

The Downstream Promoter Element DPE Appears To Be as Widely Used as the TATA Box in *Drosophila* Core Promoters

ALAN K. KUTACH AND JAMES T. KADONAGA*

Department of Biology and Center for Molecular Genetics, University of California,
San Diego, La Jolla, California 92093-0347

Received 28 February 2000/Returned for modification 28 March 2000/Accepted 31 March 2000

The downstream promoter element (DPE) functions cooperatively with the initiator (Inr) for the binding of TFIID in the transcription of core promoters in the absence of a TATA box. We examined the properties of sequences that can function as a DPE as well as the range of promoters that use the DPE as a core promoter element. By using an in vitro transcription assay, we identified 17 new DPE-dependent promoters and found that all possessed identical spacing between the Inr and DPE. Moreover, mutational analysis indicated that the insertion or deletion of a single nucleotide between the Inr and DPE causes a reduction in transcriptional activity and TFIID binding. To explore the range of sequences that can function as a DPE, we constructed and analyzed randomized promoter libraries. These experiments yielded the DPE functional range set, which represents sequences that contribute to or are compatible with DPE function. We then analyzed the DPE functional range set in conjunction with a *Drosophila* core promoter database that we compiled from 205 promoters with accurately mapped start sites. Somewhat surprisingly, the DPE sequence motif is as common as the TATA box in *Drosophila* promoters. There is, in addition, a striking adherence of Inr sequences to the Inr consensus in DPE-containing promoters relative to DPE-less promoters. Furthermore, statistical and biochemical analyses indicated that a G nucleotide between the Inr and DPE contributes to transcription from DPE-containing promoters. Thus, these data reveal that the DPE exhibits a strict spacing requirement yet some sequence flexibility and appears to be as widely used as the TATA box in *Drosophila*.

Transcription by RNA polymerase II is the target of many regulatory signals that are mediated by an array of molecules ranging from simple ions to multifunctional protein complexes. These signals are integrated at the core promoter to determine the extent to which each gene is transcribed. Thus, study of the interactions of the *cis*-acting DNA sequences and *trans*-acting proteins at the core promoter is essential to understand the diverse array of transcriptional regulatory processes that occur within living organisms (for reviews, see references 2, 15, 28, 34, 38, and 43).

The core promoter comprises the DNA sequences that direct the RNA polymerase II transcriptional machinery to the site of initiation. At present, four DNA elements have been found to be involved in core promoter function: the TATA box, the TFIIB recognition element (BRE), the initiator (Inr), and the downstream promoter element (DPE). The TATA box is an A/T-rich sequence, typically located about 20 to 30 nucleotides upstream of the transcription start site, that is bound by the TATA-binding protein (TBP) subunit of the TFIID complex (for reviews, see references 6, 31, and 37). The consensus for the TATA box is typically designated as TATAAA, although significant variation in sequences that can function as TATA elements has been observed (36, 47). In addition, the BRE, which has the consensus G/C-G/C-G/A-C-G-C-C, is located immediately upstream of the TATA element of some promoters and increases the affinity of TFIIB for the promoter (22).

The Inr was originally identified as a sequence that encompasses the transcription start site that is sufficient to direct accurate initiation in the absence of a TATA element (37–39).

Inr elements are, however, present in both TATA-containing and TATA-deficient (TATA-less) promoters. In mammalian promoters, the Inr consensus sequence is Py-Py-A₊₁-N-T/A-Py-Py (where A₊₁ is the transcription start site) (3, 20, 39), whereas in *Drosophila* promoters, the Inr consensus is T-C-A₊₁-G/T-T-T/C (1, 18, 32). It has been found that TAF_{II}150 and TAF_{II}250 play a role in the binding of TFIID to Inr elements (8, 16, 21, 42, 44).

The DPE functions cooperatively with the Inr to bind to TFIID and to direct accurate and efficient initiation of transcription in TATA-less promoters (4, 5). Thus far, the DPE has been identified in three *Drosophila* TATA-less promoters and in the TATA-less human IRF-1 promoter. In these promoters, the DPE is located about 30 nucleotides downstream of the transcription start site and appears to include a common G-A/T-C-G sequence motif. Interestingly, the addition of a DPE motif at a downstream position can compensate for the loss of transcription that occurs upon mutation of an upstream TATA box (4). In addition, photoaffinity cross-linking experiments suggested that dTAF_{II}60 and dTAF_{II}40 interact with the DPE (5). Thus, the DPE is functionally analogous to the TATA box, because both elements are recognition sites for the binding of TFIID and are functionally interchangeable for basal transcription activity. The range of sequences that can function as a DPE is not yet known. Hence, in this work, we have investigated the sequences that can function as a DPE as well as the range of promoters that use the DPE as a core promoter element. These studies have revealed, somewhat surprisingly, that the DPE sequence motif is as common as the TATA box in *Drosophila* core promoters.

MATERIALS AND METHODS

DNA templates. Minimal core promoter sequences were inserted in the same orientation into the *Xba*I and *Pst*I sites in the polylinker of pUC119. In these constructions, the *Xba*I site is upstream of the promoter, and the *Pst*I site is

* Corresponding author. Mailing address: Department of Biology, 0347, Pacific Hall, Room 2212B, University of California, San Diego, 9500 Gilman Dr., La Jolla, CA 92093-0347. Phone: (858) 534-4608. Fax: (858) 534-0555. E-mail: jkadonaga@ucsd.edu.

A

Promoter	Transcription		Sequence	
	Wt	Mut	Inr	DPE
297			TT TTAGTC	TTAAGCTGAGATCCAAAGAATAAA GTCG TGAAACTA
<i>brown</i>			TGC CCAGTT	GCCGCTCAGTCTTCCACGCGAACA GTCG ACGGCGTG
<i>caudal</i>			GTT TCAGTA	CGTGTTCGACCTGCATACTGAGTT GACG TCGCCACG
<i>Doc</i>			G ACATT	GGCATTCCACAGTCTTCGGGTGGA GACG TGTTTCTT
<i>E74A</i>			CGT TTAGTT	GTCTTTTGACTGCTGTAACGGACA GTCG CAAATTTT
<i>E74B</i>			ATT TTAGTT	TTGCTGATACCGTAAGAGATAAAT GACG TGCCCGCG
<i>engrailed</i>			CAA CTAATT	CAGTCGTTGCGCTCGATGTGAACA GACG TCGTGTTC
<i>G</i>			ACAGTC	GCGATCGAACACTCAACGAGTGCA GACG TGCCTACG
<i>glass</i>			CAA TTAGTC	GCCGTCCTTATGCTCCTCGCCAAAA GTCG CTTCTTGC
<i>I</i>			CATTA	CCACTTCAACCTCCGAAGAGATAA GTCG TGCCTCTC
<i>singed</i>			GTT TCATT	CCACTGGAGTGCACTTCGTGAGCG GTCG TTCTCTCC

B

Promoter	Transcription			Sequence	
	Wt	Mut1	Mut2	Inr	DPE
<i>labial</i>				TAA TCAGTC	ACGACTTGGTAAGCGCGCAGGCAG CACG TCGTCGTC Mut1 T GAT Mut2 ACAT
	(100)	23	5		
<i>Stellate</i>				AAG TCATAT	TTCTGTGATCAAGTGAAGTGGCAA CATG TCGAGCTC Mut1 T GAT Mut2 AGGT
	(100)	17	4		
<i>white</i>				CCT TCAGTT	GCACTTTGTGACGGGTTTCGTGAC GAAG CTCCAAGC Mut1 TCA T Mut2 TCAT
	(100)	31	5		
<i>E75A</i>				CTT TCATTG	ACTAACTGCCACTCGCAGCGCGCA GATC GTCGGCTC Mut1 TGAT Mut2 CGA T
	(100)	42	2		

FIG. 1. The distance between the Inr and DPE is strictly maintained in a variety of naturally occurring *Drosophila* core promoters. (A) In vitro transcription analysis of DPE-containing core promoters. A series of minimal core promoters were constructed with the DNA sequences indicated in the figure. Wild-type (Wt) and DPE mutant (Mut) versions of these promoter constructions were subjected to in vitro transcription and primer extension analysis. The sequences of the mutant promoter constructions as well as the quantitation of the data are given in Table 1. (B) The positioning of DPE-like sequences relative to the Inr is important for DPE function. In Mut1 promoters, DPE-like sequences with improper spacing relative to the Inr are mutated, whereas in Mut2 promoters, DPE sequences with the proper spacing relative to the Inr are mutated. The promoters were subjected to in vitro transcription and primer extension analysis, and the transcriptional activity of each mutant promoter relative to the corresponding wild-type promoter is indicated.

downstream of the promoter. The minimal promoter templates used in experiments shown in Fig. 1 include exactly the sequences shown in Fig. 1, and the sequence changes for the mutant promoters used in Fig. 1A are shown in Table 1. The upstream sequences in the pUC119 plasmid vector are 5'-AGTGAATT CGAGCTCGGTACCCGGGGATCCTCTAGA-3', where the TCTAGA sequences immediately upstream of the core promoter correspond to the *Xba*I cloning site. The minimal core promoter templates for *EF-1α F1* and *Sodh-1* and their corresponding mutant templates include sequences from -40 to +40 relative to the transcription start site. For these promoters, the sequences from -5 to +40 are shown in Fig. 5, and the remaining upstream sequences can be viewed in the *Drosophila* core promoter database website (<http://www-biology.ucsd.edu/labs/Kadonaga/DCPD.html>). The *G* promoter templates with altered spacing were as follows: *G*-3 (deletion of +19 to +21), *G*-2 (deletion of +19 and +20),

and *G*-1 (deletion of +19), *G*+1 (insertion of C between +19 and +20), *G*+2 (insertion of TC between +19 and +20), and *G*+3 (insertion of ATC between +19 and +20). The promoter sequences were described as follows: 297 (19), *brown* (11), *caudal* (26), *Doc* (9), *E74A* (7, 41), *E74B* (7, 41), *E75A* (35), *EF-1α F1* (17), *engrailed* (40), *G* (10), *glass* (27), *I* (13), *labial* (25), *singed* (30), *Sodh-1* (24), *Stellate* (23), and *white* (33).

In vitro transcription analysis. All transcription reactions were performed as previously described (45) with 200 ng of DNA supercoiled plasmid template and 5 μl (approximately 100 μg of protein) of *Drosophila* SK nuclear extract (40) in a 25-μl reaction mixture. Transcription products were detected by primer extension analysis as previously described (14). Reverse transcription products were quantified with a PhosphorImager (Molecular Dynamics). The quantitative results of the in vitro transcription data presented in Fig. 1, 2, 5, and 6 as well as

TABLE 1. Wild-type and mutant DPE-containing promoters used in this study

Promoter	Promoter sequence		Transcriptional activity of mutant promoter (% of wild type)
	Wild type	Mutant	
297	AGTCGTG	CTGATGT	<7
<i>brown</i>	AGTCGAC	ATGATAC	4
<i>caudal</i>	TGACGTC	TTCATTC	2
<i>Doc</i>	AGACGTG	CTCATGT	2
<i>E74A</i>	AGTCGCA	ATGATCA	7
<i>E74B</i>	TGACGTG	TTCATTG	<2
<i>engrailed</i>	AGACGTG	CTCATGT	2
<i>G</i>	AGACGTG	CTCATGT	6
<i>glass</i>	AGTCGCT	ATGATCT	<5
<i>I</i>	AGTCGTG	CTGATGT	3
<i>singed</i>	GGTCGTT	GTGATTT	7

in Tables 1 and 2 are derived from at least three (but typically, four or more) independent experiments. In Table 2 and Fig. 2 and 5, the standard deviations for each of the promoter activities are also reported.

Screening of the randomized promoter libraries. Partially overlapping oligonucleotides that included the *G* core promoter and flanking *Xba*I and *Pst*I sites for cloning were annealed, extended with *Escherichia coli* DNA polymerase I (Klenow) and deoxynucleoside triphosphates (dNTPs), and digested with *Xba*I and *Pst*I. The resulting DNA fragments were gel purified and ligated to *Xba*I- and *Pst*I-digested pUC119 plasmid. The oligonucleotide with the same sense as the mRNA included the *Xba*I site and *G* promoter sequences from -2 to +18. The oligonucleotide with the opposite sense from the mRNA included the *Pst*I site and *G* promoter sequences from +4 to +40. Randomized stretches of sequence were introduced by synthesizing oligonucleotides (with the opposite sense from the mRNA) with equal proportions of the four nucleotides at the positions indicated in Fig. 3A. *E. coli* was transformed with the randomized promoter libraries, and plasmid DNA was prepared from individual clones by using the Qiagen Plasmid Mini kit (catalog no. 12125) according to the suggested protocol of the manufacturer. In addition, each of the DNA samples was further purified as follows: the DNA precipitate was dissolved in Tris-EDTA (TE), extracted with phenol-chloroform-isoamyl alcohol (25:24:1), precipitated with

ethanol, and redissolved in TE. DNA concentrations were determined by UV spectrophotometry (and confirmed by agarose gel electrophoresis and staining with ethidium bromide) and then adjusted to 100 ng/ μ l. Each template was used in duplicate in vitro transcription reactions that were carried out in parallel with duplicate control transcription reactions with the wild-type *G* promoter template. DNA plasmid templates for the transcription experiments reported in Table 2 were purified by two successive CsCl equilibrium density gradients.

Construction of the *Drosophila* core promoter database. A set of 205 *Drosophila* core promoters was obtained by searching literature resources for genes with accurately mapped transcription start sites. To be included in the core promoter database, it was necessary for the transcription start site to be mapped by nuclease protection, primer extension, or multiple 5' rapid amplification of cDNA ends (RACE) clones. In cases where the reported start site overlaps a consensus Inr element, the central A nucleotide in the Inr consensus (T-C-A₄₋₁-G/T-T-C/T) was designated as the transcription start site. TATA elements were identified by visual inspection of the region upstream of -20 relative to the transcription start site for sequences conforming to the consensus T-A-T-A-A at five out of six positions. DPEs were identified by visual inspection of the positions +28 to +33 relative to the transcription start site to identify sequences matching the functional range set A/G/T-C/G-A/T-C/T-A/C/G-C/T at five out of six positions. The *Drosophila* core promoter database can be viewed at the website <http://www-biology.ucsd.edu/labs/Kadonaga/DCPD.html>.

DNase I footprint analysis. DNase I footprint probes were prepared by PCR amplification of each promoter with an unlabelled M13 universal (upstream) primer and a 5'-³²P-end-labeled M13 reverse (downstream) primer. Footprinting reactions and TFIIID purification were performed as described previously (4).

RESULTS

The Inr to DPE spacing is strictly maintained in a variety of *Drosophila* promoters. To date, only four TATA-less core promoters (*Drosophila jockey*, *Drosophila Antennapedia P2*, *Drosophila Abdominal-B*, and human IRF-1) have been found to require a DPE motif, as determined by mutational analysis of the DPE in conjunction with an in vitro transcription assay for core promoter activity (4, 5). A common feature of these DPE-containing promoters is a G-A/T-C-G motif in the +30 region. To identify DPE motifs in other TATA-less promoters, we constructed and analyzed a set of wild-type and mutant versions of 15 *Drosophila* TATA-less promoters that contain a

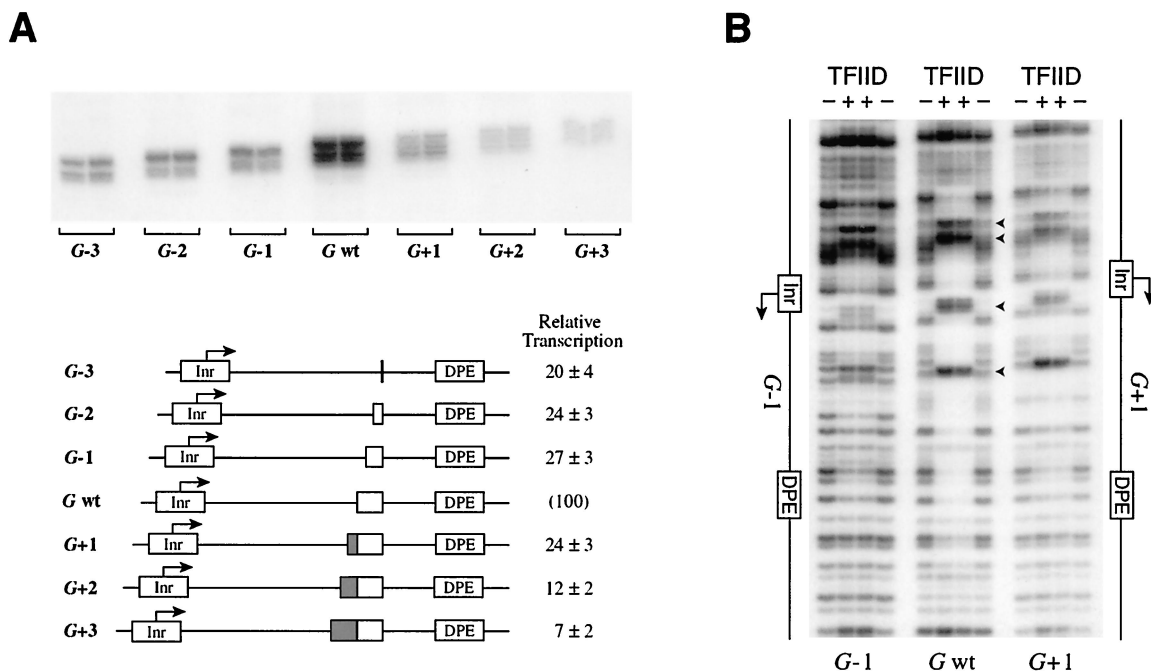


FIG. 2. A single nucleotide alteration in the spacing between the DPE and Inr reduces core promoter activity and binding of purified TFIIID. (A) In vitro transcription and primer extension analysis of a series of mutant *G* core promoters that contain 1-, 2-, or 3-nucleotide insertions or deletions between the DPE and Inr. wt, wild type. (B) DNase I footprint analysis of *G*-1, *G* wild-type, and *G*+1 core promoters with purified *Drosophila* TFIIID. Arrows indicate DNase I hypersensitive sites.

TABLE 2. Determination of a DPE functional range set^a

Promoter	DPE sequence	Relative transcription ^b
Wild-type <i>G</i>	+28 A G A C G T +33	100
Mutants		
<i>G</i> 11-1	G C A T G G	86 ± 20
<i>G</i> 11-2	T G A T C C	81 ± 20
<i>G</i> 6-1	T C A C A C	79 ± 7
<i>G</i> 6-2	G C A C C T	74 ± 27
<i>G</i> 6-3	A G T T G T	70 ± 11
<i>G</i> 6-4	T C A T G T	68 ± 7
<i>G</i> 6-5	A G A T C T	63 ± 10
<i>G</i> 6-6	A C G C A C	54 ± 7
<i>G</i> 6-7	A G A G A C	54 ± 9
<i>G</i> 6-8	A G T T G A	53 ± 4
<i>G</i> 6-9	A A C T G C	52 ± 2
<i>G</i> 6-10	G G A T G C	51 ± 6
<i>G</i> 6-11	C C A T G T	51 ± 12
DPE functional range set ^c	A C A C A C G C A T G T T G T T G T	

^a The DPE functional range set represents sequences that contribute to or are compatible with DPE function.

^b Mean ± standard deviation from four independent experiments.

^c The T nucleotide at positions +30 is included on the basis of the presence of T at this position in the DPE-containing promoters tested in Fig. 1.

G-A/T-C-G motif in the +30 region. In these experiments, 11 out of the 15 promoters exhibited a strong dependence upon the downstream G-A/T-C-G motif (13- to 60-fold reduction in transcriptional activity upon mutation) (Fig. 1A and Table 1). In contrast, the other four promoters, *labial*, *Stellate*, *white*, and *E75A*, displayed only a modest reduction (about 2.5- to 6-fold) in transcriptional activity upon mutation of their downstream G-A/T-C-G motifs (Mut1 series) (Fig. 1B).

Interestingly, the spacing between the Inr and the DPE in the 11 mutation-sensitive promoters (Fig. 1A) is identical to that of previously characterized DPE-containing promoters (*jockey*, *Antennapedia P2*, *Abdominal-B*, and IRF-1 core promoters), with the G-A/T-C-G motif positioned exactly from +29 to +32 downstream of the central A₊₁ nucleotide in the Inr. On the other hand, the *labial*, *Stellate*, *white*, and *E75A* promoters possess downstream G-A/T-C-G sequences, but not precisely at the +29 to +32 position. We therefore examined the +29 to +32 region of the *labial*, *Stellate*, and *white* promoters (and +30 to +33 in the *E75A* promoter) and found that mutation of these nucleotides (Mut2 series) significantly reduced core promoter activity (Fig. 1B). These findings indicate that the precise spacing between the Inr and DPE motifs is of critical importance for core promoter activity. In addition, the observation of +29 to +32 sequences other than G-A/T-C-G acting as DPE motifs (as in Fig. 1B) suggested that the range of sequences that can function as a DPE extends beyond the G-A/T-C-G motif that was initially found in DPE-driven core promoters.

A single nucleotide alteration in the spacing between the DPE and Inr reduces core promoter activity and binding of purified TFIID. To investigate further the importance of spacing between the DPE and Inr motifs, we constructed a series of mutant versions of the *G* promoter (derived from the *G* long interspersed nuclear element [LINE]) insertions or deletions in single nucleotide increments. In vitro transcription analysis of these templates revealed an approximately fourfold reduc-

tion of transcriptional activity as a result of a single nucleotide deletion or insertion (Fig. 2A). In addition, TFIID binding to wild-type *G*, *G*-1, and *G*+1 promoters was analyzed by DNase I footprinting (Fig. 2B). With the wild-type *G* promoter, TFIID protected the core promoter region from about -20 to +40 with DNase I hypersensitive sites at positions -11, -8, +4, and +15. With the *G*-1 and *G*+1 mutant promoters, the TFIID footprint was distinctly weaker than that seen with the wild-type promoter. These results indicate that the positioning of the DPE in the *G* promoter (with the G-A-C-G motif at precisely +29 to +32) is optimal for binding of TFIID and core promoter activity. Moreover, these findings are consistent with the strict maintenance of the +29 to +32 positioning of the DPE in naturally occurring core promoters (Fig. 1).

Determination of the range of sequences that can function as a DPE. Because the studies of the *labial*, *Stellate*, *white*, and *E75A* core promoters revealed DPE function by sequences at +29 to +32 that did not completely conform to G-A/T-C-G (Fig. 1B), we sought to explore the range of nucleotides that could function as a DPE. To this end, we performed a bio-

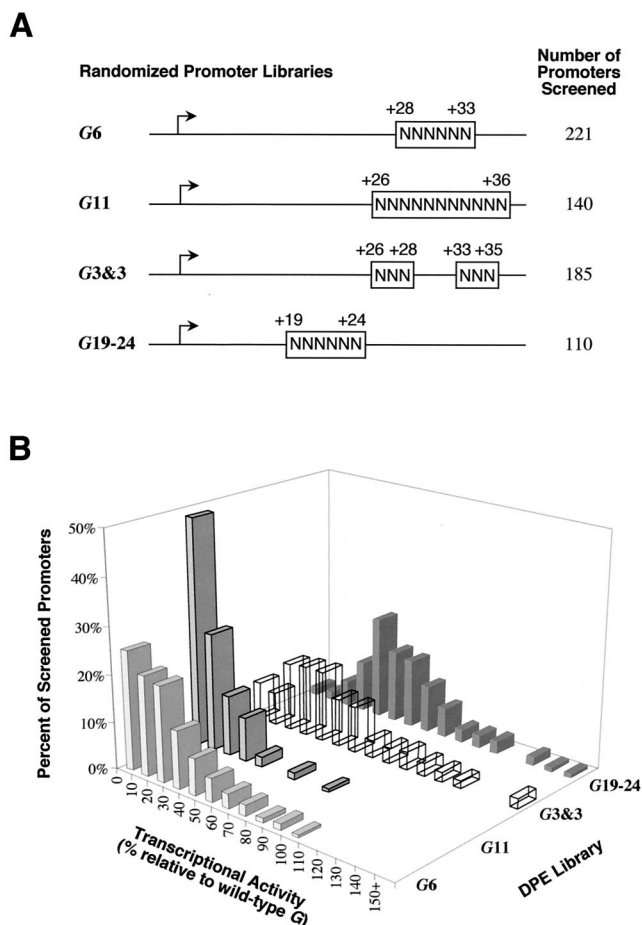


FIG. 3. Analysis of the range of sequences that can function as DPE motifs. (A) Diagram of randomized *G* core promoter libraries. Four promoter libraries were constructed with *G* core promoter sequences (-2 to +40 relative to the transcription start site), except that the portions of the sequence indicated by N's contained approximately equivalent amounts of each of the four deoxyribonucleotides. (B) Summary of the in vitro transcription screening of the randomized *G* core promoter libraries. Individual clones from each of the randomized libraries were isolated and then subjected to in vitro transcription analysis. The graph shows the distribution of transcriptional activity for each of the tested promoters relative to the wild-type *G* core promoter (100%) for each library.

chemical screen to identify sequences that possess the transcriptional activity of the DPE. First, we constructed libraries of the *G* promoter that contained random sequences instead of the wild-type sequence at different positions downstream of the Inr (Fig. 3A). Then, for each library, individual clones were subjected to *in vitro* transcription analysis, and the DNA sequences of the most active promoters were determined. The DNA sequencing additionally confirmed that the constant (i.e., not randomized) regions of the promoters remained identical to those of the wild type during the subcloning and DNA preparation procedures.

We initially screened promoters from the *G11* library, which contains a stretch of 11 random nucleotides from +26 to +36. As seen in Fig. 3B, most of the *G11*-derived core promoters exhibited low transcriptional activity. Nearly half of the 140 *G11* promoters possessed less than 10% of the activity of the wild-type promoter. These results indicate that random sequences at the location of the DPE generally do not exhibit DPE activity. The *G11* analysis led to the identification of two promoters with activity that is >50% of that of the wild-type *G* promoter.

Because the frequency of strong promoters in the *G11* library was low, we prepared libraries with shorter regions of randomized sequence. First, to focus on the core DPE sequences, we constructed the *G6* library (random nucleotides from +28 to +33) and screened 221 promoters. Then, to focus on the flanking sequences, we generated the *G3&3* library (random nucleotides from +26 to +28 and +33 to +35, with the central G-A-C-G motif intact) and screened 185 promoters. In addition, to assess the effects of sequences between the Inr and DPE, we constructed the *G19–24* library (random nucleotides from +19 to +24) and screened 110 promoters. These randomized promoter libraries are depicted in Fig. 3A.

The results of the screening of the promoter libraries are summarized in Fig. 3B. As mentioned above, the *G11* library yielded mainly weak promoters (median promoter activity = 11% of wild type). The *G6* library generally consisted of stronger promoters (median activity = 22% of wild type) than the *G11* library. The promoters from the *G3&3* library (median activity = 44% of wild type) were significantly stronger than those from the *G6* library. These results are consistent with a greater importance of the core DPE sequences relative to the flanking sequences. The analysis of the *G19–24* library (median activity = 55% of wild type) revealed a minor yet distinct contribution from sequences between the Inr and DPE to promoter strength.

The *G11* and *G6* promoters that exhibited >50% of the activity of the wild-type *G* promoter in the initial screening were then analyzed in greater detail, and the results are shown in Table 2. Notably, none of the promoters isolated from any of the libraries were stronger than the wild-type *G* promoter, which appears to be well optimized for transcriptional activity. Based on the sequences of the most active promoters obtained in the screening of the randomized libraries, a DPE functional range set was derived from the nucleotides that predominate at each position, with a bias for nucleotides that are found in the strongest promoters in the hierarchy. This functional range set represents sequences that appear to contribute to DPE-mediated transcription or to be compatible with DPE-mediated transcription. Interestingly, as seen previously in a similar analysis of the TATA box (36), a moderately broad range of sequences can function as a DPE motif.

We similarly analyzed the most active promoter constructions obtained in the screening of the *G3&3* library, in which the sequences flanking the core DPE motif were randomized. These studies yielded nine promoters with >85% activity rel-

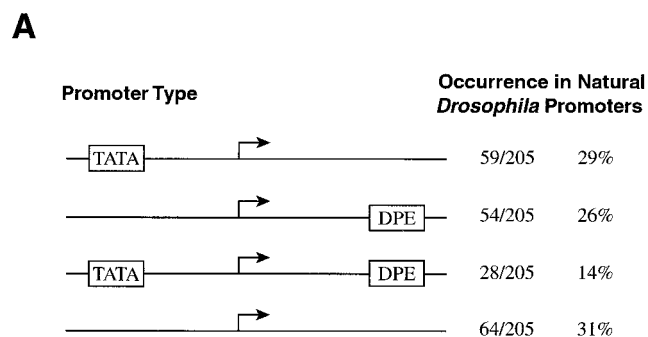


FIG. 4. The DPE appears to be present in many *Drosophila* promoters. (A) The frequency of occurrence of the DPE appears to be comparable to that of the TATA box in *Drosophila* core promoters. A *Drosophila* core promoter database was created by aligning sequences of 205 *Drosophila* core promoters with accurately determined transcription start sites. The number of promoters that appear to possess a TATA box only, a DPE only, both elements, or neither element is shown. TATA boxes were defined as sequences with at least a 5 out of 6 match with the TATAAA sequence upstream of -20 relative to the transcription start site. DPE motifs were defined as sequences with at least a 5 out of 6 match with the DPE functional range set (Table 2) at exactly +28 to +33 relative to the start site. The *Drosophila* core promoter database is available at the website <http://www-biology.ucsd.edu/labs/Kadonaga/DCPD.html>. (B) Nucleotide distributions in the upstream region of *Drosophila* core promoters. The nucleotide distributions at positions -47 to -3 relative to the transcriptional start site (+1) were analyzed for 59 TATA-only promoters, 54 DPE-only promoters, 28 TATA + DPE promoters, and 64 TATA-less and DPE-less promoters. A χ^2 test of the null hypothesis that each nucleotide is equally distributed was performed for every position. Letters over a bracket above the bars of the graph indicate the overrepresented nucleotides at positions that significantly deviate from the null hypothesis ($P < 0.001$). (C) Nucleotide distributions in the downstream region of *Drosophila* core promoters. The downstream region (from -2 to +45 relative to the start site) of *Drosophila* core promoters was analyzed as in panel B. The Inr and DPE motifs are indicated with a bracket below the graphs.

ative to the wild-type promoter. Analysis of the sequences of these promoters did not, however, reveal any notable sequence bias, except perhaps for a pyrimidine at +26 (data not shown).

Construction and analysis of the *Drosophila* core promoter database. With the DPE functional range set, we next sought to identify potential DPE-containing promoters from a database of *Drosophila* core promoters. Because of the strict spacing requirement between the Inr and DPE motifs (Fig. 1 and 2), a high degree of accuracy in the mapping of the transcription start sites was needed for the core promoters in the database. We therefore surveyed the primary literature for *Drosophila* core promoters in which the transcription start sites were mapped by nuclease protection, primer extension, or multiple 5' RACE clones. The *Drosophila* promoter database of Arkhipova (1) was a particularly useful source of literature citations. These studies yielded 205 *Drosophila* core promoters, with which we generated a *Drosophila* core promoter database (<http://www-biology.ucsd.edu/labs/Kadonaga/DCPD.html>). We then searched the database for promoters containing putative DPE and/or TATA motifs. This analysis revealed that the frequency of occurrence of putative DPE motifs (40%) is comparable to that of putative TATA box elements (43%) (Fig. 4A). Hence, in *Drosophila*, the DPE might be used as a core promoter element nearly as often as the TATA box.

Based on the presence or absence of putative TATA box and DPE motifs, we categorized the core promoters into four classes: 1, TATA only; 2, DPE only; 3, TATA plus DPE; and 4, TATA and DPE less (Fig. 4A). To gain better insight into the characteristics of these different types of core promoters, we examined the nucleotide distribution at each position (from -47 to +45 relative to the start site at +1) for promoters in each category. In the region upstream of the transcription start

B

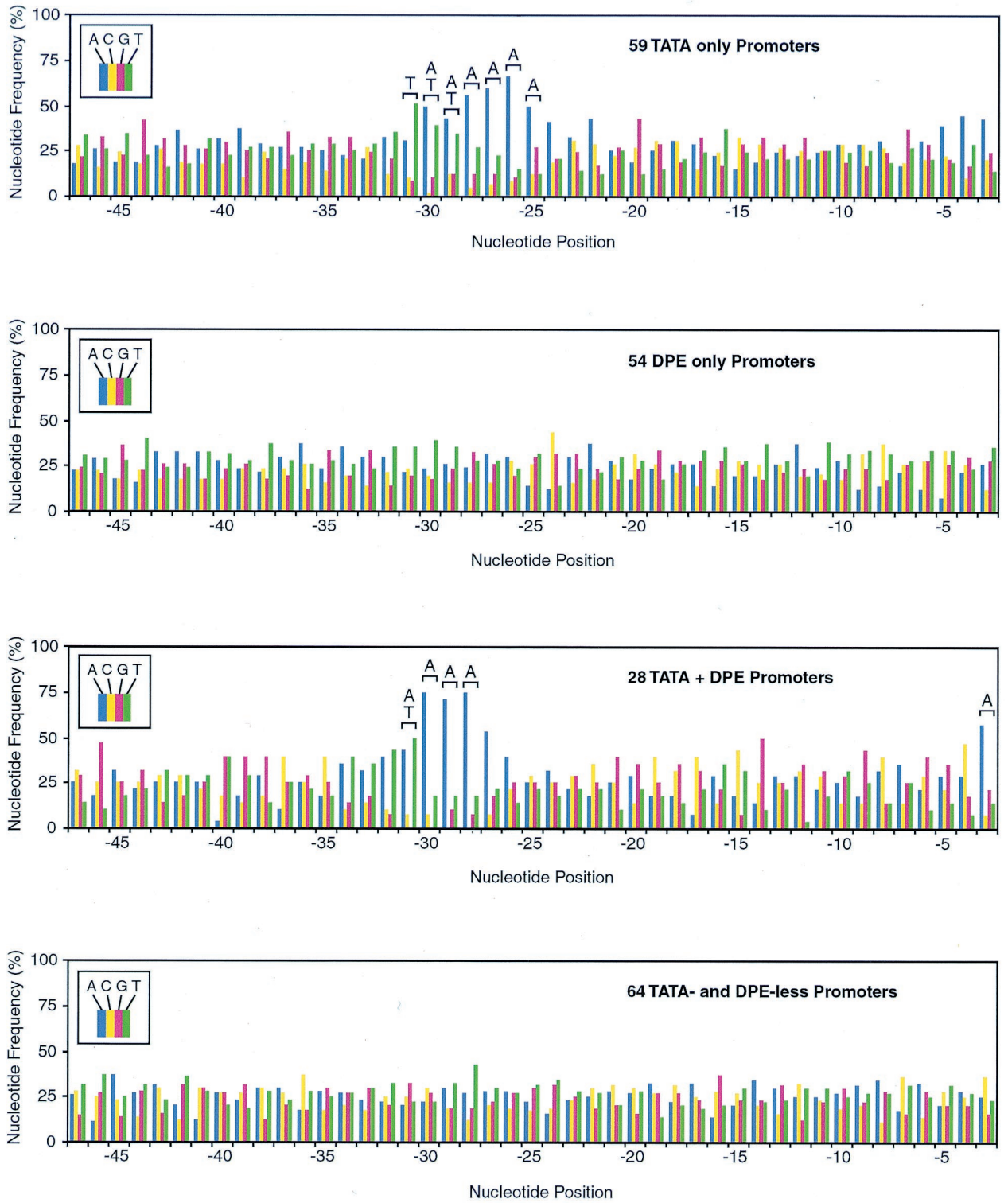


FIG. 4—Continued.

C

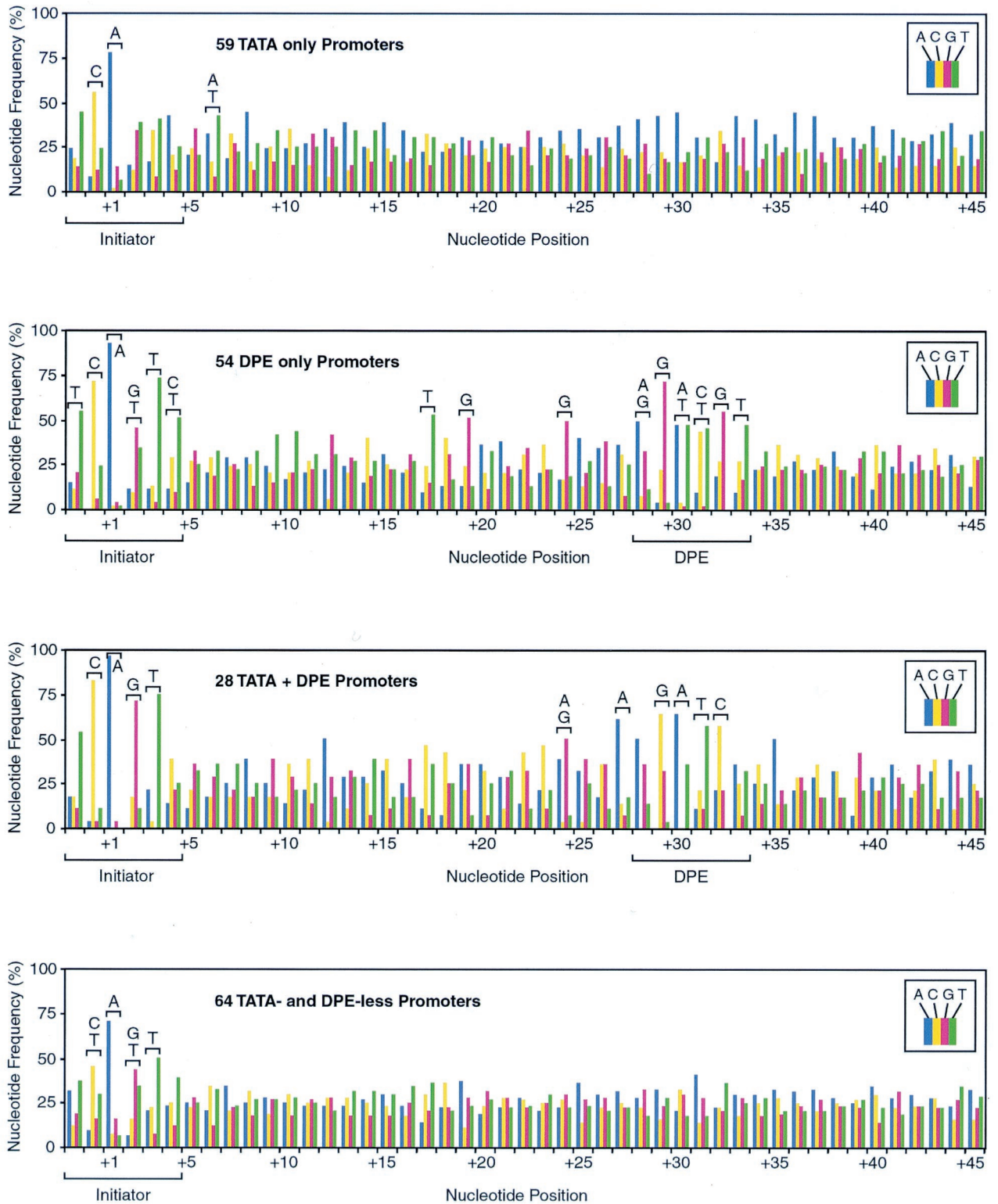


FIG. 4—Continued.

site, we observed an A/T-rich region from -31 to -25 in the TATA-only promoters as well as an A/T-rich region from -31 to -28 of the TATA plus DPE promoters (Fig. 4B). There was also an overrepresentation of A at -3 in the TATA plus DPE promoters (Fig. 4B). No upstream sequence bias was seen in either the DPE-only promoters or the TATA- and DPE-less promoters.

The statistical analysis of sequences from -2 to +45 is shown in Fig. 4C. There is a general bias for the Inr consensus, T-C-A₊₁-G/T-T-C/T, which is seen most distinctly with the DPE-only promoters. It should be noted, however, that the Inr consensus was sometimes used in the alignment of sequences in the construction of the database (see Materials and Methods), and, thus, some bias for the Inr consensus is expected. The DPE-only promoters were categorized on the basis of their conformity to the DPE functional range set, and thus, there is sequence bias in the +28 to +33 region of the DPE-only promoters. Unexpectedly, however, the nucleotide bias ($P < 0.001$) from +28 to +33 in the DPE-only promoters, A/G-G-A/T-C/T-G-T, represents only a subset of the DPE functional range set (A/G/T-C/G-A/T-C/T-A/C/G-C/T) that was used in the classification. Thus, we view the restricted set of overrepresented nucleotides to be a consensus of the DPE. Interestingly, in the DPE-only promoters, additional overrepresented nucleotides ($P < 0.001$) were observed at +17 (T), +19 (G), and +24 (G), which are in a region between the Inr and DPE motifs that was not used in the promoter classification. In addition, the TATA + DPE promoters exhibited a sequence bias ($P < 0.001$) at +24 (A/G), +27 (A), and from +29 to +32 (G-A-T-C). Lastly, with the TATA- and DPE-less promoters, we did not observe any sequence bias that might have been suggestive of other novel core promoter motifs.

The DPE functional range set identifies new DPE-containing promoters. The use of the DPE functional range set along with the *Drosophila* core promoter database led to the identification of novel, putative DPE-containing promoters (Fig. 4). We were interested, in particular, in testing whether core promoters containing sequences that conformed to the DPE functional range set, but not to the previous DPE consensus (i.e., G-A/T-C-G from +29 to +32) did indeed possess functionally important DPE motifs. To this end, we constructed and analyzed wild-type and mutant versions of the *Drosophila EF-1 α F1* and *Sodh-1* promoters (Fig. 5A). These experiments revealed that both promoters were strongly dependent upon their respective DPE motifs for transcriptional activity.

We further investigated the *EF-1 α F1* promoter because its DPE appears to differ most significantly from that of the previous consensus. First, to identify the sequences in the downstream region of the promoter that are most important for transcriptional activity, we constructed a series of mutant *EF-1 α F1* templates with triple clustered nucleotide substitutions that span from +22 to +38 (Fig. 5B). The results indicated that the sequences from +28 to +34 were the most sensitive to mutation, which is consistent with the *EF-1 α F1* downstream element functioning as a DPE. We also tested the binding of TFIID to the *EF-1 α F1* promoter. As seen in Fig. 5C, purified *Drosophila* TFIID binds to the wild type, but not to the mutant *EF-1 α F1* promoter. Notably, with the wild-type promoter, there are strong DNase I hypersensitive sites at -8 and +4 in addition to DNase I protection from about -20 to +30. These results thus indicate that the downstream core promoter sequence in the *EF-1 α F1* gene is a DPE. More generally, these experiments suggest that the DPE functional range set can be useful in the identification of new DPE-containing promoters.

The +24 position has a role in DPE promoter function. As seen in Fig. 4C, the statistical analysis of the putative DPE-containing promoters (DPE-only promoters) from the *Drosophila* core promoter database revealed sequence biases at positions +17 (T), +19 (G), and +24 (G). Moreover, we observed that there was a distinct overrepresentation of G nucleotides at +24 in experimentally confirmed DPE-containing promoters (e.g., in Fig. 1, 12 out of 15 promoters possess a G nucleotide at +24, whereas 6 out of 15 promoters have a T+17 and 7 out of 15 have a G+19). In addition, we sequenced the most active promoters (top 20%) in the G19-24 library (Fig. 3) and found that half of those promoters (11 out of 22 tested) have a G nucleotide at +24. Hence, because of the strong correlation between G+24 and DPE function, we tested the importance of a G nucleotide at +24 by mutational analysis. To this end, we constructed five core promoter templates with a mutation at +24 (Fig. 6). With the *caudal* and *I* promoters, the wild-type G+24 was mutated to a T, whereas with the 297, *E74B*, and *glass* promoters, the respective A, T, and C nucleotides at +24 in the wild-type promoters were converted to a G. These experiments revealed that the mutation of G+24 to T+24 caused about a 2- to 2.5-fold reduction in transcriptional activity, whereas the conversion of A, T, or C to a G at +24 resulted in a 2- to 4-fold increase in activity. These results suggest that a G nucleotide at +24 makes a modest yet distinct contribution to transcription from DPE-driven core promoters.

DISCUSSION

In this work, we have presented a detailed analysis of the DNA sequences that govern the function of DPE-containing core promoters. We found that the DPE is subject to strict spacing requirements. All 20 experimentally confirmed DPE motifs are located at +28 to +33 relative to the transcription start site (Fig. 1 and 5) (4, 5), and the insertion or deletion of a single nucleotide between the Inr and DPE reduces transcriptional activity and TFIID binding (Fig. 2). By in vitro transcription analysis of randomized promoter libraries, we determined the DPE functional range set, which represents sequences that contribute to or are compatible with DPE function (Fig. 3 and Table 2), and found that it can be used to identify novel DPE-containing promoters (Fig. 4 and 5). In addition, we compiled a *Drosophila* core promoter database (available at <http://www-biology.ucsd.edu/labs/Kadonaga/DCPD.html>) with which a statistical analysis of core promoter elements was performed. These studies revealed that the DPE motif appears to be approximately as common as the TATA box in *Drosophila* core promoters (Fig. 4). There is, in addition, a striking adherence of Inr sequences to the Inr consensus in DPE-containing promoters relative to DPE-less promoters (Fig. 4C). This observation is consistent with the cooperative function of the DPE and Inr motifs for TFIID binding and basal transcriptional activity (4). Furthermore, statistical and biochemical analyses indicated that a G nucleotide at +24 has a modest yet distinct role in transcription from DPE-containing promoters (Fig. 6). Thus, these experiments reveal that key features of DPE-driven core promoters are a precise spacing between the Inr and DPE, a strict adherence to the Inr consensus, a minor yet distinct contribution by G+24, and some flexibility in the sequence of the DPE.

A model for the binding of TFIID to TATA- versus DPE-containing promoters. There appear to be significant differences in the interactions of TFIID with TATA-containing and DPE-containing promoters. In Fig. 7, we present a model of TFIID engaged in two distinct core promoter interactions. In the TATA-driven promoter, some flexibility between the

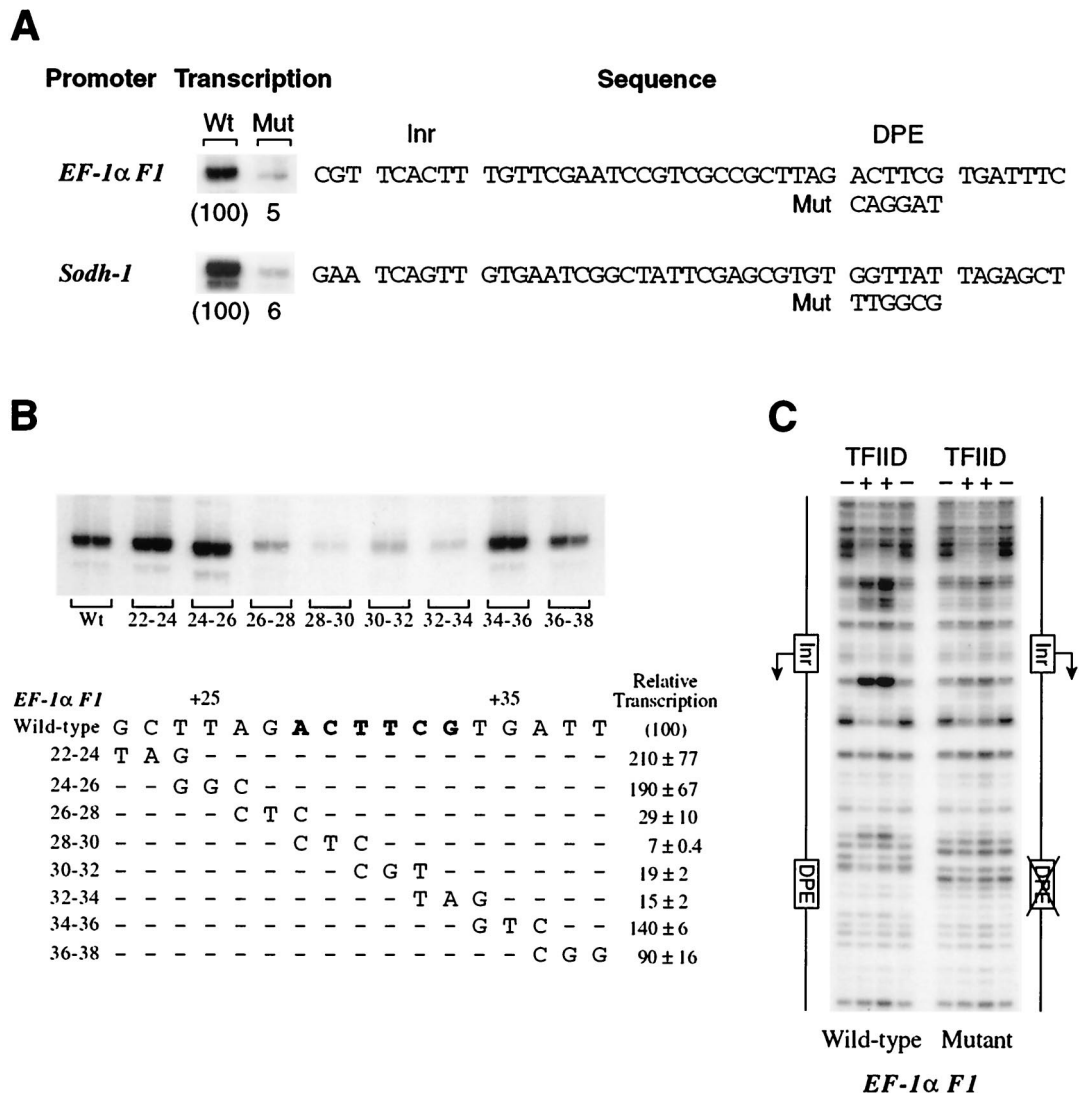


FIG. 5. The DPE functional range set identifies additional DPE-dependent promoters. (A) In vitro transcription analysis of the *EF-1 α F1* and *Sodh-1* core promoters. The sequences of the wild-type (Wt) and DPE mutant (Mut) versions of the promoters are indicated. (B) Scanning clustered point mutational analysis of the *EF-1 α F1* core promoter. A series of mutant core promoters with triple nucleotide substitutions, as indicated, were constructed and subjected to in vitro transcription and primer extension analysis. (C) DNase I footprint analysis of the *EF-1 α F1* promoter with purified *Drosophila* TFIID. The DPE mutant version of the *EF-1 α F1* promoter is identical to that used in panel A.

TATA and Inr motifs is depicted, as suggested by the variability in the distance between the TATA and Inr elements in naturally occurring promoters. In the DPE-driven promoter, the DNA is shown as following the surface of TFIID from the Inr to the DPE. This arrangement is suggested by the importance of the precise spacing between DPE and Inr (Fig. 1 and 2), the pattern of DNase I protection and hypersensitivity upon binding of purified TFIID (Fig. 2 and 5) (4), and the contribution of the G residue at +24 (Fig. 6). In addition, because we do not detect a footprint in the -20 to -35 region of TATA-less DPE-containing promoters, TBP is not depicted as bound to the DNA. It is possible, however, that there is low-affinity, non-sequence-specific binding of TBP to the upstream region that is not detectable by DNase I footprinting. Figure 7 also depicts a revised consensus for the DPE, which is based on the statistical analysis of putative DPE-containing promoters in the *Drosophila* core promoter database (Fig. 4) as well as the biochemical analysis of the +24 position (Fig. 6).

A variety of sequences can function as a DPE. The analysis of randomized promoters (Fig. 3), which yielded the DPE functional range set (A/G/T-C/G-A/T-C/T-A/C/G-C/T from +28 to +33; Table 2), revealed that a diverse collection of sequences can function as a DPE. However, when the DPE functional range set was used as the basis for the identification of putative DPE-containing promoters (Fig. 4), the distribution of nucleotides from +28 to +33 in the natural promoters (A/G-G-A/T-C/T-G-T; Fig. 4C) was only a subset of the functional range set. (It is relevant to note that only four out of the 54 DPE-only promoters in Fig. 4 are derived from LINES. Hence, LINES, which may have conserved downstream sequences other than the DPE, constitute only a minor fraction of the DPE-containing promoters in the database.) These findings are reminiscent of a similar analysis of the TATA box (36), in which it was observed that the variety of sequences that could function as TATA boxes was significantly greater than those typically used as TATA elements.

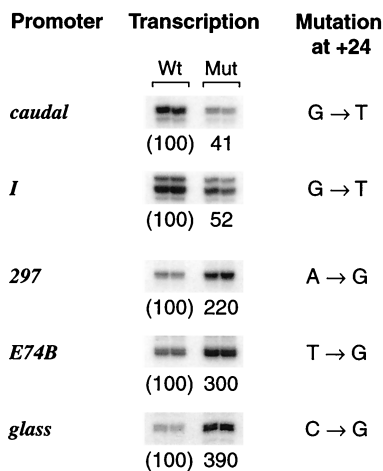


FIG. 6. The +24 position contributes to DPE promoter function. The wild-type (Wt) and +24 mutant (Mut) versions of the indicated promoters were subjected to in vitro transcription and primer extension analysis.

Why might the DPE (or TATA) consensus of natural promoters be more restricted than the range of sequences that are sufficient for transcriptional activity? It seems reasonable that a core promoter must not only perform the positive function of directing basal transcription, but it also must not contain any sequences that would have an adverse effect upon the regulation of its cognate gene. For example, some sequences might recruit undesired activators or repressors. Other sequences might interfere with the proper interactions between activators or coactivators with the basal transcriptional machinery. Thus, in this manner, the DPE consensus might reflect the need to direct basal transcription as well as to maintain the appropriate regulation of the cognate genes.

DPE motifs might be as commonly used as TATA boxes. In our analysis of the *Drosophila* core promoter database (which contains 205 core promoters), we found that approximately 40% of the promoters conformed to the DPE functional range set at five out of six positions (Fig. 4A). In comparison, about 43% of the promoters exhibited a five out of six match with the TATA consensus over a relatively broad range spanning from -47 to -19. It seems likely that many but not all of these putative DPE- or TATA-containing promoters do indeed pos-

sess functionally important DPE or TATA motifs. We also do not know how accurately the *Drosophila* core promoter database represents the distribution of TATA- versus DPE-containing promoters in the *Drosophila* genome. In spite of these uncertainties, it does appear that DPE motifs are commonly found in *Drosophila*, possibly at a frequency that is comparable to that of TATA boxes.

In addition, there are probably some DPE- and TATA-containing promoters that were not identified by the selection criteria. One such promoter is that of the *white* gene (Fig. 1B), which has only a four out of six match with the DPE functional range set. We therefore tested whether the *white* DPE is a strong DPE that does not conform to the functional range set or a weak DPE that is a poor match to the functional range set. To this end, we created a mutant version of the *white* core promoter that contains the strong DPE sequence from the *G* promoter (A-G-A-C-G-T) at +28 to +33 instead of the normal *white* DPE sequence (C-G-A-A-G-C). These experiments revealed that the mutant, DPE-optimized *white* promoter possessed six times the transcriptional activity of the wild-type *white* promoter (data not shown). Hence, the DPE in the *white* core promoter is a weak DPE that does not conform well to the functional range set.

Finally, it is interesting to note that approximately 31% of the promoters in the *Drosophila* core promoter database appear to contain neither a TATA box nor a DPE motif (Fig. 4A). Thus, there are potentially other core promoter elements to be discovered. The statistical analysis of the TATA- and DPE-less promoters did not reveal, however, any notable sequence bias. This result could be due to the set of TATA- and DPE-less promoters being a composite of different types of core promoters with different sequence biases. Alternatively, it is possible that the only core promoter motif in these promoters is the Inr element, which might act in conjunction with sequence-specific promoter binding activators to direct basal transcription, as observed with transcription factor Sp1 and the Inr (see, for example, references 12, 29, and 46).

ACKNOWLEDGMENTS

We are very grateful to Peter Geiduschek, Jessica Tyler, Jennifer Butler, Patricia Willy, Mark Levenstein, and Vassili Alexiadis for critical reading of the manuscript. We thank Scott Iyama for skillful assistance in the preparation of the *Drosophila* core promoter database, I. Arkhipova for providing computer text files of her *Drosophila* promoter database (1), and Jenny Butler for SK extracts.

A.K.K. was supported in part by a training grant from the National Institutes of Health (T32 GM07240). This work was supported by a grant from the National Institutes of Health (GM41249) to J.T.K.

REFERENCES

1. Arkhipova, I. R. 1995. Promoter elements in *Drosophila melanogaster* revealed by sequence analysis. *Genetics* **139**:1359-1369.
2. Berk, A. J. 1999. Activation of RNA polymerase II transcription. *Curr. Opin. Cell Biol.* **11**:330-335.
3. Bucher, P. 1990. Weight matrix descriptions of four eukaryotic RNA polymerase II promoter elements derived from 502 unrelated promoter sequences. *J. Mol. Biol.* **212**:563-578.
4. Burke, T. W., and J. T. Kadonaga. 1996. *Drosophila* TFIID binds to a conserved downstream basal promoter element that is present in many TATA-box-deficient promoters. *Genes Dev.* **10**:711-724.
5. Burke, T. W., and J. T. Kadonaga. 1997. The downstream core promoter element, DPE, is conserved from *Drosophila* to humans and is recognized by TAFII60 of *Drosophila*. *Genes Dev.* **11**:3020-3031.
6. Burley, S. K., and R. G. Roeder. 1996. Biochemistry and structural biology of transcription factor IID (TFIID). *Annu. Rev. Biochem.* **65**:769-799.
7. Burtis, K. C., C. S. Thummel, C. W. Jones, F. D. Karim, and D. S. Hogness. 1990. The *Drosophila* 74EF early puff contains *E74*, a complex ecdysone-inducible gene that encodes two ets-related proteins. *Cell* **61**:85-99.
8. Chalkley, G. E., and C. P. Verrijzer. 1999. DNA binding site selection by RNA polymerase II TAFs: a TAF_{II}250-TAF_{II}150 complex recognizes the initiator. *EMBO J.* **18**:4835-4845.

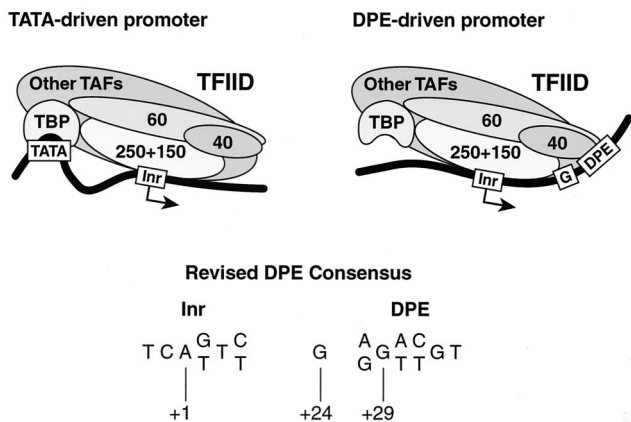


FIG. 7. A model of two distinct interactions of TFIID with TATA- versus DPE-driven core promoters. The model is discussed in the text. TAFs, TBP-associated factors.

9. **Contursi, C., G. Minchiotti, and P. P. Di Nocera.** 1995. Identification of sequences which regulate the expression of *Drosophila melanogaster* *Doc* elements. *J. Biol. Chem.* **270**:26570–26576.
10. **Di Nocera, P. P.** 1988. Close relationship between non-viral retroposons in *Drosophila melanogaster*. *Nucleic Acids Res.* **16**:4041–4052.
11. **Dreesen, T. D., D. H. Johnson, and S. Henikoff.** 1988. The brown protein of *Drosophila melanogaster* is similar to the white protein and to components of active transport complexes. *Mol. Cell. Biol.* **8**:5206–5215.
12. **Emami, K. H., W. W. Navarre, and S. T. Smale.** 1995. Core promoter specificities of the Sp1 and VP16 transcriptional activation domains. *Mol. Cell. Biol.* **15**:5906–5916.
13. **Fawcett, D. H., C. K. Lister, E. Kellett, and D. J. Finnegan.** 1986. Transposable elements controlling I-R hybrid dysgenesis in *D. melanogaster* are similar to mammalian LINES. *Cell* **47**:1007–1015.
14. **George, C. P., and J. T. Kadonaga.** 1996. Primer extension analysis of RNA, p. 133–139. *In* P. A. Krieg (ed.), *A laboratory guide to RNA: isolation, analysis, and synthesis*. J. Wiley & Sons, Inc., New York, N.Y.
15. **Hampsey, M.** 1998. Molecular genetics of the RNA polymerase II general transcriptional machinery. *Microbiol. Mol. Biol. Rev.* **62**:465–503.
16. **Hansen, S. K., and R. Tjian.** 1995. TAFs and TFIIA mediate differential utilization of the tandem *Adh* promoters. *Cell* **82**:565–575.
17. **Hovemann, B., S. Richter, U. Walldorf, and C. Cziepluch.** 1988. Two genes encode related cytoplasmic elongation factors 1 alpha (EF-1 alpha) in *Drosophila melanogaster* with continuous and stage specific expression. *Nucleic Acids Res.* **16**:3175–3194.
18. **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.
19. **Inouye, S., K. Hattori, S. Yuki, and K. Saigo.** 1986. Structural variations in the *Drosophila* retrotransposon, 17.6. *Nucleic Acids Res.* **14**:4765–4778.
20. **Javahery, R., A. Khachi, K. Lo, B. Zenzie-Gregory, and S. T. Smale.** 1994. DNA sequence requirements for transcriptional initiator activity in mammalian cells. *Mol. Cell. Biol.* **14**:116–127.
21. **Kaufmann, J., K. Ahrens, R. Koop, S. T. Smale, and R. Müller.** 1998. CTF150, a human cofactor for transcription factor IID-dependent initiator function. *Mol. Cell. Biol.* **18**:233–239.
22. **Lagrange, T., A. N. Kapanidis, H. Tang, D. Reinberg, and R. H. Ebright.** 1998. New core promoter element in RNA polymerase II-dependent transcription: sequence-specific DNA binding by transcription factor IIB. *Genes Dev.* **12**:34–44.
23. **Livak, K. J.** 1990. Detailed structure of the *Drosophila melanogaster* *Stellate* genes and their transcripts. *Genetics* **124**:303–316.
24. **Luque, T., L. Hjelmqvist, G. Marfany, O. Danielsson, M. El-Ahmad, B. Persson, H. Jornvall, and R. Gonzalez-Duarte.** 1998. Sorbitol dehydrogenase of *Drosophila*. Gene, protein, and expression data show a two-gene system. *J. Biol. Chem.* **273**:34293–34301.
25. **Mlodzik, M., A. Fjose, and W. J. Gehring.** 1988. Molecular structure and spatial expression of a homeobox gene from the labial region of the *Antennapedia*-complex. *EMBO J.* **7**:2569–2578.
26. **Mlodzik, M., and W. J. Gehring.** 1987. Expression of the *caudal* gene in the germ line of *Drosophila*: formation of an RNA and protein gradient during early embryogenesis. *Cell* **48**:465–478.
27. **Moses, K., M. C. Ellis, and G. M. Rubin.** 1989. The *glass* gene encodes a zinc-finger protein required by *Drosophila* photoreceptor cells. *Nature* **340**:531–536.
28. **Orphanides, G., T. Lagrange, and D. Reinberg.** 1996. The general transcription factors of RNA polymerase II. *Genes Dev.* **10**:2657–2683.
29. **O'Shea-Greenfield, A., and S. T. Smale.** 1992. Roles of TATA and initiator elements in determining the start site location and direction of RNA polymerase II transcription. *J. Biol. Chem.* **267**:1391–1402.
30. **Paterson, J., and K. O'Hare.** 1991. Structure and transcription of the *singed* locus of *Drosophila melanogaster*. *Genetics* **129**:1073–1084.
31. **Pugh, B. F., and R. Tjian.** 1992. Diverse transcriptional functions of the multisubunit eukaryotic TFIID complex. *J. Biol. Chem.* **267**:679–682.
32. **Purnell, B. A., P. A. Emanuel, and D. S. Gilmour.** 1994. TFIID sequence recognition of the initiator and sequences farther downstream in *Drosophila* class II genes. *Genes Dev.* **8**:830–842.
33. **Qian, S., B. Varjavand, and V. Pirrotta.** 1992. Molecular analysis of the *zeste-white* interaction reveals a promoter-proximal element essential for distant enhancer-promoter communication. *Genetics* **131**:79–90.
34. **Roeder, R. G.** 1996. The role of general initiation factors in transcription by RNA polymerase II. *Trends Biochem. Sci.* **21**:327–335.
35. **Seagraves, W. A., and D. S. Hogness.** 1990. The *E75* ecdysone-inducible gene responsible for the 75B early puff in *Drosophila* encodes two new members of the steroid receptor superfamily. *Genes Dev.* **4**:204–219.
36. **Singer, V. L., C. R. Wobbe, and K. Struhl.** 1990. A wide variety of DNA sequences can functionally replace a yeast TATA element for transcriptional activation. *Genes Dev.* **4**:636–645.
37. **Smale, S. T.** 1994. Core promoter architecture for eukaryotic protein-coding genes, p. 63–80. *In* R. C. Conaway and J. W. Conaway (ed.), *Transcription: mechanisms and regulation*. Raven Press, Ltd., New York, N.Y.
38. **Smale, S. T.** 1997. Transcription initiation from TATA-less promoters within eukaryotic protein-coding genes. *Biochim. Biophys. Acta* **1351**:73–88.
39. **Smale, S. T., and D. Baltimore.** 1989. The “initiator” as a transcription control element. *Cell* **57**:103–113.
40. **Soeller, W. C., S. J. Poole, and T. Kornberg.** 1988. In vitro transcription of the *Drosophila engrailed* gene. *Genes Dev.* **2**:68–81.
41. **Thummel, C. S.** 1989. The *Drosophila E74* promoter contains essential sequences downstream from the start site of transcription. *Genes Dev.* **3**:782–792.
42. **Verrijzer, C. P., J. L. Chen, K. Yokomori, and R. Tjian.** 1995. Binding of TAFs to core elements directs promoter selectivity by RNA polymerase II. *Cell* **81**:1115–1125.
43. **Verrijzer, C. P., and R. Tjian.** 1996. TAFs mediate transcriptional activation and promoter selectivity. *Trends Biochem. Sci.* **21**:338–342.
44. **Verrijzer, C. P., K. Yokomori, J. L. Chen, and R. Tjian.** 1994. *Drosophila* TAFII150: similarity to yeast gene *TSM-1* and specific binding to core promoter DNA. *Science* **264**:933–941.
45. **Wampler, S. L., C. M. Tyree, and J. T. Kadonaga.** 1990. Fractionation of the general RNA polymerase II transcription factors from *Drosophila* embryos. *J. Biol. Chem.* **265**:21223–21231.
46. **Zenzie-Gregory, B., A. O'Shea-Greenfield, and S. T. Smale.** 1992. Similar mechanisms for transcription initiation mediated through a TATA box or an initiator element. *J. Biol. Chem.* **267**:2823–2830.
47. **Zenzie-Gregory, B., A. Khachi, I. P. Garraway, and S. T. Smale.** 1993. Mechanism of initiator-mediated transcription: evidence for a functional interaction between the TATA-binding protein and DNA in the absence of a specific recognition sequence. *Mol. Cell. Biol.* **13**:3841–3849.