Transcription of the 'non-transcribed' spacer of Drosophila melanogaster rDNA

J.Ross Miller*, David C. Hayward and David M.Glover +

Cancer Research Campaign, Eukaryotic Molecular Genetics Research Group, Department of Biochemistry, Imperial College of Science and Technology, London SW7 2AZ, UK

Received 17 November 1982; Revised and Accepted 9 December 1982

ABSTRACT
We have detected a set of transcripts in Drosophila melanogaster cells which are homologous to repeating elements within the 'non-transcribed' spacer region of rDNA. The RNA molecules range from 240 to 1680 nucleotides, differing in length by an integral value of 240 nucleotides. We have sequenced several AluI fragments which characterise the main 240 nucleotide repeating element. We find that each of these fragments contain a segment of approximately 50 nucleotides, which is homologous to the transcription initiation site for pre-rRNA.

INTRODUCTION
The ribosomal genes of the higher eukaryotes typically occur in tandem arrays in which the transcription units are separated from each other by so called 'non-transcribed' spacer sequences. These spacers usually differ significantly from the main transcription unit in nucleotide composition and so could be readily visualised in partial denaturation mapping experiments in the electron microscope using purified rDNA (1). The spacers can also be visualised by the examination of chromatin where they are seen as quiescent regions separating actively transcribed regions on which the nascent primary rRNA transcript can be readily discerned. The primary transcripts are subsequently processed by endonucleolytic cleavage to generate mature rRNAs. The organisation of spacer elements has been most extensively studied in Xenopus laevis for which it was shown using cloned rDNA that the spacer elements were heterogeneous in length (2). Heteroduplex experiments carried out with these cloned spacers indicated that the length heterogeneity was a result of varying numbers of tandemly repetitive elements within the spacer (3). Subsequent nucleotide sequence analysis showed that there were three such repetitive regions, separated by non-repetitive areas containing a BamHI restriction site termed 'Bam islands' (4, 5). The repetitive elements in the spacer are related to the 'Bam island' sequence and in turn related to the sequence at the initiation site for transcription. Restriction mapping...
of the non-transcribed spacer of cloned DNAs of Drosophila melanogaster showed that these too contained a repetitive element characterised by multiple AluI sites spaced at regular intervals of 240 nucleotide pairs (6).

In this report we show by nucleic acid sequencing that these repetitive elements share sequence homology with sequences at the site for the initiation of rDNA transcription of the primary transcription unit (7). We also show that the repetitive elements hybridise to a 'ladder' of transcripts differing in length by 240 nucleotides, the length of the repeating unit.

MATERIALS AND METHODS

DNA Sequencing: The gel purified AluI fragments were cloned into the Smal site of M13 mp9, and sequenced as described by Sanger et al. (8) using the modifications of Schreier and Cortese (9). Computer analysis was carried out using the SEQ program (10).

RNA Gel Transfer Hybridisation: Denaturing gels were run in formaldehyde (11). Following electrophoresis the RNA was transferred to nitrocellulose (12) for hybridisation with restriction fragments labelled in vitro by nick-translation.

RESULTS AND DISCUSSION

The primary rDNA transcript in Drosophila melanogaster is an 8kb RNA molecule which as a result of endonucleolytic processing generates the mature 28S, 18S, 5.8S and 2S rRNA. These molecules are transcribed from uninterrupted rDNA units. There are, however, a high proportion of rDNA units which contain insertion sequences in their 28S rRNA genes and which are transcribed at very low levels (13). rDNA containing the major type I insertion are incapable of producing functional rRNA since they have a deletion of part of the 28S gene flanking the insertion (14). Transcripts homologous to part of the insertion can however be detected in the cytoplasm. Full length transcripts of the type II insertions on the other hand can be detected in the nucleus, but not in the cytoplasm (15). Some of these type II transcripts are linked to the proximal part of the rDNA transcription unit. It is not clear whether there is rapid processing of these insertion sequences from primary transcripts or whether transcription simply terminates at a polyA tract on the right hand junction of the insertion with the 28S gene. In the course of our investigations of these transcripts we systematically surveyed entire rDNA repeating units looking for homology to RNA. We were surprised to find that a HindIII/HaeIII
Figure 1. Restriction Map of the Spacer Region of the rDNA Clone Dm103.

(a) The 3' of the 28S gene is designated by the shaded block and the external transcribed spacer of the following transcription unit by the open block. Restriction sites are as follows: H, HindIII; A, AluI; T, TaqI; h, HaeIII.

(b) The segment corresponding to that sequenced by Long et al. (7) is expanded. The zig-zag line indicates the region found to be homologous with the repeating AluI fragments.

(c) Two of the 240 nucleotide AluI fragments are shown on the expanded scale with the regions homologous to b) in alignment.

Figure 2. Hybridisation of Embryonic Transcripts to Spacer Sequences. The arrows correspond to the migration positions of RNA molecules ranging in size from 240 nucleotides to 1.68Kb and are placed at 240 nucleotide intervals. The sizes are relative to the mobility of the HindIII fragments of pAT153, and the assumption is made that the gel system is fully denaturing. The molecules labelled 'p' and 'd' contain ETS sequences (see text for details and reference 16 for full nomenclature).
restriction fragment containing virtually all the 'non-transcribed' spacer (see Figure 1) was homologous to a set of transcripts which have not been previously described. Figure 2 shows the result of such an experiment in which the total RNA from 20 hour embryos was fractionated by electrophoresis on a denaturing gel, and then transferred to a nitrocellulose filter for hybridisation with the \textit{HindIII/HaeIII} fragment. A 'ladder' of bands is seen corresponding to RNA molecules differing in molecular length by an integral value of 240 nucleotides. In addition two other strongly hybridising bands can be seen in figure 2, designated 'p' and 'd'. These correspond to the major primary transcript and a processing intermediate respectively. Both have sequences from the external transcribed spacer (ETS), which is present within this probe. The 240 nucleotide 'ladder' cannot, however, be discerned when a \textit{TaqI/HaeIII} fragment from the ETS is used alone (see figure 1 and 5 of references (16) and (15) respectively for an example of such an experiment). The 240 nucleotide \textit{AluI} fragments from within the 'non-transcribed' spacer are homologous to this 'ladder' of transcripts. Although \textit{in vivo} transcripts of the 'non-transcribed' spacer have not been previously reported, some evidence for the transcription of these sequences in an \textit{in vitro} system has been presented by Kohorn and Rae (17).

Since the difference in length between the spacer transcripts corresponded to the length of the repeating unit as defined by regularly spaced \textit{AluI} sites, we decided to compare the nucleotide sequence of the

Figure 3. The Nucleotide Sequence of an \textit{AluI} Fragment.

a) The sequence indicated is common to four clones. Of three other clones which were sequenced one showed the following differences: an additional TA between nucleotides 24 and 25; an additional C between nucleotides 45 and 46; an additional T between nucleotide 73 and 74. Two showed the following differences: an additional A between nucleotide 100 and 101; an A replacing the T at nucleotide 133; an additional G between nucleotides 156 and 157; a T replacing the G at nucleotide 210.
b) Regions of intrasequence homology detected by the SEQ program (10), searching for homologies with greater than 75% matching.
c) Regions of dyad-symmetry detected by the SEQ program, searching for greater than 87% base pairing.
d) Map of the 240 nucleotide \textit{AluI} fragment showing cleavage sites for known restriction endonucleases. The positions of the imperfect direct sequence repeats detected by the intrasequence homology search are shown to the left of the map. The positions of sequences which show dyad-symmetry are shown to the right of the map. Broadly speaking, these fall into two groups in which the dyads are roughly clustered around nucleotides 100 and 130. The other parameters of the SEQ program were set as follows for this analysis: Mismatch, 5; After Dis, 2; Loop out, 3; GU pair, 0; Min Loop, 0; Max Loop, -1. See (10) for description of parameters.
repeat with the previously determined sequence of the initiation site for the primary transcript (7). We purified the 240 nucleotide AluI fragments from the cloned rDNA segment, Dm103 (18), cloned these into M13 and sequenced seven such clones. Four clones had an identical sequence (figure 3) and three showed minor variations of this sequence (see legend to figure 3). Two segments of the AluI fragment show considerable homology to the sequence reported by Long et al. (7). These are indicated on the line diagram in figure 3d (labelled 1 and 2) and are printed in alignment with the sequence of Long et al. in figure 4. Segment 2 extends from nucleotide 72 to 163 and

![Line diagram](image)

**Figure 4. Alignment of Sequences of the AluI fragment with the Sequence Around the Primary Transcription Initiation Site.**

The continuous sequence is from reference 7, but is numbered differently. Transcription begins at nucleotide 236. The transcribed nucleotides are overscored with a dashed line. The two homologous regions from the AluI fragment are shown below the continuous sequence and are numbered as in figure 3. Mismatched nucleotides within these homologous segments are indicated by asterisks below the lower sequence.
would seem to correspond to the distal part of the last repeating element in the 'non-transcribed' spacer. The nucleotide sequence then undergoes considerable divergence at a position 140 nucleotides upstream from the transcription initiation site. There is then, however, striking homology between nucleotides 10 and 62 of the repeating AluI fragment (segment 1 in figure 3d) and a 54 nucleotide sequence around the primary transcription initiation site. The repetition of this sequence at intervals of 240 nucleotides within the 'non-transcribed' spacer would account for the transcripts we have observed (figure 2). We imagine transcription can begin randomly at one of these sites and then either proceed to the ETS or to the next such site where further transcription might be blocked by a weak terminator signal.

The finding of homologous sequences between the 'non-transcribed' spacer and the external transcribed spacer of both *D. melanogaster* (this paper), and *Xenopus laevis* (5), indicates that this is a common organisational feature of rRNA genes. It has been suggested that such sequence elements in the 'non-transcribed' spacer can sequester RNA polymerase molecules and so serve as 'loading' sites for the enzyme. Moss (personal communication) has evidence from *in vitro* studies that the sequences in the 'non-transcribed' spacer can augment the transcription of the rDNA unit. Observations with the electron microscope of actively transcribed chromatin of certain tissues of *Xenopus* have revealed so-called 'prelude' transcripts in the spacer region, providing direct evidence that this region can be transcribed (19). It seems likely therefore that these homologous sequences found in both the 'non-transcribed' spacer and at the initiation site for transcription are important for the recognition of rDNA by RNA polymerase I.

We have used the nucleic acid sequence analysis program SEQ (10) to look for any structural features within the 240 nucleotide AluI fragment which might serve a role in recognising RNA polymerase. The only structural features of any significance which we observe are in segment 2 (see figure 3), the region immediately after which there is sequence divergence of the last repeating unit before the primary transcription initiation site. There is an indication that this region may be derived from three tandemly repeating sub-elements, each about 20 nucleotides long. Furthermore two clusters of sequences within this same region contain elements of dyad-symmetry. It will be of interest to test the functional significance of these regions in *in vitro* transcription systems. RNA polymerase I is unlike RNA polymerase II in that it has a high degree of species specificity for its template. The
D. melanogaster enzyme, for example, will transcribe D. melanogaster rDNA but not D. virilis rDNA in vitro. Since the spacer sequences of rDNA have long been known to show extensive evolutionary divergence between closely related species it will be of interest to examine the nucleotide sequences of the spacers from closely related Drosophila species capable of forming inter-specific hybrids with D. melanogaster, in order to see whether these putatively functional regions are more strongly conserved.

ACKNOWLEDGEMENTS

This work was supported by an MRC project grant. DMG holds a career development award from the Cancer Research Campaign. The Imperial Cancer Research Fund generously made their computing facilities available to us.

*Present address: Laboratoire de Différenciation Cellulaire, École de Médecine, 20 Rue de l’École-de-Médecine, CH-1211 Genève 4, Switzerland.

† To whom reprint requests should be sent

REFERENCES

Note Added In Press: since writing this manuscript we have discovered that similar experiments have been carried out by B.D. Kohorn and P.M.M. Rae (Nucl. Acids Res. 10, 6879-6886), E. Coen and G. Dover (Nucl. Acids Res. 10, 7017-7026) and A. Simeone, A. De Falco, G. Macino and E. Boncinelli (Nucl. Acids Res. in press). Kohorn and Rae (op cit.) show that the non-transcribed spacer sequences can promote transcription in vitro.