MOLECULAR AND CELLULAR BIOLOGY, Oct. 1991, p. 5171–5180
0270-7306/91/105171-10$02.00/0

Vol. 11, No. 10

Convergent Transcription Initiates from Oppositely Oriented Promoters within the 5′ End Regions of *Drosophila melanogaster* F Elements

GABRIELLA MINCHIOTTI[1] AND PIER PAOLO DI NOCERA[1,2]*

International Institute of Genetics and Biophysics, Consiglio Nazionale Delle Ricerche, Via Marconi 12, Naples 80125,[1] and Dipartimento di Biologia, Facoltà di Scienze, Università di Lecce, Lecce,[2] Italy

Received 10 June 1991/Accepted 25 July 1991

***Drosophila melanogaster* F elements are mobile, oligo(A)-terminated DNA sequences that likely propagate by the retrotranscription of RNA intermediates. Plasmids bearing DNA segments from the left-hand region of a full-length F element fused to the CAT gene were used as templates for transient expression assays in *Drosophila* Schneider II cultured cells. Protein and RNA analyses led to the identification of two promoters, F_{in} and F_{out}, that transcribe in opposite orientations. The F_{in} promoter drives the synthesis of transcripts that initiate around residue +6 and are directed toward the element. F_{in}, that probably controls the formation of F transposition RNA intermediates and gene products, is internal to the transcribed region. Sequences important for accumulation of F_{in} transcripts are included within the +1 to +30 interval; an additional regulatory element may coincide with a heptamer located downstream of this region also found in the 5′ end regions of F-like *Drosophila* retrotransposons. Analysis of the template activity of 3′ deletion derivatives indicates that the level of accumulation of F_{in} RNA is also dependent upon the presence of sequences located within the +175 to +218 interval. The F_{out} promoter drives transcription in the opposite orientation with respect to F_{in}. F_{out} transcripts initiate at nearby sites within the +92 to +102 interval. Sequences downstream of these multiple RNA start sites are not required for the activity of the F_{out} promoter. Deletions knocking out the F_{in} promoter do not impair F_{out} transcription; conversely, initiation at the F_{in} promoter still takes place in templates that lack the F_{out} promoter. At a low level, both promoters are active in cultured cells.**

The *Drosophila melanogaster* genome hosts a broad spectrum of transposable elements (15, 33). A large body of evidence supports the hypothesis that mobilization of most of these occur via retrotransposition, a process through which elements of RNA intermediates are converted into DNA copies subsequently integrated into the genome (15). *Drosophila* retrotransposons fall into two main classes. "Viral" retrotransposons are represented by copia and over a dozen structurally related sequences (15, 33, 44). These carry long terminal repeats (LTRs) and potentially encode proteins that share homology with the retroviral *gag* and *pol* gene products (15, 33). "Nonviral" or "non-LTR" retrotransposons (44, 46) are a more heterogeneous set of sequences. The group includes F (9), G (10), doc (35), and jockey elements (32); ribosomal DNA insertions (2); and I factors (3). These DNA elements lack terminal repeats, and all feature, except for type I ribosomal insertions and I factors, adenine-rich 3′ ends. Members of non-LTR retrotransposon families vary in size because they are variously truncated at the 5′ end and are flanked by target site duplications of variable length. These are also distinctive features of *LINE-1* elements, the major family of mammalian long interspersed DNA sequences (13, 37). *LINE-1* elements (18), I factors (14), F (7), G (6), jockey (32), and ribosomal DNA insertions (21) potentially encode proteins sharing multiple amino acid motifs with each other as with known reverse transcriptases. Many non-LTR retrotransposons have also coding capacity for an additional protein that features cysteine-rich motifs characteristic of nucleic acid-binding proteins cleaved off from retroviral *gag* polyproteins (5) that promote the annealing of the replication tRNA primer to retroviral RNA (31).

The marked structural difference between LTR and non-LTR retroelements suggests that distinct versions of the process of retrotransposition have coevolved in eucaryotes. In animal retroviruses and copialike elements, RNA intermediates eventually converted into DNA originate from the left LTR. Non-LTR retrotransposons structurally resemble processed pseudogenes, as they lack LTRs, the composite DNA regions generated through the reverse transcription process via the adjoining of RNA segments present at the ends of retroviral genomes (42). It has been therefore hypothesized that transposition intermediates of non-LTR retrotransposons initiate from promoters internal to their 5′ ends (14). Accordingly, a built-in RNA polymerase II-dependent promoter was identified within the 5′ terminus of jockey elements (26). Transient expression assays carried out in *Drosophila* Schneider II cells indicate that the same holds true for F elements. In addition to a promoter that transcribes in the inward direction, we demonstrate that the 5′ end region of F elements includes an additional promoter that drives transcription in the opposite orientation.

MATERIALS AND METHODS

Construction of plasmids. In all cloning procedures, incompatible termini were made blunt ended by T4 polymerase before ligation (25).

(i) F-cat plasmids. An *Sph*I-*Mlu*I fragment derived from λ F12 (9) was cloned in the Bluescript vector (Stratagene) between the *Sph*I and *Eco*RI sites. The DNA segment

* Corresponding author.

includes the first 269 bp from the left terminus of the F12 element flanked by 31 bp of genomic flanking DNA. The *Drosophila* DNA segment, recovered from this derivative as an *Hin*dIII-*Eco*RI fragment, was subsequently cloned into the *Ava*I site of pEMBL8CAT. F-cat1 and F-cat2 differ because of the orientation of the DNA insert with respect to the chloramphenicol acetyltransferase (CAT) gene (see Fig. 1). F-cat1* and F-cat2* were constructed by inserting an *Eco*RI fragment from F-cat1 into the *Sal*I site of pEMBL CAT3 (24).

(ii) 3′-end deletion derivatives of F-cat1 and F-cat2. Plasmids carrying 3′-end deletion derivatives of F-cat1 and F-cat2 were obtained by controlled exonuclease III digestion (20). The presence, in both F-cat1 and F-cat2, of unique, adjacent *Bam*HI and *Pst*I sites located, in a 5′-to-3′ orientation, between the *Drosophila* segment and the CAT gene was exploited to generate unidirectional deletions (20). About 10 μg of F-cat1 or F-cat2 DNA was digested to completion with *Bam*HI and *Pst*I and then incubated at 25°C in 60 mM Tris-HCl (pH 8.0)–5 mM $MgCl_2$ with 200 U of exonuclease III (Biolabs). Aliquots of the reaction mixture, withdrawn at 90-s intervals, were extracted once with phenol, precipitated with ethanol, and resuspended in 0.3 ml of S1 buffer (25). Samples were incubated at 25°C for 30 min with 300 U of S1 nuclease and then phenol extracted and ethanol precipitated. After treatment with T4 DNA polymerase, samples were ligated and used to transform *Escherichia coli* cells. The extent of deletion in each clone was assessed at the nucleotide level by the dideoxy-chain termination method as modified for plasmid DNA minipreparations (19).

(iii) 5′-end deletion derivatives of F-cat1 and F-cat2. 5′-end deletion derivatives were obtained by inverting the orientation of the inserts of some of the 3′ deletions. Clones were digested with *Hin*dIII, that cleaves once to the right of 3′ deletion endpoints, and *Eco*RI. *Eco*RI-*Hin*dIII fragments were cloned into the *Ava*I site of pEMBL8CAT. Recombinants carrying inserts in the desired orientation were identified by restriction and nucleotide sequence analyses.

(iv) Other clones. pGEM180 was constructed by cloning between the *Eco*RI and *Hin*dIII sites of pGEM4 an *Eco*RI-*Hin*dIII fragment from pDm-180 (17) that carries the DNA interval −180 to +34 of a *D. melanogaster* rDNA gene. In pGEM180, the rRNA promoter included in the cloned fragment transcribes toward the bacteriophage T7 promoter.

DNA transfections and CAT assays. *D. melanogaster* Schneider II cells, seeded at a density of 1×10^6 to 2×10^6 cells per ml, were transfected as previously described (8). CAT assays were carried out as reported elsewhere (8), except that 0.1 μCi of [^{14}C]chloramphenicol was used per reaction.

RNA analyses. Forty-eight hours after transfection, cells were harvested, and total RNA was prepared and analyzed by primer extension or S1 mapping as reported previously (17). Reaction products were resolved on 6% polyacrylamide–8 M urea gels. Sequencing ladders were generated by the dideoxy-chain termination method using double-stranded DNAs as templates (19) and the Sequenase kit from United States Biochemical Corp. (Cleveland, Ohio). F_{in} (5′-GGCGTTCACTTCACTCAAAACAACCG-3′) and F_{out} (5′-TTCGATCGCCGACGTGTGAAGACGTT-3′) primers are synthetic 26-mers homologous to different segments of the antisense (F_{in}, +88 to +63) or sense (F_{out}, +15 to +40) strand of F12. Transcripts initiated from the rRNA promoter in cells transfected with pGEM180 were detected by using the T7 primer (5′-TAATACGACTCACTATAGGG-3′; Q5021; Promega). The CAT primer was previously described (17).

Northern RNA analyses were carried out according to standard procedures (25). Five micrograms of poly(A)$^+$ RNA from Schneider II cells was electrophoresed onto a 1% formaldehyde–agarose gel. The gel, blotted onto a Nytran membrane, was hybridized to a 1.7-kb-long *Sph*I-*Apa*LI fragment derived from F12 that spans the 5′-end element region (residues −31 to +1690). Hybridization and washing conditions were as described elsewhere (16).

RNase protection assays were carried out by hybridizing 30 μg of total RNA from Schneider II cells to a ^{32}P-labeled antisense RNA probe obtained by transcription of pGEM-F161. This clone is a derivative of pGEM4 carrying the first 161 bases from the F12 element flanked by genomic DNA and vector sequences. The probe (225 nucleotides in length), produced by transcription with T7 polymerase of pGEM-F161 linearized with *Eco*RI, is complementary to the F sense strand.

RESULTS

Construction and assay of F-cat recombinants. Complete F elements are approximately 4.7 kb in length and potentially encode two proteins that share homology with retroviral RNA-binding proteins and reverse transcriptases (Fig. 1 [7]). Given the hypothesis that F elements propagate by the conversion into DNA of element-length RNA intermediates, a promoter directing the synthesis of F RNA must be located within the 5′ terminus of intact family members. The 5′ end region of F12, a full-length F element previously described (9), was cloned into pEMBL8CAT upstream of the bacterial CAT gene. The cloned DNA interval contains 269 bp from the left side of the element preceded by 31 bp of genomic flanking DNA (Fig. 1). The F-cat1 and F-cat2 recombinants, that differ just because of the orientation of the *Drosophila* DNA segment with respect to the CAT gene, were introduced by transfection into Schneider II *Drosophila* cells (8), and total protein extracts from cells harvested 48 h later were assayed for CAT activity. To our surprise, the cloned DNA was able to drive, in either orientation, synthesis of CAT enzyme (Fig. 1). Hybrid F-CAT transcripts in cells transfected with F-cat1 and F-cat2 originate from two promoters, both included in the F cloned DNA, that direct synthesis of convergently oriented transcripts. This conclusion was reached upon RNA analyses. Exploiting the *Hin*dIII site located in either plasmid between the cloned DNA and the CAT gene, 5′-end-labeled restriction fragments derived from F-cat1 and F-cat2 were used as probes for S1 mapping experiments. A major band of protection was observed with either probe (Fig. 2). Faint bands of higher molecular weight in lane E correspond to partial S1 digestion products visible, in prolonged autoradiogram exposures, also in the control lane F (see also Fig. 7A). The length of the S1-protected bands indicated that F-CAT transcripts initiate from two distinct sites that are approximately 90 bp apart. Similar results were obtained by S1 analysis of RNA from cells transfected with F-cat1* and F-cat2*, two derivatives of pBLCAT3 (24) in which the *Drosophila* DNA insert is in the same orientation with respect to the CAT gene as in F-cat1 and F-cat2 but is flanked by a different DNA region (data not shown). This finding rules out the possibility that the transcripts mapped originate from artifactual promoters generated by juxtaposition of F and vector sequences.

We will herein refer to the promoters that direct the synthesis of F-CAT hybrid transcripts in F-cat1 and F-cat2

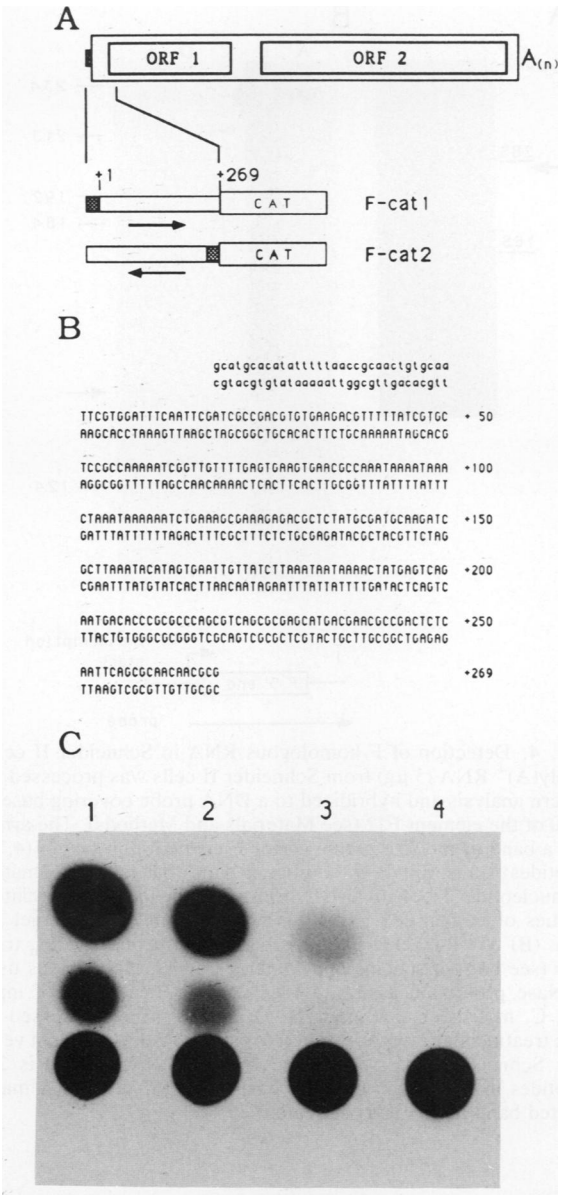

FIG. 1. (A) Structural organization and coding capacity of *Drosophila* F elements. The 5′ DNA interval from F12 (9) was cloned in either orientation in pEMBL8CAT upstream of the bacterial CAT gene. The filled box to the left of residue +1 denotes flanking genomic DNA. ORF, open reading frame. (B) Nucleotide sequence of the *Drosophila* DNA included in F-cat1 and F-cat2. Lowercase and uppercase letters mark flanking and F DNA sequences, respectively. Numbers refer to F residues. (C) Assay of CAT activity in *D. melanogaster* cells. Enzyme activity was measured in equal aliquots of Schneider II cell extracts 48 h after transfection with F-cat1 (lane 1), F-cat2 (lane 2), RSV-CAT (lane 3), and pEMBL8CAT (lane 4).

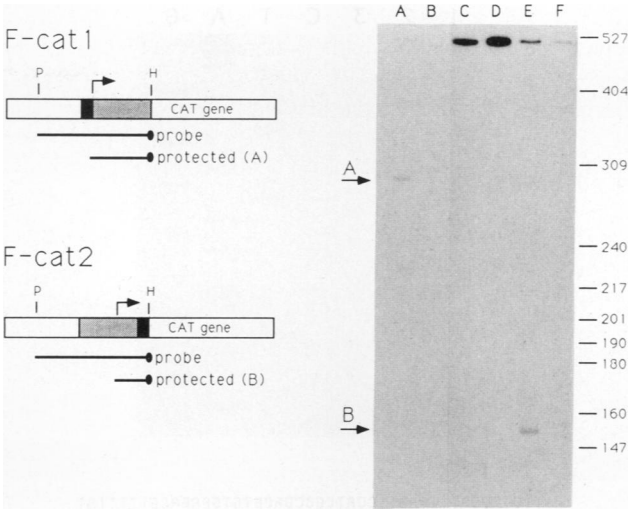

FIG. 2. S1 analysis of F-CAT RNA from transfected cells. Total RNA (20 μg) from Schneider II cells transfected with either F-cat1 (lane A) or F-cat2 (lane E) was hybridized to ^{32}P-5′-end-labeled *Hin*dIII-*Pvu*II fragments prepared from either clone. An equal amount of RNA from untransfected cells was hybridized to the probe from F-cat1 (lane B) or F-cat2 (lane F). Lanes C and D, probes from F-cat1 and F-cat2, respectively. S1-resistant hybrids were sized on a 6% acrylamide–8 M urea gel. Numbers refer to the sizes (in base pairs) of restriction fragments used as molecular size markers. Protected bands are labeled A and B as in the scheme shown on the left. Restriction sites: P, *Pvu*II, H, *Hin*dIII.

as F_{in} (oriented inward the element) and F_{out} (oriented outward the element), respectively.

F_{in} promoter. The size of the S1-protected band indicates that F_{in} transcripts initiate almost at the border with flanking genomic DNA (Fig. 2). By use of a synthetic 26-mer homologous to the antisense DNA strand of F12 from residue +63 to residue +88, the 5′ end(s) of F_{in} transcripts was mapped, by primer extension, around residue +6, i.e., the sixth residue to the right of the target site duplication that defines the F12-element termini (Fig. 3). Multiple adjacent bands of elongation may represent an experimental artifact, as often observed in primer extension assays, but may also denote heterogeneity in the site of initiation; in this view, a faint band barely visible in lane 2 of Fig. 3 may correspond to transcripts initiated around residue +13.

Additional full-size F-family members analyzed at the nucleotide level lack the first 18 (F19) and 5 (F101) residues found in F12 (9). Heterogeneity at the 5′ end, also reported for complete jockey elements (26), may underlie variable cleavages of transposition intermediates (28). Taking into account the length variation between F12 and F101, it is of interest to note that initiation at position +6 in F12 might correspond to initiation at position +1 in F101.

The primer used for the RNA extension experiment spans a DNA region highly conserved in F elements (9), yet we were unable to detect elongation products when total RNA from untransfected cells was assayed (Fig. 3, lane 3). Though at a low level, transcription of endogenous elements seems to take place in cultured cells. Northern analysis showed that transcripts that correspond in length to complete F elements occur in the cellular poly(A)$^+$ RNA population (Fig. 4A). Transcripts homologous to the sense F strand were detected by RNase protection upon hybridization of total RNA from Schneider II cells to an antisense RNA probe including bases 1 to 161 from F12 (Fig. 4B). On the basis of transfection data obtained with F-cat1, a protected band of 155 nucleotides was expected; the size of the band protected in the experiment reported in Fig. 4B (145 to 150 nucleotides) indicates that F-family members expressed in Schneider II cells define a predominant class that slightly differs in sequence content from F12.

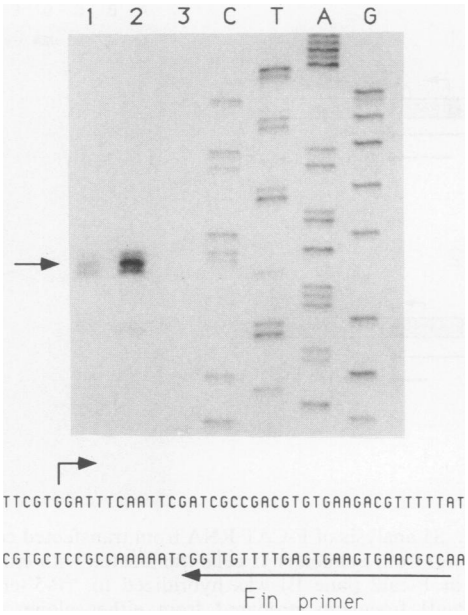

FIG. 3. Primer extension analysis of F-CAT RNA. Total RNA (20 μg) from Schneider II cells transfected with F-cat1 (lanes 1 and 2) or untransfected (lane 3) was hybridized to a ^{32}P-5′-end-labeled 26-mer (F_{in} primer) homologous to the +63 to +88 interval of the antisense F strand. Annealed primers moieties were extended, in the presence of deoxynucleoside triphosphates, by avian reverse transcriptase. Reaction products were run on a 6% acrylamide–8 M urea gel, along with sequencing ladders of the F-cat1 plasmid obtained by the dideoxy-chain termination method using the same primer. The position of the latter and the RNA start site mapped are shown along with the F 5′-end DNA at the bottom. RNA in lanes 1 and 2 was derived from cells transfected in independent experiments.

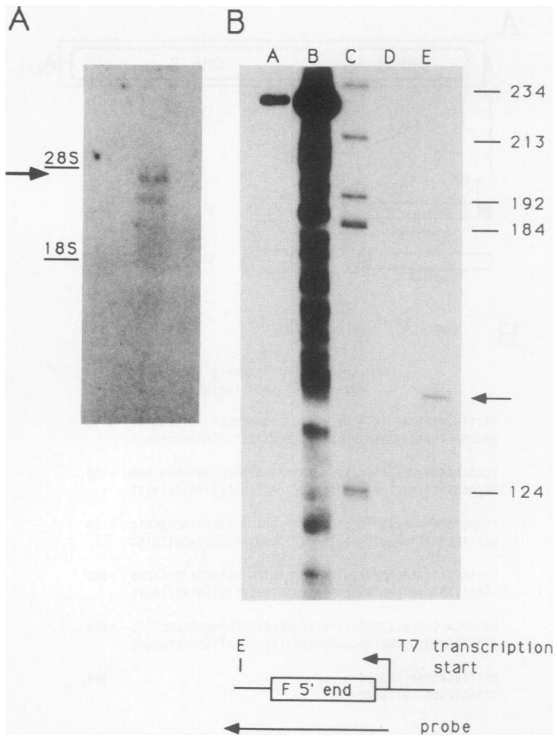

FIG. 4. Detection of F homologous RNA in Schneider II cells. (A) Poly(A)$^+$ RNA (5 μg) from Schneider II cells was processed for Northern analysis and hybridized to a DNA probe covering bases 1 to 1690 of the element F12 (see Materials and Methods). The arrow marks a band of the size predicted for F-element unit-length (4,700 nucleotides) transcripts. An additional band that is approximately 3,200 nucleotides in length also hybridizes to the probe. The relative mobilities of human 28S and 18S rRNA run on the same gel are shown. (B) A ^{32}P-labeled RNA probe complementary to F_{in} transcripts (see bottom scheme and Materials and Methods) was used for RNase protection assays. Lanes: A and B, aliquots of input probe; C, molecular size markers; D and E, probe processed by RNase treatment after hybridization to total RNA (30 μg) from yeast (D) or Schneider II (E) cells. The antisense RNA probe is 225 nucleotides in length and includes bases 1 to 161 of F12. A major protected band in lane E is indicated.

In order to identify sequences involved in the promotion of F_{in} transcription, derivatives of F-cat1 carrying different deletions of the F 5′-end region were generated by exonuclease III digestion (see Materials and Methods). The template activity of these constructs was assayed by S1 protection (Fig. 5A) and primer extension (Fig. 5B). Deletion of sequences from F-cat1 up to F residue +10 and further downstream to +15 and +22 (Fig. 5B) correlated with significant reductions of the level of F_{in} RNA; these were no longer detected with the +54 construct (Fig. 5). Though less abundant, transcripts driven by the +10 template, as by the +15 and +22 templates, are similar in length to those originated from F-cat1 (Fig. 5). This finding suggests that sequences that have a role in the selection of the site(s) of transcription initiation are included in these templates; this hypothesis is reinforced by sequence comparisons of the 5′ termini of *Drosophila* non-LTR retrotransposons (see Fig. 10A and Discussion). An extremely low level of F_{in} transcription may occur, as suggested by barely visible bands in Fig. 5, in cells transfected with the +30 template.

Changes in the accumulation of F_{in} RNA concomitant to the removal of sequences from F-cat1 likely reflect damage of the F_{in} promoter. The significant drop in the F_{in} RNA level observed with the construct +15 suggests that sequences important for transcriptional promotion are located immediately downstream of the RNA start site. However, reduced RNA levels may also be consequent to (i) removal of flanking DNA (genomic sequences and/or F12 residues 1 to 5) that may positively influence F_{in} transcription, (ii) altered transcript stability, and (iii) interference by vector sequences brought in proximity to F DNA. The apparent restimulation of transcription by the +22 deletion in Fig. 5A, not confirmed by the data shown in Fig. 5B or by primer extension assays carried out with RNAs from independent transfection sets, is likely due to incomplete S1 digestion.

An additional effect of removing bases from the F 5′ end is the appearance of sites of initiation within the vector, as indicated by upstream bands of protection in lanes +15, +22, and +30 of Fig. 5A (the highest bands in construct lanes of Fig. 5A are the same size as the corresponding input probes; they either denote far-upstream sites of initiation or represent, more likely, undigested probe molecules).

Accumulation of F_{in} transcripts is also influenced by DNA sequences located far downstream of the F 5′ terminus. This conclusion stems from the results of RNA analyses carried out with 3′ deletion derivatives of F-cat1. In these experiments, cells were cotransfected with equimolar amounts of F-cat1 derivatives and pGEM180, an internal reference template. The latter carries a *D. melanogaster* rDNA gene segment which drives faithful polymerase I dependent tran-

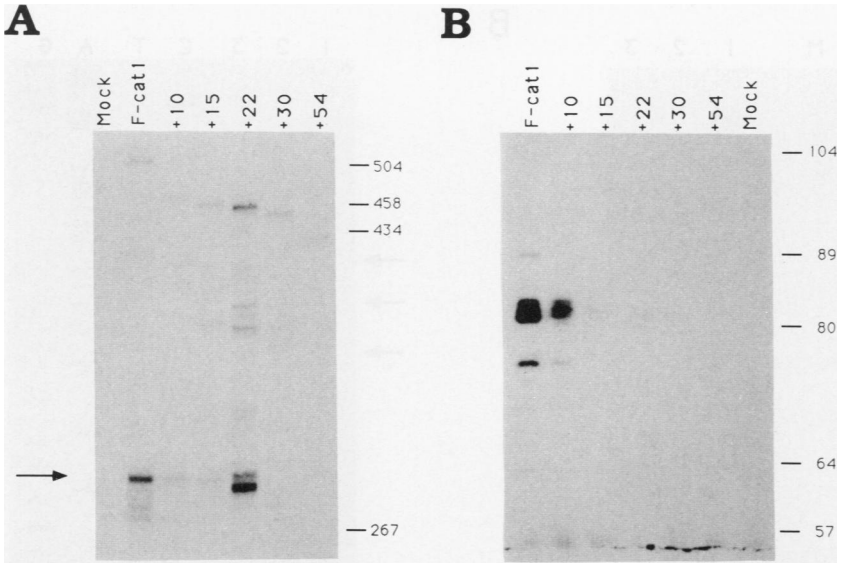

FIG. 5. Analysis of F-CAT RNA from cells transfected with F-cat1 and 5'-deletion derivatives. (A) Total RNA (20 μg) from cells transfected with either F-cat1 or 5'-end deletion derivatives was hybridized to ^{32}P-5'-end-labeled *Hin*dIII-*Pvu*II fragments derived from each recombinant (see scheme in Fig. 2). Products resistant to S1 nuclease digestion were analyzed on a 6% acrylamide–8 M urea gel. Numbers on top denote the extent of 5' truncation in each recombinant. The arrow marks protected bands corresponding to F_{in} transcripts. (B) Primer extension analysis of RNA from cells transfected with either F-cat1 or the same 5'-end deletion derivatives analyzed in panel A. The primer used is as described in the legend to Fig. 3. Numbers to the right of each autoradiogram refer to the size (in base pairs) of restriction fragments used as molecular size markers.

scription in Schneider II cells (17). A polI-dependent template was chosen as an internal control in order to avoid competition for common transcription factors between cotransfected plasmids. RNA was analyzed by primer extension using oligonucleotides specific for each of the cotransfected constructs: the F_{in} primer (Fig. 3) for the F transcripts and the T7 primer for the rRNA hybrid transcripts (see Materials and Methods). Changes in the ratio of elongated products corresponding to F and rRNA transcripts were quantitated by densitometric scanning of autoradiograms corresponding to assays carried out by analyzing RNAs from independent transfection sets. A representative experiment is shown in Fig. 6. Deletion of residues +269 to +218 led to a less than twofold decrease in the level of F_{in} RNA (Fig. 6, lane C). F_{in} RNA was approximately 6- to 8-fold and 12- to 15-fold less abundant in cells transfected with $F_{in}3'$ + 194 and $F_{in}3'$ + 175, respectively (Fig. 6, lanes D and E). No further changes were detected upon removal of the +175 to +99 region (Fig. 6, lanes F to H). Differences in the relative amounts of F-CAT hybrid RNA may reflect altered transcript stability, as well as reduced rates of transcription initiation consequent to the removal of *cis*-acting DNA elements that modulate the activity of the F_{in} promoter.

F_{out} promoter. The F 5' end contains another promoter that transcribes in the opposite orientation relative to the F_{in} promoter (Fig. 1 and 2). The site of initiation of F_{out} transcripts was determined by assaying RNA from cells transfected with F-cat2 both by S1 mapping (see diagram in Fig. 2) and by primer extension, using a synthetic 26-mer specific for F_{out} transcripts (Fig. 7). In agreement with the result in Fig. 2, a major protected band corresponding to transcripts initiated around residue +100 was observed in the S1 nuclease assay (Fig. 7A). On the other hand, three major products of elongation, corresponding to transcripts initiated around residues +92, +97, and +102, were detected assaying the same RNA preparation by primer exten-

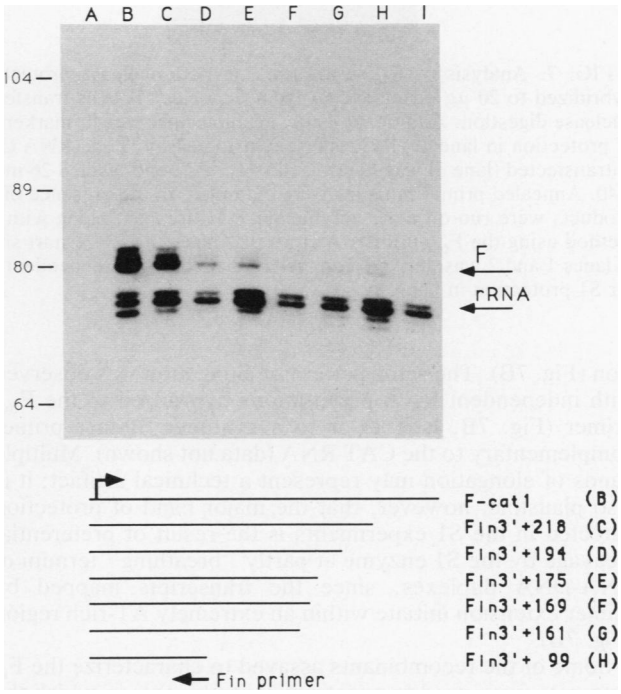

FIG. 6. Primer extension analysis of RNA from cells transfected with F-cat1 and different 3'-deletion derivatives. Total RNA (20 μg) from untransfected cells (lane A), cells cotransfected with 10 μg of each F-derived recombinant and 10 μg of pGEM180 (lanes B to H; letters mark recombinants according to the scheme shown at the bottom), and cells transfected with 10 μg of pGEM180 alone (lane I) was analyzed by primer extension. F_{in} and T7 primers were used to detect transcripts driven by F-derived constructs and pGEM180, respectively (see Materials and Methods). Numbers refer to the size (in base pairs) of restriction fragments used as molecular size markers.

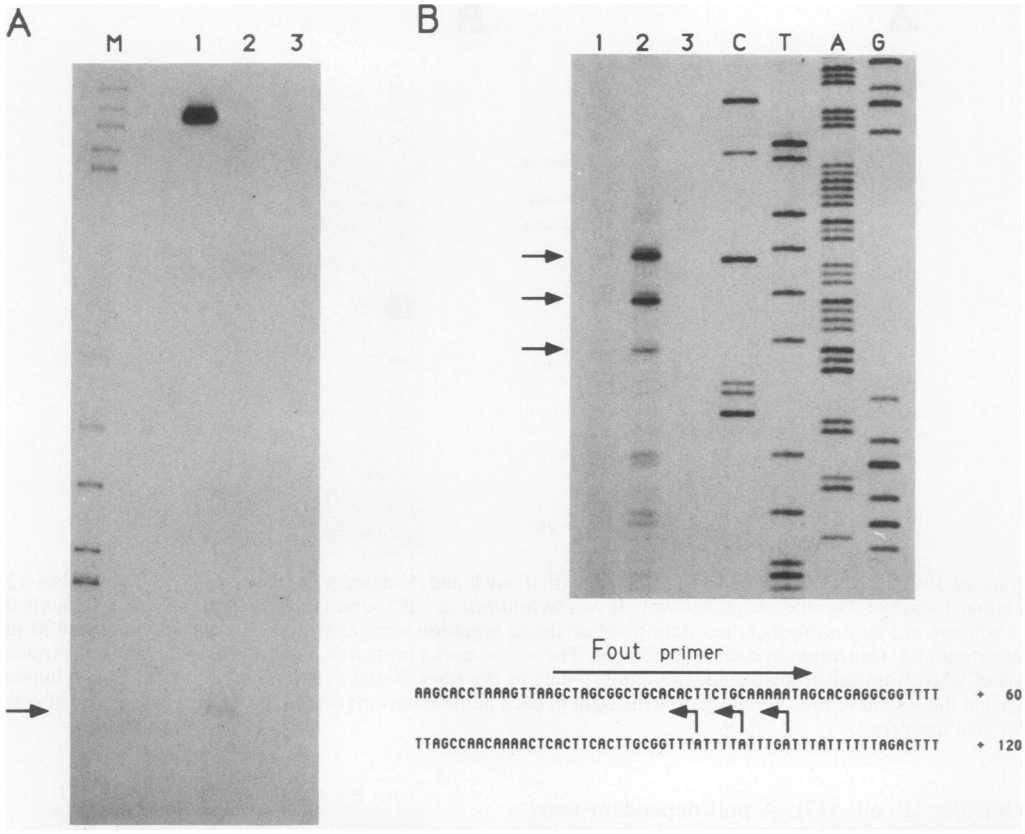

FIG. 7. Analysis of F_{out} transcripts. (A) S1 nuclease protection. A ^{32}P-5′-end-labeled *Hin*dIII-*Pvu*II fragment from F-cat2 (Fig. 2), hybridized to 20 μg of total RNA from Schneider II cells transfected with F-cat2 (lane 2) or untransfected (lane 3), was processed by S1 nuclease digestion. Additional lanes: M, molecular weight markers (marker V; Boheringer); 1, probe alone. The arrow marks a major band of protection in lane 2. (B) Primer extension assay. Total RNA (20 μg) from Schneider II cells transfected with F-cat 2 (lanes 1 and 2) or untransfected (lane 3) was hybridized to a ^{32}P-5′-end-labeled 26-mer (F_{out} primer) homologous to the F12 sense strand from residues +15 to +40. Annealed primer moieties were extended, in the presence of deoxynucleoside triphosphates, by avian reverse transcriptase. Reaction products were run on a 6% acrylamide–8 M urea gel, along with sequencing ladders of F-cat2 obtained by the dideoxy-chain termination method using the F_{out} primer. Arrows that mark the RNA start sites mapped are shown along with the F 5′-end DNA at the bottom. RNA in lanes 1 and 2 was derived from cells transfected in independent experiments; RNA in lane 2 was derived from the same preparation used for S1 protection in panel A.

sion (Fig. 7B). The same pattern of elongation was observed with independent RNA preparations hybridized to the F_{out} primer (Fig. 7B, lane 1) or to a synthetic 30-mer primer complementary to the CAT RNA (data not shown). Multiple bands of elongation may represent a technical artifact; it is also plausible, however, that the major band of protection detected in the S1 experiments is the result of preferential cleavage by the S1 enzyme at partly "breathing" termini of RNA-DNA duplexes, since the transcripts mapped by primer extension initiate within an extremely AT-rich region (Fig. 7B).

Some of the recombinants assayed to characterize the F_{in} promoter were used to construct recombinants in which the F_{out} promoter is oriented toward the CAT gene. For sake of simplicity, deletion endpoints of F_{out} derivatives keep the numbering system used so far. A representative analysis of the activity of some constructs is shown in Fig. 8. Removal of sequences located downstream of the F_{out} start sites ($F_{out}3' + 54$ and $F_{out}3' + 86$) does not impair transcription but rather correlates with an increase in CAT activity. This finding may reflect differences in the relative stability of distinct F-CAT hybrid transcripts, though it cannot be ruled out that it signals enhanced F_{out} transcription. Sequences required for the activity of the F_{out} promoter, as commonly found for *pol* II-dependent promoters, are located upstream of the sites of RNA initiation. CAT activity assays indicated that F_{out} transcription is severely impaired upon deletion of the +161 to +218 interval (Fig. 8). These results were confirmed by primer extension assays carried out with RNA from cells cotransfected with equimolar amounts of either F-cat2 or distinct F_{out} 5′-deletion derivatives and pGEM180 (Fig. 9). Prolonged exposure of the autoradiogram showed that the $F_{out}5' + 161$ construct is able to direct, though poorly, transcription initiation (Fig. 9). A fourfold difference in the CAT enzyme activity level detected in cells transfected with $F_{out}5' + 218$ and F-cat2 (lanes F and G in Fig. 8) was confirmed by the F/rRNA band ratio measured by densitometric scanning of the autoradiogram shown in Fig. 9. A four- to fivefold difference in the activity of $F_{out}5' + 218$ and F-cat2 constructs was observed in independent assays carried out at both the protein and RNA levels (data not shown); this finding suggests that regulatory sequences involved in the control of F_{out} transcription lie within the +218 to +269 DNA interval.

Though the issue was not addressed in further detail, it is possible that the accumulation of F_{in} and F_{out} transcripts is

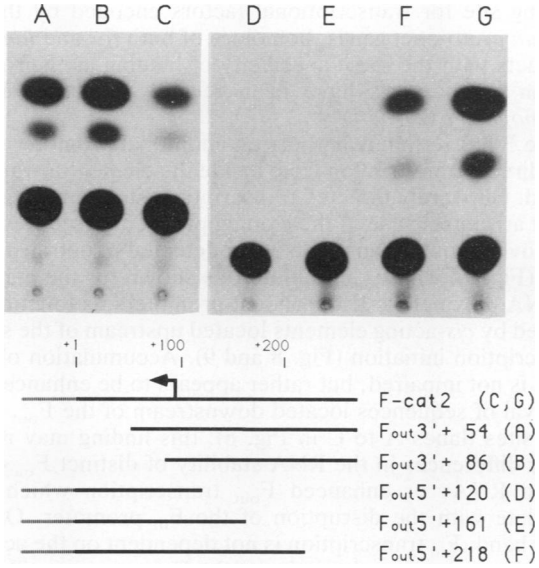

FIG. 8. Assay of CAT activity from cells transfected with F-cat2, F_{out} 5'-, and F_{out} 3'-deletion derivatives. Solid bars denote F DNA; dotted lines denote flanking genomic DNA. Letters in parentheses to the right of each construct correspond to those on top of each lane.

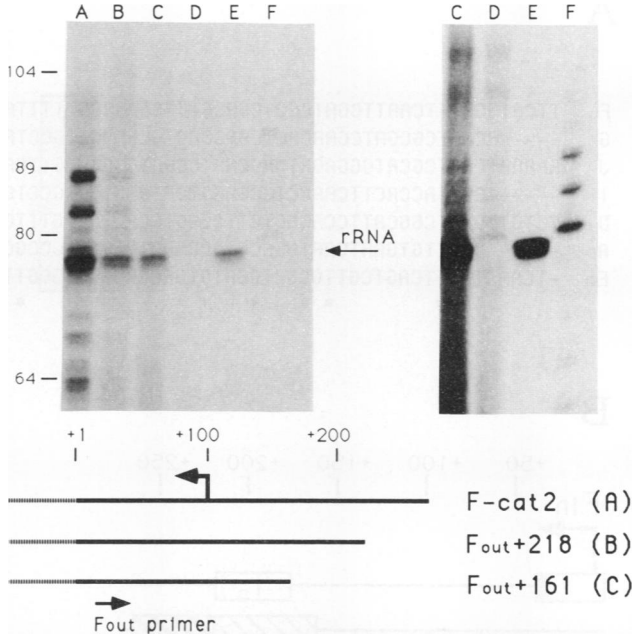

FIG. 9. Primer extension analysis of RNA from cells transfected with F-cat2 and 5'-deletion derivatives. Total RNA (20 μg) from cells cotransfected with 10 μg of pGEM180 and 10 μg of different F recombinants (lanes A to C; letters mark recombinants according to the scheme at the bottom), untransfected (lane D), or transfected with 10 μg of either pGEM180 DNA (lane E) or F-cat2 (lane F) alone was assayed by primer extension by use of the F_{out} and T7 primers. The sizes (in base pairs) of restriction fragments used as markers are indicated. A prolonged exposure of lanes C through to F is shown in the right panel.

mainly influenced by the action of common *cis*-acting regulatory elements included within the +161 to +218 DNA interval.

Outward transcription from within endogenous F elements takes place in Schneider II cells, as elongation products identical in length to those from transfected cells were detected also in untransfected cells upon prolonged film exposure (Fig. 9, lane D).

DNA regions shown to play a role in the accumulation of F_{in} and F_{out} transcripts are outlined in Fig. 10B.

DISCUSSION

F elements belong to the large evolutionarily conserved family of retroelements known as LINEs (11, 13). The structure and coding capacity suggest that LINEs propagate by the self-mediated conversion of RNA intermediates into DNA copies eventually inserted into the genome. The lack of LTRs implies that LINEs are transcribed by promoters internal to their transcriptional units. The results presented in this paper show that an internal promoter is located within the 5' end of complete F-family members. Internal promoters were identified in the 5' end region of *Drosophila* jockey elements (26) and human LINE-1 sequences (41). *Drosophila* I factors also may carry an internal promoter at their 5' end (4). Sensitivity to α-amanitin indicates that expression of endogenous jockey elements is driven by RNA polymerase II (26). The level of endogenous F transcripts did not enable us to perform similar experiments; however, the presence of adenine-rich stretches preceded by polyadenylation signals at the 3' end of family members, the detection of homologous transcripts in the poly(A)$^+$ RNA population from Schneider II cells (Fig. 4), and the translation of hybrid F-CAT mRNAs in transfected cells are all evidences that favor the hypothesis that transcription of F elements is also polII dependent.

A major site of initiation for transcripts directed from the left to the right terminus of F elements (F_{in} transcripts) was mapped around residue +6 of the F-family member analyzed in this work. Deletion of bases immediately downstream of this site led to significant reductions of the level of F_{in} transcripts (lanes +10 and +15 in Fig. 5B). The substantial drop in the RNA accumulation with the +15 construct is reminiscent of what was observed by Mizrokhi and coworkers, who reported that deletion of the first 13 bp from the 5' end of a jockey element abolished jockey transcription (26). Interestingly, the last four of these 13 residues (AAAAAAT CATTCG) are present in the F 5' end twice (bases +1 to +4 and +15 to +18; Fig. 10). The TTCG motif may be therefore part of a regulatory module partly conserved in the two retrotransposons. Intragenic sequences close to the RNA start site involved in the control of transcription initiation were also described in the murine terminal deoxynucleotidyltransferase gene and simian virus 40 major late promoter (1, 38).

Though present at lower levels, transcripts similar in length to those originated from the F-cat1 template are driven by deletion derivatives that lack up to residue +22 and possibly, as suggested by the detection of extremely weak signals, also by the +30 construct (Fig. 5). While the analysis of linker-scanner mutants may help to address directly the issue, it is plausible to assume that a second DNA region located downstream of residue +22 is involved in the promotion of F_{in} transcription. DNA sequence comparisons support the hypothesis that a heptamer sequence (consensus: GACGTGPy) found approximately at the same position with respect to the 5' terminus in F and other *Drosophila* retroelements might play a role in the transcrip-

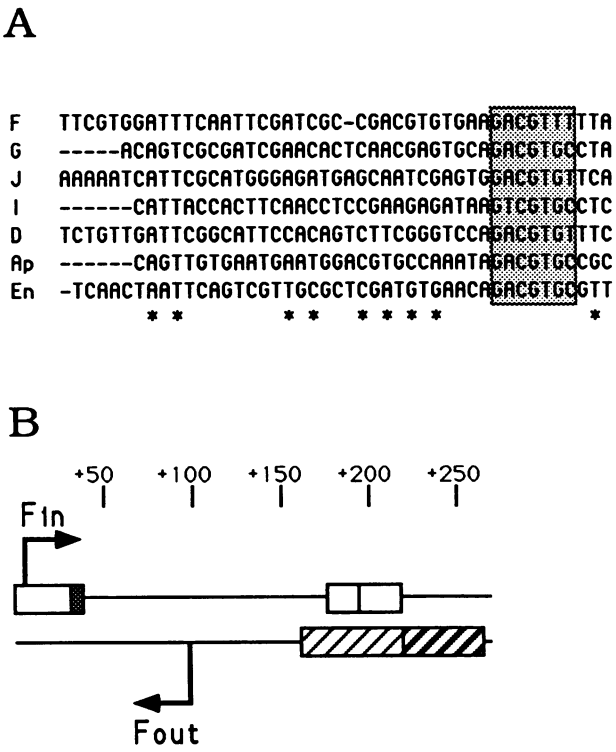

FIG. 10. Regulatory sequences in the 5′ end of F elements. (A) Alignment of sequences at the 5′ end of F, some *Drosophila* non-LTR retrotransposons, and homeotic genes. F, F element; G, G element (6); J, jockey (26); I, I factor (14); D, doc (35); Ap, *Antennapedia* (30); En, engrailed (39). Residues conserved in five or more sequences are highlighted by asterisks. A highly conserved motif is boxed (see Discussion). A dash in the F sequence was introduced to maximize homology. Dashes in other sequences represent nucleotides that are either outside the elements (G and I) or upstream of the site of RNA initiation (Ap and En). (B) DNA segments influencing the accumulation of F_{in} and F_{out} transcripts are represented by empty and hatched boxes, respectively; the filled box corresponds to the region shadowed in panel A.

tion of *Drosophila* LINEs (Fig. 10A). This motif, present once in jockey (GACGTGT, residues 36 to 42) and I factors (GTCGTGC, residues 30 to 36), is duplicated in F elements at a distance of one helix turn (GACGTGT, residues 25 to 31; GACGTTT, residues 35 to 41; Fig. 10A). The hypothesis that this heptamer is a functional promoter component is strongly supported by the finding that its removal inactivates the jockey promoter (27). Homologous motifs also occur downstream of the *cap* site of the *Antennapedia* and engrailed transcripts (Fig. 10A), within DNA regions known to be involved in the control of transcription because they interact with specific factors (30, 39).

Accumulation of F_{in} RNA varies over a 10- to 15-fold range in cells transfected with different 3′-deletion derivatives of F-cat1 (Fig. 6). Changes in the relative abundance of F RNA may reflect either altered transcript stability or reduced transcription initiation consequent to the removal of *cis*-acting regulatory DNA elements, and we cannot at the moment distinguish between these alternative possibilities. Taking into account the latter hypothesis, it is worth noticing that a major drop in the RNA level associated with the deletion of the +218 to +175 F interval (Fig. 6) may correlate with the removal of an AP-1 site (TGAGTCA) located between residues +193 and +199 (Fig. 1). AP-1 is the binding site for transcriptional factors encoded by the *fos* and *jun* proto-oncogenes; homologs of both *fos* and *jun* gene products with the same specificity of binding as their mammalian counterparts have been recently identified in *D. melanogaster* (29).

The F left terminus harbors an additional promoter (F_{out}) that directs transcription from inside the element toward the 5′ end. Outwardly directed transcription is not an artifactual event arisen because of the cloning procedure (see Results); moreover, "out" transcripts were detected in untransfected cells (Fig. 9). The F_{out} promoter, as shown for the majority of RNA polymerase II-dependent promoters, is functionally defined by *cis*-acting elements located upstream of the site of transcription initiation (Fig. 8 and 9). Accumulation of F_{out} RNA is not impaired, but rather appears to be enhanced, by removal of sequences located downstream of the F_{out} RNA start sites (lanes A to C in Fig. 8); this finding may reflect either differences in the RNA stability of distinct F_{out}-CAT hybrid RNAs or enhanced F_{out} transcription which may correlate with the disruption of the F_{in} promoter. On the other hand, F_{in} transcription is not dependent on the activity of the F_{out} promoter. Initiation at the F_{in} promoter still takes place from the $F_{in}3'$ + 99 construct that lacks the F_{out} promoter (Fig. 6).

The functional significance of outwardly oriented transcription in F elements remains to be established. A few eucaryotic transcriptional units that partly overlap at the 3′ untranslated regions have been described previously (40, 45); to our knowledge, however, adjacently located promoters that direct the synthesis of convergent, overlapping transcripts have been identified so far only in procaryotes (34, 36, 43). Gene activity can be regulated through symmetric transcription by different mechanisms. RNA polymerase complexes travelling in opposite directions along a DNA duplex might encounter a considerable degree of steric hindrance, and reduced transcription because of interference of RNA polymerase progression has been indeed documented for adjacent genes organized in opposite and convergent orientations (12, 23, 34, 43). Gene expression can be negatively regulated, via the formation of sense-antisense RNA hybrids, at the level of translation (36) as well as of nuclear export (22). Transcription from the F_{in} promoter is not significantly enhanced in the absence of F_{out} promoter activity and vice versa (Fig. 6 and 8). Factors required for the activity of both promoters are likely present in limited amounts in Schneider II cells; transcriptional interferences might be therefore overlooked in assays carried out at relatively high DNA inputs that disfavor the concomitant assembly of adjacent transcriptional complexes over the same template molecule. Accumulation of F_{in} and F_{out} transcripts is significantly reduced in cells transfected with templates that lack the +161 to +218 DNA interval. The use of common *cis*-acting elements may be a simple mechanism by which the rate of transcription at the F_{in} and F_{out} promoters is modulated in the organism.

Future investigations might clarify whether convergent transcription plays any role in the life cycle of F elements. Clues in this direction might come from a thorough investigation of the pattern of expression of F_{in} and F_{out} transcripts in the fly. It will be of interest as well to know whether the presence of "out" promoters might turn out to be a general feature of LINE-like elements.

ACKNOWLEDGMENTS

We thank Cristina Contursi for the construction of some exonuclease III derivatives and Edoardo Boncinelli for suggestions. We

are indebted to Giovanna Grimaldi for critical reading of the manuscript.

This work was supported by grants from the Ministry of Education, Progetto Finalizzato Biotecnologie e Biostrumentazione, CNR, and P. F. Ingegneria Genetica, CNR. G.M. was the recipient of a CNR fellowship.

REFERENCES

1. **Ayer, D. A., and W. S. Dynan.** 1988. Simian virus 40 major late promoter: a novel tripartite structure that includes intragenic sequences. Mol. Cell. Biol. **8**:2021–2033.
2. **Beckingham, K.** 1982. Insect rDNA, p. 206–269. *In* H. Busch and L. Rothblum (ed.), The cell nucleus, vol. X. rDNA, part A. Academic Press, Inc., New York.
3. **Bucheton, A., R. Paro, H. M. Sang, A. Pelisson, and D. J. Finnegan.** 1984. The molecular basis of I-R hybrid dysgenesis: identification, cloning and properties of the I factor. Cell **38**:153–163.
4. **Chaboissier, M. C., I. Busseau, J. Prosser, D. Finnegan, and A. Bucheton.** 1990. Identification of a potential RNA intermediate for transposition of the LINE-like element I factor in *Drosophila melanogaster*. EMBO J. **9**:3557–3563.
5. **Covey, S. N.** 1986. Amino acid sequence homology in *gag* region of reverse transcribing elements and the coat protein gene of cauliflower mosaic virus. Nucleic Acids Res. **14**:623–633.
6. **Di Nocera, P. P.** 1988. Close relationship between non-viral retroposons in *Drosophila melanogaster*. Nucleic Acids Res. **16**:4041–4052.
7. **Di Nocera, P. P., and G. Casari.** 1987. Related polypeptides are encoded by *Drosophila* F elements, I factors and mammalian L1 sequences. Proc. Natl. Acad. Sci. USA **84**:5843–5847.
8. **Di Nocera, P. P., and I. B. Dawid.** 1983. Transient expression of genes introduced into cultured cells of *Drosophila*. Proc. Natl. Acad. Sci. USA **80**:7095–7098.
9. **Di Nocera, P. P., M. E. Digan, and I. B. Dawid.** 1983. A family of oligo-adenylated transposable sequences in *Drosophila melanogaster*. J. Mol. Biol. **168**:715–728.
10. **Di Nocera, P. P., F. Graziani, and G. Lavorgna.** 1986. Genomic and structural organisation of *Drosophila melanogaster* G elements. Nucleic Acids Res. **14**:675–691.
11. **Doolittle, R. F., D. F. Feng, M. S. Johnson, and M. A. McClure.** 1989. Origins and evolutionary relationships of retroviruses. Q. Rev. Biol. **64**:1–30.
12. **Elledge, S. J., and R. W. Davis.** 1989. Position and density effects on repression by stationary and mobile DNA-binding proteins. Genes Dev. **3**:185–197.
13. **Fanning, T. G., and M. F. Singer.** 1987. LINE-1: a mammalian transposable element. Biochim. Biophys. Acta **910**:203–212.
14. **Fawcett, D. H., C. K. Lister, E. Kellett, and D. J. Finnegan.** 1986. Transposable elements controlling I-R hybrid dysgenesis in *D. melanogaster* are similar to mammalian LINEs. Cell **47**:1007–1015.
15. **Finnegan, D. J., and D. H. Fawcett.** 1986. Transposable elements in *Drosophila melanogaster*, p. 1–62. *In* N. Maclean (ed.), Oxford surveys on eukaryotic genes, vol. 3. Oxford University Press, Oxford.
16. **Franzè, A., N. Archidiacono, M. Rocchi, M. Marino, and G. Grimaldi.** 1991. Isolation and expression analysis of a human zinc finger gene (ZNF41) located on the short arm of the X chromosome. Genomics **7**:728–736.
17. **Grimaldi, G., and P. P. Di Nocera.** 1988. Multiple repeated DNA units in *Drosophila melanogaster* ribosomal DNA spacer stimulate rRNA precursor transcription. Proc. Natl. Acad. Sci. USA **85**:5502–5506.
18. **Hattori, M., S. Kuhara, O. Takenaka, and Y. Sakaki.** 1986. L1 family of repetitive DNA sequences in primates may be derived from a sequence encoding a reverse transcriptase-related protein. Nature (London) **321**:625–628.
19. **Hattori, M., and Y. Sakaki.** 1986. Dideoxy sequencing method using denatured plasmid templates. Anal. Biochem. **152**:232–238.
20. **Henikoff, S.** 1984. Unidirectional digestion with exonuclease III creates targeted breakpoints for DNA sequencing. Gene **28**:351–359.
21. **Jakubczak, J. L., Y. Xiong, and T. H. Eickbush.** 1990. Type I (R1) and type II (R2) ribosomal DNA insertions of *Drosophila melanogaster* are retrotransposable elements closely related to those of *Bombyx mori*. J. Mol. Biol. **212**:37–52.
22. **Kim, S. K., and B. J. Wold.** 1985. Stable reduction of thymidine kinase activity in cells expressing high levels of antisense RNA. Cell **42**:129–138.
23. **Lloubes, R., D. Baty, and C. Lazdunski.** 1986. The promoters of the genes for colicin production, release and immunity in the ColA plasmid: effects of convergent transcription and Lex A protein. Nucleic Acids Res. **14**:2621–2636.
24. **Luckow, B., and G. Schutz.** 1987. CAT constructions with multiple unique restriction sites for the functional analysis of eukaryotic promoters and regulatory elements. Nucleic Acids Res. **15**:5490.
25. **Maniatis, T., E. F. Fritsch, and J. Sambrook.** 1982. Molecular cloning: a laboratory manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
26. **Mizrokhi, L. J., S. G. Georgieva, and Y. V. Ilyin.** 1988. *Jockey*, a mobile *Drosophila* element similar to mammalian LINEs, is transcribed from the internal promoter by RNA polymerase II. Cell **54**:685–691.
27. **Mizrokhi, L. J., and A. M. Mazo.** 1990. Evidence for horizontal transmission of the mobile element *jockey* between distant *Drosophila* species. Proc. Natl. Acad. Sci. USA **87**:9216–9220.
28. **Mossie, K. G., M. W. Young, and H. E. Varmus.** 1985. Extrachromosomal DNA forms of *copia*-like transposable elements, F elements and middle repetitive DNA sequences in *Drosophila melanogaster*. J. Mol. Biol. **182**:31–43.
29. **Perkins, K. K., G. M. Dailey, and R. Tjian.** 1988. Novel Jun- and Fos-related proteins in Drosophila are functionally homologous to enhancer factor AP-1. EMBO J. **7**:4265–4273.
30. **Perkins, K. K., G. M. Dailey, and R. Tjian.** 1988. In vitro analysis of the *Antennapedia* P2 promoter: identification of a new *Drosophila* transcription factor. Genes Dev. **2**:1615–1626.
31. **Prats, A. C., L. Sarih, C. Gabus, S. Litvak, G. Keith, and J. L. Darlix.** 1988. Small finger protein of avian and murine retroviruses has nucleic acid annealing activity and positions the replication primer tRNA onto genomic RNA. EMBO J. **7**:1777–1783.
32. **Priimagi, A. F., L. J. Mizrokhi, and Y. V. Ilyin.** 1988. The *Drosophila* mobile element *jockey* belongs to LINEs and contains coding sequences homologous to some retroviral proteins. Gene **70**:253–262.
33. **Rubin, G. M.** 1983. Dispersed repetitive DNAs in *Drosophila*, p. 329–361. *In* J. Shapiro (ed.), Mobile genetic elements. Academic Press, Inc., New York.
34. **Saha, S., E. Haggard-Ljungquist, and K. Nordstrom.** 1987. The cox protein of bacteriophage P2 inhibits the formation of the repressor protein and autoregulates the early operon. EMBO J. **6**:3191–3199.
35. **Schneuwly, S., A. Kuroiwa, and W. J. Gehring.** 1987. Molecular analysis of the dominant homeotic *Antennapedia* phenotype. EMBO J. **6**:201–206.
36. **Simons, R. W., and N. Kleckner.** 1983. Translational control of IS 10 transposition. Cell **34**:683–691.
37. **Singer, M. F., and J. Skowronski.** 1985. Making sense out of LINES: long interspersed repeat sequences in mammalian genomes. Trends Biol. Sci. **10**:119–122.
38. **Smale, S. T., and D. Baltimore.** 1989. The "initiator" as a transcriptional control element. Cell **57**:103–113.
39. **Soeller, W. C., S. J. Poole, and T. Kornberg.** 1988. In vitro transcription of the *Drosophila engrailed* gene. Genes Dev. **2**:68–81.
40. **Spencer, C. A., D. Gietz, and R. B. Hodgetts.** 1986. Overlapping transcription units in the dopa decarboxylase region of *Drosophila*. Nature (London) **322**:279–281.
41. **Swergold, G. D.** 1990. Identification, characterization, and cell specificity of a human LINE-1 promoter. Mol. Cell. Biol. **10**:6718–6729.

42. **Varmus, H. E.** 1983. Retroviruses, p. 411–505. *In* J. Shapiro (ed.), Mobile genetic elements. Academic Press, Inc., New York.
43. **Ward, D. F., and N. E. Murray.** 1979. Convergent transcription in bacteriophage lambda: interference with gene expression. J. Mol. Biol. **133**:249–266.
44. **Weiner, A. M., P. L. Deininger, and A. Efstratiadis.** 1986. Nonviral retroposons: genes, pseudogenes, and transposable elements generated by the reverse flow of genetic information. Annu. Rev. Biochem. **55**:631–661.
45. **Williams, T., and M. Fried.** 1986. A mouse locus at which transcription from both DNA strands produces mRNAs complementary at their 3′ ends. Nature (London) **322**:275–279.
46. **Xiong, Y., and T. H. Eickbush.** 1988. The site-specific ribosomal DNA insertion element R1Bm belongs to a class of non-long-terminal-repeat retrotransposons. Mol. Cell. Biol. **8**:114–123.