Molecular Microbiology (2001) **40**(3), 509–519

MicroReview

A guided tour: small RNA function in Archaea

Patrick P. Dennis,[1]* Arina Omer[1] and Todd Lowe[2]
[1]Department of Biochemistry and Molecular Biology, University of British Columbia, 2146 Health Sciences Mall, Vancouver, BC V6T 1Z3, Canada.
[2]Department of Genetics, Stanford University, Stanford, CA 94305-5120, USA.

Summary

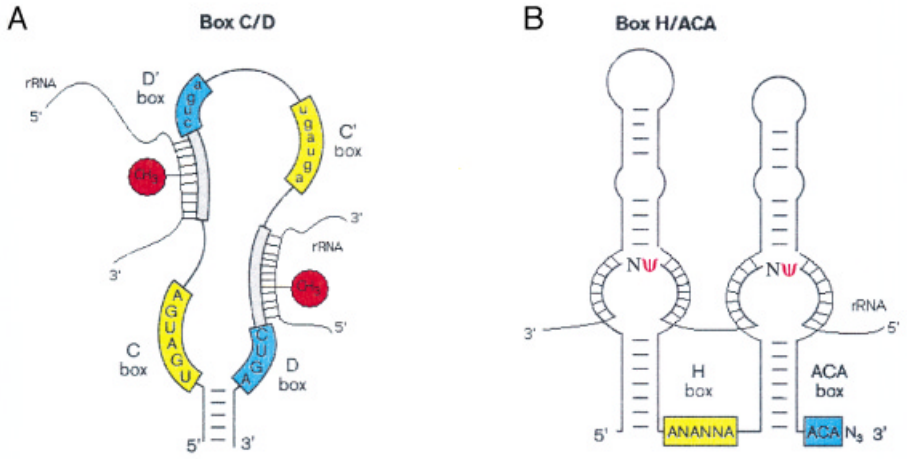
In eukaryotes, the C/D box family of small nucleolar (sno)RNAs contain complementary guide regions that are used to direct 2′-*O*-ribose methylation to specific nucleotide positions within rRNA during the early stages of ribosome biogenesis. Direct cDNA cloning and computational genome searches have revealed homologues of C/D box snoRNAs (called sRNAs) in prokaryotic Archaea that grow at high temperature. The guide sequences within the sRNAs indicate that they are used to direct methylation to nucleotides in both rRNAs and tRNAs. The number of sRNA genes that are detectable within currently sequenced genomes correlates with the optimal growth temperature. We suggest that archaeal sRNAs may have two functions: to guide the deposition of methyl groups at the 2′-*O* position of ribose, which is an important determinant in RNA structural stability, and to serve as a molecular chaperones to help orchestrate the folding of rRNAs and tRNAs at high temperature.

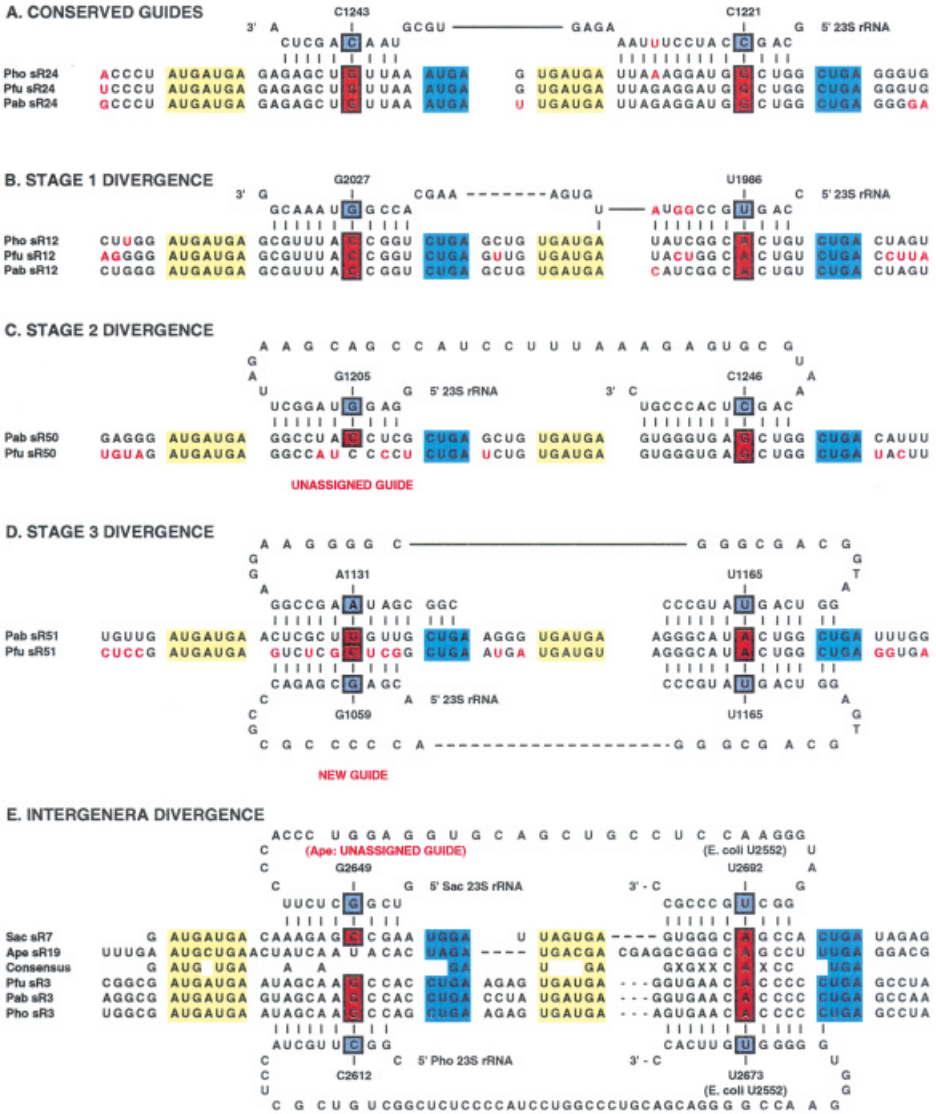
Introduction

The biosynthesis of ribosomes in eukaryotes occurs within a highly specialized organelle, the nucleolus (reviewed by Maxwell and Fournier, 1995; Bachellerie and Cavaille, 1998; Lafontaine and Tollervey, 1998; Bachellerie *et al*., 2000). The nucleolar machinery systematically integrates the transcription of rRNA genes, the maturation and folding of rRNA, and the packaging of the rRNA and r-proteins into small and large ribosomal subunits. Three of the four rRNAs (18S, 5.8S and 25-28S) are produced by endo- and exonuclease cleavage of a single long RNA polymerase I generated transcript. The fourth rRNA (5S) is transcribed independently by RNA polymerase III. During the transcription, folding and maturation processes, up to a hundred or more post-transcriptional nucleotide modifications are introduced at specific positions within the eukaryotic rRNAs. The two most frequent types of modification are methylation at the 2′-O position of ribose (about 55 and 107 in yeast and humans respectively) and isomerization of uridine to pseudouridine (about 44 and 95 in yeast and humans respectively) (Maden, 1990). Although the function of these modifications remains largely enigmatic, it is clear that virtually all of them are confined to the most conserved and functionally important regions of the rRNA.

How are the large number of ribose methylations and pseudouridine modifications targeted to specific nucleotide positions within eukaryotic rRNAs? Early clues came from the observation that the eukaryotic nucleolus contains a very large number of small nucleolar RNAs (snoRNAs), many of which contain one or more antisense elements that are complementary to sequences in rRNA (Bachellerie *et al*., 1995; Steitz and Tycowski, 1995). Almost all of these snoRNAs fall into one of two distinct classes: C/D box snoRNAs which contain C box (RUGAUGA) and D box (CUGA) motifs near their 5′ and 3′ ends, and H/ACA snoRNAs which have a common secondary structure and contain a conserved ACA trinucleotide invariably positioned three nucleotides from the end of the RNA (Balakin *et al*., 1996; see Fig. 1). The C/D box snoRNAs usually have a short-terminal hairpin, associate with several essential nucleolar proteins (fibrillarin, NOP56, NOP58 and 15.5 kDa protein; Gautier *et al*., 1997; Watkins *et al*., 2000) and function to guide methylations using the antisense elements located 5′ to either the D or D′ (internal D box-like) motifs. Methylation is directed to the rRNA nucleotide that participates in the base pair five nucleotides upstream from the start of the D or D′ box (Kiss-Laszlo *et al*., 1996; Tycowski *et al*., 1996a). There is some evidence to suggest that fibrillarin mediates methylation; it contains a weak match to an S-adenosyl methionine-binding motif and amino acid substitutions in this motif exhibit a temperature sensitive defect in global rRNA methylation (Tollervey *et al*., 1993; Wang *et al*., 2000). The NOP56 and NOP58 proteins are paralogues (the result of a gene duplication) with essential

Accepted 29 January, 2001. *For correspondence. E-mail pdennis@nsf.gov; Tel. (+1) 604 822 5975; Fax (+1) 604 822 5227.


A
Box C/D
rRNA
5'
D' box
C' box
3'
rRNA
3'
5'
C box
D box
5'
3'
B
Box H/ACA
NΨ
NΨ
rRNA
3'
5'
H box
ACA box
5'
ANANNA
ACA N3 3'


A. CONSERVED GUIDES
C1243
C1221
5' 23S rRNA
Pho sR24
Pfu sR24
Pab sR24
B. STAGE 1 DIVERGENCE
G2027
U1966
5' 23S rRNA
Pho sR12
Pfu sR12
Pab sR12
C. STAGE 2 DIVERGENCE
G1205
C1246
5' 23S rRNA
Pab sR50
Pfu sR50
UNASSIGNED GUIDE
D. STAGE 3 DIVERGENCE
A1131
U1165
5' 23S rRNA
G1059
U1165
Pab sR51
Pfu sR51
NEW GUIDE
E. INTERGENERA DIVERGENCE
(Ape: UNASSIGNED GUIDE)
(E. coli U2552)
G2649
U2692
5' Sac 23S rRNA
Sac sR7
Ape sR19
Consensus
Pfu sR3
Pab sR3
Pho sR3
5' Pho 23S rRNA
C2612
U2673
(E. coli U2552)

but uncharacterized functions (Gautier *et al*., 1997). The H/ACA box snoRNAs associate with a different set of essential proteins (including Cbf5p, Nhp2p, Nop10 and Gar1p) and exhibit complementarity to rRNA (Watkins *et al*., 1998; Henras *et al*., 1998). Duplex formation places the uridine to be modified into a conserved higher order structure termed the pseudouridylation pocket (Ganot *et al*., 1997a, b; Ni *et al*., 1997). The Gar1p protein binds directly to the H/ACA RNAs *in vitro*, and the Cbf5p protein is almost certainly the pseudouridine synthase; it has high sequence similarity to the protein that catalyses the formation of ψ55 in *Escherichia coli* tRNAs. The function of C/D box and H/ACA box snoRNAs in guiding 2′-*O*-ribose methylation and pseudouridylation has been elegantly demonstrated (Kiss-Laszlo *et al*., 1996; Ganot *et al*., 1997a; Ni *et al*., 1997). These site-specific methyl and pseudouridine modifications are believed to be involved in one or more of the following functions: enhancing RNA structural stability, providing unique features for binding of ribosomal proteins, or enhancing the activity of the rRNA within the assembled ribosome. In addition, it has been suggested that the base pairing between snoRNAs and the nascent rRNA transcript may have a chaperone function whereby the folding of the rRNA is constrained to a productive as opposed to a dead end pathway (Steitz and Tycowski, 1995; Ofengand and Fournier, 1998). In this case, the modification might be a benign by-product of the chaperone function and thus, simply a signal for dissociation of the guide RNA/rRNA complex.

In contrast to this rather complex situation, the rRNA of a typical bacterium, *E. coli*, contains only four ribose methylation and 10 pseudouridine modifications; all of these appear to be protein directed with no evidence for RNA cofactor involvement (Gustafsson *et al*., 1996; Ofengand and Fournier, 1998).

Discovery of methylation guide sRNAs in Archaea

Archaea are prokaryotic organisms that lack a nucleus but possess DNA replication, transcription and translation machineries that more closely resemble those of Eukarya than those of Bacteria (Dennis, 1997; Olsen and Woese, 1997). Two observations suggested that Archaea might contain homologues to the C/D box family of eukaryotic snoRNAs. First, all sequenced archaeal genomes contain two open reading frames (ORFs) that encode the protein homologues of the eukaryotic nucleolar proteins fibrillarin and NOP56/58 (Amiri, 1994; Omer *et al*., 2000). Second, mass spectrometry analysis of the rRNA of the hyperthermophilic archaeon *Sulfolobus solfataricus* revealed the presence of 67 sites of ribose methylation, a number similar to that found in eukaryotes (Noon *et al*., 1998). The positions of the ribose-methylated nucleotides in the 16S rRNA are currently being determined. In contrast to the large number of ribose methyl modifications, the number of pseudouridine modifications is small (about eight), similar to the number found in *E. coli* and considerably fewer than the number found in yeast or humans.

The first demonstration of archaeal C/D box small RNAs (sRNAs) took advantage of the fact that in eukaryotes, these C/D box snoRNAs are in ribonucleoprotein (RNP complexes) containing fibrillarin, NOP56 and NOP58 proteins (Omer *et al*., 2000). The archaeal genes encoding homologues of these proteins (aFIB and aNOP56) were cloned from *Sulfolobus acidocaldarius* and expressed in *E. coli*; the expressed proteins were purified and used to generate polyclonal antibodies in rabbits. Immunoprecipitates of *S. acidocaldarius* cell extract that had been fractionated on a glycerol gradient were used to selectively purify associated RNAs. The RNAs were subjected to RT-PCR and the products were used to generate cDNA libraries. Representatives of 29 different

Fig. 1. Structural features of two major classes of eukaryotic snoRNAs. The C/D box and H/ACA box snoRNAs use antisense guide elements to target ribose methylation and pseudouridylation in rRNA.
A. The C/D box snoRNAs often contain one or two regions of complementarity to rRNA that are positioned 5′ to the D or D′ box; 2′-*O*-ribose methylation is directed to the nucleotide in rRNA that participates in a Watson–Crick base pair five nucleotides upstream of the D′ or D box. Most of these RNAs have a short-terminal hairpin.
B. The H/ACA box snoRNAs contain one or two regions of hyphenated complementarity to rRNA that are 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.

Fig. 2. Evolutionary divergence of sRNA genes in Archaea. Archaeal sRNAs that guide methylation to homologous positions are aligned and the base pairing to their target sequence in 16S or 23S rRNA are indicated. Positions of nucleotide substitution in the aligned sequences are highlighted in red. Where necessary, straight lines (—) are used to connect uninterrupted sequences and stippled lines (- - -) are used to indicate sequences missing a small number of nucleotides. In (A) to (E), possible stages of evolutionary divergence are illustrated. In (A) and (B) both guides are functional and direct methylation to homologous positions. In (C), the D′ guide in *Pfu* sR50 has no apparent RNA target. In (D), the D′ guide of *Pab* and *Pfu* sR51 direct methylation to non-homologous positions. In (E), the family of archaeal sRNAs that directs methylation to position U2552 in 23S rRNA is aligned. The sites of co-variation between the D guide and the rRNA target are indicated (X) within the consensus. The D′ guide of *Ape* sR19 is unassigned, whereas the D′ guides of *Sac* sR7 and *Pyrococcus* sR3 direct methylation to non-homologous positions 23S G2649 and C2612 respectively.

Table 1. Putative or confirmed sRNA genes in archaeal genomes.

Archeal species	Number of sRNA genes	Optimum growth temperature (°C)
Crenarchaeota		
Sulfolobus acidocaldarius (*Sac*)	29[a]	75–80
Sulfolobus solfataricus (*Sso*)	13[bc]	75–80
Aeropyrum pernix (*Ape*)	23	90–95
Pyrobaculum aerophylum (*Pae*)	> 50[d]	95–100
Euryarchaeota		
Halobacterium salinarium sp. NRC1 (*Hsa*)	0[c]	37–42
Methanobacterium thermoautotrophicum (*Mth*)	0[c]	62–67
Archaeoglobus fulgidus (*Afu*)	4[c]	65–70
Methanococcus jannaschii (*Mja*)	8[c]	80–85
Pyrococcus horikoshii (*Pho*)	51	95–100
Pyrococcus abysii (*Pab*)	52	95–100
Pyrococcus furiosus (*Pfu*)	56	95–100

a. The sequence of the original set of 18 cDNA clones of *S. acidocaldarius* sRNAs and their annotated targets have been published (Omer *et al.*, 2000). An additional 11 unique cDNA clones have been isolated and sequenced but have not been further characterized (see: rna.wustl.edu/snoRNAdb/).
b. Only about two-thirds of the *S. solfataricus* genome sequence was used for search analysis. The small number of identified genes relative to the number of known ribose methylations in rRNA suggests either that a large proportion of sRNA genes may be non-canonical and have been overlooked by our search program or that some of the methylations are non-guide RNA directed.
c. In the searches for sRNA genes in archaeal genomes, we used rather conservative cut-off limits. As a consequence, it is likely that a number of authentic sRNA genes in at least some of the species have been excluded. Identification of these false negatives will require independent confirmation by primer extension, northern hybridization, or some other techniques. In our search of the *H. salanarium* genome, we screened for the top 10 candidate sRNAs (all had low confidence scores) by primer extension. None gave detectable extension products of the predicted length. The *H. salanarium* genome encodes aFIB and aNOP56 homologous proteins. This implies that the C/D box sRNAs in halophiles may be very divergent and non-canonical compared to the sRNAs in non-halophilic species.
d. In a preliminary screen of the *P. aerophylum* genome, we identified more than 50 putative sRNA genes. At present, the characterization of these sRNA are incomplete and the sequences have not yet been made available on the archaeal sRNA website.

sequences that exhibit features characteristic of eukaryotic C/D box snoRNAs have been recovered from the libraries. Although somewhat shorter than typical eukaryotic C/D box snoRNAs (a median length of 57 nucleotides compared with a median greater than 70 nucleotides in length), the *S. acidocaldarius* sRNAs exhibit the characteristic dyad structure with well-defined C, D′, C′ and D box motifs, and often contain one or two rRNA antisense guide elements positioned 5′ to the D′ and/or D box motifs. Few of the *S. acidocaldarius* C/D box sRNAs appear to have the short-terminal hairpin that is commonly present in eukaryotic C/D box snoRNAs. The presence of many of these sRNAs has been confirmed by primer extension and/or Northern hybridization analysis (for a summary of this information, see: rna.wustl.edu/snoRNAdb/).

Support for sRNA methyl guide function requires mapping the modification state of target RNAs. Unambiguous detection of positions of ribose methyl modification in large RNAs is tedious, requiring fingerprint or mass spectrometry analysis of isolated oligonucleotides. Other more convenient assays such as partial alkaline hydrolysis or dNTP concentration-dependent primer extension pause reactions are quick and simple but less reliable for unambiguous identification and detection of methyl modifications (Maden *et al.*, 1995; discussed in Bachellerie and Cavaille, 1998). Our examinations of the rRNA antisense guide elements in the 29 *S. acidocaldarius* sRNAs predict 26 different sites of rRNA methylation using the eukaryotic D/D′ box plus five base pair methylation guide rule. Two of our predictions (U52 and G1056 in 16S rRNA) are known from oligonucleotide fingerprint analysis to have some type of nucleotide modification (references cited in Omer *et al.*, 2000). To confirm our predictions, we used the dNTP concentration-dependent pause reaction, and to a lesser extent, the alkaline hydrolysis reaction; a number of the candidate sites exhibit the expected pause product, whereas others did not (see: rna.wustl.edu/snoRNAdb/). In these negative instances, it is unclear whether the assay has failed to detect the methyl modification on the RNA template, or whether the site is indeed unmodified. In eukaryotes, alteration or disruption of snoRNAs or snoRNA-encoding genes has been used to demonstrate the direct involvement of these guide RNAs in rRNA modification (Kiss-Laszlo *et al.*, 1996; Lowe and Eddy, 1999). Due to the lack of sophisticated experimental genetic systems, similar disruption or epigenetic experiments with *Sulfolobus* and other Archaea are not yet feasible.

Examination of the *S. acidocaldarius* sRNA sequences revealed that only about half of the potential guide regions exhibit complementarity to rRNA. If all rRNA methyl modifications are guide directed and if the number of modifications is similar to that of a related species, *S. solfataricus* (67 sites), then we expect that there should be between 60 and 70 C/D box sRNAs in *S. acidocaldarius*. The possible significance of guides without complementarity ('unassigned guides') is discussed below.

Mining archaeal genome sequences

A similarity search (BLASTN; Altschul *et al*., 1990) using each of the *S. acidocaldarius* sRNA genes against the non-redundant nucleotide database resulted in only two weak hits, both within the partial genome sequence of the related organism, *S. solfataricus*. This result indicated that BLASTN analysis was not an effective tool for detecting sRNA gene homologues because the sequences are short and apparently diverge very rapidly from each other (see below). As an alternative, we retrained a previously developed eukaryotic snoRNA search program with the sequences of the verified *S. acidocaldarius* sRNA genes. This program was very effective in identifying previously unknown C/D box snoRNA genes in the yeast genome (Lowe and Eddy, 1999). The retrained program was used to search the partial genome sequence of *S. solfataricus* and the complete genome sequence from nine other archaeal species (Table 1). An up-to-date list of these sRNA sequences along with annotations of their methylation target sites is available at rna.wustl.edu/snoRNAdb/. In some genomes, more than 50 putative sRNA genes were detected, whereas in other genomes, few or no high probability candidates were found.

The genomes of three closely related species of *Pyrococcus* have been sequenced and analysed for sRNA genes by combining sRNA gene searches with comparative interspecies genome analysis (Gaspin *et al*., 2000; Omer *et al*., 2000). Using this strategy, a total of 60 groups of homologous sRNA genes were identified. Fifty groups have representatives in all three species, eight are represented in only two species, and two are unique to single species. Examples of several of these homology groups are illustrated in Fig. 2; the entire set can be viewed at the snoRNA database website (rna.wustl.edu/snoRNAdb/). In each homology group, the sequence identity for end-to-end alignments (from the 5′ end of box C to the 3′ end of box D) of interspecies members is between 75% and 98%. The least conserved region is the short connector segment between the D′ and C′ boxes; the rate of divergence of the connector is roughly equal to that of the non-coding flanking sequences. The box features (C, D′, C′ and D) are well conserved with consensus motifs identical to those found in eukaryotic C/D box snoRNAs. As demonstrated in eukaryotes, the positions of the D′ and D boxes appear to guide methylation to the fifth upstream base pair, within the helix formed by the rRNA and the antisense guide. The strong intrafamily conservation of the C and D boxes plus high similarity to the eukaryotic motifs suggests that the C and D boxes play a similar role in Archaea for sRNA association with aFIB and aNOP56.

Pyrococcus sRNAs are generally between 50 and 60 nucleotides in length and exhibit the characteristic snoRNA dyad structure. More than half of the sRNAs 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. Also in contrast to eukaryotes, the predicted methylation targets are often in close proximity within the same molecule. If both guides pair simultaneously to the nascent rRNA, a stabilizing complex could be generated (see Fig. 2C–E).

Genomic context of sRNA genes

The availability of a large number of sRNA genes within sequenced genomes makes it possible to assess their location relative to protein coding genes. In eukaryotes and particularly in mammals, most snoRNA genes are encoded in the introns of translation-related protein encoding genes, and are generated by post-transcriptional processing of the pre-mRNA (Bachellerie and Cavaille, 1998). Some mammalian snoRNAs are even found in the introns of host genes that are transcribed by RNA polymerase II and processed, but apparently never translated into proteins (Tycowski *et al*., 1996b; Pelczar and Filipowicz, 1998; Smith and Steitz, 1998). In yeast, most snoRNAs are transcribed from their own promoter. These snoRNAs are generated by endo- and exonucleolytic processing of the mono- or polycistronic transcripts. One particular transcriptional unit encodes seven distinct C/D box snoRNAs (Qu *et al*., 1999).

Based on the current collection, archaeal sRNAs appear to be encoded on both strands of the DNA, usually unlinked, and distributed around the entire circular chromosome (for genomic positions of sRNA genes, see: rna.wustl.edu/snoRNAdb/). Furthermore, they are almost always positioned within the short spacer regions between protein encoding ORFs. Between 10% and 20% of the sRNA genes appear to overlap the 3′ ends of protein-encoding ORFs. The translation termination codon of the ORF often falls in the 5′ half of the sRNA, within the C box (RUGAUGA), D′ box (CUGA) or the guide region between them. A well-characterized example of an overlap with the 3′ end of an ORF is the sR1 genes from *S. acidocaldarius* and *S. solfataricus* (Fig. 3). Remarkably, the most highly conserved portion of these sRNA genes is the 3′ half that lies outside the ORF. It contains the D box antisense element that is used to

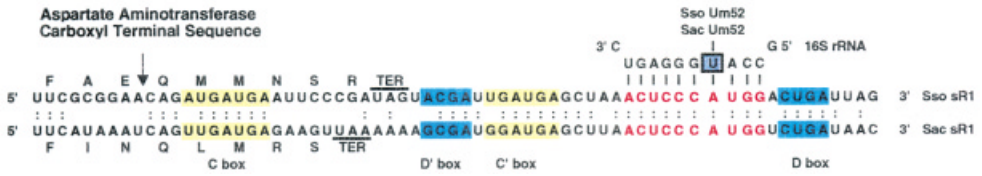


Fig. 3. Overlap between the 3′ end of the aspartate amino transferase gene and the sR1 genes in *S. solfataricus* and *S. acidocaldarius*. The RNA sequences derived from the 3′ end of the aspartate amino transferase and the sR1 genes in *S. solfataricus* (*Sso*) and *S. acidocaldarius* (*Sac*) are illustrated. The C-terminal amino acid sequence of the encoded proteins are above or below the RNA sequence; identical nucleotides in the two RNAs (:) and the approximate 5′ end of the sRNAs are indicated. The D guides that direct methylation to position U52 in both 16S rRNAs are shown in red. 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.

guide methylation to position U52 in 16S rRNA. The 5′ half of the sR1 gene is poorly conserved and is reflected in numerous amino acid replacements in the C-terminus of the encoded aspartate amino transferase protein. The translation termination codons, although offset by three nucleotides, are both located within the complementary guide region associated with the D′ box. The *Sso* sR1 D′ guide is predicted to direct methylation to position U33 in 16S rRNA, whereas the *Sac* sR1 D′ guide does not have any strong complementarity to rRNA. The D box guide of the *Pyrococcus* sR4 family also directs methylation to position U52 in 16S rRNA. In this instance, the *Pho*, *Pfu* and *Pab* sR4 genes are located entirely within the intergenic space between two ORFs that encode uncharacterized proteins.

Based on genome annotations, a number of sRNA genes would appear to overlap the 5′ end of protein-encoding ORFs. Many of these apparent overlaps may be artifactual, resulting from the incorrect assignment of the translation initiation codon that defines the 5′ end of the ORF. We have observed only one case where an sRNA is encoded completely within another gene: the *Pyrococcus* sR40 family and their homologues – including *Archaeoglobus fulgidus (Afu)* sR3 – reside within the intron in the anticodon loop of the gene encoding tRNA-Trp. Daniels and co-workers (personal communication) have suggested that these sRNAs function *in cis* to guide methylation to positions C34 and C39 within the intron-containing tRNA-Trp precursor (see below). We identified the *Pyrococcus* sR40 group of sRNAs based on their hallmark features and D′ and D guide complementarities to rRNA (D′, 16S C1252; D, 23S C1117). It remains to be

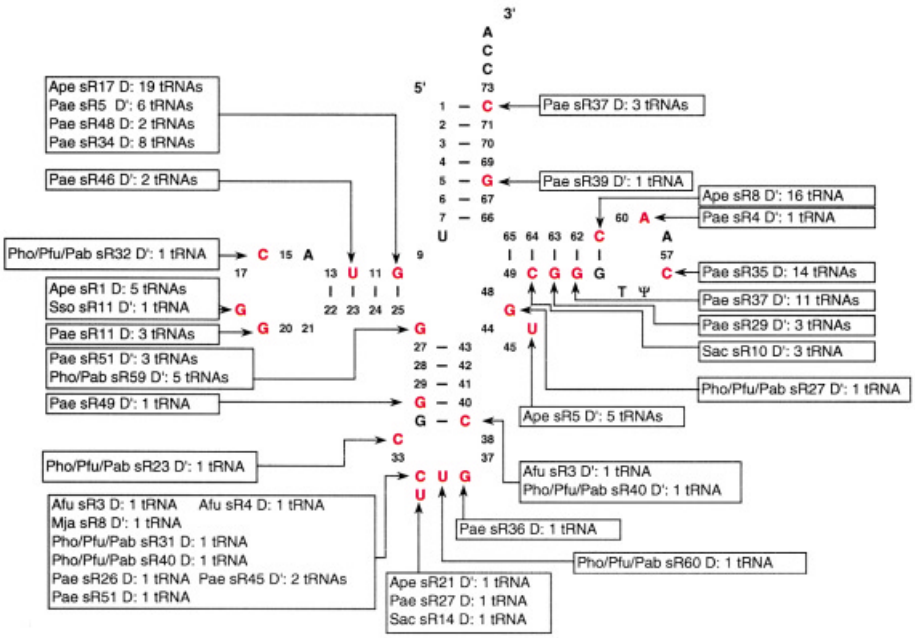


Fig. 4. Predicted methylation sites in archaeal tRNAs. The structure of a tRNA is depicted in the standard cloverleaf configuration. The positions predicted to be methylated based on sRNA guide sequence complementarity are indicated in red. Other conserved tRNA nucleotides are indicated as letters. Variable nucleotides are indicated by number only.

Table 2. Linkage between *Pyrococcus* sRNA genes.

sRNA gene pairs[a]	Orientation	*Pho*	*Pfu*	*Pab*
sR2/sR9	Opposite strand	109 bp	130 bp	116 bp
sR12/sR34	Opposite strand	1 bp	49 bp	1 bp
sR14/sR22	Opposite strand	59 bp	40 bp	56 bp
sR26/sR60	Same strand	7 bp	9 bp	11 bp
sR50/sR54	Same strand	NF[b]	34 bp	UL[c]

a. For the genomic location of sRNA genes, see: rna.wustl.edu/snoRNAdb/.
b. NF, not found; the sR50 gene is not found in *P. horikoshii.*
c. UL, unlinked; the two sRNA genes are separated by over 300 kbp.

shown whether these sRNAs truly guide methylation to the proposed tRNA and rRNA positions.

Within the *Pyrococcus* genomes there are five instances where two sRNA genes are closely linked (Table 2). For three pairs, the genes are on opposite strands and are divergently transcribed, separated by 1–130 bp. The remaining two pairs are oriented on the same strand, separated by 9–34 bp. In *S. solfataricus*, the sR10 and sR11 genes are also encoded on the same strand and separated by only five nucleotides. Primer extension and Northern hybridization analysis have shown that both the *Sso* sR10/sR11 and the *Pfu* sR26/sR60 gene pairs are co-transcribed and at least partially processed to their monomeric size (A. Omer, unpublished results; Gaspin *et al*., 2000).

The question as to how isolated sRNA genes are transcribed and how their transcripts are processed has not been addressed. Two possibilities warrant consideration. Either sRNA genes are transcribed from sRNA-specific promoters – which in some cases would probably lie within protein-encoding ORFs – or the genes are co-transcribed from mRNA promoters. In the latter instance, the sRNA product would be salvaged from an intermediate in the mRNA degradation pathway. Primer extension with sRNA-specific primers frequently reveals what appear to be minor amounts of precursors with extra nucleotides at the 5′ end (A. Omer, unpublished results).

Is growth temperature related to rRNA methylation?

The number of methylation guide sRNAs that we detected in searches of archaeal genomes correlates with the optimum growth temperature (Table 1). In the genomes *of Halobacterium salinarium*, a mesophile, and *Methanobacterium thermoautotrophicum*, a moderate thermophile, no highly probable sRNA candidate genes were detected. Both genomes encode aFIB and aNOP56 homologous proteins, implying that these organisms may possess methylation guide sRNAs. The features of these hypothetical sRNAs genes would differ substantially from those used to train the search program, and thus could have escaped detection. In the hyperthermophilic *Pyrococcus* and *Pyrobaculum* species with optimum growth temperatures approaching 100°C, our searches revealed more than 50 easily recognizable and highly canonical sRNA genes (the *Pyrococcus* data are available at: rna.wustl.edu/snoRNAdb/ the *Pyrobaculum* data are currently unpublished results). If we infer that the number of identified guide sRNAs is proportional to the true number of rRNA methylation sites, this would suggest that increased RNA methylation is important for life at high temperature. Advantages of increased rRNA methylation may include increased thermodynamic structural stability or augmented interactions with stabilizing ribosomal proteins.

In the context of RNA stabilization by increased ribose methylation, an interesting question has been raised. Do hyperthermophilic Archaea regulate the extent to which they methylate their rRNA in response to the environmental growth temperature? In a profoundly interesting study, the number of ribose-methylated nucleotides in the rRNA of *S. solfataricus* grown at 60°C, 75°C and 83°C was observed to be 62.2, 67.5 and 69.6, respectively, an increase of about 12% (Noon *et al*., 1998). At the present time, it is unclear how significant and reproducible these values are. It is also unclear whether the increase in modifications reflects regulated activation of the methylation machinery by high temperature or is simply unregulated inhibition (cold sensitivity) at low temperature. Even if the increase is concerted, its adaptive value in protecting organisms against rapid fluctuation in temperature is not immediately obvious because only newly synthesized ribosomes (not pre-existing ribosomes) would be affected. If the system is concerted and has adaptive value, it would provide credibility to the proposed chaperone function of these RNAs. It will be interesting and important to identify sites of regulated methylation, whether they exist, and to determine their role in ribosome assembly or activity.

Do sRNAs guide methylation to tRNAs?

In many archaeal sRNAs, one or both of the D′ or D box guide regions lack complementarity to 16S or 23S rRNA. We examined these unassigned guides for complementarity to other stable RNAs and noted that many exhibit complementarity to the tRNAs derived from the same organism (Omer *et al*., 2000; see also: rna.wustl.edu/snoRNAdb/). The predicted positions of methylation in tRNAs are summarized in Fig. 4. The guides are most often specific to a single tRNA sequence. For example, sRNAs from eight different species appear to target unique methylations to U34 or C34, the wobble position within the anticodon of the tRNAs. This site is known to be ribose methylated in many tRNAs, including two archaeal

Table 3. Methylation at homologous positions within small and large subunit rRNAs.

	Homologous sites conserved between:				
	Archaeal[a] sites	Archaeal[b] genera	Humans/ yeast[c]	Archaea/ Eukarya[d]	Archaea/ Eukarya/ *E. coli*[e]
SSU rRNA	83	11	13	2	1
LSU rRNA	171	31	23	7	2

a. Sites that are predicted to be methylated based upon the eukaryotic guide rule (methylation five nucleotides upstream of the D/D′ box). The criteria for guide assignment are generally a minimum of 8 bp with no more than one G/U base pairs and one mismatch. For a few double guide sRNAs where the targets are close to each other, a mismatch was permitted at positions 4, 5 or 6. In these instances, the complementarity with rRNA may form but the target base may not be methylated. For identification of target assignments, see: rna.wustl.edu/snoRNAdb/.
b. The number of homologous sites within rRNA that are targeted for methylation by an sRNA guide region in more than one archaeal genera.
c. The number of sites of rRNA methylation that are shared between yeast and humans.
d. The number of sites of rRNA methylation that are shared between yeast, human, and at least one species of Archaea.
e. The number of sites of rRNA methylations that are shared between yeast, human, at least one species of Archaea and *E. coli*.

tRNAs (tRNA-Trp and tRNA-Met in *Haloferax volcanii*; Gupta, 1984). In contrast, some sRNAs target a large number of tRNAs. For example, *Ape* sR17 is predicted to direct methylation at position G10 in 19 different tRNAs. Three *Pae* sRNAs (sR5, sR48 and sR34) appear to target methylation to the same G10 in a total of 16 different tRNAs. In the exceptional case of *Pho/Pfu/Pab* sR40 and *Afu* sR3 (the sRNAs that reside in the intron of the tRNA-Trp genes), the sRNAs are believed to be required to direct methylation to positions C34 and C39 within the precursor before the intron can be excised (C. Daniels, personal communication). After excision, the *Pyrococcus* intronic sRNAs may also guide methylation to positions C1252 in 16S and C1171 in 23S rRNA.

Although newly synthesized eukaryotic tRNAs may transit through the nucleolus prior to nuclear export, there is no evidence to date that any of their numerous nucleotide modifications are snoRNA directed. For example, none of the identified C/D box snoRNAs from yeast (Lowe and Eddy, 1999) appear to have guides that are complementary to tRNA sequences. So why would hyperthermophilic Archaea use sRNAs to guide methylation in tRNAs? Ribose methylation increases base stacking, the single most important energetic interaction in RNA (Davis, 1998). We imagine that in high-temperature Archaea, pretRNAs assume the correct tertiary structure only after ribose methyl modification by the guide-RNA-directed mechanism. Indeed, in the case of the intron containing pretRNA-Trp, *cis*-guided methylation at positions C34 and C39 appears to be required before the pre-tRNA can refold into a structure that allows intron excision and exon ligation (C. Daniels, personal communication).

Are methylation sites in rRNA phylogenetically conserved?

Based on the complementary guide sequences in the available collection of sRNAs, a count of all potential sites of ribose methylation in archaeal rRNA has been compiled (Table 3). Of the 253 archaeal sites, 44 are used in more than one archaeal genus, whereas the vast majority, 209 sites, are so far unique to a single genus. We compared all these positions of predicted archaeal ribose methylation with the 36 methylated sites conserved between yeast and humans. Only nine of these are shared between the two domains and three of the nine correspond to three of the four sites of methylation in *E. coli* rRNA (positions 16S C1402, 23S G2251 and 23S C2498; *E. coli* numbering).

Positions of methylation in rRNA are confined to the common core regions and cluster somewhat to the functionally important regions. The distribution of archaeal, yeast/human and *E. coli* sites within three core regions are illustrated in Fig. 5: the '530' translational fidelity stem loop and the 3′ terminal decoding stem in SSU rRNA, and the peptidyl transferase centre of LSU rRNA. For the nine yeast/human sites within these regions, three have not yet been predicted by sRNAs in any Archaea, four are predicted in only a single archaeal genus, and only two are predicted in more than one archaeal genus. Two of the positions, 23S G2251 and U2552, are among the three ribose methylation sites present in *E. coli* 23S rRNA. The loop region containing position G2251 has been implicated in interaction with the CCA 3′ end of the P site tRNA. The loop containing position U2552 has been implicated in an interaction with the CCA 3′ end of the A site rRNA (Moazed and Noller, 1989; Samaha *et al*., 1995; Kim and Green, 1999). Thus, it appears that with only a few notable exceptions, there is little current evidence for selective pressure in maintaining positions of guide-directed-rRNA-ribose methylation both within Archaea and between Archaea and Eukarya. This suggests to us that the number and distribution of methylation sites – that is, the number and distribution of predicted guide RNA–rRNA interactions – may be more important than their precise positioning in some critical areas. For example, of the 10 archaeal sites that

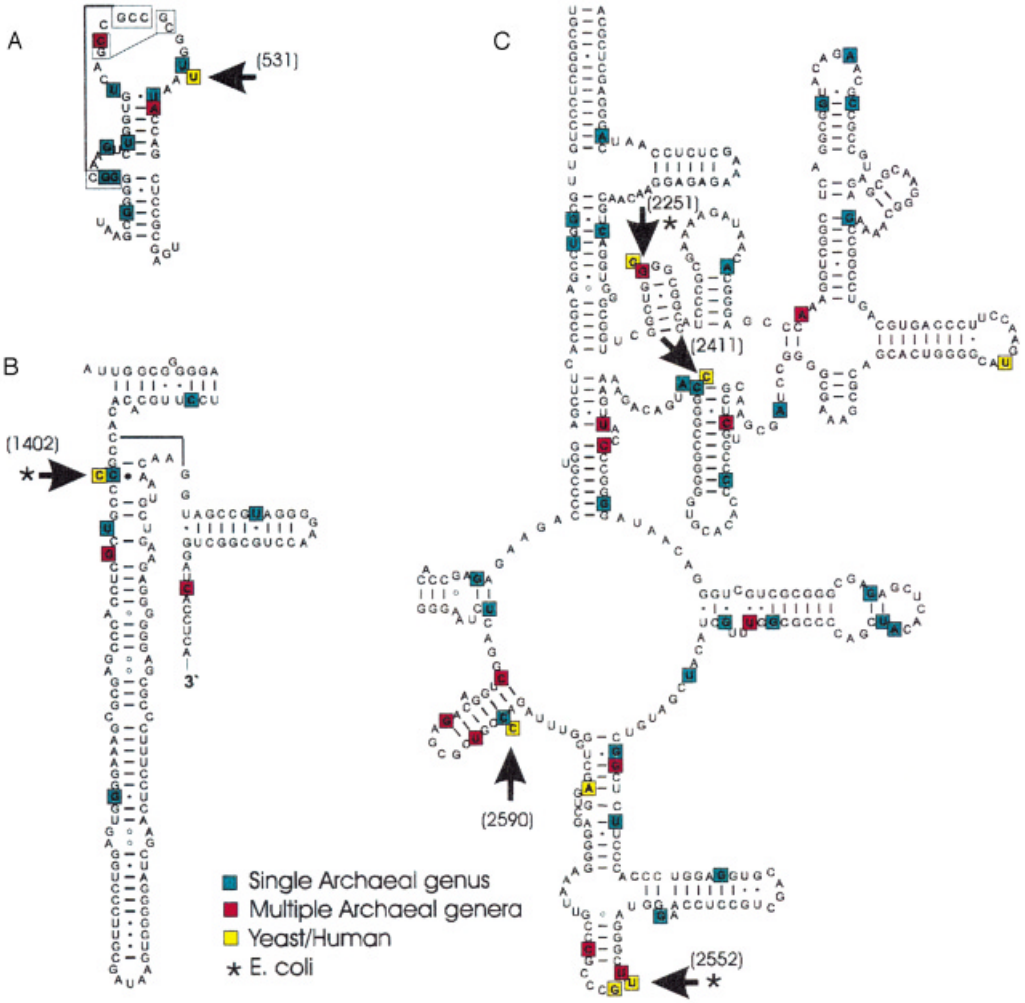


Fig. 5. Conservation of sites of 2′-*O*-ribose methylation within functionally important domains of the SSU and LSU RNAs of Eukarya, Archaea and Bacteria.The sequence and structure of the 530 translational fidelity stem loop (A) and the 3′ terminal decoding stem (B) of SSU RNA and peptidyl transferase centre (C) of LSU RNA of *S. acidocaldarius* are depicted. Potential sites of methylation in archaeal rRNAs based on guide sequence complementarity are mapped on these sequences and are boxed in green (found in only one genus) or red (found in more than one genus). The positions of methylation that are conserved between yeast and humans are boxed in yellow. An arrow indicates coincidence of conserved eukaryotic sites with one or more archaeal genera. The numbers represent the corresponding positions in *E. coli* rRNA and the asterisks (*) indicates that the sites are also methylated in *E. coli* rRNA. It is possible to map unambiguously the positions of archaeal and eukaryotic sites on the *S. acidocaldarius* rRNA sequences because almost all methylations occur in common core regions that are easily aligned. The complete structure of the SSU and LSU RNA are available on the web at rna.wustl.edu/snoRNAdb/.

are clustered in the 530 loop of 16S rRNA, where correct folding and structural stabilization are expected to be critical, only two are conserved between more than one genus. At this time, we cannot make any strong conclusions because we do not know the true positions of all or most rRNA modifications for any species of Archaea. Indeed, additional conserved sites may exist either because our identification of sRNAs is not complete or because these sites may be present but are modified in a guide-RNA-independent manner.

Evolutionary divergence of archaeal methylation guide sRNAs

The rate at which guide sequences diverge will depend on the selective advantage of the sRNA–rRNA interaction during ribosome biogenesis; high selection will minimize the probability that disruptive nucleotide substitutions become fixed in the population and visa versa. The general lack of conservation of predicted positions of methylation within rRNA between archaeal genera indicates that selection for methylation at most sites is

relatively weak. Because selection is weak, nucleotide substitutions accumulate within guides and gradually erode their ability to base pair with the more conserved rRNA target sequence. Conceptually, this process could be divided into stages that may be exemplified in several of the *Pyrococcus* sRNA families (Fig. 2). At an early stage, nucleotide substitutions may create G/U base pairs within the complementary region and mismatches that extend or shorten the length of the complementarity. In a following stage, further nucleotide substitution abolishes complementarity; the guide becomes non-specific without apparent function and is now free to explore sequence space. In a final stage, a favourable complementarity to a new region in rRNA is eventually achieved, and a new site for methylation is generated. The D′-associated guide within *Pyrococcus* sR51may be an example of this final stage (Fig. 5D); *Pfu* and *Pab* members guide methylation to G1059 and A1131 in 16S rRNA respectively. Surprisingly, there are very few guides unassigned to rRNA targets in the collection of identified *Pyrococcus* sRNAs, whereas nearly half of the cloned *S. acidocaldarius* sRNA have no clear target. This may be a function of how the sRNAs were identified: the *Pyrococcus* sRNAs were all identified computationally, based on at least one strong guide sequence; the *S. acidocaldarius* sRNAs were all identified biochemically, thus are not biased for targeting known RNAs.

We also compared the sRNAs from different archaeal genera that guide methylation to homologous positions within rRNA. For example, the D box guide of *Sac* sR7, *Ape* sR19 and *Pho/Pfu/Pab* sR3 all target methylation to position U2552 (*E. coli* numbering) that has been implicated in an interaction with the 3′ end of the A site tRNA (Moazed and Noller, 1989; Kim and Green, 1999). It is interesting that within Archaea this important region of 23S rRNA contains four separate nucleotide substitutions that are matched by compensatory substitutions in the guide regions of the sRNA. It is not possible to say whether the rRNA targets and the sRNA guides have co-evolved from a common ancestor or whether the guide sequences have independently converged on a common rRNA target nucleotide. The D′ box guides of these sRNAs show little if any sequence similarity.

To summarize, the guide regions of most archaeal sRNA genes appear to be dynamic and in constant flux. Although there may be exceptions, selection – particularly at high temperature – appears to maintain the number and general distribution of guide sRNA–rRNA interaction/methylation throughout the conserved core regions of SSU and LSU rRNA rather than the precise positions of methylation. As a result, few positions of guide-directed methylation are conserved, even across short phylogenetic distances (i.e. between genera). Furthermore, homologous sRNA genes derived from a common ancestor are difficult to identify with any degree of certainty (except within closely related species). A challenge that remains is to distinguish between the chaperone and methylation functions of these amazing sRNAs and demonstrate their respective roles in ribosome biogenesis and ribosome structure function.

Acknowledgements

This work was supported by a grant from the Medical Research Council of Canada to PDD.

References

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.

Bachellerie, J.-P., and Cavaille, J. (1998) Small nucleolar RNAs guide ribose methylations of Eukaryotic rRNAs. In *Modification and Editing of RNA.* Grosjean, H., and Benne, R. (eds). Washington DC: American Society for Microbiology, pp. 255–272.

Bachellerie, J.-P., Michot, B., Nicoloso, M., Balakin, A., Ni, J., and Fournier, M.J. (1995) Antisense snoRNAs: a family of nucleolar RNAs with long complementarities to rRNA. *Trends Biochem Sci* **20**: 261–264.

Bachellerie, J.-P., Cavaille, J., and Qu, L.-H. (2000) Nucleotide modifications in eukaryotic rRNAs: the world of small nucleolar RNA guides revisited. In *The Ribosome: Structure, Function, Antibiotics, and Cellular Interactions.* Garrett, R.A., Douthwaite, S.R., Liljas, A., Matheson, A.T., Moore, P.B., and Noller, H.F. (eds). Washington DC: American Society for Microbiology Press, pp. 191–203.

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.

Davis, D.R. (1998) Biophysical and conformational properties of modified nucleotides in RNA (nuclear magnetic resonance studies). In *Modification and Editing of RNA.* Grosjean, H., and Benne, R. (eds). Washington DC: American Society for Microbiology, pp. 85–102.

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

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.

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 56p and 58p interact with 1p and are required for ribosome biogenesis. *Mol Cell Biol* **12**: 7088–7098.

Gupta, R. (1984) Halobacterium volcanii tRNAs. Identification of 41 tRNAs covering all amino acids, and the sequences of 33 class I tRNAs. *J Biol Chem* **259**: 9461–9471.

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.

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

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.

Lafontaine, D.L.J., and Tollervey, D. (1998) Regulatory aspects of rRNA modifications and pre-rRNA processing. In *Modification and Editing of RNA.* Grosjean, H., and Benne, R. (eds). Washington DC: American Society for Microbiology, pp. 281–288.

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

Maden, B.E.H. (1990) The numerous modified nucleotides in eukaryotic ribosomal RNA. *Prog Nucleic Acid Res Mol Biol* **39**: 241–301.

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.

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

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.

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

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.

Ofengand, J., and Fournier, M.J. (1998) The pseudouridine residues in rRNA: number, location, biosynthesis, and function. In *Modification and Editing of RNA.* Grosjean, H., and Benne, R. (eds). Washington DC: American Society for Microbiology, pp. 229–253.

Olsen, G.J., and Woese, C.R. (1997) Archaeal genomes: an overview. *Cell* **89**: 991–994.

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.

Pelczar, P., and Filipowicz, W. (1998) The host gene for intronic U17 small nucleolar RNAs in mammals has no protein-coding potential. *Mol Cell Biol* **18**: 4509–4518.

Qu, L.H., Henras, A., Lu, Y.J., Zhou, H., Zhou, W.X., Zhu, Y.Q., *et al.* (1999) Seven novel methylation guide small nucleolar RNAs are processed from a common polycistronic transcript by Rat1p and RNase III in yeast. *Mol Cell Biol* **19**: 1144–1158.

Samaha, R.R., Green, R., and Noller, H. (1995) A base pair between tRNA and 23S rRNA in the peptidyl transferase centre of the ribosome. *Nature* **377**: 309–314.

Smith, C.M., and Steitz, J.A. (1998) Classification of gas5 as a multi-small-nucleolar-RNA (snoRNA) host gene and a member of the 5′-terminal oligopyrimidine gene family reveals common features of snoRNA host genes. *Mol Cell Biol* **18**: 6897–6909.

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

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.

Tycowski, K.T., Smith, C.M., Shu, M.D., and Steitz, J.A. (1996a) A small nucleolar RNA requirement for site-specific ribose methylation of rRNA in Xenopus. *Proc Natl Acad Sci USA* **93**: 14480–14485.

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

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.

Watkins, N.J., Gottschalk, A., Neubauer, G., Kastner, B., Fabrizio, P., Mann, M., and Luhrmann, R. (1998) Cbf5p, a potential pseudouridine synthase, and Nhp2p, a putative RNA-binding protein, are present together with Gar1p in all H BOX/ACA-motif snoRNPs and constitute a common bipartite structure. *RNA* **4**: 1549–1568.

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