Complete nucleotide sequence of mouse 18 S rRNA gene: comparison with other available homologs

Françoise Raynal, Bernard Michot and Jean-Pierre Bachellerie*

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

Received 23 December 1983

We present the complete sequence of mouse 18 S rRNA. As indicated by comparison with yeast, *Xenopus* and rat, the conservation of eukaryotic 18 S rRNA sequences is extensive. However, this conservation is far from being uniform along the molecule: most of the base changes and the size differences between species are concentrated at specific locations. Two distinct classes of divergent traces can be detected which differ markedly in their rates of nucleotide substitution during evolution, and should prove valuable in additional comparative analyses, both for eukaryotic taxonomy and for rRNA higher order organization. Mouse and rat 18 S rRNA sequences differ by only 14 point changes over the 1869 nucleotides of the molecule.

18 S rRNA *rRNA* *Mouse* *Eukaryote* *Phylogeny* *Sequence*

1. INTRODUCTION

Elucidation of the primary structure of rRNA in higher cells should provide new insights into the structural and functional organization of the eukaryotic ribosome and help to pinpoint specific roles of definite domains of rRNA molecules during ribosome assembly and function. The strong evolutionary conservation of rRNA structure, first indicated by heterologous nucleic acid hybridization experiments (review [1]) has been further substantiated by direct sequence analysis: the determination of the complete nucleotide sequences of 18 S rRNA genes of the yeast *Saccharomyces cerevisiae* [2] and amphibian *Xenopus laevis* [3] has revealed the presence of large regions of high homology interspersed with tracts which have extensively diverged between these two distant eukaryotes. When the comparison is extended to *Escherichia coli* 16 S rRNA [4–6] sequence homology is restricted to only a few polynucleotide tracts; but a closely homologous secondary structure folding scheme can nonetheless be proposed for a major part of the small subunit rRNA molecule in both prokaryotes and eukaryotes [7]. However, certain sequences of these two eukaryotic 18 S rRNAs definitely cannot be fitted to prokaryote homologous structural features: they correspond mainly to the regions where the major additions (as compared to *E. coli*) have occurred which are responsible for the marked enlargement of the molecule from pro- to eukaryotes. One may wonder if these areas of 18 S rRNA may play a functional role unique to eukaryotic ribosomes and more should be learned about their role and evolution by comparative analyses with additional eukaryotic sequences. We present here the complete sequence of mouse 18 S rRNA gene and analyze these data by reference to the 3 available eukaryotic homologs, yeast, *Xenopus* and the recently reported rat sequence [8].

2. MATERIALS AND METHODS

Mouse ribosomal DNA was prepared from

* To whom correspondence should be addressed

Published by Elsevier Science Publishers B.V.

PMSE2, a gift from I. Grummt, consisting of a 2 kb *Sal*I-*Eco*RI fragment of the mouse ribosomal transcription unit which encompasses the 5′-terminus of 18 S rRNA gene (see fig.1) and has been cloned in the vector pBR 322. It was obtained from λgt WES Mr 974 recombinant [11] which contains the large *Eco*RI-*Eco*RI fragment (13.6 kb) extending into external transcribed spacer and non-transcribed spacer. The 3′-terminal domain of mouse 18 S rRNA gene (downstream *Eco*RI site) was analyzed as in [9].

Isolation of DNA, restriction endonuclease analysis and sequence determination were carried out as in [9]. Chemical DNA sequencing was performed according to [12].

3. RESULTS AND DISCUSSION

3.1. *Primary structure*

The entire mouse 18 S rRNA gene sequence is shown in fig.2 (top line). Partial sequence data had been reported previously, for the 250 5′-terminal nucleotides [10] and the 231 3′-terminal nucleotides [9]. No ambiguity remained in the primary structure, due to the extensive overlappings of the sequenced DNA fragments along the gene and to confirmation by sequencing both DNA strands for most parts of the gene. Mouse 18 S rRNA is 1869 nucleotides long. When base content and size are examined by reference to the other eukaryotic 18 S rRNA sequences published so far, an increase in size accompanied by higher GC content is observed upon going from yeast to *Xenopus* to rat or mouse (table 1). Comparison of the mouse sequence with its yeast and *Xenopus* homologs (fig.2) reveals a high degree of overall homology with some very long tracts perfectly conserved between yeast, *Xenopus* and mouse: the 3 largest ones are 76, 62 and 49 nucleotides long (with starting positions in the mouse sequence:

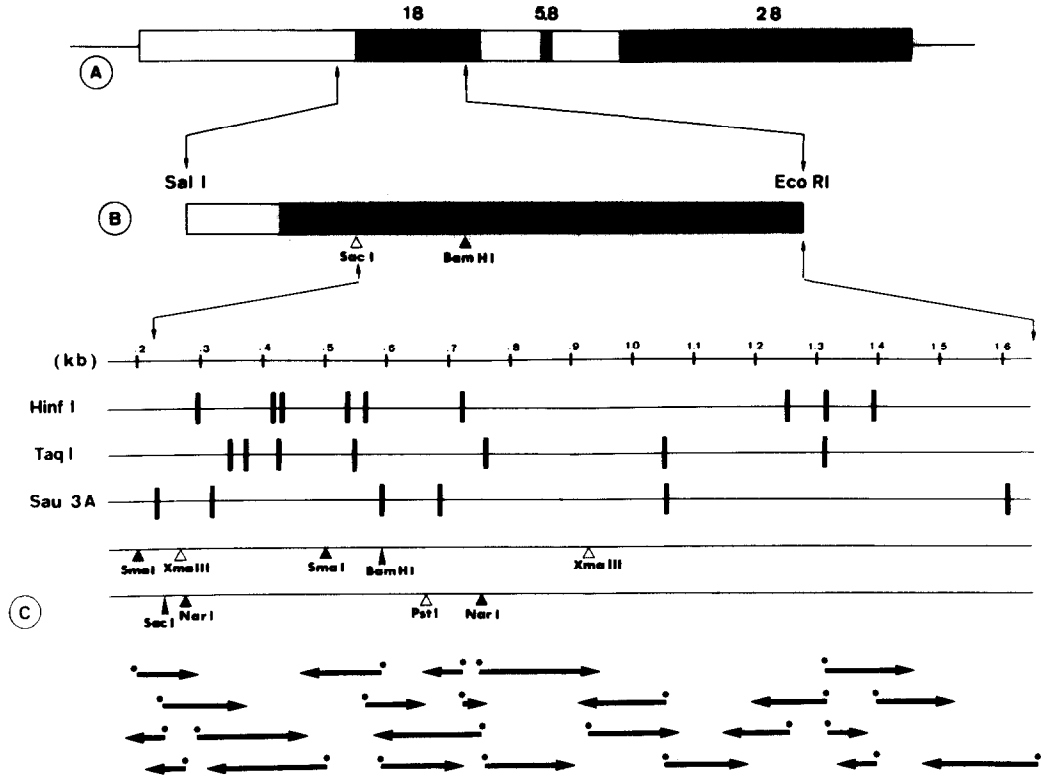

Fig.1. Restriction map of mouse 18 S rRNA gene and sequencing strategy. (A) Mouse rRNA transcription unit, (B) location of the rDNA fragment cloned in pMSE2 recombinant, (C) restriction map of the central region of 18 S rRNA gene, between *Sac*I and *Eco*RI sites (both terminal domains have been analyzed elsewhere [9,10]). Tails of horizontal arrows indicate ^{32}P-labelled 5′-ends and lengths of arrows are indicative of the extent of sequence read. Positions are numbered from the 5′-terminus of 18 S rRNA gene.

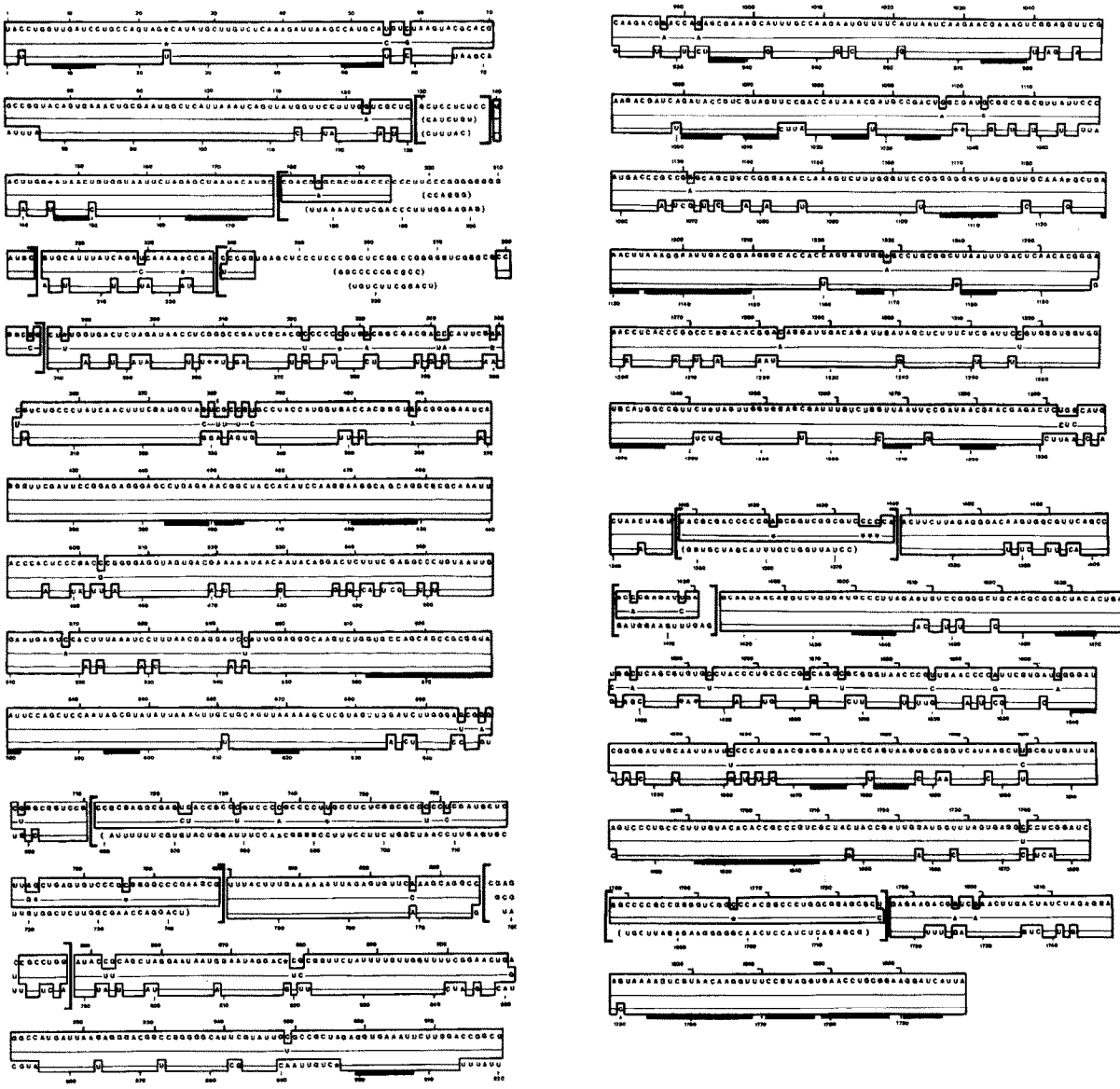

Fig.2. Mouse 18 S rRNA sequence (top line) compared with *Xenopus laevis* [3] (middle line) and the yeast *Saccharomyces cerevisiae* [2] (bottom line). Sequence tracts homologous with *Xenopus* are boxed. Nucleotides identical with the sequence immediately above are represented by a line. Deletions are denoted by a star. Segments where no unambiguous alignments could be made due to extensive divergence and size differences are shown in square brackets. Segments (equal or longer than 4 nucleotides) which are perfectly identical between eukaryotes and *E. coli* [4] are underlined by a bar. *Notes:* For yeast 18 S rRNA, the sequence of a 10 nucleotide long segment, which had been missed in the original determination [2] between position 989 and 990, has been taken from [13]. A minor rectification of the published *Xenopus* sequence, i.e., presence of an A between G^{684} and G^{685} (equivalent to mouse position 721), has also been taken into account (B.E.H. Maden, personal communication).

420, 596 and 1822, respectively). Embedded within most of these conserved tracts a large subset of oligonucleotides common to both eukaryotes and prokaryotes can be detected, which are likely to be directly involved in basic aspects of ribosomal function. They are underlined by bars in fig.3; the

Table 1

18 S rRNA base composition data

	Size (nucleotides)	G (%)	A (%)	U (%)	C (%)	G + C (%)
Mouse	1869	29.43	22.31	21.72	26.54	55.97
Rat	1869	29.05	22.52	21.83	26.6	55.65
Xenopus	1826[a]	28.20	23.71	22.51	25.57	53.77
Yeast	1799[a]	25.62	26.62	28.35	19.40	45.02

[a] See notes in fig.2 legend

largest one is 21 nucleotides long (starting position 612 in mouse). While the 3′-terminal region is particularly rich in tracts common with prokaryotes, as reported in [2,3,9], it is noteworthy that other significant blocks of homology are scattered in other domains of the molecule.

In line with [3,8], it is remarkable that base changes, insertions or deletions are far from being uniformly distributed but are mostly clustered in definite areas of the molecule.

3.2. *Divergent areas of eukaryotic 18 S rRNAs*

3.2.1. Comparison: mouse–*Xenopus*–yeast

When mouse and yeast sequences are compared, almost the entire size difference between the two species is accounted for within 8 fully divergent tracts which represent only 14.8% of the total length of the mouse molecule. The GC content of these divergent tracts is dramatically increased in mouse as compared to yeast, but a marked preference for changes from A,U to G,C is also observed within the strongly conserved regions, as summarized in table 2.

When considering mouse and *Xenopus* sequences, only 4 fully divergent tracts are observed which amount to a much smaller fraction of the molecule than in the comparison mouse *vs* yeast, i.e., 3.5% instead of 14.8%. It is noteworthy that these segments represent a subset of the divergent regions in the mouse and yeast sequences. For both conserved and divergent regions, a slight increase in GC content is observed in mouse as compared to *Xenopus* (table 2). It may be noteworthy that, in

Table 2

Comparison of mouse 18 S rRNA with yeast (A), and with *Xenopus* (B)

		Total size (nucleotides)	Relative size (% of 18 S rRNA)	G + C content (%)	Homology (%)
(A) Mouse/yeast	Divergent tracts (130–179–239–711–836–1408–1469–1748)[a]	276/207	14.8/11.5	79/48.3	no
	Conserved regions	1593/1592	–	52.0/44.6	81.5
(B) Mouse/*Xenopus*	Rapid divergent tracts (130–195–244–836)[a]	66/29	3.5/ 1.6	81.8/79.3	no
	Conserved regions	1803/1797	–	55.0/53.4	95.7

[a] In each comparison, the highly divergent tracts, which correspond to the bracketed areas in fig.2, are identified by their starting positions in mouse sequence

For each pair of values, the first one corresponds to mouse, the second one to yeast (A) or *Xenopus* (B)

terms of base composition, the divergent regions in mature mouse 18 S rRNA bear a close analogy to the most rapidly divergent domains of the ribosomal transcription unit, i.e., the RNA spacer segments which are removed from mouse rRNA precursors, during rRNA processing: these precursor-specific regions are also very rich in G + C and particularly poor in A [14].

Outside the hypervariable regions, there is 81.4% homology between mouse and yeast sequences, and 95.7% homology between mouse and *Xenopus*. From this comparison between the 3 species, it appears that 3 major kinds of sequence tracts can be distinguished in 18 S rRNAs, on the basis of marked differences in their rates of nucleotide substitution during evolution:

(i) Strongly conserved regions, which are extensively homologous between the 3 species;
(ii) Slowly divergent regions which are extensively divergent between yeast and *Xenopus* (or mouse) but are largely homologous between *Xenopus* and mouse;
(iii) Rapidly divergent regions, which are extensively divergent between *Xenopus* and mouse.

The validity of this operational distinction is further reinforced by comparing mouse sequence with its recently published rat homolog [8] and with partial sequence data on rabbit [15].

3.2.2. Comparison: mouse – rat – rabbit

Sequences of the two rodent RNAs are extremely homologous (99.2%) without any divergent tract: only 14 point changes are detectable (table 3). Such a small number of changes makes it difficult to carry out a precise and reliable analysis in terms of distribution pattern along the molecule. However, it is remarkable that even in these conditions, the 3 previously defined kinds of sequence tracts are again clearly distinguished on the basis of frequency of base changes between both rodents, as shown in table 3: in tracts which have extensively diverged between mouse and *Xenopus*, this frequency is 20-times higher than in the regions which have been conserved between mouse, *Xenopus* and yeast. In tracts which have extensively diverged between mouse and yeast but which have been mostly conserved between mouse and *Xenopus*, this frequency is about 7-times

Table 3

Changes between mouse and rat 18 S rRNA sequences

Location[a]			Mutations mouse → rat	Additions[c]	Deletions[c]	Frequency of changes in these regions (in changes per 100 nucleotides)	
In regions strongly conserved between yeast, *Xenopus* and mouse (85.2%)[b]			$^{325}C \rightarrow U$ $^{408}G \rightarrow A$ $^{1561}G \rightarrow A$	$^{287/288}U$	^{123}G	0.31	
In tracts highly divergent between:	mouse and yeast (14.8%)[b]	mouse and *Xenopus* (3.5%)[b]	$^{245}G \rightarrow C$ $^{249}U \rightarrow C$	$^{203/204}U$	^{258}G	6.06	3.26
			$^{212}U \rightarrow C$ $^{846}G \rightarrow A$ $^{1774}C \rightarrow U$	$^{282/283}A$	^{1777}G	2.38	

[a] See fig.2 and table 2 for precise positions of conserved and highly divergent regions
[b] Relative size of these regions, expressed in percent of total mouse 18 S rRNA
[c] In rat, as compared to mouse

higher than in the regions conserved between mouse, *Xenopus* and yeast (table 3).

A similar comparative analysis has been carried out between mouse 18 S rRNA and its rabbit homolog, which has been partially determined by sequencing end-labelled rRNA, i.e., the 400 and 300 nucleotides from 5′- and 3′-ends, respectively [15]. The significance of this comparison is somewhat hampered by the presence of a few sequence heterogeneities or uncertainties at some positions of the rabbit sequence (which could result, at least partly, from technical limitations inherent to direct RNA sequencing). However, the presence of a much larger number of changes between these two more distant mammals than between the two rodents makes it the more valid, if not strictly accurate. In this case again, the frequencies of change determined for the 3 kinds of areas are in agreement with our previous operational distinction: for each class, we have observed 3.7, 10 and 33 changes per 100 nucleotides, respectively. When these values are compared to those obtained in the mouse–rat comparison (table 3), it appears that the frequency of base changes for each class of segments does increase in roughly similar proportions (i.e., a factor 5–12) when switching from the mouse–rat comparison to the mouse–rabbit comparison. Assuming that the rate of nucleotide substitutions in each class of tracts has remained constant since mouse, rat and rabbit have evolved from their common ancestor, this could be taken to indicate that rabbit and rodents have diverged from their common ancestor 5–12-times earlier than the divergence of mouse and rat. This result is in line with independent determinations largely based upon tandem alignments of sequence data from several different polypeptide chains [16] suggesting a factor of about 7 between both clock dates. Since sequence determinations focussed on the slowly and rapidly evolving areas of rRNAs can be rapidly carried out in a systematic way [17], these regions now provide new sensitive indicators for eukaryote taxonomy.

ACKNOWLEDGEMENTS

We are grateful to I. Grummt for providing us with pMSE 2 recombinant. Support and encouragement of Professor Zalta were highly appreciated. We thank J. Feliu for excellent technical assistance and M. Parès for typing the manuscript. 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] Gerbi, S.A. (1982) in: The Cell Nucleus, vol.X (Busch, H. and Rothblum, L. eds) pp.351–386, Academic Press, New York.
[2] 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.
[3] Salim, M. and Maden, B.E.H. (1981) Nature 291, 205–208.
[4] Brosius, J., Palmer, M.L., Kennedy, P.S. and Noller, H.F. (1978) Proc. Natl. Acad. Sci. USA 75, 4801–4805.
[5] Carbon, P., Ehresmann, C., Ehresmann, B. and Ebel, J.P. (1978) FEBS Lett. 94, 152–156.
[6] Carbon, P., Ebel, J.P. and Ehresmann, C. (1981) Nucleic Acids Res. 9, 2325–2333.
[7] Stiegler, P., Carbon, P., Ebel, J.P. and Ehresmann, C. (1981) Eur. J. Biochem. 120, 487–495.
[8] Torczynski, R., Bollon, A.P. and Fuke, M. (1983) Nucleic Acids Res. 11, 4879–4890.
[9] Michot, B., Bachellerie, J.P., Raynal, F. and Renalier, M.H. (1982) FEBS Lett. 142, 260–266.
[10] Raynal, F., Michot, B. and Bachellerie, J.P. (1983) FEBS Lett. 161, 135–139.
[11] Grummt, I., Soellner, C. and Scholz, I. (1979) Nucleic Acids Res. 6, 1351–1369.
[12] Maxam, A.M. and Gilbert, W. (1980) Methods Enzymol. 65, 499–560.
[13] Mankin, A.S., Kopylov, A.M. and Bogdanov, A.A. (1981) FEBS Lett. 134, 11–14.
[14] Michot, B., Bachellerie, J.P. and Raynal, F. (1983) Nucleic Acids Res. 11, 3375–3391.
[15] Lockard, R.E., Connaughton, J.F. and Kumar, A. (1982) Nucleic Acids Res. 10, 3445–3457.
[16] Goodman, M., Weiss, M.L. and Czelusniak, J. (1982) Syst. Zool. 31, 376–399.
[17] Qu, L.H., Michot, B. and Bachellerie, J.P. (1983) Nucleic Acids Res. 11, 5903–5920.