Volume 10 Number 17 1982 Nucleic Acids Research

Sequence and secondary structure of mouse 28S rRNA 5'terminal domain. Organisation of the 5.8S-28S rRNA complex

Bernard Michot, Jean-Pierre Bachellerie* and Françoise Raynal

Centre de Recherche de Biochimie et de Génétique Cellulaires du CNRS, 118 Route de Narbonne, 31062 Toulouse Cedex, France

Received 22 July 1982; Accepted 14 August 1982

ABSTRACT

We present the sequence of the 5'terminal 585 nucleotides of mouse 28S rRNA as inferred from the DNA sequence of a cloned gene fragment. The comparison of mouse 28S rRNA sequence with its yeast homolog, the only known complete sequence of eukaryotic nucleus-encoded large rRNA (see ref. 1, 2) reveals the strong conservation of two large stretches which are interspersed with completely divergent sequences. These two blocks of homology span the two segments which have been recently proposed to participate directly in the 5.8S-large rRNA complex in yeast (see ref. 1) through base-pairing with both termini of 5.8S rRNA.

The validity of the proposed structural model for 5.8S-28S rRNA complex in eukaryotes is strongly supported by comparative analysis of mouse and yeast sequences : despite a number of mutations in 28S and 5.8S rRNA sequences in interacting regions, the secondary structure that can be proposed for mouse complex is perfectly identical with yeast's, with all the 41 base-pairings between the two molecules maintained through 11 pairs of compensatory base changes.

The other regions of the mouse 28S rRNA 5'terminal domain, which have extensively diverged in primary sequence, can nevertheless be folded in a secondary structure pattern highly reminiscent of their yeast's homolog. A minor revision is proposed for mouse 5.8S rRNA sequence.

INTRODUCTION

Eukaryotic rRNA is transcribed as a long precursor molecule which is converted to mature 18S, 5.8S and 28S rRNA by a series of endonucleolytic cleavages (3). Excision of 18S rRNA sequences is an early event in the ribosomal maturation process while 5.8S and 28S rRNAs remain covalently linked in a common intermediate precursor (32S rRNA in mammalian cells) until internal spacer RNA is removed by processing endonucleases. After this cleavage, it is remarkable that 5.8S and 28S rRNA remain associated by hydrogen bonding throughout ribosome cycle (4, 5).

In an attempt to better understand the structural basis for rRNA processing mechanisms in eukaryotes we have focused our attention on this late processing reaction (32S ⟶ 28S rRNA) and the formation of the

0305-1048/82/1017-5273$2.00/0

5.8S-28S rRNA junction complex in mouse cells. As a prerequisite, the primary structures of these rRNAs have been analyzed. We shall describe elsewhere (Michot et al., in preparation) the structure of 32S rRNA precursor-specific sequences and their potential role in conformational switches involved in rRNA processing.

In the present paper, we report the primary sequence of the 5'terminal 585 nucleotides of mouse 28S rRNA, a domain which has been implicated in the build up of the 5.8S-28S rRNA complex (1, 6-8).

A detailed knowledge of free 5.8S rRNA secondary structure in solution has emerged from a wide array of studies, involving experimental probing by nucleases or chemical modifications and comparative analysis of primary sequences from a variety of eukaryotic species (9-13). However, the interaction of 5.8S rRNA with large rRNA has been much less characterized, particularly due to the scarcity of primary sequence datas for eukaryotic large rRNAs in this domain of the molecule (1, 2, 14). Two major independent binding sites with 28S rRNA, each involving 20-30 nucleotides at the 5' and 3' termini of 5.8S have been suggested from thermal denaturation studies and from structure mapping of the mouse complex (7, 15-17). A secondary structure model for yeast 26S rRNA has been recently proposed (1) on the basis of complete primary sequence determination and on the assumption that main structural features had been conserved as compared with prokaryotic 23S rRNAs (18, 19). In this model, 5.8S rRNA sequences, which show significant homology with 5' end of prokaryotic 23S rRNA (20, 21), interact simultaneously with two areas of large rRNA (5'terminus and a region around position 400 from 5'terminus).

The validity of the first of the two interactions involved in the junction complex model has been recently supported by phylogenetic evidence, since an homologous structure can be proposed for contacts between 3' end of 5.8S rRNA and 5'end of 28S rRNA in mouse (6). By extending our mouse 28S rRNA sequence determinations from 5' terminus, we are able to show here that the same holds true for the second contact region involving 5'terminus of 5.8S rRNA, as demonstrated by a number of compensatory base changes between yeast and mouse. We also propose a secondary structure model for the entire 5'terminal domain of mouse 28S rRNA, which is discussed by comparison with yeast's, its only known complete homolog.

MATERIAL 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 + Bam HI) cleaved plasmid pBR 322, giving rise to a pMEB3 recombinant plasmid. Isolation and analysis of cloned DNA were carried out as described previously (6).

- DNA sequencing : all the DNA purification, 5'(^{32}P) end labelling and sequencing procedures were performed according to Maxam and Gilbert (22), with additional DE-52 cellulose chromatography before chemical DNA cleavages; we used the G (dimethylsulfate), G + A (pyridinium formate, pH 2), T + C (hydrazine, no salt) and C (hydrazine, 2 M NaCl) base specific reactions. After piperidine cleavage, reaction products were run on 20 % and 8 % acrylamide thin (0.4 mm) gels prepared in 50 mM Tris-Boric Acid (pH 8.3) 1 mM EDTA, 7 M Urea.

For most of the regions analyzed in this paper, both strands were sequenced and sufficient overlaps were obtained for separate gel readings and for adjacent fragments. No peculiarity was observed in the sequenced regions, except for two "blanks" in the sequence ladder at positions 72 (on non-coding strand) and 74 (on coding strand) from 28S rRNA 5'terminus. This was correlated to the methylation of C residues, as confirmed by sequence readings on the opposite strand.

The 5'terminus of 28S rRNA was previously mapped by S1 nuclease protection experiment (6).

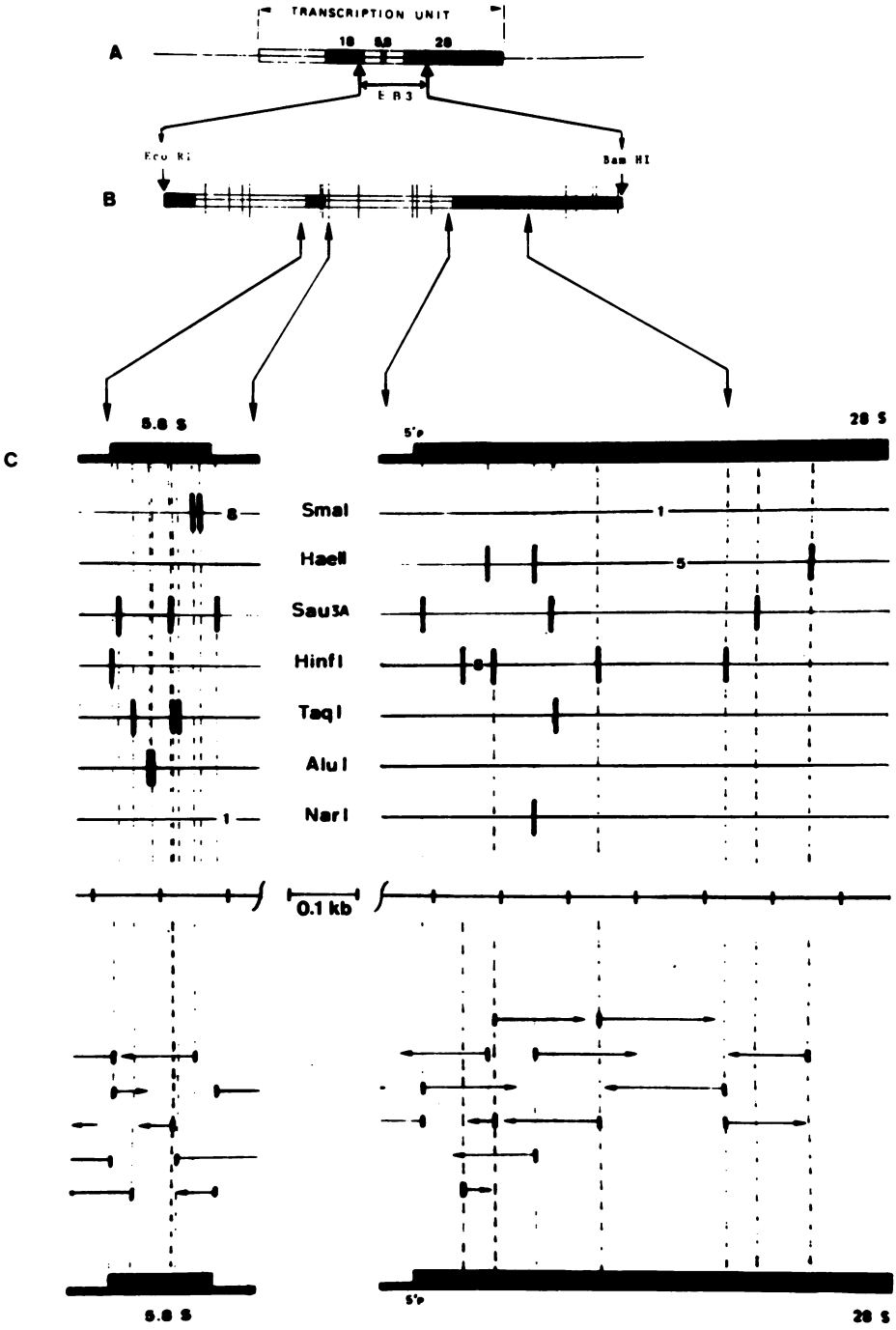
Biohazards associated with the experiments were pre-examined by the French Control Committee.

RESULTS AND DISCUSSION

A. Primary sequence :

- 5'terminal domain of mouse 28S rRNA : We recently reported the primary sequence of approximately 110 nucleotides at the 5' end of mouse 28S rRNA gene (6). We have extended this analysis of mouse 28S rRNA sequence from its 5' terminus in order to encompass the whole 5' terminal domain of the large rRNA molecule which has been proposed recently to participate in the 5.8S-26S rRNA junction complex in yeast (1).

Restriction maps of this region in mouse cloned rDNA and location of sequenced DNA fragments are depicted in Fig. 1. Mouse 28S rRNA gene sequence has been determined up to the HaeII site distal to 5'terminus (see Fig. 1b). Results are shown in Fig. 2. The comparison of mouse sequence with yeast 26S rRNA (1, 2) reveals the presence of two large blocks of high homology


TRANSCRIPTION UNIT
18
5.8
28
A
E B 3
Eco RI
Bam HI
B
C
5.8 S
5'p
28 S
8
1
SmaI
HaeII
Sau3A
HinfI
TaqI
AluI
NarI
5
1
0.1 kb
5.8 S
5'p
28 S

between these two distant species. These conserved regions - from position 5 to 113 (81 % homology) and from position 277 to 430 (83 % homology) - are separated by a long segment the sequence of which has completely diverged. An extensively divergent region is found again beyond position 430, extending through the end of the sequenced domain. This result is in line with previous indirect evidence, involving heterologous hybridizations, indicating the presence of highly conserved regions in eukaryotic rRNAs interspersed with divergent tracts (23-25). Such a pattern has been directly confirmed by recent sequence determinations in the case of small subunit eukaryotic rRNAs (26-30).

It is noteworthy that up to position 430 (in mouse 28S rRNA), the size of 5'terminal domain shows little variation between yeast and mouse, with only 12 additions, all but one located in the divergent tract separating the two blocks of homology. By contrast, the entire large rRNA molecules of the two species exhibit a much larger relative difference in size, with about 4700 nucleotides in mouse (31 and our unpublished data) as compared with 3393 nucleotides for yeast 26S rRNA (1, 2).

When mouse 28S rRNA sequence is compared to E. coli 23S rRNA (32, 19), some homologous tracts can be identified among regions conserved between yeast and mouse (see Fig. 3 for location). Their detection allows us to recognize that 64 insertions have occurred in the divergent region separating the two blocks of "eukaryotic homology" in the 5'terminal domain of mouse as compared to E. coli.

- 5.8S rRNA sequence : A primary sequence for mouse 5.8S rRNA was reported previously (10), on the basis of oligonucleotide fingerprint data and from comparative analysis with rat 5.8S rRNA (33, 9). In view of the technical limitations of this RNA sequencing approach, it seemed useful to check these data by DNA sequencing of the 5.8S rRNA gene which is located within the same mouse ribosomal DNA clone, pMEB 3, as the 5'terminal domain of 28S rRNA. Our results (Fig. 2B) agree well with the sequence proposed from the RNA sequencing study, except for the presence of only one GC doublet, instead of two, around position 50 from 5'terminus. In fact, it was not clear

Fig. 1. (A) The mouse rDNA transcription unit and the location of rDNA fragment cloned in pMEB 3 recombinant. (B) SmaI restriction map of the 3.7 kb EcoRI-BamHI region cloned into pMEB 3. (C) Detailed restriction map of the portion of pMEB 3 analyzed by DNA sequencing in this paper. The position of the 5'-end labelled fragments used for sequencing are indicated by horizontal arrows the lengths of which are indicative of the extent of sequence read.

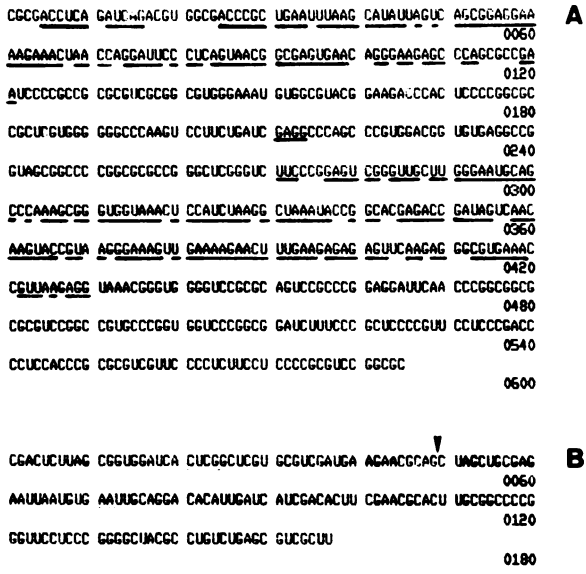


Fig. 2. (A) Sequence of the 5'terminal domain of mouse 28S rRNA. The sequence is numbered beginning at the 5'terminus which was mapped previously by S1 nuclease protection experiment (6). In the two regions showing high homology with yeast 26S rRNA sequence (1, 2), conserved positions are underlined.
(B) Sequence of mouse 5.8S rRNA. Arrow points to discrepancy with previously published data (10).

in that study (10), if the presence of a second GC doublet had been unequivocally demonstrated or if it was merely inferred by comparison with rat 5.8S rRNA, due to the detection of an homologous long partial digestion product of similar oligonucleotide qualitative content. This doublet was not detected any more in a recent study (7), involving gel analysis of mouse L cell 5' end labelled 5.8S rRNA. It is noteworthy that a similar rectification has been recently proposed (14) at the same position for Xenopus laevis 5.8S rRNA, consecutive to DNA sequencing of different rDNA clones (34, 14), as compared to previous RNA sequence established on the basis of fingerprint analysis and from strong conservation of fingerprint patterns among vertebrates (11). We therefore consider likely that this second GC doublet is not present in 5.8S rRNA, taking into account the reliability of DNA sequence determination and the lack of any indication that the cloned mouse DNA could correspond to a non-functional ribosomal gene, as discussed earlier (30). In terms of secondary structure for 5.8S rRNA, this revision has a limited impact, with a slight reduction in a

stem size (5 base pairs instead of 6) and no change at all in the corresponding hairpin loop, when considering the model of Nazar (9) which fits all known 5.8S RNA sequences (35).

B. Secondary structure :

Within the two last years, a consensus on the folding of E. coli 16S (36) and 23S rRNAs (18, 19) has emerged through phylogenetic comparison of primary sequence data and from direct topological probing. The assumption that structural features in rRNAs tend to be conserved during evolution has proven of general validity among prokaryotic species. Moreover conservation of a common basic folding pattern between pro- and eukaryotes is strongly suggested for small ribosomal subunit RNA, on the basis of recent sequence determinations (26-30, 37).

As for the eukaryotic ribosomal subunit RNA, yeast 26S species is so far the only known complete sequence (1, 2) and a secondary structure model has been proposed recently on the basis of comparative analysis with prokaryotic 23S rRNA (1). The determination of a second eukaryotic sequence covering the entire 5'terminal domain of large rRNA provides us with the opportunity to test this model, in terms of phylogenetic stability of its proposed folding pattern. The secondary structure model we have constructed for mouse 28S rRNA 5'terminal domain is shown in Fig. 3.

- 5.8S-28S rRNA junction complex : The idea that 5.8S rRNA is the eukaryotic counterpart of the 5'terminal region of prokaryotic 23S rRNA was first suggested by comparative sequence analysis (20, 21) and later reinforced by results of sequence determination of yeast large rRNA showing that 5' terminal domain of 26S rRNA and 5.8S rRNA could be folded in a binary structure highly reminiscent of prokaryotic 23S rRNA 5' domain (1). In the yeast model, 5' end of 26S rRNA is base-paired with 3' end of 5.8S rRNA while 5' end of 5.8S rRNA interacts with another region of large rRNA, located about 400 nucleotides from 5'terminus (1). We showed recently (6) that the first of these two interactions has been conserved in mouse. As shown in Fig. 3, the same holds true for the entire 5.8S-28S rRNA junction complex. The complex long double helical structure involving 5' end of 5.8S rRNA is perfectly maintained in mouse, with an identical location along the large rRNA sequence.

The biological relevance of the entire structural model for 5.8S-28S rRNA interaction is emphasized by the fact that all mutations in large rRNA interacting areas are exactly compensated by mutations in 5.8S RNA sequence which maintain base-pairing.

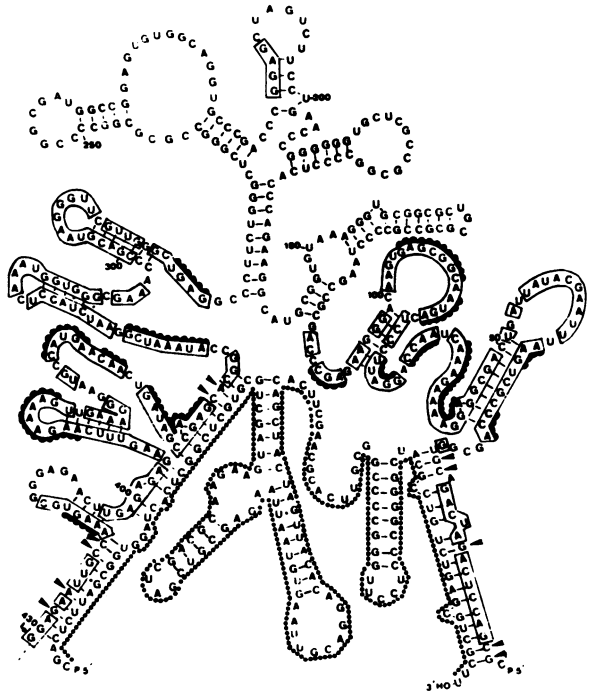


Fig. 3. Secondary structure model of the mouse 28S rRNA 5'terminal domain including junction complex with 5.8S rRNA.

28S rRNA nucleotides which are conserved as compared to 26S yeast rRNA are boxed. Among these boxed sequences, strings of nucleotides (equal or longer than 3 nucleotides) which are conserved as compared to 23S rRNA from two different strains of E. coli (32, 19) are denoted by ▬▬▬.

5.8S rRNA sequence is underlined by a string of points (••••••).

Arrows point to base changes in 28S rRNA that have been compensated by mutations in 5.8S rRNA sequence, as compared to yeast (1, 2).

- Other regions of 28S rRNA 5' domain : Their folding pattern is generally closely related to yeast's rRNA.

It is noteworthy that the region separating the two blocks of homology with yeast can also be folded into two long double helical structures, despite extensive divergence in primary sequence. The first hairpin (region 115-156) has the same length and location than its yeast's counterpart but is much more stable, with a free energy of - 41 kcal. (40) instead of - 9 kcal. for yeast. The second structural feature in this region spans positions 160-274 : it corresponds to a complex clover-leaf-like structure with a stalk constituted by a 9 base-pair long perfect duplex. Its stabi-

lity is high (free energy : - 68 kcal.). It is clearly reminiscent of the branched structure proposed in yeast 26S rRNA, but far from being closely homologous. Obviously the availability of additional large rRNA sequences between less distant eukaryotes should help to better understand the organization of this divergent region and its evolution.

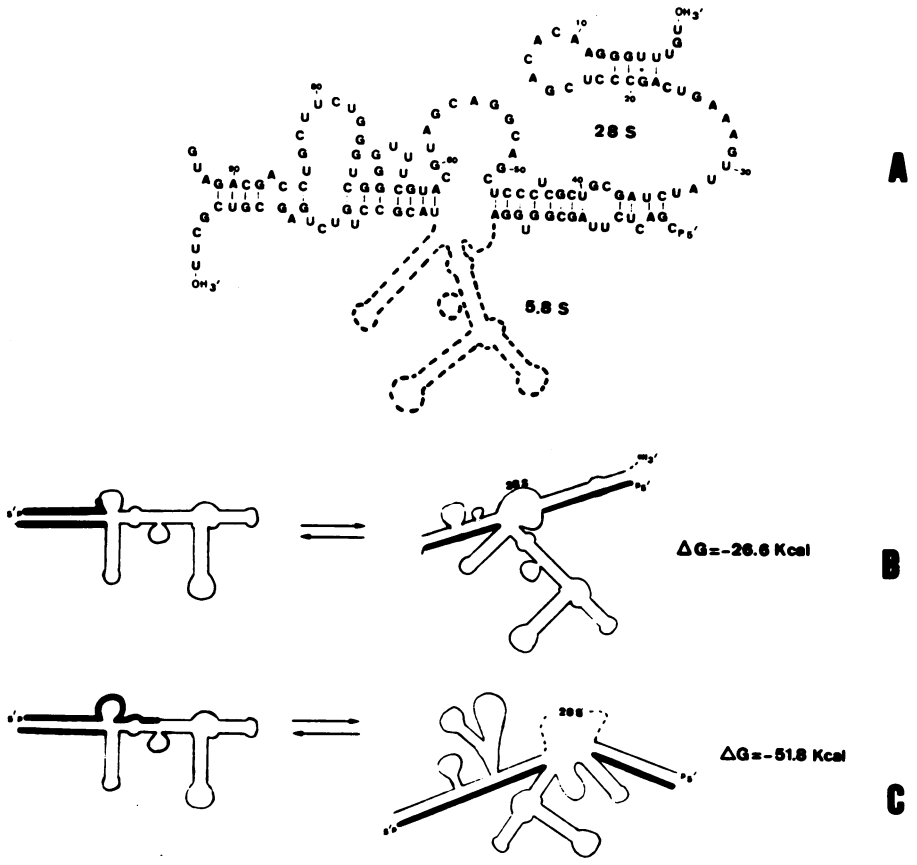


Fig. 4. Comparison of 5.8S rRNA base-pairing potential with 3' and 5'terminal domains of 28S rRNA.

(A) A duplex structure involving both termini of mouse 5.8S rRNA and the 3'terminal domain of 28S rRNA (39) can be built.

Stabilization of 5.8S rRNA structure through interaction with :

- (B) : 3'terminal domain of 28S rRNA (as depicted in (A)).
- (C) : 5'terminal domain of 28S rRNA (as proposed in Fig. 3). In (B) and (C), differences in free energy between the structure of complexed 5.8S RNA and that of free 5.8S RNA (9) have been calculated according to Tinoco et al (40). Thick lines in 5.8S RNA correspond to regions proposed to interact with 28S rRNA.

We are currently probing the secondary structure of this 5'terminal domain of mouse 28S rRNA through S1 nuclease accessibility experiments. It must be stressed that the location of S1 nuclease accessible sites within the 75 nucleotide long 5'terminal segment of 28S RNA is in full agreement with the structure model shown in Fig. 3 (Michot and Bachellerie, unpublished results).

- Potential interaction of 5.8S rRNA with 3'terminus of 28S rRNA : In view of the proposed interaction of both 3' and 5' termini of 5.8S with 3' end of 26S rRNA in Neurospora crassa (38), we have examined if such an interaction was possible in mouse and compared its stability with that involving the 5' domain of large rRNA, as shown in Fig. 3.

It is interesting to observe (Fig. 4A) that a similar interaction can also take place in mouse involving approximately the same segment of large rRNA 3' domain (between residues 35 and 90 from 3' end). This is also the case for yeast. However the stability of this structure is largely lower than that proposed in Fig. 3, both in mouse (Fig. 4B and C) and in yeast and the positions of potential base-pairings have not been strictly maintained between the two species. Although it is difficult to rule out that this interaction has a biological role, particularly in relation with reorganizations of rRNA structure taking place during ribosome assembly and functioning, the structural model proposed in Fig. 3 appears strongly favoured.

Acknowledgements

We are grateful to Professor J.P. Zalta for support and encouragement. The technical assistance of J. Feliu was highly appreciated. This work was financially supported by a fellowship from DGRST to B.M., by ADRC contract (n° 6107) and by grant from C.N.R.S. (A.T.P. : Biologie Moléculaire du Gène).

*To whom correspondence should be addressed.

REFERENCES

1. 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
2. Georgiev, O.I., Nikolaev, N., Hadjiolov, A.A., Skryabin, K.G., Zakharyev, V.M. and Bayev, A.A. (1981) Nucleic Acids Res. 9, 6953-6958
3. Perry, R.P. (1976) Ann. Rev. Biochem. 45, 605-629
4. Pene, J.J., Knight, E., Jr. and Darnell, J.E., Jr. (1968) J. Mol. Biol. 33, 609-623
5. Pace, N.R., Walker, T. and Schroeder, E. (1977) Biochemistry 16, 5321-

5328
6. Michot, B., Bachellerie, J.P., Raynal, F. and Renalier, M.H. (1982) FEBS Lett. 140, 193-197
7. Walker, T.A., Johnson, K.D., Olsen, G.J., Peters, M.A. and Pace, N.R. (1982) Biochemistry 21, 2320-2329
8. Olsen, G.J. and Sogin, M.L. (1982) Biochemistry 21, 2335-2343
9. Nazar, R.N., Sitz, T.O. and Busch, H. (1975) J. Biol. Chem. 250, 8591-8597
10. Nazar, R.N., Sitz, T.O. and Busch, H. (1976) Biochemistry 15, 505-508
11. Khan, M.S.N. and Maden, B.E.H. (1977) Nucleic Acids Res. 4, 2495-2505
12. Kelly, J.M. and Maden, B.E.H. (1980) Nucleic Acids Res. 8, 4523-4534
13. Pavlakis, G.N., Jordan, B.R., Wurst, R.M. and Vournakis, J.N. (1979) Nucleic Acids Res. 7, 2213-2238
14. Hall, L.M.C. and Maden, B.E.H. (1980) Nucleic Acids Res. 8, 5993-6005
15. Peters, M.A., Walker, T.A. and Pace, N.R. (1982) Biochemistry 21, 2329-2335
16. Nazar, R.N. and Sitz, T.O. (1980) FEBS Lett. 115, 71-76
17. Lo, A.C. and Nazar, R.N. (1982) J. Biol. Chem. 257, 3516-3524
18. Glotz, C., Zwieb, C., Brimacombe, R., Edwards, K. and Kössel, H. (1981) Nucleic Acids Res. 9, 3287-3306
19. Branlant, C., Krol, A., Machatt, M.A., Pouyet, J., Ebel, J.P., Edwards, K. and Kössel, H. (1981) Nucleic Acids Res. 9, 4303-4324
20. Nazar, R. (1980) FEBS Lett. 119, 212-214
21. Jacq, B. (1981) Nucleic Acids Res. 9, 2913-2932
22. Maxam, A.M. and Gilbert, W. (1980) Methods in Enzymol. 65, 499-560
23. Pace, N.R. (1973) Bacteriol. Rev. 37, 562-603
24. Gerbi, S.A. (1976) J. Mol. Biol. 106, 791-816
25. Gourse, R.L. and Gerbi, S.A. (1980) J. Mol. Biol. 140, 321-339
26. 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
27. Salim, M. and Maden, B.E.H. (1981) Nature 291, 205-208
28. Samols, D.R., Hagenbüchle, O. and Gage, L.P. (1979) Nucleic Acids Res. 7, 1109-1119
29. Jordan, B.R., Latil-Damotte, M. and Jourdan, R. (1980) FEBS Lett. 117, 227-231
30. Michot, B., Bachellerie, J.P., Raynal, F. and Renalier, M.H. (1982) FEBS Lett. 142, 260-266
31. Wellauer, P.K., Dawid, I.B., Kelley, D.E. and Perry, R.P. (1974) J. Mol. Biol. 89, 397-407
32. Brosius, J., Dull, T.J. and Noller, H.F. (1980) Proc. Natl. Acad. Sci. U.S.A. 77, 201-204
33. Nazar, R.N., Sitz, T.O. and Busch, H. (1975) Biochem. Biophys. Res. Comm. 62, 736-743
34. Boseley, P.G., Tuyns, A. and Birnstiel, M.L. (1978) Nucleic Acids Res. 5, 1121-1137
35. Wildeman, A.G. and Nazar, R.N. (1981) J. Biol. Chem. 256, 5675-5682
36. Woese, C.R., Magrum, L.J., Gupta, R., Siegel, R.B., Stahl, D.A., Kop, J., Crawford, N., Brosius, J., Guttel, R., Hogan, J.J. and Noller, H.F. (1980) Nucleic Acids Res. 8, 2275-2293
37. Stiegler, P., Carbon, P., Ebel, J.P. and Ehresmann, C. (1981) Eur. J. Biochem. 120, 487-495
38. Kelly, J.M. and Cox, R.A. (1981) Nucleic Acids Res. 9, 1111-1121
39. Kominami, R., Mishima, Y., Urano, Y., Sakai, M. and Muramatsu, M. (1982) Nucleic Acids Res. 10, 1963-1979
40. 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.