Chapter

Multicomponent 2′-O-Ribose Methylation Machines:
Evolving Box C/D RNP Structure and Function

Keith T. Gagnon, Guosheng Qu and E. Stuart Maxwell*

Abstract

Methylation at the 2′-O ribose position is an abundant nucleotide modification of both eukaryal and archaeal RNAs. The methyltransferase responsible for this modification is frequently a ribonucleoprotein (RNP) complex consisting of a box C/D guide RNA and associated core proteins. These RNP "machines" are responsible for the modification of numerous cellular RNAs including ribosomal RNA, spliceosomal snRNAs and transfer RNAs. This chapter will review the structure and function of both eukaryotic and archaeal box C/D RNPs. A particular focus of our discussion will be the evolving components of the box C/D RNPs and the resultant consequences upon box C/D RNP structure and function.

Introduction

Guide RNAs for nucleotide modification were first described in the eukaryotic nucleolus where they were shown to modify ribosomal RNA. Based upon conserved sequence elements, these small nucleolar RNAs (snoRNAs) were classified into two major families. The box C/D snoRNAs guide nucleotide 2′-O-methylation whereas the H/ACA snoRNAs isomerize uridine to pseudouridine. Subsequent investigations revealed that box C/D and H/ACA guide RNAs are also found in Archaea. Further characterization of both eukaryotic and archaeal guide RNAs has demonstrated that they are bound by core proteins to form ribonucleoprotein (RNP) complexes. Both RNP families accomplish nucleotide modification using a similar mechanism. Guide RNAs utilize complementary sequences to base pair with specific target RNAs, thus designating a specific nucleotide for modification. The RNA-bound core proteins catalyze the 2′-O-methyl transfer and pseudouridylation reactions. The focus of this chapter is the evolving structure and function of the box C/D RNPs. For a detailed discussion of the H/ACA RNP structure and function, the reader is referred to Chapter 22 by Grozdanov and Meier entitled "Multicomponent Machines in RNA Modification: the H/ACA Ribonucleoproteins".

Ribonucleotide Methylation and Methylation Function

Key features of ribose 2′-O-methylation indicate that this abundant nucleotide modification plays an important role in RNA folding and stability. Methylation at the ribose 2′ position stabilizes an RNA chain by inhibiting backbone cleavage and increasing the stability of base pairing and stacking interactions, thus potentially affecting the RNA's structure and ultimately function.[1,2] A number of important cellular RNAs are 2′-O-methylated by box C/D RNPs. Although the

*Corresponding Author: E. Stuart Maxwell—Department of Molecular and Structural Biochemistry, North Carolina State University, Raleigh, North Carolina, USA, 27695-7622. Email: stu_maxwell@ncsu.edu

DNA and RNA Modification Enzymes: Comparative Structure, Mechanism, Functions, Cellular Interactions and Evolution, edited by Henri Grosjean. ©2009 Landes Bioscience.

function of rRNA modification is not fully understood, disrupting box C/D snoRNA-directed 2′-O-methylation results in slowed cell growth and reduced ability of cells to adapt to environmental changes with adverse affects on ribosome biogenesis and function.[3-6] When mapped on the ribosome, these modifications cluster around functionally significant regions like the peptidyl transferase center.[4] Eukaryotic spliceosomal RNAs (snRNA) are also methylated by a variety of box C/D snoRNAs as are small Cajal RNAs (scaRNAs), RNAs localized to nuclear Cajal bodies that can contain both box C/D and H/ACA motifs.[7] Box C/D snoRNA-directed methylation of select mRNAs has been implicated in regulating RNA editing and splicing of brain mRNAs.[8,9] Computational analyses have recently revealed that alternative splice junctions may also be targets for snoRNA-guided modification as they are often complementary to a number of "orphan" box C/D snoRNA guide regions.[10] Unique to Archaea, box C/D RNAs guide methylation of tRNAs, thus potentially affecting not only tRNA folding and structure but also tRNA function in translation.[11,12]

Box C/D RNP Function

The primary function of eukaryotic and archaeal box C/D RNPs is nucleotide methylation of diverse cellular RNAs. However, other functions in RNA metabolism have been demonstrated. In eukaryotes, box C/D snoRNAs function in prerRNA processing. Select box C/D snoRNAs are essential for specific endonucleolytic cleavage events in prerRNA maturation, likely functioning as "organizers" for a trans-acting RNAse.[13-15] Several box C/D snoRNAs also play roles in prerRNA folding.[15-17] For both functions, the box C/D snoRNA utilizes complementary sequences to base pair with the prerRNA. Notably, these additional functions have not yet been observed for archaeal box C/D sRNAs.[18] This may reflect a more limited examination of archaeal box C/D sRNA populations and functions or perhaps evolving functional roles of the box C/D snoRNP in eukaryotes.

Box C/D RNAs: Diversity of Sequence and Structure

Large populations of box C/D RNAs are found in eukaryotic and archaeal organisms. In various archaeal organisms, scores of box C/D sRNAs have been identified using bioinformatic approaches and many have been experimentally verified.[19-22] However, the list of sRNAs remains small and is still limited to a handful of organisms. It appears that Archaea do not share box C/D RNA homologs with Eukarya, indicating an evolutionary ancient divergence of eukaryotic and archaeal RNAs.[19,23,24] Box C/D snoRNA populations are better defined in eukaryotes, although not nearly complete. In the unicellular eukaryote yeast, the defined box C/D snoRNA population consists of 46 species.[25,26] In humans, a larger population of over 100 box C/D snoRNAs has been identified and this number is likely to grow significantly.[24] Interestingly, the identification of brain-specific species in mammals suggests an expanding complexity of tissue-specific RNAs and perhaps snoRNA function in metazoan organisms.[9,27,28] Even more numerous may be the plant box C/D RNAs whose populations are predicted to be in the hundreds.[29] Although box C/D snoRNAs from different eukaryotic organisms can guide evolutionarily conserved modifications, most nucleotides targeted for modification are unique to a given organism, reflecting the general lack of snoRNA species conservation among eukaryotes.

The hallmark of box C/D guide RNAs are the box C (RUGAUGA) and box D (CUGA) sequence elements located at the 5′ and 3′ RNA termini, respectively (Fig. 1A). Frequently present are internal box C' and D' elements which are well conserved in archaeal sRNAs but often difficult to discern in eukaryotic snoRNAs.[30] These terminal and internal boxes establish the box C/D and C'/D' motifs, respectively. Both motifs fold into RNA elements known as kink-turns (K-turns) first revealed in U4 snRNA and archaeal ribosomes.[31,32] K-turns are characterized by an asymmetric bulge flanked by two stems and stabilized by tandem, sheared G:A pairs. The G:A pairs hydrogen bond across the bulge to generate a sharp, archetypical bend, or kink of approximately 60° in the RNA backbone.[31] Importantly, internal C'/D' motifs fold into a modified K-turn structure where canonical stem I is replaced by a loop. These modified K-turns have been

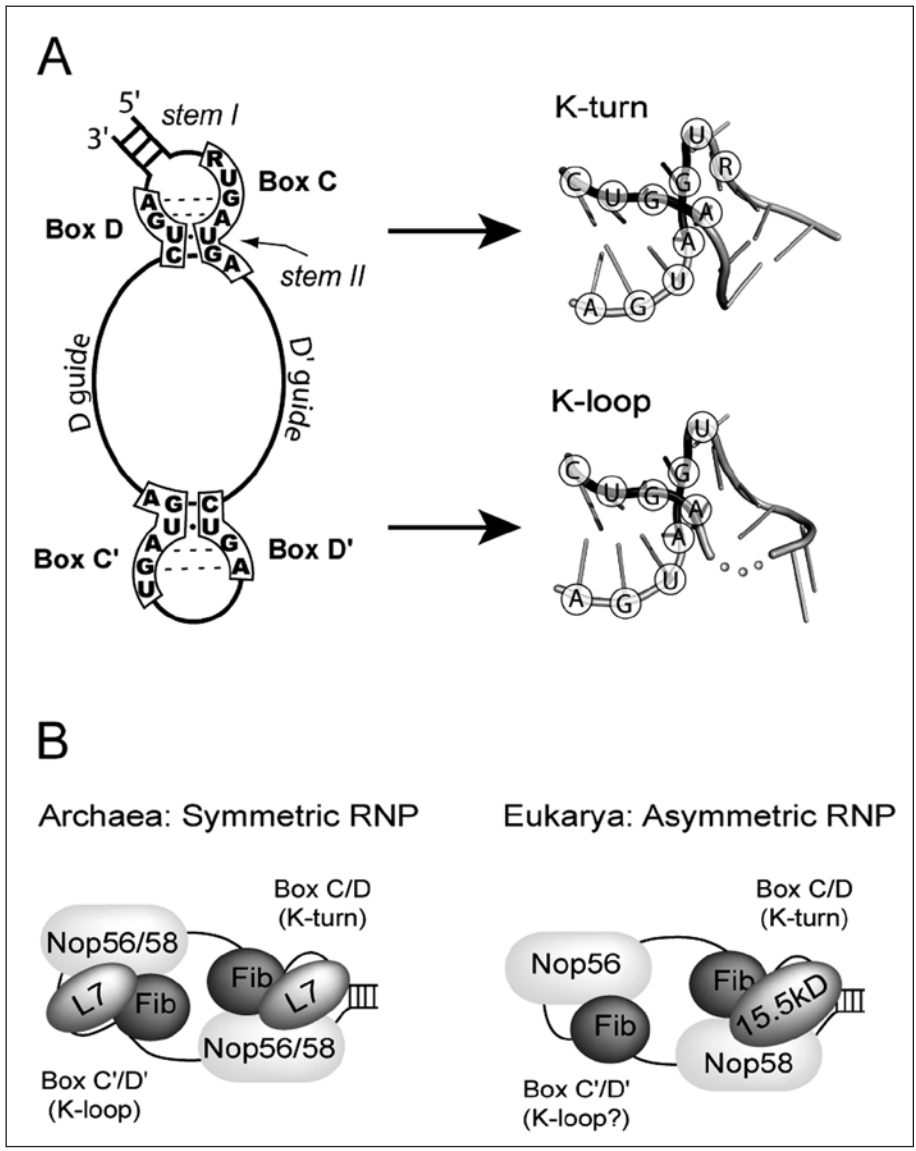

Figure 1. Archaeal and Eukaryal Box C/D RNPs. A) Secondary structural elements of box C/D RNAs and tertiary structure of the K-turn (box C/D) and K-loop (box C′/D′) motifs. Conserved box C, D, C′ and D′ nucleotides are indicated. B) RNP structure with protein distribution based upon current experimental evidence. Archaeal sRNP protein distribution based upon in vitro RNP assembly. Eukaryotic snoRNP protein distribution based upon in vivo crosslinking and in vitro protein binding. See text for specific experiments.

designated "K-loops".[33] K-turns have also been observed in mRNAs, archaeal H/ACA sRNAs and even the SAM riboswitch.[34-36] K-turn and K-loop motifs are typically protein binding platforms, important for stabilizing tertiary RNA and RNP structures. Individual box C/D RNA species are defined by their unique guide sequences located upstream of boxes D and D'. Guide sequences are

10-21 nucleotides long and complementary to their respective target RNA. It is the target RNA nucleotide which is base paired to the fifth nucleotide of the guide sequence that is specifically 2′-O-methylated by the RNP complex.[37]

Differences in size and structure between the archaeal and eukaryotic box C/D RNAs has contributed to structural and perhaps functional diversity. Archaeal sRNAs are smaller (50-70 nucleotides) and possess terminal box C/D and internal C'/D' motifs separated by minimal guide regions. Guide region length is highly conserved at 12 nucleotides in archaeal box C/D sRNAs and thus box C/D and C'/D' motif spacing is conserved.[20,38] Interestingly, circular box C/D sRNAs have been reported in some archaeal organisms.[39] In contrast, eukaryotic snoRNAs are larger in size (most often greater than 75 nucleotides) with significantly larger guide regions and associated spacer sequences between the two motifs. For those box C/D snoRNAs with hard to define or missing C'/D' motifs, the D guide sequence and associated spacer region can be quite large.[24,25,29,38] Some eukaryotic box C/D snoRNAs utilize their guide regions for prerRNA processing steps. The larger RNA size and correspondingly larger guide regions may have contributed to and even facilitated the functional diversity of box C/D snoRNPs in Eukarya.

Box C/D RNP Structure and Assembly

Mature box C/D RNAs are assembled as ribonucleoprotein complexes bound with a limited number of highly conserved core proteins (Fig. 1B). Eukaryotic box C/D snoRNPs contain four conserved core proteins: the 15.5kD protein, nucleolar proteins Nop56 and Nop58 and the methyltransferase enzyme fibrillarin.[40-42] Three highly homologous proteins, ribosomal protein L7Ae, Nop56/58 and fibrillarin, bind the archaeal box C/D sRNAs to assemble a simpler and what could be considered minimal box C/D sRNP complex.[43,44]

In vitro reconstitution of catalytically active archaeal box C/D sRNPs has revealed an order of core protein binding.[44,45] L7Ae initiates sRNP assembly by binding the K-turn and K-loop motifs of the terminal box C/D and internal C'/D' motifs, respectively.[43-45] Nop56/58 and then fibrillarin bind both the terminal box C/D and internal C'/D' motifs to assemble a "symmetric" sRNP with all three core proteins bound to both motifs.[44,45] The assembly of a symmetric RNP is essential for efficient nucleotide methylation.[45,46] Initial binding of L7Ae core protein stabilizes K-turn and K-loop structure and remodels the box C/D RNA to facilitate subsequent binding of the Nop56/58 and fibrillarin proteins.[47,48] Remodeling of the sRNA continues with binding of Nop56/58 while fibrillarin has no significant affect on RNA structure.[48] For the archaeal complex, RNA remodeling requires elevated temperature to increase RNA structure dynamics, thus facilitating core protein binding. Notably, in vitro assembly of archaeal box C/D sRNPs does not require accessory proteins for either RNA remodeling or hierarchical core protein binding.[44,45]

In contrast to the symmetric archaeal sRNP, the eukaryotic box C/D snoRNPs assemble an apparently "asymmetric" complex.[49] The 15.5kD protein initiates snoRNP assembly but appears to bind only the K-turn of the terminal box C/D motif.[50] Core proteins Nop58 and Nop56 have been predicted to bind the C/D and C'/D' motifs, respectively, based upon in vivo crosslinking experiments.[49] Only fibrillarin appears to be associated with both motifs. Unfortunately, the lack of a functional in vitro assembly system for the eukaryotic complex has hampered a more detailed analysis of box C/D snoRNP assembly and structure.

Limited knowledge of eukaryotic box C/D snoRNP assembly has nonetheless revealed a highly complex and dynamic process requiring accessory factors. Assembly of the mammalian presnoRNP requires two trans acting AAA+ ATPases, TIP48 and TIP49.[51,52] Additional processing/assembly factors for the U3 snoRNP include TGS1, La, LSm proteins and the exosome as well as nucleocytoplasmic transport factors such as PHAX, CRM1, CBC, Ran and Nopp140.[51,53] Four novel human biogenesis factors (BCD1, NOP17, NUFIP and TAF9), which are likely to be involved in the formation of the U8 presnoRNP, have also been identified.[54] Most recently, the heat shock protein Hsp90 has been implicated in orchestrating assembly of the eukaryotic complex.[55,56] Whereas archaeal sRNPs require elevated temperature (accessory factors in vivo?) to facilitate RNA remodeling required for in vitro sRNP assembly, the eukaryotic snoRNPs require multiple

accessory factors for in vitro and in vivo assembly. These accessory factors are presumed to promote RNA remodeling and facilitate sequential core protein binding, an apparently common theme of both archaeal and eukaryotic box C/D RNP biogenesis.

The vast majority of higher eukaryotic snoRNA coding sequences are positioned within introns of RNA Polymerase II protein coding host genes. A second genomic organization, prevalent in yeast and plants, is box C/D snoRNA genes transcribed from independent RNA Pol II (infrequently Pol III) promoters.[57] Archaeal box C/D sRNA genes, although not well characterized, appear to be intergenic and transcribed from independent promoters.[11,20] Transcription of intronic box C/D snoRNAs is coupled with the transcription of the host premRNA and linked to splicing.[58] Box C/D snoRNP assembly is also coupled with posttranscriptional processing, maturation and transport to the nucleolus.[59] The differences in genomic organization for the eukaryotic box C/D snoRNAs versus archaeal sRNA coding sequences perhaps reflects an evolution of gene structure for purposes of regulated expression.

Structure, Function and Evolution of the L7Ae/15.5kD Core Protein

Archaeal core protein L7Ae intitiates sRNP assembly by binding the terminal box C/D (K-turn) and internal C'/D' (K-loop) motifs.[44,45] L7Ae binding remodels sRNA structure and establishes a platform for subsequent box C/D sRNP core protein binding.[47,48,60] Eukaryotic 15.5kD protein similarly initiates snoRNP assembly by binding the terminal box C/D core motif's K-turn.[43,45,50] The differential binding of L7Ae and 15.5kD proteins to K-turn and K-loop motifs in vitro is striking as the crystal structures of both proteins are nearly superimposable and their RNA-binding domains are well conserved across both domains of life[31,61,62] (Fig. 2).

L7Ae and 15.5kD are members of the L7Ae/L30 protein family[31,61,62] (Fig. 2). Additional members of this family include rpL30e in Archaea and Rpp38, rpL30, rpL7a, SBP2 and Nhp2p proteins in Eukarya. Proteins in this closely related family are typically small and composed of an internal beta sheet surrounded by several alpha helices, a three-layer topology fold known as an alpha-beta-alpha sandwich.[32,34] They possess conserved RNA binding domains, almost uniformly recognize K-turn motifs and play critical roles in RNA stabilization and RNP assembly.[31,32,34,42,55,61-63]

Each family member is interesting from a functional standpoint. Family members in both Archaea and Eukarya function as ribosomal proteins of the large subunit. Eukaryotic ribosomal protein rpL30 is also capable of binding its own mRNA to regulate translation and ribosomal protein Rpp38 is a constituent protein of the MRP complex.[34,64] SBP2 is another mRNA-binding protein, recognizing those mRNAs possessing the SECIS RNA element important for selenocysteine incorporation into selenoproteins. It consists of multiple domains including one very similar to that found in the L7Ae/15.5kD protein.[65] An L7Ae/L30 sequence appears to have been inserted during genomic shuffling, thus conferring K-turn RNA binding capability upon SBP2.[35] The 15.5kD protein is not only a box C/D snoRNP core protein but also a component of the spliceosomal U4 snRNP where it also binds a K-turn motif in U4 and functions in snRNP assembly.[32,42,60] Eukaryotic nonhistone chromosomal protein 2 (NHP2p) is a core protein of the eukaryotic H/ACA snoRNPs and highly homologous to both archaeal L7Ae and eukaryotic 15.5kD proteins. NHP2p binds a stem loop of the box H/ACA snoRNAs and is essential for H/ACA snoRNP assembly. Notably, NHP2p stands out as being the sole L7Ae/L30 family member without clear RNA-binding specificity. Specific recruitment of NHP2 to the assembling snoRNP requires interaction with the RNA and other core proteins.[66] Its functional equivalent in the archaeal H/ACA sRNP is L7Ae, the only guide RNA core protein of both domains to be found in both the box C/D and H/ACA RNPs and binding the K-turn/K-loop motifs. Despite great similarities in sequence and folded structure, each L7Ae/L30 family member has sufficiently diverged such that its binding is specific for the K-turn of its respective cognate RNA.[31,34,35,45,50,66]

The recurring theme of L7Ae/L30 protein function is RNP formation via recognition of the K-turn motif. The binding of L7Ae/L30 proteins to a variety of RNAs provides insight into the evolutionary emergence of the L7Ae/L30 protein family and even evolution of the box C/D RNPs. The limited number of L7Ae/L30 proteins in Archaea (two) and expansion of family members in

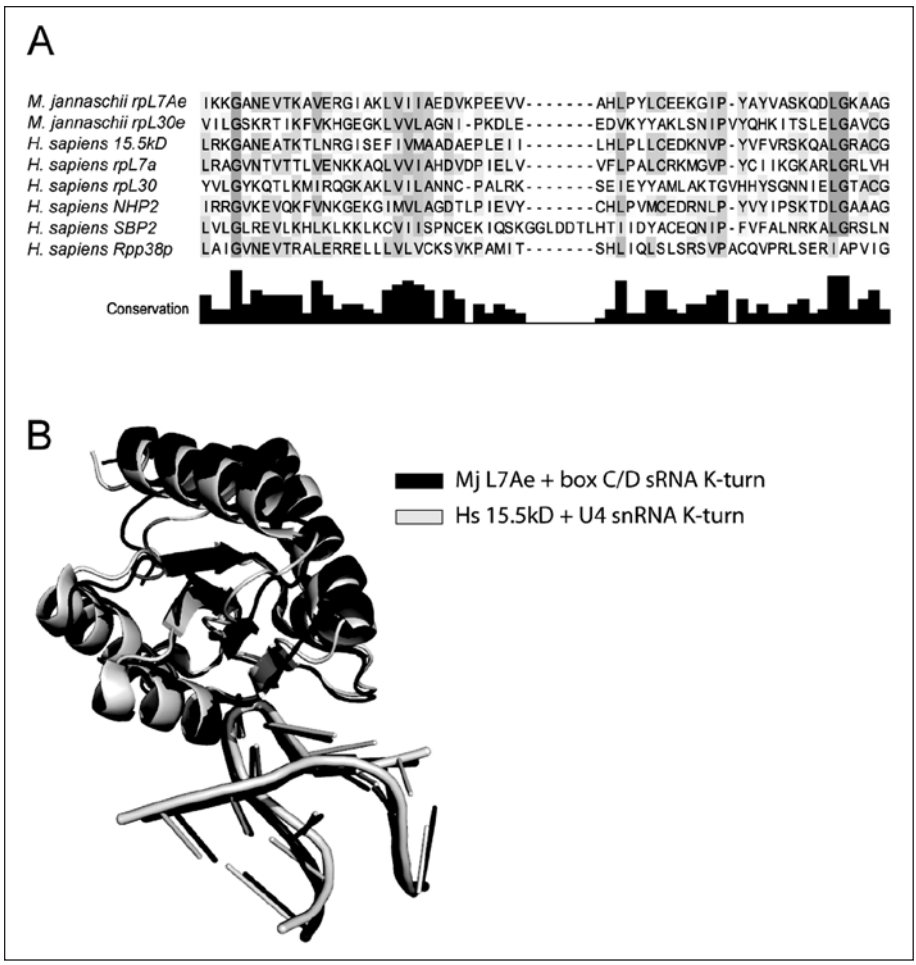

Figure 2. Conserved Sequence and Structure of the L7Ae/L30 Protein Family. A) Sequence alignment of the conserved RNA-binding domain of known L7Ae/L30 protein family members. Conservation is indicated by shaded amino acids and the bar graph alignment below. B) Superimposed structures of *M. jannaschii* L7Ae and human 15.5kD bound to their respective K-turn RNAs. *M. jannaschii* L7Ae (1RLG) is shown in black and human 15.5kD (1E7K) bound is shown in white.

eukaryotes (six) suggest a continuing evolution and diversity of protein structure and function, particularly in eukaryotic organisms. L7Ae is a component of three separate RNPs in Archaea whereas in eukaryotes these same functions are carried out by three separate but closely related family members (ribosomal protein L7a, 15.5kD, NHP2). This would suggest that L7Ae is the progenitor of the L7Ae/L30 protein family.[31,43,45,50,62] We have previously proposed that L7Ae or an L7Ae-like protein binding a K-turn motif in a primitive RNP translational apparatus may be the ancestral RNP complex for this protein family.[23] The utilization of a single archaeal core protein to bind K-turns in both archaeal box C/D and H/ACA RNPs suggests a common RNP origin for both guide RNP families early in evolution. The absence of L7Ae/L30 RNP complexes in Eubacteria implies an emergence and evolution of the protein family after divergence of Archaea and Eukarya from Eubacteria.

Evolution of L7Ae/15.5kD RNA-binding capabilities may well have facilitated evolution of box C/D RNP structure and hence function. The minimal archaeal box C/D sRNP makes use of L7Ae for both K-turn and K-loop binding, thus assembling a symmetric sRNP whose box C/D and C'/D' RNPs are spatially constrained and functionally coupled.[38,45] In contrast, the evolved binding capability of eukaryotic 15.5kD to recognize only the box C/D K-turn could have allowed greater structural and consequently functional snoRNP diversity. Lack of sequence conservation in the C'/D' motif of eukaryotic snoRNAs may reflect a concomitant loss of 15.5kD binding, resulting in spatial decoupling of the internal and terminal motifs.[38,45] Thus, modern day eukaryotic box C/D snoRNAs are less conserved in sequence and larger in size. This flexibility in snoRNA structure may have allowed the eukaryotic complexes to drift further, acquiring new functions such as chaperoning prerRNA processing events.

Structure, Function and Evolution of the NOP56 and NOP58 Core Proteins

The Nop core proteins play essential structural and functional support roles in the box C/D RNPs. A single Nop56/58 is found in Archaea while two homologs, presumably arising from gene duplication and designated Nop56 and Nop58, are present in Eukarya.[41,44] Their roles include bridging protein interactions within box C/D RNPs, RNA remodeling during RNP assembly, fibrillarin recruitment and assisting the methyltransferase reaction. Archaeal Nop56/58 helps to remodel RNA structure during in vitro box C/D sRNP assembly by restructuring guide regions and box elements after initial remodeling by L7Ae.[48] Nop56/58 interactions with other core proteins may affect RNA remodeling, perhaps helping to establish bridging interactions between the box C/D and C'/D' RNPs.[48,67] Archaeal Nop56/58 and eukaryotic Nop56 and Nop58 proteins interact with fibrillarin.[45,67,68] Evidence from in vitro assembly of the archaeal sRNP suggests that Nop56/58 and fibrillarin may bind the assembling complex as a dimmer.[45,48,67] While the methyltransferase fibrillarin clearly interacts with guide and target RNAs, its binding in the archaeal sRNP is primarily through interaction with Nop56/58.[45,48,67,69] Archaeal Nop56/58 may also assist in catalysis of the methyltransferase reaction as critical Nop56/58 amino acids are positioned adjacent to the S-adenosyl-L-methionine binding site of fibrillarin.[70]

Only a few members of the Nop protein superfamily have been well characterized. They include the box C/D RNP core proteins Nop56/58 in Archaea, Nop56 and Nop58 in Eukarya and eukaryotic Prp31 (premRNA processing factor 31).[41,45,60,67-69] Nop proteins are composed of an N-terminal domain, a central coiled-coiled domain, a Nop domain and a variable lysine-rich C-terminal tail (Fig. 3A). The N-terminal domain is not well characterized in eukaryotes but is responsible for dimerization with fibrillarin in Archaea.[67] The coiled-coil domain may mediate protein interactions with other core proteins or regulatory factors. Crystal structures of the Nop56/58-fibrillarin dimer from Archaea show that the coiled-coil domain can dimerize with itself, leading to the suggestion that this interaction could mediate protein-protein or crosstalk interactions between the box C/D and C'/D' RNPs.[67,69] Best understood is the Nop domain, the defining feature of the Nop superfamily, which comprises most of the C-terminal region. A recent U4-15.5kD-Prp31 RNP crystal structure has provided new insight into the role of this domain in RNP assembly[60] (Fig. 3B). The Prp31 Nop domain makes nearly equal contact with both U4 RNA and the 15.5kD protein, thus explaining a need for 15.5kD to be bound to U4 for Prp31 interaction.[71] In a similar manner, archaeal Nop56/58 binds a box C/D RNA only after L7Ae has first bound the K-turn or K-loop motif.[45] Thus, the Nop proteins may serve as checkpoints in RNP assembly, ensuring that the K-turn recognition protein has first bound RNA. Deletion of the Nop domain completely disrupts binding to the box C/D RNA-L7Ae complex, indicating that it is the necessary RNP assembly module of Nop protein family members.[67] The highly charged, lysine-rich C-terminal tail, also called a KKE/D repeat, remains an enigma. It is poorly conserved in sequence and length and appears to be dispensable for Nop protein function in both Eukarya and Archaea.[68,69]

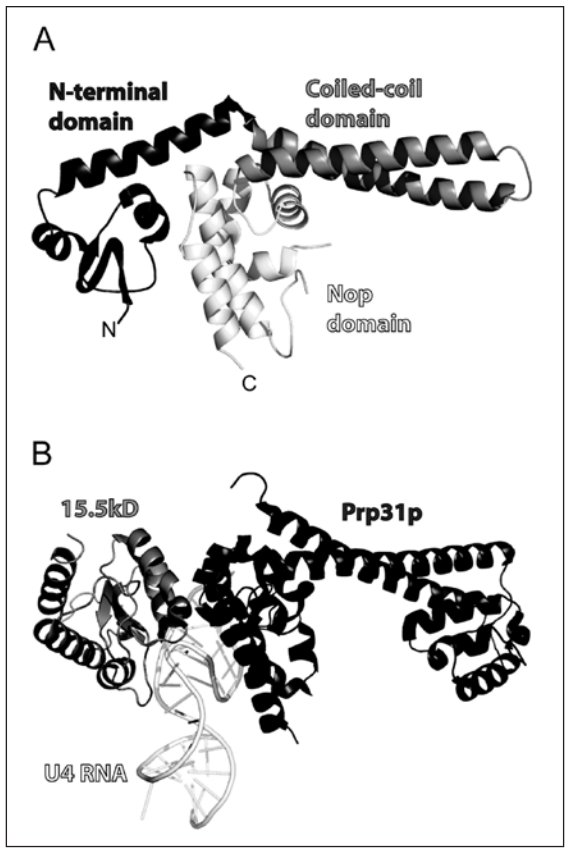

Figure 3. Nop Protein Structure and RNP Interaction. A) Crystal structure of the *Archaeoglobus fulgidus* Nop56/58 core protein (1NT2). N-terminal, coiled-coil and C-terminal (Nop) domains are shown in black, gray and white, respectively. B) Crystal structure of the human Prp31 Nop domain protein bound to the 15.5kD-U4 snRNA RNP (2OZB) through its C-terminal (Nop) domain. Prp31 is shown in black, 15.5kD in gray and the U4 K-turn in white.

Eukaryotic box C/D snoRNPs may owe much of their structural and functional diversity to evolution of the Nop56/58 core protein. In archaeal box C/D sRNPs the Nop56/58 protein binds both box C/D and C'/D' motifs.[45] In contrast, crosslinking experiments indicate that eukaryotic Nop56 and Nop58 may differentially bind the C'/D' and C/D motifs, respectively.[49] Nop56 and Nop58 are highly related, with the mouse proteins having 43% identity and 63% similarity.[41] Archaeal Nop56/58 from *Methanocaldococcus jannaschii* is 57% and 59% similar to mouse Nop56 and Nop58, respectively. Thus, gene duplication of Nop56/58 coding sequence followed by co-evolution of the two eukaryotic proteins and the box C/D RNA could contribute to the apparent asymmetric structure of eukaryotic box C/D snoRNPs.[23] As 15.5kD does not recognize the K-loop, association of Nop56 with the C'/D' motif could suggest that this Nop protein has acquired the ability to bind RNA independently of 15.5kD.[49,50] In vitro assembly of the archaeal sRNP has also shown that archaeal Nop56/58 along with fibrillarin can specifically, albeit weakly, bind the K-loop motif in the absence of L7Ae RNA.[45] The possible differential recognition of Nop56 and Nop58 proteins to K-loop and K-turn motifs, respectively, as well as the K-turn specificity of the 15.5kD protein, could also contribute to the uncoupling of the eukaryotic box C/D and C'/D' RNP complexes.

Structure, Function and Evolution of Fibrillarin

Fibrillarin is the catalytic protein of the box C/D RNPs, yet it plays only a minor role in RNP assembly. In Archaea, fibrillarin is recruited to the complex primarily through protein-protein interaction with the Nop56/58 protein.[45,48,67] In eukaryotes, fibrillarin may play a more active role in assembly. Eukaryotic fibrillarin contacts the box C/D snoRNAs and association of Nop56 requires the presence of fibrillarin.[49,72,73] Fibrillarin is recruited to the RNP at a late stage of assembly.[44,45,48,51]

Fibrillarin was originally predicted to be the methyltransferase enzyme based on its sequence similarity to other *S*-adenosyl-l-methionine (SAM)-dependent methylases[74] (Fig. 4A). Subsequent in vitro reconstitution of box C/D RNPs[44,45] and crystallographic analyses of archaeal fibrillarins[66,67,75] provided further evidence of the methlytransferase function of fibrillarin. Despite this progress, it is still unknown exactly how fibrillarin interacts with guide and target RNAs to accurately methylate the target nucleotide.

Eukaryotic and archaeal fibrillarin proteins have both common and unique features. They all share a highly conserved alpha-beta carboxy-terminal domain (CTD) in which is nested a short consensus sequence, the SAM-binding motif.[67,75,76] The CTD of *M. jannaschii* fibrillarin (Mjfib) is approximately 60% identical and 80% similar to vertebrate fibrillarins between residues 25 and 95 of the CTD, which harbors the SAM-binding motif. Even in poorly related regions outside this segment (Mjfib residues 95-227), archaeal and eukaryotic fibrillarins are about 40% identical and 65% similar[76] (Fig. 4A).

In contrast to the CTD, fibrillarin proteins have variable sequence and structure in their N-terminal domains (NTD). Eukaryotic fibrillarins often contain a glycine-arginine-rich (GAR) domain which is necessary and sufficient for nucleolar localization of eukaryotic box C/D snoRNPs.[77] However, archaeal fibrillarins lack this domain and their N terminal regions are much shorter[67,75,76] (Fig. 4A). Moreover, the fibrillarin NTD varies within archaeal species and may confer different protein binding properties upon them.[75,78] For example, the Mjfib NTD was reported to facilitate dimerization of fibrillarin molecules through specific β-strand interactions.[76] In contrast, available evidence indicates that fibrillarin from both *Archaeoglobus fulgidus* and *Pyrococcus furiosus* exist as monomers in solution and in crystalline state.[67,75]

Despite a lack of significant sequence homology, the archaeal fibrillarin CTD is structurally similar to other SAM-dependent methylases. The consensus topology for the methyltransferase catalytic domain is a seven-stranded β-sheet flanked by three α-helices on each side[76] (Fig. 4B). The CTD of MjFib forms a Rossman fold like other methyltransferases and only differs from the consensus topology by the addition of a minihelix (α_5). Fibrillarin is most closely related to other SAM-dependent RNA methyltransferases, like RrmJ from *E. coli* which catalyzes site-specific 2′-O-methylation of rRNAs, tRNAs and mRNAs independent of a guide RNA.[79] The site-specific RNA methyltransferases (MTases) related to RrmJ and snoRNA-directed RNA MTases related to fibrillarin form a closely related monophyletic clade. They possess a spatially superimposable tetrad of conserved residues localized in the heart of the substrate-binding pocket, three of which (K-D-K) are essential for activity[79,80] (Fig. 4C). This invariant triad is considered a synapomorphy, an ancient feature derived from a common ancestor that might have possessed ribose 2′-O-MTase activity. Collectively, these observations suggest that methyltransferase enzymes evolved from a common ancestor to acquire substrate-specific activities.

Fibrillarin relies upon a guide RNA and other core proteins in an assembled box C/D RNP to catalyze nucleotide-specific 2′-O-methylation.[44,45] Most other methyltranferases utilize accessory domains for substrate specificity. For example, the DNA methylase HhaI recognizes and binds its double-stranded DNA substrate by utilizing a large peripheral domain which binds the DNA and flips the target base out of the duplex for modification.[81] Evolution of fibrillarin appears to have occurred within the box C/D RNPs as well. Archaeal fibrillarins possess organism-specific NTDs while eukaryotic fibrillarins have related GAR domains.[78] Aside from affecting nucleolar localization, the GAR domain serves as an interaction domain with the SMN protein which is transiently associated with premature box C/D snoRNPs and important for assembly.[82]

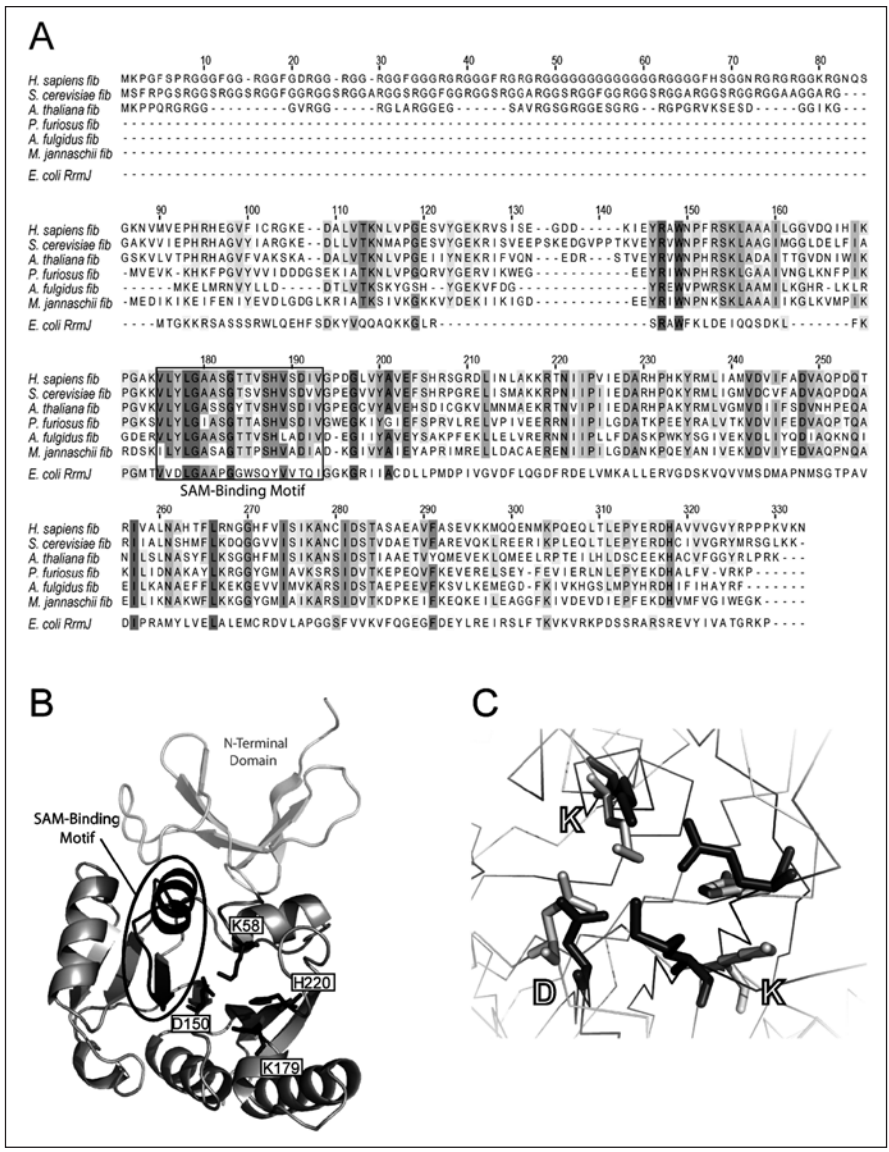

Figure 4. Conserved sequence and structure of fibrillarin. A) Sequence alignment of three eukaryotic and three archaeal fibrillarins with the *E. coli* RrmJ methyltransferase. Degree of conservation is indicated by shades of gray. The highly conserved SAM-binding motif is boxed. B) Crystal structure of *M. jannaschii* fibrillarin (1FBN). The variable N-terminal domain is light gray, the SAM-binding motif circled and highly conserved catalytic residues designated (black sticks). C) Spatial superposition of the *E. coli* RrmJ catalytic residues (black) (1EIZ) with those of *M. jannaschii* fibrillarin (light gray). The invariant catalytic triad (K-D-K) is labeled and peptide backbones are illustrated with lines.

Interestingly, eukaryotic fibrillarin may have other roles in addition to ribose methylation. Most eukaryotic box C/D snoRNPs appear to direct only one ribose methylation per snoRNA,

even though fibrillarin is believed to bind both box C/D and C'/D' motifs. Notably, box C/D snoRNPs involved only in prerRNA processing or folding, such as U3 and U8, also contain the fibrillarin core protein.[54,83] These observations suggest that eukaryotic fibrillarin may have acquired a more structural role in some RNPs and may possess other functions aside from strictly catalyzing the methyltransfer reaction.

Concluding Remarks: The Evolving Box C/D RNP Machinery

RNA-guided nucleotide modification complexes are ancient RNA:protein enzymes found in both Eukarya and Archaea. Despite their conservation in these two domains of life, the box C/D RNPs exhibit domain-specific structural and functional features indicating an evolving RNP over time. The archaeal sRNP complex can well be considered a minimal RNP composed of smaller RNAs, three core proteins, with spatially and functionally coupled box C/D and C'/D' RNPs. Known RNA targets are confined to ribosomal and transfer RNAs and its only function appears to be nucleotide modification. The sRNAs are directly transcribed from intergenic genes and assembly of the sRNP does not require, at least in vitro, accessory proteins. In contrast, the eukaryotic snoRNP is more complex both structurally and functionally. It is composed of larger RNAs, one additional core protein resulting from gene duplication, with poorly conserved C'/D' RNPs that do not appear to be spatially linked to the box C/D RNP. SnoRNP target RNAs are more diverse and RNP functions include rRNA folding and processing as well as nucleotide modification. The snoRNA genes are varied in genomic organization, often transcribed as introns and snoRNA processing is essential with RNP assembly requiring numerous assembly factors.

In this chapter, we have presented the current state of knowledge concerning the structure and function of the box C/D RNPs. Our focus has been comparison of the archaeal and eukaryotic complexes, detailing their differences to provide the reader with an overview of the evolving box C/D RNP complexes. More remains to be learned about these ancient enzymes and the coming years are certain to yield exciting and unexpected findings.

Acknowledgements

The authors would like to thank Skip Fournier, Mike and Becky Terns, Tom Meier and Yi-Tao Yu for helpful comments on our chapter. This work was supported by NSF Grant MCB 0543741 to ESM.

References

1. Helm M. Post-transcriptional nucleotide modification and alternative folding of RNA. Nucl Acids Res 2006; 34:721-733.
2. Chow C, Lamichhane TN, Mahto SK. Expanding the nucleotide repertoire of the ribosome with post-transcriptional modifications. ACS Chem Biol 2007; 2:610-619.
3. Tollervey D, Lehtonen H, Jansen R et al. Temperature-sensitive mutations demonstrate roles for yeast fibrillarin in prerRNA processing, prerRNA methylation and ribosome assembly. Cell 1993; 72:443-457.
4. Decatur W, Fournier MJ. rRNA modifications and ribosome function. Trends Biochem Sci 2002; 27(7):344-351.
5. Liang X, Hury A, Hoze E et al. 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 2007; 6(3):361-377.
6. Esguerra J, Warringer J, Blomberg A. Functional importance of individual rRNA 2′-O-ribose methylations revealed by high-resolution phenotyping. RNA 2008; 14:649-656.
7. Darzacq X, Jady BE, Verheggen C et al. Cajal body-specific small nuclear RNAs: a novel class of 2′-O-methylation and pseudouridylation guide RNAs. EMBO J 2002; 21(11):2746-2756.
8. Beal P, Maydanovych O, Pokharel S. The chemistry and biology of RNA editing by adenosine deaminases. Nucl Acids Symp Ser 2007; 51:83-84.
9. Kishore S, Stamm S. The snoRNA HBII-52 regulates alternative splicing of the serotonin receptor 2C. Science 2006; 311(5758):230-232.
10. Bazeley P, Shepelev V, Talebizadeh Z et al. SnoTARGET shows that human orphan snoRNA targets locate close to alternative splice junctions. Gene 2008; 408:172-179.

11. Dennis PP, Omer A, Lowe T. A guided tour: small RNA function in Archaea. Mol Microbiol 2001; 40:509-519.
12. Singh SK, Gurha P, Tran EJ et al. Sequential 2′-O-methylation of Archaeal pretRNA^{Trp} nucleotides is guided by the intron-encoded but trans-acting box C/D ribonucleoprotein of pretRNA. J Biol Chem 2004; 279(46):47661-47671.
13. Hughes JM, Ares Jr, M. Depletion of U3 small nucleolar RNA inhibits cleavage in the 5′ external transcribed spacer of yeast preribosomal RNA and impairs formation of 18S ribosomal RNA. EMBO J 1991; 10(13):4231-4239.
14. Morrissey JP, Tollervey D. Yeast snR30 is a small nucleolar RNA required for 18S rRNA synthesis. Mol Cell Biol 1993; 13(4):2469-2477.
15. Liang WQ, Fournier MJ. U14 base-pairs with 18S rRNA: a novel snoRNA interaction required for rRNA processing. Genes Dev 1995; 9(19):2433-2443.
16. Peculis B, Steitz J. Disruption of U8 nucleolar snRNA inhibits 5.8S and 28S rRNA processing in the Xenopus oocyte. Cell 1993; 73:1233-1245.
17. Beltrame M, Tollervey D. Base pairing between U3 and the preribosomal RNA is required for 18S rRNA synthesis. EMBO J 1995; 14(17):4350-4356.
18. Schoemaker RJ, Gultyaev AP. Computer simulation of chaperone effects of Archaeal C/D box sRNA binding on rRNA folding. Nuc. Acids Res 2006; 34(7):2015-2026.
19. Omer AD, Lowe TM, Russell AG et al. Homologs of small nucleolar RNAs in Archaea. Science 2000; 288(5465):517-522.
20. Gaspin C, Cavaille J, Erauso G et al. Archaeal homologs of eukaryotic methylation guide small nucleolar RNAs: lessons from the Pyrococcus genomes. J Mol Biol 2000; 297:895-906.
21. Huttenhofer A, Cavaille J, Bachellerie JP. Experimental RNomics: a global approach to identifying small nuclear RNAs and their targets in different model organisms. Methods Mol Biol 2004; 265:409-428.
22. Schattner P, Brooks AN, Lowe TM. The tRNAscan-SE, snoscan and snoGPS web servers for the detection of tRNAs and snoRNAs. Nuc Acids Res 2005; 33(web server):W686-W689.
23. Tran E, Brown J, Maxwell ES. Evolutionary origins of the RNA-guided nucleotide-modification complexes: from the primitive translation apparatus? Trends Biochem Sci 2004; 29:343-350.
24. Lestrade L, Weber MJ. snoRNA-LBME-db, a comprehensive database of human H/ACA and C/D box snoRNAs. Nuc Acids Res 2006; 34(database):D158-D162.
25. Samarsky DA, Fournier MJ. A comprehensive database for the small nucleolar RNAs from Saccharomyces cerevisiae. Nuc Acids Res 1999; 27(1):161-164.
26. Piekna-Przybylska D, Decatur WA, Fournier MJ. New bioinformatic tools for analysis of nucleotide modifications in eukaryotic rRNA. RNA 2007; 13(3):305-312.
27. Cavaille J, Buiting K, Keifmann M et al. Identification of brain-specific and imprinted small nucleolar RNA genes exhibiting an unusual genomic organization. Proc Natl Acad Sci USA 2000; 97(26):14311-14316.
28. Nahkuri S, Taft RJ, Korbie DJ et al. Molecular evolution of the HBII-52 snoRNA cluster. J Mol Biol 2008; 381:810-815.
29. Brown JW, Echeverria M, Qu LH et al. Plant snoRNA database. Nuc Acids Res 2003; 31(1):432-435.
30. Kiss-Laszlo Z, Henry Y, Kiss T. Sequence and structural elements of methylation guide snoRNAs essential for site-specific ribose methylation of prerRNA. EMBO J 1998; 17(3):797-807.
31. Klein DJ, Schmeing TM, Moore PB et al. The kink-turn: a new RNA secondary structure motif. EMBO J 2001; 20:4214-4221.
32. Vidovic I, Nottrot S, Hartmuth K et al. Crystal structure of the spliceosomal 15.5kD protein bound to a U4 snRNA fragment. Mol Cell 2000; 6:1331-1342.
33. Nolivos S, Carpousis AJ, Clouet-d'Orval B. The K-loop, a general feature of the Pyrococcus C/D guide RNAs, is an RNA structural motif related to the K-turn. Nucl Acids Res 2005; 33:6507-6514.
34. Mao H, White SA, Williamson JR. A novel loop-loop recognition motif in the yeast ribosomal protein L30 autoregulatory RNA complex. Nat Struct Biol 1999; 6:1139-1147.
35. Clery A, Bourguignon-Igel V, Allmang C et al. An improved definition of the RNA-binding specificity of SECIS-binding protein 2, an essential component of the selenocysteine incorporation machinery. Nucl Acids Res 2007; 35:1868-1884.
36. Montange RK, Batey RT. Structure of the S-adenosylmethionine riboswitch regulatory mRNA element. Nature 2006; 441:1172-1175.
37. Kiss-Laszlo Z, Henry Y, Bachellerie M et al. Site-specific ribose methylation of preribosomal RNA: a novel function for small nucleolar RNAs. Cell 1996; 85(7):1077-1088.
38. Tran EJ, Zhang X, Lackey L et al. Conserved spacing between the box C/D and C'/D' RNPs of the archaeal box C/D sRNP complex is required for efficient 2′-O-methylation of target RNAs. RNA 2005; 11(3):285-293.

39. Starostina NG, Marshburn S, Johnson LS et al. Circular box C/D RNAs in Pyrococcus furiosus. Proc Natl Acad Sci 2004; 101:14097-14101.
40. Wu P, Brockenbrough JS, Metcalfe AC et al. Nop5p is a small nucleolar ribonucleoprotein component required for pre18S rRNA processing in yeast. J Biol Chem 1998; 273:16453-16463.
41. Newman DR, Kuhn JF, Shanab GM et al. Box C/D snoRNA-associated proteins: two pairs of evolutionarily ancient proteins and possible links to replication and transcription. RNA 2000; 6:861-879.
42. Watkins NJ, Segault V, Charpentier B et al. A common core RNP structure shared between the small nucleolar box C/D RNPs and the spliceosomal U4 snRNP. Cell 2000; 103:457-466.
43. Kuhn J, Tran E, Maxwell ES. Archaeal ribosomal protein L7 is a functional homolog of the eukaryotic 15.5kD/Snu13p snoRNP core protein. Nuc Acids Res 2002; 30:931-941.
44. Omer A, Ziesche S, Ebhardt H et al. In vitro reconstitution and activity of a C/D box methylation guide ribonucleoprotein complex. Proc Natl Acad Sci USA 2002; 99:5289-5294.
45. Tran EJ, Zhang X, Maxwell ES. Efficient RNA 2′-O-methylation requires juxtaposed and symmetrically assembled archaeal box C/D and C'/D' RNPs. EMBO J 2003; 22:3930-3940.
46. Hardin JW, Batey RT. The bipartite architecture of the sRNA in an archaeal box C/D complex is a primary determinant of specificity. Nucl Acids Res 2006; 34:5039-5051.
47. Turner B, Melcher SA, Wilson TJ et al. Induced fit of RNA on binding the L7Ae protein to the kink-turn motif. RNA 2005; 11:1192-1200.
48. Gagnon KT, Zhang X, Agris PF et al. Assembly of the archaeal box C/D sRNP can occur via alternative pathways and requires temperature-facilitated sRNA remodeling. J Mol Biol 2006; 362:1025-1042.
49. Cahill NM, Friend K, Speckman W et al. Site-specific cross-linking analyses reveal an asymmetric protein distribution for a box C/D snoRNP. EMBO J 2002; 21:3816-3828.
50. Szewczak LB, DeGregorio SJ, Strobel SA et al. Exclusive interaction of the 15.5 kD protein with the terminal box C/D motif of a methylation guide snoRNP. Chem Biol 2002; 9:1095-1107.
51. Watkins NJ, Lemm I, Ingelfinger D et al. Assembly and maturation of the U3 snoRNP in the nucleoplasm in a large dynamic multiprotein complex. Mol Cell 2004; 16:789-798.
52. King T, Decatur WA, Bertrand E et al. A well-connected and conserved nucleoplasmic helicase is required for production of box C/D and H/ACA snoRNAs and localization of snoRNP proteins. Mol Cell Biol 2001; 21:7731-7746.
53. Boulon S, Verheggen C, Jady BE et al. PHAX and CRM1 are required sequentially to transport U3 snoRNA to nucleoli. Mol Cell 2004; 16:777-787.
54. McKeegan KS, Debieux CM, Boulon S et al. A dynamic scaffold of presnoRNP factors facilitates human box C/D snoRNP assembly. Mol Cell Biol 2007; 27:6782-6793.
55. Boulon S, Marmier-Gourrier N, Pradet-Balade B et al. The Hsp90 chaperone controls the biogenesis of L7Ae RNPs through conserved machinery. J Cell Biol 2008; 180:579-595.
56. Zhao R, Kakihara Y, Gribun A et al. Molecular chaperone Hsp90 stabilizes Pih1/Nop17 to maintain R2TP complex activity that regulates snoRNA accumulation. J Cell Biol 2008; 180:563-578.
57. Bachellerie JP, Cavaille J, Huttenhofer A. The expanding snoRNA world. Biochimie 2002; 84:775-790.
58. Hirose T, Shu MD, Steitz JA. Splicing-dependent and -independent modes of assembly for intron-encoded box C/D snoRNPs in mammalian cells. Mol Cell 2003; 12:113-123.
59. Kiss T, Fayet E, Jady BE et al. Biogenesis and intranuclear trafficking of human box C/D and H/ACA RNPs. Cold Spring Harb Symp Quant Biol 2006; 71:407-417.
60. Liu S, Li P, Dybkov O et al. Binding of the human Prp31 Nop domain to a composite RNA-protein platform in U4 snRNP. Science 2007; 316:115-120.
61. Charron C, Manival X, Clery A et al. 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 2004; 342:757-773.
62. Koonin EV, Bork P, Sander C. A novel RNA-binding motif in omnipotent suppressors of translation termination, ribosomal proteins and a ribosome modification enzyme? Nucl Acids Res 1994; 22:2166-2167.
63. Moore T, Zhang Y, Fenley MO et al. Molecular basis of box C/D RNA-protein interactions: cocrystal structure of archaeal L7Ae and a box C/D RNA. Structure 2004; 12:807-818.
64. Welting TJM, van Venrooij WJ, Pruijn GJM. Mutual interactions between subunits of the human RNase MRP ribonucleoprotein complex. Nucl Acids Res 2004; 32:2138-2146.
65. Allmang C, Carbon P, Krol A. The SBP2 and 15.5 kD/Snu13p proteins share the same RNA binding domain: identification of SBP2 amino acids important to SECIS RNA binding. RNA 2002; 8:1308-1318.
66. Wang C, Meier UT. Architecture and assembly of mammalian H/ACA small nucleolar and telomerase ribonucleoproteins. EMBO J 2004; 23:1857-1867.

67. Aittaleb M, Rashid R, Chen Q et al. Structure and function of Archaeal box C/D sRNP core proteins. Nat Struct Biol 2003; 10:256-263.
68. Gautier T, Berges T, Tollervey D et al. Nucleolar KKE/D repeat proteins Nop56p and Nop58p interact with Nop1p and are required for ribosome biogenesis. Mol Cell Biol 1997; 17:7088-7098.
69. Oruganti S, Zhang Y, Li H et al. Alternative conformations of the archaeal Nop56/58-Fibrillarin complex imply flexibility in box C/D RNPs. J Mol Biol 2007; 371:1141-1150.
70. Aittaleb M, Visone T, Fenley MO et al. Structural and thermodynamic evidence for a stabilizing role of Nop5p in S-adenosyl-L-methionine binding to fibrillarin. J Biol Chem 2004; 279:41822-41829.
71. Liu S, Rauhut R, Vornlocher H-P et al. The network of protein—protein interactions within the human U4/U6.U5 tri-snRNP. RNA 2006; 12:1418-1430.
72. Fatica A, Galardi S, Altieri F et al. Fibrillarin binds directly and specifically to U16 box C/D snoRNA. RNA 2000; 6:88-95.
73. Lafontaine DL, Tollervey D. Synthesis and assembly of the box C+D small nucleolar RNPs. Mol Cell Biol 2000; 20:2650-2659.
74. Niewmierzycka A, Clarke S. S-Adenosylmethionine-dependent methylation in Saccharomyces cerevisiae. Identification of a novel protein arginine methyltransferase. J Biol Chem 1999; 274:814-824.
75. Deng L, Starostina NG, Liu ZJ et al. Structure determination of fibrillarin from the hyperthermophilic archaeon Pyrococcus furiosus. Biochem Biophys Res Comm 2004; 315:726-732.
76. Wang H, Boisvert D, Kim K et al. Crystal structure of a fibrillarin homologue from Methanococcus jannaschii, a hyperthermophile, at 1.6Å resolution. EMBO J 2000; 19(3):317-323.
77. Snaar S, Wiesmeijer K, Jochemsen AG et al. Mutational analysis of fibrillarin and its mobility in living human cells. J Cell Biol 2000; 151:653-662.
78. Amiri KA. Fibrillarin-like proteins occur in the domain Archaea. J Bacteriol 1994; 176:2124-2127.
79. Feder M, Pas J, Wyrwicz LS et al. Molecular phylogenetics of the RrmJ/fibrillarin superfamily of ribose 2'-O-methyltransferases. Gene 2003; 302:129-138.
80. Hager J, Staker BL, Bugl H et al. Active site in RrmJ, a heat shock-induced methyltransferase. J Biol Chem 2002; 277:41978-41986.
81. Klimasauskas S, Kumar S, Roberts RJ et al. HhaI methyltransferase flips its target base out of the DNA Helix. Cell 1994; 76:357-369.
82. Jones KW, Gorzynski K, Hales CM et al. Direct interaction of the spinal muscular atrophy disease protein SMN with the small nucleolar RNA-associated protein fibrillarin. J Biol Chem 2001; 276:38645-38651.
83. Watkins NJ, Dickmanns A, Luhrmann R. Conserved stem II of the box C/D motif is essential for nucleolar localization and is required, along with the 15.5K protein, for the hierarchical assembly of the box C/D snoRNP. Mol Cell Biol 2002; 22:8342-8352.