Alternative activation of transcriptional initiators in *Drosophila melanogaster* LINE promoters

Gabriella Minchiotti[a], Pier Paolo Di Nocera[a,b,*]

[a] *Dipartimento di Biologia e Patologia Cellulare e Molecolare 'L. Califano', Università degli Studi di Napoli Federico II, Via Pansini 5, 80131 Naples, Italy*

[b] *Istituto Internazionale di Genetica e Biofisica, C.N.R., Via Marconi 12, Naples, Italy*

Received 10 April 1997; revised version received 26 May 1997

Abstract In the *Drosophila* I, F and Doc LINEs, basal transcription is ensured by the functional interaction of initiator sequences with intragenic regulatory segments (B regions) which comprise distinct functional modules. Removing the B regions, as changing their composition or location, allowed different activators to stimulate transcription from novel initiators both in Doc and F promoters. The use of distinct initiators plausibly reflects the assembly of transcriptional complexes in which TFIID assumes alternative spatial conformations. The response of I, F and Doc promoters to the same enhancer is significantly influenced by the number, position and type of core elements present.

Key words: Transcription regulation; Initiator; Downstream promoter; TFIID; Transcriptional activation

1. Introduction

The transcription of eukaryotic genes by RNA polymerase II (pol II) is regulated via two types of core promoter elements: the TATA box, which is typically located 25–30 bp upstream of the transcription start site, and the Inr (initiator), which encompasses the transcription start site [1]. The recognition of TATA-containing promoters is activated by the interaction of the TATA motif with TFIID, a multiprotein complex including a TATA-binding protein (TBP) and several tightly bound proteins called TAFs (TBP-associated factors) [2]. TFIID binding to promoter DNA provides a site of nucleation through which multiple general transcription factors (GTFs) and pol II sequentially associate to form a transcriptionally competent complex [1]. TFIID, through one or more TAFs, seems involved also in the recognition of Inr-containing promoters [3–5]. A peculiar class of Inrs is found in LINEs, evolutionarily conserved DNA elements [6] which use self-encoded proteins to reverse transcribe their own mRNA and integrate cDNA copies at new genomic locations [7]. The expression of LINEs is regulated by *cis*-acting signals located at and downstream from the RNA start site(s) [8–14]. In this report, we show that changes introduced in the promoter regions of distinct *Drosophila* LINEs allowed transcriptional activators to stimulate cryptic Inr modules. The response of different promoter constructs to the same enhancer is significantly influenced by the number, position and type of core elements present.

2. Materials and methods

2.1. Construction of plasmids

Plasmids I_A/I_B, D_A/D_B, F_A/F_B and Fα–F_A/F_B correspond to the constructs I/I, D/D, F/F, and F/F$^+$, respectively [14]. A *Bam*HI-*Bgl*II fragment spanning the Fα region from the element F12 [11] was cloned in I_A/I_B, D_A/D_B or F_A/F_B to obtain Fα–I_A/I_B, Fα-D_A/D_B, Fαi-D_A/D_B and Fα–F_A/F_B. *Sal*I–*Nco*I fragments spanning the promoter B regions and part of the chloramphenicol acetyltransferase (CAT) gene in the above constructs were replaced by a*Sal*I–*Nco*I fragment from pEMBL8CAT to obtain Fα–I_A, Fα–D_A, Fαi–D_A and Fα–F_A. Fαm–D_A is a derivative of Fα-D_A carrying a *Bam*HI–*Bgl*II fragment spanning a mutated Fα region. AP1–D_A was obtained by replacing the Fα region of Fα–D_A with a synthetic double stranded oligonucleotide which contains the AP1 site TGAGTCA. Sp1–D_A/D_B was obtained by inserting into the *Ava*I site of D_A/D_B a *Sph*I–*Eco*RI fragment from the plasmid GI CAT [15] spanning the six Sp1 sites found in the SV40 early promoter. Sp1–D_A was constructed by replacing the *Sal*I–*Nco*I fragment of Sp1–D_A/D_B spanning the B region and part of the CAT gene with the corresponding *Sal*I–*Nco*I fragment from Fα–D_A. Constructs described in Figs. 3 and 4 were obtained by replacing first the A promoter regions of Fα–I_A/I_B, Fα–D_A/D_B and Fα–F_A/F_B with double stranded oligonucleotides carrying specific base changes. Constructs in which the mutated A regions are flanked either 5′ by Fα, or 3′ by the B region, were subsequently obtained by replacing suitable restriction fragments with pEMBL8CAT sequences. Fα–$I_A/I_B{}^{+4}$ and Fα–$F_A/F_B{}^{+4}$ were obtained by cleaving Fα–I_A/I_B and Fα–F_A/F_B with *Sal*I, and treating the digested DNA with the Klenow enzyme prior to ligation and transformation. A *Sal*I–*Nco*I fragment spanning the B region of the Doc promoter and part of the CAT gene in Fα–D_A/D_B was replaced by *Sal*I–*Nco*I fragments from the constructs Dde1, Dde2 and Dde3 [14] to obtain Fα–DA/Dde1, Fα–DA/Dde2 or Fα–DA/Dde3, respectively. Incompatible termini were blunt-ended by T4 DNA polymerase before ligation.

2.2. DNA transfections, CAT assays and RNA analyses

3 ml of *D. melanogaster* S2 cells, seeded at a density of 1–2× 10^6/ml, were cotransfected, as described previously [13], with 5 μg of the plasmid of interest and 5 μg of γF-gal, a construct directing the expression of the *Escherichia coli* β-galactosidase gene. The amount of cellular lysate used in each CAT assay was normalized to the expression of γF-gal. CAT and β-galactosidase assays were performed as previously described [13]. CAT enzyme levels reported for the various constructs represent average values of 3–4 independent transfections carried out with different cell populations. Sp1–D_A and Sp1–D_A/D_B were cotransfected also with 0.1 μg of pPacSp1, an expression vector directing the synthesis of the Sp1 protein [15]. Total RNA was analyzed by primer extension as described [9]. Reaction products were resolved on 6% (w/v) polyacrylamide-8M urea gels. Sequencing ladders were generated by the dideoxy chain termination method utilizing double stranded DNA templates. In some transfections the plasmid pGEM180, which directs faithful polymerase I dependent transcription in S2 cells [9], was used as internal control. Primers used to detect transcripts directed by the different promoter constructs and pGEM180 have been previously described [9].

*Corresponding author. Fax: (+39) 81-7703285.
E-mail: dinocera@cds.unina.it

PII S0014-5793(97)00689-3

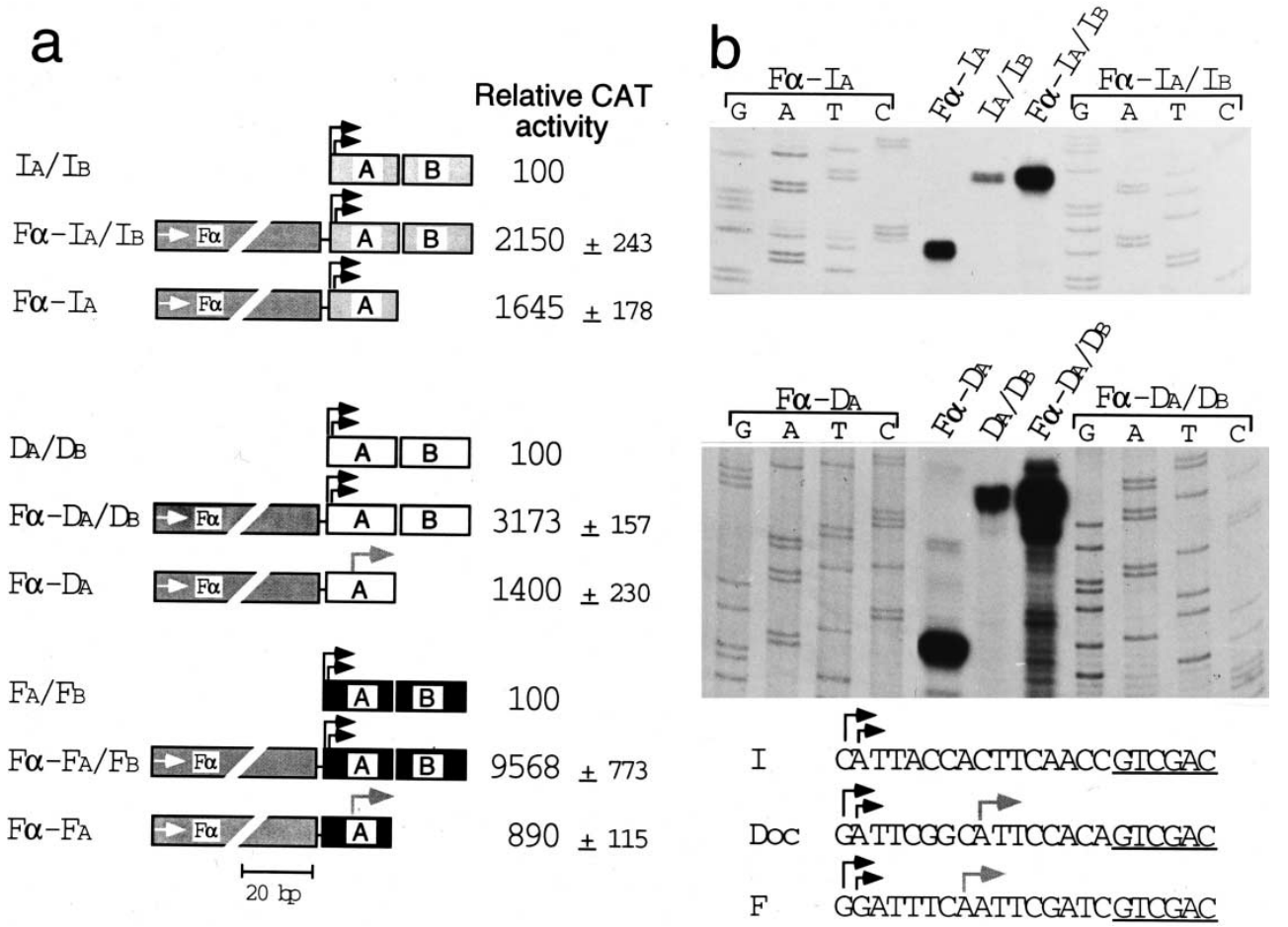

Fig. 1. Transcriptional activation by Fα. (a) DNA regions cloned upstream of the CAT gene in the different constructs are shown. (b) Total RNA (40 μg) from S2 cells transfected with 5 μg of the DNAs indicated at the top was analyzed by primer extension using a ^{32}P-5′-labeled 30-mer complementary to the CAT gene. Sequencing ladders were obtained by the dideoxy-chain termination method with the same primer. The sequence of the A regions of the I, Doc and F promoters are shown at the bottom. Black and gray arrows mark transcripts directed by the primary and secondary Inrs, respectively. *Sal*I sites created for cloning purposes [14] are underlined.

3. Results

3.1. The pattern of transcription initiation of LINE promoters can be reprogrammed by activators

In three *Drosophila* LINEs, the I factor and the Doc and F elements, core promoter elements occupy the same relative positions within two adjacent regions, called A and B, both internal to the transcribed region [14] (Fig. 1). Sequences located at the left-hand side of the A regions function as Inrs only when flanked 3′ by the B regions. The latter comprise, in each promoter, three functional elements called de1, de2 and de3 [14] (see also Fig. 5b). To study how activators stimulate this class of modular promoters, the Fα enhancer from the F LINE [11] was inserted in constructs in which either only the A regions, or both A and B regions of the I, Doc and F promoters direct the expression of the CAT gene. Plasmids were introduced by transfection into *Drosophila* S2 cells, and their expression monitored 48 h later by protein and RNA analyses. We already showed [14] that Fα stimulated ~100-fold the F promoter. The activity of I and Doc promoters was also significantly enhanced by Fα (Fig. 1). In all instances, the stimulatory effect was correlated to an increase in the accumulation of faithfully initiated transcripts (Fig. 1). Though less efficiently, constructs containing only the A promoter re-

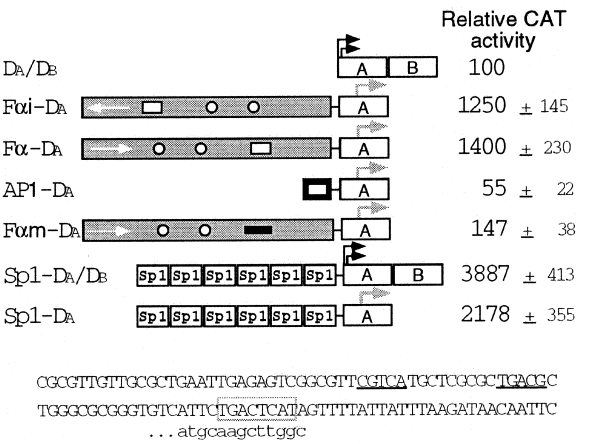

Fig. 2. Activation of the secondary Doc Inr. White arrows denote the orientation of Fα, black and gray arrows transcription initiation from the primary and secondary Doc Inr, respectively. White rectangles and circles denote AP1 and CRE I sites, respectively. Changes in Fαm–DA are denoted by a black bar. The sequence of Fα and the Fα bases deleted (dots) or mutated (lowercase letters) in Fαm-D_A are shown at the bottom. The AP1 site is boxed, the CRE I sites underlined.

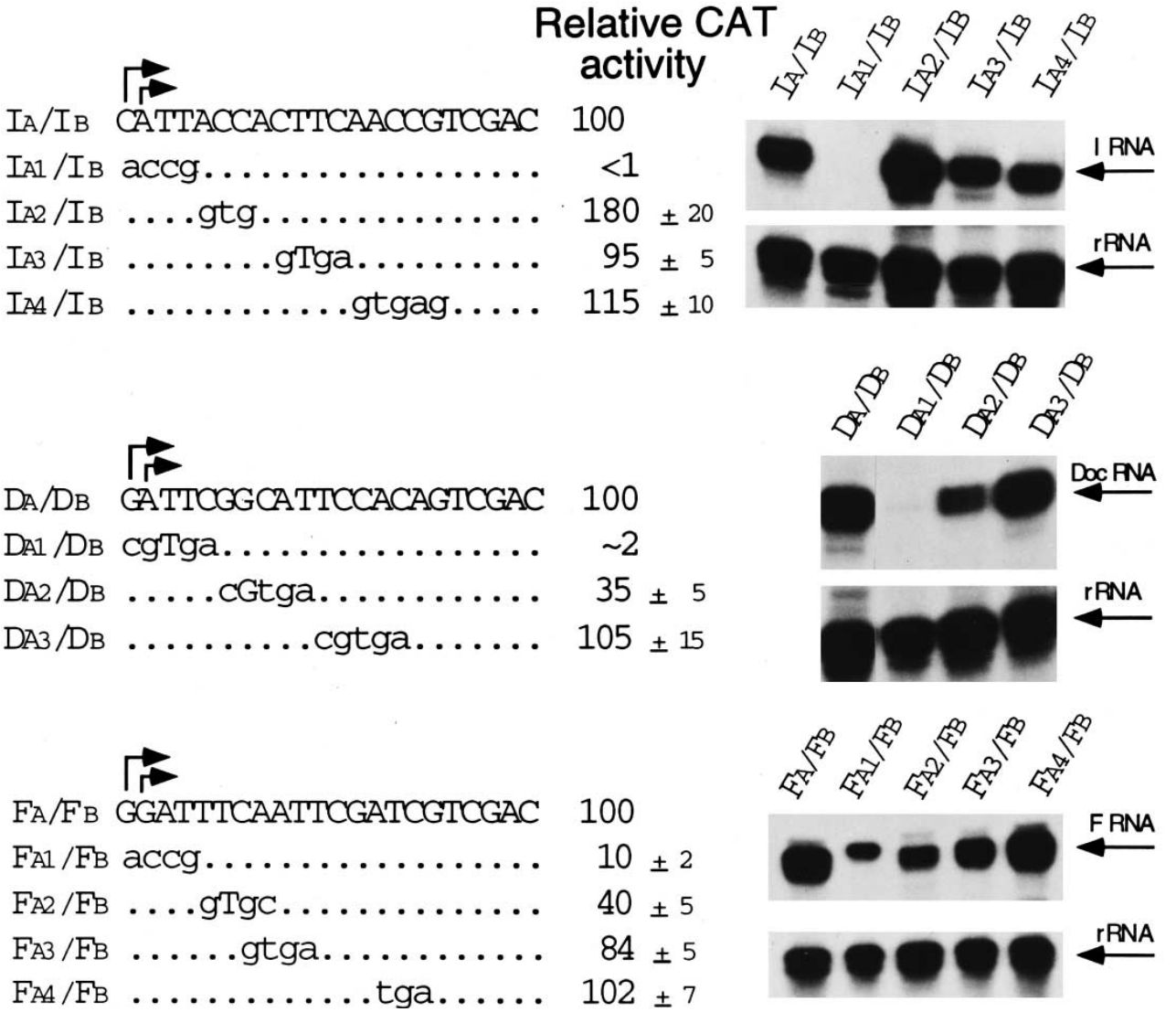

Fig. 3. Mutations introduced in the A regions affecting I, Doc and F basal transcription. In the listed constructs, A promoter regions are flanked 3′ by the corresponding B regions. Lowercase letters and dots denote base changes and sequence identities, respectively. Arrows mark the sites of RNA initiation. Total RNA (40 μg) from S2 cells cotransfected with 5 μg of the DNAs indicated at the top and 5 μg of pGEM180 was analyzed by primer extension. A distinct ^{32}P-5′-end-labeled primer was used to detect rRNA transcripts directed by pGEM180.

gion were also responsive to Fα, and directed CAT expression at levels 8–16-fold higher than basal promoters (Fig. 1). Transcripts directed by Fα–I_A/I_B and Fα–I_A initiated at the same sites (Fig. 1b). By contrast, transcription initiated both in Fα–D_A (Fig. 1b) and Fα–F_A (data not shown) at novel sites located downstream from the Doc and F Inrs, respectively (Fig. 1b). The possibility that the novel RNA start sites were activated by Fα sequences fortuitously functioning as a TATA element was ruled out by the analysis of Fαi–D_A, a construct which differs from Fα–D_A for the orientation of Fα. Fαi–D_A directed CAT expression almost as efficiently as Fα–D_A (Fig. 2), and transcription initiated at the same sites in both plasmids (data not shown). Hereafter, for clarity, the Inrs directing transcription in Doc and F constructs which contain or lack the B promoter region, will be referred to as primary and secondary, respectively.

Fα contains an AP1 site (see Fig. 2), and this site was sufficient to activate the secondary Doc Inr (construct AP1–D_A, Fig. 2; RNA data not shown). The poor template efficiency of AP1–D_A is plausibly correlated with the fact that in AP1–D_A the AP1 site and the secondary Doc Inr are only ~20 bp apart, and activators and Inrs cooperate inefficiently at such short distance [16]. Mutating the AP1 site severely reduced transcription, but still allowed selective initiation at the secondary Inr (construct Fαm–D_A, Fig. 2; RNA data not shown). Activating sequences in Fαm–D_A may coincide with one (or both) of the two ATF sites found in Fα (Fig. 2), an hypothesis supported by knowledge that ATF and AP1 sites are recognized by proteins which enhance pol II transcription through the same co-activators [17].

Sp1 is a protein particularly active in stimulating initiator-dependent transcription [16], and Sp1 sites act as transcriptional positioning elements in some TATA-less promoters [18,19]. The six Sp1 sites from the SV40 early promoter were inserted upstream of Doc promoter sequences both in Sp1–D_A and Sp1–D_A/D_B (Fig. 2). Since *Drosophila* S2 cells lack Sp1 activity [20], either clone was cotransfected with a Sp1 expression vector to detect activation. Sp1–D_A and Sp1–

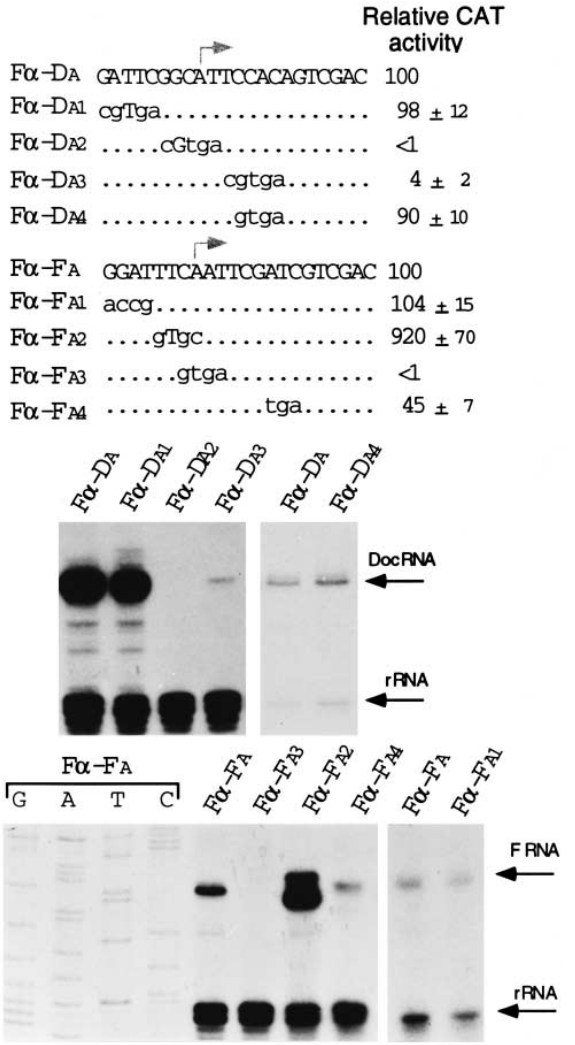

Fig. 4. Mutations introduced in the A regions of the Doc and F promoters influencing the activity of the secondary Inrs. In the listed constructs, wild-type and mutated A regions from the Doc and F promoters are flanked 5′ by Fα. Lowercase letters, dots and arrows are as in Fig. 3. Primer extension analyses were carried out as described in Fig. 3.

D_A/D_B directed CAT expression ~20- and 40-fold more efficiently than the basal Doc promoter, respectively (Fig. 2). Similarly to what observed with Fα, transcription was directed (data not shown) by the primary Inr in the presence of the Doc B region, but by the secondary Inr in its absence.

3.2. *Primary and secondary Inrs are functionally independent promoter modules*

We monitored the expression of constructs in which mutated versions of the A regions of the I, Doc or F promoters were flanked either 3′ by the corresponding B regions (Fig. 3), or 5′ by Fα (Fig. 4). In Doc and F constructs, base changes which either knocked out or severely reduced basal transcription did not affect the activation of the secondary Inrs (compare the activities of the Doc derivatives D_{A1}/D_B and Fα–D_{A1} and the F derivatives F_{A1}/F_B and Fα–F_{A1} in Figs. 3 and 4). In turn, mutations inhibiting Fα-dependent transcription did not significantly impair the accumulation of Doc and F transcripts directed by the primary Inrs (compare the activities of Fα–D_{A2}, Fα–D_{A3} and Fα–F_{A3} in Fig. 4 with those of the corresponding constructs D_{A2}/D_B, D_{A3}/D_B and F_{A3}/F_B in Fig. 3). Data support the notion that primary and secondary Inrs are functionally independent modules alternatively recruited in different promoter contexts. In the I promoter, bases crucial for transcription are restricted to the sequence CATT (Fig. 3). Since the I promoter lacks a secondary Inr, the response to Fα of the mutated A regions I_{A1} to I_{A4} shown in Fig. 3 was not investigated.

The comparison of the constructs Fα–D_{A2}, Fα–D_{A3} and Fα–D_{A4} (Fig. 4) reveals that bases crucial for the activity of the secondary Doc Inr are restricted to the sequence CATT. In the F promoter, changing the related motif AATT to TGAT inactivated the secondary Inr (Fig. 4, construct Fα–F_{A3}). By contrast, the change to CATT increased its activity ~10-fold (Fig. 4, construct Fα–F_{A2}). Data suggest that secondary Inrs are recruited on a sequence-specific base, as they match, or closely resemble, the sequence CATT. Accordingly, the I promoter lacks a secondary Inr because its basal expression is already directed by the sequence CATT, which is an optimal Inr module [21].

3.3. *The location and composition of the B regions influence both qualitatively and quantitatively the response to activators*

The insertion of 4 bp between the A and B regions reduced the response of F and I promoters to Fα ~5-fold (Fig. 5a). We cannot formally exclude that the change in sequence per se reduced the activity of the B regions. We consider this possibility unlikely, and believe that the insertion inhibited the transcription of either promoter by altering the space between the A and B regions, since previous work has shown that the B regions cooperate in a strictly space-dependent manner with primary Inrs to direct basal transcription [11,13,14]. In Fα–F_A/F_B^{+4} the B region has a marginal role in transcriptional control. Fα–F_A/F_B^{+4} is expressed roughly at the same level of Fα–F_A (compare data in Figs. 1 and 5a),

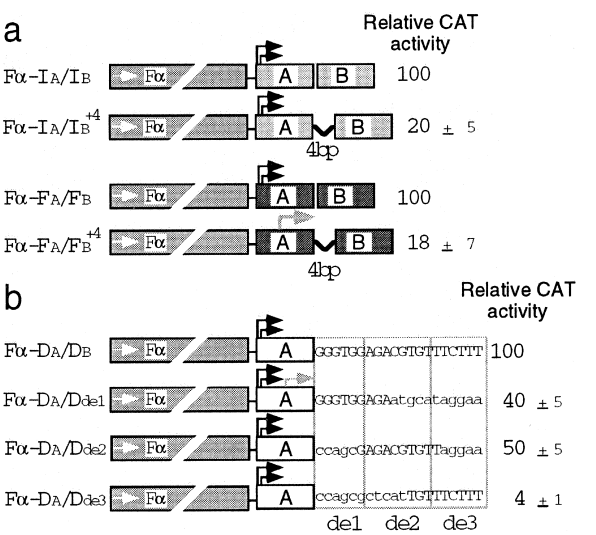

Fig. 5. (a) Misalignment of the A and B regions inhibits the activation by Fα. Black and gray arrows denote RNA initiation at the primary and secondary Inrs, respectively. (b) Activation of Doc promoters containing mutated B regions. Changes introduced in the Doc B region are in lowercase letters. Transcripts directed by the primary and secondary Doc Inrs are marked by black and gray arrows, respectively.

and $F_A/F_B{}^{+4}$ transcripts initiate at the secondary F Inr (data not shown). Dislodgment of the B region did not alter the pattern of RNA initiation in Fα–$I_A/I_B{}^{+4}$ (data not shown). It is of interest to note that Fα–$I_A/I_B{}^{+4}$ is a template ∼4-fold less active than Fα–I_A (see Figs. 1 and 5a). Though transcripts directed by these two constructs may have different stability, we believe that the B region is negatively affecting transcription in Fα–$I_A/I_B{}^{+4}$.

The B regions share the de2 module, but include different de1 and de3 modules [14]. The hypothesis that specific components of the B regions may differently affect transcriptional activation in specific promoter contexts is reinforced by the analysis of derivatives of the Doc promoter containing only de1, de2 or de3 (Fig. 5b). Primer extension analyses (data not shown) revealed that transcription was predominantly directed by the primary Inr in Fα–D_A/D_{de2}, and by the secondary Inr in Fα–$D_A/D_{de1.}$ At low levels, the primary Doc Inr was active in Fα–D_A/D_{de1}. In Fα–D_A/D_{de3}, the replacement of the B region with de3 still allowed selective initiation at the primary Inr (data not shown), but reduced dramatically CAT expression levels (Fig. 5b). While it cannot be ruled out that foreign DNA inserted to mutate de1 and de2 may have significantly altered the stability of transcripts directed by Fα–D_A/D_{de3}, data suggest that de3, as single downstream promoter element, severely antagonizes the stimulatory effect of Fα.

4. Discussion

Data presented in this work provide valuable information on the role that specific intragenic cis-acting signals may play both in the recognition and activation of TATA-less pol II promoters. The basal promoters of the I, Doc and F *Drosophila* LINEs have the same functional architecture, and can be viewed as two adjacent regions, A and B. In all, basal transcription is ensured by the cooperation of initiator sequences located in the A region with sequence elements (de1, de2 and de3) located in the B region, which stimulate transcription by providing sites of interaction for the TFIID complex [14,22]. Fα, an enhancer-like region from the F LINE, activated the three promoters by increasing the levels of faithfully initiated transcripts (Fig. 1). Templates lacking the B region were also responsive to Fα. Unexpectedly, both in F and Doc promoters, the removal of the B region allowed Fα to selectively activate novel or secondary Inrs located downstream of the physiological or primary Inrs (Fig. 1). An AP1 site within Fα played a major role in reprogramming the pattern of initiation (Fig. 2). Additional Fα sequences, plausibly ATF sites, could also activate, though at low efficiency, the secondary Doc Inr (Fig. 2). Selective Inr recruitment was also observed by using Sp1 sites as activating sequences (Fig. 2). Primary and secondary Inrs function independently from each other (Figs. 3 and 4). In order to work, the former must be located at a specific distance from a B region, or part of it [14] (Fig. 5). In contrast, secondary Inrs appear to be selected on a sequence-dependent base, as they match, or strictly resemble, the optimal Inr motif CATT. This sequence coincides with the physiological Inr of the I promoter, and this reasonably explains why the latter lacks a secondary Inr (Fig. 1).

Transcriptional complexes enabling the alternative activation of distinct Inrs must somehow differ. We hypothesize that in *Drosophila* LINE promoters the utilization of distinct Inrs is modulated by changes in the conformation of the TFIID complex. By providing contact sites for one or more TAFs [14] the B regions plausibly influence the spatial arrangement of TFIID, and dictate interactions between TFIID, the other GTFs and pol II favouring the activation of the primary Inrs. According to this view, in the absence of constraints imposed by the B regions, TFIID assumes a different conformation, and nucleates the assembly of transcriptional complexes which allow the activation of the secondary Inrs. The same holds true for templates in which either the location or the composition of the B regions was changed (Fig. 5). It is of interest to note that both modifications, in addition to influence the RNA start site selection, invariably impaired transcriptional enhancement. Upon the insertion of bases between the A and B regions, the activation of the F and I promoters dropped to levels similar to those measured in the absence the B region, or to levels 4-fold lower, respectively (compare the activities of Fα–F_A, Fα–F_A/F_B and Fα–$F_A/F_B{}^{+4}$ with those of Fα–I_A, Fα–I_A/I_B and Fα–$I_A/I_B{}^{+4}$ in Fig. 1 and 5A). The comparison of data shown in Figs. 1 and 5b reveals that derivatives of the Doc promoter containing single components of the B region were stimulated by Fα either as efficiently as the Doc A region alone, or at levels ∼10-fold lower. The F promoter was activated by Fα 2-fold less efficiently when the B region was replaced by de2, but ∼100-fold less efficiently when the B region was replaced by either de1 or de3 [14]. Doc and F de1 exhibit poor sequence homology [14]. It is therefore not surprising that these modules, when assayed as single downstream promoter elements, may differently affect transcription activation.

Sequences homologous to de1, de2 and de3 modules are located 20 to 30 bp downstream from the cap site in several eukaryotic genes [9,14,22]. In light of the results presented here, it will be of interest to investigate the relationship between the activity of other pol II transcriptional units and the number, position and type of core promoter elements present in the 5′ untranslated region. Analyses aimed at the identification of the proteins which interact with de1, de2 and de3 should clarify how these downstream promoter modules are recognized by the transcriptional apparatus, and why they can either stimulate or inhibit transcription in specific promoter contexts.

Acknowledgements: This work was supported by grants from MURST, P.F. Ingegneria Genetica of the C.N.R. and Commission of the European Communities (contract ERBSC1*CT920811).

References

[1] Zawel, L. and Reinberg, D. (1995) Annu. Rev. Biochem. 64, 533–561.
[2] Dynlacht, B.D., Hoey, T. and Tjian, R. (1991) Cell 66, 563–576.
[3] Kauffmann, J. and Smale, S. (1994) Genes Dev. 8, 821–829.
[4] Martinez, E., Chiang, C.-M., Ge, H. and Roeder, R.G. (1994) EMBO J. 13, 3115–3126.
[5] Verrijzer, C.P., Yokomori, K., Chen, J.L. and Tjian, R. (1994) Science 264, 933–941.
[6] Singer, M.F. and Skowronski, J. (1985) Trends Biol. Sci. 10, 119–122.
[7] Pelisson, A., Finnegan, D.J. and Bucheton, A. (1991) Proc. Natl. Acad. Sci. USA 88, 4907–4910.
[8] Mizrokhi, L.J. and Mazo, A.M. (1990) Proc. Natl. Acad. Sci. USA 87, 9216–9220.
[9] Minchiotti, G. and Di Nocera, P.P. (1991) Mol. Cell. Biol. 11, 5171–5180.

[10] Minakami, R., Kurose, K., Etoh, K., Furuhata, Y., Hattori, M. and Sakaki, Y. (1992) Nucleic Acids Res. 20, 3139–3145.
[11] Contursi, C., Minchiotti, G. and Di Nocera, P.P. (1993) J. Mol. Biol. 234, 988–997.
[12] Mclean, C., Bucheton, A. and Finnegan, D.J. (1993) Mol. Cell. Biol. 13, 1042–1050.
[13] Contursi, C., Minchiotti, G. and Di Nocera, P.P. (1995) J. Biol. Chem. 270, 26570–26576.
[14] Minchiotti, G., Contursi, C. and Di Nocera, P.P. (1997) J. Mol. Biol. 267, 37–46.
[15] Colgan, J. and Manley, J.L. (1992) Genes Dev. 6, 304–315.
[16] Smale, S.T. and Baltimore, D. (1989) Cell 57, 103–113.
[17] Arias, J., Alberts, A.S., Brindle, P., Claret, F.X., Smeal, T., Karin, M., Feramisco, J. and Montminy, M. (1994) Nature 370, 226–229.
[18] Kollmar, R., Sukow, K.A., Sponagle, S.K. and Farnham, P.J. (1994) J. Biol. Chem. 269, 2252–2257.
[19] Lu, J., Lee, W., Jiang, C. and Keller, E.B. (1994) J. Biol. Chem. 269, 5391–5402.
[20] Courey, A.J. and Tijan, R. (1988) Cell 55, 887–898.
[21] Weis, L. and Reinberg, D. (1992) FASEB J. 6, 3300–3309.
[22] Burke, T.W. and Kadonaga, J.T. (1996) Genes Dev. 10, 711–724.