Primary sequence of the 5′-terminal region of mouse 18 S rRNA and adjacent spacer

Implications for rRNA processing

F. Raynal, B. Michot and J.P. Bachellerie

Centre de Recherche de Biochimie et de Génétique Cellulaires du CNRS, 118, route de Narbonne, 31062 Toulouse Cédex, France

Received 18 July 1983

Among the stepwise cleavage reactions involved in the processing of rRNA precursors in mammalian cells, an early event corresponds to the removal of the so-called 'external transcribed regions' which are located upstream 18 S rRNA sequence within the primary transcript. We have determined the primary sequence of the domain of mouse pre-rRNA which encompasses this early processing site and analyzed its structural features with reference to the other eukaryotic homologs available. The potential involvement of secondary structure features of rRNA precursors in the recognition process for cleavage is discussed.

18 S rRNA sequence *Mouse* *RNA processing*

1. INTRODUCTION

In an attempt to identify the recognition signals for processing nucleases which are involved in the stepwise cleavages of rRNA precursors in eukaryotes [1], we have undertaken a structural analysis of rRNA precursors in mouse. We previously reported the nucleotide sequence of the 3′ terminal domain of 18 S rRNA gene [2], the internal transcribed spacers (ITS) 1 and 2 [3] and the 5′ terminal domain of 28 S rRNA gene [4]. This allowed us to analyze the structural features of pre-rRNA around the 4 cleavage sites which are located in this central region of rRNA gene [3,5,6]. We here analyzed the structure of mouse pre-rRNA around a fifth processing site. This site is located at the 5′ terminus of mature 18 S rRNA and corresponds to an early processing event; i.e., the removal of the external transcribed spacer (ETS). Potential signals for this reaction are discussed in terms of secondary structure folding of this area of pre-rRNA and by reference to the other eukaryotic sequences available so far, yeast [7,8], *Xenopus* [9–11] and rat [12].

2. METHODS

Recombinant DNA–mouse ribosomal DNA was prepared from PMSE 2, a gift from I. Grummt. It consists of a 2 kb *SalI–Eco*RI fragment of the mouse ribosomal transcription unit encompassing the 5′ terminus of 18 S rRNA gene (see fig.1), cloned in the vector pBR 322. It was obtained from λgt*WES* M_r 974 recombinant, which contains the large *Eco*RI–*Eco*RI fragment (13.6 kb) extending into external transcribed spacer and non-transcribed spacer of mouse plasmocytoma cells, MOPC 149 [13].

Isolation of cloned DNA, restriction endonuclease analysis and DNA sequencing were carried out as in [5]. Chemical DNA sequencing was performed as in [14].

The 5′ terminus of mature 18 S rRNA was identified by reverse transcriptase elongation [15] of a

Published by Elsevier Science Publishers B.V.
00145793/83/$3.00

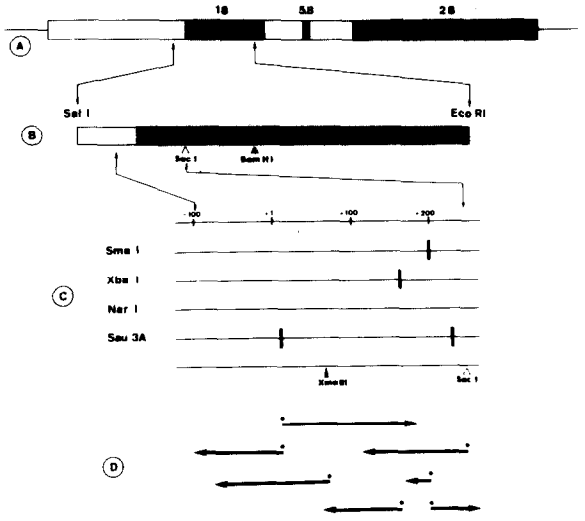

Fig.1. Restriction map of the sequenced region of mouse rDNA. (A) Mouse rRNA transcription unit. (B) Location of the rDNA fragment cloned in pMSE 2 recombinant. (C) Restriction map of the region encompassing the 5′ end of 18 S rRNA gene (positions are numbered from the 5′ terminus of 18 S gene). (D) Sequencing strategy. Starts of arrows indicate ^{32}P-labelled 5′ ends and lengths of arrows are indicative of the extent of sequence read.

5′ ^{32}P-labelled DNA primer. This primer was the coding strand of a 190 nucleotide long restriction fragment from pMSE2, limited by a *Sau*3A site (position 10 from the 5′ end of 18 S rRNA gene) and a *Sma*I site (position 200). End-labelling and strand separation were carried out as in [14]. Hybridization of DNA primer to 18 S rRNA template and reverse transcriptase elongation were performed as in [16]. Biohazards associated with the experiments were pre-examined by the French Control Committee.

3. RESULTS AND DISCUSSION

3.1. *Primary structure of the 5′ terminal region of mouse 18 S rRNA*

The 5′ terminus of 18 S rRNA gene along the DNA sequence was unambiguously identified by reverse transcriptase elongation (see section 2) of a DNA primer. On its 3′ end, this primer was limited by a *Sau*3A site (GATC). It was extended by 13 nucleotides by reverse transcription onto 18 S rRNA template (not shown). No ambiguity remained in the sequence shown in fig.2, after

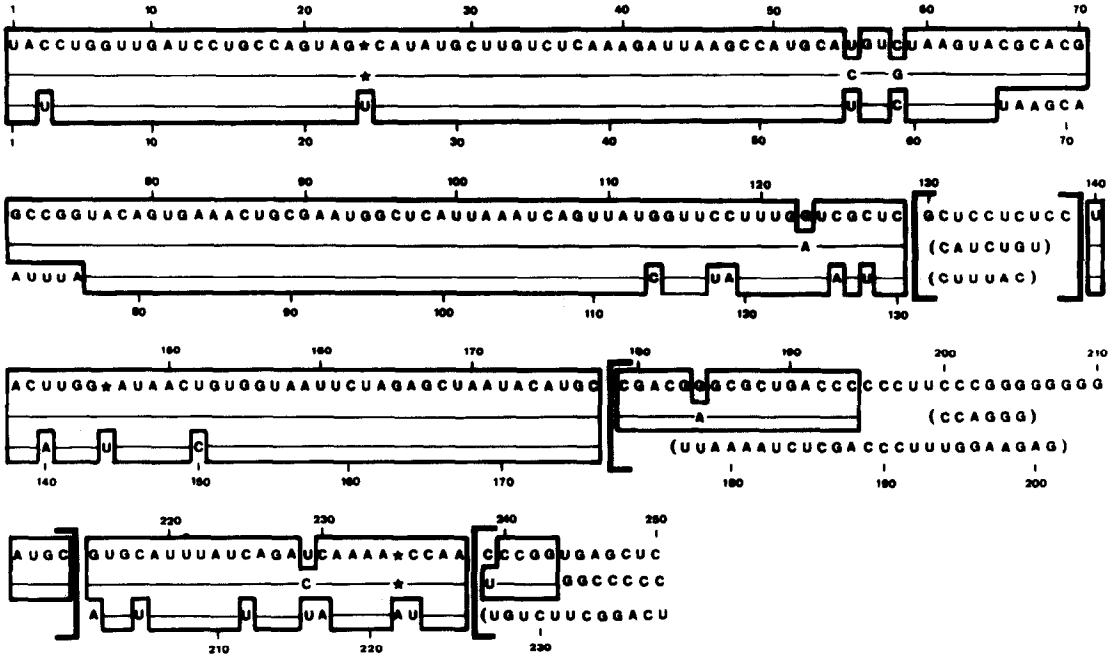

Fig.2. Sequence of the 250 5′ terminal nucleotides of mouse 18 S rRNA (top line) and comparison with *Xenopus laevis* [9], middle line, and yeast [8], bottom line. Sequence tracts which are homologous with *Xenopus* are boxed. Nucleotides identical with the sequence immediately above are represented by a line. Deletions are indicated by a star. Segments where no significant homology could be detected are shown in square brackets.

analysis on both DNA strands and extensive overlapping. Comparison with the two completely sequenced eukaryotic 18 S rRNAs, yeast [8] and *Xenopus* [9], reveals an extensive conservation between these distantly related eukaryotes. However, the rate of evolution is far from being uniform along the sequence: if the sequence homology is extremely high over long tracts, such as segment 1–129 where only 3 base differences are detected between mouse and *Xenopus*, it is abruptly interrupted over some areas of the molecule (shown in parentheses in fig.2), where most of the few changes are concentrated and where some size variation can take place. The rapid divergence of these areas is again confirmed when considering the recently reported sequence for this terminal region of rabbit 18 S rRNA [17]: all the 12 nucleotide changes between mouse and rabbit are located within these bracketted 'hot spots' (8 in segment 179–214, 4 in segment 239–250).

3.2. *Sequence of the external transcribed spacer, immediately upstream mouse 18 S rRNA gene*

As opposed to the adjacent domain of mature 18 S rRNA, this region is characterized by a strikingly unbalanced nucleotide content: the 90 nucleotide long upstream region which has been sequenced (fig.3) is very rich in pyrimidine (79% with 55% C) with only 2 A.

This region is also clearly distinct from mature rRNAs in terms of sequence variation during evolution: no homology could be detected between yeast [7], *Xenopus* [10] and rat [12]. The same holds true between yeast, *Xenopus* and mouse. However, some short conserved tracts can be seen when mouse and rat sequences are compared (fig.3), as has been reported recently for both internal transcribed spacers of these rodents [3]. Some of these common tracts could have arisen by chance within a region which seems submitted to some constraint for a high pyrimidine content in the 3 vertebrates, mouse, rat [12] and *Xenopus* [10]. However, this appears less likely for the common stretch (−5, −18) which could be more directly relevant to the excision process of ETS sequences during ribosomal maturation.

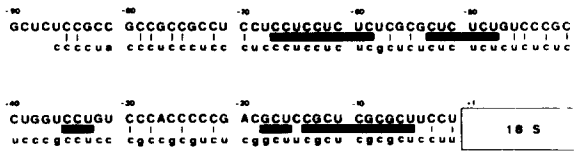

Fig.3. Sequence of the mouse ribosomal external transcribed spacer, immediately upstream 18 S rRNA (capital letters) and comparison with rat (lower-case letters) [12]. Vertical bars denote identical nucleotides. Homologous tracts (at least 3 consecutive nucleotides) are indicated by a thick horizontal bar.

It is worthwhile mentioning that this ETS domain shares common features, in terms of nucleotide content, with the ITS 2 domain which is immediately upstream 28 S rRNA sequence [3]: for both regions, the high pyrimidine content (particularly in C) also applies to rat [12] and *Xenopus* [10]. Such pyrimidine-rich tracts located immediately upstream mature rRNA sequences could represent recognition signals either during transcription or cleavage of rRNA precursors.

3.3. *Potential folding of RNA precursor around the 5′ terminus of 18 S rRNA*

In *E. coli*, rRNA precursor-specific sequences which flank immediately mature 16 S rRNA are able to form a long double helical structure whose hairpin loop contains the whole 16 S molecule [18]. Mouse pre-rRNA sequences flanking mature 18 S rRNA have been searched in view of a potential analogous interaction between ETS and ITS 1 sequences [3] immediately downstream 18 S rRNA. The most stable basepairings we could detect between these regions are shown in fig.4A. Such a folding pattern could significantly contribute to stabilizing pre-rRNA structure (ΔG = −71 kcal, as in [19]) and it is noteworthy that it contains a 9 bp long perfect duplex. Homologous regions in rat pre-rRNA [12] were examined in a similar way. The folding pattern we have come up to (fig.4B) is markedly different from the model proposed in [12]: it is also characterized by a much higher stability (ΔG = −90 kcal, whereas the contribution of the same regions is only −25 kcal in [12]). In this case again, a 9 bp long perfect duplex can be formed (fig.4B). It is interesting to note that both termini of mature rRNA, which have remained free for basepair interactions in the secondary structure models previously proposed for prokaryotic 16 S and eukaryotic 18 S rRNA [20,21], can form a 6 bp duplex (shown by bars in fig.4) in all eukaryotes (while a 5 bp duplex in identical

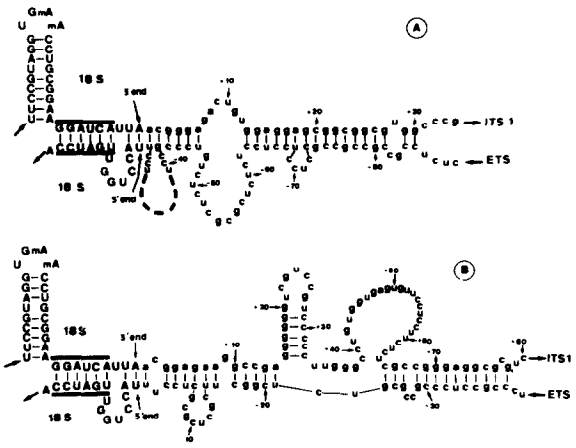

Fig.4. Potential base-pairing between 18 S rRNA flanking sequences in pre-rRNA for mouse (A) and rat (B). Terminal nucleotides of mature 18 S rRNA are represented by capital letters, while nucleotides in external (ETS) or internal (ITS 1) spacers are shown in lower-case letters. Positions in each spacer are numbered from the junction with mature 18 S rRNA sequence. Arrows indicate 5′ to 3′ direction.

positions can also be proposed in *E. coli* 16 S rRNA).

Obviously, the structures shown in fig.4 could be directly relevant to the recognition signals involved in the cleavages of rRNA precursor. However, it must be stressed that the folding patterns for the 18 S rRNA flanking regions are far from being closely homologous between both rodents: potentially interacting regions map at clearly different locations by reference with the termini of mature rRNA. Moreover, this secondary structure folding is not common to more distantly related eukaryotes: no significant basepairing can be proposed for homologous regions in xenopus [10] and only very limited complementarity is observed in yeast [22]. It is also noteworthy that the basepairing potential between ETS and ITS 1 sequences in both rodents is much lower than in *E. coli* precursors around RNase III cleavage sites.

The early cleavage reaction which generates 41 S rRNA precursor by removing ETS regions has been shown by electron microscopy analysis to occur near the 5′ terminus of mature 18 S rRNA [23]. However, it is not known so far whether this cleavage takes place precisely at this terminus or several nucleotides upstream (then followed by a further trimming of 41 S rRNA precursor). The precise identification, at sequence resolution, of the 5′ terminal region of 41 S pre-rRNA (now under way in our lab) should provide further insight into the potential significance of the structural features shown in fig.4 in terms of RNA processing signals.

ACKNOWLEDGEMENTS

We thank I. Grummt for providing us with pMSE 2 recombinant. Support and encouragement of Professor Zalta were highly appreciated. We are grateful to J. Feliu for excellent technical assistance. This work was financially supported by a grant from CNRS (ATP: Biologie Moléculaire du Gène), by ADRC contract no.6107 and by a fellowship from Ligue National Française Contre le Cancer to B.M.

REFERENCES

[1] Perry, R.P. (1976) Annu. Rev. Biochem. 45, 605–629.
[2] Michot, B., Bachellerie, J.P., Raynal, F. and Renalier, M.H. (1982) FEBS Lett. 142, 260–266.
[3] Michot, B., Bachellerie, J.P. and Raynal, F. (1983) Nucleic Acids Res. 11, 3375–3391.
[4] Michot, B., Bachellerie, J.P. and Raynal, F. (1982) Nucleic Acids Res. 10, 5273–5283.
[5] Michot, B., Bachellerie, J.P., Raynal, F. and Renalier, M.H. (1982) FEBS Lett. 140, 193–197.
[6] Bachellerie, J.P., Michot, B. and Raynal, F. (1983) Molec. Biol. Rep. 9, 79–86.
[7] Skryabin, K.G., Zacharev, V.M., Rubtsov, P.M. and Bayev, A.A. (1979) Dokl. Akad. Nauk. SSSR 247, 1278–1280.
[8] Rubstov, P.M., Musakhanov, M.M., Zakhariev, V.M., Krayev, A.S., Skryabin, K.G. and Bayev, A.A. (1980) Nucleic Acids Res. 8, 5779–5794.
[9] Salim, M. and Maden, B.E.H. (1981) Nature 291, 205–208.
[10] Maden, B.E.H., Moss, M. and Salim, M. (1982) Nucleic Acids Res. 10, 2387–2398.
[11] Furlong, J.C. and Maden, B.E.H. (1983) EMBO J. 2, 443–448.
[12] Cassidy, B.G., Subrahmanyan, C.S. and Rothblum, L.I. (1982) Biochem. Biophys. Res. Commun. 107, 1571–1576.
[13] Grummt, I., Soellner, C. and Scholz, I. (1979) Nucleic Acids Res. 6, 1351–1369.
[14] Maxam, A.M. and Gilbert, W. (1980) Methods Enzymol. 65, 499–560.

[15] Bina-Stein, M., Thoren, M., Salzman, N. and Thompson, J. (1979) Proc. Natl. Acad. Sci. USA 76, 731–735.
[16] Youvan, D.C. and Hearst, J.E. (1979) Proc. Natl. Acad. Sci. USA 76, 3751–3754.
[17] Lockard, R.E., Connaughton, J.F. and Kumar, A. (1982) Nucleic Acids Res. 10, 3445–3457.
[18] Young, R.A. and Steitz, J.A. (1978) Proc. Natl. Acad. Sci. USA 75, 3593–3597.
[19] Tinoco, I. jr, Borer, P.N., Dengler, B., Levine, M.D., Uhlenbeck, O.C., Crothers, D.M. and Gralla, J. (1973) Nature New Biol. 246, 40–41.
[20] Noller, H.F. and Woese, C.R. (1981) Science 212, 403–411.
[21] Stiegler, P., Carbon, P., Ebel, J.P. and Ehresmann, C. (1981) Eur. J. Biochem. 120, 487–495.
[22] Veldman, G.M., Klootwijk, J., Van Heerikhuizen, H. and Planta, R.J. (1981) Nucleic Acids Res. 8, 4847–4862.
[23] Wellauer, P.K. and Dawid, I.B. (1974) J. Mol. Biol. 89, 379–395.