Sequence and Spacing of TATA Box Elements Are Critical for Accurate Initiation from the β-Phaseolin Promoter*

Margaret L. Grace‡, Mahesh B. Chandrasekharan§, Timothy C. Hall§, and Alison J. Crowe‡¶

From the ‡Department of Biology, University of Washington, Seattle, Washington 98195 and the ¶Institute of Developmental and Molecular Biology and Department of Biology, Texas A&M University, College Station, Texas 77843

The β-phaseolin (phas) gene, which encodes one of the major seed storage proteins of *P. vulgaris*, is tightly regulated at the transcription level resulting in strict tissuespecific and spatial expression during embryonic development. The phas proximal promoter contains a complex arrangement of core promoter elements including three TATA boxes as well as several putative initiator elements. To delineate the respective contributions of the core promoter elements to transcription initiation we have performed site-directed mutagenesis of the phas promoter. In vivo expression studies were performed on transgenic *Arabidopsis* harboring phas promoter mutants driving expression of the β-glucuronidase (gus) reporter gene. Quantitative assessment of GUS activity in seeds bearing the promoter mutants indicated that both sequence and spacing of the TATA elements influenced the efficiency of transcription. Substitution, insertion or deletion mutations had no effect on histochemical staining patterns indicating that strict spacing requirements are not essential for correct spatial expression of phas during embryogenesis. Further evaluation of the phas promoter by in vitro transcription analysis revealed the presence of multiple TATA-dependent transcription initiation start sites. The distance between TATA elements and transcription start sites was maintained in insertion and deletion mutants through the creation of novel initiation sites, indicating that positioning of the TATA elements rather than DNA sequence was the primary determinant of start site location. We conclude that, while dispensable for proper spatial distribution, the complex architecture of the phas promoter is required to ensure high levels of accurate phas transcription initiation in the developing embryo.

The proximal promoter region of a gene contains core promoter elements that determine the basal transcription activity of the gene and typically direct the positioning of the transcription initiation start site. Three different classes of transcription initiation elements have been identified in eukaryotes: TATA boxes, initiator (Inr) elements and downstream promoter elements (DPE). To date, only TATA boxes and Inr elements have been identified in plant promoters. All three elements utilize similar mechanisms of initiation requiring RNA polymerase II and sequence-specific binding of transcription factor IID (1). However, mechanistic differences exist between the different initiation elements. TATA-containing promoters can function in the absence of an Inr, whereas DPE promoters are Inr-dependent; TATA box directed transcription occurs on average 25–30 bp downstream of the TATA box while Inr driven transcription typically originates at the adenosine residue in the +1 position of the Inr element itself (2). Sequence analysis of the *Drosophila* and human genomes indicates that only a minority of promoters contain the classical arrangement of one TATA box and one Inr element (3, 4). The remaining promoters contain varying combinations of TATA, Inr, and DPE elements. It remains unclear, however, what role this core promoter diversity plays in transcription regulation.

Originally core promoter elements were thought to mediate basal transcription whereas gene-specific upstream regulatory elements were responsible for directing regulated gene expression. However, recent studies with both plant and animal genes have demonstrated that core promoter elements can play an integral role in both environmentally induced and developmentally regulated gene expression (1). Recently, developmental stage-specific recruitment of the TATA-binding protein (TBP) has been demonstrated for the human γ-globin gene (5) and light-induced expression of the TATA-less photosystem I gene *psaDb* was found to be Inr-dependent (6). Several plant and animal promoters have been identified which contain multiple TATA boxes (7, 8). Although the benefit, if any, of having more than one TBP binding site has not been well documented, recent analysis of the tubulin promoter in soybean indicates that the two TATA boxes present in this promoter function additively to direct transcription in seedlings. Interestingly, these two TATA boxes are differentially sensitive to light conditions (8), suggesting that, in the case of tubulin, the multiple TATA boxes play a role not only in directing high levels of transcription but also provide a mechanism for titrating gene activity in response to altered environmental conditions. Finally, recent enhancer-trapping studies in *Drosophila* (9, 10) and analysis of upstream activating sequences (UAS) in yeast (11) indicate that many enhancers and UAS preferentially activate select groups of core promoters, suggesting that the composition of a core promoter may restrict its responsiveness to specific transcription regulators. These studies, along with the recent identification of several general transcription factors that regulate a distinct subset of genes in a developmentalspecific manner (12, 13) strongly suggest that core promoter elements play a more active role in determining gene expression patterns than previously expected. Core promoter diversity may thus provide an additional level of transcription control beyond that achieved with upstream regulatory elements.

We have employed the seed storage protein gene β-phaseolin
The detection of multiple transcription initiation sites. We show here that both the sequence homology to a consensus mammalian Inr element (1, 16). The AT rich proximal promoter region of Arabidopsis thaliana plants (ecotype Columbia) were transformed using vacuum infiltration (24). The T1 seeds were surface-sterilized using 50% bleach and plated on MS medium containing 1 × B5 vitamins, kanamycin (50 μg/ml), and Timentin (tetracycllin disodium and clavulante potassium: SmithKline Beecham Pharmaceuticals, Philadelphia, PA; 150 μg/ml). Two weeks after plating, putative transformants (kanamycin-resistant seedlings, T1 plants) were transferred to fresh MS medium containing 1 × B5 vitamins and antibiotics, and grown until they formed rosette leaves. The plants were subsequently transferred individually to soil pots and grown to maturity in the greenhouse. Total DNA were collected from the dried siliques of individual transformants. Several individual homozygous or heterozygous lines were established for various phas gus constructs. Genomic DNA blots were performed to determine the copy number of T-DNA insertions in the transgenic plants.

Mutagenesis—Except for TATA3, the mutations shown in Fig. 2 were introduced into the TATAA elements within the −295 to +34 region of the phas promoter present in −295phas-gus/PCR-Script (14) using Mega- primer PCR-based site-directed mutagenesis (22). Products obtained following two rounds of amplification were ligated to an SfiI-digested vector (pPCR-Script® Amp SK+; Stratagene Inc.). The TATA3 mutant was generated using a QuickChange® mutagenesis kit (Stratagene Inc.) with −295phas-gus/PCR-Script as the DNA template. Incorporation of the mutations into the specific cis-elements and the integrity of the other regions of the promoter were verified by DNA sequencing.

Phas Constructs—Promoter fragments bearing mutations in the TATA region were digested with HindIII and NcoI, and ligated to HindIII/Ncol-digested vector −295phas-gus3′/PUC19 (14). Subsequently, the constructs were digested with HindIII and EcoRI to release fragments containing the mutant −295phas promoter, reporter gene and the 3′ region. The fragments were fused to HindIII/EcoRI-digested plant transformation vector pT350K. The vector pHM301K is a derivative of pCB301 (23), modified to include a CaMV35S-nptII-nos3′ cassette and a polycloning site in an inverted orientation to that of the parent vector. −295phas-gus-3′ phas fragment (3.4 kb) was excised from the vector –295phas-gus-3′/phaspUC19 using HindIII and EcoRI. The fragment was then mobilized into HindIII and EcoRI-digested pHM301K to obtain −295phas-gus-3′/phaspHM301K. A T-DNA cassette and a polycloning site in an inverted orientation to that of the parent vector. −295phas-gus-3′ phas fragment (3.4 kb) was excised from the vector –295phas-gus-3′/phaspUC19 using HindIII and EcoRI. The fragment was then mobilized into HindIII and EcoRI-digested pHM301K to obtain −295phas-gus-3′/phaspHM301K. A T-DNA cassette and a polycloning site in an inverted orientation to that of the parent vector.
RESULTS

Multiple TATA Elements Are Not Essential for Spatial Expression but Are Required for High Levels of Transcriptional Activity from the phas Promoter—The presence of three phased TATA elements within the phas promoter prompted an evaluation of their respective contributions to tissue-specificity and spatial distribution during embryogenesis. Substitution mutations were made in individual TATA elements (Fig. 1) and the spatial distribution during embryogenesis. Substitution mutations bearing the mutations in cis-elements are denoted by the prefix m. TATA regions are denoted by bold letters, and the introduced mutations are represented by lowercase letters. The 4-bp deletion (Δ) is indicated by −. The position and sequence of the insertions are also shown (boxes). Putative Inr elements are underlined.

Multiple TATA Boxes Direct β-Phaseolin Transcription

Fig. 1. Position and mutations of the TATA region within the −295phas promoter. The black bar (top) represents the proximal −295 to +1 region of the phas promoter (P-295phas). The 34 bp of the 5′-untranslated (UTR) region are indicated by a thin black line, and the hatched box represents the upstream part of the reporter gene (gus). T1, T2, and T3 (ovals) denote the locations of the three TATA elements. The primary and secondary transcription start sites, +1 and −12, respectively, are also indicated. A list of wild type and mutant sequences studied is shown. The −295phas promoter derivatives bearing the mutations in cis-elements are denoted by the prefix m. TATA regions are denoted by bold letters, and the introduced mutations are represented by lowercase letters. The 4-bp deletion (Δ) is indicated by −. The position and sequence of the insertions are also shown (boxes). Putative Inr elements are underlined.

Introduction of 10 bp in the TATA2-gus construct caused a 53% reduction in GUS activity relative to control. The statistically significant reduction in GUS activity due to these mutations indicates that both TATA1 and TATA2 contribute to the overall high level of expression from the phas promoter. As predicted from the histochemical analysis, mutation of all three TATA boxes essentially eliminated phas promoter activity. The minimal (−4%) activity shown for mTATA2-gus plants suggests that, in vivo, a cryptic cis-element or Inr site may be utilized in the absence of TATA elements to initiate low levels of transcription. The 68% reduction in GUS activity for seeds of mTATA1-gus lines as compared with the 46% reduction in GUS activity observed for mTATA2-gus, suggested that TATA1 was functionally more important than TATA2. However, statistical analyses (two sample Student’s t test) of the pairwise comparison of GUS activity values for the seeds of transgenic lines bearing mTATA1-gus and mTATA2-gus yielded a p value of 0.1 (data not shown), indicating that any ascribed functional hierarchy to the TATA elements is statistically insignificant.

Spacing of TATA Elements Influences phas Promoter Activity—The role of TATA element spacing relative to each other and to the transcription start site was evaluated by deletion of 4 bp between TATA2 and TATA3 (mTATA4-gus; Fig. 1). As shown in Fig. 2D, the embryos harboring the deletion derivative displayed a staining pattern similar to that obtained for the embryos bearing the −295phas-gus control (A). The absence of GUS expression in the radicle is consistent with earlier findings that expression in this region is mediated by distal regulatory elements upstream of bp −295 (14). As with −295phas-gus control plants, no GUS expression could be detected in leaves of the stably transformed Arabidopsis plants harboring the TATA mutants (data not shown). Conservation of the wild type staining pattern indicates that both tissue-specific expression and modular expression of phas within the embryo are not dependent on core promoter organization.

To identify any potential contributions from TATA-independent transcription, a mutant derivative of −295phas promoter bearing mutations in all three TATA elements (mTATA123) was generated (Fig. 1). Given the importance of TATA elements in transcriptional initiation, we anticipated that debilitation of all three TATA elements would completely abrogate transcription from the phas promoter. This was essentially confirmed by the absence of GUS expression in embryos harboring mTATA123-gus, except for a faint staining in cotyledons upon prolonged incubation (>5 h) in the X-gluc substrate (Fig. 2E). This faint staining was not observed for non-transgenic embryos (Fig. 2F). This qualitative assessment suggested that while tissue-specific and spatial positioning are maintained in the TATA mutants, high levels of phas expression require the presence of at least one functioning TATA box.

In order to assess the quantitative contribution of individual TATA elements to overall promoter activity, GUS activity was evaluated in seeds bearing each of the TATA mutants. As shown in Table I, mutation of TATA1 or TATA2 yielded 33 and 54%, respectively, of the GUS activity obtained for the −295phas-gus control. The statistically significant reduction in GUS activity due to these mutations indicates that both TATA1 and TATA2 contribute to the overall high level of expression from the phas promoter. As predicted from the histochemical analysis, mutation of all three TATA boxes essentially eliminated phas promoter activity. The minimal (−4%) activity shown for mTATA123-gus plants suggests that, in vivo, a cryptic cis-element or Inr site may be utilized in the absence of TATA elements to initiate low levels of transcription. The 68% reduction in GUS activity for seeds of mTATA1-gus lines as compared with the 46% reduction in GUS activity observed for mTATA2-gus, suggested that TATA1 was functionally more important than TATA2. However, statistical analyses (two sample Student’s t test) of the pairwise comparison of GUS activity values for the seeds of transgenic lines bearing mTATA1-gus and mTATA2-gus yielded a p value of 0.1 (data not shown), indicating that any ascribed functional hierarchy to the TATA elements is statistically insignificant.

Spacing of TATA Elements Influences phas Promoter Activity—The role of TATA element spacing relative to each other and to the transcription start site was evaluated by deletion of 4 bp between TATA2 and TATA3 (mTATA4-gus; Fig. 1). As shown in Fig. 2D, the embryos harboring the deletion derivative displayed a staining pattern similar to that obtained for the embryos bearing the control −295phas-gus (Fig. 2A). However, quantitative fluorimetric MUG assays indicated that the juxtaposition of TATA2 and TATA3 caused a statistically significant 63% reduction in GUS activity relative to control. Thus, disruption of spacing between the TATA elements did not greatly affect spatial regulation during embryogenesis, but did severely inhibit overall promoter activity. To further explore how TATA box spacing contributed to phas promoter regulation, we generated two additional spacing mutants, mTATA5AT-gus and mTATA10GC-gus, which contain insertions between TATA1 and TATA2 of 5 and 10 bp, respectively (Fig. 1). Interestingly, quantitative evaluation of GUS activity in embryos harboring these mutants (Table I) suggests that the two insertion mutants have distinct effects on phas activity. Introduction of 10 bp in the mTATA10GC-gus mutant en-
prominent of which migrated at −117 bp. A third minor product of −85 bp was detected with varying efficiency, and is most likely the result of TATA-box independent transcription (Fig. 4A and data not shown). Mapping the 5′-ends of the primer extension products (Fig. 3B) confirmed that the primary transcription product (105 bp) initiated at the previously determined +1 in vivo transcription start site (21) while the 117-bp secondary product initiated at −12. After prolonged exposure, minor products could also be detected which initiated at −9, +4, and +16/+17. These findings are consistent with previous mapping studies in bean cotyledons which identified a complex pattern of phas messages predicted to originate between −12 and +5, with the predominant forms localizing around +1 and −11 (21). Substitution of hTBP with purified recombinant wheat TBP (a gift of S. Ackerman, University of Massachusetts) (32) had no effect on the location of transcription initiation, although the preference for +1 was slightly enhanced.2 These data suggest that reconstitution of faithful basal phas transcription can be achieved with a human in vitro transcription system.

**TATA Boxes Determine Preferred Transcription Initiation Start Site in Vitro**—To assess the role of the TATA elements in determining both the efficiency and accuracy of transcription start site selection, TATA substitution mutants were transcribed in vitro and the resulting RNA products were analyzed by primer extension. Mutation of TATA1 eliminated transcription initiation at the secondary initiation site (−12) while only marginally reducing transcription efficiency from +1 (Fig. 4A, lane 2) resulting in an overall decrease of −20% relative to control −295phas-gus values. Conversely, templates carrying a mutated TATA2 exhibited undetectable levels of the primary transcription product yet still maintained wild type levels of initiation at −12 (lane 3). Thus, in agreement with the in vivo analysis, both TATA1 and TATA2 were found to contribute to phas promoter activity. These results further suggested that TATA1 mediates initiation at −12 while TATA2 is principally responsible for directing initiation at the primary transcription start site situated at +1. To determine if the non-canonical TATA3 element played a role in phas initiation we analyzed primer extension products from in vitro transcribed mTATA3-gus templates. Mutation of TATA3 caused no detectable changes in either level of transcription or start site location relative to −295phas-gus (lane 4) suggesting that TATA3 was not required for transcription initiation. To assess whether the TATA3 element could functionally substitute for TATA1 or TATA2, we generated a TATA12 double mutant (mTATA12-gus). Inactivation of both TATA1 and TATA2 severely inhibited transcription from all sites (lane 5) confirming that TATA3 is unable to function as an authentic TATA box. Consistent with the minimal GUS activity observed in mTATA123-gus embryos the mutation of all three TATA boxes caused a similar overall reduction in transcription (lane 6). These data suggest that TATA3 is dispensable for phas promoter activity whereas TATA1 and TATA2 function in concert to direct initiation of multiple phas messages.

**TATA Spacing Is Critical for Start Site Selection**—To assess whether relative spacing of the phas core promoter elements influenced transcription start site selection, we analyzed the deletion and insertion mutants diagrammed in Fig. 1 by in vitro transcription. Insertion of either five (mTATA5AT-gus; Fig. 4A, lane 9) or ten (mTATA10GC-gus; Fig. 4A, lane 10) nucleotides between TATA1 and TATA2 did not affect positioning of the primary transcription start site. Together with the mTATA2 results, the finding that both insertion mutants re-

---

TABLE I

<table>
<thead>
<tr>
<th>Construct</th>
<th>N</th>
<th>Natural logarithmic scale</th>
<th>Pairwise P value</th>
<th>Mean GUS activity</th>
<th>% GUS activity</th>
</tr>
</thead>
<tbody>
<tr>
<td>~295phas-gus</td>
<td>19</td>
<td>12.74 ± 0.45</td>
<td>0.23</td>
<td>935.02</td>
<td>100.00</td>
</tr>
<tr>
<td>mTATA1-gus</td>
<td>13</td>
<td>12.63 ± 0.23</td>
<td>0.10</td>
<td>830.27</td>
<td>92.96</td>
</tr>
<tr>
<td>mTATA2-gus</td>
<td>14</td>
<td>13.13 ± 0.61</td>
<td>0.18</td>
<td>505.14</td>
<td>54.02</td>
</tr>
<tr>
<td>mTATA3-gus</td>
<td>15</td>
<td>12.69 ± 0.68</td>
<td>0.14</td>
<td>345.46</td>
<td>37.39</td>
</tr>
<tr>
<td>mTATA123-gus</td>
<td>14</td>
<td>10.39 ± 0.40</td>
<td>0.15</td>
<td>32.75</td>
<td>3.50</td>
</tr>
<tr>
<td>mTATA5AT-gus</td>
<td>14</td>
<td>13.42 ± 0.85</td>
<td>0.20</td>
<td>671.42</td>
<td>71.81</td>
</tr>
<tr>
<td>mTATA10G4-gus</td>
<td>11</td>
<td>14.26 ± 0.65</td>
<td>0.20</td>
<td>1561.63</td>
<td>167.02</td>
</tr>
</tbody>
</table>

* Construct names reflect the size of the promoter (bp) for the control (~295phas-gus) and mutated (m) derivatives; the mutant bearing the 4-bp deletion between TATA2 and TATA3 is denoted by Δ4; the mutants bearing either an insertion of 5 bp AT-rich sequence or an insertion of a 10 bp GC-stretch (Fig. 1) between TATA2 and TATA3 are denoted by 5AT and 10GC, respectively.

* N, the total number of independent transformants analyzed for each construct.

* Means were calculated using a natural logarithmic (ln) scale of GUS activities.

* Standard deviations were calculated using a natural logarithmic (ln) scale of GUS activities.

* Standard errors of the mean were calculated using a natural logarithmic (ln) scale of GUS activities.

* Standard error of the mean obtained using the two sample Student’s t test for each pairwise comparison involving a given mutant construct compared to the control.

* /H11002 p, indicates the p value obtained for each comparison using the two sample Student’s t test at 95% confidence interval.

* Mean GUS activity was calculated by back transforming the mean ln values to the actual scale of measurement (nmol 4-MU/mg protein).

* Percentage (%) GUS activity was calculated using the mean GUS activity for each construct to depict the change in GUS activity due to the mutation compared to the control.

* /H9262 Not applicable in the control.

A.

ng template DNA 0 250 500 1000 2500 5000 7500 10000 α-amanitin - - - - - - +

MW 1 2 3 4 5 6 7 8 9

118 117 bp 105 bp 72

B.

FIG. 3. Reconstitution of phas transcription in vitro. A, template-dependent transcription of phas in vitro. Increasing concentrations of ~295phas-gus DNA template (250–4000 ng DNA as indicated) was in vitro transcribed in HeLa nuclear extract supplemented with recombinant hTBP. Transcription reactions were performed in the presence (+, lane 8) or absence (−, lanes 1–7) of 2 ng/μl α-amanitin. Transcripts were detected by primer extension followed by autoradiography. The primary 105 bp and secondary 117 bp phas primer extension products are indicated by arrows. A, mapping of phas transcription initiation start sites. Primer extension products obtained exactly as in Fig. 1A, lane 4, were electrophoresed on a 6% denaturing polyacrylamide gel alongside a dideoxy sequencing ladder. The predominant transcription initiation sites are indicated by arrows. Minor sites (detected on longer exposures) are designated with closed circles. Primer extension products are indicated with arrows. Radiolabeled dX 174 DNA digested with HaeIII (Invitrogen) was used as a molecular weight marker (MW).

Table 1 shows the results of the GUS activity in seeds of plants transformed with gus fusions to the wild type and TATA-mutated ~295phas promoter. The table includes the construct names, the number of transformants analyzed (N), the natural logarithmic scale of GUS activities, the mean GUS activity, and the percentage of GUS activity compared to the control. The table also includes the standard deviations and standard errors of the mean. The data indicate that the mutation of TATA1 and TATA2 had a significant effect on the GUS activity, with a decrease in GUS activity observed in the mutants compared to the control. This suggests that TATA1 and TATA2 are important for the transcription of the phaseolin promoter.

The results also show that the mutation of TATA3 had a minimal effect on the GUS activity, indicating that TATA3 is not as important for transcription as TATA1 and TATA2. The insertion of 5 bp AT-rich sequence or an insertion of a 10 bp GC-stretch between TATA2 and TATA3 had a more significant effect on the GUS activity, with a decrease in GUS activity observed in the mutants compared to the control.

The data suggest that the most prominent start site was the wild type secondary product located at position -22, and the secondary site located at position -24. The results also indicate that the TATA1 and TATA2 sites are essential for transcription, and that the secondary start site is not as important as the primary start site.

As expected, the TATA1 site was responsible for initiation at -12, and the altered spacing between TATA1 and TATA2 caused a shift in the location of the secondary start site. The in vitro transcription of mTATA10GC-gus generated a novel, diffuse secondary product initiating at approximately -20, whereas mTATA5AT-gus transcription yielded two discrete secondary products equal to or smaller in size than the secondary product observed with the wild type promoter (lane 9). The generation of a diffuse band suggests the use of multiple start sites situated near position -23, potentially due to the lack of a consensus Inr element in this region. The absence of a start site upstream of the wild type promoter further supports the idea of multiple start sites.

In conclusion, the results suggest that TATA1 and TATA2 are essential for transcription, and that the secondary start site is not as important as the primary start site. The mutation of TATA3 had a minimal effect on the GUS activity, indicating that TATA3 is not as important for transcription as TATA1 and TATA2. The insertion of 5 bp AT-rich sequence or an insertion of a 10 bp GC-stretch between TATA2 and TATA3 had a more significant effect on the GUS activity, with a decrease in GUS activity observed in the mutants compared to the control. The results also indicate that the TATA1 and TATA2 sites are essential for transcription, and that the secondary start site is not as important as the primary start site.

Multiple TATA Boxes Direct β-Phaseolin Transcription
FIG. 4. Core promoter architecture determines transcription start site selection. A, in vitro transcription of phas promoter mutants. -295phas-gus and the indicated TATA mutant templates were in vitro transcribed in HeLa extract supplemented with hTBP. Transcripts were detected by primer extension followed by autoradiography. Radiolabeled φX 174 DNA digested with HaeIII was used as a molecular weight marker (MW). Numbers below the figure indicate levels of expression relative to wild type as determined by ImageQuant analysis of the scanned autoradiogram. B, mapping of TATA mutant transcription start sites. Primer extension products obtained as in A were electrophoresed on a 6% denaturing polyacrylamide gel alongside a dideoxy sequencing ladder. Primer extension products initiating from +1 and -12 are indicated by arrows. Asterisks denote novel primer extension products resulting from aberrant transcription initiation.
novel TATA box in the 5AT mutant (see Fig. 1). If the mTATA5AT-gus mutant utilizes this newly created TATA box (TATTTAA), which has been found to function at moderate levels in human and yeast in vitro systems (20, 33, 34) the spacing between TATA and −12 would be maintained at a distance of 23 bp. Further mutational analysis will need to be performed to ascertain which TATA box is employed in this mutant. Regardless of the exact spacing, these data clearly demonstrate that insertions between TATA1 and TATA2 do not affect positioning of primary transcription initiation, but do have a marked effect on the location of the secondary transcription product.

To confirm whether positioning of TATA2 is responsible for the start site selection at +1, we analyzed the in vitro transcription products generated by mTATAΔ4-gus that contains a 4-bp deletion between TATA2 and TATA3. In order to maintain optimal spacing between the TATA boxes and their respective initiation sites, both the −12 and +1 sites would be predicted to shift downstream. Consistent with this prediction, both the primary and secondary transcripts generated by mTATAΔ4-gus exhibited a decrease in their apparent size (lane 8). Mapping the 5′-end of the mTATAΔ4-gus primer extension products indicated that transcription initiated at +4 and −9 (Fig. 4B). By relocating the start sites 3 bp downstream of their expected locations, a distance of 25–28 bp was maintained between the TATA boxes and their respective initiation sites in the deletion mutant (Table II). These results are consistent with our initial prediction that TATA2 is critical for promoting transcription initiation at the primary start site, whereas TATA1 is primarily involved in directing transcription through the secondary start site located at −12.

In Vitro Reconstitution of in Vivo Transcription Efficiency—Comparison of total phas message levels produced from both primary and secondary transcription start sites by TATA spacing mutants in vitro with GUS activity detected for the same mutants in vivo revealed several significant differences. Most noticeably, we observed wild type levels of transcription from mTATAΔ4-gus and mTATA10GC-gus in vitro, while these same mutants exhibited significant differences in GUS activity relative to −295phas-gus in vivo (Table I). We reasoned that the observed discrepancy might reflect differing requirements for basal transcription factor concentrations in vitro versus in vivo. To assess this possibility, we conducted titration experiments on each mutant, varying the level of HeLa extract in the in vitro transcription reactions from 0.5 mg/ml, previously determined to be optimal for −295phas-gus expression (Fig. 4 and data not shown) up to 2.0 mg/ml. As shown in Fig. 5, transcription under conditions of excess HeLa (1.5 mg/ml) resulted in a significant decrease in transcription efficiency of mTATAΔ4-gus relative to wild type (lane 2) equivalent to the 63% reduction in GUS levels observed in transgenic mTATAΔ4-gus embryos (Table I). Concomitantly, mTATA10GC-gus exhibited a 2-fold increase in transcription (Fig. 5, lane 4), closely paralleling the 68% increase observed in vivo. Thus, by performing the in vitro transcription reactions in the presence of suprathermal concentrations of the basal transcription machinery we were able to successfully reconstitute the requirements for core promoter spacing observed in vivo.

Comparison of TATA substitution mutant transcription activity determined in vitro (Fig. 4) and in vivo (Table I) revealed a significant difference in TATA box requirements. Specifically, the absolute requirement for a functioning TATA2 element in order to obtain high levels of transcription activity in vitro (Fig. 4, lane 3) did not accurately reflect the moderate decrease of 46% in total GUS activity detected in Arabidopsis seeds harboring mTATA2-gus (Table I). However, the strict requirement for an intact TATA2 element in vitro was abrogated when transcription was performed under excess HeLa conditions (Fig. 5, lane 6). The increase in overall mTATA2-gus transcription levels was due to a sharp increase in transcription from −12, suggesting a shift in start site preference in the absence of TATA2. The relatively moderate effect of disrupting TATA2 in transgenic plants may similarly reflect the use of an alternate start site in vivo. Lastly, consistent with the very low level of expression detected upon disruption of all three TATA boxes in vivo (−4%), no detectable expression from −295mTATA12 or −295mTATA123 was observed under high HeLa concentrations (data not shown). These data suggest that although optimal expression of wild type −295phas-gus is obtained under limiting HeLa conditions, analysis of the promoter mutants in the presence of excess HeLa more accurately reflects the effect of disrupting TATA elements in vivo.

![Figure 5](image-url)
DISCUSSION

The core promoter is the ultimate target through which all transcription is regulated. Transcription factors must communicate with the core promoter in order to either enhance or repress transcription (35). Mutational analysis of the phas core promoter has demonstrated a requirement for multiple phased TATA boxes to direct high levels of accurate basal transcription.

Maintenance of Spatial Distribution in TATA Mutants—Sub-localization of phas message within the embryo was recently found to be under the combinatorial control of distal regulatory elements (14). The conservation of wild type histochemical staining patterns observed for the TATA mutants (Fig. 2) indicates that module-specific expression of phas during embryogenesis is not dependent on core promoter organization but instead relies on a complex array of upstream regulatory elements (14).

Reconstitution of Basal Transcription in Vitro—The reconstitution of faithful phas transcription initiation demonstrated here illustrates the utility of the HeLa in vitro transcription assay as a rapid means of assessing how core promoter architecture contributes to transcription initiation and supports previous findings that the function of basal transcription machinery is highly conserved in eukaryotes (20, 34, 36). By analyzing phas transcription activity in vitro we were able to assign specific roles for core promoter elements in transcription initiation in the absence of potential contributing effects on mRNA processing, transport or stability. In vitro and in vivo analysis of phas promoter activity demonstrated that the TATA box elements are required for basal transcription. Mapping of the primary and secondary phas transcription start sites confirmed that the HeLa in vitro transcription system accurately reconstituted the in vivo pattern of phas transcription initiation (21). The reconstitution of phas basal transcription regulation on naked DNA templates indicates that the observed requirement for core promoter organization is chromatin-independent. Comparison of in vitro transcription activity with GUS activity from transgenic Arabidopsis harboring phas TATA mutants confirmed that the in vitro transcription assay accurately reflected the in vivo dependence on multiple phased TATA boxes for strong phas activity.

TATA-dependent Transcription—In the majority of TATA-containing plant promoters, the TATA box lies 25–30-bp upstream of the transcription start site (18).3 The relative positioning of TATA elements within the phas promoter (Table II) suggested that TATA2 and TATA1 function independently to direct transcription through the primary and secondary start sites, respectively, whereas the noncanonical TATA3 element played no role in phas transcription. Our in vitro transcription analysis of TATA substitution and spacing mutants confirmed this prediction, and also revealed an interdependence between TATA1 and TATA2. As such, mutation of TATA1 eliminated transcription from the secondary start site, while also reducing the efficiency of TATA2-directed transcription at +1. Conversely, the deleterious effect of eliminating transcription at the primary start site in mTATA2 was alleviated through a compensatory increase in the secondary transcription product (Fig. 5). Together these data suggest that TATA1 and TATA2 function in concert to direct start site selection and transcription efficiency at the phas promoter.

Strict Spacing Requirement for Core Promoter Elements—Previous promoter studies of both animal and plant genes have demonstrated a strict spacing requirement between TATA box and Inr elements for accurate transcription initiation (2, 18, 37). Our data demonstrate that altering the distance between TATA elements and their respective sites of initiation has a significant effect both on accurate positioning and efficiency of transcription initiation at the phas promoter. Start site location was shifted in the phas promoter mutants in order to maintain a 25–30-bp distance between TATA box and start site. Our finding that both the primary and secondary start sites are relocated in the insertion and deletion mutants strongly suggests that it is the positioning of the TATA box that is critical for determining phas transcription start site selection. However, sequences surrounding the site of initiation may influence start site location as deletion of 4 bp in the mTATAA4-gus mutant caused the secondary transcription start site to shift 3-bp downstream to the putative Inr element located at −9 (PyPyCAPyPyPPyPy) while the primary start site relocated from +1 to a weak consensus Inr element at +4 (PyPyCAPyPyPPyPPy). Although it remains unclear if this shift of 3 rather than 4 bp is due to the presence of functional Inr elements at these sites or simply a preference for initiation at adenosine residues, these data suggest that additional sequences at the site of initiation itself may contribute to TATA-directed start site positioning.

Altered spacing had only a negligible effect on transcription efficiency through the secondary site; however, transcription from the primary start site displayed marked sensitivity to promoter architecture. Thus, deletion of 4 bp between TATA2 and the +1 transcription start site in mTATAA4-gus dramatically reduced efficiency of transcription both in vivo and under excess HeLa conditions in vitro indicating that the function of TATA2 is highly dependent on its position relative to the initiation site. Further site-directed mutagenesis analysis will be needed to determine if the putative Inr, which overlaps the primary transcription start site is indeed a functional Inr element. Interestingly, in vitro transcription efficiency at +1 was also lowered upon insertion of one-half of a helical turn upstream of TATA2 (mTATAA5AT-gus). This decrease was not due solely to increased spacing between TATA1 and TATA2 as insertion of a full helical turn at the same location (mTATA10GC-gus) enhanced transcription activity of the primary message. Therefore, the relative positioning of TATA1 and TATA2 appears to play a role in TATA2-directed initiation at the primary transcription start site, even though these insertions do not alter the distance between TATA2 and +1 (Table II). The higher level of expression observed in mTATA10GC-gus may reflect an increased binding affinity for TBP. Based on the observations of Brukn er et al. (38), it can be suggested that insertion of a series of GC residues induces intrinsic DNA curvature. Bent DNA exhibits a significantly higher binding affinity for TBP, presumably through increased exposure of the minor groove (40, 41), and thus may lead to an up-regulation of gene expression (42). Alternatively, the observed impact of TATA box spacing on transcription efficiency may reflect a requirement for TBP to bind to TATA1 and TATA2 in a cooperative manner. This hypothesis is consistent with our finding that transcription from the phas promoter displays a strong dose dependence on TBP.3 Thus, the enhanced transcription efficiency realized by having multiple TATA boxes might be achieved through cooperative binding of TBP molecules to adjacent binding sites on the same surface of the DNA helix. Crippling of TATA1/TATA2 cooperativity in the TATA1 and TATA2 mutants may cause the observed decrease in transcription at the primary start site. Regardless of mechanism, it is clear from our data that the relative positioning of TATA boxes within the phas core promoter strongly influences promoter activity.

Nucleosome Positioning at the phas Promoter—Chromatin

---

3 A. Ackerman, personal communication.
structural analysis of the *phas* promoter has implicated the TATA region in positioning a nucleosome over the start site of transcription (43). Screening of existing genomic sequences for their competitive ability to bind nucleosomes identified phased TATA boxes, similar to those found at the *phas* promoter, as one of the most stable nucleosome positioning sequences yet characterized (44). Preliminary micrococcal nuclease analysis of the *phas* promoter suggest that the TATA elements themselves are essential for maintaining a positioned nucleosome over the transcription start site in vitro. Potentially TATA3, which appears to play no significant role in transcription initiation, still contributes to *phas* gene regulation by influencing the local chromatin structure of the TATA region. Histone modification, including acetylation, is a key modulator of chromatin structure and gene activation (45, 46). Interestingly, a role for the TATA box in mediating targeted histone acetylation has recently been identified at the CUP1 promoter in yeast (47). This study suggests a direct requirement for sequence-specific TBP binding in the localized recruitment of histone acetyl transferases. We are currently exploring these and other potential mechanisms for TATA-directed chromatin regulation of the *phas* promoter.

Our data suggest that the complex core promoter architecture of the *phas* promoter is required for the robust production of multiple messages observed in vivo. As correct patterning of GUS activity was maintained in the absence of multiple functioning TATA elements, cooperativity between TATA boxes is not essential for proper spatial distribution or tissue specificity of *phas* messages. Taken together, our data support a model in which *phas* start site location is directed by individual TATA boxes, while efficiency of initiation requires the combined presence of multiple phased TATA boxes.

Acknowledgments—We thank Wang Kit Ng, Xiangyu Shi, and James Townsend for technical assistance.

REFERENCES