Proc. Natl. Acad. Sci. USA
Vol. 87, pp. 5387–5391, July 1990
Botany

Functional borders, genetic fine structure, and distance requirements of cis elements mediating light responsiveness of the parsley chalcone synthase promoter

(cis-element synergism/light induction/parsley protoplasts/point mutation/transient expression)

ANNETTE BLOCK, JEFFERY L. DANGL*, KLAUS HAHLBROCK, AND PAUL SCHULZE-LEFERT†

Max-Planck-Institut für Züchtungsforschung, Abteilung Biochemie, D-5000 Köln 30, Federal Republic of Germany

Communicated by Clarence A. Ryan, April 9, 1990

ABSTRACT **The genetic fine structure of cis-acting sequences previously shown to be necessary for light-regulated expression in the promoter of the parsley (*Petroselinum crispum*) chalcone synthase gene was analyzed. Site-directed mutations and changes in spacing between cis elements were measured in transient expression assays in parsley protoplasts. Clustered point mutations allowed assignment of functional borders. Single-base substitutions within a highly conserved cis element (box II/G box) defined a critical core of seven bases, 5′-ACGTGGC-3′. It is functionally equivalent to a second sequence-related element (box III), which could replace box II in an orientation-dependent manner. The activity of box II required the presence of another juxtaposed element (box I) at a defined distance. No distance requirement was observed between the two large separable promoter regions known to independently confer light-regulated expression. These data support our hypothesis that a cis-acting sequence that is present in a limited number of diversely regulated plant genes gains its functional capacity and specificity by combinatorial diversity involving flanking partner elements.**

Eukaryotic promoters are usually composed of multiple discrete functional modules, each containing one or more recognition sites for proteins triggering defined frequencies of transcription starts (1). It is assumed that both the sum and the interplay between such cis-acting elements explain the regulative capacity of a promoter.

Chalcone synthase (CHS) catalyzes the first committed enzymatic step of flavonoid biosynthesis in plants (2). Activity is tightly controlled at the transcriptional level and is regulated both temporally and spatially (3). Light is the primary external stimulus regulating the expression of a single CHS gene in parsley through the action of different defined wavelengths (4). This mode of expression is retained in suspension-cultured parsley cells and protoplasts (5). CHS is, therefore, an attractive candidate to study dissection of the cis elements responsible for light-mediated expression.

We have shown (6) that CHS promoter sequences are sufficient to direct light-regulated expression of chimeric gene constructs in parsley protoplasts. Various promoter constructs allowed us to define TATA-proximal (positions −100 to −226) and TATA-distal (−227 to −615) regions, which independently confer light responsiveness (7). However, when present in combination, levels of gene expression are controlled in a synergistic manner. Genomic sequencing of this promoter region revealed four light-inducible *in vivo* footprints defining four boxed sequences (I–IV, see Fig. 1). They are assembled within two light-responsive cis units (boxes I plus II and boxes III plus IV) and lie within the TATA-proximal and TATA-distal regions, respectively. Site-directed clustered mutations and deletions proved that the "footprinted" sequences are functionally necessary. Furthermore, the presence of boxes I and II together was required within a minimal TATA-proximal CHS promoter for light-regulated CHS expression.

We now address questions regarding the interplay of this set of cis elements (*i*) by defining the functional borders of the elements identified by footprints I and II, (*ii*) by genetic fine-structure analysis of the highly conserved box II/G box sequence, (*iii*) by functional exchange experiments between related sequences of boxes II and III, and (*iv*) by analyzing spacing requirements within and between the two separable light-responsive units.

MATERIALS AND METHODS

Oligonucleotide Synthesis and Cloning. Basic plasmids 061 and 041, containing either bases −226 to +150 or bases −615 to +150 of the parsley CHS[a] gene, linked by a *Bam*HI site to the coding region of the bacterial β-glucuronidase (GUS) gene in vector pRT99GUS.JD, were constructed by standard DNA manipulation procedures (6). The target for all mutations was the *Hin*dIII–*Bam*HI fragment from 041 (Fig. 1) cloned into the $pM_{a/c}$ vector system (8). All site-directed mutants were generated as described by Friedrich *et al.* (9). Single-base mutations were generated with 16-mer oligonucleotides (Applied Biosystems DNA synthesizer) containing the substituted base at a central position. Clustered mutations were generated using 30-mer oligonucleotides containing at least 12 bases of homology on either side of the mismatches. Identification of mutant clones was facilitated through a colony filter hybridization procedure using mutant oligonucleotides, ^{32}P-labeled with T4 polynucleotide kinase, as probes (10). Site-directed mutants were shuttled back into expression plasmids as *Stu* I–*Bam*HI fragments between the end-filled *Sal* I site and the *Bam*HI site of pRT99GUS.JD. All mutants were verified before transfection by dideoxynucleotide chain-termination sequencing (11).

Spacing mutants 061-B1 and 061-B2 are derivatives of the pM_c clone of mutant 061-B introducing an *Xho* I site between boxes I and II. After digestion with *Xho* I, either mungbean nuclease digestion or the Klenow polymerase end-filling reaction of the protruding 5′ ends resulted in either a 4-base-pair (bp) deletion or insertion of 5′-TCGA-3′. Spacing mutant 041-01 is a derivative of the pM_c clone of mutant 061-A, which introduces an *Xba* I site upstream of box II. Digestion with *Stu* I and *Xba* I, Klenow polymerase end-filling of the *Xba* I site, and religation gave rise to the 49-bp

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

Abbreviations: CHS, chalcone synthase; GUS, β-glucuronidase.

*Present address: Max-Delbrück-Laboratorium, Carl-von-Linne-Weg 10, D-5000 Köln 30, Federal Republic of Germany.
†To whom reprint requests should be addressed.

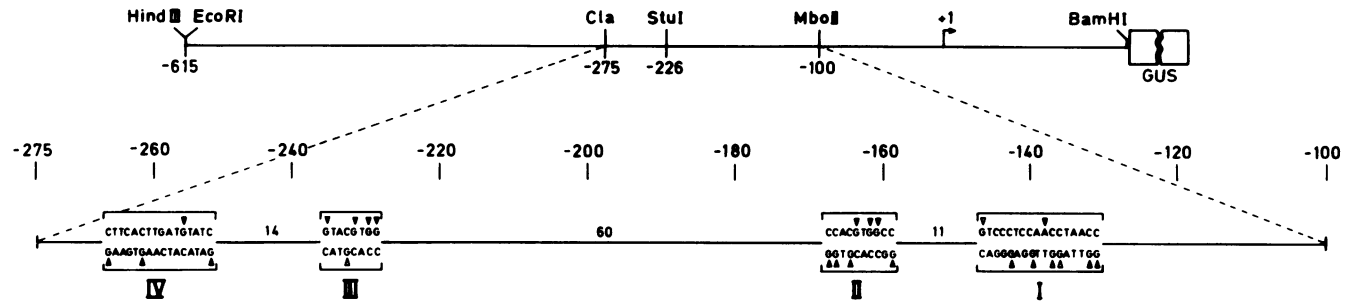

FIG. 1. Structure of expression plasmids. CHS promoter sequences of either 615 (041-based) or 226 bp (061-based) were fused to the GUS coding region. Lower line shows the positions and sequences of the four previously identified light-inducible *in vivo* footprints (I–IV). Open and solid triangles denote hypo- and hyperreactive base residues, respectively, within footprinted sequences (6, 7). Numbers on the bottom line indicate distances between boxes in bp. Landmark restriction sites for cloning, transcription start site (indicated by the arrow), and positions relative to the transcription start site (scale above lower line) are given.

deletion. This spacing mutant was shuttled back into expression plasmid pRT99GUS.JD as a *Hin*dIII–*Bam*HI fragment.

Transient Expression Analysis. Protoplasts were prepared from 5-day-old suspension-cultured parsley cells as described (5). Supercoiled plasmid DNA (20 μg per 10^6 protoplasts) was transferred into freshly prepared protoplasts by using a modification (12) of original protocols (13, 14). Each transformation assay containing 2×10^6 protoplasts was split; one part was kept in the dark, the other part was immediately irradiated with UV-containing white light (Philipps TL 20W18 lamps). After 9 hr, protoplasts were harvested and GUS enzyme activity was assayed as described (6, 15).

RESULTS

To determine the fine structure of cis-acting elements, we used two basic constructs containing either 226 or 615 bp of the CHS promoter sequence linked to the GUS reporter gene (Fig. 1). Construct 061 contains 226 bp of the CHS promoter, including footprinted sequences I and II, and defines the minimal-tested promoter fragment retaining light-regulated GUS expression (minimal promoter) (6). Construct 041, encompassing 615 bp, contains all four boxed sequences (I–IV) and directs high levels of light-regulated GUS expression (6, 7). The effect of site-directed or spacing mutations of all hybrid gene constructs was measured in transient expression assays by PEG-mediated transfer of plasmid DNA into parsley protoplasts (12).

Functional Borderline Determinations of Footprinted Sequences I and II. Clustered point mutations upstream from, between, and downstream from bordering contact residues of footprints I and II were generated and tested in the context of the minimal promoter construct 061 (Fig. 2). In Fig. 2 and see Figs. 3–5, GUS activity levels from each construct are

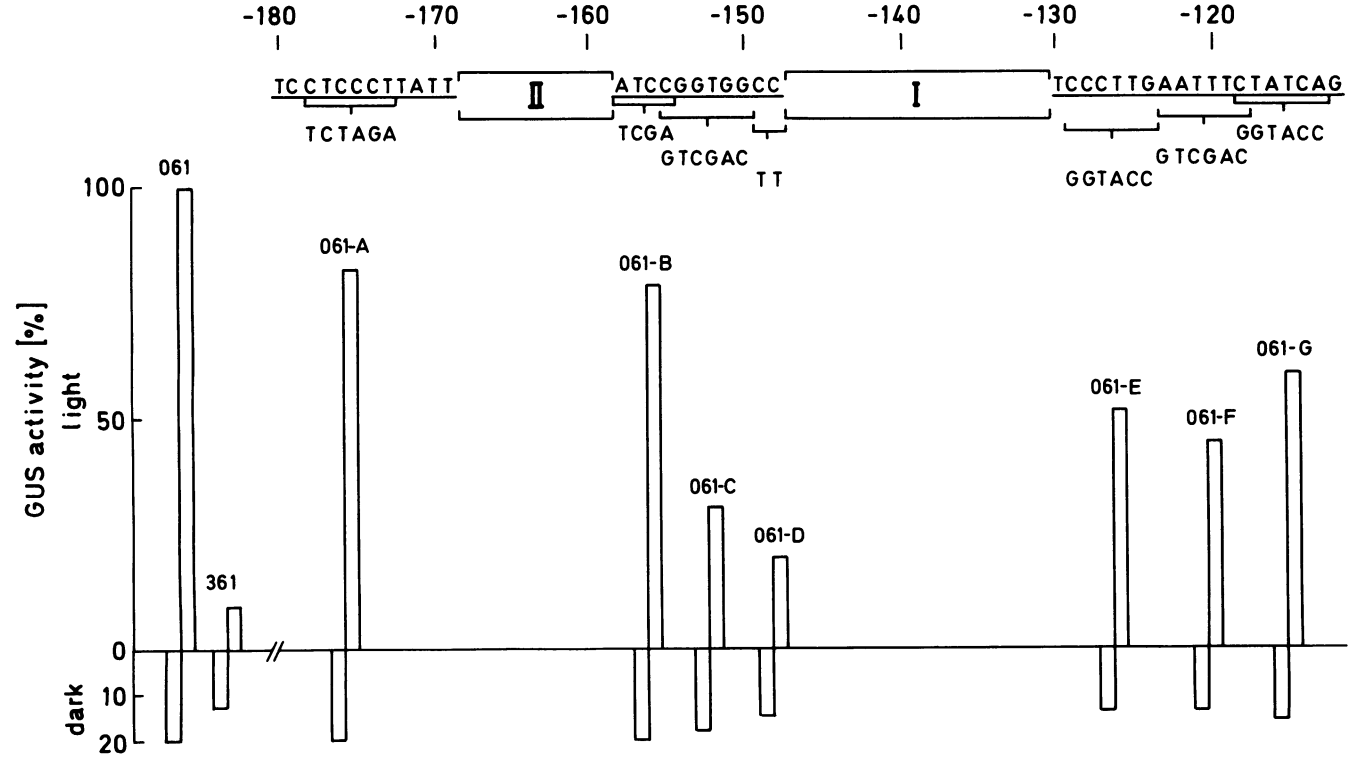

FIG. 2. Effects of clustered base substitutions. Transient expression data from at least six transformations of parsley protoplasts with each construct are given in this and all following figures. (*Lower*) Specific GUS activities (conversion rate of 4-methylumbelliferyl glucuronide to 4-methylumbelliferone) from protoplasts kept in the dark or irradiated with light are given for each promoter mutant (061-A to 061-G) as percentage of wild-type activity obtained with construct 061. This data representation is based on the mean value for 061 (278 pmol/mg, SD = 63 pmol/mg). All other SDs range from 8 to 76 pmol/mg. Construct 361, carrying 10-bp block mutations within boxes I and II, serves as a negative control (6). (*Upper*) The base replacements are shown for the coding strand.

shown as a bar graph directly below each mutation. Unambiguous borderlines were identified on both sides of footprint II: constructs 061-A and 061-B retained about 80% of light-controlled GUS expression mediated by the minimal promoter (construct 061). In contrast to the sharp functional boundaries around footprint II, mutation of sequences flanking footprint I affected light-regulated expression differently: Two upstream mutations essentially abrogated light-controlled expression (constructs 061-C and 061-D), whereas three mutations covering 17 bp downstream of box I (constructs 061-E, 061-F, and 061-G) individually reduced 061-mediated expression by 40–50%.

None of these mutations affected the dark level of GUS expression significantly. It should be noted that the two constructs 061-F and 061-G substituted for seven guanosine residues, each of which showed unaltered dimethyl sulfate reactivity *in vivo* upon light activation of the CHS gene.

These clustered point mutations confined functionally relevant sequences of box II to a small cis-acting element with sharp boundaries. Box I cannot be defined as precisely and clearly either exceeds both ends of the previously established *in vivo* footprint or overlaps with another element that was not detectable in the footprint analysis.

Single-Base Substitutions in Boxes I and II. Each base of the evolutionarily conserved box II (6) was substituted separately and the effect on light responsiveness was quantified relative to the activity of the unaltered minimal promoter (Fig. 3). All substitutions were base transitions, except the two transversions at position −168. The negative control was a 10-bp block mutation (construct 161) (6). Single-base transitions at each position between −160 and −166 virtually eliminated light controlled expression (constructs 061-4 to 061-10) and thus defined a core of functionally critical positions in box II. Upstream of this core, base substitutions either had no effect (construct 061-3) or retained at least half of the 061-mediated GUS activity (constructs 061-1 and 061-2). An intermediate effect was also observed at the 3′ end of the core (construct 061-11).

Together with the neutral effect of the flanking base substitutions in construct 061-B (Fig. 2), we conclude that critical core nucleotides of box II are confined to positions −160 through −166. The dark level of GUS expression was in all cases very low (Fig. 3 *Lower*).

Box II contains a palindrome, also a common feature of related sequences in other plant genes. Yet, the distribution of mutant phenotypes is asymmetric with respect to the center of the dyad symmetry. The best example that symmetrical changes result in different phenotypes is illustrated by the transitions at positions −162 and −167: whereas the former resulted in a complete loss of light-regulated expression, the latter was unaffected.

A question that can be answered with single-base substitutions is whether all contact residues defined by *in vivo* footprinting are functionally essential. With the exception of a hypomethylated guanosine residue at position −167, all footprinted guanosine residues within box II were found to be of functional importance. The correlation of functional relevance and light-induced alterations in dimethyl sulfate reac-

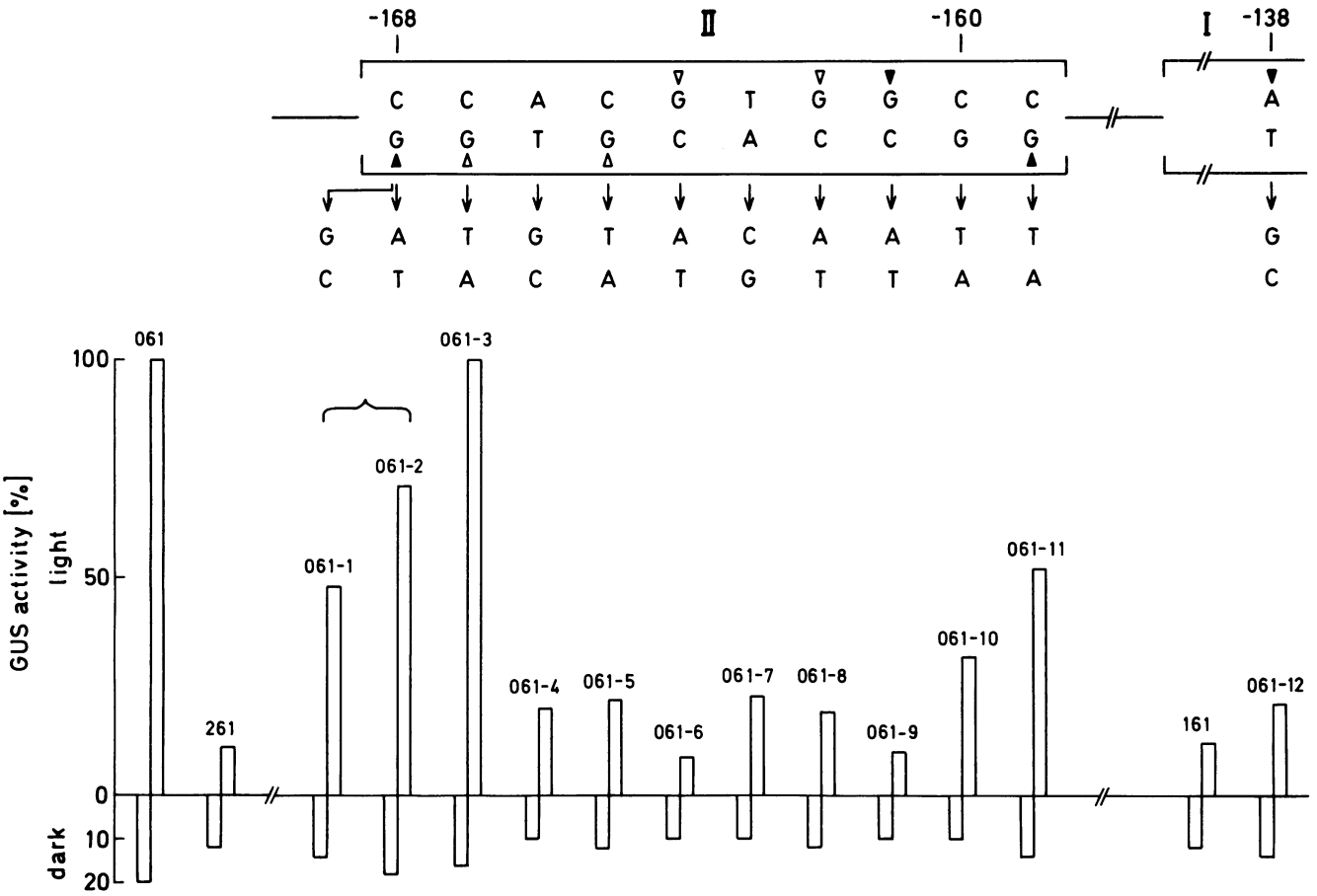

FIG. 3. Effects of single-base substitutions. (*Upper*) Single-base substitutions (061-1 to 061-12) are indicated by arrows below the boxed wild-type sequence. (*Lower*) GUS activity is expressed as a percentage of mean wild-type activity mediated by 061 (291 pmol per mg per min, SD = 59 pmol per mg per min). All other SDs range from 9 to 82 pmol per mg per min. Constructs 261 and 161, carrying 10-bp block mutations either in box II or box I, serve as negative controls. Open and solid triangles denote hypo- or hyperreactive base residues, respectively, from *in vivo* footprinting analysis (6).

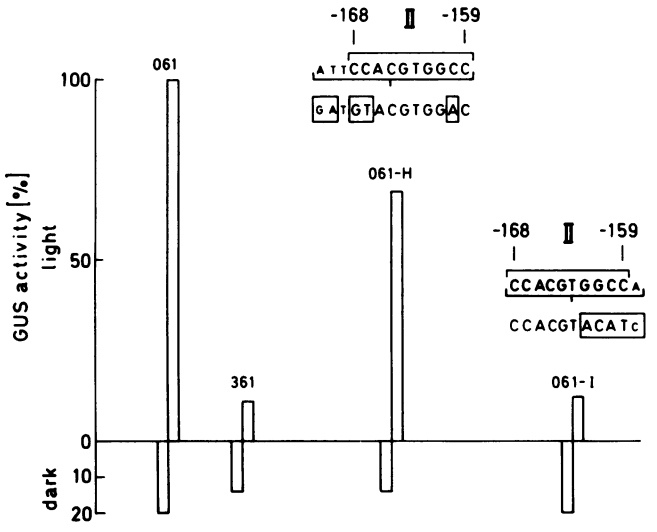

FIG. 4. Functional replacement of box II by box III. GUS activity is shown for the indicated promoter mutants containing box III either in its normal (061-H) or inverted (061-I) orientation. Boxed bases denote substitutions relative to wild-type sequences. Large letters of the DNA sequences indicate bases within, small case letters indicate bases adjacent to box II. Construct 361 serves as a negative control. Mean GUS activity of construct 061: 282 pmol per mg per min, SD = 55 pmol per mg per min. Other SDs range from 10 to 53 pmol per mg per min.

tivity was also the rationale for testing a single-base transition in box I (construct 061-12). Substitution of the unique hypermethylated adenosine residue at position −138 clearly defines a critical nucleotide within footprint I.

Functional Replacement of Box II by Box III. The sequence similarity between boxes II and III (Fig. 1) raised the question whether box III could functionally substitute for box II in the context of the minimal promoter. Two constructs were designed that replaced box II with a 13-bp fragment containing box III in either wild-type or reversed orientation. As shown in Fig. 4, this fragment restored light-inducible GUS activity almost completely, but only in the wild-type orientation. Comparison of the replaced nucleotides of box II in these two constructs with point mutation data from Fig. 3 reveals that functionally important bases are substituted in either orientation. In the normal orientation, two of the three exchanges (positions −160 and −168) affected light-controlled GUS activity as point mutations. In the reverse orientation, all four exchanges, which lie next to each other, are functionally important positions. We conclude that box III can replace box II in an orientation-dependent manner.

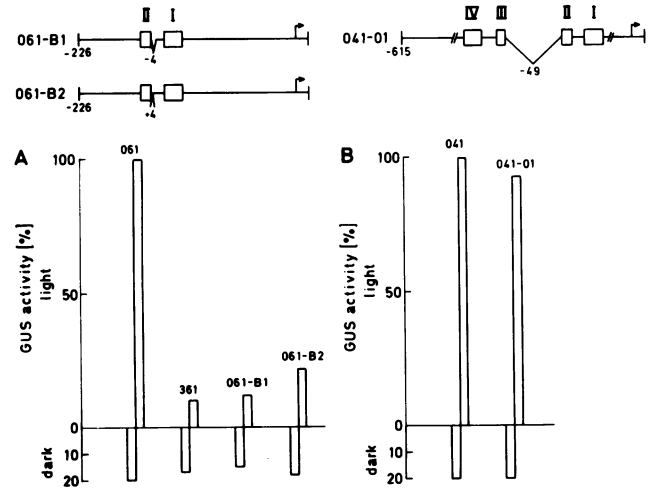

FIG. 5. Effects of spacing mutations. Transient expression data are shown for promoter mutants carrying altered distances either within (*A*) or between (*B*) the light-responsive cis units (7) of the parsley CHS promoter. GUS activity of constructs carrying a 4-bp deletion (061-B1) or insertion (061-B2) is given relative to the minimal promoter construct 061 (mean value, 263 pmol per mg per min; SD = 71 pmol per mg per min). Other SDs range from 20 to 36 pmol per mg per min. Data for the deletion construct spanning 49 bp between boxes II and III (041-01) is given relative to construct 041 containing all four boxes (mean value, 1780 pmol per mg per min; SD = 487 pmol per mg per min). The SD for construct 041-01 was 449 pmol per mg per min.

Spacing Requirements in the Parsley CHS Promoter. We next addressed the question of possible spacing requirements within a light-responsive cis unit (i.e., between boxes I and II). Since construct 061-B (Fig. 2) defines four functionally irrelevant bases outside the 3′ border of box II, we used this construct for deletion (061-B1) or insertion (061-B2) of four nucleotides at this position (Fig. 5). Both mutations essentially destroyed light-regulated GUS expression, suggesting a strict spacing requirement within the functional cis unit.

By contrast, deletion of 49 bp between boxes II and III (construct 041-01 in Fig. 5) had no appreciable effect when compared with the corresponding control construct 041. We conclude that sequences between boxes II and III can be deleted with no interference with synergistically high levels of light-dependent GUS expression.

DISCUSSION

An important goal of this study was the definition of functionally necessary bases within the evolutionarily conserved box II sequence (6, 16). On the basis of our results, we can define a core of essential nucleotides, 5′-ACGTGGC-3′, flanked by nucleotides with no or comparatively little functional relevance. We realize that we have generated only 11 of 30 possible changes in our point mutagenesis and cannot rule out the existence of additional critical bases. Green *et al.* (17) also defined a functional core of six nucleotides centered within the 14-bp sequence of the GT-1 binding site in the pea *rbcS-3A* promoter. Their analysis was confined to neighboring 2-bp alterations. The enhancer from the *Agrobacterium tumefaciens* octopine synthase gene is the only other example where mutagenesis of single nucleotides was carried out on a cis-acting element in plants (18). The authors found a total of five critical bases within a 16-bp palindrome. Interestingly, mutants showing less than 20% of wild-type activity were scattered throughout the 5′ half of the palindrome.

Palindromic sequences are a common feature of cis-acting elements. This has been interpreted as a reflection of the fact that proteins involved in sequence-specific protein–DNA interactions are often active as dimers or tetramers, binding with a twofold symmetry that matches the symmetry of the palindrome (19). The asymmetric distribution of mutant phenotypes in the box II palindrome indicates that the palindrome has no functional importance in the recognition of box II by its cognate trans-acting factor. Alternatively, binding of each subunit of a putative dimeric factor may have different requirements on the palindrome halves.

Our replacement of box II by box III sequences implies that the putative factor binding at box III can functionally replace that binding at footprint II. Since boxes I and II form a light-responsive cis unit (6) (neither alone can mediate light responsiveness), the factor(s) binding at boxes II and III must also be able to interact with the factor(s) binding at footprint I. Giuliano *et al.* (16) have shown that nuclear protein factors (G box binding factor) from *Arabidopsis* and tomato bind stably to the same dodecameric oligonucleotide containing the G box from the tomato *rbcS* gene. In addition, nuclear extracts prepared from *Antirrhinum majus*, *Nicotiana*

tabacum, Petunia hybrida, and *Arabidopsis thaliana* contain nuclear factors that bind to a G box/box II-related sequence at position −122 in the *Antirrhinum* CHS promoter (20). Neither available *in vitro* nor *in vivo* data allow discrimination between one or more factors that recognize these sequences.

Our results show clearly a spacing requirement between boxes I and II, although we cannot exclude that distances corresponding to even turns of a B helix could restore functionality. We cannot yet distinguish between the abolition of a protein–protein contact or the disruption of normal factor binding on the same face of the helix at box I and box II.

A striking feature of the parsley CHS promoter is that the two sequence-related boxes II and III are each paired with extended nonrelated sequences (boxes I and IV) at similar distances. *In vivo* footprinting of the *Arabidopsis* alcohol dehydrogenase gene showed an analogous spatial organization (21): a G box/box II version at position −217, showing a nearly identical pattern of footprinted bases, is flanked by an extended footprint at a distance that is remarkably similar to that between footprints I and II or footprints III and IV. Importantly however, the flanking footprint in the *Arabidopsis* alcohol dehydrogenase gene shows no obvious sequence similarities to boxes I or IV. Similar juxtapositions of sequences related to box II with other cis elements have been reported for the pea *rbcs* genes (18) and the *Nicotiana plumbaginifolia Cab-E* gene (22). These findings strengthen our initial speculation (6) that the G box/box II cis-element family gains its functional capacity and specificity by combinatorial diversity involving various cis-acting partners.

Finally, our present study allows a comparative evaluation of data obtained by an *in vivo* footprint and a functional mutation analysis within cis-acting sequences. We found that occasionally bases showing an altered sensitivity toward dimethyl sulfate can be functionally irrelevant and relevant bases can be insensitive. This disparity is, however, not surprising, since hypo- or hypermethylated guanosine residues are thought to be the result of an interference of the accessibility of dimethyl sulfate with hydrophilic or lipophilic protein domains (23), whereas functionally critical bases may be involved in complex phenomena such as binding geometry, binding stability, and accessibility (19, 24–26). Nevertheless, one major result is that the footprint data enabled us to target functionally important sequences with a resolution of a few nucleotides.

This work was supported by Fonds der Chemischen Industrie (to K.H.), the Fritz Thyssen Foundation (to P.S.-L.), and the National Science Foundation (to J.L.D.).

1. Dynan, W. S. (1989) *Cell* **58,** 1–4.
2. Hahlbrock, K. & Grisebach, H. (1979) *Annu. Rev. Plant Physiol.* **30,** 105–130.
3. Dangl, J. L., Hahlbrock, K. & Schell, J. (1989) in *Cell Culture and Somatic Cell Genetics of Plants*, eds. Vasil, J. K. & Schell, J. (Academic, New York), Vol. 6, pp. 156–173.
4. Ohl, S., Hahlbrock, K. & Schäfer, E. (1988) *Planta* **173,** 197–204.
5. Dangl, J. L., Hauffe, K. D., Lipphardt, S., Hahlbrock, K. & Scheel, D. (1987) *EMBO J.* **6,** 2551–2556.
6. Schulze-Lefert, P., Dangl, J. L., Becker-Andre, M., Hahlbrock, K. & Schulz, W. (1989) *EMBO J.* **8,** 651–656.
7. Schulze-Lefert, P., Becker-Andre, M., Schulz, W., Hahlbrock, K. & Dangl, J. L. (1989) *The Plant Cell* **1,** 707–714.
8. Stanssens, P., Opsomer, Ch., McKeown, Y. M., Kramer, W., Zabeau, M. & Fritz, H.-J. (1989) *Nucleic Acids Res.* **17,** 4441–4454.
9. Friedrich, K. H., Kolmar, H. & Fritz, H.-J. (1989) *Nucleic Acids Res.* **17,** 5762.
10. Wood, W. J., Gitschler, J., Lasky, L. A. & Lawn, R. M. (1985) *Proc. Natl. Acad. Sci. USA* **82,** 1585–1588.
11. Sanger, F., Nicklen, S. & Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. USA* **74,** 5463–5467.
12. Lipphardt, S., Brettschneider, R., Kreuzaler, F., Schell, J. & Dangl, J. L. (1988) *EMBO J.* **7,** 4027–4033.
13. Krens, F. H., Molendijk, Z., Wullens, G. J. & Schilperoort, R. A. (1982) *Nature (London)* **296,** 72–74.
14. Hein, R., Stabel, P., Steinbiss, A. H., Herrera-Estrella, L. & Schell, J. (1985) *Mol. Gen. Genet.* **199,** 161–168.
15. Jefferson, R. A. (1987) *Plant Mol. Biol. Rep.* **5,** 387–405.
16. Giuliano, G., Pichersky, E., Malik, V. S., Timko, M. P., Scolnik, P. A. & Cashmore, A. R. (1988) *Proc. Natl. Acad. Sci. USA* **85,** 7089–7093.
17. Green, P. J., Yong, M.-H., Cuozzo, M., Kano-Murakami, Y., Silverstein, P. & Chua, N.-H. (1988) *EMBO J.* **7,** 4035–4044.
18. Singh, K., Tokuhisa, J. G., Dennis, E. S. & Peacock, W. J. (1989) *Proc. Natl. Acad. Sci. USA* **86,** 3733–3737.
19. Takeda, Y., Ollendorf, D. H., Anderson, W. Z. & Matthews, B. W. (1983) *Science* **221,** 1020–1026.
20. Staiger, D., Kaulen, H. & Schell, J. (1989) *Proc. Natl. Acad. Sci. USA* **86,** 6930–6934.
21. Ferl, R. J. & Laughner, B. H. (1989) *Plant Mol. Biol.* **12,** 357–366.
22. Castresana, C., Garcia-Luque, I., Alonso, E., Malik, V. S. & Cashmore, A. R. (1988) *EMBO J.* **7,** 1929–1936.
23. Siebenlist, U. & Gilbert, W. (1980) *Proc. Natl. Acad. Sci. USA* **77,** 122–126.
24. Sarai, A. & Takeda, Y. (1989) *Proc. Natl. Acad. Sci. USA* **86,** 6513–6517.
25. Takeda, Y., Sarai, A. & Rivera, V. M. (1989) *Proc. Natl. Acad. Sci. USA* **86,** 439–443.
26. Pabo, C. O. & Sauer, R. T. (1984) *Annu. Rev. Biochem.* **53,** 293–321.