
Nucleotide sequence of a complete ribosomal spacer of *D. melanogaster*

Antonio Simeone, Adriana La Volpe and Edoardo Boncinelli

International Institute of Genetics and Biophysics, CNR, v. G. Marconi 10, 80125 Naples, Italy

Received 21 December 1984; Accepted 18 January 1985

ABSTRACT

We determined the nucleotide sequence of a *D. melanogaster* ribosomal DNA spacer. Sequences of various portions of different cloned ribosomal spacers have been previously reported. We extend the analysis to cover the entire nontranscribed and external transcribed regions. Comparison to other cloned ribosomal DNA gene units of this species confirms a conserved general organization of the ribosomal spacer through different size classes. *D. melanogaster* ribosomal gene units interrupted by insertions are known to be transcribed at a much lower level than the continuous gene units. Nonetheless previous sequence analysis of a region around the transcription initiation site did not reveal significant differences in rDNA genes with and without insertions. We extend such analysis to cover the last two promoter duplications in the spacer and the entire external transcribed spacer up to the 5' cleavage site of the 18S rRNA.

INTRODUCTION

Ribosomal genes (rDNA) of every higher organism studied are tandemly repeated and clustered with contiguous transcribed regions separated by nontranscribed spacers (NTS) (1). The 28S, 18S and 5.8S rRNA molecules are transcribed as parts of a large precursor molecule which is subsequently processed in several steps to yield the mature rRNA species (2, 3, 4). While the mature rRNA coding regions are largely conserved among eukaryotes, spacer regions evolved much more rapidly and diverge in all but the most closely related species (5, 6). Spacer regions comprise the NTS region and the portion of the transcribed region upstream to the 18S rRNA coding region, termed external transcribed spacer (ETS).

In *X. laevis* and *D. melanogaster*, among others, the NTS region is highly variable in length within the same locus, and the size distribution of fragments may be quite different in individual animals (7, 8). This length

heterogeneity appears to be originated by different numbers of internally repetitious portions of the spacers (7, 9, 10).

In *D. melanogaster* there are 150-250 ribosomal gene units in each nucleolus organizer present on the X chromosome and the Y chromosome. A large fraction of gene units are interrupted in the 28S rRNA coding region by noncoding DNA segments which have been called ribosomal insertions. These insertions occur in two distinct sequence types, each represented by several size classes (11, 12). rDNA repeating units containing an insertion are generally not transcribed (13, 14, 15, 16, 17, 18). Nucleotide sequences around the transcription initiation site do not differ in cloned genes with and without insertions (19).

Various portions of different cloned NTS regions have been sequenced (19, 20, 21, 22, 23, 24). We report here on the nucleotide sequences of the complete spacer, i.e. NTS and ETS regions, of a cloned *D. melanogaster* gene and extend the sequence comparison of cloned genes with and without insertions.

MATERIALS AND METHODS

rDNA clones

pDmr a56 and pDmr Y22 were obtained from P. K. Wellauer (25). Recombinant phage λ Dmr275 contains a ribosomal spacer upstream to a continuous gene unit, whereas λ Dmr231, λ Dmr241 and λ Dmr214 contain spacers upstream to gene units with type 1 insertions (19). These phage clones were kindly provided by I. B. Dawid.

D. melanogaster stocks

Stock 1 males are $\bar{X}Y, y^2 su(w^a) w^a bb/Ybb^+$. Stock 2 males are R(1)2, cv v f bb/Ybb⁺. Wild-type Oregon-R is from the University of Naples.

Genomic DNA analysis

Genomic DNA was digested, electrophoresed on agarose gels, transferred onto nitrocellulose filters according to Southern (26) and hybridized to a ³²P-nick-translated probe (27).

Terminal labelling and fine mapping

The fragments to be analyzed were end-labelled at the 5' end using T4

polynucleotide kinase (28) or at 3' end using the large fragment of DNA polymerase I of *E. coli* and the appropriate α - 32 P-dNTP. Asymmetrically labelled fragments were partially digested and separated according to Smith and Birnstiel (29).

DNA sequencing

DNA sequencing was performed as described by Maxam and Gilbert (28).

RESULTS

A generalized rDNA gene unit of *D. melanogaster* is shown schematically in Fig. 1A. Interrupted gene units (INS+) contain a type 1 or type 2 insertion positioned at about two thirds along the 28S coding region as shown in the figure. The NTS region can be subdivided into 3 portions defined operationally by determined Alu I sites (9, 23). There is a central variable portion containing integral numbers of a basic 240 bp repeated module bound by two contiguous Alu I sites. A gene unit containing 12 such repeats, actually pDmr a56, is depicted in Fig. 1A. The nucleotide sequence of several 240 bp repeats from clone a56 (23) and from other cloned units has been reported (21, 22,

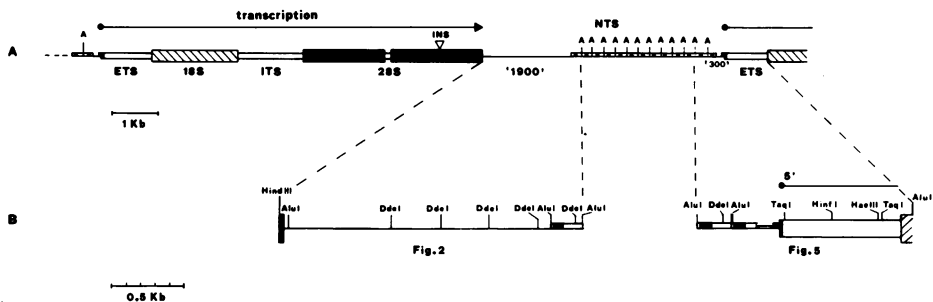


Figure 1. Schematic representation of a typical rDNA gene unit of *D. melanogaster* (A) and enlargement of two regions of the spacer (B). ITS is the internal transcribed spacer, whereas INS indicate the approximate localization of the insertion in interrupted gene units. Small black rectangles indicate the 42 bp sequence homologous to bases -18/+24 around the actual transcription initiation site. Every A in (A) indicates two contiguous Alu I sites delimiting the basic 240 bp repeat in the NTS. The gene unit shown contains in its NTS 12 copies of this repeat, as it is the case for pDmr a56. The nucleotide sequences of the regions shown in (B), on the left and on the right, are reported in Fig. 2 and 5, respectively.

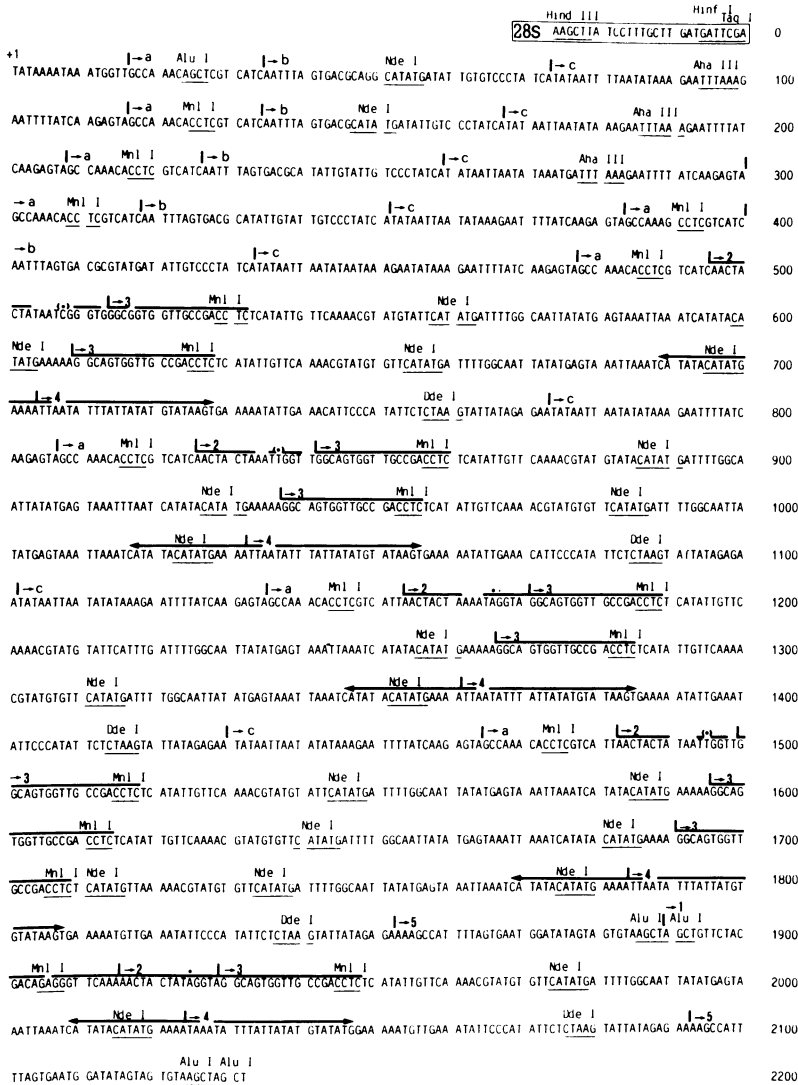


Figure 2. Nucleotide sequence of the '1900' region (position 1-1889) and of the first contiguous 240 bp repeat of clone a56. The last 27 bases of the 28S coding region are also shown, boxed. Small letters above the sequence define homology regions which are schematized in Fig. 3. Bases 1-16 can not be easily assigned to any homology region. Sequences homologous to the region of actual transcription initiation are lined above and a dot indicates the position corresponding to the 5' terminus of the transcript. Converging arrows indicate homologies to a long imperfect inverted repeat present in the 240 bp repeat.

24). From the 3' end of the transcribed region to the leftmost Alu I site in the NTS there is a spacer portion 1900 bp long in clone a56. This portion is not as variable in size as the central portion (see below), and most gene units contain the 1900 bp size class; this region will be termed the '1900' region. Finally there is a conserved NTS region, about 300 bp long, from the rightmost Alu I site in the spacer up to the transcription initiation site. It will be termed the '300' region. This region and every studied 240 bp repeat contain a 42 bp sequence homologous to nucleotides -18/+24 around the transcription initiation site. These homologous sequences are shown as small solid boxes in Fig. 1. The '1900' region together with the first contiguous 240 bp repeat and the '300' region with the last contiguous 240 bp repeat and the ETS of clone a56 are shown in greater detail in Fig. 1B.

Characterization of the '1900' region

The nucleotide sequence of the '1900' region and of the first 240 bp repeat of clone a56 is shown in Fig. 2 and schematically represented in Fig. 3. This region appears to be internally repetitious as well, showing portions of the basic 240 bp repeat intermingled with specific sequences. A specific module of about 100 bp is repeated 5 times from nucleotide 17 to nucleotide 477, whereas portions of this module are found subsequently, alternating with portions of the 240 bp repeat. The first 16 bases of the spacer, just downstream to the transcription termination site, appear unique even if the base-composition of the whole region hinders a precise assignment. The rDNA clone Y22 contains a shorter representative of the '1900' region, 1150 bp in length (9). A restriction fine mapping of the Y22 region is shown in Fig. 3 and compared to the a56 region. The overall organization appears to be the same apart from a probable deletion, in clone Y22, of a central 700 bp portion spanning two Dde I-Dde I modules of 340 bp present in clone a56.

Cloned rDNA gene units and purified uncloned rDNA from a wild-type stock contain more than one size class of the '1900' region (9). Fig. 4 shows the size distribution of this region in genomic DNA of three D. melanogaster stocks. In addition to the prevalent 1900 bp class, 2600 bp and 1150 bp classes are also abundant in agreement with the distribution observed in cloned genes. In addition to a variety of minor length classes one can observe

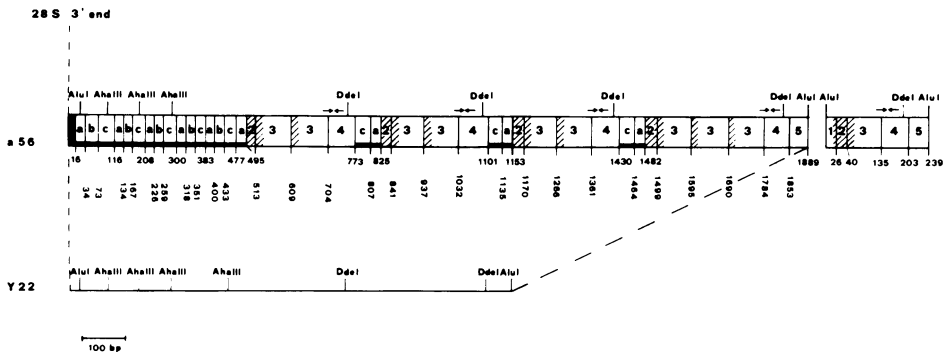


Figure 3. Schematic representation of the nucleotide sequencing data of Fig. 2, and comparison with a restriction map of clone Y22. Homologies to the transcription initiation region are etched; converging arrows indicate the position of the long inverted repeat. Numbers in boxes refer to regions within the 240 bp repeat, whereas small letters refer to specific regions of the 100 bp module repeated 5 times downstream to the transcription termination site. Numbers under the line indicate base positions (1-1889 and 1-239, for the '1900' region and the basic 240 bp repeat, respectively).

a major ± 700 bp pattern and a secondary ± 350 bp pattern. The organization of the '1900' region we studied may account for most features of the observed length class patterns.

Characterization of the '300' region and ETS in different genes

rDNA genes with insertions are present in a large percentage in every studied stock of *D. melanogaster*, both from laboratory cultures and in the wild. These genes are transcribed at a much lower level than the uninterrupted gene units. The region from -234 to +236 around the 5' of the transcribed region does not show significant differences in 4 uninterrupted cloned genes and 4 genes with a type 1 insertion (19). The reported presence of a 42 bp sequence homologous to the promoter region centered around position -272 in the '300' region and in every 240 bp repeat suggested us to extend the comparison between INS+ and INS- genes further upstream, starting from the beginning of the last 240 bp repeat. We also extended the analysis to comprise the ETS in view of the possibility that the primary transcript from interrupted genes might be prematurely degraded, for example owing to an imperfect processing at the 5' end of the 18S rRNA.

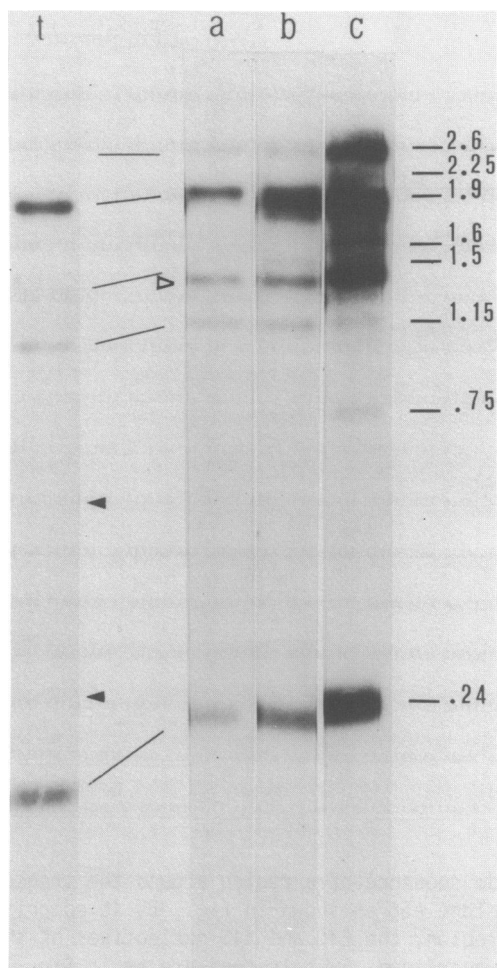


Figure 4. Genomic DNA Southern transfer of 3 stocks of *D. melanogaster* hybridized to a ribosomal spacer probe. DNA extracted from males of stock 1 (a), males of stock 2 (b) and females from wild-type stock Oregon-R (c) has been digested with Alu I. The strong 240 bp band represents the basic 240 bp repeat, whereas longer bands represent different length classes of the '1900' region. A band of about 1.3 Kb (open triangle) contains the '300' region, the ETS and 150 nucleotides of the 18S rRNA coding region. It disappears if the DNA of stock 1 (a) is further digested, e. g. with Taq I (t), originating 2 hybridization bands (solid triangles) 0.6 Kb and 0.3 Kb long.

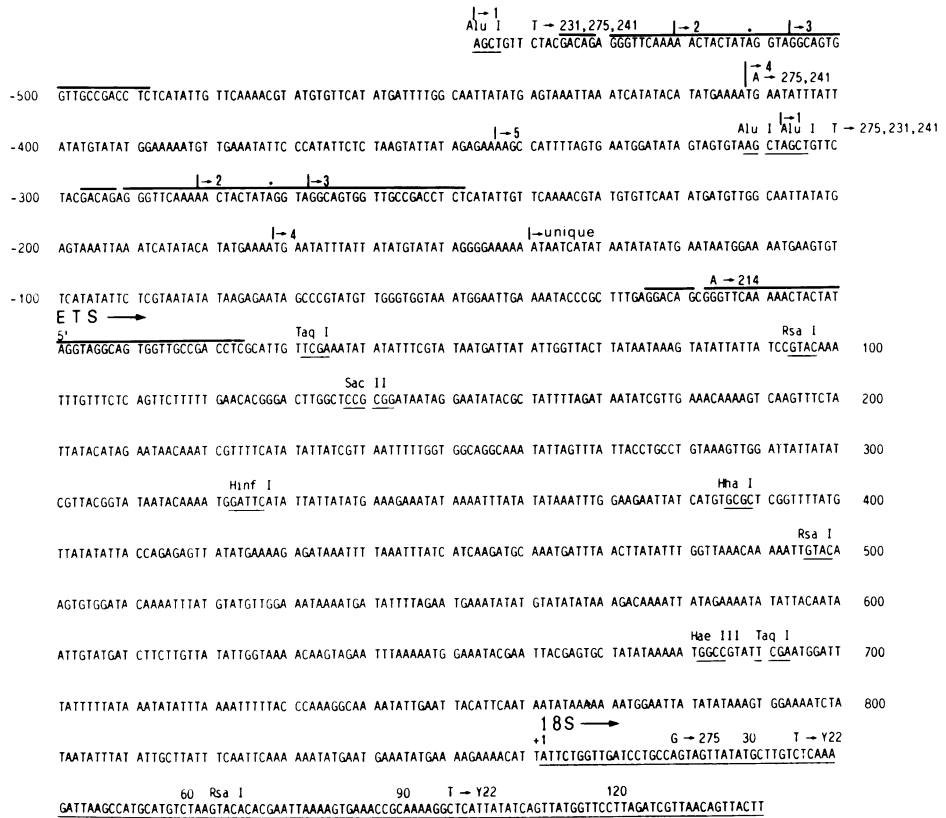


Figure 5. Nucleotide sequence of a region around the transcription initiation site contained in clone a56 as shown in Fig. 1B. It comprises the last 240 bp repeat, the '300' region, the ETS and 119 nucleotides of the 18S rRNA coding region, which are underlined. Symbols used are as in Fig. 2, whereas 'unique' denotes the region of 115 nucleotides unique to the '300' region. A few nucleotide substitutions found in clones indicated are reported above the sequence.

Fig. 5 shows the nucleotide sequence of this region in clone a56 and its comparison to one INS- clone (275) and 3 INS+ clones (214, 241 and 231). It is impossible to know whether this region in clones a56 and Y22 precedes a gene with or without insertion. No significant difference appears either upstream nor 1 Kb downstream from the transcription initiation. Moreover the number of 240 bp repeats is comparable in analyzed INS+ and INS- genes, being 5 in clone 214 and 6 in clones 275 and 241.

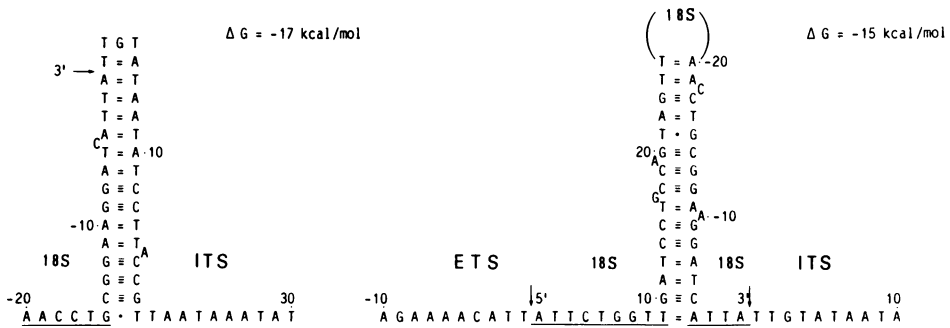


Figure 6. Alternative stem-and-loop structures for the 18S rRNA precursor processing of *D. melanogaster*. Free energy evaluation was according to Cech *et al.* (47). The arrows point to cleavage sites.

We have determined the 5' cleavage site of the 18S rRNA by S1-mapping, 5' ribonucleotide determination of the mature 18S rRNA (30) and primer extension (unpublished results). The 5' cleavage site of the 18S rRNA is CAUU.5'AUU whereas the 3' cleavage site is known to be CAUUA3'.UU (31).

DISCUSSION

The present analysis completes the nucleotide sequence of a ribosomal spacer in *D. melanogaster*, namely clone a56. DNA sequences around the transcription termination site of clone Y22 have been reported by Mandal and Dawid (20), whereas a region around the transcription initiation site from several clones has been investigated by Long *et al.* (19). Several copies of the 240 bp repeated module and the '300' regions of various cloned units have been sequenced as well (21, 22, 23, 24). We report here on the sequence of the '1900' region with the first contiguous 240 bp repeat, the '300' region with the last 240 bp repeat and the ETS through the 5' cleavage site of the 18S rRNA. Present data together with the analysis of several copies of the 240 bp repeat from clone a56 (23) represent a complete ribosomal spacer.

The NTS region

The '1900' region is internally repetitious as anticipated from restriction analysis (22, 23). Portions of the basic 240 bp repeat are repeated in the '1900' region, interspersed with specific sequences derived

from a 100 bp module. The overall organization of the '1900' region can be visualized as consisting of three subregions. Starting from nucleotide 16 downstream from the transcription termination site there is a region containing the specific 100 bp module repeated 5 times up to position 477. The central portion of the '1900' region consists of one or more copies of a 340 bp composite module containing alternating small portions of the 100 bp repeat and of the 240 bp repeat. Finally there is a 3' terminal region where truncated copies of the 240 bp repeat are tandemly arranged.

The imperfect copies of the 240 bp repeat present in the '1900' region appear as 'frozen' ancient duplications, unable to undergo the processes responsible for the redundancy variation typical of the intact repeat. Within these imperfect copies, several truncated versions of the 42 bp homology to the Pol I initiation site are present. The region -9/+24, including the nucleotide corresponding to the 5' terminus of the transcribed region, is repeated four times in the '1900' region. In all four instances the sequence CTACTATAGGTAGG is interrupted by an insertion of 2-3 bases upstream to the conserved T at position -1. This insertion individuates the promoter duplications present in the '1900' region as compared to any other occurrence in intact 240 bp repeats.

The reported structure accounts well for the observed pattern of size variation displayed by the '1900' region in various gene units. It is notable that also the 5' region of the human NTS shows size classes increasing in discrete steps of about 700 bp (32, 33). A repetitive region proximal to the 3' end of the 28S rRNA coding region and consisting of several copies of a 150 bp module, has been reported in the rat NTS (37).

The ribosomal NTS of X. laevis contains repeated 100 bp modules in repetitive region 1 upstream to the first promoter duplication, called Bam Island 1 (34, 35). The repetitive region 1 follows repetitive region 0 consisting of three 34 bp repeated modules, whereas specific sequences extend from the 3' terminus of the ribosomal transcriptional unit up to repetitive region 0. In both X. laevis and D. melanogaster the sequence organization of the spacer region proximal to the transcription termination is different from that of more distal repetitive regions. In yeast the NTS is split in two

portions (NTS1 and NTS2) by the 5S RNA coding region (36).

In the X. laevis ribosomal unit the first Bam Island marks the transition to a repetitious spacer portion containing many duplications of sequences involved in transcription initiation. There are two (sometimes more) promoter duplications, called Bam Island 1 and 2. The duplication corresponds to a region of 149 bp, from -145 to +4, including the transcription initiation site. Between the two Bam Islands and downstream from the second one there are repetitive regions (repetitive region 2 and 3, respectively) consisting of 60/81 bp spacer modules. Each of these modules contains an imperfect copy of a 42 bp sequence that is present from -72 to -114 in the gene promoter. A role of polymerase I enhancer has been proposed for this repeated 42 bp element (38, 39). It is conceivable that the highly conserved 42 bp homology present in every 240 bp repeat and in the '300' region of D. melanogaster ribosomal spacer, represents the structural and possibly the functional homologue of the X. laevis 42 bp element, even if the D. melanogaster element includes the nucleotide homologous to position +1 of the primary transcript.

Transcription initiation and the ETS region

The region of the transcription initiation, from -250 to +980, appears to be almost identical in sequence in one uninterrupted (275), three interrupted gene units (231, 214 and 241) and two unassigned genes (a56 and Y22). These findings confirm and extend previous reports and account for the failure to observe differences in transcription in vitro of genes with and without insertions (40). Taken altogether these data rule out that the biological low activity of the interrupted rDNA genes is due to primary sequence differences in the initiation region and/or that the primary transcript of these genes is unstable owing to differences in the ETS region and 5' end processing of the 18S rRNA.

The external transcribed spacer of D. melanogaster is 861 bp long, with a G+C content of 24% and several ATG triplets (Fig. 5). X. laevis ETS is 712 bp long, has a G+C content of 83% and no ATG (41). Conserved nucleotides at positions -16 (G) and -1 (T) upstream to the transcription initiation site have been reported (42). We observe that G is present at position +12 in mammals (43), X.laevis (35), X. clivii (44), yeast (36), Dictyostelium

discoideum (45) and Physarum polycephalum (46).

We have determined the 5' cleavage site of the 18S rRNA (30). The sequence at the 3' terminus of the 18S rRNA is such that the precursor RNA can form a secondary structure including the cleavage site (31). A secondary structure of comparable stability may in principle form between the 5' and 3' cleavage regions of the 18S rRNA. These alternative structures are shown in Fig. 6. Alternative pathways in the processing of rRNA precursor in D. melanogaster have been reported (4). In the main pathway the 5' cleavage of the 18S rRNA precedes the 3' cleavage, whereas in a secondary pathway both termini are generated in the same step.

ACKNOWLEDGEMENTS

We are grateful to Dr. I. B. Dawid for kindly providing phage clones. We thank Drs. P. P. Di Nocera, J. F. Pulitzer and S. Boast for constructive criticism on the manuscript. This work was supported by Progetto Finalizzato Ingegneria Genetica e Basi Molecolari delle Malattie Ereditarie of the CNR.

REFERENCES

1. Long, E. O. and Dawid, I. B. (1980). *Annu. Rev. Biochem.* 49, 727-764.
2. Perry, R. P. (1976). *Annu. Rev. Biochem.* 45, 605-629.
3. Pavlakis, G. N., Jordan, B. R., Wurst, R. M. and Vournakis, J. N. (1979). *Nucl. Acids Res.* 7, 2213-2238.
4. Long, E. O. and Dawid, I. B. (1980). *J. Mol. Biol.* 138, 873-878.
5. Fedoroff, N. V. (1979). *Cell*, 16, 697-710.
6. Treco, D., Brownell, E. and Arnheim, N. (1982). In *The Cell Nucleus* (Busch, H., ed.), vol. 12, pp. 101-126, Academic Press, New York.
7. Wellauer, P. K., Dawid, I. B., Brown, D. D. and Reeder, R. H. (1976). *J. Mol. Biol.* 105, 461-486.
8. Boncinelli, E., Borghese, A., Graziani, F., La Mantia, G., Manzi, A., Mariani, C. and Simeone, A. (1983). *Mol. Gen. Genet.* 189, 370-374.
9. Long, E. O. and Dawid, I. B. (1979). *Nucl. Acids Res.* 7, 205-215.
10. Dover, G. (1982). *Nature*, 299, 111-117.
11. Wellauer, P. K., Dawid, I. B. and Tartof, K. D. (1978). *Cell*, 14, 269-278.
12. Wellauer, P. K. and Dawid, I. B. (1978). *J. Mol. Biol.* 126, 769-782.
13. Long, E. O. and Dawid, I. B. (1979). *Cell*, 18, 1185-1196.
14. Jolly, D. J. and Thomas, C. A. (1980). *Nucl. Acids Res.* 8, 67-84.
15. Kidd, S. J. and Glover, D. M. (1981). *J. Mol. Biol.* 151, 645-662.
16. Long, E. O., Collins, M., Kiefer, B. I. and Dawid, I. B. (1981). *Mol. Gen. Genet.* 182, 377-384.
17. Jamrich, M. and Miller, O. L., Jr. (1982). *Chromosoma*, 87, 561-570.
18. Jamrich, M. and Miller, O. L., Jr. (1984). *EMBO J.* 3, 1541-1545.

19. Long, E. O., Rebbert, M. L. and Dawid, I. B. (1981). *Proc. Natl. Acad. Sci. USA* 78, 1513-1517.
20. Mandal, R. K. and Dawid, I. B. (1981). *Nucl. Acids Res.* 9, 1801-1811.
21. Kohorn, B. D. and Rae, P. M. M. (1982). *Nucl. Acids Res.* 10, 6879-6886.
22. Coen, E. S. and Dover, G. A. (1982). *Nucl. Acids Res.* 10, 7017-7026.
23. Simeone, A., de Falco, A., Macino, G. and Boncinelli, E. (1982). *Nucl. Acids Res.* 10, 8263-8272.
24. Miller, J. R., Hayward, D. C. and Glover, D. M. (1983). *Nucl. Acids Res.* 11, 11-19.
25. Dawid, I. B., Wellauer, P. K. and Long, E. O. (1978). *J. Mol. Biol.* 126, 749-768.
26. Southern, E. M. (1975). *J. Mol. Biol.* 98, 503-517.
27. Maniatis, T., Jeffrey, A. and Kleid, D. J. (1975). *Proc. Natl. Acad. Sci., U.S.A.* 72, 1184-1188.
28. Maxam, A. M. and Gilbert, W. (1977). *Proc. Natl. Acad. Sci., U.S.A.* 74, 560-564.
29. Smith, H. O. and Birnstiel, M. L. (1976). *Nucl. Acids Res.* 3, 2387-2398.
30. Simeone, A. and Boncinelli, E. (1984). *FEBS Lett.* 167, 249-253.
31. Jordan, B. R., Latil-Damotte, M. and Jourdan, R. (1980). *FEBS Lett.* 117, 227-231.
32. Krystal, M. and Arnheim, M. (1978). *J. Mol. Biol.* 126, 91-104.
33. La Volpe, A., Fianza, V., de Falco, A., Scotto, L., Simeone, A. and Boncinelli, E. (1984). *Bull. Mol. Biol. Med.* 9, 27-34.
34. Boseley, P., Moss, T., Maechler, M., Portmann, R. and Birnstiel, M. (1979). *Cell*, 17, 19-31.
35. Moss, T., Boseley, P. G. and Birnstiel, M. L. (1980). *Nucl. Acids Res.* 8, 467-485.
36. Skryabin, K.G., Eldarov, M.A., Larionov, V.L., Bayev, A.A., Klootwijk, J., de Regt V.C.H.F., Veldman, G.M., Planta, R.J., Georgiev, O.I. and Hadjiolov, A.A. (1984). *Nucl. Acids Res.* 12, 2955-2968.
37. Braga, E. A., Yussifov, T. N. and Nosikov, V. V. (1982). *Gene*, 20, 145-156.
38. Labhart, P. and Reeder, R. H. (1984). *Cell*, 38, 349-351.
39. Reeder, R. H. (1984). *Cell*, 38, 349-351.
40. Kohorn, B. D. and Rae, P. M. M. (1982). *Proc. Natl. Acad. Sci. USA* 79, 1501-1505.
41. Maden, B.E.H., Moss, M. and Salim, M. (1982). *Nucl. Acids Res.* 10, 2387-2398.
42. Sommerville, J. (1984). *Nature*, 310, 189-190.
43. Financsek, I., Mizumoto, K., Mishima, Y. and Muramatsu, M. (1982). *Proc. Natl. Acad. Sci. USA* 79, 3092-3096.
44. Bach, R., Allet, B. and Crippa, M. (1981). *Nucl. Acids Res.* 9, 5311-5330.
45. Hoshikawa, Y., Iida, Y. and Iwabuchi, M. (1983). *Nucl. Acids Res.* 11, 1725-1734.
46. Blum, B., Seebeck, T., Braun, R., Ferris, P. and Vogt, V. (1983). *Nucl. Acids Res.* 11, 8519-8533.
47. Cech, T. R., Tanner, N. K., Tinoco, I. Jr., Weir, B. R., Zucker, M. and Perlman, P. S. (1983). *Proc. Natl. Acad. Sci. USA* 80, 3903-3907.