Proc. Natl. Acad. Sci. USA
Vol. 88, pp. 102–104, January 1991
Biochemistry

Drosophila retrotransposon promoter includes an essential sequence at the initiation site and requires a downstream sequence for full activity

(gypsy/initiator/transcription/gene control)

KEVIN A. JARRELL AND MATTHEW MESELSON*

Department of Biochemistry and Molecular Biology, Harvard University, 7 Divinity Avenue, Cambridge, MA 02138

Contributed by Matthew Meselson, October 1, 1990

ABSTRACT **We describe a 98-base-pair region (−38 to +60) in the long terminal repeat of the *Drosophila* gypsy retrotransposon that is sufficient for accurate normal-level transcription. We find that, unlike most RNA polymerase II (pol II) promoters, the gypsy promoter includes downstream sequences that are required for full activity. Also unlike most pol II promoters, the gypsy promoter, which lacks a TATA motif, was found to have an essential sequence at the transcription initiation site, mutation of which abolishes transcription. These three uncommon features of the gypsy promoter may be characteristic of a subset of pol II promoters, exemplified by certain retrotransposons and developmental genes of *Drosophila* and by *Tdt*, the mouse terminal deoxynucleotidyltransferase (TdT) gene.**

Mobile genetic elements are a common feature of eukaryotic genomes. Examples are the gypsy and copia elements of *Drosophila*, the Ty elements of yeast, and the integrated murine leukemia virus (1–3). Relatively little is known regarding the detailed structure of retrotransposon promoters and enhancers and the nature, effects, and genetic specification of trans-acting factors that control retrotransposon expression. We report here that the gypsy promoter has three unusual features, at least some of which are found in certain other retrotransposons and developmental genes.

MATERIALS AND METHODS

Plasmids and Constructions. The full-length gypsy–chloramphenicol acetyltransferase (CAT) construct consists of the first 970 base pairs (bp) of the bx^{34e} gypsy ligated to the CAT gene, followed by *hsp70* DNA that provides a cleavage and polyadenylylation signal. The CAT–*hsp70* segment was derived from a *hsp70*–CAT construct described elsewhere (4). The 5′- and 3′-deletion derivatives were made by BAL 31 nuclease digestion of the full-length construct. The Δ35 construct is identical to the full-length construct, except that bp +14 to +48 were replaced by 16 bp of vector sequence. The mutated initiator construct is identical to the −38 to +60 deletion derivative of the full-length construct, except that the TCAGTT sequence at the start site has been changed to GGATCC by site-directed mutagenesis (5).

Transfection. *Drosophila* Schneider 2 cells ($\approx 5 \times 10^7$) were transfected in 10 ml with 50 μg of gypsy–CAT DNA and 5 μg of heat shock protein *hsp82–lacZ* DNA (6) by the calcium phosphate method (7). Cells were harvested after 48 hr and extracted for CAT assays with ^{14}C-labeled chloramphenicol (8). The extent of conversion of chloramphenicol to acetylated chloramphenicol was determined by scintillation counting of acetylated and nonacetylated chloramphenicol recovered from TLC plates. Each value was normalized to measurements of β-galactosidase activity (9) in the same transfection extract.

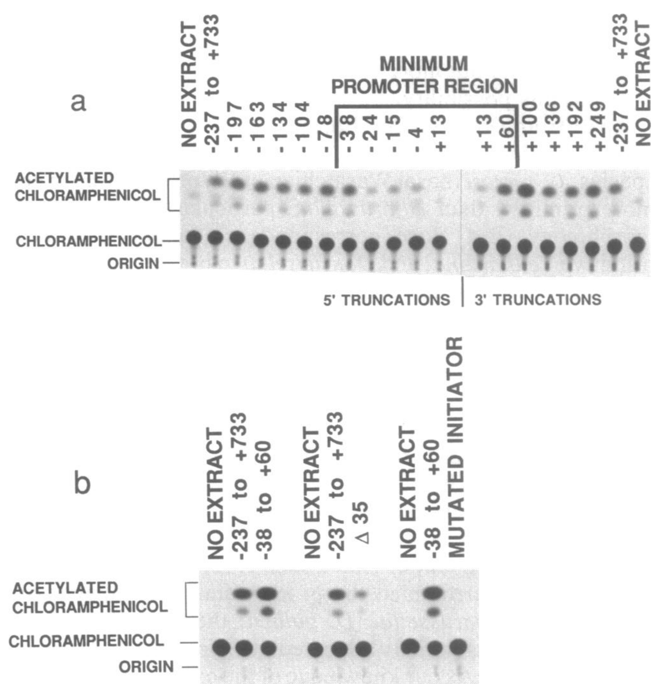

FIG. 1. Effect of 5′ and 3′ deletions on expression of a gypsy–CAT fusion gene. The full-length construct (−237 to +733) consists of the entire gypsy 5′ long terminal repeat and transcribed leader region from the bx^{34e} gypsy (11) fused to the *Escherichia coli* CAT gene. The endpoints of 16 different 5′ and 3′ derivatives are indicated. (*a*) Autoradiogram showing CAT assays from a representative experiment. All transfections, extract preparations, and CAT assays were done in parallel. In contrast to these results, Mazo *et al.* (12) report that deletion of the leader region from a gypsy–CAT construct (equivalent to our +249 truncation) decreased transcription 5-fold. The leader in their construct contained a 110-bp segment at +537 not present in the gypsy we used. Whether the apparent difference can be ascribed to this or to some other factor is not clear. (*b*) CAT assays comparing expression of the −38 to +60, Δ35, and mutated initiator constructs.

RNA Analysis. RNA was prepared from transfected cells as described (6). mRNA was selected by oligo(dT)-cellulose chromatography (10). The purified poly(A)$^+$ RNA was analyzed by primer extension (10), using a ^{32}P-5′-end-labeled

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

Abbreviations: TdT, terminal deoxynucleotidyltransferase; CAT, chloramphenicol acetyltransferase.
*To whom reprint requests should be addressed.

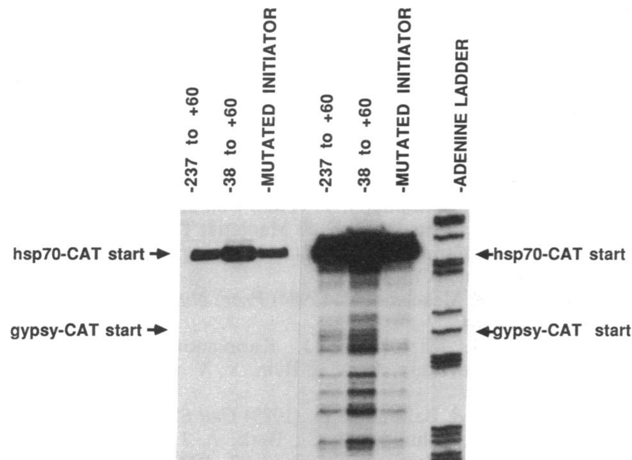

FIG. 2. Primer-extension analysis of gypsy–CAT RNA from the +60(3′) truncation, −38 to +60, and mutated initiator constructs. Two exposures of the same autoradiogram are shown (18 hr and 3 weeks). The gypsy transcription start site is seen as a doublet missing from the mutated initiator lane, as indicated. Relative to the cotransfected *hsp70*–CAT (4) control, the promoter activity of the −38 to +60 promoter is essentially the same as that of the +60(3′) truncation.

33-nucleotide oligonucleotide homologous to the CAT gene as primer.

RESULTS AND DISCUSSION

We analyzed the transcription of the full-length gypsy–CAT fusion gene and its deleted derivatives in transfected cells. Ten 5′ deletions and six 3′ deletions were tested. Fig. 1*a* shows that nearby sequences on both sides of the transcription initiation site are required for efficient expression. Deletion of upstream DNA down to −38 has no effect on promoter strength (normalized CAT activity). Deletion to −24 reduces promoter strength by ≈4-fold. Deletions to −15 or −4 have no further effect. A 5′ deletion through the origin, down to +13, reduces expression by a factor of >10. A 3′ deletion extending upstream to +13 reduces activity ≈4-fold, as does deletion of the interval +14 to +48, designated Δ35 (Fig. 1*b*). Thus, within the 98-bp region from −38 to +60

FIG. 3. Nucleotide sequence of the −38 to +60 gypsy promoter region. The sequence is numbered relative to the transcription initiation nucleotide, designated +1. The region deleted from the full-length construct to yield the Δ35 construct is underlined. The conserved TCAGTY sequence at the start site is in boldface and underlined.

there are sequences both 5′ and 3′ of the transcription initiation site that are essential for full promoter activity. The 11 deletions with endpoints outside the 98-bp region all have activities 0.6–1.3 times that of the full-length construct.

To determine whether the region −38 to +60 by itself is sufficient for transcription, we tested a construct with this sequence fused to CAT. As may be seen in Fig. 1*b*, it was as active as the full-length construct. Primer extension (Fig. 2) showed that transcription from the 98-bp promoter initiates at the expected site and that the +60 and −38 to +60 constructs produce comparable amounts of RNA, confirming the results of CAT enzyme assays.

The retention of partial activity by both the −4(5′) and +13(3′) deletions (Fig. 1*a*) suggested that the sequence at the initiation site is important for gypsy transcription. We therefore tested the effect of site-directed mutagenesis of the sequence TCAGTT at the origin of transcription, changing it to a *Bam*HI site, GGATCC. No promoter activity was detected with CAT assays or by primer-extension analysis (Figs. 1*b* and 2).

Table 1 summarizes features common to the promoters of gypsy and certain other *Drosophila* retrotransposons and developmental genes that lack the TATA motif. The sequence at the gypsy initiation site (Fig. 3) closely resembles sequences at the initiation sites of most of the other promoters listed, with a consensus of TCAGTY. Moreover, the sequence TCAGTT at the gypsy start site is a close match to the consensus hexamer TCANTC in the functional initiator of the mouse *Tdt* gene, identified by Smale *et al.* (13, 14). These authors have proposed that there is a specific class of RNA polymerase II (pol II) promoter, exemplified by the devel-

Table 1. Common features of the promoters of six *Drosophila* retrotransposons and five developmental genes

Promoter	TATA box	Start site	Initiator function	Downstream elements	Reference
gypsy (mdg4)	No	CTCAGTTC	Yes	Yes	11, 15, and this work
mdg1	No	ATCAGTTAT	ND	ND	15
mdg3	No	GTCAGTCG	ND	ND	15
B104	No	GTCAGTCA	ND	ND	16
17.6	No	TTCAGTCT	ND	ND	17
297	No	TTTAGTCT	ND	ND	18
Antennapedia	No	TTCAGTTG	ND	Yes	19
E74	No	TTTAGTTG	ND	Yes	20
engrailed	No	GTCAACTA	ND	Yes	21
Ultrabithorax	No	GGTGCCAC	ND	Yes	22
Mouse *Tdt*	No	CTCATTCT	Yes	Yes	13
Drosophila start site consensus		NTCAGTYN			15, 23, and this work
Mouse *Tdt* functional initiator		CTCANTCT			14

If the transcription initiation nucleotide has been mapped, it is underlined. In each long terminal repeat the TCAGTPY sequence is found upstream of the polyadenylylation signal. ND, not done.

opmentally controlled mouse *Tdt* gene and a number of *Drosophila* developmental genes, defined by the lack of TATA and the presence of downstream sequences near the origin of transcription that are required for efficient expression. We have shown that the gypsy promoter has both these characteristics and, remarkably, possesses a functional initiator closely homologous to that of the *Tdt* gene. Thus far, however, only the gypsy and *Tdt* promoters have been shown to include all three characteristics of this putative class of pol II promoters.

The expression of transposons is developmentally regulated, as exemplified in *Drosophila* by the tissue-specific intron splicing event that limits *P*-element transposition to the germ line (24) and by stage-specific transcription of retrotransposons (25). The similarity in promoter structure of certain retrotransposons and developmental genes suggests that a particular set of transcription factors may be found to control both retrotransposons and developmental genes.

We thank T. Maniatis and M. Matthews for valuable discussions and R. Rhew for technical assistance. This research was supported by the National Institutes of Health and by an American Cancer Society fellowship (K.A.J.).

1. Bingham, P. M. & Zachar, Z. (1989) in *Mobile DNA*, eds. Berg, D. E. & Howe, M. M. (Am. Soc. Microbiol., Washington, DC), pp. 485–502.
2. Boeke, J. D. (1989) in *Mobile DNA*, eds. Berg, D. E. & Howe, M. M. (Am. Soc. Microbiol., Washington, DC), pp. 335–374.
3. Varmus, H. & Brown, P. (1989) in *Mobile DNA*, eds. Berg, D. E. & Howe, M. M. (Am. Soc. Microbiol., Washington, DC), pp. 53–108.
4. Cohen, R. S. & Meselson, M. (1988) *Nature (London)* **332**, 856–858.
5. Jarrell, K. A., Dietrich, R. C. & Perlman, P. S. (1988) *Mol. Cell. Biol.* **8**, 2361–2366.
6. Dorsett, D., Viglianti, G. A., Rutledge, B. J. & Meselson, M. (1989) *Genes Dev.* **3**, 454–468.
7. DiNocera, P. P. & Dawid, I. B. (1983) *Proc. Natl. Acad. Sci. USA* **80**, 7095–7098.
8. Gorman, C. M., Moffat, L. F. & Howard, B. H. (1982) *Mol. Cell. Biol.* **2** (9), 1044–1051.
9. Fischer, J. A. & Maniatis, T. (1988) *Cell* **53**, 451–461.
10. Sambrook, J., Fritsch, E. F. & Maniatis, T. (1989) *Molecular Cloning: A Laboratory Manual* (Cold Spring Harbor Lab., Cold Spring Harbor, NY), 2nd Ed.
11. Freund, R. & Meselson, M. (1984) *Proc. Natl. Acad. Sci. USA* **81**, 4462–4464.
12. Mazo, A. M., Mizrokhi, L. J., Karavanov, A. A., Sedkov, Y. A., Krichevskaja, A. A. & Ilyin, Y. V. (1989) *EMBO J.* **8**, 903–911.
13. Smale, S. T. & Baltimore, D. (1989) *Cell* **57**, 103–113.
14. Smale, S. T., Schmidt, M. C., Berk, A. J. & Baltimore, D. (1990) *Proc. Natl. Acad. Sci. USA* **87**, 4509–4513.
15. Arkhipova, I. R., Mazo, A. M., Cherkasova, V. A., Gorelova, T. V., Schuppe, N. G. & Ilyin, Y. V. (1986) *Cell* **44**, 555–563.
16. Scherer, G., Telford, J., Baldari, C. & Pirrotta, V. (1981) *Dev. Biol.* **86**, 438–447.
17. Kugimiya, W., Ikenaga, H. & Saigo, K. (1983) *Proc. Natl. Acad. Sci. USA* **80**, 3193–3197.
18. Ikenaga, H. & Saigo, K. (1982) *Proc. Natl. Acad. Sci. USA* **79**, 4143–4147.
19. Perkins, K. K., Dailey, G. M. & Tjian, R. (1988) *Genes Dev.* **2**, 1615–1626.
20. Thummel, C. S. (1989) *Genes Dev.* **3**, 782–792.
21. Soeller, W. C., Poole, S. J. & Kornberg, T. (1988) *Genes Dev.* **2**, 68–81.
22. Biggin, M. D. & Tjian, R. (1988) *Cell* **53**, 699–711.
23. Hultmark, D., Klemenz, R. & Gehring, W. J. (1986) *Cell* **44**, 429–438.
24. Laski, F. A., Rio, D. C. & Rubin, G. M. (1986) *Cell* **44**, 7–19.
25. Parkhurst, S. M. & Corces, V. G. (1987) *EMBO J.* **6**, 419–424.