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 285 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 and is functionally interchangeable for basal transcription in TATA-containing and TATA-deficient (TATA-less) promoters. In mammalian promoters, the Inr consensus sequence is Py-Py-A1-N-T/A-Py-Py (where A1 is the transcription start site) (3, 20, 39), whereas in Drosophila promoters, the Inr consensus is T-C-A1-G/T-T-T/C (1, 18, 32). It has been found that TAF150 and TAF250 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 site (4). In addition, photoaffinity cross-linking experiments suggested that dTAF150 and dTAF250 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 XbaI and PstI sites in the polylinker of pUC119. In these constructions, the XbaI site is upstream of the promoter, and the PstI 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.
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 CGAGCTCAGTCTTCCACGGGTCAC-3’, where the TCTAGA sequences immediately upstream of the core promoter correspond to the XbaI cloning site. The minimal core promoter templates for EF-1α and Sodh-1 and their corresponding mutant templates include sequences from −240 to +40 relative to the transcription start site. For these promoters, the sequences from −25 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 (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 (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

<table>
<thead>
<tr>
<th>Promoter</th>
<th>Promoter sequence</th>
<th>Transcriptional activity of mutant promoter (% of wild type)</th>
</tr>
</thead>
<tbody>
<tr>
<td>297 brown</td>
<td>AGTCGACGTC</td>
<td>&lt;7</td>
</tr>
<tr>
<td>E74A caudal</td>
<td>AGTCGATGC</td>
<td>2</td>
</tr>
<tr>
<td>E74B Doc</td>
<td>AGTCGATAC</td>
<td>2</td>
</tr>
<tr>
<td>engrailed G</td>
<td>AGACGCTTG</td>
<td>&lt;2</td>
</tr>
<tr>
<td>glass I</td>
<td>AGTCGCTG</td>
<td>&lt;5</td>
</tr>
<tr>
<td>singed</td>
<td>AGTCGTTG</td>
<td>3</td>
</tr>
<tr>
<td>Pst I site and Pst I sites</td>
<td>CTGATGGC</td>
<td>6</td>
</tr>
<tr>
<td>Pst I site and Xba I</td>
<td>CTGATTC</td>
<td>7</td>
</tr>
<tr>
<td>Promoter sequences from each of the promoter activities are also reported.</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

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 XbaI and PstI sites for cloning were annealed, extended with Escherichia coli DNA polymerase I (Klenow) and deoxynucleoside triphosphates (dNTPs), and digested with XbaI and PstI. The resulting DNA fragments were gel purified and ligated to XbaI- and PstI-digested pUC119 plasmid. The oligonucleotide with the same sense as the mRNA included the XbaI site and G promoter sequences from −2 to +18. The oligonucleotide with the opposite sense from the mRNA included the PstI 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 six positions. The functional range set A/G/T-C/G-A/T-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. Table 1, 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.

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 sequence were introduced by synthesizing oligonucleotides (with the opposite sense from the mRNA) with equal proportions of the four nucleotides at the six positions. The functional range set A/G/T-C/G-A/T-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. Table 1, 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.

FIG. 2. A single nucleotide alteration in the spacing between the DPE and Inr reduces core promoter activity and binding of purified TFIID. (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 TFIID. Arrows indicate DNase I hypersensitive sites.
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 motif (13- to 60-fold reduction in transcriptional activity upon mutation) (Fig. 1A and Table 1). These results indicate that the precise spacing between the Inr and DPE motifs is of critical importance for core promoter activity. Moreover, these findings are consistent with the strict maintenance of the +29 to +32 position of the DPE in naturally occurring core promoters (Fig. 1).

**Table 2. Determination of a DPE functional range set**

<table>
<thead>
<tr>
<th>Promoter</th>
<th>DPE sequence</th>
<th>Relative transcription</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wild-type G</td>
<td>AGACGT</td>
<td>100</td>
</tr>
<tr>
<td>Mutants</td>
<td></td>
<td></td>
</tr>
<tr>
<td>G 11-1</td>
<td>GCATGG</td>
<td>86 ± 20</td>
</tr>
<tr>
<td>G 11-2</td>
<td>GTATCC</td>
<td>81 ± 20</td>
</tr>
<tr>
<td>G 6-1</td>
<td>TCACAC</td>
<td>79 ± 7</td>
</tr>
<tr>
<td>G 6-2</td>
<td>GCACCT</td>
<td>74 ± 27</td>
</tr>
<tr>
<td>G 6-3</td>
<td>AGTTGT</td>
<td>70 ± 11</td>
</tr>
<tr>
<td>G 6-4</td>
<td>TCATGT</td>
<td>68 ± 7</td>
</tr>
<tr>
<td>G 6-5</td>
<td>AGATCT</td>
<td>63 ± 10</td>
</tr>
<tr>
<td>G 6-6</td>
<td>ACGCAC</td>
<td>54 ± 7</td>
</tr>
<tr>
<td>G 6-7</td>
<td>AGAGAC</td>
<td>54 ± 9</td>
</tr>
<tr>
<td>G 6-8</td>
<td>AGTTGA</td>
<td>53 ± 4</td>
</tr>
<tr>
<td>G 6-9</td>
<td>AACTGC</td>
<td>52 ± 2</td>
</tr>
<tr>
<td>G 6-10</td>
<td>GGATGC</td>
<td>51 ± 6</td>
</tr>
<tr>
<td>G 6-11</td>
<td>CACAGC</td>
<td>51 ± 12</td>
</tr>
</tbody>
</table>

DPE functional range set

<table>
<thead>
<tr>
<th>Promoter</th>
<th>DPE sequence</th>
<th>Relative transcription</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wild-type G</td>
<td>AGACGT</td>
<td>100</td>
</tr>
<tr>
<td>Mutants</td>
<td></td>
<td></td>
</tr>
<tr>
<td>G 11-1</td>
<td>GCATGG</td>
<td>86 ± 20</td>
</tr>
<tr>
<td>G 11-2</td>
<td>GTATCC</td>
<td>81 ± 20</td>
</tr>
<tr>
<td>G 6-1</td>
<td>TCACAC</td>
<td>79 ± 7</td>
</tr>
<tr>
<td>G 6-2</td>
<td>GCACCT</td>
<td>74 ± 27</td>
</tr>
<tr>
<td>G 6-3</td>
<td>AGTTGT</td>
<td>70 ± 11</td>
</tr>
<tr>
<td>G 6-4</td>
<td>TCATGT</td>
<td>68 ± 7</td>
</tr>
<tr>
<td>G 6-5</td>
<td>AGATCT</td>
<td>63 ± 10</td>
</tr>
<tr>
<td>G 6-6</td>
<td>ACGCAC</td>
<td>54 ± 7</td>
</tr>
<tr>
<td>G 6-7</td>
<td>AGAGAC</td>
<td>54 ± 9</td>
</tr>
<tr>
<td>G 6-8</td>
<td>AGTTGA</td>
<td>53 ± 4</td>
</tr>
<tr>
<td>G 6-9</td>
<td>AACTGC</td>
<td>52 ± 2</td>
</tr>
<tr>
<td>G 6-10</td>
<td>GGATGC</td>
<td>51 ± 6</td>
</tr>
<tr>
<td>G 6-11</td>
<td>CACAGC</td>
<td>51 ± 12</td>
</tr>
</tbody>
</table>

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.

A single nucleotide alteration in the spacing between the DPE and Inr reduces core promoter activity and binding of purified TFII D. 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 reduction of transcriptional activity as a result of a single nucleotide deletion or insertion (Fig. 2A). In addition, TFII D 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, TFII D 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 TFII D 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 TFII D and core promoter activity. Moreover, these findings are consistent with the strict maintenance of the +29 to +32 position 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-

![Figure 3](http://mcb.asm.org/)  
**Figure 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 Ns 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 relative 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 nuclelease 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 (45%) (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...
FIG. 4—Continued.
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, 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-G-A/T-C/T-G-T) that was used in the classification. Thus, we viewed 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 +28 (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-1a 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-1a 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-1a 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-1a F1 downstream element functioning as a DPE. We also tested the binding of TFII D to the EF-1a F1 promoter. As seen in Fig. 5C, purified Drosophila TFII D binds to the wild type, but not to the mutant EF-1a 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-1a 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, 747B, 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), and the insertion or deletion of a single nucleotide between the Inr and DPE reduces transcriptional activity and TFII D 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 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 TFII D 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 TFII D to TATA- versus DPE-containing promoters. There appear to be significant differences in the interactions of TFII D with TATA-containing and DPE-containing promoters. In Fig. 7, we present a model of TFII D engaged in two distinct core promoter interactions. In the TATA-driven promoter, some flexibility between the
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.
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 possess 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


