Journal of Cellular Biochemistry 98:486–495 (2006)

The Vertebrate E1/U17 Small Nucleolar Ribonucleoprotein Particle

George L. Eliceiri*

Department of Pathology, St. Louis University School of Medicine, St. Louis, Missouri 63104

Abstract Each of the many different box H/ACA ribonucleoprotein particles (RNPs) present in eukaryotes and archaea consists of four common core proteins and one specific H/ACA small RNA, which bears the sequence elements H (ANANNA) and ACA. Most of the H/ACA RNPs are small nucleolar RNPs (snoRNPs), which are localized in nucleoli, and are one of the two major classes of snoRNPs. Most H/ACA RNPs direct pseudouridine synthesis in pre-rRNA and other RNAs. One H/ACA small nucleolar RNA (snoRNA), vertebrate E1/U17 (snR30 in yeast), is required for pre-rRNA cleavage processing that generates mature 18S rRNA. E1 snoRNA is encoded in introns of protein-coding genes, and the evidence suggests that human E1 RNA undergoes uridine insertional RNA editing. The vertebrate E1 RNA consensus secondary structure shows several features that are absent in other box H/ACA snoRNAs. The available UV-induced RNA-protein crosslinking results suggest that the E1 snoRNP is asymmetrical in vertebrate cells, in contrast to other H/ACA snoRNPs. The vertebrate E1 snoRNP in cells is surprisingly complex: (i) E1 RNA contacts directly and specifically several proteins which do not appear to be any of the H/ACA RNP four core proteins; and (ii) multiple E1 RNA sites are needed for E1 snoRNP formation, E1 RNA stability, and E1 RNA–protein direct interactions. J. Cell. Biochem. 98: 486–495, 2006.
© 2006 Wiley-Liss, Inc.

Key words: snoRNP asssembly; box H/ACA small nucleolar RNAs; ribosomal RNA precursor cleavage processing; pseudouridine formation; RNA-protein ultraviolet light-crosslinking; E1/U17 small nucleolar RNA

Small nucleolar ribonucleoprotein particles (snoRNPs) can be divided into two major families, which have either a specific box C/D small nucleolar RNA (snoRNA) or a box H/ACA snoRNA, and several core proteins that are common to all members of a given snoRNP class, and function primarily in nucleotide modifications of RNAs by base pairing. Box C/D small nuclear RNAs have the conserved sequence elements C (consensus 5′-PuUGANGA-3′, usually near the 5′ end of the RNA) and D (consensus 5′-CUGA-3′, typically near the 3′ terminus of the RNA), and often a second pair of similar elements named boxes C′and D′ [Eliceiri, 1999; Terns and Terns, 2002]. Most box C/D small nuclear RNPs reside in nucleoli and guide the 2′-O-ribose methylation of pre-rRNA. Some C/D RNPs are localized in nucleoplasmic Cajal bodies and direct the 2′-O-ribose methylation of spliceosomal small nuclear RNAs. A few C/D snoRNPs function in pre-rRNA cleavage processing [Eliceiri, 1999].

Box H/ACA small nuclear RNAs have the conserved sequence motifs ACA and box H (consensus 5′-ANANNA-3′) [Balakin et al., 1996; Ganot et al., 1997b]. Most box H/ACA small nuclear RNPs are nucleolar and direct the isomerization of specific uridines to pseudouridines in pre-rRNA [Ganot et al., 1997a; Ni et al., 1997]. Some H/ACA small RNPs reside in nucleoplasmic Cajal bodies, and guide the pseudouridylation of RNA polymerase-II-specific spliceosomal small nuclear RNAs [Richard et al., 2003; Henras et al., 2004b; Kiss et al., 2004]. Pseudouridines are synthesized only in segments of pre-rRNA that become mature rRNA, and cluster only in functionally important sites of rRNA and spliceosomal small nuclear RNAs.

One H/ACA snoRNA, vertebrate E1/U17 (snR30 in yeast), functions in pre-rRNA

Grant sponsor: National Institute of General Medical Sciences; Grant number: R01GM64729.

*Correspondence to: George L. Eliceiri, Department of Pathology, St. Louis University School of Medicine, 1402 South Grand Boulevard, St. Louis, MO 63104-1028. E-mail: eliceiri@slu.edu

Received 12 December 2005; Accepted 15 December 2005

DOI 10.1002/jcb.20821

© 2006 Wiley-Liss, Inc.

cleavage processing, and is essential for ribosome formation and cell viability [Morrissey and Tollervey, 1993; Mishra and Eliceiri, 1997; Atzorn et al., 2004]. Telomerase RNA, the template for the synthesis of chromosome ends, is a Cajal body H/ACA RNA; its H/ACA domain is required for accumulation, stability and 3′ end processing of telomerase RNA in cells, and for telomerase activity [Mitchell et al., 1999; Jady et al., 2004; Meier, 2005]. The function of other H/ACA RNAs, including some that are apparently tissue-specific, is unknown [Kiss et al., 2004; Meier, 2005]. H/ACA RNPs have been identified in archaea and eukaryotes, but not in eubacteria, strongly suggesting that they originated before the divergence of archaea and eukaryotes, approximately over 2 billion years ago.

H/ACA snoRNA SEQUENCE AND STRUCTURE

A sequence spanning E1 RNA sites 13 and 14 (Fig. 1) was first identified as an accessible segment, and was the only accessible sequence detected in E1 RNA (in its snoRNP form), by antisense oligodeoxynucleotide-targeted RNase H digestion of human cell nucleolar extracts [Rimoldi et al., 1993]. Box H/ACA snoRNAs have a consensus 5′-hairpin-hinge-hairpin-tail-3′ secondary structure, a box H (consensus sequence 5′-ANANNA-3′, site 13 in Fig. 1) in the hinge region, and the motif ACA located three nucleotide positions from the 3′end (site 22 in Fig. 1) [Balakin et al., 1996; Ganot et al., 1997b; Henras et al., 2004b]. As an example, see the vertebrate E2 RNA consensus secondary structure (Fig. 1) [Selvamurugan et al., 1997]. In contrast, the vertebrate E1 RNA consensus secondary structure shows features that are absent in other (RNA pseudouridylation guide) H/ACA RNAs: (i) two stem-loops in the 5′ half; (ii) a stem-loop in the hinge (labeled D in Fig. 1); and (iii) a double-stranded 5′ end, as shown by evolutionary co-variations in four base pairs of that stem (site 1) (Fig. 1) [Selvamurugan et al., 1997].

The intracellular localization of H/ACA snoRNAs depends on their H and ACA boxes [Lange et al., 1999; Narayanan et al., 1999; Ruhl et al., 2000]. Cajal body H/ACA small RNAs also have H and ACA boxes, but their localization is determined by the consensus sequence 5′-ugAG-3′ (named CAB box) in a stem-loop, in which adenine and guanine are highly conserved in the third and fourth nucleotide positions, respectively [Richard et al., 2003; Jady et al., 2004]. Thus, it appears that the CAB box Cajal body localization signal supersedes the box H and ACA nucleolar localization signals. However, localization of H/ACA RNAs in Cajal bodies requires the H and ACA boxes, in addition to the CAB box.

BIOSYNTHESIS OF snoRNAs

Various snoRNAs are generated by different mechanisms [Eliceiri, 1999; Terns and Terns, 2002]. Most of the vertebrate snoRNAs, including E1/U17 RNA, are encoded in introns of protein-coding genes, but some snoRNAs are processed from introns of non-protein-coding genes. Most of the yeast snoRNAs and some of the vertebrate snoRNAs originate from independent transcription units, but many plant and yeast snoRNAs are produced from polycistronic snoRNA precursors. In H/ACA snoRNAs, boxes H and ACA are essential: (i) for snoRNA correct 5′ and 3′ end formation, respectively; (ii) for snoRNA accumulation; and (iii) for efficient nucleolar localization of snoRNA [Ganot et al., 1997b; Bortolin et al., 1999; Lange et al., 1999; Narayanan et al., 1999; Ruhl et al., 2000]. H/ACA snoRNAs apparently transit through Cajal bodies for maturation, before going to the nucleolus [Terns and Terns, 2002].

The main E1/U17 snoRNA species of HeLa (human) cells has an extra nucleotide (a uridine) at position 19 that its gene lacks, strongly suggesting uridine insertional RNA editing in an intron-encoded RNA [Nag et al., 1993]. However, a uridine residue at this nucleotide position is not phylogenetically conserved in vertebrate E1 RNA [Selvamurugan et al., 1997].

FUNCTIONS OF H/ACA RNPs

There are many pseudouridine residues in eukaryotic rRNA, and they are only in functionally important segments of mature rRNA. Many H/ACA snoRNA species function as site-specific pre-rRNA pseudouridylation guides via snoRNA:pre-rRNA base pairing, which is essential for pseudouridine formation [Ganot et al., 1997a; Ni et al., 1997]. Pseudouridylation guide H/ACA RNAs have a “pseudouridylation pocket,” which consists of two sequences (3–10 bases long), one on each side of the distal

segment of an internal loop structure, that are complementary to two sequences in the target RNA which are separated by two unpaired nucleotides, one of them being a pseudouridylation site on the target RNA [Ganot et al., 1997a]. The H and ACA boxes, and the 5′ and 3′ hairpin domains, are essential for rRNA pseudouridylation [Bortolin et al., 1999]. Box ACA or H lie 14–16 nucleotides downstream of the modification site in the pseudouridylation pocket; these boxes appear to be molecular measuring devices, since the distance from them deter-

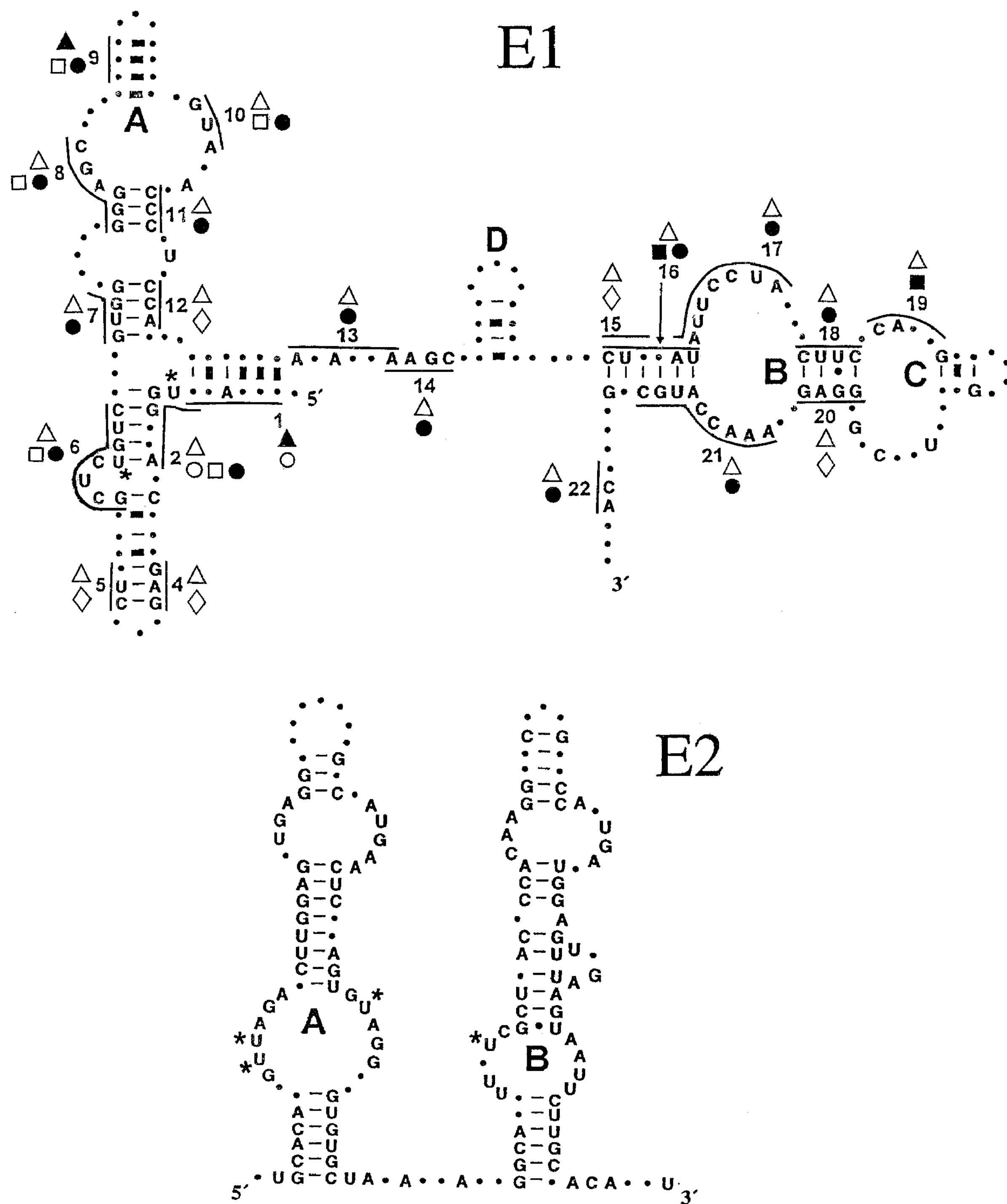

Fig. 1

mines the pseudouridylation site [Ganot et al., 1997a; Ni et al., 1997].

E1/U17 snoRNA is a box H/ACA snoRNA [Kiss and Filipowicz, 1993; Ruff et al., 1993], and is the vertebrate homolog of yeast snR30 snoRNA [Atzorn et al., 2004]. E1 RNA lacks a pre-rRNA pseudouridylation pocket that is present in all of the rRNA pseudouridylation guide snoRNAs [Ganot et al., 1997a]. In contrast to the other H/ACA snoRNAs, E1 RNA is required for pre-rRNA cleavage processing that produces mature 18S rRNA [Mishra and Eliceiri, 1997; Atzorn et al., 2004]. Sites 17 and 21 of E1 RNA (Fig. 1) were first identified as conserved sequences with conserved base pairing between the 5′-AU-3′ nucleotides at the 5′ end of site 17 and the 3′-UA-5′ residues near the 3′ end of site 21, both sequences and structure evolutionarily conserved in vertebrate E1 RNA [Mishra and Eliceiri, 1997; Selvamurugan et al., 1997; Ruhl et al., 2000]. Site 17 and all but the last two nucleotide positions of site 21 were later identified as sequences essential for 18S rRNA cleavage processing (named m1 and m2 boxes, respectively), which are conserved from yeast to vertebrates, and the m2 box is located seven nucleotide positions upstream of the ACA box [Atzorn et al., 2004]. Oligodeoxynucleotide-targeted specific RNA degradation shows that E1 RNA site 21 is one of the very few segments known to be accessible in E1 RNA (in its snoRNP form) in frog cells [Mishra and Eliceiri, 1997].

Psoralen-crosslinking in human cells first showed a direct contact: (i) of four conserved E2 RNA uridines (shown by asterisks in conserved E2 RNA loops A and B, Fig. 1, which in human E2 RNA are nucleotide positions 12, 13, 60, and 90) with nucleolar large RNA that is most likely pre-rRNA; and (ii) of E2 snoRNA with one specific site in the 28S rRNA sequence of pre-rRNA [Rimoldi et al., 1993]. This was shown later to be exactly the four E2 RNA sequences and very close to the pre-rRNA sequences which base pair with each other for the two pre-rRNA pseudouridylation reactions guided by E2 RNA [Ganot et al., 1997a]. These results emphasize the physiological significance of the direct E1 RNA–pre-rRNA interactions revealed by psoralen-crosslinking in human cells. First, psoralen-crosslinking in human cells shows direct contact of E1 RNA with two segments of pre-rRNA: (i) near the upstream end (within human nucleotide positions 697–1,163) of the 5′ external transcribed spacer; and (ii) near the middle (within human nucleotide positions 664–1,021) of the 18S rRNA sequence [Rimoldi et al., 1993]. Second, psoralen-crosslinking in human cells also demonstrates direct association of two conserved E1 RNA uridines (one at site 6 and another at site 2, which in human E1 RNA means nucleotide positions 43 and 8, respectively, shown by asterisks in Fig. 1) with nucleolar large RNA that is most likely pre-rRNA [Rimoldi et al., 1993]. The long distance of these E1 RNA-pre-rRNA contacts from the cleavage sites that generate 18S rRNA, suggest that these interactions may be part of a role of E1 RNA in folding, or as a chaperone, of pre-rRNA. Interestingly, E1 RNA site 6 is one of the very few segments known to be accessible in E1 RNA (in its snoRNP form) in frog cells [Mishra and Eliceiri, 1997].

The yeast snR30 snoRNP contains a pseudouridine synthase (Cbf5p), and E1 snoRNP, its vertebrate homolog, is assumed to contain one (NAP57), although this has not be proven experimentally. However, several observations support the conclusion that E1 RNA does not guide pseudouridine formation. First, the distal segments of internal loop structures (where pseudouridylation pockets are usually

Fig. 1. Models of vertebrate E1 and E2 RNA consensus sequences and structures, and E1 RNA sites needed for UV-induced E1 RNA-protein crosslinking, E1 snoRNP formation and E1 RNA stability in cells. The consensus sequences and structures are based on comparative sequence phylogeny, including co-variation analysis [Mishra and Eliceiri, 1997; Selvamurugan et al., 1997]; that of E1 has been revised using sequence data from Cervelli et al. [2002]. Only nucleotides conserved in vertebrate E1 and E2 RNAs are shown; the appearance of more sequence conservation in the E2 RNA model is most likely due to the fact that it is based on sequences from fewer organisms. Sequence co-variations are indicated by black rectangles. Lines span the E1 RNA sites that were detected by base substitutions [Ruhl et al., 2000; Smith et al., 2005]. Asterisks show nucleotides that can be psoralen-crosslinked in cells to nucleolar large RNA, most likely pre-rRNA [Rimoldi et al., 1993]. Conserved E1 RNA segments which are involved in E1 RNA-protein UV-crosslinking, E1 snoRNP formation, and/or E1 RNA stability in cells are apparently sequences (white triangles) or structures (black triangles). These E1 RNA sites are needed for crosslinking to proteins A (white circles), B (white squares), C (black squares) or E (black circles). The E1 RNA sites which in cells are needed for crosslinking to proteins are also necessary for E1 snoRNP formation and E1 RNA stability, except for site 1, which is not known to be required for E1 snoRNP formation [Ruhl et al., 2000; Smith et al., 2005]. Some E1 RNA sites in cells are involved in E1 snoRNP formation and E1 RNA stability, but are not known to be needed for crosslinking to proteins (white diamonds).

located) have many non-conserved nucleotides in vertebrate E1 RNA (loops labeled A–C in Fig. 1, E1 RNA). In contrast, pseudouridylation sites are flanked by conserved sequences. Second, there are no known pseudouridylation sites in any known RNA which could be potentially guided by E1 RNA. Finally, in the consensus secondary structure of vertebrate E1 RNA, sites 6 and 2 (preceding paragraph) are not in the distal segments of internal loop structures, in contrast to the pseudouridylation pockets of pseudouridylation guide snoRNAs (Fig. 1).

H/ACA snoRNA–PROTEIN INTERACTIONS

Core Proteins Common to All H/ACA RNPs

All of the tested box H/ACA RNPs, including the telomerase RNP, have the same four core proteins, which are evolutionarily highly conserved: Gar1p; yeast Cbf5p (rodent NAP57, human dyskerin); Nhp2p (ribosomal protein L7Ae in archea); and Nop10p [Henras et al., 1998; Lafontaine et al., 1998; Watkins et al., 1998; Henras et al., 2004b; Meier, 2005]. These four proteins are essential for cell viability and, except for Gar1p, they are required for box H/ACA snoRNA stability in cells. A complex, consisting of only an H/ACA RNA and the four H/ACA RNP core proteins, is apparently sufficient for RNA pseudouridylation [Wang et al., 2002]. Gar1p is a protein needed for global pseudouridylation of pre-rRNA and stable association of H/ACA snoRNAs with pre-rRNA in yeast, but does not share any conserved sequence motifs with pseudouridine synthases [Bousquet-Antonelli et al., 1997].

Cbf5p/NAP57 is a protein accepted to be a pseudouridine synthase, based on its sequence and on the effect of its mutation [Lafontaine et al., 1998; Meier, 2005]. In addition to its functions as an enzyme and a key component of H/ACA RNPs, Cbf5p/NAP57 may play other roles, such as with centromeres, nuclear cap-binding complex of pre-mRNAs and small nuclear RNAs, nuclear organization, rRNA transcription and RNA *trans*-splicing [Meier, 2005]. Mutations in the human ortholog of this protein cause the X-linked form (the most common and severe form) of the inherited disease named dyskeratosis congenita [Meier, 2005].

Little is known about the protein-binding sites in H/ACA RNAs. For example, archaeal H/ACA snoRNAs have kink-turns, and the core protein L7Ae binds to this kink-turn [Rozhdestvensky et al., 2003; Henras et al., 2004b]. In contrast, most eukaryotic H/ACA RNAs, including E1 snoRNA (Fig. 1), lack sequences that could fold into canonical kink-turns.

There are some apparent differences in the direct contacts among the H/ACA RNP core proteins and the H/ACA RNA, and in the order of those interactions, when testing different experimental systems developed from various organisms. First, in cell-free systems, *Saccharomyces cerevisiae* purified proteins: (i) Gar1p and Nhp2p each binds directly to H/ACA snoRNA [Bagni and Lapeyre, 1998; Henras et al., 2001]; and (ii) Cbf5p, Gar1p, and Nop10p form a complex in the absence of Nhp2p and H/ACA snoRNA [Henras et al., 2004a]. Second, in vitro, archaeal purified proteins: (i) L7Ae (the homolog of eukaryotic Nhp2p) and Cbf5p each interacts directly and independently with H/ACA snoRNA; (ii) Cbf5p contacts directly and independently Nop10p and Gar1p; and (iii) L7Ae does not associate with other H/ACA RNP core proteins in the absence of H/ACA snoRNA [Baker et al., 2005; Charpentier et al., 2005). Third, immunoprecipitation of recombinant rodent proteins with antibodies to NAP57 shows that: (i) association with Nhp2p requires binding of Nop10p to NAP57; (ii) this Nop10p-NAP57-Nhp2p trimer interacts with H/ACA snoRNA; and (iii) Gar1p contacts NAP57 [Wang and Meier, 2004]. Finally, in a human cell extract, four proteins, whose approximate sizes (about 60, 29, 23, and 14 kDa) are compatible with being the four H/ACA RNP core proteins, each contacts human E1/U17 RNA directly [Dragon et al., 2000].

Non-Core Proteins That Associate With H/ACA RNPs

In yeast, several additional proteins associate with H/ACA RNPs or H/ACA RNP core proteins, and some have been shown to be required for accumulation and/or localization of H/ACA sno RNAs, and/or proteins. First, the phylogenetically conserved phosphoprotein Nopp140: (i) associates with NAP57; (ii) is the only protein known to interact with mature H/ACA RNPs; (iii) is not a stable component of H/ACA RNPs; and (iv) may be a chaperone of H/ACA snoRNPs [Yang et al., 2000; Meier, 2005]. Nopp140 may play a wide range of other roles, such as in rRNA transcription and the induction of intranuclear membrane cisternae in endometrial cells

[Meier, 2005]. Second, Sen1p, an essential RNA helicase, interacts with box H/ACA snoRNAs and is required for processing of snoRNAs [Ursic et al., 1997]. Third, the nucleolar protein Ssb-1p associates with H/ACA snoRNAs snR10 and snR11 [Clark et al., 1990]. Fourth: (i) the essential proteins Naf1p and Shq1p form a complex; (ii) both proteins contact Cbf5p and Nhp2p directly; (iii) Naf1p also binds Gar1p and Nop10p; (iv) neither protein is a stable component of H/ACA RNPs; (v) both proteins are required for accumulation of H/ACA RNAs; and (vi) Naf1p is also necessary for accumulation of Cbf5p, Gar1p, and Nop10p [Dez et al., 2002; Fatica et al., 2002; Yang et al., 2002, 2005]. Thus, Naf1p and Shq1p may be involved in the assembly of H/ACA RNPs [Dez et al., 2002; Fatica et al., 2002; Yang et al., 2002]. Fifth, direct binding of the double-stranded RNA nuclease Rnt1p to Gar1p is required for the nuclear import of H/ACA RNP core proteins [Tremblay et al., 2002]. Finally, the nucleoplasmic helicase Rvb2 is required for production of H/ACA snoRNAs and for localization of H/ACA RNP core protein Gar1p [King et al., 2001].

Proteins Specific to One H/ACA RNP

H/ACA RNPs with unique functions are expected to contain or interact with some unique proteins. The only sequence-identified protein known to be a unique, integral component of an H/ACA RNP is the reverse transcriptase TERT of the telomerase RNP. The yeast snR30 snoRNP (E1/U17 in vertebrates) [Atzorn et al., 2004] has seven proteins, whose molecular masses are 10, 23, 25, 38, 46, 48, and 65 kDa, based on biochemical isolation, meaning that the composition of this snoRNP includes at least three snR30-specific proteins [Lübben et al., 1995].

E1 snoRNP in Vertebrate Cells

The E1 snoRNP differs from other box H/ACA snoRNPs in human cells: (i) it sediments faster, and as a more compact RNP, than other H/ACA snoRNPs in glycerol gradients; and (ii) its release from larger structures is more resistant to high ionic strength than that of other H/ACA snoRNPs [Rimoldi et al., 1993].

Different proteins contact E1 RNA directly in frog oocytes. Sequence-specific UV-crosslinking at, or primarily at, uridine residues in frog cells detects primarily five protein bands crosslinked to E1 RNA, plus several additional protein bands less intensely labeled with a crosslinked E1 RNA radioactive UMP [Smith et al., 2005]. Protein A (~62 kDa) is not common to all H/ACA snoRNAs because it appears as a strong signal with E1 RNA but not with E2 RNA [Smith et al., 2005]. Protein C (~20 kDa) is not common to all H/ACA snoRNAs because it is detectable with E1 RNA but not with E3 RNA [Smith et al., 2005]. It has been estimated that the vertebrate H/ACA snoRNP core proteins migrate in denaturing polyacrylamide gel electrophoresis as proteins of about 60 (Cbf5p), 29 (Gar1p), 23 (Nhp2p), and 14 (Nop10p) kDa [Dragon et al., 2000]. Then, proteins B (~50 kDa) and D (~19 kDa) might not be any of the known vertebrate core H/ACA proteins, and the possibility that protein E (~13 kDa) may be Nop10p cannot be ruled out now. Neither of these five proteins could be Nopp140 because of their smaller sizes. It would not be surprising for E1 RNA to contact unique proteins since, other than the H and ACA boxes, the sequences evolutionarily conserved in vertebrate E1 RNA are E1-specific [Selvamurugan et al., 1997]. Identification of a UV-crosslinked protein in microinjected frog oocytes by mass spectroscopy would be very difficult because this analysis requires five orders of magnitude more protein than what is needed to detect RNA-protein UV-crosslinking in cells.

Figure 1 shows a model of vertebrate E1 RNA based on the available results. In microinjected frog oocytes, the UV-crosslinking of E1 RNA at uridine residues: (i) to protein A (~62 kDa) depends on two sites at the 5′ end of the RNA: sites 1 and 2; (ii) to protein B (~50 kDa) depends on five sites found in two stem-loops located upstream of the central hinge of the RNA: sites 2, 6, and 8–10; (iii) to protein C (~20 kDa) depends on two sites at the 3′ stem-loop of E1 RNA: sites 16 and 19; and (iv) of protein E (~13 kDa) depends on 14 sites scattered throughout the length of E1 RNA: sites 2, 6–11, 13, 14, 16–18, 21, and 22 [Smith et al., 2005]. Similar results have not been reported in any other snoRNP or H/ACA RNP.

It has been suggested that yeast H/ACA snoRNPs may be symmetrical particles in which each of the two hairpins in the snoRNA binds a copy of the four core proteins [Watkins et al., 1998; Terns and Terns, 2002; Meier, 2005]. In contrast, the dependence of UV-crosslinking of proteins A and B on the 5′ half of E1 RNA and of protein C on the 3′ half of E1 RNA in cells strongly suggest that the frog E1 snoRNP is, at least dynamically, an asymmetric particle

[Smith et al., 2005]. In a human cell extract, the 3′-terminal stem-loop with box ACA of E1/U17 RNA is sufficient for protein binding to form an RNP [Dragon et al., 2000). Based on the results of proteins A, B, and E in frog oocytes, the RNP assembled in the human cell extract may not have a full, physiological set of proteins.

The effects of these base substitutions do not appear to be caused by changes in E1 RNA conformation. Substitution of three bases (in sites 10, 13, 19, and 22) or four bases (in site 14) in conserved single-stranded segments would not be expected to have major global effects on the conformation of E1 RNA. Indeed, 18 out of the 22 E1 RNA substitution mutants tested have no detectable effect on the conformation of deproteinized E1 RNA [Ruhl et al., 2000]. However, each of these 22 mutants exhibits in cells abnormalities in E1 snoRNP profile, E1 RNA-protein contacts, and/or E1 RNA stability [Ruhl et al., 2000; Smith et al., 2005].

One possibility is that each of these E1 RNA sites contacts directly a given protein. For example, some ribosomal proteins interact directly with many segments of rRNA [Brodersen et al., 2002]. Alternatively, some of these E1 RNA sites may be needed for the proper snoRNP conformation required for protein binding. Upon base substitution at one E1 RNA site, it is most likely that its effects in cells on E1 snoRNP pattern and E1 RNA stability are linked to its effect on E1 RNA-protein crosslinking. At each E1 RNA site, a substitution mutation could alter the E1 snoRNP gel electrophoresis pattern because of decreased protein content of the snoRNP, or change in the conformation of the snoRNP, or both. The following observations may be useful in reference to this question. Mutation at E1 RNA sites 5, 8, 12, 17, 19, and 20 result in two, or at least two, E1 snoRNP bands each [Ruhl et al., 2000]. How could two bands be produced? It is possible that: (i) the levels of accessible cellular pools of some of the proteins that react with E1 RNA may be limiting; and (ii) some E1 RNA mutants may have decreased affinity for such proteins. Then, in these mutants, it is more likely that two E1 snoRNP bands of distinct electrophoretic mobility are generated by different protein content rather than by the same protein content and two different E1 snoRNP conformations.

One of the questions is whether the snoRNP pattern and snoRNA-protein crosslinking depend on RNA sequence or structure at each conserved E1 RNA site. They apparently depend on RNA structure at E1 RNA sites 1 and 9, because these sites are conserved structures with non-conserved sequences in the vertebrate E1 RNA consensus sequence and secondary structure [Selvamurugan et al., 1997]. They apparently depend on RNA sequence at E1 RNA sites 10, 13, 14, 19, and 22, and at most of sites 17 and 21, because these sites are conserved single-stranded sequences in the vertebrate E1 RNA consensus sequence and structure [Selvamurugan et al., 1997].

The snoRNP pattern and snoRNA-protein crosslinking appear to depend on RNA sequence at sites 2, 6, 8, 11, 18, and 20, because of the following observations. First, E1 RNA site 2 is base-paired to site 6, site 8 to site 11, and site 18 to site 20 in the consensus vertebrate E1 RNA secondary structure [Selvamurugan et al., 1997]. Mutation at sites 2 and 6, mutation at sites 8 and 11, as well as mutation at sites 18 and 20, generate non-overlapping E1 snoRNP profiles in each pair [Ruhl et al., 2000]. Second, E1 RNA: (i) site 2, but not site 6, is needed for crosslinking to protein A; (ii) site 8, but not site 11, is necessary for crosslinking to protein B; and (iii) site 18, but not site 20, is needed for crosslinking to protein E [Smith et al., 2005].

E1 RNA sites 5, 7, and 16 (and 15) are base-paired to sites 4, 12, and 21, respectively, in the consensus structure of vertebrate E1 RNA [Selvamurugan et al., 1997]. E1 RNA-protein crosslinking appears to depend on E1 RNA sequence at sites 7 and 16, because: (i) site 7, but not site 12, is needed for crosslinking to protein E (~13 kDa); and (ii) site 16, but not site 21, is necessary for crosslinking to protein C (~20 kDa) [Smith et al., 2005]. E1 RNA-protein crosslinking depends most likely on E1 RNA sequence: (i) at site 21 because this sequence is highly conserved from yeast to mammals [Selvamurugan et al., 1997; Atzorn et al., 2004]; and (ii) at the site 4 conserved three-base sequence because this dependence is apparently on sequence at conserved three-base sequences (sites 10–13) and two-base sequences (sites 15, 19, and 22) [Smith et al., 2005]. The E1 snoRNP pattern depends most likely on RNA sequence at sites 5, 12, and 15 because, if the dependence in these sites were on RNA structure, the more heterodisperse snoRNP pattern of mutation at each of these sites should be seen also after mutation at sites 4, 7, and 21 (on the opposite strand, Fig. 1), respectively, but it is not [Ruhl et al., 2000].

The following comments can be made in reference to requirements for E1 snoRNP formation in cells. First, association of E1 RNA with proteins C and E may be involved because: (i) E1 RNA site 19 is necessary to crosslink protein C; (ii) sites 7, 11, 13, 14, 17, 18, 21, and 22 are needed to crosslink protein E; and (iii) all nine of these sites are necessary to produce a normal E1 snoRNP pattern [Ruhl et al., 2000; Smith et al., 2005]. Second, contact of E1 RNA with protein A may not be needed, because E1 RNA site 1 is necessary to crosslink this protein, but not to generate a normal E1 snoRNP profile [Ruhl et al., 2000; Smith et al., 2005].

Interaction of E1 RNA with proteins A, C, and E is apparently needed for E1 RNA stability in cells, because: (i) E1 RNA site 1 is necessary for crosslinking to protein A; (ii) site 19 is needed for crosslinking to protein C; (iii) sites 7, 11, 13, 14, 17, 18, 21, and 22 are necessary for crosslinking to protein E; and (iv) all 10 of these sites are also needed for E1 RNA stability [Ruhl et al., 2000; Smith et al., 2005].

E1 RNA-protein contacts, E1 snoRNP formation, and E1 RNA stability in cells, all three depend on many evolutionarily conserved E1 RNA sites. Except for the H and ACA boxes, those conserved E1 RNA sites are E1-specific, rather than generic for box H/ACA snoRNAs [Selvamurugan et al., 1997; Ruhl et al., 2000; Smith et al., 2005]. Many conserved E1 RNA sites are needed for E1 RNA-protein contacts, E1 snoRNP formation, and E1 RNA stability in cells, but few such RNA elements are known for other snoRNAs and H/ACA RNAs.

Base substitutions in each of 20 of the conserved E1 RNA sites shown in Figure 1 produces a different abnormal phenotype in cells, in terms of E1 RNA-protein contact pattern, E1 snoRNP electrophoretic profile, and/or E1 RNA stability. That is, each of these E1 RNA sites has a different role in these cellular events. These findings, plus the separation of these conserved E1 RNA regions by evolutionarily non-conserved segments, indicate that these E1 RNA sites are separate elements. Thus, E1 RNA has at least 20 distinct elements, apparently two structures and 18 sequences. These observations indicate an unexpected complexity in the vertebrate E1 snoRNP.

FUTURE DIRECTIONS

The E1 snoRNP exhibits high complexity in cells: (i) E1 RNA contacts directly several proteins which apparently are not any of the H/ACA RNP four core proteins; (ii) many proteins [Smith et al., 2005] and multiple E1 RNA sites are needed for E1 snoRNP formation, E1 RNA stability [Ruhl et al., 2000], and E1 RNA–protein interactions [Smith et al., 2005]. Many interesting questions remain to be elucidated in cells, including: (i) the exact role of E1/U17/snR30 snoRNA in pre-rRNA cleavage processing, and the mechanism of this function; (ii) the identification, structure, and function of each of the proteins which interact specifically with E1 RNA or the E1 snoRNP; (iii) the identification of the nucleotide and amino acid positions required for the various specific E1 RNA-protein contacts; (iv) the mechanism of assembly of all of the components of the E1 snoRNP; (v) the identification of the E1 RNA role which involves direct interaction of E1 RNA with a site near the upstream end of the 5′ external transcribed spacer of pre-rRNA (which in human pre-rRNA are within nucleotide positions 697–1,163) [Rimoldi et al., 1993]; (vi) the identification of the E1 RNA role which involves direct contact of a uridine at site 6 and a uridine at site 2 of E1 RNA (which in human E1 RNA are nucleotide positions 43 and 8, respectively) with nucleolar large RNA that is most likely pre-rRNA [Rimoldi et al., 1993]; (vii) to test whether the E1 snoRNP is, at least dynamically, an asymmetric particle [Smith et al., 2005]; and (viii) to test whether the uridine residue at position 19 of human E1 RNA is generated by uridine insertional editing [Nag et al., 1993]. Since the E1/U17/snR30 snoRNP does not appear to be involved in RNA pseudouridylation, is a pseudouridine synthase a component of the vertebrate E1 snoRNP and, if so, why? In view of the surprising complexity of the E1 snoRNP known thus far, the search for these answers promises to be exciting.

ACKNOWLEDGMENTS

I apologize for not citing many relevant publications because of reference limitations. I thank the National Institute of General Medical Sciences for grant R01GM64729.

REFERENCES

Atzorn V, Fragapane P, Kiss T. 2004. U17/snR30 is a ubiquitous snoRNA with two conserved sequence motifs essential for 18S rRNA production. Mol Cell Biol 24:1769–1778.

Bagni C, Lapeyre B. 1998. Gar1p binds to the small nucleolar RNAs snR10 and snR30 in vitro through a nontypical RNA binding element. J Biol Chem 273: 10868–10873.

Baker DL, Youssef OA, Chastkofsky MI, Dy DA, Terns RM, Terns MP. 2005. RNA-guided RNA modification: Functional organization of the archaeal H/ACA RNP. Genes Dev 19:1238–1248.

Balakin AG, Smith L, Fournier MJ. 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.

Bortolin ML, Ganot P, Kiss T. 1999. Elements essential for accumulation and function of small nucleolar RNAs directing site-specific pseudouridylation of ribosomal RNAs. EMBO J 18:457–469.

Bousquet-Antonelli C, Henry Y, Gélugne JP, Caizergues-Ferrer M, Kiss T. 1997. A small nucleolar RNP protein is required for pseudouridylation of eukaryotic ribosomal RNAs. EMBO J 16:4770–4776.

Brodersen DE, Clemons WM, Carter AP, Wimberly BT, Ramakrishnan V. 2002. Crystal structure of the 30 S ribosomal subunit from *Thermus thermophilus*: Structure of the proteins and their interactions with 16 S RNA. J Mol Biol 316:725–768.

Cervelli M, Cecconi F, Giorgi M, Annesi F, Oliverio M, Mariottini P. 2002. Comparative structure analysis of vertebrate U17 small nucleolar RNA (snoRNA). J Mol Evol 54:166–179.

Charpentier B, Muller S, Branlant C. 2005. Reconstitution of archaeal H/ACA small ribonucleoprotein complexes active in pseudouridylation. Nucleic Acids Res 33:3133–3144.

Clark MW, Yip ML, Campbell J, Abelson J. 1990. SSB-1 of the yeast *Saccharomyces cerevisiae* is a nucleolar-specific, silver-binding protein that is associated with the snR10 and snR11 small nuclear RNAs. J Cell Biol 111:1741–1751.

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

Dragon F, Pogačić V, Filipowicz W. 2000. In vitro assembly of human H/ACA small nucleolar RNPs reveals unique features of U17 and telomerase RNAs. Mol Cell Biol 20: 3037–3048.

Eliceiri GL. 1999. Small nucleolar RNAs. Cell Mol Life Sci 56:22–31.

Fatica A, Dlakic M, Tollervey D. 2002. Naf1p is a box H/ACA snoRNP assembly factor. RNA 8:1502–1514.

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

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

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

Henras A, Dez C, Noaillac-Depeyre J, Henry Y, Caizergues-Ferrer M. 2001. Accumulation of H/ACA snoRNPs depends on the integrity of the conserved central domain of the RNA-binding protein Nhp2p. Nucleic Acids Res 29:2733–2746.

Henras AK, Capeyrou R, Henry Y, Caizergues-Ferrer M. 2004a. Cbf5p, the putative pseudouridine synthase of H/ACA-type snoRNPs, can form a complex with Gar1p and Nop10p in absence of Nhp2p and box H/ACA snoRNAs. RNA 10:1704–1712.

Henras AK, Dez C, Henry Y. 2004b. RNA structure and function in C/D and H/ACA s(no)RNPs. Curr Opin Struct Biol 14:335–343.

Jady BE, Bertrand E, Kiss T. 2004. Human telomerase RNA and box H/ACA scaRNAs share a common Cajal body-specific localization signal. J Cell Biol 164:647–652.

King TH, Decatur WA, Bertrand E, Maxwell ES, Fournier MJ. 2001. 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 21:7731–7746.

Kiss T, Filipowicz W. 1993. Small nucleolar RNAs encoded by introns of the human cell cycle regulatory gene RCC1. EMBO J 12:2913–2920.

Kiss AM, Jady BE, Bertrand E, Kiss T. 2004. Human box H/ACA pseudouridylation guide RNA machinery. Mol Cell Biol 24:5797–5807.

Lafontaine DLJ, Bousquet-Antonelli C, Henry Y, Caizergues-Ferrer M, Tollervey D. 1998. The box H + ACA snoRNAs carry Cbf5p, the putative rRNA pseudouridine synthase. Genes Dev 12:527–537.

Lange TS, Ezrokhi M, Amaldi F, Gerbi SA. 1999. Box H and box ACA are nucleolar localization elements of U17 small nuclear RNA. Mol Biol Cell 10:3877–3890.

Lübben B, Fabrizio P, Kastner B, Lührmann R. 1995. Isolation and characterization of the small nucleolar ribonucleoprotein particle snR30 from *Saccharomyces cervisiae*. J Biol Chem 270:11549–11554.

Meier UT. 2005. The many facets of H/ACA ribonucleoproteins. Chromosoma 114:1–14.

Mishra RK, Eliceiri GL. 1997. Three small nucleolar RNAs that are involved in ribosomal RNA precursor processing. Proc Natl Acad Sci USA 94:4972–4977.

Mitchell JR, Cheng J, Collins K. 1999. A box H/ACA small nucleolar RNA-like domain at the human telomerase RNA 3′ end. Mol Cell Biol 19:567–576.

Morrissey JP, Tollervey D. 1993. Yeast snR30 is a small nucleolar RNA required for 18S rRNA synthesis. Mol Cell Biol 13:2469–2477.

Nag MK, Thai T, Ruff EA, Selvamurugan N, Kunnimalaiyaan M, Eliceiri GL. 1993. Genes for E1, E2 and E3 small nucleolar RNAs. Proc Natl Acad Sci USA 90:9001–9005.

Narayanan A, Lukowiak A, Jády BE, Dragon F, Kiss T, Terns RM, Terns MP. 1999. Nucleolar localization signals of box H/ACA small nucleolar RNAs. EMBO J 18:5120–5130.

Ni J, Tien AL, Fournier MJ. 1997. Small nucleolar RNAs direct site-specific synthesis of pseudouridine in ribosomal RNA. Cell 89:565–573.

Richard P, Darzacq X, Bertrand E, Jady BE, Verheggen C, Kiss T. 2003. A common sequence motif determines the Cajal body-specific localization of box H/ACA scaRNAs. EMBO J 22:4283–4293.

Rimoldi OJ, Raghu B, Nag MK, Eliceiri GL. 1993. Three new small nucleolar RNAs that are psoralen crosslinked

in vivo to unique regions of pre-rRNA. Mol Cell Biol 13:4382–4390.

Rozhdestvensky TS, Tang TH, Tchirkova IV, Brosius J, Bachellerie JP, Huttenhofer A. 2003. Binding of L7Ae protein to the K-turn of archaeal snoRNAs: A shared RNA binding motif for C/D and H/ACA box snoRNAs in archaea. Nucleic Acids Res 31:869–877.

Ruff EA, Rimoldi OJ, Raghu B, Eliceiri GL. 1993. Three small nucleolar RNAs of unique nucleotide sequences. Proc Natl Acad Sci USA 90:635–638.

Ruhl DD, Pusateri ME, Eliceiri GL. 2000. Multiple conserved segments of E1 small nucleolar RNA are involved in the formation of a ribonucleoprotein particle in frog oocytes. Biochem J 348:517–524.

Selvamurugan N, Joost OH, Haas ES, Brown JW, Galvin NJ, Eliceiri GL. 1997. Intracellular localization and unique conserved sequences of three small nucleolar RNAs. Nucleic Acids Res 25:1591–1596.

Smith JL, Walton AH, Eliceiri GL. 2005. UV-crosslinking of E1 small nucleolar RNA to proteins in frog oocytes. J Cell Physiol 203:202–208.

Terns MP, Terns RM. 2002. Small nucleolar RNAs: Versatile trans-acting molecules of ancient evolutionary origin. Gene Expr 10:17–39.

Tremblay A, Lamontagne B, Catala M, Yam Y, Larose S, Good L, Elela SA. 2002. A physical interaction between Gar1p and Rnt1pi is required for the nuclear import of H/ACA small nucleolar RNA-associated proteins. Mol Cell Biol 22:4792–4802.

Ursic D, Himmel KL, Gurley KA, Webb F, Culbertson MR. 1997. The yeast SEN1 gene is required for the processing of diverse RNA classes. Nucleic Acids Res 25:4778–4785.

Wang C, Meier UT. 2004. Architecture and assembly of mammalian H/ACA small nucleolar and telomerase ribonucleoproteins. EMBO J 23:1857–1867.

Wang C, Query CC, Meier UT. 2002. Immunopurified small nucleolar ribonucleoprotein particles pseudouridylate rRNA independently of their association with phosphorylated Nopp140. Mol Cell Biol 22:8457–8466.

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

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

Yang PK, Rotondo G, Porras T, Legrain P, Chanfreau G. 2002. The Shq1p.Naf1p complex is required for box H/ACA small nucleolar ribonucleoprotein particle biogenesis. J Biol Chem 277:45235–45242.

Yang PK, Hoareau C, Froment C, Monsarrat B, Henry Y, Chanfreau G. 2005. Cotranscriptional recruitment of the pseudouridylsynthetase Cbf5p and of the RNA binding protein Naf1p during H/ACA snoRNP assembly. Mol Cell Biol 25:3295–3304.