5'-Cleavage site of *D. melanogaster* 18 S rRNA

A. Simeone and E. Boncinelli

International Institute of Genetics and Biophysics, via Marconi 10, Naples, Italy

Received 9 December 1983

We determined the nucleotide sequence of the DNA region around the 5'-terminus of 18 S rRNA in two cloned rDNA gene units of *Drosophila melanogaster*. The 5'-base is within a sequence CATTATT which is present also at the 3'-terminus of the 18 S rRNA coding region. In this case it is known that the situation is CATTA3'·TT. With various methods we determined that the precise 5'-cleavage site is CATT·5'ATT.

18 S ribosomal RNA *5'-Terminus* *3'-Terminus* *Processing site* *S1 protection*

1. INTRODUCTION

Small and large subunit rRNAs are cotranscribed in both prokaryotic and eukaryotic cells in long precursors which are then processed to produce mature rRNAs [1–3]. rRNA precursors and mature rRNAs are analogous in all cells.

The initial steps in rRNA processing are especially well characterized in *Escherichia coli*. Here primary cleavages are made by RNase III at very stable double-stranded stems enclosing large loops containing the 16 S or 23 S rRNA sequences [4,5]. The resultant RNAs bear extra nucleotides at their termini; the mature 5'- and 3'-termini are formed by secondary cleavages which occur after the intermediates have bound ribosomal proteins. Although the mechanisms of eukaryotic pre-rRNA processing are largely unknown, published studies have suggested some differences between the pre-rRNA processing in eukaryotic cells and *E. coli*.

The spacer sequences flanking the small and large subunit rRNAs in yeast [6], *Xenopus laevis* [7], and mouse [8] cannot form long stable stems as in *E. coli*. Stems of much lower stability are possible, and it has been suggested (at least for yeast) that these may be analogous to the *E. coli* stems. The initial cleavages of pre-rRNA in *E. coli* occur while the pre-rRNA is being transcribed, whereas the initial cleavages of higher organisms' pre-rRNA occur after the completion of transcription.

Fig.1 shows the organization of a rDNA gene unit of *D. melanogaster*. Nontranscribed spacers (NTS) separate two adjacent transcription units. A typical transcription unit comprises, in turn, an external transcribed spacer (ETS) of about 850 base pairs, the 18 S rRNA coding region, an internal transcribed spacer (ITS) which includes the coding regions for the 5.8 S RNA and the 2 S RNA, and the 28 S coding region. In several gene units of *D. melanogaster* the 28 S coding region is interrrupted at about two-thirds from the 5'-terminus by an rDNA insertion. Transcription begins at the 5'- of the ETS and proceeds to the 3'- of the 28 S coding region [10]. Processing of the pre-rRNA precursor proceeds along at least two alternative pathways [11]. According to one pathway the cleavage originating at the 5'-terminus of mature 18 S rRNA precedes the 3'-cleavage of the same mature rRNA, whereas according to the other pathway the termini originate at the same time.

We report here the nucleotide sequence of the DNA region containing the 5'-terminus of the mature 18 S rRNA of *D. melanogaster*. The sequence around the 5'-nucleotide shows a striking homology to the 3'-terminus of the same molecule: the heptanucleotide CATTATT is present at both locations [9]. On the other hand, 5'-

Published by Elsevier Science Publishers B.V.
00145793/84/$3.00

and 3'-cleavage sites differ from each other for one base. It is conceivable to speculate on the role of this short sequence in the processing mechanisms.

2. MATERIALS AND METHODS

2.1. *Construction of clones and DNA sequencing*

Plasmids pDmra56 and pDmrY22, containing *D. melanogaster* rDNA gene units, were originally obtained from P.K. Wellauer [15]. *Hae*III–*Alu*I fragments containing the boundary region between external transcribed spacer and 18 S rRNA coding region from these plasmids were subcloned into the *Sma*I site of a pBR322 derivative, namely pK011 [16].

DNA sequences were determined according to [17] after 5'-end labeling.

2.1. *S1 nuclease mapping*

Cytoplasmic RNA was extracted from *D. melanogaster* embryos, 18 S rRNA purified as in [18], and hybridized to the coding strand of the *Hae*III–*Rsa*I DNA fragment shown in fig.1, end-labeled at the 5'-end with ^{32}P. Hybridization was in 80% formamide–0.4 M NaCl–0.02 M Pipes (pH 6.4)–0.001 M EDTA at a temperature slowly decreasing from 85°C to 14°C in about 8 h. The hybrids were diluted 30-fold with ice-cold 0.28 M NaCl–0.05 M sodium acetate (pH 4.6)–0.0045 M $ZnSO_4$–20 μg of carrier ssDNA from salmon sperm per ml and were digested with various concentrations of S1 nuclease [19] at 14°C or 37°C for 3 h. The resulting hybrids were precipitated with ethanol and run on a 10% acrylamide sequencing gel along G- and (G + A)-specific sequencing reactions of the original *Hae*III–*Rsa*I DNA fragment.

2.3. *Determination of the 5'-terminal nucleotide of 18 S rRNA*

Purified 18 S rRNA was labeled at the 5'-end as for DNA fragments and subjected to complete digestion with S1 nuclease. Digests were separated on PEI–cellulose in parallel with the 4 pN nucleotide markers using 1 N acetic acid as solvent in an ascending chromatography [20].

3. RESULTS AND DISCUSSION

The structure of a typical ribosomal unit of *D. melanogaster* is shown in fig.1. The maturation process leading to the formation of stable RNA products from such long precursors is largely unknown. Alternative pathways have been suggested in the maturation of 18 S rRNA, but no information is available about the nature and location of regulatory signals for the involved machinery.

We have determined the nucleotide sequence of the ETS and 18 S rRNA boundary region in two cloned rDNA units from *D. melanogaster* (fig.2).

Comparison of this sequence with published sequences of 18 S rRNA of different species allows identification of the 5'-terminus of *Drosophila* 18 S rRNA as 5'ATT. Two points emerge from such analysis. This is the first example in which the 5'-end base is A whereas the other RNA molecules start with U. The inferred 5'-terminus of the 18 S

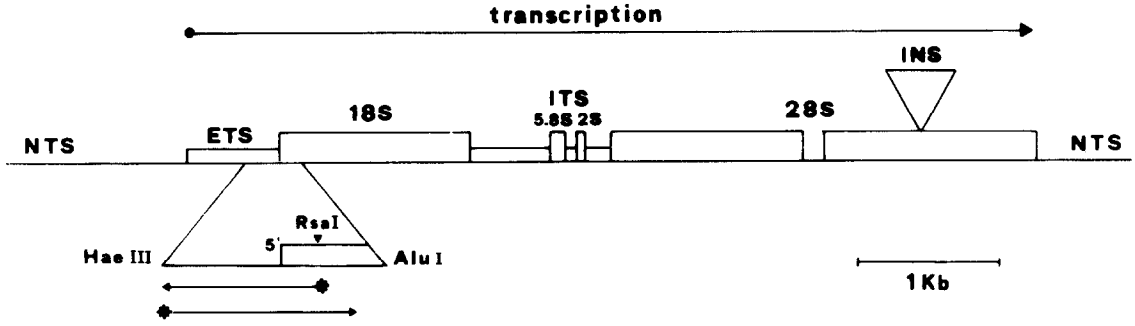


Fig.1. Schematic representation of a rDNA gene unit of *D. melanogaster*. Recognition sites of relevant restriction enzymes are shown. The sequencing strategy is shown below the map; asterisks indicate ^{32}P-labeled 5'-ends. NTS, nontranscribed spacer; ETS, external transcribed spacer; ITS, internal transcribed spacer; INS depicts rDNA insertion. Coding regions are designated according to the name of their product (large open boxes).

```
-60                  -40                  -20
AATATTTATA TTGCTTATTT CAATTCAAAA AATATGAATG AAATATGAAA          DROSOPHILA


-10      5'                +20                   +40
AGAAAACATT ATTCTGGTTG ATCCTGCCAG TAGTTATATG CTTGTCTCAA          DROSOPHILA

           TACCTCGTTG ATCCTGCCAG TAGC-ATATG CTTGTCTCAA          XENOPUS

           TACCTCGTTG ATCCTGCCAG TAGC-ATATG CTTGTCTCAA          RABBIT

           TACCTCGTTG ATCCTGCCAG TAGC-ATATG CTTGTCTCAA          MOUSE

           TATCTCGTTG ATCCTGCCAG TAGTCATATG CTTGTCTCAA          YEAST


                  +60            -20        -10      3'
AGATTAAGCC ATGCATGTCT ---------- AACCTGCGGA AGGATCATTA  TTGTA   DROSOPHILA

AGATTAAGCC ATGCACGTGT ---------- AACCTGCGGA AGGATCATTA         XENOPUS

AGATTAAGCC ATGCATGTCT ---------- AACCTGCGGA AGGATCATTA         RABBIT

AGATTAAGCC ATGCATGTCT ---------- AACCTGCGGA AGGATCATTA         MOUSE

AGATTAAGCC ATGCATGTCT ---------- AACCTGCGGA AGGATCATTA         YEAST
```

Fig.2. Comparison of *D. melanogaster* 18 S rRNA coding sequence at the 5′- and 3′-domains with *Xenopus laevis*, rabbit, mouse and yeast 18 S rRNA coding sequences. Noncoding strand is shown. For *Drosophila*, 60 nucleotides of the ETS and the first 5 nucleotides of the ITS are also shown. Bases differing from a common consensus are underlined. Spacer sequences of different species are completely divergent and are not reported here.

rRNA is located within a heptanucleotide (CATTATT) which is precisely repeated at the boundary between the 18 S rRNA 3′-terminus and the ITS [9]. The 3′-terminus of *D. melanogaster* 18 S rRNA, as in other organisms, is CAUUA3′ [12–14]; the inferred nucleotide sequence of the 5′-end of the same molecule seems to be 5′AUU, implying that an identical sequence located at both ends of this region is differently recognized and/or utilized in maturation.

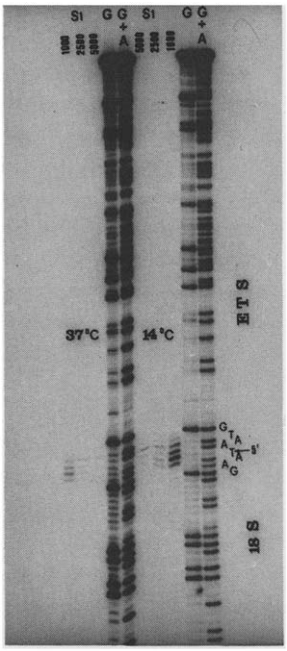
To test this hypothesis, we have analyzed by S1 mapping [19] the actual 5′-end of the mature 18 S rRNA of *D. melanogaster*. The scheme and the result of this experiment are shown in fig.3. A purified *Hae*III–*Rsa*I DNA fragment spanning the 5′-end of the 18 S rRNA coding region (fig.1) was terminally labeled at its 5′-ends by polynucleotide kinase and the two strands separated by gel electrophoresis [17]. The coding strand was hybridized to purified 18 S rRNA and subjected to nuclease S1 digestion.

S1 resistant hybrids were run on a polyacrylamide sequencing gel along with sequence ladders of the same coding strand. This experiment confirms that the 5′-terminus of the mature 18 S rRNA is in fact AUU, as inferred from the comparison to other species rRNAs.

Direct analysis of the 5′-terminal nucleotide of the 18 S rRNA, shown in fig.4, indicates that it is A.

From our experiments it emerges that the repeated heptanucleotide CATTATT present at both ends of the 18 S rRNA coding region is recognized with different specificity in the maturation process. Actually the 5′-cleavage site is CAUU·5′AUU, whereas the 3′-cleavage site is CAUUA3′·UU.

It is impossible to assess from this analysis



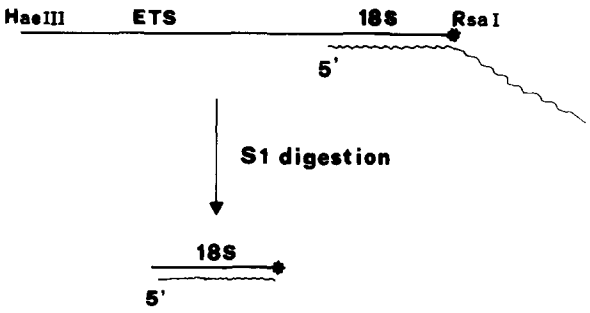


Fig.3. S1 nuclease mapping of the cleavage site at the 5'-end of 18 S rRNA of *D. melanogaster*. Below is shown a scheme of the experiment: annealing was performed between purified 18 S rRNA and the coding strand terminally labeled at the *RsaI* site. The sequence ladder of the same DNA fragment is shown together with S1 protected DNA fragment. Figures below S1 indicate the units used.

whether this sequence is sufficient for the 18 S rRNA maturation or only a part of a complex recognition signal including both nucleotide primary sequence and secondary RNA structures. Constructions in which these two heptanucleotides can be displaced or modified should help to identify, in in vivo experiments, their respective functional roles.

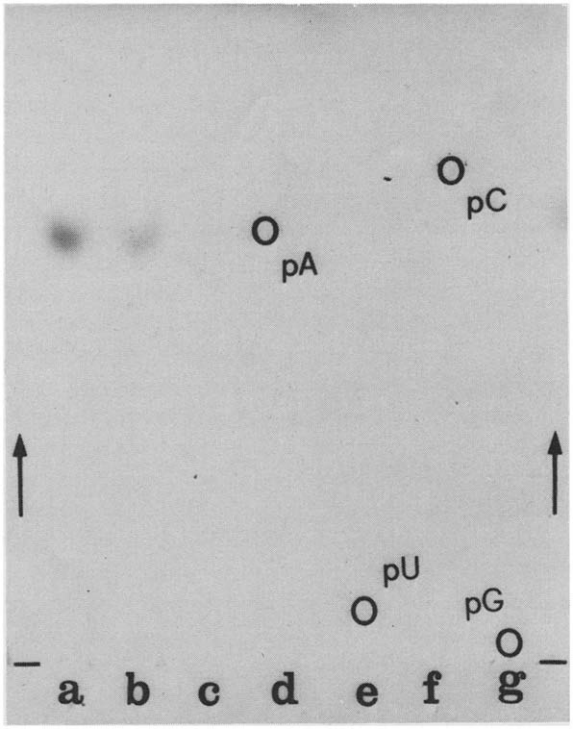


Fig.4. PEI–cellulose chromatography of nuclease S1 digested 18 S rRNA labeled at the 5'-end. (a–c) Different amounts of digested rRNA; (d–g) pN nucleotide markers as indicated.

ACKNOWLEDGEMENT

This work was supported by Progetto Finalizzato Ingegneria Genetica e Basi Molecolari delle Malattie Ereditarie of the CNR.

REFERENCES

[1] Abelson, J. (1974) Annu. Rev. Biochem. 48, 1035–1069.
[2] Hadjolov, A.A. and Nikolaev, N. (1976) Prog. Biophys. Mol. Biol. 31, 95–144.
[3] Perry, R.P. (1976) Annu. Rev. Biochem. 45, 605–629.
[4] Bram, R.J., Young, R.A. and Steitz, J.A. (1980) Cell 19, 393–401.
[5] Young, R.A. and Steitz, J.A. (1978) Proc. Natl. Acad. Sci. USA 75, 3593–3598.

[6] Veldam, G.M., Klootwijk, J., Heerikhuizan, H. and Planta, R.J. (1981) Nucleic Acids Res. 9, 4847–4862.
[7] Maden, B.E.H., Moss, M. and Salim, M. (1982) Nucleic Acids Res. 10, 2387–2398.
[8] Goldman, W.E., Goldberg, G., Bowman, L.H., Steinmetz, D. and Schlessinger, D. (1983) Moll. Cell. Biol. 3, 1488–1500.
[9] Jordan, B.R., Latil-Damotte, M. and Jourdan, R. (1980) FEBS Lett. 117, 227–231.
[10] Glover, D.M. (1981) Cell 26, 297–298.
[11] Long, E.O. and Dawid, I.B. (1980) J. Mol. Biol. 138, 873–878.
[12] Lockard, R.E., Connaughton, J.F. and Kumar, A. (1982) Nucleic Acids Res. 10, 3445–3457.
[13] Bowman, L.H., Goldman, W.E., Goldberg, G.I., Hebert, M.B. and Schlessinger, D. (1983) Mol. Cell. Biol. 3, 1501–1510.
[14] Torczynski, R., Bollon, A.P. and Fuke, M. (1983) Nucleic Acids Res. 11, 4879–4890.
[15] Long, E.O. and Dawid, I.B. (1980) Annu. Rev. Biochem. 49, 727–764.
[16] Simeone, A., De Falco, A., Macino, G. and Boncinelli, E. (1982) Nucleic Acids Res. 10, 8263–8272.
[17] Maxam, A.M. and Gilbert, W. (1980) Methods Enzymol. 65, 499–560.
[18] Boncinelli, E. and Furia, M. (1979) Molec. Gen. Genet. 176, 81–85.
[19] Berk, A.J. and Sharp, P.A. (1978) Proc. Natl. Acad. Sci. USA 75, 1274–1278.
[20] Randerath, K. and Randerath, E. (1967) Methods Enzymol. 12A, 323–350.