Complex patterns of transcription of a *Drosophila* retrotransposon *in vivo* and *in vitro* by RNA polymerases II and III

Irina R. Arkhipova*

Department of Molecular and Cellular Biology, Harvard University, 7 Divinity Avenue, Cambridge, MA 02138-2092, USA and Engelhardt Institute of Molecular Biology, Russian Academy of Sciences, Moscow, Russia

Received June 19, 1995; Revised and Accepted September 19, 1995

ABSTRACT

The mdg1 retrovirus-like retrotransposon of *Drosophila melanogaster* was found to possess a complex promoter which can be transcribed by both RNA polymerases II and III (pol II and pol III). Pol III transcription, which is not typical of protein-coding genes, is driven by the sequences located in the long terminal repeat (LTR) of mdg1, predominantly within the transcribed region and is initiated 10 bp upstream from the regular pol II RNA start site. The pol III RNA start site is observed not only in *in vitro* transcription reactions, but also in total RNA isolated from tissue culture cells, larvae, pupae and adult flies. A possible role of pol III transcription in mechanisms controlling the expression of full-length mdg1-encoded transcripts in the developing fly, which are apparently relaxed in cell culture, is discussed.

INTRODUCTION

It is generally accepted that retrotransposons, including both retrovirus-like (long terminal repeat or LTR-containing) and LINE-like (non-LTR) elements, are RNA polymerase II (pol II) transcription units (for review, see refs 1-5). Short retrotransposons known as SINEs do not code for proteins and use an internal RNA polymerase III (pol III) promoter for their efficient transcription and subsequent propagation throughout the genome via reverse transcriptional pathway (for review, see ref. 6). In contrast, active retrotransposons (both LTR- and non-LTR-type) give rise to mRNAs that provide both the substrate and the enzymatic activities for retrotransposition. The proteins are expressed from genome-sized or spliced subgenomic mRNAs several kilobases in length. Transcripts of such length are not typical for pol III, which usually terminates transcription at short oligo(T) stretches (7) that are common enough in eukaryotic DNA to be present at least once within a few hundred base pairs. *Drosophila melanogaster* possesses a diverse set of both retrovirus-like and LINE-like retrotransposons (reviewed in refs 2,8); SINE-like elements are not characteristic for this organism. As a rule, promoters of these retrotransposons lack TATA motifs (for

review, see ref. 1). A number of LINE-like *D.melanogaster* retrotransposons (jockey, I and F elements), which initiate their transcription at the first nucleotide of the element, possess an internal promoter and α -amanitin experiments have shown that pol II is indeed the transcribing polymerase (9-11). Very few retrovirus-like retrotransposons have been assayed for pol II/III transcription: their mRNA start sites are located within LTRs, preserving both upstream and downstream promoter elements after a retrotransposition cycle and there is little reason to expect that they can be transcribed by anything other than pol II. The gypsy element, for instance, is the one that has been analysed in this respect and the results were consistent with pol II rather than pol III transcription (11).

The promoter of the mdg1 LTR-retrotransposon, which has no TATA box, has been studied in detail (12). Although the level of its transcription can be modulated by upstream regulatory elements present within LTRs, the basal pol II transcription requires only the RNA start site element (initiator or Inr) and a downstream element located at +25/+30 bp with respect to the RNA start site. Later, such downstream elements were shown to be widespread in *Drosophila* promoters and the TATA box could be found only in a half of them (13). In addition, mdg1 LTR was also able to drive expression of the reporter gene in an inverted orientation, thus representing a bidirectional promoter (12); full-length antisense mdg1 transcripts have been detected in early studies (14,15). In the present study, it is demonstrated that the mdg1 promoter is even more complex than previously thought. Its transcription may be driven both by pol II and pol III, opening additional possibilities for controlling expression of retrotransposon-encoded transcripts.

MATERIALS AND METHODS

Preparation of transcriptionally active extracts

Two independent methods were used for preparation of nuclear extracts from *D.melanogaster* KcO cells (12 1) growing in a serum-free medium. The first method involves extraction of nuclei with 0.36 M $(\text{NH}_4)_2\text{SO}_4$, precipitation with 2.3 M $(\text{NH}_4)_2\text{SO}_4$, as described in the original procedure of Parker and Topol (16) modified by Soeller *et al.* (17), and subsequent

*Correspondence address: Department of Molecular and Cellular Biology, Harvard University, 7 Divinity Avenue, Cambridge, MA 02138-2092, USA

concentration and partial purification of transcription factors by second precipitation with $(\text{NH}_4)_2\text{SO}_4$, dialysis into potassium glutamate and polyvinyl alcohol (PVA) precipitation (18); this is further referred to as the PVA extract for convenience only. The second method (9) does not involve salt extraction but releases soluble transcription factors from nuclei by centrifuging the nuclear pellet at high speed (the so-called soluble nuclear fraction or SNF). The protein concentration was typically 15 mg/ml in the PVA extract and 25 mg/ml in the SNF; the extracts were stored in aliquots at -70°C . A commercially available *in vitro* eukaryotic transcription kit (Stratagene), prepared from *Drosophila* embryos, was also used in some experiments. Although the efficiency of *in vitro* transcription varied for different types of extracts, no major qualitative differences in transcription patterns were observed.

Plasmids

A detailed description of the mdg1-chloramphenicol acetyltransferase (CAT) fusion constructs is given in ref. 12. Briefly, the templates used in this study were the entire LTR sequence (catX), 5'-terminally deleted constructs (-146, -20, -8, +15) which contain the transcribed sequences up to the position +112, the minimal pol II promoter fragment (-8/+34) and the two mutagenized plasmids differing from the wild-type sequence at +27/+31 by a 5 bp replacement of the +30 element which abolishes mdg1-CAT activity *in vivo* (TACACT→GGATCC, constructs -8/+34 M and -8/+112 M). The numbering refers to the positions with respect to the pol II RNA start site, as determined originally by comparison of different reverse transcription and S1-digestion products to sequence ladders generated from heterologous markers (12,19). Its precise position, as determined from the homologous sequence ladders in the present study, differs from the previously reported RNA start site by 5 bp and corresponds to the C within the initiator sequence TCAGT, which is the most common RNA start site for *Drosophila* protein-coding genes (13,20). For convenience, the original numbering is preserved in the names of constructs; the actual position should be corrected by subtraction of 5 bp. The mdg3-CAT construct (-109/+47) and the pUSVL-CAT promoterless vector are described in ref. 12.

In vitro transcription and primer extension analysis

Transcription reactions and the analysis of resulting products were performed essentially as described in ref. 21, with modifications as in ref. 9. Transcription was carried out in 12.5 mM HEPES (K^+), pH 7.6, 0.1 M potassium glutamate, 6.25 mM MgCl_2 , 0.05 mM EDTA, 5% (v/v) glycerol, 0.5 mM dithiothreitol and 2 U/ μl RNasin (Promega). Supercoiled plasmid DNA was added to a concentration of 20 $\mu\text{g}/\text{ml}$ and the protein concentration was typically 5 mg/ml in a final volume of 25 μl . Specific inhibitors of transcription were included into the reaction mix at the appropriate concentrations indicated in the text. Transcription was allowed to proceed for 30 min at 25°C . The commercial kit was used according to the manufacturer's instructions. The *Kr* DNA template and the corresponding primer from the *in vitro* transcription kit were also used in control experiments with the extracts prepared from Kc0 cells. The $5'$ - ^{32}P -labeled primer used for reverse transcription analysis of mdg1 RNA was TGCGAG-TGTACATTCCAG and the transcripts from the mutant mdg1 templates were analyzed with the primer TGCGGGATCCCAT-

TCCAG. mdg3-CAT contains only 47 bp of the transcribed region and was analysed with the CAT primer (CAACGGTGG-TATATCCAGTGATTTT), as was the pUSVL-CAT vector. The annealed primers were extended with four dNTPs and 20 U/ μl Mo-MLV reverse transcriptase (GIBCO BRL) in the presence of actinomycin D at 37°C for 1 h. The radiolabeled extension products were analyzed in 6% polyacrylamide-urea gels. The products of dideoxy chain termination reactions obtained with the Sequenase kit (USB, version 2.0) using the same $5'$ - ^{32}P -labeled primers were run in parallel as size markers.

UV crosslinking and DNase I protection assays

The binding reactions were performed as described in ref. 12. The binding was allowed to proceed for 20 min in a 20 μl volume at 0°C . UV-crosslinking was performed according to ref. 22. $5'$ -labeled and annealed complementary oligonucleotides containing the wild-type and mutant downstream element (12, see also above) were used as probes. For comparison, annealed complementary oligonucleotides composed of $(\text{CAA})_5$ and $(\text{TAA})_5$ with the same GATC overhang were used. Poly[dG-dC].poly[dG-dC] (100 $\mu\text{g}/\text{ml}$; Sigma) was used as a non-specific competitor. The binding reactions were subjected to UV-irradiation at 254 nm for 5 min and 4 μl 0.4% Sarkosyl, 2 mM PMSF were added to dissociate many of the noncovalently bound proteins from DNA. No nuclease digestions were performed, as the probes were short enough to produce negligible effects on protein mobility. The photoadducts were separated through a discontinuous SDS-PAGE (8%) and visualized by autoradiography.

For DNase I footprinting, the binding reactions were scaled up to 100 μl . The $3'$ - ^{32}P -end-labeled antisense and sense strand probes were prepared by filling in the *Hind*III and *Xma*I ends, respectively, in the *Hind*III-*Xma*I digest of the -8/+112 construct and purified through polyacrylamide gel electrophoresis. The *Hind*III-*Xma*I digest of the -95/+112 construct was used to prepare a probe for footprinting of the upstream region. Footprinting was performed essentially as described in ref. 23. For reference, the G+A Maxam-Gilbert sequencing reactions were performed on the same fragments and run in parallel.

Isolation and analysis of *Drosophila* RNA

Total RNA was isolated from Schneider 2 tissue culture cells by the hot phenol extraction procedure as described in ref. 14. To assay inhibition of pol II transcription *in vivo*, α -amanitin was added to a concentration of 4 $\mu\text{g}/\text{ml}$ 24 h prior to RNA isolation. RNA from Canton-S, w^a and *Adh^{fn4} pr cn* (24) fly strains was prepared by phenol extraction and LiCl precipitation. The $5'$ - ^{32}P -labeled wild-type mdg1 oligonucleotide (see above) and the oligonucleotide complementary to the +75/+93 region of the *white* gene (GAAGCGAGAGGAGTTTTGG) were used for primer extension. The RNA blot used in Figure 9 was kindly provided by O. Danilevskaya. The short 0.25 kb transcript was detected with several mdg1 LTR probes, including double-stranded nick-translated internal 250 bp *Eco*RI LTR fragment and single-stranded wild-type oligonucleotide labeled either at the $5'$ end with $[\gamma\text{-}^{32}\text{P}]\text{ATP}$ and T4 polynucleotide kinase or at the $3'$ end by extension with the Klenow fragment of DNA polymerase I in the presence of dATP, dTTP and $[\alpha\text{-}^{32}\text{P}]\text{dCTP}$, using the 250 bp *Eco*RI fragment as a template. The latter procedure results in addition of 10 nucleotides, including three labeled cytidines, to the $3'$ end of the original oligonucleotide.

Figure 1. Drug sensitivity of mdg1 (–146/+112 construct) and *Kr* transcription *in vitro*. Transcription reactions were carried out by the SNF (left panel) or PVA extracts (right panel) in the absence (–) or presence of 4, 40 or 100 µg/ml α -amanitin (as indicated by corresponding subscript numbers) or 200 U/ml tagetin (T). Specific products were analyzed by primer extension. The major pol III mdg1 RNA start site is indicated by an arrow, the major pol II *Kr* RNA start sites, by asterisks.

Alignment of nucleotide sequences

Sequences were analyzed with the aid of the GCG sequence analysis software package (25) running on a SUN Sparcstation. The programs PILEUP and PRETTY were used for alignment and display, respectively.

RESULTS

In vitro transcription pattern and α -amanitin sensitivity of mdg1

To study the sequence requirements of transcription from the mdg1 promoter and to compare the *in vivo* and *in vitro* transcription patterns, transcriptionally active extracts were obtained from cultured Kc0 cells by two different methods (conventional salt extraction and preparation of the soluble nuclear fraction or SNF, as described in Materials and Methods).

In vitro transcriptional analysis of mdg1-CAT fusion constructs (12) revealed that the majority of *in vitro* transcripts are initiated not at the RNA start site observed in the earlier *in vivo* experiments (12,19), but rather at an unexpected position (appearing either as a doublet or as a triplet) centered ~10 bp upstream (Fig. 1), with an additional start site observed 10 bp farther upstream in some experiments. The utilization of the latter two sites may vary, depending on the method of extract preparation and its concentration. The intensity of the proximal site is higher in SNF extracts, while the intensity of the distal site increases with lowering the extract concentration, which might be indicative of non-specific initiation (data not shown). The proximal RNA start site predominates *in vivo* (see Fig. 7 for comparison of *in vitro* and *in vivo* patterns), although the distal site could also be observed in Schneider 2 cells.

Figure 2. *In vitro* transcription patterns of *Kr*, mdg1 and its various mutant derivatives by the SNF. The numbering refers to the position with respect to the pol II RNA start site, as indicated in ref. 12. M denotes replacement of 5 bp in the wild-type downstream element at the positions +27/+31. Pluses and minuses correspond to the presence and absence of 4 µg/ml α -amanitin, respectively. A, C, G and T, chain termination products of the –8/+34 construct including the upstream polylinker sequence. The RNA start sites are indicated as in Figure 1.

Transcription of the control *Kr* promoter under the same conditions (temperature, template and protein concentrations, etc.) is properly initiated *in vitro* at the same positions as *in vivo* and is strongly inhibited by the low concentration (4 µg/ml) of the fungal toxin α -amanitin (Fig. 1, *Kr* α_4). The same is true for *in vitro* transcription of the *white* gene (not shown). In contrast, mdg1 *in vitro* transcription from either site is not sensitive to low, as opposed to high, concentrations of the drug (Fig. 1, compare α_4 and α_{100}). Such α -amanitin sensitivity is characteristic of genes transcribed by RNA polymerase III and is often used to discriminate between different polymerases (26).

mdg1 promoter is transcribed by RNA polymerase III

To obtain a more straightforward evidence of mdg1 transcription by pol III, a specific pol III inhibitor, tagetin (27), was used in the *in vitro* transcription assays. The experiments show that mdg1 transcription *in vitro* from the major start sites is clearly inhibited by tagetin, while transcription from the control pol II *Kr* promoter is practically insensitive to the drug (Fig. 1, T). Thus, the pattern of response to two different drugs is entirely consistent with pol III transcription of mdg1 promoter.

Promoter elements responsible for pol III transcription of mdg1

A set of plasmid constructs, which represent the mdg1 promoter mutagenized or truncated to various extents, was used in transcription reactions with Kc0 extracts to explore the sequence requirements for pol III initiation *in vitro*. The major pol III RNA start site located 10 bp upstream from the regular pol II start site was observed in the –146/+112, –8/+112 (Fig. 2), complete LTR (catX) and –20/+112 constructs (not shown). The patterns of minor upstream sites in the –8/+112 and –20/+112 constructs were slightly different from the more complete constructs (Fig.

Figure 3. Nucleotide sequence alignment between the promoter regions of *mdg1*, *D.melanogaster* 5S RNA (Dm5S) and *X.laevis* 5S RNA (Xl5S) (GenBank accession nos X59545, M25016 and M35175, respectively) displaying regions of homology. Pol III RNA start sites are double-underlined, the pol II RNA start site of *mdg1* is in boldface and indicated by a triangle. Uppercase letters denote the nucleotides which are identical between at least two sequences and coincide with the computer-generated consensus. Vertical lines correspond to deletion breakpoints designated as -20, -8, +15 and +34 in the deletion constructs, respectively; five nucleotides at +27/+31 replaced by site-directed mutagenesis in the constructs -8/+34M and -8/+112M are italicized. Arrows indicate DNase I cleavages induced by protein binding (downward arrows, cleavages in the minus or antisense strand; upward arrows, cleavages in the plus strand, the sequence of which is shown here) and the bracket designates the footprinted area approximately corresponding in location to the region occupied by TFIIC in the 5S promoter (Fig. 4; ref. 30).

2), apparently due to the influence of juxtaposed vector sequences at the 5' boundary which produce a similar effect in other plasmids described below. Interestingly, the same pol III start site is preserved in the +15/+112 construct which does not retain any of the sequences surrounding the RNA start site (Fig. 2; see also Fig. 3). This is in general agreement with the properties of pol III transcription, which is known to depend primarily on internal control regions that specify the position of transcription initiation at some distance upstream (28).

Use of the constructs in which the +30 region essential for pol II transcription *in vivo*, as determined by CAT activity assays (12), was subjected to site-directed mutagenesis (-8/+34 M and -8/+112M), resulted in elimination of the major pol III start site (Fig. 2). This indicates that the internal control regions for pol II and pol III transcription may overlap, although the interacting proteins might be different in each case. The -8/+34 construct with an intact +30 element but with the sequences +34/+112 deleted still utilizes the major pol III start site, although with lower efficiency and slightly altered specificity, demonstrating that the more remote downstream regions contribute to the efficiency and accuracy of pol III transcription. This agrees with the footprinting data (see below) and with the structure of known pol III internal control regions which are multipartite and extend up to 90 bp downstream (29,30).

Sequence similarity to the promoter regions of 5S RNA genes

The ability of the *mdg1* LTR to function as a pol III promoter stimulated the search for sequence homologies to known pol III promoters. While no significant matches to the promoter elements of known tRNA, U6 or 7SL RNA genes could be found, the regions adjacent to the RNA start site and further downstream displayed a relatively good overall similarity to the promoters of 5S RNA genes (Fig. 3). The pairwise similarity scores between *mdg1* and the 5S RNA genes of *D.melanogaster* and *X.laevis*, as determined by the program PILEUP, were ~0.5; the score between the 5S RNA genes for these two organisms was 0.7 and the scores between *mdg1* and different *Drosophila* tRNA genes averaged 0.3 and never exceeded 0.4. The regions of detectable similarity spanned the RNA start sites and the more downstream sequences, including those that were known to interfere with pol III promoter activity of the 5S genes, especially in *Drosophila* (31–33), as well as *mdg1* (see above).

The downstream region is protected from DNase I digestion

Pre-incubation of the end-labeled downstream promoter fragment [antisense or (-) strand] with increasing concentrations of

Figure 4. DNase I footprinting patterns of the *mdg1* transcribed region in the construct -8/+112 (A and B) and of the entire promoter region in the construct -95/+112 (C). The concentration of the protein extract increases from 2.5–5 to 10 mg/ml, as indicated by the triangle; N, naked DNA. Arrows indicate DNase I cleavages induced by the addition of the extract (their positions on the promoter sequence are also indicated by arrows in Fig. 3). The brackets designate the areas protected from DNase I digestion, roughly corresponding to the position of the binding sites for TFIIC (lower bracket; see Fig. 3) and TFIIA (upper bracket) on the 5S promoter (30). The G+A Maxam–Gilbert sequencing reactions of the same labeled fragment were run in parallel as size markers.

the nuclear extract prior to DNase I digestion results in a large protected area extending from the RNA start site 75–80 bp into the transcribed region (brackets on Figs 3 and 4). This large footprinted area can be subdivided into two domains with a DNase I-sensitive region between them (Figs 3 and 4A and C); this region is nevertheless protected in comparison with naked DNA, which exhibits a local increase in DNase I sensitivity at this position (+29/+30). A few novel DNase I cleavage sites appear that are not present in the naked DNA and are apparently induced by bound nuclear proteins (arrows on Figs 3 and 4). The sense strand is protected more uniformly (Fig. 4B). The pattern of DNase I sensitivity remains the same when the -95/+112 fragment is used as a probe (Fig. 4C), indicating that it reflects specific protein

Figure 5. Proteins that can be visualized in PVA extracts from Kc0 cells after UV-crosslinking (+UV) to the corresponding 5'-³²P-labeled and annealed complementary oligonucleotide probes (see Materials and Methods). W and M denote the wild-type and mutant oligonucleotides, respectively; T corresponds to (TAA)₅ and C, to (CAA)₅. The first four lanes (-UV) represent non-irradiated controls. Positions of the molecular weight markers are indicated on the right.

binding rather than generalized protection. In addition, DNase I footprinting of the upstream promoter region in this fragment does not show extensive protected areas, although some protection may be noticed in the region corresponding to the usual location of the TATA box in TATA-containing promoters (Fig. 4C). This may probably indicate participation of TFIID, which is considered to be a universal transcription factor involved in transcription by all three eukaryotic RNA polymerases (reviewed in refs 34,35), in *mdg1* transcription, although the *mdg1* promoter can be classified as TATA-less (12,19).

Extensive protection of the downstream region is a characteristic feature of pol III promoters and includes the sites of interaction with pol III transcription factors TFIIA and TFIIC (28,32; for review, see refs 29,30). Transcription of *Drosophila* 5S RNA genes is apparently also mediated by TFIIA-like activity and includes additional internal control regions closer to the RNA start site (33). The pattern observed in Figure 4 is highly reminiscent of DNA-protein interactions over boxes A and C of the 5S gene promoter (29,30). Although relatively crude protein fractions were used for footprinting in the present study; the picture was not dramatically different when some of the known pol III promoters were footprinted with crude extracts or affinity-purified proteins (e.g. 36,37). In addition, these areas are not protected in the mutant probe (-8/+112 M, data not shown), suggesting that the observed DNA-protein interactions may indeed be related to transcription.

UV crosslinking patterns differ for the wild-type and mutant downstream oligonucleotides

Additional evidence for the involvement of the downstream region in DNA-protein interactions was obtained in the experiment in which the annealed complementary oligonucleotides containing the wild-type and mutant downstream elements were crosslinked to DNA-binding proteins present in transcriptionally active Kc0 extracts. In addition to the bands that are also observed for the mutant and control (YAA)₅ oligonucleotides, the wild-type oligonucleotide displays strong crosslinking to a number of

Figure 6. *In vitro* transcription patterns of *mdg3*-CAT (-109/+47) and the CAT vector (pUSVL-CAT) plasmids. α and - denote the presence and absence of 4 μg/ml α-amanitin, respectively. The -109/+47 plasmid was also used for generating the sequencing ladder.

bands ~35 and 40–45 kDa in size (Fig. 5). The fact that these proteins are not seen with the mutant probe lacking transcriptional activity suggests that they might possibly be involved in transcription. It is interesting that the control oligonucleotide (CAA)₅ is crosslinked to a completely different subset of proteins, which differs not only from the wild-type and mutant promoter oligonucleotides but also from the control (TAA)₅ oligonucleotide. This implies the existence of nuclear proteins able to bind differentially to these relatively simple sequences. No specific crosslinked proteins can be visualized in non-irradiated control binding reactions.

Other retrotransposons are transcribed *in vitro* and *in vivo* by RNA polymerase II

It was of interest to find out whether the dual polymerase specificity of the *mdg1* promoter represents a common phenomenon among *Drosophila* retrotransposons. Promoters of two other retrotransposons (*mdg3*, which also belongs to the LTR-containing subclass and *jockey*, which is a LINE-like retrotransposon), for which RNA start sites *in vivo* have been reported previously (11,12), were analysed in this respect. Neither of them exhibited *in vitro* any differences from the initiation patterns observed *in vivo* and transcription from both promoters was sensitive to low α-amanitin concentrations (*mdg3*, Fig. 6; *jockey*, see ref. 9 and data not shown). Thus, *mdg1* appears to be an exception rather than a rule among retrotransposons with respect to the difference observed between *in vivo* and *in vitro* transcription patterns.

mdg1 transcription by pol III occurs *in vivo*

To test the possibility that *mdg1* transcription by pol III represents an artefactual phenomenon that can be observed only *in vitro* and does not involve promoters of other retrotransposons and cellular genes, total RNA was isolated from Schneider 2 cells as well as from larvae, pupae and adults of several *D. melanogaster* strains. Figure 7 demonstrates that both pol III RNA start sites can be utilized with variable efficiency *in vivo* in tissue culture cells,

Figure 7. Primer extension analysis of *mdg1* *in vitro* transcripts (SNF; -146/+112 construct) and of total RNA isolated from Schneider 2 (Sh2) tissue culture cells and from adult flies of the Canton-S (CS), Oregon-R (OR) and *w^a* strains. The asterisk and arrows denote the pol II and pol III RNA start sites, respectively. T and α denote the addition of 50 U/ml taqetatin and 4 μg/ml α-amanitin, as described in Materials and Methods. The products of sequencing reactions using the -146/+112 template were run in parallel as size markers.

while the proximal one predominates in adult flies. In addition, an impressive difference between the fly strains and cultured cells is the presence of the transcript initiated downstream, which is sensitive to 4 μg/ml α-amanitin *in vivo* in Schneider 2 cells (Fig. 7). This latter transcript apparently originates from the regular *mdg1* pol II start site. Although the addition of α-amanitin *in vivo* does not lead to complete inhibition of pol II transcription, it is nevertheless reduced very strongly in comparison with the levels of pol III transcripts which remain practically unaffected. The control *Kr* promoter *in vivo* exhibits ~5-fold drop in activity in response to α-amanitin (not shown).

The fact that the same pol III RNA start sites are observed *in vitro* with purified *mdg1* plasmids and corresponding primers and *in vivo* with preparations of total *Drosophila* RNA strongly argues against fortuitous homology of the primers to some unrelated cellular RNAs as well as against non-specific initiation *in vitro*.

An interesting observation is that *mdg1* is transcribed from the pol III RNA start sites at relatively low levels in late larvae, pupae and newly eclosed Canton-S adults, as opposed to mature flies (Fig. 8); the intensity is reversed with respect to an additional RNA start site observed at the far upstream position (>200 bp in the 5' direction). No major changes in intensity can be noticed at these stages for the transcript of the *white* gene (Fig. 8).

A short non-polyadenylated *mdg1* transcript is present in flies

In addition to the full-length 7 kb *mdg1* transcript, a short transcript of the same polarity ~0.25 kb in length (Fig. 9) was detected by Northern blot analysis of total RNA isolated from adult flies of two *D.melanogaster* strains (Oregon R and *gt-1*) with the same probe that was used in primer extension experiments (Figs 7 and 8). This transcript was not seen in the poly(A)-containing fraction (not shown) and its length is more

Figure 8. Primer extension analysis of total RNA isolated from third-instar larvae (L), early pupae (P), newly eclosed adults (E) and mature adults (A). The RNA preparations were divided equally between the primers corresponding to the *white* gene (four middle lanes, L–A) and to *mdg1* (four right lanes, L–A); the *white* transcript is initiated at its authentic RNA start site, as determined by the parallel sequencing ladder. The pol III RNA start site of *mdg1* is indicated by an arrow.

characteristic of the short non-polyadenylated pol III transcripts rather than pol II transcripts usually having a poly(A) stretch ~0.2 kb in length. It is reasonable to suggest that this short transcript is generated by pol III and its presence in two different fly strains indicates that it is completely element-specific and the adjacent genomic sequences are not involved in its production.

A variety of *mdg1* transcripts of different lengths, often strain-specific, have been observed with different probes, mostly in the poly(A)-containing RNA fraction of tissue culture cells or different fly strains at various developmental stages (14,38–40). However, transcripts shorter than 0.8 kb have not been reported. Although *mdg1* transcript abundance and patterns can be rather complicated and variable between strains, there seems to be a general trend to suppress the formation of full-length pol II transcripts (or otherwise prevent their expression) in flies. The present study indicates that this control might be relaxed in tissue culture cells, where both pol II and pol III start sites can be observed simultaneously.

DISCUSSION

The present study demonstrates that the promoter of the *mdg1* *Drosophila* retrotransposon possesses a dual RNA polymerase specificity, with both pol II and pol III being able to initiate transcription at the polymerase-specific neighboring RNA start sites located 10 bp apart. The results of DNA–protein binding experiments and the similarity to some of the known pol III

Figure 9. Northern blot analysis of total *D. melanogaster* RNA isolated from Oregon R (OR) and gt-1 (gt) strains (58). The membrane was probed with an antisense mdg1 oligonucleotide probe. The sizes of the RNA markers are indicated on the right.

promoters are also consistent with the involvement of pol III in transcription. Moreover, pol III transcription can be observed both *in vitro* and *in vivo*, including tissue culture cells and different fly strains, although it is not known whether pol II and pol III transcripts originate from the same or different genomic copies.

There is evidence that pol II can interfere with the function of promoters through which it transcribes (41–43). Similarly, elevated levels of pol III transcription from upstream sites might compete with pol II mdg1 transcription at the downstream position, leading to changes in the amounts of full-length pol II transcripts encoded by the retrotransposon. Pol III interference from an upstream promoter has not been reported, although there is an example of a distal Alu element pol III promoter relieving transcriptional interference from even more upstream pol II promoters (44).

Closely linked pol II/III transcription units in eukaryotic cells are not unprecedented, but the reported examples are scarce. In mammals, superimposed pol II and pol III promoters which initiate transcription from the same RNA start site have been reported for the *c-myc* gene in injected *Xenopus* oocytes and in transcriptionally active HeLa extracts (45). Pol III transcription was implicated in regulation of *c-myc* expression; however, it was observed only at high template concentrations and its existence *in vivo* was not confirmed (46,47). Although most of the snRNA genes are not usually transcribed by pol III, the requirements for pol II/III transcription in snRNA genes seem to be relatively loose, since a 7 bp mutation was sufficient to convert the human U6 gene into a pol III transcription unit (48). The promoter of a *Xenopus* U6 snRNA seems to represent a *bona fide* dual pol II/III transcription unit upon injection into oocytes (49).

Promoters of retrotransposons which have been analyzed in detail often possess a complex structure. The lepidopteran LTR-retrotransposon TED, in addition to a promoter utilized in uninfected cells, also contains a baculovirus-specific promoter which can be used during viral infection (50). Some promoters are able to drive transcription in both directions, producing antisense

RNAs (12,51–54). Such diversity argues against any conserved mechanism for appearance of additional promoters in retrotransposons. Rather, it might arise in individual cases (including specific tissues and developmental stages) as one of the cellular controls over expression levels of retrotransposon-encoded genome-sized mRNAs, as suggested for F and micropia elements (53,54), especially in the germ line to restrict inheritable retrotransposition.

An analogous role could be suggested for pol III transcription of mdg1. Promoter interference by pol III could be a convenient way to control expression of full-length pol II transcripts. In this respect, it is appropriate to mention that the 5' leader region of mdg1 possesses an unusual structure among retrotransposons: in addition to two short open reading frames preceding the major ORFs, it also contains sites of nuclear protein binding and 3' end processing regions which lead to formation of prematurely terminated transcripts, predominantly at the larval stage (38). The overall impression is that high levels of full-length mdg1 expression might be contained in a number of ways, including transcriptional interference, premature termination and formation of antisense transcripts and these ways may vary in a stage-, tissue- and strain-specific manner. The presence of a short 0.25 kb mdg1 transcript in different fly strains could be indicative of its functional significance. It is worth noting that the distribution of mdg1 in *D. melanogaster* chromosomes shows under-representation on the X chromosome, this being consistent with negative selection acting on mdg1 insertions, while the chromosomal distribution of mdg3 fits the neutral model (55). No known mutations have been associated with mdg1 insertion. It is also a constituent of heterochromatic repeat units in *D. melanogaster* (56).

Although some of the yeast retrotransposons integrate specifically in the vicinity of pol III-transcribed promoters (reviewed in ref. 57), none of them are known to be transcribed by pol III. It is not clear whether acquisition of a pol III promoter could provide any selective advantage for a retrotransposon not belonging to the SINE class.

The sequences within the transcribed region of mdg1 and the proteins which bind to them apparently play an essential role in transcription by either pol II or pol III, although different subsets of proteins may be involved in each case. While their complete identification and investigation represents a rather complicated problem due to the dual nature of the promoter, it appears likely that these multiple components constitute a dynamic system which may be sensitive to quantitative changes in their ratios and relative affinities. The complexity of the problem is increased by differences between *Drosophila* cell lines in which the extent of mdg1 amplification and expression varies greatly (for example, Kc0 cells contain relatively few mdg1 copies and these are transcribed at rather low levels).

In conclusion, mdg1 appears to represent an interesting exception among promoters of protein-coding genes, possessing a dual RNA polymerase specificity both *in vivo* and *in vitro*. The existence of two characteristic RNA start sites allows one to distinguish between pol II and pol III transcription *in vivo*. There is no doubt that pol II transcription of mdg1 occurs *in vivo*, since it is initiated at the consensus *Drosophila* pol II RNA start site, is sensitive to low concentrations of α -amanitin *in vivo* and functionally active proteins are expressed from mRNAs transcribed from the mdg1 promoter in full-length elements as well as in mdg1-CAT fusion constructs. There is also little doubt that pol III transcription of mdg1 can occur *in vivo*, since the characteristic pol III RNA start site, which is defined by the pattern of drug

sensitivity both *in vivo* and *in vitro*, can be observed in total RNA preparations from tissue culture cells and flies at different stages of development. The functional significance of *mdg1* transcription by pol III, however, is still unclear and deserves further investigation. It also remains to be seen how widespread are such phenomena and what role, if any, could they play in spatial and temporal regulation of eukaryotic gene expression.

ACKNOWLEDGEMENTS

I wish to thank M. Meselson for his continuous support and advice throughout the course of this work and preparation of the manuscript. I would also like to thank O. Danilevskaya, Y. Ilyin and N. Lyubomirskaya for providing some of the materials, critical reading of the manuscript and valuable comments. This research was supported by the National Institutes of Health grant GM-22274 to M. Meselson.

REFERENCES

- Arkipova, I.R. and Ilyin, Y.V. (1992) *BioEssays*, **14**, 161–168.
- Arkipova, I.R., Lyubomirskaya, N.V. and Ilyin, Y.V. (1995) *Drosophila Retrotransposons*. R.G. Landes Co., Austin, TX, p. 134.
- Bingham, P.M. and Zachar, Z. (1989) In Berg, D.E. and Howe, M.M. (eds), *Mobile DNA*. American Society for Microbiology, Washington, DC, pp. 485–502.
- Hutchison, C.A. III, Hardies, S.C., Loeb, D.D., Shehee, W.R. and Edgell, M.H. (1989) In Berg, D.E. and Howe, M.M. (eds), *Mobile DNA*. American Society for Microbiology, Washington, DC, pp. 593–617.
- Varmus, H. and Brown, P. (1989) In Berg, D.E. and Howe, M.M. (eds), *Mobile DNA*. American Society for Microbiology, Washington, DC.
- Deininger, P. (1989) In Berg, D.E. and Howe, M.M. (eds), *Mobile DNA*. American Society for Microbiology, Washington, DC, pp. 619–636.
- Bogenhagen, D.F. and Brown, D.D. (1981) *Cell*, **24**, 261–270.
- Finnegan, D.J. (1992) In Lindsley, D.L. and Zimm, G.G. (eds), *The Genome of Drosophila melanogaster*. Academic Press, New York, pp. 1096–1107.
- Kamakaka, R.T., Tyree, C.M. and Kadonaga, J.T. (1991) *Proc. Natl. Acad. Sci. USA*, **88**, 1024–1028.
- McLean, C., Bucheton, A. and Finnegan, D.J. (1993) *Mol. Cell. Biol.*, **13**, 1042–1050.
- Mizrokhi, L.J., Georgieva, S.G. and Ilyin, Y.V. (1988) *Cell*, **54**, 685–691.
- Arkipova, I.R. and Ilyin, Y.V. (1991) *EMBO J.*, **10**, 1169–1177.
- Arkipova, I.R. (1995) *Genetics*, **139**, 1359–1369.
- Arkipova, I.R., Gorelova, T.V., Ilyin, Y.V. and Schuppe, N.G. (1984) *Nucleic Acids Res.*, **12**, 7533–7548.
- Ilyin, Y.V., Chmeliauskaite, V.G. and Georgiev, G.P. (1980) *Nucleic Acids Res.*, **8**, 3439–3457.
- Parker, C.S. and Topol, J. (1984) *Cell*, **36**, 357–369.
- Soeller, W.C., Poole, S.J. and Kornberg, T. (1988) *Genes Dev.*, **2**, 68–81.
- Wampler, S.L., Tyree, C.M. and Kadonaga, J.T. (1990) *J. Biol. Chem.*, **265**, 21 223–21 231.
- Arkipova, I.R., Mazo, A.M., Cherkasova, V.A., Gorelova, T.V., Schuppe, N.G. and Ilyin, Y.V. (1986) *Cell*, **44**, 555–563.
- Cherbas, L. and Cherbas, P. (1993) *Insect Biochem. Mol. Biol.*, **23**, 81–90.
- Kadonaga, J.T. (1990) *J. Biol. Chem.*, **265**, 2624–2631.
- Gilmour, D.S., Dietz, T.J. and Elgin, S.C.R. (1990) *Mol. Cell. Biol.*, **10**, 4233–4238.
- Ausubel, F.M., Brent, R., Kingston, R.E., et al. (1992) *Short Protocols in Molecular Biology*. Second edn, John Wiley & Sons, New York.
- Sass, H. and Meselson, M. (1991) *Proc. Natl. Acad. Sci. USA*, **88**, 6795–6799.
- Genetics Computer Group (1991) *Program manual for the GCG package*. Version 7, April 1991. Madison, WI.
- Hames, B.D. and Higgins, S.J. (1984) *Transcription and Translation: a Practical Approach*. IRL Press, Oxford, p. 328.
- Steinberg, T.H., Mathews, D.E., Durbin, R.D. and Burgess, R.R. (1990) *J. Biol. Chem.*, **265**, 499–505.
- Sakonju, S., Brown, D.D., Engelke, D., Ng, S.-Y., Shastry, S. and Roeder, R.G. (1981) *Cell*, **23**, 665–669.
- Geiduschek, E.P. and Tocchini-Valentini, G.P. (1988) *Ann. Rev. Biochem.*, **57**, 873–914.
- Geiduschek, E.P. and Cassavetti, G.A. (1992) In McKnight, S.L. and Yamamoto, K.R. (eds), *Transcriptional Regulation*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, pp. 247–280.
- Nickol, J. and Rau, D.C. (1992) *J. Mol. Biol.*, **228**, 1115–1123.
- Pieler, T., Hamm, J. and Roeder, R.G. (1987) *Cell*, **48**, 91–100.
- Sharp, S.J. and Garcia, A.D. (1988) *Mol. Cell. Biol.*, **8**, 1266–1274.
- Sharp, P.A. (1992) *Cell*, **68**, 819–821.
- Hernandez, N. (1993) *Genes Dev.*, **7**, 1291–1308.
- Newman, A.J., Ogden, R.C. and Abelson, J. (1983) *Cell*, **35**, 117–125.
- Camier, S., Gabrielsen, O., Baker, R. and Sentenac, A. (1985) *EMBO J.*, **4**, 491–500.
- Cherkasova, V.A., Surkov, S.N. and Ilyin, Y.V. (1991) *Nucleic Acids Res.*, **19**, 3213–3219.
- Kuvakina, A.I., Nurminsky, D.I., Kogan, G.L. and Gvozdev, V.A. (1988) *Genetika (Russ.)*, **24**, 1234–1240.
- Parkhurst, S.M. and Corces, V.G. (1987) *EMBO J.*, **6**, 419–424.
- Cullen, B.R., Lomedico, P.T. and Tu, G. (1984) *Nature*, **307**, 241–245.
- Proudfoot, N.J. (1986) *Nature*, **322**, 562–565.
- Corbin, V. and Maniatis, T. (1989) *Nature*, **337**, 279–282.
- Wu, J., Grindlay, G.J., Bushel, P., Mendelsohn, L. and Allan, M. (1990) *Mol. Cell. Biol.*, **10**, 1209–1216.
- Chung, J., Sussman, D.J., Zeller, R. and Leder, P. (1987) *Cell*, **51**, 1001–1008.
- Bentley, D.L. and Groudine, M. (1988) *Cell*, **53**, 245–256.
- Bentley, D.L., Brown, W.L. and Groudine, M. (1989) *Genes Dev.*, **3**, 1179–1189.
- Lobo, S.M. and Hernandez, N. (1989) *Cell*, **58**, 55–67.
- Mattaj, J.W., Dathan, N.A., Parry, H.D., Carbon, P. and Croll, A. (1988) *Cell*, **55**, 435–442.
- Friesen, P.D. and Nissen, M.S. (1990) *Mol. Cell. Biol.*, **10**, 3067–3077.
- Contursi, C., Minchiotti, G. and DiNocera, P.P. (1993) *J. Mol. Biol.*, **234**, 988–997.
- Friesen, P.D., Rice, W.C., Miller, D.W. and Miller, L.K. (1986) *Mol. Cell. Biol.*, **6**, 1599–1607.
- Minchiotti, G. and DiNocera, P.P. (1991) *Mol. Cell. Biol.*, **11**, 5171–5180.
- Lankenau, S., Corces, V.G. and Lankenau, D.-H. (1994) *Mol. Cell. Biol.*, **14**, 1764–1775.
- Biemont, C., Lemeunier, F., Garcia Guerreiro, M.P., Brookfield, J.F., Gautier, C., Aulard, S. and Pasyukova, E.G. (1994) *Genet. Res.*, **63**, 197–212.
- Nurminsky, D.I., Shevelyov, Y.Y., Nuzhdin, S.V. and Gvozdev, V.A. (1994) *Chromosoma*, **103**, 277–285.
- Voytas, D.F. and Boeke, J.D. (1993) *Trends Genet.*, **9**, 421–427.
- Hogan, N.C., Slot, F., Traverse, K.L., Garbe, J.C., Bendena, W.G. and Pardue, M.-L. (1995) *Genetics*, **139**, 1611–1621.