ABSTRACT OF PhD THESIS

RNA-guided post-transcriptional modification of spliceosomal small nuclear RNAs

written by
Beáta Erika Jády

Supervisors:
Prof. Tamás Kiss
Prof. Ferenc Solymosy

Institute of Plant Biology
Biological Research Center, Hungarian Academy of Sciences
Szeged, Hungary
and
Laboratoire de Biologie Moléculaire Eucaryote du CNRS
Université Paul Sabatier Toulouse III
Toulouse, France

2001

INTRODUCTION

In addition to the four major ribonucleotides, adenosine (A), guanosine (G), cytidine (C) and uridine (U), 96 post-transcriptionally modified nucleotides have been described in ribonucleic acids (*1*). The modified nucleotides are synthesised after synthesis of the primary RNA transcript by covalent modification of the base or the ribose moiety of the classical ribonucleotides. In eukaryotic ribosomal RNAs (rRNAs), about 2-3% of the total nucleotides are modified. In rRNAs, the most abundant modifications are the 2'-*O*-ribose methylation and formation of pseudouridines. Many sites of modification are evolutionarily conserved, and the number of modified nucleotides in rRNAs has increased during evolution (*2*). Ribose methylated nucleotides and pseudouridines occur in diverse local sequence and secondary structure environments, and the mechanism of recognition of the substrate nucleotides for both modifications had been a mystery for a long time. With the discovery of the function of guide small nucleolar RNAs (snoRNAs), a unifying process was identified for recognition of ribose methylation and pseudouridylation sites. Site-specific formation of ribose methylated nucleotides and pseudouridines in rRNAs takes place on the 45/35S precursor rRNA (pre-rRNA) and selection of the modification sites are directed by snoRNAs. SnoRNAs belong to two distinct families. The box C/D snoRNAs, most of which direct ribose methylation of rRNAs, feature two conserved sequence elements, the C and the D boxes with the conserved sequences of UGAUGA and CUGA, respectively (figure 1.), (*3, 4*). In the middle of these RNAs, a C and a D box-like element, termed the C' and D' boxes are located. Selection of a ribosomal nucleotide for ribose methylation requires transient base-pairing between the guide snoRNA and the pre-rRNA. The 10-21 nucleotide long base-paired region encompassing the substrate nucleotide is immediately followed by the D or D' box. The ribosomal nucleotide that is located opposite to the fifth nucleotide upstream of the D or D' box of the snoRNA is selected for ribose methylation. Each methylation guide snoRNA is associated with a set of proteins, and they function as small nucleolar ribonucleoproteins

(snoRNPs). Base-pairing between box C/D snoRNAs and the pre-rRNA is recognised by the fibrillarin (yeast Nop1p) snoRNP core protein, which is the methyltransferase enzyme catalysing the formation of 2'-O-ribose methylated nucleotides in rRNAs (*5*).

The pseudouridine is formed by isomerisation of the uridine residue. During this process, the bond between N1 of uracil and C1' of ribose is disrupted, the base is rotated 180° around the N3-C6 axis and finally, a new C5-C1' bond is formed (*6*). The majority of the members of the box H/ACA snoRNA family select uridines for pseudouridylation (*7, 8*). These snoRNAs feature a conserved hairpin-hinge-hairpin-tail secondary structure with the box H sequence element (ANANNA, N is any nucleotide) in the hinge region and the ACA motif located 3 nucleotides upstream from their 3' ends (figure 1.). The recognition motifs of box H/ACA snoRNAs are located on the opposite strands of an internal loop in either the 5'- or the 3'-terminal hairpin. These regions, referred to as the pseudouridylation pocket, transiently base-pair with the pre-rRNA. The short base-paired regions flank two unpaired ribosomal nucleotides from which the one closer to the 5' end of the pre-rRNA is a uridine selected for pseudouridylation. Similarly to box C/D snoRNAs, box H/ACA snoRNAs function in form of snoRNPs, which carry the pseudouridine synthase. The human dyskerin (yeast Cbf5) snoRNP core protein is the pseudouridine synthase in snoRNA-guided rRNA modification (*9-11*).

The ubiquitous U3 and U14 snoRNAs, yeast snR10, snR30 and mammalian U8, U17 and U22 snoRNAs have been demonstrated to function in the nucleolytic processing of rRNAs (*12*). How snoRNPs function in processing of the pre-rRNA is largely unknown. It has been proposed, that snoRNAs functioning in processing of the pre-rRNA are essential chaperones required for the correct folding of the pre-rRNA, however, it cannot be excluded that they are directly involved in cleavages (*13*).

Introns of eukaryotic nuclear precursor messenger RNAs (pre-mRNAs) are removed by two successive trans-esterification reactions (splicing) within a large ribonucleoprotein complex called the spliceosome (*14*). The spliceosome is formed by assembly of ribonucleoproteins (snRNPs) and several non-snRNP splicing factors. The majority of pre-

mRNAs are spliced by the major spliceosome containing the U1, U2, U4, U5 and U6 small nuclear RNAs (snRNAs). However, removal of metazoan introns with AU and AC 5' and 3' splice site consensus sequences is catalysed by a minor spliceosome carrying the U11, U12, U4atac, U6atac and U5 snRNAs (*15*). The spliceosomal snRNAs can be divided into two groups according to their biogenesis. The U1, U2, U4, U5, U11, U12 and U4atac snRNAs are synthesised by RNA polymerase II, whereas the U6 and U6atac snRNAs are transcribed by RNA polymerase III (*16*). Biogenesis of the RNA polymerase II-transcribed spliceosomal snRNAs includes a cytoplasmic phase, but maturation of the RNA polymerase III-specific U6 and U6atac snRNAs takes place entirely in the nucleus.

Compared to rRNAs, less is known about the synthesis and function of modified nucleotides in spliceosomal snRNAs. In human spliceosomal snRNAs, there are 30 2'-*O*-ribose methylated, 3 base methylated nucleotides and 23 pseudouridines (*17*). Like in the case of rRNAs, modified nucleotides in spliceosomal snRNAs are clustered in the conserved and functionally important regions (*18*). The U2 snRNA carries the largest number of modified nucleotides. Modification in the U2 snRNA is essential for both snRNP assembly and pre-mRNA splicing in *Xenopus* oocytes (*19*).

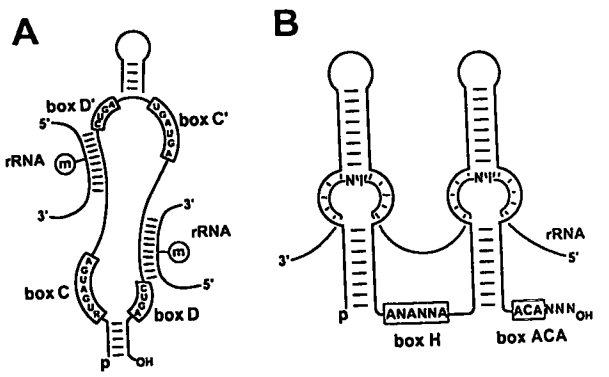

Figure 1. Schematic representation of the secondary structure of snoRNAs (*20*). **A.** Box C/D snoRNAs. **B.** Box H/ACA snoRNAs.

OBJECTIVES

1. Small nucleolar RNAs are synthesised in the nucleoplasm and transported to the nucleolus where they function in rRNA processing and nucleotide modification. We investigated the sequences and structural elements involved in the intranuclear transport of box H/ACA snoRNAs by assaying the localisation of injected fluorescent RNAs in nuclei of *Xenopus* oocytes.

2. We investigated whether small nucleolar RNAs function in maturation of non-ribosomal RNAs by identification and characterisation of novel small nucleolar RNAs.

3. The nucleolus is a functionally highly specialised subnuclear compartment where synthesis, post-transcriptional nucleotide modification and nucleolytic processing of rRNAs takes place. We examined whether nucleotide modification of the RNA polymerase III-transcribed U6 spliceosomal snRNA occurs inside the nucleolus.

4. Factors directing 2'-O-ribose methylation and pseudouridylation of the RNA polymerase II-synthesised U1, U2, U4 and U5 spliceosomal snRNAs was largely unknown. Only one nucleotide modification factor, the yeast Pus1p protein was identified to function in pseudouridylation of the U44 residue in the yeast U2 snRNA. We identified and characterised the first nucleotide modification factor for a higher eukaryotic, RNA polymerase II-specific snRNA.

RESULTS

Nucleolar localisation of box H/ACA snoRNAs.

1. Upon microinjection into the nucleus of *Xenopus* oocytes, in vitro transcribed, fluorescently labelled human box H/ACA snoRNAs localised directly to the nucleolus. We found that cis-acting elements of box H/ACA snoRNAs essential for nucleolar localisation are the H and ACA boxes and a hairpin structure bringing the two conserved sequence elements close to each-other.

2. We demonstrated that microinjected human telomerase RNA transiently localises to the nucleolus. Importantly, vertebrate telomerase RNAs possess a box H/ACA snoRNA domain at their 3' ends (*21*). We found that nucleolar targeting of the human telomerase RNA is dependent on the presence of the ACA box motif in the box H/ACA snoRNA domain suggesting that the box H/ACA snoRNA domain is responsible for transient nucleolar localisation of the telomerase RNA.

Characterisation of the U83 and U84 box C/D snoRNAs.

1. We identified two novel box C/D snoRNAs, the U83 and U84, that lacked a significant complementarity to any of the known non-coding RNAs. The U83 and U84 snoRNAs, unlike snoRNAs functioning in maturation of rRNAs, were not detected in precursor ribosomal particles, suggesting that the U83 and U84 snoRNAs do not function in any aspects of rRNA biogenesis.

2. The U83 and U84 snoRNAs are able to direct 2'-*O*-ribose methylation of artificial substrate RNAs. We believe that both U83 and U84 snoRNAs are ribose methylation guide snoRNAs directing modification of a non-ribosomal, thus far unidentified RNA.

Post-transcriptional nucleotide modification of the U6 spliceosomal small nuclear RNA.

1. We identified a novel guide snoRNA, the mgU6-53 snoRNA, which directs 2'-O-ribose methylation of the U6 snRNA at position A53.

2. We demonstrated that the nucleolus contains all the trans-acting factors that are responsible for the accurate and efficient synthesis of the eight 2'-O-ribose methylated nucleotides and the three pseudouridines carried by the RNA-polymerase III-transcribed U6 spliceosomal snRNA. Since all 2'-O-ribose methylation and pseudouridylation factors recognised short sequences located around the target nucleotides, we concluded that modification of the U6 snRNA is directed exclusively by snoRNAs. Based on our results, we proposed that the U6 snRNA is transiently localised to the nucleolus during its maturation to undergo snoRNA-mediated post-transcriptional nucleotide modification. Indeed, it has been recently shown that microinjected U6 snRNA cycles through the nucleolus in *Xenopus* oocytes (22).

Post-transcriptional nucleotide modification of the RNA polymerase II-specific spliceosomal snRNAs.

1. We identified and characterised a novel, evolutionarily conserved guide RNA, the U85 snRNA, which is composed of a box C/D and a box H/ACA snoRNA domain. We demonstrated that the U85 hybrid RNA directs 2'-O-ribose methylation of the C45 and pseudouridylation of the U46 residues in the human U5 spliceosomal snRNA. The U85 is the first example of an snRNA that directs modification of an RNA-polymerase II-transcribed RNA and that functions both in pseudouridylation and 2'-O-ribose methylation.

2. Surprisingly, the U85 RNA is not detectable in the nucleolus. Instead, it accumulates exclusively in Cajal bodies (manuscript in preparation). Cajal bodies have been

identified as nucleoplasmic organelles about a hundred years ago (*23*). However, the cellular function of Cajal bodies has remained largely elusive.

3. We have recently identified four novel 2'-*O*-ribose methylation and pseudouridylation guide RNAs, which, similarly to the U85 snRNA, reside in Cajal bodies of human HeLa cells. The members of this novel class of small RNAs have been termed small Cajal body-specific RNAs (scaRNAs). Intriguingly, the novel scaRNAs are predicted to guide post-transcriptional modification of the RNA polymerase II-transcribed U1, U2, U4 or U5 spliceosomal snRNAs. These results strongly indicate that post-transcriptional nucleotide modification of the RNA polymerase II-specific spliceosomal snRNAs are directed by scaRNAs at least in some positions. Therefore, we propose that the nucleoplasmic Cajal body is the cellular locale for post-transcriptional maturation of the RNA polymerase II-synthesised spliceosomal snRNAs (manuscript in preparation).

REFERENCES

1. Limbach, P. A., Crain, P. F., and McCloskey, J. A. (2001) in *http://medlib.med.utah.edu/RNAmods/*.
2. Maden, B. E. (1990) *Prog Nucleic Acid Res Mol Biol 39*, 241-303.
3. Kiss-Laszlo, Z., Henry, Y., Bachellerie, J. P., Caizergues-Ferrer, M., and Kiss, T. (1996) *Cell 85*, 1077-88.
4. Cavaille, J., Nicoloso, M., and Bachellerie, J. P. (1996) *Nature 383*, 732-5.
5. Niewmierzycka, A., and Clarke, S. (1999) *J Biol Chem 274*, 814-24.
6. Ofengand, J., Bakin, A., Wrzesinski, J., Nurse, K., and Lane, B. G. (1995) *Biochem Cell Biol 73*, 915-24.
7. Balakin, A. G., Smith, L., and Fournier, M. J. (1996) *Cell 86*, 823-34.
8. Ganot, P., Caizergues-Ferrer, M., and Kiss, T. (1997) *Genes Dev 11*, 941-56.
9. Heiss, N. S., Knight, S. W., Vulliamy, T. J., Klauck, S. M., Wiemann, S., Mason, P. J., Poustka, A., and Dokal, I. (1998) *Nat Genet 19*, 32-8.
10. Lafontaine, D. L., Bousquet-Antonelli, C., Henry, Y., Caizergues-Ferrer, M., and Tollervey, D. (1998) *Genes Dev 12*, 527-37.
11. Zebarjadian, Y., King, T., Fournier, M. J., Clarke, L., and Carbon, J. (1999) *Mol Cell Biol 19*, 7461-72.

12. Smith, C. M., and Steitz, J. A. (1997) *Cell 89*, 669-72.
13. Steitz, J. A., and Tycowski, K. T. (1995) *Science 270*, 1626-7.
14. Will, C. L., and Luhrmann, R. (2001) *Curr Opin Cell Biol 13*, 290-301.
15. Tam, W. Y., and Steitz, J. A. (1996) *Cell 84*, 801-11.
16. Zieve, G. W., and Sauterer, R. A. (1990) *Crit Rev Biochem Mol Biol 25*, 1-46.
17. Gu, J., Patton, J. R., Shimba, S., and Reddy, R. (1996) *Rna 2*, 909-18.
18. Ares, M., Jr., and Weiser, B. (1995) *Prog Nucleic Acid Res Mol Biol 50*, 131-59.
19. Yu, Y. T., Shu, M. D., and Steitz, J. A. (1998) *Embo J 17*, 5783-95.
20. Kiss, T. (2001) *Embo J 20*, 3617-22.
21. Mitchell, J. R., Cheng, J., and Collins, K. (1999) *Mol Cell Biol 19*, 567-76.
22. Lange, T. S., and Gerbi, S. A. (2000) *Mol Biol Cell 11*, 2419-28.
23. Cajal, S. R. y. (1903) *Trab. Lab. Invest. Biol. (Madrid) 2*, 129-221.

PUBLICATIONS

1. Narayanan A, Lukowiak A, Jady BE, Dragon F, Kiss T, Terns RM, Terns MP. (1999) Nucleolar localization signals of box H/ACA small nucleolar RNAs. ***EMBO J.*** **18**, 5120-30. (impact factor: 13.999)

2. Ganot P, Jady BE, Bortolin ML, Darzacq X, Kiss T. (1999) Nucleolar factors direct the 2'-O-ribose methylation and pseudouridylation of U6 spliceosomal RNA. ***Mol. Cell. Biol.*** **19**, 6906-17. (impact factor: 9.666)

3. Jady BE, Kiss T. (2000) Characterisation of the U83 and U84 small nucleolar RNAs: two novel 2'-O-ribose methylation guide RNAs that lack complementarities to ribosomal RNAs. ***Nucleic Acids Res.*** **28**, 1348-54. (impact factor: 5.396)

4. Jady BE, Kiss T. (2001) A small nucleolar guide RNA functions both in 2'-O-ribose methylation and pseudouridylation of the U5 spliceosomal RNA. ***EMBO J.***, **20**, 541-51. (impact factor: 13.999)

Cumulative impact factor: 43.060

ACKNOWLEDGEMENTS

I am highly indebted to Prof. Tamás Kiss, my PhD supervisor, for his excellent scientific guidance, precious friendship and valuable help.

Köszönöm Solymosy Ferenc professzornak, témavezetőmnek szakmai segítségét és tanácsait.

Je remercie à Marie-Line Bortolin et Philippe Ganot pour les conseils et pour l'aide qu'ils m'ont donné pendant notre collaboration. Merci á Xavier Darzacq pour son amitié, aide et traduction, que j'apprécie beaucoup.

Köszönöm munkatársaimnak, Kiss Mihály Arnoldnak, Boros Évának, Antal Máriának és Kiriné Borka Ildikónak értékes segítségüket.

Je remercie á Vera Atzorn et Bruno Faucon pour leur confiance et leur aide.

Merci á tous les membres de Laboratoire de Biologie Moléculaire Eucaryote pour m'accueillir.

Hálás vagyok a Növénybiológiai Intézet munkatársainak segítségéért.

I thank the referees, Prof. Pál Venetianer and Prof. Philippe Bouvet for critical reading of my PhD thesis.

I gratefully acknowledge the French Government, the Institut Française, the Magyar Tudományos Akadémia and the OTKA Foundation for providing me financial support during my PhD studies. I especially thank Bob Kaba (Institut Française) for his help.

I am grateful to the professors, teachers and members of the administration of the Szegedi Tudományegyetem and the Université Paul Sabatier.

Hálásan köszönöm családomnak és barátaimnak a nap mint nap nyújtott szeretetüket és bizalmukat. Külön köszönettel tartozom édesanyámnak és húgomnak szeretetükért és támogatásukért.