Volume 12 Number 10 1984 Nucleic Acids Research

Secondary structure of mouse 28S rRNA and general model for the folding of the large rRNA in eukaryotes

B.Michot, N.Hassouna and Jean-Pierre Bachellerie*

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

Received 30 March 1984; Accepted 30 April 1984

ABSTRACT

We present a secondary structure model for the entire sequence of mouse 28S rRNA (1) which is based on an extensive comparative analysis of the available eukaryotic sequences, i.e. yeast (2, 3), Physarum polycephalum (4), Xenopus laevis (5) and rat (6). It has been derived with close reference to the models previously proposed for yeast 26S rRNA (2) and for prokaryotic 23S rRNA (7-9). Examination of the recently published eukaryotic sequences confirms that all pro- and eukaryotic large rRNAs share a largely conserved secondary structure core, as already apparent from the previous analysis of yeast 26S rRNA (2). These new comparative data confirm most features of the yeast model (2). They also provide the basis for a few modifications and for new proposals which extend the boundaries of the common structural core (now representing about 85 % of E. coli 23S rRNA length) and bring new insights for tracing the structural evolution, in higher eukaryotes, of the domains which have no prokaryotic equivalent and are inserted at specific locations within the common structural core of the large subunit rRNA.

INTRODUCTION

Elucidation of the primary and secondary structure of rRNAs is an obvious prerequisite for identifying its tertiary structure folding and topographical organization within the ribosome and consequently for unraveling the precise functional roles of definite regions of the rRNA molecule, though involvements in RNA-protein or RNA-RNA interactions during the ribosome cycle. Considerable progress has been achieved in this area for the E. coli ribosome.

Through a combination of direct experimental data and of extensive comparative analyses, detailed models have been derived for the secondary structure of 23S rRNA (7-9) which is strongly conserved during the evolution of prokaryotes, with most base-pairings maintained despite mutations, through the presence of compensatory base-changes in opposite positions.

Our structural understanding of the eukaryotic ribosome is much less advanced and until recently yeast 26S rRNA was the only complete sequence

available for a large subunit rRNA in eukaryotes (2, 3). Its comparison with prokaryotic models revealed an extensive conservation of secondary structure while the conserved domains contain a number of inserted regions which account for the increased length of this molecule in yeast as compared to E. coli (2). Whereas conserved features are likely to be involved in a number of basic functions common to pro- and eukaryotic ribosomes, the existence of additional eukaryote-specific domains - which have dramatically increased in size in higher eukaryotes - together with the presence of a larger number of ribosomal proteins poses an intriguing problem as to their potential roles, structural organization within the ribosome and mode of variation during the evolution of eukaryotes.

Comparative analyses can now be carried out among several eukaryotic 28S rRNAs, due to the recent publication of four other complete sequences, i.e. slime mold Physarum polycephalum (4), amphibian Xenopus laevis (5), and two rodents, rat (6) and mouse (1). As a consequence, some insight can be gained into the process of size increase of the large subunit rRNA during the evolution of eukaryotes, as shown elsewhere (1) and more generally into the secondary structure folding potential of the entire molecule, without being restricted to the domains common to pro- and eukaryotes. In this paper, we present a model for the secondary structure of mouse 28S rRNA, based on comparative evidence, and discussed by reference to the models previously proposed for yeast 26S rRNA (2) and for E. coli 23S rRNA (7-9). The folding potentials of the few eukaryotic-specific domains of the molecule are also examined for the other eukaryotes.

METHODS

The mouse 28S rRNA sequence (1) has been compared with the four other complete eukaryotic sequences available so far, i.e. yeast S. carlsbergensis (2), slime mold Physarum polycephalum (4), amphibian Xenopus laevis (5) and rat (6), and with E. coli 23S rRNA (10, 8). Outside the few areas where major size differences have taken place during the evolution (1), eu- and pro-karyotic sequences can be unambiguously aligned along a large fraction of the molecule, due to the presence of a number of conserved tracts : their folding potential in eukaryotic 28S rRNA can therefore be examined by close reference to the models previously derived for E. coli 23S rRNA (7-9).

In a preliminary stage, the prokaryotic models were tested by using the more recently prokaryotic sequences (11, 12). This additional compara-

tive sequence analysis provided a strong support for most of the stems of a "consensus" model, particularly for some base-pairings for which previous comparative proofs were weak or inexistent (see Results).

In a second stage, the secondary structure model proposed for most of yeast 26S rRNA (2) was critically reexamined in view of both the "consensus" prokaryotic structure and of the potential for homologous base-pairing of the recently available eukaryotic sequences. For the regions of the molecule for which no folding pattern was proposed on the basis of the yeast sequence and more generally for all the domains which have extensively varied in size during the evolution of eukaryotes, a systematical examination was carried out on all eukaryotic sequences after alignment for maximal homology (1). Catalogs of potentially base-paired regions for the remaining unstructured domains were established using the HELCAT computer program (13) and selection of potentially homologous helical features was carried out according to the strategy described by Noller et al. (9). The resulting secondary structures were in turn compared to the prokaryotic models (7-9). In some cases, the presence of common structural features at equivalent positions in all species provided a means to extend the boundaries of the common structural core, and consequently to better delimitate the domains of variable size. Partial sequence data available for other eukaryotes (14-18) were also considered for testing several secondary structure features, as indicated in Figures and Table 1.

RESULTS AND DISCUSSION

As described elsewhere (1), the comparison of the prokaryotic and eukaryotic sequences of the large subunit rRNA and the examination of their potential for secondary structure folding reveal that all these molecules share a largely conserved structural core which is interrupted, at specific locations, by a few divergent domains which have undergone large variations in size during the evolution of eukaryotes. Both types of domains are clearly delineated in the entire secondary structure model displayed in Fig. 1.

1. Common structural core :

The comparative evidences for the proposed base-pairings are listed in Table 1. All these stems have their clear counterpart in prokaryotic 23S rRNA (7-9), except for helices 9, 11, 56b and 83b (moreover the correspondence is far from being close for helices 84-86). Almost all the helices of interkingdom occurrence are supported by comparative criteria among proka-

ryotes only : in some cases, particularly for stems 18, 22, 23, 65, 73, 76, 80-82 and 84-86, the previous comparative evidences (7-9), which were not clear-cut, are strongly reinforced when considering the recently published Anacystis nidulans (11) and tobacco chloroplast (12) 23S rRNA sequences. Moreover, among the few universal stems which are not proven by the sole examination of the available prokaryotic sequences, i.e.# 10, 25, 30, 39, 52, 55 and 58, it is noteworthy that helices 30 and 39 do receive a comparative support from the recent eukaryotic sequences.

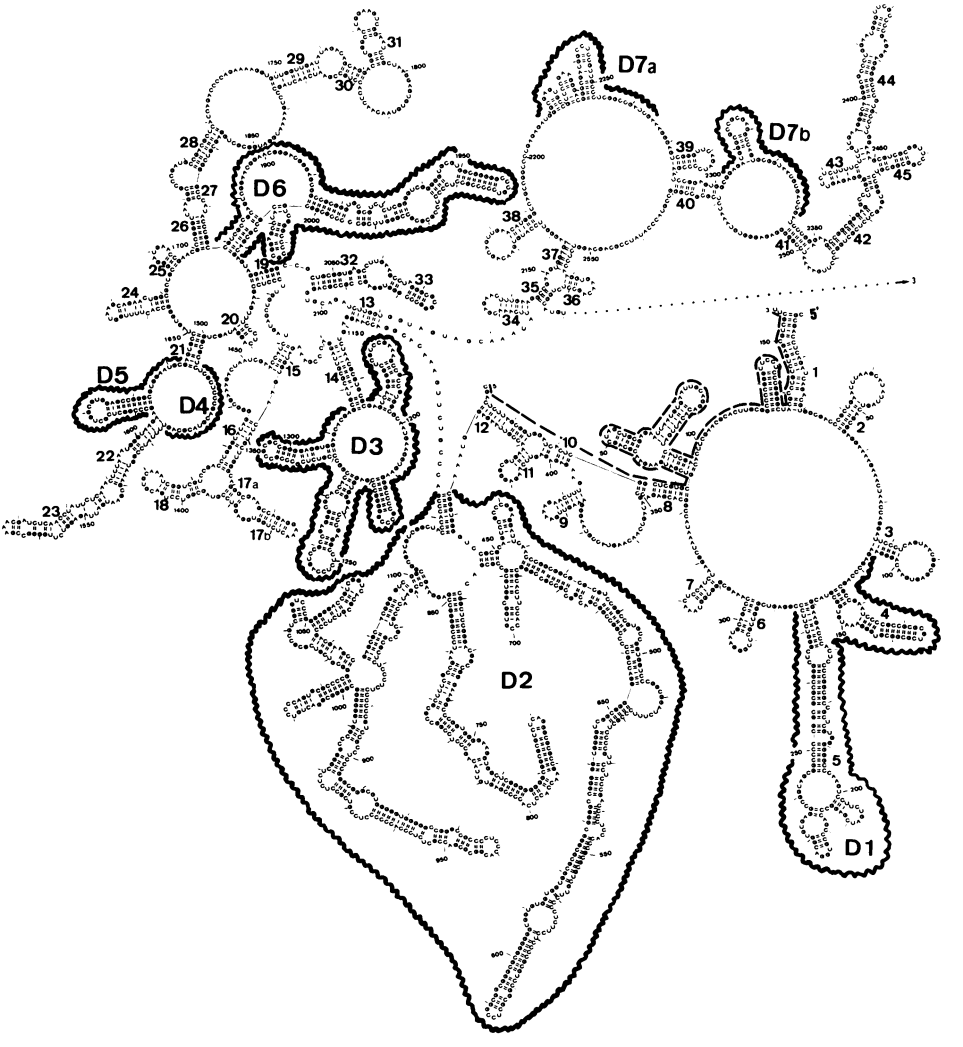


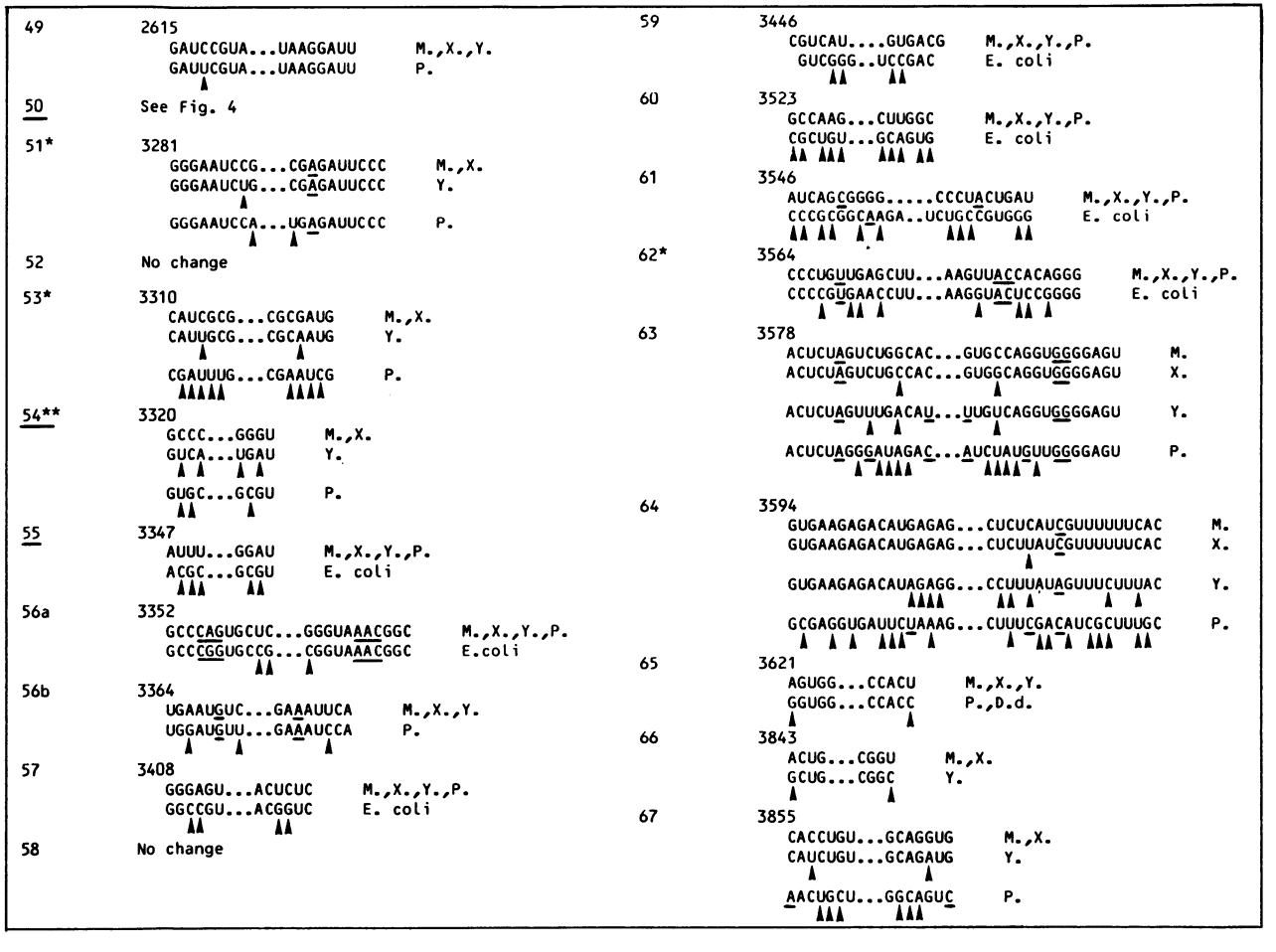
Fig. 1 : Secondary structure model of mouse 28S rRNA.
The 5' and 3' halves of the molecules are represented on the left and the right pages respectively. The base-paired association with 5.8S rRNA, which involves the 5'terminal domain of 28S rRNA, is also shown (19). (The 5.8S sequence is denoted by a thick broken overline). The thick wavy overlines delineate the divergent domains where major size differences have been restricted during the evolution. Helical stems of the conserved structural core are numbered from the 5' end (see Table 1 for available evidences from comparative sequence analysis).

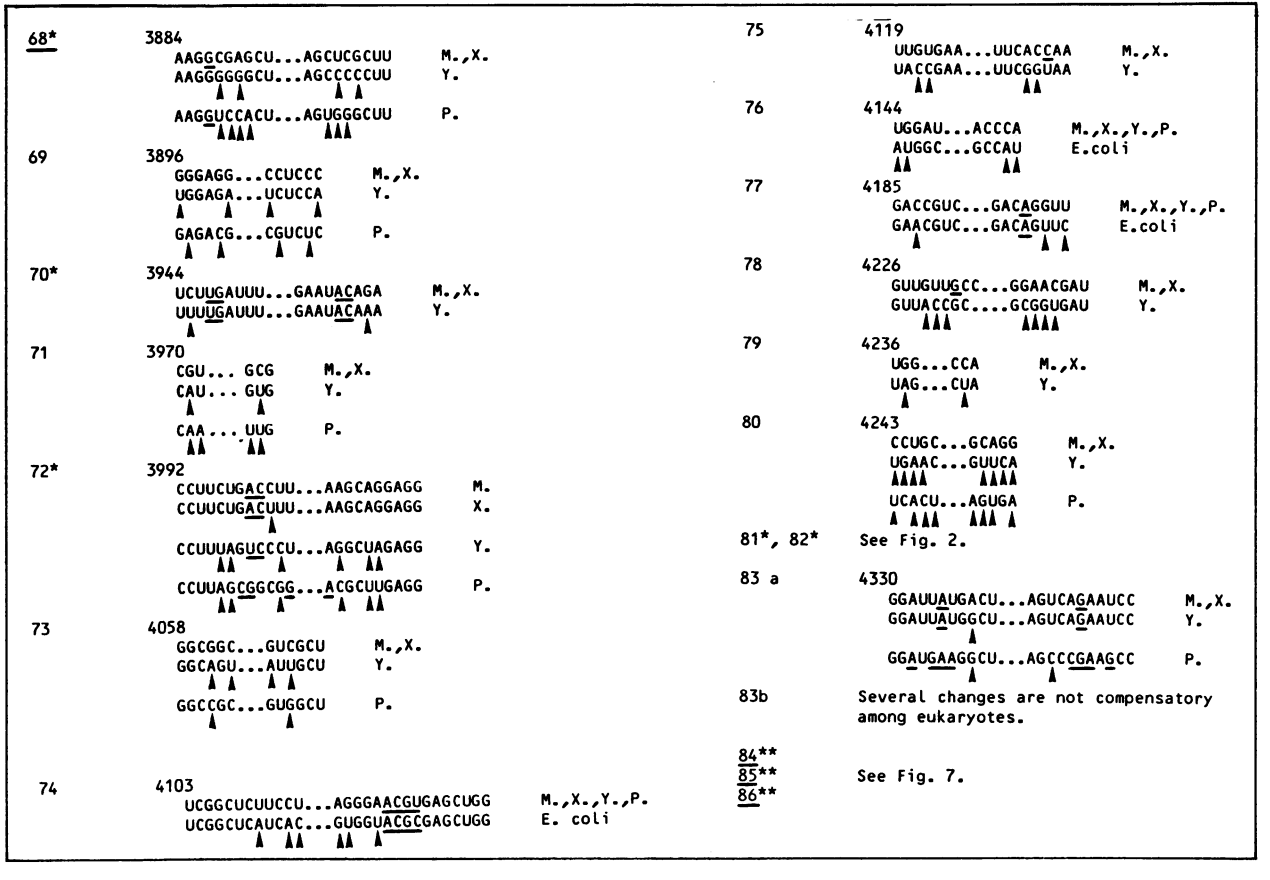
Table 1 : Comparative evidence for base-pairings within the conserved core of 28S rRNA.

Helix numberings refer to Fig. 1. New helices as compared to the yeast 26S rRNA model (2) are denoted by underlined numbers (the double asterisk indicates that a markedly different base-pairing was proposed for yeast; the absence of asterisk means that no folding model had been proposed for this area of yeast rRNA). A single asterisk indicates the presence of only minor modifications of the base-pairings proposed for yeast. A bracketted + denotes the presence of compensatory changes between pro-and eukaryotes. Unpaired or bulged nucleotides are underlined and compensatory changes are indicated by arrow-heads. Sequences are designated M. : mouse, Y. : yeast, Saccharomyces carlsbergensis (2), P. : Physarum polycephalum (4), X. : Xenopus laevis (5), L.v. : Lytechinus variegatus (18), Dd.: Dictyostelium discoideum (14).

Helix #	
1	See (19-21) for details.
2	See (20) for details
3	77 UUCCCU...AGGGAA M.,X.,P.,L.v. UUGCCU...AGGCAA Y.
6	283 GGGUUGC...GCAGCCC M.,X. GAGUUGU...GCAGCUC Y. GGGUAGU...GCUGCUC L.v.
7	310 GGGUGG...CCAUCU M.,X.,Y. AGGUGG...CCAUCU P.
8	Several changes between yeast, mouse and physarum, See (19, 21)
9, 10, 11	No change.
12	Several changes between yeast, mouse and physarum, See (19, 21).
13**	1143 ACCAAGG...UCUUGGU M.,X.,Y.,P. GUCAGCG...UGCUGAC E. coli
14	Several changes involving 6 bp, See Fig. 6 for details.
15	1338 GAUGGU...ACCAUC M.,X.,Y. AACGAC...GUCGUU P.
16	1347 CUAUGCUUGGG...CCUGGGUAUAG M.,X. CUAUGCCUGAA...UUUGGGUAUAG Y. CUAUGCUUGAG...CUUGAGUAUAG P.
17a	1360 GGGC...GUCC M.,X. GGGU...GCUC Y. GGCC...GGUC P.
17b	1367 GCCAGAG...CUCUGGU M.,X.,Y. GCCAGGA...UCUUGGU P.
18	1396 GCGGUCCUGA...UCGGUCGU M.,X. GCGGUUCUGA...UCGAUCGU Y. GCAGUACUGA...UCGUUUGU P.

19	1473 CCCUCC...GGAGGG M.,X. CCUGCC...GGCAGG Y.
20	1485 UCC...GGA M.,X.,Y.,P. UAU...GUA E. coli
21, 22**	See Fig. 8 for details.
23**	See Fig. 8 for details
24, 25	No change
26	1700 GCCGGGU...ACCCGGC M.,X. GUAGAGU...ACUCUAC Y. GCCCGGU...ACCGGCC P.
27	1709 AGGC...GCCC M.,X. AGGU...ACCU Y. GGUU...AACC P.
28	1721 CCGACGCUC...GAGCGUCGG M.,X. UACACGCUC...AAGCGUGUU Y. ACAUUGGCA...AACCAGUGC P.
29	1749 GUUGGUUGAU...AUCAACUAGC M.,X. GUUAGUUCAU...AUGAACUAGC Y. AUUGCCUCAU...AUGGGGUAGU P.
30 (+)	1766 GCAGG...CCUGC M.,X. GCCGG...CCGGC Y. GCUGG...CCAGC P.
31	1772 CGGUGGCC...GGAAUCCG M.,X.,Y. CAGUGGCC...GGAAUCUG P.
32*, 33*	See Fig. 2.
34	2125 UUCAAA...UUUGAA M.,X.,Y.,P. AGCGGG...CCCGCU E. coli
35	2145 GCCG...UGGC M.,X. ACUG...UAGU Y.
36	2559 AGGU...GCCU M.,X.,Y. AGAU...GUCU P.
37**	2152 UGGAGA...UCUCCA M. UGGAGG...UCUCCA X. UGGGGA...UCUCCA Y.
38*	2163 UCCAUGU...ACAUGGG M.,X. UCCACGU...ACGUGGG Y. UCCUCAA...UUGGGGG P.
39 (+)	2273 UCGGG...CCCGA M.,X. UCCGG...CCGGA Y. GCGGG...CCUGC P.
40	See Fig. 3 for details
41-45	See text and Fig. 3 for details
46	2581 UGGA...UCCA M.,X. UAGA...UCUA Y.
47	2590 GUAGGUA...UACCUAC M.,X. GUAGAUA...UAUCUAC Y.
48	No change


49
2615
GAUCCGUA...UAAGGAUU M.,X.,Y.
GAUUCGUA...UAAGGAUU P.
50
See Fig. 4
51*
3281
GGGAAUCCG...CGAGAUUCCC M.,X.
GGGAAUCUG...CGAGAUUCCC Y.
GGGAAUCCA...UGAGAUUCCC P.
52
No change
53*
3310
CAUCGCG...CGCGAUG M.,X.
CAUUGCG...CGCAAUG Y.
CGAUUUG...CGAAUCG P.
54**
3320
GCCC...GGGU M.,X.
GUCA...UGAU Y.
GUGC...GCGU P.
55
3347
AUUU...GGAU M.,X.,Y.,P.
ACGC...GCGU E. coli
56a
3352
GCCCAGUGCUC...GGGUAAACGGC M.,X.,Y.,P.
GCCCGGUGCCG...CGGUAAACGGC E.coli
56b
3364
UGAAUGUC...GAAAUUCA M.,X.,Y.
UGGAUGUU...GAAAUCCA P.
57
3408
GGGAGU...ACUCUC M.,X.,Y.,P.
GGCCGU...ACGGUC E. coli
58
No change
59
3446
CGUCAU....GUGACG M.,X.,Y.,P.
GUCGGG..UCCGAC E. coli
60
3523
GCCAAG...CUUGGC M.,X.,Y.,P.
CGCUGU...GCAGUG E. coli
61
3546
AUCAGCGGGG.....CCCUACUGAU M.,X.,Y.,P.
CCCGCGGCAAGA..UCUGCCGUGGG E. coli
62*
3564
CCCUGUUGAGCUU...AAGUUACCACAGGG M.,X.,Y.,P.
CCCCGUGAACCUU...AAGGUACUCCGGGG E. coli
63
3578
ACUCUAGUCUGGCAC...GUGCCAGGUGGGGAGU M.
ACUCUAGUCUGCCAC...GUGGCAGGUGGGGAGU X.
ACUCUAGUUUGACAU...UUGUCAGGUGGGGAGU Y.
ACUCUAGGGAUAGAC...AUCUAUGUUGGGGAGU P.
64
3594
GUGAAGAGACAUGAGAG...CUCUCAUCGUUUUUUCAC M.
GUGAAGAGACAUGAGAG...CUCUUAUCGUUUUUUCAC X.
GUGAAGAGACAUAGAGG...CCUUUAUAGUUUCUUUAC Y.
GCGAGGUGAUUCUAAAG...CUUUCGACAUCGCUUUGC P.
65
3621
AGUGG...CCACU M.,X.,Y.
GGUGG...CCACC P.,D.d.
66
3843
ACUG...CGGU M.,X.
GCUG...CGGC Y.
67
3855
CACCUGU...GCAGGUG M.,X.
CAUCUGU...GCAGAUG Y.
AACUGCU...GGCAGUC P.


68*
3884
AAGGCGAGCU...AGCUCGCUU M.,X.
AAGGGGGGCU...AGCCCCCUU Y.
AAGGUCCACU...AGUGGGCUU P.
69
3896
GGGAGG...CCUCCC M.,X.
UGGAGA...UCUCCA Y.
GAGACG...CGUCUC P.
70*
3944
UCUUGAUUU...GAAUACAGA M.,X.
UUUUGAUUU...GAAUACAAA Y.
71
3970
CGU... GCG M.,X.
CAU... GUG Y.
CAA... UUG P.
72*
3992
CCUUCUGACCUU...AAGCAGGAGG M.
CCUUCUGACUUU...AAGCAGGAGG X.
CCUUUAGUCCCU...AGGCUAGAGG Y.
CCUUAGCGGCGG...ACGCUUGAGG P.
73
4058
GGCGGC...GUCGCU M.,X.
GGCAGU...AUUGCU Y.
GGCCGC...GUGGCU P.
74
4103
UCGGCUCUUCCU...AGGGAACGUGAGCUGG M.,X.,Y.,P.
UCGGCUCAUCAC...GUGGUACGCGAGCUGG E. coli
75
4119
UUGUGAA...UUCACCAA M.,X.
UACCGAA...UUCGGUAA Y.
76
4144
UGGAU...ACCCA M.,X.,Y.,P.
AUGGC...GCCAU E.coli
77
4185
GACCGUC...GACAGGUU M.,X.,Y.,P.
GAACGUC...GACAGUUC E.coli
78
4226
GUUGUUGCC...GGAACGAU M.,X.
GUUACCGC....GCGGUGAU Y.
79
4236
UGG...CCA M.,X.
UAG...CUA Y.
80
4243
CCUGC...GCAGG M.,X.
UGAAC...GUUCA Y.
UCACU...AGUGA P.
81*, 82*
See Fig. 2.
83 a
4330
GGAUUAUGACU...AGUCAGAAUCC M.,X.
GGAUUAUGGCU...AGUCAGAAUCC Y.
GGAUGAAGGCU...AGCCCGAAGCC P.
83b
Several changes are not compensatory among eukaryotes.
84**
85**
86**
See Fig. 7.

We will restrict a further discussion to the cases of helices which either are not supported by comparative criteria or represent new folding proposals as compared to the previous yeast model (2).

No evidence is available for helical stems 9-11, 25, 52, 58 and 83b. However, for stems 25, 52 and 58 the base-pairing is possible in all species (but is not proven since the sequence has remained unchanged).

As for the base-pairing modifications within the conserved core by reference to the yeast model, they are mostly based on the availability of additional comparative proofs, provided by the recently published eukaryotic sequences : two major changes in a couple of long range interactions (helices 13 and 37) are proposed, together with a few rearrangements of the details of some elementary structural features (helices with underlined numbers in Table 1). Moreover we present a folding pattern for a domain for which no proposal had been made so far and which seems likely to be part of the conserved structural core (namely helices 39 to 45).

Helix 13 involves a long range base-pairing which deliminates a very large domain of 28S rRNA (about 1 kb in mouse). A markedly different interaction was proposed previously (2) in yeast for the corresponding oligonucleotides (yeast positions 651-657 and 1434-1440). However this alternate base-pairing is not validated by comparative analysis of the other eukaryotic sequences. Moreover, it is noteworthy that helix 13 as shown in Fig. 1 has its exact counterpart (positions 579-585/1255-1261) in the E. coli structure proposed by Noller et al. (9) : this is clearly apparent from the presence of some conserved sequence features in the vicinal single-stranded regions and by reference to the proximal conserved base-pairings (helices 14 and 32). This interaction, which involves two sequence tracts which are invariant in eukaryotes, is supported by compensatory changes between pro- and eukaryotes. An alternative base-pairing has been proposed for these segments in the two other prokaryotic models (7, 8) : it involves E. coli positions 578-584/805-811 and could provide the basis for a conformational switch. However, it is not similarly supported by comparative criteria in eukaryotic sequences. As for helix 37, evidence is even stronger, not only among prokaryotes but also among eukaryotes, with two base-pairs compensatorily changed between mouse, xenopus and yeast.

Comparative support for the slight modifications proposed for a few stem structures of the common core appears either in Table 1 or may be detailed in Figures. The consensus folding pattern that emerges for helices 32-33 and 81-82 is represented in Fig. 2, as a typical example.

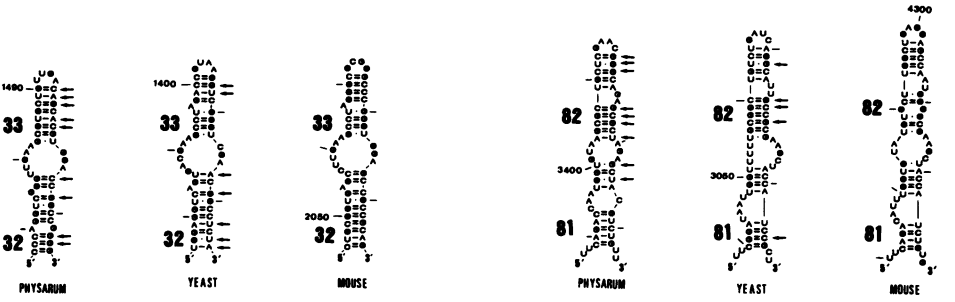
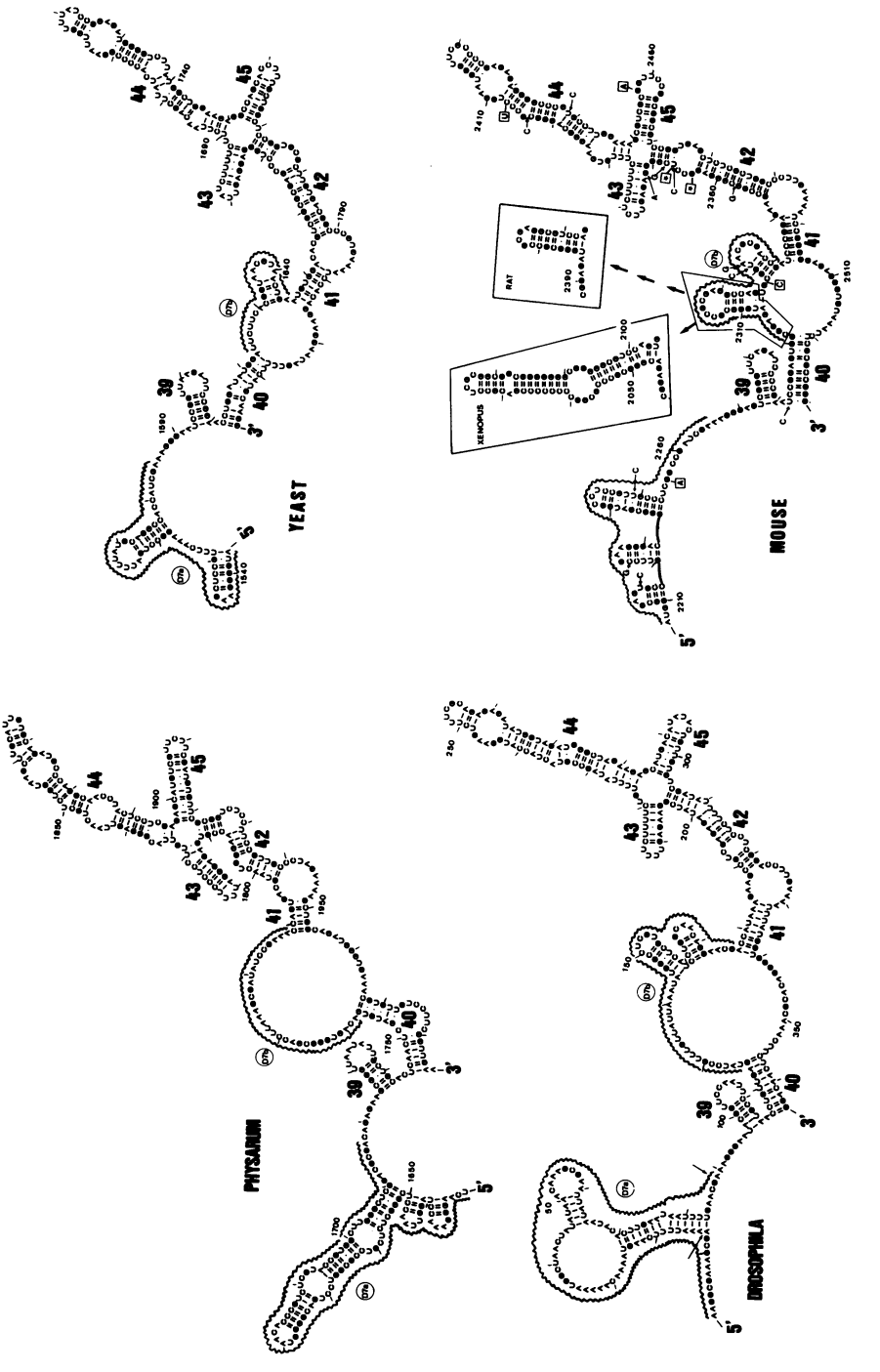


Fig. 2 : Example of detailed consensus folding patterns.
Minor changes as compared to the yeast model (2) are derived from the presence of a number of compensatory base changes in the other eukaryotic sequences by reference to mouse (denoted by arrows).

The case of the area encompassing helices 39-45 (Fig. 3) deserves more comments since no model had been proposed so far for this domain of eukaryotic large subunit rRNA. It is remarkable that a number of homologous secondary structure features mapping at equivalent positions relative to each other can be proposed for all eukaryotic sequences, while the structural conservation is interrupted over two highly divergent tracts (called D7a and D7b) which have undergone marked variations in size during the evolution. Examination of E. coli models (7-9) shows that helices 39 and 40 have their counterpart in 23S rRNA, with significant conservation of some primary sequence features at identical positions, either within loop 39 or within the single-stranded tract immediately upstream stem 39. The existence of stem 39, for which no comparative proof was apparent from the prokaryotic sequences (E. coli positions : 1385-1389/1398-1402) is validated by several compensatory changes among eukaryotes (Fig. 3). Helix 40 is definitely established by comparing prokaryotic sequences, particularly when including the more recently published ones (11, 12). This stable medium-range base-pairing is also conserved in all eukaryotes, although some irregularity is tolerated in yeast and Physarum (presence of a central bulge). A close comparison of the various sequences folded in Fig. 3 also provides strong evidence for helix 41, which is somewhat truncated in Physarum. The homology is also obvious for features 42-45, despite the presence of a few minor local differences. It must be stressed that a characteristic four-branched structure involving stems 43 and 45 can be proposed in all eukaryotes, as it is the case in one of the prokaryotic models (9). In conclusion, stems 39 to 45 appear to represent a part of the


YEAST
MOUSE
RAT
XENOPUS
PHYSARUM
DROSOPHILA

structural core of the large subunit rRNA, which is conserved among pro-and eukaryotes. Such an assignment has allowed a more precise delimitation of the two intervening size-variable D7a and D7b domains which have diverged much more rapidly during the evolution.

The rate of nucleotide change during the evolution of eukaryotes is far from being uniform along the entire conserved structural core. It is important to note that within some secondary structure features a particularly strong constraint has been exerted for maintaining definite primary sequence motifs or peculiarities in local conformation (such as bulged nucleotides) which are likely to play specific roles in a variety of intermolecular recognition processes involved in the ribosomal function as proposed earlier (9). Conversely in other helical stems extensive changes in primary sequence are tolerated provided the base-pairing potential is preserved. Assuming that the rates of nucleotide substitution have remained similar among different lineages during evolution, these conserved areas display nonetheless a sufficient degree of variability to represent valuable phylogenetic indicators (they amount to a total length that is about 15 times larger than the classical 5S rRNA index). Point changes have been scored in these areas between all pairs of eukaryotic 28SrRNA sequences and the corresponding degrees of divergence have been determined. All these values obtained for each possible pair of species (Table 2) are consistent with a unique phylogenetic tree topology - when mutational distances are equated to elapsed times since divergence from a common ancestor - which would indicate that Physarum polycephalum has diverged very early from the eukaryotic mainstream. The extensive comparison of all the secondary structure features of the common core among these species again indicates that Physarum features are the least related to the other eukaryotes. However, rather than being indicative of an early phylogenetic divergence, this clear departure could also merely result from substantial differences in the rates of fixation of mutations and in the processes of concerted evolu-

Fig. 3 : A new folding proposal for a part of the common structural core of eukaryotic 28S rRNA.
This region contains two size-variable domains (D7a and D7b) which are denoted by a wavy overline. Helical stems of the common structural core are numbered as in Fig. 1. For Drosophila melanogaster, the sequence shown corresponds to a partial determination (22) for the region of rRNA precursor which contains the central break of 26S rRNA (arrows denote the presumed 3'and 5' ends of 26S α and 26 Sβ). For rat and xenopus, the divergence, as compared to mouse, is restricted to a subarea of D7b domain which is represented in insets. Outside this subarea, only point changes are found between the vertebrates (xenopus : unboxed letters - rat : boxed letters).

Table 2 : Matrix of differences within the conserved core among eukaryotic 28S rRNAs.

	M.	R.	X.	Y.	P.
Mouse	-	0.27 %	1.44 %	12.7 %	21.7 %
Rat	5	-	1.60 %	12.8 %	21.7 %
Xenopus	27	30	-	12.4 %	21.9 %
Yeast	238	240	233	-	24.4 %
Physarum	408	409	411	458	-

For each pair of species, nucleotide changes have been scored within the twelve tracts (totalling 1880 nucleotides) where all eukaryotic sequences can be aligned without interruption, as described in (1). Percentages of divergence are indicated in the upper-right half of the matrix.

tion of the ribosomal genes (particularly in relation to the extrachromosomal status of ribosomal genes in Physarum).

2. Domains of variable size :

Conclusions that can be reached on the secondary structure folding of the few domains which have undergone major size variations during the evolution of eukaryotes (see Fig. 1 for location) are obviously less definitive. However, the availability of three vertebrate sequences - i.e. two closely related species (rat/mouse) and a much more distant one (amphibian Xenopus laevis) - provides a means to test structural models by comparative criteria and to gain further insight into the process of size increase of 28S rRNA in higher eukaryotes (1). We have recently proposed definite folding patterns for two divergent domains of small size, D6 and D9, as well as for one of the largest, i.e. D8 (1). In that latter case, our proposal was clearly substantiated by direct E.M. observations of vertebrate 28S rRNAs (23). While the case of D1 will be discussed in greater detail elsewhere, by means of a much more extensive comparative analysis carried out through direct rRNA sequencing in a variety of eukaryotes (L.H. Qu and J.P. Bachellerie, in preparation), we will examine here the status of all other D domains.

As shown in Fig. 4, the most stable secondary structure for D8 domain in lower eukaryotes does not bear a clear relatedness with its vertebrate counterpart - a Y-shaped giant stem structure (1) - except for the presence of long folded-back structures, while its prokaryotic equivalent is either considerably shorter (E. coli) or absent (as for A.n. or T.c.

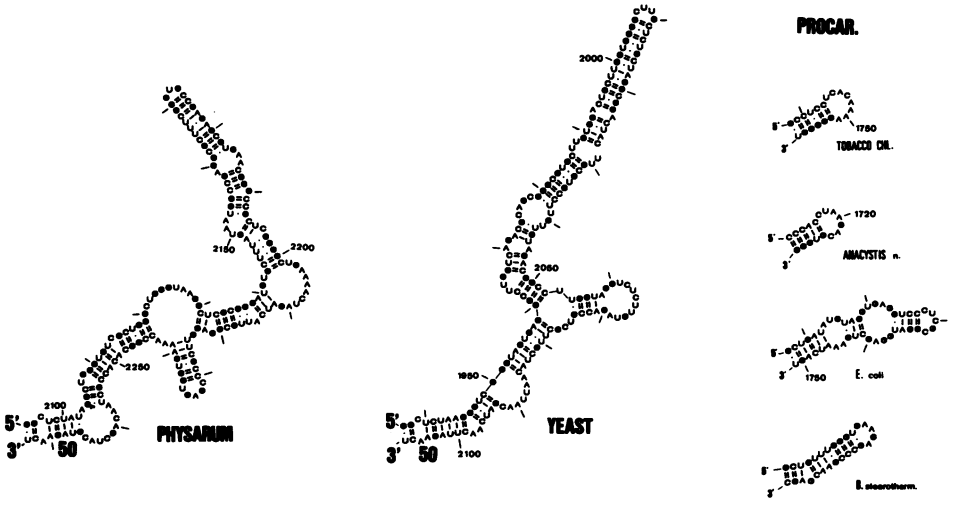


Fig. 4 : Secondary structure of the D8 domain.
The folding potential of this area in the 2 lower eukaryotes is compared to prokaryotic sequences (Helix 50 is a part of the conserved core). The folding of this domain in vertebrates is shown elsewhere - Fig. 9 in (1).

where this region is restricted to the conserved helix 50). Despite the clear analogy in the structures proposed for the two lower eukaryotes, the lack of residual sequence homology prevents from deriving any firm conclusion in relation to a clear conservation of some elementary structural features within this domain.

Similar observations emerge from the comparison of yeast and Physarum folding patterns in the largest of the divergent domains, D2 (Fig. 5). However it must be stressed that the stem involving the very distal segments of D2 is conserved in the two lower eukaryotes and in vertebrates despite a number of base changes (It maps at perfectly identical locations in all cases : on the 5'side it is separated from helix 12 by 5 unpaired nucleotides and, on the 3'side, it is immediately vicinal to a long sequence tract conserved in all eukaryotes). It is also clearly apparent that the entire D2 domains in the 3 vertebrates share analogous folding potentials ; the 3 long base-paired branches are present in all cases and the large size differences in this domain (277 additional nucleotides in rat as compared to xenopus) are mostly accommodated by differences in the length of these stems. Such folding patterns, which are particularly stable are also entirely consistent with mappings of secondary structures by direct E.M. obser-

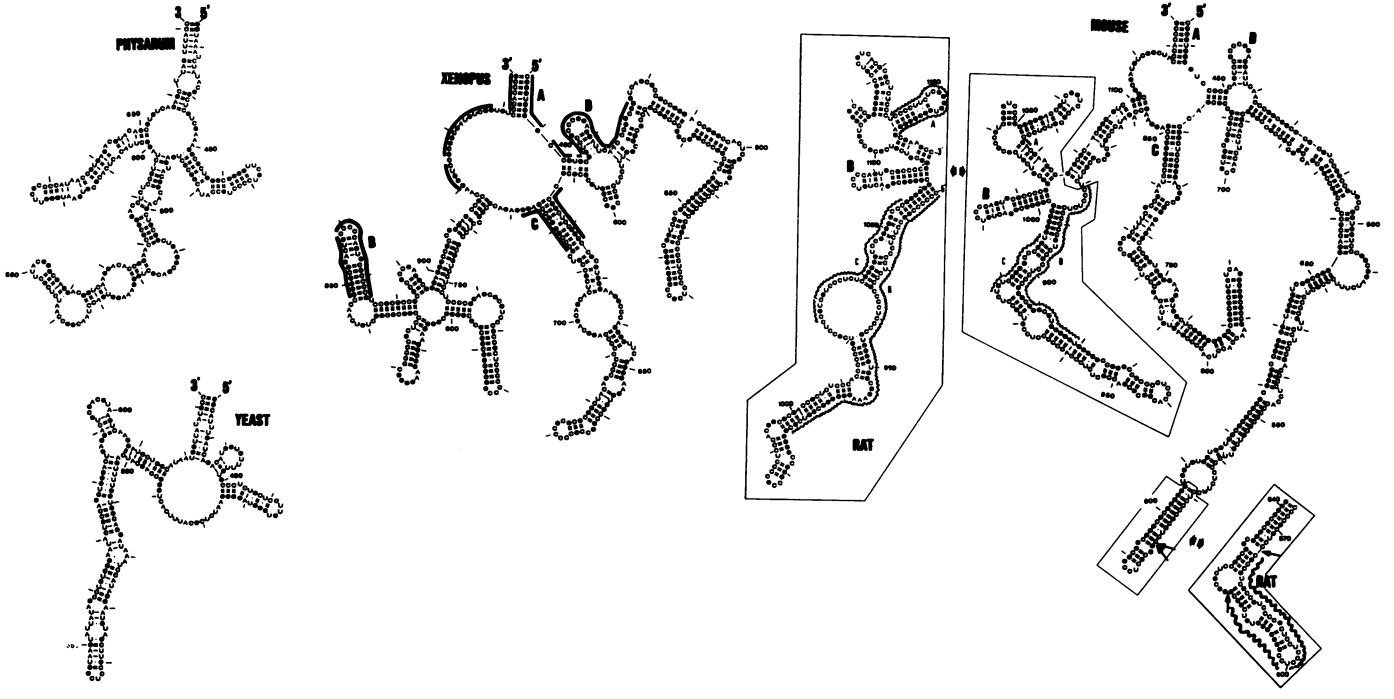


Fig. 5 : Secondary structure model of the D2 domain in eukaryotes.
For rat, differences with mouse are restricted to 2 areas which are shown in insets (the wavy lines denote segments which have extensively varied in sequence and size between both rodents. The 43 nucleotide long rat insert is represented in the bottom-right inset, with the site of insertion in mouse indicated by 2 arrows). For xenopus, the thick overline denote sequence tracts common to the 3 vertebrates (with thick letters referring to identical structures in xenopus and mouse).

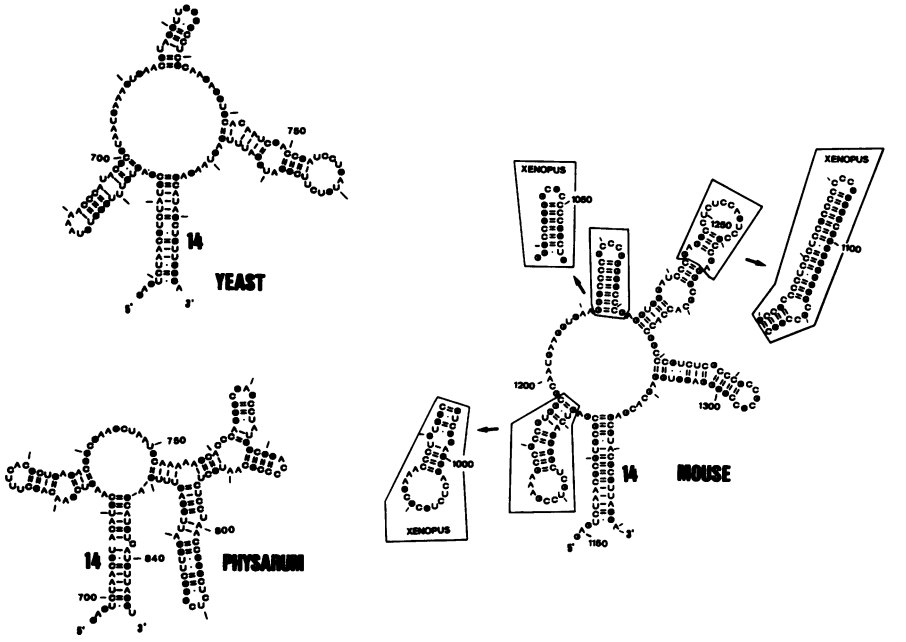


Fig. 6 : Folding of the D3 size-variable domain.
This divergent domain is closed by a conserved 14 bp stem (helix 14). For xenopus, differences with mouse are restricted to the areas represented in insets.

vation of vertebrate 28S rRNAs (23) : the characteristic "triple loop", seen in all vertebrates at similar locations, i.e. about 450 nucleotides from the 5'terminus of 28S rRNA (assignment of 5',3' polarity was incorrect in that work), is likely to correspond to the three-branched structures proposed in Fig. 5 which should not be denatured in the conditions used for E.M. analysis. Its relative size in xenopus and rat, as estimated from the E.M. data, is roughly in line with the lengths of the proposed stems in both species. It is noteworthy that D2 domain in rodents encompasses four small areas where major differences between mouse and rat are concentrated, with one of them corresponding in rat to a perfect 43 nucleotide long insert into the mouse sequence located at the very tip of one of the three long stems (Fig. 5), as discussed elsewhere (1).

Size variations among vertebrates accommodated through amputation or extension of the terminal regions of long GC-rich stems are also apparent in D6, D8 and D9 domains, as discussed in (1), as well as for D3 and D12, as shown in Fig. 6 and 7 respectively. While there are only a few cases of

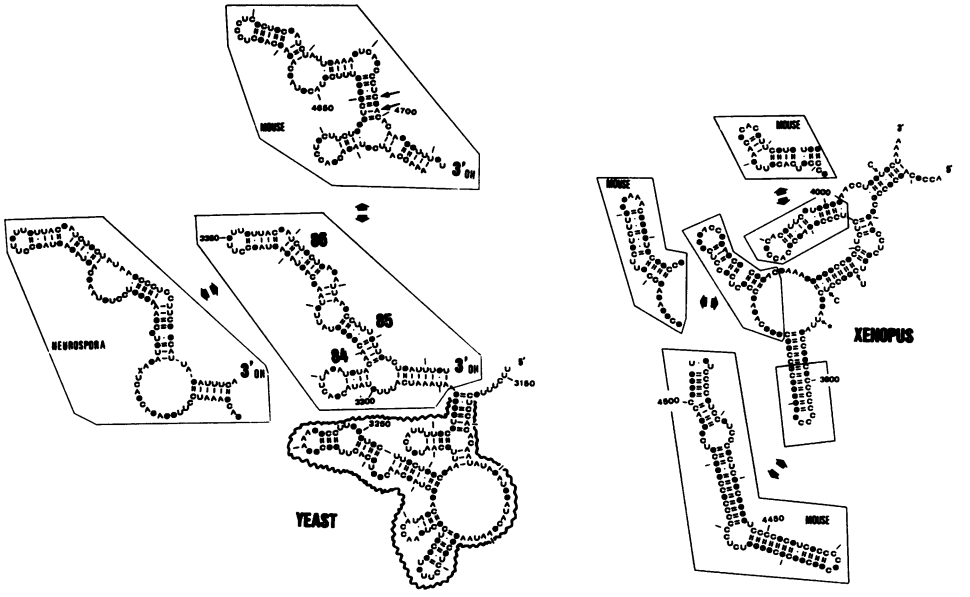


Fig. 7 : Folding of the 3'terminal area of eukaryotic 28S rRNA : Only yeast 26S rRNA is folded in its entirety. The mouse and Neurospora crassa (17) versions of the boxed area are represented in insets (with arrows in the mouse insets showing positions of compensatory changes as compared to xenopus), while the wavy overline denotes the D12 divergent domain. The counterpart of this overlined region is also shown for xenopus : outside the boxed regions - which are highly divergent between xenopus and mouse (mouse structure in insets) - only point changes (superimposed letters) are found between both vertebrates.

stem extension in xenopus as compared to rodents (see one inset of Fig. 6 and also Fig. 3), the trend for size increase from amphibian to rodents is almost general as shown above for D2 and elsewhere (1) for D6, D8 and D9, and as confirmed for D12 (Fig. 7). The comparison of mouse and xenopus folding patterns suggests that some elementary structural features within D3 and D12 may be conserved in vertebrates despite divergence in primary sequence (see 2 insets in Fig. 6). Moreover the conservation extends to lower eukaryotes for the stem in D3 (Fig. 6) which is 3' vicinal to helix 14 (and separated from it by two unpaired purines). A prokaryotic equivalent may be identified for this stem (E. coli, positions 604-624) which therefore could be included in the common structural core.

In divergent domain D5 (Fig. 8), a stem common to all eukaryotes can be formed immediately adjacent to helix 22; however evidence of compensa-

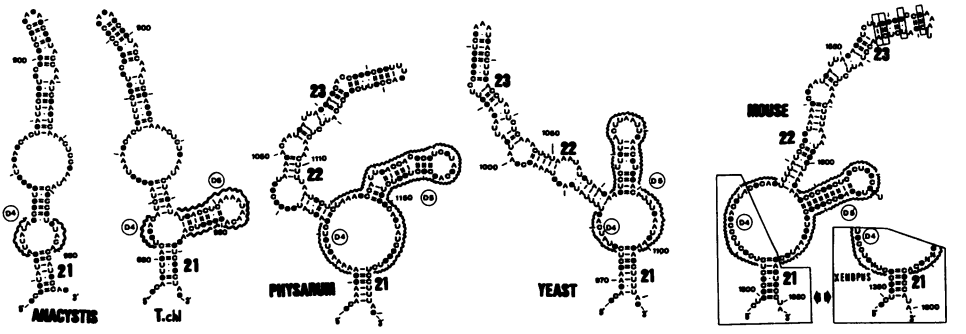


Fig. 8 : Secondary structure model of the D4-D5 domains.
The two size-variable areas are denoted by a wavy line. Adjacent features of the common core are numbered as in Fig. 1. For xenopus, the divergence with mouse is restricted to the area shown in inset (outside this area, only 2 point changes - superscribed letters - are found). Boxes base-pairings in mouse are compensated in yeast (helix 23). The folding of a prokaryotic (Anacystis nidulans) or prokaryote-related (Tobacco chloroplast) sequence is also shown.

tory changes is not straightforward from the sole sequences available so far. An homologous structural feature is present in tobacco chloroplast 23S rRNA at this position, but it is clearly missing in strictly prokaryotic sequences, such as E. coli or A. nidulans.

In the case of domain D12, it is not possible to unambiguously identify a correspondence, between lower and higher eukaryotes, for most of the structural features except for the conservation of a base-pairing potential between both termini of this domain in yeast and vertebrates (Fig. 7). Although the stem involving the very last 3'terminal nucleotides of 28S rRNA can be proposed not only for rodents but also for lower eukaryotes (Fig. 7), it can only be regarded so far as tentative, particularly when considering that it has no stable prokaryotic counterpart. Concerning the structure of both termini of the molecule, it must be remembered that while 5.8S rRNA does represent the eukaryotic counterpart of the 5'end of 23S rRNA (24, 25, 9, 8, 19), the very distal oligonucleotides (about 10 nucleotides at each terminus), which are involved in a base-pairing rejoining both ends of the prokaryotic molecule, appear to have no equivalent in eukaryotes, with the composite 5.8S/28S rRNA molecule slightly truncated at both ends.

3. Structural domains and functional sites in large subunit rRNA.

The conservation of a closely homologous structural core during the evolution obviously suggests that these common structural features parti-

cipate in basic functions shared by pro- and eukaryotic ribosomes. A number of them in prokaryotic 23S rRNA have been identified as binding sites for some ribosomal proteins, 5S rRNA and other components of the protein synthesis machinery (7-9) and are therefore likely to be involved in homologous interactions in eukaryotes as discussed earlier (2, 5). Much more intriguing is the potential functional significance of the so-called divergent domains which have undergone a major and concerted increase in size during the evolution of higher eukaryotes. As already mentioned (1), these domains of mature 28S rRNA are clearly related to the transcribed spacers which are removed during rRNA processing and in some species one of these spacer-like regions may happen to be eliminated from mature rRNA : the equivalent of D12 domain in plant chloroplasts, like maize's (8) or tobacco's (12) precisely corresponds to the 23S-4.5S rRNA intergenic spacer which is excised from pre rRNA. An analogous situation is found for D7a : its counterpart in Drosophila is the "central break" region which is removed from 26S rRNA precursor to give rise to 26S α and 26S β halves (22), as shown in Fig. 3. One may expect that the availability of additional eukaryotic 28S rRNA sequences and of direct informations related to their molecular environment within the eukaryotic large ribosomal subunit will provide further insight on the role of these rapidly evolving domains.

ACKNOWLEDGMENTS

We are grateful to H. Noller for communicating a manuscript before publication and to F. Michel for providing computer programs and for generous assistance in their adaptation. The support of Pr. J.P. Zalta was appreciated. We thank M. Parès for typing the manuscript. This work was supported by a grant from C.N.R.S. ("ATP Organisation du Génome Eucaryote"), by an ADRC contract (6107), by a fellowship from the Ligue Nationale Française Contre le Cancer to B.M. and a schlolarship (Ministère de la Coopération) to N.H.

*To whom correspondence should be addressed

REFERENCES

1. Hassouna, N., Michot, B. and Bachellerie, J.P. (1984) submitted.
2. 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.
3. 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.
4. Otsuka, T., Nomiyama, H., Yoshida, H., Kukita, T., Kuhara, S. and Sakaki, Y. (1983) Proc. Natl. Acad. Sci. USA 80, 3163-3167.

5. Ware, V.C., Tague, B.W., Clark, C.G., Gourse, R.L., Brand, R.C. and Gerbi, S. (1983) Nucleic Acids Res. 11, 7795-7817.
6. Chan, Y.L., Olvera, J. and Wool, I.G. (1983) Nucleic Acids Res. 11, 7819-7831.
7. Glotz, C., Zwieb, C., Brimacombe, R., Edwards, K. and Kössel, H. (1981) Nucleic Acids Res. 9, 3287-3306.
8. Branlant, C., Krol, A., Machatt, M.A., Pouyet, J., Ebel, J.P., Edwards, K. and Kössel, H. (1981) Nucleic Acids Res. 9, 4303-4324.
9. Noller, H.F., Kop, J., Wheaton, V., Brosius, J., Gutell, R.D., Kopylov, A., Dohme, F., Herr, W., Stahl, D.A., Gupta, R. and Woese, C.R. (1981) Nucleic Acids Res. 9, 6167-6189.
10. Brosius, J., Dull, T.J. and Noller, H.F. (1980) Proc. Natl. Acad. Sci. USA 77, 201-204.
11. Kumano, M., Tomioka, N. and Sugiura, M. (1983) Gene 24, 219-225.
12. Takaiwa, F. and Sugiura, M. (1972) Eur. J. Biochem. 124, 13-19.
13. Michel, F., Jacquier, A. and Dujon, B. (1982) Biochimie 64, 867-881.
14. Gourse, R.L., Thurlow, D.L., Gerbi, S.A. and Zimmermann, R.A. (1981) Proc. Natl. Acad. Sci. USA 78, 2722-2726.
15. Rae, P.M.M., Kohorn, B.D. and Wade, R.P. (1980) Nucleic Acids Res. 8, 3491-3504.
16. Roiha, H. and Glover, D.M. (1981) Nucleic Acids Res. 9, 5521-5532.
17. Kelly, J.M. and Cox, R.A. (1981) Nucleic Acids Res. 9, 1111-1121.
18. Hindenach, B.R. and Stafford, D.W. (1984) Nucleic Acids Res. 12, 1737-1747.
19. Michot, B., Bachellerie, J.P. and Raynal, F. (1982) Nucleic Acids Res. 10, 5273-5283.
20. Qu, L.H., Michot, B. and Bachellerie, J.P. (1983) Nucleic Acids Res. 11, 5903-5920.
21. Ursi, D., Vandenberghe, A. and De Wachter, R. (1983) Nucleic Acids Res. 11, 8111-8120.
22. Delanversin, G. and Jacq, B. (1983) C.R. Acad. Sci., ser. III, 296, 1041-1044.
23. Schibler, U., Wyler, T. and Hagenbüchle, O. (1975) J. Mol. Biol. 94, 503-517.
24. Nazar, R. (1980) FEBS Lett. 119, 212-214.
25. Jacq, B. (1981) Nucleic Acids Res. 9, 2913-2932.