Volume 8 Number 16 1980 Nucleic Acids Research

Coding and spacer sequences in the 5.8S - 2S region of Sciara coprophila ribosomal DNA

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 30 June 1980

ABSTRACT

The sequence of 436 nucleotides around the region coding for 5.8S RNA in the Sciara coprophila rDNA transcription unit (1) has been determined. Regions coding for 5.8S and 2S RNAs have been identified ; they are 80 - 90 % homologuous to the corresponding Drosophila sequences and are separated by a 22 nucleotide long spacer. This sequence as well as the two before the 5.8 and after the 2S coding region are very different from the corresponding Drosophila sequences. The main features reported in the Drosophila study (2) are however also found, i.e. all three spacers are very rich in A-T ; the sequence of the internal spacer allows base pairing ; 5.8S and 2S RNAs can pair through their 3' and 5' terminal regions respectively. The features previously proposed as processing sites in the Drosophila case are thus all found in Sciara in spite of very different spacer sequences.

INTRODUCTION

In Drosophila melanogaster (3) the structure of the large ribosomal RNA molecule is somewhat different from that found in most eukaryotes : the mature 26S (or 28S) rRNA is made up of four hydrogen bonded polynucleotides : 26Sa, 26Sb, 5.8S and 2S. In Drosophila these are the result of maturation cleavages which remove the central portion of the molecule, a 400 nucleotide long region close to the 5' end of 26S and a short (28 nucleotides) spacer which separates 5.8S and 2S (2,3,4) (see Fig. 1). The 5.8S-2S region in rDNA thus makes it possible to examine a short transcribed piece of DNA which codes for two conserved and three non-conserved regions. The processing event which removes from precursor ribosomal RNA the 5.8S-2S spacer is particularly well defined and is a late event which occurs in the cytoplasm (3). Sequencing studies on Drosophila 5.8S and 2S RNAs, DNA sequencing of the corresponding rDNA and studies of complex formation between these molecules and 26S RNA showed that the 123 nucleotide long 5.8S and the 30 nucleotide long 2S Drosophila molecules correspond respectively to the 5' part and the normally 160 nucleotide long 5.8S RNA and led to predictions on the structure of the precursor RNA molecule

and on the features recognized by processing enzymes (2).

Sciara coprophila, a dipteran quite distant in evolutionary terms from Drosophila melanogaster, has been shown to contain a short (130 nucleotide long) 5.8S RNA (1). It is thus likely that the organization of its rDNA is similar to that of Drosophila. We have determined the DNA sequence of the corresponding region in a cloned rRNA gene from this organism (1) to examine whether the features identified in Drosophila are also present in this case. Most of these are found in the sequence of spacer regions which we expect to diverge considerably. Thus if some features are conserved in spite of this sequence divergence this will be a strong indication that they are biologically significant.

METHODS

Restriction enzymes Eco RI, Hpa I and Pst I were purchased from Boehringer ; Hind III was from New England Biolabs. T4 nucleotide kinase was from Boehringer, bacterial alkaline phosphatase and terminal transferase from Bethesda Research Labs. α- and γ-^{32}P ATP's were purchased from Amersham.

Sciara ribosomal DNA was prepared from recombinant plasmid pBC2 kindly provided by R. Renkawitz. Whole plasmid DNA was labelled with polynucleotide kinase (after digestion with Hpa I restriction enzyme and dephosphorylation) or with terminal transferase in the presence of cobalt ion after digestion with Pst I restriction enzyme. The labelled DNA was then recut with either Eco RI or Hind III restriction enzymes and fractionated by agarose gel electrophoresis. The relevant bands, previously identified by analytical multiple digestion experiments, were cut out and electro eluted in dialysis bags (5).

DNA sequencing was performed essentially as described by Maxam and Gilbert (6). 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 all four cleavage steps. Reaction products were run on 20 %, 8 % and 6 % thin (0.3 mm) gels (7) which were autoradiographed with Ilford fast tungstate screens at - 70°. 150 to 200 bases could normally be read from these gels.

RESULTS

Fig. 1 shows a map of plasmid pBc 2 with a blow up of the region of interest to us. The Hpa I site indicated by Renkawitz et al. (1) near the 5' end of the 5.8S coding region is located 27 nucleotides upstream ; a Pst I site

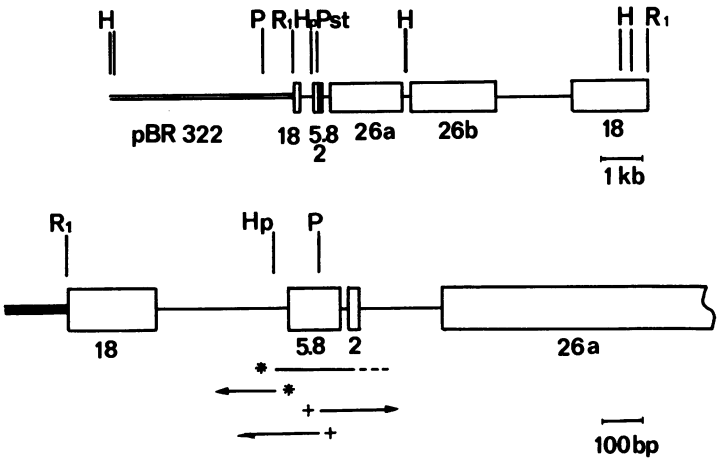

Fig. 1 : Restriction map of plasmid pBc 2.

Top : complete map (1) ; Bottom : blow up of transcribed spacer region. Ribosomal RNA coding sequences are boxed ; pBR 322 sequences are indicated by a thick line and spacer regions by a thin line. Cleavage sites are indicated for Eco RI (RI), Pst I (P), Hpa I (Hp) and Hind III (H) restriction enzymes. The position type of labelling (* : 5', + : 3') and extent of sequence read on the fragments used for sequencing are indicated below the map.

is present in the 5.8S coding sequence. The sequence of the region was determined by the Gilbert-Maxam chemical degradation methods using the fragments indicated on Fig. 1. Most of the sequence (except nucleotides of the leftward transcribed spacer) was determined on both strands ; two "compression artefacts" which would otherwise have led to errors of one or two nucleotides were thus detected and corrected. A representative sequencing gel is shown in Fig. 2 and the sequence determined is displayed in Fig. 3.

The nucleotide sequence of Sciara 5.8S and 2S rRNAs has not been reported although the 5.8S molecule is known to be approximately 130 nucleotides long in this species (1). However the homology with the Drosophila molecules is evident and the sequences can be quite unambiguously aligned (Fig. 3).

DISCUSSION

The first conclusion from these results is that the general organization of the 5.8S-2S rDNA region in Sciara coprophila is similar to that found in Drosophila melanogaster (2,3), i.e. the normally 160 nucleotide long 5.8S molecule is split into a 123 nucleotide long "m5.8S" (using the nomenclature introduced in (2)) and a 30 nucleotide long 2S molecule, and these two coding

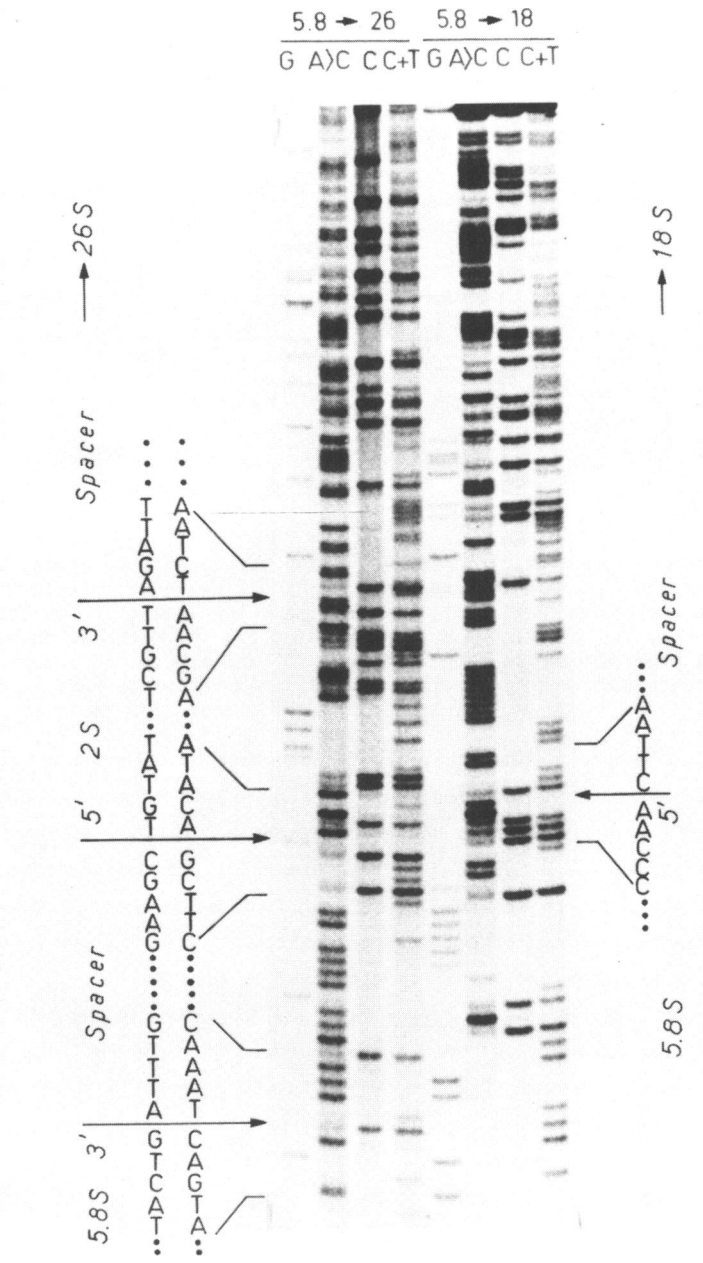

Fig. 2 : Maxam-Gilbert (6) sequencing gel showing
a) left : 3' end of the 5.8S sequencing internal spacer, 2S and external spacer (coding strand). b) right : 5' end of the 5.8S sequence, external spacer towards 18S DNA.

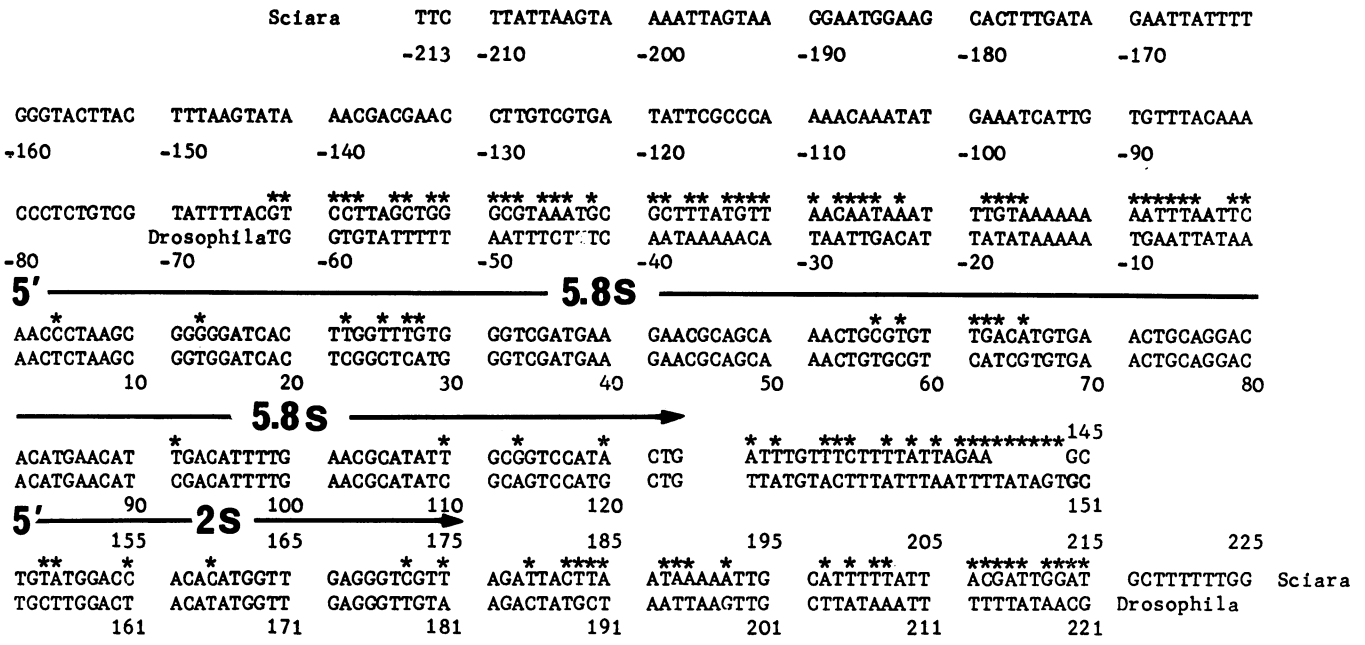

Fig. 3 : Sequence of the 5.8S-2S coding region in Sciara coprophila.

The non-coding strand of the sequence determined is shown in the orientation (18S) spacer, 5.8S, spacer, 2S, spacer (26S). The sequence is numbered starting at nucleotide 1 of the 5.8S coding region.

The sequence determined in Drosophila melanogaster (2) is shown below, differences are denoted by stars. The two sequences have been aligned according to the very strong homology found in coding sequences, the position of the "gap" in the internal spacer is largely arbitrary.

regions are separated by a short (22 bp) spacer. Recent results obtained on small ribosomal RNAs of a number of organisms suggest in fact that dipterans do contain "split" 5.8S molecules (i.e. 5.8S and 2S) whereas other insects contain longer 5.8S species which however differ from their vertebrate counterpart in the 3' terminal region of the molecule (Pavlakis, G.N. and Vournakis, J.N., private communication).

A comparison of the Drosophila and Sciara sequences (Fig. 3) shows that the 5.8S and 2S coding regions are quite strongly conserved. The overall homology is 87 % and 80 % respectively. This was expected since the restriction enzyme maps are generally quite similar in the coding regions for several insect rDNAs (1,8).

The differences found in the sequence of the 5.8S molecule are quite compatible with the secondary structure model proposed for Drosophila 5.8S (2). They either occur in regions considered to be single stranded or change marginally the stability of proposed double helical regions (Fig. 4). The two base changes at position 56 and 58 do disrupt the stem thought to exist in this position ; however an equivalent paired region can be formed with a small shift of the paired strands (Fig. 4). Altogether the pattern of differences between Drosophila and Sciara 5.8S RNAs supports the secondary structure model previously proposed. The sequence deduced from the DNA data for Sciara 2S RNA is also closely homologuous to the Drosophila sequence.

If we now compare the spacer sequences it is apparent that they are quite divergent and show very little sequence homology. However the important feature detected in the Drosophila study is also found in the Sciara sequences, i.e. all three spacers show localized A-T rich regions immediately adjacente to the conserved sequences.

The most important result from the Drosophila study in terms of possible structure of "processing sites" was the proposal of a particular secondary structure for the transcription product of the 5.8S-spacer-2S region. On the basis of sequence data and complex formation experiments it was proposed that a paired region is formed between the 3' end of the 5.8S and the 5' end of the 2S sequences respectively, and that the spacer loops out as an A-U rich region possibly with some internal base pairing. Other features found in this region were a direct repeat (UGCUG) at the position where cleavage occurs and a palindrome ($U_3A_2U_2A_2U_3$) in the predicted loop (Fig. 5) (2). The Sciara sequence can be arranged in a similar structure. The paired region between the 5.8 and 2S sequences is preserved : the four base changes found in this stem occur in such a way as to maintain (actually increase) the stability of the

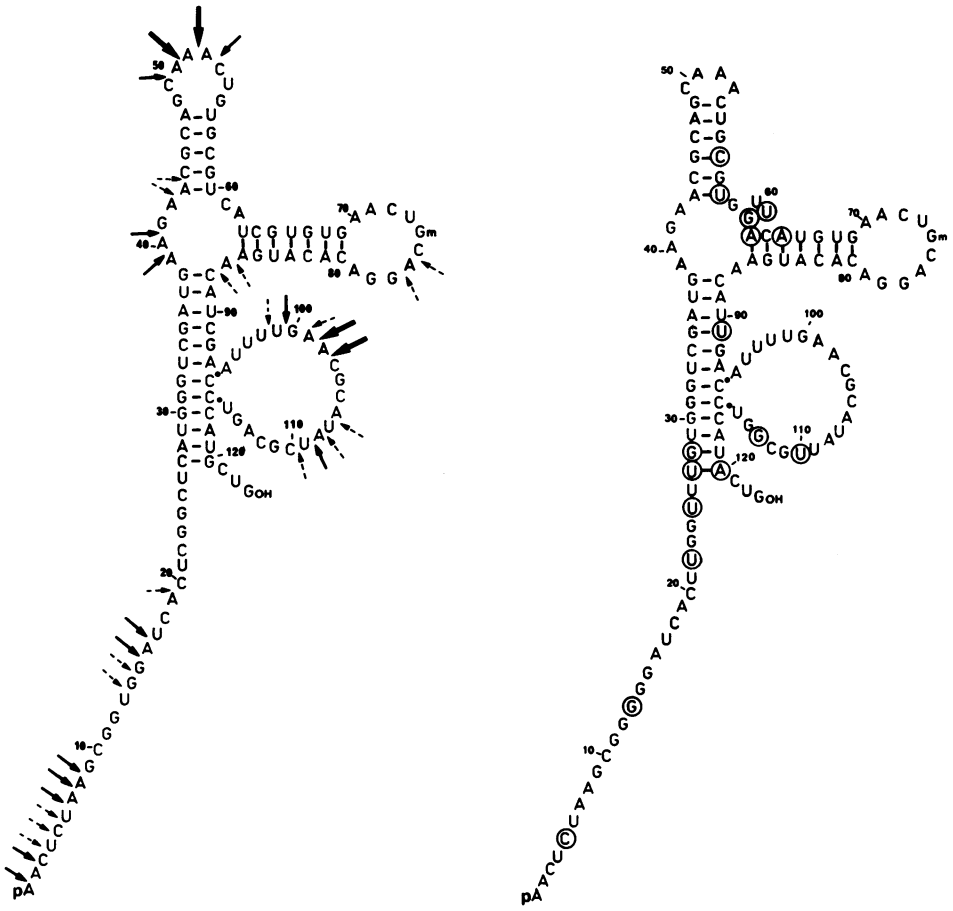

Fig. 4 : Comparison of Drosophila and Sciara 5.8S RNAs. Fig. 4A shows the model previously proposed for Drosophila 5.8S RNA on the basis of S1 accessibility experiments (2) ; Fig. 4B shows the Sciara sequence arranged in a similar pairing scheme. These nucleotides in the Sciara sequence which are different from their Drosophila counterpart are circled. The paired stem at the top of the model has been modified in the Sciara model by a two nucleotide displacement of the second strand.

helix. The spacer sequence can form a short helical region (of marginal stability, as in the Drosophila case) and the looped out part forms an imperfect palindrome ($U_2CU_4AU_2$). The direct repeat at the cleavage point is only three nucleotides long in this case (CUG) and probably not significant. In spite of these differences the general structure of this region is very similar and the conservation of the base-paired stem in particular is quite striking.

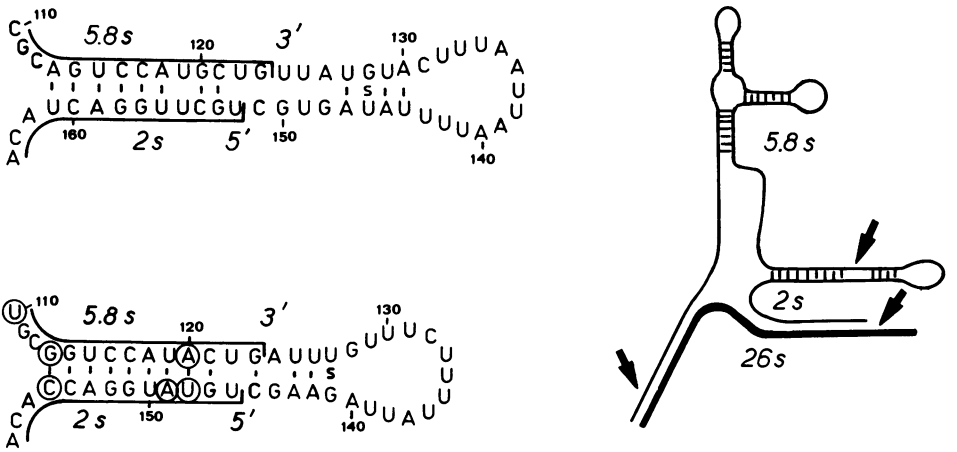

Fig. 5 : 5.8S-2S base pairing and precursor rRNA structure 5.8S-2S base pairing is shown for the Drosophila (top) and Sciara (bottom) sequences. Possible pairing in the internal spacer is also shown. These nucleotides in the Sciara 5.8S and 2S sequences which are different from their Drosophila counterparts are circled.

A possible structure for the precursor rRNA is shown on the left ; the cleavage points which generate 5.8S and 2S RNAs are shown by arrows.

In conclusion the sequence of the 5.8S-2S coding region in Sciara coprophila rDNA shows that in spite of sequence differences in the coding regions and almost complete divergence of the transcribed spacers the main features found in the case of Drosophila rDNA and tentatively identified as characteristics of processing sites are present. This makes it very likely that they have functional significance.

ACKNOWLEDGMENTS

We thank Drs R. Renkawitz for providing plasmid pBC2 as well as some pBC2 DNA and J. Vournakis for communication of unpublished information. This work was supported by a grant from the CNRS (ATP 4257). Biohazards associated with these experiments were previously examined by the French Control Committee.

REFERENCES

1. Renkawitz, R. and Gerbi, S.A. (1979) Molec. Gen. Genet. 173, 1-13
2. Pavlakis, G.N., Jordan, B.R., Wurst, R.M. and Vournakis, J.N. (1979) Nucl. Acids Res. 7,2213-2237
3. Jordan, B.R., Jourdan, R. and Jacq, B. (1976) J. Mol. Biol. 101, 85-105

4. Long, E.O. and Dawid, I.B. (1980)
J. Mol. Biol. 138, 873-878
5. Jordan, B., Latil-Damotte, M. and Jourdan, R. (1980) Febs Lett., in press.
6. Maxam, A.M. and Gilbert, W. (1977) Proc. Natl. Acad. Sci. USA 74, 560-564
7. Sanger, F. and Coulson, A.R. (1978) Febs Lett. 87, 107-110
8. Beckingham, K. and White, R. (1980) J. Mol. Biol. 137,349-373.