View metadata, citation and similar papers at core.ac.uk
brought to you by CORE
provided by Elsevier - Publisher Connector

Cell, Vol. 109, 145–148, April 19, 2002, Copyright ©2002 by Cell Press

Small Nucleolar RNAs: An Abundant Group of Noncoding RNAs with Diverse Cellular Functions

Minireview

Tamás Kiss[1]
Laboratoire de Biologie Moléculaire Eucaryote du CNRS
Université Paul Sabatier
118 route de Narbonne
31062 Toulouse Cedex
France
Biological Research Center
Hungarian Academy of Sciences
Szeged
Hungary

Small nucleolar RNAs represent an abundant, evolutionarily ancient group of noncoding RNAs which possess impressively diverse functions ranging from 2′-O-methylation and pseudouridylation of various classes of RNAs, through nucleolytic processing of rRNAs to the synthesis of telomeric DNA.

The nucleolus, the most prominent organelle in the interphase nucleus, provides the cellular locale for the synthesis and processing of cytoplasmic ribosomal RNAs (rRNAs). Besides precursor rRNAs (pre-rRNAs) at various stages of processing, the nucleolus also contains an enormous number of 60 to 300 nucleotide long, metabolically stable RNAs, called small nucleolar RNAs (snoRNAs), which associate with a set of proteins to form small nucleolar RNPs (snoRNPs). During the last decade, studies on snoRNAs have revealed many novel, unexpected cellular functions for noncoding RNAs and changed long-held principles about eukaryotic gene expression. Several hundreds of snoRNAs have been reported from a broad variety of organisms, making snoRNAs the most abundant group of noncoding RNAs. Some snoRNAs play essential roles in the nucleolytic processing of rRNAs, but the majority of them function as guide RNAs in the post-transcriptional synthesis of 2′-O-methylated nucleotides and pseudouridines in rRNAs, small nuclear RNAs (snRNAs) and probably other cellular RNAs, including even mRNAs. Another vertebrate snoRNA-like RNA, the telomerase RNA, directs the synthesis of telomeric DNA that provides stability for chromosomes. The biogenesis of snoRNAs is also interesting. Instead of being transcribed from independent genes, most vertebrate snoRNAs are processed from introns of pre-mRNAs, demonstrating that introns may code for functional RNAs.

Two Major Classes of snoRNAs

After identification of several dozens of snoRNAs, it became apparent that they fall into two major classes which posses distinctive, evolutionarily conserved sequence elements (reviewed by Kiss, 2001). One group of snoRNAs contains the box C (RUGAUGA) and D (CUGA) motifs, whereas members of the other family carry the box H (ANANNA) and ACA elements (Figure 1). In both classes of snoRNAs, short stems bring the conserved boxes close to one another to constitute the structural core motifs of the snoRNAs (indicated in red), which coordinate the specific binding of two distinct sets of snoRNP proteins. The box C/D core motif directs the binding of fibrillarin, Nop56p, Nop58p, and 15.5 kDa/Snu13p snoRNP proteins. The 15.5 kDa/Snu13 protein interacts directly with the box C/D motif that forms a kink-turn (K-turn) secondary structure motif, a widespread protein binding motif found in many RNAs (Vidovic et al., 2000; Klein et al., 2001). The box H/ACA snoRNAs are associated with dyskerin/Cbf5p, Gar1p, Nhp2p, and Nop10p. It is still uncertain which of these snoRNP proteins bind(s) directly to the H/ACA core structure.

The Majority of snoRNAs Function in 2′-O-methylation and Pseudouridylation of Various Classes of RNAs

Biogenesis of functional rRNAs, tRNAs, and snRNAs includes the posttranscriptional covalent modification of many carefully selected ribonucleotides. The modified nucleotides contribute to the correct function of tRNAs, rRNAs, and snRNAs, although in most cases their precise function is still speculative. Site-specific synthesis of the most abundant modified nucleotides, the 2′-O-methylated nucleotides and pseudouridines that are found in diverse sequence and structural contexts in rRNAs and snRNAs, is directed by snoRNAs. While methylation of the 2′-hydoxyl groups of the correct target nucleotides is directed by box C/D snoRNAs, conversion of uridines to pseudouridine is guided by box H/ACA snoRNAs. Both classes of guide snoRNAs specify the sites of modification by forming direct base pairing interactions with substrate RNAs (Figure 1). The 2′-O-methylation guide snoRNAs establish a long (10–21 bp) helix, whereas the pseudouridylation guide snoRNAs form two short (3–10 bp) duplexes with the appropriate target sequence. In the interaction of the snoRNA and the substrate RNA, the target nucleotide destined for 2′-O-methylation or pseudouridylation occupies an invariant position relative to the conserved box elements of the snoRNAs. The 2′-O-methylated nucleotides are located 5 nucleotides upstream of the D or D′ box of the snoRNA and the pseudouridine residues are found about 15 nucleotides upstream of the H or ACA box. The snoRNP proteins utilize this structural information to select the correct target nucleotide. It is almost certain that the 2′-O-methyl transfer and the uridine-to-pseudouridine isomerization reactions are catalyzed by the fibrillarin and dyskerin/Cbf5 snoRNP proteins, respectively (Wang et al., 2000; Hoang and Ferré-D'Amaré, 2001).

Guide snoRNAs function in 2′-O-methylation and pseudouridylation of the RNA polymerase (pol) I-transcribed 18S, 5.8S, and 28S rRNAs and the pol II- and pol III-specific spliceosomal snRNAs (Tycowski et al., 1998; Jády and Kiss, 2001; Kiss, 2001). In rRNAs and the pol III-specific U6 spliceosomal snRNA, most, if not all, 2′-O-methylated nucleotides and pseudouridines are synthesized by snoRNPs (Kiss, 2001; Ganot et al., 1999).

[1]Correspondence: tamas@ibcg.biotoul.fr

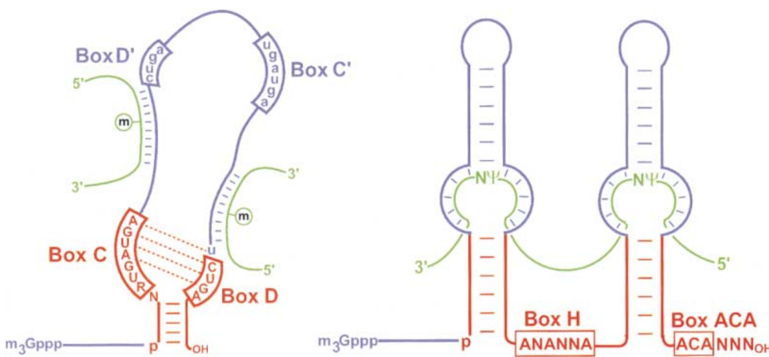

Figure 1. Structure and Function of Box C/D and Box H/ACA snoRNAs

In 2′-O-methylation guide snoRNAs, the box C and D motifs and a short 5′, 3′-terminal stem constitute a kink-turn (K-turn) structural motif that is specifically recognized by the 15.5 kDa snoRNP protein. The C′ and D′ boxes represent internal, frequently imperfect copies of the C and D boxes. Dashed lines indicate nucleotides interacting in the C and D boxes (Watkins et al., 2000; Klein et al., 2001). The pseudouridylation guide snoRNAs fold into a "hairpin-hinge-hairpin-tail" structure and contain the H and ACA boxes. The box C/D 2′-O-methylation guide snoRNAs and the substrate RNAs form a 10–21 base pair double helix in which the target residue is positioned exactly five nucleotides upstream of the D or D′ box. The 5′ and/or 3′ hairpin of the box H/ACA pseudouridylation guide snoRNAs contains an internal loop, called the pseudouridylation pocket, that forms two short (3–10 bp) duplexes with nucleotides flanking the unpaired substrate uridine that is located about 15 nucleotides from the H or ACA box of the snoRNA. Although each box C/D and H/ACA snoRNA could potentially direct two modification reactions, apart from a few exceptions, the majority of snoRNAs possess only one functional 2′-O-methylation or pseudouridylation domain. The snoRNA core motifs that are essential and sufficient for the correct processing and nucleolar accumulation of snoRNAs are highlighted in red. Regions that do not contribute to the metabolic stability of snoRNAs are blue. SnoRNAs transcribed independently by RNA pol II contain 5′ leader sequences and carry the trimethylguanosine cap structure (m_3Gppp). Substrate RNAs are green. Nucleotides destined for pseudouridylation (Ψ) and 2′-O-methylation (circled m) are marked.

To what extent snoRNAs participate in modification of the RNA pol II-specific U1, U2, U4, and U5 spliceosomal RNAs is still unclear. So far, 12 putative guide snoRNAs have been implicated in 2′-O-methylation and/or pseudouridylation of mammalian pol II-specific snRNAs, suggesting that modification of spliceosomal snRNAs is mostly achieved by snoRNPs (Jády and Kiss, 2001; Hüttenhofer et al., 2001; Kiss, 2001).

Recently, it has been shown that the pol III-transcribed U6 snRNA transiently travels through the nucleolus, presumably, to undergo snoRNA-mediated modifications (Lange and Gerbi, 2000). In contrast to U6 that does not leave the nucleus during its maturation, the pol II-synthesized U1, U2, U4, and U5 snRNAs are exported to the cytoplasm where they associate with Sm proteins and undergo 3′ end processing and cap hypermethylation. The newly assembled snRNPs are reimported into the nucleus and before accumulating in speckles, they transiently associate with conserved nucleoplasmic organelles, the Cajal (coiled) bodies (Sleeman and Lamond, 1999). Surprisingly, all guide snoRNAs demonstrated or predicted to direct modification of the RNA pol II-specific U1, U2, U4, and U5 snRNAs, instead of residing in the nucleolus, have been found to specifically accumulate in Cajal bodies (Kiss, 2001). This strongly suggests that modification of pol II-specific spliceosomal snRNAs may occur in Cajal bodies. The finding that some box C/D and H/ACA snoRNAs can accumulate and probably function outside of the nucleolus clearly indicates that we have to recalibrate our previous "nucleolus-centric" thinking about snoRNAs.

Functional Diversity of snoRNAs

Typically, snoRNAs are processed from longer precursor RNAs by exonucleolytic activities. While most vertebrate snoRNAs are excised from liberated introns of pre-mRNAs, yeast and plant snoRNAs are frequently processed from long polycistronic snoRNA transcripts (Weinstein and Steitz, 1999). The snoRNP proteins bound to the box C/D and H/ACA core motifs direct the correct 5′ and 3′ end formation of precursor snoRNAs and provide metabolic stability for the mature snoRNAs. The internal regions of box C/D or the distal parts of the 5′ and 3′ hairpins of box H/ACA snoRNAs encompassing the modification guide sequences have little or no effect on snoRNA accumulation (Figure 1; regions indicated in blue). These regions can therefore evolve rapidly by accumulating point mutations during evolution. This provides a great evolutionarily flexibility and functional diversity for the snoRNA-guided RNA modification systems.

Those regions of box C/D and H/ACA snoRNAs that are not crucial for accumulation may also contain *trans*-acting elements required for the nucleolytic processing rather than 2′-O-methylation or pseudouridylation of the 18S, 5.8S, and 28S rRNAs. The U3, U14, U22 box C/D and the snR30 box H/ACA snoRNAs are essential for production of the 18S rRNA, whereas the U8 box C/D snoRNA is required for the excision of the 5.8S and 28S rRNAs from the pre-rRNA (reviewed by Yu et al., 1999). Since none of the U3, U14, U22, snR30, and U8 snoRNPs have been demonstrated to possess endonucleolytic or exonucleolytic activities, they are assumed to orchestrate the appropriate structural reorganization of the pre-rRNA during processing.

Vertebrate Telomerase Is a Box H/ACA snoRNP

Telomerase is an RNP reverse transcriptase that maintains telomere length by adding telomeric DNA repeats onto the ends of eukaryotic chromosomes. The RNA component of the enzyme contains a short template motif that is copied into telomeric DNA by the associated telomerase reverse transcriptase (TERT). In yeast, telomerase is associated with Sm proteins like spliceosomal snRNPs, but in vertebrates, it is in fact a box H/ACA snoRNP. While the 5′-terminal template region of vertebrate telomerase RNAs (TRs) shares structural properties with TRs from lower eukaryotes, the 3′-terminal region folds into the characteristic "hairpin-hinge-hairpin-tail" structure of box H/ACA snoRNAs and binds the four box H/ACA snoRNP core proteins. The box H/ACA domain, besides directing 3′ end formation and providing stability for

TR, also contains elements that are essential for TERT binding and therefore, for telomerase function. The functional importance of the box H/ACA domain of telomerase is further demonstrated by the finding that mutations in the box H/ACA domain of human TR or in the associated dyskerin snoRNP protein result in a multisystem genetic disease, dyskeratosis congenita (Mitchell et al., 1999; Vulliamy et al., 2001). Indeed, dyskeratosis congenital patients have significantly shorter telomers than normal individuals, indicating that pathology of this disease is consistent with compromised telomerase function.

Brain-Specific snoRNAs

Expression of vertebrate intron-encoded snoRNAs is inevitably regulated by the transcription of their host genes. SnoRNAs directing modification of rRNAs and snRNAs are encoded within introns of constitutively expressed housekeeping genes and they accumulate in all tissues (Weinstein and Steitz, 1999). Recently, several snoRNAs have been reported that are expressed exclusively or predominantly in the central nervous system (Cavaillé et al., 2000; Runte et al., 2001). At least six human brain-specific box C/D snoRNAs are encoded within introns of an extremely complex, paternally imprinted transcription unit, IC-*SNURF-SNRPN*, that spans more than 460 kb and contains at least 150 exons (Runte et al., 2001). While exons 1–3 and 4–10 encode the SNURF protein and the SmN spliceosomal protein, respectively, the downstream exons lack apparent open reading frames. Instead, the downstream introns encode at least six box C/D snoRNAs. Unusually, two snoRNAs, called MBII-85 and MBII-52, are encoded in 24 and 47 tandemly arranged exon-intron repeats. The paternally imprinted IC-*SNURF-SNRPN* locus has been implicated in a serious neurogenetic disorder, the Prader-Willi syndrome (PWS). In mouse, paternally inherited deletions of the IC-*SNURF-SNRPN* locus result in symptoms highly reminiscent of PWS, raising the intriguing possibility that the snoRNAs encoded in this region may be involved in PWS.

None of the brain-specific snoRNAs show sequence complementarity to rRNAs, snRNAs, or other stable RNAs. However, the MBII-52 box C/D snoRNA carries a conserved 18 nt long sequence that is complementary to the mouse and human serotonin receptor 5-HT_{2C} mRNA (Cavaillé et al., 2000). The MBII-52 snoRNA could direct 2′-O-methylation of an adenosine residue that is known to be partially deaminated to inosine in the serotonin receptor mRNA. Since 2′-O-methylation inhibits adenosine deamination, the MBII-52 snoRNA might have an important regulatory function. Interestingly, the second intron of the human serotonin receptor 5-HT_{2C} gene also encodes a brain-specific box H/ACA snoRNA of unknown function (Cavaillé et al., 2000).

Evolutionary Origin of the snoRNA-Mediated Modification Systems

It is hypothesized that contemporary life forms arose from a primordial RNA world in which RNA molecules both stored genetic information and catalyzed biochemical reactions. Synthesis of the four 2′-O-methylated nucleotides and the eleven pseudouridines in *E. coli* rRNAs, however, relies exclusively on site-specific protein enzymes and no evidence supports the existence of modification guide RNAs in Eubacteria. This may suggest that the snoRNA-targeted modification mechanism evolved from a protein-based system rather than a prebiotic RNA world. It has been found that box C/D snoRNA-like 2′-O-methylation guide RNAs and homologs of fibrillarin, Nop56/58p, and 15.5 kDa snoRNP proteins are present in prokaryotic Archaea, indicating that the snoRNA-based modification system was established about 2 to 3 billion years ago already in the common ancestor of Archaea and Eucarya (Omer et al., 2000; Kuhn et al., 2002). Besides directing modification of rRNAs, archaeal guide RNAs also function in 2′-O-methylation of some tRNAs (Omer et al., 2000). The archaeal 2′-O-methylation guide RNAs exhibit all hallmarks of eukaryotic box C/D snoRNAs. They possess the conserved C, D, C′, and D′ boxes, guide regions complementary to rRNAs, and they select the target nucleotides on the basis of the "box D/D′ plus five" rule. The enormous structural and functional similarity of eukaryotic and archaeal box C/D snoRNPs indicates that molecular evolution of the snoRNA-based modification systems is highly constrained by the fact that it requires simultaneous covariation of dozens or even hundreds of snoRNAs with the snoRNP core proteins.

Most probably, primordial guide snoRNAs derived from *cis*-acting rRNA or tRNA sequences which acquired the ability to function as *trans*-acting factors. The L7 protein, the archaeal homolog of eukaryotic 15.5 kDa/Snu13p box C/D snoRNP core protein seems to have a dual function. It binds to the core motif of box C/D snoRNAs as well as to the 23S large rRNA (Kuhn et al., 2002). The mammalian 15.5 kDa protein and its yeast (Snu13p) and archaeal (L7) homologs share sequence similarity with ribosomal proteins L30 and S12, which all bind to a common, evolutionarily widespread RNA structural motif, the K-turn (Klein et al., 2001). Taken together, these findings lend further support to the idea that box C/D snoRNPs might have evolved from primordial ribosomes. Intriguingly, the Nhp2p core protein of box H/ACA snoRNPs also shows significant similarity with the L30 and S12 ribosomal proteins and the 15.5 kDa/Snu13p/L7 box C/D snoRNP protein, although the mode of interaction of Nhp2p with box H/ACA snoRNAs remains unknown. Finally, the mammalian 15.5 kDa protein also binds to the U4 spliceosomal snRNA, raising the possibility that the U4 snRNP might have evolved from a primordial box C/D snoRNP (Watkins et al., 2000).

Perspectives

Recently, many "orphan" box C/D and H/ACA snoRNAs have been identified, which lack complementarities to rRNAs, snRNA, tRNAs, or other known stable RNAs (Hüttenhofer et al., 2001; Kiss, 2001). This may indicate that the substrate repertoire of 2′-O-methylation and pseudouridylation guide snoRNAs is much broader than has been demonstrated to date. Of course, it is also possible that some "orphan" snoRNAs function in cellular processes other than RNA modification. It seems that only a fraction of cellular snoRNAs, mostly those present in high copies in the cell, have been identified so far. We can thus envisage the discovery of many novel snoRNAs. Exploring the functional diversity of the unexpectedly complex world of snoRNAs will undoubtedly be a great challenge for the future.

Selected Reading

Cavaillé, J., Buiting, K., Kiefmann, M., Lalande, M., Brannan, C.I., Horsthemke, B., Bachellerie, J.P., Brosius, J., and Hüttenhofer, A. (2000). Proc. Natl. Acad. Sci. USA *97*, 14311–14316.

Ganot, P., Jády, B.E., Bortolin, M.-L., Darzacq, X., and Kiss, T. (1999). Mol. Cell. Biol. *19*, 6906–6917.

Hoang, C., and Ferré-D'Amaré, A.R. (2001). Cell *107*, 929–939.

Hüttenhofer, A., Kiefmann, M., Meier-Ewert, S., O'Brien, J., Lehrach, H., Bachellerie, J.P., and Brosius, J. (2001). EMBO J. *20*, 2943–2953.

Jády, B.E., and Kiss, T. (2001). EMBO J. *20*, 541–551.

Kiss, T. (2001). EMBO J. *20*, 3617–3622.

Klein, D.J., Schmeing, T.M., Moore, P.B., and Steitz, T.A. (2001). EMBO J. *20*, 4214–4221.

Kuhn, J.F., Tran, E.J., and Maxwell, E.S. (2002). Nucleic Acids Res. *30*, 931–941.

Lange, T.S., and Gerbi, S.A. (2000). Mol. Biol. Cell *11*, 2419–2428.

Mitchell, J.R., Wood, E., and Collins, K. (1999). Nature *402*, 551–555.

Omer, A.D., Lowe, T.M., Russell, A.G., Ebhardt, H., Eddy, S.R., and Dennis, P.P. (2000). Science *21*, 517–522.

Runte, M., Hüttenhofer, A., Gross, S., Kiefmann, M., Horsthemke, B., and Buiting, K. (2001). Hum. Mol. Genet. *10*, 2687–2700.

Sleeman, J.E., and Lamond, A.I. (1999). Curr. Biol. *9*, 1065–1074.

Tycowski, K.T., You, Z.H., Graham, P.J., and Steitz, J.A. (1998). Mol. Cell *2*, 629–638.

Vidovic, I., Nottrott, S., Hartmuth, K., Lührmann, R., and Ficner, R. (2000). Mol. Cell *6*, 1331–1342.

Vulliamy, T., Marrone, A., Goldman, F., Dearlove, A., Bessler, M., Mason, P.J., and Dokal, I. (2001). Nature *413*, 432–435.

Wang, H., Boisvert, D., Kim, K.K., Kim, R., and Kim, S.H. (2000). EMBO J. *19*, 317–323.

Watkins, N.J., Ségault, V., Charpentier, B., Nottrott, S., Fabrizio, P., Bachi, A., Wilm, M., Rosbach, M., Branlant, C., and Lührmann, R. (2000). Cell *103*, 457–466.

Weinstein, L.B., and Steitz, J.A. (1999). Curr. Opin. Cell Biol. *11*, 378–384.

Yu, Y.-T., Scharl, E.C., Smith, C.M., and Steitz, J.A. (1999). In The RNA World, R.F. Gesteland, T.R. Cech, and J.F. Atkins eds. (Cold Spring Harbor Laboratory Press: Cold Spring Harbor, New York), pp. 487–524.