Structure of mouse rRNA precursors. Complete sequence and potential folding of the spacer regions between 18S and 28S rRNA

Bernard Michot, Jean-Pierre Bachellerie* and Françoise Raynal

Centre de Recherche de Biochimie et de Génétique Cellulaires du C.N.R.S., 118, route de Narbonne, 31062 Toulouse Cedex, France

Received 22 March 1983; Revised and Accepted 26 April 1983

ABSTRACT

We have determined the complete nucleotide sequence of the regions of mouse ribosomal RNA transcription unit which separate mature rRNA genes. These internal transcribed spacers (ITS) are excised from rRNA precursor during ribosome biosynthesis. ITS 1, between 18S and 5.8S rRNA genes,is 999 nucleotides long. ITS 2, between 5.8S and 28S rRNA genes, is 1089 nucleotides long. Both spacers are very rich in G + C, 70 and 74 % respectively. Mouse sequences have been compared with the other available eukaryotes : while no homology is apparent with yeast or xenopus, mouse and rat ITS sequences have been largely conserved, with homologous segments interspersed with highly divergent tracts. Homology with rat is much more extensive for ITS 1 than for ITS 2. Tentative secondary structure models are proposed for the folding of these regions within rRNA precursor; they are closely related in mouse and rat.

INTRODUCTION

The formation of mature 18S, 5.8S and 28S rRNA in eukaryotes involves the processing of a long common precursor molecule (45S RNA - 12.5 kbases - in mammalian cells) through a series of endonucleolytic cleavages (1). Within rRNA primary transcript, mature rRNA sequences are separated by two long segments of precursor-specific sequences, called internal transcribed spacers : ITS 1, between 18S rRNA and 5.8S rRNA, and ITS 2, between 5.8S and 28S rRNA. The characterization of these regions should bring insight into the recognition signals involved in rRNA processing steps and help to understand the role of precursor-specific sequences in the control of ribosome production. Eukaryotic ribosomal ITS have been completely sequenced so far in yeast (2, 3, 4), xenopus (5), and recently in rat (6). Whereas the primary structure of terminal domains of mature rRNAs has been strongly conserved during evolution (6-11), the comparison of these ITS sequences revealed no significant homology between the three species, except for a few nucleotides adjacent to the junctions with mature rRNAs (6). In the present study, we report the

complete primary sequence of both internal transcribed spacer regions from mouse rRNA gene. Contrarily to what observed so far between more distant eukaryotic species, large domains of ITS sequences appear conserved when mouse and rat are compared. Tentative secondary structure models have been derived and compared for both rodents.

MATERIALS AND METHODS

- Recombinant DNA : The 3.7 kb EcoRI-BamHI fragment of mouse ribosomal DNA containing the 3'terminal domain of 18S rRNA, internal transcribed spacers, 5.8S rRNA and 5'terminal domain of 28S rRNA was inserted into the (EcoRI + BamHI) cleaved plasmid pBR 322 giving rise to a pMEB 3 recombinant plasmid. Isolation, restriction endonuclease analysis and sequencing of cloned DNA were carried out as described previously (11, 12). Chemical DNA sequencing was performed according to Maxam and Gilbert (13). Termini of mature rRNAs adjacent to internal transcribed spacers were identified as reported previously (9, 11, 12). Biohazards associated with the experiments were pre-examined by the French Control Committee.

RESULTS AND DISCUSSION

The DNA sequence :

The restriction map of pMEB 3 recombinant mouse rDNA clone is shown in Fig. 1 for the different enzymes used for generating DNA fragments labelled at one 5' terminus. Determinations were confirmed by sequencing both strands for about 80 % of DNA length, especially whenever any indication of band compression, due to secondary structure effects on sequencing gels, was detected on one strand. Results were corroborated by extensive overlaps along the whole sequence, in order to ensure that no very small DNA fragment limited by identical restriction sites had been overlooked. No ambiguity remained in the sequences shown in Fig. 2. The sequences of adjacent mature mouse RNA domains have been reported recently for 18S rRNA (9), 5.8S and 28S rRNA (11) and junctions with ITS regions have been identified at sequence resolution.

Internal transcribed spacer 1 is 999 nucleotide long whereas ITS 2 contains 1089 nucleotides.

General features of ITS sequences

Both mouse ITS sequences share common features : they are very rich in G + C, 70 % and 74.3 % for ITS 1 and ITS 2 respectively (Fig. 3). Moreover, it seems worthwhile mentioning that, for each ITS, the number of G equals exactly the number of C (350:350 in ITS 1, 407:407 in ITS 2). Both ITS are parti-

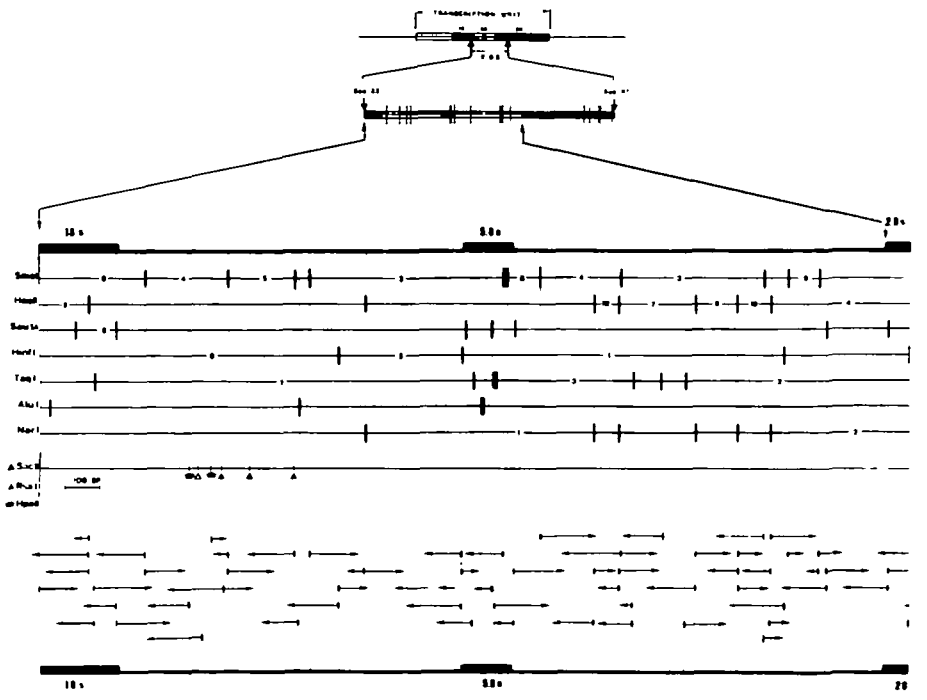

Fig. 1. Top - The mouse rDNA transcription unit and the location of rDNA fragment cloned in pMEB 3 recombinant, with SmaI restriction map (vertical bars).
Bottom - Detailed restriction map of the 18S-28S intergenic region and sequencing strategy. Starts of arrow indicate (^{32}P) labelled 5' ends. The lengths of arrows are indicative of the extent of sequence read. For HpaII, RsaI and SacII, the only sites which are shown correspond to those used for preparing labelled DNA fragments.

cularly deficient in A : 7.1 % and 6.2 % for ITS 1 and ITS 2 respectively. These general base composition features clearly distinguish ITS sequences from intervening 5.8S rRNA and from adjacent domains of 18S and 28S mature rRNAs (Fig. 3). Transitions in base composition occur abruptly at the mature rRNA/ITS boundaries. When large subdomains (longer than 100 nucleotides) of each ITS are considered, base contents do not differ dramatically from entire spacer values. However, the 160 nucleotide long 3'terminal region of ITS 2 possesses unique composition features : this domain, which is located immediately upstream the 5'terminus of mature 28S rRNA, is particularly rich in

pyrimidine (71.2 %, instead of 56.6 % for total ITS 2) with a very high C content (50 %).

Some polypurine and polypyrimidine tracts significantly larger than expected on the sole basis of random chance occurrence are present in both ITS (Table 1), with two very long polypyrimidine stretches in ITS 1 (26-nucleotide long and 22-nucleotide long starting at 176 and 826 respectively).

Both ITS sequences were also searched for the presence of alternating purine-pyrimidine sequences which can allow the appearance of left-handed

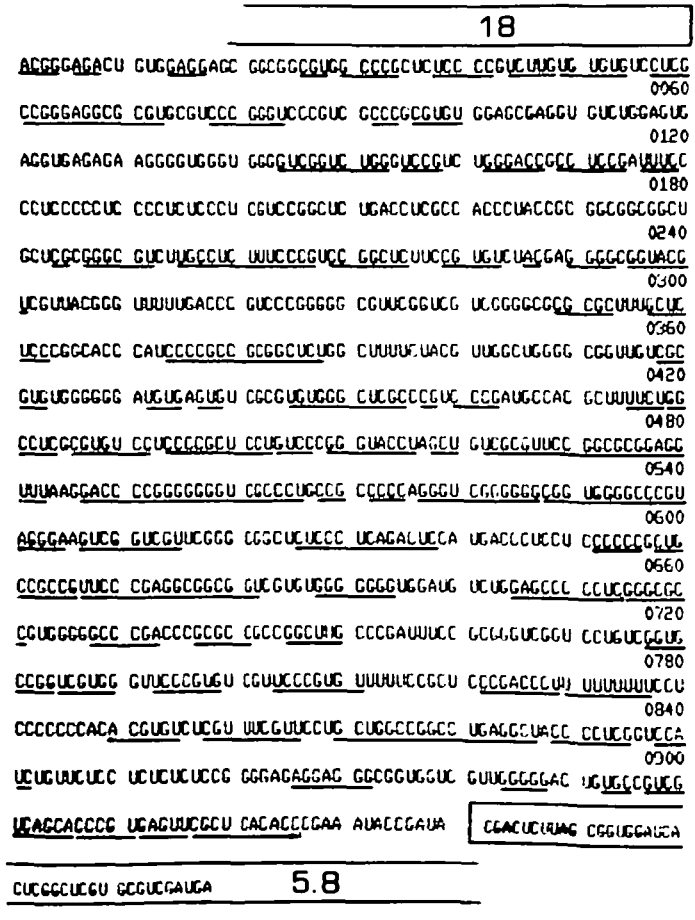

Fig. 2A. Sequence of the 18S-5.8S intergenic region of mouse rDNA. Boxed regions depict 18S and 5.8S rRNA sequences. The sequence (non-coding strand) is numbered from the first nucleotide following the 3' terminus of 18S rRNA.

helical conformations in DNA (14). As shown in Table 1, the frequency and size of such alternating stretches does not exceed what would be expected for a random distribution. The same holds true for the presence of internal homologous regions within each ITS : a computer analysis of the size and location of homologous tracts shows no indication of any internal repetition

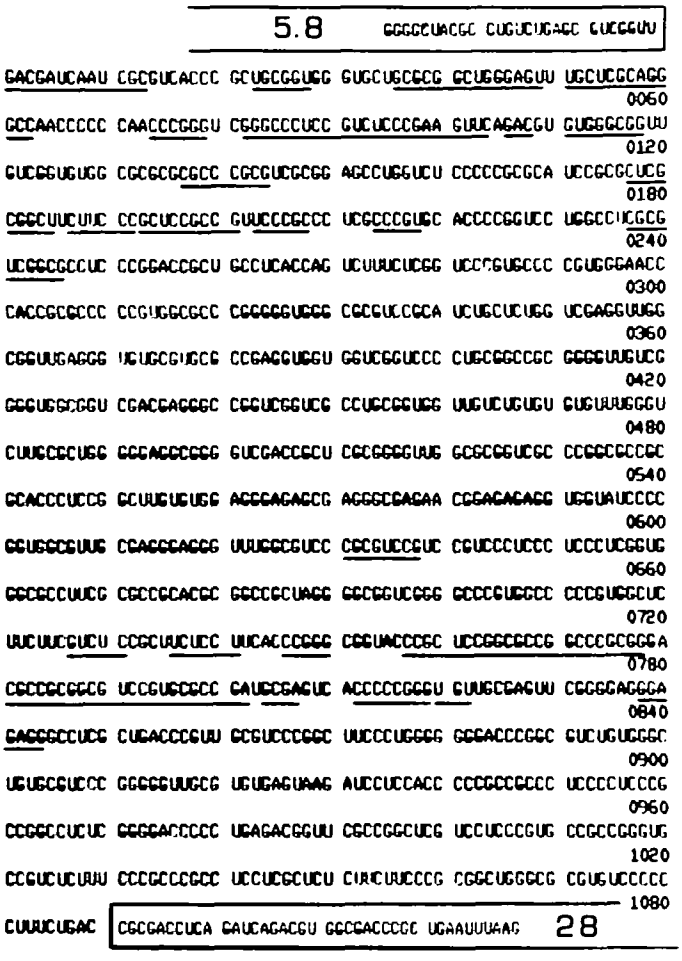

Fig. 2B. Sequence of the 5.8S-28S intergenic region of mouse rDNA. Boxed regions depict 5.8S and 28S rRNA sequences. The sequence (non-coding strand) is numbered from the first nucleotide following the 3'terminus of 5.8S rRNA.

For both sequences, blocks of perfect homology with rat internal transcribed spacers (6) are underlined.

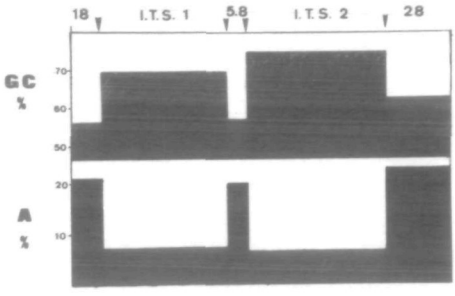

Fig. 3. Nucleotide composition of internal transcribed spacers and adjacent domains of mature rRNA in mouse pre-rRNA. Composition is expressed as percent of total ITS 1, 5.8S rRNA (11) and ITS 2 nucleotides. For 18S rRNA and 28S rRNA, the 231 3'terminal nucleotides (9) and the 500 5'nucleotides (11) were considered respectively.

for each spacer. Moreover, both mouse ribosomal ITS do not share statistically significant sequence homology.

The potential presence of tRNA genes within mouse rRNA ITS was also examined : sequences surrounding a GUUC motif (GTΨC invariant in tRNA) were searched for secondary structure folding consistent with a tRNA structure. In all cases (9 GUUC in ITS 1 and 4 in ITS 2), no tRNA-like structure was apparent, contrarily to what is observed in mitochondrial or prokaryotic rDNAs (15, 16). This result is in line with previous reports showing the absence of tRNA coding sequences within ribosomal ITS of two other eukaryotes, yeast (2-4) and xenopus (5).

It may be worthwhile mentioning that two relatively short open reading frames are present within mouse ITS 1, extending from nucleotide 464 to 640

Table 1 : Distribution of polypurine, polypyrimidine and alternating purine/pyrimidine tracts in mouse ribosomal ITSs. Expected numbers were calculated on the basis of random chance occurrence considering the nucleotide composition of each mouse ribosomal ITS.

	ITS 1			ITS 2		
	size (nucleotides)	observed number	expected number	size (nucleotides)	observed number	expected number
longest polypyrimidine tracts	26	1	.0006	14	1	0.35
	22	1	.006	13	2	0.62
	15	1	.26	12	1	1.1
	13	1	.79			
longest polypurine tracts	13	1	.013	14	1	0.01
	11	1	.077	10	1	0.28
	7	4	2.4	9	2	0.63
longest stretches of alternating purine/pyrimidine	10	1	0.87	12	1	0.24
	8	1	3.5	10	1	0.97
				8	2	4

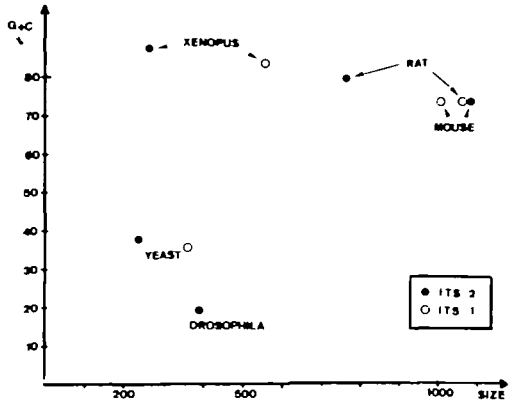

Fig. 4. Size and base content of eukaryotic ribosomal ITS : yeast (2-4), Drosophila melanogaster (B. Jacq, personal communication), Xenopus laevis (5) and rat (6).

and from 698 to 880 respectively. Conversely, the only AUG triplet in ITS 2 (position 802) is not followed by an in-phase termination codon.

Phylogenetic comparisons

Whereas mature rRNAs have been strongly conserved during evolution, ribosomal ITS regions have evolved much more rapidly and show dramatic differences in size and G + C content among distant eukaryotes (Fig. 4).

It is remarkable that, for the four eukaryotic species now available for complete comparison of their ITS, the G + C content of ITS 1 is very close to the G + C content of ITS 2 for a given species. Moreover, for the three vertebrates, ITS have all a high G + C content (ranging from 70 to 88 %) contrarily to what is observed in yeast and an insect (ITS 2 sequence in Drosophila melanogaster, B. Jacq personal communication). Sequence data previously reported for portions of Drosophila melanogaster ITS 1 (17, 18) and for portions of both ITS in Sciara coprophila (19) also indicate a very low G + C content in these two insects.

When rat, xenopus and yeast ITS sequences are compared no significant homology can be detected (6). Earlier studies, involving heterologous nucleic acid hybridizations, had pointed previously to the high degree of variability of ITS sequences during evolution, even between related xenopus species (20).

When mouse ribosomal ITS sequences are compared with yeast and xenopus a complete lack of homology is also observed, except for a short segment located immediately downstream 5.8S rRNA which is conserved between the two vertebrates. We have recently proposed for this segment a direct role in the excision process of ITS 2 from 32S rRNA precursor, through extensive base-pairing with U3 nucleolar RNA (21).

As opposed to what has been found so far between more distant eukaryo-

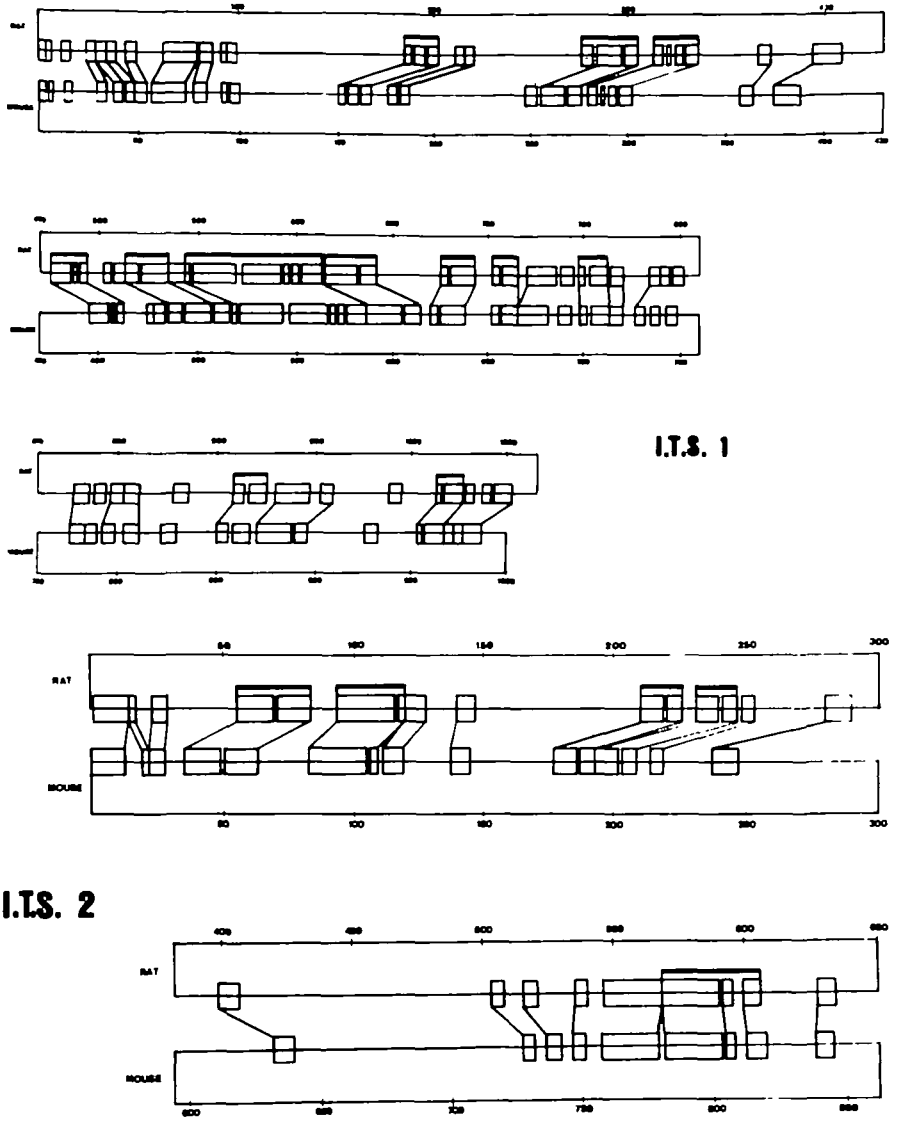

Fig. 5. Location of blocks of homology in mouse and rat rDNA internal transcribed spacers.
(A) : ITS 1. (B) : ITS 2.
Pairs of homologous blocks in both species are represented by boxes connected by a line. Their positions along each ITS sequence are indicated on the upper (rat) or lower (mouse) lines respectively. Strings of in-phase adjacent blocks of homology are overlined by thick bars. Large portions of ITS 2 where no significant block of homology could be detected are not shown here.

tes, the comparison of mouse and rat ITS sequences reveals the presence of extensive homologies along the spacers of the two rodents (Fig. 2 A, B, Fig. 5). It seems noteworthy that the distribution of homologous sequences is far from being uniform along the ribosomal spacers.

Firstly, it appears clearly that ITS 1 sequence has been conserved to a much larger extent than ITS 2 (Fig. 5) : large domains of ITS 2 (particularly regions 247-630, 845-1089 in mouse) amounting to about 70 % of total length are devoid of any significant homology. On the contrary, regions of mouse ITS 1 which are more than 75 % homologous with rat represent about 70 % of total spacer length. The conservation is particularly high in the central region of ITS 1, with segment 475-615 (in mouse sequence) about 90 % homolog to its rat counterpart.

Secondly, it is remarkable that long domains of very high homology are frequently interspersed with regions of very low or no homology. Moreover the size of these divergent tracts has not been generally conserved, with occasional very large variations : for example, the counterpart of a 10 nucleotide long divergent tract in mouse (starting position 616 in ITS 1) is a 38 nucleotide long segment in rat (starting position 643). There is not a unique trend in the size variations of these divergent regions : some are larger in mouse, others are larger in rat. As a result, the total length of ITS 1 is slightly different in both species (999 nucleotides in mouse vs. 1066 in rat). However a much larger discrepancy is observed for ITS 2 size : 1089 nucleotides in mouse vs. 765 in rat. From the location of homologous regions (see Fig. 5), it is obvious that most of this extra length may be ascribed to the central domain of mouse ITS 2, a 384 nucleotide long region starting position 247 (its counterpart in rat sequence is only 108 nucleotide long, starting position 290). It is noteworthy that the probing of this region in mouse chromosomal rDNA by Southern blot hybridization did not reveal any difference with pMEB 3 cloned mouse rDNA (in preparation). A large increase in size is also observed for the 3'terminal divergent region of mouse ITS 2 (246 nucleotide long, starting position 844) as compared with rat (133 nucleotide long, starting position 633). However, despite extensive differences in sequence and size in this area of ITS 2, a long domain immediately upstream the 5' terminus of 28S rRNA shares common features in both species : the 125 nucleotide long terminal segment of rat ITS 2 contains 80 % pyrimidine (with 55 % C) while the 160 nucleotide long terminal segment of mouse ITS 2 is also very rich in pyrimidine (71 %, with 50 % C).

The pattern of interspersion of highly homologous tracts with divergent

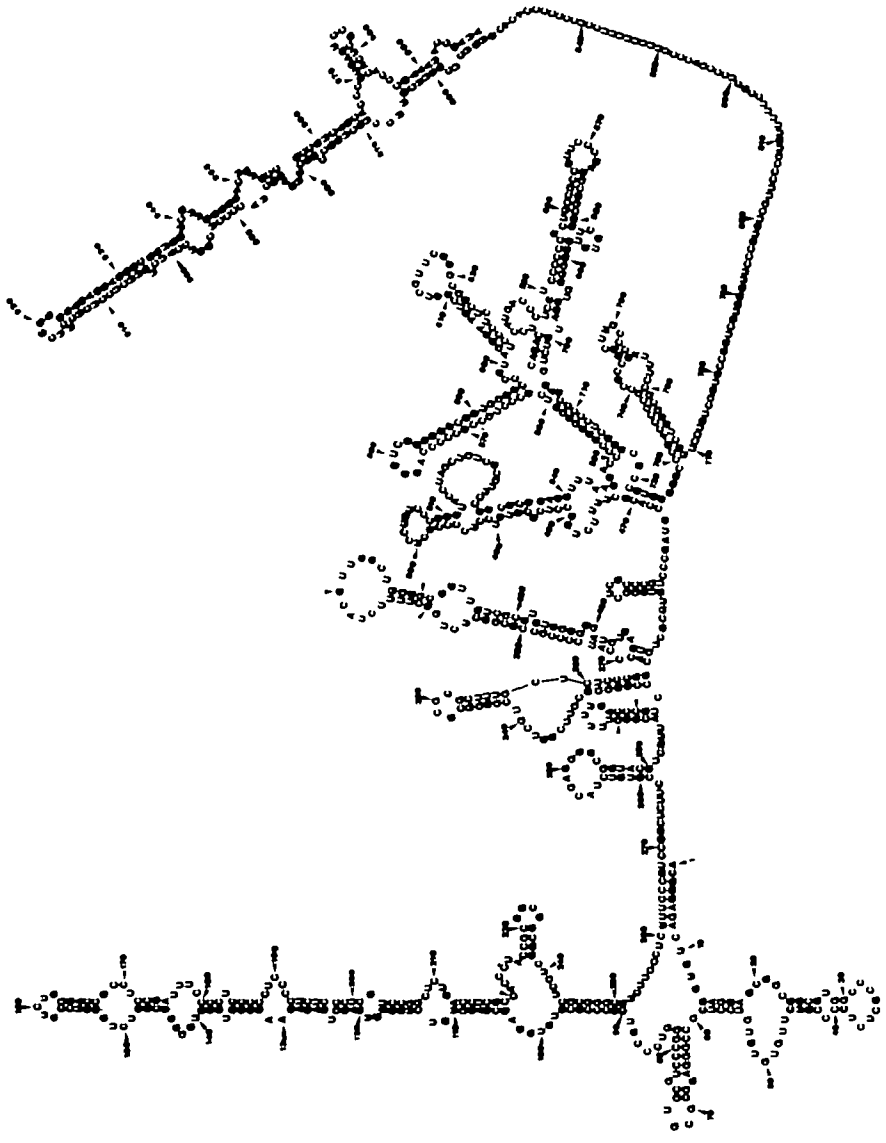

Fig. 6A : Potential secondary structure folding of mouse ribosomal ITS 1.

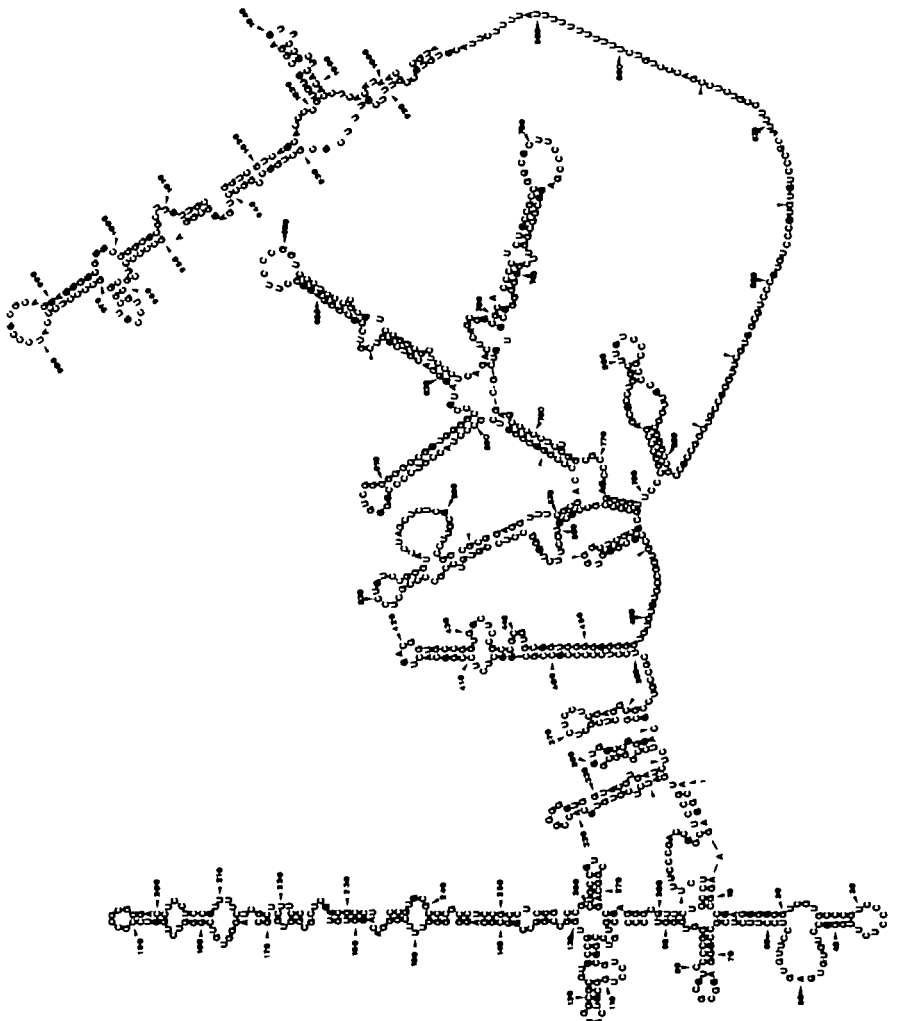

Fig. 6B : Potential secondary structure folding of rat ribosomal ITS 1 using the sequence published by Busch & coworkers (6).

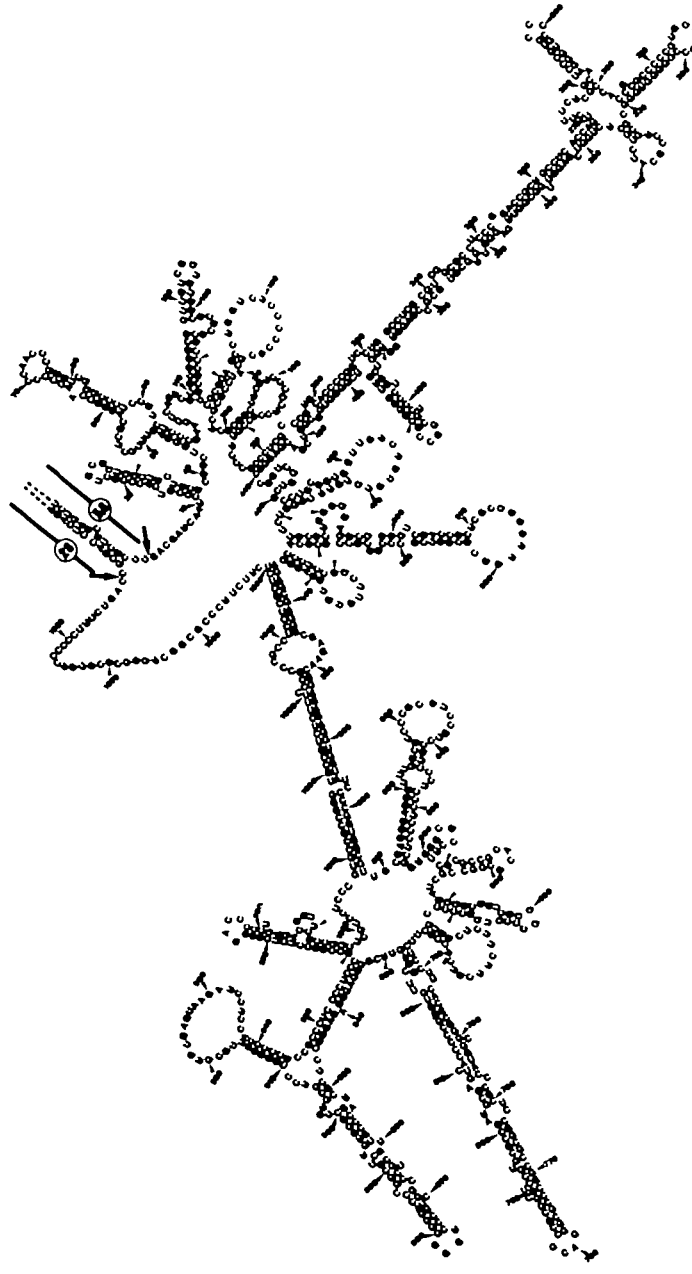

Fig. 7A : Potential secondary structure folding of mouse ribosomal ITS 2. Adjacent regions of mature 5.8S and 28S rRNA, which are held together by stable base-pairings (11), are also shown.

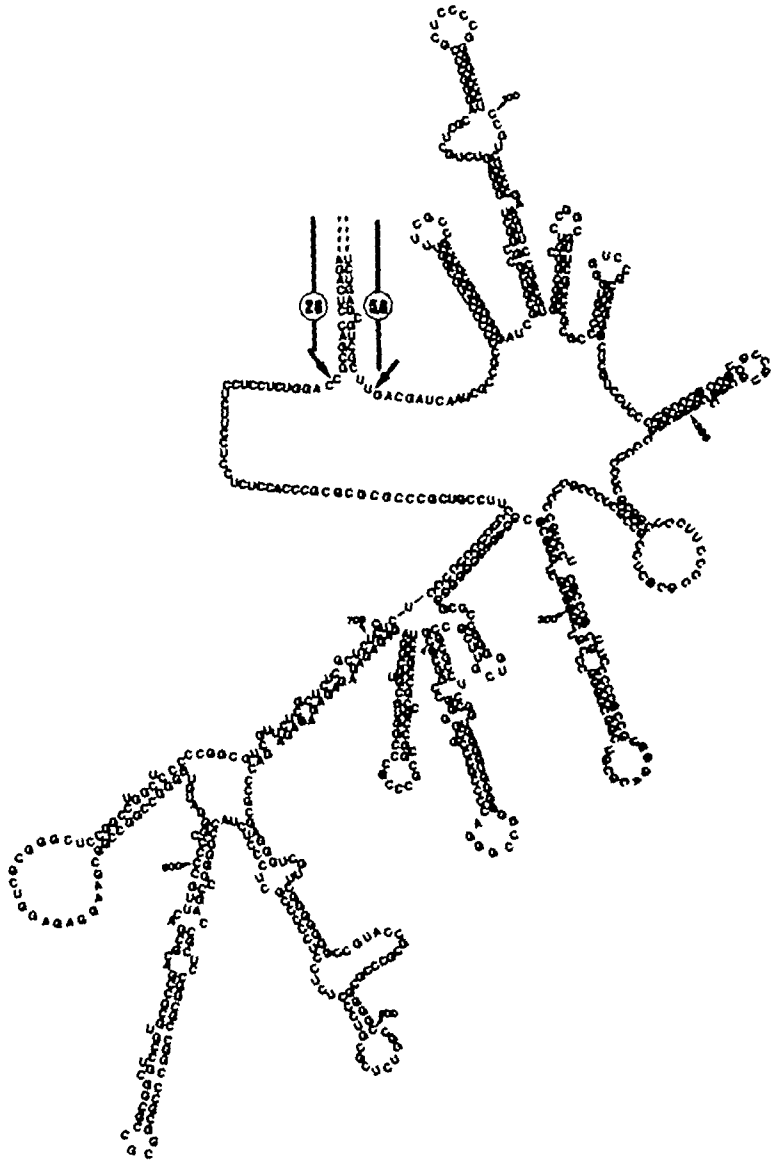

Fig. 7B : Homologous folding of rat ITS 2 using the recently published sequence (6).

regions between mouse and rat ITS is clearly reminiscent of what has been observed for mature rRNAs when phylogenetic comparisons are carried out between very distant species, like yeast, xenopus and mouse (22, 9, 11). Quanti-

tatively, it must be noted however that overall homology for ITS sequences between the two rodents is still lower than for mature rRNA sequences between mouse and yeast (9, 11 and our unpublished results).

Finally the pattern of distribution of homologies between mouse and rat suggests that definite areas of ITS which are submitted to a stronger selective pressure could have some functional roles, probably in the control of ribosome biogenesis.

Secondary structure of ITS region :

As expected from their high G + C values and from a balanced content in G and C, both mouse ITS show a high potential for folding into stable secondary structures. Obviously, their large size makes it difficult to select a folding pattern on the sole basis of sequence data. The tentative structure models shown in Fig. 6 and 7 were constructed according to the following set of rules :

- For duplex of roughly similar stability (23), short range base pairings were always given precedence over alternate interactions involving more distant regions.
- Long range base-pairing (i.e. with an intervening region longer than 25 nucleotides) were not considered as primary determinants of the folding pattern, unless favoured by a prior "shortening" of the intervening region through short range base-pairing(s).
- Topological knots were not allowed for the overall structure (i.e. a tract located between two base-paired segments could not interact with outer regions), in view of the absence of knots in the secondary structure of E. coli 16S mature rRNA (24).
- Rat ITS sequences (6) were processed in a similar way. Considering that secondary structure features in rRNA have been even more strongly conserved than primary sequences during evolution (24, 8, 9, 11), whenever alternate structures could be proposed for one species, selection was effected after comparing with the other species, so as to maximize the structural homology.

Obviously, these tentative models will have to be refined in the light of further comparative sequence analyses and of structure probing experiments to come. Nevertheless, it is remarkable that, in their present form, the folding patterns which have emerged for mouse and rat are closely related, with most pairs of homologous tracts along ITS sequences mapping at equivalent positions by reference to secondary structure features. Some major characteristics seem worthwhile mentioning here :

- ITS 1 (Fig. 6) : three large domains are "clipped" by long-range interactions.

It is noteworthy that the long range duplex limiting the first of these three domains (nucleotides 1-268 in mouse) involves in both species the hexanucleotide immediately adjacent to the 3'terminus of mature 18S rRNA, which could be relevant to the recognition process for pre rRNA cleavage. Despite two base changes in this hexanucleotide, a perfect pairing is still maintained in rat, not by compensatory changes in equivalent positions (as judged from the location of blocks of homology) but through the presence of a matching sequence ten nucleotide positions downstream in an adjacent divergent tract.

The second domain (nucleotides 467-725 in mouse) corresponds to the most highly conserved area of ITS 1. This is also the most energetically stable part of the structure, particularly for the branched subdomain corresponding to nucleotides 550-717 (with $\Delta G = -0.81$ kcal per nucleotide, according to Tinoco's rules (23)). Some of its structural features are supported by comparative sequence analysis with rat : despite two base changes (^{567}G (mouse) $\longrightarrow^{593}A$ (rat) and ^{571}C (mouse) $\longrightarrow^{597}U$ (rat)) a 14 bp stem is maintained through an equivalent change of a GU and a GC pairs into a AU and GU pairs respectively. As for stem involving residues 601-610 and 621-631 in mouse, the only two modifications correspond to a compensatory base change ($^{606}A : {}^{625}U \longrightarrow {}^{633}C : {}^{680}G$ in rat). This stem is dramatically extended in rat, consecutive to the presence of a much longer divergent tract which can be folded into a very stable helix starting at position 644.

- ITS 2 (Fig. 7) : in both species, the 3'half of ITS 2 is condensed into a highly structured domain closed by a series of extensive long range interactions. In the 5'half, a number of long hairpins also contribute to bring in close vicinity both distal regions of ITS 2, which could be a prerequisite for the build-up of the binary complex between mature 5.8S and 28S rRNA, the structure of which was recently demonstrated by comparative sequence analysis between mouse (11) and yeast (2).

From examination of xenopus sequences, some complementarity between ITS 1 and ITS 2 regions flanking 5.8S rRNA was reported (25) which could provide recognition signals for rRNA processing. Such a base-pairing should also stabilize the transient interaction which is likely to occur during transcription between 5' and 3' termini of 5.8S rRNA before 5'terminal sequences of 28S rRNA are available for formation of the binary 5.8S-28S rRNA complex (26). Such a complementarity is not observed in mouse ITS sequences.

Since ITS 2 region appears as an eukaryotic insertion within the ancestral gene coding for large rRNA (25, 26), its removal from 32S pre-rRNA appears related to the excision of introns, except that ends of mature rRNAs

are maintained in a stable duplex instead of covalent religation. A better knowledge of the structural basis for this excision process should be gained from additional ITS 2 sequence data from closely related mammalian species, particularly in relation with the specific role we recently proposed for U3 RNA in this reaction, through base-pairing with distal regions of ITS 2 (21).

Acknowledgements :
We are grateful to Professor J.P. Zalta for support and encouragement. The help of J. Feliu in the preparation of plasmid DNA was highly appreciated. This work was financially supported by a grant from CNRS (ATP : "Biologie Moléculaire du Gène"), by ADRC contract n° 6107 and by a fellowship from Ligue Nationale Française contre le Cancer to B.M.

Abbreviations used :
ITS, internal transcribed spacer.

*To whom correspondence should be addressed.

REFERENCES

1. Perry, R.P. (1976) Ann. Rev. Biochem. 45, 605-629
2. Skryabin, K.G., Kraev, A.S., Rubstov, P.M. and Baev, A. (1979) Dokl. Akad. Nauk. SSR 247, 761-765
3. Veldman, G.M., Brand, R.C., Klootwijk, J. and Planta, R.J. (1980) Nucleic Acids Res., 8, 2907-2920
4. Veldman, G.M., Klootwijk, J., Van Heerikhuizen, H. and Planta, R.J. (1981) Nucleic Acids Res. 9, 4847-4862
5. Hall,L.M.C. and Maden, B.E.H. (1980) Nucleic Acids Res. 8, 5993-6005
6. Subrahmanyam, C.S., Cassidy, B., Busch, H. and Rothblum, L.I. (1982) Nucleic Acids Res. 10, 3667-3680
7. Hagenbüchle, O., Santer, M., Steitz, J.A. and Mans, R.J. (1978) Cell 13, 551-563
8. Stiegler, P., Carbon, P., Ebel, J.P. and Ehresmann, C. (1981) Eur. J. Biochem. 120, 487-495
9. Michot, B., Bachellerie, J.P., Raynal, F. and Renalier, M.H. (1982) FEBS Lett. 142, 260-266
10. Lockard, R.E., Connaughton, J.F. and Kumar, A. (1982) Nucleic Acids Res. 10, 3445-3457
11. Michot, B., Bachellerie, J.P. and Raynal, F. (1982) Nucleic Acids Res. 10, 5273-5283
12. Michot, B., Bachellerie, J.P., Raynal, F. and Renalier, M.H. (1982) FEBS Lett. 140, 193-197
13. Maxam, A.M. and Gilbert, W. (1980) Methods in Enzym., 65, 499-560
14. Arnott, S., Chandrasekaran, R., Birdsall, D.L., Leslie, A.G.W. and Ratliff, R.L. (1980) Nature 283, 743-745
15. Young, R.A., Macklis, R. and Steitz, J.A. (1979) J. Biol. Chem. 254, 3264-3271
16. Eperon, I.C., Anderson, S. and Neirlich, D.P. (1980) Nature 286, 460-467
17. Pavlakis, G.N., Jordan, B.R., Wurst, R.M. and Vournakis, J.N. (1979) Nucleic Acids Res. 7, 2213-2237

18. Jordan, B.R., Latil-Damotte, M. and Jourdan, R. (1980) FEBS Lett. 117, 227-231
19. Jordan, B.R., Latil-Damotte, M. and Jourdan, R. (1980) Nucleic Acids Res. 8, 3565-3576
20. Brown, D.D., Wensink, P.C. and Jordan, E. (1972) J. Mol. Biol. 63, 57-73
21. Bachellerie, J.P., Michot, B. and Raynal, F. (1983) Molec. Biol. Rep. in press
22. Salim, M. and Maden, B.E.H. (1981) Nature 291, 205-208
23. Tinoco, I. Jr., Borer, P.N., Dengler, B., Levine, M.D., Uhlenbeck, O.C., Crothers, D.M. and Gralla, J. (1973) Nature (London), New Biol. 246, 40-41.
24. Noller, H.F. and Woese, C.R. (1981) Science 212, 403-411
25. Boseley, P.G., Tuyns, A. and Birnsteil, M.L. (1978) Nucleic Acids Res. 5, 1121-1137
26. Kelly, J.M. and Maden, B.E.H. (1980) Nucleic Acids Res. 8, 4523-4534
27. Veldman, G.M., Klootwijk, J., De Regt, V.C.H.F., Planta, R.J., Branlant, C., Krol, A. and Ebel, J.P. (1981) Nucleic Acids Res. 9, 6935-6952
28. Branlant, C., Krol, A., Machatt, M.A., Pouyet, J., Ebel, J.P., Edwards, K. and Kossel, H. (1981) Nucleic Acids Res. 9, 4303-4323.

Note added in proof :

After this manuscript was submitted, partly comparable findings on sequence variation between the transcribed ribosomal spacers of related species have been reported, for Xenopus laevis and Xenopus borealis, by Furlong and Maden (EMBO Journal (1983), 2, 443-448). It is remarkable that among the few short sequence tracts which are homologous between both frog species (three in ITS 1, four in ITS 2), two of them are conserved in mouse : a 14 nucleotide tract CCCGGGUACCUAGC (starting position 506 in mouse ITS 1) is perfectly homologous to Xenopus "tract 1"; "tract 1" in Xenopus ITS 2 is also strongly conserved in mouse, with a pair of identical sequences, GCGCGGCUGGG and UCGCAGGG (starting position 36 and 54 respectively in mouse ITS 2) also present in rat. Considering the very low extent of overall homology within the entire ITS sequences, on one hand between both Xenopus species, on the other hand between mouse or rat and Xenopus, this observation favours the interpretation that at least some of these conserved ITS sequences are functionnally significant and do not represent mere evolutionary relics conserved by chance within regions undergoing rapid divergence.