Eriksson S, Rosenlund M, Bernander R, Nilsson P: Global analysis of mRNA stability in the archaeon *Sulfolobus*. *Genome Biol* 2006, 7:R99.
143. Mitchell P, Petfalski E, Tollervey D: The 3' end of yeast 5.8S rRNA is generated by an exonuclease processing mechanism. *Genes Dev* 1996, 10:502-513.
144. Mitchell P, Petfalski E, Shevchenko A, Mann M, Tollervey D: The exosome: a conserved eukaryotic RNA processing complex containing multiple 3'-->5' exoribonucleases. *Cell* 1997, 91:457-466.
145. Hasenohrl D, Konrat R, Blasi U: Identification of an RNase J ortholog in *Sulfolobus solfataricus*: implications for 5'-to-3' directional decay and 5'-end protection of mRNA in Crenarchaeota. *RNA* 2011, 17:99-107.
146. Clouet-d'Orval B, Rinaldi D, Quentin Y, Carpousis AJ: Euryarchaeal beta-CASP proteins with homology to bacterial RNase J Have 5'- to 3'-exoribonuclease activity. *J Biol Chem* 2010, 285:17574-17583.
147. Clouet-d'Orval B, Phung DK, Langendijk-Genevaux PS, Quentin Y: Universal RNA-degrading enzymes in Archaea: Prevalence, activities and functions of beta-CASP ribonucleases. *Biochimie* 2015, 118:278-285.
148. Martens B, Amman F, Manoharadas S, Zeichen L, Orell A, Albers SV, Hofacker I, Blasi U: Alterations of the transcriptome of *Sulfolobus acidocaldarius* by exoribonuclease aCPSF2. *PLoS One* 2013, 8:e76569.
149. Koonin EV, Wolf YI, Aravind L: Prediction of the archaeal exosome and its connections with the proteasome and the translation and transcription machineries by a comparative-genomic approach. *Genome Res* 2001, 11:240-252.
150. Walter P, Klein F, Lorentzen E, Ilchmann A, Klug G, Evguenieva-Hackenberg E: Characterization of native and reconstituted exosome complexes from the hyperthermophilic archaeon *Sulfolobus solfataricus*. *Mol Microbiol* 2006, 62:1076-1089.
151. Lorentzen E, Walter P, Fribourg S, Evguenieva-Hackenberg E, Klug G, Conti E: The archaeal exosome core is a hexameric ring structure with three catalytic subunits. *Nat Struct Mol Biol* 2005, 12:575-581.
152. Lorentzen E, Dziembowski A, Lindner D, Seraphin B, Conti E: RNA channelling by the archaeal exosome. *EMBO Rep* 2007, 8:470-476.
153. Hasenohrl D, Lombo T, Kaberdin V, Londei P, Blasi U: Translation initiation factor a/eIF2(-gamma) counteracts 5' to 3' mRNA decay in the archaeon *Sulfolobus solfataricus*. *PNAS* 2008, 105:2146-2150.
154. Belasco JG: All things must pass: contrasts and commonalities in eukaryotic and bacterial mRNA decay. *Nat Rev Mol Cell Biol* 2010, 11:467-478.
155. Caffarelli E, Fatica A, Prislei S, De Gregorio E, Fragapane P, Bozzoni I: Processing of the intron-encoded U16 and U18 snoRNAs: the conserved C and D boxes control both the processing reaction and the stability of the mature snoRNA. *EMBO J* 1996, 15:1121-1131.
156. Matera AG, Terns RM, Terns MP: Non-coding RNAs: lessons from the small nuclear and small nucleolar RNAs. *Nat Rev Mol Cell Biol* 2007, 8:209-220.
157. Watkins NJ, Leverette RD, Xia L, Andrews MT, Maxwell ES: Elements essential for processing intronic U14 snoRNA are located at the termini of the mature snoRNA sequence and include conserved nucleotide boxes C and D. *RNA* 1996, 2:118-133.
158. Xia L, Watkins NJ, Maxwell ES: Identification of specific nucleotide sequences and structural elements required for intronic U14 snoRNA processing. *RNA* 1997, 3:17-26.

159. Balakin AG, Lempicki RA, Huang GM, Fournier MJ: *Saccharomyces cerevisiae* U14 small nuclear RNA has little secondary structure and appears to be produced by post-transcriptional processing. *J Biol Chem* 1994, 269:739-746.
160. Leader DJ, Clark GP, Boag J, Watters JA, Simpson CG, Watkins NJ, Maxwell ES, Brown JW: Processing of vertebrate box C/D small nucleolar RNAs in plant cells. *Eur J Biochem* 1998, 253:154-160.
161. Tran EJ, Zhang X, Maxwell ES: Efficient RNA 2'-O-methylation requires juxtaposed and symmetrically assembled archaeal box C/D and C'/D' RNPs. *EMBO J* 2003, 22:3930-3940.
162. Schierling K, Rosch S, Rupprecht R, Schiffer S, Marchfelder A: tRNA 3' end maturation in archaea has eukaryotic features: the RNase Z from *Haloferax volcanii*. *J Mol Biol* 2002, 316:895-902.
163. Carte J, Wang R, Li H, Terns RM, Terns MP: Cas6 is an endoribonuclease that generates guide RNAs for invader defense in prokaryotes. *Genes Dev* 2008, 22:3489-3496.
164. Phung DK, Rinaldi D, Langendijk-Genevaux PS, Quentin Y, Carpousis AJ, Clouet-d'Orval B: Archaeal beta-CASP ribonucleases of the aCPSF1 family are orthologs of the eukaryal CPSF-73 factor. *Nucleic Acids Res* 2013, 41:1091-1103.
165. Petkovic S, Muller S: RNA circularization strategies *in vivo* and *in vitro*. *Nucleic Acids Res* 2015, 43:2454-2465.
166. Ho CK, Shuman S: Bacteriophage T4 RNA ligase 2 (gp24.1) exemplifies a family of RNA ligases found in all phylogenetic domains. *PNAS* 2002, 99:12709-12714.
167. Torchia C, Takagi Y, Ho CK: Archaeal RNA ligase is a homodimeric protein that catalyzes intramolecular ligation of single-stranded RNA and DNA. *Nucleic Acids Res* 2008, 36:6218-6227.
168. Darnell R: CLIP (cross-linking and immunoprecipitation) identification of RNAs bound by a specific protein. *Cold Spring Harb Protoc* 2012, 2012:1146-1160.
169. Puttaraju M, Been MD: Generation of nuclease resistant circular RNA decoys for HIV-Tat and HIV-Rev by autocatalytic splicing. *Nucleic Acids Symp Ser* 1995:49-51.
170. Mackie GA: Stabilization of circular rpsT mRNA demonstrates the 5'-end dependence of RNase E action *in vivo*. *J Biol Chem* 2000, 275:25069-25072.
171. Cavaille J, Bachellerie JP: Processing of fibrillarin-associated snoRNAs from pre-mRNA introns: an exonucleolytic process exclusively directed by the common stem-box terminal structure. *Biochimie* 1996, 78:443-456.
172. Hury J, Nagaswamy U, Larios-Sanz M, Fox GE: Ribosome origins: the relative age of 23S rRNA Domains. *Orig Life Evol Biosph* 2006, 36:421-429.
173. Petrov AS, Bernier CR, Hershkovits E, Xue Y, Waterbury CC, Hsiao C, Stepanov VG, Gaucher EA, Grover MA, Harvey SC, et al: Secondary structure and domain architecture of the 23S and 5S rRNAs. *Nucleic Acids Res* 2013, 41:7522-7535.
174. Cate JH, Yusupov MM, Yusupova GZ, Earnest TN, Noller HF: X-ray crystal structures of 70S ribosome functional complexes. *Science* 1999, 285:2095-2104.
175. Hansen JL, Schmeing TM, Moore PB, Steitz TA: Structural insights into peptide bond formation. *PNAS* 2002, 99:11670-11675.
176. Dong ZW, Shao P, Diao LT, Zhou H, Yu CH, Qu LH: RTL-P: a sensitive approach for detecting sites of 2'-O-methylation in RNA molecules. *Nucleic Acids Res* 2012, 40:e157.
177. Hanahan D: Studies on transformation of *Escherichia coli* with plasmids. *J Mol Biol* 1983, 166:557-580.

178. Sambrook J, Fritsch EF, T M: Molecular cloning: A laboratory manual 2nd Ed.; 1989.
179. Mullis K, Faloona F, Scharf S, Saiki R, Horn G, Erlich H: Specific enzymatic amplification of DNA *in vitro*: the polymerase chain reaction. *Cold Spring Harb Symp Quant Biol* 1986, 51 Pt 1:263-273.
180. Saiki RK, Gelfand DH, Stoffel S, Scharf SJ, Higuchi R, Horn GT, Mullis KB, Erlich HA: Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase. *Science* 1988, 239:487-491.
181. Wallace RB, Shaffer J, Murphy RF, Bonner J, Hirose T, Itakura K: Hybridization of synthetic oligodeoxyribonucleotides to phi chi 174 DNA: the effect of single base pair mismatch. *Nucleic Acids Res* 1979, 6:3543-3557.
182. Enright AJ, Van Dongen S, Ouzounis CA: An efficient algorithm for large-scale detection of protein families. *Nucleic Acids Res* 2002, 30:1575-1584.
183. Katoh K, Misawa K, Kuma K, Miyata T: MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Res* 2002, 30:3059-3066.
184. Crooks GE, Hon G, Chandonia JM, Brenner SE: WebLogo: a sequence logo generator. *Genome Res* 2004, 14:1188-1190.
185. Will S, Reiche K, Hofacker IL, Stadler PF, Backofen R: Inferring noncoding RNA families and classes by means of genome-scale structure-based clustering. *PLoS Comput Biol* 2007, 3:e65.
186. Laemmli UK: Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature* 1970, 227:680-685.
187. Bradford MM: A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Anal Biochem* 1976, 72:248-254.

6. Appendix

C/D box sRNAs of *S. acidocaldarius*, *M. maripaludis*, *M. kandleri*, *T. tenax*, *I. hospitalis*, *N. equitans*. The start position of the RNAs is indicated, the sequence can be found in [121]. The genes upstream and downstream of the C/D box sRNA genes in the same orientation are depicted with their distance to the C/D box sRNA genes. For the upstream gene positive numbers indicate that an overlap exists. For the downstream gene negative numbers indicate that an overlap exists. Additionally, conserved sequences in the C/D box sRNA 50 nt upstream region that were identified in the bioinformatic analyses are depicted. For *N. equitans* the analyses of the genetic context of the C/D box sRNA genes was performed before [9]. The italic written *I. hospitalis* C/D box sRNAs were not included in the computationally screening for putative promoter sequences.

C/D box sRNA	position	orientation	us gene	ds gene	Conserved sequences in the upstream region
Sac-sR101	2.496	pos	Saci_0003 (+9)	neg strand (OL)	TTTATTAAT-49/ GATATATG-34
Sac-sR9	42.135	pos	neg strand	Saci_0057 (+32)	CTTTAAAT-37
Sac-sR102	49.050	pos	Saci_0065 (+6)	neg strand (tRNA)	TTATTAA-29
Sac-sR12	86.093	pos	Saci_0114 (-23)	Saci_0116 (+23)	TTTATTAAG-30
Sac-sR24	97.321	pos	Saci_0125 (-2)	Saci_0127 (+31)	CTTAAAAAA-46
Sac-sR103	130.023	neg	pos strand	Saci_0159 (+25)	
SacsR104	179.594	neg	pos strand	Saci_0217 (+5)	
Sac-sR10	217.035	pos	Saci_0258 (-31)	Saci_0260 (+76)	
Sac-sR105	308.043	pos	neg strand	neg strand (OL)	ATTATAAAA-9
Sac-sR106	325.553	pos	Saci_0376 (-7)	neg strand (OL)	
Sac-sR107	328.391	pos	neg strand (OL)	neg strand (OL)	
Sac-sR13	345.685	pos	neg strand	neg strand	
Sac-sR6	368.184	pos	Saci_0428 (+25)	neg strand (OL)	
Sac-sR108	371.689	pos	Saci_0433 (-191)	neg strand	
Sac-sR7	393.907	neg	Saci_0467 (-203)	Saci_0645 (+32)	ATTTAAAT-30/ GATATATG-11
Sac-sR5	449.515	neg	Saci_0552 (+7)	pos strand (OL)	
Sac-sR20	458.752	pos	neg strand (tRNA)	neg strand (tRNA)	
Sac-sR110	490.870	neg	Saci_0614 (+2)	Saci_0613 (+44)	GTTTAAAA-33
Sac-sR111	559.377	neg	pos strand	pos strand (OL)	AATATATA-31
Sac-sR112	572.735	neg	Saci_0714 (+27)	Saci_0713 (-6)	
Sac-sR113	631.930	pos	neg strand	Saci_0789 (+65)	ATTTAAAC-37
Sac-sR114	669.612	neg	Saci_0837 (+8)	pos strand	
Sac-sR115	679.118	pos	Saci_0843 (+9)	Saci_0844 (+28)	TTTTATAAG-31
Sac-sR22	701.895	pos	Saci_0875 (+28)	Saci_0877 (+32)	CTTATAAAA-16
Sac-sR116	704.342	pos	Saci_0878 (+8)	Saci_0879 (+124)	CTTAATAAA-23/ TATATATC-49
Sac-sR117	714.983	pos	in Saci_0888	in Saci_0888	
Sac-sR118	755.543	pos	Saci_0943 (-89)	neg strand (OL)	
Sac-sR27	782.596	pos	Saci_0977(-96)	neg strand (OL)	GTTAATAAT-36/ TTTTATAAA-25
Sac-sR17	794.012	neg	Saci_0994 (+8)	pos strand (OL)	CTTATTAAA-16
Sac-sR28	808.273	pos	Saci_1007(+8)	Saci_1009(+27)	
Sac-sR119	830.429	neg	Saci_1032 (-39)	pos strand (OL)	ATTATAAAA-14
Sac-sR23	1.025.710	pos	neg strand	neg strand	ATTTAAAA-10/-31

Sac-sR120	1.046.530	neg	Saci_1230 (-189)	Saci_1229 (+51)	
Sac-sR121	1.062.319	neg	pos strand	Saci_1247 (+1)	
Sac-sR2	1.075.501	pos	in Saci_1262&1264	in Saci_1262&1264	
Sac-sR122	1.079.312	neg	Saci_1269 (-62)	tRNA-Ile (+95)	ATTAATAAT-18
Sac-sR14	1.117.506	neg	Saci_1314 (-243); C/D RNA 38 on neg strand	Saci_1311 (+43)	GTTTATAAG-32/ CTTATAAAT-19/ AATATATT-8
Sac-sR3	1.117.684	pos	neg strand	neg strand	
Sac-sR25	1.122.973	pos	neg strand	neg strand	
Sacs-R123	1.126.134	pos	Saci_1320 (+9)	Saci_1321 (+54)	TTTTAAAT-35/ GTTTTTAAT-43
Sac-sR124	1.131.270	neg	pos strand	Saci_1324 (-1)	CATATATT-30
Sac-sR125	1.152.390	pos	Saci_1350 (+29)	Saci_1351 (+25)	
Sac-sR126	1.152.447	pos	Saci_1350 (-29)	Saci_1351 (-41)	
Sac-sR127	1.165.286	pos	neg strand	neg strand (OL)	
Sac-sR128	1.172.733	pos	Saci_1371 (+6)	neg strand (OL)	ATTTATAAG-31
Sac-sR129	1.212.717	pos	neg strand	Saci_1420 (+5)	
Sac-sR1	1.220.557	pos	Saci_1428 (+19)	neg strand (OL)	GTTATTAAG-41
Sac-sR19	1.388.929	pos	Saci_1627 (+6)	Saci_1629 (+30)	CTTTAAAT-23
Sac-sR29	1.649.006	pos	neg strand	neg strand	
Sac-sR130	1.745.018	neg	Saci_1934 (-175)	Saci_1933 (+794)	
Sac-sR21	1.993.123	neg	pos strand	pos strand	
Sac-sR8	2.175.803	neg	Saci_2328 (+6)	pos strand	GATATATA-44
Sacs-R131	2.179.421	neg	Saci_2334 (-237)	pos strand (OL)	CTTTATAAA-29
Sac-sR18	2.179.504	pos	Saci_2332 (-135)	neg strand	
Sac-sR16	2.196.893	pos	neg strand	Sac-sR132 (+21)	
Sac-sR132	2.196.970	pos	Sac-sR16 (-21)	Saci_2352 (+134)	
Sac-sR4	2.217.167	neg	Saci_2370 (-67)	Saci_2368 (+8)	GATATATG-20
Sac-sR11	131.786	pos	Saci_0162 (+8)	neg strand	
Sac-sR109	450.909	pos	neg strand	Saci_0553 (+12)	
Sac-sR26	1.175.549	neg	pos strand	Saci_1374 (+50)	
Sac-sR15	1.400.391	pos	neg strand	Saci_1644 (+111)	
MmasR106	418.406	pos	neg strand	MmarC5_0461 (+88)	
MmasR105	675.145	neg	pos strand	pos strand	
MmasR107	729.113	neg	pos strand	MmarC5_0760 (+71)	
MmasR101	783.580	pos	MmarC5_0812 (-7)	neg strand	ATTTAAAC-12
MmasR102	984.364	pos	neg strand	MmarC5_1018 (+117)	GTTTAAAA-27/ GATATATA-41
MmasR104	988.104	pos	neg strand	MmarC5_1023 (+131)	TATATATC-13
MmasR103	1.244.136	pos	neg strand	neg strand	ATTTAAAT-38/ AATATATA-11/ TATATATT-24
MKAsR1	1.815	pos	neg strand	neg strand	
MKAsR2	43.044	neg	pos strand	pos strand	
MKAsR3	44.728	pos	neg strand (OL)	MK0048 (-3)	

MKAsR4	69.985	neg	pos strand	MK0074 (+25)	
MKAsR5	91.904	neg	pos strand	pos strand (OL)	
MKAsR6	111.152	pos	MK0115 (+9)	neg strand	
MKAsR7	137.488	pos	neg strand	neg strand (OL)	GTTAATAAA-46
MKAsR8	139.707	pos	MK0139 (+55)	MK0140 (+80)	
MKAsR9	146.560	pos	neg strand	MK0147 (-26)	
MKAsR10	147.313	pos	MK0147 (+9)	neg strand	
MKAsR11	170.141	neg	pos strand	MK0173 (-6)	
MKAsR12	170.850	pos	MK0174 (+7)	MK0175 (+12)	
MKAsR13	172.746	neg	MK0178 (-119)	pos strand (OL)	
MKAsR14	181.467	pos	MK0188 (+56)	MK0189 (-44)	
MKAsR15	203.475	pos	neg strand	neg strand (OL)	
MKAsR16	205.387	neg	MK0214 (-33); MKAsR17 (+8)	MK0213 (+21)	
MKAsR17	205.450	neg	MK0214 (+31)	MKAsR16 (-8)	
MKAsR18	219.379	neg	pos strand	pos strand (OL)	
MKAsR19	227.948	pos	MK0239 (+20)	neg strand	
MKAsR20	262.147	pos	neg strand (OL)	MK0273(+51)	
MKAsR21	271.135	neg	MK0280(-72); MkasR22(-16)	pos strand	
MKAsR22	271.216	neg	MK0280(+10)	MkasR21(+16)	
MKAsR23	274.089	pos	MK0282 (+4)	MK0283 (+355)	
MKAsR24	325.490	neg	MK0334 (-51); MkasR25 (-31)	pos strand (OL)	
MKAsR25	325.578	neg	MK0334 (+58)	MkasR24 (+31)	
MKAsR26	327.754	pos	neg strand	MK0336 (+28)	TTTTAAAC-29
MKAsR27	333.169	neg	in MK0340 (5'end)	in MK0340 (5' end)	
MKAsR28	346.603	neg	MK0358 (-83)	Mk0357 (+257)	
MKAsR29	360.606	neg	pos strand	pos strand (OL)	
MKAsR30	361.155	neg	MK03752 (+40)	pos strand	
MKAsR31	383.257	neg	pos strand	pos strand (OL)	
MKAsR32	384.946	neg	pos strand	pos strand	
MKAsR33	409.873	pos	neg strand	neg strand (OL)	
MKAsR34	459.532	neg	MK0498 (+10)	MK0497 (+5)	
MKAsR35	462.530	pos	MK0500 (-322)	MK0501 (-2)	
MKAsR36	509.783	pos	neg strand	MK0546 (+38)	
MKAsR90	1.118.640	neg	neg strand	neg strand	
MKAsR38	518.987	neg	MK0556(-363); MKAsR39(-216)	pos strand	
MKAsR39	519.280	neg	MK0556(-70)	pos strand; MKAsR38(+216)	
MKAsR40	519.358	neg	MK0556 (+9)	pos strand; MKAsR39(+10)	
MKAsR41	520.778	pos	neg strand	MkasR42 (+1)	
MKAsR42	520.845	pos	MkasR41 (-1)	MK0588 (+116)	
MKAsR43	524.051	pos	Mka0560 (+3)	MKAsR44 (+2)	
MKAsR44	524.115	pos	MKAsR43(-2)	MKAsR45 (-3)	
MKAsR45	524.175	pos	MKAsR44 (+3)	MKA0561 (+66)	
MKAsR46	537.577	pos	MK0577(-9)	neg strand	

MKAsR47	561.189	pos	MK0603(+10)	MK0604 (+49)	
MKAsR48	582.356	neg	pos strand	MK0622 (-29)	
MKAsR49	603.202	pos	MK0644 (+10)	MK0645 (+86)	
MKAsR50	605.153	pos	neg strand	MKAsR51 (+12)	
MKAsR51	605.246	pos	MKAsR51 (-12)	MK0647 (+173)	
MKAsR52	627.789	neg	MK0669(+10)	pos strand	
MKAsR53	700.385	neg	MK0735(+37)	pos strand	
MKAsR54	729.573	pos	MK0768(-10)	MK0769(+80)	
MKAsR55	751.600	pos	neg strand	neg strand (OL)	
MKAsR56	755.226	neg	neg strand	neg strand (OL)	
MKAsR57	766.793	neg	MK0806(-128)	MK0805(+5)	
MKAsR58	772.696	pos	MK0811(-86)	neg strand (OL)	
MKAsR59	776.072	pos	MK0815(+35)	MK0816(+79)	
MKAsR60	779.256	pos	neg strand	MK0820(+119); MkasR61 neg strand	
MKAsR61	779.375	neg	pos strand	MK0819(+72); MkasR60 pos strand	
MKAsR62	790.585	neg	MK0831(-78)	MK0830(+5)	
MKAsR63	825.703	pos	MK0873 (+9)	MK0874(+205)	
MKAsR64	830.184	pos	MK0878(-36)	MK0879(-22)	
MKAsR65	839.861	neg	MK0892(+33)	pos strand (OL)	
MKAsR66	843.279	pos	MK0894(+3)	neg strand (OL)	
MKAsR67	858.313	pos	neg strand	MK0903(+18)	
MKAsR68	879.751	pos	MK0921(+9)	MK0922(+242); neg strand MKAsR69	
MKAsR69	879.992	neg	pos strand; pos strand MKAsR68	pos strand	
MKAsR70	892.485	pos			
MKAsR71	900.585	pos	neg strand	MK0942(+922)	
MKAsR72	902.819	neg	MK0943 (-13)	pos strand	
MKAsR73	915.132	neg	pos strand	pos strand (OL)	
MKAsR74	961.063	pos	MK1000(+8)	MK1001(+122)	
MKAsR75	963.170	neg	MK1003(-177)	pos strand	
MKAsR76	993.170	pos	MK1032(-16)	neg strand (OL)	
MKAsR77	1.003.081	pos	MK1039(+1)	Mk1040 (+14)	
MKAsR78	1.022.622	neg	MKAsR79(-6)	MK1054(+208)	
MKAsR79	1.022.697	neg	pos strand	MKAsR78(+6)	
MKAsR80	1.023.968	pos	MK1055(-159)	neg strand	
MKAsR81	1.043.749	neg	MK1076(+11)	pos strand	
MKAsR82	1.048.100	pos	MK1080(-122)	neg strand (OL)	
MKAsR83	1.055.620	neg	MK1086(-14)	MK1085 (+4)	
MKAsR84	1.066.036	neg	MK1096(-203)	MK1095 (+288)	
MKAsR85	1.067.563	neg	MK1097 (-2)	MK1096(+143)	
MKAsR86	1.073.432	neg	pos strand	MK1102(+202)	
MKAsR87	1.077.825	pos	neg strand	neg strand (OL)	
MKAsR88	1.094.725	neg	pos strand	MK1120 (+784)	CTTTTTAAG-43
MKAsR89	1.104.331	neg	pos strand	MK1128(+579)	
MKAsR37	515.598	neg	pos strand	pos strand (OL)	

MKAsR91	1.157.353	pos	MK1168(-198)	neg strand	
MKAsR92	1.164.281	pos	MK1173 (-19)	MK1174(+1182); MkasR93neg strand	
MKAsR93	1.165.328	neg	pos strand	pos strand; MKAsR92pos strand	
MKAsR94	1.166.529	pos	MK1174 (-342)	neg strand	
MKAsR95	1.178.423	neg	MK1183(-148)	MK1182(+576)	
MKAsR96	1.182.419	pos	MK1186(-395)	MK1187(+21)	
MKAsR97	1.183.540	neg	pos strand (OL)	pos strand (OL)	
MKAsR98	1.187.688	pos	MK1192 (-159)	MK1193(-16)	
MKAsR99	1.204.769	pos	MK1214(+2)	MK1215(+5)	
MKAsR100	1.221.416	neg	MK1236(-84)	MK1235(+8)	
MKAsR101	1.231.253	pos	MK1248(-59)	MKAsR102(+360)	
MKAsR102	1.231.699	pos	MKAsR101(-360)	neg strand (OL)	
MKAsR103	1.232.371	pos	neg strand	neg strand	
MKAsR104	1.233.144	pos	neg strand	MKAsR105 neg strand	
MKAsR105	1.233.469	neg	MK1251(-902)	MKAsR104 pos strand	
MKAsR106	1.234.397	neg	MK1251(+27)	MKAsR106	
MKAsR107	1.263.413	pos	neg strand	MK1269(+543)	
MKAsR108	1.383.423	neg	pos strand	pos strand	
MKAsR109	1.386.498	pos	neg strand	MK1363(+964)	
MKAsR110	1.415.669	neg	pos strand	pos strand (OL)	
MKAsR111	1.417.369	neg	MKsR112(-10)	MK1390(+38)	
MKAsR112	1.417.441	neg	MKsR113(-2)	MKsR111(+10)	
MKAsR113	1.417.506	neg	pos strand	MKsR112(+2)	
MKAsR114	1.492.448	pos	neg strand	neg strand (OL)	
MKAsR115	1.493.125	pos	neg strand	MK1467(+44)	
MKAsR116	1.495.316	pos	MK1470(+17)	Mk1471(+43)	
MKAsR117	1.506.611	pos	neg strand	neg strand (OL)	
MKAsR118	1.508.796	neg	pos strand	pos strand (OL)	
MKAsR119	1.510.576	neg	pos strand	pos strand (OL)	
MKAsR120	1.531.147	neg	pos strand	MK1517 (-57)	
MKAsR121	1.561.232	pos	MK1542(+3)	neg strand (OL)	
MKAsR122	1.587.806	neg	MKAsR127(-61)	neg strand (OL)	
MKAsR127	1.587.927	neg	MK1575(+4)	MKAsR122(+61)	
MKAsR123	1.605.995	pos	MK1595(-9)	neg strand (OL)	
MKAsR124	1.607.286	neg	pos strand	MK1597(+3)	
MKAsR125	1.607.334	pos	neg strand	MK1598(+2)	
MKAsR126	1.633.733	pos	MK1632(-6)	MK1633(+17)	
TtesR101	1.969	pos	TTX0003(+9)	TTX0004(+58)	
TtesR102	12.706	pos	TTX0014(+9)	TTX0015(-36)	
TtesR103	103.182	pos	TTX0108(+9)	neg strand (OL)	
TtesR104	331.054	neg	pos strand	pos strand (OL)	
TtesR105	486.306	neg	pos strand (OL)	pos strand	
TtesR106	525.863	pos	TTX0595(+34)	neg strand (OL)	
TtesR107	599.878	neg	TTX0674(-388)	pos strand (OL)	CTTATTAAG-34

TtesR108	655.152	neg	TTX0757(-48)	pos strand	GTTTTTAAC-26
TtesR109	666.213	neg	pos strand	pos strand	
TtesR110	667.982	pos	TTX0767(-6)	neg strand (OL)	
TtesR111	673.077	pos	neg strand	TTX0776(+61)	
TtesR112	697.736	pos	neg strand	neg strand (OL)	TTTTAAAT-17
TtesR113	716.477	pos	neg strand	neg strand (OL)	
TtesR114	752.641	neg	TTX0867(-147)	pos strand	ATTTAAAC-25
TtesR115	754.470	pos	neg strand	neg strand	GTTAAAAAG-11
TtesR116	803.113	pos	TTX0927(-6)	TTX0928 (+26)	
TtesR117	813.942	pos	TTX0939(-13)	neg strand (OL)	GATATATT-22
TtesR118	818.001	pos	TTX0943(-101)	TTX0944(+53)	ATTTAAAT-27
TtesR119	851.950	neg	pos strand	pos strand	TTTTAAAG-42/ GTTAAAAAT-30
TtesR120	894.377	neg	TTX1036(+8)	TTX1035(+28)	
TtesR121	905.167	neg	TTX1049(-2)	pos strand (OL)	
TtesR122	969.195	neg	pos strand	pos strand (OL)	TTTTAAAA-30
TtesR123	1.002.233	pos	neg strand	neg strand	
TtesR124	1.065.840	pos	neg strand	TXX1227(+135)	
TtesR125	1.263.500	neg	TTX1434 (-28)	pos strand (OL)	
TtesR126	1.263.843	neg	TTX1435(-63); TtesR127 pos strand	TTX1434(-13)	
TtesR127	1.263.867	pos	TtesR126 neg strand	neg strand	GTTATTAAC-27
TtesR128	1.299.283	neg	TTX1469(-69); TtesR129 on pos strand	TTX1468(+44)	ATTAATAAG-47
TtesR129	1.299.298	pos	neg strand	neg strand (OL)	ATTAATAAA-21
TtesR130	1.301.114	neg	pos strand	TTX1471(+150)	
TtesR131	1.326.226	neg	pos strand	pos strand (OL)	
TtesR132	1.333.258	pos	TTX1512(+54)	TTX1513(+251)	
TtesR133	1.379.655	pos	neg strand	neg strand (OL)	
TtesR134	1.403.849	neg	tRNA29(-6)	TTX1599(+25)	
TtesR135	1.428.798	neg	pos strand	pos strand (OL)	TTTAATAAC-28
TtesR136	1.431.446	pos	neg strand	TTX1632(+22)	ATTATAAAC-42
TtesR137	1.434.598	pos	neg strand	TTX1636(+59)	
TtesR138	1.439.580	neg	pos strand	pos strand (OL)	
TtesR139	1.447.402	pos	TTX1651(-72)	TTX1652(+19)	ATTTAAAT-24
TtesR140	1.483.460	neg	TTX1693(-5)	TTX1692(+23)	TTTTAAAG-21
TtesR141	1.498.610	pos	TTX1716(+42)	TTX1717(+15)	GTTATAAAG-27
TtesR142	1.505.892	neg	pos strand (tRNA36)	TTX1726(-42)	GTTAAAAAG-12
TtesR143	1.506.158	neg	TTX1727(-18)	pos strand (OL; tRNA36)	
TtesR144	1.515.458	pos	neg strand	neg strand (OL)	CTTTAAAG-27
TtesR145	1.584.610	neg	pos strand	TTX1815(+58)	AATATATA-29
TtesR146	1.584.811	neg	pos strand	TTX1815(+268)	TTTTAAAG-43
TtesR147	1.622.122	pos	neg strand (tRNA41)	TTX1870(+68)	GTTTTTAAT-24
TtesR148	1.688.879	neg	pos strand	TTX1942(+47)	

TtesR149	1.764.548	pos	neg strand	TTX2022(+37)	ATTTATAAC-24
TtesR150	1.805.855	neg	pos strand	pos strand (OL)	CATATATC-8
TtesR151	1.811.852	neg	pos strand	pos strand (OL)	
TtesR152	1.823.468	pos	neg strand	TTX2094(+29)	TTTTAAAA-26
IHOsR01	55	neg	IHOsR201 (-5)		
IHOsR201	118	neg	Igni0001 (-25)	IHOsR01 (+5)	
IHOsR401	11.979	neg	Igni0013 (+34)	neg strand (OL)	
IHOsR05	23.662	pos			
IHOsR202	28.224	neg	Igni0028(+18)	Igni0027(-11)	
IHOsR06	40.116	pos	Igni0041(+34)	neg strand (IHOsR07)	
IHOsR07	40.383	neg	Igni0042(-119)	pos strand (IHOsR06)	ATTATTAAT-37
IHOsR203	54.007	pos	neg strand	Igni0057(+38)	
IHOsR204	62.476	pos	Igni0065(+3)		
IHOsR205	62.543	pos			
IHOsR09	62.672	pos		neg strand (OL; IHOsR11)	
IHOsR11	62.791	neg	tRNA-Ser (-15)	pos strand (OL; IHOsR09)	
IHOsR12	66.676	neg	pos strand	Igni0069(+36)	ATTAAAAAC-12
IHOsR206	69.724	neg	Igni0073(-140)	Igni0072(+3)	
IHOsR301	74.304	pos	neg strand	Igni0078(+37)	
IHOsR402	118.510	neg	Igni0131(-35)	pos strand (OL)	
IHOsR21	151.585	pos			
IHOsR302	152.454	pos	Igni0168(+17)	tRNA-Ser(+35)	
IHOsR22	159.973	neg	pos strand	pos strand	
IHOsR303	164.456	neg	pos strand	pos strand(OL)	
IHOsR500	170.182	pos	Igni0193 (-8)	Igni0194 (-60)	
IHOsR25	195.336	pos	Igni0220(+8)	Igni0221(+30)	
IHOsR26	202.285	pos	neg strand	neg strand (IHOsR404)	GTTAAAAAA-18
IHOsR404	202.410	neg	Igni0231(+5)	pos strand (IHOsR26)	
IHOsR27	205.142	neg	Igni0235(-7)	Igni0234(-3)	
IHOsR29	223.701	neg	Igni0259(+8)	Igni0258(+37)	
IHOsR31	234.954	pos	neg strand	neg strand (OL)	
IHOsR33	236.348	neg	pos strand	Igni0274(+45)	
IHOsR414	240.785	neg	Igni0279(+23)	Igni0278(+6)	
IHOsR34	246.915	neg	pos strand	pos strand	
IHOsR405	261.330	neg	pos strand (OL)	CRISPR (+262)	
IHOsR416	264.647	neg	Igni0310(+37)	pos strand (OL)	
IHOsR37	266.603	neg	Igni0313(+13)	Igni0312(-21)	
IHOsR39	279.453	pos	neg strand	Igni0323 (+39)	
IHOsR44	310.006	neg	Igni0353(compl.)	pos strand (OL)	
IHOsR304	318.319	pos	neg strand	neg strand (OL)	
IHOsR47	330.631	neg	pos strand	Igni0374(+42)	
IHOsR406	343.953	neg	Igni0391(+37)	pos strand (OL)	
IHOsR52	365.070	neg	Igni0412(+4)	Igni0411(-36)	

IHOsR53	366.866	neg	Igni0414(+9)	pos strand (OL)	
IHOsR54	374.701	neg	Igni0419 (compl OL with 5' end)	Igni0419 (compl OL with 5'end)	
IHOsR55	376.312	pos	Igni0421(+34)	neg strand (OL)	
IHOsR57	382.819	pos	neg strand (compl OL)	neg strand (compl OL)	
IHOsR58	384.028	neg	Igni0432(-21)	Igni0431(+33)	
IHOsR62	392.697	pos	neg strand (OL)	Igni0442(+11)	
IHOsR422	393.805	pos	Igni0442(+8)	Igni0043(+36)	
IHOsR423	405.608	pos	Igni0459(-63)	Igni0460(+54)	
IHOsR64	439.888	neg	Igni0492(+43)	Igni0491(+17)	
IHOsR425	483.959	neg	Igni0540(+2)	Igni0539(+48)	
IHOsR305	507.289	neg	Igni0565(-90)	Igni0564(+145)	
IHOsR73	509.045	pos	neg strand	neg strand (RNaseP)	
IHOsR76	526.477	pos	Igni0585(+54)	neg strand (OL)	
IHOsR77	532.112	pos	Igni0594(+8)	Igni0595(+27)	
IHOsR306	542.947	pos	neg strand	neg strand (OL)	
IHOsR80	555.504	pos	neg strand	neg strand (OL)	
IHOsR81	574.543	neg	IHOsR82 (+33)	pos strand	
IHOsR82	574.577	neg	pos strand	IHOsR81(-33)	TTTTAAAC-28
IHOsR56	576.811	neg	pos strand	Igni0639(+46)	
IHOsR83	582.489	pos	Igni0645(+36)	Igni0646(+68)	
IHOsR307	598.273	pos	Igni0666(-9)	Igni0667(+178)	
IHOsR86	617.384	pos	Igni0684(+34)	Igni0685(+88)	
IHOsR430	642.533	neg	Igni0713 (+54=compl)	pos strand (OL)	
IHOsR431	651.005	pos	neg strand (OL compl)	Igni0722(+46)	
IHOsR90	664.146	pos	Igni0737(-176)	neg strand (OL; IHOsR91)	
IHOsR91	664.240	neg	pos strand	pos strand (OL; IHOsR90)	
IHOsR408	704.236	neg	pos strand	pos strand (OL)	
IHOsR92	705.374	pos	Igni0797(+8)	Igni0798(+60)	
IHOsR308	720.626	pos	neg strand	Igni0813(+48)	
IHOsR93	722.261	pos	Igni0814(+8)	neg strand (OL)	
IHOsR94	723.587	pos	neg strand	Igni0816 (-55=compl)	
IHOsR70	726.164	neg	Igni0819(-3)	pos strand (OL)	
IHOsR95	728.035	neg	IHOsR96(+27)	Igni0821(+41)	
IHOsR96	728.061	neg	pos strand	IhosR95(-26)	
IHOsR437	739.729	pos	Igni0829(+9)	Igni0830 (+59)	
IHOsR173	742.938	neg	Igni0836(+34)	Igni0835(-28)	TTTTATAAA-20
IHOsR102	761.483	neg	pos strand (OL)	pos strand	
IHOsR439	766.788	neg	Igni0867(+9)	Igni0866(-59)	
IHOsR103	773.536	pos	Igni0872(+10)	Igni0873(-43)	
IHOsR104	783.685	pos	Igni0884(+8)	Igni0885(+28)	
IHOsR441	800.964	neg	pos strand	Igni0900 (-69)	
IHOsR409	814.261	neg	Igni0917(+34)	pos strand (OL)	

IHOsR109	823.952	pos	neg strand	Igni0929(+124)	
IHOsR444	824.774	pos	Igni0930(+9)	Igni0931(+34)	
IHOsR110	826.209	neg	Igni0932(+8)	pos strand (OL)	
IHOsR309	842.749	pos	Igni0950(+9)	5S rRNA (+31)	
IHOsR114	882.324	neg	Igni0991(+8)	pos strand (Leu tRNA)	
IHOsR410	890.706	neg	pos strand	pos strand (OL)	
IHOsR310	926.603	pos	neg strand	neg strand	
IHOsR449	945.621	pos	Igni1061(+10)	Igni1062(+11)	
IHOsR121	949.333	pos	neg strand	Thr tRNA (+41)	
IHOsR122	950.324	pos	Thr tRNA (-812)	Igni1068(+295)	
IHOsR411	962.502	neg	pos strand (OL; IHOsR124)	pos strand (OL)	
IHOsR124	962.500	pos	neg strand (OL; IHOsR411)	neg strand (OL)	
IHOsR451	969.070	pos	neg strand	neg strand (OL)	
IHOsR127	971.307	pos	Igni1089(+8)	neg strand (OL)	
IHOsR128	974.460	neg	Igni1092 (+51; compl)	Igni1091 (+21)	
IHOsR129	975.593	pos	Igni1093(+9)	Igni1094(+488)	
IHOsR412	986.285	neg	Igni1105(+16)	pos strand (OL)	
IHOsR130	993.353	neg	Igni1113(+48)	Igni1112(+90)	
IHOsR131	1.000.659	pos	Igni1119(-2)	IHOsR311 (-7)	
IHOsR311	1.000.717	pos	IHOsR131 (+7)	Leu tRNA (+32)	
IHOsR132	1.002.831	pos	neg strand	Igni1122(+32)	
IHOsR135	1.022.036	neg	pos strand (OL)	Igni1143(-24)	
IHOsR312	1.027.798	pos	neg strand	neg strand (OL)	
IHOsR413	1.036.429	neg	pos strand (OL)	pos strand (OL)	
IHOsR208	1.040.939	neg	Igni1169(+9)	Igni1168(+30)	
IHOsR313	1.066.890	neg	Igni1197(-9)	Igni1196(+101)	
IHOsR415	1.085.370	pos	Igni1218(+8)	Igni1219(+31)	
IHOsR142	1.090.102	pos	neg strand	Igni1224(+69)	
IHOsR144	1.094.106	pos	neg strand	neg strand (OL)	
IHOsR207	1.106.046	neg	Igni1246(+9)	pos strand (OL)	
IHOsR314	1.115.629	pos	neg strand	Igni1255(+44)	
IHOsR147	1.126.304	pos	Igni1264	Igni1264	
IHOsR148	1.126.886	pos	Igni1264(+40)	neg strand (OL)	
IHOsR149	1.130.035	pos	Igni1268(-151)	neg strand	
IHOsR417	1.139.715	pos	Igni1282(+2)	Igni1283(+27)	
IHOsR152	1.160.287	neg	Igni1305(-14)	pos strand (OL)	
IHOsR154	1.174.567	pos	Igni1320(+1)	Igni1321(+35)	
IHOsR155	1.192.320	neg	Igni1339(+22)	Val tRNA (+38)	
IHOsR157	1.198.413	neg	Igni1347(+2)	Igni1346(+249)	
IHOsR418	1.206.880	neg	Igni1352(-12)	pos strand (OL)	
IHOsR160	1.219.644	neg	pos strand	Leu tRNA (+26)	
IHOsR165	1.250.593	neg	pos strand (Ile tRNA)	Igni1389(-42)	CTTATAAAT-30
IHOsR472	1.258.549	pos	Igni1398(-5)	neg strand (OL)	
IHOsR473	1.266.795	neg	leu tRNA (-22)	Igni1405(+28)	

IHOsR166	1.273.845	pos	Igni1415(+59)	Igni1416(+29)	
IHOsR474	1.283.922	neg	Igni1429(+8)	Igni1428(-35)	
IHOsR169	1.294.955	neg	Igni1440(+11)	pos strand (OL)	
NeqsR01	45.776	pos			CTTTTTAAT-45
NeqsR02	46.979	neg			
NeqsR03	54.085	neg			
NeqsR04	90.954	neg			
NeqsR05	113.628	neg			
NeqsR06	145.004	neg			
NeqsR07	146.019	neg			CTTTAAAT -45
NeqsR08	164.168	neg			
NeqsR09	248.239	neg			ATTTAAAA-23/ TTTATTAAG-12/ ATTTATAAA-46
NeqsR10	282.973	neg			CTTTAAAT-22/ATTTAAAG-8
NeqsR11	297.634	pos			CATATATC-40
NeqsR12	316.045	neg			TTTTAAAT-27/ TTTATTAAT-10
NeqsR13	318.915	pos			ATTTATAAA-29
NeqsR14	323.094	neg			
NeqsR15	325.371	pos			ATTTATAAT-29
NeqsR16	328.949	neg			ATAAAAAA-17
NeqsR17	334.552	pos			TATATATG-13
NeqsR18	337.505	neg			CTTTATAAA-40
NeqsR19	359.412	pos			
NeqsR20	362.755	pos			
NeqsR21	371.917	neg			CTTTAAAA-27
NeqsR22	375.406	neg			
NeqsR23	382.458	pos			TTTTAAAA-15/ TTTTATAAA-30
NeqsR24	384.218	pos			
NeqsR25	401.223	pos			
NeqsR26	403.464	pos			AATATATT-21

Erklärung des Eigenanteils

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

Omer Alkhnbashi (AG von Rolf Backofen, Albert-Ludwigs-Universität Freiburg)
Hat als Kollaborationspartner die bioinformatischen Analysen zur Identifizierung von möglichen C/D box sRNA Promotoren und Zirkularisierungssignalen durchgeführt.

Lauren Lui (Labor von Todd Lowe, University of California) und **Prof. Patrick Dennis** (Janelia Research Campus, Howard Hughes Medical Institute)
Haben als Kollaborationspartner die Vorhersagen der C/D box sRNA Modifikationsziele in rRNAs erstellt.

Danksagung

Mein Dank gilt vor allem Herrn Dr. Lennart Randau für die Möglichkeit meine Doktorarbeit in seiner Arbeitsgruppe durchführen zu dürfen, die herausragende Betreuung und das ständige Interesse am Verlauf der Arbeit.

Ein weiterer Dank gilt Herrn Prof. Dr. Torsten Waldminghaus für die Übernahme der Zweitkorrektur sowie Herrn Prof. Dr. Roland K. Hartmann und Herrn Prof. Dr. Martin Thanbichler für die Zusage meiner Prüfungskommission anzugehören.

Für die Unterstützung während meiner Doktorarbeit möchte ich mich auch bei meinem Thesis Advisory Committee, bestehend aus Prof. Dr. Torsten Waldminghaus, Prof. Dr. Roland K. Hartmann, Prof. Dr. Peter Graumann und Prof. Dr. Sonja-Verena Albers bedanken.

Außerdem danke ich dem kompletten Laborteam für die Unterstützung und das angenehme Arbeitsklima, während meiner Doktorarbeit.

Zuletzt möchte ich mich bei meiner Familie und meinen Freunden bedanken, für die häusliche Unterstützung, die aufmunternden Worte und die schöne Zeit außerhalb des Labors.