SEQUENCE OF THE 3'-TERMINAL PORTION OF *DROSOPHILA MELANOGASTER* 18 S rRNA AND OF THE ADJOINING SPACER

Comparison with corresponding prokaryotic and eukaryotic sequences

Bertrand R. JORDAN, Michèle LATIL-DAMOTTE and Roselyne JOURDAN

Centre de Biochimie et de Biologie Moléculaire, Centre National de la Recherche Scientifique, 31 chemin Joseph Aiguier, 13274 Marseille Cedex 2, France

Received 2 June 1980

1. Introduction

Ribosomal RNA sequences are conserved to a large extent among widely different eukaryotic species and even between eukaryotes and prokaryotes. This was first indicated by the finding that rRNA from one organism can hybridize to DNA from another [1]. A detailed study [2] showed that most of this cross-hybridization is due to particular regions of rDNA, ~10% of the 28 S sequence and 20% of the 18 S sequence. These regions appear to be very similar, with a divergence of a few % as judged from the melting temperature of the hybrids [2]. Recent sequence data has demonstrated that the 3'-terminal region of 18 S (16 S) rRNA in particular is very strongly conserved: the last 20 nucleotides in a variety of eukaryotes and in *Escherichia coli* are almost identical [3]; direct sequencing of a portion of a cloned *Bombyx mori* rDNA transcription unit [4] shows that the last 50 nucleotides of this molecule are nearly identical to those of *E. coli* 16 S RNA.

Our interest in the structure of the rDNA transcription unit of *Drosophila melanogaster* has led us to determine the sequence of a corresponding region in this gene which includes the 3'-terminal portion of the 18 S RNA coding region and the beginning of the transcribed spacer. This sequence makes it possible to determine the extent of homology between *D. melanogaster*, *B. mori* and *E. coli* sequences in this region. It also allows us to examine whether structures similar to those found at the boundaries of the 5.8 S–2 S region (A–T-rich 'processing sites') [5] are present at the beginning of the transcribed spacer.

2. Methods

Restriction enzymes were purchased from Boehringer (*Eco*RI, *Pst*I) or New England Biolabs (*Msp*I, *Hae*III). T4 nucleotide kinase was from Boehringer, bacterial alkaline phosphatase (used without further purification) and terminal transferase from Bethesda Research Labs. [α-^{32}P]ATP and [γ-^{32}P]ATP were purchased from Amersham.

Ribosomal DNA was prepared from recombinant plasmid cDm 238 [6] which contains slightly more than a complete transcription unit without insertion [5]. In some cases restriction enzyme digestion and end-labelling were performed on whole plasmid DNA which was subsequently recut with a second restriction enzyme and fractionated on a preparative agarose gel to obtain fragments labelled at one end only; in other cases a suitable fragment was first isolated on a preparative gel, then labelled, digested and fractionated again. Fragments were electroeluted in dialysis bags filled with electrophoresis buffer, filtered on glass fiber filters after addition of SDS to 0.2% and ethanol precipitated after addition of LiCl to 0.8 M. Labelling with T4 nucleotide kinase was performed as in [7]; labelling with terminal transferase was done in the presence of cobalt ion [8] and was followed by a 'chase' with unlabelled ATP, an important step to avoid doublet bands in the sequencing gel due to incomplete addition of A residues. Under these conditions even recessed 3'-ends such as those generated by *Eco*RI can be labelled to a sufficient extent to allow sequence determination.

Sequencing was performed essentially as in [7]: the G, A > C, C and C + T reactions were used with the following variations on published procedures: saturated NaCl (instead of 5 M NaCl) was used in the C reaction, and 0.83 M redistilled piperidine (instead of 0.5 M) was used in the cleavage step. Reaction products were run on 20%, 8% and 6% thin (0.3 mm) gels [9]; 150–200 bases could normally be read from these gels.

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

3. Results

Fig. 1 shows a map of plasmid cDm 238 [5] with a blow up of the region of interest. Sequencing was performed using fragments labelled at either the RI site or the *Msp*I site present in the 3′-end of the 18 S coding region (fig.1). Fragments were labelled at the 5′-end with polynucleotide kinase or at the 3′-end with terminal transferase so that the whole sequence could be determined on both strands. Fig. 2 shows a sequencing gel displaying the 3′-end of the 18 S coding region and the beginning of the transcribed spacer. The sequence obtained is shown in fig.3. It includes the 3′-terminal 227 nucleotides of 18 S and the first 50 of the adjoining transcribed spacer. The location of the 3′-terminus of 18 S has not been directly determined but is obvious because of the homology with the *B.mori* and *E.coli* sequences (below).

4. Discussion

A comparison of the *D.melanogaster* and the *B. mori* 18 S coding sequences shows their similarity (fig.3). Region 1–80 and 121–227 (numbered along the *Drosophila* sequence leftward from the 3′-end) are almost identical with only a few occasional differences; in contrast, the regions between 81–121 are very different and show little homology. It must be stressed that the extent of homology is surprisingly high. *B.mori* and *D.melanogaster* are evolutionarily

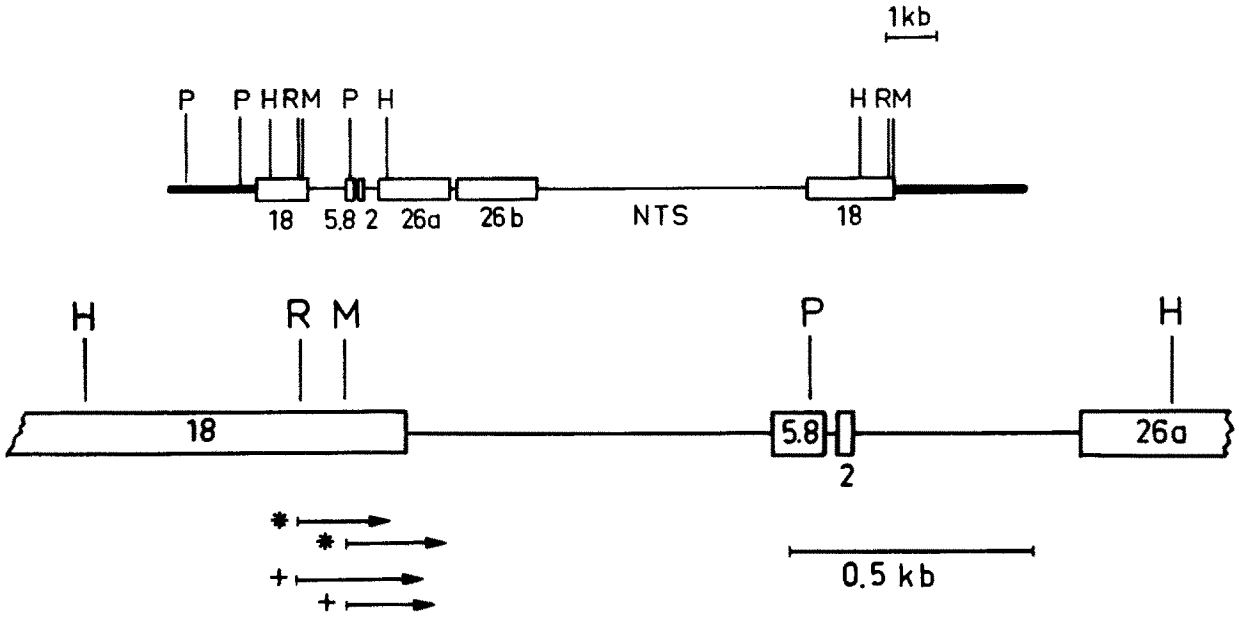


Fig.1. Restriction map of plasmid cDm 238 (top complete map, bottom blow up of transcribed spacer region). Ribosomal RNA coding sequences are boxed, vector ColE1 sequences are indicated by a thick line and spacer regions by a thin line. Cleavage sites are indicated for the following restriction enzymes: *Eco*RI (R), *Pst*I (P), *Msp*I (M), and *Hae*III (H). For these two last enzymes only the sites present in the region enlarged at the bottom of the figure are shown; other sites exist in the 26 S and non-transcribed spacer regions. The point at which the right hand 18 S region is linked to ColE1 DNA has been determined by DNA sequencing and lies 55 nucleotides after the *Msp*I site (data not shown). The position, type of labelling (*:5′, +:3′) and extent of sequence read on the fragments used for sequencing are indicated below the map.

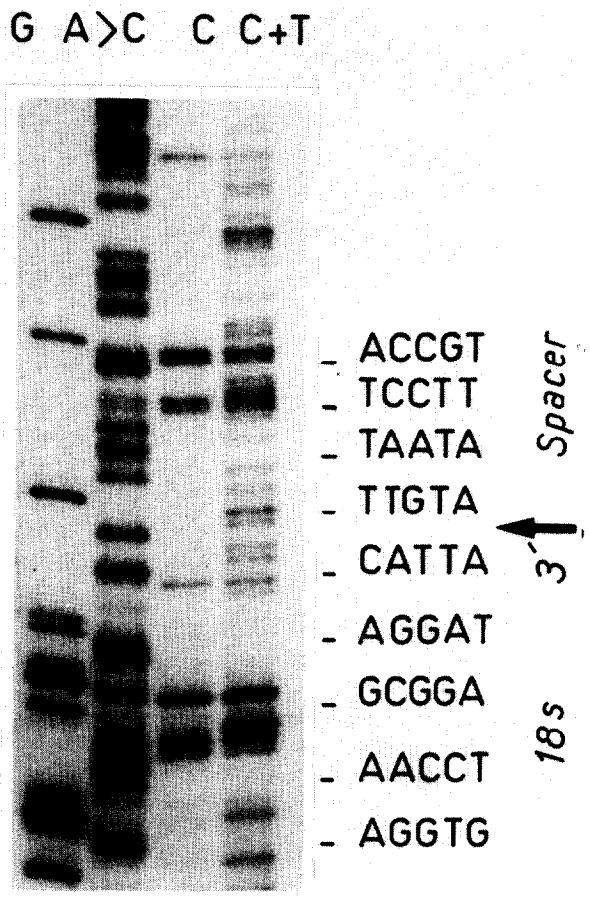


Fig.2. Portion of a 6% sequencing gel displaying the 3'-terminal sequence of 18 S RNA and the beginning of the transcribed spacer. The fragment was labelled with polynucleotide kinase at the *Msp*I site 100 nucleotides downstream from the first nucleotides shown.

quite distant and the restriction enzyme maps of their rDNAs are quite different: although both contain a unique RI site in the same position of the 18 S coding region, the *Sma*I and *Hin*dIII sites located in *D.melanogaster* 18 and 26 S RNA coding regions are absent or displaced in *B.mori*; the *Bam*HI sites present in the 26 S coding region in *B.mori* are absent in *D. melanogaster* [5,10,11]. Moreover, the 5.8 S RNA molecule in *B.mori* is not split [12] as in *D.melanogaster*, i.e., there is no 2 S RNA and no 5.8 S–2 S spacer in the rDNA. The extent of conservation of this terminal 18 S sequence is therefore highly significant and suggestive of a particular role. The stem and loop structure proposed to exist in the 3'-end of *B. mori* 18 S [4] is equally possible with the *D. melanogaster* molecule since the sequences are identical with the exception of an additional G in the loop. It is also interesting to note that the sequence comparison clearly indicates a number of additions between nucleotides 90–120; evolution in that region has definitely involved addition/deletion events along with substitutions.

When an alignment between the *E.coli* 16 S sequence [13,14] and the *D.melanogaster* sequence is attempted (fig. 4) very strong homology is found in nucleotides 1–45 of the coding region. The homology between the rest of the sequences is quite low except for the region from 160–195 which is nearly identical to sequence 135–170 in 16 S RNA. Again a number of additions have occurred between the two homologuous regions. Thus the regions of homology between *E.coli* 16 S RNA and *D.melanogaster* 18 S RNA are included in those found between *D.melanogaster* and *B.mori*. As shown for other eukaryotes [3] the homology between *E.coli* and *D.melanogaster* does not include the CCUCC pyrimidine tract present in *E.coli* position 4–9 which is implicated in mRNA binding [15,16]. An alternative interaction has been postulated between eukaryotic mRNAs and an AGGAAG sequence located at position 7–13 of the eukaryotic rRNA sequence, which is also present in *D.melanogaster* 18 S RNA. Any implication of this sequence in mRNA binding is however purely conjectural since there is no direct evidence for such an interaction.

Sequencing of the 5.8 S–2 S region in *D.melanogaster* which codes for two conserved RNAs, one complete excised region and the ends of two other excised regions [5] led us to propose that localized A–U-rich regions in the precursor rRNA may be one of the structural features recognized by processing enzymes. Such a structure is found here in the transcribed spacer which follows the 3'-end of the 18 S coding region. The last 30 nucleotides of the sequence determined are 93% A–T; moreover it is apparent from sequencing gels (fig.5) that this is a localized feature; i.e., after ~70 nucleotides the sequence returns to a balanced base composition. Thus we believe that localized, nearly pure A–U regions extending for several tens of nucleotides are one of the features which single out on precursor rRNA those regions which must be removed during processing.

Although the coding regions of *D.melanogaster* and *B.mori* are extremely similar, the transcribed spacers are completely divergent. Recent sequencing studies on the 5.8 S–2 S coding region of *Sciara coprophila* (B. R. J., M. L.-D., R. J., in preparation)

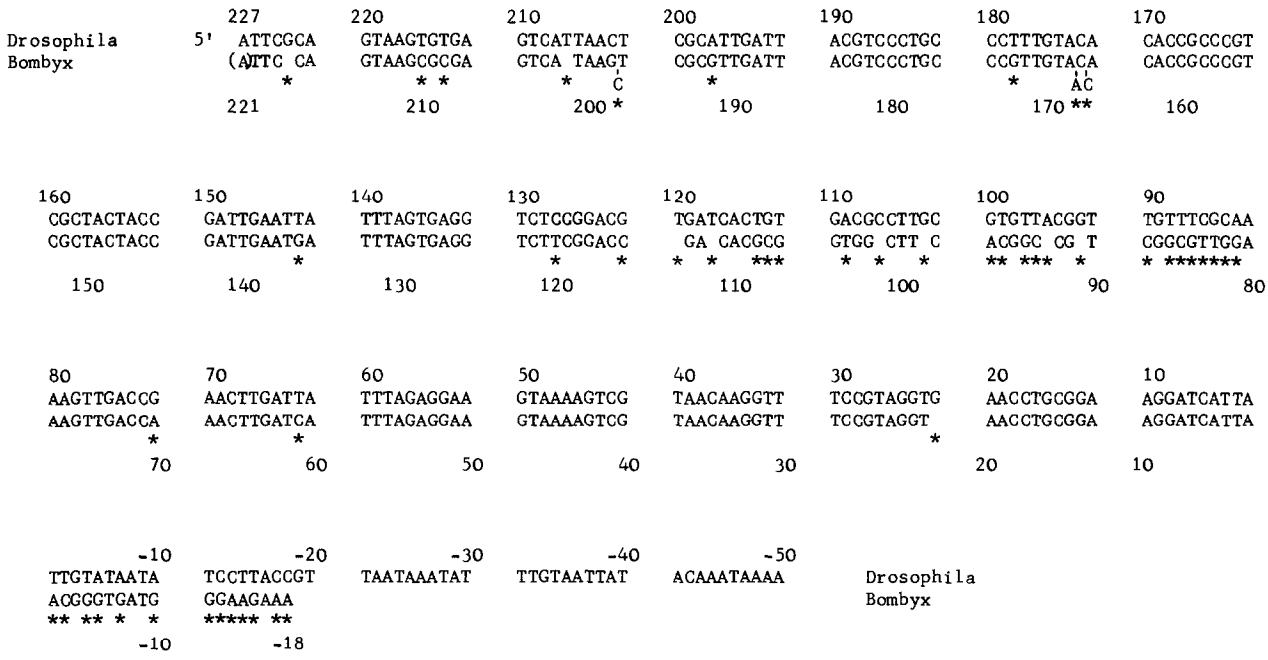


Fig.3. Sequence of the 3'-end of the 18 S rRNA gene of *Drosophila melanogaster* and the first 50 nucleotides of the adjacent spacer region. The sequence is numbered beginning at the 3'-end of the gene proceeding 5' into the gene. The strand shown is the non-coding strand. The *D.melanogaster* sequence (top line) is compared with the *Bombyx mori* sequence [4] (bottom line). Differences between the sequences (substitution, additions or deletions) are denoted by a star.

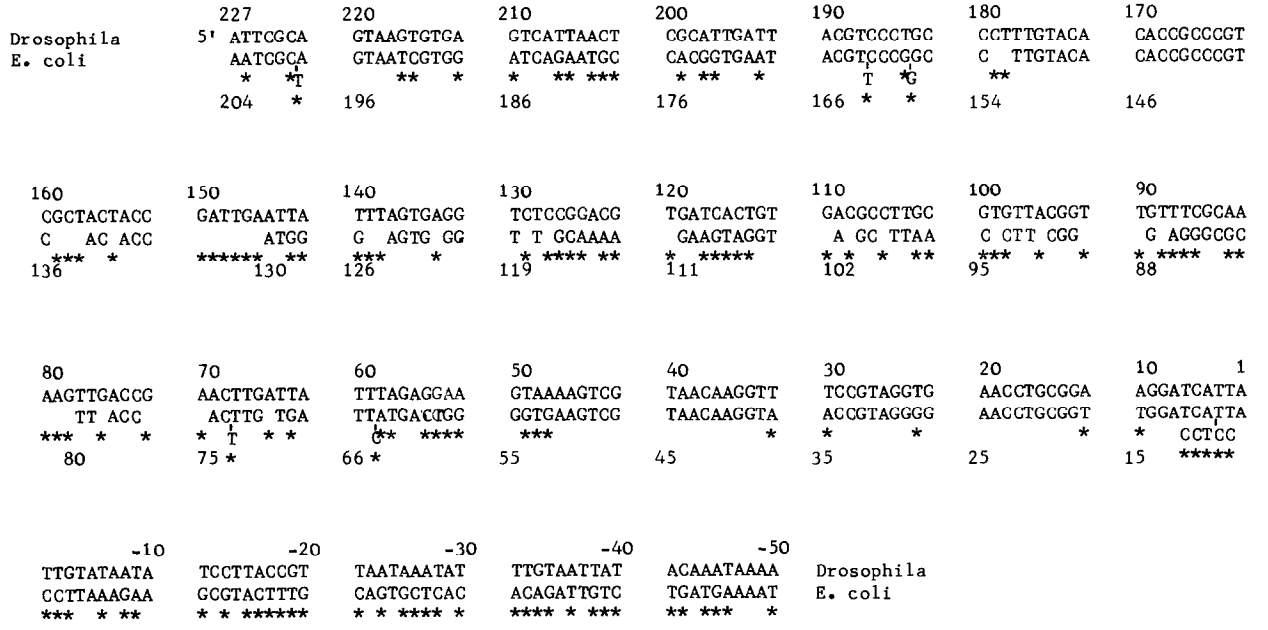
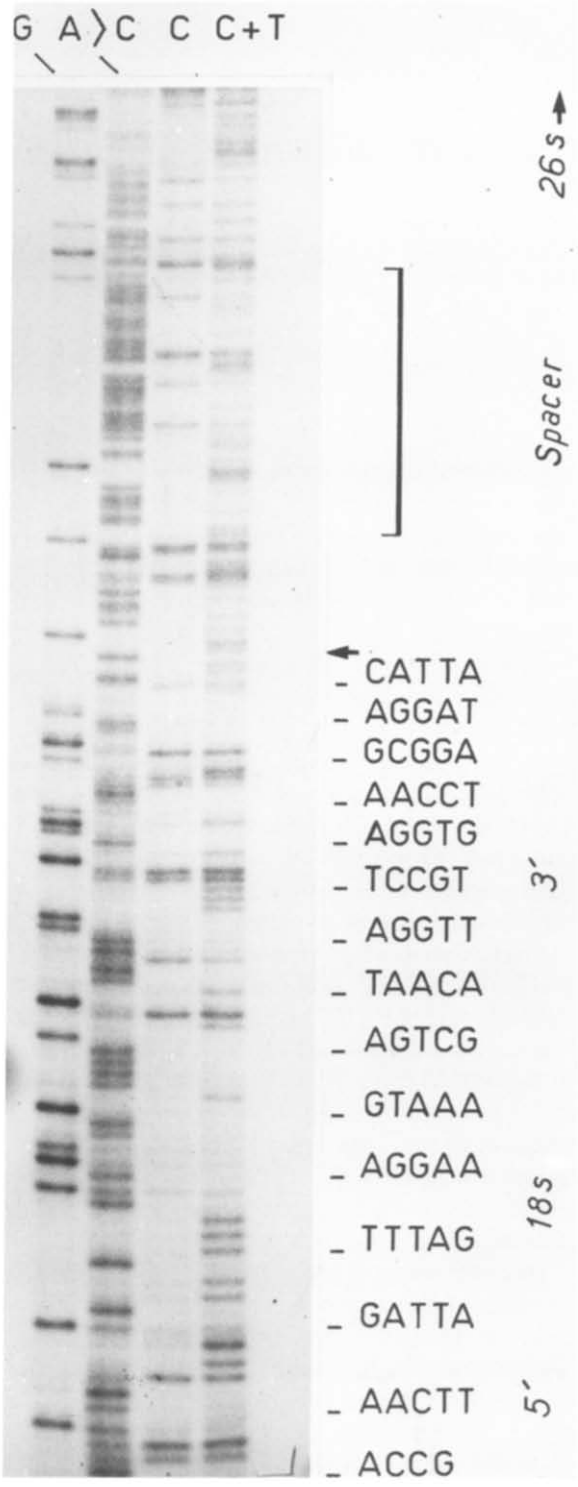


Fig.4. Comparison of *D.melanogaster* and *E.coli* [12,13] sequences. Conventions as in fig.3.



yield similar results; i.e., nearly identical coding regions and completely divergent (albeit still A–U-rich) spacers.

Acknowledgements

We thank Dr D. M. Glover for plasmid cDm 238. This work was supported by the Centre National de la Recherche Scientifique (ATP no. 4257).

References

[1] Bendich, A. J. and Mc Carthy, B. J. (1970) Proc. Natl. Acad. Sci. USA 65, 349–356.
[2] Gerbi, S. A. (1976) J. Mol. Biol. 106, 791–816.
[3] Hagenbüchle, O., Santer, M., Steiz, J. A. and Mans, R. J. (1978) Cell 1, 551–563.
[4] Samols, D. R., Hagenbüchle, O. and Gage, L. P. (1979) Nucleic Acids Res. 7, 1109–1119.
[5] Pavlakis, G. N., Jordan, B. R., Wurst, R. M. and Vournakis, J. N. (1979) Nucleic. Acids Res. 7, 2213–2237.
[6] Glover, D. M., White, R. L., Finnegan, D. L. and Hogness, D. S. (1975) Cell 5, 149–155.
[7] Maxam, A. M. and Gilbert, W. (1977) Proc. Natl. Acad. Sci. USA 74, 560–564.
[8] Roychoudhury, R., Jay, E. and Wu, R. (1976) Nucleic Acids Res. 3, 863–877.
[9] Sanger, F. and Coulson, A. R. (1978) FEBS Lett. 87, 107–110.
[10] Glover, D. M. (1977) Proc. Natl. Acad. Sci. USA 74, 4932–4936.
[11] Manning, R. F., Samols, D. R. and Gage, L. P. (1978) Gene 4, 153–166.
[12] Jordan, B. R., Jourdan, R. and Jacq, B. (1976) J. Mol. Biol. 101, 85–105.
[13] Brosius, J., Palmer, M. L., Kennedy, P. J. and Noller, H. F. (1978) Proc. Natl. Acad. Sci. USA 75, 4801–4805
[14] Young, R. A., Macklis, R. and Steitz, J. A. (1979) J. Biol. Chem. 254, 3264–3271.
[15] Shine, J. and Dalgarno, L. (1974) Proc. Natl. Acad. Sci. USA 71, 1342–1346.
[16] Steitz, J. A. and Jakes, K. (1975) Proc. Natl. Acad. Sci. USA 72, 4734–4738.

Fig. 5. 8% sequencing gel of the *D.melanogaster* 18 S spacer region showing the localized A–T-rich region found at the beginning of the spacer. The 18 S sequence shown is the non-coding strand (the last 20 nucleotides were read from another gel); the A–T-rich region is shown by a bracket.