SEQUENCE OF THE 3′-TERMINAL DOMAIN OF MOUSE 18 S rRNA

Conservation of structural features with other pro- and eukaryotic homologs

Bernard MICHOT, Jean-Pierre BACHELLERIE*, Françoise RAYNAL and Marie-Hélène RENALIER
Centre de Recherche de Biochimie et de Génétique Cellulaires du CNRS, 118, route de Narbonne, 31062 Toulouse Cédex, France

Received 2 April 1982

1. Introduction

Elucidation of the assembly process and function of eukaryotic ribosomes is largely dependent upon determination of the primary structure of rRNAs and their precursors. The observation, first obtained through heterologous nucleic acid hybridizations [1–3], that portions of rRNA sequence have been extensively conserved through evolution suggests that common critical functions in all organisms may be served by definite regions of these molecules.

A variety of experimental approaches have improved our knowledge of the topographical organization of rRNA in ribosomal subunits and provided strong evidence for a direct role of rRNA molecule at different stages of ribosome functioning, like reversible subunit association [4–6], tRNA [7,8] and 5 S rRNA binding [9] or mRNA selection [10,11]. Due both to its known location at subunit interface and to direct experimental evidence [6,13,14], the 3′-terminal domain of small subunit rRNA appears to be more directly involved in these functions. A more precise knowledge of its role should be gained from comparative RNA sequence analysis which has proved valuable for establishing secondary structure models for prokaryotic rRNAs [15–17]. Among eukaryotes, yeast [18] and *Xenopus* [19] are the only complete 18 S rRNA sequences published so far; their comparison has revealed extensive stretches of high homology interspersed with heterologous tracts while conserved sequences with prokaryotes are clearly restricted to the 3′-terminal region of the molecule. A remarkable conservation of secondary structure features of these two eukaryotic rRNAs was also apparent from comparison with *E. coli* 16 S rRNA [20].

Here, we have sequenced the 231 3′-terminal nucleotides of mouse 18 S rRNA. The first complete sequence of this domain of a mammalian 18 S rRNA shows a striking homology (97%) with the other vertebrate sequence available, *Xenopus laevis* [19]. The comparison of the mouse sequence with other prokaryotic [21] or eukaryotic [22,23] sequences reinforces the view of a critical functional role of this 3′-terminal domain of small subunit rRNA molecule; this is again indicated by the conservation of a common basic RNA folding pattern from prokaryotes [15,16] to mouse, as reported earlier for yeast and *Xenopus* [20].

2. Methods

The 3.7 kilobase *Eco*RI–*Bam*HI fragment of mouse ribosomal DNA containing the 3′-terminal portion of 18 S rRNA, internal-transcribed spacers and 5′-terminal region of 28 S rRNA was inserted into the (*Eco*RI + *Bam*HI)-cleaved plasmid pBR322, giving rise to a pMEB3 recombinant plasmid. Isolation of cloned DNA and detailed restriction mapping were done as in [37]. All the DNA purification, 5′-^{32}P-end labelling and sequencing procedures were performed according to [25] with additional DE52-cellulose chromatography before chemical DNA cleavages.

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

* To whom correspondence should be addressed

Published by Elsevier Biomedical Press
00145793/82/0000–0000/$02.75

3. Results and discussion

3.1. *Determination of primary structure*

Detailed restriction maps of cloned 3.7 kilobase *Eco*RI–*Bam*HI mouse ribosomal DNA fragment were constructed for the enzymes *Sma*I, *Hin*fI, *Nar*I, *Hae*II, *Taq*I, *Sau*3A and *Alu*I. The region encompassing 3′-end of 18 S rRNA coding sequences is shown in fig.1. The complete sequence of mouse internal transcribed spacers will be described elsewhere (in preparation). Due to availability of a short nucleotide sequence from the 3′-terminus of mouse 18 S rRNA [26,27], the experimental mapping of the terminus of 18 S rRNA coding sequences was not necessary, as was con-

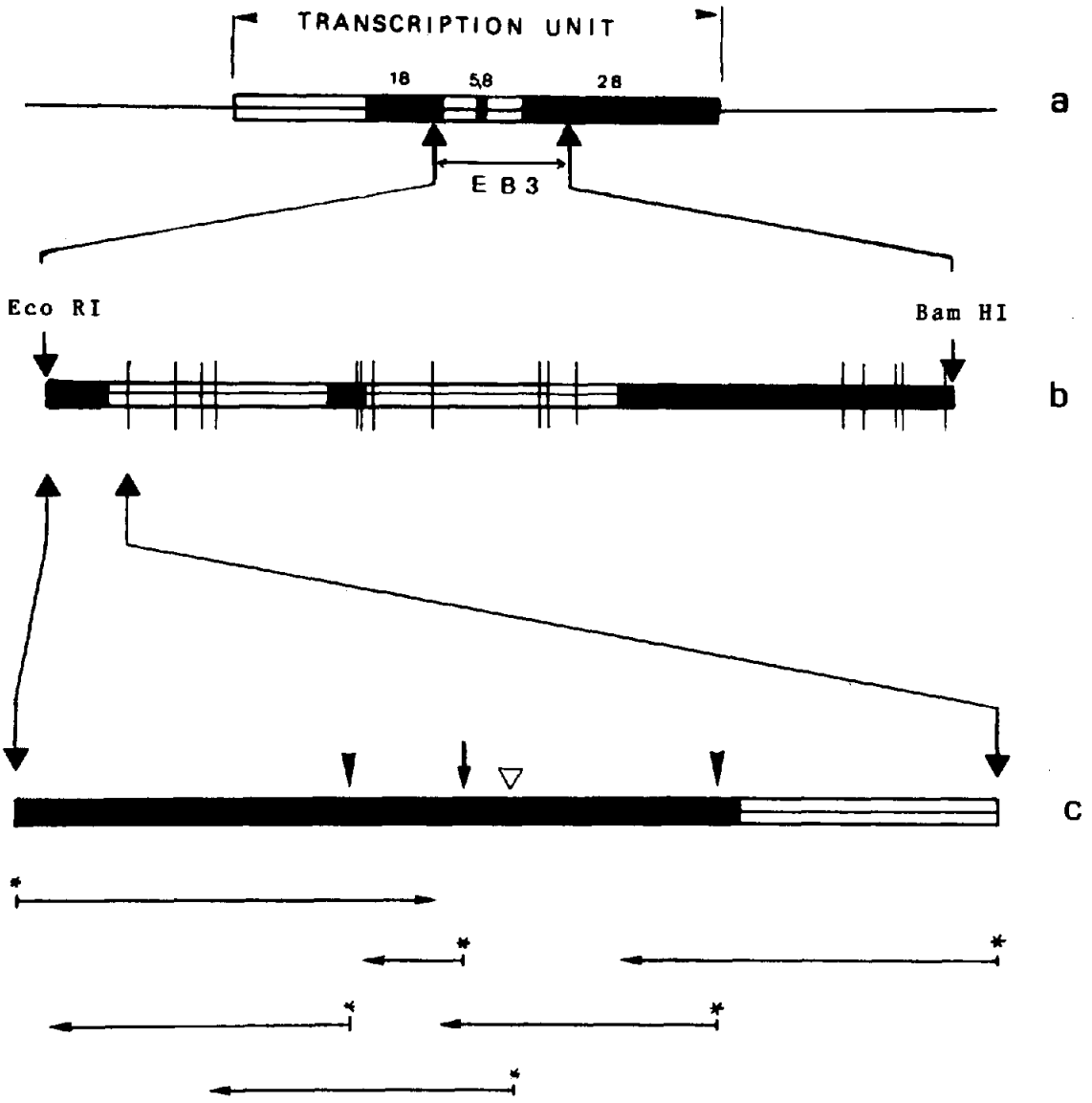

Fig.1. Restriction map of cloned mouse rDNA fragments encompassing 18 S rRNA 3′-terminal region: (a) location of the 3.7 kilobase *Eco*RI–*Bam*HI region within mouse ribosomal transcription unit; (b) *Sma*I restriction map of *Eco*RI–*Bam*HI fragment cloned into pMEB3; (c) expanded map of the 312 kilobase *Eco*RI–*Sma*I fragment containing 18 S rRNA 3′-terminus. The position of the 5′-end-labelled fragments used for sequencing are shown by horizontal arrows the lengths of which are indicative of the extent of sequence read: *Sau*3A (▾); *Hae*II (↓); *Taq*I (▿) sites.

firmed later by the strong phylogenetical conservation of this region (below).

The sequence of 231 3′-terminal nucleotides of mouse 18 S RNA gene is shown in fig.2. As indicated by the positions and readable lengths of 5′-end labelled fragments used for sequencing (fig.1c), most of the sequence was actually confirmed by independent determinations on both strands.

Our data are in full agreement with a previous partial determination involving reverse transcription of in vitro polyadenylated mouse 18 S rRNA [26]. Since reverse transcriptase cannot proceed through m_2^6A–m_2^6A residues, sequence determinations in that work was limited to the 20 3′-terminal nucleotides.

However, a significant number (9) of differences are observed with the sequence of 77 nucleotides extending from 3′-terminus of mouse sarcoma 18 S rRNA in [27]. Considering the high reliability of the DNA sequence determination (fig.3) which was carried out from more than one labelled end, these discrepancies

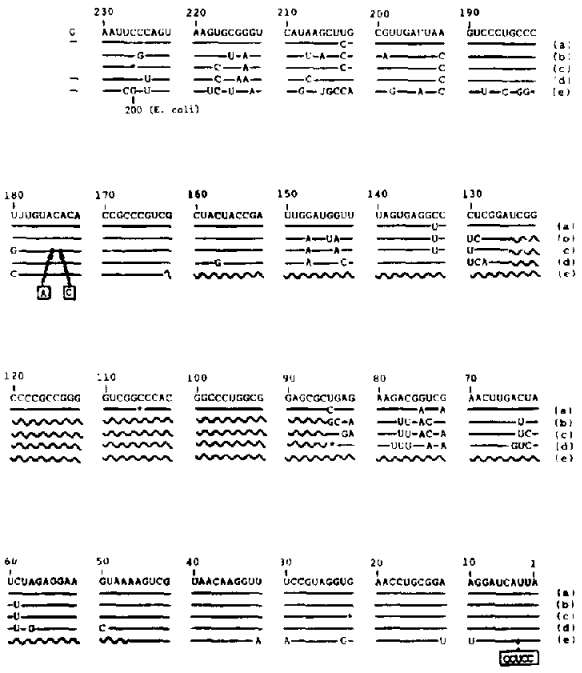

Fig.2. Portions of sequencing gels showing mouse 18 S rRNA coding sequences between positions 5–90 from 3′-terminus. This fragment (coding strand) was labelled at the *Sau*3A site proximal to 3′-terminus of 18 S rRNA. Arrows point to differences with the sequence of this segment reported in [27]. Sequence was read on a 20% acrylamide/7 M urea gel (a) and a 8% acrylamide/7 M urea gel (b).

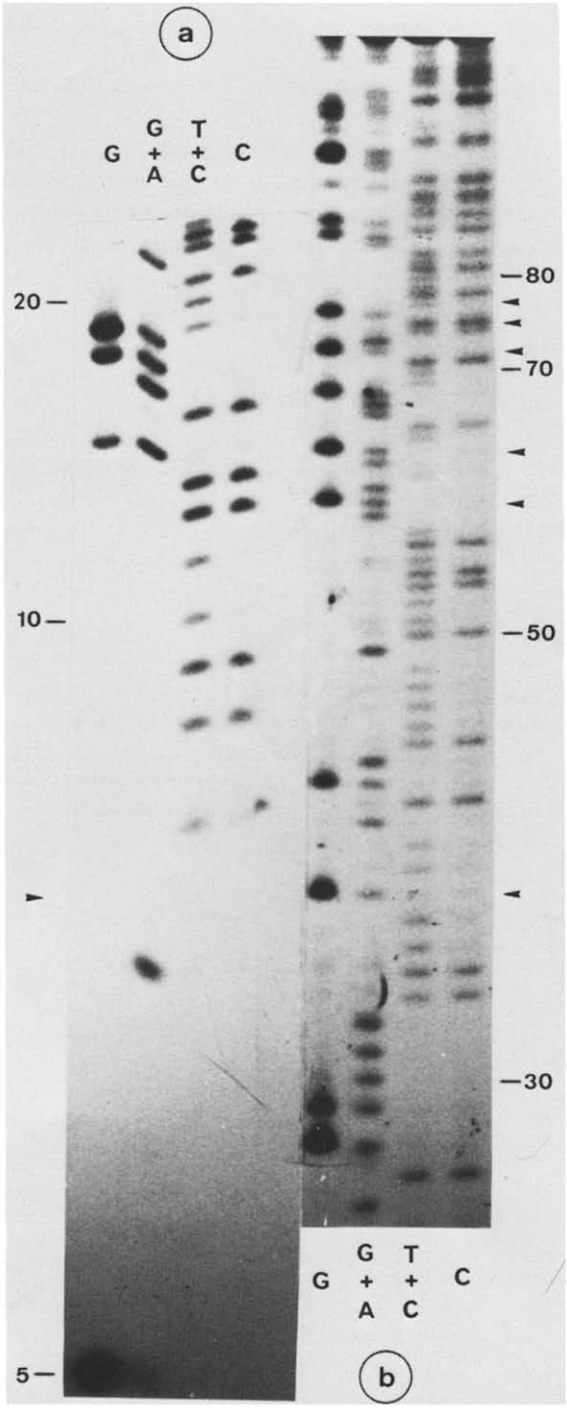

Fig.3. Sequence of the 3′-end of mouse 18 S rRNA gene. The mouse rRNA sequence (upper line) is numbered beginning at the 3′-end of the gene, and compared with that of homologous region of other small subunit rRNAs. For each sequence, positions identical with mouse are represented by a straight line, deletions by a star, and regions that have undergone extensive divergence by an undulating line: (a) *Xenopus laevis* [19]; (b) *Drosophila melanogaster* [23]; (c) *Bombyx mori* [22]; (d) *Saccharomyces cerevisiae* [18]; (e) *Escherichia coli* [21].

could correspond either to our analyzing a fragment of a non-functional ribosomal gene or to artefacts in the RNA chemical sequencing experiments [27]. This latter possibility appears more likely from a series of observations. Firstly, most of the differences are located in an area (segment 56–77, distal to 3′-end-labelled RNA-terminus) where resolution is rather low in the direct RNA sequencing approach: 5 differences correspond to nucleotides undetected in [27] while they have identified as G nucleotides 37, 59 and 63 (C, here). They also found an extra G (position 7) that we have not detected; this nucleotide was not detected any more when mouse 18 S rRNA was reverse-transcribed [26] and does not exist in another independently cloned mouse rDNA fragment [28], as indicated by the presence of the *Sau*3A restriction site (recognition sequence GATC) in identical location as compared with our pMEB3 cloned fragment. This G is not present either in the rat 18 S rRNA as demonstrated by sequencing of a 32 nucleotide-long rRNA fragment (segment 5–36) after in vitro terminal labelling [29] or by homochromatography fingerprinting of total T1 RNase oligonucleotides of 18 S rRNA [30]. More generally, all the differences introduced by the chemically derived rRNA sequence would indicate a strikingly higher number of mutations between *Xenopus* and mouse in this '1–77' region (11; cf. 2 in fig.2), with some of them being located within sequences which, like segment 32–46, have been perfectly conserved from *E. coli* to *Xenopus*, and very probably too, to rat, as suggested by the presence of T1 RNase oligonucleotides UAAAAGp and UAACAAGp in Novikoff hepatoma 18 S rRNA [30].

The phylogenetic conservation seems to rule out that these sequence discrepancies rely upon our sequencing of a non-functional ribosomal gene (such a gene would be expected to have diverged more extensively than a functional one, due to the lack of selective pressure). This conclusion has been substantiated more directly by the finding that very detailed restriction maps established for chromosomal mouse rDNA through Southern blot hybridizations (in preparation) did not reveal any difference with the present cloned rDNA fragment.

3.2. *Sequence homology*

Previous sequence determinations on the small subunit rRNA 3′-end have demonstrated the conserva-

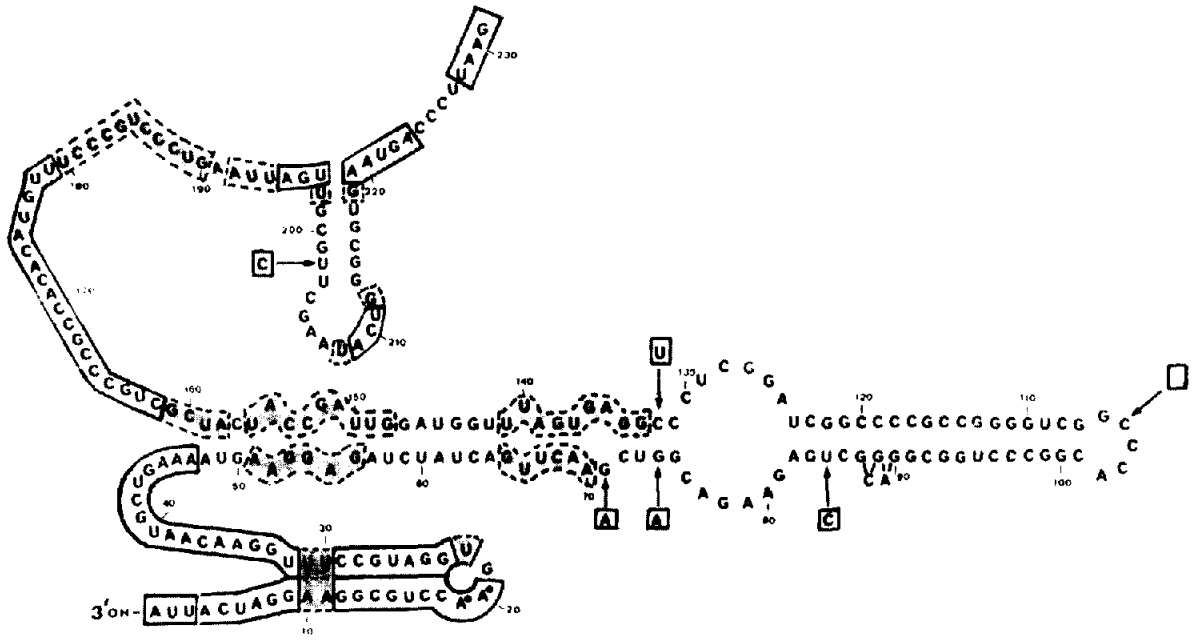

Fig.4. Secondary structure model of the mouse 18 S rRNA 3′-terminal region. Model for mouse RNA was built by reference to *E. coli* 16 S rRNA secondary structure model [15,16]. Boxes show phylogenetically conserved sequences, either in both pro- and eukaryotes (unshaded) or only in eukaryotes (shaded). Boxed single bases (with arrows) indicate point mutations as compared with another vertebrate, *Xenopus laevis* [19].

tion of large stretches within this region both in prokaryotes [15,31] and in a series of eukaryotes, ranging from yeast [18] to insects [22,23] and amphibia [19]. The mouse sequence was compared with its published counterparts in prokaryotic or eukaryotic systems, after alignment for maximum homology (fig.2). Mutations that have occurred in this region during evolution are not randomly distributed but are rather concentrated in clusters interspersed by highly conserved sequences. Regions 1–49 and 162–190 have remained perfectly identical from yeast to mouse (with the sole exception of one mutation and two additions between nucleotides 173–180 in *Bombyx mori* [22]) with very long stretches of these segments common to both prokaryotes and eukaryotes. The location of the major blocks of homology, either between all living systems or only between eukaryotes is summarized in fig.4. All the organelle small ribosomal subunit RNA sequences available so far [32–35], including mouse mitochondrial 12 S rRNA [35] show several deviations from the 'universally' conserved blocks of sequence depicted in fig.4 (unshaded boxes).

When the mouse sequence is compared to the other vertebrate sequence available (*Xenopus* [19]) a striking conservation is observed: only 5 mutations and 1 addition have occurred in the 231 nucleotide sequence (97% homology). Out of these 6 base changes, 5 are located in the region between nucleotides 71–132, an area which has undergone rather extensive divergence, even between eukaryotes as reported for two insects [22,23].

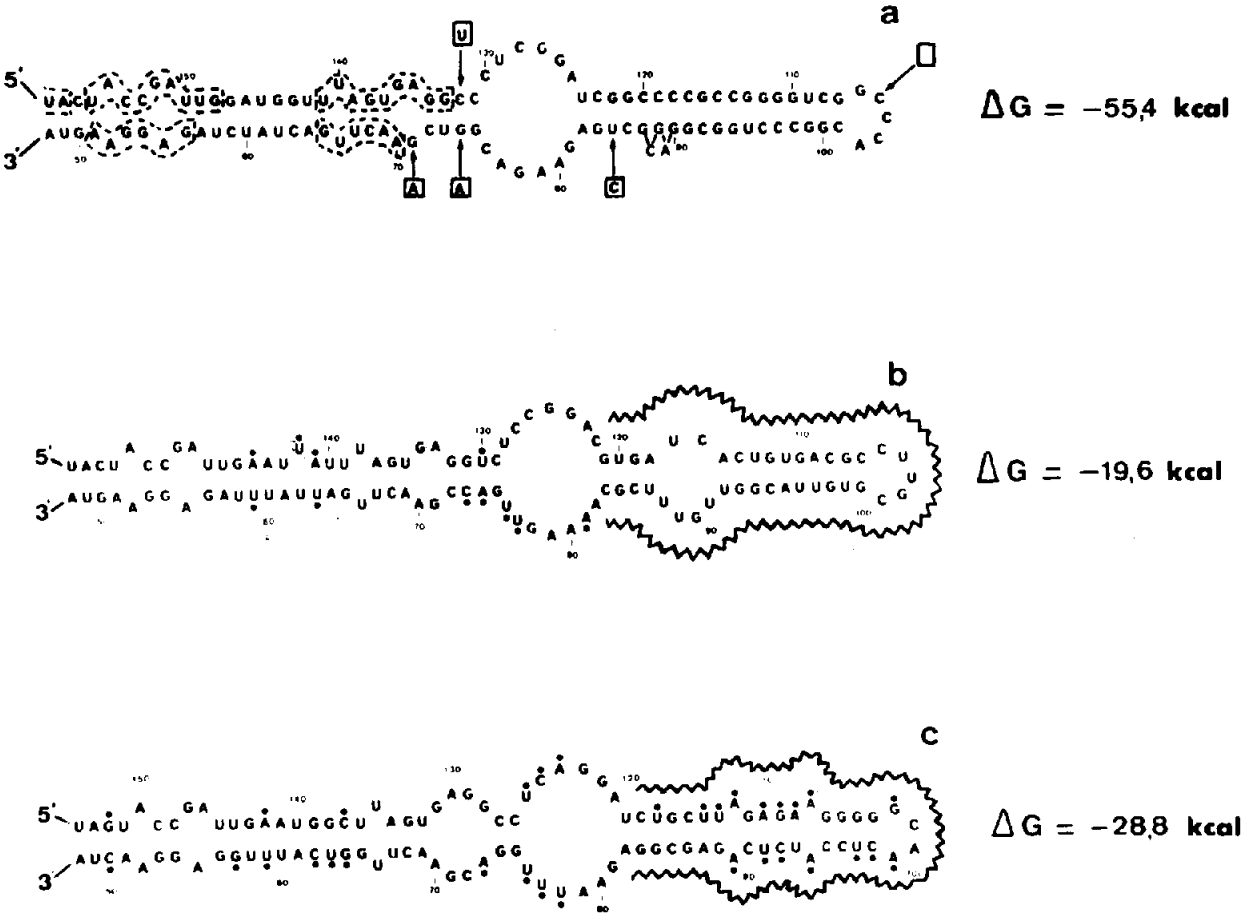

Fig.5. Comparison of mouse structural feature (II) with other eukaryotic rRNAs. A potential helix in the segment 48–159 is depicted for mouse (a), *Drosophila* (b) and *Saccharomyces cerevisiae* (c). The highly variable part of the structure is indicated by an undulating line, while, in the more evolutionary stable region, changes with mouse sequence are indicated by asterisks. Free energies were determined according to [36].

3.3. *Folding of the RNA chain*

Secondary structure models for prokaryotic 16 S rRNA have been proposed on the combined basis of comparative sequence analysis and from direct experimental evidence [15,16]. A large number of secondary structure features appear to be conserved from *E. coli* to yeast and *Xenopus* 18 S rRNA [20]. Fig.4 shows that the 3′-terminal domain of mouse 18 S rRNA can also be folded in a structure very similar to prokaryotic 16 S rRNA.

As expected from the perfect conservation of segment 1–49 in all eukaryotes examined so far, our data simply confirm that 'colicin E3' prokaryotic hairpin (I) has been preserved from pro- to eukaryotes, through two compensatory base changes (GC → AU). The other major sequence invariant (for all systems except organelle rRNAs) is the unpaired 162–179 segment which has been suggested to participate in the tRNA binding domain at the ribosomal P site [14].

The scarcity of eukaryotic rRNA sequences available so far made it difficult to propose with a high degree of confidence a definite folding pattern for the rather divergent regions of the molecule [20]. In this regard, structure (II), between nucleotides 48–159, is particularly interesting because it spans the more divergent region of this domain for the previously sequenced eukaryotic rRNAs. Despite extensive variations from prokaryotic sequence, an homologous long helical structure can be built for mouse rRNA, which is much more stable than its *E. coli* counterpart [15,16]. Other eukaryotic sequences in this domain of the molecule can also be arranged in similar structures (fig.5). The more divergent part of this variable segment, i.e., region 82–124, can be folded in a terminal GC-rich helix which is particularly stable for both vertebrates, mouse and *Xenopus*. On the contrary its insect homologs [22,23] are poorly stable. None of the few mutations that have occurred between *Xenopus* and mouse in structure (II) domain alters the base-pairing possibilities (AU → G^{74}–C^{132} and GC → G^{122}–U^{84}, respectively).

A better understanding of the role of these evolutionary changes in this 'variable' domain, in terms of rRNA folding and/or interactions with other ribosomal components will emerge from comparative analysis of additional primary sequences of a wide range of eukaryotic genera.

Acknowledgements

We thank J. Feliu for his technical assistance in the isolation of plasmid DNA. The support of Professor J. P. Zalta was highly appreciated. This work was financially supported by a fellowship from DGRST to B. M. and by ADRC contract (no. 6107).

References

[1] Pace, N. R. (1973) Bacteriol. Rev. 37, 562–603.
[2] Gerbi, S. A. (1976) J. Mol. Biol. 106, 791–816.
[3] Gourse, R. L. and Gerbi, S. A. (1980) J. Mol. Biol. 140, 321–339.
[4] Van Duin, J., Kurland, C. G., Dondon, J., Grunberg-Manago, M., Branlant, C. and Ebel, J. P. (1976) FEBS Lett. 62, 111–114.
[5] Santer, M. and Shane, S. (1977) J. Bacteriol. 130, 900–910.
[6] Chapman, N. M. and Noller, H. F. (1977) J. Mol. Biol. 109, 131–149.
[7] Noller, H. F. and Chaires, J. B. (1972) Proc. Natl. Acad. Sci. USA 69, 3115–3118.
[8] Ofengand, J. (1980) in: Ribosomes (Chambliss, G. et al. eds) pp. 497–530, University Park Press, Baltimore MD.
[9] Azad, A. A. (1979) Nucleic Acids Res. 7, 1913–1929.
[10] Shine, J. and Dalgarno, L. (1974) Proc. Natl. Acad. Sci. USA 71, 1342–1346.
[11] Steitz, J. A. and Jakes, K. (1975) Proc. Natl. Acad. Sci. USA 72, 4734–4738.
[13] Herr, W., Chapman, N. M. and Noller, H. F. (1979) J. Mol. Biol. 130, 433–449.
[14] Taylor, B. H., Prince, J. B., Ofengand, J. and Zimmermann, R. A. (1981) Biochemistry 20, 7581–7588.
[15] 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.
[16] Stiegler, P., Carbon, P., Zucker, M., Ebel, J. P. and Ehresmann, C. (1981) Nucleic Acids Res. 9, 2153–2172.
[17] Glotz, C., Zwieb, C. and Brimacombe, R. (1981) Nucleic Acids Res. 9, 3287–3306.
[18] 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.
[19] Salim, M. and Maden, B. E. H. (1981) Nature 291, 205–208.
[20] Stiegler, P., Carbon, P., Ebel, J. P. and Ehresmann, C. (1981) Eur. J. Biochem. 120, 487–495.
[21] Brosius, J., Palmer, M. L., Kennedy, P. J. and Noller, H. F. (1978) Proc. Natl. Acad. Sci. USA 75, 4801–4805.
[22] Samols, D. R., Hagenbüchle, O. and Gage, L. P. (1979) Nucleic Acids Res. 7, 1109–1119.
[23] Jordan, B. R., Latil-Damotte, M. and Jourdan, R. (1980) FEBS Lett. 117, 227–231.
[25] Maxam, A. M. and Gilbert, W. (1980) Methods Enzym. 65, 499–560.

[26] Hagenbüchle, O., Santer, M., Steitz, J. A. and Mans, R. J. (1978) Cell 13, 551–563.
[27] Azad, A. A. and Deacon, N. J. (1980) Nucleic Acids Res. 8, 4365–4375.
[28] Bowman, L. H., Rabin, B. and Schlessinger, D. (1981) Nucleic Acids Res. 9, 4951–4966.
[29] Alberty, H., Raba, M. and Gross, H. J. (1978) Nucleic Acids Res. 5, 425–434.
[30] Choi, Y. C. and Busch, H. (1978) Biochemistry 17, 2551–2560.
[31] Carbon, P., Ebel, J. P. and Ehresmann, C. (1981) Nucleic Acids Res. 9, 2325–2333.
[32] Schwartz, Z. and Kössel, H. (1980) Nature 283, 739–742.
[33] Köchel, H. G. and Küntzel, H. (1981) Nucleic Acids Res. 9, 5689–5696.
[34] Eperon, I. C., Anderson, S. and Nierlich, D. P. (1980) Nature 286, 460–467.
[35] Van Etten, R. A., Walberg, M. W. and Clayton, D. A. (1980) Cell 22, 157–170.
[36] 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.
[37] Michot, B., Bachellerie, J. P., Raynal, F. and Renalier, M. H. (1982) FEBS Lett. 140, 193–197.