The EMBO Journal vol.7 no.7 pp.1929–1936, 1988

Both positive and negative regulatory elements mediate expression of a photoregulated CAB gene from *Nicotiana plumbaginifolia*

Carmen Castresana[1,2,5], Isabel Garcia-Luque[1,2,3], Elena Alonso[1,2], Vedpal S.Malik[1,4] and Anthony R.Cashmore[2]

[1]Laboratory of Cell Biology, The Rockefeller University, 1230 York Avenue, New York, NY 10021, and [2]Plant Science Institute, Department of Biology, University of Pennsylvania, Philadelphia, PA 19104, USA

Present addresses: [3]IATA CSIC Valencia, Spain, [4]Phillips Morris Research Center, Richmond, VA, USA and [5]Laboratorium voor Genetica, Rijksuniversiteit Gent, B-9000 Gent, Belgium

Communicated by M.van Montagu

We have analyzed promoter regulatory elements from a photoregulated CAB gene (*Cab-E*) isolated from *Nicotiana plumbaginifolia*. These studies have been performed by introducing chimeric gene constructs into tobacco cells via *Agrobacterium tumefaciens*-mediated transformation. Expression studies on the regenerated transgenic plants have allowed us to characterize three positive and one negative *cis*-acting elements that influence photoregulated expression of the *Cab-E* gene. Within the upstream sequences we have identified two positive regulatory elements (PRE1 and PRE2) which confer maximum levels of photoregulated expression. These sequences contain multiple repeated elements related to the sequence -ACCGGCCCACTT-. We have also identified within the upstream region a negative regulatory element (NRE) extremely rich in AT sequences, which reduces the level of gene expression in the light. We have defined a light regulatory element (LRE) within the promoter region extending from −396 to −186 bp which confers photoregulated expression when fused to a constitutive nopaline synthase ('*nos*') promoter. Within this region there is a 132-bp element, extending from −368 to −234 bp, which on deletion from the *Cab-E* promoter reduces gene expression from high levels to undetectable levels. Finally, we have demonstrated for a full length *Cab-E* promoter conferring high levels of photoregulated expression, that sequences proximal to the *Cab-E* TATA box are not replaceable by corresponding sequences from a '*nos*' promoter. This contrasts with the apparent equivalence of these *Cab-E* and '*nos*' TATA box-proximal sequences in truncated promoters conferring low levels of photoregulated expression.

Key words: chlorophyll *a*/*b* binding protein genes/promoter analysis/chimeric constructs/transgenic plants/regulatory elements

Introduction

CAB polypeptides bind chlorophyll *a* and *b* and exist as a complex in the thylakoid membranes of chloroplasts. This complex absorbs light and the resulting excitation energy is transferred to photosystems I and II (Glazer, 1983; Thornber, 1985). CAB proteins are encoded by a family of nuclear genes in all plant species examined (Corruzi *et al.*, 1983; Dunsmuir *et al.*, 1983; Karlin-Neumann *et al.*, 1985; Pichersky *et al.*, 1985; Leutwiler *et al.*, 1986; Castresana *et al.*, 1987). They are synthesized in the cytosol as soluble precursors and then imported into the chloroplasts, where they are cleaved to their mature form (Apel and Kloppstech, 1978; Cumming and Bennett, 1981; Schmidt *et al.*, 1981).

Photoregulated expression of CAB genes has been shown to be transcriptionally regulated. This process is mediated via the photoreceptor phytochrome as indicated by red light induction and reversion by far red light (Apel, 1979; Cumming and Bennett, 1981; Thompson *et al.*, 1983; Tobin, 1981; Viro and Kloppstech, 1982). It has been demonstrated that the regulatory elements responsible for these expression characteristics are localized within the 5′ non-coding region of these genes. Chimeric constructs containing bacterial reporter genes fused to the 5′-flanking region of certain CAB genes have been introduced into tobacco and petunia plants (Lamppa *et al.*, 1985; Simpson *et al.*, 1985; Nagy *et al.*, 1986). In this manner these constructs have been shown to confer photoregulated expression.

The existence of *cis*-acting regulatory elements localized in distinct promoter regions has been described for light-regulated genes (Kuhlemeier *et al.*, 1987a). Analysis of two different CAB genes, isolated from pea (AB 8.0) (Simpson *et al.*, 1986) and wheat (*Cab-1*) (Nagy *et al.*, 1987), has revealed the presence of a positive light regulatory element within the first 400 bp of the promoter. Within this same promoter region a 'silencer' element has been described for AB 8.0 (Simpson *et al.*, 1986). Further upstream sequences are necessary to confer maximum levels of transcription to the pea AB 8.0 gene (Simpson *et al.*, 1985). However, these upstream sequences have not been characterized in any detail, in contrast to the more extensive studies carried out with the promoter sequences extending from −500 to −100 bp from the cap site of these genes.

We are interested in defining the different promoter elements that are involved in mediating the photoregulated expression of CAB genes. In pursuing these interests we have prepared for the *Cab-E* gene from *N.plumbaginifolia* (Castresana *et al.*, 1987) a series of 5′-end and internal promoter deletions. Expression studies on transgenic plants containing chimeric constructs prepared from these deletions have allowed us to characterize multiple regulatory elements that are present in distinct regions of the promoter.

Results

Construction of promoter cloning vectors

To facilitate the characterization of promoter regulatory sequences, we have constructed two plasmids designated

©IRL Press Limited, Oxford, England

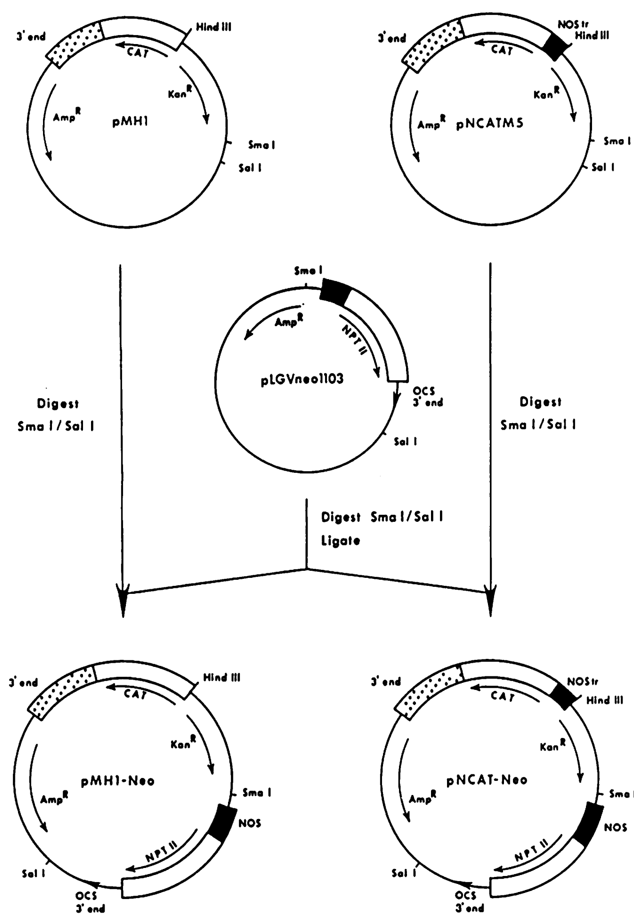

Fig. 1. Construction of intermediate cloning vectors. The plasmid pLGV1103 was restricted with *SmaI/SalI*. The fragment containing the NPT(II) gene fused to the nopaline synthase promoter and the 3′ end of the octopine synthase gene, was purified from an agarose gel and ligated to pMH-1 and pNCATM5 (previously restricted with *SmaI/SalI* enzymes).

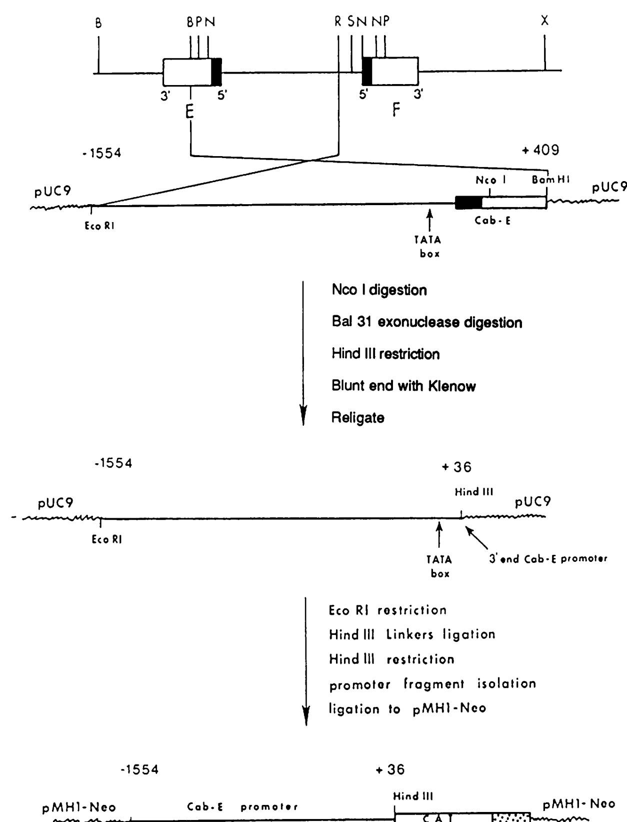

Fig. 2. Construction of the *Cab-E* promoter−CAT fusion. An *EcoRI/BamHI* DNA fragment containing 409 bp of the coding region of the *Cab-E* gene and a 5′-flanking fragment extending to −1554 bp from the cap site of the gene (Castresana *et al.*, 1987), was subcloned in pUC9. The resulting plasmid was linearized at the internal *NcoI* restriction site, digested with *Bal*31 exonuclease and then restricted with *HindIII*. The ends were repaired with the Klenow fragment of DNA polymerase I and then the plasmid was recircularized. The plasmids with the *HindIII* site restored were analyzed for the extent of the 3′ deletions by polyacrylamide gel sizing. Plasmids were sequenced from the *HindIII* site to define the end point of the deletion. A plasmid containing a 3′ deletion to position +36 from the cap site of the gene was selected, restricted with *EcoRI* and the ends repaired with Klenow. After ligation to *HindIII* linkers followed by restriction with *HindIII*, the promoter fragment was purified from an agarose gel and cloned into the unique *HindIII* site of pMH1-Neo. Correct orientation in pMH1-Neo was determined by *PstI/EcoRI* double digest.

pMH1-Neo and pNCAT-Neo (Figure 1), by cloning a kanamycin resistance marker into the *SmaI*−*SalI* restriction sites of plasmids pMH1 and pNCATM5 (Timko *et al.*, 1985). pMH1 contains the coding region of the bacterial gene chloramphenicol acetyl transferase (CAT) to which we had previously fused several promoter fragments from a pea RBCS (small subunit of ribulose 1,5-bisphosphate carboxylase) gene and conferred light-regulated expression in transformed plant tissue. As pMH1 does not contain any marker to select for transformed plants, we isolated a kanamycin gene under control of the '*nos*' promoter (nopaline synthase) from pLGV1103 (Simpson *et al.*, 1985), and cloned this fragment into pMH1.

The pNCAT-Neo plasmid was constructed in a similar manner. It contains a truncated '*nos*' promoter (extending to −145 bp 5′ of the mRNA cap site) fused to the coding region of the CAT gene. pNCAT-Neo can be used to test whether promoter regulatory sequences lacking a TATA box can confer light-regulated expression to a heterologous promoter that normally mediated constitutive expression.

Mapping Cab-E transcripts

S1 protection experiments were carried out to define the cap site for the transcripts derived from the *Cab-E* gene used in this study. A 392-bp *NcoI/DraI* fragment was 5′ end-labeled in the *NcoI* site (147 bp into the coding region of *Cab-E*). The fragment was hybridized for 15 h at 45°C to 25 μg of total RNA isolated from *N.plumbaginifolia* and then treated with S1 nuclease. A DNA protected fragment of 206 bp was obtained (results not shown) defining the cap site at 59 bp upstream from the first ATG of the coding region. Some S1-resistant products of ~180−185 bp were also obtained. These fragments are presumed to result from hybridization of the probe to transcripts derived from the other CAB genes present in the *N.plumbaginifolia* genome. These CAB genes must diverge from *Cab-E* in the region 33−38 bp upstream from the ATG codon.

Construction of a Cab-E promoter−CAT fusion and analysis of transgenic plants

We prepared a chimeric construct by fusing the 5′-flanking region from the *N.plumbaginifolia Cab-E* gene to the coding region of the CAT gene.

To isolate a promoter sequence we carried out *Bal*31 digestion from the coding region of *Cab-E* (Figure 2) and

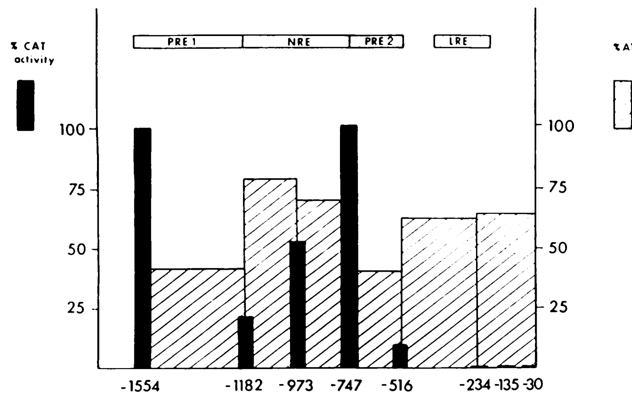

Fig. 3. 5′ End deletion analysis of the *Cab-E* promoter. Promoter deletions were analyzed by fusion to the coding region of the bacterial CAT gene and subsequent analyses for CAT activity in transgenic plants. The 5′ end point and the promoter strength assigned to each deletion is indicated. The data represent the average CAT activity determined for 10 independent transformed plants, and are expressed as a relative percentage of the activity determined for the wild-type promoter (deletion −1554 bp). The percentage of AT sequences corresponding to the different promoter deletions are also presented. The regulatory elements PRE1, PRE2, NRE and LRE are defined in the text.

selected a fragment extending from −1554 to +36 bp. A chimeric construct was prepared by inserting this promoter fragment into pMH1-Neo. This was then transferred to *Agrobacterium* and introduced into tobacco cells by leaf disc transformation as described in Materials and methods.

We initially examined seven independent transgenic plants to evaluate variability in CAT activity, NPT II activity and gene copy number (results not shown). We observed that CAT activity varied between individual transgenic plants. Variation in NPT II activity was not so acute, due possibly to the kanamycin selection. We estimated that the number of copies for both genes (CAT and NPT II) ranged from one to 10 for individual plants. Little correlation was observed between CAT activity and either gene copy number or NPT II activity. To accommodate the differences in CAT activity, we have assayed 10 independent transformant plants for each construct and we have reported the average results.

Distribution of regulatory sequences in the 5′-flanking region of Cab-E

To identify regulatory sequences within the *Cab-E* promoter we have prepared a series of chimeric constructs by fusing different 5′ promoter deletions to the CAT gene contained in pMH1-Neo. The deletions were obtained either by *Bal*31 exonuclease digestion, or by using convenient restriction sites within the promoter. The nature of the deletion as well as the activity determined for each construct is shown in Figure 3.

Chimeric constructs which contain promoter fragment extending either to −1554 or to −747 confer equal levels of CAT activity to transgenic plants; these levels correspond to the maximum determined. The activity decreases at least 5-fold when the sequences located between −1554 and −1182 are deleted. Further deletions from −1182 to −973 and to −747 result in an increase of promoter strength of 2.5- and 5-fold respectively, relative to the −1182 construct. These results suggest the presence of a positive regulatory element(s) within the promoter region lying from −1554 to −1182 (PRE1), and a negative regulatory element(s) in the region extending from −1182 to −747 (NRE). Activity is decreased at least 10-fold when the sequences between −747 to −516 are deleted, suggesting the presence in this region of a second positive regulatory element(s) (PRE2). The deletion to −516 still contains sequences with ability to direct low levels of CAT gene expression. Further deletions to −234, −135 or −30, result in a complete loss of detectable CAT activity.

To assess in more detail the significance of the promoter elements identified, we prepared a series of promoter internal deletions and analyzed these by plant transformation. Figure 4C shows that, somewhat independent of the 5′ end of the *Cab-E* promoter fragment utilized, deletion of internal sequences lying from −516 to −234 bp causes almost a complete loss of CAT activity (constructs 3 and 7). Even more striking, deletion of sequences located between −368 and −234 bp produced a complete loss of CAT activity (construct 4). In contrast, when upstream sequences extending from −747 to −516 bp are deleted, only a small reduction is observed in the level of activity conferred by the promoter fragment extending to −973 bp (construct 2). We note that construct 2 confers significantly higher levels of CAT activity than does construct 8. This result appears to be a contradiction of our definition of sequences between −973 and −747 as comprising part of a NRE. We suggest that the resolution of this apparent inconsistency is that there resides within the NRE, sequences that act as positive regulatory elements but that the effect of these sequences only becomes apparent in the absence of both PRE1 and PRE2. Our 5′ deletion studies (Figure 3) clearly indicate that in the presence of PRE2 the overall effect of sequences between −973 and −747 is that of a NRE element.

Sequence analysis of the Cab-E promoter

We previously reported the complete sequence of *Cab-E* including its promoter extending to −921 bp upstream from the cap site of the gene (Castresana *et al.*, 1987). We show here the sequence of the complete 5′-flanking region used in this study, extending to −1554 bp from the cap site of the gene (Figure 5).

To further characterize DNA sequences which might be responsible for regulatory characteristics of the different promoter regions defined, we have analyzed the base composition of the complete 5′-flanking region studied. The complete DNA sequence examined is 62% AT rich with the distribution of nucleotides being very asymmetric (Figure 3). It is observed that the two upstream regions between −1554 to −1182 and −747 and −516 containing PRE1 and PRE2 are the only regions within the promoter which contain a high percentage of GC bases (Figure 3). Interestingly, the promoter region extending from −747 to −516 (PRE2) contains the sequence CCCAC repeated six times. The same sequence is repeated five times in the complementary strand of the positive element located between −1554 and −1182 bp (PRE1). Some of these repeated sequences share a more extensive homology reaching a maximum of 12 nt in the sequence -ACCGGCCCACTT- (Figure 5 and Table I). The heptanucleotide CCGGCCC contained in this sequence is repeated a total of seven times in the two regulatory elements.

In contrast to the positive elements PRE1 and PRE2, the upstream negative element (NRE) is extraordinarily rich in AT bases with the sequences between −1182 and −973 containing 83% (Figure 3).

Cab-E promoter confers light-regulated expression

To determine whether the various promoter mutants confer light-regulated expression to the CAT gene, as well as to define the point of initiation of transcription, S1 mapping analysis was performed on total RNA prepared from transformed plants grown either in the light or dark (Figure 4A).

Plants growing under a 14 h light/10 h dark photoperiod were transferred to the dark for 4 days and then placed for 24 h under continuous white light. No CAT mRNA was

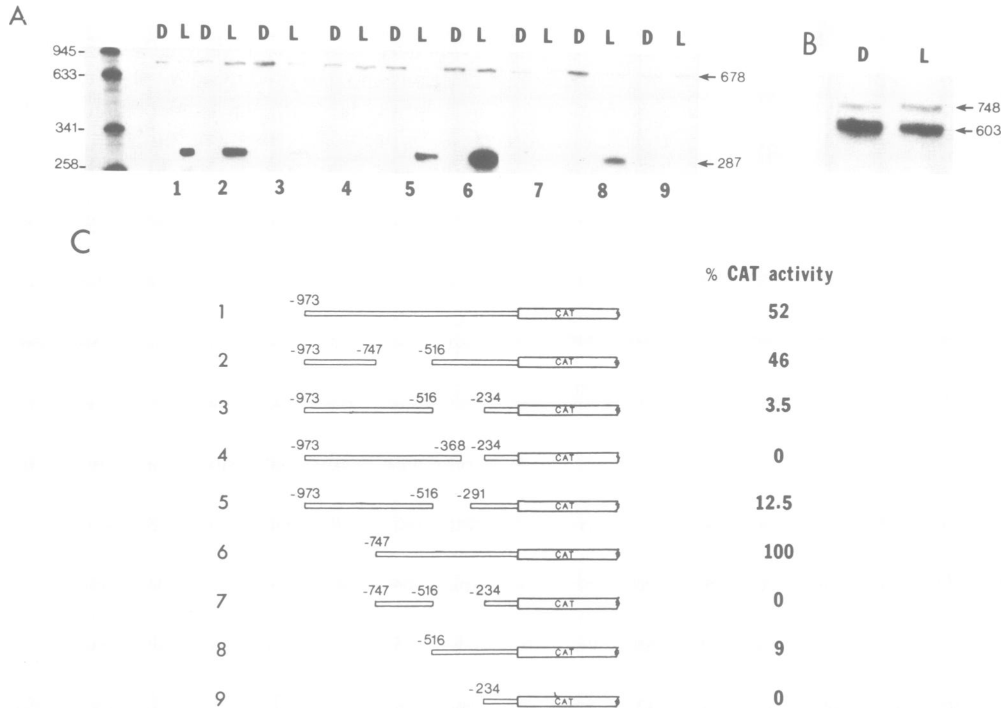

Fig. 4. Analysis of light-regulated expression conferred to the CAT gene by the different *Cab-E* promoter deletion mutants. **Panel A**. S1 nuclease digestion of 75 μg of total RNA obtained from transformed plants placed for 4 days in the dark (D) and then returned to the light for 24 h (L). A 5′ end-labeled probe containing 240 bp of the CAT coding region and extending to −391 bp in the *Cab-E* promoter was utilized. The size of the protected fragments was estimated by reference to the mobility of the plasmid pUC18 restricted with *Sau*3A and end-labeled using the Klenow fragment of DNA polymerase I (first lane on the left). **Panel B**. S1 nuclease digestion of 100 μg of total RNA obtained from transformed plants containing construct number 1. A 748-bp 5′ end-labeled DNA fragment containing 565 bp of the coding region of the NPT II gene (neomycin phosphotransferase) and extending to −145 bp in the '*nos*' (nopaline synthase) promoter was utilized as a probe. **Panel C**. Diagram of the deletion mutants. The CAT activity is expressed as described in the legend to Figure 5.

```
-1554
CCAGGTGAACCCGTAACTAGTTTGTTGTATTCGCCTCTGCTGATAGGTGATTATAACCTCCTCAAAAGCAATTTCTACTCCATTTCACCTATAAAATAATCAGAAAACTTAAGTTATATA   -1435

CATCAGTAAATAAATTTTACACCATAAGATAAAAATTGCTTTCGCAACGGTTAAGGGTGGCATGTGGGCCGGGGCCGGTTCTAAGTGGGCTTTACGGGCCCGATCCTAAGTAGGCCCAGT   -1315

CCTAAGTGGGCCGGTCCTAAGTGGTCCCGGGTTTCGCAGGCTTCTTGTTGTAATCGGCCTTGGATTGGGACCACGAACTAACGGTCCCGGGTTAAGTGGGCCGGTCCCGGGCCTAAGTGG   -1195
          -1182
GCCCAACGAATACTTTCTATTTTTTAAAAATAATTTATAGAAGTTAGAGAAAAAAAAATGAAAATAAAAATATTTAAGGCAATTCCTTGTAAATTATATTATAGAATTGTGACCTAAATTT   -1075
                                                                                                            -973
TTTAATTCAAATTTAAAGATAAAAATATTGTAAAGAGGTATTCAAAGCAATGTGTTATATATATATATATATATATACTAAGTGTATAGTATATAAGCTATAATTATATATATCTTAAGA   -955

TGTATATATAGTATTATAGTATAGTATAGTAATCTTAACATGTATATATAGCTATAAAAAGTATGGGGTTAAAACAAAGTTGGGAAAGGTTATTTTATAAATTGCCAACGGCTATTTTAG   -835
                                                                                          -747
CAGGTAAAACGGCCATATTTTAAATGCCATAACGGCTATAATGTGGCAGATTTATTTTTTAAAAAAACTAACCGTTGGGCCCGAATAGGCCTTTTTAGGACCGCTTGAACCGGCCCACTT   -715

CCCAGCCGGTCCCGGTCTCGCGGGCCTCGCCTATGGAACCGACCCACTACCCAGCCCACCTCCCCACGGTCCCGATCCTATTCGGTTAGAACCGTCTAGGCCCACCGCCCATTTGGGCTT   -595
                                                                                    -516
GCGGTCTTGGGCCGGACCTGAACCTAACCGGCCCACATGCCATCCTTACTAACGGTAAATAACTTAGAAGTTATTGTATACGTATGATCGAGCTGTTGGACTTGTAGTATCAAACTTTCA   -475

ATGACGCATCAAAATTAATTATGGTAGCTTCGCGTTGGGACACTTGTACATGCATTAACTTGATTTCAATTTCTTTTTTAAAAATATTTGTCTATTGTCAATTTACCACTCGTACTTGAA   -355

GTGGGCCTATTTGACAGGTCAGCTAAATACAGAAGTGTATGAACAATGCGTGGCCAAGAGTAACTCTTATGCTAAAGACAAGTGGATATTATATTGCAATAATCCACAATCAGACGTGGC   -235
-234                                                                                                     -135
AAATTTGGATTGGCTATAAGAGAGCAAATCTTCATTAGGTAAGTTTTTTAAACATAAAAAGTATCTAAAAAAAATCTTGTCATGTTTAACGGTGCTGAACTTTGCCAAATGGACAAGAATG   -115
                                                                                        -30
CAAAAGGTTAAAATTGCAATCCACCAATTGAAAAGTAGATATAGATACTCAAGGATAAGGGTCTTTGGGCCTGTAAAGCCATTTATATACACTTAGTGCAAAGCCCATGAAACTCAAGCC   +6

TCAAATCAACTCTTTCTTTTTGTGCATTCAAGAGTTATCATTTTACTCCTACA
```

Fig. 5. Nucleotide sequence of the 5′-flanking region of *Cab-E* gene. Nucleotides are numbered with the cap site designated +1. TATA and CAAT boxes are underlined. Direct and inverted repeat sequences represented in Table I are also underlined. Sequences overlined represent sequences similar to boxes II and III (Kuhlemeier *et al.*, 1987b) and the G box (ACGTGGCA; Giuliano *et al.*, 1988); see text. 5′ end points of the promoter deletions analyzed are indicated.

detected after the 4-day dark period. However, when the plants were returned to the light, CAT mRNA was detected in those plants for which we had previously detected CAT enzymatic activity (Figure 4, panel A and C). The mRNA increase was in proportion to the level of CAT activity previously determined.

We observed for all the chimeric CAT constructs examined, the same initiation point for transcription (Figure 4A). The S1 protection experiments showed the presence of an S1-resistant product of ~287 bp which would correspond to a cap site 26 bp downstream from the TATA box, in the same region identified for the wild-type *Cab-E* gene.

In these studies we included as a control a 5′ end-labeled probe for transcripts derived from the NPTII gene. This gene is fused to the '*nos*' promoter and, as expected, the level of expression is approximately the same in dark-grown and light-grown tissue (Figure 4B).

Table I. Multiple repeated sequences present in the two upstream positive regulatory elements (PRE1 and PRE2)

−1360	C	–	–	–	–	–	–	–	–	–	A	–	−1373
−1354	–	–	–	–	–	–	–	–	C	G	G	C	−1362
−1339	T	A	A	A	–	–	–	–	–	–	–	–	−1352
−1298	–	–	–	–	–	–	–	–	–	–	–	–	−1311
−1208	–	–	–	–	–	–	–	–	–	–	–	–	−1221
−1187	T	T	G	–	–	–	–	–	–	–	–	–	−1200
−726	A	C	C	G	G	C	C	C	A	C	T	T	−715
−678	–	–	–	–	A	–	–	–	–	–	–	A	−667
−665	C	–	–	A	–	–	–	–	–	–	C	–	−654
−657	–	–	–	T	C	–	–	–	–	–	G	G	−646
−619	C	T	A	–	–	–	–	–	–	–	C	G	−608
−577	T	–	–	–	–	–	–	–	–	A	G	A	−590
−568	–	–	–	–	–	–	–	–	–	–	A	–	−557

A dash represents homology.

Fusion of Cab-E promoter sequences to a constitutive 'nos' promoter

To evaluate whether the CAB promoter sequences could confer light-regulated expression to a heterologous promoter that normally mediates constitutive expression, a DNA fragment extending from −396 to −186 bp from the cap site of *Cab-E* was cloned into the unique *Hin*dIII restriction site of the plasmid pNCATNeo. This plasmid contains a truncated '*nos*' promoter fused to the coding region of the CAT gene. We also fused to the same truncated '*nos*' promoter a large *Cab-E* promoter fragment obtained by *Bal*31 deletion and extending from −1554 to −112 bp. This large fragment was also fused to a *Cab-E* truncated promoter extending from −135 to +36 bp; this construct essentially restores the wild-type −1554 bp *Cab-E* promoter with the addition of a small duplication of the sequence −135 to −112 (Figure 6D).

Expression in transformed plants was determined by assaying for CAT activity and by S1 analysis on total RNA. Figure 6D shows that no CAT activity is detected when the CAT gene is under the control of either the '*nos*' or *Cab-E* truncated promoters (Figure 6D, constructs 1 and 4). However, activity is observed when either of the two *Cab-E* promoter sequences (−396 to −186 and −1554 to −112) were fused to the truncated '*nos*' promoter (constructs 2 and 3).

Transformed plants containing the −396 to −186 *Cab-E* promoter, showed no detectable CAT mRNA when grown in the dark. When these plants were placed in the light, significant levels of CAT mRNA were detected (lane 2). These results demonstrated that this *Cab-E* promoter fragment acts as a light regulatory element (LRE) in that

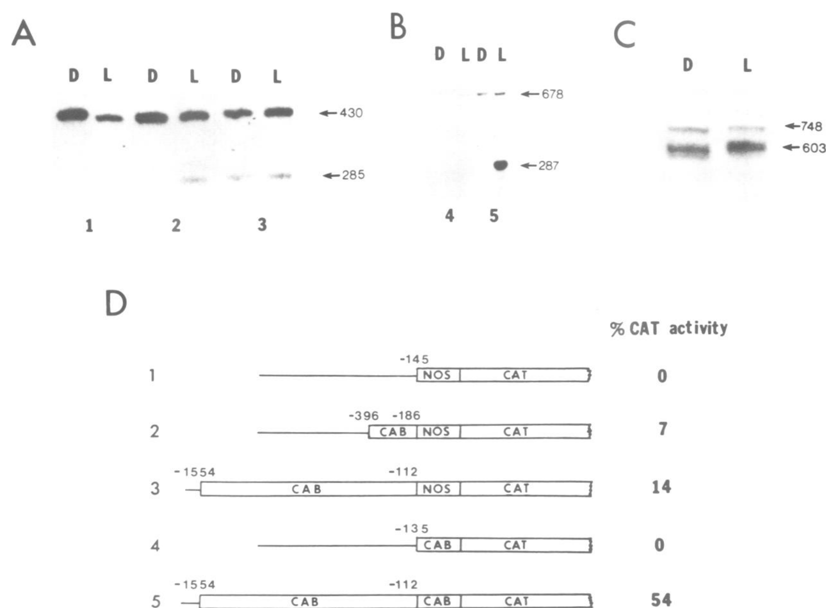

Fig. 6. Analysis of light-regulated expression of chimeric constructs containing different promoter fusions. **Panel A**. S1 nuclease digestion of 100 μg of total RNA isolated from transformed plants containing constructs 1, 2 or 3. The probe utilized contained 240 bp of the coding region of the CAT gene, and extended to −145 bp in the '*nos*' promoter. **Panel B**. S1 nuclease digestion of total RNA isolated from transformed plants containing constructs 4 or 5. The probe utilized was the same as described in **A**. **Panel C**. S1 nuclease digestion of total RNA isolated from transformed plants containing construct 5. The probe utilized was the same as described in **B**. **Panel D**. Diagram of the promoter fusions analyzed. The CAT activity assigned is expressed as described in the legend to Figure 5. Light-regulated expression was assayed using RNA obtained from transformed plants placed for 4 days in the dark (D), and then for 24 h in the light (L). The size of the protected fragments was determined as described in the legend to Figure 4.

it confers photoregulated expression to the '*nos*' truncated promoter.

Plants containing the larger *Cab-E* promoter fragment (−1554 to −112) fused to the truncated *Cab-E* promoter similarly showed no detectable CAT mRNA in the dark (lane 4), in contrast to the levels found when these plants were returned to the light (lane 5). Of particular interest, plants containing the same *Cab-E* promoter fragment fixed to the truncated '*nos*' promoter, showed significant levels of CAT mRNA when grown in the dark (lane 3). Equivalent level of expression was maintained when these plants were transferred to the light (lane 3). Thus, whereas the large (−1554 to −112) *Cab-E* promoter fragment confers the expected photoregulated expression when fused to the truncated CAB promoter, the same fragment confers constitutive expression when fused to the truncated '*nos*' promoter.

In these studies we also included as a control a 5′ end-labeled probe for transcripts derived from the NPT II gene fused to the '*nos*' promoter (Figure 6C).

Discussion

The 5′-flanking region of the *Cab-E* promoter from *N.plumbaginifolia* has been used to define promoter elements responsible for photoregulated expression. In these studies we have defined multiple positive and negative regulatory elements that determine both the level and the nature of the *Cab-E* gene expression.

Positive regulatory elements exist far upstream in the Cab-E promoter

Within the upstream sequences we have identified two positive regulatory elements (PRE1 and PRE2) which confer maximum levels of expression. 5′ Promoter deletions which eliminate either PRE1 or both PRE1 and PRE2 produce a reduction in gene expression of 5- and 10-fold respectively.

The existence of upstream sequences that exert an increase in the level of expression have previously been noticed for a CAB gene from pea (Simpson *et al.*, 1985). In contrast, our results differ from those recently reported by Nagy *et al.* (1987) using a CAB gene from wheat. Within this gene, promoter sequences extending to −357 bp from the cap site, are sufficient to confer maximum levels of photoregulated expression in transgenic tobacco plants.

The two upstream positive elements identified here (PRE1 and PRE2), are relatively GC-rich sequences containing multiple repeated elements related to the sequence -ACCGGCCCACTT- (Table I). In view of their predominance these sequences are likely to contribute to the regulatory characteristics of these positive elements. Somewhat related sequences have been found in other promoters in regions known to enhance the level of gene expression. For instance, the embryo-specific gene encoding the α' subunit of β-conglycinin from soybean plants contains five 6-bp repeats (A$\overset{A}{\underset{C}{G}}$CCCA) within a 100-bp promoter fragment, the presence of which raises the level of gene expression at least 20-fold (Chen *et al.*, 1986). Similarly, the 35S promoter of the CaMV (cauliflower mosaic virus) contains the sequence CCAC, and its complement GTGG, repeated a total of 10 times within a region which confers maximum level of expression (Kay *et al.*, 1987; Ow *et al.*, 1987). Furthermore, the GC-rich repeated sequence that we have described here, is similar to the GC-rich box CCGCCC found in a number of animal viral and cellular promoters, including the SV40 early promoter, where this sequence is repeated six times (Benoist and Chambon, 1981). This sequence is the binding site for the Sp1 factor which plays an important role in activation of transcription (Kadonaga *et al.*, 1986).

PRE1 is a positive regulatory element which, in our studies, affects the expression of *Cab-E*. However, we note that in the genome of *N.plumbaginifolia* from whence these sequences were derived, it is more likely that the primary effect of PRE1 is on *Cab-F* (Figure 2). This conclusion is based simply on the fact that PRE1 is closer to *Cab-F* (~800 bp) than it is to *Cab-E* (~1400 bp). Thus it is likely that PRE1 and PRE2 are structurally and functionally related PREs acting on divergently oriented genes. In accordance with this suggestion, is the divergent orientation of the repeated sequence found in the two elements.

Somewhat irrespective of the precise regulatory role of PRE1 and PRE2 in the *N.plumbaginifolia* genome, in the experiments that we have reported here it is quite clear that both elements affect the level of expression mediated by the *Cab-E* promoter.

That PRE1 and PRE2 contain divergently oriented repeated sequences (Table I) presumed to contribute to the regulatory characteristics of these elements, is suggestive that these elements are enhancer sequences. We believe this possibility is likely to be correct and it can be readily tested. However, in relation to this suggestion we do note that at least in the content of the *Cab-E* promoter, these elements are not acting as general enhancer sequences. This conclusion is based on the observation that the high level of CAT activity observed for the −973 truncated *Cab-E* promoter containing PRE2, is totally lost when the LRE fragment between −368 to −234 is deleted. Thus in this construct, fusion to PRE2 to a truncated *Cab-E* promoter yields an inactive promoter. Whether or not this dependence of the PRE2 sequences on the LRE sequences is a unique characteristic of the *Cab-E* promoter is not presently known. This can be tested by fusion of the positive regulatory elements to other truncated promoters.

A negative regulatory element exists far upstream in the Cab-E promoter

We have also identified within the upstream promoter region a NRE lying between the two positive elements PRE1 and PRE2. A 5′ deletion extending to −1182 bp which eliminates PRE1, produces a 5-fold reduction in the level of gene expression, while deletion of the negative element located between −1182 and −747 bp increases exponentially 5-fold, restoring maximum levels of expression.

The NRE described for *Cab-E* is extraordinarly rich in AT sequence. AT-rich sequences have been described for yeast promoters (Struhl, 1985). However, in this case the sequences, which exist as poly(dA−dT), were identified as positive elements necessary for constitutive expression. Whether the regulatory characteristics of the *Cab-E* NRE are largely a reflection of the high AT content is not known at this point.

Plant NREs have been described for a CAB gene from pea (Simpson *et al.*, 1986) and recently for a RBCS gene, also from pea (Kuhlemeier *et al.*, 1987b). Both genes are light-regulated, but the negative elements, which are located

between −347 and −100 bp from the cap site of the genes, have been characterized by their ability to silence the expression of constitutive promoters either in roots (Simpson *et al.*, 1986) or in leaf tissue when placed in the dark (Kuhlemeier *et al.*, 1987b). The negative element that we describe here for the *Cab-E* promoter, reduces the level of gene expression in leaf tissue in the light, as demonstrated by an enhancement of expression when this element is deleted. This negative element also differs from previously described elements, with respect to its far-upstream location.

With the NRE, as with the PREs, we have not addressed directly the question of whether or not these sequences can confer photoregulated expression. However, we believe it more likely that the far-upstream positive and negative regulatory elements simply serve to modulate levels of expression.

A positive element that mediates light-regulation

Positive light regulatory elements (LREs) have previously been characterized for light-regulated CAB and RBCS genes from pea. DNA sequences which promote photoregulated expression when fused to constitutive truncated promoters have been localized within promoter regions extending from −400 to −100 bp from the cap site in the genes examined (Fluhr *et al.*, 1986; Simpson *et al.*, 1986; Kuhlemeier *et al.*, 1987b; Nagy *et al.*, 1987).

Several lines of evidence indicated that a similar LRE resides within the *Cab-E* promoter. With the 5′ deletion experiments, we describe a truncated *Cab-E* promoter extending to −516 bp which directs low levels (9% of maximum) of CAT gene expression in light-grown tissue. Within this promoter fragment, we have defined a 132-bp element (extending from −368 to −234 bp) which on deletion from the −973 bp *Cab-E* promoter, reduces gene expression from high levels to undetectable levels. Furthermore, when a fragment from the *Cab-E* promoter, extending from −396 to −186, was fused to a truncated *nos* promoter, photoregulated expression was conferred.

By comparative analysis of the 5′-flanking region from different RBCS genes we have revealed the presence of a conserved sequence of ~100 bp located between −427 and −178 bp from the first ATG codon. Within these promoter regions the sequence -ACGTGGCA- is highly conserved. We have also shown that both tomato and *Arabidopsis* plants contain a nuclear protein factor that binds to this sequence (Giuliano *et al.*, 1988). This same sequence -ACGTGGCA- is located within the *Cab-E* promoter in a similar relative position (−241 bp from the cap site) and within a region we have demonstrated to be essential for gene expression. Kuhlemeier *et al.* (1987b) have characterized conserved sequences present in boxes I, II, II*, III and III*, in the promoters of pea RBCS genes. Sequences similar to boxes II and III are present immediately upstream of the G box sequence in the tobacco *Cab-E* gene (Figure 5). Similarly, in pea RBCS genes, box II* occurs immediately upstream from the G box, prompting us to suggest that factors binding to these sequences may interact (Giuliano *et al.*, 1988). This suggestion is strengthened by our observation that these sequences are similarly juxtaposed in the *Cab-E* gene. Furthermore, the fact that the G box and sequence II (and related sequences) are present within LREs of both CAB and RBCS genes, suggests a fundamental similarity in the regulation of expression of these genes. Consistent with this thought is our observation that the factor that binds to the G box in the tomato *RbcS-3A* gene, apparently also binds to the G box in the LRE fragment from the *Cab-E* gene, as binding to the *Cab-E* fragment is competed by a synthetic *RbcS-3A* G box oligomer (U.Schindler, unpublished observation). These data suggest that the G box sequences in the two genes may indeed be functionally equivalent.

Sequences proximal to the Cab-E TATA box have distinct functional characteristics

As discussed, in order to assess whether the element −396 to −186 could confer light regulation to a constitutive promoter, we fused this element to a truncated '*nos*' promoter (extending to −145 bp 5′ of the mRNA cap site). The fusion promoter gave rise to levels of CAT activity (7% of maximum activity) in light-grown plants which was very near that found for a similarly truncated *Cab-E* promoter (the *Cab-E* promoter truncated to −516 bp gave rise to 9% of maximum CAT activity). Furthermore, the fusion promoter conferred photoregulated expression. In these experiments, sequences within the truncated '*nos*' promoter behave in a manner not significantly distinct from the corresponding sequences within the *Cab-E* promoter.

In striking contrast to these results, when a large *Cab-E* promoter element (−1554 to −112) was fused to the truncated '*nos*' promoter the resulting fusion promoter displayed functional characteristics quite distinct from those observed when the same element was fused to a similarly truncated *Cab-E* promoter (−135 to +36). This *Cab−nos* fusion promoter showed only 14% of maximum CAT activity, contrasting with the 54% activity observed for the corresponding *Cab−Cab* promoter fusion, and with the 100% maximum activity observed for the −1554 bp 'wild-type' *Cab-E* promoter. Furthermore, by analysis of mRNA this *Cab−nos* fusion promoter was shown to promote expression in the dark at levels similar to that found in the light, while the *Cab−Cab* fusion promoter directed the expected photoregulated expression.

From the above results we conclude that there must exist sequences in the truncated *Cab-E* promoter (−135 to +36), lacking in the truncated '*nos*' promoter, that are required to mediate the high levels of expression promoted by the upstream PREs. Furthermore, we conclude that there are also sequences in the truncated *Cab-E* promoter, again lacking in the truncated '*nos*' promoter, which in the presence of the upstream sequences are required for photoregulated expression. These sequences may reflect unique positive, or possibly negative, elements in the truncated *Cab-E* promoter. Alternatively, they might simply reflect an absence of some sequences in the '*nos*' promoter. In the latter context we note that the '*nos*' promoter lacks both a conventional TATA and CAAT box (Ebert *et al.*, 1987). In the former context, we note that we have previously characterized a sequence GATA which commonly resides as 2−3 repeats between the CAAT box and TATA box of type I PSII CAB genes (Castresana *et al.*, 1987). What role, if any, these sequences might play in mediating high levels of photoregulated expression of CAB genes is presently unknown.

An additional conclusion from the experiments with the fusion promoters is that there reside sequences in the large promoter fragment (−1554 to −112), lacking in the smaller *Cab-E* promoter fragment (−396 to −186), which on fusion

to the '*nos*' promoter mediate expression in the dark as well as in the light. We believe that these sequences are probably the same PRE1 and PRE2 sequences that mediate high levels of photoregulated expression in both the full-length 'wild-type' *Cab-E* promoter, and the *Cab – Cab* fusion promoter. As discussed, the fact that these sequences do not mediate expression in the dark on fusion to the truncated *Cab-E* promoter, must reflect a significant difference between the two truncated promoters.

Materials and methods

Construction of chimeric constructs

Chimeric constructs were prepared using standard DNA procedures described by Maniatis *et al.* (1982). Plasmid DNA was prepared according to Birnboim and Doly (1979). *Escherichia coli* JM83 and HB101 were used for all the '*in vitro*' DNA transformation experiments.

DNA sequencing

The sequence of the *Cab-E* promoter was determined by the method of Maxam and Gilbert (1980). The structure of all chimeric constructs was confirmed by partial sequencing using the same procedure.

Ti-mediated transfer

Intermediate cloning vectors containing different chimeric constructs were transferred to *A.tumefaciens* harboring the Ti plasmid pGV3850 (Zambryski *et al.*, 1983), by triparental mating (Van Haute *et al.*, 1983). The structures of the co-integrates in *Agrobacterium* were analyzed by the method of Dhaese *et al.* (1979). *Agrobacteria* containing the co-integrates, were used to transfer the chimeric constructs to *Nicotiana tabacum* SR1 cells by leaf disc transformation (Horsch *et al.*, 1985). Transformed discs were maintained in MS medium (Murashige and Skoog, 1962) containing hormones to stimulate shoot formation (1 mg/l 6-BAP and 0.1 mg/l α NAA) and kanamycin sulfate (100 μg/ml) to select transformants. Transformed shoots were selected after rooting in MS medium containing kanamycin sulfate (50 μg/ml), and transferred to soil.

Analysis of transformed plants

Transformed plants growing under a 14 h light/10 h dark photoperiod were analyzed between 7 and 8 h after the dark period. Leaves of ~10 cm were harvested for analysis ~4 weeks after transfer of the plants to soil.

Neomycin phosphotransferase (NPT II) activity was measured according to Reiss *et al.* (1984). Chloramphenicol acetyl transferase (CAT) activity was determined as described by An (1986). In both cases 50 μg of plant protein was used per assay. For Southern blot experiments (Southern, 1975), DNA from transformed plants was isolated according to the procedure described by Dellaporta *et al.* (1983).

For total RNA preparations, leaves were ground to a fine powder in liquid nitrogen using a pestle and mortar. Guanidinium buffer (5 M guanidinium thiocyanate, 25 mM sodium citrate, 0.5% sarcosyl, 2 mM EDTA, 1 M β-mercaptoethanol and 50 mM Tris – HCl pH 7.6) was then added at a ratio of 4 ml/g of fresh tissue. The supernatant was clarified by centrifugation for 10 min at 5000 *g*, and extracted with an equal volume of phenol/chloroform. After vortexing the solution was centrifuged for 30 min at 10 000 *g*. The upper phase was removed and centrifuged again for 10 min at 5000 *g*. The nucleic acids were precipitated by addition of 0.1 vol 3 M NaOAc pH 5.6 and two vols of ethanol and resuspended in sterile water. RNA was then precipitated in 2 M LiCl overnight at 4°C. The insoluble RNA was pelleted by centrifugation for 10 min at 10 000 *g*, washed with 70% ethanol and resuspended in sterile water. After a second precipitation with NaOAc and ethanol, the RNA was finally dissolved in water and kept in aliquots at −80°C. Yields of 1 mg of RNA per gram of fresh tissue were routinely obtained. Total plant RNA was used to carry out S1 mapping analysis as previously described (Cashmore, 1984).

Acknowledgements

We thank N.Hoffman, T.Ueda, E.Pickersky and R.Donald for critical reading of this manuscript. This work was supported by grants to A.R.C. from the Department of Agriculture and the Department of Energy. C.C. was supported by a NATO post-doctoral fellowship.

References

An,G. (1986) *Plant Physiol.*, **81**, 86 – 91.
Apel,K. (1979) *Eur. J. Biochem.*, **97**, 183 – 188.
Apel,K. and Kloppstech,K. (1978) *Eur. J. Biochem.*, **85**, 581 – 588.
Benoist,C. and Chambon,P. (1981) *Nature*, **290**, 304 – 310.
Birnboim,H.C. and Doly,J. (1979) *Nucleic Acids Res.*, **7**, 1513 – 1523.
Cashmore,A.R. (1984) *Proc. Natl. Acad. Sci. USA*, **81**, 2960 – 2964.
Castresana,C., Staneloni,R., Malik,V.S. and Cashmore,A.R. (1987) *Plant Mol. Biol.*, **10**, 117 – 126.
Chen,Z.-L., Schuler,M.A. and Beachy,R.N. (1986) *Proc. Natl. Acad. Sci. USA*, **83**, 8560 – 8564.
Coruzzi,G., Boglie,R., Cashmore,A.R. and Chua,N.-H. (1983) *J. Biol. Chem.*, **258**, 1399 – 1402.
Cumming,A.C. and Bennett,J. (1981) *Eur. J. Biochem.*, **118**, 71 – 80.
Dellaporta,S.L., Wood,J. and Hicks,J.B. (1983) *Plant Mol. Biol. Rep.*, **1**, 19 – 21.
Dhaese,P., De Greve,H., Decraemer,H., Schell,J. and van Montagu,M. (1979) *Nucleic Acids Res.*, **7**, 1837 – 1849.
Dunsmuir,P., Smith,S.M. and Bedbrook,J. (1983) *J. Mol. Appl. Genet.*, **2**, 285 – 300.
Ebert,P.R., Ha,S.B. and An,G. (1987) *Proc. Natl. Acad. Sci. USA*, **84**, 5745 – 5749.
Fluhr,R., Kuhlemeier,C., Nagy,F. and Chua,N.-H. (1986) *Science*, **232**, 1106 – 1112.
Glazer,A.N. (1983) *Annu. Rev. Biochem.*, **52**, 125 – 128.
Giuliano,G., Pichersky,E., Malik,D.S., Timko,M.P., Scolnik,T.A. and Cashmore,A.R. (1988) *Proc. Natl. Acad. Sci. USA*, in press.
Horsch,R.B., Fry,J.E, Hoffman,N.L., Eichholtz,D., Rogers,S.D. and Fraley,R.T. (1985) *Science*, **227**, 1229 – 1231.
Kadonaga,J.T., Jones,K.A. and Tjian,T. (1986) *Trends Biochem. Sci.*, **11**, 20 – 23.
Karlin-Neumann,G.A., Kohorn,B.D., Thornber,J.P. and Tobin,E.M. (1985) *J. Mol. Appl. Genet.*, **3**, 45 – 61.
Kay,R., Chan,A., Daly,M. and McPherson,J. (1987) *Science*, **32**, 1299 – 1302.
Kuhlemeier,C., Green,P. and Chua,N.-H. (1987a) *Annu. Rev. Plant Physiol.*, **37**, 221 – 257.
Kuhlemeier,C., Fluhr,R., Green,P. and Chua,N.-H. (1987b) *Genes Dev.*, **1**, 247 – 255.
Lamppa,G., Nagy,F. and Chua,N.-H. (1985) *Nature*, **316**, 750 – 752.
Leutwiler,L.S., Meyerowitz,E.M. and Tobin,E.M. (1986) *Nucleic Acids Res.*, **14**, 4051 – 4076.
Maniatis,T., Fritsch,E.F. and Sambrook,J. (1982) *DNA Cloning. A Laboratory Manual*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
Maxam,A.M. and Gilbert,W. (1980) *Methods Enzymol.*, **65**, 499 – 560.
Murashige,T. and Skoog,F. (1962) *Physiol. Plant.*, **15**, 473 – 497.
Nagy,F., Kay,S.A., Boutry,M., Hsu,M.-Y. and Chua,N.-H. (1986) *EMBO J.*, **5**, 1119 – 1124.
Nagy,F., Boutry,M., Hsu,M.-Y., Wong,M. and Chua,N.-H. (1987) *EMBO J.*, **6**, 2537 – 2542.
Ow,D.W., Jacobs,J.D. and Howell,S.H. 1987) *Proc. Natl. Acad. Sci. USA*, **84**, 4870 – 4874.
Pichersky,E., Bernatzky,R., Tanksley,S.D., Breidenbach,R.B., Kausch, A.P. and Cashmore,A.R. (1985) *Gene*, **40**, 247 – 258.
Reiss,B., Sprengel,R., Will,H. and Schaller,H. (1984) *Gene*, **30**, 211 – 218.
Schmidt,G.W., Bartlett,S.G., Grossman,A.R., Cashmore,A.R. and Chua,N.-H. (1981) *J. Cell Biol.*, **91**, 468 – 478.
Simpson,J., Timko,M.P., Cashmore,A.R., Schell,J., van Montagu,M. and Herrera-Estrella,L. (1985) *EMBO J.*, **4**, 2723 – 2729.
Simpson,J., Schell,J., van Montagu,M. and Herrera-Estrella,L. (1986) *Nature*, **323**, 551 – 554.
Southern,E.M. (1975) *J. Mol. Biol.*, **98**, 503 – 517.
Struhl,K. (1985) *Proc. Natl. Acad. Sci. USA*, **82**, 8419 – 8423.
Thornber,J.P. (1985) *Annu. Rev. Plant Physiol.*, **26**, 127 – 158.
Thompson,W.F., Everett,M., Polans,N.O., Jorgensen,R.A. and Palmer,J.D. (1983) *Planta*, **158**, 487 – 500.
Timko,M.P., Kaush,A.P., Castresana,C., Fassler,J., Herrera-Estrella,L., Van den Broeck,G., van Montagu,M., Schell,J. and Cashmore,A.R. (1985) *Nature*, **318**, 579 – 582.
Tobin,E.M. (1981) *Plant Mol. Biol.*, **1**, 35 – 51.
Van Haute,E., Joos,H., Maes,M., Warren,G., van Montagu,M. and Schell,J. (1983) *EMBO J.*, **2**, 411 – 418.
Viro,M. and Kloppstech,K. (1982) *Plant*, **154**, 18 – 23.
Zambryski,P., Joos,H., Genetello,C., Leemans,J., van Montagu,M. and Schell,J. (1983) *EMBO J.*, **2**, 2143 – 2150.

Received on December 12, 1987; revised on April 5, 1988