The Box C/D RNPs: Evolutionarily Ancient Nucleotide Modification Complexes

Keith Gagnon, Xinxin Zhang, and E. Stuart Maxwell*

Department of Molecular and Structural Biochemistry, North Carolina State University,

Raleigh, North Carolina 27695

*Corresponding author:

E. Stuart Maxwell
Department of Molecular and Structural Biochemistry
Box 7622
North Carolina State University
Raleigh, North Carolina 27695

Tel: (919) 515-5803
Email: stu_maxwell@ncsu.edu
Fax: (919) 515-2047

The 1990's witnessed an explosion of small RNAs found localized to the eukaryotic nucleolus. Included in these small non-coding RNAs were the box C/D small nucleolar RNAs (snoRNAs) which base pair with ribosomal RNA to direct site-specific, nucleotide 2'-O-methylation. Continuing investigations rapidly expanded the box C/D snoRNA population, identified new functions, defined the limited set of core proteins which bind the box C/D RNAs to assemble ribonucleoprotein (RNP) complexes, and even discovered box C/D RNAs and core proteins in Archaea. *In vitro* ribonucleoprotein assembly systems along with emerging crystal structures are now providing insight into the molecular details of both box C/D RNP structure and methylation function. This chapter reviews recent advances in box C/D RNA and RNP biology, concluding with a discussion of future directions and the experimental challenges that face investigators in this exciting and rapidly advancing field of RNA-guided nucleotide modification.

I. Introduction

Small RNAs were first identified in eukaryotic nuclei in the 1960's when a handful of abundant species were phenol-extracted from isolated nuclei and characterized on sucrose gradients and polyacrylamide gels (reviewed in (1)). Subsequent nucleotide analysis revealed a base composition distinct from GC-rich ribosomal RNA leading to their designation as U-rich small nuclear RNAs or the U snRNAs (2,3). It quickly became apparent that the U snRNAs are partitioned in the nucleoplasmic and nucleolar compartments (4-8) and sequencing in the 1970s defined U3 as the first nucleolar box C/D RNA (9-12). The box C/D snoRNA population remained small until the early 1990's with the discovery of numerous intronic snoRNAs encoded within pre-messenger RNA introns (reviewed in (13)). Discovery of box C/D RNAs in archaeal organisms at the end of the decade indicated an evolutionarily ancient origin (14-16). The populations of eukaryotic and archaeal box C/D RNAs has continued to grow and now numbers in the hundreds.

Nucleolar localization of the box C/D snoRNAs implied a role in ribosome biogenesis and experiments in *Saccharomyces cerevisiae* as well as *Xenopus laevis* demonstrated the essentiality of the major U3 box C/D snoRNA for ribosomal RNA processing (7,17-19). Several additional snoRNAs were identified as crucial for pre-rRNA processing, although most tested snoRNA species revealed no growth phenotype when deleted in yeast (20-22). Critical to understanding the primary function of box C/D snoRNAs in nucleotide modification was the observation that they possessed regions of complementarity with rRNA and that these complementary regions corresponded to sites of nucleotide 2'-O-methylation (23). The list of target RNA modified by the box C/D RNAs has now expanded to include the eukaryotic splicing snRNAs, some pre-messenger RNAs, and archaeal tRNAs (15,24-26).

In the last decade, investigations have focused upon identifying box C/D RNA-binding proteins. Highly homologous eukaryotic and archaeal core proteins have been defined and the

identification of eukaryotic accessory proteins suggests that snoRNP assembly may be a complex and regulated process (27-33). *In vitro* assembly systems for the archaeal box C/D sRNP complex have been established and used to define box C/D RNP assembly pathways as well as reveal RNP structure and examine the mechanisms of box C/D RNA-guided nucleotide methylation function (34-36). These studies, coupled with emerging crystal structures of the core proteins and RNA:protein complexes, are now leading to a more detailed understanding of this RNA:protein enzyme and its function as an RNA-guided nucleotide modification complex.

II. Diversity of Box C/D RNA Populations

a. Box C/D RNA Nomenclature

Box C/D RNAs are abundant in both Eukarya and Archaea, numbering in the hundreds. Following the original nomenclature, most animal and plant homologs are designated U(n) with the species number reflecting the approximate order of discovery. Homologs identified in other organisms usually adopt the same designation. However, there are organisms which use a non-conventional nomenclature. Yeast snoRNAs are designated by a number preceded by the prefix snR(n) (37). In Archaea, the snoRNA-like small RNAs or sRNAs are numbered and preceded by sR(n) (14). Other examples of non-standard nomenclature include the protozoa *Trypanosoma brucei* (TBxCsyCz, x = chromosome number, y = cluster number, z = RNA number) and *Euglena gracilis* (Eg-mx, x = RNA number) (38,39). Different naming schemes have also been employed for tissue-specific or disease related box C/D snoRNAs , such as the mouse brain-specific box C/D snoRNA II-52 (MBII-52) or the box C/D snoRNA linked to neonatal lethality in Prader-Willi syndrome (Pwcr1/MBII-85) (40,41).

b. Box C/D RNA Structure

Box C/D RNAs are defined by their highly conserved nucleotide box C (RUGAUGA) and box D (CUGA) sequences positioned near the 5' and 3' termini, respectively (Figure 1), and frequently possess internal C' and D' boxes (42-44). The C' and D' nucleotide boxes are typically present and well conserved in the archaeal sRNAs but often degenerate and difficult to identify, or even missing, in the eukaryotic snoRNAs (14-16,45). Individual box C/D RNAs are defined by their unique guide sequences which are located immediately adjacent and upstream of boxes D and D'. These guide sequences base pair to the different target RNAs and determine the site of box C/D RNA action on the target RNA (44,46-48). Archaeal box C/D sRNAs are typically smaller (50-70 nucleotides) than the eukaryotic box C/D snoRNAs (75 to 150 nucleotides) (15,16,37,49-51). The highly conserved archaeal guide regions of 12 nucleotides most often constitute the entire spacer regions between the box C/D and C'/D' motifs and account for the smaller size of the archaeal sRNAs (52). The well defined box C/D and C'/D' motifs fold to establish kink turn or "K-turn" and "K-loop" elements, respectively. The K-turn consists of canonical stem I and internal stem II separated by an asymmetric bulge which possesses tandem sheared G-A pairs essential for K-turn stability (30,53,54). The internal K-loop lacks stem I and is typically replaced with a small loop. Ultimately, both the K-turn and K-loop serve as core protein binding sites to assemble the box C/D and C'/D' RNP complexes (34-36). The recent discovery of circularized box C/D sRNAs in *Pyrococcus furiosus* suggests a novel processing pathway that may augment box C/D sRNA stability in hyperthermophiles (55).

c. Diversity of Box C/D RNA Populations

Appreciation of box C/D RNA diversity in eukaryotic and archaeal organisms has come

with the characterization of RNA populations in a variety of organisms in both kingdoms. In eukaryotes, approximately two thirds of the yeast snoRNAs involved in rRNA maturation are conserved and found in plants and humans (56). The yeast box C/D snoRNA population has been well defined and currently numbers 46 RNAs (57). This compares with more than double that number of human box C/D RNAs involved in rRNA maturation, but does not include additional RNAs modifying other cellular target RNAs (58). Thus, the box C/D snoRNA populations of metazoan organisms are characteristically more complex with many of these additional RNAs unique to specific organisms. Box C/D homologs in different organisms may possess unique guide sequences reflecting organism-specific rRNA sequences yet modify corresponding nucleotides in the respective ribosomal RNAs. However, many box C/D snoRNAs modify rRNA nucleotides or nucleotides of other target RNAs that are unique to that organism (38). Thus, different eukaryotes typically possess a unique set of snoRNAs characteristic of that organism. While the majority of box C/D snoRNAs are involved in nucleotide methylation of rRNA nucleotides, conserved box C/D snoRNAs such as U3, U8, and U14 function in animals, plants, and fungi in non-modification roles such as facilitating pre-rRNA folding or cleavage during precursor processing (13). To date, there appears to be no shared box C/D RNA homologs between Archaea and Eukarya. Similarly, box C/D sRNA populations among Crenarchaeota and Euryarchaeota are typically distinct with respect to the specific organism examined (15,16,59).

d. Box C/D RNA Identification

Hundreds of box C/D RNAs have been defined in numerous eukaryotic and archaeal organisms. Techniques to identify and define box C/D RNAs include both biochemical and

computational approaches. Biochemical approaches include isolation of small RNAs fractionated on gradients or immunoprecipitated with antibodies against box C/D RNP proteins, followed by cDNA cloning and sequencing (60-62). Immunoprecipitation of box C/D RNPs recently identified 66 previously unknown box C/D snoRNAs in the protist *Euglena gracilis* (39). Computational approaches have utilized a variety of search engines and algorithms to analyze available databases using box C and D consensus sequences (Table 1), as well as appropriately positioned guide sequences that exhibit complementarity to cellular target RNAs (63,64). Other sequence analysis-based searches have used snoRNA or intronic sequence databases to search for homologs in particular organisms (50,65). Computational approaches can identify potential box C/D RNAs that are expressed at low levels or in a tissue-specific manner whose detection by biochemical approaches could prove difficult. The utility of computational approaches was recently demonstrated by the identification of 62 novel box C/D RNAs from the trypanosome *Leishmania major* and 50 novel snoRNAs in *Caenorhabditis elegans* (66,67). Both candidate and experimentally confirmed RNAs are found in an array of databases which report ever increasing populations of archaeal and eukaryotic box C/D RNAs (Table 1). Both biochemical approaches have identified "orphan" box C/D RNAs where the associated guide sequence is not complementary to any know cellular RNA sequence thus raising the question as to not only their potential target RNAs but also their possible function(s).

III. Box C/D RNA Functions and Target RNAs

a. Folding and Cleavage of Pre-rRNA

The base pairing of box C/D RNA guide sequences to their respective target RNAs

determines the site of RNA function. The first identified eukaryotic snoRNAs were localized in the nucleolus, thus implying a role in pre-rRNA processing and/or ribosome biogenesis (7,18,68,69). Early experiments indeed demonstrated the importance of U3 (snR17), U8, U14, snR10, and snR30 box C/D snoRNAs for pre-rRNA cleavage (20,22,70-73). Those snoRNAs required for rRNA processing are typically essential RNAs, whereas those snoRNAs guiding nucleotide modification are not. The box C/D snoRNAs required for pre-rRNA cleavage at specific sites in yeast is based upon snoRNA gene disruption experiments. Similar observations have been made for *Xenopus laevis* rRNA processing when these same essential snoRNA species are depleted in oligonucleotide knockout experiments (13,17,74,75).

Some box C/D snoRNAs such as U3, U8, and U14 have been shown to be important for pre-rRNA folding and cleavage. These snoRNAs are considered to have "chaperone" function as their base pairing to target rRNAs facilitates correct pre-RNA folding required for precursor maturation. U3 hydrogen-bonds with the 5' ETS of the rRNA precursor and U14 base pairs with 18S rRNA (76-78). Both snoRNAs are essential for pre-RNA cleavage and production of mature 18S rRNA. U8 is required for proper folding of 5.8S and 28S rRNAs within the precursor transcript and is essential for 28S rRNA production (74,79,80). Interestingly, U14 not only guides 18S rRNA processing using its C'/D' guide sequence, but also guides nucleotide methylation using its box C/D guide sequence (75). Recent work utilizing computer simulations has suggested that snoRNAs may play important roles in long range rRNA folding (81). To date, eukaryotic box C/D RNAs have not been implicated in chaperone folding or precursor processing functions for any RNA other that pre-ribosomal RNA.

b. 2'-O-Methylation of Diverse RNA Targets

The primary role of most box C/D RNAs is to guide the modification of a targeted nucleotide's ribose sugar, catalyzing addition of a methyl group at the 2' hydroxyl position (Figure 1). The nucleotide targeted for methylation is positioned within the RNA duplex formed by the hydrogen bonded target:guide RNA sequences and base paired to the fifth nucleotide upstream from the snoRNA's D or D' box (44,46,47). The number of nucleotides modified by the box C/D RNAs in eukaryotes and archaea numbers in the hundreds and thus explains the large box C/D populations found in eukaryotic and archaeal organisms. Approximately 100 ribose methylations are found in human rRNA, about 120 in plant *Arabidopsis thaliana* rRNA, and 55 in yeast rRNA (82-84). The occurrence of methylated nucleotides in archaeal rRNA is also considerable with 67 identified in *Sulfolobus solfataricus* and at least 26 and 93 predicted in *Sulfolobus acidocaldarius* and *Pyrococcus abyssi*, respectively (15,16,85). The specific rRNA nucleotides methylated vary from organism to organism, but these modifications tend to be clustered in conserved core and functional regions of the rRNA (84). Disruption of one or a few modifications at specific nucleotides often does not result in an obvious phenotype (84), although complete disruption of nucleotide methylation in yeast has severe effects upon rRNA maturation and ribosome function (86).

The types of RNA targeted for methylation by box C/D RNAs extends beyond the ribosomal RNAs to include archaeal tRNAs and the eukaryotic splicing snRNAs. Many of the modifications in archaeal rRNA and tRNAs are predicted by newly defined box C/D sRNAs and their associated D and D' guide sequences (15). Also found in Archaea and Eukarya are "orphan" RNAs where the guide sequence(s) of an identified C/D RNA is not complementary to

any sequence in a known cellular RNA (48,64). In eukaryotes, specific box C/D RNAs guide nucleotide methylation of U1, U2, U4, U5, U6 and U12 snRNAs (24,87,88) (see Chapter 9 by Karijolich et al.). Specific members of the box C/D RNA family are also found localized to the nuclear Cajal bodies and designated small Cajal RNAs or scaRNAs. These scaRNAs are unique in that they possess both box C/D and box H/ACA RNA motifs and direct the 2'-O-methylation and pseudouridylation of the RNA Pol II-transcribed snRNAs U1, U2, U4, and U5 within the Cajal bodies (24).

c. Additional Roles and Targets for Box C/D RNAs

Investigation of specific tissues has suggested a larger population of box C/D RNAs than originally anticipated. Analysis of small RNAs in mouse, rat and human brain has identified many brain-specific box C/D snoRNAs (40,89,90). Only the mouse/human brain-specific box C/D snoRNA MBII-52/HBII-52 has been mapped to a putative target. This box C/D snoRNA appears to target an A-to-I editing site in the serotonin receptor 2C (5-HT_{2C}) pre-mRNA. Methylation of this site potentially modulates pre-mRNA editing and alternative splicing, thereby resulting in changes to serotonin signaling capacity (25). Alternatively, it has been suggested that the HBII-52 snoRNA uses its guide region sequences to affect alternative splicing of the 5-HT_{2C} pre-mRNA by masking splicing silencers which then leads to the inclusion of an exon in the final spliced serotonin receptor mRNA (26). The loss of MBII-85/HBII-85 is implicated in Prader-Willi syndrome (91). Furthermore, regulation of MBII-52 and MBII-48 snoRNA levels has been noted in the early phase of memory consolidation during learning, thus suggesting a possible role for this snoRNA in higher brain function (92).

IV. Box C/D RNP Structure and Nucleotide Methylation Function

a. Eukaryotic Box C/D Core Proteins and snoRNP Structure

Box C/D snoRNAs bind a small set of core proteins to establish ribonucleoprotein (RNP) complexes (Figure 2). Terminal box C/D and internal C'/D' snoRNA motifs serve as protein binding sites for RNP assembly and it is the bound core proteins that carry out the nucleotide 2'-O-methytransferase reaction. Four well conserved core proteins bind eukaryotic box C/D snoRNAs. They are the 15.5kD protein (Snu13p in yeast), nucleolar proteins 56 (Nop56) and 58 (Nop58), and the methyltransferase fibrillarin. The 15.5kD protein directly binds the box C/D K-turn, but not box C'/D' motifs, initiating assembly of box C/D snoRNPs (30,93-95). The use of nucleotide analog interference mapping (NAIM) has suggested that 15.5kD's inability to bind C'/D' motifs may lie in slight structural differences that arise when a K-turn possesses a terminal loop instead of a stem II structure (96).

The Nop56 and Nop58 core proteins are highly homologous and essential for nucleotide modification and ribosome biogenesis (27,97). Cross-linking experiments have shown that Nop58 interacts with box C whereas Nop56 interacts with box C' (94). This suggests that Nop56 and Nop58 core proteins bind the K-turn or K-loops, respectively, although their asymmetric distribution in the snoRNP complex has not yet been confirmed in RNP assembly experiments. Binding of both Nop56 and Nop58 requires stem II of the box C/D motif (98). These core proteins are likely to be RNP structural proteins, although their role in the methylation reaction or its regulation cannot be ruled out at this time.

The methyltransferase fibrillarin interacts with both Nop56 and Nop58 and can be cross-linked to both the box C/D and C'/D' motifs (27,94). Evidence establishing fibrillarin as the methyltransferase enzyme includes the presence of an S-adenosylmethionine (AdoMet)-dependent methyltransferase-like domain that is required for methylation activity (99,100) and the observation that disruption of the yeast fibrillarin gene results in the loss of pre-rRNA methylation (86).

b. Archaeal Box C/D Core Proteins and In Vitro sRNP Assembly

Archaeal box C/D sRNAs bind three core proteins which are highly homologous to the eukaryotic core proteins (29,34,35). Ribosomal protein L7, homologous to eukaryotic 15.5kD, recognizes both box C/D and C'/D' motifs to initiate box C/D and C'/D' RNP assembly. A single Nop56/58 core protein, highly homologous to the eukaryotic Nop56 and Nop58 pair, also binds both box C/D and C'/D' motifs. The third core protein fibrillarin binds both complexes to complete box C/D and C'/D' RNP assembly. Much of what is presently known about box C/D RNP structure and assembly comes from examination of *in vitro* assembly of the archaeal sRNP using *in vitro* synthesized box C/D sRNAs and recombinant sRNP core proteins. These *in vitro* assembly studies revealed an order of assembly with L7 initiating RNP formation followed by Nop56/58 and then fibrillarin binding (34-36). Binding of the L7 core protein induces structural changes in both box C/D and C'/D' motifs (101-103). These structural changes are required for subsequent Nop56/58, but not fibrillarin, binding. Nop56/58 and fibrillarin form a stable dimer in the absence of the sRNA and/or L7 and evidence suggests that it is the dimer that binds the assembling box C/D and C'/D' RNP complexes (35,103-105). In contrast to the "asymmetric" eukaryotic snoRNP complex, the binding of all three core proteins to the box C/D and C'/D'

motifs assembles a “symmetric” sRNP (Figure 2).

c. Emerging Core Protein and RNP Crystal Structures

Emerging crystal structures of core proteins, the K-turn, and sRNA:core protein sub-complexes are beginning to reveal the detailed architecture of the box C/D RNP. The ease of expressing soluble, recombinant archaeal core proteins has greatly facilitated these studies and all three archaeal core proteins have yielded crystal structures at detailed resolution. Crystal structures of free L7 core protein, the K-turn, and the L7:K-turn RNP have shown that L7 binding induces a conformational change or “remodels” the sRNA (54,95,101,102,106). This induced fit has been confirmed in fluorescence resonance energy transfer (FRET) experiments (102). Both the L7 and 15.5kD core proteins lock the K-turn into a tightly kinked structure characterized by stacked stem I and stem II helices bent at approximately 60°, hence the designation of kink-turn for this RNA element (53,107). Particularly important for binding is the requirement of a pyrimidine nucleotide from box C (**U**GAUGA) which is extended and inserted into a cleft of the L7 protein. A co-crystal structure of *Archaeoglobus fulgidus* Nop56/58 and fibrillarin has provided the molecular details of this core protein dimer (104). S-adenosylmethionine is bound to the methyltransferase domain of fibrillarin and biophysical studies have indicated that Nop56/58 binding to fibrillarin helps stabilize Ado-Met in its binding pocket (108). A recent crystal structure of the Nop56/58 homolog hPrp31 bound to the 15.5kD:U4 K-turn RNP has revealed that Nop56/58 and hPrp31 bind the K-turn-L7/15.5kD protein complex through their Nop domains and require interaction with both the K-turn and the L7/15.5kD protein (109). Despite crystallization of all the individual components of the complex, a crystal structure of the fully assembled box C/D sRNP has thus far proved elusive.

d. Investigating Methylation Function Using In Vitro Assembled Archaeal Box C/D sRNP

The archaeal *in vitro* assembly system has enabled investigations of box C/D sRNP methylation function with respect to sRNP structure and the role of the sRNA and individual core proteins. The assembled box C/D sRNP is catalytically active and guides site-specific methylation of target RNAs using both D and D' guide sequences. Target RNA methylation is nucleotide-specific and dependent upon fibrillarin binding and the presence of S-adenosylmethionine (34). Further studies have shown that efficient methylation of D and D' guide RNAs requires both box C/D and C'/D' RNPs be juxtaposed in a fully assembled sRNP (35). The juxtaposed box C/D and C'/D' RNPs are dependent upon the highly conserved spatial positioning of each RNP separated by 12 nucleotides (52). This suggests molecular interactions between the two complexes and/or induced sRNA remodeling is important for methylation function. This is not likely the case for the eukaryotic box C/D snoRNPs that guide the methylation using both D and D' guide sequences where the spatial positioning of box C/D and C'/D' RNPs is not conserved and is often quite distant. Finally, both *in vivo* and *in vitro* studies have examined the base pairing interactions of the target RNA with the sRNA guide sequence (110,111). A minimum RNA:RNA duplex of 9-11 nucleotides is required and Watson-Crick pairing is essential. Interestingly, the ability of the *in vitro* complex to methylate target nucleotides positioned within thermally stable double-stranded secondary structures may suggest an ability of the core complex to facilitate target RNA melting necessary for base pairing with the sRNA guide sequence.

V. Box C/D RNP Biogenesis

a. Genomic Organization of Eukaryotic Box C/D snoRNA Genes

The genomic organization of eukaryotic box C/D RNAs falls into two major categories. Some box C/D snoRNA genes are independently transcribed from snoRNA-specific promoters with the primary transcript possessing a single or multiple snoRNA species. However, the majority of box C/D snoRNAs are encoded within introns of host genes and transcribed by RNA polymerase II as part of the host precursor transcript. Fungi, plants, trypanosomes and unicellular organisms possess primarily independently-transcribed snoRNA genes, although snoRNAs encoded within host gene introns are present. In metazoan organisms, snoRNA coding sequences are overwhelmingly positioned within host gene introns (reviewed in (13) and (48)). For both snoRNA categories, the primary transcripts undergo post-transcriptional processing to produce mature snoRNA species.

b. Independently-Transcribed and Intronic Eukaryotic Box C/D snoRNA Genes

Box C/D snoRNA genes that are independently-transcribed using a snoRNA-specific promoter may be found as single or clustered species. U3 is the best characterized box C/D snoRNA that is independently-transcribed as a single snoRNA. U3 genes have been identified and characterized in fungi, protists, vertebrates, insects and plants (reviewed in (13)). They are most often found in multiple copies and dispersed throughout the genome, although occurrence of pseudogenes in vertebrates can make determination of gene copy number difficult. (112-114). Box C/D snoRNA genes in yeast, plants, and trypanosomes are often organized into clusters and are transcribed as a polycistronic precursor using an snoRNA-specific promoter (83,115-118). These polycistronic transcripts are subsequently processed to produce the mature individual box

C/D snoRNAs.

The predominant box C/D snoRNA gene organization in metazoans, and particularly vertebrates, is that of encoding within introns of host RNA Pol II transcripts. Typically, the intronic snoRNAs are located in protein-coding pre-mRNA introns. Intronic box C/D snoRNAs are frequently found in multiple introns of common host genes with each intron limited to a single encoded snoRNA. Intronic box C/D snoRNAs are often found in the same host gene in different organisms but this organization is not universal, and frequently an snoRNA species is found in the same host gene but in different introns (reviewed in (13)). Host genes often possess isomers of the same intronic snoRNA species (119-121). Host genes also encode different snoRNA species in different introns ((122,123) and reviewed in (13)). The intronic snoRNAs are characteristically found within host genes that encode proteins involved in ribosome biogenesis or protein synthesis. Ribosomal protein genes very often contain intronic snoRNAs (13,48). The positioning of intronic snoRNAs within protein coding genes important for ribosome biogenesis and protein synthesis suggests a possible coordination of the snoRNA and ribosome biogenesis pathways. Of particular interest are the intronic box C/D snoRNAs encoded with pre-mRNAs that do not encode a protein such as the UHG genes of human and *Drosophila melanogaster* (124-128). In these cases, the host precursor transcript simply serves as a carrier for the encoded intronic snoRNAs.

c. Archaeal Box C/D sRNA Genes

Most of the limited information concerning archaeal box C/D sRNA genes comes from an analysis of several hyperthermophile *Pyrococcus* genomes (16). Box C/D sRNA genes are

primarily intergenic although some do overlap upstream and/or downstream open reading frames on the same DNA strand. They typically exhibit very little clustering, although in *S. solfataricus* and *P. fulgidus* two sRNA genes (sR10/sR11 and sR26/sR60, respectively) are separated by only a few nucleotides and appear to be co-transcribed (15). In *P. abyssi*, the box C/D sRNA genes are preferentially located in non-coding regions of the genome, again with little clustering. Interestingly, the box C/D sR40 of *Pyroccocus* and *A. fulgidus* is found within the intron of the $tRNA^{Trp}$ gene and it is this sRNA that is responsible for 2'-O-methylation of two nucleotides in the $tRNA^{Trp}$ itself (129-131). Nothing is presently known about the RNA polymerase promoters responsible for sRNA transcription as obvious promoter elements are not observed and detailed analysis of possible sRNA precursor transcription has not been examined.

d. Transcription and Processing of Independently Transcribed Box C/D snoRNAs

Independently transcribed box C/D snoRNA genes possess their own promoter and terminator regions as well as enhancer elements flanking the coding sequence. They typically possess promoters similar to protein coding genes and are typically transcribed by RNA Polymerase II (13,132). Transcription of U3 is driven by a Pol II promoter in vertebrates and possesses several conserved sequence elements. These include TATA-like boxes, proximal and distal sequence elements (PSE and DSE), and a "U3 box" specific for U3 at the DSE (113,133-135). Plant U3 genes also possess TATA-like boxes but are transcribed by RNA Pol III due to the shorter spacing between the TATA box and upstream sequence elements (USE) (136,137). The U3 transcripts of vertebrates are capped at the 5' terminus with trimethylguanosine (TMG) whereas those of plants are O-methyl capped (138). Other Pol III transcribed box C/D snoRNA genes include snR52 in yeast (139) and several plant box C/D snoRNAs clustered with tRNA

genes. The plant snoRNA genes are co-transcribed with tRNA as a tRNA-snoRNA precursor using the tRNA gene's RNA Pol III promoter (140). Yeast promoters of box C/D snoRNA genes that are independently transcribed by Pol II often contain A/T rich stretches and a TATA box. Upstream Rap1p or Abf1p binding sites that are typical for yeast ribosomal protein genes are also present sometimes (115). Box C/D snoRNA promoters for independent transcripts in plants are not well characterized but do contain putative TATA boxes (116). Independently-transcribed box C/D snoRNAs are often polycistronic. The nascent transcripts are capped with trimethylguanosine (TMG) but undergo 5' end processing involving endonucleases to produce processing intermediates that are matured by trimming exonucleases. In yeast, the ortholog of bacterial endonuclease RNase III, Rnt1p, and the exonucleases Rat1p and Xrn1p are responsible for snoRNA processing and trimming (115,141-143).

Termination of Pol II-transcribed genes can result in either polyadenylated or non-polyadenylated transcripts. The box C/D snoRNAs are not polyadenylated but current understanding of 3' end formation is complicated by the overlapping machinery involved in these two pathways and shared components which include the core cleavage and polyadenylation factor (CPF) complex and the exosome (144). 3' End formation of independently transcribed box C/D snoRNA precursors involves factors Nrd1p, Nab3p, and Sen1p which appear to be specific for non-poly(A) termination and 3' end pre-snoRNA processing (145). Nrd1p and Nab3p recognize and bind specific sequence elements upstream of the termination signal of box C/D snoRNA transcripts, although their recognition of these sequences alone is not sufficient to prevent polyadenylation and direct proper 3' end formation (146). Nrd1p interacts with the exosome, cap-binding complex, and Pol II and is therefore implicated in coordinating

transcription and 3' end formation with exosome processing (147). 3' End formation is also dependent upon the co-transcriptional assembly of the box C/D snoRNP (148,149). Thus, termination and 3' end formation may rely on a monitoring mechanism whereby Nrd1p, Nab3p, and Sen1p are involved in detecting the assembled snoRNP upstream of the termination signal and directing a bypass of transcript polyadenylation while activating exosome-mediated 3' end processing.

e. Transcription and Processing of Intronic Box C/D snoRNAs

Box C/D snoRNA genes, particularly those of metazoan organisms, are positioned within introns of host genes that may be protein or non-protein coding. Intronic snoRNAs are excised from the intron during host precursor splicing (reviewed in (13)). The intronic snoRNA host genes are primarily driven by promoters containing terminal oligopyrimidine tracts or TOP promoters, which are characteristic of a broad family of protein-coding genes involved in ribosome biogenesis and protein synthesis (128,150-152). Processing of intronic box C/D snoRNAs requires specific positioning of the snoRNA within the host intron, approximately 50 nucleotides upstream of the branchpoint (153,154). Recently, a splicing factor designated intron binding protein 160 (IBP160) has been identified that defines the snoRNA distance from the branchpoint (155). IBP160 binds the intron approximately 35-40 nucleotides upstream of the branch point in a sequence-independent manner. Intronic snoRNA processing also requires snoRNP assembly. Immunoprecipitation experiments with fibrillarin antibodies have shown that the snoRNP complex is assembled while the snoRNA is still part of the unspliced, host pre-mRNA transcript. Accordingly, the box C/D snoRNP was found to assemble in the C1 splicing complex where IBP160 exerts its function (156). Thus, it appears that IBP160 may be a key

factor linking intronic box C/D snoRNP assembly and intronic snoRNA processing with host pre-mRNA splicing events.

Upon pre-mRNA splicing, box C/D snoRNAs are excised from the intron and trimmed to a mature snoRNA using exonucleases. In yeast, an RNase III enzyme Rnt1p debranches the lariat intron. Rnt1p cleavage requires interaction with Nop1p (fibrillarin) of the assembled box C/D snoRNP (157). This cleavage provides entry sites for exonucleolytic trimming by the 5'-3' exonucleases Rat1p and Xrn1p and the 3'-5' trimming activities of Rrp6p and the exosome (158,159). Using a minor pathway, intron-encoded snoRNAs can be directly excised from unspliced pre-mRNA using an endonuclease followed by exonucleolytic trimming (160,161). Box C/D snoRNAs that are processed directly from unspliced pre-mRNAs appear to be released by a mechanism similar to that observed for independently-transcribed snoRNAs (157).

f. Box C/D snoRNP Transport

Box C/D snoRNA transcription, processing, and snoRNP assembly occurs in the nucleoplasm before snoRNP transport into the nucleolus. Transport to the nucleolus requires the box C/D core motif and all four core proteins for nucleolar localization (162-166). Experiments indicate that in higher eukaryotes, box C/D snoRNPs require transit through Cajal bodies for maturation before entering the nucleolus (165). In Hela cells, independently-transcribed U3 snoRNA precursors are found in the nucleoplasm as large multiprotein processing complexes containing factors linked to RNA processing (hRrp46, LSm 2-8, La protein), snoRNP assembly (Tip48, Tip49, Nopp140), and RNA transport (PHAX, CRM1) (167). The snRNA export factor PHAX is required for U3 transport to the Cajal body where it is m^3G-capped and undergoes

CRM1-dependent transport to the nucleolus (168,169). Intronic and uncapped box C/D snoRNAs may use a variation of this nucleolar transport pathway. U14 nucleolar localization is impaired when PHAX is depleted and some intronic and uncapped box C/D snoRNAs can bind PHAX in the absence of the cap binding complex (167,169).

Future Directions and Experimental Challenges

a. Box C/D RNA Diversity, Targets, and Functions

Computational approaches will continue to define box C/D RNA populations in both archaeal and eukaryotic organisms. Improvements in bioinformatics approaches coupled with increasing genome data bases will identify new box C/D RNAs to reveal new populations and how these populations vary between organisms. These analyses may reveal conserved box C/D RNAs implying conserved functions in different organisms. More certain is the identification of new RNAs unique to given organisms. Defining new RNAs and their potential targets may well suggest novel functions. Comparison of archaeal and eukaryotic organism-specific populations should also provide insight into the rate of evolution for this class of small non-coding RNAs and how rapidly box C/D RNA diversity (and function?) is evolving.

Biochemical approaches will be required to identify box C/D RNAs that do not exhibit well conserved box sequence elements. Defining these RNAs will likely require either biochemical fractionation of small non-coding RNA populations or immunoprecipitation of these RNAs using antibodies against the RNP core proteins. The observation of myriad RNAs bound

to the L7 core protein in Archaea may suggest a greater diversity of C/D box RNAs than originally anticipated (170). Non-conventional box C/D RNAs may be associated with novel box C/D RNA-binding proteins implying novel RNP structures or perhaps functions. Indeed, it is a formal possibility that the protein(s) of these RNPs is the functional component of the complex working independently of the RNA's guide sequence. Certainly this is the case for the archaeal H/ACA core proteins that function as a pseudouridylate synthase for a tRNA nucleotide in the absence of bound H/ACA RNA (171). Evidence does not indicate that the box C/D core proteins can methylate nucleotides in the absence of the RNA but what other "moonlighting" functions might the box C/D core or accessory proteins carry out? Certainly this is the case for the L7/15.5kD core protein where it functions in Archaea as a ribosomal protein and in eukaryotes as a U4 snRNP protein (29,30). Are there other functions for the core proteins?

Both computational and biochemical approaches are certain to identify not only new box C/D RNAs but also new orphan RNAs with unknown targets. The occurrence of orphan box C/D RNAs raises the formal possibility that the function of these orphan RNAs is not determined by their guide sequence. Thus, novel functions could be expected for these box C/D RNAs and RNPs. A particular challenge will be defining tissue-specific and development-specific box C/D RNAs. It is these unique and perhaps rare box C/D RNAs that are most likely to exhibit novel functions as do the human/mouse brain-specific box C/D RNAs. Identification of tissue-specific RNAs will require biochemical approaches isolating box C/D RNAs specifically expressed in a given tissue or developmental stage and identification through cloning and sequencing approaches. It will be interesting to determine how wide spread is the tissue-specific expression of the box C/D RNAs and ultimately if these same RNAs are specifically expressed in the same

tissue of different eukaryotes or within different tissues of a given eukaryote. These uniquely expressed RNAs may well modify nucleotides that, in turn, regulate other RNA metabolic processes such as observed for the brain-specific box C/D RNAs. Thus, the effect of box C/D RNA-guided modification may extend well beyond methylation of a given nucleotide to simply effect target RNA structure or function.

b. Box C/D RNP Structure and Methylation Function

Archaeal *in vitro* assembly systems have facilitated detailed examinations of box C/D sRNP structure and methylation function. However, important questions concerning this minimal RNA:protein enzyme remain. What are the inter-RNP interactions of juxtaposed box C/D and C'/D' RNPs (protein-protein contacts?) which are required for efficient nucleotide modification and how do they affect RNP structure and, in turn, methylation function? As with eukaryotic snoRNPs, do the archaeal complexes function in more than nucleotide methylation? One would predict that, at the very least, chaperone activity to assist in RNA folding should be anticipated.

Most archaeal box C/D sRNPs guide the methylation of two target nucleotides, sometimes positioned within different target RNAs. What is the mechanism of target substrate interaction with the respective guide sequences? Can a box C/D sRNP bind two target RNAs simultaneously to carry out guided methylation from both complexes or is there interplay between the two RNPs such that only one is catalytically active at any given time? Recent work indicates the importance of the RNP core proteins for target RNA binding (110). What are these target RNA:protein interactions and how do they enhance substrate turnover? Do the core

proteins assist in unfolding highly structured RNAs and how is this accomplished as none of the core proteins appear to possess helicase activity as evidenced by the lack of known helicase domains? What is the topology of the target RNA in an RNA duplex when bound to the guide sequence? A target RNA:guide sequence of 12 nucleotides is equivalent to a full turn of an A form RNA. This would suggest that this RNA duplex is distorted or untwisted to accommodate target RNA binding to the sRNP. What is the function of the known RNA-binding motifs on both fibrillarin and Nop56/58 and do these function in target RNA binding and positioning of the target nucleotide at the catalytic site? Many questions remain as to the function of specific domains of each core protein and their roles in sRNP assembly and methylation function. The structural details obtained from a crystal structure of this sRNP both free and bound with target RNA substrate will help to address some of these questions. More challenging will be enzymatic analyses of the methylation mechanism(s). Assembled sRNPs have thus far exhibited low efficiencies of methylation. Perhaps additional accessory proteins will be required for full activity of *in vitro* assembled sRNPs. Certainly, *in vivo* genetic systems will ultimately be important to confirm *in vitro* results.

The archaeal sRNP is certain to serve as a minimal model complex for examination of the more structurally and functionally diverse eukaryotic snoRNP. The diversity of the eukaryotic complex poses interesting questions as well as specific challenges. How similar is the eukaryotic box C/D snoRNP both structurally and functionally with respect to the archaeal sRNP? Although there are only a few double guide eukaryotic box C/D snoRNP, are they also constrained by spatial positioning of the box C/D and C'/D' RNP for methylation activity? What is the stoichiometry of the snoRNP core proteins on each motif, and specifically, are Nop56 and

Nop58 differentially distributed on the box C/D and C'/D' RNPs as indicated by crosslinking experiments? If so, then how have these core proteins evolved their RNA binding domains to recognize the two very similar but different RNA motifs? A similar question can be asked for the L7/15.5 kD homologs where the archaeal protein recognizes K-turn and K-loop elements whereas the eukaryotic homolog binds only the K-turn. Has this evolution of core protein binding capability allowed more evolutionary drift in the C'/D' motifs that are presently observed, perhaps at the same time promoting greater variation in eukaryotic box C/D snoRNA structural and functional diversity? Because of the diversity in size and folded structures of the eukaryotic box C/D snoRNAs, do individual snoRNPs possess unique proteins distinct for a given RNA species? Are there specific snoRNP proteins (either core or accessory) that are characteristic of the non-methylation functions? These snoRNA-specific proteins might include helicases for chaperone or RNA folding activity as has been demonstrated for the Dbp4 protein of the yeast U14 snoRNP (172).

A particular challenge will be the establishment of *in vitro* RNP assembly systems for the eukaryotic box C/D snoRNP. Such systems can directly address specific structural and functional questions. However, the eukaryotic complex is likely to be difficult to assemble *in vitro* as it appears to require an array of accessory proteins or assembly factors for snoRNP formation. Of course crystal structures of the assembled eukaryotic snoRNP will be valuable but are unlikely in the near future. In contrast to Archaea, genetic systems are now available to study snoRNP structure and function *in vivo*. Both genetic manipulations in yeast as well as transfection approaches in cell culture presently provide tools to address many questions. Of particular interest will be comparison of archaeal and eukaryotic complexes to reveal the

evolution of this RNA:protein enzyme and perhaps suggest how these RNPs acquired new functions.

c. Box C/D RNP Biogenesis

Both intronic and independently-transcribed eukaryotic snoRNA genes have been characterized in a number of eukaryotic organisms for a fundamental understanding of snoRNA transcription. Still lacking however is an understanding of snoRNA transcriptional regulation and how snoRNA biogenesis may be coordinated with other biosynthetic pathways, most specifically ribosome biogenesis. Is snoRNA transcription and processing coordinately regulated with pre-rRNA transcription and/or ribosome assembly? If so, are there snoRNA transcription or processing factors involved in this coordination? Are the clustered snoRNA genes, either intronic snoRNAs in common host genes or polycistronic transcripts, organized for coordinated function? In contrast, transcription of the archaeal snoRNA genes has not been characterized and much remains to be learned. What are the sRNA promoters and which elements are shared with other sRNA genes or are unique to a given sRNA species? Are the sRNA genes coordinately expressed with overlapping genes or independently transcribed and what are the precursor sRNA transcripts, if any?

With respect to box C/D RNP assembly, what are the eukaryotic proteins needed to assemble the complex and are there corresponding assembly factors in Archaea? For the eukaryotic snoRNP, how do these assembly factors interact if at all with the spliceosome and/or the transcription complex for coordinate regulation? Thus far, only IBP60 has been shown to coordinate splicing with snoRNP synthesis (155). What are the accessory proteins and do they

remain bound to the complex to affect RNP structure or function? Is snoRNP assembly completed in the nucleoplasm or are there additional assembly steps in the Cajal bodies or even the cell cytoplasm? Do all snoRNP transit through the Cajal bodies or are there several transport pathways to the nucleolus? Does this transport utilize specific trafficking proteins? Do independently-transcribed and intronic snoRNAs utilize different assembly and trafficking pathways? Finally, how are the snoRNAs and snoRNPs turned over? While it is apparent that box C/D RNAs are relatively stable small non-coding RNAs, nothing is known about their degradation pathway. What are the nucleases involved in snoRNA turnover and are they common to other RNA degradation pathways? Furthermore, is snoRNA/snoRNP turnover coordinated with repression of ribosome synthesis or other metabolic pathways? These and many more questions await further investigation. Results from these investigations are certain to provide new and unexpected insights into the assembly, structure, and function, of this evolutionarily ancient RNA:protein complex.

Acknowledgments

This work was supported by NSF Grant MCB 0543741 to ESM.

Bibliography

1. Busch, H., Reddy, R., Rothblum, L. and Choi, Y.C. (1982) snRNAs, snRNPs, and RNA Processing. *Ann. Rev. Biochem.*, **51**, 617-654.
2. Muramatsu, M., Hodnett, J.L. and Busch, H. (1966) Base composition of fractions of nuclear and nucleolar ribonucleic acid obtained by sedimentation and chromatography. *J. Biol. Chem.*, **241**, 1544-1550.
3. Hodnett, J.L. and Busch, H. (1968) Isolation and characterization of uridylic acid-rich 7 S ribonucleic acid of rat liver nuclei. *J. Biol. Chem.*, **243**, 1368-1375.
4. Nakamura, T., Prestayko, A.W. and Busch, H. (1968) Studies on nucleolar 4 to 6 S ribonucleic acid of Novikoff hepatoma cells. *J. Biol. Chem.*, **243**, 1368-1375.
5. Wienberg, R.A. and Penman, S. (1969) Small molecular weight monodisperse nuclear RNA. *J. Mol. Biol.*, **38**, 289-304.
6. Zapisek, W.F., Saponara, A.G. and Enger, M.D. (1969) Low molecular weight methylated ribonucleic acid species from Chinese hamster ovary cells. I. Isolation and characterization. *Biochemistry*, **8**, 1170-1181.
7. Prestayko, A.W., Tonato, M. and Busch, H. (1970) Low molecular weight RNA associated with 28 s nucleolar RNA. *J. Mol. Biol.*, **47**, 505-515.
8. Prestayko, A.W., Tonato, M., Lewis, B.C. and Busch, H. (1971) Heterogeniety of nucleolar U3 ribonucleic acid of the Novikoff hepatoma. *J. Biol. Chem.*, **246**, 182-187.
9. Reddy, R., Ro-Choi, T.S., Henning, D. and Busch, H. (1974) Primary sequence of U-1 nuclear ribonucleic acid of Novikoff hepatoma ascites cells. *J. Biol. Chem.*, **249**, 6486-6494.
10. Reddy, R., Henning, D. and Busch, H. (1979) Nucleotide sequence of nucleolar U3B RNA. *J. Biol. Chem.*, **253**, 11097-11105.
11. Shibata, H., Reddy, R., Henning, D., Ro-Choi, T.S. and Busch, H. (1974) Low molecular weight nuclear RNA. The 3'-terminal sequence of the U2 RNA. *Mol. Cell. Biochem.*, **4**, 3-19.
12. Shibata, H., Ro-Choi, T.S., Reddy, R., Choi, Y.C., Henning, D. and Busch, H. (1975) The primary nucleotide sequence of nuclear U-2 ribonucleic acid. The 5'-terminal portion of the molecule. *J. Biol. Chem.*, **250**, 3909-3920.
13. Maxwell, E.S. and Fournier, M.J. (1995) The small nucleloar RNAs. *Ann. Rev. Biochem.*, **64**, 897-934.
14. 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.
15. Dennis, P.P., Omer, A. and Lowe, T. (2001) A guided tour: small RNA function in Archaea. *Mol. Microbiol.*, **40**, 509-519.
16. 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.
17. Borovjagin, A.V. and Gerbi, S.A. (1999) U3 small nucleolar RNA is essential for cleavage at sites 1, 2 and 3 in pre-rRNA and determines which rRNA processing pathway is taken in Xenopus oocytes. *J. Mol. Biol.*, **286**, 1347-1363.
18. Zieve, G. and Penman, S. (1976) Small RNA species of the HeLa cell: metabolism and subcellular localization. *Cell*, **8**, 19-31.
19. Savino, R. and Gerbi, S.A. (1990) In vivo disruption of Xenopus U3 snRNA affects

ribosomal RNA processing. *EMBO J.*, **9**, 2299-2308.
20. Tollervey, D. (1987) A yeast small nuclear RNA is required for normal processing of pre-ribosomal RNA. *EMBO J.*, **6**, 4169-4175.
21. Zagorski, J., Tollervey, D. and Fournier, M.J. (1988) Characterization of an SNR gene locus in *Saccharomyces cerevisiae* that specifies both dispensable and essential small nuclear RNAs. *Mol. Cell. Biol.*, **8**, 3282-3290.
22. 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.
23. Bachellerie, J., Michot, B., Nicoloso, M., Balakin, A., Ni, J. and Fournier, M. (1995) Antisense snoRNAs: a family of nucleolar RNAs with long complementarities to rRNA. *Trends in Biochem. Sci.*, **20**, 261-264.
24. Darzacq, X., Jady, B.E., Verheggen, C., Kiss, A.M., Bertrand, E. and Kiss, T. (2002) Cajal body-specific small nuclear RNAs: a novel class of 2'-*O*-methylation and pseudouridylation guide RNAs. *EMBO J.*, **21**, 2746-2756.
25. Vitali, P., Basyuk, E., Le Meur, E., Bertrand, E., Muscatelli, F., Cavaille, J. and Huttenhofer, A. (2005) ADAR2-mediated editing of RNA substrates in the nucleolus is inhibited by C/D small nucleolar RNAs. *J. Cell Biol.*, **169**, 745-753.
26. Kishore, S. and Stamm, S. (2006) The snoRNA HBII-52 regulates alternative splicing of the serotonin receptor 2C. *Science*, **311**, 230-232.
27. 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.
28. Newman, D., Kuhn, J., Shanab, G. and Maxwell, E. (2000) Box C/D snoRNA-associated proteins: Two pairs of evolutionarily ancient proteins and possible links to replication and transcription. *RNA*, **6**, 861-879.
29. Kuhn, J., Tran, E. and Maxwell, E.S. (2002) Archaeal ribosomal protein L7 is a functional homolog of the eukaryotic 15.5kD/Snu13p snoRNP core protein. *Nuc. Acids Res.*, **30**, 931-941.
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 nucleolar box C/D RNPs and the spliceosomal U4 snRNP. *Cell*, **103**, 457-466.
31. Schimmang, T., Tollervey, D., Kern, H., Frank, R. and Hurt, E. (1989) A yeast nucleolar protein related to mammalian fibrillarin is associated with small nucleolar RNAs and is essential for viability. *EMBO J.*, **8**, 4015-4024.
32. Amiri, K. (1994) Firillarin-like proteins occur in the domain Archaea. *J. Bacteriology*, **176**, 2124-2127.
33. King, T., Decatur, W., Bertrand, E., Maxwell, E. and Fournier, M. (2001) A well-connected and conserved nucleoplasmic helicase is required for the production of box C/D and H/ACA snoRNAs and localization of snoRNP proteins. *Mol. Cell. Biol.*, **21**, 7731-7746.
34. Omer, A., 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. USA*, **99**, 5289-5294.
35. 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.
36. 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.*, **33**, 295-306.
37. Riedel, N., Wise, J.A., Swerdlow, H., Mak, A. and Guthrie, C. (1986) Small nuclear RNAs from *Saccharomyces cerevisiae*: Unexpected diversity in abundance, size, and molecular complexity. *Proc. Natl. Acad. Sci. USA*, **83**, 8097-8101.
38. Liang, X.H., Uliel, S., Hury, A., Barth, S., Doniger, T., Unger, R. and Michaeli, S. (2005) A genome-wide analysis of C/D and H/ACA-like small nucleolar RNAs in *Trypanosome brucei* reveals a trypanosome-specific pattern of rRNA modification. *RNA*, **11**, 619-645.
39. Russell, A.G., Schnare, M.N. and Gray, M.W. (2006) A large collection of compact box C/D snoRNAs and their isoforms in *Euglena gracilis*: Structural, functional and evolutionary insights. **357**, 1548-1565.
40. Cavaille, J., Buiting, K., Keifmann, 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. USA*, **97**, 14311-14316.
41. De los Santos, T., Schweizer, J., Rees, C.A. and Francke, U. (2000) Small evolutionarily conserved RNA, resembling C/D box small nucleolar RNA, is transcribed from PWCR1, a novel imprinted gene in Prader-Willi deletion region, which is highly expressed in brain. *Am. J. Hum. Genet.*, **67**, 1067-1082.
42. 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.
43. 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.
44. Kiss-Laszlo, Z., Henry, Y., Bachellerie, M., 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.
45. Kiss-Laszlo, Z., Henry, Y. and Kiss, T. (1998) Sequence and structural elements of methylation guide snoRNAs essential for site-specific ribose methylation of pre-rRNA. *EMBO J.*, **17**, 797-807.
46. Nicoloso, M., Qu, L.H., Michot, B. and Bachellerie, J.P. (1996) Intron-encoded, anti-sense 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.
47. Tollervey, D. (1996) Small nucleolar RNAs guide ribosomal RNA methylation. *Science*, **273**, 1056-1057.
48. Bachellerie, J.P., Cavaille, J. and Huttenhofer, A. (2002) The expanding snoRNA world. *Biochimie*, **84**, 775-790.
49. Brown, J.W., Clark, G.P., Leader, D.J., Simpson, C.G. and Lowe, T. (2001) Multiple snoRNA gene clusters from Arabidopsis. *RNA*, **7**, 1817-1832.
50. Chen, C.L., Liang, D., Zhou, H., Zhuo, M., Chen, Y.Q. and Qu, L.H. (2003) The high diversity of snoRNAs in plants: identification and comparative study of 120 snoRNA genes from *Oryza sativa*. *Nuc. Acids Res.*, **31**, 2601-2613.
51. 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.
52. Tran, E.J., Zhang, X., Lackey, L. and Maxwell, E.S. (2005) Conserved spacing between the box C/D and C'/D' RNPs of the archaeal box C/D sRNP complex is required for efficient 2'-*O*-methylation of target RNAs. *RNA*, **11**, 285-293.
53. 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.
54. Moore, T., Zhang, Y., Fenley, M.O. and Li, H. (2004) Molecular basis of box C/D RNA-protein interactions: cocrystal structure of archaeal L7Ae and a box C/D RNA. *Structure*, **12**, 807-818.
55. Starostina, N.G., Marshburn, S., Johnson, L.S., Eddy, S.R., Terns, R.M. and Terns, M.P. (2004) Circular box C/D RNAs in *Pyrococcus furiosus*. *Proc. Natl. Acad. Sci. USA*, **101**, 14097-14101.
56. Samarsky, D.A. and Fournier, M.J. (1999) A comprehensive database for the small nucleolar RNAs from *Saccharomyces cerevisiae*. *Nuc. Acids Res.*, **27**, 161-164.
57. Piekna-Przybylska, D., Decatur, W.A. and Fournier, M.J. (2007) New bioinformatic tools for analysis of nucleotide modifications in eukaryotic rRNA. *RNA*, **13**, 305-312.
58. Lestrade, L. and Weber, M.J. (2006) snoRNA-LBME-db, a comprehensive database of human H/ACA and C/D box snoRNAs. *Nuc. Acids Res.*, **34**, D158-D162.
59. 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.
60. Huttenhofer, A., Cavaille, J. and Bachellerie, J.P. (2004) Experimental RNomics: a global approach to identifying small nuclear RNAs and their targets in different model organisms. *Methods Mol. Biol.*, **265**, 409-428.
61. Aspegren, A., Hinas, A., Larsson, P., Larsson, A. and Soderbom, F. (2004) Novel non-coding RNAs in *Dictyostelium discoideum* and their expression during development. *Nuc. Acids Res.*, **32**, 4646-4656.
62. Yang, C., Zhou, H., Luo, J. and Qu, L. (2005) Identification of 20 snoRNA-like RNAs from the primitive eukaryote, *Giardia lamblia*. *Biochem. Biophys. Res. Commun.*, **328**, 1224-1231.
63. Schattner, P., Brooks, A.N. and Lowe, T.M. (2005) The tRNAscan-SE, snoscan and snoGPS web servers for the detection of tRNAs and snoRNAs. *Nuc. Acids Res.*, **33**, W686-W689.
64. Yang, J., Zhang, X., Huang, Z., Zhou, H., Huang, M., Zhang, S., Chen, Y. and Qu, L. (2006) snoSeeker: an advanced computational package for screening of guide and orphan snoRNA genes in the human genome. *Nuc. Acids Res.*, **34**, 5112-5123.
65. Fedorov, A., Stombaugh, J., Harr, M.W., Yu, S., Nasalean, L. and Shepelev, V. (2005) Computer identification of snoRNA genes using a Mammalian Orthologous Intron Database. *Nuc. Acids Res.*, **33**, 4578-4583.
66. Liang, X.H., Hury, A., Hoze, E., Uliel, S., Myslyuk, I., Apatoff, A., Unger, R. and Michaeli, S. (2007) Genome-wide analysis of C/D and H/ACA-like small nucleolar RNAs in *Leishmania major* indicates conservation among trypanosomatids in the repertoire and in their rRNA targets. *Eukaryot. Cell*, **6**, 361-377.
67. Huang, Z.P., Chen, C.J., Zhou, H., Li, B.B. and Qu, L.H. (2007) A combined computational and experimental analysis of two families of snoRNA genes from *Caenorhabditis elegans*, revealing the expression and evolution pattern of snoRNAs in

nematodes. *Genomics*, **89**, 490-501.
68. Crouch, R.J., Kanaya, S. and Earl, P.L. (1983) A model for the involvement of the small nucleolar RNA (U3) in processing eukaryotic ribosomal RNA. *Mol. Biol. Rep.*, **9**, 75-78.
69. Bachellerie, J.P., Michot, B. and Raynal, F. (1983) Recognition signals for mouse pre-rRNA processing: A potential role for U3 nucleolar RNA. *Mol. Biol. Rep.*, **9**, 79-86.
70. Kass, S., Tyc, K., Steitz, J.A. and Sollner-Webb, J. (1990) The U3 small nucleolar ribonucleoprotein functions in the first step of preribosomal RNA processing. *Cell*, **60**, 897-908.
71. Hughes, J.M. and Ares Jr., M. (1991) Depletion of U3 small nucleolar RNA inhibits cleavage in the 5' external transcribed spacer of yeast pre-ribosomal RNA and impairs formation of 18S ribosomal RNA. *EMBO J.*, **10**, 4231-4239.
72. Lempicki, R.A., Jarmolowski, A., Huang, G.Y., Li, H.V. and Fournier, M.J. (1990) Mutations in conserved domains of U14 RNA impair 18S ribosomal RNA production in *Saccharomyces cerevisiae*. *Mol. Biol. Rep.*, **14**, 119-120.
73. Li, H.D., Zagorski, J. and Fournier, M.J. (1990) Depletion of U14 small nuclear RNA (snR128) disrupts production of 18S rRNA in *Saccharomyces cerevisiae*. *Mol. Cell. Biol.*, **10**, 1145-1152.
74. Peculis, B. and Steitz, J. (1993) Disruption of U8 nucleolar snRNA inhibits 5.8S and 28S rRNA processing in the Xenopus oocyte. *Cell*, **73**, 1233-1245.
75. Dunbar, D.A. and Baserga, S.J. (1998) The U14 snoRNA is required for 2'-*O*-methylation of the pre-18S rRNA in Xenopus oocytes. *RNA*, **4**, 195-204.
76. Beltrame, M. and Tollervey, D. (1995) Base pairing between U3 and the pre-ribosomal RNA is required for 18S rRNA synthesis. *EMBO J.*, **14**, 4350-4356.
77. Borovjagin, A.V., and Gerbi, S.A. (2005) An evolutionary intra-molecular shift in the preferred U3 snoRNA binding site on pre-ribosomal RNA. *Nuc. Acids Res.*, **33**, 4995-5005.
78. Liang, W.Q. and Fournier, M.J. (1995) U14 base-pairs with 18S rRNA: a novel snoRNA interaction required for rRNA processing. *Genes Dev.*, **9**, 2433-2443.
79. Peculis, B.A. (1997) The sequence of the 5' end of the U8 small nucleolar RNA is critical for 5.8S and 28S rRNA maturation. *Mol. Cell. Biol.*, **17**, 3702-3713.
80. Michot, B., Joseph, N., Mazan, S. and Bachellerie, J.P. (1999) Evolutionarily conserved structural features in the ITS2 of mammalian pre-rRNAs and potential interactions with the snoRNA U8 detected by comparative analysis of new mouse sequences. *Nuc. Acids Res.*, **27**, 2271-2282.
81. Schoemaker, R.J. and Gultyaev, A.P. (2006) Computer simulation of chaperone effects of Archaeal C/D box sRNA binding on rRNA folding. *Nuc. Acids Res.*, **34**, 2015-2026.
82. Maden, B.E. (1990) The numerous modified nucleotides in eukaryotic ribosomal RNA. *Prog. Nucleic Acid Res. Mol. Biol.*, **39**, 241-303.
83. Brown, J.W., Echeverria, M., Qu, L.H., Lowe, T.M., Bachellerie, J.P., Huttenhofer, A., Kastenmayer, J.P., Green, P.J., Shaw, P. and Marshall, D.F. (2003) Plant snoRNA database. *Nuc. Acids Res.*, **31**, 432-435.
84. Decatur, W.A. and Fournier, M.J. (2002) rRNA modifications and ribosome function. *Trends Biochem. Sci.*, **27**, 344-351.
85. 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.

86. 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.
87. Tycowski, K.T., Aab, A. and Steitz, J.A. (2004) Guide RNAs with 5' caps and novel box C/D snoRNA-like domains for modification of snRNAs in metazoan. *Curr. Biol.*, **14**, 1985-1995.
88. Tycowski, K., You, Z., Graham, P. and Steitz, J. (1998) Modification of U6 spliceosomal RNA is guided by other small RNAs. *Mol. Cell*, **2**, 629-638.
89. 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.
90. Cavaille, J., Seitz, H., Paulsen, M., Ferguson-Smith, A.C. and Bachellerie, J.P. (2002) Identification of tandemly-repeated C/D snoRNA genes at the imprinted human 14q32 domain reminiscent of those at the Prader-Willi/Angelman syndrome region. *Hum. Mol. Genet.*, **11**, 1527-1538.
91. Ding, F., Prints, Y., Dhar, M.S., Johnson, D.K., Garnacho-Montero, C., Nicholls, R.D. and Francke, U. (2005) Lack of Pwcr1/MBII-85 snoRNA is critical for neonatal lethality in Prader-Willi syndrome mouse models. *Mamm. Genome*, **16**, 424-431.
92. Rogelj, B., Hartmann, C.E., Yeo, C.H., Hunt, S.P. and Giese, K.P. (2003) Contextual fear conditioning regulates the expression of brain-specific small nucleolar RNAs in hippocampus. *Eur. J. Neurosci.*, **18**, 3089-3096.
93. 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.
94. Cahill, N., Friend, K., Speckman, W., Li, Z., Terns, R., Terns, M. and Steitz, J.A. (2002) Site-specific cross-linking analyses reveal an asymmetric distribution for a box C/D snoRNP. *EMBO J.*, **21**, 3816-3828.
95. Charron, C., Manival, X., Clery, A., Senty-Segault, V., Charpentier, B., Marmier-Gourrier, N., Branlant, C. and Aubry, A. (2004) The archaeal sRNA binding protein L7Ae has a 3D structure very similar to that of its eukaryal counterpart while having a broader RNA-binding specificity. *J. Mol. Biol.*, **342**, 757-773.
96. Szewczak, L.B., Gabrielsen, J.S., DeGregorio, S.J., Strobel, S.A. and Steitz, J.A. (2005) Molecular basis for RNA kink-turn recognition by the h15.5K small RNP protein. *RNA*, **11**, 1407-1419.
97. 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.
98. Watkins, N.J., Dickmanns, A. and Luhrmann, R. (2002) Conserved stem II of the box C/D motif is essential for nucleolar localization and is required, along with the 15.5K protein, for the hierarchical assembly of the box C/D snoRNP. *Mol. Cell. Biol.*, **22**, 8342-8352.
99. Wang, H., Boisvert, D., Kim, K., Kim, R. and Kim, S.H. (2000) Crystal structure of a fibrillarin homologue from Methanococcus jannaschii, a hyperthermophile, at 1.6Å resolution. *EMBO J.*, **19**, 317-323.
100. 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.
101. Suryadi, J., Tran, E.J., Maxwell, E.S. and Brown II, B.A. (2005) The crystal structure of Methanocaldococcus jannaschii multifunctional L7Ae RNA-binding protein reveals an induced-fit interaction with the box C/D RNAs. *Biochemistry*, **44**, 9657-9672.
102. Turner, B., Melcher, S.A., Wilson, T.J., Norman, D.G. and Lilley, D.M.J. (2005) Induced fit of RNA on binding the L7Ae protein to the kink-turn motif. *RNA*, **11**, 1192-1200.
103. Gagnon, K.T., Zhang, X., Agris, P.F. and Maxwell, E.S. (2006) Assembly of the archaeal box C/D sRNP can occur via alternative pathways and requires temperature-facilitated sRNA remodeling. *J. Mol. Biol.*, **362**, 1025-1042.
104. Aittaleb, M., Rashid, R., Chen, Q., Palmer, J., Daniels, C. and Li, H. (2003) Structure and function of Archaeal box C/D sRNP core proteins. *Nat. Struct. Biol.*, **10**, 256-263.
105. Zhang, X., Champion, E.A., Tran, E., Brown II, B.A., Baserga, S.J. and Maxwell, E.S. (2006) The coiled-coil domain of the Nop56/58 core protein is dispensable for sRNP assembly but is critical for archaeal box C/D sRNP-guided nucleotide methylation. *RNA*, **12**, 1092-1103.
106. Hama, T. and Ferre-D'Amare, A.R. (2004) Structure of protein L7Ae bound to a k-turn derived from an archaeal box H/ACA sRNA at 1.8Å resolution. *Structure*, **12**, 893-903.
107. Vidovic, I., Nottrot, 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.
108. Aittaleb, M., Visone, T., Fenley, M.O. and Li, H. (2004) Structural and thermodynamic evidence for a stabilizing role of Nop5p in S-adenosyl-L-methionine binding to fibrillarin. *J. Biol. Chem.*, **279**, 41822-41829.
109. Liu, S., Li, P., Dybkov, O., Nottrott, S., Hartmuth, K., Luhrmann, R., Carlomagno, T. and Wahl, M.C. (2007) Binding of the human Prp31 Nop domain to a composite RNA-protein platform in U4 snRNP. *Science*, **316**, 115-120.
110. Appel, C.D. and Maxwell, E.S. (2007) Structural features of the guide:target RNA duplex required for archaeal box C/D sRNA-guided nucleotide 2'-O-methylation. *RNA*, **in press**.
111. 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.
112. Dahlberg, J.E. and Lund, E. (1988) In Birnstiel, M. (ed.), *Structure and Function of the Major and Minor snRNAs*. Springer-Verlag, Berlin.
113. Stroke, I.L. and Weiner, A.M. (1985) Genes and pseudogenes for rat U3A and U3B small nuclear RNA. *J. Mol. Biol.*, **184**, 183-193.
114. Mazan, S., Michot, B. and Bachellerie, J.P. (1989) Mouse U3-RNA-processed pseudogenes are nonrandomly integrated into genomic DNA. Implications for the process of retrogene formation. *Eur. J. Biochem.*, **181**, 599-605.
115. Qu, L.H., Henras, A., Lu, Y.J., Zhou, H., Zhou, W.X., Zhu, Y.Q., Zhao, J., Henry, Y., Caizergues-Ferrer, M. and Bachellerie, J.P. (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.
116. 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.
117. Dunbar, D.A., Chen, A.A., Wormsley, S., and Baserga, S.J. (2000) The genes for small nucleolar RNAs in *Trypanosoma brucei* are organized in clusters and are transcribed as a

polycistronic RNA. *Nuc. Acids Res.*, **28**, 2855-2861.
118. Uliel, S., Liang, X.H., Unger, R. and Michaeli, S. (2004) Small nucleolar RNAs that guide modification in trypanosomatids: repertoire, targets, genome organization, and unique features. *Int. J. Parasitol.*, **34**, 445-454.
119. Cecconi, F., Mariottini, P., Loreni, F., Pierandrei-Amaldi, P., Campioni, N. and Amaldi, F. (1994) U17XS8, a small nucleolar RNA with a 12 nt complementarity to 18S rRNA and coded by a sequence repeated in the six introns of *Xenopus Laevis* ribosomal protein S8 gene. *Nuc. Acids Res.*, **22**, 732-741.
120. Liu, J. and Maxwell, E.S. (1990) Mouse U14 snRNA is encoded in an intron of the mouse cognate hsc70 heat shock gene. *Nuc. Acids Res.*, **18**, 6565-6571.
121. Pellizzoni, L., Crosio, C., Campioni, N., Loreni, F. and Pierandrei-Amaldi, P. (1994) Different forms of U15 snoRNA are encoded in the introns of the ribosomal protein S1 gene of *Xenopus laevis*. *Nuc. Acids Res.*, **22**, 4607-4613.
122. Prislei, S., Michienzi, A., Presutti, C., Fragapane, P. and Bozzoni, I. (1993) Two different snoRNAs are encoded in introns of amphibian and human L1 ribosomal protein genes. *Nuc. Acids Res.*, **21**, 5824-5830.
123. Xia, L., Lui, J., Sage, C., Trexler, B., Andrews, M. and Maxwell, E. (1995) Intronic U14 snoRNAs of Xenopus laevis are located in two different parent genes and can be processed from their introns during early oogenesis. *Nuc. Acids Res.*, **23**, 4844-4849.
124. Tycowski, K.T. and Steitz, J.A. (2001) Non-coding snoRNA host genes in Drosophila: expression strategies for modification guide snoRNAs. *Eur. J. Cell Biol.*, **80**, 119-125.
125. Huang, Z., Zhou, H., He, H., Chen, C., Liang, D. and Qu, L. (2005) Genome-wide analyses of two families of snoRNA genes from *Drosophila melanogaster*, demonstrating the extensive utilization of introns for coding of snoRNAs. *RNA*, **11**, 1303-1316.
126. Tycowski, K., Shu, M. and Steitz, J. (1996) A mammalian gene with introns instead of exons generating stable RNA products. *Nature*, **379**, 464-466.
127. Bortolin, M. and Kiss, T. (1998) Human U19 intron-encoded snoRNA is processed from a long primary transcript that possesses little potential for protein coding. *RNA*, **4**, 445-454.
128. Pelczar, P. and Filipowicz, W. (1998) The host gene for intronic U17 small nucleolar RNAs in mammals has no protein-coding potential and is a member of the 5'-terminal oligopyrimidine gene family. *Mol. Cell. Biol.*, **18**, 4509-4518.
129. Singh, S.K., Gurha, P., Tran, E.J., Maxwell, E.S. and Gupta, R. (2004) Sequential 2'-*O*-methylation of Archaeal pre-tRNA^{Trp} nucleotides is guided by the intron-encoded but trans-acting box C/D ribonucleoprotein of pre-tRNA. *J. Biol. Chem.*, **279**, 47661-47671.
130. Bortolin, M., Bachellerie, J. and Clouet-d'Orval, B. (2003) In vitro RNP assembly and methylation guide activity of an unusual box C/D RNA, cis-acting archaeal pre-tRNATrp. *Nuc. Acids Res.*, **31**, 6524-6535.
131. Clouet-d'Orval, B., Bortolin, M., Gaspin, C. and Bachellerie, J. (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. *Nuc. Acids Res.*, **29**, 4518-4529.
132. Steinmetz, E.J., Warren, C.L., Kuehner, J.N., Panbehi, B., Ansari, A.Z. and Brow, D.A. (2006) Genome-wide distribution of yeast RNA polymerase II and its control by Sen1 helicase. *Mol. Cell*, **24**, 735-746.
133. Suh, D., Busch, H. and Reddy, R. (1983) Isolation and characterization of a human U3

small nucleolar RNA gene. *Biochem. Biophys. Res. Commun.*, **137**, 1133-1140.
134. Ach, R.A. and Weiner, A.M. (1991) Cooperation between CCAAT and octamer motifs in the distal sequence element of the rat U3 small nucleolar RNA promoter. *Nuc. Acids Res.*, **19**, 4209-4218.
135. Savino, R., Hitti, Y. and Gerbi, S.A. (1992) Genes for *Xenopus laevis* U3 small nuclear RNA. *Nuc. Acids Res.*, **20**, 5435-5442.
136. Kiss, T. and Solymosy, F. (1990) Molecular analysis of a U3 RNA gene locus in tomato: transcription signals, the coding region, expression in transgenic tobacco plants and tandemly repeated pseudogenes. *Nuc. Acids Res.*, **18**, 1941-1949.
137. Connelly, S., Marshallsay, C., Leader, D., Brown, J.W. and Filipowicz, W. (1994) Small nucleolar RNA genes transcribed by either RNA polymerase II or RNA polymerase III in monocot plants share three promoter elements and use a strategy to regulate gene expression different from that used by their dicot plant counterparts. *Mol. Cell Biol.*, **14**, 5910-5919.
138. Shimba, S., Buckley, B., Reddy, R., Kiss, T. and Filipowicz, W. (1992) Cap structure of U3 small nucleolar RNA in animal and plant cells is different. Gamma-monomethyl phosphate cap structure in plant RNA. *J. Biol. Chem.*, **267**, 13772-13777.
139. Roberts, D.N., Stewart, A.J., Huff, J.T. and Cairns, B.R. (2003) The RNA polymerase III transcriptome revealed by genome-wide localization and activity-occupancy relationships. *Proc. Natl. Acad. Sci. USA*, **100**, 14695-14700.
140. Kruszka, K., Barneche, F., Guyot, R., Ailhas, J., Meneau, I., Schiffer, S., Marchfelder, A. and Echeverria, M. (2003) Plant dicistronic tRNA-snoRNA genes: a new mode of expression of the small nucleolar RNAs processed by RNase Z. *EMBO J.*, **22**, 621-632.
141. Petfalski, E., Dandekar, T., Henry, Y. and Tollervey, D. (1998) Processing of the precursors to small nucleolar RNAs and rRNAs requires common components. *Mol. Cell Biol.*, **18**, 1181-1189.
142. Chanfreau, G., Rotondo, G., Legrain, P. and Jacquier, A. (1998) Processing of a dicistronic small nucleolar RNA precursor by the RNA endonuclease Rnt1. *EMBO J.*, **17**, 3726-3737.
143. Chanfreau, G., Legrain, P. and Jacquier, A. (1998) Yeast RNase III as a key processing enzyme in small nucleolar RNAs metabolism. *J. Mol. Biol.*, **284**, 975-988.
144. Kim, M., Vasiljeva, L., Rando, O.J., Zhelkovsky, A., Moore, C. and Buratowski, S. (2006) Distinct pathways for snoRNA and mRNA termination. *Mol. Cell*, **24**, 723-734.
145. Steinmetz, E.J., Conrad, N.K., Brow, D.A. and Corden, J.L. (2001) RNA-binding protein Nrd1 directs poly(A)-independent 3'-end formation of RNA polymerase II transcripts. *Nature*, **413**, 327-331.
146. Carroll, K.L., Pradhan, D.A., Granek, J.A., Clarke, N.D. and Corden, J.L. (2004) Identification of cis elements directing termination of yeast nonpolyadenylated snoRNA transcripts. *Mol. Cell. Biol.*, **24**, 6241-6252.
147. Vasiljeva, L. and Buratowski, S. (2006) Nrd1 interacts with the nuclear exosome for 3' processing of RNA polymerase II transcripts. *Mol. Cell*, **21**, 239-248.
148. Morlando, M., Greco, P., Dichtl, B., Fatica, A., Keller, W. and Bozzoni, I. (2002) Functional analysis of yeast snoRNA and snRNA 3'-end formation mediated by uncoupling of cleavage and polyadenylation. *Mol. Cell Biol.*, **22**, 1379-1389.
149. Morlando, M., Ballarino, M., Greco, P., Caffarelli, E., Dichtl, B. and Bozzoni, I. (2004) Coupling between snoRNP assembly and 3' processing controls box C/D snoRNA

biosynthesis in yeast. *EMBO J.*, **23**, 2392-2401.
150. 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.
151. Levy, S., Avni, D., Hariharan, N., Perry, R.P. and Meyuhas, O. (1991) Oligopyrimidine tract at the 5' end of mammalian ribosomal protein mRNAs is required for their translational control. *Proc. Natl. Acad. Sci. USA*, **88**, 3319-3323.
152. Amaldi, F. and Pierandrea-Amaldi, P. (1997) In Jeanteur, P. (ed.), *Progress in Molecular and Subcellular Biology*. Springer-Verlag, Berlin, Vol. 18, pp. 1-17.
153. Hirose, T. and Steitz, J.A. (2001) Position within the host intron is critical for efficient processing of box C/D snoRNAs in mammalian cells. *Proc. Natl. Acad. Sci. USA*, **98**, 12914-12919.
154. Vincenti, S., De Chiara, V., Bozzoni, I. and C., P. (2007) The position of yeast snoRNA-coding regions within host introns is essential for their biosynthesis and for efficient splicing of the host pre-mRNA. *RNA*, **13**, 138-150.
155. Hirose, T., Ideue, T., Nagai, M., Hagiwara, M., Shu, M.D. and Steitz, J.A. (2006) A spliceosomal intron binding protein, IBP160, links position-dependent assembly of intron-encoded box C/D snoRNP to pre-mRNA splicing. *Mol. Cell*, **23**, 673-684.
156. Hirose, T., Shu, M.D. and Steitz, J.A. (2003) Splicing-dependent and -independent modes of assembly for intron-encoded box C/D snoRNPs in mammalian cells. *Mol. Cell*, **12**, 113-123.
157. Giorgi, C., Fatica, A., Nagel, R. and Bozzoni, I. (2001) Release of U18 snoRNA from its host intron requires interaction of Nop1p with the Rnt1p endonuclease. *EMBO J.*, **20**, 6856-6865.
158. Kiss, T. and Filipowicz, W. (1995) Exonucleolytic processing of small nucleolar RNAs from pre-mRNA introns. *Genes Dev.*, **9**, 1411-1424.
159. Allmang, C., Kufel, J., Chanfreau, G., Mitchell, P., Petfalski, E. and Tollervey, D. (1999) Functions of the exosome in rRNA, snoRNA and snRNA synthesis. *EMBO J.*, **18**, 5399-5410.
160. Caffarelli, E., Fatica, A., Prislei, S., De Gregorio, E., Fragapane, P. and Bozzoni, I. (1996) Processing of the intron-encoded U16 and U18 snoRNAs: the conserved C and D boxes control both the processing reaction and the stability of the mature snoRNA. *EMBO J.*, **15**, 1121-1131.
161. Villa, T., Ceradini, F., Presutti, C. and Bozzoni, I. (1998) Processing of the intron-encoded U18 small nucleolar RNA in the yeast *Saccharomyces cerevisiae* relies on both exo- and endo-nucleolytic activities. *Mol. Cell Biol.*, **18**, 3376-3383.
162. Samarsky, D., Fournier, M., Singer, R. and Bertrand, E. (1998) The snoRNA box C/D motif directs nucleolar targeting and also couples snoRNA synthesis and localization. *EMBO J.*, **17**, 3747-3757.
163. Lange, T.S., Borovjagin, A., Maxwell, E. and Gerbi, S. (1998) Conserved boxes C and D are essential nucleolar localization elements of U14 and U8 snoRNAs. *EMBO J.*, **17**, 3176-3187.
164. Lange, T.S., Ezrokhi, M., Borovjagin, A.V., Rivera-Leon, R., North, M.T. and Gerbi, S.A. (1998) Nucleolar localization elements of *Xenopus laevis*, U3 small nucleolar RNA. *Mol. Cell Biol.*, **9**, 2973-2985.
165. Narayanan, A., Speckman, W., Terns, R. and Terns, M. (1999) Role of the box C/D motif

in localization of small nucleolar RNAs to coiled bodies and nucleoli. *Mol. Cell Biol.*, **10**, 2131-2147.
166. 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.
167. Watkins, N.J., Lemm, I., Ingelfinger, D., Schneider, C., Hobbach, M., Urlaub, H. and Luhrmann, R. (2004) Assembly and maturation of the U3 snoRNP in the nucleoplasm in a large dynamic multiprotein complex. *Mol. Cell*, **16**, 789-798.
168. Verheggen, C., Lafontaine, D., Samarsky, D., Mouaikel, J., Blanchard, J.M., Bordonne, R. and Bertrand, E. (2002) Mammalian and yeast U3 snoRNPs are matured in specific and related nuclear compartments. *EMBO J.*, **21**, 2736-2745.
169. Boulon, S., Verheggen, C., Jady, B.E., Girard, C., Pescia, C., Paul, C., Ospina, J.K., Kiss, T., Matera, A.G., Bordonne, R. *et al.* (2004) PHAX and CRM1 are required sequentially to transport U3 snoRNA to nucleoli. *Mol. Cell*, **16**, 777-787.
170. Dennis, P.P. and Omer, A. (2005) Small non-coding RNAs in Archaea. *Curr. Opin. Microbiol.*, **8**, 685-694.
171. Roovers, M., Hale, C., Tricot, C., Terns, M., Terns, R., Grosjean, H. and Droogmans, L. (2006) Formation of the conserved pseudouridine at position 55 in archaeal tRNA. *Nuc. Acids Res.*, **34**, 4293-4301.
172. Liang, W., Clark, J. and Fournier, M.J. (1997) The rRNA-processing function of the yeast U14 small nucleolar RNA can be rescued by a conserved RNA helicase-like protein. *Mol. Cell. Biol.*, **17**, 4124-4132.

Figure Legends

Table 1. Box C/D RNA Databases and Search Engines

Figure 1. Archaeal and Eukaryotic Box C/D RNAs and Their Target RNAs

The secondary structures of archaeal and eukaryotic box C/D RNAs are shown with box C, D, C', and D' nucleotide sequences designated in bold. Representative target RNAs are base paired with the D and D' guide sequences and the nucleotides targeted for methylation indicated. Box C/D and C'/D' motifs are designated as well as those structural differences characteristic of the archaeal and eukaryotic box C/D RNAs. Know target RNAs of the archaeal and eukaryotic box C/D RNAs are shown below.

Figure 2. Archaeal and Eukaryotic Box C/D Ribonucleoprotein Complexes

Archaeal and eukaryotic box C/D RNAs are shown with the RNP core proteins bound to the box C/D and C'/D' motifs. The binding of all three archaeal core proteins to both the box C/D and C'/D' motifs assembles a symmetric sRNP. The differential distribution of the eukaryotic 15.5kD, Nop56, and Nop58 core proteins binding the box C/D and C'/D' motif establishes an asymmetric snoRNP. (Crosslinking experiments have implied the differential distribution of Nop58 and Nop56 on the eukaryotic box C/D and C'/D' motifs, respectively (94)).

Table 1

DATABASE	WEBSITE	COMMENTS / UTILITY	REFERENCES
snoRNA Database	http://lowelab.ucsc.edu/snoRNAdb/	Box C/D snoRNAs from *S. cerevisiae*, *A. thaliana*, and currently 8 Archaea are sorted by name, target and genome locus.	Lowe and Eddy, 1999; Omer et al., 2000; Brown et al., 2001
Yeast snoRNA Database	http://people.biochem.umass.edu/fournierlab/snornadb/main.php	Box C/D and H/ACA snoRNA database specific for yeast. Provides interactive 3-D view of rRNA modifications.	Samarsky and Fournier, 1999; Piekna-Przybylska et al., 2007
snoRNA-LBME-db	http://www-snorna.biotoul.fr/	Human box C/D and H/ACA snoRNA database. Find snoRNAs by name, target sequence or modification, or genomic location.	Lestrade and Weber, 2006
Plant snoRNA Database	http://bioinf.scri.sari.ac.uk/cgi-bin/plant_snorna/home	Database of currently 18 plant species in a tabulated and downloadable sequence format.	Brown et al., 2003
The snoRNP Database	http://www.mbio.ncsu.edu:8001/home.html	Collection of 6,349 small nucleolar RNAs and 302 associated proteins. Includes genus and species, accession numbers, loci, molecular family/class, and citations.	
Human snoRNA Database	http://www.trex.uqam.ca/~snorna/	Contains the sequences of currently 463 human snoRNAs. Provides 2-D structure and energy information imported from mFold.	
sno/scaRNAbase	http://gene.fudan.sh.cn/snoRNAbase.nsf	Database consisting of 1979 sno/scaRNAs from 85 organisms. Contains sequence, target site, accession number, references and allows users to perform BLAST searches.	Xie et al., 2007
RNAdb 2.0	http://research.imb.uq.edu.au/RNAdb	Comprehensive database of mammalian non-protein-coding RNAs (ncRNAs). Provides nucleotide sequences and annotations for tens of thousands of ncRNAs.	Pang et al., 2005, 2007

ALGORITHM / SEARCH ENGINE	WEBSITE	FUNCTION / UTILITY	REFERENCES
snoSeeker	http://genelab.zsu.edu.cn/snoseeker/	Searches for box C/D snoRNAs in the genomic alignment of two or more organisms.	Yang et al., 2006
snoScan	http://lowelab.ucsc.edu/snoscan/	Searches mammalian, yeast, or archael genomic sequences for box C/D snoRNA genes with a probabilistic model starting from a query or target RNA sequence.	Lowe and Eddy, 1999; Schattner et al., 2005
SNO.pl	http://hsc.utoledo.edu/bioinfo/eid/	Searches for conserved structures characteristic of C/D box snoRNAs within the Mammalian Orthologous Intron Database (MOID).	Fedorov et al., 2005

Figure 1

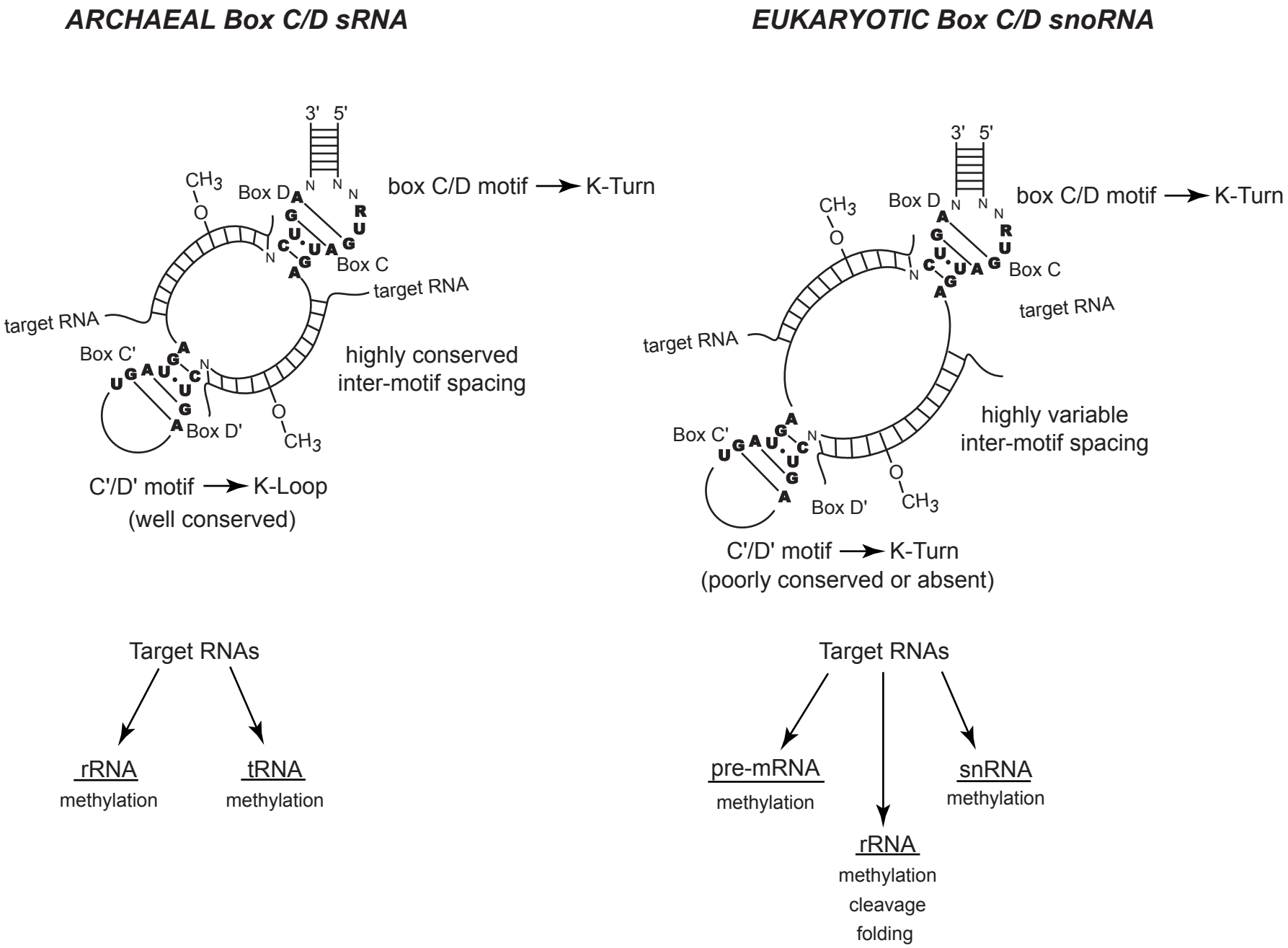
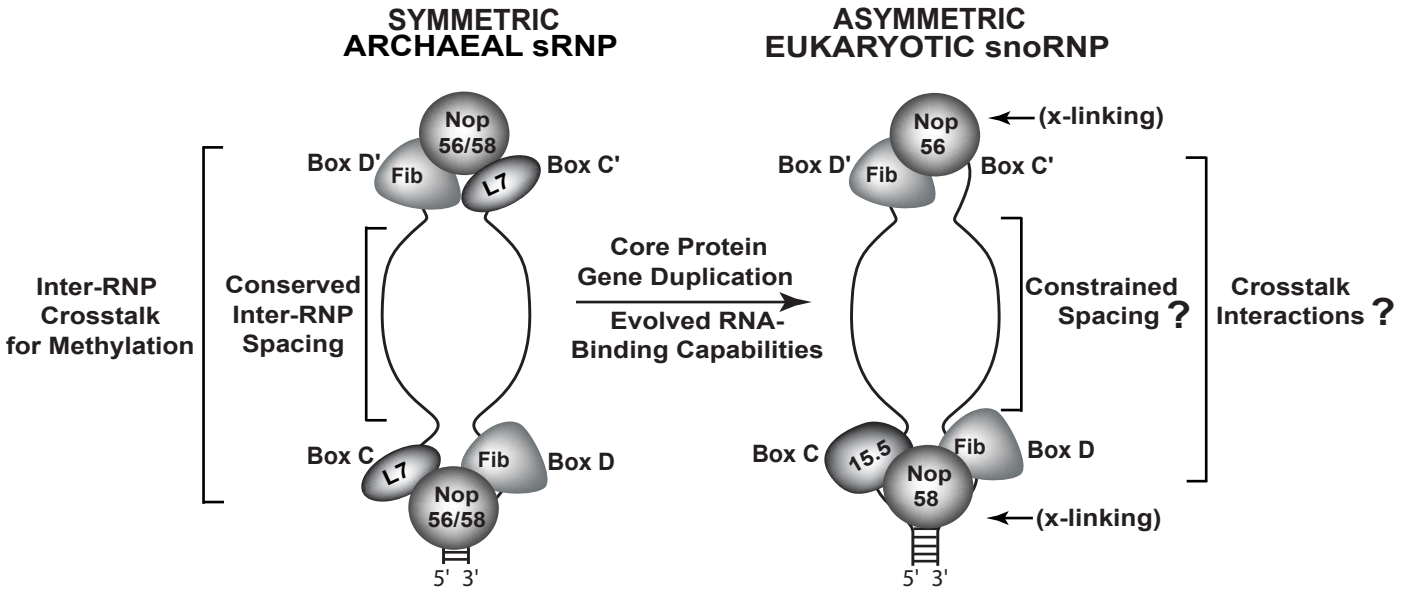

ARCHAEAL Box C/D sRNA
EUKARYOTIC Box C/D snoRNA
3' 5'
Box D
box C/D motif → K-Turn
Box C
target RNA
highly conserved
inter-motif spacing
Box C'
Box D'
C'/D' motif → K-Loop
(well conserved)
highly variable
inter-motif spacing
C'/D' motif → K-Turn
(poorly conserved or absent)
Target RNAs
rRNA
methylation
tRNA
methylation
pre-mRNA
methylation
snRNA
methylation
rRNA
methylation
cleavage
folding

Figure 2


SYMMETRIC
ARCHAEAL sRNP
ASYMMETRIC
EUKARYOTIC snoRNP
Nop 56/58
Box D'
Fib
L7
Box C'
Box C
L7
Fib
Box D
Nop 56/58
5' 3'
Inter-RNP Crosstalk for Methylation
Conserved Inter-RNP Spacing
Core Protein Gene Duplication
Evolved RNA-Binding Capabilities
Nop 56
(x-linking)
Box D'
Fib
Box C'
Box C
15.5
Fib
Box D
Nop 58
(x-linking)
5' 3'
Constrained Spacing ?
Crosstalk Interactions ?