REGULATORY ELEMENTS REQUIRED FOR LIGHT-MEDIATED EXPRESSION OF THE *PETROSELINUM CRISPUM* CHALCONE SYNTHASE GENE

BERND WEIßHAAR, ANNETTE BLOCK, GREGORY A. ARMSTRONG, ANETTE HERRMANN, PAUL SCHULZE-LEFERT and KLAUS HAHLBROCK

MPI für Züchtungsforschung, Abteilung Biochemie, D-5000 Köln 30, Germany

Summary

Chalcone synthase (CHS) catalyzes the committed enzymatic step in flavonoid biosynthesis. In parsley (*Petroselinum crispum*), CHS is encoded by a single gene locus. Transcriptional activation of the gene in response to UV-containing white light has been demonstrated. Analysis of the CHS gene promoter by *in vivo* footprinting revealed four short sequences, designated Boxes I, II, III, and IV, which contain guanosine residues with altered reactivity to the methylating agent dimethylsulfate in UV-treated *versus* untreated parsley cells.

Studies were performed to characterize the functional components of the CHS gene promoter using a parsley protoplast transient expression system. By deletion and block-mutation analyses it was shown that Boxes I and II act together as a *cis*-acting unit and are necessary components of the minimal, light-responsive CHS gene promoter. The Box II sequence, which is similar to the conserved G Box sequence defined in promoters of ribulose 1,5-bisphosphate carboxylase small subunit (RBCS) genes, has been subjected to detailed analysis by site-directed mutagenesis. The heptameric sequence 5′-ACGTGGC-3′ has been defined as the critical core of Box II required for light induction in the context of the CHS gene minimal promoter. Box II is functionally equivalent to a second, sequence-related element (Box III) that can replace Box II in an orientation-dependent manner.

Chimaeric promoter-fusion constructs to the GUS reporter gene demonstrated that Boxes I and II, together constituting a *cis*-acting unit, are necessary and sufficient for light-mediated activation of the CHS gene promoter.

Introduction

Throughout their development and life cycle, organisms are exposed to various biotic and abiotic stress factors. Plants have evolved an array of mechanisms to protect themselves against these stresses. As a response to the potentially damaging effects of the abiotic stress factor UV light, plants accumulate UV-absorbing substances in the exposed tissue. The main compounds deposited,

Key words: promoter structure, chalcone synthase, *cis*-regulatory elements, light induction, ACGT factors.

which absorb light of 230 to 380 nm, are flavonoids (Hahlbrock *et al.* 1982). Flavonoids are a diverse class of substances that occur in all higher plants in the structurally related forms of anthocyanins, flavones, flavanones, flavonols and isoflavones, among others. The basic biochemical structure in all cases is a chalcone scaffold (Fig. 1). In addition to their role as UV protectants, flavonoids were also found to be involved in flower and fruit pigmentation (Harborne and Turner, 1984) and as phytoalexins in plant-pathogen interactions (Dixon, 1986). The increased synthesis of flavonoids in response to UV light is preceded by the transient, coordinated expression of the enzymes catalyzing the formation of these compounds (Chappell and Hahlbrock, 1984; Scheel *et al.* 1987). Chalcone synthase catalyzes the committed enzymatic step in the flavonoid-specific branch of phenylpropanoid metabolism (Hahlbrock and Grisebach, 1979). CHS enzymatic activity leads to the stepwise condensation of three acetate residues from malonyl-CoA with 4-coumaroyl-CoA to give 4,2′,4′6′-tetrahydroxy-chalcone (Heller and Hahlbrock, 1980; Ebel and Hahlbrock, 1982). The main substrate of CHS, 4-coumaroyl-CoA, is provided by the general phenylpropanoid metabolism (Fig. 1).

Extensive studies using cell suspension cultures and intact leaves of parsley have revealed that flavonoids are only present in minor amounts in dark-grown cells or etiolated leaves. Upon irradiation with UV light or with UV-containing white light, flavonoids accumulated in the vacuole of cultured parsley cells (Matern *et al.* 1983). In leaves exposed to UV light the vacuolar accumulation is restricted in a tissue-specific manner to epidermal cells (Schmelzer *et al.* 1988).

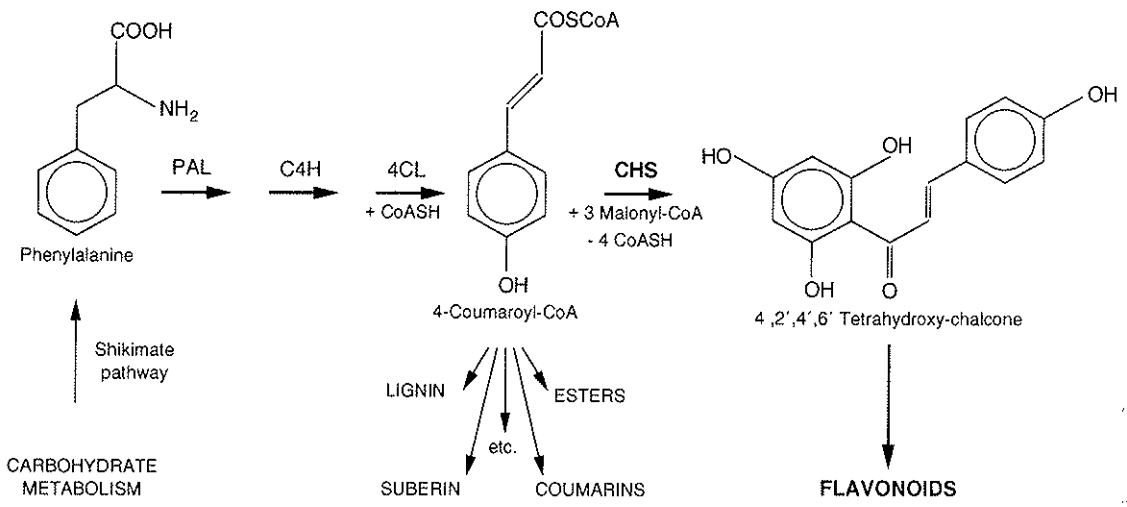

Fig. 1. Schematic representation of a selected portion of the flavonoid biosynthetic pathway. The enzymes of general phenylpropanoid metabolism and the key step towards flavonoid biosynthesis are shown: phenylalanine ammonia-lyase (PAL), cinnamic acid 4-hydroxylase (C4H), 4-coumarate:CoA ligase (4CL), chalcone synthase (CHS). The branch-point reaction for the formation of flavonoids is catalyzed by CHS and is specifically induced by UV light in parsley.

The need for various phenylpropanoid products varies during plant development and in response to changing environmental conditions, including stresses like pathogen attack or short-wavelength light. As a consequence, direction and intensity of metabolic fluxes through general phenylpropanoid and subsequent biochemical pathways are strictly regulated (see Hahlbrock and Scheel (1989) for review of phenylpropanoid metabolism). We focus our interest on the regulatory events necessary for light-dependent gene expression. Because of its clear response to light and its central role in flavonoid biosynthesis, CHS is a key candidate for molecular dissection of regulatory elements mediating light-dependent gene regulation in the biochemically well defined parsley system.

Induction of CHS gene expression by light

In many plant species CHS gene expression is strongly induced by more than one signal, e.g. floral development or light in *Petunia*, elicitor or light in french bean and soybean (see Dangl *et al.* 1989, for review). In cultured parsley cells, maximal CHS expression is UV-dependent, although blue light, red light, diurnal rhythm and developmental state of the tissue have additional modulating effects (Kreuzaler *et al.* 1983; Ohl *et al.* 1989). Cloning of cDNAs encoding CHS (Reimold *et al.* 1983) permitted experiments to determine the kinetics of CHS mRNA accumulation *in vivo* (Kreuzaler *et al.* 1983). Maximal transcriptional activity of the respective gene, as measured by run-on transcription in isolated nuclei, occurs a few hours after onset of irradiation with UV-containing white light (Chappell and Hahlbrock, 1984). There is a lag period of approximately 2 h before the first increase in transcriptional activity of the CHS gene is detectable. This lag period is modulated by light quality and can be circumvented by a pretreatment with blue light. This effect lasts for at least 20 h after blue light treatment (Ohl *et al.* 1989).

CHS transcripts are located in the epidermis of irradiated leaves, as shown by *in situ* hybridization experiments using the CHS cDNA as a probe (Schmelzer *et al.* 1988). The same tissue-specific localization was found for CHS protein and for the end products of flavonoid biosynthesis. Importantly, the kinetics of CHS induction in these *in planta* studies paralleled the results obtained with parsley cell suspension cultures.

CHS is encoded by one gene in parsley

Genes encoding CHS have been structurally characterized in various plant species. The deduced amino acid sequences have been well conserved during evolution (Niesbach-Klösgen *et al.* 1987). In parsley, CHS is encoded by a single genetic locus (Herrmann *et al.* 1988). Two allelic forms have been cloned and sequenced. Allele *PcCHS*^{a}, in contrast to the allele *PcCHS*^{b}, contains a 927-bp long transposon-like insertion at position −586 relative to the transcription start site, which is defined as position +1. No other differences were detected between

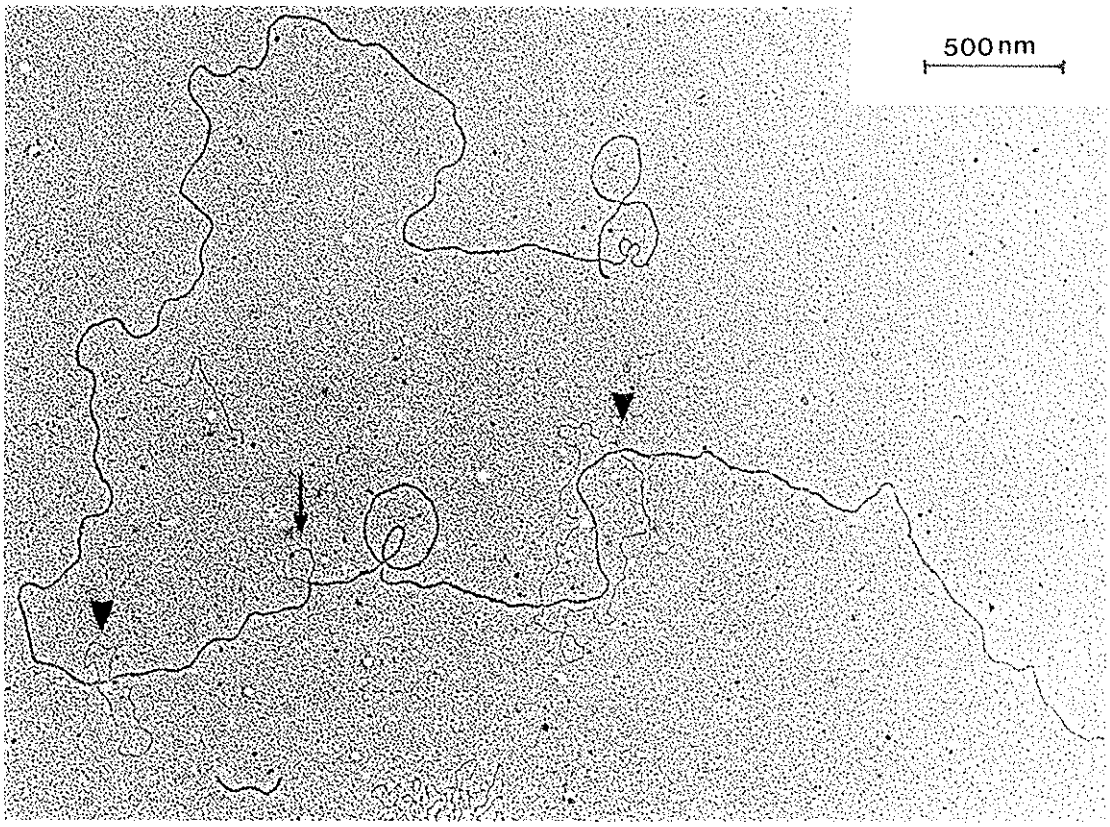

Fig. 2. Heteroduplex analysis of genomic λ-clones containing fragments from *PcCHS*^{a} and *PcCHS*^{b}. The arrow marks the position of the DNA loop corresponding to the unique segment in *PcCHS*^{a}, the arrowheads indicate loops caused by the different size of fragments adjacent to the λ-vector arms. The heteroduplex analysis was performed according to Davis *et al.* (1971).

the two alleles, either by heteroduplex analysis (Fig. 2) or by comparison of the available sequence information. Since both alleles showed light-inducible CHS expression in cell cultures derived from homozygous plants, the insertion in the promoter of allele *PcCHS*^{a} does not abolish correct light inducibility (Herrmann *et al.* 1988). The gene structure (Fig. 3) resembles that of most known genes encoding CHS in other species. One intron is present at a position conserved during evolution. Upon splicing of the primary transcript, a cysteine codon (position 65) is formed at the splice site (Niesbach-Klösgen *et al.* 1987). This cysteine codon was found in all CHS genes so far examined (Schröder and Schröder, 1990; Epping *et al.* 1990).

Parsley protoplasts: a transient assay system for inducible gene expression

Parsley protoplasts retain the responsiveness of dark-grown suspension culture

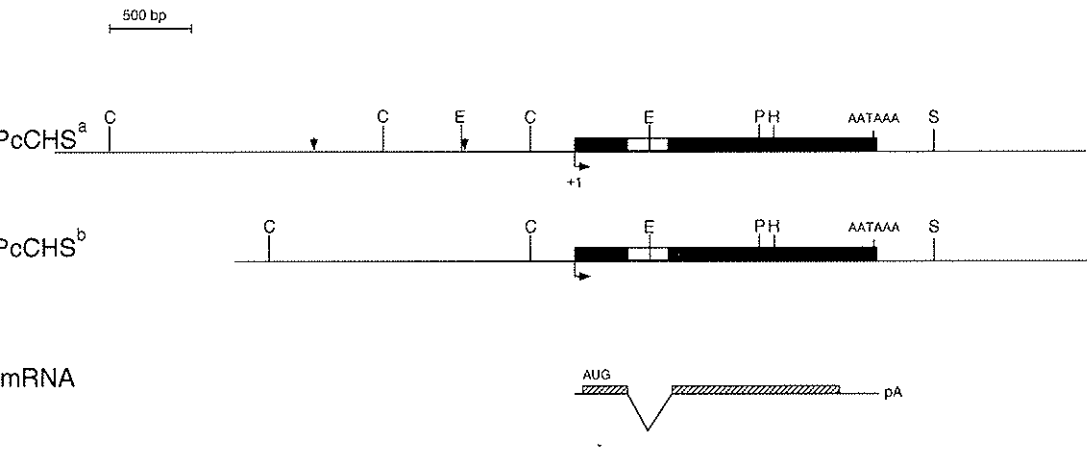

Fig. 3. Diagram of the parsley CHS gene. Some structural features of the two alleles of the parsley chalcone synthase gene, *PcCHS*^{a} and *PcCHS*^{b}, and of the common mRNA are shown. Filled boxes indicate the two exons, the open box marks the 263-bp intron. The region between the two arrowheads in *PcCHS*^{a} represents a transposon-like insertion. From the two mapped start sites (Herrmann *et al.* 1988) only the major one is shown and designated +1 as reference for nucleotide positions in the CHS gene promoter. The cutting sites of the following restriction enzymes are indicated: E, *Eco*RI; C, *Cla*I; H, *Hin*dIII; P, *Pst*I; S, *Sma*I. Positions of the putative poly(A) addition signal, of the poly(A) tail, and of the AUG are given. The hatched boxes on the bar representing the mRNA indicate the coding region.

cells to light or fungal elicitor (Dangl *et al.* 1987). The transcriptionally regulated genes involved in the formation of flavonoid glycosides and coumarin phytoalexins, respectively, are not activated during protoplasting or by treatment of the protoplasts with polyethylene glycol (PEG) to introduce foreign DNA. Thus, chimaeric gene constructs can be introduced into parsley protoplasts to analyze the effects of promoter manipulations on the expression of a reporter gene (Lipphardt *et al.* 1988). The application of the *Escherichia coli* β-D-glucuronidase (GUS) reporter gene (Jefferson *et al.* 1987) proved very useful in the parsley protoplast system (Schulze-Lefert *et al.* 1989*a*). This reporter gene offered the advantage of low background and rapid and sensitive fluorometric detection of the product formed by the GUS enzyme.

Chimaeric CHS gene promoter constructions (see below) were translationally fused to the GUS reporter gene in the background of a pRT99-based vector, pRT99 GUS JD (Schulze-Lefert *et al.* 1989*a*), which also contains a cauliflower mosaic virus (CaMV) 35S promoter-driven neomycin phosphotransferase (NPTII) gene. The construction of pUC-based GUS expression vectors also allowed the testing of elicitor-activated promoters (van de Löcht *et al.* 1990) under further optimized experimental conditions, e.g. linearization of the plasmid DNA (Ballas *et al.* 1988). The basic CHS promoter–GUS fusions tested in the pUC-based vectors respond to light qualitatively in the same fashion as in the more complex pRT99-based vectors (B. Weißhaar and K. Hahlbrock, unpublished results).

To broaden the usage of this system, the vector pUC GUS BT-2 (Fig. 4) was constructed that allows the easy cloning of *cis*-acting regulatory elements 5′ to a heterologous core promoter (see below). An important part of such a vector is the target promoter chosen. A 35S promoter truncated to −90 was used in experiments showing evidence that the binding site of the nuclear factor GT-1 confers light responsiveness in transgenic plants (Lam and Chua, 1990). The region between −46 and −90 of the 35S promoter, which contains an activation sequence factor 1 (ASF 1) binding site (positions −83 to −63, Katagiri *et al.* 1989), has been shown to have modulating effects on other promoter elements (Fang *et al.* 1989). For this reason, in pUC GUS BT-2 the CaMV 35S promoter from positions −46 to +8 is used as the target for *cis*-acting element-dependent transcription initiation. Because in this vector a transcriptional fusion is used, the sequence surrounding

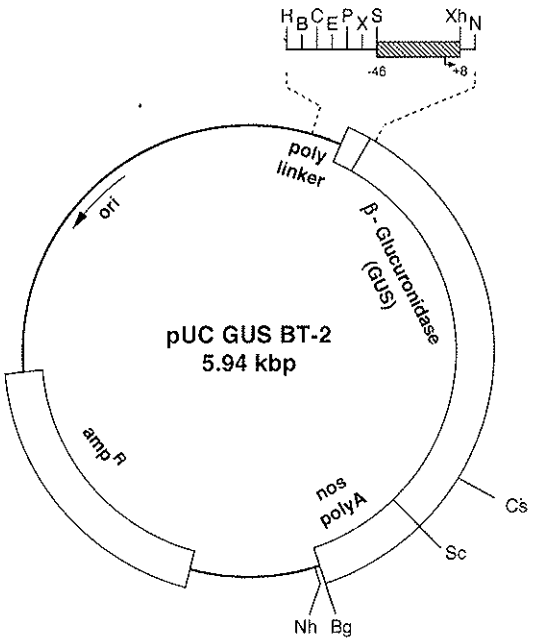

Fig. 4. Schematic representation of pUC GUS BT-2. The nucleotide sequences of pUC GUS BT-2 have the following origin: The modified ATG and the GUS coding sequence between the *Xho*I site and the *Csp*45I site originate from the plasmid pRT103 (Töpfer *et al.* 1988). The rest of the GUS coding sequence, the nos poly(A) addition sequence, and the pUC9 vector backbone are derived from pUC GUS (van de Löcht *et al.* 1990). Sequences including the polylinker in front of the promoter, the promoter itself and the polylinker after the nos poly(A) site were introduced as synthetic oligonucleotides. The length of the CaMV 35S promoter fragment was chosen according to Fang *et al.* (1989). The *Nhe*I and *Bgl*II sites, which are *Xba*I and *Bam*HI compatible, respectively, allow the easy subcloning of the hybrid promoter-GUS construction into other vectors (e.g. plant transformation vectors). All cloning steps were controlled by sequencing. The actual sequence of the whole cassette is available upon request. The hatched region represents the 35S promoter part, the numbers indicate the positions relative to the start site. Abbreviations for restriction enzyme sites are: H, *Hin*dIII; B, *Bam*HI; C, *Cla*I; E, *Eco*RI; P, *Pst*I; X, *Xba*I; S, *Sal*I; Xh, *Xho*I; N, *Nco*I; Cs, *Csp*45I; Sc, *Sac*I; Bg, *Bgl*II; Nh, *Nhe*I.

the ATG start codon of the GUS open reading frame was introduced in a modified version following the Kozak rules for optimized eukaryotic translational initiation (Kozak, 1983). Details of the construction are given in the legend to Fig. 4.

Genomic footprinting to the parsley CHS gene promoter

To gain insight into the mechanisms governing light-regulated gene expression, experiments were initiated to identify *cis*-acting elements within the parsley CHS gene promoter. Of special interest were elements that are necessary and sufficient for the light-dependent activation of this promoter. Characterization of light-induced protein-DNA interactions in the parsley CHS gene promoter provided a necessary basis for later experiments to define the DNA sequences and protein factors involved in these interactions.

To find possible *cis*-acting sequences, the method of *in vivo* footprinting (Church and Gilbert, 1984; Nick and Gilbert, 1985) was applied. Intact cells were exposed to the strong methylating agent dimethylsulfate (DMS). The concentration of DMS and the length of the treatment were optimized to allow partial methylation of guanosine residues at the N^7 position of the guanine ring. This chemical reaction is believed to be influenced by protein molecules that are in close contact with the DNA, thereby leading to hyper- or hypomethylation of G residues at the protein binding site or in the vicinity thereof. The product of this chemical reaction is a DNA molecule that is cleavable at the position of the modified residue by piperidine. Differences in the cleavage at certain G residues are then compared between *in vivo* -treated and control DNA.

To increase the sensitivity of detection, a restriction fragment containing the promoter region of interest is enriched by centrifugation of appropriately cut genomic DNA through a sucrose density gradient (Schulze-Lefert *et al.* 1989*a*). After creating a defined end point in the DNA fragment by a second restriction enzyme cut, the Maxam-Gilbert piperidine reaction is carried out to yield a series of genomic DNA fragments, starting at the reference cut and ending at the various methylated G residues. The position of the reference cut is crucial for the display of the resulting sequence ladder. The DNA fragments are then separated on a polyacrylamide gel and detected after electroblotting by indirect end labeling. A strand-specific probe of high specific activity is prepared by primer elongation on a single-stranded template (Church and Gilbert, 1984).

Suspension-cultured parsley cells homozygous for the *CHS*^{a} allele were used in the genomic footprinting experiments, excluding the problem of detection of multiple sequences by the probe. This problem may arise not only when heterozygous eukaryotic cells are used, but also when gene families are analyzed. The entire CHS gene promoter region from +40 to −615 (see Fig. 3) was analyzed for the appearance of light-induced *in vivo* footprints. Four short sequences showing differential reactivity to DMS in dark-grown *versus* UV-irradiated cells were detected (Schulze-Lefert *et al.* 1989*a*; Schulze-Lefert *et al.* 1989*b*). These differences were taken as indications of light-inducible DNA-protein interactions.

The four regions were named Boxes I through IV. They were defined by the outermost residues showing altered reactivity to DMS *in vivo* (see Fig. 5 for a summary of these results).

A time course of light-induced changes in the *in vivo* reactivity of the G residues to DMS more clearly showed the light-dependence of the four footprints in the CHS gene promoter. The footprints appeared approximately one hour after the onset of irradiation and were maintained during the period investigated (Schulze-Lefert *et al.* 1989*a*, 1989*b*). The notion that the four boxes are of functional relevance is strengthened by the fact that the timing of appearance of the footprints is in good agreement with the previously established kinetics of transcriptional activation of the CHS gene (Chappell and Hahlbrock, 1984; Ohl *et al.* 1989).

The four boxes are functionally necessary for light-dependent CHS gene promoter activity

The development of a transient expression system in parsley protoplasts allowed the rapid analysis of the parsley CHS gene promoter in the homologous system. A *PcCHS*^{a} promoter fragment, containing sequences up to −615, was fused translationally to the GUS reporter gene. Serial deletions of 5′ sequences from the basic construct 041 (Fig. 6) showed that sequences starting at position −226 relative to the transcriptional start site (construct 061) retain all the necessary information to direct light-regulated GUS expression. Further deletion to −100 resulted in a complete loss of light responsiveness. Therefore, construct 061 operationally defined the promoter sequences from −226 to +147 as the minimal light-responsive promoter (Schulze-Lefert *et al.* 1989*a*; see below and Fig. 9). By comparing constructs 061 and 071 it became clear that sequences between positions −100 and −226, which contain Box I and Box II, might include an element(s) needed for light-dependent expression (Fig. 6). To address this question, clustered point mutations were introduced into Boxes I and II. Ten basepairs were chosen from each sequence element and mutated to unrelated sequences creating diagnostic restriction sites. Each block mutation, when tested in the context of the minimal promoter, abolished light-regulated GUS expression in the transient assay system. It is important to note that mutation of either sequence element results in loss of light responsiveness (Fig. 6). Therefore, Box I and Box II are necessary *cis*-acting elements for the light response in the context of the minimal CHS promoter (Schulze-Lefert *et al.* 1989*a*). Since both sequences had to be intact for inducibility of expression, they were defined as a light-responsive *cis*-acting unit (Unit 1).

This interpretation was supported by the results of additional experiments carried out by Block *et al.* (1990). In the region between Box I and Box II a restriction site was introduced by site-directed mutagenesis. The changes in the promoter sequence at these positions were shown not to affect light inducibility. This restriction site was then used for the deletion or introduction of four

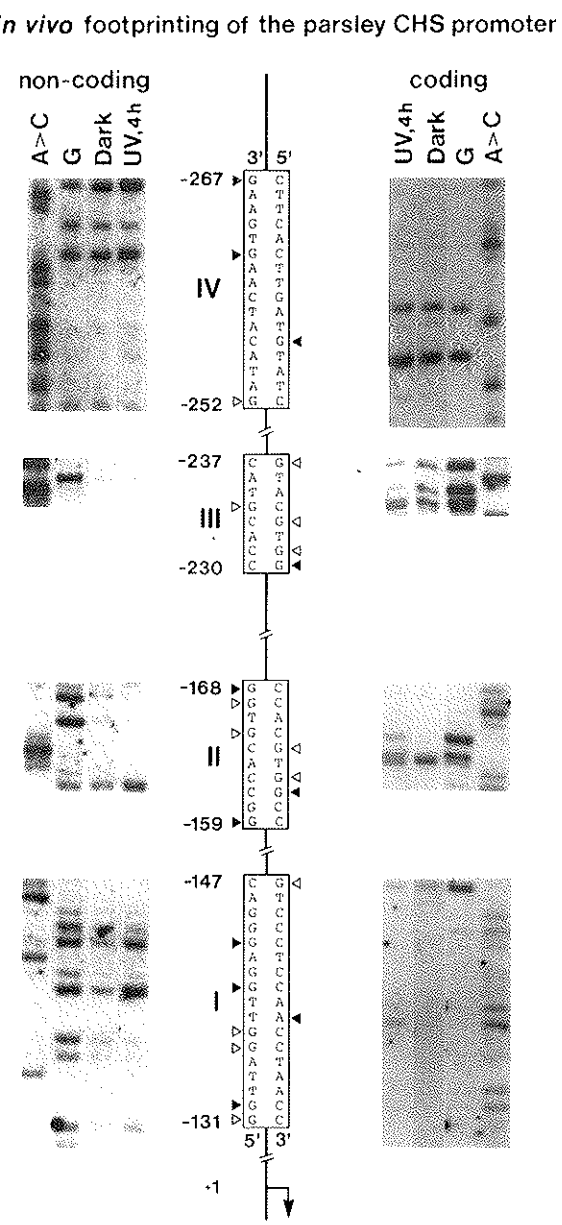

Fig. 5. Genomic footprints defining four protein contact regions in the parsley CHS gene promoter. The results from Schulze-Lefert *et al.* (1989*a*,*b*) are summarized. The filled triangles point to G residues showing hypermethylation, the open triangles indicate hypomethylation. The nucleotide sequences of the boxes and their positions are given. Results of an *in vivo* footprinting experiment comparing light-treated (UV, 4 h) and control (dark) cells are shown for the coding and the non-coding strands. The lanes labeled A>C and G show the Maxam-Gilbert reactions performed with cloned DNA.

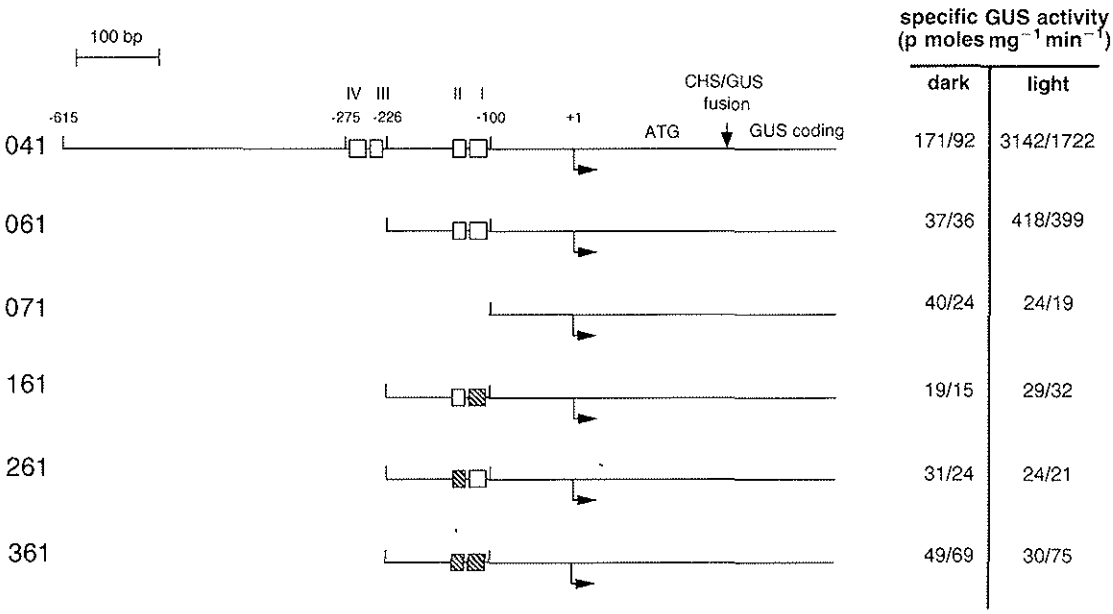

Fig. 6. GUS expression data from various CHS gene promoter-GUS deletion and block-mutation constructs. Results in this Figure are taken from Schulze-Lefert *et al.* (1989*a*). Names of CHS gene promoter-GUS fusions are listed on the left and the results from GUS assays on the right. The amount of fluorescent 4-methylumbelliferone formed is given as specific activity (pmoles product formed min^{-1}) of GUS enzyme mg^{-1} protein and is taken as a measure for promoter activity. The results from two typical experiments, each represented by the average of two transformations, are shown. End points of the CHS gene promoter sequences and their positions relative to the transcriptional start site are displayed. Boxed regions and roman numbers mark the positions of the *in vivo* footprints; hatched boxes indicate block mutations.The same translational fusion is shared by all constructs; 20 amino acids are encoded by *CHS* sequences and 10 amino acids are encoded by polylinker nucleotides.

nucleotides. Both mutations essentially destroyed light-regulated GUS expression, suggesting a strict spacing requirement within the functional *cis*-acting unit.

Functional redundancy in the parsley CHS gene promoter

The light-responsive unit contained in construct 061 generated only partial GUS activity when compared with the longer construct 041. The presence of Box III and Box IV further upstream in the promoter motivated the construction of CHS gene promoter-GUS fusions carrying these upstream elements in the context of various other parts of the CHS gene promoter (Schulze-Lefert *et al.* 1989*b*). Fig. 7 shows schematically the various promoter constructions and the results from GUS assays with these constructs. The data demonstrate that a second light-responsive element is contained in the CHS gene promoter in addition to Unit 1. In analogy to

Unit 1, Boxes III and IV were named Unit 2. Interestingly, a striking sequence similarity exists between Box II and Box III (see below). The presence of Unit 2 compensates partly for the loss of Unit 1 (see construct 351). Unit 2 can enlarge the light responsiveness of Unit 1 (compare constructs 051 and 061) and the 49-basepair fragment encompassing Boxes III and IV is required for the light response mediated through CHS gene promoter sequences from −226 to −615 (constructs 091 and 0101). Unit 2 in combination with sequences farther upstream is able to partially replace Unit 1 (compare constructs 341, 0101 and 061 in Fig. 7). It was concluded that Unit 2 is a weak light-responsive *cis*-acting unit. It is clear from these data that the elevated response to light of Unit 1 (construct 061) in conjunction with further upstream regions (construct 341) does not regenerate the complete activity of the full-length promoter construct 041. Therefore, the synergistic action of the two *cis*-active units controls the expression of CHS (Schulze-Lefert *et al.* 1989*b*). Interestingly, the elements are separated by approximately full numbers of turns of the B-DNA helix (see below; Fig. 9), although recent experiments have shown that 49 basepairs between Unit 1 and Unit 2 could be deleted without effect on light-regulated expression

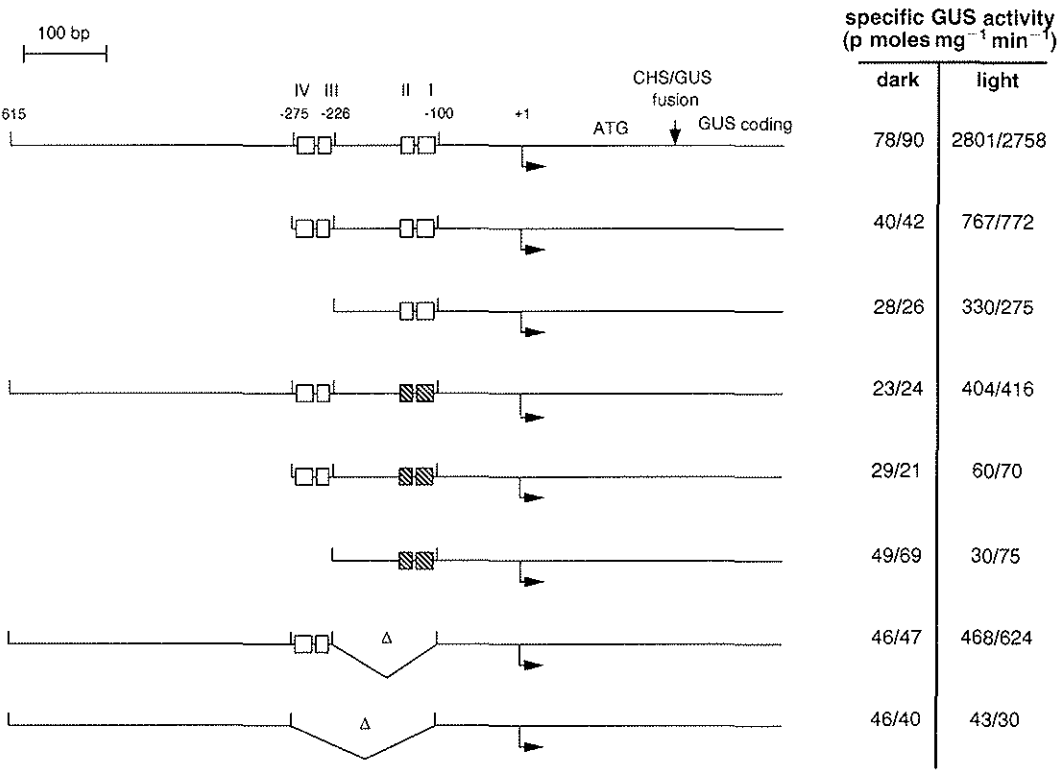

Fig. 7. Summary of GUS expression data demonstrating the existence of a second light responsive unit. Results are taken from Schulze-Lefert *et al.* (1989*b*). See Fig. 6 for description of symbols and data calculation. Deletions within the promoter region in constucts 0101 and 091 are indicated.

(Block *et al.* 1990). Taken together, these and other data show that parsley CHS expression in response to light is regulated by at least two separable light-responsive *cis*-acting units and one upstream region that is able to enhance the response of each of the units (Schulze-Lefert *et al.* 1989*b*).

Unit 1 is sufficient for light-dependent expression

The data so far discussed demonstrate that Unit 1 is necessary for light responsiveness in the context of the minimal CHS promoter. But is this Unit able to confer light inducibility on a heterologous promoter? To answer this question, the vector pUC GUS BT-2 was used (Fig. 4). A synthetic oligonucleotide comprising the sequence of Unit 1 was cloned in both orientations into the polylinker of pUC GUS BT-2 as single copy, as a dimer, and as a tetramer. The basic construct containing only the 35S promoter start site was shown to be minimally active in the transient expression system. GUS expression directed by this construct is not increased after irradiation of the protoplasts. Possibly, the expression is slightly repressed by light. As a control, a derivative of this vector was made, lacking the 35S promoter sequences (construct pUC GUS BT-Δ; Fig. 8). This construct served as a negative control in transient expression experiments and is transcriptionally inactive; it shows only the endogeneous background in the fluorometric determination of the product formed by the GUS enzyme.

The results obtained with the oligonucleotide-containing constructs showed that either orientation of Unit 1 is able to confer light responsiveness on a heterologous transcription start site (Weißhaar *et al.* 1991). From this it can be concluded that Unit 1 is not only necessary, but also sufficient for directing light-dependent expression in the homologous system (see Fig. 9). As mentioned above, co-action of both Box I and Box II at a defined distance is needed for light-mediated expression (Block *et al.* 1990). This finding is reinforced by additional results

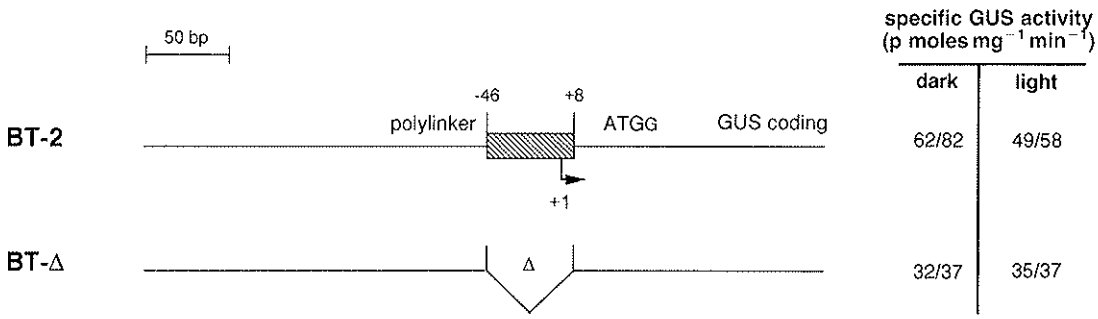

Fig. 8. Schematic delineation and results from GUS assays of pUC GUS BT-2 and its deletion derivative pUC GUS BT-Δ. The CaMV 35S promoter fragment contained in pUC GUS BT-2 (BT-2) is represented by the hatched region. These sequences were deleted from plasmid pUC GUS BT-2 to give pUC GUS BT-Δ (BT-Δ). See Fig. 6 for description of other symbols and data calculation.

demonstrating that tetramers of either of these boxes cloned 5′ to the truncated 35S promoter in pUC GUS BT-2 do not confer light inducibility on the constructs (Weißhaar, B. and Hahlbrock, K., unpublished results).

Box II of the parsley CHS gene promoter belongs to the G Box family of *cis*-acting elements

Many features of the transcriptional machinery of higher eukaryotes are functionally related and evolutionarily conserved. In the case of Box II (5′-CCACGTGGCC-3′) from the parsley CHS gene promoter, similar sequences have been found in many other promoters in a variety of plant species, including light-regulated promoters such as those of several CHS and RBCS genes. Also, in promoters known not to be light-regulated, sequences similar to Box II were found (Schulze-Lefert *et al.* 1989*a*). In the parsley CHS gene promoter, this sequence element is present in two related copies. Both Box II and Box III were detected by *in vivo* footprinting and found to be components of Units 1 and 2, respectively. Boxes II and III are functionally related as well in parsley (see below).

A sequence similar to Box II was originally identified as a conserved region among promoters of RBCS genes (Giuliano *et al.* 1988). These authors identified a factor in nuclear extracts from *Arabidopsis* and tomato, which binds to the conserved sequence, the so called G Box. More recently, a G Box binding factor (GBF) has been characterized in fractionated *Arabidopsis* extracts by gel retardation and *in vitro* DNAse I footprinting (DeLisle and Ferl, 1990). The G Box sequence is of functional relevance in the expression of RBCS genes, as demonstrated by mutation of the G Box present in the *Arabidopsis rbcS-1A* promoter. Also in this promoter, a second sequence (I Box) required for full RBCS gene promoter activity in transgenic plants has been identified (Donald and Cashmore, 1990).

Single base substitutions within Box II define a functional core of 7 bases

The Box II sequence has been recently subjected to detailed analysis by site-directed mutagenesis (Block *et al.* 1990). The goal of these experiments was to define the functionally relevant bases within the Box II/G Box *cis*-acting element. Each base within the 10-basepair region defined as Box II (Schulze-Lefert *et al.* 1989*b*) was substituted separately. All mutations were tested for their ability to direct light-dependent GUS expression in the context of the minimal promoter construct 061 (Fig. 9). The specific GUS activity yielded by the reference construct 061 was set to 100 % when evaluating the results (Fig. 10a). Single base mutations at each position between −160 and −166 virtually eliminated light-controlled expression (constructs 061/4 to 061/10). Upstream and downstream of this core, base substitutions had either no effect or retained at least half of the GUS expression of the controls (constructs 061/2, /3, and /11). The data show the heptameric sequence 5′-ACGTGGC-3′ to be the critical core of Box II required

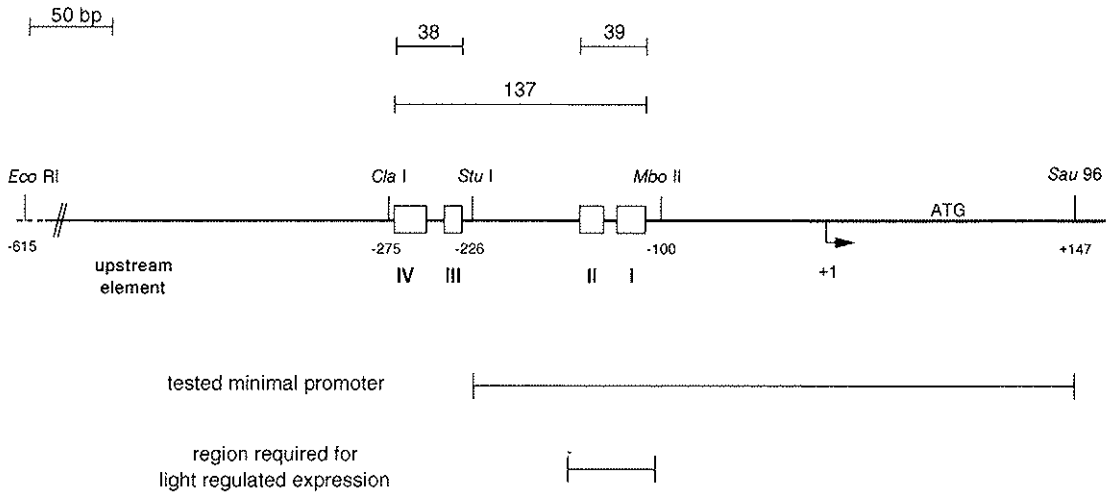

Fig. 9. Functional representation of the parsley chalcone synthase promoter. The positions and spacing of the promoter elements defined thus far in the parsley *CHS (PcCHS*^{a}*)* promoter are indicated. Positions of restriction sites used in the construction of the CHS gene promoter-GUS fusions are marked. The broken line at the left end of the drawing shows the position of the transposon-like element in the *PcCHS*^{a} allele. Above, the sizes and the distances of the *cis*-acting units are given in basepairs. As borders, the outermost nucleotides with changed reactivity to DMS in the *in vivo* footprinting experiments are given. At the bottom, the extension of the minimal promoter used in transient expression experiments and the region encompassed by the synthetic oligonucleotide cloned into pUC GUS BT-2 are indicated.

for light-induction in the context of the CHS minimal promoter. Together with data defining the borders of Box II by the introduction of clustered point mutations, it was concluded that flanking nucleotides had little or no relevance, provided that Box I or other *cis*-acting elements were not affected (Block *et al.* 1990). Obviously, it cannot be ruled out that point mutations other than those tested (11 out of 30 possibilities were examined) may lead to the definition of additional functionally important bases in Box II. In general, the data yielded by these analyses fully agree with the results of the *in vivo* footprinting experiments (see Block *et al.* 1990 for discussion).

Functional replacement of Box II by Box III

The nucleotide sequence of Box II is related to that of Box III, raising the question as to whether these sequences are functionally equivalent. The construction of the GUS-fusion plasmids 061/H and 061/I addressed this question (Block *et al.* 1990). As shown in Fig. 10b, replacement of Box II by Box III in the wild-type 5′ to 3′ orientation resulted in the restoration of light inducibility, whereas the reverse orientation did not. Comparison of the nucleotides changed by replacing Box II with Box III in this experiment with the results of the point-

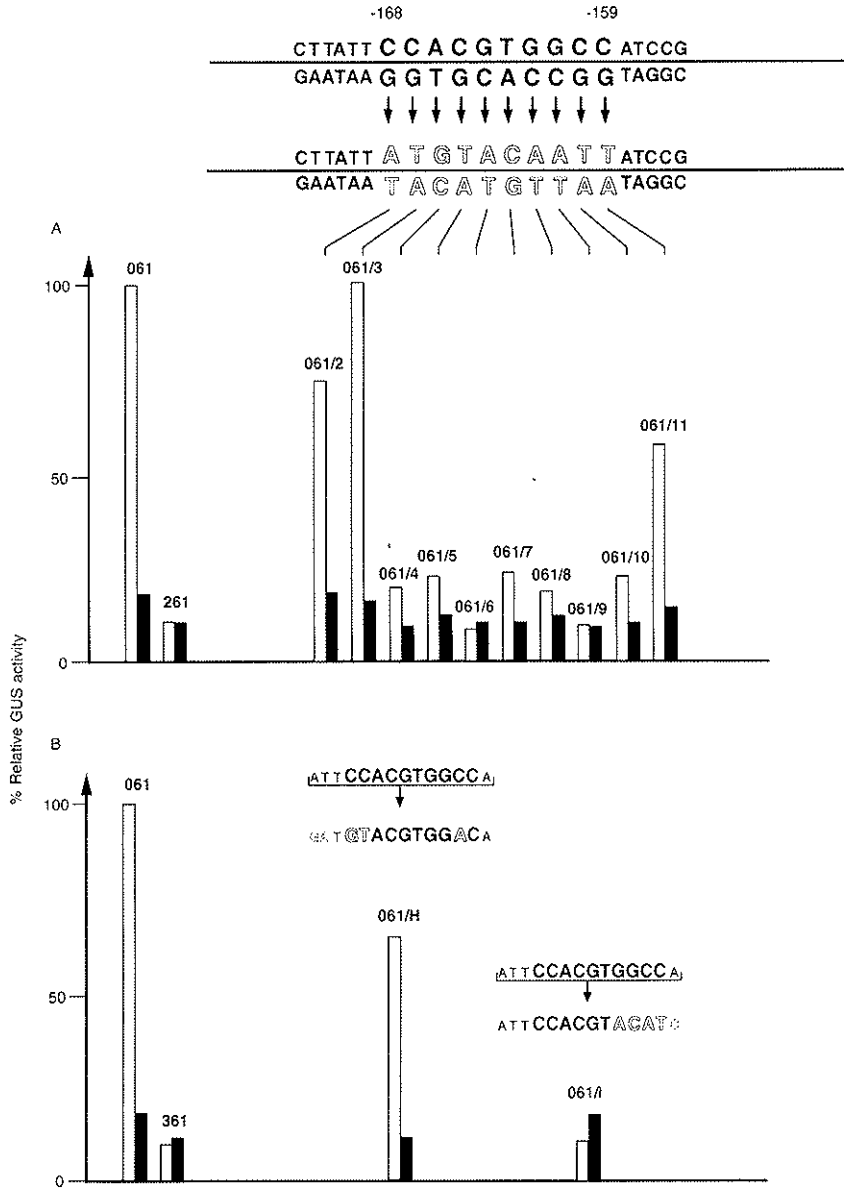

Fig. 10. Results of point-mutation analyses of Box II and of replacement experiments of Box II by Box III. Data presented are taken from Block *et al.* (1990). The nucleotides constituting the wild-type sequence of Box II are represented in bold and surrounding sequences are given in smaller characters. Mutated positions are indicated by outline letters. All constructs were made in the 061 background and the names of the respective 061 derivatives are listed. The open bars show the specific GUS activity for the given construct relative to the reference construct 061 after light treatment of the transformed protoplasts. Filled bars show the corresponding data for the dark control of each transformation assay. The data were calculated from at least six independent transformation experiments.

mutation series revealed that functionally important bases were substituted in both cases. It was concluded that Box III can functionally substitute for Box II in an orientation-dependent manner (Block *et al.* 1990). It is interesting to note that in Unit 1, Box II is located 5′ to Box I although Box III is located 3′ to Box IV. Since the individual *cis*-acting elements were unable to mediate light responsiveness, the interaction between factors binding to Boxes I and II (and to Box III in the construct 061/H) is necessary. This result suggests that the asymmetry of nucleotides surrounding the Box II/Box III palindrome influences the binding of protein factor(s). The asymmetry may be mandatory for the proper interaction with Box I, the other *cis*-acting element of the unit conferring light responsiveness.

How are differential effects of related *cis*-acting elements controlled by *trans*-acting factors?

Box II contains a palindromic sequence, as is often observed to be the case in *cis*-acting elements. It has been interpreted as a reflection of the fact that DNA-binding proteins are in many cases active as homodimers or tetramers, binding symmetrically to DNA matching the symmetry of the binding site (Takeda *et al.* 1983).

On the other hand, the functional core of Box II is asymmetrical. In animal systems it was hypothesized that the formation of heterodimers between different polypeptides constituting *trans*-acting factors may increase dramatically the combinatorial possibilities for regulation (Busch and Sassone-Corsi, 1990), thereby offering one possible explanation of asymmetric binding sites. If this concept holds true for plant *cis*- and *trans*-acting elements, it might explain the finding of the asymmetrical core region of Box II and the results of the replacement experiments with Box II and Box III.

An interesting observation was made when the heptameric sequence defined as the critical core was compared with some other published sequences of *cis*-acting elements (Fig. 11). This heptameric sequence is present not only in the G Box (Giuliano *et al.* 1988), but also in the *cis*-acting element called *hex* in histone gene promoters (Tabata *et al.* 1989) and in the abscisic acid responsive element (ABRE) defined in *rab* gene promoters (Marcotte *et al.* 1989; Mundy *et al.* 1990). This family of related elements may be named common plant regulator (CPR).

The data collected so far indicate a diverse class of similar *cis*-acting elements and cognate binding factors that are involved in the regulation of various promoters. In fact, the occurrence of the CPR sequence in genes controlled by anaerobic conditions, developmental state (e.g. *adh* or *patatin*, respectively; see Schulze-Lefert *et al.* 1989*a*) or in the promoter of a parsley polyubiquitin gene (P. Kawalleck, I. Somssich, K. Hahlbrock and B. Weißhaar, unpublished data) strongly supports the argument in favour of an involvement of this family of *cis*-acting elements and cognate binding factors in promoter activation by several different kinds of stimulus.

It remains to be shown whether the results of the point-mutation analysis of Box

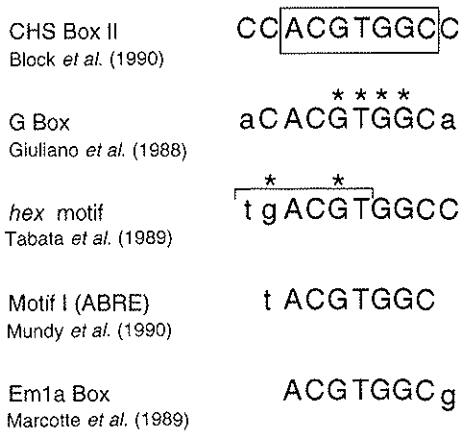

Fig. 11. Sequence elements identical to the functional core of Box II. The boxed region in the Box II sequence indicates the critical core. In the sequence surrounding the *hex* motif the nucleotides defined as the hexamer are marked. The asterisk indicates nucleotide positions that were mutated simultaneously and used as controls in the references given. In addition to the heptameric sequences identical to the Box II core there are also others matching 6 out of 7 positions (see for example Box III, Fig. 10)

II are transferable to the CPR family in a more general fashion. Interestingly, the mutations introduced into the *hex* motif or the *Arabidopsis adh*-promoter G Box, which abolish factor binding and/or *cis*-activation, all change the heptameric sequence at one position at least (Tabata *et al.* 1989; Donald *et al.* 1990). In the parsley minimal CHS promoter the specificity is brought about by a *cis*-acting Unit (Box I plus Box II), which is sufficient for light regulation. A sequence called I Box and the G Box are both important for expression from the *Arabidopsis rbcS-1A* promoter (Donald and Cashmore, 1990). In the case of the ABA-responsive promoter of the wheat *Em* gene a second element (Em2) was identified in addition to Em1a in the 50-basepair region shown to confer ABA responsiveness on a truncated 35S promoter (Marcotte *et al.* 1989). Also, the artificial combination of regulatory *cis*-acting elements from the *RBCS-3A* gene and from a heat shock gene created a novel specificity, namely a light-dependent heat shock response (Strittmatter and Chua, 1987).

The co-action of two *cis*-acting elements as one unit (and the synergistic action of two *cis*-active units), together with the hypothesis of heterodimer formation of *trans*-acting factors binding to the class of elements similar to Box II, may explain the experimental results obtained for the light response of the parsley CHS gene promoter and serve as a working hypothesis.

We and others have so far not detected differences in the binding of nuclear factors to the Box II/G Box sequence between extracts prepared from light-treated *versus* untreated or dark-adapted plant sources (Armstrong, G. A., Weißhaar, B. and Hahlbrock, K., unpublished, Giuliano *et al.* 1988). One possible

explanation of these findings is that the gel retardation assay used in these studies is not sensitive to important differences between the bound proteins. Either the differences were lost during extract preparation or modified versions of the same binding activity differ functionally. The question of how Box II is involved in light-dependent gene regulation remains to be answered. The cloning of factor(s) binding to this or related sequences would help to address this question.

We thank R. Töpfer for the vector pRT103, and I. Somssich, P. Schweizer, and O. da Costa e Silva for critically reading the manuscript. This material is based in part upon work supported by the North Atlantic Treaty Organisation under a grant awarded in 1989 to G. A. and by a fellowship of the Fritz Thyssen Stiftung to P. S.-L.

References

BALLAS, N., ZAKAI, N., FRIEDBERG, D. AND LOYTER, A. (1988). Linear forms of plasmid DNA are superior to supercoiled structures as active templates for gene expression in plant protoplasts. *Pl. molec. Biol.* **11**, 517–527.

BLOCK, A., DANGL, J. L., HAHLBROCK, K. AND SCHULZE-LEFERT, P. (1990). Functional borders, genetic fine structure, and distance requirements of cis elements mediating light responsiveness of the parsley chalcone synthase promoter. *Proc. natn. Acad. Sci. U.S.A.* **87**, 5387–5391.

BUSCH, S. J. AND SASSONE-CORSI, P. (1990). Dimers, leucine zippers and DNA-binding domains. *Trends Genet.* **6**, 36–40.

CHAPPELL, J. AND HAHLBROCK, K. (1984). Transcription of plant defense genes in response to UV light or fungal elicitor. *Nature* **311**, 76–78.

CHURCH, G. M. AND GILBERT, W. (1984). Genomic sequencing. *Proc. natn. Acad. Sci. U.S.A.* **81**, 1991–1995.

DANGL, J. D., HAHLBROCK, K. AND SCHELL, J. (1989). Regulation and structure of chalcone synthase genes. In *Cell structure and somatic cell genetics of plants*, vol. 6 (ed. J. U. Vasil and J. Schell), pp. 155–173. Academic Press, New York.

DANGL, J. D., HAUFFE, K. D., LIPPHARDT, S., HAHLBROCK, K. AND SCHEEL, D. (1987). Parsley protoplasts retain differential responsiveness to UV light and fungal elicitor. *EMBO J.* **6**, 2551–2556.

DAVIS, R. W., SIMON, M. AND DAVIDSON, N. (1971). Electron microscope heteroduplex methods for mapping regions of base sequence homology in nucleic acids. *Methods Enzymol.* **21**, 413–428.

DELISLE, A. J. AND FERL, R. J. (1990). Characterisation of the *Arabidopsis Adh* G-Box binding factor. *Pl. cell* **2**, 547–557.

DIXON, R. A. (1986). The phytoalexin response: Elicitation, signaling and control of host gene expression. *Biol. Rev.* **61**, 239–291.

DONALD, R. G. K. AND CASHMORE, A. R. (1990). Mutation of either G Box or I Box sequences profoundly affects expression from the *Arabidopsis rbcS-1A* promoter. *EMBO J.* **9**, 1717–1726.

DONALD, R. G. K., SCHINDLER, U., BATSCHAUER, A. AND CASHMORE, A. R. (1990). The plant G box promoter sequence activates transcription in Saccharomyces cerevisiae and is bound *in vitro* by a yeast activity similar to GBF, the plant G box binding factor. *EMBO J.* **9**, 1727–1735.

EBEL, J. AND HAHLBROCK, K. (1982). Biosynthesis. In *The Flavonoids* (eds. Harborne, J. B. and Mabry, T. J.), pp. 641–675. Chapmann and Hall, London.

EPPING, B., KITTEL, M., RUHNAU, B. AND HEMLEBEN, V. (1990). Isolation and sequence analysis of a chalcone synthase cDNA of *Matthiola incana* R. *Br. (Brassicaceae). Pl. molec. Biol.* **14**, 1061–1063.

FANG, R.-X., NAGY, F., SIVASUBRAMANIAM, S. AND CHUA, N.-H. (1989). Multiple cis regulatory

elements for maximal expression of the cauliflower mosaic virus 35S promoter in transgenic plants. *Pl. Cell* **1**, 141–150.

GIULIANO, G., PICHERSKY, E., MALIK, V. S., TIMKO, M. P., SCOLNIC, P. A. AND CASHMORE, A. R. (1988). An evolutionary conserved protein binding sequence upstream of a plant light-regulated gene. *Proc. natn. Acad. Sci. U.S.A.* **85**, 7089–7093.

HAHLBROCK, K. AND GRISEBACH, H. (1979). Enzymatic controls in the biosynthesis of lignin and flavonoids. *Rev. Pl. Physiol.* **30**, 105–130.

HAHLBROCK, K., KREUZALER, F., RAGG, H., FAUTZ, E. AND KUHN, D. N. (1982). Regulation of flavonoid and phytoalexin accumulation through mRNA and enzyme induction in cultured plant cells. In *Biochemistry of Differentiation and Morphogenesis* (ed. L. Jaenicke) pp. 34–43. Springer, Berlin.

HAHLBROCK, K. AND SCHEEL, D. (1989). Physiology and molecular biology of phenylpropanoid metabolism. *Annu. Rev. Plant Physiol. Plant Mol. Biol.* **40**, 347–369.

HARBORNE, J. B. AND TURNER, B. L. (1984). *Plant Chemosystematics*. Academic Press, London.

HELLER, W. AND HAHLBROCK, K. (1980). Highly purified 'flavanone synthase' from parsley catalyzes the formation of naringenin chalcone. *Archs Biochem. Biophys.* **200**, 617–619.

HERRMANN, A., SCHULZ, W. AND HAHLBROCK, K. (1988). Two alleles of the single-copy chalcone synthase gene in parsley differ by a transposon-like element. *Molec. gen. Genet.* **212**, 93–98.

JEFFERSON, R. J. (1987). Assaying chimaeric genes in plants: The GUS gene fusion system. *Pl. molec. Biol. Rep.* **5**, 387–405.

KATAGIRI, F., LAM, E. AND CHUA, N.-H. (1989). Two tobacco DNA-binding proteins with homology to the nuclear factor CREB. *Nature* **340**, 727–729.

KOZAK, M. (1983). Comparison of initiation of protein synthesis in prokaryotes, eucaryotes, and organelles. *Microbiol. Rev.* **47**, 1–45.

KREUZALER, F., RAGG, H., FAUTZ, E., KUHN, D. N. AND HAHLBROCK, K. (1983). UV-induction of chalcone synthase mRNA in cell suspension cultures of *Petroselinum hortense*. *Proc. natn. Acad. Sci. U.S.A.* **80**, 2591–2593.

LAM, E. AND CHUA, N.-H. (1990). GT-1 binding site confers light responsive expression in transgenic tobacco. *Science* **248**, 471–474.

LIPPHARDT, S., BRETTSCHNEIDER, R., KREUZALER, F., SCHELL, J. AND DANGL, J. D. (1988). UV-inducible transient expression in parsley protoplasts identifies regulatory *cis*-elements of a chimeric *Antirrhinum majus* chalcone synthase gene. *EMBO J.* **7**, 4027–4033.

MARCOTTE, W. R., JR, RUSSELL, S. H. AND QUATRANO, R. S. (1989). Abscisic acid-responsive sequences from the *Em* gene of wheat. *Pl. Cell* **1**, 969–976.

MATERN, U., HELLER, W. AND HIMMELSPACH, K. (1983). Conformational changes of apigenin 7-*O*-(6-*O*-malonylglucoside), a vacuolar pigment from parsley, with solvent composition and proton concentration. *Eur. J. Biochem.* **133**, 439–448.

MUNDY, J., YAMAGUCHI-SHINOZAKI, K. AND CHUA, N.-H. (1990). Nuclear proteins bind conserved elements in the abscisic acid-responsive promoter of a rice *rab* gene. *Proc. natn. Acad. Sci. U.S.A.* **87**, 1406–1410.

NICK, H. AND GILBERT, W. (1985). Detection *in vivo* of protein-DNA interactions within the *lac* operon of *Escherichia coli*. *Nature* **313**, 795–798.

NIESBACH-KLÖSGEN, U., BARTIN, E., BERNHARDT, J., ROHDE, W., SCHWARZ-SOMMER, Z., REIF, H. J. AND SAEDLER, H. (1987). Chalcone synthase genes in plants: A tool to study evolutionary relationships. *J. molec. Evol.* **26**, 213–225.

OHL, S., HAHLBROCK, K. AND SCHÄFER, E. (1989). A stable blue-light-derived signal modulates ultraviolet-light-induced activation of the chalcone-synthase gene in cultured parsley cells. *Planta* **177**, 228–236.

REIMOLD, U., KRÖGER, M., KREUZALER, F. AND HAHLBROCK, K. (1983). Coding and 3′ non-coding nucleotide sequence of chalcone synthase mRNA and assignment of amino acid sequence of the enzyme. *EMBO J.* **2**, 1801–1805.

SCHEEL, D., DANGL, J. L., DOUGLAS, C., HAUFFE, K. D., HERRMANN, A., HOFFMANN, H., LOZOYA, E., SCHULZ, W. AND HAHLBROCK, K. (1987). Stimulation of phenylpropanoid pathways by environmental factors. In *Plant Molecular Biology* (ed. von Wettstein, D. and Chua, N.-H.) pp. 315–316, Plenum Press.

SCHMELZER, E., JAHNEN, W. AND HAHLBROCK, K. (1988). *In situ* localization of light-induced

chalcone synthase mRNA, chalcone synthase, and flavonoid end products in epidermal cells of parsley leaves. *Proc. natn. Acad. Sci. U.S.A.* **85**, 2989–2993.

SCHRÖDER, J. AND SCHRÖDER, G. (1990). Stilbene and chalcone synthases: Related enzymes with key functions in plant-specific pathways. *Z. Naturforsch.* **45c**, 1–8.

SCHULZE-LEFERT, P., BECKER-ANDRE, M., SCHULZ, W., HAHLBROCK, K. AND DANGL, J. L. (1989*b*). Functional architecture of the light-responsive chalcone synthase promoter from parsley. *Pl. Cell* **1**, 707–714.

SCHULZE-LEFERT, P., DANGL, J. L., BECKER-ANDRE, M., HAHLBROCK, K. AND SCHULZ, W. (1989*a*). Inducible *in vivo* DNA footprints define sequences necessary for UV light activation of the parsley chalcone synthase gene. *EMBO J.* **8**, 651–656.

STRITTMATTER, G. AND CHUA, N.-H. (1987). Artificial combination of two cis-regulatory elements generates a unique pattern of expression in transgenic plants. *Proc. natn. Acad. Sci. U.S.A.* **84**, 8986–8990.

TABATA, T., TAKASE, H., TAKAYAMA, S., MIKAMI, K., NAKATSUKA, A., KAWATA, T., NAKAYAMA, T. AND IWABUCHI, M. (1989). A protein that binds to a *cis*-acting element of wheat histone genes has a leucine zipper motif. *Science* **245**, 965–967.

TAKEDA, Y., OLLENDORF, D. H., ANDERSON, W. F. AND MATTHEWS, B. W. (1983). DNA-binding proteins. *Science* **221**, 1020–1026.

TÖPFER, R., PRÖLS, M., SCHELL, J. AND STEINBIß, H. H. (1988). Transient gene expression in tobacco protoplasts: II. Comparision of the reporter gene systems for CAT, NPT II, and GUS. *Pl. Cell Rep.* **7**, 225–228.

VAN DE LÖCHT, U., MEIER, I., HAHLBROCK, K. AND SOMSSICH, I. (1990). A 125 bp promoter fragment is sufficient for strong elicitor-mediated gene activation in parsley. *EMBO J.* **9**, 2945–2950.

WEIßHAAR, B., ARMSTRONG, G. A., BLOCK, A., DA COSTA E SILVA, O. AND HALLBROCK, K. (1991). Light-inducible and constitutively expressed DNA-binding proteins recognizing a plant promoter element with functional relevance in light responsiveness. *EMBO J.*, in press.