RNA guided Nucleotide Modification of Ribosomal and Non-ribosomal RNAs in Archaea

by

Sonia Madlen Ziesche

Haupt-Diplom Biology, Darmstadt Institute of Technology, 1996

A THESIS SUBMITTED IN PARTIAL FULFILMENT OF
THE REQUIREMENTS FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

in

THE FACULTY OF GRADUATE STUDIES

(Department of Biochemistry and Molecular Biology)

We accept this thesis as conforming
to the required standard

THE UNIVERSITY OF BRITISH COLUMBIA

July 2004

Abstract

Archaea use ribonucleoprotein (RNP) machines similar to those found in the eukaryotic nucleolus to methylate ribose residues in nascent ribosomal RNA. The archaeal complex required for this 2'-O-ribose-methylation consists of the C/D box sRNA guide and three proteins, the core RNA binding aL7a protein, the aNop56 protein and the methyltransferase fibrillarin protein. These RNP machines were reconstituted *in vitro* from purified recombinant components, and were shown to have methylation activity when provided with a simple target oligonucleotide, complementary to the sRNA guide sequence (Omer *et al.*, 2002). The accuracy in directing methylation to the correct nucleotide in the target of the *in vitro* reconstituted C/D box RNP was shown.

To obtain a better understanding of the versatility and specificity of this reaction, the activity of reconstituted particles on more complex target substrates (including 5S rRNA, $tRNA^{Gln}$ and "double guide" oligonucleotides that exhibit either direct or reverse complementarity to both the D' and D box guides) has been examined. The natural 5S rRNA and $tRNA^{Gln}$ substrates were efficiently methylated *in vitro,* providing that the complementarity between guide and target was approximately ten base pairs in length, and lacked mismatches. Maximal activity of double guide sRNAs required that both methylation sites be present in *cis* on the target RNA. These experiments defined the minimum number of components and conditions required to achieve *in vitro*, RNA guide directed 2'-O-ribose methylation of ribosomal and non-ribosomal target RNA.

Table of Contents

List of Figures and Figure Legends

List of Abbreviations

ATCC American Type Culture Collection

BSA Bovine serum albumin

cpm counts per minute

DEPC diethylpyrocarbonate

dH_2O deionized water

DTT dithiothreitol

DSM Deutsche Sammlung von Mikroorganismen

EDTA ethylenediamine tetraacetic acid

EMSA Electro mobility shift assay

h hour

IPTG isopropyl-β-D-thiogalactopyranoside

kD kilodalton

min minute

NMR nuclear magnetic resonance

nt nucleotide

MTase Methyltransferase

MALDI matrix assisted laser desorption ionisation

OD optical density

ORF open reading frame

PAGE Polyacrylamide gel electrophoresis

PCR Polymerase Chain Reaction

Ψ Pseudouridine

rDNA ribosomal DNA

rRNA ribosomal RNA

rpm revolution per minute

S Svedberg unit of sedimentation coefficient

Sac *Sulfolobus acidocaldarius*

SDS sodium dodecyl sulphate

snoRNA small nucleolar RNA

snoRNPs small nucleolar ribonucleoprotein particles

sp species

sR small RNA

Sso *Sulfolobus solfataricus*

Taq *Thermus aquaticus*

TBE Tris•Borate EDTA solution

Tm melting temperature

Tris Tris(hydroxymethyl)aminoethane

tRNA transfer RNA

U unit

UV ultraviolet

V Volt

v/v volume per volume

W Watt

w/v weight per volume

Acknowledgements

I would like to thank Patrick P. Dennis for giving me the opportunity to study at UBC and for exposing me to a variety of different projects.

I am especially grateful to George Mackie and the other members of my advisory committee, Natalie Strynadka and Michel Roberge, for giving me guidance and support during my PhD project.

I would also like to thank Anthony Russell for sharing his knowledge about protein purification with me and generally for the good times and fruitful scientific discussions we had, especially during our coffee breaks. I thank all the other members of the former Dennis lab for all their support and advice. In particular, I am grateful to Arina Omer and Maria Zago for keeping me company in the secluded basement of the Biochemistry building as well as for valuable discussions, technical advice, and assistance.

I would also like to thank Carolyn Brown from Medical Genetics and all the great people of the Brown lab, who accepted a biochemist so readily into their midst and supported me in the end phase of my thesis.

Very special and warm thanks go to Holger Hoos and Jehannine Austin for all their loving support during challenging times.

I would also like to thank Lindt/Spruengli and Urban Fare, as the providers of fine European Chocolate that kept me going during my writing-up phase.

This thesis is dedicated to my parents Monika and Bernhard Dieter Ziesche, who always supported me in all my endeavours - even if they seemed to occur at the end of the world.

"Die Wissenschaft fängt eigentlich erst da an,

interessant zu werden, wo sie aufhört."

"Actually, science only starts

to get interesting where it ends."

(Justus von Liebig)

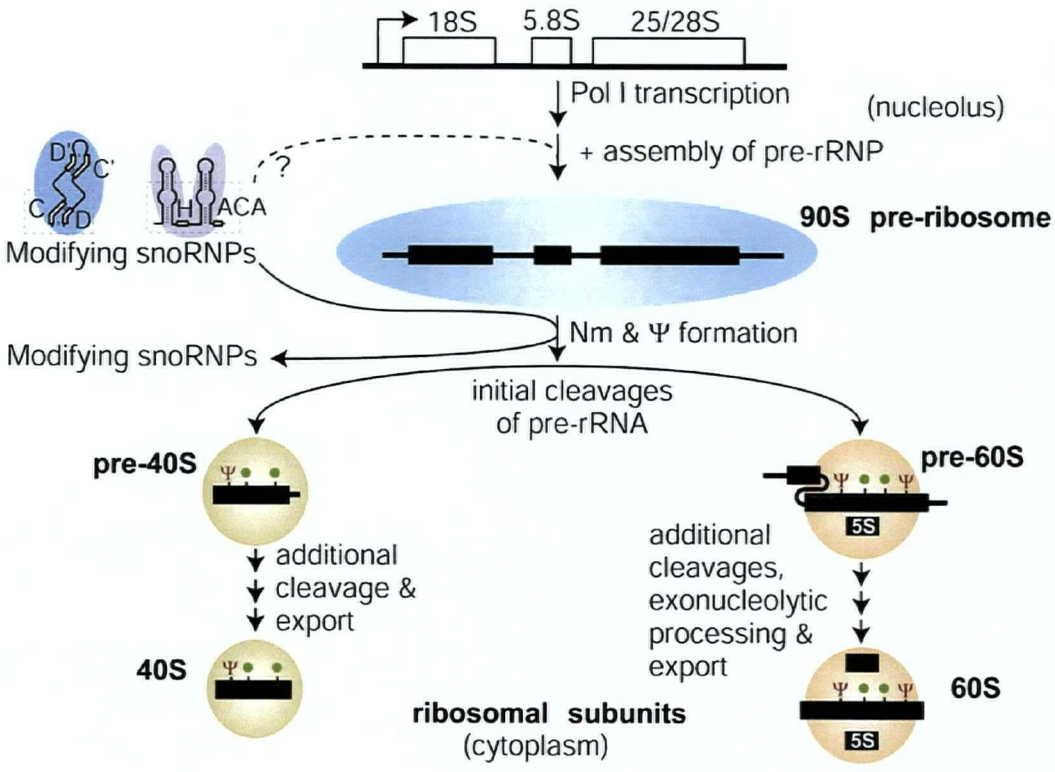
1 Introduction

1.1 Biogenesis of eukaryotic ribosomes

The biogenesis of ribosomes in the eukaryotic nucleolus is a remarkably complex process, involving an intricate series of processing steps of the ribosomal RNA (rRNA) precursor. First, the 18S, 5.8S and 28S rRNA are transcribed as one contiguous unit by RNA polymerase I (Figure 1-1). The fourth rRNA (5S) is transcribed independently by RNA polymerase III, presumably in the nucleoplasm (Pombo *et al.*, 1999). Maturation of the rRNA primary transcript requires covalent modification of the small subunit (SSU) and large subunit (LSU) rRNA moieties and a series of endo- and exonucleolytic cleavages that liberate the mature rRNAs. There are two prevalent types of modifications: 2'-O-ribose methylation and pseudouridylation, each involving more than 100 sites per eukaryotic ribosome (Lane *et al.*, 1995).

1.1.1 Posttranscriptional modification of ribosomal RNA

Posttranscriptional ribosomal RNA modification is common in all branches of the tree of life - Eukaryotes, Eubacteria and Archaea. There are three basic types of modification found in rRNA: base methylation, ribose methylation, and pseudouridylation.



Adapted from Decatur WA and Fournier MJ, J Biol Chem. 2003 Jan 10;278(2):695-8.

Figure 1-1 Modification of rRNA by snoRNPs

An overview of eukaryotic ribosome biogenesis is illustrated. The scheme shown is generic. Modifying snoRNPs act on pre-rRNA in the nucleolus during or after transcription.

1.1.1.1.1 Base methylation

Base methylation is the best conserved modification, both in total number and position among all species, with bacteria containing slightly more than the 10 commonly found in eukaryotes. Base methylation takes place late in ribosome maturation, and occurs only in highly conserved rRNA sequences. Base methylation within SSU rRNA in

prokaryotes is not essential (Krzyzosiak *et al.*, 1987), but it is thought to improve protein translation efficiency (Raue *et al.*, 1988).

1.1.1.1.2 Ribose methylation

Ribose methylation takes place on the nascent rRNA transcript. Because most ribose methylation takes place early in rRNA processing, it is hypothesized to be important for rRNA folding or association with chaperone proteins that may aid in folding. No single site of ribose methylation has been found to be essential (Weinstein and Steitz, 1999), although global rRNA demethylation caused by mutation in the Nop1p protein, the yeast homologue of fibrillarin, (a protein essential for methylation), severely impairs growth (Tollervey *et al.*, 1993). In hyperthermophiles, ribose methylation may also be important in thermostability of rRNA and other structural RNA molecules (Noon *et al.*, 1998). In Eukaryotes and Archaea 2'-O-ribose methylation is mediated by RNA-protein complexes (RNPs). The RNPs contain a small nucleolar RNA (snoRNA) and several associated proteins. The snoRNA provides the guide function, and snoRNP proteins catalyze the modification reaction. In contrast to Eukaryotes and Archaea, the rRNA of a typical Eubacterium possesses only a handful of 2'-O-ribose methylations; each of these modifications seems to be mediated by site-specific enzymes that do not contain an RNA co-factor.

1.1.1.1.3 Pseudouridylation

The exact function of pseudouridines on a molecular level is not known, but through its unique ability to coordinate a structural water molecule via its free N1-H, pseudouridine exerts a subtle but significant "rigidifying" influence on the nearby sugar-

phosphate backbone and also enhances base stacking. Several studies on eukaryotes and prokaryotes show that deletion of pseudouridine-synthesising genes has a negative impact on cell growth and protein synthesis rate (Lecointe *et al.*, 1998; Raychaudhuri *et al.*, 1998). In Eubacteria each pseudouridine modification appears to be modified by a site-specific pseudouridine synthase without any RNA cofactor involved.

Conversely, in Eukaryotes pseudouridylation is guided by H/ACA box snoRNAs, assembled in complex ribonucleoprotein particles. H/ACA box snoRNAs have recently been discovered in Archaea, suggesting that they might use a similar mechanism to Eukaryotes to guide pseudouridylation in rRNA.

1.2 The two families of rRNA modification guides

Except for the RNA component of RNase MRP, all snoRNAs to date fall into two major classes, the C/D box and the H/ACA box snoRNAs, based on the presence of short consensus sequence motifs (Balakin *et al.*, 1996).

Most members of the two snoRNA families guide 2'-O-ribose methylation and pseudouridylation, respectively. However a few members are involved in pre-rRNA cleavage instead (Tollervey, 1996).

In eukaryotes, the C/D box snoRNPs contain four common proteins: Fibrillarin, which catalyzes the methyltransfer reaction; Nop56p and Nop58p, two proteins that have substantial sequence similarity with each other; and 15.5 kD/Snu13p, which is found in both the C/D box snoRNPs and the U4 spliceosomal snRNP (Watkins *et al.*, 2000).

1.2.1 C/D box snoRNAs

C/D box snoRNAs contain one or two pairs of small distinguishing sequence elements called box C (RUGAUGA, where R can be either purine) and D (CUGA) and a less conserved copy of the C motif, termed box C', and an additional box D motif, termed box D'(Bachellerie *et al.*, 2002; Cavaille and Bachellerie, 1998; Kiss, 2001; Tycowski *et al.*, 1996).

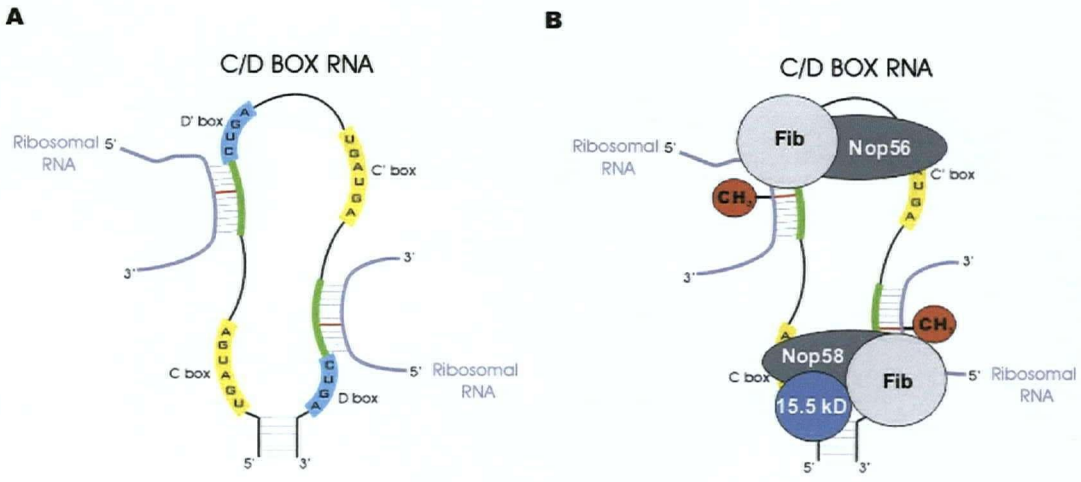


Figure 1-2 Structural features of eukaryotic C/D box snoRNA

(A) C/D box snoRNAs involved in ribose methylation contain one or two long 10-21 bp stretches of exact complementarity to ribosomal RNA, and four conserved box features: C, C', D and D' boxes. (B) The C and D box sequence motifs are required for snoRNA nucleolar localization, accumulation, and association with the ribonucleoprotein particle complexes (RNPs). (B) The proteins of the C/D box snoRNPs are arranged asymmetrically. aL7a binds to the terminal core motif, followed by Nop58 and fibrillarin(Fib). Nop56 and fibrillarin are binding to the C'/D' motif. The position of 2'-O-methylation of rRNA is within the helix formed by the complementary guide sequence of the snoRNA, and precisely 5 nt upstream of D or D' box.

C/D boxes occur near the 5' and 3'ends of the RNA and C'/D' boxes are located internally (Figure 1-2A). The methylation guide sequence is located up-stream of the box D or D' element (Bachellerie *et al.*, 1995; Kiss, 2001). The C and D box motifs are generally brought together in a typical 5'-3' terminal stem-box structure, involving the 4-5nt at both termini. This is critical for snoRNA biogenesis and nucleolar localization (Bachellerie, 1998).

The C and D boxes form a secondary structure motif, designated as the k-turn (Klein *et al.*, 2001; Vidovic *et al.*, 2000; Watkins *et al.*, 2000).

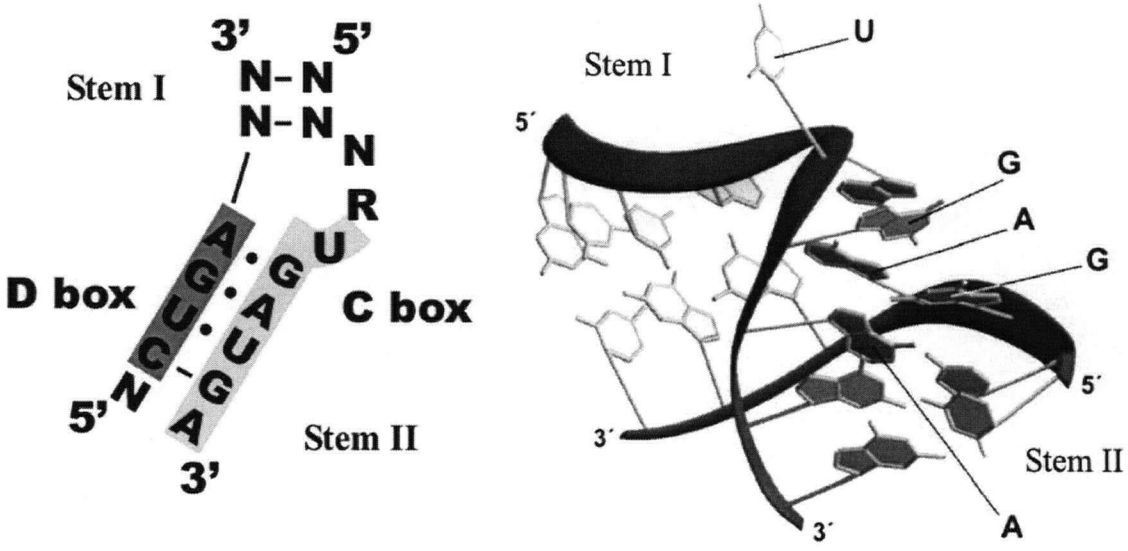


Figure 1-3 Consensus k-turn motif of C/D box snoRNAs

The consensus k-turn motif for canonical C/D box sRNAs is shown at the left. At the right the C/D box core motif is folded into a kink- or k-turn.

The k-turn motif is highly conserved within C/D box snoRNAs and is formed between C and D boxes via interaction of short inverted repeats present at the 5' and 3' ends of snoRNAs (Watkins *et al.*, 2000). Characteristic for this motif are two helical stems separated by an internal three base loop. The internal loop is always asymmetrical and purine-rich and contains three unpaired nucleotides at one side with a terminal U base. The second helix exhibits two sheared G-A base pairs usually followed by a U-U base pair and a regular Watson–Crick base pair (Klein *et al.*, 2002) (Figure 1-3). k-turn motifs have not only been found in C/D box snoRNAs, but in various eukaryal and prokaryal RNAs including ribosomal RNAs (Klein *et al.*, 2002; Winkler *et al.*, 2001) and U3 and U8, C/D box snoRNAs that do function as methylation guides.

Immediate upstream from D and/or D' box, snoRNAs exhibit antisense elements - sequences 10-21 nt in length that are complementary to rRNA containing the site of 2' O-ribose methylation (Cavaille *et al.*, 1996; Kiss-Laszlo *et al.*, 1996; Nicoloso *et al.*, 1996) (Figure 1-2 A). Methylation is directed to the rRNA nucleotide that participates in a Watson-Crick base pair five nucleotides up-stream from the start of the D or D' box; this is the N plus five rule (Ni *et al.*, 1997b; Schimmang *et al.*, 1989; Wu *et al.*, 1998).

In eukaryotes, expression of an artificial C/D box snoRNA carrying an appropriate antisense element is sufficient to target a novel ribose methylation on the predicted pre-rRNA nucleotide. This shows that the antisense element associated with D (or D') box is the sole determinant of the site of methylation (Cavaille *et al.*, 1996).

1.2.2 Proteins associated with C/D box snoRNAs

Both types of snoRNAs (C/D and H/ACA) function as small ribonucleoprotein particles (snoRNPs).

The snoRNA provides the guide function and an integral snoRNP protein catalyzes the modification reaction. The total number of proteins in a particle is not known, nor is it known if all modifying snoRNPs are identical except for the RNA component. However methylating snoRNPs contain four common, evolutionarily conserved, core proteins, fibrillarin (Nop1p), Nop56p, Nop58p and 15.5 kD protein (Snu13p).

Eukarya	**Archaea**
Fibrillarin (Nop1p)	Fibrillarin (aFib)
Nop56p	aNop56 (aNop5)
Nop58p	
15.5 kD (Snu13p)	aL7a

Table 1-1 Core Proteins associated with C/D Box snoRNAs

Homologues of snoRNP core proteins have been identified in Archaea. Archaea contain a single orthologue of the eukaryotic Nop56p and Nop58p proteins, termed aNop56 (or aNop5).

1.2.2.1 Fibrillarin

Fibrillarin is generally accepted to be the 2'-O-methyltransferase. It exhibits

amino acid sequence motifs characteristic of S-Adenosyl Methionine (SAM)-dependent methyltransferases (Niewmierzycka and Clarke, 1999). In support of this line of reasoning, point mutations in the methylase-like element have been shown to inhibit globally ribose methylation in yeast cells (Tollervey *et al.*, 1993). In addition, the crystal structure of fibrillarin from *Methanococcus jannaschii,* an archaeal orthologue that contains the methylase signature elements, showed that most of the protein adopts a three dimensional structure similar to many known SAM-dependent methylases (Wang *et al.*, 2000).

1.2.2.2 Nop56p and Nop58p

Nop56p and Nop58p are highly related to each other by sequence (Filippini *et al.*, 2000; Newman *et al.*, 2000). Both proteins were found to be essential and localized to the nucleolus. The two proteins share 45% sequence identity and contain a KKD/E signature sequence that is also present in other nucleolar proteins (Gautier *et al.*, 1997). Nop56p and Nop58p also display a significant sequence similarity with a newly identified 61kD protein. The 61 kD protein is associated with the activation of the splicesome U4/U6-specific proteins; Protein 61K directly contacts the 5' portion of U4 snRNA via a novel RNA-binding domain (Nottrott *et al.*, 2002).

Recent work has revealed the asymmetric distribution of the C/D snoRNP core proteins upon the terminal C/D box core and internal C'/D' motifs (Cahill *et al.*, 2002; Szewczak *et al.*, 2002). *In vivo* crosslinking showed that Nop58p and Nop56p are differentially bound to the C/D and C'/D' box motifs, respectively (Figure 1-2 B).

1.2.2.3 15.5 kD/ Snu13p

Most recently, the fourth core protein has been identified. Designated the 15.5 kD protein in humans (Watkins *et al.*, 2000), Snu13p protein, the yeast homolog of mammalian 15.5 kD protein, binds specifically to the C/D structural motif, the k-turn. The eukaryotic 15.5 kD/Snu13p protein has two functions and it is also part of the U4/U6.U5 tri-snRNP complex (Nottrott *et al.*, 1999), where it binds to the same structural motif in U4 snRNA. A previous crystal structure of the human 15.5 kD/Snu13p protein in complex with U4 snRNA showed that it interacts almost exclusively with the bulged loop situated between two short stems (Watkins *et al.*, 2000); similar binding is expected between Snu13p and the C/D box snoRNAs.These findings suggest that the eukaryotic spliceosomal and modification machinery are related and share homologous components (Nottrott *et al.*, 1999).

1.2.2.4 Transiently associated proteins

Other proteins have been found to transiently interact with the snoRNPs.

1.2.2.4.1 p50/p55

Recently, a related pair of evolutionary conserved proteins, p50 (Rvb2) and p55 (Rvb1), were observed to bind to model C/D box snoRNA *in vitro.* Both proteins have DNA helicase activity and are linked to replication and transcription. It is suggested that they function at an early stage of snoRNP biogenesis (Filippini *et al.*, 2000; Newman *et al.*, 2000) and are involved in the coordination of snoRNA processing and snoRNP assembly with replication and/or transcriptional events in the nucleus. Studies of the

Saccharomyces cerevisiae orthologue of p50 (Rvb2, Tih2p, and other names) showed, that it is required for production of C/D snoRNAs *in vivo* and, surprisingly, H/ACA snoRNAs as well. Homologs for both snoRNA-associated protein pairs occur in Archaea, strengthening the hypothesis that the C/D box RNA elements and their interacting proteins are of ancient evolutionary origin.

1.2.2.4.2 Srp40p (Nopp140)

Using co-immunoprecipitation experiments, it has been shown that the nucleolar and coiled-body protein Nopp140 (Srp40p) interacts with both classes - C/D box and H/ACA box snoRNPs. It has been proposed that Nopp140 functions as a chaperone of snoRNPs in yeast and vertebrate cells (Yang *et al.*, 2000).

1.2.2.4.3 Survival of Motor Neuron (SMN) protein

While it has been shown that the SMN protein is involved in spliceosome biogenesis and pre-mRNA splicing, there is increasing evidence indicating that SMN may also perform important functions in the nucleolus. Polyclonal anti-SMN antibodies show a striking co-localization with fibrillarin, in both nucleoli and Cajal body (CBs). These studies raise the possibility that SMN may serve a function in rRNA maturation/ribosome synthesis similar to its role in spliceosome biogenesis (Verheggen *et al.*, 2001; Wehner, 2002).

1.2.3 H/ACA box snoRNAs

The H/ACA snoRNAs share a common secondary structure consisting of two large hairpin domains linked by a hinge, and followed by a short tail (Figure 1-4).

Conserved motifs, termed H (ANANNA, where N stands for any nucleotide) and ACA (a trinucleotide always found three nucleotides away from the 3' end) are located in the hinge and tail respectively (Balakin *et al.*, 1996; Ganot *et al.*, 1997a; Ganot *et al.*, 1997b).

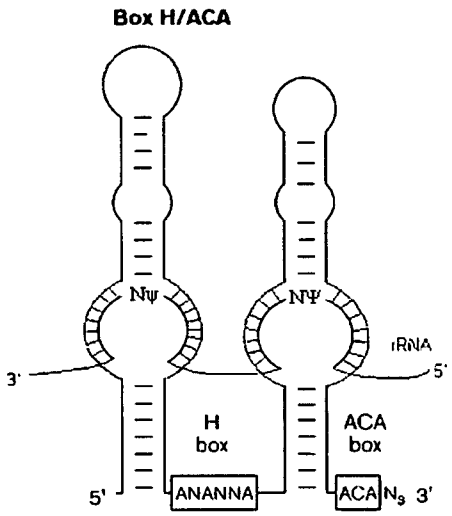


Figure 1-4 H/ACA box snoRNA

Structural features of the eukaryotic H/ACA snoRNAs are shown. The H/ACA box snoRNAs contain one or two regions of hyphenated complementarity to rRNA that lie within the bulge regions of the 5' or 3' helices; base pairing to rRNA positions the uridine nucleotide to be modified in a pocket between the hyphenated regions of rRNA-snoRNA complementarity.

Like the C/D elements, the H and ACA box and neighbouring duplexes are required for processing of snoRNA precursors, protein binding, and localization (Filipowicz and Pogacic, 2002; Kiss, 2002). Substrate targeting involves base pairing through two short guide sequences in the internal loop of one or both of the two large hairpin domains (Ganot *et al.*, 1997a; Ofengand, 1998) and a distance measurement of ~

14-15 nucleotides from the H or ACA box (Ganot *et al.*, 1997a; Ni *et al.*, 1997b). In Eukarya, in some cases, a k-turn motif can be folded manually. This is, however, energetically less stable than alternate structures not exhibiting the k-turn (Rozhdestvensky *et al.*, 2003).

1.2.4 Proteins associated with H/ACA snoRNAs

Proteins common to H/ACA snoRNPs include evolutionarily conserved protein Cbf5p (dyskerin), Gar1p, Nhp2p and Nop10p, all of which are essential for the pseudouridylation reaction.

Eukarya	**Archaea**
Cbf5p (dyskerin) (=pseudouridylase?)	Cbf5 homologue
Gar1p	Gar1p homologue
Nhp2p	aL7a
Nop10p	Nop10p homologue

Table 1-2 Core Proteins associated with H/ACA Box snoRNAs

All four orthologues of the eukaryotic pseudouridylation (Ψ) snoRNP proteins have been identified in Archaea (Watanabe and Gray, 2000).

1.2.4.1 Cbf5p (dyskerin)

Cbf5p (centromere-binding factor 5) is likely to correspond to the catalytic component of H/ACA snoRNA guided modification. It appears to be closely related to

TruB/Pus4p, which catalyzes the Ψ55 pseudouridine formation in the T loop of virtually all tRNAs (Samuelsson and Olsson, 1990). Cbf5p and *Escherichia coli* TruBp, a known Ψ synthase, share the "KP" and "XLD" conserved sequence motifs found in the catalytic domains of three distinct families of known and putative Ψ synthases. The crystal structure of TruB bound to RNA suggests that both enzymes recognize RNA in a similar manner (Hoang and Ferre-D'Amare, 2001).

Genes for archaeal Cbf5p homologues have also been identified (Bult *et al.*, 1996; Klenk *et al.*, 1997; Smith *et al.*, 1997) but their function is currently unclear. Putative Cbf5p homologues from archaeal kingdom Crenarchaeota are composed of two fragmented peptides encoded by two overlapping or adjacent open reading frames (ORFs) (Watanabe *et al.*, 2002).

1.2.4.2 Gar1p

Gar1p is a 25-kDa nucleolar protein that is essential for yeast cell growth and has been shown to be associated with every H/ACA snoRNA (Balakin *et al.*, 1996; Ganot *et al.*, 1997b). Although the Gar1p protein is required for the accumulation of mature 18S rRNA (Girard *et al.*, 1992; Girard *et al.*, 1994), its nucleolar function as a component of numerous non-essential snoRNP particles is largely unknown. Gar1p plays a specific and fundamental role in global pseudouridylation of yeast rRNAs (Bousquet-Antonelli *et al.*, 1997). Gar1p, fibrillarin, and several nucleolar proteins involved in pre-rRNA processing, all contain similar domains. These domains are rich in glycine and arginine residues and are referred to as the GAR domain (Girard *et al.*, 1992). Curiously, Gar1p has two GAR domains, one at each end of the protein, that together represent almost 40% of the total

protein. However, the two GAR domains of Gar1p are not necessary for cell growth nor are they required to target a reporter protein to the nucleolus (Girard *et al.*, 1994).

1.2.4.3 Nhp2p

In *S. cerevisiae*, Snu13p and Nhp2p exhibit significant similarity to ribosomal protein L30, raising the possibility of a further link between ribosome and snoRNP production. Snu13p and Nhp2p share a common archaeal homologue, ribosomal protein aL7a, pointing to a common evolutionary origin for the two families of modification guides.

1.2.4.4 Nop10p

Nop10p is a small nucleolar protein that is specifically associated with H/ACA snoRNAs. It is essential for normal 18S rRNA production and rRNA pseudouridylation by the ribonucleoprotein particles containing H/ACA snoRNAs (H/ACA snoRNPs). Nop10p is probably necessary for the stability of these RNPs (Henras *et al.*, 1998).

1.2.4.5 Other proteins associated with H/ACA box RNAs

1.2.4.5.1 Naf1p

Recently, the essential *Saccharomyces cerevisiae* Naf1p protein was found to interact in a two-hybrid assay with two core protein components of mature H/ACA snoRNPs, Cbf5p and Nhp2p (Ito *et al.*, 2001). Naf1p is localized to the nucleus and is not a stable component of the H/ACA snoRNPs, but it is required for the accumulation of all H/ACA box snoRNAs. It is suggested that the Naf1p is involved in H/ACA snoRNP

biogenesis, trafficking, and/or function. (Dez *et al.*, 2002) and is promoting snoRNP assembly (Fatica *et al.*, 2002).

1.3 Other targets for eukaryotic snoRNA guides

Intriguingly, an increasing number of C/D box snoRNAs devoid of rRNA antisense elements have been recently identified (Cavaille *et al.*, 2000; Cavaille *et al.*, 2001; Huttenhofer *et al.*, 2001; Jady and Kiss, 2000; Massenet, 1998) and indicate that snoRNPs are involved in processes other than the modification of rRNA.

1.3.1 Modification guides for spliceosomal snRNAs

Biogenesis of functional spliceosomal small nuclear RNAs (snRNAs) includes the post-transcriptional covalent modification of numerous internal nucleotides. Mammalian U1, U2, U4, U5 and U6 snRNA contain a very substantial number of 2'-O-ribose methylations and pseudouridylations, amounting collectively to 30 and 24, respectively (Massenet, 1998). The modified nucleotides are confined to snRNA regions involved in formation of RNA–RNA or RNA–protein interactions that are crucial for spliceosome function, strongly suggesting that they have beneficial effects on the efficiency and/or fidelity of pre-mRNA splicing. Modification of snRNAs occurs in nucleoplasmic Cajal bodies, where modification guide RNAs accumulate. Thus they were named small Cajal body-specific RNAs (scaRNAs). The snoRNAs that guide modification of U6, are structurally and functionally indistinguishable from snoRNAs directing modification of the 18S, 5.8S and 28S rRNAs (Filipowicz and Pogacic, 2002; Kiss, 2001; Terns and

Terns, 2002). In fact, one 2'-O-methylation guide snoRNA was found to function in both U6 and 28S rRNA modification (Tycowski *et al.*, 1998).

A chimeric C/D-H/ACA snoRNA (a 'Siamese' snoRNA), U85, has also been discovered in humans and *Drosophila* (Jady and Kiss, 2001). This unique snoRNA possesses the box elements of both classes of snoRNAs and associates with both fibrillarin and Garlp. *In vitro* and *in vivo* pseudouridylation and 2'-O-methylation experiments provide evidence that this snoRNA guides 2'-O-methylation and pseudouridylation at two adjacent positions in U5 snRNA. Current observation suggests that other chimeric C/D-H/ACA snoRNA guides for the 2'-O-methylation and pseudouridylation of vertebrate snRNAs might exist. Several putative 2'-*O*-methylation and pseudouridylation guide snoRNAs with significant sequence complementarities to the U1, U2, U4 and U5 snRNAs have been identified (Huttenhofer *et al.*, 2001). The new snoRNAs possess the potential to select known 2'-*O*-ribose methylated nucleotides and pseudouridines in these spliceosomal snRNAs.

1.3.2 Telomerase RNA

Recently, it was discovered that vertebrate telomerase RNA contains a typical H/ACA domain and that snoRNPs and telomerase share evolutionary conserved proteins (Dez *et al.*, 2001; Pogacic *et al.*, 2000). No presumptive RNA target for pseudouridylation has been identified for the H/ACA domain of vertebrate telomerase RNA, but it has been shown that the human telomerase RNA H/ACA domain is essential *in vivo* for accumulation, 3' end processing, and telomerase activity (Mitchell *et al.*, 1999).

1.3.3 Orphan guide snoRNA

An increasing number of ubiquitously expressed snoRNA belonging to the C/D box or H/ACA box RNA family in mammals are devoid of antisense elements to rRNA or snRNA have been reported (Huttenhofer *et al.*, 2001; Jady and Kiss, 2000). These snoRNAs are termed orphan guide snoRNAs.

A recent expressed sequence tag (EST) screen for mouse small non-mRNAs, identified 15 C/D and 11 H/ACA orphan snoRNAs, the biological role of which remains elusive (Huttenhofer *et al.*, 2001). Many RNA species that are not directly related to ribosome biogenesis (including several mRNAs) have been reported to localize transiently to the nucleolus. The 5' terminal processing of some pre-tRNAs in yeast by RNase P also takes place in the nucleolus (Pederson, 1998) (Bertrand *et al.*, 1998). This suggests that a host of cellular RNAs might be targeted by these intriguing orphan modification guides. A search for potential target sites in the three stable non-coding RNAs trafficking through the nucleolus (telomerase RNA, RNase P and SRP RNA) was negative for all the orphan snoRNAs reported so far. Alternatively, some orphan snoRNAs could be exclusively involved in pre-rRNA cleavage instead of modifications, similar to C/D snoRNA U3, U8 and U22 (Hughes, 1996; Sharma and Tollervey, 1999; Tollervey, 1996) and yeast H/ACA snR10 and snR30 (Morrissey and Tollervey, 1993; Tollervey, 1987). Remarkably, the vertebrate homologue of yeast snR30, U17, also involved in pre-rRNA cleavage, has a 3' domain exhibiting hallmarks of a pseudouridylation guide for a still unidentified target, suggesting a dual function of this snoRNA. Dual functions have previously been demonstrated for a few modification

guides targeting rRNA, such as H/ACA snoRNA E2 and E3 and C/D snoRNA U14, which all have an additional role in pre-rRNA cleavage (Enright *et al.*, 1996; Liang and Fournier, 1997).

1.3.4 Brain specific snoRNAs, genomic imprinting and mRNA targets

Recently, a collection of brain-specific snoRNAs whose expression is paternally imprinted has been characterized (Cavaille *et al.*, 2000). In a search for small RNAs expressed in mouse brain, three novel C/D box snoRNAs and one H/ACA box snoRNA were found. Interestingly, genes encoding most of the identified RNAs map within a region implicated in the neurogenetic disease, Prader-Willi-syndrome (PWS). Genes in the PWS region are parentally imprinted, with only alleles being inherited from the father being expressed. The loss of imprinted gene expression results in severe developmental and neurobehavioral problems. The novel snoRNAs are intron-encoded like all guide snoRNAs previously reported in vertebrates. However, the snoRNAs containing introns and flanking exons are tandemly repeated several dozen times, while the adjacent exons do not seem to encode any protein. The biological role of the brain-specific snoRNAs is not known. Lack of complementarity to rRNA and their tissue-specific expression argue against the possibility that they are involved in rRNA modification. The three brain C/D box snoRNAs also show no significant potential to base-pair with known cellular small RNAs, but one of them (MBII-52/HBII-52) has a guide region with an 18 nt phylogenetically conserved complementarity to the serotonin receptor 5-HT_{2C} mRNA that is expressed in the brain (Cavaille *et al.*, 2000). Based on this observation, it is possible that a variety of mRNAs might be substrates for guided modification. Moreover,

in the mRNA-coding region (Xie *et al.*, 1996; Yu *et al.*, 1991) this complementarity matches a conserved region which is critical, because it is subject to both alternative splicing (Canton *et al.*, 1996), and adenosine-to-inosine editing at four vicinal sites (Burns *et al.*, 1997). The serotonin receptor mRNA transcripts undergo alternative splicing and adenosine to inosine editing at four different sites, to yield proteins with different signal transduction potentials. Remarkably, the mRNA nucleotide potentially targeted for ribose methylation by MBII-52 corresponds precisely to one of the four edited adenosines. Ribose methylation of this adenosine should dramatically hamper its deamination to inosine (Yi-Brunozzi *et al.*, 1999). Ribose methylation at the editing site could play a regulatory role, since adenosine to inosine editing is tissue-specific and most abundant in brain mRNA (Paul and Bass, 1998). It could also interfere with the utilization of the upstream alternate splice site present in the serotonin mRNA. An alternatively spliced variant of the receptor is most abundant in the choroid plexus brain area where guide MBII-52 is absent (Canton *et al.*, 1996).

1.4 SnoRNAs involved in cleavage events

In addition to mediating modification reactions, a few snoRNAs are required for processing of pre-rRNA. The corresponding snoRNAs interact directly with rRNA, but the actual function of the snoRNA in processing is not known in most cases. In yeast the U3, U14, snR10, snR30, and MRP RNA are directly involved in processing of the primary transcript and with the exception of snR10, are essential for cell viability (Maxwell and Fournier, 1995). The most abundant and well-characterized snoRNA is U3. A large snoRNP complex (Mougey *et al.*, 1993), termed the processome, interacts with

the 5' external transcript spacer (ETS) on the primary transcript via essential interactions between the 5'ETS and the U3 snoRNA (Beltrame and Tollervey, 1992; Beltrame *et al.*, 1994) and is believed to contribute to the correct folding and presentation of the precursor rRNA substrate (Tollervey and Kiss, 1997).

The third group of snoRNAs includes only a single member, the RNA component of RNase MRP. Of the known snoRNAs, the MRP RNA is the only one possessing catalytic activity (Lygerou *et al.*, 1994) and is related by homology to the RNA component of RNase P (Morrissey and Tollervey, 1995). RNase MRP is responsible for the endonucleolytic cleavage of pre-rRNA, upstream of the 5.8S rRNA (Morrissey and Tollervey, 1995). Protein components of RNase MRP are also shared with the RNase P complex, that endonucleolytically generates the mature 5' end of tRNAs (Ursic *et al.*, 1997).

1.5 Biogenesis and assembly of snoRNPs

Both snoRNA families are closely related by their unusual genomic organization and modes of biosynthesis. The DNA coding units for the snoRNAs occur in both traditional and novel genetic arrangements. Some are transcribed from independent promoters which serve mono- or polycistronic snoRNA coding units. Others are encoded within introns of protein or protein-like genes. Regardless of the diverse genomic organization, snoRNA synthesis appears to involve a number of pathways with common steps: 1) folding of the precursor to form a C/D box or H/ACA box protein binding motif; 2) binding of protein(s) to this motif; 3) processing of the precursor to the mature RNA;

4) partial or complete assembly of the snoRNP particle; and 5) transport to the nucleolus (Allmang *et al.*, 1999)

The snoRNAs encoded within introns of vertebrates are not independently transcribed but processed from the pre-mRNA introns, in most cases by exonucleolytic digestion of the debranched lariat. In yeast *S. cerevisiae,* only few guide snoRNA are intronic and most are synthesised from independent mono-, di-, or polycistronic RNA transcripts processed by endo-and exonucleases (Chanfreau *et al.*, 1998a; Chanfreau *et al.*, 1998b). Occurrence of clusters of multiple different snoRNA genes, transcribed as polycistronic precursors, sometimes encoding both C/D and H/ACA box types, from which individual snoRNAs are processed, is also widespread in higher plants and in *Trypanosoma* (Barneche *et al.*, 2001; Leader *et al.*, 1997; Leader *et al.*, 1998; Liang *et al.*, 2001; Qu *et al.*, 2001; Xu *et al.*, 2001). U 86, a novel yeast C/D snoRNA which has a human intronic homologue, is encoded within an open reading frame and its synthesis appears to be alternative to that of the cotranscribed mRNA (Filippini *et al.*, 2001). Most genes hosting intronic snoRNA guides for rRNA modification, code for proteins involved in ribosome biosynthesis or function, which suggests that this particular gene organization might provide a regulatory link between partners in the same biological process. However, several intronic snoRNA guides are hosted by different genes in different eukaryotes, and an increasing number of vertebrate host genes lack any direct relationship with translation, some even being devoid of protein-coding potential. A common feature of these snoRNAs is that their production is alternative to that of the mRNA, suggesting an important regulatory role for all the factors involved in the processing reaction.

1.6 Archaea

While the Archaea are prokaryotes, they are as distinct from simple bacteria as they are from the complex eukaryotes. Phylogenetic diversity in the domain Archaea is subdivided into two main lineages: Euryarchaeota, that include the methanogens and the extreme halophiles and Crenarcheaota that includes the sulfur-dependent thermophilic acidophiles (Fox *et al.*, 1980).

The Archaea physically appear bacteria-like since they lack a cell nucleus and complex subcellular organelles, and possess circular chromosomes, multigene transcription units and bacterial-like metabolic enzymes. In contrast, their machineries for DNA replication, transcription and translation are generally eukaryotic-like (Dennis, 1997).

1.6.1 Ribosome biogenesis in Archaea

Within Archaea there is a significant variation in the structure of rRNA transcript units. In Euryarchaeota, the units are generally bacterial-like, containing tRNA genes in the spacer region and a cotranscribed 5S rRNA gene. In Crenarchaeota, the units are more eukaryal-like, generally containing only 16S and 23S rRNA genes. The 5S rRNA gene is relocated and independently transcribed. As in Eubacteria, archaeal primary transcripts have conserved inverted repeat sequences in the spacer regions flanking the 16S and 23S rRNA that base-pair to form helices that are the sites for precursor RNA cleavage and subsequent rRNA maturation (Dennis, 1997).

A first assay of the extent of rRNA 2'-O-methylation in the archaeon *Sulfolobus solfataricus,* has revealed a high number of methylations - 67. This is very similar to the number of methylations of eukaryal rRNA (Noon *et al.*, 1998), pointing to the potential existence of an RNA-guided site selection system in Archaea. In contrast, a typical bacterium (*E. coli*) contains only four 2'-O-methylations and 10 pseudouridylations. Each of these modifications appears to be catalyzed by a site-specific protein enzyme, ribose methylase or pseudouridine synthase, without any RNA co-factor (Caldas *et al.*, 2000; Gustafsson *et al.*, 1996; Ofengand, 2000).

1.6.2 Archaeal modification guides

Archaea contain orthologous components of eukaryotic ribose-methylation and pseudouridylation snoRNPs (Dennis *et al.*, 2001; Omer *et al.*, 2003). Modifications are known to occur at many methylation sites in rRNA and tRNA and at six predicted pseudouridylation sites in rRNA. The snoRNA-like guide RNAs are called sRNAs and the modifying complexes are sRNPs.

1.6.3 C/D box sRNA in Archaea

Archaeal homologues of C/D box snoRNAs have been discovered in two independent screens. 18 sRNAs harbouring C/D hallmarks have been characterized in the crenarchaeote *Sulfolobus acidocaldarius* (Omer *et al.*, 2000) through analysis of a cDNA library generated from small RNAs immunoprecipitated from total cell extracts with antibodies raised against cloned archaeal homologues of fibrillarin and Nop56/58. A previously developed eukaryotic search program (Lowe and Eddy, 1999) was retrained

on the structural traits of experimentally identified *S. acidocaldarius* sRNAs. 13 high scoring candidates were detected, 10 of which were confirmed in the closely related archaeon *Sulfolobus solfataricus*. More than 50 candidates were also predicted in each of the three sequenced *Pyrococcus* genomes, most of which were confirmed independently (Gaspin *et al.*, 2000; Omer *et al.*, 2000).

A further 23 untested candidates in *Aeropyrum pernix* were predicted, but only 8 and 4 candidates in euryarchaea *Methanocaldococcus jannaschii* and *Archaeoglobus fulgidus*, respectively. One additional presumptive methylation guide for rRNA was detected in an experimental screen for small non-messenger RNAs in *Archaeoglobus fulgidus* (Tang *et al.*, 2002a), suggesting that the extent of rRNA 2'-O-methylation may widely vary among Archaea, even amongst hyperthermophiles.

Archaeal C/D snoRNA are slightly shorter than their eukaryotic counterparts, probably reflecting the size constraints in compact archaeal genomes.

Many of the sRNAs, particularly those in *Pyrococcus*, are "double guide" sRNAs predicted to guide methylation to two different sites, usually within the same molecules (i.e., 16S or 23S rRNA) and often in close proximity. This suggests the simultaneous formation of two guide duplexes, which could reflect chaperone function in the control of pre-rRNA folding. In contrast to eukaryotic C/D box RNAs, the archaeal sRNA appear to guide methylation not only to rRNA, but also to different positions within various tRNAs (Armbruster and Daniels, 1997; Dennis *et al.*, 2001). In many archaeal sRNAs, one or both of the D' or D box guide regions lack complementarity to 16S or 23S rRNA, but exhibit complementarity to tRNA. Remarkably, the presence of a typical C/D box methylation guide has been identified in pre-tRNA-Trp introns from numerous

euryarchaeal species(Armbruster and Daniels, 1997). The general mechanism of reaction has been investigated using a reconstituted system, and the results demonstrate that the intronically encoded C/D box RNA directs ribose methylation in cis to two sites in the mature tRNA halves (Clouet d'Orval *et al.*, 2001).

Guide regions of sRNAs appear to evolve more rapidly than the corresponding target regions in rRNA. As a consequence, most positions of methylation are not conserved between different genera. In the limited number of cases where methylation sites are conserved between genera, it is uncertain whether this is due to common ancestry or convergent evolution of the respective guide regions (Dennis *et al.*, 2001). Known genes for archaeal C/D sRNAs are not clustered but widely distributed in the genome. They are encoded on both DNA strands, usually mapping within the small portion of archaeal genomes to segments between open reading frames. A few C/D sRNAs, however slightly overlap the upstream and/or downstream protein-coding region (Dennis *et al.*, 2001; Gaspin *et al.*, 2000; Omer *et al.*, 2000).

Archaeal C/D box sRNA are able to assemble into functional RNA-protein complexes in the eukaryotic nucleus and direct site-specific 2'-O-methylation of eukaryotic rRNA (Speckmann *et al.*, 2002).

1.6.4 H/ACA box sRNA in Archaea

Genes encoding homologues of three H/ACA snoRNP core proteins Gar1p, Nop10p and Cbf5p are present in archaeal genomes (Watanabe and Gray, 2000). Recently, in an experimental screen, four candidates for H/ACA box sRNA have been identified in the archaeon *Archaeoglobus fulgidus* (Tang *et al.*, 2002a). The guide sRNAs

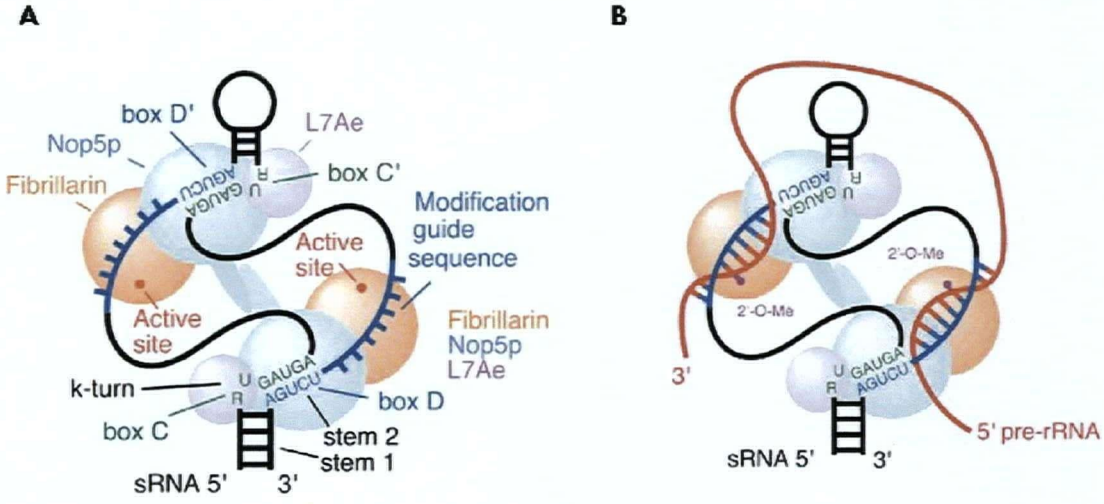
are predicted to guide a total of six rRNA pseudouridylations. The presence of the six pseudouridines has been experimentally verified in 16S and 23S *A.fulgidus* rRNAs. Several non-messenger RNAs with a high degree of secondary structure were also identified in the genomes of AT-rich hyperthermophilic Archaea by computational search of GC-rich regions associated with a screen based on comparative analysis in phylogenetically related species (Klein *et al.*, 2002). One of the small RNAs detected in three *Pyrococcus* species, three-hairpin Pf7, is unambiguously related to one of the H/ACA box sRNAs identified in *A. fulgidus* (Afu-4 sRNA), both structurally and by its potential to guide pseudouridylation (Rozhdestvensky *et al.*, 2003).

1.6.5 Archaeal proteins involved in modification

Genomic and biochemical analyses showed that archaeal box C/D sRNPs contain three proteins. These include a homologue for Fibrillarin, a single protein that is homologous to Nop58p and Nop56p, and a ribosomal protein aL7a that is homologous to 15.5 kD/Snu13p. Interestingly aL7a is also related, to the middle portion of the protein sequence for Nhp2p, one of the H/ACA snoRNA core proteins. This finding suggests that aL7a may be able to substitute for the Nhp2p protein in archaeal H/ACA sRNAs. The *in vitro* interaction of H/ACA box sRNA with aL7a from the archaeon *Archaeoglobus fulgidus* has been analysed by band shift assays (Rozhdestvensky *et al.*, 2003). The H/ACA RNA motif required for aL7a binding exhibits a structure, designated as the k-turn, which is also present in all C/D box sRNAs. These findings suggest a triple role for aL7a protein in Archaea, e.g. in ribosomes as well as H/ACA and C/D box sRNP biogenesis and function by binding to the k-turn motif.

1.6.5.1 Co-crystal structure of aNop56 and fibrillarin

The co-crystal structure of aNop56 in complex with fibrillarin from the archaeon *A. fulgidus* and the methyl donor S-adenosyl-L-methionine has been recently determined (Aittaleb *et al.*, 2003). The archaeal sRNAs and eukaryotic snoRNAs generally contain both C/D and related C'/D' regions, flanked by a modification guide sequence, giving a pseudo symmetric structure. There is also symmetry in protein structure (Figure 1-5).



Adapted from Fatica, A. and Tollervey, D. Nat Struct Biol. 2003 Apr;10(4):237-9

Figure 1-5 Potential structure of archaeal C/D box sRNP

(A) A schematic view of the potential structure of archaeal C/D box sRNP and a model for methylation site selection is illustrated. The archaeal sRNAs (black; with the modification guide in blue) contain C/D and C'/D' boxes. Two copies of Nop5p (aNop56) homodimerize. (B) The target RNA (red) is 2'-O-methylated at the site (indicated by a purple dot) which is determined within the duplex formed with the guide sRNA - 5 nt from the D or D' box.

Two copies of aNop56 homodimerize via a coiled-coil domain, linking the complex at C/D and C'/D'. In eukaryotic snoRNPs a heterodimer between the closely related

Nop58p and Nop56p proteins is predicted. Without the target RNA sequences, the single-stranded guide sequence adjacent to the D box will be bound across the aNop56–Fibrillarin surface and presented to potential substrates. Following substrate recognition, the target RNA will form a duplex with the guide sequence (Figure 1-5 B). The site of 2'-O-ribose methylation is determined solely by its position within a duplex (formed within the guide sRNA) relative to the active site of Fibrillarin. This duplex is likely to be stacked onto stem II, formed by the C/D or C'/D' box elements, therefore resulting in a continuous helix up to the k-turn. Accurate positioning of stem II within the complex is likely to be assisted by binding of aL7a to the sharp k-turn in the loop formed between boxes C and D. aL7a binds only the loop of the k-turn, while the stem II is required for binding of aNop56 and fibrillarin (Aittaleb *et al.*, 2003). The aNop56–fibrillarin structure suggests that stem II of C/D box must be bound by aNop56. The interface between fibrillarin and the N-terminal region of aNop56 is extensive, largely non-polar, and exhibits surface complementarity features consistent with stable relative positioning (Aittaleb *et al.*, 2003). Consistent with this observation, a point mutation in fibrillarin that lies within this interaction region inhibited methylation in yeast and in an *in vitro* system derived from *Sulfolobus* (Omer *et al.*, 2002; Tollervey *et al.*, 1993). It is, therefore, likely that the positions of both the snoRNA-target RNA duplex and the active site in fibrillarin are tightly constrained relative to the C/D box bound to aNop56. The net result is that only the 2'-hydroxyl in the target RNA that is positioned five base pairs, (or one half helical turn), from the D box is available to the active site for modification. Any mismatches would distort the snoRNA-target helix, altering the position of this 2'-hydroxyl relative to the active site in Fibrillarin. This presumably contributes to the

normally strict requirement for complete complementarity.

1.6.5.2 *In vitro* reconstitution of a C/D box methylation guide ribonucleoprotein complex

Recently, RNA-guided methylation has been demonstrated *in vitro,* using a reconstituted archaeal sRNP and a short rRNA fragment as substrate (Omer *et al.*, 2002). The archaeal proteins aFib, aNop56 and aL7a proteins (homologues of the eukaryotic proteins fibrillarin, Nop56/58, and 15.5 kD/Snu13p respectively) from *Sulfolobus solfataricus* have been expressed and purified ribonucleoprotein complex has been assembled from the three proteins and an *in vitro* transcript of the *Sulfolobus* sR1 sRNA (Omer *et al.*, 2002). When provided with S-adenosylmethionine and a fragment of rRNA complementary to the D box guide, this reconstituted complex was active in site specific methylation. The presence of fibrillarin is essential for methylation. Mutant proteins having amino acid replacements in the SAM-binding motif of fibrillarin can assemble into an RNP complex, but the resulting complexes are defective in methylation activity (Omer *et al.*, 2002). This demonstration shows that that an sRNP complex containing only the core proteins and guide RNA is sufficient to catalyze methylation.

Further investigations with *in vitro* assembly of an archaeal C/D box sRNP from *M. jannaschii* using recombinant core proteins has yielded an RNA:protein enzyme that guides methylation from both the terminal C/D box core and the internal C'/D' RNP complexes (Tran *et al.*, 2003). Efficient ribonucleotide 2'-O-methylation requires that both the C/D and C'/D' box RNPs function within the full-length sRNA molecule. In contrast to the eukaryotic snoRNP complex, where the core proteins are distributed

asymmetrically on the C/D and C'/D' box motifs, all three archaeal core proteins bind both motifs symmetrically (Tran *et al.*, 2003).

1.7 Project Objectives

We have recently demonstrated RNA-guided methylation *in vitro* with a reconstituted archaeal sRNP and a short 16S rRNA fragment as substrate. The main objectives of this research were to optimize the *in vitro* methylation assay, to show that site-specific methylation of the rRNA fragment was achieved, and to obtain a better understanding of the versatility and specificity of this reaction. Also, the activity of reconstituted particles was examined on more complex target substrates, including 5S rRNA and $tRNA^{Gln}$ and "double guide" oligonucleotides that exhibit either direct or reverse complementarity to both the D' and D box guides.

This research should spur attempts to establish homologous eukaryotic systems and *in vitro* systems for other RNA modifications (e.g. pseudouridylation).

2 Materials and Methods

2.1 Materials

2.1.1 Prokaryotic strains and plasmids

Sulfolobus strains

Sulfolobus solfataricus and *Sulfolobus acidocaldarius* were used to obtain genomic DNA (from Deutsche Sammlung von Mikroorganismen, Braunschweig, Germany)

Table 2-1 *Sulfolobus* strains

Strain	ATCC/DSM number	Reference
Sulfolobus solfataricus	ATCC 35092 / DSM 1617 / P2	(Zillig, 1980)
Sulfolobus acidocaldarius	ATCC 33909 / DSM 639	(Brock, 1972)

***E.coli* strains**

E. coli strain JM109 and NovaBlue were used for the propagation of plasmid vectors. BL21(DE3)pLysS and BL21(DE)pLysE were used for expression of recombinant proteins. Their genotypes are listed in Table 2-2 .

Table 2-2 *E.coli* strains

Strain	Genotype	Reference
JM109	*e14⁻(McrA⁻) recA1 endA1 gyrA96 thi-1 hsdR17($r_K^- m_K^+$) supE44 relA1 .(lac-proAB) [F' traD36 proAB lacIq Z .ΔM15]*	(Yanisch-Perron *et al.*, 1985)
NovaBlue	*endA1 hsdR17 (r_{K12}-m_{K12}+) supE44 thi-1 recA1 gyrA96 relA1 lac [F' $proA^+B^+$ $lacI^qZ\Delta M15$ $Tn10(Tc^R)$]*	Novagen
BL21(DE3)pLysS	*F^- ompT $hsdS_B$ (rB^- mB^-) gal dcm (DE3) pLysS (Cm^R)*	(Studier and Moffatt, 1986; Studier, 1991)
BL21(DE3)pLysE	*pLysE F^- ompT $hsdS_B$ (rB^- mB^-) gal dcm (DE3) pLysE (Cm^R)*	(Studier and Moffatt, 1986; Studier, 1991)

E.coli plasmid vectors

Table 2-3 Plasmid vectors

Vector	Resistance	Reference
pGEM 3+	Ampicillin	Promega, Madison, WI USA.
pET3d (expression vector)	Ampicillin	Novagen, Madison, WI USA
PET28a (expression vector)	Kanamycin	Novagen, Madison, WI USA

Enzymes

Restriction enzymes and DNA Ligase were purchased from Invitrogen (Carlsbad, CA) or Amersham/Pharmacia (Piscataway, NJ). P1 Nuclease was obtained from Roche Diagnostics (Basel, Switzerland). The Mega-Short Script kit was obtained from Ambion (Austin, TX).

Chemicals

All radioactive nucleotides were purchased from Perkin Elmer. Non-radioactive nucleotides where purchased from Amersham/Pharmacia. Oligonucleotides were synthesized by Sigma Genosys or Invitrogen. Yeast extract, Bactotryptone and Bacto Agar were supplied by Difco Laboratories (Detroit, MI). All other chemicals were purchased from Fisher Scientific, Aldrich or Pharmacia. TLC plates were purchased from Merck, Darmstadt. The Sequenase Version 2.0 Sequencing Kit was purchased from United States Biochemical / Amersham. The pGem Vector kit was supplied by Promega. The pET system was obtained from Novagen. Dr. Arina Omer kindly provided the plasmid clone 27 containing sR10 and the plasmid pD1282.

2.1.2 Media, buffers, solutions

2× YT medium:

Per liter:

Deionized H_2O, 16 g Bacto-Tryptone, 10 g Bacto-Yeast extract, 5 g NaCl.

Medium was adjusted with 5N NaOH to pH 7.0 and autoclaved.

SOC Medium

Per liter:

Deionized H_2O, 20 g Bacto-Tryptone, 5 g Bacto-Yeast extract, 0.5 g NaCl.

Medium was autoclaved and 20 ml of a sterile 1M Glucose solution was added.

Medium for *S. acidocaldarius*

0.1 % w/v Bacto-Tryptone, 0.1% w/v Bacto-Yeast extract, 800 μM $MgSO_4 \bullet 7H_2O$, 2.25 mM $CaCl_2$, 20 mM $(NH_4)_2SO_4$, 22.5 mM KH_2PO_4 and 0.1% v/v trace metal mix. Medium was adjusted with H_2SO_4 to a pH of 3.6 prior to inoculation and continuously adjusted to maintain this pH value during the growth of the culture.

Trace Metal mix

3.5 μM $CuSO_4 \bullet 7H_2O$, 15 μM $NaMoB_4 \bullet 2H_2O$, 15 μM $VOSO_4 \bullet 2H_2O$, 30 μM $CuCl_2 \bullet 2H_2O$, 75 μM $ZnSO_4 \bullet 7H_2O$, 875 mM $MnCl_2 \bullet 4H_2O$, and 1.25 mM $Na_2B_4O_7 \bullet 10H_2O$.

Medium for *S. solfataricus*

20mM KH_2PO_4, 100 mM $(NH_4)_2SO_4$, 10 mM $MgSO_4 \bullet 7H_2O$, 5 mM $CaCl_2$, 30 mM glutamic acid, 1% Bacto-Yeast extract, 2% glucose. The pH was adjusted to 3.5 with H_2SO_4.

TE buffer

10 mM Tris-HCl (pH8), 1 mM EDTA,

TES buffer

10 mM Tris-HCl (pH8), 1mM EDTA, 0.1 M NaCl

10 ×TBE buffer

0.89 M Tris-HCl, 0.89 M boric acid, 20 mM EDTA (pH 8)

Renaturation buffer R

25 mM sodium phosphate buffer, pH 7, 100 mM NaCl

Binding buffer D

25 mM sodium phosphate buffer, pH 7, 100 mM NaCl and 1mM $MgCl_2$

6 × Sequencing loading dye

0.09% Bromophenol blue, 0.09% xylene cyanol, 60% glycerol, 60 mM EDTA

Protein loading buffer

100 mM Tris-HCl pH 6.8, 20% Glycerol, 4 % SDS, 5% β-mercaptoethanol, 0.2% bromophenol blue

Sonication buffer

50 mM Tris-HCl pH 8.0, 1 mM EDTA, 1 mM DTT, 15 mM $MgCl_2$

SDS destain Solution

40% methanol, 10% acetic acid

Coomassie Blue Stain Solution

0.25% w/v Coomassie Blue R-250, 40% methanol, 10% acetic acid

DEPC Water

For RNAse-free water, DEPC was added to double-distilled H_2O to a concentration of 0.05%, stirred overnight at room temperature and autoclaved.

2.1.3 Oligonucleotides

Oligonucleotides that were directly used as partially single-stranded templates for run-off transcripts were precipitated with isopropanol and resuspended in TES buffer.

Table 2-4 Oligonucleotides

Name	Sequence	Description
OSZ102	5' CCGATATCCATGGTGAAAA TATACCTAATTGA 3'	Complementary to 5' end of Sso Nop – contains a NcoI site for cloning into pET28
OSZ103	5' CCGAATTCTCACTTTCTTT TACCTCTTCTCT 3'	Complementary to 3'end of Sso Nop – contains an EcoRI site for cloning into pET28
OSZ120	5' TGGTCCCTCTTGTCACTGCG ACCTACTGACCTATAGTGAGTC GTATTA 3'	Used in combination with T7 primer to directly transcribe the region of C1914 of 23S rRNA (Sac sR2 and sR1/sR2 hybrid target)
OSZ121	5' TAATACGACTCACTATAGGG AGTGATGAGACGAGCGATA 3'	Complementary to 5'end of Sac sR2
OSZ122	5' TTCTTCAGTTCGCTGTGAC 3'	Complementary to 3'end of Sac sR2
OSZ173	5' CACCCTCAGCCATCCTTTA ATCATCACTCATTTAACAGC TCTCTCATCATAGGGTCCTAT AGTGAGTCGTATTAAATT 3'	Used in combination with T7 primer to directly transcribe sR24 from partially single stranded DNA template; T7 promoter is underlined.
OSZ174	5'GTGAGCTGTTACGCACTCTTT AAAGGATGGCTGCTTCCC TATAGTGAGTCGTATTAAAT TT 3'	Used in combination with T7 primer to directly transcribe 23S rRNA from1213 to 1250 from partially single stranded DNA; T7 promoter is underlined.
OSZ176	5'TAGCCGGGCAGGGATTCGAA3'	Complementary to 3'end of $tRNA^{Gln}$
OSZ180	5' GGGTTCGAAACGAGTCCGG GTGTTACCCGCCTATAGTGAGT CGTATTA 3'	Used in combination with T7 primer to directly transcribe the region of C32 from Sac 5S rRNA.

OSZ182	5' TAATACGACTCACTATAGG GATTTAAATGATGAGCTTG 3'	Complementary to 5'end of sR11 from Sso.
OSZ183	5'AAATGGATACTCAGCGGCTA3'	Complementary to 3'end of sR11 from Sso
OSZ190	5' TAATACGACTCACTATAGG GAATGATGTGGAATCCGGGA 3'	Complementary to 5'end of sR10 from Sac
OSZ192	5' TCTTTAAAGGATGGCTGCT TCTAAGCCTACCCTATAGTGAG TCGTATTA 3'	Used in combination with T7 primer to directly transcribe D 1221 target for Pyrococcus double guide (sR24) from partially single stranded DNA
OSZ193	5' GGGTGAGCTGTTACGCACT CCCCTATAGTGAGTCGTATTA 3'	Used in combination with T7 primer to directly transcribe D' 1243 target for Pyrococcus double guide (sR24)
OSZ194	5'TAATACGACTCACTATAGC CCACCCGGTCATAGTGAGC 3'	Complementary to 5'end of 5S rRNA of Sac
OSZ196	5'CCCATCCCAGCTCCAGAGAGG 3'	Complementary to 3'end of 5S rRNA of Sac
OSZ198	5'GTTTGCTCTGAGCTTTGGCC 3'	Complementary to 3'end of sR14 from Sac
OSZ199	5'TAATACGACTCACTATAGG CTGTGAAGACGCTAGACTT 3'	Complementary to 5'end of sR14 from Sac
OSZ200	5'TAATACGACTCACTATAGAGC CGGGTAGTCTAGTAGTCAAGG 3'	Complementary to 5' end of $tRNA^{Gln}$. Contains a mismatch at position 18, to introduce a mutation from G18 to A18.
OSZ203	5'TAATACGACTCACTATAGGGA TTTAAATGATGATGAGCTTGAC TACTTA 3'	Complementary to 5' end of sR11. Contains a mismatch at position +5 from the D' box, to introduce a mutation from C5 to T5.
OSZ206	5' ATAATCAGCGCTCGCGCTT 3'	Complementary to 3'end of sR10 from Sac
OSZ207	5'GGCCCCCGGGAATTCCGAG CCGGGTAGTCTAGTGGTC 3'	Complementary to 5'end of tRNAGln (YTG) from Sso. Contains Eco RI site for cloning into pGem3+.

OSZ208	5'TCCCCCGGGAAGCTTTAGC CGGGCAGGGATTCGAA 3'	Complementary to 3' end of tRNAGln (YTG) from Sso. Contains HindIII site for cloning into pGem3+.
OSZ216.2	5'GTTTGCTCTGAGCTCTGGCCCT TCACATGAGTCAGTCTAAGTCT AGCGTCTTCACAGCCCTATAGT GAGTCGTATTAAATT-3'	Used in combination with T7 primer to directly transcribe Sac sR14 from partially single stranded DNA.
OSZ218	5'TATAATCAGCGCTCGCGCTT TTTGTCATCATTCTCAGCACCC GGACTCCACATCATTCCCTATA GTGAGTCGTATTA 3'	Used in combination with T7 primer to directly transcribe an optimized Sac sR10 from partially single stranded DNA.
OSZ219.2	5'TAATACGACTCACTATAGGA TGATGACAAAGAGCCGAA3'	Complementary to the 5'end of Sac sR7
OSZ220	5'CTCTATCAGTGGCTGCCCAC3'	Complementary to the 3'end of Sac sR7
OSZ221	5'CTCCAGGGTGGGAAGAGCCG ACATCGATGTAGCCTATAGTGA GTCGTATTA3'	Used in combination with T7 primer to directly transcribe region of D' box target (23S rRNA Sac C2649) from partially single stranded DNA.
OSZ222	5'CCTTTAACGGGCGGGCAGCCC CTACCCTTGGCCTATAGTGAGT CGTATTA3'	Used in combination with T7 primer to directly transcribe region of D box target (23S Sac U2692) from partially single stranded DNA.
OSZ228	5' TAATACGACTCACTATAGAC AGATGATGAATTCCCGATAG 3'	Complementary to 5'end for sR1/sR2 hybrid cloned in plasmid pPD1282.
OSZ229	5' AATCAGTTCGCAGTGACCT AG 3'	Complementary to 3'end for sR1/sR2 hybrid cloned into plasmid pPD1282.
OSZ233	5' TAATACGACTCACTATAGGC TACATCGATGTCGGCTCT 3'	Complementary to 5'end of target region of Sac sR7 in 23S RNA
OSZ234	5'CTTTAACGGGCGGGCAGCCC 3'	Complementary to 3'end of target region of Sac sR7 in 23S rRNA
OSZ235	5'TAATACGACTCACTATGGGGT GCCTATCGTGGTTTA 3'	Complementary to 5' end of C1914 region in 23S rRNA (target for Sac sR2 and sR1/sR2 hybrid)

OSZ236	5'ACCCTTTCGGGCTAGCGGGG3'	Complementary to 3' end for C1914 region in 23S rRNA
OSZ238	5' CTTTAACGGAAGAGCCGACT ACCCTTGGAGGCAGCTGCACCT CCAGGGTGGGGCGGGCAGCCC ATCGATGTACCTATAGTGAGTC GTATTAAATT 3'	Used in combination with T7 primer to directly transcribe the target for Sac sR7, containing the methylation sites in a "direct" order.
AO63.1	5' GTAATACGACTCACTATAG GGATAAGCCATGGGAG 3'	Forward primer used to PCR-amplify the DNA template for transcription of sR1 target RNA.
AO65	5'TATTTAGGTGACACTATAGG TTAGCCACGTGTTACTCAGCC 3'	Reverse primer used to PCR-amplify the DNA template for transcription of sR1 target RNA.
AO66	5' AGAATTCCCATGGACGCGC GATGTCAAAAGCTAG 3'	Complementary to 5'end of Sso aL7a ORF. Contains NcoI site for cloning into pET3d vector.
AO67	5' TTAGGATCCTTAACTTGAAG TTTTACCTTTAATC 3'	Complementary to 3'end of Sso aL7a ORF. Contains BamHI site for cloning into pET3d vector.
AO70	5' AAAGATCTCCATGGCTGA AGTAATTACCGTAAAAC 3'	Complementary to the 5'end of Sso aFib ORF. Contains NcoI site for cloning into pET3d vector.
AO71	5' TTAGGATCCCTACCCTTTA TATTTGCTAAGAAC 3'	Complementary to 3'end of Sso aFib ORF. Contains BamHI site for cloning into pET3d vector.
HE63	5'GTTATCAGACCATGGG 3'	Complementary to 3'end of sR1 from Sac
T7	5'TAATACGACTCGCTATAG 3'	Complementary to T7 RNA Polymerase promoter region

* T7 promoter region are underlined

2.2 Methods

2.2.1 Cell culturing of *S. solfataricus* and *S. acidocaldarius*

S. solfataricus and *S. acidocaldarius* were grown at 80° C in a well-aerated 2 litre Perkin-Elmer culture vessel in their appropriate medium. Cells were harvested when the OD_{600} was between 0.4-0.6. Cells were rapidly cooled to 10°C, pelleted by centrifugation at 5,000 rpm in a Sorvall GS-3 rotor for 10 min. Cell paste was either stored at –70°C or resuspended in sonication buffer.

2.2.2 DNA extraction from *Sulfolobus*

1.5 ml of culture was harvested by centrifugation in a tabletop centrifuge and the supernatant discarded. The cells were lysed by adding 0.4 ml of Lysis buffer (40mM Tris-HCl, 20 mM EDTA pH 8.0) and incubated for 30 min at 37°C. 0.5 ml 5% SDS with 100 mg/ml RNase A were added. The suspension was twice extracted with phenol/chloroform. The aqueous phase was transferred into a new microtube that contained 15 µl 5 M NaCl and 1 ml of 95% ethanol was added. The suspension was mixed until the DNA formed threads and then spooled on a sealed pasteur pipette. The DNA was rinsed with 70% cold ethanol and dried by standing the tip up at room temperature for 15 min. The DNA was recovered by introducing the pasteur pipette into a new microtube, adding 200 µl of TE buffer and shaking overnight at 4°C.

2.2.3 Cloning of $tRNA^{Gln}$

$tRNA^{Gln}$ (YTG) was amplified by PCR from *S. solfataricus* genomic DNA using oligonucleotides OSZ207 and OSZ208, cloned between the EcoRI and HindIII sites of pGEM3+ (Promega) and propagated into the *E.coli* strain NovaBlue (Novagen). Individual colonies were picked and used to verify the insert sequences for $tRNA^{Gln}$ (TTG) and $tRNA^{Gln}$ (CTG).

2.2.4 Cloning of the *S. solfataricus* aL7a, aNop56 and aFib genes

The gene encoding *S. solfataricus* aL7a (accession number: S75397, gi:7440709) was amplified by PCR using oligonucleotides AO66 and AO67, cloned between the NcoI and BamHI site of pET3d and transformed into the *E.coli* strain BL21(DE3)pLysE, for overexpression. Cells would only grow if 0.2% glucose were added to the YT media.

PCR amplification of aNop56 (accession number: AKK41215, gi: 13814119) was carried out with oligonucleotides OSZ102 and OSZ103. The PCR fragment was cloned into the NcoI and EcoRI site of pET28a and transformed into the *E.coli* strain BL21 (DE3).

aFib (NP_342426, gi:15897821) was amplified by PCR using oligonucleotides AO70 and AO71, cloned between the NcoI and BamHI site of pET3d and transformed into the *E.coli* strain BL21(DE3)pLysS .

Clones were subjected to restriction enzyme digest. Clones containing inserts of the correct length were sequenced to confirm the insert sequence.

2.2.5 Expression and Purification of Recombinant Proteins in *E.coli*

Expression

Cultures were grown to an OD_{600nm} of 0.4-0.6 and protein expression was induced by adding 0.5 mM IPTG. Cells were grown overnight at room temperature and harvested by low speed centrifugation, washed and resuspended in their appropriate buffer: buffer A for aL7a: 50 mM Bis-Tris buffer, pH 6.5, 50mM NaCl, 10 mM DTT, buffer B for aNop56: 50 mM Tris-HCl buffer, pH 7.5, 200 mM NaCl, 10 mM DTT and buffer C for aFib: 50 mM Tris-HCl buffer, pH 8.5, 50 mM NaCl, 10 mM DTT. Cells were disrupted by sonication and the clear lysate (10 μg/μl) was heated for 5 minutes at 65°C followed by centrifugation (Figure 2-1). Proteins were further purified by ion-exchange chromatography.

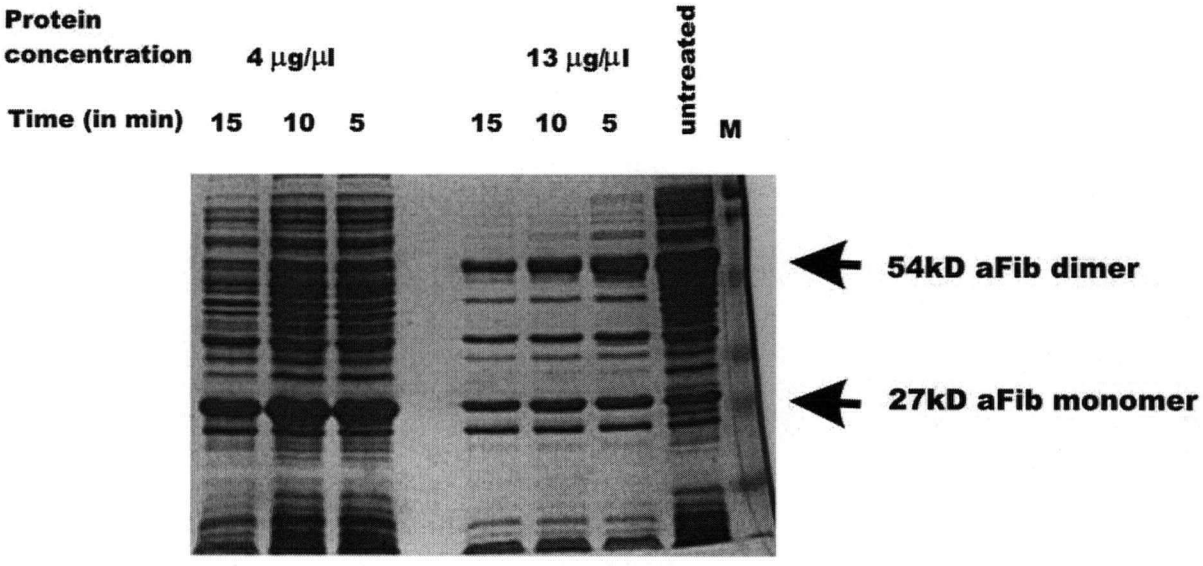


Figure 2-1 Purification of aFib by heat precipitation of *E.coli* extract

An example of *S. solfataricus* protein (aFib) expressed in *E. coli* is shown. Total protein extracts from *E.coli* with concentrations of 4 µg/µl and 13 µg/µl were heated for 5, 10 and 15 min. The supernatant was separated on SDS PAGE gel. The lane with protein molecular weight standards (M) and protein sample that has not been heated (untreated) are indicated. At high protein concentrations aFib forms dimers. In further purifications concentrations of 10 µg/µl were used.

Chromatography

aL7a: The heat soluble protein was loaded on a 10 ml bed DEAE-Sepharose column (Pharmacia), equilibrated with buffer A. The protein was collected in 4ml fractions in the flow-through, concentrated (Centricon, Millipore) to an end volume of 500 µl and applied to a size-exclusion column (Superdex 75, 10/30, Pharmacia). Fraction containing the purified aL7a protein were pooled and concentrated (Figure 2-2).

aNop56: The heat soluble protein was loaded on a 10 ml bed DEAE-Sepharose column (Pharmacia), equilibrated with buffer B. The bound aNop56 protein was eluded by a step gradient in 4 ml fractions (salt concentrations: 200 mM, 400 mM and 700 mM NaCl). The majority of the aNop56 protein would elute in the 400 mM NaCl fractions. Protein was concentrated (Centricon, Millipore) to an end volume of 500 µl and applied to a size-exclusion column (Superdex 75, 10/30, Pharmacia). Fractions containing the purified aNop56 protein were pooled and concentrated (Figure 2-2).

aFib: The heat soluble protein was loaded on a 10 ml bed DEAE-Sepharose column (Pharmacia), equilibrated with buffer C. The protein was collected in 4 ml fractions in the flow-through, concentrated (Centricon, Millipore) to an end volume of 500 µl and applied to a size-exclusion column (Superdex 75, 10/30, Pharmacia). Fractions containing the purified aFib protein were pooled and concentrated (Figure 2-2). The purified recombinant proteins were frozen in liquid nitrogen and stored at -80°C.

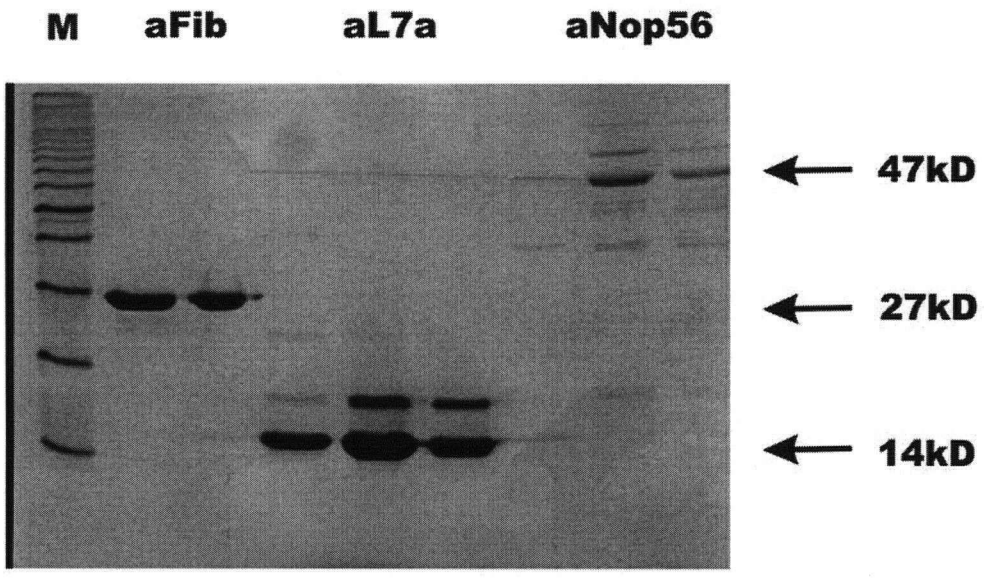


Figure 2-2 Purified proteins aFib, aL7a and aNop56

S. solfataricus proteins aFib, aL7a and aNop56 were expressed in *E.coli* and purified as described in Materials and Methods. Protein samples from the size-exclusion chromatography have been separated on SDS PAGE gel. The lane with protein molecular weight standards is indicated (M) and the sizes of the purified proteins in kD are shown.

2.2.6 Protein Assay

Protein concentrations in the purified fractions were determined using the BioRad Microassay which is based on the Bradford Assay. The purified protein samples were diluted to 800 µl volume with subsequent incubation for 10 min with 200 µl dye reagent concentrate. BSA was used to develop a standard protein concentration curve.

2.2.7 *In vitro* transcription of guide and target RNAs

The Mega-Short Script kit from Ambion (Austin, TX) was used for *in vitro* transcription, following the manufacturer's recommendations. The reaction was carried out overnight at 37°C.

2.2.7.1 Standard target and guide

***S. acidocaldarius* sR1 and 16S rRNA fragment**

sR1 was amplified by PCR from a plasmid containing the *S. acidocaldarius* sR1 gene using primer T7 and HE63.

The target RNA, a 112 bp fragment of rDNA spanning position U52, was amplified from *S. acidocaldarius* genomic DNA using primers AO 63.1 and AO 65. The products were purified by PAGE and used as template for T7 transcription.

2.2.7.2 Full length tRNA targets and guides

***S. solfataricus* sR11, *S. acidocaldarius* sR14and tRNA^{Gln} target**

sR11 was amplified by PCR from *S.solfataricus* genomic DNA using primers OSZ182 and OSZ183. The resulting fragment was precipitated with isopropanol and used as template for T7 transcription.

The mutant sR11 was amplified by PCR from *S. solfataricus* genomic DNA using primers OSZ203 and OSZ 183. The fragment was used as template for T7 transcription.

sR14 was amplified by PCR from *S. acidocaldarius* genomic DNA using the

primers OSZ198 and OSZ199. The product was precipitated with isopropanol and used as template for the T7 polymerase reaction. The T7 template to transcribe mutant sR14 was generated from a partially single-stranded DNA template by annealing the T7 promoter primer to the oligonucleotide OSZ216.2.

T7 template used to transcribe $tRNA^{Gln}$ was generated by PCR using the respective templates pGEM_$tRNA^{Gln}$(UUG) and pGEM_$tRNA^{Gln}$(CUG) and the primers OSZ208 and T7.

The G18/A mutant tRNA was generated by PCR using *S. solfataricus* genomic DNA and the primers OSZ200 and OSZ176. The fragment was sequenced and used as template for the T7 polymerase reaction.

2.2.7.3 Full length ribosomal RNA target and guide

***S. acidocaldarius* sR10 and 5S rRNA target**

sR10 was amplified by PCR from the plasmid clone 27 (obtained from Dr. Omer) containing sR10 from *S.acidocaldarius* using primers OSZ190 and OSZ206.

Optimized sR10 was generated from a partially single-stranded DNA template by annealing the T7 promoter to the oligonucleotide OSZ218.

5S rDNA was amplified from *S. acidocaldarius* genomic DNA by using primers OSZ194 and OSZ196. The resulting fragment was precipitated with isopropanol and used as template for T7 transcription reaction.

A T7 template used to transcribe a 30 nt long RNA containing the only the region spanning C32 of 5S rRNA was generated from a partially single-stranded DNA template, by annealing the T7 promoter primer to the oligonucleotide OSZ180.

2.2.7.4 Double guides and targets

S. acidocaldarius sR7 inverted guide and targets

sR7 was amplified by PCR from *S .acidocaldarius* genomic DNA using primers OSZ219.2 and OSZ220. The fragment was precipitated and used as template for T7 transcription reaction.

The two targets for Sac sR7, containing methylation sites C649 and C2692 of 23S rRNA, were generated by using partially single-stranded DNA templates by annealing the T7 promoter to the oligonucleotides OSZ221 and OSZ222 respectively.

Inverted target: The target molecule containing both methylation sites on one molecule in an inverted order was amplified from *S. acidocaldarius* genomic DNA by using primers OSZ233 and OSZ234. The product was precipitated with isopropanol and used as template in the T7 polymerase reaction.

Direct target: The target molecule containing both methylation sites on one molecule in direct order was generated by annealing the T7 promoter to oligonucleotide OSZ238 and using the partially single stranded DNA as template for the T7 polymerase reaction.

Pyrococcus sR24 direct guide and target

Pyrococcus sR24 was generated by using a partially single-stranded DNA template by annealing the T7 promoter to the oligonucleotide OSZ173.

To generate the target RNA, spanning the two predicted sites of methylation (position C1243 and C1221) in 23S rRNA of *Pyrococcus horikoshii*, a 37 nt long

partially single-stranded DNA template (OSZ 174) with an annealed T7 promoter was used as template for a T7 transcription reaction. To produce the D box target (C1221) and D' box target (C1243), partially single-stranded templates were created by annealing the T7 promoter to the oligonucleotides OSZ192 (D box target) and OSZ193 (D'box target) respectively.

2.2.7.5 Engineered guide

Engineered sR1/sR2 hybrid sRNA, *S. acidocaldarius* sR2 sRNA and target

sR1/sR2 hybrid RNA contains an 11 nucleotide-substitution of the *S. solfataricus* D box guide region with that of *S. acidocaldarius* sR2. This replacement changes the target methylation site from U52 in 16S rRNA to position C1914 in 23S rRNA; site C1914 has been previously confirmed to be ribose-methylated in *S. acidocaldarius* but not normally methylated in *S. solfataricus*. Plasmid pPD1282 (Omer AD) and oligonucleotides OSZ228 and OSZ229 were used to generate the DNA template for the run-off transcription of the sR1/sR2 hybrid RNA.

sR2 was amplified by PCR from *S.acidocaldarius* genomic DNA using primers OSZ121 and OSZ122. The fragment was precipitated and used as template for T7 transcription reaction.

To generate the target RNA spanning the region of C1914 in 23S rRNA of *Sulfolobus* a 28 nt long partially single-stranded DNA template (OSZ 120) with an annealed T7 promoter was used as template for a T7 transcription reaction.

To ensure the integrity of all transcripts from every reaction (20 µl), 3 µl were removed and ([α–^{32}P] GTP or [α–^{32}P] CTP) were added prior to incubation, to uniformly label the transcripts. The transcripts were separated on an 8% or 6 % polyacrylamide gels and visualized by Phosphor Imager.

Table 2-5 Guide and target RNAs

Guide and target RNAs used in this study. Complementary regions in guide and target are underlined. The residue predicted to be methylated in the target RNA is bold.

Guide RNA		Target RNA	
Sac sR1	5′CAGUUGAUGAGAAGUUAAAAAAGCGA UGGAUGAGCUUAACUCCCAUGGUCUGAU AAC 3′	Sac 16S rRNA /U52	5′GGAUAAGCCA**U**GGGAGUCUUACACUCCCGGGUAAG GGAGUGUGGCGGACGGCUGAGUAACACGUGGCUAACC UACCCUCGGGACGGGGAUAACCCCGGGAAACUGGGGA UAA 3′
Sso sR11	5′UCGGAUUUAAAUGAUGAGCUUGACCA CUUAUGAAGCUAGAUGAUAUAAUAAAGG UAGCCGCUGAGUAUCCAUUU	tRNA Gln(YUG) /G18	5′AGCCGGGUAGUCUAGU**G**GUCAAGGAUCCAGGGCUY UGGCCCCUGGGACCAGGGUUCGAAUCCCUGCCCGGCU A 3′
Sso Mut sR11	5′UCGGAUUUAAAUGAUGAGCUUGACUA CUUAUGAAGCUAGAUGAUAUAAUAAAGG UAGCCGCUGAGUAUCCAUUU	tRNA Gln(YUG) G18-> A18	5′AGCCGGGUAGUCUAGU**A**GUCAAGGAUCCAGGGCUY UGGCCCCUGGGACCAGGGUUCGAAUCCCUGCCCGGCU A 3′
Sac sR14	5′GCUGUGAAGACGCUAGACUUAGACUG ACUCAUGAUGAAGGGCCAAAGCUCAGAG CAAAC 3′	tRNA Gln(UUG) /U34	5′AGCCGGGUAGUCUAGUGGUCAAGGAUCCAGG GCU**U**UGGCCCCUGGGACCAGGGUUCGAAUCCCU GCCCGGCUA 3′
Sac Mut sR14	5′GCUGUGAAGACGCUAGACUUAGACUG ACUCAUGAUGAAGGGCCAGAGCUCAGAG CAAAC 3′	tRNA Gln(CUG) /C34	5′AGCCGGGUAGUCUAGUGGUCAAGGAUCCAGG GCU**C**UGGCCCCUGGGACCAGGGUUCGAAUCCCU GCCCGGCUA 3′
Sac sR10	5′GAAUGAUGUGGAAUCCGGGAUCUGAG AAUGAUGACAAAAAGCGCGAGCGCUGAU UAUA 3′	Sac 5S rRNA /C32	5′CCCACCCGGUCAUAGUGAGCGGGUAACACC**C**GGAC UCGUUUCGAACCCGGAAGUUAAGCCGCUCACGUCAGA GGGGCCGUGGGAUCCGAGAGGGCCCGCAGCCUCUCUG AGCUGGGAUGG 3′
Sac sR10	5′GAAUGAUGUGGAAUCCGGGAUCUGAG AAUGAUGACAAAAAGCGCGAGCGCUGAU UAUA 3′	Sac 5S rRNA frgmt	5′GCGGGUAACACC**C**GGACUCGUUUCGAACCC 3′
Opt. Sac sR10	5′GAAUGAUGUGGAGUCCGGGUGCUGAG AAUGAUGACAAAAAGCGCGAGCGCUGAU UAUA 3′	Sac 5S rRNA /C32	5′CCCACCCGGUCAUAGUGAGCGGGUAACACC**C**GGAC UCGUUUCGAACCCGGAAGUUAAGCCGCUCACGUCAGA GGGGCCGUGGGAUCCGAGAGGGCCCGCAGCCUCUCUG AGCUGGGAUGG 3′
Opt. Sac sR10	5′GAAUGAUGUGGAGUCCGGGUGCUGAG AAUGAUGACAAAAAGCGCGAGCGCUGAU UAUA 3′	Sac 5S rRNA frgmt	5′GCGGGUAACACC**C**GGACUCGUUUCGAACCC 3′
Sac sR7	5′GAUGAUGACAAAGAGCCGAAUGGAUU AGUGACAUCUAAUUUUGUGGGCAGCCAC UGAUAGAG 3′	Sac 23S rRNA /G2649 /U2692	G2649: 5′CUACAUCGAUGUCG**G**CUCUUCCCACCCUGGAG 3′ U2692: 5′CCAAGGG**U**AGGGGCUGCCCGCCCGUUAAAGG 3′

Sac sR7	5'GAUGAUGACAAAGAGCCGAAUGGAUU AGUGACAUCUAAUUUUGUGGGCAGCCAC UGAUAGAG 3'	Sac 23S rRNA G2649 + U2692	G2649 and U2692 (inverted) 5'GCUACAUCGAUGUCG**G**CUCUUCCCACCCUGGA GGUGCAGCUGCCUCCAAGGG**U**AGGGCUGCCCGCC CGUUAAAG 3'
Sac sR7	5'GAUGAUGACAAAGAGCCGAAUGGAUU AGUGACAUCUAAUUUUGUGGGCAGCCAC UGAUAGAG 3'	Sac 23S rRNA U2692 + G2649	U2692 and G2649 (direct) 5'GCUACAUCGAUGGGG**U**AGGGCUGUCCCACCCUGGA GGUGCAGCUGCCUCCAAUCGG**C**UCUCCCGCC CGUUAAAG 3'
Pho sR24	5'CCCUAUGAUGAGAGAGCUGUUAAAUG AGCGGUGAUUAAAGGAUGGCUGGCUGAG GGUG 3'	Pho 23S rRNA	C1221 5'UAGGCUUAGAAGCAG**C**CAUCCUUUAAAGA 3' C1243 5'GGAGUGCGUAA**C**AGCUCACCC 3'
Pho sR24	5'CCCUAUGAUGAGAGAGCUGUUAAAUG AGCGGUGAUUAAAGGAUGGCUGGCUGAG GGUG 3'	Pho 23S rRNA C1221 + C1243	5'GAAGCAG**C**CAUCCUUUAAAGAGUGCGUAA**C**AGCUC AC 3'
Engineered sR1/ sR2	5'CAGAUGAUGAAUUCCCGAUAGUACGA UUGAUGAGCUAGGUCACUGCGAACUGAU UAG 3'	Sso/Sac 23S rRNA /C1914	5'UCAGUAGGUCG**C**AGUGACAAGAGGGACCU 3'
sR2	5'GAGUGAUGAGACGAGCGCUAACAGAG AGAGUGAAGAGGUCACUGCGAACUGAAG AAA 3'	Sso/Sac 23S rRNA /C1914	5'UCAGUAGGUCG**C**AGUGACAAGAGGGACCU 3'

2.2.8 *In vitro* methylation assay

Equimolar amounts (720 pmol) of guide RNA and target RNA were mixed (final volume 120 µl, in renaturing buffer R: 25 mM phosphate buffer, pH7/ 100 mM NaCl), denatured by incubating for 2 min at 95°C, and renatured by cooling rapidly to 55°C. The RNAs were added at 0°C to 75 µl containing aFIB, aNOP56, aL7a (6 pmol each), and tritiated-S-adenosyl methionine (360 pmol, 3.9 Ci/mmol, Amersham Pharmacia) in binding buffer D (25 mM phosphate buffer, pH 7, 100 mM NaCl and 1mM MgCl2). Aliquots (20 µl) were removed and transferred to 70oC in a Thermalcycler (Eppendorf);

after 3, 6, 10, 20 and 45 minutes the individual 20 µl samples were removed and precipitated at 0°C for 15 min with 5% trichloroacetic acid. The precipitates were collected on 0.2 µm nitrocellulose filters (Millipore), dried and radioactivity was measured by scintillation counting in a Beckman LS6000IC scintillation system. The scintillation cocktail used was ScintiVerse (Fisher Scientific), containing 1,2,4-trimethylbenzene.

2.2.9 Thin Layer Chromatography

Target and guide RNA were transcribed *in vitro* and mixed in a 20 µl reaction in the standard methylation assay (except that [methyl-^{3}H] SAM specific activity was 72.15 Ci/mmol; 60 pmol per assay) and incubated at 70°C for 1 h. The RNAs were extracted with phenol/chlorophorm from the reaction mixes, precipitated and resuspended in 4 µl of 50 mM Na actetate, pH 5.2 and digested with 0.01 unit of P1 nuclease (Roche Diagnostics) for 12 h at 37°C. Products were mixed with 5nmol each pAm, pCm, pGm, pUm, pUmG, pCmC and pCmA standards (Dharmacon Research, Lafayette CO) before two-dimensional TLC separation (Keith, 1995). TLC analysis was carried out on thin-layer cellulose plates (10 cm × 10 cm, Merck, Darmstadt) by using the following chromatographic system: first dimension, isobutyric acid/concentrated NH_4OH/H_2O (66:0.5:33.5; vol/vol/vol); second dimension 0.1 M sodium phosphate, pH 6.8/ammonium sulfate/*n*-propyl alcohol (100:60:2; vol/wt/vol). Unlabeled standard was detected by UV irradiation and ^{3}H labeled spots where identified by autoradiography using standard 2D-TLC maps as described (Keith, 1995).

2.2.10 Standard Molecular Biology Techniques

Polymerase Chain Reaction (PCR)

PCR was used to amplify genes from *S. acidocaldarius* and *S. solfataricus* genomic DNA. Reactions were either performed in a ERICOM Twin Block™ Thermocycler or an Eppendorf Mastercycler® gradient Thermocycler. Cycling parameters varied depending on the particular experiment; however, they usually included a 95°C denaturing step (45 sec), an annealing step at the theoretical Tm-5°C (30 sec - 1 min), and polymerase extension step (72°C for 1 min). The theoretical Tm was calculated as 4×(#G+C) nts + 2×(#A+T) nts. The annealing temperature used was the lowest Tm-5°C for the pair of oligonucleotides used. Taq polymerase (Invitrogen) or ThermalAce polymerase (Invitrogen) were used for amplification.

Isolation of plasmid DNA from *E.coli* cells

Plasmid DNA was essentially isolated using the procedure for Alkaline Plasmid Mini-Preparation as described in (Sambrook, 1989). Alternatively the QIAprep Spin Miniprep kit from Qiagen was used, following the recommendations of the manufacturer.

DNA Sequencing

Sequencing was done either using the Thermo Sequenase Radiolabeled Terminator Cycle Sequencing Kit (USB, Cleveland, OH, USA), following the manufacturer's recommendations or performed at the NAPS Unit (Nucleic Acid Protein

Service) at the University of British Columbia) with an Applied Biosystems PRISM 377 automated sequencer and Applied Biosystems BigDye™ v3.1 Terminator Chemistry .

2.2.11 Gel Electrophoresis

Native Gel

Agarose

Agarose slab gels (1%-2%) were routinely employed for resolving DNA fragments over 100 bp in length. Gels were poured and samples were electrophoretically separated using the Pharmacia GNA-200 Gel Electrophoresis Apparatus in 0.5× TBE, adding ethidium bromide to both gel and buffer for visualization of nucleic acid by UV light irradiation. Samples were electrophoresed at 100-150 V on a Pharmacia 500/400 Power Supply.

Non-denaturing Polyacrylamide

DNA fragments under 100 bp were resolved on 6-10% non-denaturing polyacrylamide gels, in 0.5×TBE buffer, using the BioRad Mini Protean II system. DNA fragments were visualized by ethidium bromide staining for 10 min.

Agarose and non-denaturing polyacrylamide gels were also used for the isolation and purification of restriction enzyme generated linear DNA fragments or for purification of oligonucleotides. The ethidium bromide stained bands were excised from the gel, gel slices put into Spectrum Spectra/Por ® molecular porous membrane dialysis tubing in 0.5×TBE buffer and electroeluted at 100 V for 1h-3 h, the duration depending on the size

of the fragment. The eluted DNA was extracted with phenol/chloroform, ethanol or isopropanol precipitated, dried and resuspended in dH_2O.

1kb and 100 bp Base-Pair Ladder (BioRad) were used as size markers.

Non-denaturing polyacrylamide gels were also used in EMSA and to verify the integrity of RNA transcripts. After resolving the RNA-protein complex in a 6-10% gel, gels were transferred to Whatman Filter Paper, covered with PVDC wrap (plastic wrap) and dried on the BioRad Model 583 Gel Dryer.

Perfect RNA markers (Novagen) for the 0.1-1 kb range and decade RNA markers (Ambion, TX) for the 10-100 nt range were used as size makers were appropriate. The markers were prepared following the manufacturer's directions.

Denaturing Gel

6% or 8% polyacrylamide gels containing 8M urea were used to separate DNA sequencing products. Samples were boiled for 2 min in sequencing sample buffer. Samples were separated by electrophoresis in 0.5×TBE buffer using a Pharmacia ECPS 3000/150 power Pack at 32 W constant power (1800-2000 V). Gels were transferred to Whatman Filter paper and dried on the BioRad gel Dryer.

SDS PAGE

Purified proteins were separated by size using SDS PAGE and visualized by Coomassie blue staining of the gel. Purified samples were diluted in SDS protein sample buffer (Sambrook, 1989) and heated at 90°C for 3 min prior to loading the gel. Gels were prepared according to the method of Laemmli using a 5% acrylamide stacking gel and

either 10, 12 or 15 % acrylamide separating gel, depending on the size range of the protein to be resolved (Laemmli, 1970). All gels were prepared using a BioRad 30% Acrylamide solution (Acrylamide/Bis ratio of 37.5:1). Gels were poured and protein samples were electrophoretically separated using the BioRad Mini Protean II Electrophoresis system run at a constant voltage of 100-150 V using Tris-Glycine gel running buffer. Following electrophoresis, gels were rinsed briefly with water, incubated for 5-10 min in destain solution. Gels were then stained for 10-20 min in stain solution followed by destaining.

Images of gels were obtained using DiaMed Transilluminator and Imager 7500 Digital Imager.

2.2.12 Visualization of radioactive gels

Radioactive gels were visualized using a Typhoon Imager 8600 (Molecular Dynamics) and ImageQuant image analyzing software.

2.2.13 Bioinformatics tools

RNA Secondary Structure Prediction

RNA secondary structure predictions were done using the program mfold (Zuker, 2003). MFold predicts optimal and suboptimal secondary structures for an RNA or DNA molecule using the most recent energy minimization method by Zuker. The web based tool can be accessed at: http://www.bioinfo.rpi.edu/applications/mfold/old/rna/form1.cgi .

BLAST search

BLAST (Basic Local Alignment Search Tool) was used to search for sequence similarity of DNA or protein in all available databases. BLAST uses a heuristic algorithm that seeks local as opposed to global alignments and is therefore able to detect relationships among sequences that share only isolated regions of similarity (Altschul et al., 1990). BLAST and variation of BLAST can be accessed at the NCBI (National Center for Biotechnology Information) web page at: http://www.ncbi.nlm.nih.gov/BLAST/ .

3 Results[1]

3.1 The *in vitro* methylation assay

3.1.1 Development of an *in vitro* methylation assay

Eukaryotic C/D box snoRNAs associate with four common core proteins; however, the total number of proteins in a particle is not known. The core proteins are Fibrillarin, Nop56p, Nop58p and 15.5 kD protein.

Fibrillarin has long been associated with C/D box snoRNAs in Eukaryotes and is generally accepted to be the 2'-O-methyltransferase (Tollervey *et al.*, 1993; Wang *et al.*, 2000). Nop56p and Nop58p are related to each other and both proteins have been shown to interact with snoRNAs (Cahill *et al.*, 2002; Filipowicz and Pogacic, 2002). The 15.5 kD protein belongs to a larger family of related RNA-binding proteins that includes human L7a, S12 and yeast L30 ribosomal proteins. (Mao *et al.*, 1999; Nottrott *et al.*, 1999; Watkins *et al.*, 2000). The 15.5 kD protein binds to the characteristic asymmetric stem-loop structure - the k-turn- that includes the canonical C/D elements (Watkins *et al.*, 2000).

In complex eukaryotic cells guide-containing RNP complexes are transient and unstable. Their molecular architecture has been difficult to establish or reconstitute. The best progress in purification and functional characterization of these complexes to date has

[1] "A version of this chapter has been accepted for publication. Ziesche, S.M., Omer, A.D. and Dennis, P.P. (2004) RNA guided Nucleotide Modification of Ribosomal and Non-ribosomal RNAs in Archaea, Mol. Microbiol."

involved purification of yeast C/D complexes that retained methylation activity (Galardi *et al.*, 2002). Are there simpler systems where the biochemical, structural and functional complexes of the RNP can be more easily addressed? The first demonstration that Archaea possess guide RNAs that direct methylation to rRNA came from biochemical studies with the archaeon *Sulfolobus acidocaldarius* to which I was a contributor (Omer *et al.*, 2000; Omer *et al.*, 2002).

Analysis of archaeal genomes has also revealed the presence of C/D box core protein coding sequences, including fibrillarin and Nop56/58 (Amiri, 1994; Omer *et al.*, 2000). A single archaeal aNop56 takes the place of the eukaryotic Nop56/Nop58 protein pair (Omer *et al.*, 2002; Tang *et al.*, 2002b). Notably absent has been a defined homolog of the 15.5 kD core protein. However, database searches in the *Sulfolobus solfataricus* genome and other archaeal genomes revealed a high degree of sequence similarity between the eukaryotic 15.5 kD protein and the archaeal protein aL7a. The *Sulfolobus* aL7a protein has been annotated as ribosomal protein. These observations suggest that the archaeal 2'-O-methylation complex might be simpler than its eukaryotic counterpart, with three proteins rather than four.

The *S. solfataricus* genes were cloned and recombinant protein were expressed in *E. coli* (see Material and Methods). The ability to form an RNP together with C/D box sRNA that is active *in vitro* in site-directed methylation of a fragment of target RNA was tested and the complex was active in targeting S-adenosyl methionine dependent, site-specific 2'-O- methylation (Omer *et al.*, 2002).

sR1 guide sRNA

Sulfolobus acidocaldarius sR1 sRNA (Sac sR1) was selected as the model sRNA for *in vitro* sRNP assembly. *S. acidocaldarius* sR1 sRNA contains well-defined and highly conserved C (AUGAUGA) and D (CUGA) box motifs located near the 5' and 3' ends, respectively (Figure 3-1). Internally, *S. acidocaldarius* sR1 contains C' and D' box motifs, giving the RNA a characteristic dyad repeat structure. The C/D motif can be structured according to the k-turn consensus. The C'/D' motif deviates from the consensus by forming a loop instead of stem I, and by exhibiting a Watson-Crick base pair CG, instead of a non-canonical base pair in stem II (Figure 3-1).

The D box guide of *S. acidocaldarius* sR1 sRNA recognizes a region of the 16S rRNA *in vivo*. Methylation occurs at position U52 in the 16S rRNA. Southern hybridizations have confirmed the existence of a single-copy *S. acidocaldarius* sR1 encoding sequence within *S. acidocaldarius* genomic DNA (Omer *et al.*, 2000).

S. acidocaldarius sR1 sRNA has a homologous sRNA in the related archaeon, *Sulfolobus solfataricus*, termed Sso sR1. *S. solfataricus* sR1 sRNA is also predicted to guide methylation to the same U52 position in 16S rRNA. In addition, the D' box of *S. solfataricus* sR1 sRNA is predicted to methylate position U605 and/or position U33 in 16S rRNA, although neither of these sites have been confirmed. Both *S. acidocaldarius* sR1 and *S. solfataricus* sR1 sRNAs overlap (by 21 and 18 nt, respectively), the C-terminal end of the aspartate aminotransferase gene coding region (Figure 3-1).

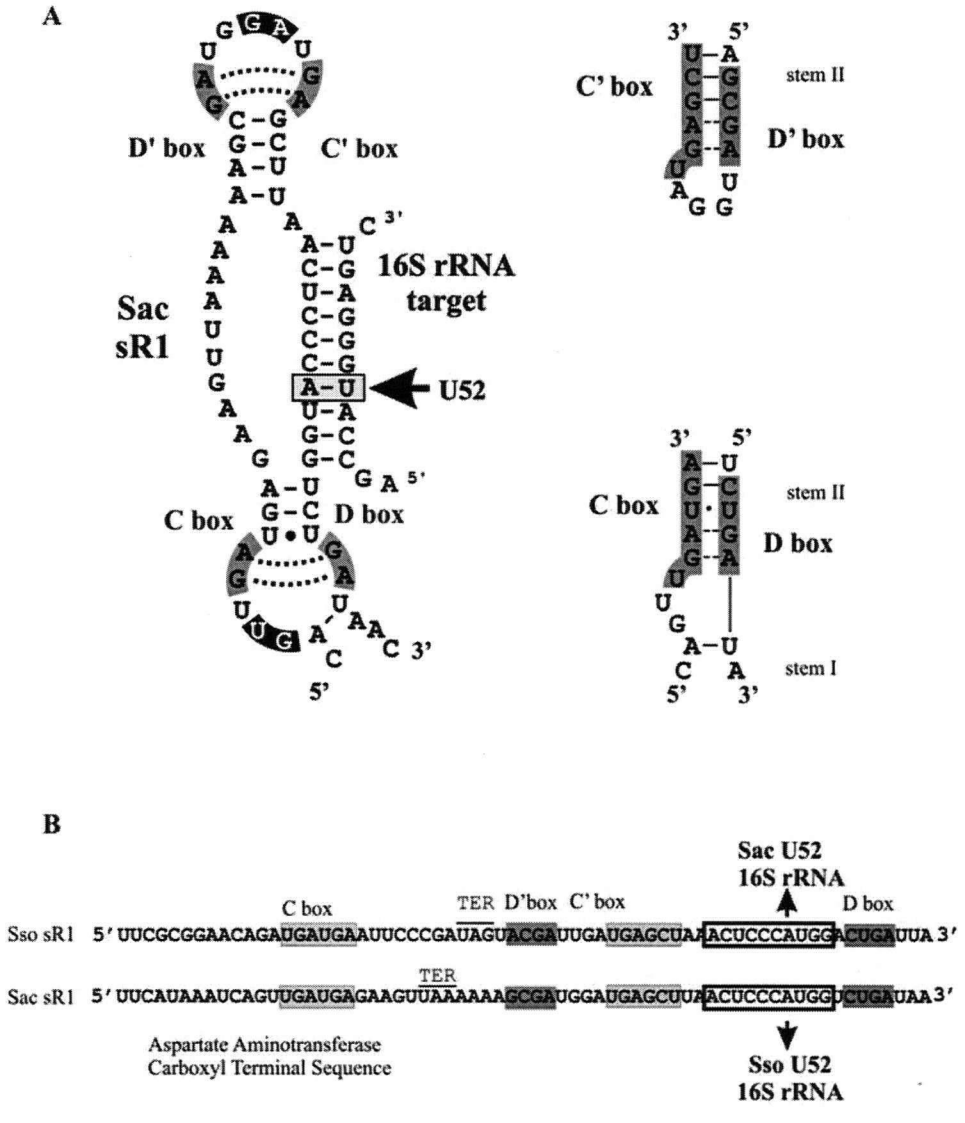


Figure 3-1 *S. acidocaldarius* sR1 sRNA

(A) A model of the secondary structure of *S. acidocaldarius* sR1 sRNA is depicted. The predicted k-turn consensus of the C/D and C'/D' motif is shown at the right. The C'/D' motif derivates from the consensus by forming a Watson-Crick base pair CG instead of a non-canonical base pair in stem II. (B) The RNA sequence derived from the 3' end of the aspartate amino transferase and the sR1 genes in *S. solfataricus* (Sso) and *S. acidocaldarius* (Sac) are illustrated The termination codon of the aspartate amino transferase is overlined. The conserved boxes of the sRNA are shaded. The D guides that direct methylation to position U52 in both 16S rRNAs is boxed. The D' guide of Sso sR1 appears to direct methylation to position 16S U33 (not shown), whereas the D'guide of Sac sR1 lacks complementarity to rRNA.

The most highly conserved portion of the sRNA genes is in the 3'-half that lies outside the ORF. It contains the D box antisense element. The 5'-half of sR1 sRNA gene is poorly conserved which is reflected in numerous amino acid replacements in the C-terminus of the aspartate aminotransferase protein. The translation termination codons are both located within the complementary guide region associated with the D' box (Dennis *et al.*, 2001; Omer *et al.*, 2000).

Target

The target RNA (29 nucleotides) was designed to contain the region of *Sulfolobus* 16S rRNA that is recognized *in vivo* by the D box guide of the *S. acidocaldarius* sR1 sRNA; methylation occurs at position U52. Methylation at the predicted U52 positions in *S. solfataricus* and *S. acidocaldarius* 16S rRNA has been verified by dNTP concentration-dependent primer extension (Omer *et al.*, 2000). In this assay, reverse transcription shows characteristic pausing at most 2'-O-methylation sites in an dNTP concentration-dependent manner, a pause observed at the predicted site of methylation, indicating a likely 2'-O-ribose methyl (Maden *et al.*, 1995).

***In vitro* methylation assay**

The *S. acidocaldarius* sR1 sRNP complex was assembled *in vitro* by incubating *in vitro* transcribed *S. acidocaldarius* sR1 sRNA, with recombinant *S. solfataricus* sRNP core proteins aL7a, aNop56, and aFib. In the presence of target RNA and SAM

(radiolabeled at the methyl position), radioactivity was incorporated into acid-insoluble material (Figure 3-2).

3.1.2 Optimization of the assay

Reaction incubation temperature, time, concentration of the three core proteins (aL7a, aNop56 and aFib), and concentration of guide and target RNAs were varied in order to find the optimal *in vitro* assay conditions (Figure 3-2). Different reaction temperatures were used (37°C, 42°C, 55°C, 65°C, 70°C, 75°C). The reaction requires elevated temperatures around 70°C (data not shown). The elevated temperature might induce conformational changes in the core proteins and/or the sRNA that are critical for protein binding. Assuming that a sufficient concentration of suitable substrate is available, increasing the concentration of the enzyme will increase the reaction rate. The initial rate of the methylation reaction is linear (as a function of protein concentration) and reaches a plateau at protein concentrations higher then 6 pmoles (Figure 3-3); this observation suggests other factors became limiting at high protein concentrations.

It was determined that a reaction containing equimolar amounts (120 pmol) of guide RNA and target RNA, and 6 pmol of each of the recombinant protein aFIB, aNop56, aL7a produced the best results, by maximizing the activity and product formation per pmol of protein used. In previous experiments, lower protein concentrations were used. The reaction was carried out at 70°C in a Thermalcycler with a heated lid to minimize condensation of the reaction mix. The methylation depends both on the addition of a suitable complementary target RNA (data not shown) and on the presence of aFib in the complex. In additional control reactions where only RNAs and

either single proteins, or combinations of only two proteins were added, radioactive methyl incorporation into acid-insoluble material was not detected (data not shown).

Within the limits of the determination of protein concentration, the amount of product formed (about 12 pmol per reaction) was approximately two times the molar amount of aFib present (6 pmol per reaction). This implies that under these reaction conditions, each molecule of aFib was, on average, able to participate in the methylation of two target RNAs. In contrast, when the amount of protein was doubled (12 pmol of aFib per reaction), only a slight increase in product production could be detected (14 pmol per reaction), implying other limiting factors in the reaction, such as misfolded and thus unproductive RNA.

The linear incorporation of labeled methyl groups reached a plateau after 20-30 min (Figure 3-2). This plateau appears to be largely the consequence of the degradation of the RNA and SAM cofactor at 70°C required for catalysis, rather than a consequence of inactivation of the three proteins. Supplementation of the reaction after 45 min with additional guide and target RNAs has been shown to result in a 50-100% increase in production of methylated target, whereas supplementation with the three proteins has no effect (data not shown). The mechanism for enzymatic turnover is currently not clear. Either the RNP complexes are disassembled after each methylation and reassembled before the next round of methylation, or multiple target RNA are able to associate and disassociate from single preassembled RNP complex. The supplementation experiments suggest that some reassembly occurs.

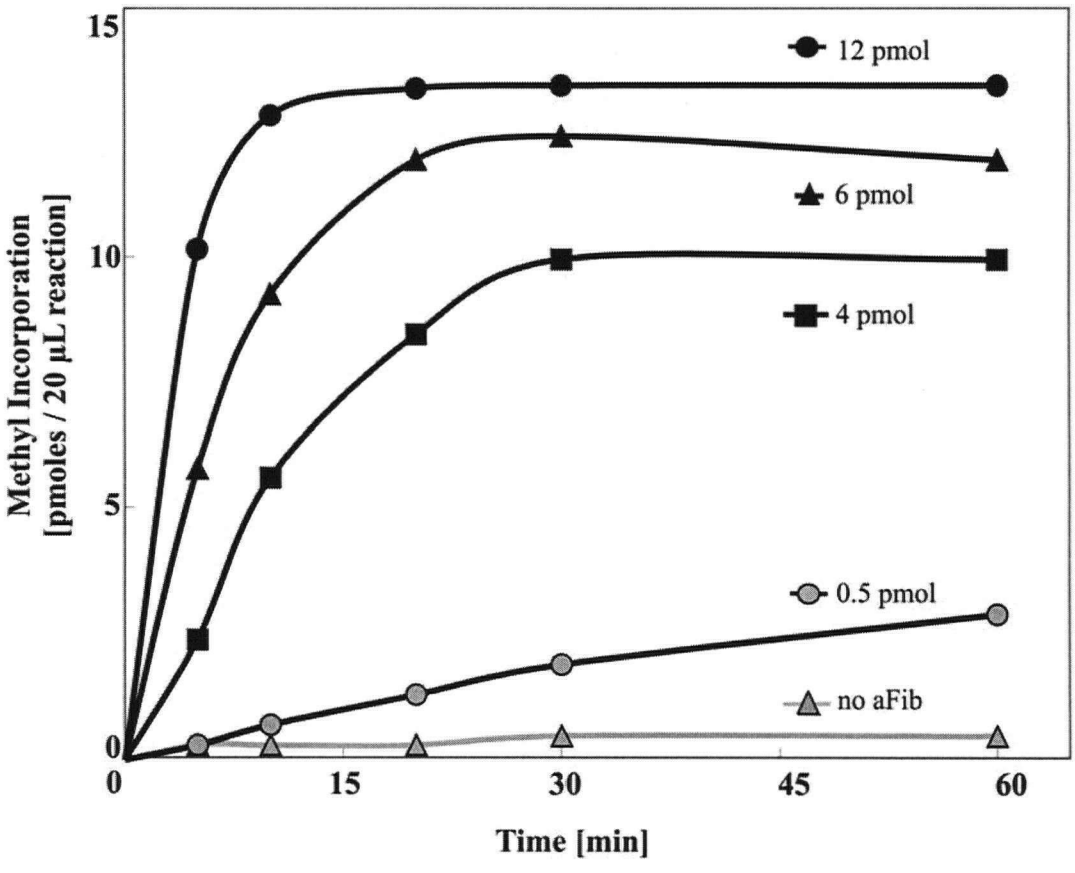


Figure 3-2 Effect of protein concentration on the efficiency of the methylation reaction

RNP guide-dependent methyl incorporation into U52 of 16S rRNA is illustrated. The RNP complex was assembled by renaturing *in vitro* transcribed *S. acidocaldarius* sR1 sRNA with target RNA (Table 2-5) and mixing various amount of aFIB, aNop56, and aL7a: (0.5 pmol, 4 pmol, 6 pmol, 12 pmol) and [^{3}H]-methyl-SAM at 0°C. The mixture (120 µl containing 720 pmoles of sR1 guide RNA, 720 pmoles of target RNA, 360 pmoles of radioactive SAM, and 3, 24, 36, 72 pmoles of each of the three recombinant proteins) was divided into five 20 µl reactions and transferred to 70°C. At various time intervals, single 20 µl reactions were removed and precipitated at 0°C with 5% trichloroacetic acid. The precipitates were collected on nitrocellulose filters, dried, and radioactivity was determined by scintillation counting.

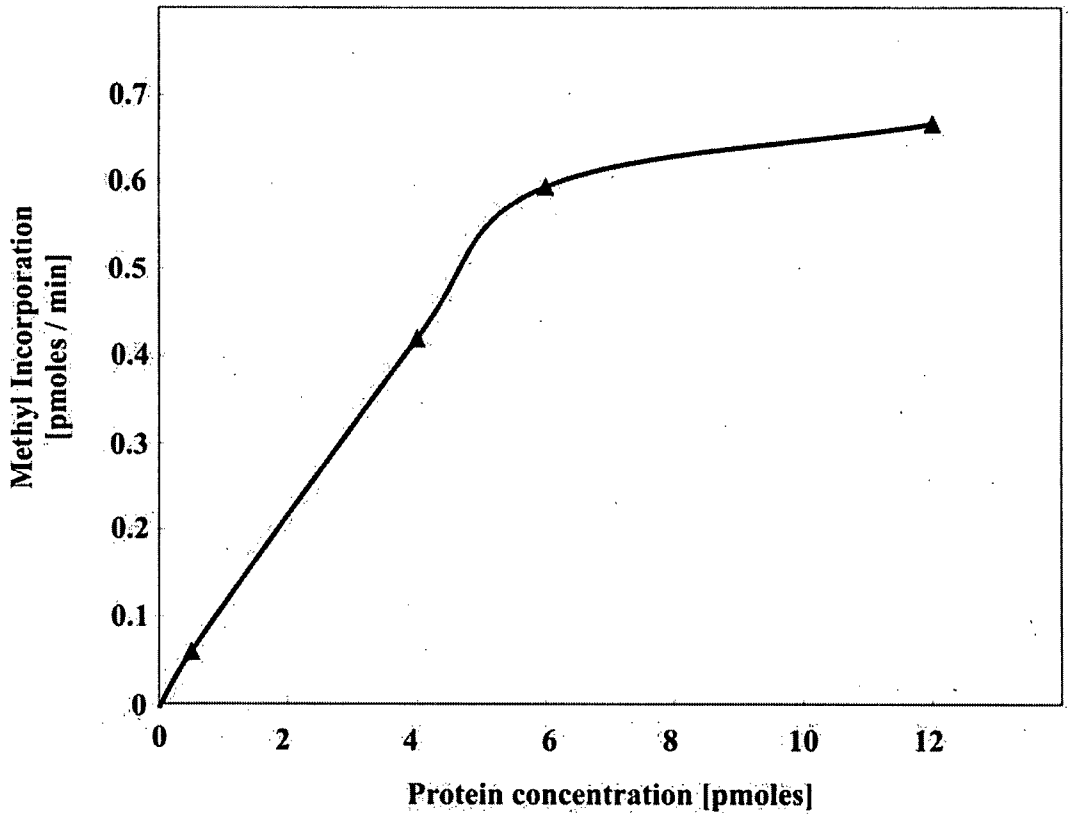


Figure 3-3 Initial rates as function of protein concentration

The initial rates (pmoles/min) of the methylation reaction (U52) as a function of protein concentration (equal amounts of the three proteins aFib, aL7a and aNop56) are illustrated. Initial rates are linearly dependent on the protein concentration. A plateau is reached at concentrations higher then 6 pmoles of proteins.

3.1.3 Specificity of the methylation reaction

Based on the N plus five rule, the site of methyl modification within the target RNA is predicted to occur at the nucleotide position corresponding to U52 in 16S rRNA (Lafontaine and Tollervey, 1998; Maxwell and Fournier, 1995). This residue within the target RNA is part of a unique UG dinucleotide. To show that methylation is specific to the U52 position, target RNA was incubated in the presence of [^{3}H]-methyl SAM in the standard methylation assay and then subjected to digestion with nuclease P1. Nuclease P1 has a single-strand-specific endonuclease activity, but only slowly digests phosphodiester bonds adjacent to residues methylated at the ribose under conditions where non methylated nucleotides are completely hydrolyzed. The hydrolysis products were mixed with 2'-O-ribose methylated NMPs (pAm, pCm, pGm and pUm) and pUmG standards, and separated by two-dimensional thin-layer chromatography (Figure 3-4).

The position of individual standards was observed by UV shadowing and superimposed on the image obtained by tritium imaging. The tritium radioactivity was uniquely found in the pUm and pUmG spots, as further confirmed by [^{3}H]-radioactivity scintillation counting of all individual UV-detectable spots (pUm, 1129 cpm, pUmG, 5955 cpm, pAm, 23 cpm, pCm, 37 cpm and pGm, 28 cpm,). Omission of any of the RNA or protein components from the reaction mixture resulted in background levels of incorporation (data not shown).

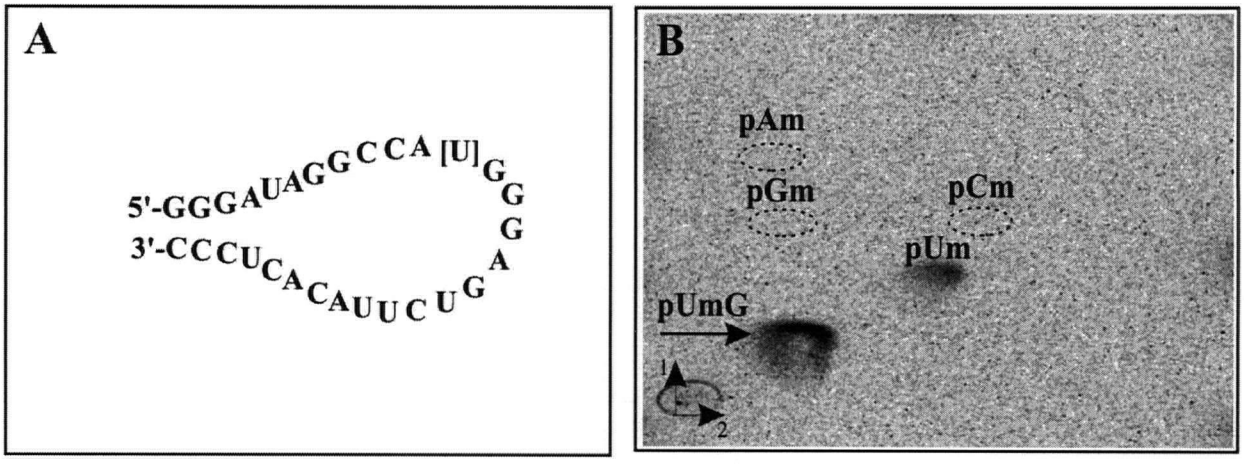


Figure 3-4 Two-dimensional thin-layer chromatography

Thin-layer chromatographic separation of the hydrolysis products of the target RNA is shown. (A) The standard 29 nt long target RNA contains a single unique UG dinucleotide. The U residue in this dinucleotide is the expected site of methylation and corresponds to position U52 in 16 S rRNA. (B) Guide and target RNAs were mixed in a 20 µl reaction in the standard methylation assay using [methyl-^{3}H]SAM and incubated at 70°C for 1 h. The RNA was extracted, digested with 0.01 units of P1 nuclease for 12 h at 37°C, and products were mixed with 5 nmol each of pAm, pCm, pGm, pUm, and pUmG standards before the two-dimensional TLC separation. Unlabeled standards were detected by UV shading, and radioactivity was detected by ^{3}H-imaging. The spots corresponding to the UV-detectable standards were excised from the plate and each subjected to scintillation counting (pUm, 1129 cpm, pUmG, 5955 cpm, pAm, 23 cpm, pCm, 37 cpm and pGm, 28 cpm,). Omission of any of the RNA or protein components from the reaction mixture results in background level of incorporation of radioactivity into acid-insoluble material.

3.2 Non-ribosomal target RNA

3.2.1 tRNA methylation

3.2.1.1 Archaeal sRNAs guide methylation to tRNAs

In many archaeal sRNAs, one or both of the D' or D box guide regions lack complementarity to 16S and 23S rRNA, but exhibit complementarity to tRNA. In all cases, the predicted position of methylation within the tRNA corresponds to one of the 22 sites of documented ribose methyl modification, and never to the remaining positions where ribose methyl modification has never been observed (Figure 3-5).

To determine if unmodified tRNA could serve as a substrate for guide-directed methylation *in vitro*, I used sR11 sRNA from *S. solfataricus*, and sR14 sRNA from *S. acidocaldarius* to investigate modification of *Sulfolobus* tRNA (the tRNA sequences from *S. acidocaldarius* and *S. solfataricus* are identical in the target region of the guide RNAs).

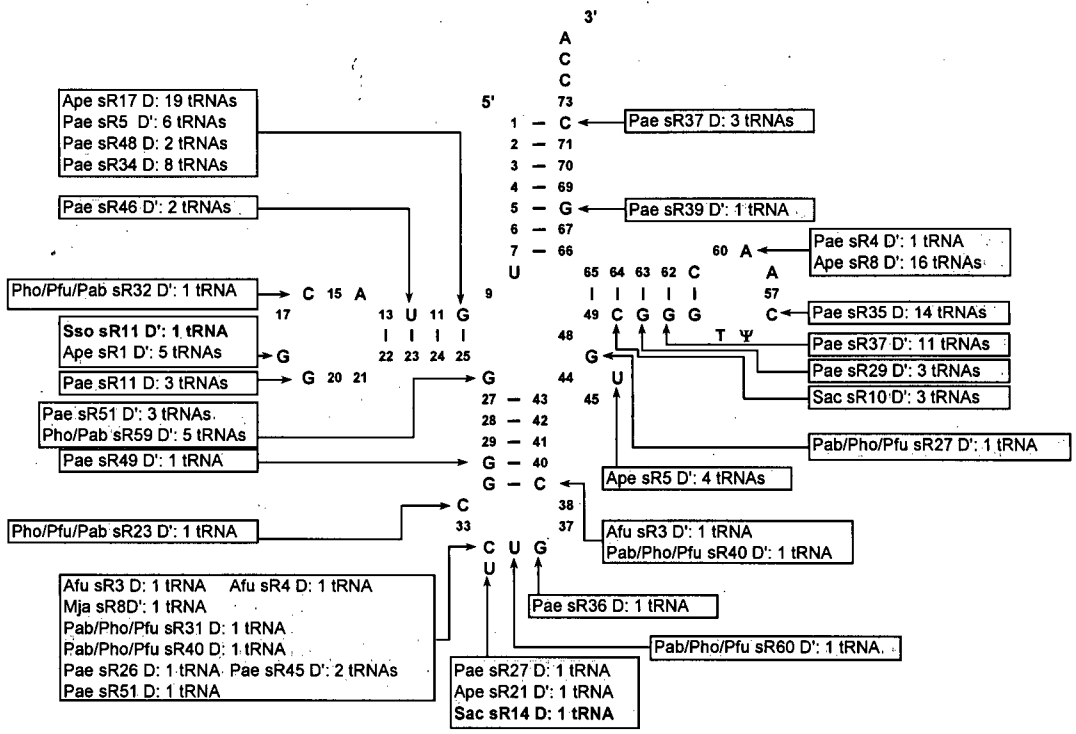


Figure 3-5 Predicted methylation sites in tRNA of Archaea

Predicted methylation sites in archaeal tRNAs are shown. The positions predicted to be methylated are based on sRNA guide sequence complementarity. The corresponding guide sRNA are shown in boxes, with archaeal species being indicated by three letter abbreviations (*Ape*: *Aeropyrum pernix*; *Afu*: *Archaeoglobus fulgidus*; *Pae*: *Pyrobaculum aerophilum*; *Pab*: *Pyrococcus abyssi*; *Pfu*: *Pyrococcus furiosus*; *Pho*: *Pyrococcus horikoshii*; *Mja*: *Methanocaldococcus jannaschii* (formally in the genus *Methanococcus*); *Sac*: *Sulfolobus acidocaldarius*; *Sso*: *Sulfolobus solfataricus*). sRNAs (*Sso* sR11 and *Sac* sR14) used in this study are shown in bold. The structure of tRNA is depicted in the standard cloverleaf configuration. Conserved tRNA nucleotides are indicated as letters. Variable nucleotides are indicated by numbers only.

3.2.1.2 D' guide of *S. solfataricus* sR11 guides methylation of G18 in $tRNA^{Gln}$

In *S. solfataricus*, the sR10 and sR11 sRNA genes are co-transcribed and generated by nucleolytic processing (A.D. Omer, unpublished). The D box of *S. solfataricus* sR10 sRNA is predicted to guide methylation of residue A2082 of the 23S rRNA while the D' box guide of sR11 is predicted to guide methylation to position G18 in $tRNA^{Gln}$. The C/D and the C'/D' motifs of sR11 sRNA can be structured according to the k-turn consensus motif, with the exception that the C'/D' motif contains a non-canonical GA pairing in Stem II rather than a typical Watson-Crick pairing. In addition, *S. solfataricus* sR11 sRNA can form a 6 base pair terminal stem (Figure 3-6).

The G18 residue is highly conserved in all tRNAs (Dirheimer, 1995; Giege *et al.*, 1993), and is responsible for the formation of the L-shaped three dimensional structure by D-loop/T-loop interaction through the tertiary base pairs G18-Ψ55 and G19-C56 (Kim, 1974; Robertus, 1974).

To determine if the *S. solfataricus* sR11 sRNA could direct methylation to position G18 of a full length $tRNA^{Gln}$ target, the sRNA was assembled along with aL7A, aNop56 and aFib into an RNP complex.

In the presence of radioactive SAM and unmodified $tRNA^{Gln}$, the assembled complex was able to direct methylation as evidenced by the incorporation of radioactivity into acid insoluble material (Figure 3-7). In contrast, the complex was inactive on a G18A mutant form of $tRNA^{Gln}$. Similarly, a complex assembled using a C to U mutant at position five in the D' box guide of sR11 was also inactive on wildtype $tRNA^{Gln}$.

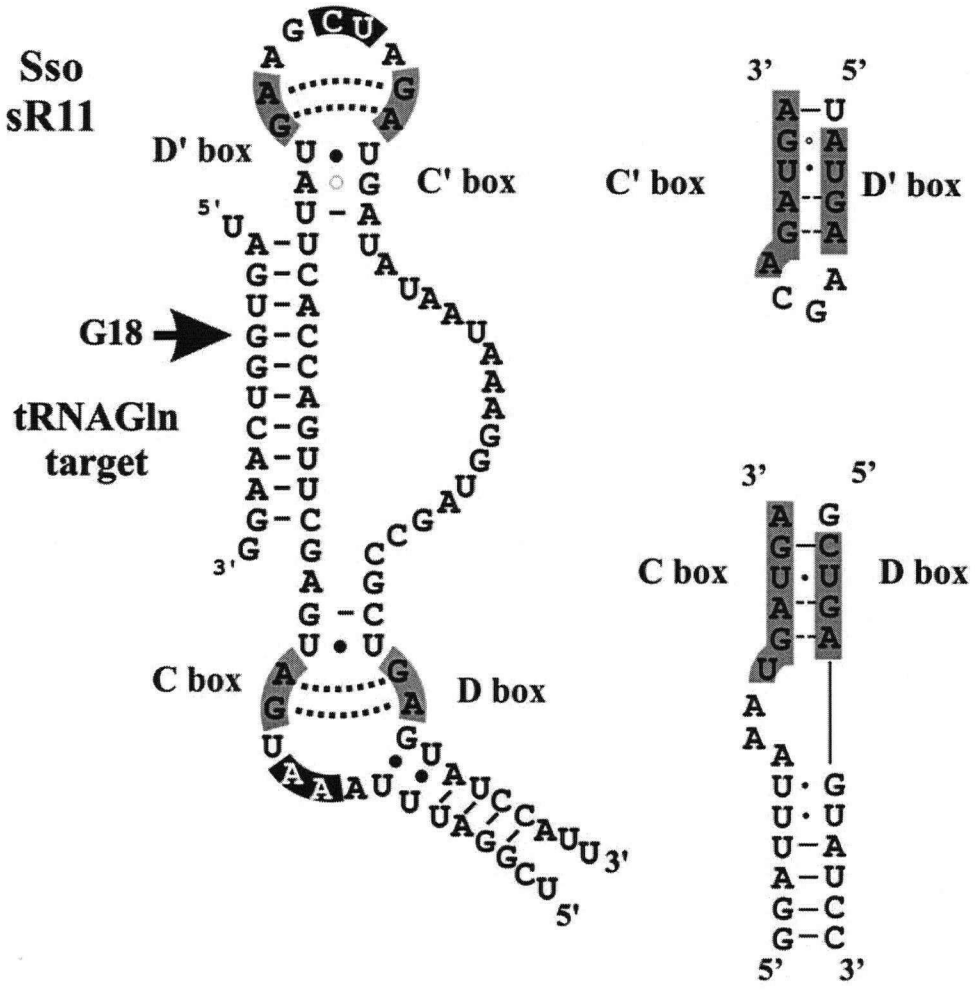


Figure 3-6 *S. solfataricus* sR11 sRNA

A secondary structure model of *S. solfataricus* sR11 sRNA is depicted. The predicted k-turn consensus of the C/D and C'/D' motif is shown on the right. The D' box guide shows a ten base pair complementarity to the D-loop region of $tRNA^{Gln}$. The residue G18 is predicted to be methylated.

However, when a canonical base pair was reestablished at the site of methylation by combining the C to U mutation in *S. solfataricus* sR11 sRNA guide with the G18A mutation in the $tRNA^{Gln}$ target, nearly full activity was recovered. This suggests that the sR11 RNP is capable of directing accurate methylation to position G18 of unmodified full length $tRNA^{Gln}$. The amount of product formed in the sR1 sRNA control reaction was lower (about 4-5 pmol per reaction) than anticipated. The quantity of product for sR11 and tRNA(G18) and the compensatory mutation reaction were slightly lower (4 and 3 pmol per reaction respectively). This might be due to the quality of the purified aNop56 protein used in the reaction.

There is no evidence to date that any of the numerous nucleotide modifications in eukaryotic tRNA are snoRNA-directed. It has been shown that ribose methylation increases base stacking, suggesting that at high temperature, archaeal pre-tRNAs may assume the correct tertiary structure only after ribose methyl modification by the guide-RNA- directed mechanism.

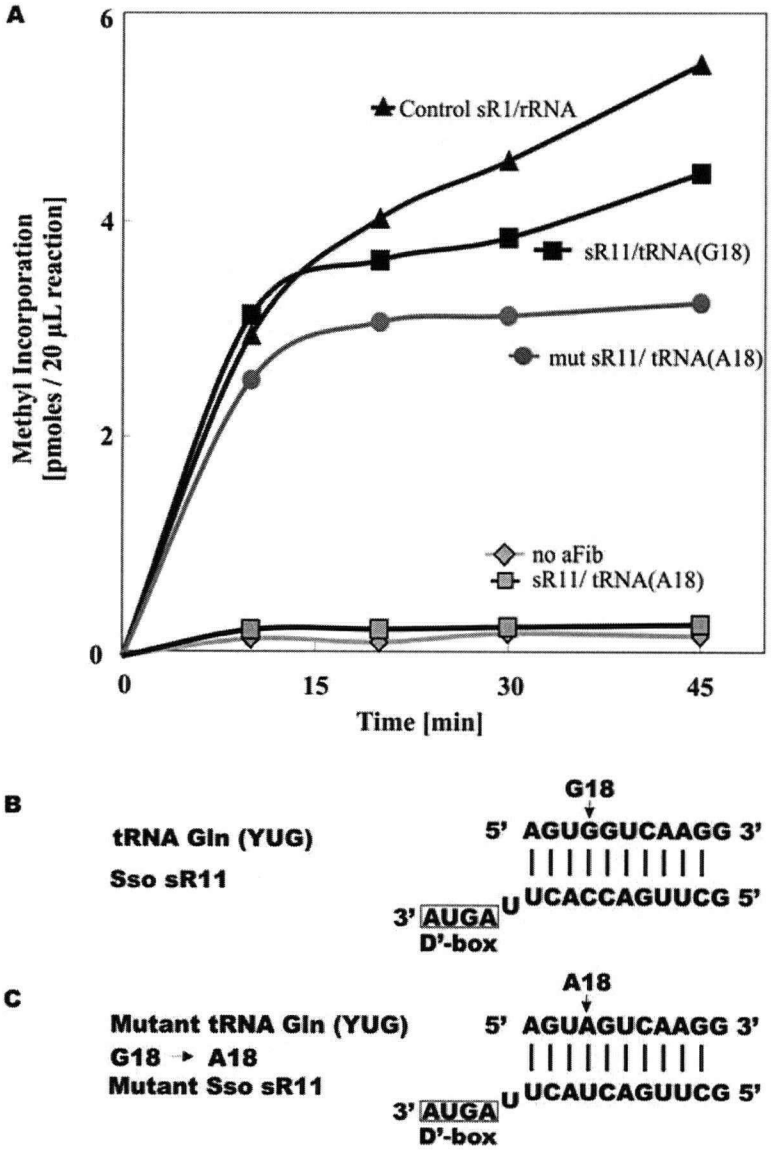


Figure 3-7 (see next page for legend.)

Figure 3-7 RNP guide-dependent methyl incorporation into G18 of $tRNA^{Gln}$

(A) RNP complex was assembled by renaturing *in vitro* transcribed *S. solfataricus* sR11 sRNA and $tRNA^{Gln}$ (Table 2-5) and mixing with aFIB, aNop56, aL7a, and [^{3}H]-methyl-SAM at 0°C. The mixture (120 µl containing 720 pmoles of sR11 guide RNA, 720 pmoles of $tRNA^{Gln}$, 360 pmoles of radioactive SAM, and 24 pmol of each of the three recombinant proteins) was divided into five 20 µl reactions and transferred to 70°C. At various time intervals, single 20 µl reactions were removed and precipitated at 0°C with 5% trichloroacetic acid. The precipitates were collected on nitrocellulose filters, dried and radioactivity was determined by scintillation counting. Nucleotide substitutions were introduced into the *S. solfataricus* sR11 D box guide (C to U at the N plus five position relative to the start of the D box) and the $tRNA^{Gln}$ target (G to A at position corresponding to G18). Methylation assays, as described before, were carried out using (i) wild-type guide and mutant target and (ii) compensatory mutant guide and mutant target. As a control reaction *S. acidocaldarius* sR1 and target rRNA with and without aFib protein in the reaction were used. The structure of *S. solfataricus* sR11 guide RNA duplex with $tRNA^{Gln}$ target RNA is shown in (B). The structure of the mutant guide and the mutant target is shown in (C). The tRNA nucleotide paired to the fifth nucleotide upstream from the D' box is denoted by an arrow with indication of the position in mature tRNA using the universal numbering system for tRNAs.

3.2.1.3 *S. acidocaldarius* sR14 guides methylation to the wobble base in $tRNA^{Gln}$

Many tRNA species have a 2'-O-ribose methylated nucleoside at the first (wobble) position of the anticodon; this modification is believed to stabilize the codon-anticodon interaction (Satoh *et al.*, 2000). A comprehensive examination of archaeal sRNAs has revealed 16 instances of potential methyl modification at position 34 of the anticodon in various tRNAs; 13 of the predicted modifications are at C nucleotides and three are at U nucleotides (Dennis *et al.*, 2001). For example, the D box guide in *S. acidocaldarius* sR14 sRNA (Sac sR14) is predicted to methylate the wobble base U34 in $tRNA^{Gln}$ (UUG), but not the corresponding position C34 in the isoacceptor $tRNA^{Gln}$ (CUG) (Figure 3-8).

To test the ability of *S. acidocaldarius* sR14 sRNA to guide methylation of tRNA at position U34, the guide RNA was assembled into an RNP complex and provided with radiolabeled SAM and an *in vitro* transcript of $tRNA^{Gln}$ (UUG). Radioactivity was efficiently incorporated into acid insoluble material, indicating that the complex was active in directing methylation to a full length tRNA substrate (Figure 3-9).

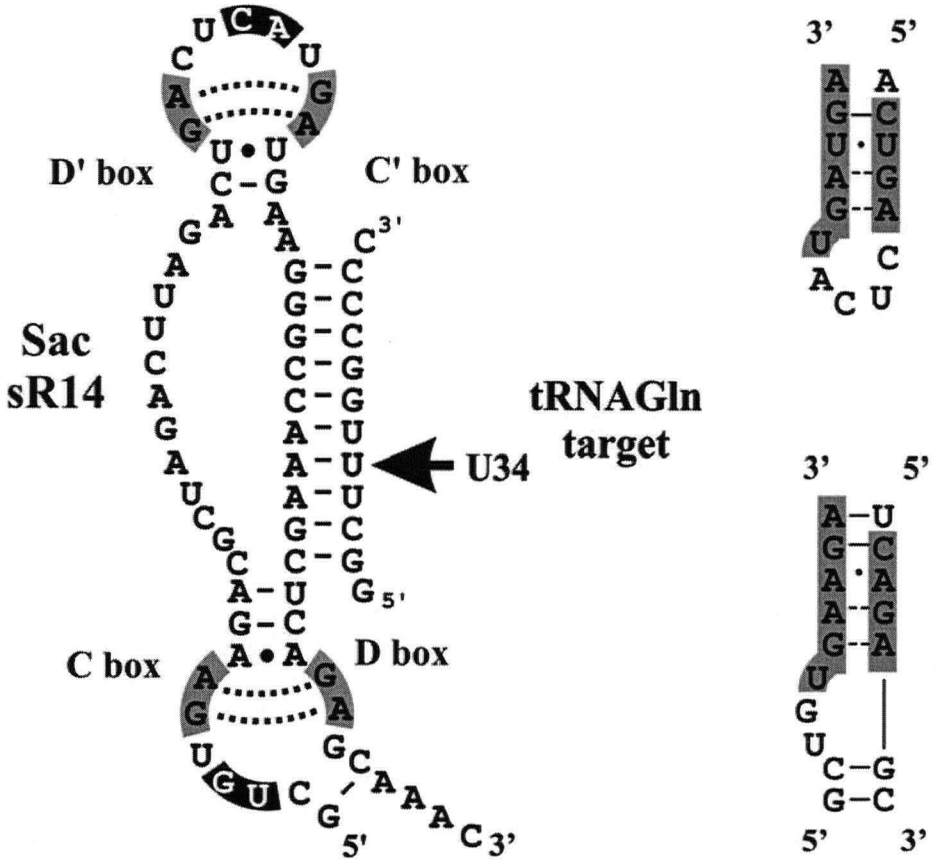


Figure 3-8 *S. acidocaldarius* sR14 sRNA

A secondary structural model of *S. acidocaldarius* sR14 sRNA is depicted. The predicted k-turn consensus of the C/D and C'/D' motif is shown on the right. The D box guide displays a ten base pair complementarity to the anticodon region of $tRNA^{Gln}$ (UUG). The wobble base U34 is predicted to be methylated.

In contrast, either an A to G mutant at position 5 of the D box guide of *S. acidocaldarius* sR14 sRNA or the use of the isoacceptor $tRNA^{Gln}$ (CUG) as target, reduced incorporation to background levels. Activity was fully restored by re-establishing a canonical base pair at the site of methylation using the *S. acidocaldarius* sR14 A to G mutant with the $tRNA^{Gln}$ (CUG) target.

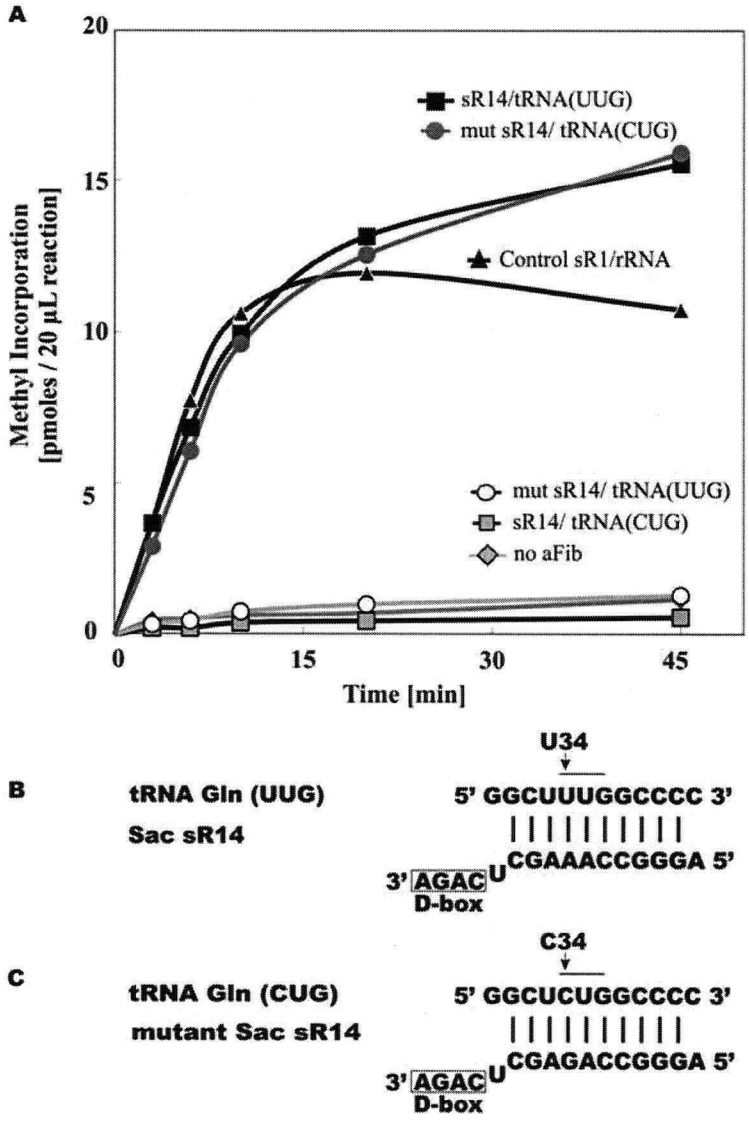


Figure 3-9 (see next page for legend.)

Figure 3-9 RNP guide-dependent methyl incorporation into $tRNA^{Gln}$ anticodon.

(A) RNP complex was assembled by renaturing *in vitro* transcribed *S. acidocaldarius* sR14 sRNA and $tRNA^{Gln}$(UUG) and mixing with aFIB, aNop56, aL7a and [^{3}H]-methyl-SAM at 0°C. The mixture (120 µl containing 720 pmoles of sR14 guide RNA, 720 pmoles of $tRNA^{Gln}$, 360 pmoles of radioactive SAM and 36 pmol of each of the three recombinant proteins) was divided into five 20 µl reactions and transferred to 70°C. At various time intervals, single 20 µl reactions were removed and precipitated at 0°C with 5% trichloroacetic acid. The precipitates were collected on nitrocellulose filters, dried, and radioactivity was determined by scintillation counting. A Nucleotide substitution was introduced into the *S. acidocaldarius* sR14 D box guide (A to G at the N plus five position relative to the start of the D box) and $tRNA^{Gln}$ (CUG) was used as tRNA target. Methylation assays, as described were carried out using (i) sR14 wildtype guide and $tRNA^{Gln}$ (UUG) target, (ii) *S. acidocaldarius* sR14 mutant guide and $tRNA^{Gln}$ (CUG) target, (iii) sR14 wildtype guide and $tRNA^{Gln}$ (CUG) target, and (iv) *S. acidocaldarius* sR14 mutant guide and $tRNA^{Gln}$ (UUG) target and (v) *S. acidocaldarius* sR1 and rRNA target as control with and without aFib protein in the reaction. The incorporation kinetics for the six reactions is illustrated. The structure of *S. acidocaldarius* sR14 guide RNA duplex with $tRNA^{Gln}$ (UUG) target RNA is shown in (B). The structure of the mutant guide and $tRNA^{Gln}$ (CUG) target is shown in (C). The tRNA nucleotide paired to the fifth nucleotide upstream from the D box is denoted by an arrow with indication of the position in mature tRNA using the universal numbering system for tRNAs. The anticodon is overlined.

3.3 Full length ribosomal target

3.3.1 5S rRNA methylation

5S rRNA is an integral component of the large ribosomal subunit found in all known organisms, with the exception of the small ribosomes of fungal and animal mitochondria (Szymanski *et al.*, 2002). 5S rRNA is normally the smallest RNA constituent of a ribosome and enhances protein synthesis by stabilization of the ribosome structure; its exact role in protein synthesis is still not known.

In *S. acidocaldarius* the D' box guide of sR10 exhibits a short interrupted eight base pair complementarity to the region spanning residue C32 of 5S rRNA, with a conspicuous mismatch at position +9 of the D' box guide (Figure 3-10).

Oligonucleotide fingerprinting and MALDI mass spectrometry have shown that the 5S rRNA of *S. acidocaldarius* is modified at position C32 (Kirpekar *et al.*, 2000). To determine if *S. acidocaldarius* sR10 sRNA might be involved in this modification, I assembled an RNP complex and tested its ability to guide methylation to full length 5S rRNA.

3.3.2 Full length 5S rRNA

After combining full length 5S rRNA transcript, *S. acidocaldarius* sR10 sRNA, and the three proteins, radioactive methyl incorporation into acid-insoluble material was not detected above the control reaction that contained *S. solfataricus* sR1 sRNA, target RNA and proteins with the aFib protein omitted (Figure 3-11).

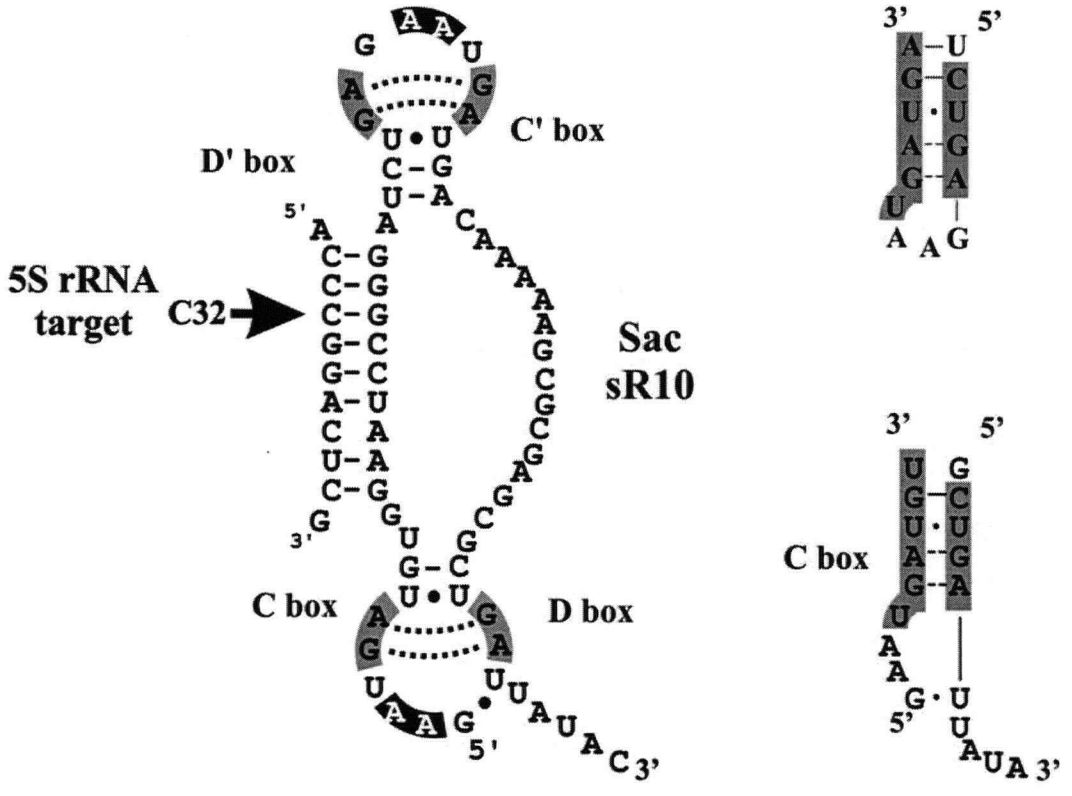


Figure 3-10 *S.acidocaldarius* sR10 sRNA

A secondary structural model of *S. acidocaldarius* sR10 sRNA is illustrated. The predicted k-turn consensus of the C/D and C'/D' motif is shown at the right. The D' box guide exhibits a short eight base pair complementarity to the region spanning residue C32 region of 5S rRNA, with a mismatch at position +9 of the D' box guide.

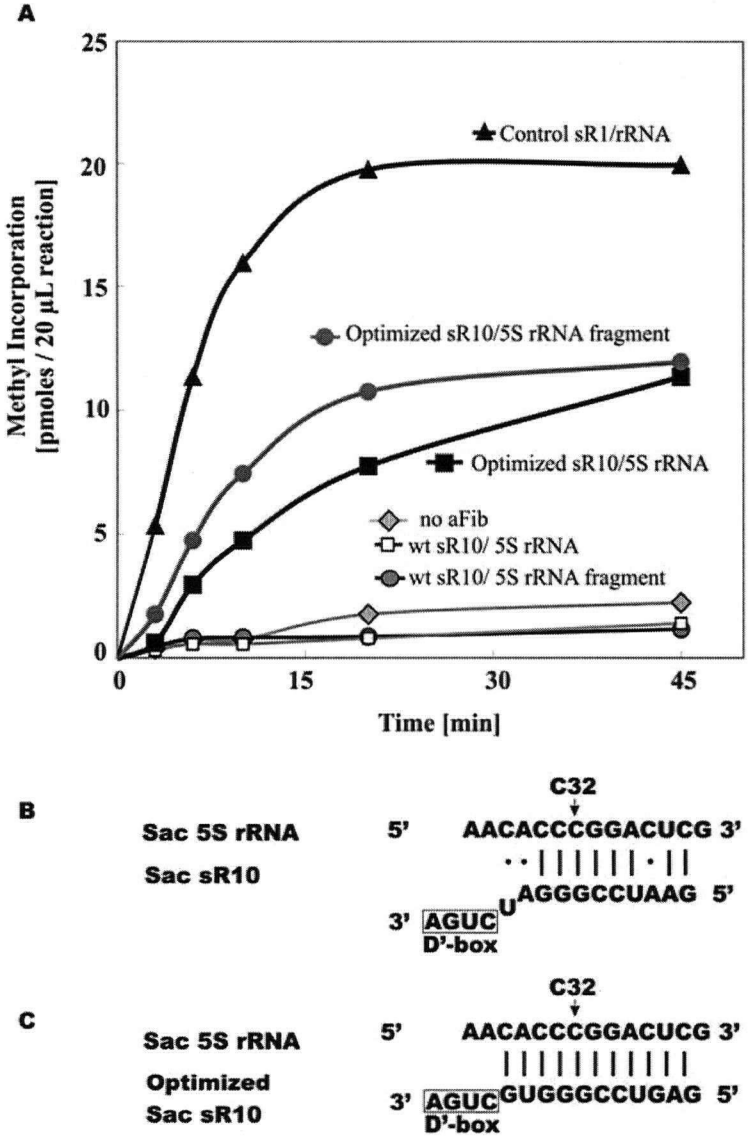


Figure 3-11 (see next page for legend.)

Figure 3-11 RNP guide-dependent methyl incorporation into 5S rRNA

(A) Methylation assays were carried out as described in the legend to Figure 3-9 . The following guides and targets (Table 2-5) were used: (i) *in vitro* transcribed *S. acidocaldarius* sR10 sRNA and full length 5S rRNA from *S. acidocaldarius*, (ii) *in vitro* transcribed *S. acidocaldarius* sR10 sRNA and 30 nt long 5S rRNA fragment containing the region of C32, (iii) *in vitro* transcribed optimized sR10 (nucleotide substitutions were introduced at position +1, + 2 and +9 upstream of the D' box), and full length 5S rRNA, and (iv) optimized sR10 and 30 nt long 5S rRNA fragment. *S. acidocaldarius* sR1 sRNA and rRNA target were used as controls with and without aFib protein in the reaction. The incorporation kinetics for the six reactions are illustrated. (B) The structure of *S. acidocaldarius* sR10 guide RNA duplex with *S. acidocaldarius* 5S rRNA target RNA is shown. Mismatches at position +1, +2 and +9 relative to the D' box in 5S rRNA are indicated by bullets.(C) The structure of the optimized sR10 and 5S rRNA target is shown.

3.3.2.1 Secondary structure prediction of *S. acidocaldarius* 5S rRNA

The secondary and three-dimensional architecture of the 5S rRNA substrate might influence the efficiency of the methylation reaction. 5S rRNA molecules have a high degree of predicted secondary structure and stability. To predict where in the structure of 5S rRNA residue C32 was situated, the secondary structure of 5S was examined. The secondary structure of *S. acidocaldarius* 5S rRNA has been reported previously by Shen and Hagerman (Shen and Hagerman, 1994). However, a closer investigation of the sequence used in this study led me to believe that their identification of the strain as *S. acidocaldarius* was probably incorrect and that instead the 5S rRNA of *S. solfataricus* had been examined. There has been considerable confusion about the specific identification of *Sulfolobus* strains employed in studies in the past. It had been reported that the 5S rRNA from the two species *S. acidocaldarius* and *S. solfataricus* are identical (Dams *et al.*, 1983; Stahl *et al.*, 1981). However, the sequences of 5S rRNA of *S. acidocaldarius* and *S. solfataricus* were later found to differ at 17 of 124 positions (Durovic *et al.*, 1994).

To predict the secondary structure of 5S rRNA from *S. acidocaldarius* the program mfold (Zuker, 2003) was used. In addition, the nucleotide sequence of 5S rRNA is highly conserved throughout nature, and phylogenetic analysis alone provided an initial model for its secondary structure.

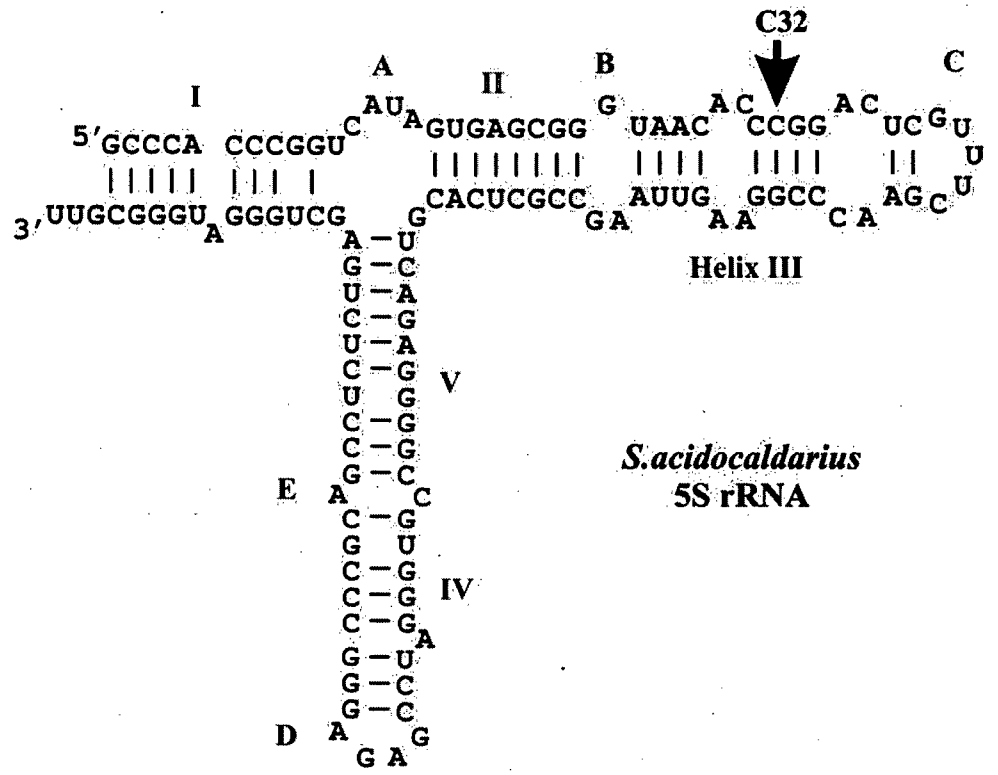


Figure 3-12 Predicted secondary structure of *S. acidocaldarius* 5S rRNA

Predicted secondary structure of 5S rRNA from *S. acidocaldarius.* The sequence is presented in a standard 3-arm form (Stahl et al., 1981), with helices I through V indicated. Residue C32 is buried in helix III and its position is indicated by an arrow.

In agreement with secondary structures of 5S rRNA determined for other species, the structure consists of five helices (I–V), with residue C32 buried within helix III and two hairpin loops (C and E), two internal loops (B and D) and a hinge region (A), organized into three helix junctions (Szymanski *et al.*, 2003). Altogether the 5S rRNA exhibits a high degree of predicted secondary structure.

3.3.2.2 5S rRNA fragment

To test if the secondary and three-dimensional architecture of the 5S rRNA substrate would influence the efficiency of the methylation reaction, a 29 nt long fragment of 5S RNA containing the C32 target was used in the methylation assay (Figure 3-13). Even under these conditions, radioactive methyl incorporation was still not above background (Figure 3-11).

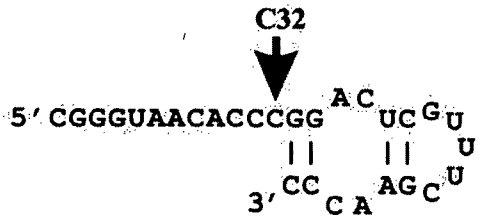


Figure 3-13 Structure of the 29 nt long 5S rRNA fragment

A predicted secondary structures of the 29 nt 5S rRNA fragment. The residue predicted to be methylated (C32 in 5S rRNA) is indicated by an arrow.

3.3.2.3 Optimized *S. acidocaldarius* sR10 guide

The results with wildtype *S. acidocaldarius* sR10 sRNA suggest that the lack of activity may be due to the short length of the complementarity and the internal mismatch at position 9. To confirm this, T to G, A to T, and A to G mutations were introduced to at positions 1, 2, and 9, respectively, within the D' box guide of *S. acidocaldarius* sR10 sRNA. This eliminated the guide target mismatch at position 9, and extended the

complementarity from eight to eleven base pairs. The mutant sR10 sRNA exhibited a level of activity on both the oligonucleotide fragment, and on the full length 5S rRNA targets (Figure 3-11) that was about 50% of the control *S. acidocaldarius* sR1 sRNA mediated methylation activity.

3.4 Engineered sR1/sR2 hybrid

A key feature of the guide RNA modification systems is that few if any constraints are placed on the substrate beyond complementarity to the guide RNA sequence; this suggests that artificial guides can be constructed to direct methylation to new sites in an RNA target molecule. In yeast it has been shown that methylation can be shifted to novel sites by either moving the D or D' box relative to the antisense guide sequence or by equipping a snoRNA with a new guide sequence (Liu *et al.*, 2001; Ni *et al.*, 1997a). Methylation has been directed to rRNA nucleotides which are normally not ribose methylated and to sites that usually are. In some cases methylation impairs growth, indicating that a functionally important nucleotide has been altered.

The *S. acidocaldarius* and *S. solfataricus* sR1 sRNAs both guide methylation to position U52 in 16S rRNA using the D box guide. The sR2 sRNA from *S. acidocaldarius* guides methylation to position C1914 in 23S rRNA. Methylation at this position has been confirmed *in vivo* in *S. acidocaldarius* but the position is not methylated in *S. solfataricus* (Omer et al., 2000). To determine the plasticity and flexibility of the guide-directed methylation system, I replaced the D box guide sequence of the *S. solfataricus* sR1 with the D box guide sequence of sR2 sRNA of *S. acidocaldarius.* Thus the sR1/sR2 engineered mutant contains a replacement of the indigenous sR1 D box guide region of *S. solfataricus* with the guide region of *S. acidocaldarius* sR2 RNA and therefore should guide methylation to a novel site, C1914 in 23S rRNA in *S. solfataricus*.

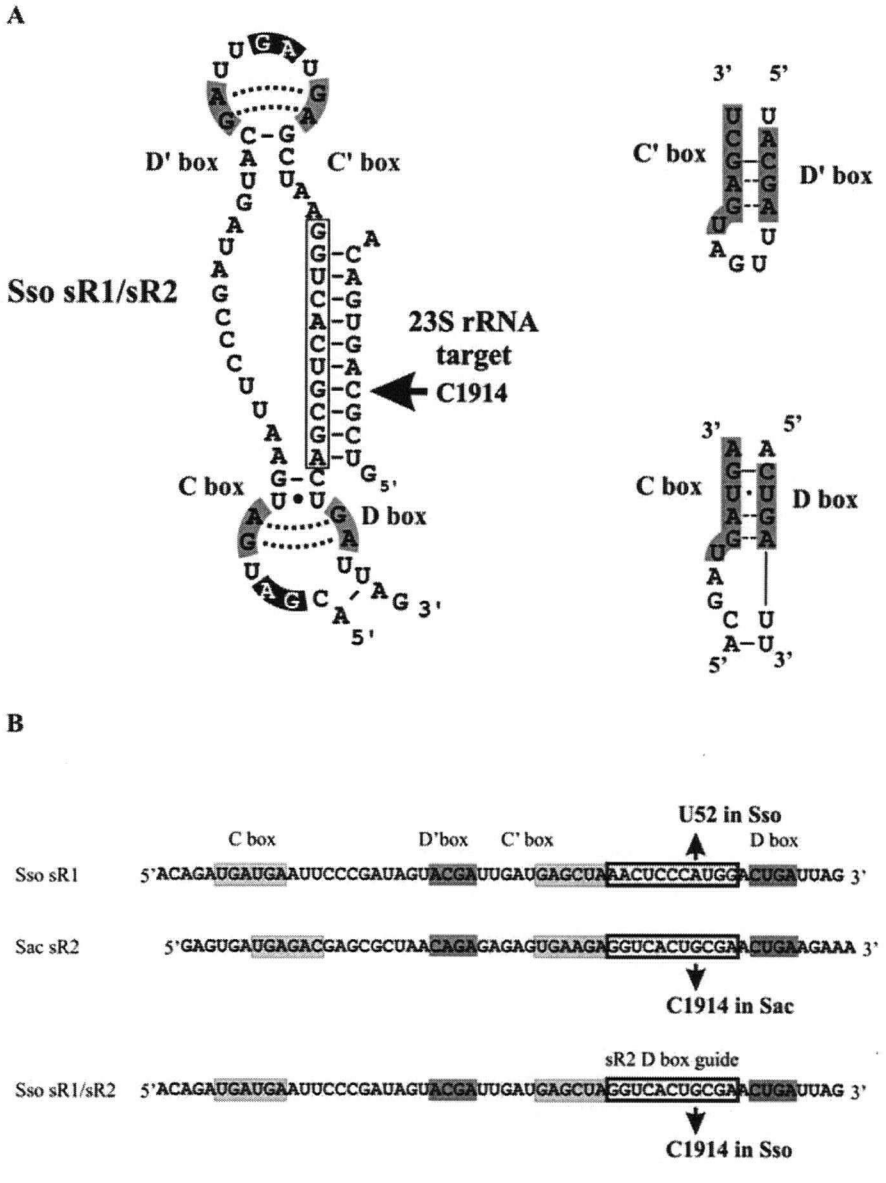


Figure 3-14 Engineered *Sulfolobus* sR1/sR2 hybrid

(A) The predicted secondary structure of the engineered *Sulfolobus* sR1/sR2 hybrid (Sso sR1/sR2) is illustrated. The predicted k-turn consensus of the C/D and C'/D' motif is shown at the right. The C'/D' motif exhibits a Watson-Crick base pair instead of the non canonical basepair in stem II. (B) The nucleotide sequences of *S. solfataricus* sR1 (Sso sR1), *S. acidocaldarius* sR2 (Sac sR2) and *S. solfataricus* sR1/sR2 are shown. In *S. solfataricus* sR1/sR2 the eleven nucleotide D box guide sequence has been replaced by the D box guide sequence from *S. acidocaldarius* sR2, changing the target residue from U52 in 16S rRNA to C1914 in 23S rRNA.

A 29 nt long target containing the region of C1914 in 23S rRNA was assembled with the engineered sR1/sR2 hybrid and the three core proteins in the presence of SAM (Figure 3-15). Methylation occurred and exhibited an intermediate level of activity compared to the control reaction with sR1 sRNA, but similar activity to the reaction using the natural *S. acidocaldarius* sR2 guide and the target containing C1914. This result shows that new sites of RNA modification can readily be specified in the guide RNA modification systems. No constraints seem to be placed on the substrate beyond complementarity to the guide RNA sequence, allowing the construction of artificial guides. Ultimately, it will be necessary to confirm that the reconstituted system is an accurate representation of the *in vivo* system. The engineered guide presented here is a prime candidate for epigenetic expression in the recently developed genetic system for *S. solfataricus* in order to probe the *in vivo* function of the methylation guide machinery (Sowers, 2003). In *S. solfataricus*, residue C1914 in 23S rRNA is not methylated. With the introduction and expression of the engineered sRNA in transformant *S. solfataricus* cells, methylation at residue C1914 should occur. Since the site is known to be methylated in *S. acidocaldarius*, it is expected that the engineered methylation at this site in *S. solfataricus* will not be harmful or will not impair growth of the cells.

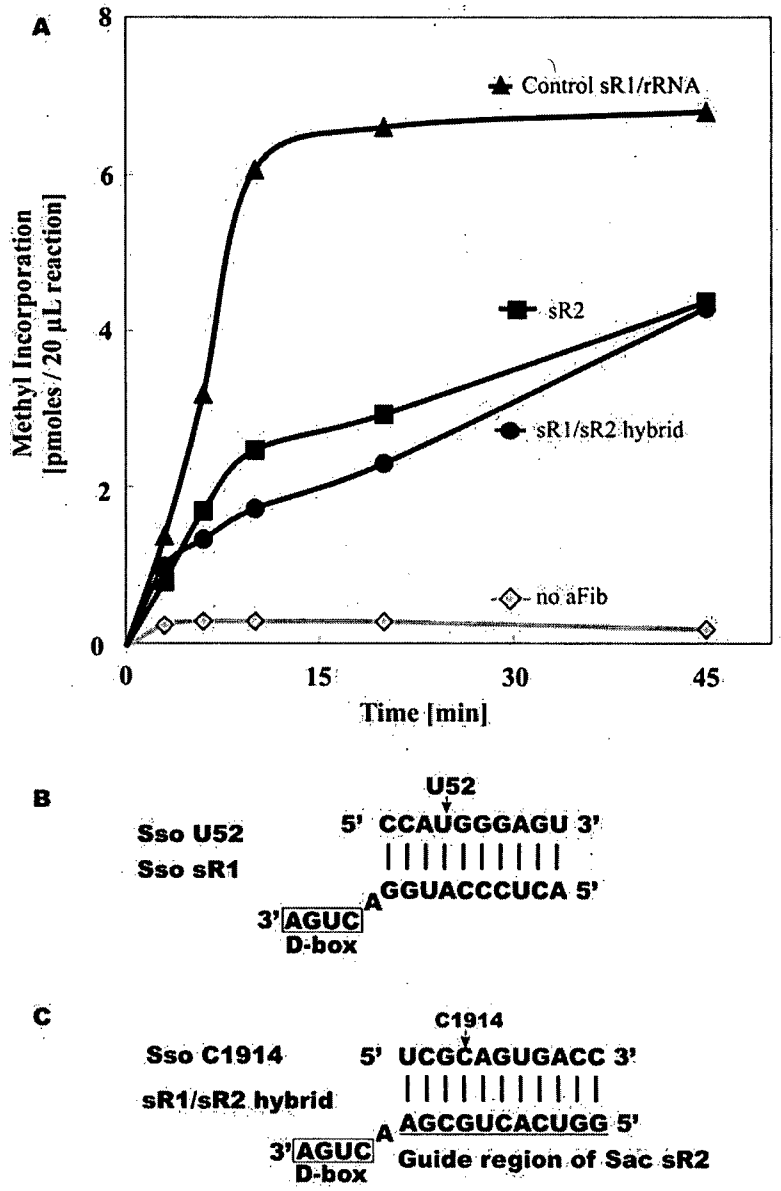


Figure 3-15 (see next page for legend.)

Figure 3-15 RNP guide-dependent methyl incorporation with sR1/sR2 hybrid

(A) Methylation assays, as described in the legend to Figure 3-9 , were carried out using an engineered *S. solfataricus* sR1/sR2 guide sRNA and a 28 nt long target containing the C1914, the site of methylation in 23S rRNA in *Sulfolobus* (Table 2-5). In *S. solfataricus* sR1/sR2 sRNA the *S. solfataricus* sR1 D box guide is replaced by the D box guide of *S. acidocaldarius* sR2 sRNA. The methylation assay was carried out with the following controls: *S. acidocaldarius* sR2 sRNA and the *S.acidocaldarius* sR2 28 nt target, and *S. acidocaldarius* sR1 sRNA and *S. acidocaldarius* sR1 rRNA target. (B) The structure of *S. solfataricus* sR1 guide RNA duplex with D box target is illustrated. (C) sR1/sR2 hybrid with the D box target of sR2 in 23S rRNA is shown.

3.5 Double Guide Methylation

3.5.1 *Pyrococcus* double guide

More than half of the sRNAs identified in the genomes of the hyperthermophilic Euryarchaeota *Pyrococcus furiosus, P. abyssi* and *P. horikoshii* have rRNA antisense elements associated with both the D' and D box motifs, whereas only about 20% of eukaryotic C/D box snoRNAs have double guide function (Kiss-Laszlo *et al.*, 1996; Omer *et al.*, 2000). Moreover, the predicted methylation targets in the *Pyrococcus* sRNAs are often in close proximity within either the 16S or the 23S rRNA molecule. The two target site sequences can be in either a direct or reverse orientation relative to the sRNA guide sequences. In the direct orientation, the region within the rRNA that is complementary to the D box guide is located 5' to the region complementary to the D' box guide, and simultaneous interactions between the sRNA guides and rRNA target, would form a hyphenated anti-parallel duplex structure. Conversely, in the reverse orientation, the region within the rRNA that is complementary to the D box guide is located 3' to the region complementary to the D' box guide, and simultaneous interactions between the sRNA guides and rRNA targets would form a pseudoknot structure. In these instances it is uncertain if both of the guides are active, and if they function simultaneously to form an extended, hyphenated and more stable structure.

The *Pyrococcus* sR24 sRNA has D and D' box antisense elements that are predicted to guide methylation in a direct orientation to positions C1221 and C1243

respectively in 23S rRNA (Figure 3-16). The simultaneous interaction of the guide and target sequences is expected to form a hyphenated anti-parallel duplex structure.

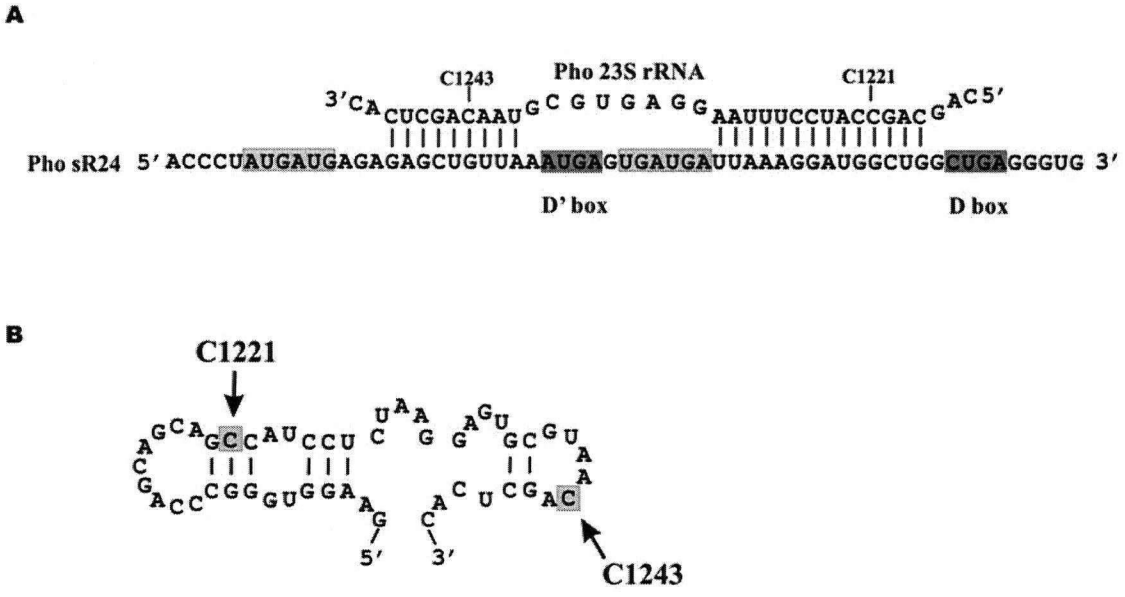


Figure 3-16 *Pyrococcus* sR24 direct double guide

(A) The sequence of *Pyrococcus* sR24 sRNA is shown, and the conserved motifs are boxed and shaded. sR24 sRNA is predicted to guide methylation to C1243 and C1221 in 23S rRNA in a direct manner; the sequence of the target RNA is illustrated.(B) The potential secondary structure of the 50 nt long *in vitro* transcript used in the *in vitro* methylation assay is depicted. *Pyrococcus* sR24 is predicted to guide methylation to C1221 and C1243 in 23S rRNA. The target residues are boxed and shaded.

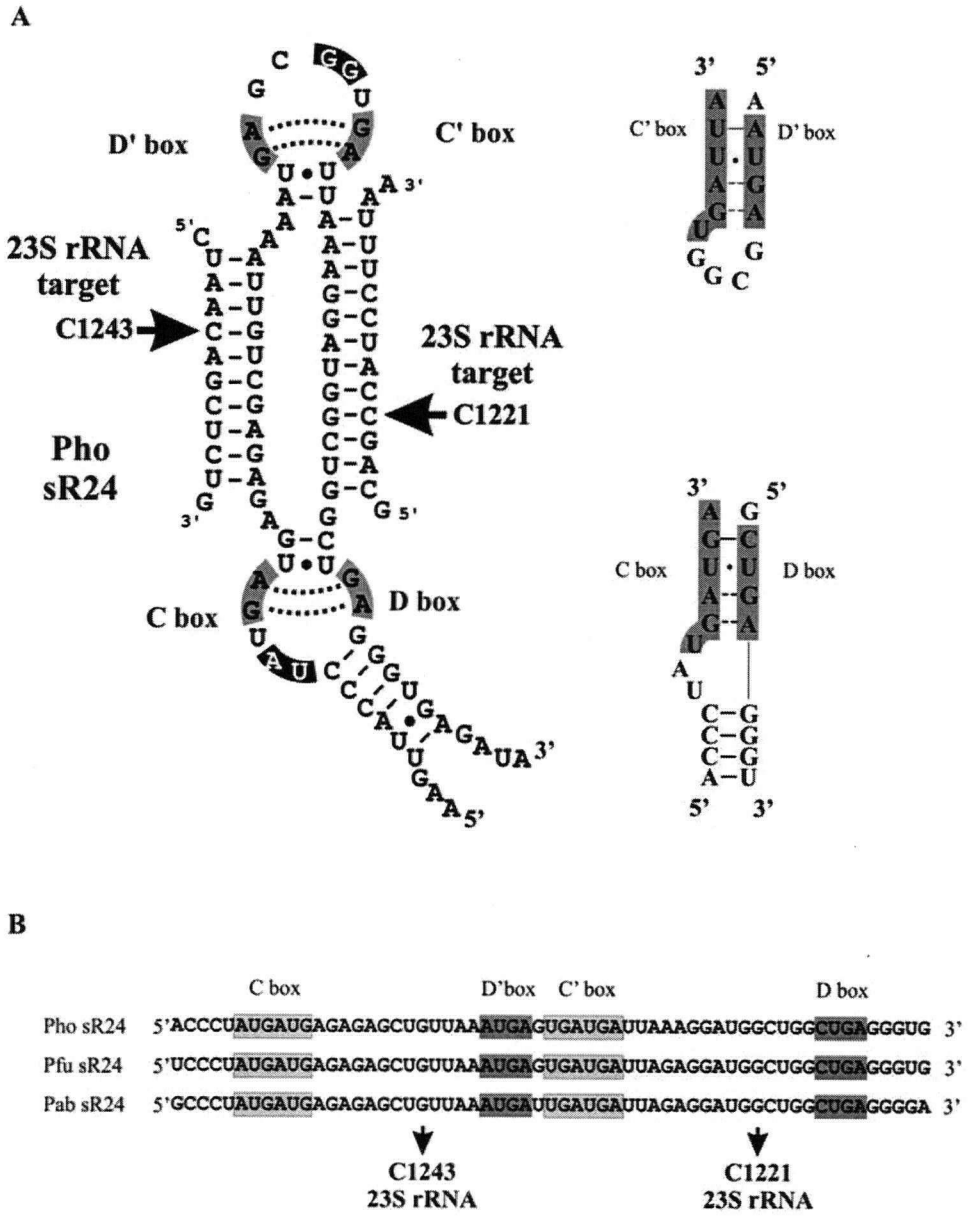


Figure 3-17 *Pyrococcus* sR24 direct double guide

(A) A secondary structural model of Pho sR24 sRNA is depicted. The predicted k-turn consensus of the C/D and C'/D' motif is shown on the right. The D box guide exhibits a fourteen base pair long complementarity to the region spanning residue C1221 of 23S rRNA, and the D' box guide displays ten base pair complementarity to the region spanning residue C1243 of 23S rRNA. PhosR24 also forms a six base pair long stem at the C/D box motif. (B) Sequences of the three *Pyrococcus* sR24 sRNAs are shown, and the conserved boxes are shaded.

To understand the function of these direct double guide interactions, I tested the methylation activity of sR24 sRNA containing RNP complexes on target oligonucleotides containing either one or both of the predicted sites of methylation (Figure 3-16 B).

To my surprise, the RNP complex was virtually inactive on the C1221 and C1243 single target oligonucleotides, when added either individually or together in the methylation reaction (Figure 3-18). Maximum activity was detected when both the C1221 and C1243 targets were present in *cis* on a single oligonucleotide substrate. When C1221 and C1243 targets were present on separate molecules the initial rate of activity reached only approximately 25% of the single oligonucleotide substrate. This suggests that the sR24 has evolved to function efficiently as a double guide RNA and requires simultaneous base pairing of the two guide regions to a single target molecule. The activity of the sR24 double guide was about 40% of the control sR1 sRNA activity.

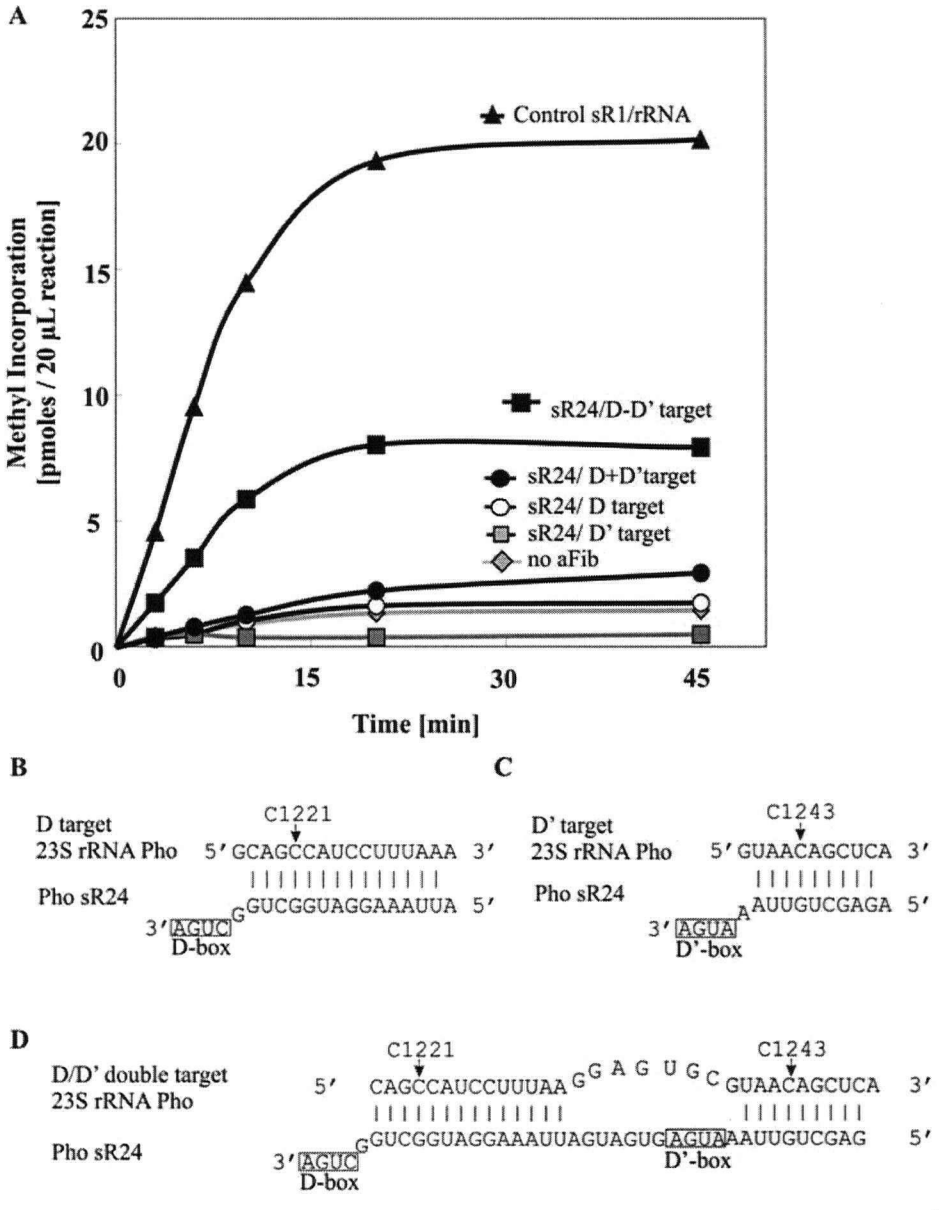


Figure 3-18 (see next page for legend.)

Figure 3-18 RNP guide-dependent methyl incorporation with double guide sRNA

(A) Methylation assays were carried out as described in the legend to Figure 3-9, using the following guide RNAs and targets (Table 2-5): (i) *in vitro* transcribed *Pyrococcus* sR24 and a 30 nt long target containing the D box site of methylation (C1221 of 23S rRNA), (ii) *Pyrococcus* sR24 and a 21 nt long target containing the D' box site of methylation (C1243 of 23S rRNA), (iii) *Pyrococcus* sR24 and two separate targets containing the D and D' box site of methylation, (iv) *Pyroccocus* sR24 and a 37nt long target containing D and D' box site of methylation on one molecule, and (v) sR1 and rRNA target as control with and without aFib protein in the reaction. The incorporation kinetics for the six reactions are illustrated.(B) Structure of *Pyrococcus* sR24 guide RNA duplex with D box target and (C) the D' box target in 23S rRNA are illustrated. (D) The structure of the *Pyrococcus* sR24 guide RNA and the duplex with D and D' box are shown.

There are a number of possible explanations for the reduced activity obtained when using *Pyrococcus* RNA. One explanation is that the reduction in activity could be due to the heterologous proteins from *Sulfolobus* used in this experiment. Although both organisms belong to the domain Archaea and are thermophiles, *Sulfolobus* belongs to the phylum of Crenarchaeota, whereas *Pyrococcus* belongs to the Hyperthermophiles in the phylum of Euryarchaeota. An alternative explanation for the reduced activity is that it could be due to the more complex nature of the double guide reaction, where the two targets have to bind to the two antisense elements of the sRNA simultaneously.

3.5.1.1 Specificity of Double Guide

Residues C1221 and C1243 are predicted to be methylated, guided by *Pyrococcus* sR24 sRNA. These residues are found in a CC and a CA dinucleotide, respectively. To assign the sites of methylation in the *Pyrococcus* double target, two-dimensional thin-layer chromatography was performed. The oligonucleotide substrate with both targets was incubated in the presence of [^{3}H]-methyl SAM in the standard methylation assay and then subjected to digestion with nuclease P1 under conditions where non methylated nucleotides are completely hydrolyzed. The hydrolysis products were mixed with 2'-O-ribose methylated NMPs (pAm, pCm, pGm, and pUm) as well as pCmC and pCmA standards and separated by two-dimensional thin-layer chromatography. The position of individual standards was observed by UV shading and superimposed to the image obtained by tritium imaging (Figure 3-19). The tritium radioactivity was uniquely found only in the pCm and pCmC and the pCmA spots.

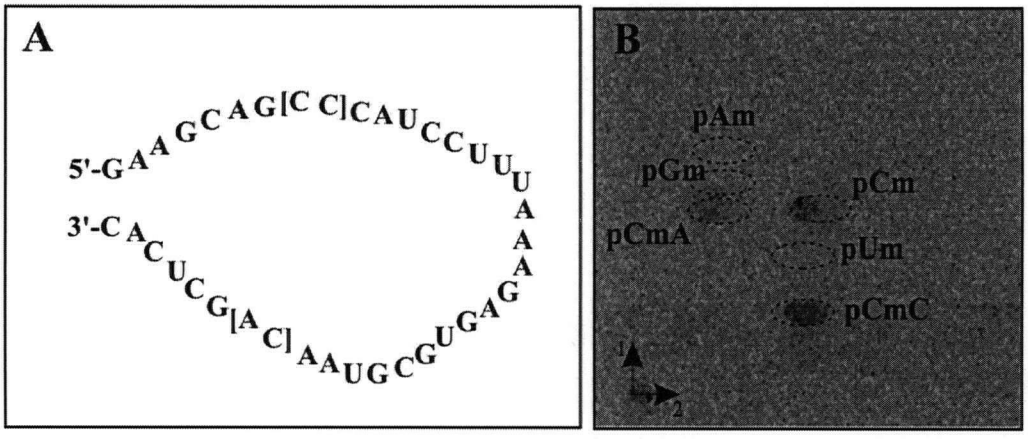


Figure 3-19 Two-dimensional thin-layer chromatography - *Pyrococcus* sR24 sRNA

Thin-layer chromatographic separation of the hydrolysis products of the sR24 sRNA target RNA. (A) The standard 37 nt long target RNA containing the CC and CA dinucleotide is illustrated. The C residues are the expected site of methylation and correspond to position C1221 and C1243 in 23S rRNA. (B) Guide and target RNAs were mixed in 20 µl reaction in the standard methylation assay using [methyl-^{3}H]SAM and incubated at 70°C for 45 min. The RNA was extracted, digested with 0.05 unit of P1 nuclease for 16 h at 37°C, and products were mixed with 10 nmol each of of pAm, pCm, pGm, pUm, pCmC, and pCmA standards before the two-dimensional TLC separation. Unlabeled standards were detected by UV shadowing, and radioactivity was detected by ^{3}H-imaging. Omission of any of the RNA or protein components from the reaction mixture results in background level of incorporation of radioactivity into acid-insoluble material (data not shown).

3.5.2 Inverted double guide

The *S. acidocaldarius* sR7 sRNA has D and D' box antisense elements that are predicted to guide methylation in a reverse orientation to positions U2692 and G2649, respectively, in 23S rRNA. The simultaneous interaction of the guide and target sequences is expected to form a pseudoknot structure (Figure 3-20).

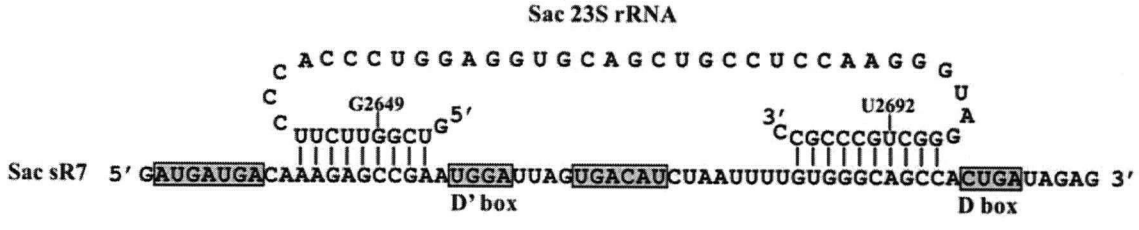


Figure 3-20 *S. acidocaldarius* sR7 inverted double guide

The sequence of *S. acidocaldarius* sR7 is shown, and the conserved motifs are boxed and shaded. *S. acidocaldarius* sR7 sRNA is predicted to guide methylation to G2649 and U2692 in 23S rRNA in a reverse orientation; the target sequence is partially illustrated. In this reverse orientation, simultaneous interactions between the guides and targets forms a pseudoknot structure.

S. acidocaldarius sR7 sRNA shows several uncommon features. First, the C' and D' motifs contain exceptions to the consensus motifs of these boxes. The resulting k-turn deviates from the consensus by forming a GC Watson-Crick base pair instead of a non-canonical base pair in stem II of the predicted k-turn. Second, the spacing between the C' and D box is, at 22 nucleotides, unusually long. Typically the distance between these boxes is 11 to 16 nt in archaeal sRNA. Last, no terminal stem pairing can be predicted;

this might be due to the fact that only incomplete RNA was cloned, and the genomic sequence of *S. acidocaldarius* is currently not available.

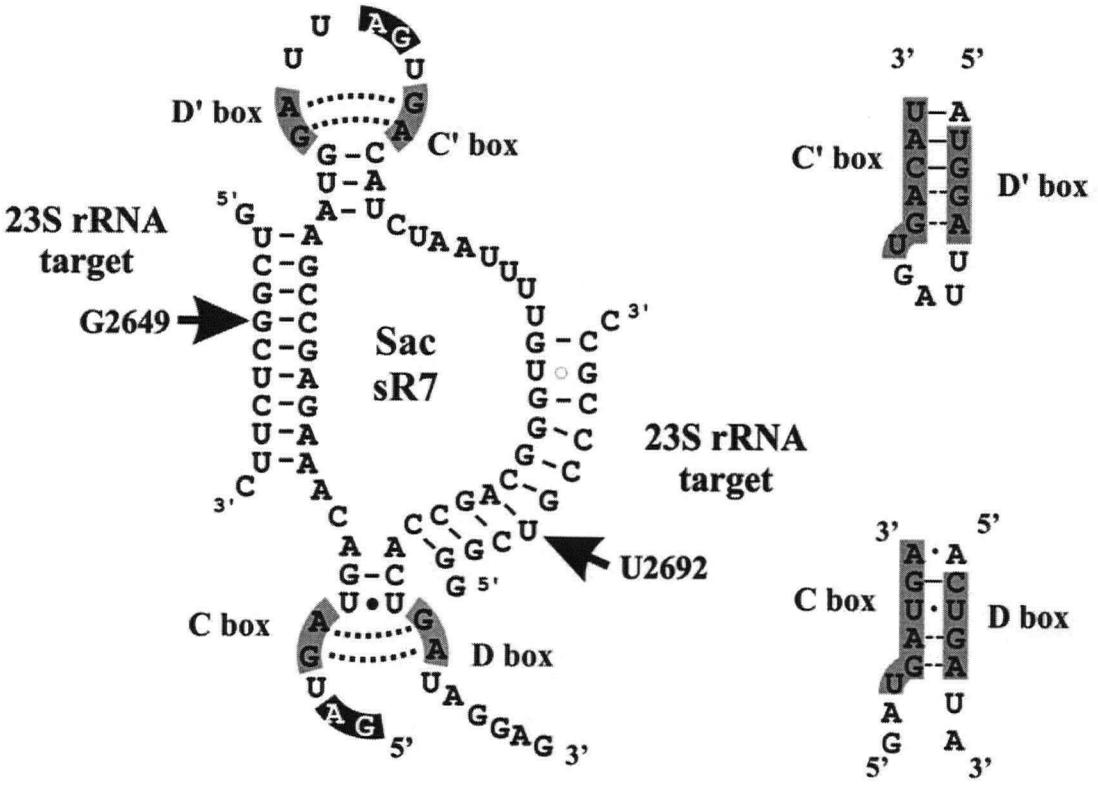


Figure 3-21 *S. acidocaldarius* sR7 sRNA

A secondary structural model of *S. acidocaldarius* sR7 sRNA is illustrated. The predicted k-turn consensus of the C/D motif is shown on the right. The C'/D' motif deviates from the consensus by forming a GC Watson-Crick base pair instead of a non-canonical base pair in stem II. The D box guide exhibits a ten base pair long complementarity to the region spanning residue U2692 of 23S rRNA, and the D' box guide displays nine base pair complementarity to the region spanning residue G2649 of 23S rRNA.

To understand the function of these reverse double guide interactions, I tested the methylation activity of sR7 containing RNP complexes on target oligonucleotides containing either one or both of the predicted sites of methylation (Figure 3-23).

As with the *Pyrococcus* sR24 double guide, maximal activity was observed with a target containing both target sites in *cis* ( sR7/D-D' target). However the overall efficiency of the reaction was rather poor compared to the control reaction. To my surprise, the U2692 single target oligonucleotide activity ( sR7/D target) reached 50% of the activity of the target with both sites in *cis*, whereas the RNP complex showed virtually no activity on the G2649 single target ( sR7/D' target). Consequently, intermediate active was observed when both U2692 and G2649 single target oligonucleotides were added together ( sR7/D+D' target) in the reaction.

To determine if the reverse orientation of the target influences the methylation efficiency, the target was mutated to resolve the pseudoknot structure. The sequence complementary to the D guide box sequence was substituted by the D' antisense element and the D' box guide antisense element was replaced by the sequence complementary to the D box guide, resulting in a direct target (Figure 3-23 E). Still, methylation was very limited ( sR7/D'-D target) and did not show more activity than the target containing both sites in a reverse order.

The D box guide of sR7 guides methylation to the homologous position U2692 (U2552 - *E.coli* numbering), that has been implicated in an interaction with the 3'end of the A site in tRNA (Kim and Green, 1999; Moazed and Noller, 1989). This region shows a high degree of secondary structure, and a ten base pair long stem with a four nucleotide bulge and a terminal loop is predicted between the antisense element of the D box guide

and the D' box guide (Figure 3-22). The limited activity observed with the reverse double guide and the direct double guide in the assay with *S. acidocaldarius* sR7 (Figure 3-23) may reflect the requirement for an RNA helicase in the reaction.

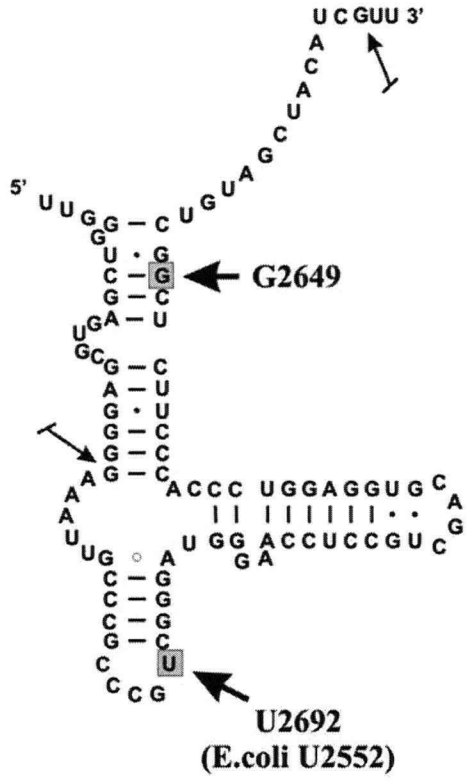
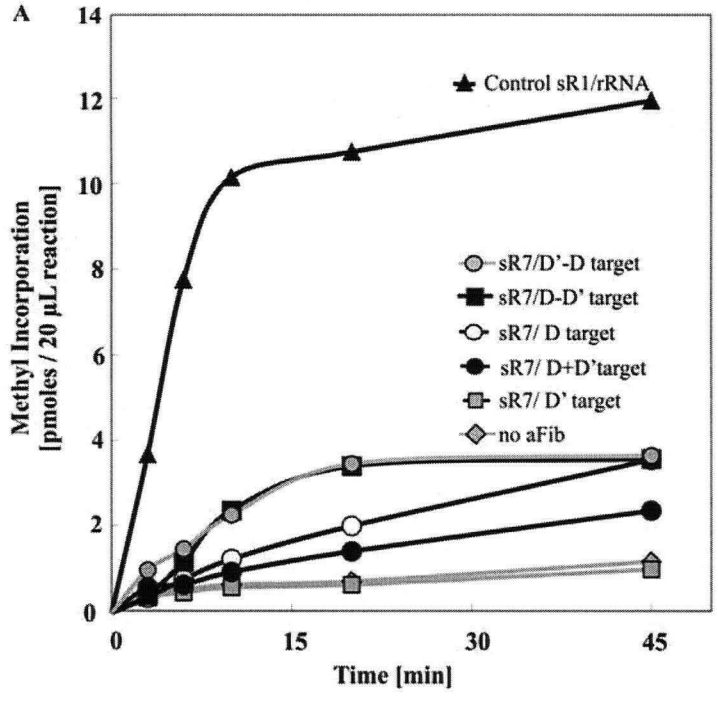
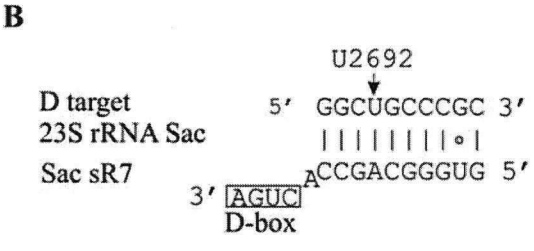
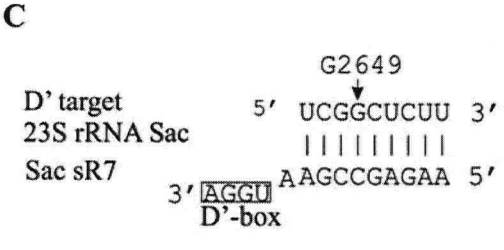


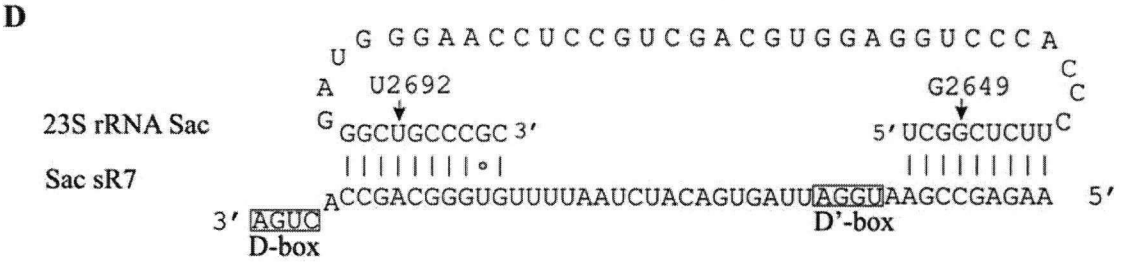
Figure 3-22 *S. acidocaldarius* sR7 target

The sequence and structure of the 23S rRNA target region of *S. acidocaldarius* sR7 sRNA are depicted. *S. acidocaldarius* sR7 sRNA is predicted to guide methylation in a reverse orientation to G2692 and the homologous position U2692 (U2552 in *E.coli* numbering). The targets residues are boxed and shaded. The thin arrows indicate the start and end of the *in vitro* transcript used in the methylation assay.


A
Methyl Incorporation [pmoles / 20 µL reaction]
Time [min]
Control sR1/rRNA
sR7/D'-D target
sR7/D-D' target
sR7/ D target
sR7/ D+D'target
sR7/ D' target
no aFib


B
U2692
D target
23S rRNA Sac
Sac sR7
5′ GGCUGCCCGC 3′
3′ AGUC
ACCGACGGGUG 5′
D-box


C
G2649
D' target
23S rRNA Sac
Sac sR7
5′ UCGGCUCUU 3′
3′ AGGU
AAGCCGAGAA 5′
D'-box


D
23S rRNA Sac
Sac sR7
U2692
G2649
3′ AGUC
D-box
D'-box

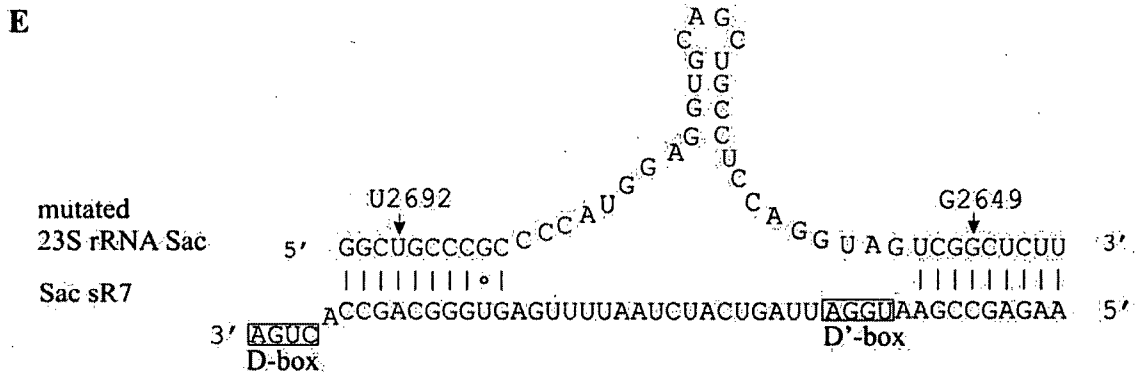


Figure 3-23 RNP guide-dependent methyl incorporation with inverted double guide sR7

(A) Methylation assays were carried out as described in the legend to Figure 3-9, using the following guide RNAs and targets (Table 2-5): (i) *in vitro* transcribed *S. acidocaldarius* sR7 and a 31 nt long target containing the D box site of methylation (G2649 of 23S rRNA), (ii) *S. acidocaldarius* sR7 and a 32 nt long target containing the D' box target (U2692), (iii) *S. acidocaldarius* sR7 and two separate targets containing the D and D' box site of methylation , (iv) *S. acidocaldarius* sR7 and a 74 nt long target containing D and D' box site of methylation on one molecule, (v) *S. acidocaldarius* sR7 and a 74 nt long target with the D and D' box exchanged, resulting in a direct guide and (vi) sR1 and rRNA target as control with and without aFib protein in the reaction. The incorporation kinetics for the six reactions is shown.
(B) Structure of *S. acidocaldarius* sR7 guide RNA duplex with D box target and (C) the D' box target in 23S rRNA are illustrated.(D) The structure of the *S. acidocaldarius* sR7 guide RNA and the duplex with D and D' box are shown in and the *S. acidocaldarius* sR7 guide RNA and (E) the duplex with the exchanged D and D' box, resulting in a direct guide is illustrated.

3.5.3 Conserved Methylation Sites

Previous studies analysed the distribution of methylation sites within the secondary structure of the small and large rRNAs of Eukaryotes and Archaea (Dennis *et al.*, 2001; Maden, 1990). Modifications cluster in the universally conserved functional regions of rRNAs. Most 2'*O*-methylated residues are conserved, with almost no differences among the more than 100 sites. Yeast rRNA contains 55 2'-*O*-methylated residues, 18 of which correlate with those in higher eukaryotes. There are three methylated sugar moieties found in *Escherichia coli* 23S rRNA; two of which (G2251 and U2552) seem to be universally conserved.

The D box guide of sR7 guides methylation to the homologous position U2692 (U2552 - *E.coli* numbering), which has been implicated in an interaction with the 3'end of the A site tRNA (Kim and Green, 1999; Moazed and Noller, 1989).

sR3 sRNA from three *Pyrococcus* species also guides methylation to this position and are potential double guides with a D' target in close proximity. However, the D' box guides/targets do not show sequence similarities to the D' box guide sequence of *S. acidocaldarius* sR7 sRNA, but instead show complementarity to the region of C2612 - 62 nt up stream of the antisense element of the D box (Figure 3-24). This could imply that the D' box is more important for stability of the sRNP complex, rather than for guiding methylation to this position. Double guides could therefore be involved in direct folding of nascent rRNAs.

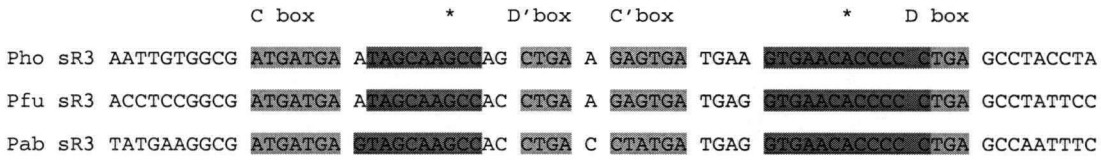


Figure 3-24 Alignment of sR3 from three *Pyrococcus* species

sR3 sRNA is predicted to guide methylation from its D box antisense element to the homologous position U2592 in 23S rRNA. The D' box is predicted to methylate a position is close proximity - C2612 in 23S rRNA. The conserved box elements are shaded in light grey, the antisense element complementary to the target sequence is shaded in dark grey. The asterisks indicate the nucleotide 5 nt up-stream of the guide box, which forms a Watson-Crick base pair with the nucleotide that is being methylated.

To determine if there are homologous sRNAs to *S. acidocaldarius* sR7 in other archaea, whose sequences have been published, I performed a BLASTN search against the non-redundant nucleotide database. Since short sequences often do not find any significant matches to the database entries under the standard nucleotide-nucleotide BLAST settings, the significance threshold of the expected (E) value parameter was set to be less stringent (1000) and the default word size parameter was set to seven (a lower setting then the default of ten in the standard nucleotide BLAST search). I recovered a hit against an intergenic region of the *Sulfolobus tokodaii* genome (score = 46.1, bits, E-value of 0.005). Searches with short sequences can be virtually identical and have relative high E values, because the calculation of the E-value takes the length of the sequence into account. However, after close examination of the region, I was able to identify canonical C and D boxes and C' and D' boxes and designated this new putative C/D sRNA *S. tokodaii* sR7 (Figure 3-25).

The *S. tokodaii* sR7 sRNA gene is encoded intergenically, between the genes ST9989 $tRNA^{Met}$ and ST0264, a 221aa long hypothetical protein. A BLASTP and a PSI (Position-Specific Iterated) BLAST search did not reveal any homologous proteins in other organisms, nor could it detect any known domains in this hypothetical protein.

S. tokodaii sR7 sRNA was examined for potential guide sequences with *S. tokodaii* rRNA, and regions of complementarity to 23S rRNA were identified adjacent to the D and D' box. Similar to *S. acidocaldarius* sR7, *S. tokodaii* sR7 sRNA is predicted to target methylation to positions U2692 and G2649 in 23S rRNA. *S. tokodaii* sR7 sRNA shows an overall high sequence similarity to *S. acidocaldarius* sR7 sRNA, except for the 5 nt insertion between the C' box and the D box, that is not present in *S. tokodaii* sR7 sRNA. No significant hits against other sequenced genomes were recovered; this might be due to the limitation of the BLAST search.

```
                C box          *    D'box   C'box                    *     D box
Sac sR7 -----G ATGATGA CAAAGAGCCGAA TGGA T TAGTGA CATCTAATTTTGTGGGCAGCCA CTGA TAGAG
Sto sR7 CCCATT ATGATGA CAAAGAGCCGAA TGGA A ATGTGA CAGCT-----TGTGGGCAGCCA CTGA TGAAA
```

Figure 3-25 Alignment of sR7 sRNA from *S. acidocaldarius* and *S. tokodaii*.

The alignment of sR7 of *S. acidocaldarius* and *S. tokodaii* is shown. The conserved boxes are shaded in light grey, the antisense elements complementary to the target are in dark grey. The asterisks indicate the nucleotide 5 nt up stream of the guide boxes.

3.6 Novel sRNA in *Sulfolobus*

S. acidocaldarius sRNA homologues in *Sulfolobus*

S. acidocaldarius sR7 sRNA has been identified in a biochemical study with *Sulfolobus acidocaldarius*. Antibodies against aFib and aNop56 protein from *S. acidocaldarius* were used to immunoprecipitate complexes containing these proteins from cell extracts. Nearly 30 different cDNA clones with features characteristic of eukaryotic C/D box RNAs were recovered from the RNA present in these precipitates (Omer *et al.*, 2000). The finding of sR7 sRNA in *S. tokodaii* prompted searches for other unidentified homologues of *S. acidocaldarius* sRNAs in the recently sequenced genomes of the genera *Sulfolobus*. Thirteen sRNAs had been previously identified computationally in the unfinished genome of *S. solfataricus* (Omer *et al.*, 2000).

Only four genomes from crenarchaeota are completely sequenced and published to date - *Sulfolobus tokodaii (Sto)*, *Sulfolobus solfataricus (Sso)*, *Aeropyrum pernix (Ape)* and *Pyrobaculum aerophilum (Pae).* The genome of a fifth crenarchaeon, *Sulfolobus acidocaldarius (Sac)*, is currently being sequenced.

Fifteen new putative sRNA homologues could be detected: five in *S. solfataricus* and ten in *S. tokodaii. S. solfataricus* sR1 sRNA had been predicted in a previous screen (Omer *et al.*, 2000). In three cases (sR2, sR17, and sR26), previously unknown potential methylation sites could be identified.

	S. acidocaldarius (Sac)		*S. solfataricus (Sso)*		*S. tokodaii (Sto)*	
sRNA	D	D'	D	D'	D	D'
sR1	16S U52	?	16S U52	16S U606 16S U33	16S U52	?
sR2	23S C1914	?	--	--	23S C1914 $tRNA^{Phe}$	?
sR3*	23S G2739	?	23S G2739 snoRNA_2	?	23S G2739	?
sR4	23S G1995	?	--	--	--	--
sR5	16S G1056	?	--	--	--	--
sR6	23S G2666	?	--	--	--	--
sR7	23S U2692	23S G2692	--	--	23S U2692	23S G2692
sR8	23S G334	23S U2972	--	--	--	--
sR9	?	16S G26	?	16S G26	--	--
sR10	23S C2539	$tRNA^{Gly}$ C50	--	--	--	--
sR11	?	23S A 2618	--	--	--	--
sR12	23S A1134	23S G1114	--	--	--	--
sR13	23S C2746	23S G385	--	--	--	--
sR14	$tRNA^{Gln}$ U34	?	$tRNA^{Gln}$U 34 snoRNA_6	--	$tRNA^{Gln}$ U34	?
sR15	?	?	--	--	--	--
sR16	?	?	--	--	--	--
sR17	?	?	?	?	23S U1043 16S U689	?
sR18	?	23S G140	--	--	?	23S G140
sR19	16S G672	?	--	--	--	--
sR20	?	23S C2065	?	23S C2065	?	23S C2065

sR21	?	?	--	--	--	--
sR22	23S U2893 16S U1310	23S C2916	--	--	23S U2893 16S U1310	23S C2916
sR23	?	?	--	--	--	--
sR24	?	?	--	--	--	--
sR25	?	?	--	--	--	--
sR26	?	?	$tRNA^{Lys}$ U34	?	$tRNA^{Lys}$ U34	?
sR27	?	?	--	--	--	--
sR28	23S C473	?	--	--	--	--
sR29	?	?	--	--	--	--

Table 3-1 Result of *S. acidocaldarius* sRNA BLAST search

Results of a BLAST search against the non-redundant nucleotide database with 29 sRNA from *S. acidocaldarius*, that had been biochemically identified (Omer *et al.*, 2000). *S. acidocaldarius* sRNAs for which putative homologues in other *Sulfolobus* genera have been identified have been shaded. "D" or "D'" indicates the box adjacent to complementarity. Predicted target ribose methylation sites are specified. In cases where no target could be identified, a "?" is shown."*" indicates that a putative homologue sRNA has been identified in *S. shibatea*. *S. solfataricus* sR3sRNA has been annotated as snoRNA_2, and *S. solfataricus* sR14 sRNA has been annotated as snoRNA_6 in the *Sulfolobus solfataricus* genome.

sR1 sRNA

A BLAST search with *S. acidocaldarius* sR1 resulted in no hits against the genome of *S. tokodaii*. However, the *S. tokodaii* genome contains an orthologue to the aspartate aminotransferase gene of *S. acidocaldarius* and *S. solfataricus*. After examination of the C-terminus and its adjacent sequence, a sR1 sRNA homologue could be identified in *S. tokodaii*. The newly predicted sR1 sRNA was named *S. tokodaii* sR1 sRNA (Figure 3-26).

S. tokodaii sR1 sRNA overlaps the C-terminal end of the aspartate aminotransferase coding gene region by 21 nt. It contains the D box antisense element

that is used to guide methylation to position U52 in 16S rRNA. Although *S. tokodaii* sR1 sRNA shows striking similarity to the D' box antisense element of *S. solfataricus* sR1 (only one mismatch at position +7), no target could be predicted for the D' box guide.

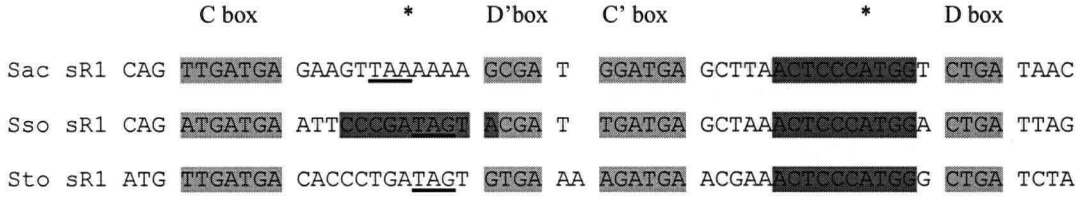


Figure 3-26 Alignment of sR1 sRNA from three *Sulfolobus* species

sR1 from *S.acidocaldarius, S. solfataricus* and *S. tokodaii* are aligned. The conserved boxes are shaded in light grey. The translation termination codon of the aspartate amino transferase is underlined. The antisense element of the sRNA is shown in dark grey. The asterisk indicates the nucleotide 5 nt up-stream of the D' box.

sR2 sRNA

A BLAST search with *S. acidocaldarius* sR2 sRNA resulted in a weak hit (Score: 34.2 bits, E-value: 18) partially overlapping the N terminus of a long conserved hypothetical protein (ST0964) in *S. tokodaii*. After examining the hit, conserved C/ D boxes and C'/D' boxes could be identified, and the new putative sRNA was designated *S. tokodaii* sR2 (Figure 3-27). *S. tokodaii* sR2 posses a 9 nt long D box antisense element while *S. acidocaldarius* sR2 posses a 11 nt D box antisense element for 23S rRNA. The predicted site of methylation is residue C1914 in 23S rRNA in both cases. In addition, the D box guide of *S. tokodaii* sR2 sRNA has a 10 nt base pair complementarity with a region of $tRNA^{Phe}$ (GAA). According to the N plus 5 rule, methylation is predicted for the C residue in the extra loop of the $tRNA^{Phe}$. This site has previously not been shown

to be methylated in other tRNA. It will be necessary to confirm this prediction experimentally by demonstrating that this site in $tRNA^{Phe}$(GAA) is methylated *in vivo*. No target has been assigned to the D' box guide.

```
            C box              D'box      C' box          *    D box
Sac sR2 GAG TGATGA GACGAGCGCTAA CAGA GAGAG TGAAGA GGTCACTGCGAA CTGA AGAAA
Sto sR2 TAG TGATGA GACGAGCGCACA CTGA TGAGA TGAAGA CGTCACTGCGTG CTGA TAAGA
```

Figure 3-27 Alignment of sR2 sRNA from *S. acidocaldarius* and *S. tokodaii*

sR2 sRNA from *S.acidocaldarius* and *S. tokodaii* are aligned. The conserved boxes are shaded in light grey. The antisense element of the sRNA is shown in dark grey. The asterisk indicates the nucleotide 5 nt up-stream of the D' box.

S. tokodaii sR2 sRNA is nearly completely encoded within another gene and overlaps the beginning of the assigned reading frame of a long conserved hypothetical protein - a putative endonuclease III (Figure 3-28). The predicted initiation codon lies partially in the C box of the sRNA. Endonuclease III has homologues in several other archaeal genomes that possess dissimilar sequence at the beginnings of the protein, but otherwise show clear similarity in the rest of the protein sequence.

```
Sso_endoIII     TTGATACAGGTACTCCAAATCTGGTTAAACATGTATTAAGTGAACTATAATGGTTCGTAA 60
Sto_endoIII     GTGATG-AGACGAGCGCACACTGATGAGA------TGAAGACGTCACTGCGTGCTGATAA 53
                 ****  **     *  *  *** * * *      * ***    *  *    * *  ***

Sso_endoIII     AATACTTGACACATTGCTGAAAATATTTGAAAATAATAAAAGCATATTGAAGGAAAAAGG cont 120
Sto_endoIII     GATAATTCAAAAACTTATCCAAGAATTTGAGAAGAATAAAGAGCTCTTAAGACAAGCTGG.cont 113
                 *** ** * * * *  *  **  ****** ** ******    * ** *   **   **
```

Figure 3-28 Alignment of the 5'end of the putative endonuclease III gene of *Sulfolobus solfataricus* and *Sulfolobus tokodaii*

The 5'end of the putative endonuclease III gene of *S. solfataricus* and *S. tokodaii* are aligned. The conserved boxes of *S. tokodaii* sR2 sRNA are shaded. The translation initiation codons of the endonuclease III are underlined. Identical nucleotides are marked by an asterisk.

sR3 sRNA

A BLAST search against *S. acidocaldarius* sR3 sRNA resulted in a hit (Score: 48 bits, E-value: 0.001) against an intergenic region between ST1295 (a 387aa long hypothetical type II DNA topoisomerase VI subunit A) and ST1296 (a 131aa long hypothetical translation initiation factor 5a) in the genome of *S. tokodaii* (Figure 3-29).

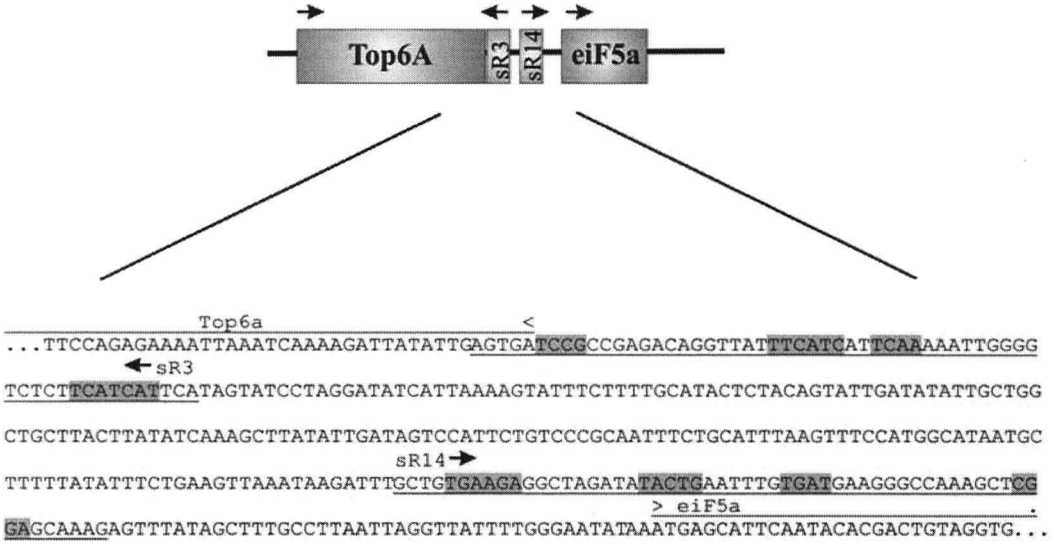


Figure 3-29 Intergenic region between the Top6A and ei5Fa gene

The arrangement of genes surrounding the sR3 and sR14 genes in *S. tokodaii* is depicted. Genes are not drawn to scale. The intergenic sequence between the topoisomerase VI subunit A (Top6A) and the translation initiation factor 5a (eiF5a) from *S. tokodaii* is shown. The C-terminal end of the Top6A gene and the N-terminus of the eiF5a genes are overlined; The sRNA genes are underlined with the conserved boxes shaded. sR3 is encoded on the minus strand - indicated by a left hand arrow. sR14 is encoded on the plus strand - indicated by a right hand arrow. sR3 and sR14 are 175 nt apart from each other.

The putative sRNA was designated *S. tokodaii* sR3. As with *S. acidocaldarius* sR3, *S. tokodaii* sR3 has a D box antisense element, that shows complementarity to 23S rRNA - the predicted methylation site is G2739 (Figure 3-30).

A second hit (Score: 40.1 bits, E-value: 0.29) in the genome of *S. solfataricus* has been previously reported (Omer *et al.*, 2000) and has been annotated in the genome of *S. solfataricus* as snoRNA_2. snoRNA_2 is encoded intergenically between the DNA topoisomerase VI subunit A and initiation factor 5a genes. In the same intergenic region, the *S. acidocaldarius* sR14 homologues sRNA is located, annotated as snoRNA_6.

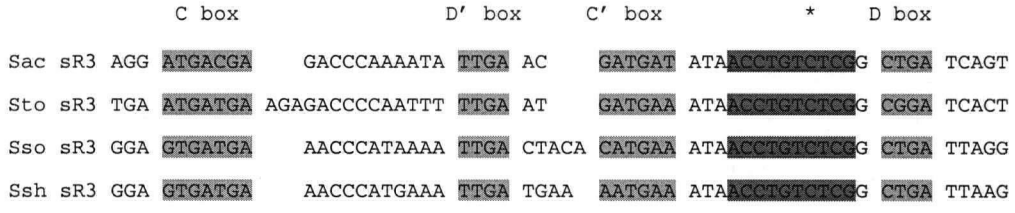


Figure 3-30 Alignment of sR3 from four *Sulfolobus* species

sR3 from *S.acidocaldarius, S. tokodaii, S. solfataricus* and *S. shibatea* are aligned. The conserved boxes are shaded in light grey. The antisense element of the sRNA is in dark grey. The asterisk indicates the nucleotide 5 nt up-stream of the D' box. sR3 from *S. solfataricus* has been assigned as snoRNA_2 in the *Sulfolobus solfataricus* genome.

S. acidocaldarius sR3 also "had a hit" near the *Sulfolobus shibatae* top6B topoisomerase II gene (Score: 40.1 bits, E-value: 0.28). This putative sRNA had been predicted previously (Omer *et al.*, 2000).

The genomic location of *S. acidocaldarius* sR3 is not known, since the complete *S. acidocaldarius* sequence is not available.

sR9 sRNA

A BLAST search (Score: 44.1 bits, E-value: 0.022) against *S. acidocaldarius* sR9 revealed a previously unidentified sRNA in *S. solfataricus* that is encoded intergenically. The new *S. solfataricus* sRNA was designated *S. solfataricus* sR99, since *S. solfataricus* sR9 has been previously assigned. *S. acidocaldarius* sR9 and the newly identified *S. solfataricus* sR99 sRNA are both predicted to guide methylation from their D' box antisense element to residue G926 in 16S rRNA (Figure 3-31).

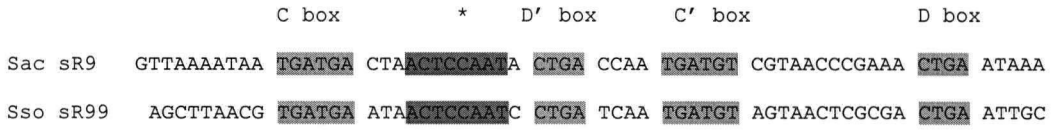


Figure 3-31 Alignment of sR9 sRNA from *S. acidocaldarius* and *S. solfataricus*

The alignment of sR9 and sR99 from *S. acidocaldarius* and *S. solfataricus* respectively is shown. The conserved boxes are shaded in light grey. The antisense element of the sRNA is shown in dark grey. The asterisk indicates the nucleotide 5 nt up-stream of the D' box.

sR14 sRNA

A BLAST search against *S. acidocaldarius* sR14 gene (shown in this study to guide methylation to the anticodon in tRNA), gave a hit (Score: 46.1 bits, E-value: 0.005) in the genome of *S. tokodaii* (Figure 3-32). I named this new sRNA *S. tokodaii* sR14. *S. tokodaii* sR14 is located in the same intergenic region as *S. tokodaii* sR3 sRNA - between a hypothetical type II DNA topoisomerase VI subunit A, and a hypothetical translation initiation factor 5a, but the two sRNAs are not linked together (Figure 3-29). Like *S. acidocaldarius* sR14, *S. tokodaii* sR14 sRNA is predicted to guide methylation of U34 in $tRNA^{Gln}$ from its D box antisense element. Interestingly, no hit was obtained against the

genome of *S. solfataricus.* However, the intergenic region between DNA topoisomerase VI and the inititiation factor 5a in *S. solfataricus* contains a homologue of *S. acidocaldarius* sR14 - annotated as snoRNA_6. The sequence of sR14 is well conserved only in the 3' end.

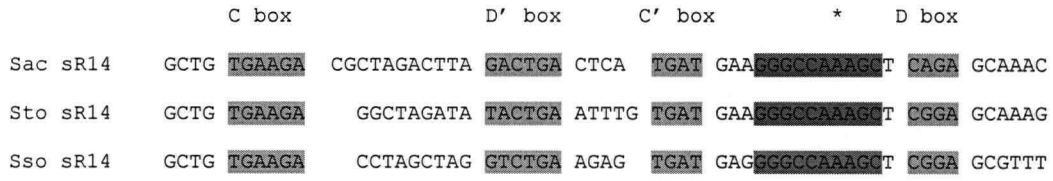


Figure 3-32 Alignment of sR14 sRNA from three *Sulfolobus* species

sR14 from *S.acidocaldarius, S. tokodaii* and *S. solfataricus* are aligned. The conserved boxes are shaded in light grey. The antisense element of the sRNA is shown in dark grey. The asterisk indicates the nucleotide 5 nt up-stream of the D' box. sR14 sRNA from *S. solfataricus* has been assigned as snoRNA_6 in the *Sulfolobus solfataricus* genome.

sR17 sRNA

A Blast search against *S. acidocaldarius* sR17 revealed a hit (Score: 36.2 bits, E-value: 0.002) in the *S. tokodaii* genome. The putative sRNA overlaps the C terminus of ST0488, a 301 long conserved hypothetical protein, by 8 nucleotides. A BLASTP search revealed that this hypothetical protein might be a homologue to the ribosomal protein S6 modification protein. In my initial search for *S. acidocaldarius* sR17 homologues, no significant hits towards the genome of *S. solfataricus* were obtained. However, after examination of the C terminus of the *S. solfataricus* ribosomal S6 modification protein, a putative homologous sRNA was discovered. The genomic location of the *S. acidocaldarius* sR17 gene is not known, since the genome is not available. sR17 sRNAs

from all three Sulfolobus species show dissimilarities in their D box and D' box antisense elements (Figure 3-33).

A target for *S. acidocaldarius* sR17 has not been identified. The homologous *S. tokodaii* sR17 sRNA shows a D box antisense element with a short six base pair complementarity towards 16S rRNA (residue A689) and an eight base pair complementarity towards 23S rRNA - the residue predicted to be methylated is A1043. Both *S. acidocaldarius* sR17 and *S. solfataricus* sR17 D box guide regions show complementarities with mismatches towards the same 23 sRNA region, and either a short sequence complementarity or a complementarity with multiple mismatches, towards the 16S rRNA region. Although the D' box guides of the sR17 genes exhibit a higher sequence similarity then in their D box guides, no target could be identified for the D' box guide of any of these sRNAs.

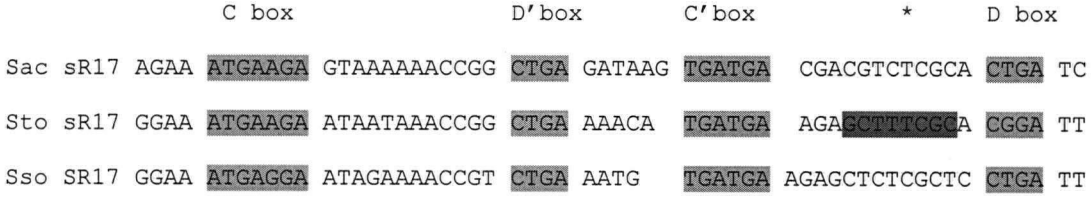


Figure 3-33 Alignment of sR17 sRNA from three *Sulfolobus* species

Alignment of sR17 sRNA from *S. acidocaldarius*, *S. tokodaii,* and *S. solfataricus*. No target could be identified for the D' box guide. A short complementarity to 16S and 23S rRNA could only be detected for *S. tokodaii* sR17 the complementarity for 23S rRNA is indicated in dark grey.

sR18 sRNA

A BLAST search with *S. acidocaldarius* sR18 sRNA gave a very strong hit (Score: 52.0 bits, E-value: 8e-05) in the genome of *S. tokodaii*. I termed this putative new sRNA *S. tokodaii* sR18 (Figure 3-34). The *S. tokodaii* sR18 sRNA gene overlaps the C-terminal end of ST2224, a 283aa long conserved hypothetical protein coding region, by 32 nt. A similar protein could be identified in *S. solfataricus*, but the regions of similarity are restricted to the core of the protein, and no putative sRNA could be detected in the C-terminal region of the *S. solfataricus* protein. *S. acidocaldarius* sR18 sRNA is predicted to direct methylation with its D' box antisense element to residue G140 in 23S rRNA, with a mismatch in the complementarity at position + 7 from the D' box.

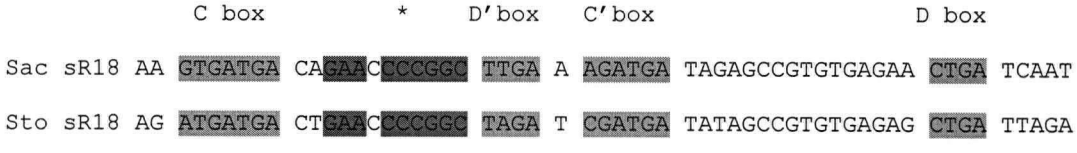


Figure 3-34 Alignment of sR18 sRNA from *S. acidocaldarius* and *S. tokodaii*.

sR18 from S.acidocaldarius and S. tokodaii are aligned. sR18 shows a nine base pair long complementarity with a mismatch at position 7 from its D' box guide to 23S rRNA. The complementarity is indicated in dark grey. No potential target could be identified for the D box guide sequence.

sR20 sRNA

A BLAST search with *S. acidocaldarius* sR20 showed a very strong hit (Score: 105 bits, E-value: 5e-21) in the genome of *S. solfataricus* and a much weaker hit (Score: 34.2 bits, E-value: 16) in the genome of *S. tokodaii*. The two putative sRNAs were named *S. solfataricus* sR20 and *S. tokodaii* sR20. *S. acidocaldarius* sR20 and *S. solfataricus* sR20 show a remarkably degree of similarity (Figure 3-35).

The D' box guide antisense element shows complementarity to 23S rRNA and the residue predicted to be methylated is C2065. Although the D box guide antisense element is preserved in all three *Sulfolobus* sR20 sRNAs, no potential target could be identified.

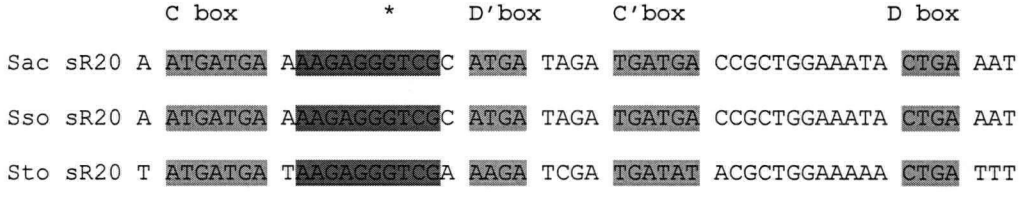


Figure 3-35 Alignment of sR20 from three *Sulfolobus* species

sR20 sRNA from *S. acidocaldarius*, *S. solfataricus*, and *S. tokodaii* are aligned. The complementarity is indicated in dark grey. The asterisk indicates the nucleotide 5 nt up-stream of the D' box.

In the genome of *S. solfataricus* Sso sR20 is located on the minus strand, intergenically between the genes for $tRNA^{Phe}$ and $tRNA^{Gly}$. *S. tokodaii* sR20 sRNA is located between ST0405 (a 220aa long hypothetical protein) and the gene for $tRNA^{Gly}$.

sR22 sRNA

S. acidocaldarius sR22 is a predicted double guide. The D box guide antisense element of *S. acidocaldarius* sR22 displays a ten basepair complementarity to 23S rRNA in which residue U2893 is predicted to be methylated. It also shows a ten basepair complementarity to 16S rRNA with one mismatch; residue U1310 is predicted to be methylated. The D'box guide antisense element displays an eleven basepair complementarity to the region of residue C2916 in 23S rRNA. A BLAST search with *S. acidocaldarius* showed a weak hit (Score: 24 bits, E-value: 21) in the genome of *S. tokodaii*. The new sRNA was named *S. tokodaii* sR22 (Figure 3-36). *S. tokodaii* sR22 sRNA is located in an intergenic region between ST1710 (a 146aa long hypothetical transcriptional regulator) and ST1711 (587aa long hypothetical cell division control protein), and shows high similarity to the *S. acidocaldarius* guide regions. Therefore, the predicted target sites are residue U2893 in 23S rRNA and/or residue U1310 in 16S rRNA for the D box guide and C2916 in 23S rRNA for the D' box guide element.

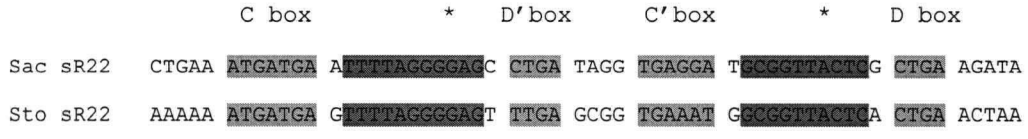


Figure 3-36 Alignment of sR22 from *S. acidocaldarius* and *S. tokodaii*

sR22 sRNA from *S. acidocaldarius* and *S. tokodaii* are aligned. The complementarity is indicated in dark grey. The asterisk indicates the nucleotide 5 nt up-stream of the D' box.

sR26 sRNA

Two new sRNA where identified in a homology search for *S. acidocaldarius* sR26 sRNA (Figure 3-37).

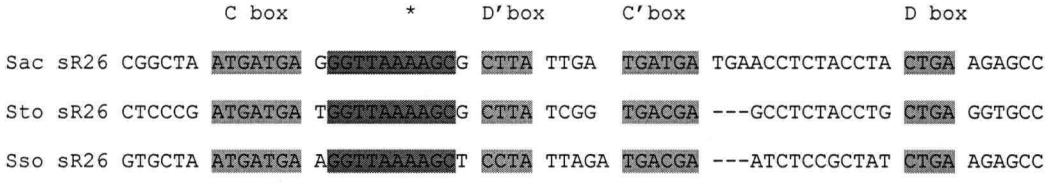


Figure 3-37 Alignment of sR26 from three *Sulfolobus* species

sR26 from *S. acidocaldarius* , *S. tokodaii* and *S. solfataricus* are aligned. The complementarity is indicated in dark grey. The asterisk indicates the nucleotide 5 nt up-stream of the D' box.

In the genomes of *S. tokokodaii* (Score: 40.1 bits, E-value: 0.34) and *S. solfataricus* (Score: 36.2 bits, E-value: 5.4) a BLAST search revealed two sRNA, that I designated *S. tokodaii* sR26 and *S. solfataricus* sR26 sRNA, respectively

The *S. tokodaii* sR26 sRNA gene is situated intergenically between a hypothetical enolase (ST1212) and a 238aa long hypothetical protein (ST1213). The *S. solfataricus* sR26 sRNA gene is located between a conserved hypothetical protein (SSO0911) with sequence similarity to the *S. tokodaii* hypothetical protein ST1213, and a hypothetical protein with no homologues in any other organism (SSO6778).

No target had been previously assigned to sR26. A genomic search for potential targets for sR26 sRNA revealed that the D' box guide antisense possesses a ten basepair complementarity to the $tRNA^{Lys}$ (UUU) anticodon region. The predicted site of methylation is the wobble base at U34.

4 Discussion

In Eukaryotes the site-specific formation of one of the most prevalent rRNA modified nucleotides, 2'-O- methylation of the ribose, is directed by the family of C/D box snoRNAs. These snoRNAs exert their function through formation of a canonical guide RNA-target RNA duplex at the modification site. Initial studies identified fibrillarin as the signature protein cofactor associated with C/D box snoRNAs, but the repertoire of associated proteins has now been expanded to include the two paralogous proteins, Nop56p and Nop58p, and 15.5 kD/Snu13p, which is found in both the C/D box snoRNPs and the U4 spliceosomal snRNP. Conversely, the few 2'-O-methylations found in the typical bacterium *E. coli* appear to be catalyzed by site- specific protein enzymes, without any RNA cofactor.

Archaea are prokaryotic organisms that are distinct from Bacteria and are believed to be related to earlier Eukaryotes. By utilizing biochemical and computational methods, the presence of up to 50 or more distinct C/D box small RNAs has been demonstrated in several species of hyperthermophilic Archaea (Omer *et al.*, 2000). The archaeal sRNAs not only guide methylation of numerous rRNA residues during ribosome biogenesis, but also of 21 different positions within various tRNAs. The genome of the archaeon *Sulfolobus solfataricus* encodes three proteins designated aFib, aNop56 and aL7a, that are homologues of the human nucleolar proteins fibrillarin, Nop56p and Nop58p and 15.5 kD. The *Sulfolobus* aL7a protein has been annotated as a ribosomal protein although it exhibits greater sequence similarity to the 15.5 kD protein than to either of the paralogous human L7a or S12 ribosomal proteins.

4.1 Methylation assay and specificity

A *Sulfolobus* box C/D sRNP complex has been reconstituted *in vitro* from purified recombinant components, and was shown to have methylation activity when provided with a simple target oligonucleotide, complementary to the sRNA D box guide sequence. Up to 2 pmoles of product were formed for each pmole of protein present in the reaction (within the limits of the accuracy of protein concentration estimate), indicating that the catalytic system can carry out multiple turnovers. Nucleotide mapping demonstrated that methylation was targeted specifically to the predicted position in the 16S rRNA substrate.

The core catalytic complex in Archaea is much simpler than anticipated and requires only three proteins: the core RNA-binding aL7a protein, the aNop56 protein and the aFib protein which appears to possess the methyltransferase activity (Omer *et al.*, 2002). It seems surprising that at least in the *in vitro* methylation, no RNA helicase is required. The absence of a helicase requirement, may be related to the high temperature of the *in vitro* reaction or to the short length of the guide-target interaction. Reactions performed with full length 5S rRNA and a short oligonucleotide substrate show minimal differences in efficiency. It remains a possibility that for more complex and dynamic *in vivo* rRNA interactions, a helicase might be necessary for recycling of the snoRNA-rRNA and snoRNA-protein interactions that occur during the processing of rRNA precursors. In addition, helicases might facilitate changes in intramolecular base pairing

of rRNA, to expose binding sites for ribosomal or non-ribosomal proteins required for pre-rRNA processing.

It has been shown that efficient methylation requires that the C/D box and the C'/D' box function in *cis* (Tran *et al.*, 2003) . There are suspected conformational changes in the sRNA which occur upon protein binding. Both protein-protein and protein-RNA contacts are important in assembly. While the order of protein addition to the sRNP is known, the actual stoichiometry is presently unclear. aNop56 associates with the sRNA following aL7a binding and is then joined by aFib (Omer *et al.*, 2002). The binding of aNop56 is enhanced in the presence of aFib, suggesting that these proteins likely associate prior to sRNA binding. aNop56 clearly can bind alone but *in vivo* aNop56 and aFib likely exist as a heterodimer. When both proteins are present, they are probably added in a single step.

4.2 The importance of being methylated

An important question concerns the function of extensive rRNA methyl modification in Eukaryotes and Archaea compared with the few modified nucleotides in Prokaryotes. The conservation of this complex machinery responsible for directing the modifications (all clustered in the functional core of the rRNA) indicates that these modifications must be important for insuring efficient ribosome subunit assembly or ribosome function. Although genetic depletion of single or multiple guide snoRNA genes in yeast had no obvious effect on cell growth (Lowe and Eddy, 1999), inactivation of the snoRNA associated methyl transferase protein (fibrillarin) results in a lethal phenotype

(Tollervey *et al.*, 1993), suggesting that the total lack of 2'-O-ribose methylation may be lethal to the cell.

In hyperthermophilic Archaea, studies have suggested increased levels of ribose methylation as a function of culture temperature in both tRNA and rRNA (Noon *et al.*, 1998). It is believed that ribose methylation contributes to the stabilization of secondary structure in RNA. In mesophiles, RNA structure is stabilized by elevated G/C contents of stems, which in turn loosely correlates with optimal growth temperature and with overall level of posttranscriptional modification (Galtier and Lobry, 1997). However, in the case of thermophiles, the G/C content in stems is maximized and modification becomes more important. The effect of modifications has been determined directly by studying the correlation between culture temperature for a given organism and increase in RNA melting temperature, or has been implied from an increase in the level of specific modifications, such as 2'-O- methylation of the ribose in tRNA and rRNA, as a consequence of increased culture temperature (Noon *et al.*, 1998).

Thermodynamically, 2'-O-methylated nucleosides favor the C3'-endo ribose pucker, which minimizes steric interference of the base and C3'-phosphate with the ribose 2'-O-methyl group (Kawai *et al.*, 1992). In a pento-furanose ring, this conformation affects the backbone torsion angle and correlates with a low anti-glycosidic bond angle. The 2'-O-methyl group decreases the entropic value of forming a duplex structure and therefore will stabilize the base-paired form (Yokoyama *et al.*, 1981). Ribose methylation of specific nucleotides could, therefore, be a common and frequently occurring method of stabilizing functionally important secondary structure in RNA of archaeal thermophiles.

4.3 Non-ribosomal targets - tRNA

More than 80 modified ribonucleosides have been identified in tRNAs (Rozenski *et al.*, 1999); a small number of these modifications occurs in the tRNA of almost all organisms (Motorin, 1998). In addition to this conserved core, Archaea, Bacteria and Eukaryotes each make domain specific post-transcriptional modifications to their tRNAs. Although the specific function of many tRNA modifications remains elusive, studies have generally indicated that modifications are required for optimal growth and translation; this implies that, all of these modifications are believed to play important rolls in the stabilization of the secondary and tertiary structure of the tRNA or its function in protein synthesis.

Methylation is the simplest modification known. Both nucleobase and sugar moiety of a nucleoside may be methylated (Agris, 1996). 2'-O-methylation of the sugar stabilizes the ribose C3'-endo form (Kawai *et al.*, 1992). The local conformational rigidity conferred by this modification may affect RNA stability and RNA-protein interactions. 2'-O-methylation of the ribose at position 34, the wobble nucleotide of the anticodon, has been reported to be important for translation efficiency (Satoh *et al.*, 2000) and is believed to contribute to specificity and stability in the codon-anticodon interaction.

A global analysis of the guide sequences in all available archaeal C/D box sRNAs, predicts that many sRNAs target methylation to 21 different sites in pre-tRNAs (Dennis *et al.*, 2001). In all instances, the predicted position of methylation corresponds to one of the 21 sites of documented ribose methyl modification in tRNA, but never to

positions where ribose methyl modification has never been observed. In some cases a single sRNA directs methylation to only a single species of tRNA whereas in other instances a single sRNA can direct methylation to up to 19 different tRNAs, based on conservation of the target complementarity within the tRNA sequences (Dennis *et al.*, 2001). The most frequently targeted position is C34 or U34, the wobble position in the anticodon of the tRNA. These observations suggest that non-ribosomal RNAs may be substrates for sRNA guide-directed methylation in Archaea. To test this proposal, the archaeal C/D box sRNAs sR11 from *S. solfataricus* and sR14 from *S. acidocaldarius* were used as models to investigate sRNA mediated methylation in archaeal tRNA. The respective sRNAs were reconstituted into RNP particles *in vitro* and tested for the ability to guide methylation to position G18 and U34 in *Sulfolobus* $tRNA^{Gln}$.

4.3.1 C'/D' motif guided methylation to tRNA - *S.solfataricus* sR11 sRNA

The well characterized *S. acidocaldarius* sR1 containing sRNP uses a D box guide to direct methylation to position U52 in 16S rRNA. A similar experimental protocol utilizing *S. solfataricus* sR11 containing sRNP complex, was used to demonstrate that the D' box guide can direct methylation to G18 in tRNAGln. Although the sR11 sRNA possesses a variation of the classical k-turn fold in its internal C'/D' motif, it is clearly able to form an active k-turn at this motif as judged by methylation activity. Methylation was detected in the C'/D' guided substrate, but inhibited when an A to G mutation at the position five nucleotides from the D' box was introduced; this

mutation disrupted the crucial Watson-Crick pairing at the site of methylation. Methylation was restored to near wildtype level when a G18A tRNA mutation restoring the Watson -Crick base pairing was used together with the mutated sRNA; this supports the idea that the methylation is site-specific and requires a Watson Crick base pair at the site of methylation).

In Archaea, the box C'/D' and box C/D motifs are well conserved. Both motifs have a similar mechanism by which they select the site for 2'-O- methylation - the modified nucleotide is positioned five nucleotides upstream from the D or D' box. The internal C'/D' box probably adopts the same RNA fold as the well characterized terminal C/D box element, resulting in an symmetric arrangement of proteins within the C/D box sRNP. This implies that the method of recruiting proteins for the specific modification of RNA is most likely identical for both the C/D and C'/D' motifs.

The situation might be somewhat different in Eukaryotes, where C' and D' sequences are often less conserved and not easily identified in snoRNAs. Of particular note is the lack of strict conservation of the GA nucleotides of the eukaryotic C'/D' motif, which are crucial for establishing the k-turn structure. This suggests that in Eukaryotes the C'/D' motif does not adopt a canonical k-turn structure, and therefore will not bind the 15.5 kD protein. Consistent with this line of reasoning, recent work has revealed the asymmetric distribution of core proteins upon the terminal C/D and internal C'/D' motifs (Cahill *et al.*, 2002; Szewczak *et al.*, 2002). Nucleotide modification experiments have shown that the 15.5 kD core protein binds exclusively to the terminal C/D box motif and that its binding is primarily responsible for the stability of the C/D box snoRNAs *in vivo*.

Comparison of the archaeal and eukaryotic box C/D RNPs reveals two structurally distinct complexes. The eukaryotic snoRNP is comprised of four core proteins compared with three for archaeal sRNPs. Duplication of an ancestral gene has resulted in the two paralogous genes for Nop56p and Nop58p. *In vivo* cross-linking experiments revealed that these core proteins Nop56p and Nop58p differentially recognize the C'/D' and C/D motif respectively. Nop56p has evolved such that the 15.5 kD protein is no longer necessary for the Nop56p and fibrillarin binding to the C'/D' motif. These results suggest that despite using homologous proteins for RNP construction, archaeal and eukaryotic C/D box RNAs may be noticeably different. The archaeal complex appears at least in some instances to be more symmetrical than the typical eukaryotic complex.

4.3.2 tRNA (Gm18) Methyltransferase

Methylation of the guanosine residue at position 18 (Gm18) is widely distributed in tRNAs of Bacteria, Eukaryotes, Archaea, and plant mitochondria (Sprinzl *et al.*, 1998). There is no evidence that Prokaryotes possess an RNA-based methylation system; enzymatic activity of tRNA (Gm18) methyltransferase has been shown in *E.coli*, and the tRNA (Gm18) methyltransferase protein has been purified from *T. thermophilus*. It has been suggested that the recognition element for the enzyme consists of the D loop of the tRNA and the presence of G18 and G19, although there clearly has to be another determinant or anti-determinant in the recognition element, since all *E.coli* tRNAs carry these features, and only 13 tRNA species (out of 45 sequenced) contain the nucleoside modification.

In yeast, several enzymes responsible for the post-transcriptional methylation of tRNA have been characterized and named Trm (tRNA-methyltransferase). There is no evidence to date that any of the numerous nucleotide modifications in eukaryotic tRNAs are snoRNA-directed. A search of the yeast *S. cerevisiae* genome for candidate C/D snoRNAs able to target yeast tRNA 2'-O-methylation was negative. No likely tRNA targets could be identified either for any of the mammalian C/D snoRNAs devoid of rRNA or snRNA complementarities, suggesting that in yeast 2'-O-methylation of residues in tRNA is enzyme based, without any RNA cofactors. In fact, the 2'-O-tRNA MTase Trmp3 has been shown to be required for site-specific formation of methyl G at position 18 in the D-loop of intron-less tRNA (Cavaille *et al.*, 1999). Interestingly, in the same study, five putative RNA 2'-O-ribose methylases were identified in four Archaea species from the branch of Euryarchaeota. However, none of these putative methylases has been characterized and their specificity is unknown.

Although the majority of structurally characterized SAM- dependent methyltransferases (MTases) adopt a very similar three-dimensional fold, they show minimal sequence conservation apart from several motifs defining the AdoMet-binding face of the common catalytic domain; this makes it virtually impossible to define the substrate based on sequence similarity.

4.3.3 Importance of 2'-O methylation of G18 in tRNA

Residues G18G19 are universally conserved residues located in the D loop of the tRNA and make important contributions to the formation of the L-shaped three-

dimensional structure by virtue of the D-loop/T-loop interaction through the tertiary base pairs G18-Ψ55 and G19-C56. In *E.coli*, the absence of the methyl group at G18 in tRNA has no apparent effect on growth or on efficiency of decoding during translation (Persson *et al.*, 1997). Similarly, Gm18 is not essential in *S. cerevisiae*, and disruption of the gene encoding for Trm3, the methylase responsible for modification at this position in yeast, does not alter cell growth or translation fidelity (Cavaille *et al.*, 1999). However, disrupted yeast strains were slightly more resistant to paromomycin, an aminoglycoside antibiotic that interferes with protein synthesis by causing codon misreading by binding to 30S ribosomal A-site RNA (Fourmy *et al.*, 1996). The molecular basis of increased resistance to paromomycin is not understood. Artificial ribose methylation of G18 in tRNAs which are naturally devoid of methylation in this position has no detectable effects on either amino acid acceptor activity or melting temperature of the tRNA (Kumagai *et al.*, 1982). On the other hand, an extensive change in elution profiles of $tRNA^{Ser}$ from rat hepatoma cells that naturally lack Gm18, but otherwise are identical to normal rat liver $tRNA^{Ser}$, has been observed (Randerath *et al.*, 1981). This finding suggests that the tRNA lacking the modification has an altered tertiary structure. It is, therefore, likely that 2'-O methylation of ribose G18 might be a factor in stabilizing the tertiary structure of the tRNA. The G18 in tRNA forms hydrogen bonds with residue Ψ55 in the T-loop; these hydrogen bonds constitute an important part of the connections between the D and T-loops that stabilize the L form tertiary structure of the tRNA molecule. The 2'-*O*-methylation of the ribose of G18 may contribute to the rigidity of the nucleoside and thus stabilize the Gm18-Ψ55 base pair and the interaction between the D- and T-loop in tRNA molecules.

4.3.4 Methylation of the wobble position

tRNA position 34 reads the third codon base, and non-Watson-Crick wobble base pairing is allowed at this position. This position is also commonly modified on either the base or the ribose, with many tRNA species having a 2'-O-methylated nucleoside at the first position of the anticodon. 2'-O-methylation has never been found in the second or third positions of the anticodon of the tRNA from any organism.

The most frequent residues targeted by the predicted archaeal guide RNAs are C34 and U34, the wobble position of the anticodon of the tRNA. Position 34 is known to be methylated in many tRNAs from Prokaryotes and Eukaryotes. For example Gm34 is present in all eukaryotic $tRNA^{Phe}$, except those from tumor cells (Kuchino *et al.*, 1982). And Cm34 is present in tRNAs specific for leucine, serine, tryptophan, and methionine, whereas Um34 has only been detected in eukaryotic $tRNA^{Gln}$ and has been predicted in Archaea.

Modification of position 34, the wobble nucleotide of the anticodon, has been reported to be important for translation efficiency and fidelity (Curran, 1998; Satoh *et al.*, 2000). Modifications in the anticodon region of tRNA also has been shown to dramatically alter codon specificity (Curran, 1998) and aminoacyl-tRNA synthetase recognition (Giege *et al.*, 1998). Although modification of the anticodon is common among Eukaryotes and Prokaryotes, little is known about the enzymes involved. In yeast, Trm7p has been shown to catalyse the formation of 2'-O-methylribose in the yeast tRNA anticodon loop (Pintard *et al.*, 2002b). Trm7p belongs to a group of three yeast proteins (Trmp7, Spb1p and Mrm2p) that exhibit sequence similarities to the *E.coli* heat shock

protein and rRNA MTase FtsJ/RrmJ. The unique enzyme FtsJ/RrmJ methylates the peptidyltransferase centre of the large rRNA and has been shown to be able to methylate tRNA *in vitro* (Bugl, 2000). The *in vitro* ability of FtsJ/RrmJ to recognize and methylate certain tRNAs could represent a dual substrate specificity of FtsJ/RrmJ.

However, in an effort to determine the *in vivo* methylation target, no unambiguous differences in elution profiles of tRNA nucleosides prepared from wildtype and FtsJ/RrmJ deletion strains could be detected (Bugl, 2000). It still has to be determined which individual isoacceptor tRNA species, if any, serves as efficient methyl acceptor for FtsJ/RrmJ *in vivo*.

The *S. acidocaldarius* sR14 sRNA and $tRNA^{Gln}$ were used as a model to investigate modification of the wobble base by guide sRNA in *Sulfolobus*. The sRNP complex assembled with *S. acidocaldarius* sR14 sRNA was capable of methylating $tRNA^{Gln}$ (UUG) as substrate, but the RNP complex was inactive when the isoacceptor $tRNA^{Gln}$ (CUG) was used as substrate. Methylation with the isoacceptor $tRNA^{Gln}$ (CUG) is detected when a mutated *S. acidocaldarius* sR14 is used that restores Watson-Crick base pairing at the site of methylation, confirming that a Watson-Crick base pair is required at this position.

Archaeal $tRNA^{Trp}$

S. acidocaldarius sR14 sRNA guides methylation to the anticodon in a *trans* target. A particular interesting example for a *cis*-acting sRNA is intronically encoded in the $tRNA^{Trp}$ of many euryarchaeal species. The intron has been shown to be essential for tRNA maturation (Armbruster and Daniels, 1997). The guide sequence within the intron

is required for methylation of two positions of the anticodon within the exon regions of the precursor tRNA. The unusually large $tRNA^{Trp}$ is so far the sole archaeal intron displaying the hallmarks of a C/D box sRNA. A search of yeast tRNA introns for C/D box motifs was negative.

4.3.5 Why do Archaea utilize sRNAs for tRNA modification?

No evidence for guided RNA modification of tRNA has been found in Bacteria and Eukaryotes. Instead, all known tRNA MTase carry out site-specific modification of one or several strictly defined nucleosides and are believed to recognize their target directly by using both, sequence and three-dimensional structure (Grosjean *et al.*, 1996). Many of these enzymes have been identified and characterized in bacteria, yeast and higher Eukaryotes, but few enzymes have been identified or characterized in Archaea (Bjork, 1987; Garcia, 1998). This might suggest that at high temperatures, many archaeal tRNAs may not fold into a three dimensional structure that is sufficiently stable to be recognized by a protein enzyme. To overcome the absence of stable three dimensional structure as a determinant for modification, Archaea appear to employ, at least in some instances, sRNA methylation guides that direct ribose modification to critical positions within precursor tRNAs. Many of these modifications are believed to play a role in stabilizing higher order structure within the tRNA.

4.4 Modification of 5S rRNA

Posttranscriptional modifications of nucleotides are extremely rare in 5S rRNA. None are reported from Bacteria, whereas pseudouridine is present in several lower eukaryal RNAs (Szymanski *et al.*, 2002). However, ribose methyl modifications to position C32 in 5S rRNA have been reported in both *Sulfolobus acidocaldarius* and *Pyrodictium occultum* (Bruenger *et al.*, 1993). Protein methyltransferase enzymes responsible for these modifications have not been identified.

In *S. acidocaldarius*, the D' box guide of *S. acidocaldarius* sR10 sRNA displays a short eight base pair complementarity to the region spanning residue C32 region of 5S rRNA, with a noticeable mismatch at position +9 of the D' box guide. Methylation could only be detected after extension of the complementarity and elimination of the mismatch in the duplex. My *in vitro* results suggest that *S. acidocaldarius* sR10 sRNA has little or no activity on native 5S rRNA. The absence of activity appears to be due to the limited complementarity between the guide and the target.

The modified C is located in a region that is clearly conserved with respect to both sequence and position in other organisms; however, *Sulfolobus acidocaldarius* and *Pyrodictium occultum* are the only organisms where a 2'-O-methylation in 5S rRNA has been reported. The exact role of the modification is not known, but it is possible that its presence provides thermal stability to the 5S rRNA molecule. Methylation of the 2'-hydroxyl group of the sugar stabilizes the C3'-endo ribose conformation in the cytidine nucleotide, favoring an A-type helical conformation of the RNA.

The absence of an identified methyl transferase enzyme and of demonstrable *in vitro* activity of sR10 sRNA on a 5S rRNA substrate, make it unclear if this modification is sRNA guided or protein-mediated. Searches for other sRNAs in *Sulfolobus* with more extended complementarity to the C32 region of 5S rRNA have thus far failed to identify any other candidates. There are numerous ways to explain this puzzle. Two of the simplest are that *S. acidocaldarius* contains a specialized protein methyl transferase to make this modification and that the limited complementarity between the sR10 sRNA and the 5S rRNA is fortuitous or that the sR10 sRNA is active *in vivo*, but not *in vitro*. Another possibility is that sR10 sRNA may have been involved in this modification in the past but now has lost activity because of the accumulation of mutations in the guide region. The presumed gain and loss of guide function in archaeal sRNAs through the accumulation of mutations in guide regions has been documented (Dennis *et al.*, 2001). If sR10 sRNA has lost the ability to direct methylation to position C32, it implies that either sR10 sRNA initially worked in concert with a protein methyltransferase which now functions alone or that a new methyltransferase evolved to carry out this modification making the sR10 sRNA guide function redundant.

In addition to the limited complementarity to 5S rRNA the *S. acidocaldarius* sR10 D' box guide also shows a 12 base pair complementarity in its D' box guide to a region of $tRNA^{Gly}$ - residue C50 is predicted to be methylated, and a 9 base pair complementarity in its D box guide to the C2539 region in 23S rRNA. Since my *in vitro* results suggest that *S. acidocaldarius* sR10 sRNA has little activity on 5S rRNA, it is therefore likely that the only natural targets for this sRNA are 23S rRNA and $tRNA^{Gly}$.

4.5 Double guide

It has been estimated that at least 50% of archaeal sRNAs guide methylation using both 5' and 3' guide regions, whereas only about 20% of human and yeast snoRNAs have been reported to be double-guides. The two antisense elements of a given sRNA almost always match sequences within the same rRNA and their cognate target sites are generally very close to each other along the rRNA sequence, in most cases less than 50 nt apart. For a few *Pyrococcus* sRNAs with a pair of antisense elements matching targets much further apart in the rRNA primary sequence, the two rRNA targets appear, in most cases, remarkably close to each other in the rRNA secondary structure, such as for sR36 and sR8 (121 and 285 nt apart in 16S rRNA, respectively) and sR28 and sR55 (132 and 182 nt in 23S rRNA apart, respectively). This noticeable proximity of the two targets rRNA suggests that formation of the sRNA-target duplex in *Pyrococcus* might take place simultaneously. This would be in contrast to what seems likely in the few eukaryal C/D box snoRNA carrying a pair of antisense guide elements, in which there is no apparent correlation between the locations of methylation sites for double guide snoRNAs.

In Archaea, the two target site sequences can be in either a direct or reverse orientation, relative to the sRNA guide sequences. sR24 sRNA from *Pyrococcus* and sR7 sRNA from *S. acidocaldarius* were chosen as examples to examine the function of a direct and a reverse guide, respectively.

4.5.1 Direct Double Guide - *Pyrococcus* sR24 sRNA

The sRNP complex assembled on native sR24 sRNA is capable of methylating substrates using either the terminal C/D box or the internal C'/D' box guide. However, maximum activity was detected when both sites were present in *cis*, suggesting that the two guides work in concert to stabilize the complex. Nucleotide mapping showed that both target sites are methylated. The order of the interaction between the two target sequences is not known.

The formation of this bipartite RNA duplex involving a total of 22 base pairs provides thermodynamic stability to the sRNA-rRNA complex and could facilitate necessary RNA-protein and protein-protein interactions within the sRNP. Biochemical and mutagenesis studies with truncated sRNA constructs have shown that efficient methylation requires the juxtaposed and symmetrically assembled C/D and C'/D' RNPs (Tran *et al.*, 2003). aL7a binds the C/D and C'/D' box motif in a cooperative manner which probably reflects induced changes in sRNA structure or protein-protein interactions following initial binding of the protein. In addition, the symmetric nature of the aFib-aNop56 complex and the C/D sRNA has been demonstrated to be functionally important (Rashid *et al.*, 2003). Simultaneous binding of the target to the sRNA guides might stabilize the symmetric structure of the complex. The widespread occurrence of sRNA carrying two guide regions in hyperthermophilic Archaea could therefore reflect the constraints related to the extreme environment these organisms live in and/or the compactness of their genome.

4.5.2 Are double guides RNA chaperones?

It has been long proposed that in addition to the methylation guide function, snoRNAs have a chaperone function in pre-rRNA folding (Bachellerie *et al.*, 1995; Dennis *et al.*, 2001; Maxwell and Fournier, 1995) . Ribose methylation takes place early in ribosome biogenesis, on the elongating pre-rRNA transcript. The folding of the nascent rRNA must be largely constrained by formation of numerous rRNA/snoRNA guide duplexes. It has been suggested that as a result snoRNAs have an intrinsic chaperone function in pre-rRNA folding in addition to their methylation guide function (Steitz and Tycowski, 1995). The base pairing of guide RNA and pre-rRNA during the assembly process may direct the efficient localized folding of the rRNA, thereby preventing the formation of non-productive structures that block or delay the ribosome assembly or may stabilize rRNA tertiary structure during ribosome biogenesis. Higher growth temperatures may require more stabilizing of the RNA structure and more chaperones to direct a correct folding process.

This idea has been given further support with the recent identification of U106 snoRNA in human (Vitali *et al.*, 2003). This snoRNA possesses two conserved 9-10 nt long antisense elements matching two 18S rRNA segments devoid of 2'-O-methylated nucleotides. Interestingly, the two sites map very close to each other within the rRNA secondary structure, potentially supporting a role as RNA chaperone.

In summary ribose methylation guide RNA may play an important role in rRNA folding. The direct double guide might, therefore, illuminate the functional significance of the link between chaperone and methylation function.

4.5.3 Inverted Double Guide - *S. acidocaldarius* sR7 sRNA

To probe the function of reverse double guides, *S. acidocaldarius* sR7 sRNA was used. Although the highest methylation activity was detected in the target containing both methylation sites on one molecule, methylation activity was poor compared to the control reaction. An alteration in the target substrate removed the necessity to form the pseudoknot structure and instead results in a "direct guide" but did not improve the overall methylation activity (Figure 4-1).

The low activity of *S. acidocaldarius* sR7 might be due to secondary structure concerns in the sRNA. *S. acidocaldarius* sR7 sRNA shows several unusual features. The lack of a terminal stem or a terminal base pair might contribute significantly to the instability of sR7 sRNA and might prevent the formation of a stable k-turn in the C/D motif. *S. acidocaldarius* sR7 was identified in a biochemical screen for sRNA in *S. acidocaldarius*; therefore, it is possible that the missing stem arose from a shortened sequence in the isolated clone and does not reflect the complete sequence of *S. acidocaldarius* sR7 sRNA. Preliminary data from the *Sulfolobus acidocaldarius* sequencing project indicate that indeed *S. acidocaldarius* sR7 sRNA contains a four base pair stem.

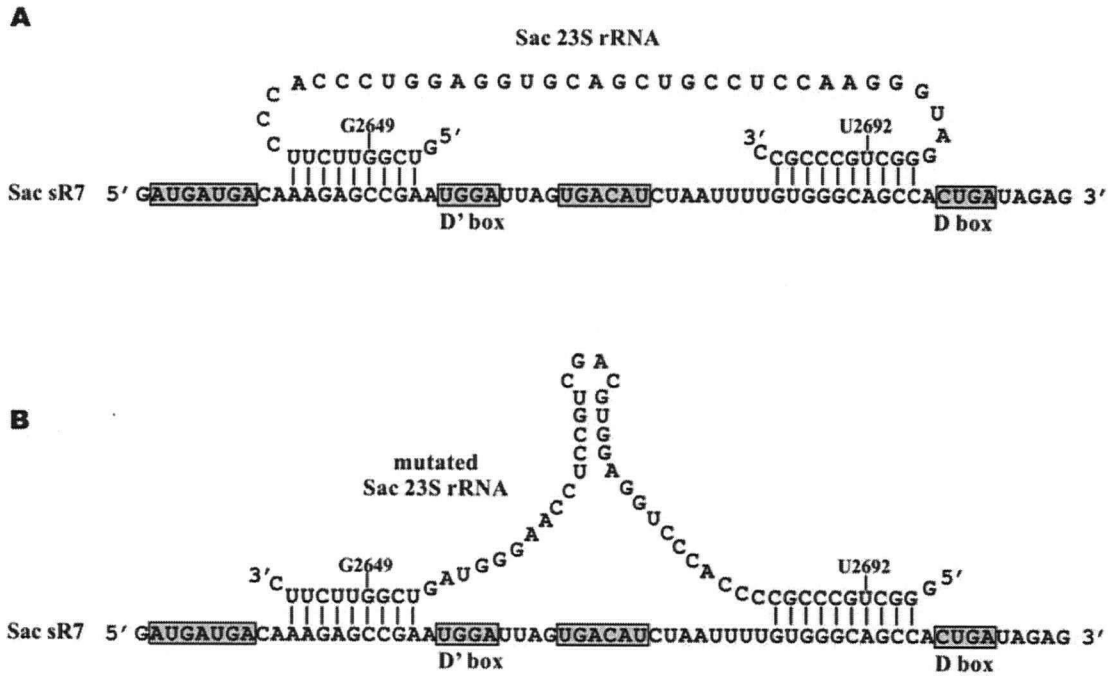


Figure 4-1 *S.acidocaldarius* sR7 sRNA with targets

(A) The sequence of *S.acidocaldarius* sR7 is illustrated, and the conserved motifs are boxed and shaded. sR7 sRNA is predicted to guide methylation to G2649 and U2692 in 23S rRNA in a reverse orientation.(B) The mutated target, resolving the pseudoknot and resulting in a direct guide is shown.

Another secondary structure concern might lie in the structure of the double target. The sequence between the two antisense elements is able to form an eleven base pair long hairpin with a high G and C content and a four nucleotide loop. This would lead to a highly stable secondary structure. A simultaneous binding of the antisense element of the substrate to the guide regions of *S. acidocaldarius* sR7 sRNA might prevent or facilitate the correct folding of the sRNA in this case.

Another remarkable feature of *S. acidocaldarius* sR7 is the unusual long 22 nucleotide spacing between the C' and D box of the sRNA. This unusual long sequence

element however does not extend the complementarity to the target. Usually the spacing between C' and D box in archael sRNAs is very compact (15-17 nucleotide long). Comparison to the newly identified *S. tokodaii* sR7 sRNA homologue reveals that *S. acidocaldarius* sR7 sRNA contains a five nucleotide insertion between the internal C' box and the terminal D box. It can only be speculated that, if at all, this insertion might influence stability of the sRNA or efficiency of the methylation reaction. Further experiments will have to be conducted to probe *S. acidocaldarius* sR7 stability for example by adding a terminal stem and/or deleting the insertion between the boxes. It would also be of interest to test if the secondary structure influences the efficiency of the double guide reaction, by simultaneous binding to the sRNA. This could be accomplished by eliminating the internal helical structure of the target and replacing it with a sequence that is less prone to form a stable secondary structure.

4.5.4 Homologous guide RNAs

In *S. acidocaldarius* sR7 sRNA the natural D box guide directs methylation to U2692 (U2552 in *E.coli* numbering) in 23S rRNA. This residue is highly conserved among Bacteria, Archaea and Eukaryotes and is located in the peptidyltransferase centre of the ribosome. Several homologous sRNAs have been identified in other archaeal species, as well as a potential snoRNA guiding this modification in yeast. Yeast snoRNA snR52 (alternative name Z13) has been predicted to direct methylation of U2918 (U2552 in *E.coli* numbering) using an 11 nt long D' box guide. This prediction could not be confirmed experimentally, although this site is known to be methylated in yeast (Brand *et*

al., 1977; Maden, 1990). The snoRNA snR52 is a predicted double guide. The D box guide has been shown to mediate methylation of residue A420 in 18S rRNA. However, a snR52 disruption mutant in yeast did not show the loss of modification at the second predicted position, U2918 (Lowe and Eddy, 1999) , suggesting that this modification might be guide RNA independent.

4.5.5 Is the methylation at U2692 guide RNA independent?

Prokaryotes contain only a fraction of the rRNA modifications found in Eukaryotes and Archaea, and do not appear to contain guide RNAs. Thus, a few site-specific enzymes may accomplish these modifications without guide RNAs. Conservation of such enzymes in Eukaryotes and Archaea would obviate the need for a number of specific guide RNAs.

In *E.coli* the heat shock protein FtsJ/RrmJ has been shown to be responsible for the methylation of essential residue U2552 in 23S rRNA. The rRNA target has a peculiar structure that is composed of a hairpin and a loop of five nucleotides, with the modified nucleotide being located at the 5' end of the loop. Interestingly, FtsJ/RrmJ has been shown to also be able to methylate tRNA *in vitro* (Bugl, 2000). FtsJ/RrmJ homologues are found in Bacteria, Archaea, and Eukaryotes. The smallest FtsJ/RrmJ homologues enzymes are found in Archaea where the proteins are 210 amino acids long. Among 145 completely sequenced microorganisms, only approximately half (77) of them encode putative FtsJ/RrmJ homologous, eight of which are in Archaea. However, no homologous gene could be detected in the Crenarchaota branch of Archaea. If the rRNA methylation

mediated by FtsJ/RrmJ is conserved, it could be performed by methylases lacking obvious sequence similarities with FtsJ/RrmJ. Surprisingly, the human JM23 protein encoded by the X chromosome has 35% identity over 192 amino acids with FtsJ/RrmJ, suggesting that, at least in certain higher organisms, rRNA methylation could be performed by an enzyme closely related to the *E. coli* FtsJ/RrmJ.

In yeast, a group of three proteins (Trmp7, Mrm2p and Spb1p) exhibit sequence similarity with FtsJ/RrmJ. Trmp7 has been shown to be responsible for the methylation of the wobble base in the anticodon of tRNAs (Pintard *et al.*, 2002b). Spb1p and Mrm2p have been recently shown to methylate the conserved residues U2918 of 28S rRNA and U2791 of 21S rRNA of mitochondria, respectively (Bonnerot *et al.*, 2003; Pintard *et al.*, 2002a). Methylated ribose U2791 is one of only three bases that are modified in the minimally modified 21S rRNA of yeast mitochondria (Sirum-Connolly *et al.*, 1995) and located in the peptidyl transferase centre of the 21S rRNA.

4.5.6 Importance of methylation at position U2692

The importance of modified nucleotides in rRNA has long been suggested, since their locations correlate well with universally conserved nucleotides and are known to cluster near the functional center of the ribosome (Brimacombe *et al.*, 1993). Residue U2692 (U2552 in *E.coli* numbering) is one of three universally occurring ribose methylations within the universally conserved peptidyltransferase center of domain V, at the site of interaction with the 3' terminal end of the tRNA (the CCA region). The fact that there are very few modifications located at the heart of the ribosome and that these

are conserved from Prokaryotes to Eukaryotes and cellular organelles, has led to the speculation that they must fulfill some essential functions.

While modifications in 16S rRNA are not essential for the proper assembly of the small ribosomal subunits in reconstitution experiments (Krzyzosiak *et al.*, 1987), only *in vivo* modified LSU rRNA can be used for the *in vitro* assembly of functional large subunits (Green and Noller, 1996). This evidently indicates a requirement of posttranscriptional modifications in the assembly of the large ribosomal subunits. So far, however, only a few modified nucleotides in rRNAs have been assigned a role in ribosome assembly or peptidyltransferse activity (Green and Noller, 1996).

One of these required modifications could be the ribose methylation of the conserved U2552. Interestingly, the position equivalent to *E.coli* U2552 on the cytoplasmic rRNA of yeast, U2918, is only methylated once the rRNA is assembled into pre-ribosomal particles (Brand *et al.*, 1977). Methylation of U2552 is not essential for the function of the ribosome, but mutations that alter U2552 cause severely reduced growth rates in *E.coli* (Bonnerot *et al.*, 2003; Porse and Garrett, 1995). Furthermore, significant reduction in peptidyl transferase activity has been associated with these mutations, indicating that the presence of methylated U2552 in the A site of the ribosome plays an important role in protein translation (Porse and Garrett, 1995). Moreover, lack of methylation of this residue in 21S mitochondrial rRNA leads to genomic instability, a phenotype that is typically associated with mutations that affect the mitochondrial translation apparatus (Pintard *et al.*, 2002a). It has been reported that the so-called 2550-loop, which contains Um2552 and G2553 plays an essential role in translation by base pairing with C75 of the tRNA 3'terminal region. Mutations of G2553, not U2552 abolish

the interaction with tRNA and thus the peptidyl transferase activity. Since mutation at position U2552 has little effect, it is conceivable that it is not the nature of the nucleotide that is important, but rather the presence of a methyl group at that position. Recently, the solution structure of the A-loop of 23S rRNA has demonstrated that methylation at U2552 modifies the configuration of C2556 and U2555, which in turn, affects the tertiary interaction of the A-loop within ribosome structure (Blanchard and Puglisi, 2001)

Recent studies in yeast have demonstrated that the methyltransferase Spb1 targets methylation to the highly conserved residue U2918 (U2552 in *E.coli* numbering) *in vivo* and *in vitro*, and that in the absence of the MTase function of Spb1, this position is methylated by a snR52-dependent mechanism (Bonnerot *et al.*, 2003). However, strains that depend on the snoRNA guided methylation show growth disadvantage and defects in ribosome synthesis, suggesting that in the sn52-dependent mechanism, methylation might be less efficient or delayed. The importance of modification at this residue for ribosome function and assembly might explain why a redundant mechanism is maintained in Eukaryotes. With the identification of yeast Spb1 and *E.coli* FtsJ/RrmJ homologues in Archaea, it seems plausible that a similar redundant mechanism might be present in some Archaea.

4.6 Engineered Guide sR1/sR2 hybrid

The engineered sR1/sR2 hybrid sRNA was assembled with its target into an RNP complex and showed the same methylation activity as the control *S. acidocaldarius* sR2 sRNA on the target. This confirms that ribose methylation can be directed to a novel nucleotide position in rRNA by using artificially constructed sRNAs containing the appropriate sequence complementary to the target site. In the guide RNA modification systems no constraints are placed on the substrate beyond complementarity to the guide RNA sequence, thus allowing the construction of artificial guides. The *in vitro* methylation assay provides a powerful tool for studying the function of the methylation complex.

However, it will eventually be necessary to confirm that the reconstituted system is an accurate representation of the *in vivo* system of the cell. Modification studies in yeast have shown that methylation can be shifted to new sites in rRNA *in vivo* by either moving the D or D' box relative to the antisense guide sequence or by equipping snoRNA with new sequences (Liu *et al.*, 2001; Ni *et al.*, 1997a). Methylation has been directed to rRNA nucleotides which are normally not ribose methylated, and to sites that usually are. In some cases methylation impairs the growth rate of the cell, indicating that a functionally important nucleotide has been altered.

Genetic systems have recently become available for *Sulfolobus solfataricus* (Sowers, 2003). Therefore, epigenetic expression to probe the *in vivo* function of the methylation guide complex in Archaea has become feasible. As a next step, the engineered sR1/sR2 hybrid will be transformed into *S. solfataricus* cells. Assuming that

the N plus five rule is correct, these mutant sRNAs should produce new sites of *in vivo* methylation in rRNA corresponding to position C1914 in 23S rRNA, a site that is normally not methylated in *S. solfataricus* but is methylated in *S. acidocaldarius*. To determine that the correct site in rRNA is methylated, the concentration-dependent primer extension pause assay can be used. Once the *in vivo* system is established, the method can be expanded to study the *in vivo* structure/function of aFib, aNop56, aL7a and other sRNAs.

4.7 *S. acidocaldarius* sRNA Homologues in other *Sulfolobus* species

The first demonstration that Archaea possess guide RNAs that direct methylation of rRNA came with biochemical studies in *Sulfolobus acidocaldarius* (Omer *et al.*, 2000). Nearly thirty different clones with features characteristic of eukaryotic C/D box snoRNAs were recovered.

To find homologues of the verified *S. acidocaldarius* sRNA genes, I ran BLASTN on each cDNA clone against the non redundant nucleotide database and recovered fourteen hits against sequences in other *Sulfolobus* species. These findings are in contrast to a previous report that BLASTN is not an effective tool for detecting sRNA gene homologues, because the sequences are short and were thought to diverge rapidly from each other (Dennis *et al.*, 2001). The earlier inability to detect sRNAs by BLASTN search was due to the fact that genomic data from *Sulfolobus* species genome projects were not publicly available at that time, or that the cut off limits in the previous BLASTN search were set too conservatively, e.g., when the default settings of the program were used. Because of the earlier failure to detect homologues by BLASTN search, a previously developed eukaryotic snoRNA search program was retrained with the verified *S. acidocaldarius* sRNA genes and used to screen the available archaeal genome sequences. With this search, thirteen sRNA were predicted in the partially sequenced genome of *S. solfataricus* (Omer et al., 2000). Except for *S. solfataricus* sR1 sRNA, none of these sRNAs were homologs to the original set of *S. acidocaldarius* sRNAs. With the new BLASTN search, I was able to identify five previously unknown sRNAs in the now

completed genome of *S. solfataricus*. In addition, nine sRNA candidates in the recently sequenced and published genome of *S. tokodaii* were identified.

Four sRNA candidates are present in all three *Sulfolobus* species (sR3, sR17, sR20 and sR26). sR3 sRNA also had been identified previously in *S. shibatae* by BLASTN search. sR17 sRNA is of particular interest. No potential methylation target has been identified for *S. acidocaldarius* sR17 and *S. solfataricus* sR17 sRNA. sR17 sRNA from *S. tokodaii* exhibits short complementarities to 23S rRNA and 16S rRNA. The general lack of conservation of predicted positions of methylation within rRNA within this archaeal generus implies that selection for methylation at most sites is relatively weak. Because selection is weak, nucleotide substitutions can accumulate within guides. Once a particular guide sequence has lost its target specificity through the accumulation of nucleotide substitutions, it would become free to explore available sequences in order to identify more favorable interactions with rRNA or other types of RNA. It is possible that sR17 sRNA is a "snap shot" of evolution and that sR17 sRNA in *S. acidocaldarius* and *S. solfataricus* lost its previous target or that *S. tokodaii* found a target site through accumulation of nucleotide substitution. Observations of this kind support the view that methylation target selection is an ongoing, dynamic process.

4.8 Future Perspectives

In this work, RNA guide methylation *in vitro* with a reconstituted archaeal sRNP was demonstrated. Accurate, site-specific methylation of rRNA fragments was achieved with a sRNP complex formed by incubating a cognate sRNA with the three core proteins aL7a, aNop56, and aFib. The use of the *in vitro* methylation assay was expanded to demonstrate methylation on more complex targets, including $tRNA^{Gln}$ and 5S rRNA. Furthermore, double guide function and specificity in direct or reverse order, both common in Archaea, were examined.

The mechanism of the *in vitro* methylation reaction

The *in vitro* reconstitution of the core complex used in this study sheds some light on structure and function of the methylation guide complex in Archaea. But the detailed mechanism of the catalytic cycle has not been elucidated. A more detailed description of the assembly process, the kinetic parameters of the reaction and whether the complex is capable of multiple turnovers will be necessary to understand the biochemistry of the complex methylation reaction. An important question concerns the assembly of the complex. Is the sRNP complex partially or completely disassembled after each methylation, and reassembled prior to the next round of methylation or are multiple target RNAs able to associate and dissociate from a single pre-assembled sRNP complex? The issue can, to some extent, be addressed by examining the methylation reaction in competition experiments where different guide and target RNAs are employed. In the first reaction, a primary set of guide and target RNAs are assembled under conditions

where the amount of core proteins is the limiting factor. A second set of competitor guide and target RNAs are added prior to the start of the reaction by the addition of SAM. In a second reaction, the primary and competitor guide and target RNAs are assembled together. Following the kinetics of total methylation as well as methylation of the primary and competitor RNAs in the two reactions should provide insight relating to the mechanistic pathway employed by the RNP complex during the catalytic cycle.

Crystal structure of the core methylation guide sRNP complex

The crystal structure of aNop56 and fibrillarin from the archaeon *A. fulgidus* has been solved (Aittaleb *et al.*, 2003). The structure gives insight into the assembly and function of box C/D sRNPs. However, to fully understand the methylation process, determination of the structure of the sRNP complex in the presence of the substrate RNA will most likely be required. Studies in this project demonstrated the stable formation and functional activity of sRNP from *S. solfataricus*. Because of the preliminary characterization of the sRNP complex from *S. solfataricus* this complex is a good candidate for further structural investigations of archaeal sRNPs by X-ray crystallography. Knowledge of the complete structure of the RNA-protein complexes should provide a better understanding of the interaction between the individual components of the complex.

Identification of additional components

In this study, simple substrates and small full length targets were successfully methylated *in vitro* by the core RNP complex. However, *in vivo* many RNA substrates

will typically have a more complex structure and be assembled in larger ribonucleoprotein complexes, suggesting that additional factors, most likely including RNA helicases, might be needed for substrate recognition and function. For example, many helicases are required for ribosome synthesis in eukaryotes; however, none has been assigned to a specific role in rRNA modification.

In addition, in eukaryotes the proteins p50 and p55 are known to transiently associate with C/D snoRNP complexes. Homologues of these proteins have been identified in Archaea. It will be of interest to investigate the role of these transiently associated proteins in sRNP. One hypothesis might be that these proteins are necessary for sRNP assembly and biogenesis. To get a full understanding of the methylation machinery and its function in ribosome biogenesis, it will be necessary to identify additional factors interacting with sRNP.

***In vivo* analysis of the methylation guide machinery.**

The *in vitro* methylation assay made it possible to study the structure and function of the archaeal methylation complex. However, it will be necessary to verify that the *in vitro* system is a correct reflection of the *in vivo* system. The construction and the verified methylation activity of the engineered sRNA hybrid is the first step towards the investigation of the *in vivo* structure by epigenetic gene expression in *S. solfataricus*. The engineered sR1/sR2 hybrid will be transformed into *Sulfolobus* cells, and specific methylation of the *in vivo* target site will have to be verified by primer extension assay.

Analysis of double guides

Double guide function has been investigated in this study and shown to direct methylation to both predicted sites of methylation; however, several question remain unanswered. Are both sites methylated in a single reaction cycle, or are the sites methylated consecutively? Are both sites of the same strand methylated, or does remain one site unmethylated after the other site has been methylated? Some of these questions could be answered by using a target RNA in the methylation assay that is pre-methylated at either the D or D' box target site and to monitor the reaction kinetics for these reactions. Another interesting question remains whether double guides are used to direct folding of nascent rRNAs.

Evolutionary origin of the guide snoRNA mediated modification

There is no evidence to support existence of modification guide RNAs in eubacteria. Instead, all known bacterial RNA methyltransferases (MTase) carry out site-specific modification of one or several strictly defined nucleosides and are believed to recognize their target directly. One possible explanation would be that Bacteria once possessed an RNA dependent modification machinery and that this mechanism was lost in favour of a site-specific protein methyltransferase. Another possibility is that the snoRNA targeted modification mechanism evolved from a protein-based system rather than a pre-biotic RNA world. The occurrence of guide RNAs in both Eukaryotes and Archaea indicates that the process of RNA-guided nucleotide modification is an ancient mechanism that predates the divergence of Eukaryotes and Archaea. In yeast, the Spb1 protein has been characterized as one of the first site-directed MTases of rRNA. It will be

interesting to investigate if Archaea have kept the same redundancy and utilize specific MTases in addition to the guide RNA, dependent system to methylate important specific residues in rRNA and tRNA. So far no MTase from Archaea has been characterized, but several potential homologues have been identified. It will be interesting to further study these homologues and characterize these potential MTases in terms of substrate specificity and target recognition.

In summary

This study has contributed to the understanding of the complex mechanisms behind RNA modifications. Many more doors for further investigation have been opened and the study of RNA modifications promises to remain an exciting field for many more years.

5 Bibliography

Agris, P.F. (1996) The importance of being modified: roles of modified nucleosides and Mg2+ in RNA structure and function. *Prog Nucleic Acid Res Mol Biol* **53**: 79-129.

Aittaleb, M., Rashid, R., Chen, Q., Palmer, J.R., Daniels, C.J., and Li, H. (2003) Structure and function of archaeal box C/D sRNP core proteins. *Nat Struct Biol* **10**: 256-263.

Allmang, C., Petfalski, E., Podtelejnikov, A., Mann, M., Tollervey, D., and Mitchell, P. (1999) The yeast exosome and human PM-Scl are related complexes of 3' --> 5' exonucleases. *Genes Dev* **13**: 2148-2158.

Altschul, S.F., Gish, W., Miller, W., Myers, E.W., and Lipman, D.J. (1990) Basic local alignment search tool. *J Mol Biol* **215**: 403-410.

Amiri, K.A. (1994) Fibrillarin-like proteins occur in the domain Archaea. *J Bacteriol* **176**: 2124-2127.

Armbruster, D.W., and Daniels, C.J. (1997) Splicing of intron-containing tRNATrp by the archaeon Haloferax volcanii occurs independent of mature tRNA structure. *J Biol Chem* **272**: 19758-19762.

Bachellerie, J.P., Nicoloso, M., Qu, L.H., Michot, B., Caizergues-Ferrer, M., Cavaille, J., and Renalier, M.H. (1995) Novel intron-encoded small nucleolar RNAs with long sequence complementarities to mature rRNAs involved in ribosome biogenesis. *Biochem Cell Biol* **73**: 835-843.

Bachellerie, J.P., Cavaille, J., and Huttenhofer, A. (2002) The expanding snoRNA world. *Biochimie* **84**: 775-790.

Bachellerie, J.P., and Cavaille J. (1998) Small nucleolar RNAs guide ribose methylations of eukaryotic rRNAs. In *Modification and Editing of RNA: The Alteration of RNA Structure and Function.* Grosjean, H., Benne, R (ed). Washington, DC: ASM Press, pp. 255-272.

Balakin, A.G., Smith, L., and Fournier, M.J. (1996) The RNA world of the nucleolus: two major families of small RNAs defined by different box elements with related functions. *Cell* **86**: 823-834.

Barneche, F., Gaspin, C., Guyot, R., and Echeverria, M. (2001) Identification of 66 box C/D snoRNAs in Arabidopsis thaliana: extensive gene duplications generated multiple isoforms predicting new ribosomal RNA 2'-O-methylation sites. *J Mol Biol* **311**: 57-73.

Beltrame, M., and Tollervey, D. (1992) Identification and functional analysis of two U3 binding sites on yeast pre-ribosomal RNA. *EMBO J* **11**: 1531-1542.

Beltrame, M., Henry, Y., and Tollervey, D. (1994) Mutational analysis of an essential binding site for the U3 snoRNA in the 5' external transcribed spacer of yeast pre-rRNA. *Nucleic Acids Res* **22**: 4057-4065.

Bertrand, E., Houser-Scott, F., Kendall, A., Singer, R.H., and Engelke, D.R. (1998) Nucleolar localization of early tRNA processing. *Genes Dev* **12**: 2463-2468.

Bjork, G., Ericson, JU, Gustafsson, CE, Hagervall, TG, Jonsson, YH, and Wikstrom, PM. (1987) Transfer RNA modification. *Annu Rev Biochem.*

Blanchard, S.C., and Puglisi, J.D. (2001) Solution structure of the A loop of 23S ribosomal RNA. *Proc Natl Acad Sci U S A* **98**: 3720-3725.

Bonnerot, C., Pintard, L., and Lutfalla, G. (2003) Functional redundancy of Spb1p and a snR52-dependent mechanism for the 2'-O-ribose methylation of a conserved rRNA position in yeast. *Mol Cell* **12**: 1309-1315.

Bousquet-Antonelli, C., Henry, Y., Gelugne, J.-P., Caizergues-Ferrer, M., and Kiss, T. (1997) A small nucleolar RNP protein is required for pseudouridylation of eukaryotic ribosomal RNAs. *EMBO J.* **16**: 4770-4776.

Brand, R.C., Klootwijk, J., Van Steenbergen, T.J., De Kok, A.J., and Planta, R.J. (1977) Secondary methylation of yeast ribosomal precursor RNA. *Eur J Biochem* **75**: 311-318.

Brimacombe, R., Mitchell, P., Osswald, M., Stade, K., and Bochkariov, D. (1993) Clustering of modified nucleotides at the functional center of bacterial ribosomal RNA. *Faseb J* **7**: 161-167.

Brock, T.D., Brock, K. M., Belly, R. T., Weiss, R. L. (1972) Sulfolobus: a new genus of sulfur-oxidizing bacteria living at low pH and high temperature. *Arch. Mikrobiol.* **84**: 54-68.

Bruenger, E., Kowalak, J.A., Kuchino, Y., McCloskey, J.A., Mizushima, H., Stetter, K.O., and Crain, P.F. (1993) 5S rRNA modification in the hyperthermophilic archaea Sulfolobus solfataricus and Pyrodictium occultum. *Faseb J* **7**: 196-200.

Bugl, H., Fauman, E.B., Staker, B.L., Zheng, F., Kushner, S.R., Saper, M.A., Bardwell, J.C., and Jakob,U. (2000)

RNA methylation under heat shock control. *Mol Cell* **6**: 349-360.

Bult, C.J., White, O., Olsen, G.J., Zhou, L., Fleischmann, R.D., Sutton, G.G., Blake, J.A., FitzGerald, L.M., Clayton, R.A., Gocayne, J.D., Kerlavage, A.R., Dougherty, B.A., Tomb, J.F., Adams, M.D., Reich, C.I., Overbeek, R., Kirkness, E.F., Weinstock, K.G., Merrick, J.M., Glodek, A., Scott, J.L., Geoghagen, N.S., and Venter, J.C. (1996) Complete genome sequence of the methanogenic archaeon, Methanococcus jannaschii. *Science* **273**: 1058-1073.

Burns, C.M., Chu, H., Rueter, S.M., Hutchinson, L.K., Canton, H., Sanders-Bush, E., and Emeson, R.B. (1997) Regulation of serotonin-2C receptor G-protein coupling by RNA editing. *Nature* **387**: 303-308.

Cahill, N.M., Friend, K., Speckmann, W., Li, Z.H., Terns, R.M., Terns, M.P., and Steitz, J.A. (2002) Site-specific cross-linking analyses reveal an asymmetric protein distribution for a box C/D snoRNP. *EMBO J* **21**: 3816-3828.

Caldas, T., Binet, E., Bouloc, P., and Richarme, G. (2000) Translational defects of Escherichia coli mutants deficient in the Um(2552) 23S ribosomal RNA methyltransferase RrmJ/FTSJ. *Biochem Biophys Res Commun* **271**: 714-718.

Canton, H., Emeson, R.B., Barker, E.L., Backstrom, J.R., Lu, J.T., Chang, M.S., and Sanders-Bush, E. (1996) Identification, molecular cloning, and distribution of a short variant of the 5-hydroxytryptamine2C receptor produced by alternative splicing. *Mol Pharmacol* **50**: 799-807.

Cavaille, J., Nicoloso, M., and Bachellerie, J.P. (1996) Targeted ribose methylation of RNA in vivo directed by tailored antisense RNA guides. *Nature* **383**: 732-735.

Cavaille, J., and Bachellerie, J.P. (1998) SnoRNA-guided ribose methylation of rRNA: structural features of the guide RNA duplex influencing the extent of the reaction. *Nucleic Acids Res* **26**: 1576-1587.

Cavaille, J., Chetouani, F., and Bachellerie, J.P. (1999) The yeast Saccharomyces cerevisiae YDL112w ORF encodes the putative 2'-O-ribose methyltransferase catalyzing the formation of Gm18 in tRNAs. *RNA* **5**: 66-81.

Cavaille, J., Buiting, K., Kiefmann, M., Lalande, M., Brannan, C.I., Horsthemke, B., Bachellerie, J.P., Brosius, J., and Huttenhofer, A. (2000) Identification of brain-specific and imprinted small nucleolar RNA genes exhibiting an unusual genomic organization. *Proc Natl Acad Sci U S A* **97**: 14311-14316.

Cavaille, J., Vitali, P., Basyuk, E., Huttenhofer, A., and Bachellerie, J.P. (2001) A novel brain-specific box C/D small nucleolar RNA processed from tandemly repeated introns of a noncoding RNA gene in rats. *J Biol Chem* **276**: 26374-26383.

Chanfreau, G., Legrain, P., and Jacquier, A. (1998a) Yeast RNase III as a key processing enzyme in small nucleolar RNAs metabolism. *J Mol Biol* **284**: 975-988.

Chanfreau, G., Rotondo, G., Legrain, P., and Jacquier, A. (1998b) Processing of a dicistronic small nucleolar RNA precursor by the RNA endonuclease Rnt1. *EMBO J* **17**: 3726-3737.

Clouet d'Orval, B., Bortolin, M.L., Gaspin, C., and Bachellerie, J.P. (2001) Box C/D RNA guides for the ribose methylation of archaeal tRNAs. The tRNATrp intron guides the formation of two ribose-methylated nucleosides in the mature tRNATrp. *Nucleic Acids Res* **29**: 4518-4529.

Curran, J.F. (1998) Modified nucleosides in translation. In *Modification and Editing of RNA*. Grosjean, H., Benne, R (ed). Washington D.C.: American Society for Microbiology, pp. 493-516.

Dams, E., Londei, P., Cammarano, P., Vandenberghe, A., and De Wachter, R. (1983) Sequences of the 5S rRNAs of the thermo-acidophilic archaebacterium Sulfolobus solfataricus (Caldariella acidophila) and the thermophilic eubacteria Bacillus acidocaldarius and Thermus aquaticus. *Nucleic Acids Res* **11**: 4667-4676.

Dennis, P.P. (1997) Ancient ciphers: translation in Archaea. *Cell* **89**: 1007-1010.

Dennis, P.P., Omer, A., and Lowe, T. (2001) A guided tour: small RNA function in Archaea. *Mol Microbiol* **40**: 509-519.

Dez, C., Henras, A., Faucon, B., Lafontaine, D., Caizergues-Ferrer, M., and Henry, Y. (2001) Stable expression in yeast of the mature form of human telomerase RNA depends on its association with the box H/ACA small nucleolar RNP proteins Cbf5p, Nhp2p and Nop10p. *Nucleic Acids Res* **29**: 598-603.

Dez, C., Noaillac-Depeyre, J., Caizergues-Ferrer, M., and Henry, Y. (2002) Naf1p, an essential nucleoplasmic factor specifically required for accumulation of box H/ACA small nucleolar RNPs. *Mol Cell Biol* **22**: 7053-7065.

Dirheimer, G., Keith, G, Dumas, P and Westhof, E (1995) Primary, Secondary, and Tertiary Structures of tRNAs. In *tRNA: Structure, Biosynthesis and Function.* Soell, D., RajBhandary, U.L. (ed). Washington: American Society of Microbiology, pp. 93-126.

Durovic, P., Kutay, U., Schleper, C., and Dennis, P.P. (1994) Strain identification and 5S rRNA gene characterization of the hyperthermophilic archaebacterium Sulfolobus acidocaldarius. *J Bacteriol* **176**: 514-517.

Enright, C.A., Maxwell, E.S., Eliceiri, G.L., and Sollner-Webb, B. (1996) 5'ETS rRNA processing facilitated by four small RNAs: U14, E3, U17, and U3. *RNA* **2**: 1094-1099.

Fatica, A., Dlakic, M., and Tollervey, D. (2002) Naf1 p is a box H/ACA snoRNP assembly factor. *RNA* **8**: 1502-1514.

Filipowicz, W., and Pogacic, V. (2002) Biogenesis of small nucleolar ribonucleoproteins. *Curr Opin Cell Biol* **14**: 319-327.

Filippini, D., Bozzoni, I., and Caffarelli, E. (2000) p62, a novel Xenopus laevis component of box C/D snoRNPs. *Rna* **6**: 391-401.

Filippini, D., Renzi, F., Bozzoni, I., and Caffarelli, E. (2001) U86, a novel snoRNA with an unprecedented gene organization in yeast. *Biochem Biophys Res Commun* **288**: 16-21.

Fourmy, D., Recht, M.I., Blanchard, S.C., and Puglisi, J.D. (1996) Structure of the A site of Escherichia coli 16S ribosomal RNA complexed with an aminoglycoside antibiotic. *Science* **274**: 1367-1371.

Fox, G.E., Stackebrandt, E., Hespell, R.B., Gibson, J., Maniloff, J., Dyer, T.A., Wolfe, R.S., Balch, W.E., Tanner, R.S., Magrum, L.J., Zablen, L.B., Blakemore, R., Gupta, R., Bonen, L., Lewis, B.J., Stahl, D.A., Luehrsen, K.R., Chen, K.N., and Woese, C.R. (1980) The phylogeny of prokaryotes. *Science* **209**: 457-463.

Galardi, S., Fatica, A., Bachi, A., Scaloni, A., Presutti, C., and Bozzoni, I. (2002) Purified box C/D snoRNPs are able to reproduce site-specific 2'-O-methylation of target RNA in vitro. *Mol Cell Biol* **22**: 6663-6668.

Galtier, N., and Lobry, J.R. (1997) Relationships between genomic G+C content, RNA secondary structures, and optimal growth temperature in prokaryotes. *J Mol Evol* **44**: 632-636.

Ganot, P., Bortolin, M.L., and Kiss, T. (1997a) Site-specific pseudouridine formation in preribosomal RNA is guided by small nucleolar RNAs. *Cell* **89**: 799-809.

Ganot, P., Caizergues-Ferrer, M., and Kiss, T. (1997b) The family of box ACA small nucleolar RNAs is defined by an evolutionarily conserved secondary structure and ubiquitous sequence elements essential for RNA accumulation. *Genes Dev* **11**: 941-956.

Garcia, G., and Goodenough-Lashua, DM (1998) Mechanisms of RNA-Modifying and Editing Enzymes. In *Modification and Editing of RNA*. Grosjean, H., Benne, R (ed). Washington, DC: American Society of Microbiology, pp. 135-168.

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

Gautier, T., Berges, T., Tollervey, D., and Hurt, E. (1997) Nucleolar KKE/D repeat proteins Nop56p and Nop58p interact with Nop1p and are required for ribosome biogenesis. *Mol Cell Biol* **17**: 7088-7098.

Giege, R., Puglisi, J.D., and Florentz, C. (1993) tRNA structure and aminoacylation efficiency. *Prog Nucleic Acid Res Mol Biol* **45**: 129-206.

Giege, R., Sissler, M., and Florentz, C. (1998) Universal rules and idiosyncratic features in tRNA identity. *Nucleic Acids Res* **26**: 5017-5035.

Girard, J., Lehtonen, H., Caizergues-Ferrer, M., Amalric, F., Tollervey, D., and Lapeyre, B. (1992) GAR1 is an essential small nucleolar RNP protein required for pre-rRNA processing in yeast. *EMBO J.* **11**: 673-682.

Girard, J., Bagni, C., Caizergues-Ferrer, M., Amalric, F., and Lapeyre, B. (1994) Identification of a segment of the small nucleolar ribonucleoprotein- associated protein GAR1 that is sufficient for nucleolar accumulation. *J. Biol. Chem.* **269**: 18499-18506.

Green, R., and Noller, H.F. (1996) In vitro complementation analysis localizes 23S rRNA posttranscriptional modifications that are required for Escherichia coli 50S ribosomal subunit assembly and function. *RNA* **2**: 1011-1021.

Grosjean, H., Edqvist, J., Straby, K.B., and Giege, R. (1996) Enzymatic formation of modified nucleosides in tRNA: dependence on tRNA architecture. *J Mol Biol* **255**: 67-85.

Gustafsson, C., Reid, R., Greene, P.J., and Santi, D.V. (1996) Identification of new RNA modifying enzymes by iterative genome search using known modifying enzymes as probes. *Nucleic Acids Res* **24**: 3756-3762.

Henras, A., Henry, Y., Bousquet-Antonelli, C., Noaillac-Depeyre, J., Gelugne, J.P., and Caizergues-Ferrer, M. (1998) Nhp2p and Nop10p are essential for the function of H/ACA snoRNPs. *EMBO J* **17**: 7078-7090.

Hoang, C., and Ferre-D'Amare, A.R. (2001) Cocrystal structure of a tRNA Psi55 pseudouridine synthase: nucleotide flipping by an RNA-modifying enzyme. *Cell* **107**: 929-939.

Hughes, J.M. (1996) Functional base-pairing interaction between highly conserved elements of U3 small nucleolar RNA and the small ribosomal subunit RNA. *J Mol Biol* **259**: 645-654.

Huttenhofer, A., Kiefmann, M., Meier-Ewert, S., O'Brien, J., Lehrach, H., Bachellerie, J.P., and Brosius, J. (2001) RNomics: an experimental approach that identifies 201 candidates for novel, small, non-messenger RNAs in mouse. *EMBO J* **20**: 2943-2953.

Ito, T., Chiba, T., Ozawa, R., Yoshida, M., Hattori, M., and Sakaki, Y. (2001) A comprehensive two-hybrid analysis to explore the yeast protein interactome. *PNAS* **98**: 4569-4574.

Jady, B.E., and Kiss, T. (2000) Characterisation of the U83 and U84 small nucleolar RNAs: two novel 2'-O-ribose methylation guide RNAs that lack complementarities to ribosomal RNAs. *Nucleic Acids Res* **28**: 1348-1354.

Jady, B.E., and Kiss, T. (2001) A small nucleolar guide RNA functions both in 2'-O-ribose methylation and pseudouridylation of the U5 spliceosomal RNA. *EMBO J* **20**: 541-551.

Kawai, G., Yamamoto, Y., Kamimura, T., Masegi, T., Sekine, M., Hata, T., Iimori, T., Watanabe, T., Miyazawa, T., and Yokoyama, S. (1992) Conformational rigidity of specific pyrimidine residues in tRNA arises from posttranscriptional modifications that enhance steric interaction between the base and the 2'-hydroxyl group. *Biochemistry* **31**: 1040-1046.

Keith, G. (1995) Mobilities of modified ribonucleotides on two-dimensional cellulose thin-layer chromatography. *Biochimie* **77**: 142-144.

Kim, D.F., and Green, R. (1999) Base-pairing between 23S rRNA and tRNA in the ribosomal A site. *Mol Cell* **4**: 859-864.

Kim, S., Sussman, JL, Suddath, FL, Quigley, GJ, McPherson, A, Wang, AH, Seeman, NC, and RICH, A (1974)
The general structure of transfer RNA molecules. *Proc Natl Acad Sci U S A* **71**: 4970-4974.

Kirpekar, F., Douthwaite, S., and Roepstorff, P. (2000) Mapping posttranscriptional modifications in 5S ribosomal RNA by MALDI mass spectrometry. *RNA* **6**: 296-306.

Kiss, T. (2001) Small nucleolar RNA-guided post-transcriptional modification of cellular RNAs. *EMBO J* **20**: 3617-3622.

Kiss, T. (2002) Small nucleolar RNAs: an abundant group of noncoding RNAs with diverse cellular functions. *Cell* **109**: 145-148.

Kiss-Laszlo, Z., Henry, Y., Bachellerie, J.P., Caizergues-Ferrer, M., and Kiss, T. (1996) Site-specific ribose methylation of preribosomal RNA: a novel function for small nucleolar RNAs. *Cell* **85**: 1077-1088.

Klein, D.J., Schmeing, T.M., Moore, P.B., and Steitz, T.A. (2001) The kink-turn: a new RNA secondary structure motif. *EMBO J* **20**: 4214-4221.

Klein, R.J., Misulovin, Z., and Eddy, S.R. (2002) Noncoding RNA genes identified in AT-rich hyperthermophiles. *Proc Natl Acad Sci U S A* **99**: 7542-7547.

Klenk, H.P., Clayton, R.A., Tomb, J.F., White, O., Nelson, K.E., Ketchum, K.A., Dodson, R.J., Gwinn, M., Hickey, E.K., Peterson, J.D., Richardson, D.L., Kerlavage, A.R., Graham, D.E., Kyrpides, N.C., Fleischmann, R.D., Quackenbush, J., Lee, N.H., Sutton, G.G., Gill, S., Kirkness, E.F., Dougherty, B.A., McKenney, K., Adams, M.D., Loftus, B., Venter, J.C., and et al. (1997) The complete genome sequence of the hyperthermophilic, sulphate-reducing archaeon Archaeoglobus fulgidus. *Nature* **390**: 364-370.

Krzyzosiak, W., Denman, R., Nurse, K., Hellmann, W., Boublik, M., Gehrke, C.W., Agris, P.F., and Ofengand, J. (1987) In vitro synthesis of 16S ribosomal RNA containing single base changes and assembly into a functional 30S ribosome. *Biochemistry* **26**: 2353-2364.

Kuchino, Y., Borek, E., Grunberger, D., Mushinski, J.F., and Nishimura, S. (1982) Changes of post-transcriptional modification of wye base in tumor-specific tRNAPhe. *Nucleic Acids Res* **10**: 6421-6432.

Kumagai, I., Watanabe, K., and Oshima, T. (1982) A thermostable tRNA (guanosine-2')-methyltransferase from Thermus thermophilus HB27 and the effect of ribose methylation on the conformational stability of tRNA. *J Biol Chem* **257**: 7388-7395.

Laemmli, U.K. (1970) Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature* **227**: 680-685.

Lafontaine, D.L., and Tollervey, D. (1998) Birth of the snoRNPs: the evolution of the modification-guide snoRNAs. *Trends Biochem Sci* **23**: 383-388.

Lane, B.G., Ofengand, J., and Gray, M.W. (1995) Pseudouridine and O2'-methylated nucleosides. Significance of their selective occurrence in rRNA domains that function in ribosome-catalyzed synthesis of the peptide bonds in proteins. *Biochimie* **77**: 7-15.

Leader, D.J., Clark, G.P., Watters, J., Beven, A.F., Shaw, P.J., and Brown, J.W. (1997) Clusters of multiple different small nucleolar RNA genes in plants are expressed as and processed from polycistronic pre-snoRNAs. *EMBO J* **16**: 5742-5751.

Leader, D.J., Clark, G.P., Boag, J., Watters, J.A., Simpson, C.G., Watkins, N.J., Maxwell, E.S., and Brown, J.W. (1998) Processing of vertebrate box C/D small nucleolar RNAs in plant cells. *Eur J Biochem* **253**: 154-160.

Lecointe, F., Simos, G., Sauer, A., Hurt, E.C., Motorin, Y., and Grosjean, H. (1998) Characterization of Yeast Protein Deg1 as Pseudouridine Synthase (Pus3) Catalyzing the Formation of Psi 38 and Psi 39 in tRNA Anticodon Loop. *J. Biol. Chem.* **273**: 1316-1323.

Liang, W.Q., and Fournier, M.J. (1997) Synthesis of functional eukaryotic ribosomal RNAs in trans: development of a novel in vivo rDNA system for dissecting ribosome biogenesis. *Proc Natl Acad Sci U S A* **94**: 2864-2868.

Liang, X.H., Liu, L., and Michaeli, S. (2001) Identification of the first trypanosome H/ACA RNA that guides pseudouridine formation on rRNA. *J Biol Chem* **276**: 40313-40318.

Liu, B., Ni, J., and Fournier, M.J. (2001) Probing RNA in vivo with methylation guide small nucleolar RNAs. *Methods* **23**: 276-286.

Lowe, T.M., and Eddy, S.R. (1999) A computational screen for methylation guide snoRNAs in yeast. *Science* **283**: 1168-1171.

Lygerou, Z., Mitchell, P., Petfalski, E., Seraphin, B., and Tollervey, D. (1994) The POP1 gene encodes a protein component common to the RNase MRP and RNase P ribonucleoproteins. *Genes Dev* **8**: 1423-1433.

Maden, B.E. (1990) The numerous modified nucleotides in eukaryotic ribosomal RNA. *Prog Nucleic Acid Res Mol Biol* **39**: 241-303.

Maden, B.E., Corbett, M.E., Heeney, P.A., Pugh, K., and Ajuh, P.M. (1995) Classical and novel approaches to the detection and localization of the numerous modified nucleotides in eukaryotic ribosomal RNA. *Biochimie* **77**: 22-29.

Mao, H., White, S.A., and Williamson, J.R. (1999) A novel loop-loop recognition motif in the yeast ribosomal protein L30 autoregulatory RNA complex. *Nat Struct Biol* **6**: 1139-1147.

Massenet, S., Mougin, A, Branlant C (1998) Posttranscriptional modifications in the U snRNAs. In *Modification and Editing of RNA: The Alteration of RNA structure and Function.* Grosjean, H., Benne, R (ed). Washington, DC: ASM Press, pp. 201-228.

Maxwell, E.S., and Fournier, M.J. (1995) The small nucleolar RNAs. *Annu Rev Biochem* **64**: 897-934.

Mitchell, J.R., Cheng, J., and Collins, K. (1999) A box H/ACA small nucleolar RNA-like domain at the human telomerase RNA 3' end. *Mol Cell Biol* **19**: 567-576.

Moazed, D., and Noller, H.F. (1989) Interaction of tRNA with 23S rRNA in the ribosomal A, P, and E sites. *Cell* **57**: 585-597.

Morrissey, J.P., and Tollervey, D. (1993) Yeast snR30 is a small nucleolar RNA required for 18S rRNA synthesis. *Mol Cell Biol* **13**: 2469-2477.

Morrissey, J.P., and Tollervey, D. (1995) Birth of the snoRNPs: the evolution of RNase MRP and the eukaryotic pre-rRNA-processing system. *Trends Biochem Sci* **20**: 78-82.

Motorin, Y., Grosjean, H (1998) Appendix 1: Chemical structures and classification of posttranslationally modified nucleosides in RNA. In *Modification and Editing of RNA*. Grosjean, H., Benne, R (ed). Washington DC: American Society of Microbiology, pp. 543-549.

Mougey, E.B., O'Reilly, M., Osheim, Y., Miller, O.L., Jr., Beyer, A., and Sollner-Webb, B. (1993) The terminal balls characteristic of eukaryotic rRNA transcription units in chromatin spreads are rRNA processing complexes. *Genes Dev* **7**: 1609-1619.

Newman, D.R., Kuhn, J.F., Shanab, G.M., and Maxwell, E.S. (2000) Box C/D snoRNA-associated proteins: two pairs of evolutionarily ancient proteins and possible links to replication and transcription. *Rna* **6**: 861-879.

Ni, J., Samarsky, D.A., Liu, B., Ferbeyre, G., Cedergren, R., and Fournier, M.J. (1997a) SnoRNAs as tools for RNA cleavage and modification. *Nucleic Acids Symp Ser*: 61-63.

Ni, J., Tien, A.L., and Fournier, M.J. (1997b) Small nucleolar RNAs direct site-specific synthesis of pseudouridine in ribosomal RNA. *Cell* **89**: 565-573.

Nicoloso, M., Qu, L.H., Michot, B., and Bachellerie, J.P. (1996) Intron-encoded, antisense small nucleolar RNAs: the characterization of nine novel species points to their direct role as guides for the 2'-O-ribose methylation of rRNAs. *J Mol Biol* **260**: 178-195.

Niewmierzycka, A., and Clarke, S. (1999) S-Adenosylmethionine-dependent methylation in Saccharomyces cerevisiae. Identification of a novel protein arginine methyltransferase. *J Biol Chem* **274**: 814-824.

Noon, K.R., Bruenger, E., and McCloskey, J.A. (1998) Posttranscriptional modifications in 16S and 23S rRNAs of the archaeal hyperthermophile Sulfolobus solfataricus. *J Bacteriol* **180**: 2883-2888.

Nottrott, S., Hartmuth, K., Fabrizio, P., Urlaub, H., Vidovic, I., Ficner, R., and Luhrmann, R. (1999) Functional interaction of a novel 15.5kD [U4/U6.U5] tri-snRNP protein with the 5' stem-loop of U4 snRNA. *EMBO J* **18**: 6119-6133.

Nottrott, S., Urlaub, H., and Luhrmann, R. (2002) Hierarchical, clustered protein interactions with U4/U6 snRNA: a biochemical role for U4/U6 proteins. *EMBO J* **21**: 5527-5538.

Ofengand, J., Rudd, K, (2000) Bacterial, archaea, and organellar RNA pseudouridnes and methylated nucleotides and their enzymes. In *Ribosome: Structure, Function, Antibiotics, and Cellular Interactions*. Garrett, R., Douthwaite S, Liljas, A, Matheson, A, Moore, PB, Noller HF (ed). Washington, DC: ASM Press, pp. 175-190.

Ofengand, J.a.F.M.J. (1998) The pseudouridine residues of rRNA:number,location, biosynthesis and function. In *Modification and Editing of RNA: The Alteration of RNA Structure and Function*. Grosjean, H., Benne, R (ed). Washington, DC: ASM Press, pp. 229-253.

Omer, A.D., Lowe, T.M., Russell, A.G., Ebhardt, H., Eddy, S.R., and Dennis, P.P. (2000) Homologs of small nucleolar RNAs in Archaea. *Science* **288**: 517-522.

Omer, A.D., Ziesche, S., Ebhardt, H., and Dennis, P.P. (2002) In vitro reconstitution and activity of a C/D box methylation guide ribonucleoprotein complex. *Proc Natl Acad Sci U S A* **99**: 5289-5294.

Omer, A.D., Ziesche, S., Decatur, W.A., Fournier, M.J., and Dennis, P.P. (2003) RNA-modifying machines in archaea. *Mol Microbiol* **48**: 617-629.

Paul, M.S., and Bass, B.L. (1998) Inosine exists in mRNA at tissue-specific levels and is most abundant in brain mRNA. *EMBO J* **17**: 1120-1127.

Pederson, T. (1998) The plurifunctional nucleolus. *Nucleic Acids Res* **26**: 3871-3876.

Persson, B.C., Jager, G., and Gustafsson, C. (1997) The spoU gene of Escherichia coli, the fourth gene of the spoT operon, is essential for tRNA (Gm18) 2'-O-methyltransferase activity. *Nucleic Acids Res* **25**: 4093-4097.

Pintard, L., Bujnicki, J.M., Lapeyre, B., and Bonnerot, C. (2002a) MRM2 encodes a novel yeast mitochondrial 21S rRNA methyltransferase. *EMBO J.* **21**: 1139-1147.

Pintard, L., Lecointe, F., Bujnicki, J.M., Bonnerot, C., Grosjean, H., and Lapeyre, B. (2002b) Trm7p catalyses the formation of two 2'-O-methylriboses in yeast tRNA anticodon loop. *EMBO J* **21**: 1811-1820.

Pogacic, V., Dragon, F., and Filipowicz, W. (2000) Human H/ACA small nucleolar RNPs and telomerase share evolutionarily conserved proteins NHP2 and NOP10. *Mol Cell Biol* **20**: 9028-9040.

Pombo, A., Jackson, D.A., Hollinshead, M., Wang, Z., Roeder, R.G., and Cook, P.R. (1999) Regional specialization in human nuclei: visualization of discrete sites of transcription by RNA polymerase III. *EMBO J* **18**: 2241-2253.

Porse, B.T., and Garrett, R.A. (1995) Mapping important nucleotides in the peptidyl transferase centre of 23 S rRNA using a random mutagenesis approach. *J Mol Biol* **249**: 1-10.

Qu, L.H., Meng, Q., Zhou, H., Chen, Y.Q., Liang-Hu, Q., Qing, M., Hui, Z., and Yue-Qin, C. (2001) Identification of 10 novel snoRNA gene clusters from Arabidopsis thaliana. *Nucleic Acids Res* **29**: 1623-1630.

Randerath, E., Gopalakrishnan, A.S., Gupta, R.C., Agrawal, H.P., and Randerath, K. (1981) Lack of a specific ribose methylation at guanosine 17 in Morris hepatoma 5123D tRNASer1IGA. *Cancer Res* **41**: 2863-2867.

Rashid, R., Aittaleb, M., Chen, Q., Spiegel, K., Demeler, B., and Li, H. (2003) Functional requirement for symmetric assembly of archaeal box C/D small ribonucleoprotein particles. *J Mol Biol* **333**: 295-306.

Raue, H.A., Klootwijk, J., and Musters, W. (1988) Evolutionary conservation of structure and function of high molecular weight ribosomal RNA. *Prog Biophys Mol Biol* **51**: 77-129.

Raychaudhuri, S., Conrad, J., Hall, B.G., and Ofengand, J. (1998) A pseudouridine synthase required for the formation of two universally conserved pseudouridines in ribosomal RNA is essential for normal growth of Escherichia coli. *RNA* **4**: 1407-1417.

Robertus, J.D., Ladner, J.E., Finch, J.T., Rhodes, D., Brown, R.S., Clark, B.F.C., and Klug, A. (1974) Structure of yeast phenylalanine tRNA at 3 A resolution. *Nature* **250**: 546-551.

Rozenski, J., Crain, P.F., and McCloskey, J.A. (1999) The RNA Modification Database: 1999 update. *Nucleic Acids Res* **27**: 196-197.

Rozhdestvensky, T.S., Tang, T.H., Tchirkova, I.V., Brosius, J., Bachellerie, J.P., and Huttenhofer, A. (2003) Binding of L7Ae protein to the K-turn of archaeal

snoRNAs: a shared RNA binding motif for C/D and H/ACA box snoRNAs in Archaea. *Nucleic Acids Res* **31**: 869-877.

Sambrook, J.F., EF. and Maniatis, T. (1989) *Molecular cloning - a laboratory manual*: Cold Spring Harbor Press.

Samuelsson, T., and Olsson, M. (1990) Transfer RNA pseudouridine synthases in Saccharomyces cerevisiae. *J Biol Chem* **265**: 8782-8787.

Satoh, A., Takai, K., Ouchi, R., Yokoyama, S., and Takaku, H. (2000) Effects of anticodon 2'-O-methylations on tRNA codon recognition in an Escherichia coli cell-free translation. *RNA* **6**: 680-686.

Schimmang, T., Tollervey, D., Kern, H., Frank, R., and Hurt, E.C. (1989) A yeast nucleolar protein related to mammalian fibrillarin is associated with small nucleolar RNA and is essential for viability. *EMBO J* **8**: 4015-4024.

Sharma, K., and Tollervey, D. (1999) Base pairing between U3 small nucleolar RNA and the 5' end of 18S rRNA is required for pre-rRNA processing. *Mol Cell Biol* **19**: 6012-6019.

Shen, Z., and Hagerman, P.J. (1994) Conformation of the central, three-helix junction of the 5 S ribosomal RNA of Sulfolobus acidocaldarius. *J Mol Biol* **241**: 415-430.

Sirum-Connolly, K., Peltier, J.M., Crain, P.F., McCloskey, J.A., and Mason, T.L. (1995) Implications of a functional large ribosomal RNA with only three modified nucleotides. *Biochimie* **77**: 30-39.

Smith, D., Doucette-Stamm, L., Deloughery, C., Lee, H., Dubois, J., Aldredge, T., Bashirzadeh, R., Blakely, D., Cook, R., Gilbert, K., Harrison, D., Hoang, L., Keagle, P., Lumm, W., Pothier, B., Qiu, D., Spadafora, R., Vicaire, R., Wang, Y., Wierzbowski, J., Gibson, R., Jiwani, N., Caruso, A., Bush, D., and Reeve, J. (1997) Complete genome sequence of Methanobacterium thermoautotrophicum deltaH: functional analysis and comparative genomics. *J. Bacteriol.* **179**: 7135-7155.

Sowers, K.R., P. H. Blum, and S. DasSarma (2003) Genetics of Archaea. In *Methods for General and Molecular Microbiology*. Washington, D.C.: ASM Press.

Speckmann, W.A., Li, Z.H., Lowe, T.M., Eddy, S.R., Terns, R.M., and Terns, M.P. (2002) Archaeal guide RNAs function in rRNA modification in the eukaryotic nucleus. *Curr Biol* **12**: 199-203.

Sprinzl, M., Horn, C., Brown, M., Ioudovitch, A., and Steinberg, S. (1998) Compilation of tRNA sequences and sequences of tRNA genes. *Nucleic Acids Res* **26**: 148-153.

Stahl, D.A., Luehrsen, K.R., Woese, C.R., and Pace, N.R. (1981) An unusual 5S rRNA, from Sulfolobus acidocaldarius, and its implications for a general 5S rRNA structure. *Nucleic Acids Res* **9**: 6129-6137.

Steitz, J.A., and Tycowski, K.T. (1995) Small RNA chaperones for ribosome biogenesis. *Science* **270**: 1626-1627.

Studier, F.W., and Moffatt, B.A. (1986) Use of bacteriophage T7 RNA polymerase to direct selective high-level expression of cloned genes. *J Mol Biol* **189**: 113-130.

Studier, F.W. (1991) Use of bacteriophage T7 lysozyme to improve an inducible T7 expression system. *J Mol Biol* **219**: 37-44.

Szewczak, L.B., DeGregorio, S.J., Strobel, S.A., and Steitz, J.A. (2002) Exclusive interaction of the 15.5 kD protein with the terminal box C/D motif of a methylation guide snoRNP. *Chem Biol* **9**: 1095-1107.

Szymanski, M., Barciszewska, M.Z., Erdmann, V.A., and Barciszewski, J. (2002) 5S Ribosomal RNA Database. *Nucl. Acids. Res.* **30**: 176-178.

Szymanski, M., Barciszewska, M.Z., Erdmann, V.A., and Barciszewski, J. (2003) 5 S rRNA: structure and interactions. *Biochem J* **371**: 641-651.

Tang, T.H., Bachellerie, J.P., Rozhdestvensky, T., Bortolin, M.L., Huber, H., Drungowski, M., Elge, T., Brosius, J., and Huttenhofer, A. (2002a) Identification of 86 candidates for small non-messenger RNAs from the archaeon Archaeoglobus fulgidus. *Proc Natl Acad Sci U S A* **99**: 7536-7541.

Tang, T.H., Rozhdestvensky, T.S., d'Orval, B.C., Bortolin, M.L., Huber, H., Charpentier, B., Branlant, C., Bachellerie, J.P., Brosius, J., and Huttenhofer, A. (2002b) RNomics in Archaea reveals a further link between splicing of archaeal introns and rRNA processing. *Nucleic Acids Res* **30**: 921-930.

Terns, M.P., and Terns, R.M. (2002) Small nucleolar RNAs: versatile trans-acting molecules of ancient evolutionary origin. *Gene Expr* **10**: 17-39.

Tollervey, D. (1987) A yeast small nuclear RNA is required for normal processing of pre-ribosomal RNA. *EMBO J* **6**: 4169-4175.

Tollervey, D., Lehtonen, H., Jansen, R., Kern, H., and Hurt, E.C. (1993) Temperature-sensitive mutations demonstrate roles for yeast fibrillarin in pre-rRNA processing, pre-rRNA methylation, and ribosome assembly. *Cell* **72**: 443-457.

Tollervey, D. (1996) Trans-acting factors in ribosome synthesis. *Exp Cell Res* **229**: 226-232.

Tollervey, D., and Kiss, T. (1997) Function and synthesis of small nucleolar RNAs. *Curr Opin Cell Biol* **9**: 337-342.

Tran, E.J., Zhang, X., and Maxwell, E.S. (2003) Efficient RNA 2'-O-methylation requires juxtaposed and symmetrically assembled archaeal box C/D and C'/D' RNPs. *EMBO J* **22**: 3930-3940.

Tycowski, K.T., Shu, M.D., and Steitz, J.A. (1996) A mammalian gene with introns instead of exons generating stable RNA products. *Nature* **379**: 464-466.

Tycowski, K.T., You, Z.H., Graham, P.J., and Steitz, J.A. (1998) Modification of U6 spliceosomal RNA is guided by other small RNAs. *Mol Cell* **2**: 629-638.

Ursic, D., Himmel, K.L., Gurley, K.A., Webb, F., and Culbertson, M.R. (1997) The yeast SEN1 gene is required for the processing of diverse RNA classes. *Nucleic Acids Res* **25**: 4778-4785.

Verheggen, C., Mouaikel, J., Thiry, M., Blanchard, J.M., Tollervey, D., Bordonne, R., Lafontaine, D.L., and Bertrand, E. (2001) Box C/D small nucleolar RNA trafficking involves small nucleolar RNP proteins, nucleolar factors and a novel nuclear domain. *EMBO J* **20**: 5480-5490.

Vidovic, I., Nottrott, S., Hartmuth, K., Luhrmann, R., and Ficner, R. (2000) Crystal structure of the spliceosomal 15.5kD protein bound to a U4 snRNA fragment. *Mol Cell* **6**: 1331-1342.

Vitali, P., Royo, H., Seitz, H., Bachellerie, J.P., Huttenhofer, A., and Cavaille, J. (2003) Identification of 13 novel human modification guide RNAs. *Nucleic Acids Res* **31**: 6543-6551.

Wang, H., Boisvert, D., Kim, K.K., Kim, R., and Kim, S.H. (2000) Crystal structure of a fibrillarin homologue from Methanococcus jannaschii, a hyperthermophile, at 1.6 A resolution. *EMBO J* **19**: 317-323.

Watanabe, Y., and Gray, M.W. (2000) Evolutionary appearance of genes encoding proteins associated with box H/ACA snoRNAs: cbf5p in Euglena gracilis, an early diverging eukaryote, and candidate Gar1p and Nop10p homologs in archaebacteria. *Nucleic Acids Res* **28**: 2342-2352.

Watanabe, Y., Yokobori, S., Inaba, T., Yamagishi, A., Oshima, T., Kawarabayasi, Y., Kikuchi, H., and Kita, K. (2002) Introns in protein-coding genes in Archaea. *FEBS Lett* **510**: 27-30.

Watkins, N.J., Segault, V., Charpentier, B., Nottrott, S., Fabrizio, P., Bachi, A., Wilm, M., Rosbash, M., Branlant, C., and Luhrmann, R. (2000) A common core RNP structure shared between the small nucleoar box C/D RNPs and the spliceosomal U4 snRNP. *Cell* **103**: 457-466.

Wehner, K., Ayala, L, Kim, Y, Young, PJ, Hosler, BA, Lorson, CL, Baserga, SJ, and Francis, JW (2002) Survival motor neuron protein in the nucleolus of mammalian neurons. *Brain Res* **945**: 160-173.

Weinstein, L.B., and Steitz, J.A. (1999) Guided tours: from precursor snoRNA to functional snoRNP. *Curr Opin Cell Biol* **11**: 378-384.

Winkler, W.C., Grundy, F.J., Murphy, B.A., and Henkin, T.M. (2001) The GA motif: an RNA element common to bacterial antitermination systems, rRNA, and eukaryotic RNAs. *RNA* **7**: 1165-1172.

Wu, P., Brockenbrough, J.S., Metcalfe, A.C., Chen, S., and Aris, J.P. (1998) Nop5p is a small nucleolar ribonucleoprotein component required for pre-18 S rRNA processing in yeast. *J Biol Chem* **273**: 16453-16463.

Xie, E., Zhu, L., Zhao, L., and Chang, L.S. (1996) The human serotonin 5-HT2C receptor: complete cDNA, genomic structure, and alternatively spliced variant. *Genomics* **35**: 551-561.

Xu, Y., Liu, L., Lopez-Estrano, C., and Michaeli, S. (2001) Expression studies on clustered trypanosomatid box C/D small nucleolar RNAs. *J Biol Chem* **276**: 14289-14298.

Yang, Y., Isaac, C., Wang, C., Dragon, F., Pogacic, V., and Meier, U.T. (2000) Conserved composition of mammalian box H/ACA and box C/D small nucleolar ribonucleoprotein particles and their interaction with the common factor Nopp140. *Mol Biol Cell* **11**: 567-577.

Yanisch-Perron, C., Vieira, J., and Messing, J. (1985) Improved M13 phage cloning vectors and host strains: nucleotide sequences of the M13mp18 and pUC19 vectors. *Gene* **33**: 103-119.

Yi-Brunozzi, H.Y., Easterwood, L.M., Kamilar, G.M., and Beal, P.A. (1999) Synthetic substrate analogs for the RNA-editing adenosine deaminase ADAR-2. *Nucleic Acids Res* **27**: 2912-2917.

Yokoyama, S., Inagaki, F., and Miyazawa, T. (1981) Advanced nuclear magnetic resonance lanthanide probe analyses of short-range conformational interrelations controlling ribonucleic acid structures. *Biochemistry* **20**: 2981-2988.

Yu, L., Nguyen, H., Le, H., Bloem, L.J., Kozak, C.A., Hoffman, B.J., Snutch, T.P., Lester, H.A., Davidson, N., and Lubbert, H. (1991) The mouse 5-HT1C receptor

contains eight hydrophobic domains and is X-linked. *Brain Res Mol Brain Res* **11**: 143-149.

Zillig, W., Stetter, K. O., Wunderl, S., Schulz, W., Priess, H., Scholz, J. (1980) The Sulfolobus-"Caldariella" group: taxonomy on the basis of the structure of DNA-dependent RNA polymerase. *Arch. Microbiol.* **125**: 259-269.

Zuker, M. (2003) Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res* **31**: 3406-3415.