



THE UNIVERSITY *of* EDINBURGH

Edinburgh Research Explorer

The 5' untranslated region of the I factor, a long interspersed nuclear element-like retrotransposon of *Drosophila melanogaster*, contains an internal promoter and sequences that regulate expression

Citation for published version:

MCLEAN, C, BUCHETON, A, FINNEGAN, DJ & Finnegan, D 1993, 'The 5' untranslated region of the I factor, a long interspersed nuclear element-like retrotransposon of *Drosophila melanogaster*, contains an internal promoter and sequences that regulate expression' *Molecular and Cellular Biology*, vol 13, no. 2, pp. 1042-1050., 10.1128/MCB.13.2.1042

Digital Object Identifier (DOI):

[10.1128/MCB.13.2.1042](https://doi.org/10.1128/MCB.13.2.1042)

Link:

[Link to publication record in Edinburgh Research Explorer](#)

Document Version:

Publisher final version (usually the publisher pdf)

Published In:

Molecular and Cellular Biology

Publisher Rights Statement:

RoMEO blue

General rights

Copyright for the publications made accessible via the Edinburgh Research Explorer is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

Take down policy

The University of Edinburgh has made every reasonable effort to ensure that Edinburgh Research Explorer content complies with UK legislation. If you believe that the public display of this file breaches copyright please contact openaccess@ed.ac.uk providing details, and we will remove access to the work immediately and investigate your claim.



The 5' Untranslated Region of the I Factor, a Long Interspersed Nuclear Element-Like Retrotransposon of *Drosophila melanogaster*, Contains an Internal Promoter and Sequences That Regulate Expression

CAROL McLEAN,¹ ALAIN BUCHETON,² AND DAVID J. FINNEGAN^{1*}

Institute of Cell and Molecular Biology, University of Edinburgh, King's Buildings, Edinburgh EH9 3JR, Scotland,¹ and Centre de Génétique Moléculaire, Centre National de la Recherche Scientifique, 91198 Gif-sur-Yvette Cedex, France²

Received 3 August 1992/Returned for modification 6 October 1992/Accepted 20 November 1992

The I-R system of hybrid dysgenesis in *Drosophila melanogaster* is controlled by a long interspersed nuclear element-like retroposon, the I factor. Transposition of the I factor occurs at a high frequency only in the ovaries of females produced by crossing males of inducer strains that contain functional I factors with females of reactive strains that lack them. In this study, the 5' untranslated region of the I factor was joined to the chloramphenicol acetyltransferase gene, and activity was assayed in transfected *D. melanogaster* tissue culture cells and transformed flies. The results have identified a promoter that lies within the first 186 bp of the I factor. Deletion analysis shows that nucleotides +1 to +40 are sufficient for high promoter activity and accurate transcription initiation. This region contains sequences that are found in a class of RNA polymerase II promoters that lack both a TATA box and CpG-rich motifs. In transformed flies, high levels of expression from nucleotides +1 to +186 are confined to the ovaries of reactive females, suggesting that the promoter is involved in the tissue and cytotypic specificity of transposition.

The I factor is the transposable element that controls the I-R system of hybrid dysgenesis in *Drosophila melanogaster*. Strains of *D. melanogaster* can be categorized as being one of two types with respect to I-R dysgenesis: inducer strains that are estimated to have 10 to 15 complete I factors, or reactive strains that have none (8). Both inducer and reactive strains have defective copies of the I factor located in pericentromeric chromatin (8, 13). I-R hybrid dysgenesis is seen in the female progeny from crosses between inducer males and reactive females (7). These females, called SF females, lay a normal number of eggs, but as many as 100% may fail to hatch. This sterility correlates with a high frequency of I factor transposition in the germ line. The reciprocal cross between inducer females and reactive males produces RSF daughters that are fertile and show a much reduced rate of I-factor transposition (49). I factors are stable in inducer strains in which transposition occurs only at a very low frequency (23). Transposition has never been observed in males or in the somatic tissues of females (46, 49).

The structure of the I factor is similar to that of long interspersed nuclear elements (LINEs), (19), which are repeated elements found in a wide range of mammals (10) and in several other eukaryotic species such as *Bombyx mori* (9, 70), trypanosomids (21, 31, 42), *Neurospora crassa* (32), and *Zea mays* (60). They are distinct from other transposable elements in that they lack terminal repeats and have an adenine-rich sequence at the 3' end of the plus strand. This strand normally encodes two open reading frames (ORFs). The 5' ORF (ORF1) encodes a polypeptide with cysteine fingers like those in retroviral *gag* genes, and the 3' ORF (ORF2) encodes putative RNase H and reverse transcriptase domains. For some LINE-like elements, the reverse tran-

scriptase domain has been shown to be functional (20, 25, 38).

Analysis of the structure and coding capacity of LINEs has led to a model for their transposition in which a full-length RNA, corresponding to the plus strand, is reverse transcribed and the DNA copy is integrated at a new site (53, 60). Introns inserted within the I factor are removed during transposition, confirming that an RNA intermediate is involved (28, 47). A candidate for this RNA has been identified, the presence and abundance of which correlate with the frequency of transposition (11). It is most abundant in SF ovaries, is rarer in RSF ovaries, is barely detectable in inducer females, and is undetectable in female somatic tissues and in males. Chaboissier et al. (11) concluded that I-factor transposition is controlled, at least in part, at the level of transcription or RNA stability.

The 5' end of the putative I-factor transposition intermediate lies at or very close to position +1 (11). There is no evidence that the I factor depends on flanking DNA sequences for transcription, as functional elements that are inserted downstream of sequences that apparently lack promoter properties have been described (8, 44, 53). These results suggest that the promoter for the transposition intermediate should lie entirely within the I factor. Such a promoter would be unlike classical RNA polymerase II promoters but more like those recognized by RNA polymerase III. Polymerase III is, however, unlikely to transcribe the I factor, as the template strand contains several oligo(T) sequences which would serve as RNA polymerase III terminators (4), and the first 200 bp lack any regions similar to polymerase III promoter consensus sequences (19). This finding suggests that the I factor contains an internal RNA polymerase II promoter (19). Two other *D. melanogaster* LINE-like elements, jockey and F (39, 40), have been shown to have internal promoters, and at least

* Corresponding author.

jockey is transcribed by RNA polymerase II. An RNA polymerase II promoter internal to the human LINE element LHs has also been described (66).

The results of the experiments reported in this paper demonstrate that there is a promoter within the 186-bp 5' untranslated region (UTR) of the I factor that initiates transcription at nucleotide +1. This promoter is most active in ovaries from reactive strains, and expression is reduced in the inducer state. This finding suggests that tissue specificity and genetic control of I-factor transposition reflect the activity of the promoter responsible for synthesis of the transposition intermediate.

MATERIALS AND METHODS

Drosophila stocks. The reactive line used as a host in P-element transformation was W^K (w^- ; +; +) (36). The inducer lines were line 3 (*SM6/Cy*; *Dcxf/Sb*), having cross over suppressors on chromosomes 2 and 3, and line 125, which carries the Muller-5 X chromosome [= *Basc*, *In(1)sc^{SIL}sc^{RR+S}*, *sc^{S1}sc⁸ w^a B*] (35). All stocks were maintained at 25°C.

Cell culture. *D. melanogaster* Schneider line 2 cells (59) were maintained at 23°C in Schneider's *Drosophila* medium (revised) (Imperial Laboratories) supplemented with 5% fetal calf serum (Sera-lab), 100 U of penicillin per ml, and 100 µg of streptomycin (GIBCO) per ml.

Construction of plasmids. pCAT.1 was constructed by inserting the *Bam*HI-*Sac*I fragment of pC4CAT (68) containing the chloramphenicol acetyltransferase (CAT) gene and simian virus 40 (SV40) polyadenylation sequences into the *Bam*HI-*Sac*I site of pUC18, followed by insertion of the *Pst*I linker, GTACTGCAGTACAGCT, into the *Sac*I site, thus abolishing that site. Promoter fragments were synthesized by the polymerase chain reaction (PCR), using oligonucleotides with an *Xba*I site at the 5' end for the plus-strand primer and a *Bam*HI site at the 5' end for the minus-strand primer and 20 nucleotides complementary to the desired region of the I-factor 5' UTR (see reference 19 for sequence). The PCR products were cloned into the *Xba*I-*Bam*HI sites of M13mp18, sequenced by the dideoxy method, and subcloned into the *Xba*I-*Bam*HI site of pCAT.1. The promoter fragment in p186rp was made by PCR except that the plus-strand primer had a *Bam*HI site and the minus-strand primer had an *Xba*I site and nucleotide +3 was a G rather than a T so that there was no methionine codon in the reversed promoter. p12G was made by annealing two complementary oligonucleotides, CTAGACAGTACCACTTCG and GATCCGAAGTGGTACTGT, which gives a double-stranded product with an *Xba*I 5' plus-strand overhang and a *Bam*HI 5' minus-strand overhang. This double-stranded product was ligated into pCAT.1 and subjected to plasmid sequencing with the primer CAACGGTGGTATATCCAGT GATTTT, which is complementary to a region within the 5' end of the CAT gene. p40λT (see Fig. 2) was made by inserting a 118-bp λcI857 *Sau*3A fragment (coordinates 33327 to 33206 on the minus strand [14]) into the *Bam*HI site of p40T and then subjected to plasmid sequencing as described above.

The P-element transformation vector p186W8 was made by ligating the *Pst*I fragment of p186T.1, containing the I-factor promoter, CAT gene, and SV40 polyadenylation sequences, into the *Pst*I site of pW8 (33) in an orientation such that transcription of I is opposite that of the white (*w*) selectable marker gene and P promoter. The control trans-

formation plasmid pCATW8 was made as described above except the *Pst*I fragment was from pCAT.1.

Transfections and CAT assays. *D. melanogaster* Schneider line 2 cells (approximately 7×10^7) in 7 ml of medium were transfected with 10 µg of test construct and 5 µg of internal control plasmid p-194.70Z (69), using the calcium phosphate method (22). Plasmid p-194.70Z contains the *Escherichia coli lacZ* gene under control of the *D. melanogaster hsp70* promoter. After 48 h, cells were heat shocked at 37°C for 20 min and then returned to 23°C for 90 min. Extracts were prepared, and CAT assays performed as described by Gorman et al. (22) except that 12.5 nCi of [¹⁴C]chloramphenicol (Amersham) was used per reaction. Transfection efficiency was estimated by measuring the β-galactosidase activity of the extract as described by Sun et al. (65), and the CAT activity was normalized accordingly.

Primer extension. Primer extensions were carried out as described by Kamakaka et al. (29), using a *Drosophila* embryo transcription extract from Stratagene. The primer used was the same as described above for plasmid sequencing. RNA polymerase II activity was inhibited by including 4 µg of α-amanatin per ml.

Transgenic lines. P-element transformation was performed by embryo injection as described by Rubin and Spradling (58), using the helper plasmid pΠ25.7wc (30). Inducer derivatives of reactive transformed lines were made by crossing males with the genotype w^a *M5B*/+; *Cy*/+; *Sb*/+ (produced by crossing males of line 3 with females from line 125) with homozygous female transformants. SF female progeny with the genotype w^a *M5B*/ w^- ; *Cy*/+; *Sb*/+ (w^+ *CAT*) were then selected and mated with W^K males. Heterozygous (w^+ *CAT*) progeny that were otherwise wild type were selected and used to establish homozygous lines by sibling matings.

Drosophila CAT assays. Ten males and ten females of each stage, ten pairs of ovaries and ten carcasses of adult flies, or several hundred 0- to 2-h embryos were collected, homogenized in 100 to 200 µl of 0.25 M Tris (pH 7.8), passed through five freeze-thaw cycles to break open the cells, and then incubated at 65°C for 5 min to inactivate inhibitors of the CAT enzyme. Cell debris was removed by centrifugation in a microcentrifuge, and the supernatant was used as the CAT extract. CAT assays were performed as described above, and the activity was normalized to the concentration of protein measured from a 5-µl sample by the Bradford assay (5, 55).

RESULTS

Detection of I-factor promoter activity and 5' deletion analysis. Nucleotides +1 to +186 of the I factor (see reference 19 for sequence) were selected as a putative promoter region, as they make up the entire 5' UTR. This sequence was amplified by PCR from plasmid pI954, as this construct contains a complete I factor that can transpose and induce hybrid dysgenesis when introduced into reactive flies by P-element transformation (53). The amplified product was first cloned into M13mp18 to check its sequence and then subcloned into plasmid pCAT.1, containing the reporter CAT gene, to form p186T.1 (Fig. 1) (see Materials and Methods for construction of pCAT.1).

Promoter activity was tested by transient assays in *Drosophila* Schneider line 2 cells. Figure 2 shows that p186T.1 directs high CAT activity relative to the low level of activity from pCAT.1. The same pattern of expression was seen when the *Pst*I fragments in p186T.1 and pCAT.1 were inverted (data not shown), indicating that this result was not

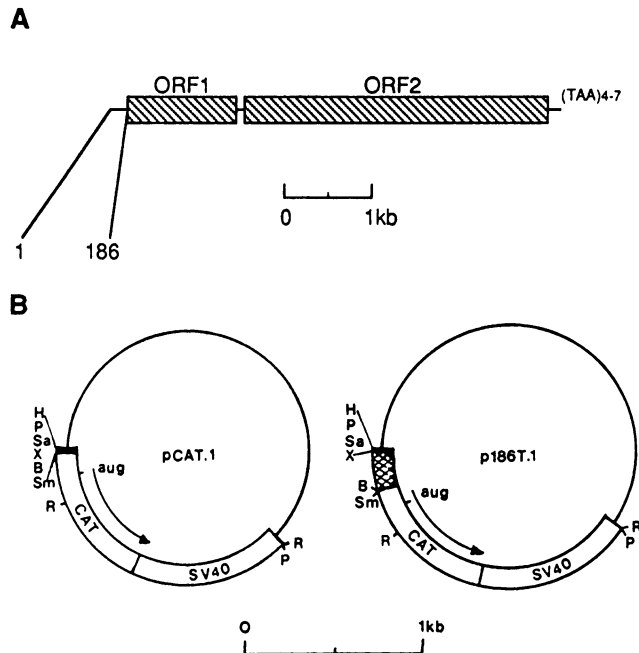


FIG. 1. (A) Structure of the I factor illustrating the region selected as the putative promoter. ORFs are indicated as striped boxes; UTRs are indicated as single lines. (B) Restriction maps of pCAT.1 and p186T.1. Abbreviations for restriction sites: B, *Bam*HI; H, *Hind*III; P, *Pst*I; R, *Eco*RI; Sa, *Sal*I; Sm, *Sma*I; X, *Xba*I. aug represents the methionine codon of the CAT gene, and the arrow indicates the direction of CAT transcription. SV40 sequences include the polyadenylation signal of the small t antigen. I-factor sequences are represented by a crossed box.

due to sequences immediately upstream from the I-factor sequences. The 5' UTR of the I factor therefore contains a promoter. This region does not contain divergent promoters as was found in the F element (39) because nucleotides 1 to 186 of the I factor do not direct CAT activity when inserted into pCAT.1 in the opposite direction (plasmid p186rp; Fig. 2).

Several other fragments (illustrated in Fig. 2) with progressive 3' deletions were synthesized by PCR and cloned into pCAT.1. Deletion of sequences to position +100 (p100T) has very little effect on CAT activity. However, deletion to +40 (p40T) increases activity twofold, suggesting that sequences between positions +40 and +100 suppress promoter activity. This increase is also seen when 118 bp of phage λ DNA are inserted in p40T between the promoter and the CAT gene, forming p40T λ . This finding makes it unlikely that this increase in activity is due to the formation of an enhancer-like sequence at the CAT/I factor boundary in p40T or to a more favorable location of a cryptic promoter within the vector sequences.

Deletion to position +28 reduces CAT activity twofold relative to that from p40T, suggesting that sequences between +28 and +41 play a positive role in transcription from the I-factor promoter. Further deletion to +12 reduces CAT activity almost to background levels, suggesting that the 5' limit of the I-factor promoter lies between positions +12 and +28.

Comparison of the I-factor promoter with TATA-less promoters and LINES. The majority of RNA polymerase II promoters require either an upstream TATA box or CpG-rich sequences for transcription initiation. The apparent

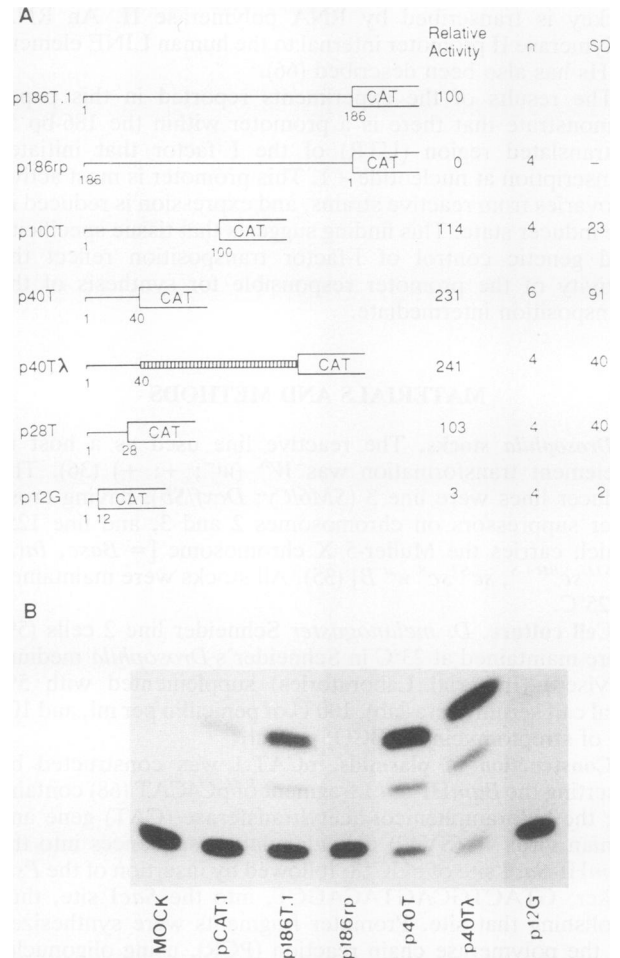


FIG. 2. Promoter activity from the I-factor 5' UTR. (A) Line drawings showing the extents of the various deletion end points. λ c1857 sequences are shown as the vertical striped box. Relative activity indicates the CAT expression level relative to that of p186T.1 (taken as 100% in each experiment) after normalization to the transfection efficiency and subtraction of the CAT activity directed by pCAT.1. The standard deviation (SD) of n independent transfections is also listed. (B) Autoradiograph of the CAT activity of some of the constructs illustrated in panel A.

independence of the I-factor promoter from upstream sequences places it within a third group of RNA polymerase II promoters that lack both a TATA box and a CpG-rich sequence but require sequences at and downstream of the initiation site for basal transcription. This group of promoters includes those of LINE-like elements and some developmentally regulated genes such as *D. melanogaster* Antennapedia (*Antp*) and murine terminal deoxynucleotidyl-transferase (*TdT*) (48, 62, 63). One would predict that these promoters share sequence motifs involved in transcription initiation which, in the I factor, should lie from +1 to +40, as p40T directs the highest levels of CAT expression. Figure 3 shows a comparison of nucleotides +1 to +40 of I with the corresponding region in other genes with this type of promoter. Two conserved motifs exist: one having the consensus CA(G/T)T and lying at the transcription initiation site or start of the element, and one centered around position +30 and having the consensus AGACGTGPyPy. The more prox-

```

I      CAGTACCACCTTCAACCTCCGAGAGATAAGTCGTGCCTCT
jockey# AATCAATTCACATGGGAGATGAGCAATCGAGTGGACGTGTTTAC
F#     GGATTTCAATTCGATCGCCGACGTGTGAAGACGTTTTTAT
G      ACAGTCGCGATCGAACACTCAACGAGTGCAGACGTGCCTAC
doc    GACATTCGGCATTCCACAGTCTTCGGGTGGAGACGTGTTTCT
R1Bm   TGACTTCGCGCTCGGCCCTTGGTCGAGGACAGACGTGCCTTC
R1Dm                   CGGACGTGTTTTTC
R1Dv   CTCAGTTCGTTTCAGACAGTCTTGGGAACAGACGTGT
Antp P2* GTTCAGTGTGTAATGAATGGACGTGCCAAATAGACGTGCCGCC
enDm*  CAACTAATTCAGTCGTTGGCGCTCGATGTGAACAGACGTGCGTGT
    
```

FIG. 3. Comparison of the first 40 bp of some LINE-like elements and TATA-less non-CpG-rich promoters. The conserved motifs around the transcription start site and position +30 are shown in boldtype, and the transcription initiation site(s) are underlined. + and - denote limits that have been tested for promoter function, with + signifying promoter function and - indicating loss of promoter activity; 5' and 3' indicated the direction from which these deletions were made. #, the assay was done in vivo; *, the assay was done in vitro. References (in parentheses) are as follows: I (19), jockey (52), F (16), G (15), doc (17), R1Bm (70), R1Dm (26), R1Dv (D. virilis) (54), Antp P2 (48), and enDm (64).

imal location of this motif in R1Dm could be because this element is a 5' truncated copy.

Plasmid p28T contains nucleotides +1 to +28 of the I factor (Fig. 2). It lacks the AGTCGTGPyPy motif and demonstrates a twofold drop in CAT activity, suggesting that this conserved motif plays a positive role in but is not essential for expression from the I-factor promoter in tissue culture. The sequence immediately downstream of position +28 in plasmid p28T is GGATCCCCG. This sequence bears no resemblance to the wild-type motif and should therefore be unable to substitute for the wild-type function.

The consensus CA(G/T)T is found at or near the transcription start site in several other TATA-less and TATA-containing promoters (1, 24), with transcription often initiating at the C. The motif is usually preceded by a T. The 5' UTR of the I factor, however, does not depend on a T at position -1 for promoter activity, as the constructs assayed here have an A at this position.

The 5' ends of all I factors sequenced to date are highly conserved, with the only difference in the first 100 bp or so being at position +3, which can be either a T or a G (6, 19). To test the effect of this difference and the importance of the CA(G/T)T motif, the CAT activities directed from the four constructs in Fig. 4 were compared. p186G, which has the sequence CAGT, directs on average 80% more CAT activity than does p186T.1, which has CATT, although the standard deviation is high. Mutation of the CAGT motif to TAGA in p186Δ4.1 reduced CAT activity slightly. Replacement of nucleotides 1 to 4 by ACCG in construct p186Δ4.2, effectively deleting the first four bases, abolishes promoter activity. Thus, the CA(G/T)T motif is essential for expression from the I-factor promoter, although some variation at positions +1 and +4 can be tolerated. Results for two further constructs, p40G and p40Δ4, that have a 3' limit at +40 are similar to those for p186G and p186Δ4.2 (Fig. 4).

Plasmids identical to those described above but with the PstI fragment containing the promoter-CAT-poly(A) signal inverted gave similar results, indicating that the effects observed are not due to sequences immediately upstream from the I-factor sequences (data not shown).

Transcription of the I factor initiates at nucleotide +1. The site at which transcription initiates in the promoter contained in p40G was determined in vitro by using a Drosophila embryo transcription extract (see Materials and Methods). p40G was chosen because it retains both conserved motifs and directs highest CAT activity. Primer extension analysis showed that initiation is at position +1 and is sensitive to a low concentration (4 μg/ml) of α-amanatin, consistent with

Construct	Sequence	Relative Activity	n	SD
p186T.1	catt ————— CAT 186	100		
p186G	cagt ————— CAT 186	180	4	90
p186Δ4.1	taga ————— CAT 186	130	2	4
p186Δ4.2	accg ————— CAT 186	0	4	-
p40G	cagt — CAT 40	434	4	179
p40Δ4	accg — CAT 40	0	4	-

FIG. 4. Mutation of the CA(G/T)T motif. The line drawings indicate the extent of the deletions, with the sequence of the first four nucleotides given. Relative activity indicates the CAT expression level relative to that of p186T.1 (taken as 100%) after normalization to the transfection efficiency and subtraction of the CAT activity directed by pCAT.1. The standard deviation (SD) of n independent transfections is also listed. For p186Δ4.2 and p40Δ4, no standard deviation is given because the CAT activity from both of these promoters is slightly less than that of pCAT.1.

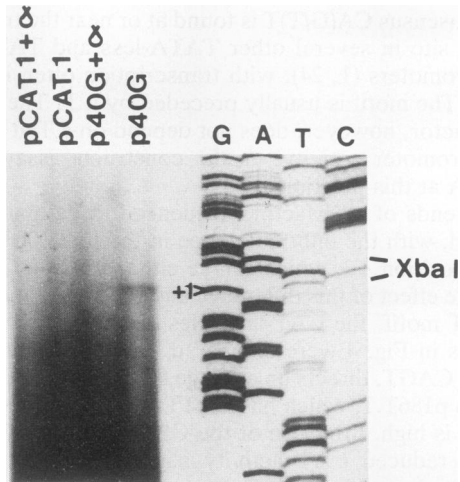


FIG. 5. Primer extension analysis of transcripts initiating from p40G and pCAT.1 in a *Drosophila* in vitro transcription extract. The transcription start site maps to position +1, indicated on the size markers that are sequence products of an M13 clone that contains the *Xba*I-*Eco*RI (promoter/CAT 5' end) fragment of p40T. The *Xba*I site is indicated on the size markers. The same primer was used for both primer extension and sequencing. Reactions that include 4 μ g of α -amanatin per ml are indicated as + α .

transcription being initiated from an internal promoter by RNA polymerase II (Fig. 5).

Establishment of inducer and reactive transformed lines. The promoter in the 5' UTR of the I factor was tested for activity in reactive and inducer flies by making transformed lines. The P-element transformation vectors p186W8 and pCATW8 (see Materials and Methods) were injected into the *D. melanogaster* reactive strain *W^K* (36), together with a

source of transposase. Several lines homozygous for independent insertion events were established from each construct. The expression of the I-factor promoter on the p186W8 transgene in a reactive background should mimic its behavior in a dysgenic female, as in both situations it is in a maternal environment in which regulation is not established.

Isogenic inducer lines were established for five of the p186W8 transformants by mating transformed females with inducer males that had marked balancer chromosomes (see Materials and Methods). The maternal genotype was recovered in the progeny of the resultant SF female and homozygous lines established. These lines differ from their reactive ancestors only by the presence of I factors spread by transposition in the SF germ line and chromosome 4, which was not marked or balanced. Their inducer state was verified by Southern hybridization to detect the presence of full-length I factors and in some cases by crossing putative inducer males with reactive females and confirming that they gave sterile SF female progeny (data not shown).

Developmental profile of CAT expression in transformed lines. The levels of CAT expression in females and males were measured at various developmental stages for five lines transformed with pCATW8 and five isogenic reactive and inducer lines transformed with p186W8. Four of the pCATW8 lines had no detectable CAT activity. The other line showed a low level of CAT activity in late pupae (6 pmol of [¹⁴C]chloramphenicol acetylated per min per mg of protein). In contrast, each of the p186W8 transformants showed CAT expression in each stage tested (Fig. 6 and Table 1). Figure 6 illustrates the change in CAT activity for two pairs of reactive and inducer lines. A steady increase in CAT activity was found throughout development in both males and females whether reactive or inducer; however, after eclosion, reactive females displayed a much greater increase in CAT activity. This burst of activity was not seen in males

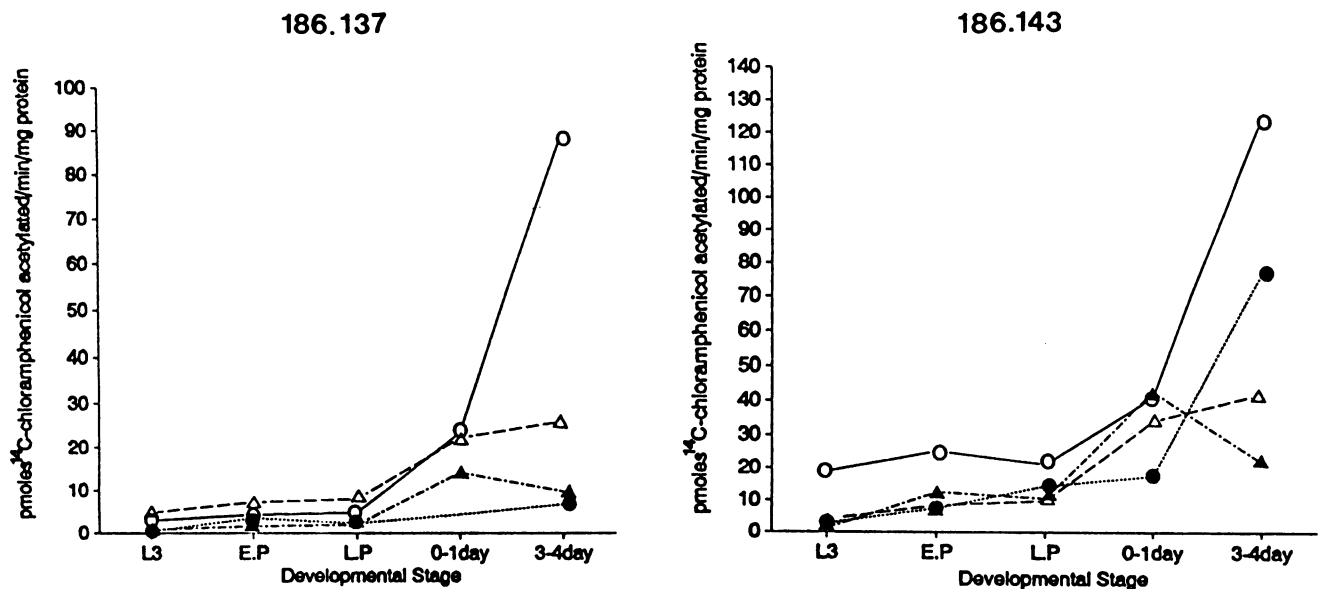


FIG. 6. Developmental profile expression from the I-factor promoter in inducer and reactive backgrounds. The two graphs show the CAT activities in different developmental stages, normalized for the protein concentration, for two independent lines transformed with p186W8, 186.137 and 186.143. L3, wandering third-instar larvae; E.P, early pupae (1 to 2 days old); L.P, late pupae (3 to 4 days old); 0-1day, adults 0 to 1 day after eclosion; 3-4day, adults 3 to 4 days after eclosion. Each point represents the CAT activity from a single assay of an extract made from 10 pooled individuals. Open and filled circles represent reactive and inducer females, respectively; open and filled triangles represent reactive and inducer males, respectively.

TABLE 1. CAT activities of reactive and inducer p186W8 transformants

LINE	CAT activity (pmol of [¹⁴ C]chloramphenicol acetylated/min/mg of protein)				Ratio, ovary/carcass activity
	Males	Carcass	Ovaries	Embryos	
186.132	15 ^a	15 ^a	160 ^a	ND ^b	11:1 ^a
186.132 ^c	28 ^d	15 ^d	92 ^d	ND	6:1 ^d
186.137	25	19	445		23:1
	7 ^a	1 ^a	536 ^a	496 ^a	536:1 ^a
186.137 ¹	10	4	16	ND	4:1
186.138	78	50	229	ND	4.6:1
186.138 ¹	168 ^d	78 ^d	84 ^d	ND	1.1:1 ^d
186.143	42	61	707		12:1
	36 ^a	55 ^a	928 ^a	378 ^a	17:1 ^a
186.143 ¹	22	34	55	ND	1.6:1
186.148	39	109	684		6:1
	29 ^a	72 ^a	1,259 ^a	306 ^a	17:1 ^a
186.148 ¹	59	52	37	ND	0.7:1

^a Assays done approximately 36 generations postestablishment. All other reactive assays were done 20 generations postestablishment.

^b ND, not done.

^c ¹ indicates the inducer state.

^d Assay done eight generations postestablishment. All other inducer assays were done 18 generations postestablishment.

and was reduced or absent in inducer females. A similar pattern of expression was found in all five lines, although the absolute levels of CAT activity differed from one line to another, indicating that the I-factor promoter is sensitive to position effect. This finding could explain the observation that the ability of an I factor to confer the inducer phenotype can increase upon transposition to a new site (45).

The I-factor promoter is most active in reactive ovaries. Ovaries were dissected from carcass in 3- to 4-day-old adult females, and both tissues were assayed for CAT activity. Figure 7 shows that the CAT activity from ovaries of reactive adult females greatly exceeds that of the carcass

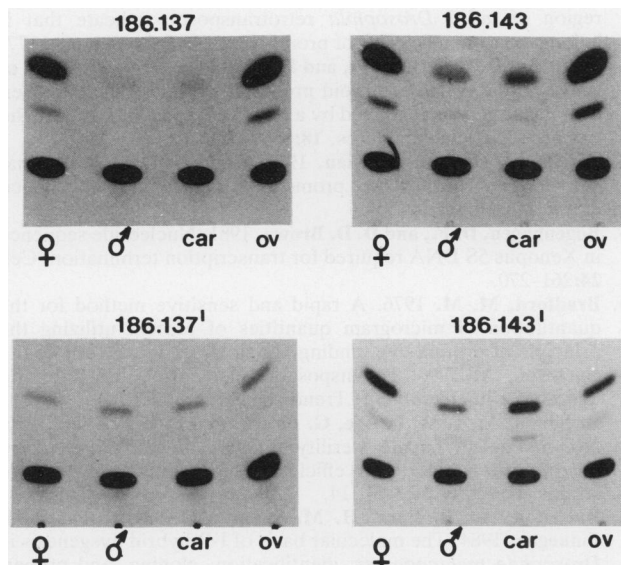


FIG. 7. Expression from the I-factor promoter in ovary and carcass tissues of reactive and inducer females of lines 186.137 and 186.143. ¹, inducer state; ♀, females 3 to 4 days after eclosion; ♂, males 3 to 4 days after eclosion; car, carcass tissue of 3- to 4-day-old females; ov, ovary tissues of 3- to 4-day-old females.

tissues, the activity of which is similar to that of males. This enhanced expression is absent or reduced in isogenic inducer females, suggesting that the inducer state regulates expression of the I-factor promoter. Table 1 lists the CAT expression values, normalized for the concentration of protein in the extracts of isogenic reactive and inducer lines. The data show that expression in reactive ovaries is on average around 15-fold higher than in carcass and that the activity in carcass is similar to that in males and inducer ovaries. The first 186 nucleotides of the I factor therefore contain sequences that contribute to the ovary and cytotype specificity of I-factor expression and consequently transposition.

The CAT activity of 0- to 2-h embryos was measured for some of the reactive lines and was found to be much higher than carcass CAT activity (Table 1). The majority of transcripts and some proteins present at this stage of development are synthesized by the nurse cells of the ovaries and then passed to the oocyte (reviewed in reference 37). High CAT activity in 0- to 2-h embryos suggests that the I-factor promoter is active in nurse cells.

DISCUSSION

I factors have an internal promoter within their 5' UTR.

The results presented in this report demonstrate that the I factor has within its 5' UTR a promoter that is completely internal to the transcription unit. The sensitivity of the promoter to α -amanatin is consistent with transcription by RNA polymerase II. The internal nature of the I-factor promoter distinguishes it from the more familiar class II promoters that have a TATA box, located 25 to 30 bp upstream of the initiation site, or an upstream CpG island. Instead, the promoter of the I factor and other LINE-like elements lies in the same group as do the TATA-less, non-CpG-rich RNA polymerase II promoters of several developmentally regulated genes, such as murine *TdT* (62), human porphobilinogen deaminase (*PBGD*) (2), *D. melanogaster Antp* (48), engrailed (*en*) (64), Ultrabithorax (*Ubx*) (3), and *E74* (67), and promoters of several retrovirus-like transposons and LINE-like elements such as *mdg-1* (1), *gypsy* (27), *jockey* (40), and *F* (39). Accurate initiation of transcription from promoters of this type requires sequences at and downstream from the start site but is often independent of upstream sequences. Two sequence motifs were found to be conserved between the 5' ends of the I factor, other insect LINE-like elements, and the *Antp P2* and *en* promoters. One of these is at the transcription initiation site, and the other is around position +30. Mutational analysis of the motif CA(G/T)T, situated at positions +1 to +4, showed it to be important as complete removal abolished CAT expression.

Deletion of sequences between +28 and +41 that contain the second conserved motif, AGACGTGPyPy, indicates that this motif stimulates but is not essential for expression of the I-factor promoter (Fig. 2). This finding contrasts with deletion analysis of this motif in the *D. melanogaster Antp P2*, *en*, and *jockey* promoters (41, 48, 64). Complete removal of this motif and two bases in the 5' direction abolishes transcription from *Antp P2* and *en* promoters in vitro (see Fig. 3 for deletion limits) (48, 64), and deletion of this motif from *jockey* abolishes promoter activity, as measured by CAT activity in transient tissue culture assays (41). The results presented here indicate that this motif is absolutely required by some but not all promoters of this type. Similar properties have been found for the conserved motif ACAG, situated around position +30 in some other TATA-less non-CpG-rich promoters (1). Deletion of this motif from the

D. melanogaster E74, *Ubx*, and *mdg-1* promoters abolishes transcription (1, 3, 67), whereas deleting the same sequence from the *TdT*, *PBGD*, and *gypsy* promoters either reduces expression or has no effect (2, 27, 62). It therefore appears that although the sequence of this internal motif is conserved among promoters, it may not perform precisely the same function in each case.

Control of I-factor activity. The results of previous experiments have shown that transposition of I elements is regulated in at least two ways. There is genetic control that limits transposition in inducer strains and that is disrupted in the female progeny of a dysgenic cross, and there is tissue-specific control that restricts transposition to the female germ line. The results reported here suggest that sequences within the first 186 bp of the I factor are involved in both control systems, since the promoter within this region is active only at high levels in the ovaries and is less active in ovaries of females of an inducer strain than in the ovaries of an isogenic reactive strain (Table 1). Lachaume et al. (34) have recently demonstrated that a larger portion of the I factor gives similar tissue and cytotypic expression when fused to the *lacZ* reporter gene. Our results are consistent with these data and more closely define the region of the I factor involved in regulation. Higher levels of expression in ovaries than in other tissues could be due to sequences that enhance expression in ovaries, sequences that repress expression in other tissues, or a combination of the two. Plasmid p40T, carrying bases 1 to 40, gave higher levels of CAT activity in tissue culture cells than did plasmid p100T, carrying bases 1 to 100, suggesting that sequences between bases 41 and 100 may inhibit somatic expression (Fig. 2).

Although the putative I-factor transposition intermediate has not been detected either in males or in female carcass (11), we detected CAT activity in these tissues. This activity is low relative to that in ovaries and may be due to accumulation of CAT protein translated from rare transcripts. Alternatively, there may be sequences downstream of position 186 that further suppress transcription in these tissues or reduce the stability of transcripts that are formed.

Reduced expression from the 1-186 fragment in inducer females must be caused directly or indirectly by a molecule encoded by complete I factors or by I factors themselves. A candidate for an I-factor-encoded regulatory molecule is the product of ORF1, as this is a nucleic acid-binding protein (43). This protein may interact with the 5' UTR to suppress promoter activity.

A second system of hybrid dysgenesis, P-M hybrid dysgenesis (18), is controlled by P transposable elements. These are also subject to both genetic and tissue-specific controls (56, 57), as high levels of transposition are seen only in the germ line of progeny of crosses between males of a strain carrying functional elements (a P strain) and females of a strain without P elements (an M strain). Genetic control of P elements is mediated by P-element-encoded proteins just as I elements appear to be controlled by I-element-encoded proteins. There are differences in the mechanisms for tissue-specific control in the two systems, however. I-factor transposition is restricted to ovaries at least partly as a result of transcriptional control, whereas P transposition is restricted to the germ line because of somatic inhibition of a splicing event required for production of P transposase (12, 61).

The CAT activity in ovaries of inducer strains is on average 20-fold lower than that in reactive strains for all lines except 186.132¹ and 186.138¹ (Table 1). At the time of these experiments, these two lines had been established for 8 generations, and the other inducer lines had been established

for 18 generations. Eight generations may be insufficient time to establish the inducer state, as low levels of SF sterility and presumably I-factor transposition are perpetuated in the descendants of SF females for six or more generations after the dysgenic cross (50). This finding may indicate that a critical level of regulatory molecule is required before transcription and transposition are reduced to the level seen in stable inducer lines. The time taken to attain this level of the autoregulator will depend on I-factor copy number and the strength of the promoters, which in turn will be influenced by position effect. A similar delay in establishment of the P cytotypic has been reported (51).

Females of line 186.143¹ displayed higher CAT activity than do their brothers or males from the isogenic reactive strain 186.143 (Fig. 6) even though this inducer line had been established for 36 generations. This is probably not because the full inducer state has not been reached, however, as CAT activity in the ovaries of these females is reduced to a level similar to that in carcass (Table 1). The higher CAT activity in 186.143¹ females might therefore be caused by enhanced expression in females due to a position effect.

One intriguing observation is that CAT activity in the three reactive lines tested at 20 and 36 generations increases in ovaries and decreases in female carcass and males during this interval (Table 1). These changes are slight but if significant might indicate that the I-factor promoter takes several generations to establish maximum strength in reactive ovaries. If this is the case, the full potential of the I-factor promoter would never be observed under dysgenic conditions, as I-factor autoregulatory molecules would check this effect in the descendants of SF females.

ACKNOWLEDGMENTS

This work was supported by a Science and Engineering Research Council postgraduate grant and a project grant from the Medical Research Council.

We are grateful to G. Brown and F. Johnston for photography.

REFERENCES

1. Arkhipova, I. R., and Y. V. Ilyin. 1991. Properties of promoter region of *mdg1* *Drosophila* retrotransposon indicate that it belongs to a specific class of promoter. *EMBO J.* **10**:1169-1177.
2. Beaupain, D., J. F. Eléouët, and P. H. Roméo. 1990. Initiation of transcription of the erythroid promoter of the porphobilinogen deaminase gene is regulated by a *cis*-acting sequence around the cap site. *Nucleic Acids Res.* **18**:6509-6515.
3. Biggin, M. D., and R. Tjian. 1988. Transcription factors that activate the *Ultrathorax* promoter in developmentally staged extracts. *Cell* **53**:699-711.
4. Bogenhagen, D. F., and D. D. Brown. 1981. Nucleotide sequence in *Xenopus* 5S DNA required for transcription termination. *Cell* **24**:261-270.
5. Bradford, M. M. 1976. A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Anal. Biochem.* **72**:248-254.
6. Bucheton, A. 1990. I transposable elements and I-R hybrid dysgenesis in *Drosophila*. *Trends Genet.* **6**:16-21.
7. Bucheton, A., J. M. Lavigne, G. Picard, and P. L'Heritier. 1976. Non-mendelian female sterility in *Drosophila melanogaster*: quantitative variation in efficiency of inducer and reactive strains. *Heredity* **36**:305-314.
8. Bucheton, A., R. Paro, H. M. Sang, A. Pélisson, and D. J. Finnegan. 1984. The molecular basis of I-R hybrid dysgenesis in *Drosophila melanogaster*: identification, cloning, and properties of the I factor. *Cell* **38**:153-163.
9. Burke, W. D., C. C. Calalang, and T. H. Eickbush. 1987. The site-specific ribosomal insertion element type II of *Bombyx mori* (R2Bm) contains the coding sequence for a reverse transcriptase-like enzyme. *Mol. Cell. Biol.* **7**:2221-2230.

10. **Burton, F. H., D. D. Loeb, C. F. Voliva, S. L. Martin, M. H. Edgell, and C. A. Hutchison III.** 1986. Conservation throughout mammalia and extensive protein-encoding capacity of the highly repeated DNA long interspersed sequence one. *J. Mol. Biol.* **187**:291-304.
11. **Chaboissier, M.-C., I. Busseau, J. Prosser, D. J. 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.
12. **Chain, A. C., S. Zollman, J. C. Tseng, and F. A. Laski.** 1991. Identification of a cis-acting sequence required for germline specific splicing of the P element ORF2-ORF3 intron. *Mol. Cell. Biol.* **11**:1538-1546.
13. **Crozatier, M., C. Vauray, I. Busseau, A. Péllisson, and A. Bucheton.** 1988. Structure and genomic organisation of I elements involved in I-R hybrid dysgenesis in *Drosophila melanogaster*. *Nucleic Acids Res.* **16**:9199-9123.
14. **Daniels, D. L., J. L. Schroeder, W. Szybalski, F. Sanger, A. R. Coulson, G. F. Hong, D. F. Hill, G. B. Peterson, and F. R. Blattner.** 1983. Appendix II, p. 519-661. In R. W. Hendrix, J. W. Roberts, F. W. Stahl, and R. A. Weisberg (ed.), *Lambda II*. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
15. **Di Nocera, P. P.** 1988. Close relationship between non-viral retroposons in *Drosophila melanogaster*. *Nucleic Acids Res.* **16**:4041-4052.
16. **Di Nocera, P. P., M. E. Digar, and I. B. Dawid.** 1983. A family of oligo-adenylated terminated transposable sequences in *Drosophila melanogaster*. *J. Mol. Biol.* **168**:715-727.
17. **Driver, A., S. F. Lacey, T. E. Cullingford, A. Mitchelson, and K. O'Hare.** 1989. Structural analysis of *Doc* transposable elements associated with mutations at the *white* and *suppressor of forked* loci of *Drosophila melanogaster*. *Mol. Gen. Genet.* **220**:49-52.
18. **Engels, W. R.** 1989. P elements in *Drosophila melanogaster*, p. 437-484. In D. E. Berg and M. M. Howe (ed.), *Mobile DNA*. American Society for Microbiology, Washington, D.C.
19. **Fawcett, D. H., C. K. Lister, E. Kellet, and D. J. Finnegan.** 1986. Transposable elements controlling I-R hybrid dysgenesis in *D. melanogaster* are similar to mammalian LINES. *Cell* **47**:1007-1015.
20. **Gabriel, A., and J. D. Boeke.** 1991. Reverse transcriptase encoded by a retrotransposon from the trypanosomatid *Crithidia fasciculata*. *Proc. Natl. Acad. Sci. USA* **88**:9794-9798.
21. **Gabriel, A., T. J. Yen, D. C. Schwart, C. L. Smith, J. D. Boeke, B. Sollner-Webb, and D. W. Cleveland.** 1990. A rapidly rearranging retrotransposon within the minixion gene locus of *Crithidia fasciculata*. *Mol. Cell. Biol.* **10**:615-624.
22. **Gorman, C., L. Moffat, and B. Howard.** 1982. Recombinant genomes which express chloramphenicol acetyltransferase in mammalian cells. *Mol. Cell. Biol.* **2**:1044-1051.
23. **Harada, K., K. Yukuhiro, and T. Mukai.** 1990. Transposition rates of movable genetic elements in *Drosophila melanogaster*. *Proc. Natl. Acad. Sci. USA* **87**:3248-3252.
24. **Hultmark, D., R. Klemenz, and W. J. Gehring.** 1986. Translational and transcriptional control elements in the untranslated leader of the heat shock gene *hsp22*. *Cell* **44**:429-438.
25. **Ivanov, V. A., A. A. Melnikov, A. V. Siunov, I. I. Fodor, and Y. V. Ilyin.** 1991. Authentic reverse-transcriptase is coded by jockey, a mobile *Drosophila* element related to mammalian LINES. *EMBO J.* **10**:2489-2495.
26. **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.
27. **Jarrel, K. A., and M. Meselson.** 1991. *Drosophila* retrotransposon promoter includes essential sequences at the initiation site and requires a downstream sequence for full activity. *Proc. Natl. Acad. Sci. USA* **88**:102-104.
28. **Jensen, S., and T. Heidmann.** 1991. An indicator gene for detection of germline retrotransposition in transgenic *Drosophila* demonstrates RNA-mediated transposition of the LINE I elements. *EMBO J.* **10**:1927-1937.
29. **Kamakaka, R. T., C. M. Tyree, and J. T. Kadanoga.** 1991. Accurate and efficient RNA polymerase II transcription with a soluble nuclear fraction derived from *Drosophila* embryos. *Proc. Natl. Acad. Sci. USA* **88**:1024-1028.
30. **Karess, R. E., and G. M. Rubin.** 1984. Analysis of P transposable element functions in *Drosophila*. *Cell* **38**:135-146.
31. **Kimmel, B. E., O. K. Ole-Moiyoi, and J. R. Young.** 1987. Ingi, a 5.2-kb dispersed sequence element from *Trypanosoma brucei* that carries half of a smaller mobile element at either end and has homology to mammalian LINES. *Mol. Cell. Biol.* **7**:1465-1475.
32. **Kinsey, J. A., and J. Helber.** 1989. Isolation of a transposable element from *Neurospora crassa*. *Proc. Natl. Acad. Sci. USA* **86**:1929-1933.
33. **Klemenz, R., U. Weber, and W. J. Gehring.** 1987. The *white* gene as a marker in a new P-element vector for gene transfer in *Drosophila*. *Nucleic Acids Res.* **15**:3947-3957.
34. **Lachaume, P., K. Bouhidel, M. Mesure, and H. Pinon.** 1992. Spatial and temporal expression of the I factor during oogenesis in *Drosophila melanogaster*. *Development* **115**:729-735.
35. **Lindsley, D. L., and E. M. Grell.** 1972. Genetic variation of *Drosophila melanogaster*. Carnegie Institute of Washington, Washington, D.C.
36. **Lüning, K. G.** 1981. Genetics of inbred *Drosophila melanogaster*. *Hereditas* **95**:181-188.
37. **Mahowald, A. P., and M. P. Kambyzellis.** 1980. p. 141-224. In M. Ashburner and T. R. F. Wright (ed.), *The genetics and biology of Drosophila*, vol. 4. Academic Press, London.
38. **Mathias, S. L., A. L. Scott, H. H. Kazazian, J. D. Boeke, and A. Gabriel.** 1991. Reverse-transcriptase encoded by a human transposable element. *Science* **254**:1808-1810.
39. **Minchiotti, G., and P. P. Di Nocera.** 1991. Convergent transcription initiates from oppositely orientated promoters within the 5' end regions of *Drosophila melanogaster* F elements. *Mol. Cell. Biol.* **11**:5171-5180.
40. **Mizrokhi, L. J., S. G. Georgieva, and Y. V. Ilyin.** 1988. jockey, a mobile *Drosophila* element similar to mammalian LINES, is transcribed from an internal promoter by RNA polymerase II. *Cell* **54**:685-691.
41. **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.
42. **Murphy, N. B., A. Pays, P. Tebabi, H. Coquelet, M. Guyaux, M. Steinert, and E. Pays.** 1987. *Trypanosoma brucei* repeated elements with unusual structural and transcriptional properties. *J. Mol. Biol.* **195**:855-871.
43. **Paterson, T., and D. J. F. Finnegan.** Unpublished data.
44. **Péllisson, A.** 1981. The I-R system of hybrid dysgenesis in *Drosophila melanogaster*: are I factor insertions responsible for the mutator effect of the I-R interaction? *Mol. Gen. Genet.* **183**:123-129.
45. **Péllisson, A., and J. C. Bregliano.** 1981. The I-R system of hybrid dysgenesis in *Drosophila melanogaster*: construction and characterisation of a non-inducer stock. *Biol. Cell* **40**:159-164.
46. **Péllisson, A., and J. C. Bregliano.** 1987. Evidence for rapid limitation of the I element copy number in a genome submitted to several generations of I-R hybrid dysgenesis in *Drosophila melanogaster*. *Mol. Gen. Genet.* **207**:306-313.
47. **Péllisson, A., D. J. Finnegan, and A. Bucheton.** 1991. Evidence for retrotransposition of the I factor, a LINE element of *Drosophila melanogaster*. *Proc. Natl. Acad. Sci. USA* **88**:4907-4910.
48. **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.
49. **Picard, G.** 1976. Non mendelian female sterility in *Drosophila melanogaster*: heredity and transmission of I factor. *Genetics* **83**:107-123.
50. **Picard, G.** 1978. Non mendelian female sterility in *Drosophila melanogaster*: sterility in stocks derived from the genotypically inducer or reactive offspring of SF and RSF females. *Biol. Cell* **31**:245-254.
51. **Preston, C. R., and W. R. Engels.** 1989. Spread of P transposable elements in inbred lines of *Drosophila melanogaster*. *Prog. Nucleic Acid Res. Mol. Biol.* **36**:71-85.

52. Priimägi, 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.
53. Pritchard, M. A., J.-M. Dura, A. Péllisson, A. Bucheton, and D. J. Finnegan. 1988. A cloned I-factor is fully functional in *Drosophila melanogaster*. *Mol. Gen. Genet.* 214:533–540.
54. Rae, P. M. M., B. D. Kohorn, and R. P. Wade. 1980. The 10kb *Drosophila virilus* 28S rDNA intervening sequence is flanked by a direct repeat of 14 base pairs of coding sequence. *Nucleic Acids Res.* 8:3491–3504.
55. Read, S. M., and D. H. Northcote. 1981. Minimisation of variation in the response to different proteins of the Coomassie blue G dye binding assay of protein. *Anal. Biochem.* 116:53–64.
56. Rio, D. C. 1990. Molecular mechanisms regulating *Drosophila* P element transposition. *Annu. Rev. Genet.* 24:543–578.
57. Rio, D. C. 1991. Regulation of *Drosophila* P element transposition. *Trends Genet.* 7:282–287.
58. Rubin, G. M., and A. C. Spradling. 1982. Genetic transformation of *Drosophila* with transposable element vectors. *Science* 218:348–353.
59. Schneider, I. 1972. Cell lines derived from late embryonic stages of *Drosophila melanogaster*. *J. Embryol. Exp. Morphol.* 27:353–367.
60. Schwarz-Sommer, Z., L. Leclercq, E. Ciobel, and H. Saedler. 1987. *Cin4*, an insert altering the structure of the *A1* gene in *Zea mays*, exhibits properties of non-viral retrotransposons. *EMBO J.* 6:3873–3880.
61. Siebel, C. W., and D. C. Rio. 1990. Regulated splicing of the *Drosophila* P transposable element third intron *in vitro*: somatic repression. *Science* 248:1200–1208.
62. Smale, S. T., and D. Baltimore. 1989. The “initiator” as a transcription control element. *Cell* 57:103–113.
63. Smale, S. T., M. C. Schmidt, A. J. Berk, and D. Baltimore. 1990. Transcriptional activation by Sp1 as directed through TATA or initiator: specific requirements for mammalian transcription factor IID. *Proc. Natl. Acad. Sci. USA* 87:4509–4513.
64. Soeller, W. C., S. J. Poole, and T. Kornberg. 1988. *In vitro* transcription of the *Drosophila engrailed* gene. *Genes Dev.* 2:68–81.
65. Sun, X.-H., J. T. Lis, and R. Wu. 1988. The positive and negative transcriptional regulation of the *Drosophila Gapdh-2* gene. *Genes Dev.* 2:743–753.
66. Swergold, G. D. 1990. Identification, characterization, and cell specificity of a human LINE-1 promoter. *Mol. Cell. Biol.* 10:6718–6729.
67. Thummel, C. S. 1989. The *Drosophila E74* promoter contains essential sequences downstream from the start site of transcription. *Genes Dev.* 3:782–792.
68. Thummel, C. S., A. M. Boulet, and H. D. Lipshitz. 1988. Vectors for *Drosophila* P-element-mediated transformation and tissue culture transfection. *Gene* 74:445–456.
69. Xiao, H., and J. T. Lis. 1988. Germline transformation used to define key features of heat-shock response elements. *Science* 239:1139–1142.
70. 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.