J Mol Evol (1989) 28:403–417

Journal of Molecular Evolution

Sequence and Secondary Structure of the Central Domain of *Drosophila* 26S rRNA: A Universal Model for the Central Domain of the Large rRNA Containing the Region in which the Central Break May Happen

Guy de Lanversin and Bernard Jacq

Laboratoire de Génétique et Biologie Cellulaires, CNRS, Case 907, 13288 Marseille Cedex 9, France

Summary. An 890-bp sequence from the central region of *Drosophila melanogaster* 26S ribosomal DNA (rDNA) has been determined and used in an extensive comparative analysis of the central domain of the large subunit ribosomal RNA (lrRNA) from prokaryotes, organelles, and eukaryotes. An alignment of these different sequences has allowed us to precisely map the regions of the central domain that have highly diverged during evolution. Using this sequence comparison, we have derived a secondary structure model of the central domain of *Drosophila* 26S ribosomal RNA (rRNA). We show that a large part of this model can be applied to the central domain of lrRNA from prokaryotes, eukaryotes, and organelles, therefore defining a universal common structural core. Likewise, a comparative study of the secondary structure of the divergent regions has been performed in several organisms. The results show that, despite a nearly complete divergence in their length and sequence, a common structural core is also present in divergent regions. In some organisms, one or two of the divergent regions of the central domain are removed by processing events. The sequence and structure of these regions (fragmentation spacers) have been compared to those of the corresponding divergent regions that remain part of the mature rRNA in other species.

Key words: Large subunit ribosomal RNA — Central domain — RNA secondary structure — Divergent region — Fragmentation spacer — Evolution

Offprint requests to: B. Jacq

Introduction

The structure and function of ribosomes have been studied extensively for many years through various technical approaches, including biophysical and biochemical methods, genetics, and molecular biology (see Wittmann 1985 for a recent review). One of the most recent areas of progress has been the elucidation of the primary structure of RNA and protein components from the ribosome. If, at the level of ribosomal protein sequences, our knowledge is essentially restricted to *Escherichia coli,* the situation is completely different with respect to rRNA sequences: More than 350 different 5S rRNA have been sequenced (Erdmann and Wolters 1986), and several complete sequences are available for the large RNAs of the small and large ribosomal subunits (srRNA and lrRNA, respectively) from a number of evolutionarily distant organisms (see Huysmans and de Wachter 1986, for a compilation of 57 srRNA, and Table 1 for a representative set of 33 lrRNA). At the same time that this information was accumulating, many attempts were made to fold the sequences into secondary structure models, using experimental, theoretical, and comparative approaches. In the case of prokaryotic lrRNA (Noller 1984; Brimacombe and Stiege 1985) and srRNA from prokaryotes and eukaryotes (Gutell et al. 1985), the models that have been proposed by different groups have now converged toward a consensus model, although some minor discrepancies still exist. As far as eukaryotic lrRNA are concerned, agreement between the models from different sources has not yet reached the same level. Different explanations can be proposed: first, the determination of

eukaryotic rRNA sequences is more recent than that of prokaryotic ones, and the size of lrRNA is larger in eukaryotes (5 kb in human instead of 2.9 kb in *E. coli* for instance); second, the comparative approach was used to a lesser extent and was complicated by the presence of several regions of variable length and sequence in eukaryotes (Michot et al. 1984); third, direct secondary structure probing experiments are lacking in eukaryotes, except for some specific regions, namely the 5.8S region and the L25 binding region (Walker et al. 1982; El-Baradi et al. 1985).

In the present study, we have determined an 890-bp sequence from the *Drosophila* 26S rRNA gene, covering the region in which the limits of the central gap created by a nuclear processing event have been mapped (de Lanversin and Jacq 1983). This sequence was compared to several corresponding sequences from prokaryotic and eukaryotic organisms. The resulting alignments were used in an extensive search for a common secondary structure model of the central domain of the lrRNA, for which conflicting models have been proposed. The central domain is particularly interesting in terms of rRNA evolution, because it contains three divergent regions, some of which are eliminated by processing events in lower eukaryotes. Our results lead to a unified general model in which prokaryotes and eukaryotes show a common structural core, and where the regions of size and sequence variations are precisely mapped both at the primary and secondary structure levels. Moreover, we propose some new universal interactions supported by evolutionary compensating base changes (CBCs) that are present in the central domain of prokaryotic, chloroplastic, eukaryotic, and some mitochondrial lrRNA.

Materials and Methods

The plasmid PY22 carrying a complete 11.5-kb rRNA transcription unit from *Drosophila melanogaster* (Dawid et al. 1978) was used in this study. The methods for preparing plasmid DNA and purifying specific DNA fragments have been described previously (de Lanversin and Jacq 1983). The sequencing was done according to Maxam and Gilbert (1980), and the chemical degradation products were run on 8% and 15% acrylamide–bisacrylamide denaturing gels.

DNA sequence editing, restriction site and base composition analysis, and dot-matrix comparisons of sequences have been achieved using Apple II and Macintosh Pascal programs (Malthiery et al. 1984; Bellon 1988). Comparisons of sequence data to EMBL and GENBANK libraries and establishment of hairpin catalogs and of partial secondary structure models have been done using banks and programs stored in the CITI-2 center (Université René Descartes, Paris). In order to perform a comparative sequence analysis of the central domain of lrRNA, a set of 33 sequences representative of bacterial, mitochondrial, chloroplastic, and eukaryotic lrRNA was first selected (Table 1). In all these sequences, we only considered the region equivalent to domain III of *E. coli* 23S rRNA or domain IV of eukaryotic 26–28S rRNA. Throughout this paper, we will call this region the central region or domain because it always contains the middle of the sequence and because in eukaryotes this domain is surrounded by three domains on the 5′ side (domains I–III) and three domains on the 3′ side (domains V–VII). The 5′ and 3′ limits of the central domain were determined as follows. For each sequence, the presence of two secondary features characteristic of the border was verified, i.e., helix A on the 5′ side and the long-range pairing constituted by helix B on the 3′ side (see Figs. 3 and 5). The G belonging to the conserved trinucleotide GUA located 7–11 nucleotides upstream of helix A was used as a 5′ landmark for the central domain (this trinucleotide is present at this position in all sequences listed in Table 1). On the 3′ side, it has been shown that the domain following the central one starts 2–4 nucleotides after the 3′ end of helix B (Michot et al. 1984; Noller 1984). As a 3′ limit for the central domain, we arbitrarily chose the nucleotide that in each sequence is equivalent to the C immediately following helix B in *E. coli*. These two 5′ and 3′ landmarks are indicated by vertical arrows located, respectively, at the top and the bottom of Fig. 3, and the corresponding limits in the different sequences are indicated in the middle column of Table 1. A computer-aided primary structure alignment was then made between different central domain sequences, and a final refinement was made by aligning regions involved in conserved secondary structure elements (Fig. 3). This precise multiple alignment is the basis of our secondary structure modeling comparative approach that has been detailed elsewhere (de Lanversin et al. 1987).

Results and Discussion

Analysis of the Central Region from Drosophila 26S rDNA and Comparison with Homologous Regions from Other Organisms

An 890-bp DNA sequence covering the entire central domain of the *Drosophila* 26S rRNA gene, as well as part of the two adjacent domains, has been determined using the sequencing strategy shown in Fig. 1 and as presented in Fig. 2. At the RNA level, this sequence includes the central break region of the 26S rRNA precursor (de Lanversin and Jacq 1983). The total G+C content of this sequence is only 41.2%, a value that is slightly above that of 38% found for the entire 26S rRNA (Tartof and Perry 1970). This percentage is far from being uniform along the 890 nucleotides, because two regions, centered around residues 113 and 355, are very A+U rich when compared to the remaining part of the sequence. Transcribed spacers in *Drosophila* rRNA have a higher A+U content than the mature sequences (Pavlakis et al. 1979; Jordan et al. 1980), and it is interesting to note that the second A+U-rich region encompasses the central gap of the 26S rRNA that can be considered as an internal transcribed spacer (de Lanversin and Jacq 1983). However, as exemplified by the first A+U-rich region, which is present in the mature RNA, not all A+U-rich regions from *Drosophila* rRNA are removed by processing events.

Heterologous hybridization experiments showed

Table 1. List of sequences used for the comparative study of the central domain of the large rRNA

Species		Central domain	References
Archaebacteria			
Halobacterium halobium	(H.h.)	1357–1707	Mankin and Kagramanova 1986
Halococcus morrhuae	(H.m.)	1376–1727	Leffers et al. 1987
Methanobacterium thermoautotrophicum	(M.t.)	1412–1811	Leffers et al. 1987
Methanococcus vannielii	(M.v.)	1347–1740	Jarsch and Böck 1985
Desulfurococcus mobilis	(D.mo.)	1434–1838	Leffers et al. 1987
Thermoproteus tenax	(T.t.)	1398–1794	Kjems et al. 1987
Eubacteria			
Escherichia coli	(E.c.)	1266–1646	Brosius et al. 1980
Bacillus subtilis	(B.s.)	1304–1689 (3511–3896)	Green et al. 1985
Anacystis nidulans	(A.n.)	1275–1651	Douglas and Doolittle 1984
Chloroplasts			
Zea mays chloroplast	(Z.m. cp)	1297–1761	Edwards and Kössel 1981
Nicotiana tabacum chloroplast	(N.t. cp)	1284–1678	Takaiwa and Sugiura 1982
Chlorella ellipsoidea chloroplast	(C.e. cp)	1589–2045	Yamada and Shimaji 1987
Mitochondria			
Paramecium primaurelia mitochondria	(P.p. mt)	1434–1688	Seilhamer et al. 1984
Paramecium tetraurelia mitochondria	(P.t. mt)	1432–1686	Seilhamer et al. 1984
Schizosaccharomyces pombe mitochondria	(S.p. mt)	1229–1466	Lang et al. 1987
Oenothera berteriana mitochondria	(O.b. mt)	1505–2070	Manna and Brennicke 1985
Zea mays mitochondria	(Z.m. mt)	1445–2351	Dale et al. 1984
Ascomycota			
Saccharomyces carlsbergensis	(S.ca.)	1445–1826	Veldman et al. 1981
Saccharomyces cerevisiae	(S.ce.)	1446–1826	Georgiev et al. 1981
Kinetoplastida			
Crithidia fasciculata	(C.f.)	1539–2209 (2510–3180)	Spencer et al. 1987
*Trypanosoma brucei**	(T.b.)	14–ND	Campbell et al. 1987
Myxomycota			
Physarum polycephalum	(P.p.)	1534–2032	Otsuka et al. 1983
Dictyostelium discoideum	(D.d.)	1683–2159 (6414–6890)	Ozaki et al. 1984
Nemathelminthes			
Caenorhabditis elegans	(C.e.)	1512–1695 (4153–4611)	Ellis et al. 1986
Arthropoda			
*Artemia salina**	(A.s.)	ND	Nelles et al. 1984
*Drosophila melanogaster**	(D.m.)	216–720	This paper
*Sciara coprophila**	(S.c.)	14–ND	Ware et al. 1985
*Bombyx mori**	(B.m.)	383–ND	Fujiwara and Ishikawa 1986
Chordata			
Xenopus laevis	(X.l.)	1854–2362	Ware et al. 1983
Mus musculus	(M.m.)	2116–2578	Hassouna et al. 1984
Rattus norvegicus	(R.n.)	2201–2665	Chan et al. 1983
		2188–2660	Hadjiolov et al. 1984
Homo sapiens	(H.s.)	2338–2803	Gonzalez et al. 1985
Spermatophyta			
Oryza sativa	(O.s.)	1442–1871	Takaiwa et al. 1985

The name, the abbreviation used throughout this study (in parentheses), the 5′ and 3′ limits of the central domain, and the reference of the publication of the corresponding sequence are indicated for each species. The 5′ and 3′ limits of the central domain were determined as indicated in the Materials and Methods section. The numbering of the corresponding nucleotides in the different sequences is indicated in the middle column. The numbers in parentheses refer to the limits of the central domain in the case where the published sequence was larger than that of the lrRNA alone. In the four species indicated by an *, the complete sequence of either the lrRNA or the central domain was not available, and the numbering therefore does not reflect the actual length of the molecule (ND: not determined)

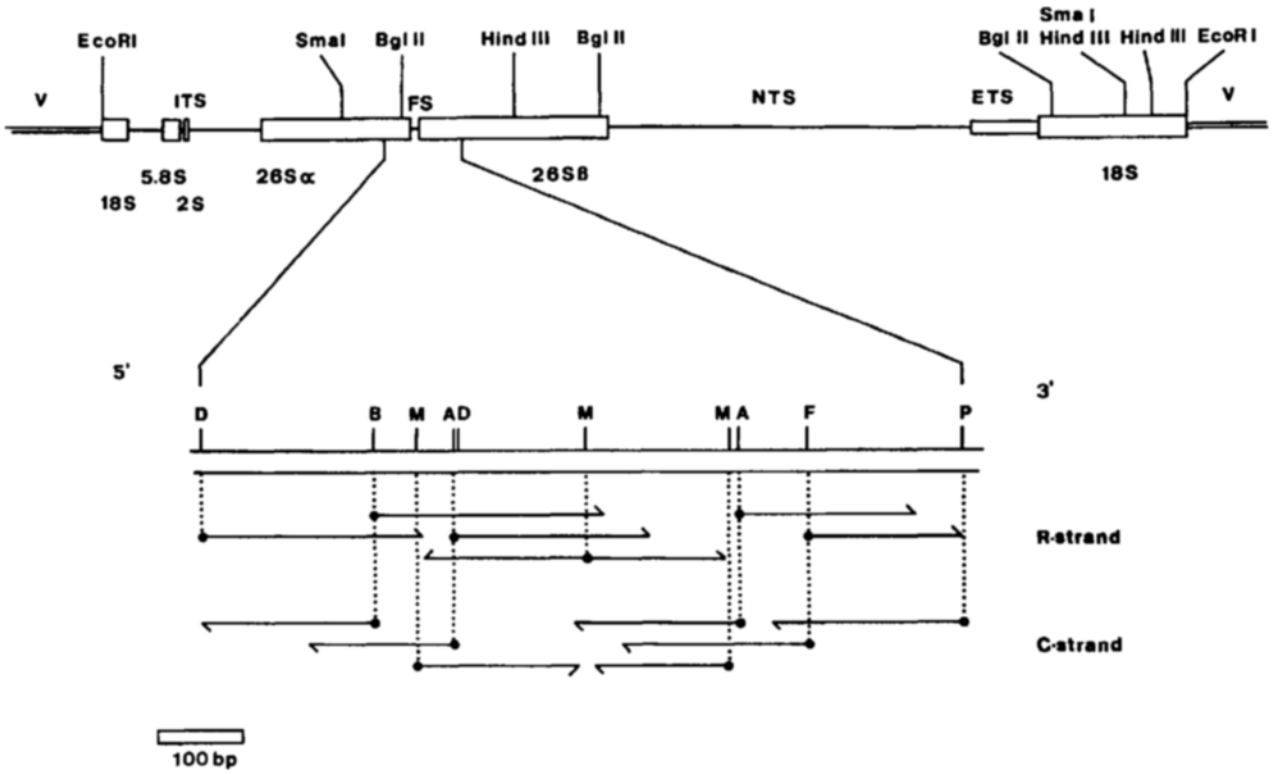

Fig. 1. Restriction map and sequencing strategy of the central region from *Drosophila* 26S rDNA. At the top of the figure is shown a physical and restriction map of the PY22 plasmid containing a normal *Drosophila melanogaster* ribosomal unit (without insertion in the 26S gene) cloned in the PmB9 vector. *Abbreviations:* ETS, external transcribed spacer; FS, fragmentation spacer; ITS, internal transcribed spacer; NTS, nontranscribed spacer; V, vector. The DNA region that has been sequenced is shown expanded at the bottom of the figure. Vertical bars indicate the position of the restriction enzymes used for sequencing. Horizontal arrows show the direction and extent of DNA segments that were sequenced. R strand, RNA-like strand; C strand, complementary strand. Restriction enzyme sites are indicated by the following abbreviations: A, AvaII; B, BglII; D, DdeI; F, HinfI; M, MnlI; P, HpaII.

the existence of evolutionarily conserved regions within eukaryotic rDNA (Sinclair and Brown 1971; Gerbi 1976). Sequence comparisons of rRNA genes from various origins later confirmed this finding and also showed that regions that have been strongly conserved during evolution are separated by regions of variable length and sequence (e.g., Ware et al. 1983; Hassouna et al. 1984). Three such divergent regions can be identified when the sequence of the central domain from *Drosophila* 26S rRNA is aligned with corresponding prokaryotic and eukaryotic sequences (Fig. 3). Two of them (D7a and D7b) were identified previously in a comparative analysis of the mouse 28S rRNA sequence (Hassouna et al. 1984). The D7b variable region was also identified in a comparison between archaebacterial sequences (Leffers et al. 1987). The presence of another variable region (D7c) in the central domain of eukaryotic 28S rRNA has recently been suggested in a study of the lrRNA of *Crithidia fasciculata* (Spencer et al. 1987). The sequence comparison of Fig. 3 strongly supports the existence of D7c in the central domain of the lrRNA; this conclusion is further reinforced by alignments (data not shown) performed with corresponding sequences from archaebacteria, eubacteria, and eukaryotic cytoplasmic lrRNA (see Table 1 for a compilation of the sequences that were used). These sequence data also favor the idea that, in a comparison between prokaryotes and eukaryotes, no more than three divergent regions can be found in the central domain of lrRNA (Fig. 4).

We have also used sequences from organelles to see if additional divergent regions could be detected in the central domain. The *Zea mays* (Edwards and Kössel 1981) and *Chlorella ellipsoidea* (Yamada and Shimaji 1987) chloroplast 23S sequences exhibit, respectively, a 65- and 43-nucleotide insertion in the central domain when compared to the *E. coli* sequence. The positions of these insertions (arrowheads in Figs. 3 and 5) clearly identify two new divergent regions (tentatively called D7d and D7d') between chloroplasts and prokaryotes or eukaryotes.

In the case of mitochondria, conclusive phylogenetic evidence was hard to obtain in order to decide about the number and location of divergent sequences in the central domain between mito-

chondria and prokaryotes or eukaryotes. The reasons for this are, first, the size of mitochondrial lrRNA ranges from 1250 nucleotides in trypanosomatids to more than 3800 nucleotides in plants (reviewed in Curgy 1985); and second, this extreme variation in size is complicated by an apparent lack of sequence conservation among mitochondria, and also with respect to prokaryotic or eukaryotic organisms.

Processing Events Occurring in the Central Domain of Eukaryotic lrRNA

It has been known for some time that the lrRNA of several lower eukaryotes contains a "hidden break" located near the center of the molecule (see Ishikawa 1977 for a review). In the case of *Drosophila* 26S rRNA, this cleavage is not an artifact of rRNA extraction and purification, but rather a true maturation step, because newly synthesized rRNA molecules are not dissociable upon denaturation (Jordan 1975). Moreover, pulse-chase experiments, coupled with fractionation of *Drosophila* culture cells into nuclear and cytoplasmic fractions have shown that this maturation step takes place in the nucleus (Jordan et al. 1976). Recent studies on trypanosomatid protozoa (White et al. 1986; Campbell et al. 1987; Spencer et al. 1987) have revealed that the lrRNA from these species is submitted to a complex pattern of processing events giving rise to six polynucleotide chains, and that two cleavages are occurring in the central domain of the RNA molecule.

RNA mapping experiments and sequence data are now available from some organisms in which the central domain of the lrRNA is submitted to processing cleavages. In a crustacean, *Artemia salina* (Nelles et al. 1984), and in three insects, *Drosophila melanogaster* (de Lanversin and Jacq 1983), *Sciara coprophila* (Ware et al. 1985), and *Bombyx mori* (Fujiwara and Ishikawa 1986), the central processing event can be unambiguously mapped to the D7a region. In the two trypanosomatids examined so far (Campbell et al. 1987; Spencer et al. 1987), the two central cleavages concern both the D7b and D7c regions, but not the D7a region. It is interesting to note that in all these cases, the processing events lead not only to a break of a phosphodiester bond in the polynucleotide chain, but also to the elimination of nucleotides from the precursor molecule. These regions that appear to be dispensable in several species represent examples of "fragmentation spacers," i.e., sequences that can be removed by processing events from the mature rRNA, leading to a product that is fragmented in two or more pieces, usually detectable upon analysis in denaturing conditions. In *S. coprophila* (Ware et al. 1985), 19 nucleotides of the D7a region are removed from the

```
   1 TAAGGAGTGTGTAACAACTCACCTGCCGAAGCAACTAGCCCTTAAAATGGATGGCGCTTA
  61 AGTTGTATACCTATACATTACCGCTAAAGTAGATGATTTATATTACTTGTGATATAAATT
 121 TTGAAACTTTAGTGAGTAGGAAGGTACAATGGTATGCGTAGAAGTGTTTGGCGTAAGCCT
                                  ↓
 181 GCATGGAGCTGCCATTGGTAGAGATCTTGGTGGTAGTAGCAAATAATCGAATGAGACCTT
 241 GGAGGACTGAAGTGGAGAAGGGTTTCGTGTGAACAGTGGTTGATCACGAGTTAGTCGGTC
 301 CTAAGTTCAAGGCGAAAGCGAAAATTTTCAAGTAAAACAAAAATGGCTAACTATATAAAC
 361 AAAGCGAATTATAATACACTTGAATAATTTTGAACGAAAGGGAATACGGTTCCAATTCCG
 421 TAACCTGTTGAGTATCCGTTTGTTATTAAATATGGGCCTCGTGCTCATCCTGGCAACAGG
 481 AACGACCATAAAGAAGCCGTCGAGAGATATCGGAAGAGTTTTCTTTTCTGTTTTATAGCC
 541 GTACTACCATGGAAGTCTTTCGCAGAGAGATATGGTAGATGGGCTAGAAGAGCATGACAT
 601 ATACTGTTGTGTCGATATTTTCTCCTCGGACCTTGAAAATTTATGGTGGGGACACGCAAA
                                                              ↓
 661 CTTCTCAACAGGCCGTACCAATATCCGCAGCTGGTCTCCAAGGTGAAGAGTCTCTAGTCG
 721 ATAGAATAATGTAGGTAAGGGAAGTCGGCAAATTAGATCCGTAACTTCGGGATAAGGATT
 781 GGCTCTGAAGATTGAGATAGTCGGGCTTGATTGGGAAACAATAACATGGTTTATGTGCTC
 841 GTTCTGGGTAAATAGAGTTTCTAGCATTTATGTTAGTTACTTGTTCCCGG
```

Fig. 2. Nucleotide sequence of the central domain and adjacent regions of *Drosophila* 26S rDNA. This sequence (RNA-like strand) was determined using the strategy depicted in Fig. 1. The unique BglII site of the sequence is boxed. The vertical arrows at the top and the bottom of the sequence indicate, respectively, the 5' and 3' limits of the central domain of the 26S molecule (see Figs. 3 and 5).

precursor, whereas in *Trypanosoma brucei* (Campbell et al. 1987), as many as 157 and 75 nucleotides, respectively, are removed from the D7b and D7c regions. From these different results it can be concluded that, as far as the central domain of the lrRNA is concerned, all three regions defined as divergent on the basis of sequence comparisons may be submitted to processing events (Fig. 4). However, maturation in all three divergent regions of the same organism has not yet been reported.

In addition to the aforementioned examples, several other instances of central breaks and multiple cleavages of the lrRNA have been reported in lower eukaryotes (see Ishikawa 1977 for a review), in plant chloroplasts (Rozier et al. 1979), and possibly, in some vertebrates (Leipoldt and Engel 1983). In these cases, it would be interesting to know if the cleavages map to one (or more) of the three divergent regions D7a, b, and c.

It has recently been noted that the G+C content of the two main divergent regions D2 and D8 of the lrRNA (located in eukaryotic domains II and V, respectively) appear to be closely related to each other during evolution (Michot and Bachellerie 1987). This seems to be true also for the D7a and D7b regions from organisms where no processing event occurs in the central domain (for instance, 71.5% and 71% in man, 50% and 51.2% in *Caenorhabditis elegans* for the G+C content of D7a and D7b, respectively). However, in organisms where a processing event is affecting either D7a or D7b, there

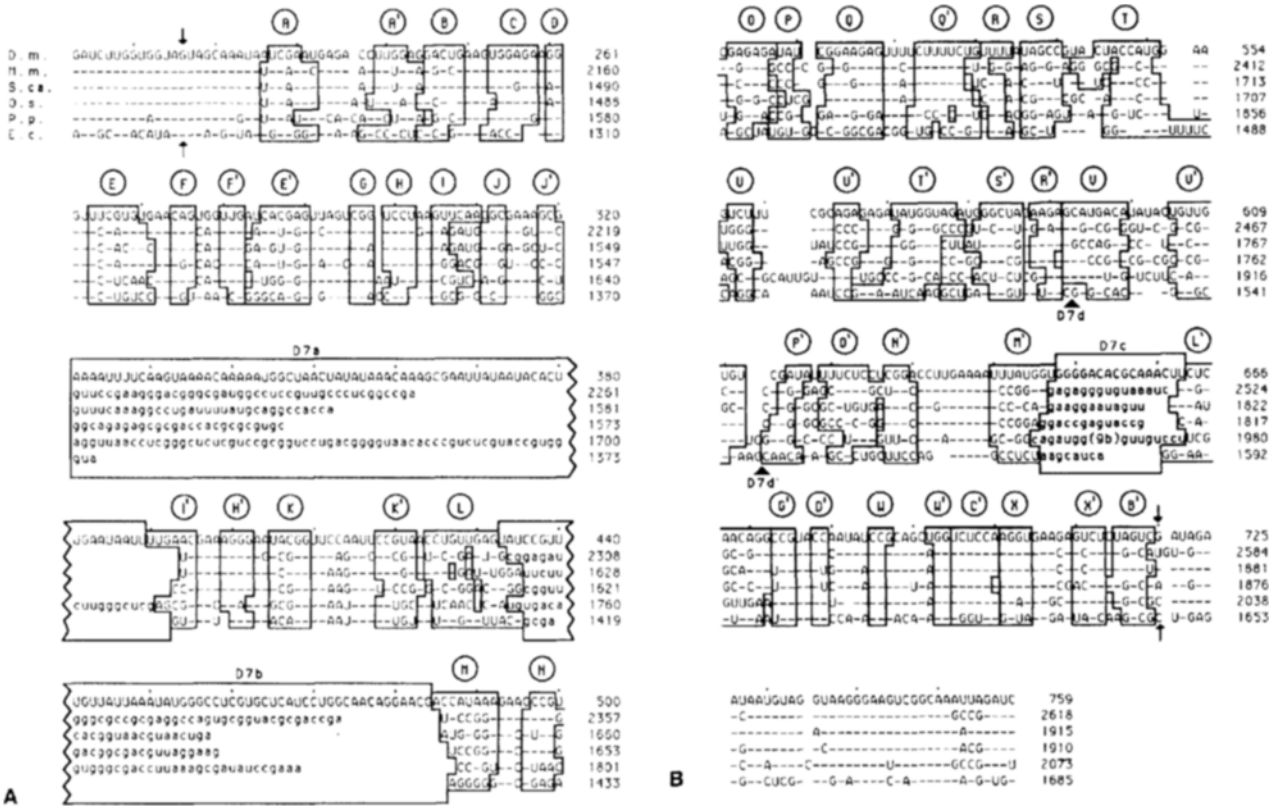

Fig. 3. Sequence alignment of the central domain regions of 23–28S rRNAs from *D. melanogaster* (D.m.), mouse (M.m.), yeast (S.ca.), rice (O.s.), *P. polycephalum* (P.p.), and *E. coli* (E.c.). The sequences were aligned on the basis of conserved primary structure and secondary structural features (shown in boxes). Identical nucleotides between sequences are indicated by hyphens, and gaps introduced to maximize similarity are indicated by blanks. Numbering of the sequences is that of the corresponding central regions in the different 23–28S rRNAs (see Table 1), except for *Drosophila* in which the numbering corresponds to the partial 26S sequence shown in Fig. 2. The vertical arrows near the beginning and the end of the sequences point, respectively, to the 5′ and 3′ limits of the central domain (listed in Table 1). Shaded boxes correspond to the three divergent regions, D7a, b, and c, for which the corresponding nucleotides in the different sequences have been arbitrarily aligned without gaps. The "(9b)" in the D7c region of *P. polycephalum* indicates a sequence tract (GUGAAACGU) longer than the corresponding *Drosophila* sequence. The D7d and D7d′ indicated by black arrowheads below the sequences in Fig. 3B correspond to the position of divergent regions found in chloroplast sequences (see text). Nucleotides in open boxes correspond to paired tracts (A/A′, B/B′, . . .) proven by comparative evidence (see Fig. 5 for the corresponding *Drosophila* secondary structure model).

is a marked difference in their respective G+C content, and the region that is removed always has the lowest G+C content: In *Drosophila* the D7a region (removed) is only 21.7% G+C, whereas the D7b region (not removed) is 46.2%. In *C. fasciculata* and *T. brucei* the values for the D7a region (not removed) are 52.7% and 46% G+C, respectively, whereas those for the D7b region (removed) are 43.5% and 31.6% G+C.

A Unified Secondary Structure Model for the Central Domain of Prokaryotic and Eukaryotic lrRNA

The recent determination of several lrRNA sequences from bacteria, mitochondria, chloroplasts, and eukaryotic cytoplasm has led to secondary structure models for these molecules (Branlant et al. 1981; Glotz et al. 1981; Veldman et al. 1981; Clark et al. 1984; Hadjiolov et al. 1984; Michot et al. 1984; Noller 1984; Brimacombe and Stiege 1985; Jarsch and Böck 1985; Manna and Brennicke 1985; Ellis et al. 1986; Gorski et al. 1987; Lang et al. 1987; Leffers et al. 1987). Large portions of these models are now in good agreement, but the secondary structure of some specific regions of the molecule is still a matter of debate. In an effort to clarify the secondary structure of the central domain of the lrRNA molecule, we have undertaken an extensive comparative approach, analyzing data from the *Drosophila* sequence reported herein and a total of 32 lrRNA sequences from archaebacteria, eubacteria, chloroplasts, mitochondria, and eukaryotic cytoplasm (Table 1). Our results have led to a general model for this region that presents several interesting features:

A common structural core is present in all species (except in some mitochondria), which concerns

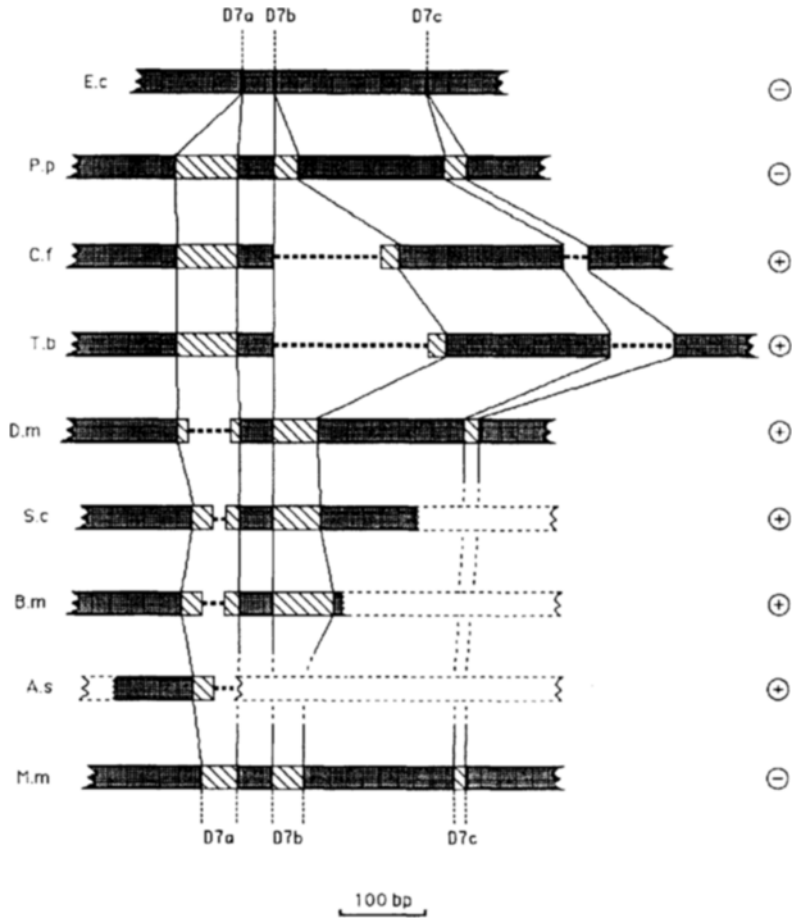

Fig. 4. Diagrammatic representation of the central domain from different lrRNAs. Abbreviations are as in Table 1. Shaded areas correspond to regions of the common structural core. Cross-hatched regions correspond to divergent regions present in the mature rRNAs. Interrupted bold lines correspond to the divergent regions that in some organisms are eliminated from the mature rRNAs (fragmentation spacers). Encircled − and + signs on the right part of the figure refer, respectively, to the unfragmented, or fragmented, character of the lrRNA molecule in the corresponding species. Because only partial sequences were available for *Sciara coprophila*, *Bombyx mori*, and *Artemia salina*, the sequence of the central domain that is not determined is represented by dotted lines.

the parts of the sequences that are not classified as divergent according to the alignment (Fig. 3).

This common core contains several structural elements that have already been proposed in previous models of either prokaryotic or eukaryotic lrRNA; these are shown here to be universal.

The existence of a new helix (D in Figs. 3 and 5) is proposed, which is supported by the present comparative analysis.

For the first time, a comparative detailed folding pattern of the three divergent regions is presented. Despite complete sequence divergence, some structural features are shared by several species and therefore define a subset of common structural core in these regions.

In the following two sections, we present and discuss separately the conserved and divergent regions of this model.

Secondary Structure of the Conserved Regions

The conserved region of the central domain of the lrRNA is composed of 24 helices, named A–W in the *Drosophila* model of Fig. 5; an equivalent for each of them can be found in all prokaryotic, chloroplastic, eukaryotic, and in some mitochondrial sequences tested. These helices are indicated in Fig. 3, in an alignment of six selected sequences from different phylogenetic origins.

According to this model, the central domain is closed by a T-shaped structure composed of helices B, C, and X, and is preceded by the short hairpin A. The existence of these helices has been proposed in previous studies and is confirmed here.

This is the first report, however, of the short helix D. It may be 2–3 bp long and, in the majority of cases, consists of a 5′-GG . . . CC-3′ pairing. Two lines of evidence support the existence of helix D in all species: First, in an extensive search on many sequences (see Table 1), we found no example where this helix cannot be formed. Second and more important, five sequences show CBCs in this helix, changing the GG . . . CC pairing to an AG . . . CU in two cases (S.ca. and O.s.; see Fig. 3), and to a GG . . . CU in the other three (C.f., T.b., and P.p. mt; data not shown). On the basis of this strong comparative evidence, we propose that helix D be incorporated in the common structural core of the central domain.

Helices E, F, and G were present in the human model (Gorski et al. 1987) and those of several bacteria (Noller 1984; Jarsch and Böck 1985; Leffers et al. 1987). We show here that they are also present

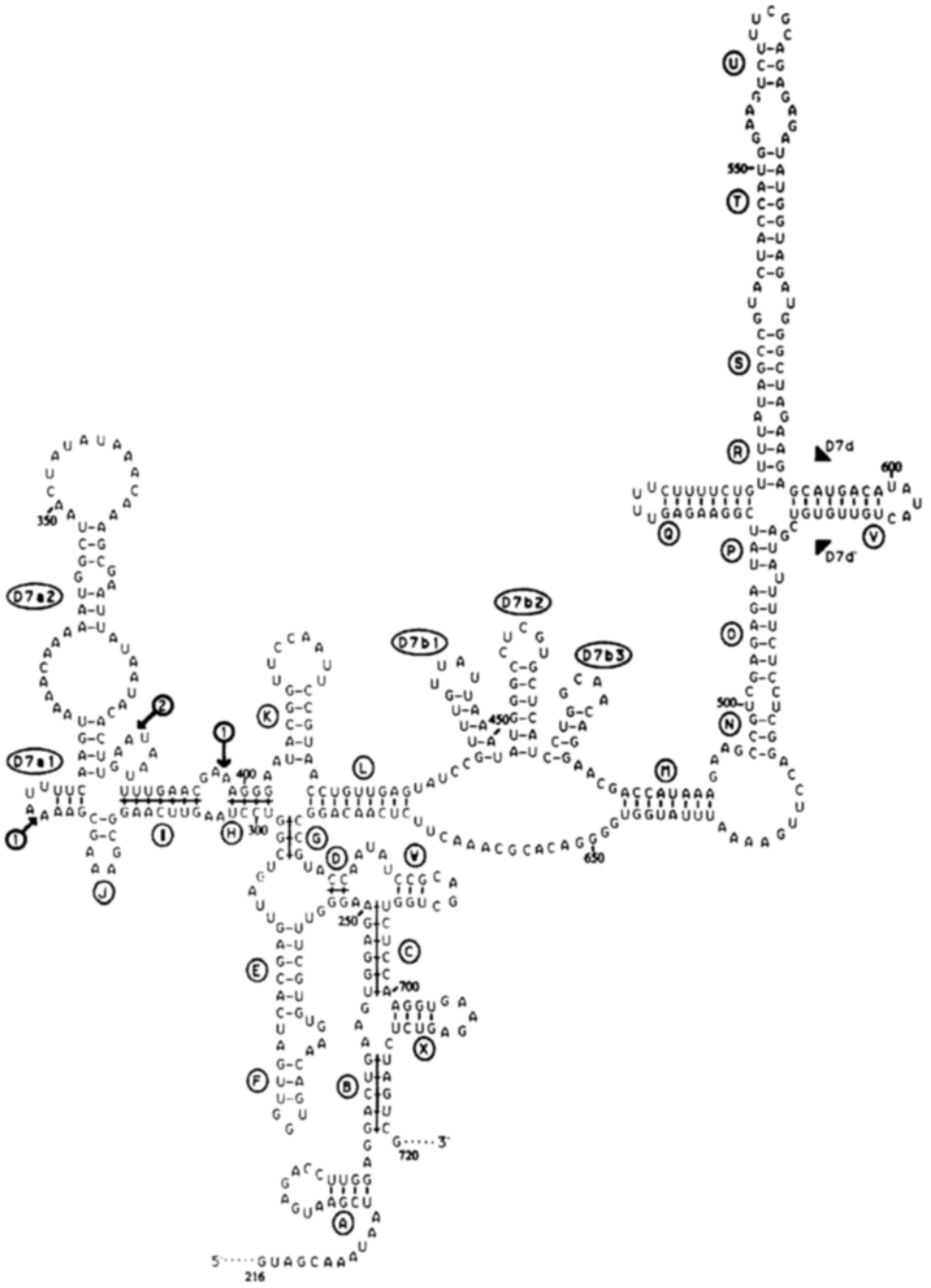

Fig. 5. Secondary structure model of the central domain from *Drosophila* 26S rRNA. The portion of the sequence represented here corresponds to the central region comprised between vertical arrows in Figs. 2 and 3. Numbering of the sequence is as in Fig. 2, with a stroke at every 50th nucleotide (250, 300, . . . , 700). Each helix of the conserved structural core is indicated by an encircled capital letter as in Fig. 3. The five helices of the divergent regions are indicated by ovals (D7a1, D7a2, . . .). Black arrowheads on the right part of the figure correspond to the position of divergent regions found in chloroplast sequences (see text and Fig. 3). The 3′ and 5′ limits of 26S α and β, as determined by de Lanversin and Jacq (1983) are indicated by an encircled 1 with arrow on the left part of the figure, whereas the encircled 2 indicates an alternative proposal for the 5′ extremity of 26S β (Ware et al. 1985). Bold lines crossing the base pairs of helices B, C, D, G, H, and I indicate the regions that are supposed to be involved in the association of the α and β moieties of the 26S molecule.

in all eukaryotes tested. The fact that numerous CBCs can be found in different organisms unambiguously establishes the existence of these three helices (see Figs. 3 and 5).

The presence of the short helix W (see Figs. 3 and 5) was suggested in some prokaryotic models (Noller 1984; Jarsch and Böck 1985). Although this structure can be formed in both organelles and eukaryotic organisms, it was sometimes not retained because it is only 2 bp long and nearly always consists of a CC . . . GG pairing. However, we found that in the sequence of the mitochondrial lrRNA from *Paramecium aurelia* (Seilhamer et al. 1984), *Aspergillus nidulans* (Kochel and Kuntzel 1982), and *Xenopus laevis* (Roe et al. 1985), a CU . . . AG pairing can be formed at this position, which is in favor of the existence of this structure. It has also to be noted that in all but two cases, the pairing of helix W can be extended toward the top of the loop (Fig. 5), either by a classical G–U pair or by a G–A pair, a type of association for which evidence is accumulating (reviewed in Noller 1984).

In our model we propose three other helices as universal (H, I, and J) that are located in a region of the molecule where 37 consecutive nucleotides were left impaired in Noller's (1984) model of *E.*

coli 23S rRNA. Previous eukaryotic models are in nearly complete disagreement in this region, probably because of the presence of the divergent region D7a in these organisms. A pairing scheme for this region was proposed for *E. coli* (helix D in Vester and Garrett 1984) and the archaebacteria *Desulfurococcus mobilis* (helix 52 in Leffers et al. 1987), in which a long helix is interrupted by one or two small bulged regions. In a precise alignment of different eukaryotic and prokaryotic sequences (Fig. 3), we found that, despite the presence of the D7a region in eukaryotes, the model of prokaryotic helix D or 52 can be universally extended. Comparative evidence based on the presence of CBCs and of conserved nucleotides at invariant positions (Fig. 6) strengthens the validity of helices H, I, and J that may now be incorporated in the universal common core of the central domain. As a consequence, we propose that the D7a region is absent in *E. coli*, contrary to what was previously stated (Hassouna et al. 1984).

Helix K was proposed in several earlier models and is confirmed here. The long-range pairings constituted by helices L and M delimit a region comprising two divergent domains (D7b and D7c) that are facing each other (Figs. 5 and 7). These helices are confirmed by previous and present (Fig. 3) comparative evidence. However, the length of helix M was found to be very variable, ranging from only 3 bp in *Z. mays* chloroplasts to as much as 12 bp in *Dictyostelium discoideum* (Fig. 7), suggesting that this helix may be part of the divergent regions D7b and D7c.

The remaining part of the model (helices N–V) has been arranged in a long cruciform structure in which the cross bar is constituted by helices Q and V (Fig. 5). This arrangement was originally proposed in some prokaryotic (Noller et al. 1981; Noller 1984) and eukaryotic (Michot et al. 1984; Ellis et al. 1986) models. Alternative models for this region were proposed for *E. coli* (Brimacombe and Stiege 1985) and several eukaryotic organisms (Veldman et al. 1981; Clark et al. 1984; Hadjiolov et al. 1984). Our results (Figs. 3 and 5, and other data not shown here) do not support the latter models; rather, the former models are to be preferred in this region of the central domain.

In summary, for the secondary structure of the entire central domain, we have found that all helices described in the Santa Cruz/Urbana model of *E. coli* 23S rRNA (Noller 1984) can be formed in nearly all species tested. After this analysis was entirely completed, an lrRNA secondary structure compilation was published (Gutell and Fox 1988). Some features of the common core that were not present in Noller's model of the central domain have now been incorporated in this new version, such as helices H, I, and J in eubacteria, archaebacteria, chloroplasts, and mitochondria. However, it has to be noted that helix D is never proposed in any organism, and that helix J is never proposed in any eukaryotic organism. In the present work, we have presented comparative evidence (now documented in all species with the exception of some mitochondrial sequences) that Noller's model as revised by Gutell and Fox can be improved by the universal occurrence of helices D, H, I, and J.

Secondary Structure of the Divergent Regions

Sequence alignments such as the one shown in Fig. 3 have allowed us to precisely map the limits of the three regions D7a, b, and c that are highly divergent between species. Moreover, we present a revised model for the common structural core that allows us to localize precisely the position of the divergent regions and to propose a secondary structure for them. Two important features emerge from this analysis: First, in all cases examined (see below and Figs. 6 and 7), divergent regions can be folded into independent structures that do not affect the folding of the conserved structural common core. Second, in all species, these divergent structures are always "inserted" at the same three specific sites of the model of the conserved core.

The first divergent region, D7a, is absent from eubacteria, archaebacteria, and chloroplasts and seems therefore to be specific for cytoplasmic rRNAs from eukaryotes, in which it is always present. In the latter case, the D7a structure is always localized between helices J and I, immediately after the 3′ end of helix J, and can be folded in two consecutive helices (Fig. 6). Helix D7a1 is always small and is formed of a stem of 2–4 bp closed by a terminal loop of three to six nucleotides. The second helix D7a2 has a very variable size, but never forms a branched structure: In rice, it forms a 3-bp-long stem closed by a loop of six nucleotides, whereas in *Physarum polycephalum* it forms a very long helix of 60 nucleotides interrupted by two internal loops (Fig. 6). In a previous comparative study of the central region from four eukaryotic rRNAs (Michot et al. 1984), these helices were correctly predicted in two cases. In the recent compilation of Gutell and Fox (1988), containing nine eukaryotic cytoplasmic lrRNA sequences, no secondary structure model was proposed for the D7a region. We show here (Fig. 6) that the short helix D7a1 and the variable helix D7a2 are present in all eukaryotic sequences tested. Although the sequence of the D7a2 helix is highly variable, supplementary evidence for its existence can be obtained from related organisms where sequence similarity is still detectable in this region and where CBCs are indeed present, as is the case for vertebrates (Fig. 6).

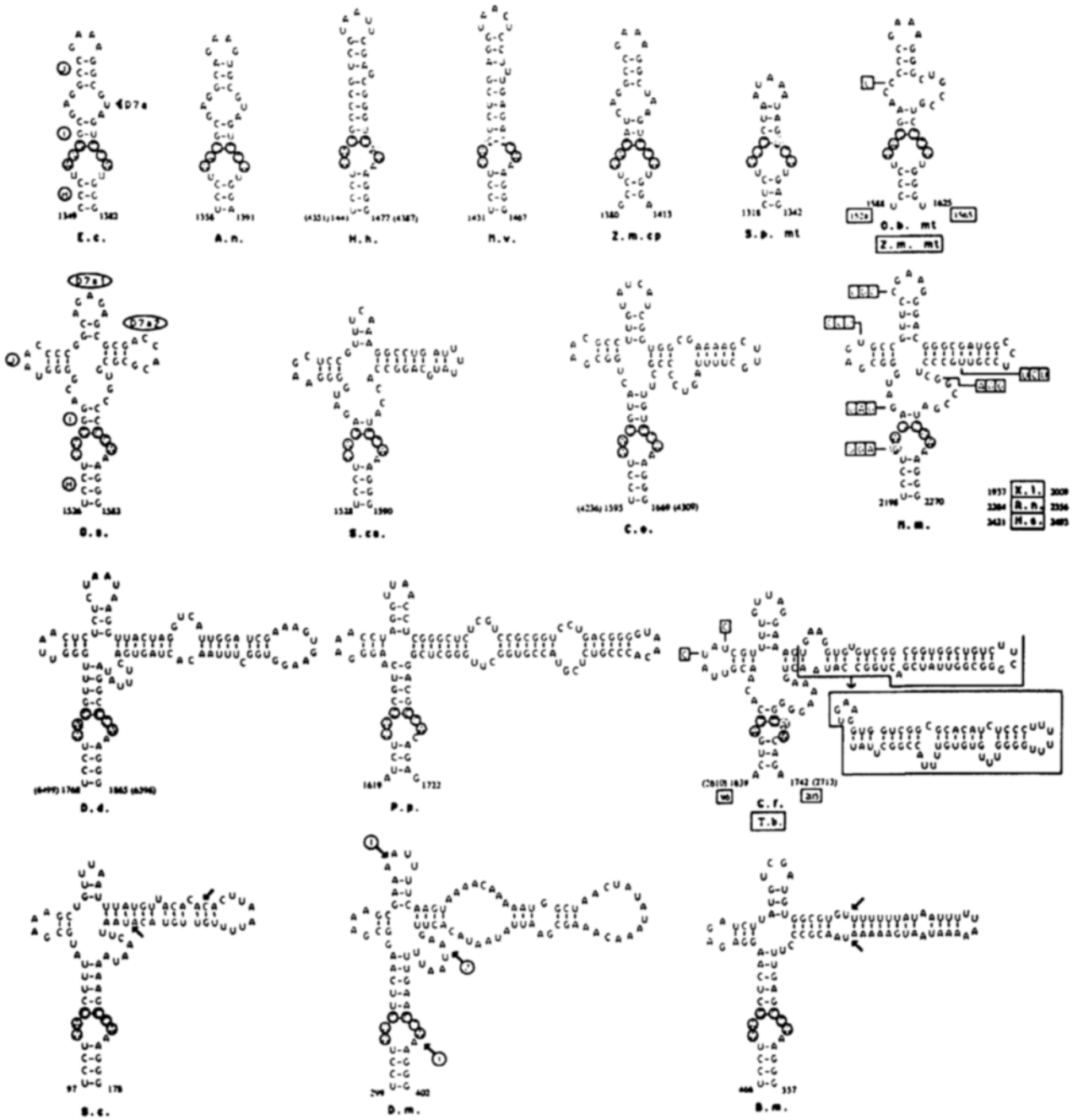

Fig. 6. Folding of the D7a domain and flanking regions in prokaryotes, organelles, and eukaryotes. Abbreviations of species and numbering conventions are as in Table 1. Sequences from prokaryotes and organelles are shown on the first row. For each species the H, I, and J helices of the conserved structural core (see Figs. 3 and 5) and the secondary structure of the D7a region (if present) are represented. Each model starts at the first nucleotide immediately following the 3′ end of helix G. The position of the divergent D7a region found in eukaryotes is indicated by a solid triangle on the *E. coli* model. In prokaryotes and organelles, the H, I, and J helices are drawn vertical, as in the *E. coli* (E.c.) model. In eukaryotes, H and I are drawn vertical, and J is drawn horizontal as in the *O. sativa* (O.s.) model. All eukaryotic D7a1 helices are drawn vertical (at the top of the model), and D7a2 helices are drawn horizontal (right part of the model), as is shown for the *O. sativa* (O.s.) model. The six nucleotides that are highly conserved in all sequences are circled, a dotted circle indicating a change relative to the consensus sequence. In the case where more than one species is represented on a secondary structure model (last models of the first, second, and third rows), the abbreviations, numbering, and nucleotide changes of the supplementary models are boxed. Nucleotide changes occurring in *X. laevis* (X.l.), *Rattus norvegicus* (R.n.), and human (H.s.) relative to the mouse (M.m.) sequence, are indicated in this order on the corresponding model. The boxed inset on the *C. fasciculata* (C.f.) model corresponds to the structure of the D7a2 helix in *T. brucei* (T.b.). The models of three insect sequences, in which part of the D7a region is eliminated by a processing event are represented on the last row. The positions of the breaks introduced in the RNA are indictaed by arrows. Conventions for arrows on the *D. melanogaster* (D.m.) model are as in Fig. 5.

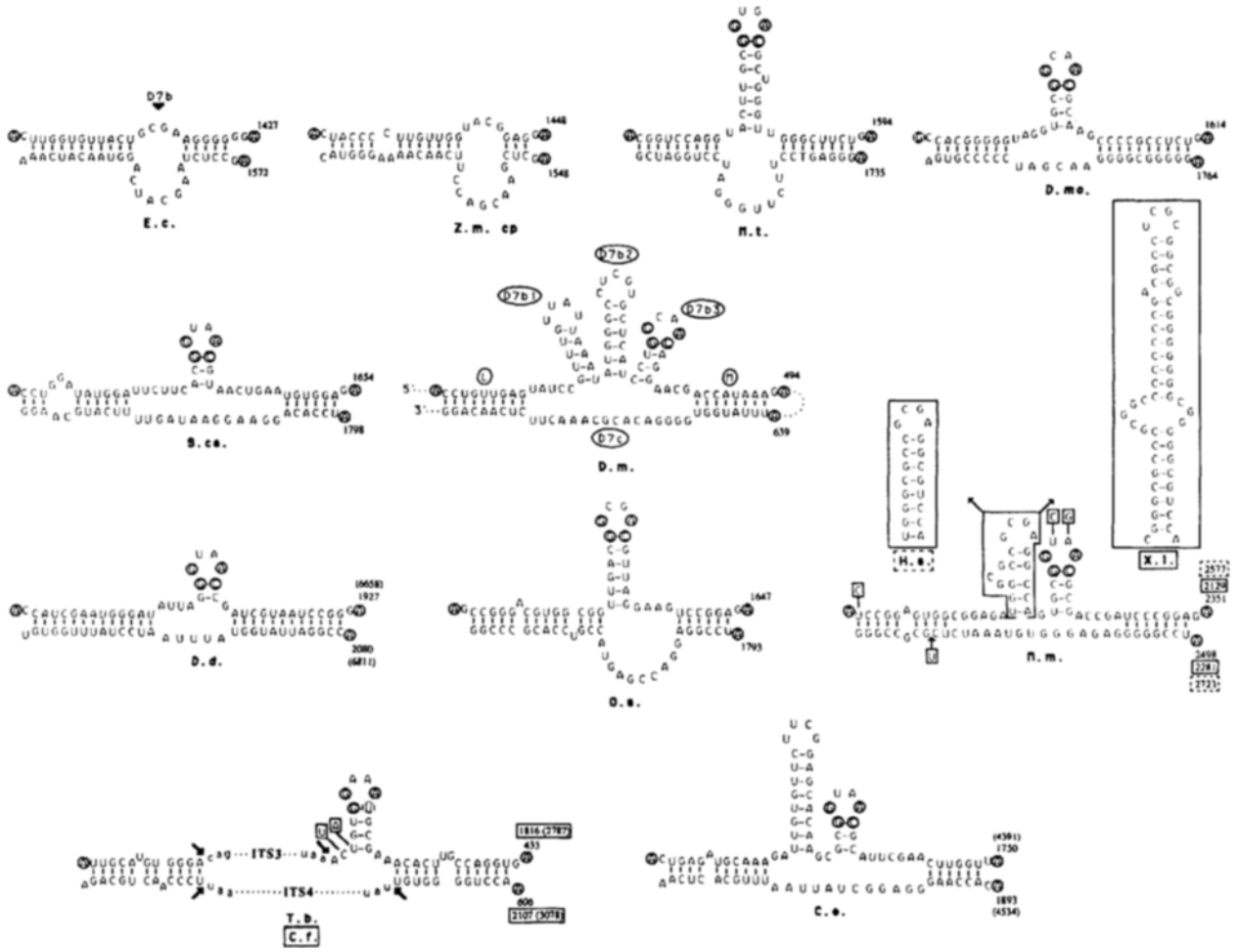

Fig. 7. Folding of the D7b and D7c domains and flanking regions in prokaryotes, organelles, and eukaryotes. For each sequence the L and M helices of the conserved structural core (see Figs. 3 and 5), the D7c region, and the different helices of the D7b region (if present), as is shown for the *D. melanogaster* (D.m.) model, are represented. The polarity of the sequence is indicated on this model, where dotted lines indicate parts of the sequence that have not been represented. The upper part of each model starts at a conserved A immediately preceding the 5′ end of helix M and stops at a conserved A following helix M (position 494 in *Drosophila*). The lower part of each model starts at a conserved A preceding the 3′ partner of helix M (position 639 in *Drosophila*) and stops at the position immediately preceding the 3′ partner of helix G. General conventions for abbreviation of species, sequence numbering, conserved nucleotides, and representation of more than one sequence on the same model are as in Fig. 6. The position of the divergent D7b region found in archaebacteria and eukaryotes is indicated by a solid triangle on the *E. coli* model. The two boxed insets on the mouse (M.m.) model correspond to the structure of the D7b2 helix in human (H.s.) and *X. laevis* (X.l.). The model of two trypanosomatid sequences in which part of the D7b and the entire D7c region (respectively, ITS3 and ITS4 in the model) are eliminated by processing events is represented at the bottom left of the figure. The positions of the breaks introduced in the RNA are indicated by arrows.

As already discussed, the D7a region is the target of the unique processing event occurring in the central domain of insect 26S rRNA. Depending on the organism being considered, a fragmentation spacer of variable size is removed from the D7a region during this maturation step (de Lanversin and Jacq 1983; Ware et al. 1985; Fujiwara and Ishikawa 1986). It is noteworthy that the D7a2 helix, which is always very A–U rich, is removed (in part or totally) in all three cases so far examined (Fig. 6).

The other two divergent regions D7b and D7c face each other on both sides of an internal loop located between helices L and M of the conserved structural core (Figs. 5 and 7). D7b is always longer than D7c (except in *E. coli* and in *Z. mays* chloroplasts). No convincing comparative evidence was found in favor of a pairing between D7b and D7c, therefore suggesting a separate folding pattern for each of them. Although in some species (e.g., *Drosophila*) a potential pairing is possible between the 5′ and the 3′ ends of D7c, it is not supported by comparative evidence. Consequently, the D7c sequence from all species was left single-stranded in the model of Fig. 7.

The D7b region is absent in eubacteria and chloroplasts; it is absent in some archaebacteria (e.g., *Halococcus morrhuae*), and present in others (e.g., *Methanococcus vannielii, D. mobilis,* and *Methanobacterium thermoautotrophicum*). It appears in all cytoplasmic ribosomes from eukaryotes where it can

form 1, 2, or 3 helical regions (D7b1, D7b2, and D7b3; see Fig. 7). The D7b3 helix is present in some archaebacteria and all eukaryotes, and is located on the 3′ side of the divergent region. It can be recognized by the conserved GGYRAC hexanucleotide sequence containing the terminal loop of four nucleotides (Fig. 7). In all multicellular animals, a second helix of variable length (D7b2) precedes helix D7b3. A third helix (D7b1), unique to *Drosophila,* is present on the 5′ side of the divergent domain (Fig. 7). The trypanosomatid organisms *C. fasciculata* and *T. brucei* represent an interesting example in which D7b and D7c regions are unusually long and are removed by processing events (see Fig. 4). The D7b3 helix can be formed in *C. fasciculata* (Spencer et al. 1987; Fig. 7) and in *T. brucei* (Fig. 7), and is present in the mature RNA. Although more comparative evidence is needed for these organisms, it seems that the remaining part of D7b, which is eliminated from the rRNA as a fragmentation spacer, may form a long and branched D7b2 helix (ITS3; Campbell et al. 1987), but no D7b1 helix. It is also intriguing that in these organisms, maturation cleavages occur in two regions facing each other in the secondary structure model of the rRNA (Campbell et al. 1987; Spencer et al. 1987; Fig. 7), leaving open the possibility that a unique mechanism is involved in this processing step.

The model proposed here for D7b is in general good agreement with that proposed by Gutell and Fox (1988). However, a proposal is made herein for the D7b region of *X. laevis, C. elegans, P. polycephalum, C. fasciculata,* and *T. brucei,* for which no model was proposed in that compilation. The D7d and D7d′ regions that, as far as we know, have been detected only in chloroplasts can be unambiguously placed before and after helix V, respectively (arrowheads in Figs. 3 and 5). These two regions can be folded in two independent small domains (data not shown).

Compatibility of the Model with Experimental Data

It has been known for some time that the region of *E. coli* 23S rRNA encompassing roughly helices C–J of the present model is the binding site for the L23 ribosomal protein; this RNA–protein complex is located at the A site of the peptidyl transferase center of the ribosome (Stöffler et al. 1971; Garrett et al. 1974; Branlant et al. 1975; Grant et al. 1979; Vester and Garrett 1984). More recently, it has been shown that the ribosomal protein L25 of yeast (the putative evolutionary counterpart of *E. coli* L23) binds to a yeast 26S rRNA fragment that is the structural equivalent of the L23 binding site. The yeast L25 protein is also able to bind to *E. coli* 23S rRNA at the L23 site (El-Baradi et al. 1985). Chemical modification, enzymatic cleavage, and RNA–RNA binding experiments indicate that the rRNA can form a very compact and resistant structure even in the absence of the L23/25 protein (Vester and Garrett 1984; El-Baradi et al. 1985). It was therefore proposed that the RNA–RNA interactions in this region may be appreciably more extensive than envisaged in the previous secondary structure models (El-Baradi et al. 1985). In our model, the part of the rRNA central domain that is supposed to interact with L23/25 has a higher percentage of paired residues and therefore, in agreement with the above proposal, exhibits a tighter structure than all previous models. It is perhaps surprising that in eukaryotes, a divergent region (D7a) is present within an evolutionarily and functionally conserved binding site. However, nucleotides from helices I and J in *E. coli* are probably not part of the L23 binding site, because they do not seem to exhibit an altered chemical reactivity in the presence of L23 (Vester and Garrett 1984). If prokaryotic L23 and eukaryotic L25 RNA binding sites are very similar (El-Baradi et al. 1985), it is possible that nucleotides from the D7a region (located between helices J and I in our eukaryotic model) are not protected by L23/25 and are not essential to the peptidyl transferase activity. The fact that the D7a region is eliminated from the RNA in several species is in agreement with this proposition.

The structure of psoralen-crosslinked 26S rRNA from *D. melanogaster* has been examined by electron microscopy (Wollenzien et al. 1978). In all 26S rRNA molecules, a characteristic hairpin in which the central break is located was reproducibly observed. Considering our secondary structure model shown in Fig. 5, we propose that a long hairpin could be formed if one assumes a co-stacking of helices B, C, D, G, L, M, N, O, P, R, S, T, and U, and that this hairpin corresponds potentially to the structure observable in the electron microscope.

The model of Fig. 5 also allows us to examine at the sequence level the molecular basis for the association between the two halves of *Drosophila* 26S rRNA. Heat-dissociation studies (Shine and Dalgarno 1973) have shown that *Drosophila* 26S rRNA is dissociated in two polynucleotide chains (α and β) at a low temperature (T′m = 46°C) and over a narrow range in 0.1 M NaCl, suggesting that a relatively short region of base pairing is involved in the α–β association. Previous models of prokaryotic and eukaryotic lrRNA agree on a separate folding of the different domains of the molecule (Michot et al. 1984; Noller 1984). This implies that the determinants for the association between the two halves of *Drosophila* 26S rRNA are essentially present in the central domain that contains the respective 3′ and 5′ extremities of 26S α and β. After the removal

from the precursor of the sequence of the central gap (roughly corresponding to the D7a2 domain of Fig. 5), it can be seen that the two parts of the RNA remain held together by helices B, C, D, G, H, and I only (Fig. 5), which represent a total of 27 bp. This value explains the low T'm of 26S rRNA and fits very well with that proposed by Wollenzien et al. (1978); on the basis of estimates of psoralen incorporation and crosslinking efficiencies, these authors concluded that the partial complementarity between the two rRNA halves may be as small as 25 bp.

Concluding Remarks

A minimum of 13 highly divergent regions was previously identified in a comparative analysis of complete lrRNA sequences from prokaryotes and eukaryotes (Hassouna et al. 1984) and this number is higher if chloroplastic and mitochondrial sequences are taken into account. In several species, some of these regions appear to be dispensable and represent examples of fragmentation spacers. In the central domain of lrRNA, the D7a region on one hand and the D7b/D7c region on the other are fragmentation spacers in several arthropods and in trypanosomatids, respectively. Other domains of eukaryotic lrRNA also contain divergent regions that can behave as fragmentation spacers in some organisms (Clark 1987; Spencer et al. 1987). Despite the fact that only a limited number of cases have been examined so far, it seems clear that fragmentation spacers always map to divergent regions. However it is difficult to understand why a given divergent region will be removed by processing in some species and will be present in the mature rRNA from other species. A search for primary sequence or secondary structure signals in the spacer or near the coding/spacer boundaries did not reveal clear consensus features; it only showed that fragmentation spacers from different species have an A+U percentage that is never lower than 55% in all cases so far examined. It could be that the elimination of fragmentation spacers reflects either a greater accessibility to nucleases, or the existence of nucleases with low specificity in some organisms. Another intriguing possibility is that some sequences may have intrinsic "auto-processing" properties, a proposal that could be experimentally tested.

The question of a function for divergent regions is still open to debate. Two opposite alternatives are: these sequences have no function at all but are tolerated because they do not disrupt ribosome function (Clark 1987); or they are involved in diversified control mechanisms that could be specific for some phylogenetic branches (Michot and Bachellerie 1987). In this respect, it would be interesting to know if there is any relationship between the presence of "extra" sequences in eukaryotic rRNAs and the fact that eukaryotic ribosomal subunits contain more ribosomal proteins than their prokaryotic counterparts. That fragmentation spacers are absent from mature ribosomes in eukaryotes strongly suggests that this particular class of divergent sequences is not involved in protein synthesis. Nevertheless, a role for fragmentation spacer sequences in ribosome assembly or transport cannot be completely excluded.

It has been suggested (Clark 1987) that the divergent regions in today's rRNAs might be the remnants of base-paired regions that were the site of fusion of the different RNA segments constituting the primitive ribosome. Several lines of evidence support this idea (Clark 1987), but it is still very difficult to determine whether the elimination of a fragmentation spacer (corresponding to these primitive links) is also a primitive event or, rather, a more recently acquired species-specific feature. An answer to this question clearly requires the determination of more sequences and secondary structure models for the fragmentation spacer regions from organisms where the occurrence of genuine rRNA breaks has been proven.

Acknowledgments. We are grateful to Dr. R. Griffin-Shea and Pr. R. Rosset for critical reading of the manuscript and helpful suggestions. This work was supported by CNRS.

References

Bellon B (1988) Apple Macintosh programs for nucleic and protein sequence analyses. Nucleic Acids Res 16:1837–1846

Branlant C, Krol A, Sriwidada J, Ebel JP, Sloof P, Garret R (1975) A partial localisation of the binding sites of the 50S subunit proteins L1, L20 and L23 on 23S ribosomal RNA of *Escherichia coli*. FEBS Lett 25:195–201

Branlant C, Krol A, Machatt MA, Pouyet J, Ebel JP, Edwards K, Kössel H (1981) Primary and secondary structures of *Escherichia coli* MRE 600 23S ribosomal RNA. Comparison with models of secondary structure for maize chloroplast 23S rRNA and for large portions of mouse and human 16S mitochondrial rRNAs. Nucleic Acids Res 9:4303–4324

Brimacombe R, Stiege W (1985) Structure and function of ribosomal RNA. Biochem J 229:1–17

Brosius J, Dull TJ, Noller HF (1980) Complete nucleotide sequence of a 23S ribosomal RNA gene from *Escherichia coli*. Proc Natl Acad Sci USA 77:201–204

Campbell DA, Kubo K, Clark CG, Boothroyd JC (1987) Precise identification of cleavage sites involved in the unusual processing of trypanosome ribosomal RNA. J Mol Biol 196:113–124

Chan YL, Olvera J, Wool IG (1983) The structure of rat 28S ribosomal ribonucleic acid inferred from the sequence of nucleotides in a gene. Nucleic Acids Res 11:7819–7831

Clark CG (1987) On the evolution of ribosomal RNA. J Mol Evol 25:343–350

Clark CG, Tague BW, Ware VC, Gerbi SA (1984) *Xenopus laevis* 28S ribosomal RNA: a secondary structure model and its evolutionary and functional implications. Nucleic Acids Res 12:6197–6220

Curgy JJ (1985) The mitoribosomes. Biol Cell 54:1–38
Dale RMK, Mendu N, Ginsburg H, Kridl JC (1984) Sequence analysis of the maize mitochondrial 26S ribosomal RNA gene and flanking regions. Plasmid 11:141–150
Dawid IB, Wellauer PK, Long EO (1978) Ribosomal DNA in *Drosophila melanogaster* I. Isolation and characterization of cloned fragments. J Mol Biol 126:749–768
de Lanversin G, Jacq B (1983) Séquence de la région de la coupure centrale du précurseur de l'ARN ribosomique 26S de Drosophile. C R Acad Sci Ser III 296:1041–1044
de Lanversin G, Pillay DTN, Jacq B (1987) Sequence studies on the soybean chloroplast 16S–23S rDNA spacer region. Comparison with other angiosperm sequences and proposal of a generalized RNA secondary structure model for the intergenic regions. Plant Mol Biol 10:65–82
Douglas SE, Doolittle WF (1984) Complete nucleotide sequence of the 23S rRNA gene of the cyanobacterium, *Anacystis nidulans*. Nucleic Acids Res 12:3373–3386
Edwards K, Kössel H (1981) The rRNA operon from *Zea mays* chloroplasts: nucleotide sequence of 23S rDNA and its homology with *E. coli* 23S rDNA. Nucleic Acids Res 9:2853–2869
El-Baradi TTAL, Raué HA, de Regt VCHF, Verbree EC, Planta RJ (1985) Yeast ribosomal protein L25 binds to an evolutionary conserved site on yeast 26S and *E. coli* 23S rRNA. EMBO J 4:2101–2107
Ellis RE, Sulston JE, Coulson AR (1986) The rDNA of *C. elegans*: sequence and structure. Nucleic Acids Res 14:2345–2364
Erdmann VA, Wolters J (1986) Collection of published 5S, 5.8S and 4.5S ribosomal RNA sequences. Nucleic Acids Res 14: r1–r59
Fujiwara H, Ishikawa H (1986) Molecular mechanism of introduction of the hidden break into the 28S rRNA of insects: implication based on structural studies. Nucleic Acids Res 14:6393–6401
Garrett RA, Müller S, Spierer P, Zimmermann RA (1974) Binding of 50S ribosomal subunit proteins to 23S RNA of *Escherichia coli*. J Mol Biol 88:553–557
Georgiev OI, Nikolaev N, Hadjiolov AA, Skryabin KG, Zakharyev VM, Bayev AA (1981) The structure of the yeast ribosomal RNA genes. 4. Complete sequence of the 25S rRNA gene from *Saccharomyces cerevisiae*. Nucleic Acids Res 9: 6953–6958
Gerbi SA (1976) Fine structure of ribosomal RNA I. Conservation of homologous regions within ribosomal RNA of eukaryotes. J Mol Biol 106:791–816
Glotz C, Zwieb C, Brimacombe R, Edwards K, Kössel H (1981) Secondary structure of the large subunit ribosomal RNA from *Escherichia coli, Zea mays* chloroplast, and human and mouse mitochondrial ribosomes. Nucleic Acids Res 9:3287–3306
Gonzalez IL, Gorski JL, Campen TJ, Dorney DJ, Erickson JM, Sylvester JE, Schmickel RD (1985) Variation among human 28S ribosomal RNA genes. Proc Natl Acad Sci USA 82:7666–7670
Gorski JL, Gonzalez IL, Schmickel RD (1987) The secondary structure of human 28S rRNA: the structure and evolution of a mosaic rRNA gene. J Mol Evol 24:236–251
Grant PG, Strycharz WA, Jaynes EN, Cooperman BS (1979) Antibiotic effects on the photoinduced affinity labeling of *Escherichia coli* ribosomes by puromycin. Biochemistry 18: 2149–2154
Green CJ, Stewart GC, Hollis MA, Vold BS, Bott KF (1985) Nucleotide sequence of the *Bacillus subtilis* ribosomal RNA operon, rrnB. Gene 37:261–266
Gutell RR, Fox GE (1988) A compilation of large subunit RNA sequences presented in a structural format. Nucleic Acids Res 16:r175–r269
Gutell RR, Weiser B, Woese CR, Noller HF (1985) Comparative anatomy of 16S-like ribosomal RNA. Prog Nucleic Acid Res Mol Biol 32:155–216
Hadjiolov AA, Georgiev OI, Nosikov VV, Yavachev LP (1984) Primary and secondary structure of rat 28S ribosomal RNA. Nucleic Acids Res 12:3677–3693
Hassouna N, Michot B, Bachellerie JP (1984) The complete nucleotide sequence of mouse 28S rRNA gene. Implications for the process of size increase of the large subunit rRNA in higher eukaryotes. Nucleic Acids Res 12:3563–3583
Huysmans E, de Wachter R (1986) Compilation of small ribosomal subunit RNA sequences. Nucleic Acids Res 14:r73–r118
Ishikawa H (1977) Evolution of ribosomal RNA. Comp Biochem Physiol 58B:1–7
Jarsch M, Böck A (1985) Sequence of the 23S rRNA gene from the archaebacterium *Methanococcus vannielii*: evolutionary and functional implications. Mol Gen Genet 200:305–312
Jordan BR (1975) Demonstration of intact 26S ribosomal RNA molecules in *Drosophila* cells. J Mol Biol 98:277–280
Jordan BR, Jourdan R, Jacq B (1976) Late steps in the maturation of *Drosophila* 26S ribosomal RNA: generation of 5.8S and 2S RNAs by cleavages occurring in the cytoplasm. J Mol Biol 101:85–105
Jordan BR, Latil-Damotte M, Jourdan R (1980) Sequence of a 3'-terminal portion of *Drosophila melanogaster* 18S rRNA and of the adjoining spacer. Comparison with corresponding prokaryotic and eukaryotic sequences. FEBS Lett 117:227–231
Kjems J, Leffers H, Garrett RA, Wich G, Leinfelder W, Böck A (1987) Gene organization, transcription signals and processing of the single ribosomal RNA operon of the archaebacterium *Thermoproteus tenax*. Nucleic Acids Res 15:4821–4835
Kochel HG, Kuntzel H (1982) Mitochondrial L-rRNA from *Aspergillus nidulans*: potential secondary structure and evolution. Nucleic Acids Res 10:4795–4801
Lang BF, Cedergren R, Gray MW (1987) The mitochondrial genome of the fission yeast, *Schizosaccharomyces pombe*. Sequence of the large-subunit ribosomal RNA gene, comparison of potential secondary structure in fungal mitochondrial large-subunit rRNAs and evolutionary considerations. Eur J Biochem 169:527–537
Leffers H, Kjems J, Ostergaard L, Larsen N, Garrett RA (1987) Evolutionary relationships amongst archaebacteria. A comparative study of 23S ribosomal RNAs of a sulphur-dependent extreme thermophile, an extreme halophile and a thermophilic methanogen. J Mol Biol 195:43–61
Leipoldt M, Engel W (1983) Hidden breaks in ribosomal RNA of phylogenetically tetraploïd fish and their possible role in the diploïdization process. Biochem Genet 21:819–841
Malthiery B, Bellon B, Giorgi D, Jacq B (1984) Apple II Pascal programs for molecular biologists. Nucleic Acids Res 12:569–579
Mankin AS, Kagramanova VK (1986) Complete nucleotide sequence of the single ribosomal RNA operon of *Halobacterium halobium*: secondary structure of the archaebacterial 23S rRNA. Mol Gen Genet 202:152–161
Manna E, Brennicke A (1985) Primary and secondary structure of 26S ribosomal RNA of *Oenothera* mitochondria. Curr Genet 9:505–515
Maxam AM, Gilbert W (1980) Sequencing end-labeled DNA with base-specific chemical cleavages. Methods Enzymol 65: 499–560
Michot B, Bachellerie JP (1987) Comparisons of large subunit rRNAs reveal some eukaryote-specific elements of secondary structure. Biochimie 69:11–23
Michot B, Hassouna N, Bachellerie JP (1984) Secondary structure of mouse 28S rRNA and general model for the folding

of the large rRNA in eukaryotes. Nucleic Acids Res 12:4259–4279

Nelles L, Van Broeckhoven C, de Wachter R, Vandenberghe A (1984) Location of the hidden break in large subunit ribosomal RNA of *Artemia salina*. Naturwissenschaften 71:634–635

Noller HF (1984) Structure of ribosomal RNA. Annu Rev Biochem 53:119–162

Noller HF, Kop JA, Wheaton V, Brosius J, Gutell RR, Kopylov AM, Dohme F, Herr W, Stahl DA, Gupta R, Woese CR (1981) Secondary structure model for 23S ribosomal RNA. Nucleic Acids Res 9:6167–6189

Otsuka T, Nomiyama H, Yoshida H, Kukita T, Kuhara S, Sakaki Y (1983) Complete nucleotide sequence of the 26S rRNA gene of *Physarum polycephalum*: its significance in gene evolution. Proc Natl Acad Sci USA 80:3163–3167

Ozaki T, Hoshikawa Y, Iida Y, Iwabuchi M (1984) Sequence analysis of the transcribed and 5′ non-transcribed regions of the ribosomal RNA gene in *Dictyostelium discoideum*. Nucleic Acids Res 12:4171–4184

Pavlakis GN, Jordan BR, Wurst RM, Vournakis JN (1979) Sequence and secondary structure of *Drosophila melanogaster* 5.8S and 2S rRNAs and of the processing site between them. Nucleic Acids Res 7:2213–2238

Roe BA, Ma DP, Wilson RK, Wong JFH (1985) The complete nucleotide sequence of the *Xenopus laevis* mitochondrial genome. J Biol Chem 260:9759–9774

Rozier C, Rocipon M, Mache R (1979) Post-maturation of the plastid ribosomal RNA in the plant kingdom. J Mol Evol 13: 271–279

Seilhamer JJ, Gutell RR, Cummings DJ (1984) *Paramecium* mitochondrial genes II. Large subunit rRNA gene sequence and microevolution. J Biol Chem 259:5173–5181

Shine J, Dalgarno L (1973) Occurrence of heat-dissociable ribosomal RNA in insects: the presence of three polynucleotide chains in 26S RNA from cultured *Aedes aegypti* cells. J Mol Biol 75:57–72

Sinclair JH, Brown DD (1971) Retention of common nucleotide sequences in the ribosomal deoxyribonucleic acid of eukaryotes and some of their physical characteristics. Biochemistry 10:2761–2769

Spencer DF, Collings JC, Schnare MN, Gray MW (1987) Multiple spacer sequences in the nuclear large subunit ribosomal RNA gene of *Crithidia fasciculata*. EMBO J 6:1063–1071

Stöffler G, Daya L, Rak KH, Garrett RA (1971) Ribosomal proteins. XXVI. The number of specific protein binding sites on 16S and 23S RNA of *Escherichia coli*. J Mol Biol 62:411–414

Takaiwa F, Sugiura M (1982) The complete nucleotide sequence of a 23S rRNA gene from tobacco chloroplasts. Eur J Biochem 124:13–19

Takaiwa F, Oono K, Iida Y, Sugiura M (1985) The complete nucleotide sequence of a rice 25S rRNA gene. Gene 37:255–259

Tartof KD, Perry RP (1970) The 5S RNA genes of *Drosophila melanogaster*. J Mol Biol 51:171–183

Veldman GM, Klootwijk J, de Regt VCHF, Planta RJ, Branlant C, Krol A, Ebel JP (1981) The primary and secondary structure of yeast 26S rRNA. Nucleic Acids Res 9:6935–6952

Vester B, Garrett RA (1984) Structure of a protein L23–RNA complex located at the A-site domain of the ribosomal peptidyl transferase centre. J Mol Biol 179:431–452

Walker TA, Johnson KD, Olsen GJ, Peters MA, Pace NR (1982) Enzymatic and chemical structure mapping of mouse 28S ribosomal ribonucleic acid contacts in 5.8S ribosomal ribonucleic acid. Biochemistry 21:2320–2329

Ware VC, Tague BW, Clark CG, Gourse RL, Brand RC, Gerbi SA (1983) Sequence analysis of 28S ribosomal DNA from the amphibian *Xenopus laevis*. Nucleic Acids Res 11:7795–7817

Ware VC, Renkawitz R, Gerbi SA (1985) rRNA processing: removal of only nineteen bases at the gap between 28S α and 38S β rRNAs in *Sciara coprophila*. Nucleic Acids Res 13: 3581–3597

White TC, Rudenko G, Borst P (1986) Three small RNAs within the 10 kb trypanosome rRNA transcription unit are analogous to domain VII of other eukaryotic 28S rRNAs. Nucleic Acids Res 14:9471–9489

Wittmann HG (1985) Structure of ribosomes. In: Hardesty B, Kramer G (eds) Structure, function and genetics of ribosomes. Springer-Verlag, New York, pp 1–27

Wollenzien PL, Youvan DC, Hearst JE (1978) Structure of psoralen-crosslinked ribosomal RNA from *Drosophila melanogaster*. Proc Natl Acad Sci USA 75:1642–1646

Yamada T, Shimaji M (1987) An intron in the 23S rRNA gene of the *Chlorella* chloroplasts: complete nucleotide sequence of the 23S rRNA gene. Curr Genet 11:347–352

Received May 2, 1988/Revised August 1, 1988

Note added in proof. After this paper was accepted for publication, a study of the primary and secondary structures of *Drosophila melanogaster* rRNAs has been published (Tautz D, Hancock JM, Webb DA, Tautz C, Dover GA (1988) Mol Biol Evol 5:366–376; Hancock JM, Tautz D, Dover GA (1988) Mol Biol Evol 5:393–414). The primary structure of the central part of the 26S rRNA presented in this study is in general good agreement with the one reported herein, except for four differences: Nucleotide 356 in Fig. 2 is a C in our study and a G in Tautz's paper; three supplementary nucleotides appear in the latter study, an A, a G and a C, respectively located after nucleotides 213, 236, and 888 of Fig. 2. These differences could correspond to variant copies of the rRNA genes. At the secondary structure level, helices G and D7a are different in the two studies, and helices F, W, D, H, I, J, and D7bl of our model are absent in Hancock's proposal.