Biogenesis of small nuclear RNPs

Tamás Kiss

Laboratoire de Biologie Moléculaire Eucaryote du CNRS, UMR5099, Université Paul Sabatier, IFR109, 118 route de Narbonne, 31062 Toulouse CEDEX 4, France
Biological Research Centre, Hungarian Academy of Sciences, Szeged, Hungary
(e-mail: tamas@ibcg.biotoul.fr)

The nucleus contains a large number of metabolically stable, 60-450-nt-long small nuclear RNAs (snRNAs) that exist in the form of ribonucleoproteins (snRNPs) (Yu et al., 1999). Each snRNP is composed of an snRNA and a set of associated RNP proteins. Based on their function and intranuclear localization, mammalian snRNPs can be classified into three major classes. The U1, U2, U4, U5 and U6 major and the U11, U12, U4atac and U6atac minor spliceosomal snRNPs function in removal of pre-mRNA introns and are predominantly nucleoplasmic (Lamond and Spector, 2003). More than 200 small nucleolar RNPs (snoRNPs) are present in the nucleolus, where they function mainly in the nucleolar maturation of cytoplasmic 18S, 5.8S and 28S rRNAs (Kiss, 2001; Terns and Terns, 2002). Finally, a recently discovered group of nuclear RNAs, termed small Cajal body RNPs (scaRNPs), accumulate in Cajal (coiled) bodies. The scaRNAs direct the site-specific 2′-O-ribose methylation and pseudouridylation of the RNA polymerase (Pol)-II-transcribed spliceosomal snRNAs (Darzacq et al., 2002). Biogenesis of snRNPs is a complex process that usually includes four major steps: (1) synthesis of a large precursor snRNA (pre-snRNA); (2) nucleolytic processing of the nascent pre-snRNA into mature-sized snRNA; (3) introduction of site-specific covalent nucleotide modifications; and (4) packaging of snRNA with RNP proteins. However, despite the apparent similarities, the various classes of snRNP follow different biosynthetic pathways and most steps of snRNP biogenesis can be linked to distinct subcellular compartments.

Biogenesis of spliceosomal snRNPs

With the exception of U6, the major spliceosomal snRNAs (U1, U2, U4 and U5) are synthesized by Pol II. The Pol-II-specific snRNAs acquire a
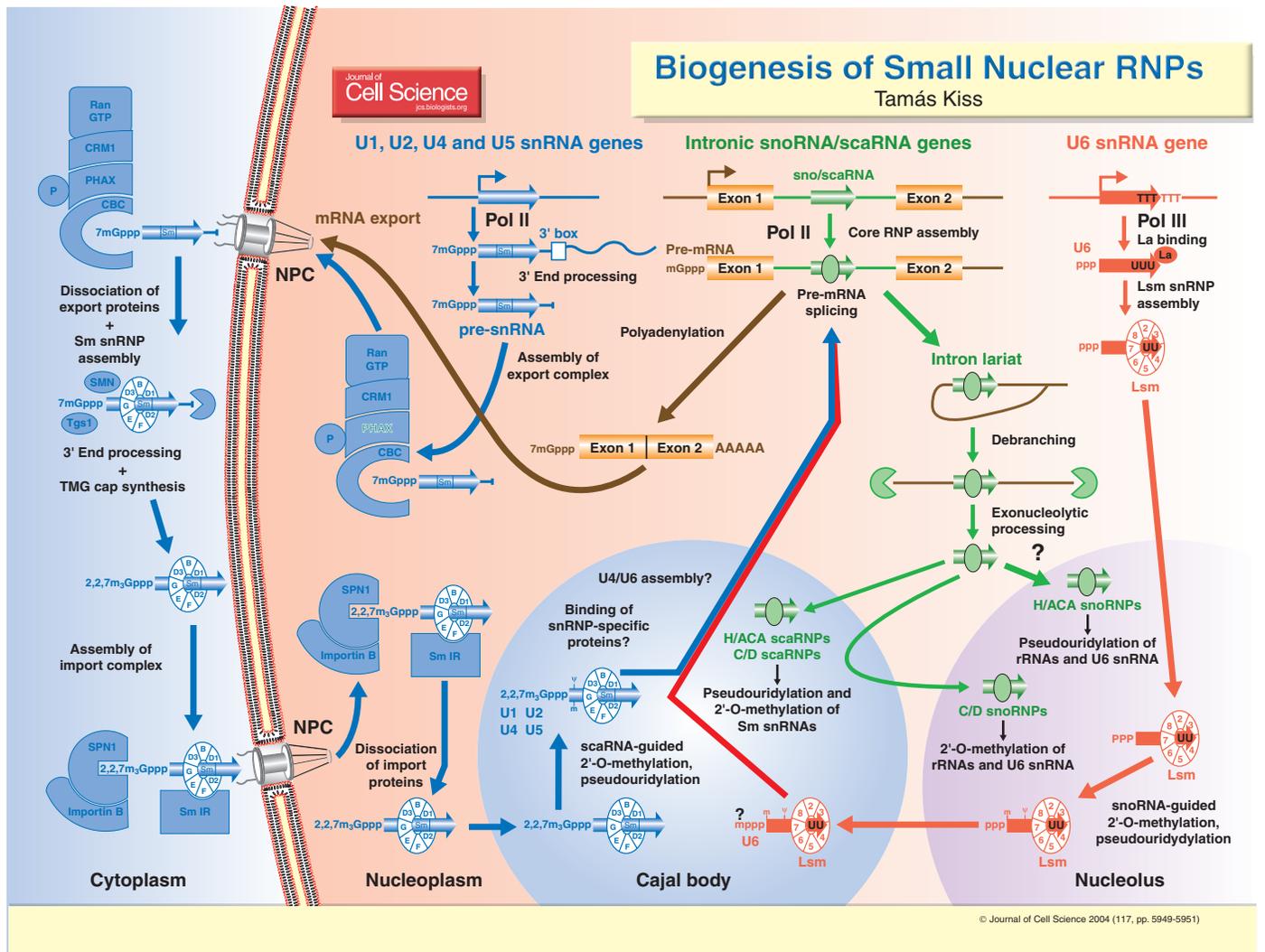
© Journal of Cell Science 2004 (117, pp. 5949-5951)

(See poster insert)

monomethyl guanosine (7mGpppG) cap structure and their primary transcripts are extended far beyond the 3′ end of the mature RNA. Stable pre-snRNA intermediates carrying only a short (5-10 nt) 3′ overhang are generated by a transcription-linked processing event that requires a cis-acting signal sequence (3′ box) and phosphorylation of the C-terminal domain of the largest subunit of Pol II (Medlin et al., 2003). The pre-snRNA is assembled into a large export complex that includes the 7mG-cap-binding complex (CBC), the phosphorylated adaptor for RNA export (PHAX) and the CRM1/RanGTP complex. After export to the cytoplasm through the nuclear pore complex (NPC), dephosphorylation of PHAX and GTP hydrolyses of Ran induce disassembly of the snRNA export complex (Ohno et al., 2000).

In the cytoplasm, the survival of motor neurons (SMN) protein complex facilitates the assembly of pre-snRNA with a heteroheptameric ring of seven Sm proteins – B/B′, D1, D2, D3, E, F and G – to form the snRNP core complex (Paushkin et al., 2002). The Sm proteins bind to the conserved Sm-binding site (Sm, consensus RAUU^U/$_G$UUGR) of Pol-II-specific spliceosomal snRNAs. Binding of Sm proteins is a prerequisite for hypermethylation of the 7mG cap to 2,2,7-trimethylguanosine (TMG) by the Tgs1 methyltransferase, as well as for exonucleolytic removal of the 3′ trailer sequences. The newly synthesized TMG cap and the associated Sm proteins provide the nuclear localization signal for the cytoplasmic core snRNP. Re-importation of Sm snRNPs into the nucleus is mediated by three factors: snurportin-1 (SPN1), which recognizes the TMG cap of the snRNA; importin β, which associates with SPN1; and an unidentified import receptor that interacts with Sm proteins (Sm IR) (Palacios et al., 1997).

After re-entering the nucleoplasm, the import proteins dissociate and the newly imported snRNPs transiently accumulate in Cajal bodies (Sleeman and Lamond, 1999). In the Cajal body, the U1, U2, U4 and U5 snRNAs undergo site-specific pseudouridylation (Ψ) and 2′-O-methylation (m) directed by scaRNAs, which specifically accumulate in Cajal bodies (Jády et al., 2003). In addition to the common Sm core proteins, the mature U1, U2, U4 and U5 snRNPs also contain species-specific RNP proteins. Since the snRNP-specific proteins appear to concentrate in Cajal bodies, they are believed to associate with core snRNPs in this organelle.

Finally, mature spliceosomal snRNPs accumulate in the interchromatin region in structures called splicing speckles. Usually, speckles are located close to actively transcribed genes. They probably provide storage sites for snRNPs or function in snRNP recycling (Lamond and Spector, 2003). It is important to note that several laboratories have reported low levels of nucleolar localization of Sm snRNPs under various conditions (Gerbi et al., 2003; Sleeman and Lamond, 1999). Unfortunately, the functional importance of transient accumulation of Sm snRNAs in the nucleolus remains conjectural. Biogenesis of the Pol-II-specific U11, U12 and U4atac minor spliceosomal snRNPs probably follows the maturation pathway of major Sm snRNPs.

In contrast to the Pol-II-specific Sm snRNAs, biogenesis of the Pol-III-transcribed U6 snRNA is confined to the nucleus. Before accumulating in the nucleoplasmic splicing speckles, the maturing U6 snRNA visits the nucleolus and Cajal body. Synthesis of U6 snRNA terminates within a short U stretch that serves as a transcription termination signal for Pol III. The 3′-terminal U stretch of the newly synthesized U6 RNA transiently associates with the La protein, which provides stability for the nascent RNA and facilitates snRNP assembly (Wolin and Cedervall, 2002). Later, the La protein is replaced by the doughnut-like heteromer of seven Sm-like proteins (Lsm2, Lsm3, Lsm4, Lsm5, Lsm6, Lsm7 and Lsm8) (Achsel et al., 1999). The structure of the Lsm ring is not yet known.

During its maturation, the U6 snRNA also undergoes site-specific pseudouridylation and 2′-O-methylation. However, in contrast to Sm snRNAs, which are modified by scaRNPs in the Cajal body, 2′-O-methylation and probably also pseudouridylation of the U6 snRNA is directed by snoRNAs, which reside in the nucleolus. This indicates that U6 cycles through the nucleolus to undergo snoRNA-mediated modifications (Ganot et al., 1999; Lange and Gerbi, 2000). Binding of the Lsm protein complex is essential for targeting of U6 snRNA to the nucleolus and, afterward, to the Cajal body (Gerbi and Lange, 2002). The mature U6 and U4 snRNPs exist in a common U4-U6 di-snRNP that also binds the U5 snRNP to form the U4-U6-U5 tri-snRNP. The recent finding that SART3/p110, a protein factor facilitating U4-U6 assembly, is concentrated in the Cajal body suggests that assembly of the U4-U6 snRNP may take place in this organelle (Stanek et al., 2003). It is unclear whether U5 snRNP joins the U4-U6 di-snRNP in the Cajal body or later in the nucleoplasm. The mature U6 snRNA carries a γ-monomethyl-phosphate (mpppG) cap structure (Singh and Reddy, 1989), but it remains unknown at what stage of U6 biogenesis the mpppG cap is synthesized.

Biogenesis of snoRNPs and scaRNPs

Although a few mammalian snoRNPs (U3, U8, U14, U17 and U22) function in the nucleolytic processing of pre-rRNA, the great majority direct the site-specific 2′-O-methylation and pseudouridylation of the 18S, 5.8S and 28S rRNAs and the U6 snRNA (Kiss, 2001; Terns and Terns, 2002). Likewise, most Cajal-body-specific scaRNPs function as 2′-O-methylation and pseudouridylation guide RNAs directing modification of the Pol-II-specific U1, U2, U4 and U5 spliceosomal snRNAs (Darzacq et al., 2002). All 2′-O-methylation guide RNAs, both snoRNAs and scaRNAs, carry the conserved C, C′ (consensus UGAUGA), D and D′ (CUGA) box elements, which direct binding of four common box C/D RNP proteins: fibrillarin (a methyltransferase), Nop56p, Nop58p and 15.5 kDa protein. The pseudouridylation guide snoRNAs and scaRNAs possess the conserved H (ANANNA) and ACA boxes and are associated with the dyskerin (a pseudouridine synthase), Nhp2, Nhp10 and Gar1 RNP proteins (Filipowicz and Pogacic, 2002).

Except for the U3, U8 and U13 box C/D

snoRNAs, which are transcribed from independent genes, all mammalian snoRNAs and scaRNAs are encoded within introns of split genes. The intronic sno/scaRNAs are processed from the spliced and debranched host introns by exonucleolytic trimming. It appears that binding of the sno/scaRNP proteins occurs on the host pre-mRNA. Moreover, the splicing machinery bound to the host pre-mRNA may facilitate sno/scaRNP assembly (Hirose et al., 2003). In the released host intron, exonucleolytic excision of the intronic snoRNA occurs close to the site of transcription and splicing of the host pre-mRNA (Samarsky et al., 1998). The bound RNP proteins probably protect the sno/scaRNA from processing exonucleases.

Before accumulating in the nucleolus, intronic box C/D snoRNAs transit through Cajal bodies. The recent finding that final 3′ end trimming and synthesis of the TMG cap of the U3 box C/D snoRNA that is transcribed by Pol II from an independent gene occurs in the Cajal body suggests that the final steps of box C/D snoRNP maturation take place in this organelle (Verheggen et al., 2002). So far, no transient Cajal-body-specific accumulation has been reported for box H/ACA snoRNPs. Therefore, it remains uncertain whether this class of snoRNA is transported directly into the nucleolus or, like box C/D snoRNAs, they also transit through Cajal bodies.

The nucleolar accumulation of the two classes of snoRNAs is determined by the conserved C/D and H/ACA box elements themselves (Terns and Terns, 2002). However, the final destination of scaRNAs, although they carry C/D and/or H/ACA boxes, is the Cajal body. The box H/ACA scaRNAs contain a cis-acting localization signal, the CAB box, that is responsible for targeting these RNAs into the Cajal body (Richard et al., 2003). Interestingly, the CAB box Cajal body localization signal is also present in human telomerase RNA, which has a box-H/ACA-RNA-like domain and

accumulates in the Cajal body (Jády et al., 2004; Zhu et al., 2004). Presumably, box H/ACA scaRNPs and the telomerase RNP share a common CAB-box-associated trans-acting protein factor responsible for Cajal-body-specific accumulation of these RNPs. The molecular mechanism directing the Cajal-body-specific accumulation of box C/D scaRNAs is still unknown.

References

Achsel, T., Brahms, H., Kastner, B., Bachi, A., Wilm, M. and Lührmann, R. (1999). A doughnut-shaped heteromer of human Sm-like proteins binds to the 3′-end of U6 snRNA, thereby facilitating U4/U6 duplex formation in vitro. *EMBO J.* **18**, 5789-5802.

Darzacq, X., Jády, B. E., Verheggen, C., Kiss, A. M., Bertrand, E. and Kiss, T. (2002). Cajal body-specific small nuclear RNAs: a novel class of 2′-O-methylation and pseudouridylation guide RNAs. *EMBO J.* **21**, 2746-2756.

Filipowicz, W. and Pogacic, V. (2002). Biogenesis of small nucleolar ribonucleoproteins. *Curr. Opin. Cell Biol.* **14**, 319-327.

Ganot, P., Jády, B. E., Bortolin, M. L., Darzacq, X. and Kiss, T. (1999). Nucleolar factors direct the 2′-O-ribose methylation and pseudouridylation of U6 spliceosomal RNA. *Mol. Cell. Biol.* **19**, 6906-6917.

Gerbi, S. A. and Lange, T. S. (2002). All small nuclear RNAs (snRNAs) of the [U4/U6.U5] tri-snRNP localize to nucleoli: identification of the nucleolar localization element of U6 snRNA. *Mol. Biol. Cell* **13**, 3123-3137.

Gerbi, S. A., Borovjagin, A. V. and Lange, T. S. (2003). The nucleolus: a site of ribonucleoprotein maturation. *Curr. Opin. Cell Biol.* **15**, 318-325.

Hirose, T., Shu, M. D. and Steitz, J. A. (2003). Splicing-dependent and -independent modes of assembly for intron-encoded box C/D snoRNPs in mammalian cells. *Mol. Cell* **12**, 113-123.

Jády, B. E., Darzacq, X., Tucker, K. E., Matera, A. G., Bertrand, E. and Kiss, T. (2003). Modification of Sm small nuclear RNAs occurs in the nucleoplasmic Cajal bodies following import from the cytoplasm. *EMBO J.* **22**, 1878-1888.

Jády, B. E., Bertrand, E. and 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.

Kiss, T. (2001). Small nucleolar RNA-guided post-transcriptional modification of cellular RNAs. *EMBO J.* **20**, 3617-3622.

Lamond, A. I. and Spector, D. L. (2003). Nuclear speckles: a model for nuclear organelles. *Nat. Rev. Mol. Cell Biol.* **4**, 605-612.

Lange, T. S. and Gerbi, S. A. (2000). Transient nucleolar localization of U6 small nuclear RNA in *Xenopus laevis* oocytes. *Mol. Biol. Cell* **11**, 2419-2428.

Medlin, J. E., Uguen, P., Taylor, A., Bentley, D.

L. and Murphy, S. (2003). The C-terminal domain of pol II and a DRB-sensitive kinase are required for 3′ processing of U2 snRNA. *EMBO J.* **22**, 925-934.

Ohno, M., Segref, A., Bachi, A., Wilm, M. and Mattaj, I. W. (2000). PHAX, a mediator of U snRNA nuclear export whose activity is regulated by phosphorylation. *Cell* **101**, 187-198.

Palacios, I., Hetzer, M., Adam, S. A. and Mattaj, I. W. (1997). Nuclear import of U snRNPs requires importin beta. *EMBO J.* **16**, 6783-6792.

Paushkin, S., Gubitz, A. K., Massenet, S. and Dreyfuss, G. (2002). The SMN complex, an assemblyosome of ribonucleoproteins. *Curr. Opin. Cell Biol.* **14**, 305-312.

Richard, P., Darzacq, X., Bertrand, E., Jády, B. E., Verheggen, C. and Kiss, T. (2003). A common sequence motif determines the Cajal body-specific localisation of box H/ACA scaRNAs. *EMBO J.* **22**, 4283-4293.

Samarsky, D. A., Fournier, M. J., Singer, R. H. and Bertrand, E. (1998). The snoRNA box C/D motif directs nucleolar targeting and also couples snoRNA synthesis and localization. *EMBO J.* **17**, 3747-3757.

Singh, R. and Reddy, R. (1989). Gamma-monomethyl phosphate: a cap structure in spliceosomal U6 small nuclear RNA. *Proc. Natl. Acad. Sci. USA* **86**, 8280-8283.

Sleeman, J. E. and Lamond, A. I. (1999). Newly assembled snRNPs associate with coiled bodies before speckles, suggesting a nuclear snRNP maturation pathway. *Curr. Biol.* **9**, 1065-1074.

Stanek, D., Rader, S. D., Klingauf, M. and Neugebauer, K. M. (2003). Targeting of U4/U6 small nuclear RNP assembly factor SART3/p110 to Cajal bodies. *J. Cell Biol.* **160**, 505-516.

Terns, M. P. and Terns, R. M. (2002). Small nucleolar RNAs: versatile trans-acting molecules of ancient evolutionary origin. *Gene Expr.* **10**, 17-39.

Verheggen, C., Lafontaine, D. L., Samarsky, D., Mouaikel, J., Blanchard, J. M., Bordonne, R. and Bertrand, E. (2002). Mammalian and yeast U3 snoRNPs are matured in specific and related nuclear compartments. *EMBO J.* **21**, 2736-2745.

Wolin, S. L. and Cedervall, T. (2002). The La protein. *Annu. Rev. Biochem.* **71**, 375-403.

Yu, Y. T., Scharl, E. C., Smith, C. M. and Steitz, J. A. (1999). The growing world of small nuclear ribonucleoproteins. In *The RNA World* (ed. Gesteland, R. F., Cech, T. R. and Atkins, J. F.), pp. 487-524. Cold Spring Harbor, New York: Cold Spring Harbor Laboratory Press.

Zhu, Y., Tomlinson, R. L., Lukowiak, A. A., Terns, R. M. and Terns, M. P. (2004). Telomerase RNA accumulates in Cajal bodies in human cancer cells. *Mol. Biol. Cell* **15**, 81-90.