Identification of Sequences Which Regulate the Expression of Drosophila melanogaster Doc Elements

(Received for publication, June 22, 1995, and in revised form, September 7, 1995)

Cristina Contursi, Gabriella Minchiotti, and Pier Paolo Di Nocera
From the Dipartimento di Biologia e Patologia Cellulare e Molecolare "L. Califano," Università degli Studi di Napoli "Federico II," Via Pansini 5, 80131 Napoli, Italy and Istituto Internazionale di Genetica e Biofisica, C.N.R., Via Marconi 12, Napoli, Italy

Long interspersed nuclear elements (LINEs) are mobile DNA elements which propagate by reverse transcription of RNA intermediates. LINEs lack long terminal repeats, and their expression is controlled by promoters located inside to the transcribed region of unit-length DNA copies. Doc elements constitute one of the seven families of LINEs found in Drosophila melanogaster. Plasmids in which the chloramphenicol acetyltransferase (CAT) gene is preceded by DNA segments from different Doc family members were used as templates for transient expression assays in Drosophila S2 cells. Transcription is initiated at the 5' end of Doc elements inside hexamers fitting the consensus (C/G)AYTCG and is regulated by a DNA region which is located 20 base pairs (bp) downstream from the RNA start site(s). The region includes a sequence (RGACGTGY motif, or DE2) which stimulates transcription in other Drosophila LINEs, and two adjacent elements, DE1 and DE3. Moving the downstream region either 4 bp away from, or 5 bp closer to the RNA start site region inhibited transcription. Sequences located 200 bp downstream from the Doc 5' end repressed CAT expression in an orientation- and position-dependent manner. The inhibition reflects impaired translation of the CAT gene possibly consequent to the interaction of specific Doc RNA sequences with a cellular component.

Doc is one of 50 or more mobile DNA elements that have been identified in the fruit fly Drosophila melanogaster (Finnegan and Fawcett, 1986). Doc elements lack terminal repeats, instead terminating at the 3' end in runs of adenine residues flanked by polyadenylation signals. They differ in size, being variously truncated at the 5' end, and are flanked by target site duplications which vary in length from 10 to 14 bp (Schneuwly et al., 1987; Driver et al., 1989). Complete family members are 4.7 kb in length and potentially encode a putative nucleic acid binding protein and a reverse transcriptase (O'Hare et al., 1991). The structure and coding capacity of Doc is typical of LINEs, nomadic DNA sequences conserved in evolution from protozoa to man (Doolittle et al., 1989; Xiong and Eickbush, 1990). LINEs, also known as type II retrotransposons, use self-encoded proteins to reverse transcribe their own mRNA and integrate cDNA copies at new locations in the genome. This hypothesis has been experimentally supported by the analysis of transgenic flies carrying intron-marked Drosophila I factors (Pelisson et al., 1991; Jensen and Heidmann, 1991) and baby hamster kidney cells transfected with mouse LINE-1 elements (Evans and Palmiter, 1991).

Mammalian genomes harbor >10⁵ LINEs that belong to a single superfamily (Singer and Skowronski, 1985; Hutchinson et al., 1989). By contrast, distinct LINE families, each including 50–80 members, coexist in D. melanogaster. In addition to Doc, six other families of LINE elements have been described so far in this organism, including the I factor (Fawcett et al., 1986), F (Di Nocera and Casari, 1987), G (Di Nocera, 1988), and jockey (Priimagi et al., 1988) elements, and type I and type II ribosomal DNA insertions (Jacubczak et al., 1990).

LINEs differ markedly from other mobile DNA sequences that also propagate by the retrotranscription of RNA intermediates such as copia-like elements in D. melanogaster (Finnegan and Fawcett, 1986) and the Ty element (Boeke et al., 1985) in Saccharomyces cerevisiae. These elements, also known as viral retrotransposons, resemble the integrated genomes of retroviruses as they carry LTRs. LINEs lack LTRs, and their expression is controlled by promoters which are located within the transcribed region (Mizrokhi et al., 1988; Swergold, 1990; Minchiotti and Di Nocera, 1991; Minakami et al., 1992; Contursi et al., 1993; McLean et al., 1993).

By means of transient transfection assays we monitored the expression of constructs in which the reporter CAT gene was under the control of various Doc DNA segments in Drosophila Schneider II (S2) cells. We show that distinct cis-acting DNA elements, clustered in a 50-bp long DNA region located at the 5' end of unit-length Doc copies, cooperate to control RNA initiation. In addition, we found that sequences located 200 bp downstream from the 5' end inhibit the expression of the reporter CAT gene in a position- and orientation-dependent manner. The inhibition appears to be due to reduced translation rather than to impaired synthesis of CAT mRNA.

MATERIALS AND METHODS
Plasmids—The clones suN, 6N, and 11N, in which the 5' end regions of the elements suff suDoc6, and Doc11, respectively, precede the CAT gene, were obtained by cloning Sali-Nael (suN), Xhol-Nael (6N), and EcoRl-Nael (11N) fragments, isolated, respectively, from pDoc, λDoc6, and λDoc11, into the Aval site of pEMBLBLCAT. pDoc, λDoc6, and λDoc11, described in Driver et al. (1989), were kindly provided by Dr. Kevin O’Hare. The clones suA, 11A, and 6A were obtained by cloning an EcoRl-Alul fragment from suN (suA), or Alul fragments from either...
6N (6A) or 11N (11A), into the BamHI site of pEMBL8CAT. Plasmids su132 and su218 carry DNA fragments that span the 5' end region of suN1 and terminate 3' at residue 132 or 218, respectively. These fragments were synthesized by polymerase chain reaction experiments in which suN DNA was amplified by using as primers the reverse sequencing primer (Biols) and either the D132 or the D218 oligomers, which were complementary to residues 112–132 and 202–218, respectively. The amplified DNA was digested with XbaI, and the fragments were gel-purified and cloned into the Aval site of pEMBL8CAT.

Doc DNA called δ corresponds to a HindIII-NaeI fragment spanning residues 219–267 of suN2δ. δ was cloned in the plasmid suA, either in the direct (1) or reverse (2) orientation. Plasmids s26 and Doc6–21 were constructed by first annealing the complementary oligonucleotides SB1/ SB2 (s26) and 6B1/6B2 (Doc6–42). The double-stranded products, which are blunt at one end and give SalI plus-strand overhangs at the other end, were cloned between the Aval and Sal site of pEMBL8CAT. Plasmids s47 and Doc6–42 were similarly obtained by first annealing the complementary oligonucleotides SB1/ SB2 (s47) and 6B1/6B2 (Doc6–42). The double-stranded DNA fragments, which are blunt at one end and give SalI plus-strand overhangs at one end and were blunt at the other end, were cloned between the SalI and HindIII sites of s26 and Doc6–21, respectively. To obtain s47 + 4 and Doc6–42 + 4, s47 and Doc6–42 DNAs were digested with SalI, and the SalI termini were made blunt-ended by the Klenow enzyme. To construct s47 – 5, s47 was digested with SalI, and Sall terminal were made blunt-ended with the Klenow terminal nucleases. Mutations were introduced in the s47 backbone by annealing complementary oligonucleotides carrying specific base changes. Double-stranded oligonucleotides were cloned into s26 either between the SalI and HindIII sites (s47a, oligonucleotides 47a1 and 47a2, 47b oligonucleotides 47b1 and 47b2, 47c oligonucleotides 47c1 and 47c2), or between the Aval and Sal sites (s47t1; oligonucleotides 47t1a and 47t1b; 47t2, oligonucleotides 47t2a and 47t2b). Constructs s26Δ and Doc6Δ were obtained by exchanging the SalI-NcoI fragments, which span the Doc DE-DE array and the 5' end of the CAT gene, between s26 and Doc6–42. Oligonucleotides used are as follows, listed in a 5' to 3' configuration: D132, GTTAGTTTGGCATACTGCAC; D218, CATTGGTCTAGGCGCAC; SA1, GAATGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTG
Regulatory Sequences in Doc Elements

Fig. 2. CAT expression is repressed by Doc sequences in S2 cells. A, CAT activity detected in cells transfected with suN, su-218, and su-132 is shown at the top. The δ region and the ε and η subregions, are diagrammed in the middle. ATGs found within δ, ε, and η subregions are highlighted. The sequence of the region 208–267 of suFα1, is reported at the bottom. Boxed residues correspond to δ sequences. The Ddel I site at the boundary between the ε and η subregions is underlined; ATGs triplets are in uppercase letters. B, effect of δ, ε, and η sequences on CAT expression directed by suFα1, and RSV promoters. The orientation of δ, ε, and η in each construct is denoted by an arrow. In this panel, as in Panel A, S2 cells, transfected with 5 μg of test plasmid and 5 μg of the internal control γF-gal, were assayed for CAT activity. Relative enzymatic activities are expressed as described in the legend to Fig. 1. C, Northern analysis of suA1 and suA2 transcripts. Total RNA (30 μg) from S2 cells co-transfected with 5 μg of γF-gal and 5 μg of either suA1 (lane 1) or suA2 (lane 2) was analyzed by Northern blot using as probes 32P-labeled DNA fragments spanning CAT and β-galactosidase sequences (see "Materials and Methods"). Bands corresponding to CAT transcripts are indicated by an arrow. Upper hybridization bands correspond to β-galactosidase transcripts directed by γF-gal.

shown). Despite the extensive changes, plasmid 6N, which carries residues 1–262 of Doc6, directed CAT expression nearly at the same levels of suN. In contrast, the construct 11N, which carries residues 1–260 of Doc11, was ~10-fold less efficient than suN in directing the expression of the CAT gene (Fig. 1).

An Alu site conserved in the Doc 5′ UTR was used to construct three clones (Fig. 1, suA, 11A, and 6A) in which ~50 bp from the 5′ end of each element preceded the CAT gene. These constructs directed CAT expression at higher (5–8-fold) levels than the parental ones, suN, 11N, and 6N (Fig. 1).

Information sufficient to direct transcription is thus restricted in Doc to a relatively small DNA interval. This contrasts what has been reported for the F element, in which basal transcription is stimulated by sequences located far downstream from the 5′ end region (Contursi et al., 1993).

Doc Sequences Inhibit CAT Expression in an Orientation- and Position-dependent Manner—Data shown in Fig. 1 suggest that the Alu-Nael interval, which was removed in plasmids of the A series, contains sequences which inhibit CAT expression. Via polymerase chain reaction cloning (see "Materials and Methods"), we obtained two 3′ deletion derivatives of suN in which Doc DNA extended 3′ to residue 218 (construct su218) or 132 (construct su132). The removal of the region 219–267 was sufficient to delete inhibitory sequences from suN (Fig. 2A). Construct su132 directed CAT expression ~2-fold less efficiently than su218. This may denote either that positive DNA sequences are located between residues 132 and 218, or that transcripts directed by the two constructs have different stability.

One interpretation of the data is that the region 219–267, which we will call herein δ, hosts a silencer-like element. Silencers generally inhibit transcription in a position and orientation independent manner (Laurenson and Rine, 1992). δ, cloned in either orientation upstream of a copy of the D. melanogaster hsp70 promoter transcribing the CAT gene, did not reduce CAT expression (data not shown). Subsequently, δ was cloned into the plasmid suA, either upstream (δsuA) or downstream (suAδ1 and suAδ2) of the suFα1 promoter. We found that CAT expression was inhibited only when δ flanked 3′, in direct orientation, the suFα1 promoter (Fig. 2B). CAT levels were reduced, again in an orientation-dependent fashion (control data not shown) in cells transfected with RSVα1, a construct in which δ had been cloned between the CAT gene and the Rous sarcoma virus (RSV) promoter (Fig. 2B).

The position- and orientation-dependent mode of action suggests that inhibitory sequences within δ act at the RNA, rather than at the DNA level. This hypothesis is supported by Northern blot analyses showing that suA1 transcripts accumulated at levels 2–3-fold higher than suA2 transcripts (Fig. 2C). Northern data also ruled out that the inhibition is associated to enhanced degradation of CAT transcripts.

Upstream AUG triplets may inhibit translation initiation at natural AUGs (Liu et al., 1984), and an ATG is present within δ (see Fig. 2A). However, by assaying derivatives of both suA and RSVCAT in which the CAT gene is flanked 5′ by either the left-hand (ε) or the right-hand (η) portion of δ (Fig. 2A), we observed CAT inhibition with the ε, but not with the η subregion, which includes the δ ATG (Fig. 2B). The ε subregion extends 5′ to residue 208 and carries an ATG, not present in δ (Fig. 2A). We cannot therefore rule out that data obtained with suA1 and RSVε reflect an inhibitory action of the ε ATG, which is the codon for the initiating methionine of the hypothetical Doc open reading frame 1, on the translation of the CAT mRNA.

Functional Organization of the Doc Promoter Region—Sites of transcription initiation in suFα2, Doc6, and Doc11 were determined by RNA primer extension analyses by using total RNA from cells transfected with suA, 6A, 11A, and an oligomer complementary to the CAT coding region as a primer. To set unambiguously the RNA 5′ ends, reaction products were electrophoresed along with sequencing ladders of the three plasmids (Fig. 3). Transcription initiates predominantly at 6 and 7 in suFα2 and at 1 and 2 in Doc6 (in both elements 1 refers to the first residue flanking the target site). Faint bands of elongation may be artifactual or mark minor sites of RNA initiation at ~20, 9, and 19 in suFα2 and at ~19, ~3, ~5, and 16 in Doc6. A faint band, corresponding to transcripts initiating at 17 in Doc11, was the only discrete product of extension detected with RNA from cells transfected with 11A.

The organization of the cis-acting elements involved in the control of Doc transcription was investigated by assaying plasmids in which the CAT gene is under the control of different versions of the suFα2 promoter region. A construct carrying the interval 1–26 directed CAT expression, but could not drive faithful RNA initiation (Fig. 4, construct s26). A 10-fold increase in CAT levels, and a correct transcription pattern, was observed by using as template a construct carrying the interval
Fig. 3. Primer extension analysis of Doc-CAT RNAs. Total RNA (40 μg) from S2 cells transfected with 11A, suA, or 6A was hybridized to a 32P-5′-end-labeled 30-mer (CAT primer) complementary to the CAT gene sense strand. Annulled primer mobilities 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 11A, suA, and 6A templates obtained by the dideoxy chain termination method using the CAT primer. Predominant RNA start sites, marked by arrows, are shown along with the DNA sequence of the 5′ end of each element at the bottom. Lowercase letters denote flanking DNA: target sites setting the 5′ boundaries of su(f)S2 and Doc are underlined. Residue 2 in the su(f)S2 element is A and not G as previously reported (Driver et al., 1989).

1–47 (Fig. 4, s47). In s26 and s47, the dinucleotide TT, found in su(f)S2 at residues 25–26, is replaced by GA. The change, which created in s47 a Sall site between residues 22 and 27 (Fig. 4A), did not perturb promoter function, as judged by comparing the expression of suA and s47 (data not shown). By repairing the termini of s47 DNA after Sall cleavage with either the Klenow enzyme or the mung bean nuclease, we obtained constructs in which the interval 23–47 of su(f)S2 was moved either 4 bp away from (s47 + 4) or 5 bp closer to (s47 − 5) the RNA start site(s). Both space changes severely reduced CAT expression (Fig. 4A). Primer extension data showed that faithful RNA initiation was abolished in the construct s47 + 4 (Fig. 4B).

The interval 25–47 spans a DNA sequence (AGACGTGT, residues 33–42; see Fig. 4A) which is conserved in all Drosophila LINEs (consensus RGACGTGY; see Fig. 6). The transcriptional pattern of the construct 47b indicates that this sequence has a key role in Doc transcription (Fig. 4, A and B). The reduced template activity of both s47 + 4 and s47 − 5 may be due to the novel position of the RGACGTGY motif, which is located in all Drosophila LINEs at the same distance from the RNA start site(s) (see Fig. 6). However, the situation is more complex, since we found that base changes introduced either to the left (residues 28–32, construct s47a) or to the right of the AGACGTGT sequence (residues 43–47, construct s47c) also impaired promoter function (Fig. 4, A and B). Results thus indicate that three adjacent downstream sequences, herein called DE1, DE2, and DE3 (Fig. 4A) stimulate transcription of the su(f)S2 element.

Sequences spanning the RNA start sites also have a critical role in transcriptional promotion. The RNA start site regions of su(f)S2 (GATTGC, residues 6–11) and Doc6 (CAGTCG, residues 1–6) fit the consensus (C/G)AYTCG (RNA start sites are underlined; see Fig. 3). The notion that the (C/G)AYTCG motif spans an initiator (Inr) module is supported by the analysis of the construct 47in1. This clone, in which residues 6–10 of su(f)S2 have been changed from GATTG to CAGTG, directed CAT expression 50-fold less efficiently than the parental construct s47 (Fig. 4A). We have also constructed a derivative of s47 (s47in2) in which residues 1–7 of su(f)S2 have been replaced by the heptamer TGCCTTC, which is the sequence found at the same relative position in Doc11 (see Fig. 6). Base changes introduced in s47in2 did not impair CAT expression (Fig. 4A). This result suggests that the template activity of Doc11 does not correlate with base changes in the Inr region, but possibly reflects the negative interference of flanking genomic DNA.

The organization of promoter sequences is similar, on the whole, in su(f)S2 and Doc6 (Fig. 5). The construct Doc6–21, which carries the interval 1–21 of Doc6, directed CAT expression 50-fold less efficiently than Doc6–42, a construct in which Doc6 DNA extended 3′ to residue 42 (Fig. 5). Faithful RNA initiation was observed with Doc6–42, but not with Doc6–21 (data not shown). By mutating residues 17 and 22 of Doc6, a Sall site was also introduced into Doc6–42. Sall sites in s47 and Doc6–42 are at the same distance from the RNA start site region (Fig. 5). Similarly to what was observed with the s47 + 4 construct (Fig. 4), the insertion of 4 bp in the Doc6 promoter region significantly reduced CAT expression (Fig. 5, construct Doc6–42 + 4) and abolished faithful RNA initiation (data not shown). The DNA region dislodged in Doc6–42 + 4 corresponds to the DE1–DE3 array. Interestingly, while DE2...
Analysis of the Doc6 promoter.

Residues 1–42 from Doc6 flanking the CAT gene in Doc6–42 are shown at the top. In the clones listed below, sequence identities are denoted by dashes. Residues 1–47 from su(f)S2, flanking the 5′ end, are shown at the bottom. In Doc6–42, su(f)S2, residues from s47 are in uppercase letters. The Sail sites in Doc6–42 are shown. Relative CAT activities and S.D. were calculated as described in the legend to Fig. 1.

(with a single base pair change) and DE3 are conserved in Doc6 and su(f)S2, the DE1 region varies (Figs. 5 and 6). Comparison of the template activities of Doc6–42, s47, and two chimeric clones (Doc6/s and s/Doc6) in which regulatory sequences of Doc6 and su(f)S2 have been exchanged, suggests that the two DE1–DE3 arrays are functionally equivalent (Fig. 5).

Doc Elements Lack an Antisense Promoter—In addition to a sense promoter (Finit) transcribing toward the 3′ end, the 5′ end region of the D. melanogaster F element also hosts an antisense promoter (Fat) located ∼ 100 bp downstream from the Fin RNA start site (Minchiotti and Di Nocera, 1991; Contursi et al., 1993). The relatedness of Doc and F elements (see O’Hare et al. (1991)) prompted the search for an antisense promoter in the 5′ end region of Doc. suN-inv, a construct in which the DNA region cloned in suN (Fig. 1) is reversed with respect to the CAT gene, directed CAP expression ∼60 times less efficiently than F-cat2, a construct in which the CAT gene is flanked 5′ by the 267/1 region from the F element (see Minchiotti and Di Nocera, 1991). We were unable to map, by primer extension and RNase protection experiments, specific sites of initiation of CAP transcripts within Doc DNA (data not shown). Deleting DNA from either end of the suN-inv insert did not stimulate CAP expression (data not shown).

From these results we conclude that the 5′ end region of Doc elements lacks an antisense promoter. CAP activity directed by the construct suN-inv likely results from the translation of heterogeneous transcripts initiated within vector sequences.

**DISCUSSION**

In many polII genes, transcriptional signals are located downstream from the CAP site (Ayer and Dynan, 1988; Soeller et al., 1988; Perkins et al., 1988; Thummel, 1989; Nakatani et al., 1990; Harihara et al., 1991; Friedel and Searles, 1992). LINEs represent an interesting system for studying the nature and organization of intragenic promoters, because their expression is regulated by DNA elements that are largely internal to the transcriptional unit. Data presented in this work add knowledge to the picture emerging from analyses carried out with other Drosophila LINEs in many respects.

In spite of significant sequence divergence, the 5′ end regions of su(f)S2 and Doc6, two “full-length” members of the Drosophila Doc LINE family, directed CAP expression at comparable levels in S2 cells (Fig. 1). Transcription is initiated in the 5′ end region of the DNA tract specifying the consensus (C/G)AYTCG motif (Fig. 3). A related motif, CATTCG, is found at the 5′ end of the Drosophila jockey elements (see Fig. 6), and its deletion abolished jockey-dependent expression in S2 cells (Mizrokhi et al., 1988). All of these hexamers share an A at the RNA start site flanked by pyrimidines, a sequence which is the core motif in many Inr (Weis and Reinberg, 1992; Jhavery et al., 1994). The notion that the (C/G)AYTCG motif overlaps an Inr motif in many Inr (Weis and Reinberg, 1992; Jahavery et al., 1994). The notion that the (C/G)AYTCG motif overlaps a Inr motif in many Inr (Weis and Reinberg, 1992; Jahavery et al., 1994). The notion that the (C/G)AYTCG motif overlaps a Inr site has not been experimentally verified (see O’Hare et al., 1991).

Three copies of the sequence (C/G)AYTCG, and one copy of the sequence CATTCCG, flank 3′ the RNA start site regions of Doc6 and su(f)S2, respectively (Fig. 6). None of these repeats primes transcription (Fig. 3). Selectivity is not dictated by base content, since the sequence CATTCCG is found both in su(f)S2 and Doc6, but transcription is initiated within the su(f)S2 hexamer only. We favor the hypothesis that a functional hierarchy is imposed by the nature of the trans-acting factors interacting with Doc promoters. According to this view, sequence redundancy may facilitate the binding of a protein to the Doc 5′ end. In transcriptionally competent complexes, however, such a protein would be able to recognize only one target as Inr because of stereo-specific interactions with one or more factors bound to downstream promoter elements. The fidelity and the efficiency of transcription, both in su(f)S2 and Doc6, is controlled by a DNA region located ∼ 20 bp downstream from the RNA start sites (Figs. 4 and 5). Sequences found at nearly the same place within the 5′ UTR of the HIV-1 enhance the activity of the HIV-1 Inr in a distance-independent fashion (Zenie-Gregory et al., 1993). In contrast, the distance between the Inr and 3′ flanking sequences is critical in the Drosophilia mdg1 (Arkhipova and Ilyin, 1991) and the adenovirus IVa2 (Chen et al., 1994) promoters. Space changes are similarly not tolerated in the Doc promoter (Figs. 4 and 5). Although it cannot be formally excluded that negatively acting sequences had been inserted (or created) both in s47 + 4 and Doc6–42 + 4, and crucial ones had been deleted in s47 – 5, we believe that transcription in all of these constructs is inhibited by loss of protein-protein interactions due to the misalignment of promoter modules. Site-directed mutations introduced in the su(f)S2 promoter revealed that multiple DNA elements located downstream from the 5′ promoter stimulated transcription. The key element is a DNA sequence (RGACGTCG motif or DE2) which is conserved in all Drosophila LINEs at a fixed distance from the RNA start site(s) (Fig. 6). Similarly to what has been
reported for jockey (Mizokhi and Mazo, 1990) and F (Contursi et al., 1993) elements, DE2 is absolutely required for Doc transcription (Fig. 4). To a less but significant extent, base changes introduced within the adjacent DE1 and DE3 regions also reduced CAT expression (Fig. 4). This finding is novel and suggests that the functional organization of Drosophila LINE promoters may be more complex than predicted from previous transient transfection assays. Sequences similar to DE1 or DE3 are found neither in other LINEs (Fig. 6) nor in Antenna-pedia and engrailed, two Drosophila genes in which transcription is also regulated by RGACGTGY motifs located in the 5’ UTR (Soeller et al., 1988; Perkins et al., 1988). Since the analysis of Doc6 and su(f)D3 chimeric constructs established that DE1–DE3 arrays are functionally equivalent (Fig. 5), the peculiar substitution of DE1 sequences in Doc6 with a copy of the RNA start site region (Fig. 6) leads to speculate that the same protein may bind both to the Inr and the DE1 region. This hypothesis is supported by knowledge that some Inrs are recognized by transcription factors with multiple binding specificities (Seto et al., 1991; Du et al., 1993). On the other hand, it cannot be ruled out that two distinct proteins interact with the Inr and DE1 in the su(f) promoter.

Sequences analogous to DE1 and DE3 are plausibly present in other LINEs. In the 1 factor, an Inr-like element spanning residues 1–4 is flanked 3’ by a RGACGTGY motif spanning residues 29–36 (Fig. 6). In contrast with data reported in this work, deletion of the regions corresponding to DE2 and DE3 in Doc reduced only 2-fold the transcription of the 1 factor (McLean et al., 1993). This finding, and the inability of the Inr factor region 1–20 to direct faithful transcription in S2 cells,2 favor the notion that in this LINE sequences crucial for transcription are located in the region corresponding to DE1 in Doc. The relative contribution of downstream cis-acting elements to transcriptional promotion may thus vary among Drosophila LINEs, and plausibly rely, because of the heterogeneity of DE1 and DE3 regions, on the recruitment of distinct trans-acting factors.

Significant similarities between the 5’ UTRs of Doc and other Drosophila LINEs are restricted to matches in the promoter region. Therefore it is not surprising that loc, although closely related to the F element at the gene products level (O’Hare et al., 1991), lacks an antisense promoter.

Selective accumulation of LINE transcripts in specific developmental stages or cell types has been described (Mizokhi et al., 1988; Chabosier et al., 1990; Lachaume et al., 1992; Martin and Branciforte, 1993; Minchiotti et al., 1994). Less is known, however, about the synthesis of LINE proteins (McMillan and Singer, 1993; Branciforte and Martin, 1994; Trelgan and Martin, 1995). Data reported in this study suggest that the expression of Doc elements may be regulated at the gene products level. Sequences found at the boundary between the 5’ UTR and the open reading frame 1 region inhibited CAT expression in an orientation- and position-dependent manner (Fig. 2, A and B). Northern RNA bloting data (Fig. 2C) substantiated the hypothesis that inhibitory sequences act at a post-transcriptional level, and ruled out that reduced CAT production is due to enhanced degradation of CAT mRNA. We cannot exclude that ATG triplets present within Δ and ε regions inhibit CAT expression by impairing translation of the CAT mRNA. This hypothesis is, however, weakened by results obtained with the Doc η subregion (Fig. 2B). We favor the hypothesis that, by interacting with a cellular component, transcripts carrying δ (or η) sequences are somehow compartmentalized, thereby reducing their translation. The effect of inhibitory sequences was more pronounced in transcripts driven by the RSV promoter (Fig. 2B). This may reflect differences in the folding of inhibitory sequences and consequently in the ability to interact with a cellular component, within distinct RNAs.

A DNA region exhibiting ~70% homology to the ε-δ interval is found in the F element between residues 184 and 243. A 3’ deletion derivative lacking this region (Fin3’ + 175) is transcribed ~10-fold less efficiently than F-cat1, a construct in which F DNA extended 3’ to residue 267 (Minchiotti and Nocera, 1991). This result is correlated to the removal of sequences located between 193 and 207 which stimulate F basal transcription (Contursi et al., 1993). However, Fin3’ + 175 directs CAT expression only 2-fold less efficiently than F-cat1.3 This observation indirectly suggests that inhibitory sequences may be conserved in F. Germ line transformation experiments will eventually clarify whether this type of post-transcriptional regulation operates in vivo and if the inhibition is enhanced (or relieved) in specific cell types.

At the moment, genetic conditions which trigger transposition are known only for the Drosophila LINE 1 factor (Bucheton, 1990). In this respect, the isolation of a fly stock in which members of the Doc family transpose at a high rate (Pasyukova and Nazhdin, 1993) provides the base for analyses aimed at the characterization of the mechanisms that control the expression and mobilization of the Doc retrotransposon in the organism.

Acknowledgments—We are indebted to Kevin O’Hare for the gift of Doc clones. We thank Giovanna Grimaldi, Carmelo Bruni, and Virginia H. Goekjian for critical reading of the manuscript.

REFERENCES

Arkhipova, I. I., and Ilyin, Y. V. (1991) EMBO J. 10, 1169–1177
Ayer, D. E., and Dynan, W. S. (1988) Mol. Cell. Biol. 8, 2121–2133
Boeke, J. D., Garfinkel, D. J., Styles, C. A., and Fink, G. R. (1985) Cell 40, 491–500
Branciforte, D., and Martin, S. L. (1994) Mol. Cell. Biol. 14, 2584–2592
Bucheton, A. (1990) Trends Genet. 6, 1–5
Chabosier, M. C., Busseau, I., Prasser, J., Finnegan, D. J., and Bucheton, A. (1990) EMBO J. 9, 3557–3563
Chen, H., Vinnakkot, K., and Flint, S. J. (1994) Mol. Cell. Biol. 14, 676–685
Contursi, C., Minchiotti, G., and Di Nocera, P. P. (1993) J. Mol. Biol. 224, 398–407
Di Nocera, P. P. (1988) Nucleic Acids Res. 16, 4011–4025
Di Nocera, P. P., and Casari, G. (1987) Proc. Natl. Acad. Sci. U.S.A. 84, 7883–7887
Di Nocera, P. P., and Dawid, I. B. (1988) Proc. Natl. Acad. Sci. U.S.A. 85, 7095–7098
Di Nocera, P. P., Feng, D. F., Johnson, M. S., and McClure, M. A. (1989) Q. Rev. Biol. 64, 1–30
Driver, A., Lacey, S. F., Culliford, T. E., Mitchelson, A., and O’Hare, K. (1989) Mol. Gen. Genet. 220, 49–52
Du, H., Roy, A. L., and Roeder, R. G. (1993) EMBO J. 12, 501–511
Evans, J. P., and Palmer, R. D. (1991) Proc. Natl. Acad. Sci. U.S.A. 88, 8792–8795
Fawcett, D. H., Lister, C. K., Kellett, E., and Finnegan, D. J. (1986) Cell 47, 1007–1015
Finnegan, D. J., and Fawcett, D. H. (1986) Oxf. Surv. Eukaryotic Genes 3, 1–62
Fridell, Y. C., and Searles, L. L. (1992) Mol. Cell. Biol. 12, 4571–4577
Grimaldi, G., and Di Nocera, P. P. (1988) Proc. Natl. Acad. Sci. U.S.A. 85, 5502–5506
Harinaran, N., Kelley, D. E., and Perry, R. P. (1991) Proc. Natl. Acad. Sci. U.S.A. 88, 9799–9803
Hattori, M., and Sakaki, Y. (1986) Anal Biochem. 152, 232–238
Hutchison, C., III, Hardies, S. C., Loeb, D. D., Shehee, R. W., and Edgell, M. H. (1988) in Mobile DNA (Berg, D. E., and Howe, M. M., eds) pp. 593–617, American Society for Microbiology, Washington, D.C.
Jakucazki, J. L., Xiong, Y., and Eidkush, T. H. (1990) J. Mol. Biol. 212, 37–52
Jakubczak, J. L., Xiong, Y., and Eickbush, T. H. (1990) J. Mol. Biol. 234, 4571–4577
Jensen, S., and Heidmann, T. (1991) EMBO J. 10, 1927–1937
Lachaume, P. K., Bouhidel, K., Mesure, M., and Pinon, H. (1992) Development 115, 729–735
Laurensen, P., and Rine, J. (1992) Microbiol. Rev. 56, 543–560
Liu, C. C., Simonsen, C. C., and Levinson, A. D. (1984) Nature 309, 62–65
Martin, S. L., and Branciforte, D. (1991) Mol. Cell. Biol. 11, 5593–5592
McLean, C., Bucheton, A., and Finnegan, D. J. (1993) Mol. Biol. Cell. 13, 1042–1050
McMillan, J. P., and Singer, M. F. (1993) Proc. Natl. Acad. Sci. U.S.A. 90, 11533–11537
Minakami, R., Kurose, K., Furuhata, Y., Hattori, M., and Sakaki, Y. (1992) Nucleic Acids Res. 20, 3139–3145
Minchiotti, G., and Di Nocera, P. P. (1991) Mol. Cell. Biol. 11, 5171–5180

2 G. Minchiotti and P. P. Di Nocera, unpublished observations.

3 G. Minchiotti and P. P. Di Nocera, unpublished results.
Minchiotti, G., Contursi, C., Graziani, F., Gargiulo, G., and Di Nocera, P. P. (1994) Mol. & Gen. Genet. 245, 152–159
Mizrokhi, L. J., and Mazo, A. M. (1990) Proc. Natl. Acad. Sci. U. S. A. 87, 9216–9220
Mizrokhi, L. J., Georgieva, S. G., and Ilyin, Y. V. (1988) Cell 54, 685–691
Nakatani, Y., Brenner, M., and Freese, E. (1990) Proc. Natl. Acad. Sci. U. S. A. 87, 4289–4293
O’Hare, K., Alley, M. R. K., Cullingford, T. E., Driver, A., and Sanderson, M. J. (1991) Mol. & Gen. Genet. 225, 17–24
Pasyukova, E. G., and Nazhdin, S. V. (1993) Mol. & Gen. Genet. 240, 302–306
Pelisson, A., Finnegan, D. J., and Bucheton, A. (1991) Proc. Natl. Acad. Sci. U. S. A. 88, 4907–4910
Perkins, K. K., Dailey, G. M., and Tijan, R. (1988) Genes & Dev. 2, 1615–1626
Priimagi, A. F., Mizrokhi, L. J., and Ilyin, Y. V. (1988) Gene (Amst.) 70, 253–262
Sambrook, J., Fritsch, E. F., and Maniatis, T. (1989) Molecular Cloning: A Laboratory Manual, 2nd Ed., Cold Spring Harbor Laboratory, Cold Spring Harbor, NY
Schalet, A. (1986) Mutat. Res. 163, 115–144
Schneuwly, S., Kurowa, A., and Gehring, W. J. (1987) EMBO J. 6, 201–206
Seto, E., Shi, Y., and Shenk, T. (1991) Nature 354, 241–245
Singer, M. F., and Skowronski, J. (1985) Trends Biol. Sci. 10, 119–122
Soeller, W. C., Poole, S. J., and Karnberg, T. (1988) Genes & Dev. 2, 68–81
Swergold, G. D. (1990) Mol. Cell. Biol. 10, 6718–6729
Thummel, C. S. (1989) Genes & Dev. 3, 782–792
Thummel, C. S., Boulet, A. M., and Lipshitz, H. D. (1988) Gene (Amst.) 74, 445–456
Trelogan, S. A., and Martin, S. L. (1995) Proc. Natl. Acad. Sci. U. S. A. 92, 1520–1524
Weis, L., and Reinberg, D. (1992) FASEB J. 6, 3300–3309
Xiong, Y., and Eickbush, T. H. (1990) EMBO J. 9, 3353–3362
Zenzie-Gregory, B., Sheridan, P., Jones, K. A., and Smale, S. T. (1993) J. Biol. Chem. 268, 15823–15832

Regulatory Sequences in Doc Elements