Proc. Natl. Acad. Sci. USA
Vol. 78, No. 3, pp. 1513–1517, March 1981
Biochemistry

Nucleotide sequence of the initiation site for ribosomal RNA transcription in *Drosophila melanogaster*: Comparison of genes with and without insertions

(rRNA precursor/ribosomal DNA/rDNA insertion/DNA sequence determination/RNA sequence determination)

ERIC O. LONG, MARTHA L. REBBERT, AND IGOR B. DAWID

Laboratory of Biochemistry, National Cancer Institute, National Institutes of Health, Bethesda, Maryland 20205

Communicated by Joseph G. Gall, November 24, 1980

ABSTRACT **The sequence of 470 nucleotides surrounding the initiation site for rRNA transcription in *Drosophila melanogaster* has been determined. The precise initiation site was determined first by measuring the DNA fragment protected by the rRNA precursor against digestion by the single-strand specific nuclease S1 and second by direct sequence determination of the first 13 nucleotides of the rRNA precursor. Because >80% of rRNA precursor molecules have been shown previously to bear pppA or ppA 5′ termini, we assume that they represent the primary transcription product. Short sequence homologies exist with the initiation regions for rRNA transcription of *Xenopus laevis* and *Saccharomyces cerevisiae*. We have determined the nucleotide sequence of the initiation region in four cloned ribosomal genes from *D. melanogaster* which are not interrupted and in four cloned ribosomal genes in which the 28S rRNA coding region is interrupted by a 5-kilobase type 1 insertion. Three uninterrupted genes and three interrupted genes have identical sequences in the entire analyzed region. The remaining two genes have a single identical base substitution at position −17. We have shown previously that interrupted ribosomal genes in *D. melanogaster* are not effectively transcribed. Because the nucleotide sequences of the region where transcription initiates are identical in genes with or without insertions, we postulate that the presence of the insertion itself may be responsible for the inactivity of the interrupted genes.**

The genes for rRNA occur as tandemly repeated units in *Drosophila melanogaster* as in many other organisms (1). A precursor rRNA molecule is synthesized and then processed to mature rRNAs, which include 5.8S, 18S, and 28S RNA molecules (2, 3). The precursor rRNA molecule in *D. melanogaster* is 8 kilobases (kb) long and its coding region has been mapped in the rDNA by electron microscopy and restriction enzyme analysis (4, 5). Levis and Penman (6) have shown that this 8-kb rRNA molecule carries pppA or ppA at more than 80% of its 5′ termini. Therefore, one may confidently assume that the 8-kb RNA is the primary transcription product of the ribosomal genes.

In *Drosophila* and other dipteran insects, but not in most other animals studied, a large fraction of the rRNA genes are interrupted in the 28S RNA coding region by noncoding DNA segments which have been called "ribosomal insertions" (1). In *D. melanogaster*, these insertions occur in two distinct sequence types, each represented by several size classes (7, 8). Active rRNA genes visualized by electron microscopy show dense RNA polymerase packing (9). If interrupted rRNA genes were transcribed at a normal rate, there would be thousands of nascent rRNA chains containing insertion sequences in each cell nucleus. However, we found that RNA molecules encoded by the major type 1 and type 2 insertion sequences were present at a frequency of less than one copy per nucleus. From this observation we concluded that interrupted rRNA genes do not contribute significantly, if at all, to rRNA production (10, 11).

The present work had two aims. First, we wanted to determine the precise initiation site of rRNA transcription in *D. melanogaster* and to compare the nucleotide sequence of the initiation region with the known analogous sequence in *Xenopus laevis* and yeast ribosomal DNA (rDNA) (12–14). Second, we wished to compare the sequences in the initiation region of interrupted and uninterrupted genes. Because the former are transcribed at a much lower level than the latter, we wondered whether the primary DNA sequence might be different in the region of transcription initiation.

METHODS

DNA Isolation and Sequence Determination. A library of *D. melanogaster* Oregon R DNA, randomly sheared and inserted into λSep6 by poly(dA)·poly(dT) tails was generously provided by L. Prestidge and D. S. Hogness. Phages containing rDNA were selected (15) and characterized by restriction mapping, using the known maps of rDNA as a guide (4, 16). R-Loop mapping (17) was done in some cases, and heteroduplex molecules between different phages were prepared and analyzed as described (18). The cloned *Eco*RI fragment containing a continuous rDNA repeat, pDmr Y22, has been described (4). DNA sequence was determined according to Maxam and Gilbert (19) with restriction fragments labeled at either a 5′ end or 3′ end. 5′-End-labeling was achieved as described (19) by treatment with bacterial alkaline phosphatase (Worthington, further purified by DEAE-cellulose chromatography, a kind gift of B. Sollner-Webb), removal of the enzyme by two phenol extractions, and incubation with [γ-^{32}P]ATP and T4 polynucleotide kinase (Boehringer Mannheim). The 5′ end to be sequenced was isolated by a secondary restriction enzyme digestion and electrophoresis in a polyacrylamide gel (19). Labeling of *Taq* I restriction sites at their 3′ end was achieved by adding the first missing deoxyribonucleotide to the recessed strand with reverse transcriptase. In a final volume of 50 μl, 5 pmol of 3′ ends was incubated with 36 units of avian myeloblastosis virus reverse transcriptase (kindly provided by J. W. Beard, Life Sciences, St. Petersburg, FL) and 250 pmol of [α-^{32}P]dCTP (2000 Ci/mmol; 1 Ci = 3.7×10^{10} becquerels; Amersham or New England Nuclear) in 10 mM Tris·HCl, pH 8.0/10 mM $MgCl_2$/50 mM KCl/1 mM dithiothreitol at 37°C for 1 hr. The reverse transcriptase and the unincorporated nucleotides were removed by two phenol extractions and chromatography over a Sephadex

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U. S. C. §1734 solely to indicate this fact.

Abbreviations: rDNA, ribosomal DNA; kb, kilobase(s).

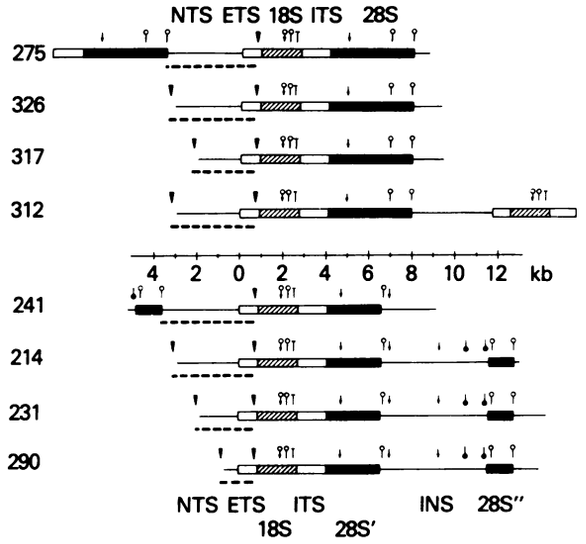

FIG. 1. Maps of *D. melanogaster* rRNA genes cloned in the λSep6 vector. The top four recombinant molecules, λDmr275, -326, -317, and -312, represent uninterrupted ribosomal genes. The bottom four recombinant molecules represent ribosomal genes interrupted in the sequence representing 28S rRNA by a type 1 insertion of 5 kb. The regions representing the nontranscribed spacer (NTS), the external transcribed spacer (ETS), 18S rRNA, the internal transcribed spacer (ITS), 28S rRNA, and the insertion (INS) are indicated above and below the gene maps. These maps have been obtained by a combination of restriction endonuclease mapping and heteroduplex and R-loop analysis in the electron microscope. The right arm of the λSep6 vector is at the right in all these maps. The sites for the following restriction endonucleases are indicated: *Eco*RI ⊤; *Hin*dIII, ♀; *Sma* I, ↓; *Bam*HI, ♁. Only the *Hae* III sites flanking the NTS are shown (▼). Spacer fragments that were used in the nucleotide sequencing analysis were obtained from a *Hin*dIII/*Hae* III double digest of the entire λDmr recombinant molecule and are indicated by a dashed line under each map. In the cases in which the NTS was not intact, *Hae* III fragments that contained a short segment of the λSep6 left arm were isolated.

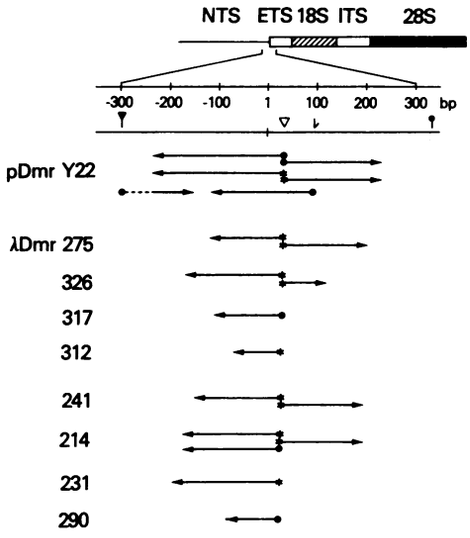

FIG. 2. Strategy for determining the sequence of the transcription initiation region. The initiation region is shown enlarged under the map of an entire ribosomal gene. The sites for the following restriction endonucleases are indicated: *Alu* I, ⍒; *Taq* I, ▽; *Rsa* I, ↓; and *Hin*fI, ⍑. Base pairs (bp) are numbered in both directions, starting from the first transcribed nucleotide. Restriction fragments labeled at their 5′ end are indicated by ● and those labeled at their 3′ end are indicated by *; the extent of the analyzed segment is indicated by the length of the arrow. The plasmid pDmr Y22 contains a single rDNA repeating unit cleaved by *Eco*RI (see ref. 4).

G-50 column. Up to 75% of the 3′ ends were labeled. This end-labeling procedure was also very specific. *Taq* I/*Hae* III restriction fragments have been labeled in this way and subjected to sequence determination without further purification because all the label was present at the *Taq* I end. Electrophoresis of chemically treated fragments was in 0.4-mm-thick 20%, 14%, and 8% polyacrylamide gels (20).

Nuclease S1 Mapping. About 10^5 cpm of a restriction fragment labeled at the 5′ end of the coding strand (3×10^6 cpm/pmol) was hybridized at 41°C overnight with 6 μg of nuclear RNA enriched for 8-kb rRNA precursor in 250 μl of 80% formamide/0.3 M NaCl/40 mM Tris·HCl, pH 8.0/1 mM EDTA with 3 μg of denatured sheared salmon DNA and 3 μg of intact pBR322 DNA as carriers. Five 50-μl aliquots were diluted into 550 μl of S1-buffer (0.3 M NaCl/50 mM sodium acetate, pH 4.8/0.5 mM $ZnSO_4$) and treated with up to 25,000 units of nuclease S1 (Miles), which corresponds to about 830 units according to Vogt (21), for 1 hr at 37°C. The digestion was stopped by addition of EDTA to 5 mM, $NaDodSO_4$ to 0.1%, and 5 μg of *Escherichia coli* tRNA. After two extractions with phenol and one ethanol precipitation, 100–500 cpm per sample was loaded on sequencing polyacrylamide gels.

RNA Isolation and Sequence Determination. Nuclear RNA from *D. melanogaster* embryos was prepared as described (10). rRNA precursor molecules were enriched by centrifugation through sucrose gradients and labeled at their 5′ end as described for DNA fragments. The 5′-terminal rRNA fragments were isolated by hybridization to nitrocellulose-bound *Taq* I spacer fragment in 0.75 M NaCl/50 mM sodium phosphate, pH 6.8/5 mM EDTA/Denhardt's solution (22)/0.5% $NaDodSO_4$ containing 0.5 mg of unlabeled cytoplasmic RNA per ml, at 60°C for 22 hr. The filter was washed extensively as described (10), treated with pancreatic RNase (0.4 μg/ml) in 0.3 M NaCl/30 mM sodium citrate at room temperature for 20 min, and washed again extensively. The hybridized RNA was eluted by incubation at 100°C for 90 sec in 5 mM Tris·HCl, pH 7.5/1 mM EDTA with 25 μg of *E. coli* tRNA. The 38-base-long end-labeled RNA fragment was electrophoresed in an 8% polyacrylamide/8 M urea gel, extracted, and analyzed for sequence by the method of Donis-Keller *et al.* (23) and Lockard *et al.* (24). One modification was that the buffer for U2 RNase was diluted 1:1 (10 mM sodium citrate, pH 5.0/3.5 M urea/0.5 mM EDTA containing 0.25 mg of tRNA per ml) because 1 vol of RNase U2 was added. This caused the U2 RNase to cleave some of the GpN bonds also. The samples were electrophoresed in a formamide/20% polyacrylamide sequencing gel (12) that had been pre-electrophoresed until the bromophenol blue dye reached the bottom of the gel (R. Lockard, personal communication).

Complete nuclease P1 digests were electrophoresed on Whatman 3 MM paper in parallel with the four pN nucleotide markers. Complete T2, T1, and pancreatic RNase digests were electrophoresed on Whatman DE-81 paper at pH 3.5 and pH 1.7 (25).

RESULTS

We wanted to compare the nucleotide sequence of the transcription initiation region of ribosomal genes with or without insertions. rDNA molecules were isolated from a library of *D. melanogaster* DNA linked to the vector λSep6; the structures of eight such molecules used in this study are shown in Fig. 1. As a starting point for the sequence determination, we used another recombinant molecule carried by the plasmid pMB9.

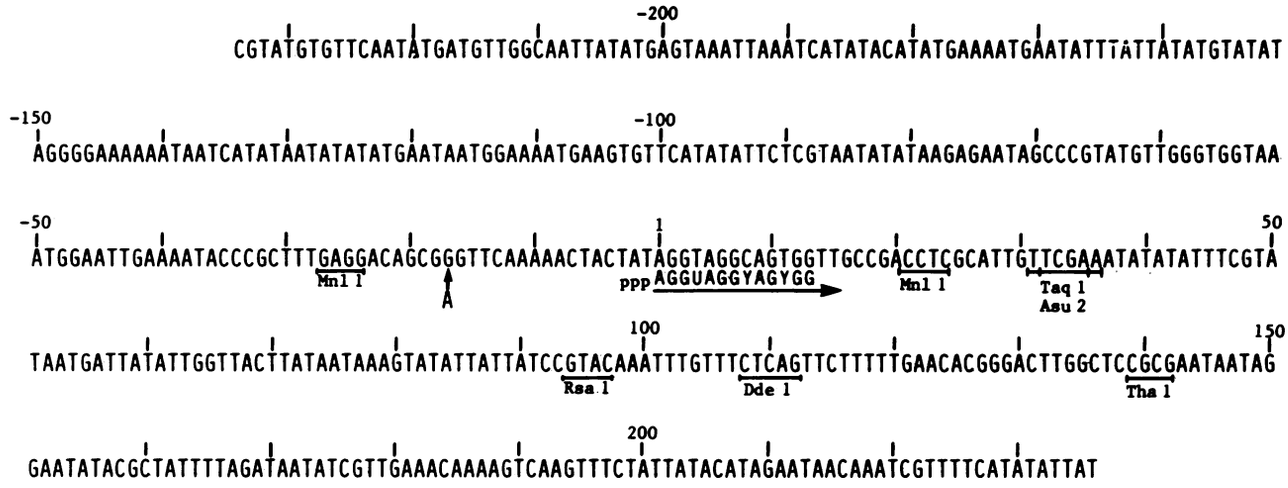

FIG. 3. Nucleotide sequence of the rRNA transcription initiation region. The nucleotide sequence of the noncoding strand determined from the plasmid pDmr Y22 is shown. The nucleotide sequence determined at the 5′ end of the 38S rRNA precursor is shown under the corresponding DNA sequence, with an arrow indicating the direction of transcription. All restriction sites found by a computer search (26) are shown except those for *Eco*RI*. The positions of the sites for *Taq* I, *Rsa* I, and *Tha* I, but not for *Mnl* I, *Asu* 2, and *Dde* I, have been determined experimentally. The nucleotide sequence determined in the recombinant molecules λDmr275, -326, -317, -241, -231, and -290 (see Fig. 2) was identical to the sequence shown here. The two recombinant molecules λDmr312 and -214 contained a single base change at position −17 (A instead of G, shown by vertical arrow). The only ambiguity in the sequence is at position 144.

This clone, pDmr Y22, contains an 11-kb *Eco*RI fragment derived from two adjacent ribosomal genes (4). The structure of this *Eco*RI fragment corresponds to the *Eco*RI fragment derived from λDmr312 (Fig. 1). It is not known whether the initiation region on this rDNA fragment is derived from a gene with or without insertion. More than 230 base pairs were determined in both directions from the initiation site (Fig. 2 and 3).

In order to map the site of transcription initiation, the *Taq* I restriction fragment (see Fig. 2) was labeled at the 5′ end of the coding strand, hybridized to precursor rRNA, and digested with single strand-specific nuclease S1. The protected fragment was electrophoresed in a sequencing gel in parallel with the same *Taq* I fragment cleaved by base-specific reactions (Fig. 4). It was necessary to use high concentrations of nuclease S1 to reduce the short ladder of partial digestion products to a single prominent band. The shortest visible band probably represents "nibbling" of the first paired nucleotides by nuclease S1 because a triphosphorylated 5′ nucleotide forms a weak base pair. When a correction is made for the different migration of nuclease S1-generated and chemically cleaved fragments (see legend), the first transcribed nucleotide is found to be T, in agreement with the 5′ terminal A of the rRNA precursor (6). This experiment mapped the first transcribed nucleotide 33 bases upstream from the *Taq* I site.

The position of transcription initiation was also determined by direct sequence determination of the 5′ terminus of the 8-kb rRNA precursor. The enzymatically labeled 5′-terminal oligonucleotide was sequenced (Fig. 5). There is some ambiguity in the determination of the first three nucleotides because of two weak bands that were nearly undetectable in the ladder and a band in the T1 lane that probably represents the slowly migrating 2′,3′-cyclic phosphate intermediate (25). To resolve this ambiguity, the 5′-terminal sequence was determined independently by paper electrophoresis of digestion products of the terminally labeled oligonucleotide (25). A complete RNase P1 digest was clearly identified as pA. Complete T2, T1, and pancreatic RNase digests were electrophoresed on DEAE-paper at both pH 1.7 and pH 3.5 and could be clearly identified as pAp, pApGp, and pApGpGpUp, respectively. Therefore, the 5′-terminal sequence of *D. melanogaster* precursor rRNA is 5′ A-G-G-U-A-G-G-Y-A-G-Y-G-G (Y represents a pyrimidine). This sequence matches perfectly the first 13 transcribed nucleotides determined by DNA sequencing and nuclease S1 protection.

The spacers of *D. melanogaster* rDNA are very A+T-rich and contain very few restriction sites (ref. 5; Fig. 3). The overall G+C content of the initiation region is also low, about 20%, except for four discrete G+C-rich regions around positions −64, −25, +14, and +133 (Fig. 6). A large number of repeats and dyad symmetries has been found in the sequence by using a computer program (26). These features have no known significance and are not shown.

The sequence of the initiation region has been determined in four uninterrupted rRNA genes and in four genes carrying a 5-kb type 1 insertion (see Fig. 1). The extent of the regions sequenced is shown in Fig. 2. The sequence in three uninterrupted genes and in three interrupted genes was identical to the sequence of pDmr Y22 (Fig. 3). One uninterrupted and one interrupted gene had a single identical base change at position −17, a G replaced by an A in both cases.

DISCUSSION

The identification of the initiation site for rRNA transcription in *D. melanogaster* is based on the following data. First, >80% of rRNA precursor molecules carry either 5′ pppA or 5′ ppA termini (6). Second, nuclease S1 mapping with a restriction fragment labeled at the 5′ end of the coding strand and 8-kb precursor rRNA showed that the first protected nucleotide is a T located 33 bases upstream from the *Taq* I restriction site. Third, direct sequence determination on 5′ end-labeled rRNA precursor identified 13 nucleotides that matched the initiation sequence determined by DNA sequencing and nuclease S1 protection.

The sequence of about 235 nucleotides in either direction from the initiation site has been determined. From about 80 nucleotides preceding the initiation site and further upstream the sequence is A+T-rich and most likely extends into the repetitive portion of the nontranscribed spacer. We have previ-

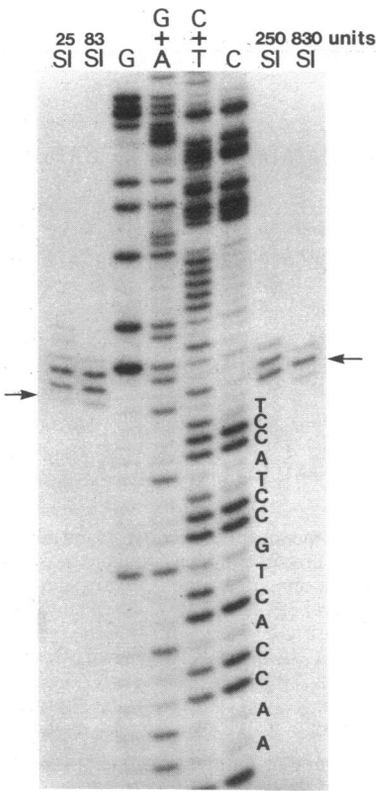

FIG. 4. Determination of the transcription initiation site by nuclease S1 protection. The *Taq* I–*Alu* I fragment from the plasmid pDmr Y22 (see Fig. 2), labeled at the 5′ end of the coding strand, was hybridized to the 8-kb rRNA precursor molecule. Aliquots were digested with increasing amounts of nuclease S1 and electrophoresed in a 20% polyacrylamide gel in parallel with the same end-labeled *Taq* I–*Alu* I fragment cleaved by base-specific chemical reactions. The arrows point to the band that was prominent at the highest nuclease S1 concentration. A shift in migration by 1.5 nucleotides is expected between the chemically cleaved and the nuclease S1-protected DNA strands because the sequencing reaction eliminates the modified nucleotide and leaves a 3′ phosphate group (12, 27).

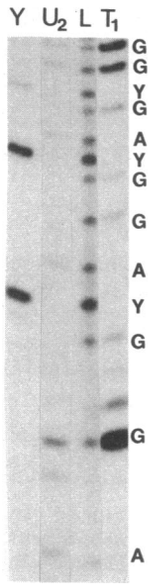

FIG. 5. Nucleotide sequence at the 5′ terminus of the 8-kb rRNA precursor. The 8-kb rRNA precursor was end-labeled at the 5′ end with polynucleotide kinase. The purified 5′-terminal oligonucleotide was digested with several concentrations of U2, T1, and pancreatic (Y) RNases. The ladder (L) was obtained by incubating an aliquot in 90% formamide at 100°C for 4 hr. All samples were electrophoresed in a formamide/20% polyacrylamide gel. The figure represents a montage of the relevant lanes from the gel. The sequence of the first four nucleotides was determined independently as A-G-G-U (see text).

ously identified repeated 250-base-pair elements in the spacer, defined by regular *Alu* I restriction sites, and have shown by electron microscopy that the repeated sequence begins within less than 150 base pairs from the initiation site (5). The region downstream from the initiation site is also rich in A+T. Four blocks of G+C-rich segments occur around the initiation site.

Comparison of our sequence with the sequence determined for the region of transcription initiation in *Xenopus laevis* and yeast rDNA (12–14) shows little homology (Fig. 7). By computer analysis (26), several other homologous sequences up to 9 nucleotides long were found located in completely different relative positions; they are not statistically significant. In the case of genes transcribed by RNA polymerase II, widely conserved sequences at a constant distance from the cap site (presumed initiation site) have been found and are an essential requirement for proper transcription (28). rDNA sequences from other species and functional assays will be necessary to determine whether a similar conserved sequence is required for initiation by RNA polymerase I.

Comparison of the sequences of different rDNA molecules from *D. melanogaster* showed that the initiation region has a conserved sequence. In particular, cloned fragments derived from the X chromosome [genes with 5-kb type 1 insertions are present only in the X chromosome (7)] and the Y chromosome [the clone pDmr Y22 carries a gene derived from the Y chromosome (4)] have identical sequences. More interesting was the finding that three cloned rDNA fragments with uninterrupted 28S RNA coding regions and three fragments carrying 5-kb insertions had identical sequences. Two cloned rDNA fragments, one interrupted and one uninterrupted, had a single nucleotide substitution at position −17, in both cases A replacing G. The fact that two independent clones with the identical substitution have been found shows that this particular base pair is polymorphic in *D. melanogaster* rDNA, unless both changes resulted from a mutation in *E. coli*.

The findings reported here rule out the possibility that the biological inactivity of the interrupted rDNA repeats is due to a modified initiation region that would not allow proper recognition by RNA polymerase I. The difference between interrupted and uninterrupted rDNA repeats with respect to *in vivo* transcription might therefore be explained by the presence of the insertion itself. Three possible explanations of the phenomenon can be proposed.

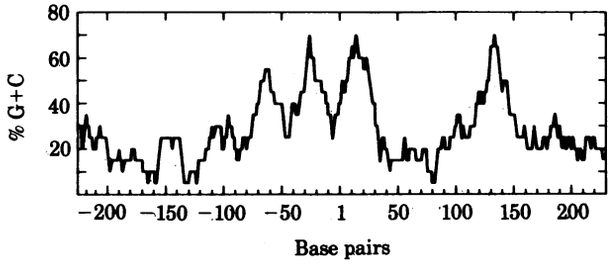

FIG. 6. G+C content in the DNA sequence of the rRNA transcription initiation region. Each point on the graph represents the percentage of G+C pairs in a sequence of 20 base pairs around each nucleotide.

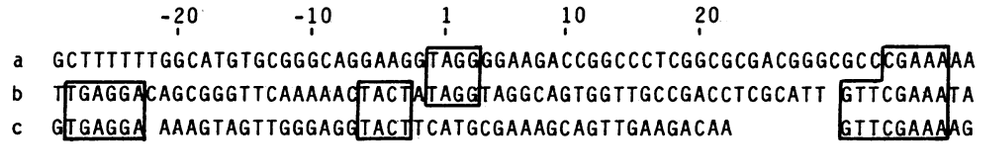

FIG. 7. Comparison of the DNA sequences at the initiation site for rRNA transcription in *Xenopus laevis* (*a*) (12), *D. melanogaster* (*b*), and *Saccharomyces cerevisiae* (*c*) (13, 14).The sequence of the noncoding strand is shown. The sequences have been aligned at the site of transcription initiation (position 1) and at the regions of homology, shown by boxes. Shifts were allowed to maximize homology.

(*i*) The insertion could act as a transcription terminator. We have discussed this possibility (10) and consider it unlikely. No prematurely terminated precursor rRNA molecules of the predicted size could be found. Our interpretation that interrupted rRNA genes are transcriptionally inactive is supported by the observation in the electron microscope of long silent DNA stretches interspersed with active rRNA genes in *D. melanogaster* (9) and *D. hydei* (29).

(*ii*) There could be a control region for transcription outside of the region sequenced. An internal control region has been demonstrated for the *Xenopus* 5S rRNA gene, located 50–80 bases downstream from the initiation site (30, 31). Our sequence extends more than 200 bases from the initiation site but the possibility remains of a control element located further in the gene, even at the point where insertions occur, 6.8 kb from the initiation site.

(*iii*) The presence of an insertion could disrupt a particular chromatin structure necessary for initiation or elongation of transcription. Such an effect would have to be polar (upstream) because interrupted and uninterrupted rDNA repeats are interspersed (18). The possibility of a structural effect of a DNA region on the activity of a potential initiation site at a considerable distance upstream has also been raised in the case of immunoglobulin genes (32). We have tested whether the chromatin structure might be different in ribosomal insertion DNA sequences and rRNA coding sequences by measuring their DNase I sensitivity in nuclei of *D. melanogaster* cultured cells. No substantial difference was observed in the DNase I sensitivity of rRNA coding sequences, nontranscribed spacer sequences, and type 1 or type 2 insertion sequences (C. Wu, personal communication; our unpublished data). It is still possible, however, that a structural difference exists but cannot be detected by this assay.

Besides the mechanism by which insertions prevent expression of rRNA genes, other questions remain unanswered. What is the function, if any, of these insertion elements, and why have interrupted nonfunctional genes been conserved? We tend to think that these insertion elements exist for reasons unrelated to rRNA synthesis and that they are tolerated because a sufficient number of uninterrupted genes remains in *Drosophila*. Excision and integration (conservative transposition) of these elements within the rRNA genes could explain why they were not eliminated by unequal crossing over and also why the nucleotide sequence of interrupted genes has not diverged from uninterrupted genes.

We are grateful to D. S. Hogness and L. Prestidge for providing the *D. melanogaster* DNA library. We thank R. Lockard, D. Schümperli, B. Sollner-Webb, and especially M. Rosenberg for advice on sequence determination techniques, C. Queen for carrying out the computer analysis, and D. Schümperli and M. Singer for useful comments on the manuscript. E.O.L. was supported by a fellowship from the Swiss National Fund for Scientific Research.

1. Long, E. O. & Dawid, I. B. (1980) *Annu. Rev. Biochem.* **49**, 727–764.
2. Pavlakis, G. N., Jordan, B. R., Wurst, R. M. & Vournakis, J. N. (1979) *Nucleic Acids Res.* **7**, 2213–2238.
3. Long, E. O. & Dawid, I. B. (1980) *J. Mol. Biol.* **138**, 873–878.
4. Dawid, I. B., Wellauer, P. K. & Long, E. O. (1978) *J. Mol. Biol.* **126**, 749–768.
5. Long, E. O. & Dawid, I. B. (1979) *Nucleic Acids Res.* **7**, 205–215.
6. Levis, R. & Penman, S. (1978) *J. Mol. Biol.* **121**, 219–238.
7. Wellauer, P. K., Dawid, I. B. & Tartof, K. D. (1978) *Cell* **14**, 269–278.
8. Wellauer, P. K. & Dawid, I. B. (1978) *J. Mol. Biol.* **126**, 769–782.
9. McKnight, S. L. & Miller, O. L., Jr. (1976) *Cell* **8**, 305–319.
10. Long, E. O. & Dawid, I. B. (1979) *Cell* **18**, 1185–1196.
11. Long, E. O., Rebbert, M. L. & Dawid, I. B. (1981) *Cold Spring Harbor Symp. Quant. Biol.* **45**, in press.
12. Sollner-Webb, B. & Reeder, R. H. (1979) *Cell* **18**, 485–499.
13. Klemenz, R. & Geiduschek, E. P. (1980) *Nucleic Acids Res.* **8**, 2679–2689.
14. Valenzuela, P., Bell, G. I., Venegas, A., Sewell, E. T., Masiarz, F. R., DeGennaro, L. J., Weinberg, F. & Rutter, W. J. (1977) *J. Biol. Chem.* **252**, 8126–8135.
15. Benton, W. D. & Davis, R. W. (1977) *Science* **196**, 180–182.
16. Roiha, H. & Glover, D. M. (1980) *J. Mol. Biol.* **140**, 341–355.
17. White, R. L. & Hogness, D. S. (1977) *Cell* **10**, 177–192.
18. Wellauer, P. K. & Dawid, I. B. (1977) *Cell* **10**, 193–212.
19. Maxam, A. M. & Gilbert, W. (1980) *Methods Enzymol.* **65**, 499–560.
20. Sanger, F. & Coulson, A. R. (1978) *FEBS Lett.* **87**, 107–110.
21. Vogt, V. M. (1973) *Eur. J. Biochem.* **33**, 192–200.
22. Denhardt, D. T. (1966) *Biochem. Biophys. Res. Commun.* **23**, 641–646.
23. Donis-Keller, H., Maxam, A. M. & Gilbert, W. (1977) *Nucleic Acids Res.* **4**, 2527–2538.
24. Lockard, R. E., Alzner-Deweerd, B., Heckman, J. E., MacGee, J., Tabor, M. W. & RajBhandary, U. L. (1978) *Nucleic Acids Res.* **5**, 37–56.
25. Brownlee, G. G. (1972) in *Laboratory Techniques in Biochemistry and Molecular Biology*, eds. Work, T. S. & Work, E. (North-Holland, Amsterdam).
26. Queen, C. & Korn, L. J. (1980) *Methods Enzymol.* **65**, 595–609.
27. Tomizawa, J., Ohmori, H. & Bird, R. (1977) *Proc. Natl. Acad. Sci. USA* **74**, 1865–1869.
28. Corden, J., Wasylyk, B., Buchwalder, A., Sassone-Corsi, P., Kedinger, C. & Chambon, P. (1980) *Science* **209**, 1406–1414.
29. Glätzer, K. H. (1979) *Chromosoma* **75**, 161–175.
30. Sakonju, S., Bogenhagen, D. F. & Brown, D. D. (1980) *Cell* **19**, 13–25.
31. Bogenhagen, D. F., Sakonju, S. & Brown, D. D. (1980) *Cell* **19**, 27–35.
32. Perry, R. P., Kelley, D. E., Coleclough, C., Seidman, J. G., Leder, P., Tonegawa, S., Matthyssens, G. & Weigert, M. (1980) *Proc. Natl. Acad. Sci. USA* **77**, 1937–1941.