The EMBO Journal vol.9 no.11 pp.3415-3427, 1990

Photoregulated gene expression may involve ubiquitous DNA binding proteins

Ulrike Schindler and Anthony R.Cashmore

Plant Science Institute, Department of Biology, University of Pennsylvania, Philadelphia, PA 19104, USA

Communicated by M.van Montagu

Several promoter elements have previously been shown to influence the expression of the *cab-E* gene in *Nicotiana plumbaginifolia*. Here we demonstrate, by electrophoretic mobility shift and methylation interference assays, that a complex pattern of protein-DNA interactions characterizes this promoter. Among the multiple proteins identified, we focused on five different factors which either occupied important regulatory elements and/or were present in relatively large amounts in nuclear extracts. All of these proteins were distinguished on the basis of their recognition sequence and other biochemical parameters. One, GBF, interacted with a single sequence within the *cab-E* promoter homologous to the G-box found in many photoregulated and other plant promoters. A second factor, GA-1, bound to the GATA element which is located between the CAAT and TATA boxes of the *cab-E* and all other LHCII Type I CAB promoters. GA-1 also interacted *in vitro* with the I-boxes of the *Arabidopsis rbcS-1A* promoter and the as-2 site of the CaMV 35S promoter. Two other factors, GC-1 and AT-1, bound to multiple recognition sites localized within the GC-rich and AT-rich elements, respectively. GT-1, a protein which interacts with promoters of other light-regulated genes, bound to seven distinct sites distributed throughout the *cab-E* promoter.

Key words: CAB promoter/DNA binding motifs/light regulation/transacting factors: GBF, GA-1, GT-1, AT-1, GC-1

Introduction

The transcriptional control of gene expression commonly depends on an interplay between multiple sequence-specific DNA binding proteins and their cognate promoter elements (see reviews by Johnson and McKnight, 1989; Mitchell and Tjian, 1989). Based on studies in yeast and mammalian systems, it has been suggested that promoters may be composed of multiple elements or modules, each contributing to what may often be a complex pattern of transcriptional regulation (Dynan, 1989; Jones *et al.*, 1988). This modular organization explains the presence of the same distinct element within multiple unrelated promoters. It also allows a gene to be regulated in response to several different stimuli, as in the case of the metallothionein promoters (Lee *et al.*, 1987). Moreover, a large number of expression patterns can be generated by unique combinations of relatively few regulatory elements.

Recent work indicates that plant promoters are also composed of different modules. Several regulatory elements have been identified using either transgenic plants or transient expression assays (reviewed by Benfey and Chua, 1989; Kuhlemeier *et al.*, 1987a; Schell, 1987; Willmitzer, 1988). Considerable effort is now underway to characterize DNA binding proteins which specifically interact with these elements.

The transcription of light-regulated plant genes, including those encoding chlorophyll a/b binding proteins (CAB) and ribulose-1,5-bisphosphate carboxylase small subunit (RBCS), is regulated by both light quality and quantity (for review see Tobin and Silverthorne, 1985). This transcriptional control is likely to be mediated by specific interaction of nuclear factors with *cis*-acting promoter sequences. Nuclear proteins interacting with conserved DNA motifs present in promoters of photoregulated genes have been described. GT-1, a factor isolated from pea nuclei, binds to multiple motifs, designated box II, II*, II**, III, III* and III**, within the pea *rbcs-3A* promoter (Green *et al.*, 1987 and 1988). Several studies imply that GT-1 may be involved in both positive and negative regulation of the *rbcs-3A* gene (Kuhlemeier *et al.*, 1987b, 1988a and 1989; Strittmatter and Chua, 1987).

Another plant DNA binding protein that has been characterized is GBF, a protein present in tomato and *Arabidopsis* nuclei, which interacts with the G-box motif found in many RBCS promoters (Giuliano *et al.*, 1988a). Mutation of this binding site in the context of the *Arabidopsis rbcS-1A* promoter dramatically reduces the expression of a reporter gene in transgenic tobacco plants (Donald and Cashmore, 1990). A similar sequence, indispensable for UV-light induction of the parsley chalcone synthase gene, is bound *in vivo* by a nuclear protein (Schulze-Lefert *et al.*, 1989). *In vitro* binding experiments with the chalcone synthase promoter of *Antirrhinum majus* led to the identification of a homologous element occupied by a protein which was designated CG-1 (Staiger *et al.*, 1989). The relationship between GBF and CG-1, which interact with almost identical DNA motifs, is presently unknown. G-box-like elements are not restricted to photoregulated genes. An identical DNA motif, present in the promoter of the *Arabidopsis* alcohol dehydrogenase gene (*adh*), is bound *in vitro* by a nuclear protein and *in vivo* binding studies suggest that this element is only occupied in transcriptionally active tissue (McKendree *et al.*, 1990). Another protein, AT-1, has been characterized and shown to interact with AT-rich sequences within promoters of several light-regulated genes. An interesting feature of AT-1 is that binding of the factor to DNA is modulated by phosphorylation (Datta and Cashmore, 1989). Multiple other plant nuclear proteins interacting with AT-rich promoter regions have been identified (de Brujin *et al.*, 1989; Deikman and Fischer, 1988; Holdsworth and Laties, 1989a and b; Jensen *et al.*, 1988; Jofuku *et al.*, 1987; Metz *et al.*, 1988; Stougaard *et al.*, 1987). The relationship between these various nuclear

proteins that bind to AT-rich promoter sequences has not been determined.

In continuing our studies on the mechanism of transcriptional control of CAB gene expression we have turned our attention to characterizing the nature of the interactions between nuclear proteins and promoter sequences. In addition to light-induction, the CAB genes have been shown to be controlled by endogenous stimuli causing circadian oscillation in transcription (Giuliano *et al.*, 1988b; Nagy *et al.*, 1988) and furthermore the expression of these genes is developmentally regulated in a tissue specific manner (Fluhr and Chua, 1986; Fluhr *et al.*, 1986; Simpson *et al.*, 1985 and 1986). This complex expression pattern could result from the integration of different signals mediated by a combination of several nuclear proteins interacting with distinct DNA modules.

We have previously reported for the *cab-E* gene of *Nicotiana plumbaginifolia*, deletion studies which revealed several *cis*-acting promoter elements influencing the expression of a reporter gene in transgenic tobacco plants. A light regulatory element residing between −396 and −186 bp, as well as more distal positive and negative regulatory elements, were identified (Castresana *et al.*, 1988). By electrophoretic mobility shift assays (Fried and Crother, 1981; Garner and Revzin, 1981) and methylation interference experiments, we have now characterized multiple DNA binding factors interacting with this *cab-E* promoter. Five different proteins have been distinguished based on competitive binding studies. Four of these factors were shown to interact with multiple sites within the promoter. No pronounced differences in DNA binding activity were observed for any of these factors between extracts isolated from light-grown and dark-adapted plants. Furthermore, many of these factors appear to be closely related, if not identical, to factors that mediate the expression of non-photoregulated genes. Models involving common factors in the regulation of light-induced gene expression are discussed. Both the number of proteins, and the multiplicity of binding sites for the individual factors, provide a picture of photoregulated gene expression that is substantially more complex than hitherto described.

Results

General strategy

In order to investigate the interaction of nuclear proteins with various regulatory elements of the *cab-E* promoter, we subcloned DNA fragments, ranging in size from 30 to 200 bp, and subjected them to electrophoretic mobility shift assays using tobacco, tomato and pea nuclear extracts (Fried and Crother, 1981; Garner and Revzin, 1981). DNA−protein complexes were obtained with all promoter segments leading to the identification of multiple DNA binding activities. Little variation was observed between tomato and tobacco extracts, however some of those protein−DNA interactions were not detected in extracts prepared from pea nuclei. In this report we focused on five different factors which we distinguished on the basis of their DNA recognition sequences. These factors were chosen either because they specifically interacted with promoter elements shown to be important by expression studies employing site-directed mutagenesis (P.Pringmann and A.R.Cashmore, in preparation; Donald and Cashmore, 1990; Gidoni *et al.*, 1989) or because they were particularly abundant. Where appropriate, binding studies with promoter fragments derived from genes other than the *cab-E* gene have been included.

GBF binds to a G-box-like sequence within the cab-E promoter

The *cab-E* promoter fragment A14 (Figure 1A) contains an 8 bp motif similar to the G-box element previously shown to bind the nuclear factor GBF (Giuliano *et al.*, 1988a) and to be required for expression from the *Arabidopsis rbcS-1A* promoter (Donald and Cashmore, 1990). In protein binding studies with fragment A14 and tomato nuclear extracts, we observed a DNA−protein complex which was specifically competed by an oligonucleotide derived from the tomato *rbcS-3A* G-box (Figure 1B, lane 3), but not by a mutant derivative (lane 4) or oligonucleotides and promoter fragments containing other protein binding sites (lanes 5−8). These data imply that GBF also interacts with the *cab-E* promoter *in vitro*. Furthermore, the methylation interference pattern performed with fragment A14 (Figure 1C) is strikingly similar to the one obtained with the *Arabidopsis rbcS-1A* G-box (Figure 1D; Donald *et al.*, 1990). Figure 1C shows that fragments cleaved at the G residues at positions −236, −237, −239 and −242, located within the *cab-E* G-box motif, are absent in the bound fraction and therefore critical for protein binding. Methylation of the G residues on the complementary DNA strand resulted only in partial binding inhibition (data not shown). Based on competitive binding studies, the *cab-E* element is bound less strongly than the G-box residing within the tomato *rbcS-3A* or *Arabidopsis rbcS-1A* promoters (data not shown). This weaker binding is probably due to the two mismatches within the 8 bp core recognition sequence (Figure 1D).

G-1 interacts with GATA boxes located between the TATA box and the CAAT box of the cab-E promoter

Previous studies on the *Petunia cab22R* gene (Gidoni *et al.*, 1989) and in our laboratory (P.Pringmann and A.R.Cashmore, in preparation) demonstrated that a mutation within the conserved GATA element, localized between the TATA and CAAT boxes (Castresana *et al.*, 1987; Grob and Stueber, 1987), resulted in a five-fold reduction of expression. The three GATA motifs of the *cab-E* promoter reside within fragment A7 (Figure 2A). To avoid interference by GT-1, which also bound to this fragment (see below), an oligonucleotide bearing the *cab-E* GATA element (−81 to −55) was used for binding studies (GATA321, Figure 2B). In mobility shift assays with tobacco nuclear extract, this wild-type oligonucleotide was bound very strongly, even in the presence of high concentrations of non-specific DNA (Figure 2C, lane 2). The protein complex was competed by the oligonucleotide itself (Figure 2C, lane 6), but not by any of the *cab-E* promoter fragments bearing binding sites for other proteins (data not shown). Similar results were obtained with extracts prepared from tomato and pea nuclei (data not shown).

To investigate whether all three GATA sequences were essential for this protein interaction, double mutant oligonucleotides bearing only one of the three GATA motifs (GATA--1, GATA-2-, GATA3--, Figure 2B) were included in the binding reactions. All three derivatives competed less

effectively than the wild-type oligonucleotide for GA-1 binding (Figure 2C, lanes 3–5). No competition at all was observed when a triple mutant oligonucleotide GATA***, (Figure 2B) was included in the binding reaction (Figure 2C, lane 7). From these experiments we conclude that a factor, which we designate GA-1, specifically recognizes and interacts with the GATA element and that a single GATA-box is sufficient for this interaction although the affinity increases dramatically when all three motifs are present.

GA-1 also interacts with the I-box sequences present in RBCS promoters and the as-2 site of the CaMV 35S promoter

We also investigated whether GA-1 interacted with the GATA-box-like sequences present in RBCS promoters. These motifs, designated I-boxes (Giuliano *et al.*, 1988a) or sequence 2 (Manzara and Gruissem, 1988), reside further upstream than their counterparts in CAB genes. For the *Arabidopsis rbcS-1A* promoter, it was shown that mutation of the I1 or I2 element decreased the expression of a reporter gene in transgenic tobacco plants (Donald and Cashmore, 1990). Oligonucleotides derived from the I1 and I2 motifs of the *Arabidopsis rbcS-1A* promoter (Figure 2B) were included in binding reactions with a *cab-E* wild-type GATA321 oligonucleotide as radiolabeled probe. Both I-boxes competed to the same extent as the *cab-E* double mutants bearing a single GATA motif only (Figure 2C, lanes 8 and 9). In the reverse experiment using I1 (data not shown) or I2 (Figure 2D) as radiolabeled probes, the protein–DNA complex was efficiently competed by the wild-type GATA321 oligonucleotide (lane 4), whereas the triple mutant GATA*** had no effect (lane 3). To demonstrate a requirement for the sequence GATA for GA-1 binding to the I2 oligonucleotide, an I2 mutant (I2m, Figure 2B) with a 4 bp substitution was included in the binding reaction. This oligonucleotide failed to compete with the wild-type I2

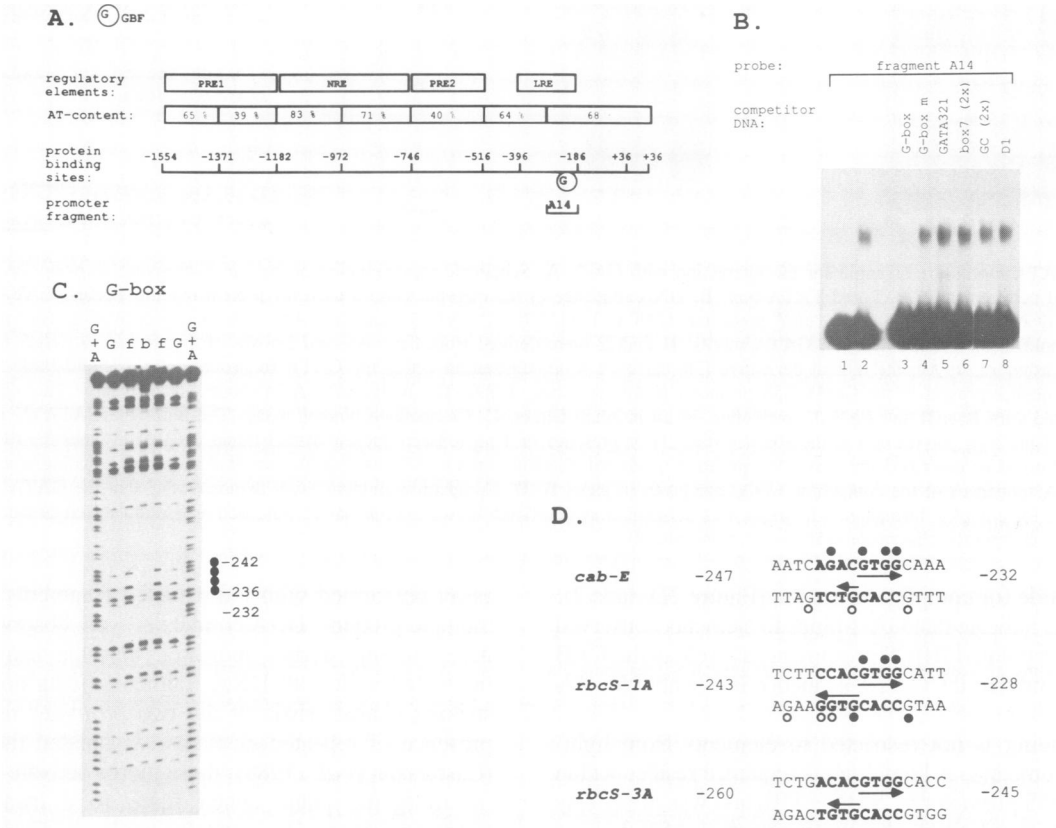

Fig. 1. GBF binds to a G-box like motif within the *cab-E* promoter. **A.** Schematic representation of the *cab-E* promoter showing the positive (PRE1 and PRE2), negative (NRE) and light regulatory (LRE) elements. The AT content of the promoter regions is given below. The G-box-like motif is located within promoter fragment A14 (−301 to −186). **B.** Competitive binding assay. Radiolabeled fragment A14 was incubated with 2 μg poly(dI.dC) in the absence (lane 1) and presence (lanes 2–8) of 2 μg tomato nuclear extract. Lanes 3–8, unlabeled promoter fragments or oligonucleotides (0.1 pmol each), containing different protein binding sites were included in the binding reactions, as indicated. The DNA sequence of the G-box oligonucleotide (lane 3) was derived from the tomato *rbcS-3A* motif (panel D), the mutant derivative (lane 4) carries a 4 bp substitution (GTGG replaced by TGTT), all other competitor DNAs (lanes 5–8) are defined in the legends to Figures 2, 3, 4 and 6. **C.** Methylation interference experiment. Fragment A14, cloned into the *Sma*I site of pUC18, was 5′ end-labeled at the *Eco*RI site. The partially methylated DNA was incubated with 15 μg tomato nuclear extract. Free (f) and protein-complexed (b) DNA bands were electrophoretically separated, eluted, and cleaved with piperidine. The cleavage products were analyzed on an 8% sequencing gel. The reactions were run in parallel with a G and G+A specific sequencing ladder of the same fragment (Maxam and Gilbert, 1980). Filled circles indicate the G residues critical for protein binding. **D.** Comparison of the *cab-E*, *Arabidopsis rbcS-1A* and tomato *rbcS-3A* G-box motifs and their methylation interference patterns. Filled circles indicate the methylguanines that completely inhibited protein binding, methylation of positions marked with open circles only partially impaired protein binding. The arrows indicate the palindromic structure of the motif. The core binding sequence of GBF is given in bold face.

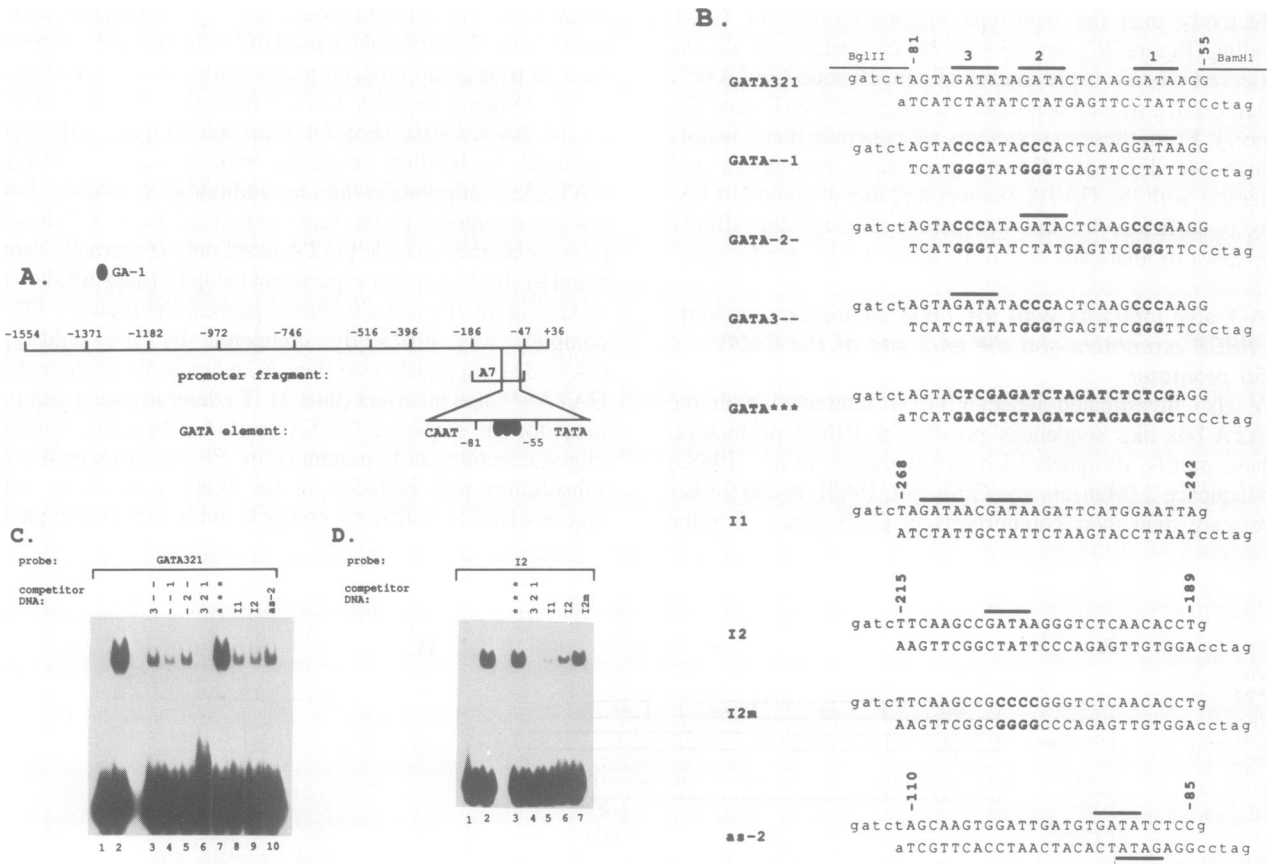

Fig. 2. The GATA motifs are essential for the interaction with GA-1. **A.** Schematic representation of the *cab-E* promoter showing the GATA element located between the CAAT and TATA box. **B.** DNA sequence of the oligonucleotides used for protein binding studies. GATA321 contains the three wild-type GATA motifs of the *cab-E* promoter. GATA--1, GATA-2-, GATA3-- and GATA*** are mutant derivatives carrying basepair substitutions (bold face) within two or all GATA motifs. I1 and I2 were derived from the conserved I-boxes of the *Arabidopsis rbcS-1A* promoter (Donald and Cashmore, 1990), the mutant derivative I2m carries a 4 bp substitution within the GATA sequence of the I2 motif (bold face). as-2 represents a sequence present in the CaMV 35S promoter (Lam and Chua, 1989). The individual GATA motifs are overlined. The oligonucleotides were synthesized with *Bam*HI and *Bgl*II 5′ overhangs, given in small letters. **C.** Competitive binding assay. The wild-type GATA321 oligonucleotide was incubated with 2 μg poly(dI.dC) in the absence (lane 1) or presence of 1 μg tobacco nuclear extract (lanes 2–10). In lanes 3–10 the binding assay was performed as described for lane 2 but different unlabeled oligonucleotides (0.5 pmol each) were included in the binding reactions, as indicated. DNA sequences of the competitor DNAs are given in panel B. **D.** Competitive binding assay demonstrating that the GATA motif of the I2 box is essential for complex formation. The assay was performed as described above, except that I2 was used as a radiolabeled probe. The DNA sequence of the competitor oligonucleotides is given in panel B.

oligonucleotide for complex formation (Figure 2D, lane 7). These results indicate that GA-1 binds to elements conserved in the photoregulated RBCS promoters, as well as the CAB promoters, and that the GATA sequence is essential for this interaction.

GA-1 binding is not restricted to elements from light-regulated promoters. A homologous sequence resides within the CaMV 35S promoter (−85 to −110) and was previously shown to be occupied by the nuclear protein ASF-2 (Lam and Chua, 1989). Competitive binding studies demonstrated that the binding behavior of an oligonucleotide (as-2, Figure 2B) derived from this region was similar to that observed with other oligonucleotides bearing a single GATA motif (Figure 2C, lane 10).

GC-1, a protein interacting with GC-rich positive regulatory elements within the cab-E promoter

Two GC-rich regions (Figure 3A, fragment B1 and E2) shown to potentiate the activity of the *cab-E* promoter in transgenic plants are located at positions −1371 to −1182 and −746 to −516 (Castresana *et al.*, 1988). In view of these regulatory properties we searched for protein factors that interacted with this GC-rich region. In a mobility shift assay performed with radiolabeled fragments B1 and E2, multiple protein−DNA complexes were observed. Competitive binding studies indicated that the same protein(s) interacted with both DNA fragments (data not shown). A striking characteristic of the two promoter regions is the presence of repeated elements, designated the GC-motifs (Castresana *et al.*, 1988); these motifs activate transcription in transgenic plants when fused to a truncated *cab-E* promoter (Alonso,E., Castresana,C. and Cashmore,A.R., unpublished). For binding experiments we concentrated on one of these elements, flanked by two *Sma*I sites and located at positions −1226 to −1207. This 21 bp fragment [GC(1×), Figure 3A] was subcloned, ligated to form two tandem copies [GC(2×)] and subjected to mobility shift assays. Using tobacco (Figure 3B) or tomato (data not shown) nuclear extracts, two protein−DNA complexes, b1 and b2 (Figure 3B) were resolved. Both complexes were efficiently competed for by the homologous DNA fragment (Figure 3B, lane 6), whereas no competition was obtained with oligonucleotides or promoter fragments bearing binding sites for other proteins (lanes 3–5, 7 and 8) or pUC18 DNA (lane 9). Furthermore, the dimer GC-motifs also specifically competed for complex formation with labeled fragments B1

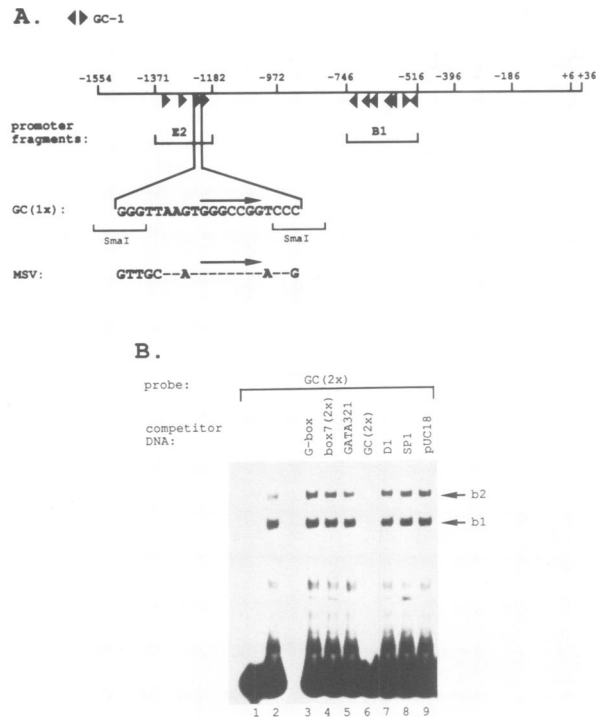

Fig. 3. The GC-repeats are bound by GC-1. **A.** Schematic representation of the *cab-E* promoter showing the location and orientation of the GC-repeats (solid triangles). The DNA sequence of GC(1×) (−1226 to −1207), used for binding studies, is given and compared with a similar sequence present in the MSV promoter (Mullineaux *et al.*, 1984). **B.** Competitive binding assay. Lane 1 shows the radiolabeled GC(2×) only. Lanes 2−9 represent binding assays in the presence of 5 μg tobacco nuclear extract and 3 μg poly(dI.dC). Lanes 3−8, competitor oligonucleotide or promoter fragments containing recognition sites for different proteins (0.3 pmol each), indicated above each lane, were included in the binding reactions. The competitor DNAs are defined in the legends to Figures 1, 2, 4 and 6; pUC18 DNA (0.3 pmol) was included as a negative control (lane 9).

and E2 (data not shown), inferring that the repeats present within these fragments were occupied by the same protein. Since we have not been able to separate the proteins involved in the two complexes obtained with GC(2×), and since both complexes show identical sensitivities with competing probes, we propose that they are mediated by the same protein(s), which we name GC-1.

The GC-motifs share considerable homology with the DNA binding site of the mammalian transcription factor SP1 (Gidoni *et al.*, 1984). However Figure 3B shows that an oligonucleotide bearing the SP1 recognition site does not compete for GC-1 (lane 8), indicating that these proteins have different DNA binding specificities.

Similarly, a DNA sequence present in the maize streak virus promoter (MSV; Mullineaux *et al.*, 1984) is almost identical to the GC-motifs (Figure 3A). Previously it was shown that a maize nuclear protein specifically interacts with this MSV fragment and a polymer of this region potentiates the expression of the truncated 35S promoter in a transient assay system (Fenoll *et al.*, 1988). It remains to be determined whether this factor is related to GC-1, although binding studies (data not shown) indicated that this fragment competed for binding to GC-1.

AT-1, a protein specifically interacting with the AT-rich NRE region of the cab-E promoter

The AT-rich region between −1182 and −972 of the *cab-E* promoter (fragment D1, Figure 4A) was shown to have a negative influence on expression of a reporter gene in transgenic tobacco plants (Castresana *et al.*, 1988). Three DNA motifs residing within this element (Figure 4A) resemble the AT-1 recognition sequence which was recently shown to be bound by AT-1, a factor characterized by its phosphorylation-modulated DNA binding activity (Datta and Cashmore, 1989). When fragment D1 (Figure 4A) was incubated with crude pea nuclear extracts, two to three pronounced protein−DNA complexes (b1, b2 and b3) were resolved; the number depending upon the amount of extract used (Figure 4B). Similar results were obtained with tobacco and tomato nuclear extracts (data not shown). The protein(s) involved in the formation of the fast-migrating complex b1 were distinct from the ones forming b2 and b3 based on the following criteria: first, they were precipitated by 50% (Figure 4D and E) but not by 40% ammonium sulfate, which selectively yields the proteins associated with b2 and b3 (Figure 4C); second, they were not retained on a Q-Sepharose column in contrast to the proteins associated with b2 and b3; and third, in an experiment where the proteins were renatured after SDS−PAGE, they were recovered in a low molecular weight fraction (<30 kd), in contrast to the proteins forming b2 and b3 which were in the 40−45 kd fraction (data not shown).

In the course of our experiments we focused on the protein(s) involved in the formation of complexes b2 and b3, since their binding was more specific. A competitive binding assay (Figure 4C) showed that, with the exception of fragment D1 (lane 3), all promoter fragments (lanes 4−7) bearing different protein binding sites failed to compete for the formation of complexes b2 and b3. The same result was obtained with pUC18 DNA (lane 8). These two complexes may be generated by the same protein as after SDS−PAGE the proteins forming b2 and b3 were recovered in the same fraction, indicating that the molecular weights of the protein(s) involved are very similar (data not shown). Similarly, the two activities have not been separable by chromatographic procedures. It is possible that the complex b3 corresponds to an oligomer of the protein associated with complex b2, as the former complex is only observed at relatively high extract concentration.

To investigate whether the two subfragments D2 and D3, containing one and two AT-1 boxes respectively (Figure 4A), would exhibit different binding affinities, a cross-competition assay was performed. The results displayed in Figure 4D (D2 probe) and Figure 4E (D3 probe) showed that the binding pattern of both subfragments was similar, however fragment D3, containing two AT-1 boxes, competed more efficiently (Figure 4D and E, lane 5) than D2 with only a single AT-1 box (lane 4). In both cases the binding affinity was slightly reduced in comparison with the complete fragment D1 (compare Figure 4C, D and E). When an oligonucleotide containing the *rbcS3.6* AT-1 box was included in the binding reaction, the protein complexes b2 and b3, formed with probes D1, D2 or D3, were abolished (data not shown). We conclude that AT-1 interacts with the three AT-1 motifs giving rise to protein complexes b2 and b3.

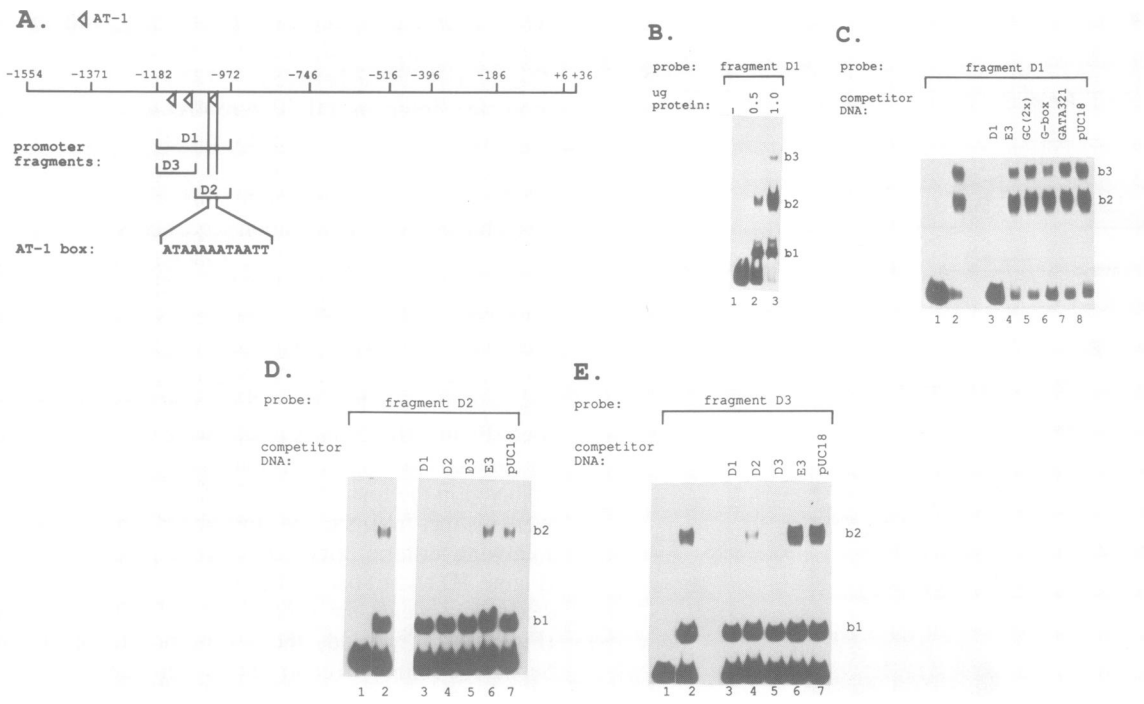

Fig. 4. Three AT-1 boxes reside within the AT-rich promoter region. **A.** Schematic representation of the *cab-E* promoter showing the three AT-1 boxes (open triangles) and fragment D1 (−1182 to −972), D2 (−1087 to −972) and D3 (−1182 to −1083). The DNA sequence of one of those motifs is given below. **B.** Mobility shift assay. Fragment D1 was incubated with increasing amount of pea nuclear extract (50% ammonium sulfate precipitate), indicated in μg above each lane, in the presence of 2 μg poly(dI.dC); b1, b2 and b3 indicate the position of the protein complexed DNA bands. Lane 1 contains the free DNA fragment. **C.** Competitive binding assay. Lane 1, free fragment D1; lanes 2−8, binding reaction in the presence of 1 μg pea nuclear extract (40% ammonium sulfate precipitate) and 2 μg poly(dI.dC); lanes 3−7, oligonucleotides or promoter fragments containing binding sites for different proteins (0.1 pmol each) were included in the binding reaction. The competitor DNAs are defined in the legends to Figures 1, 2, 3 and 5; pUC18 DNA (0.1 pmol) was included as a negative control (lane 8). **D** and **E.** Competitive binding assay. In each set of 7 lanes, the first shows the free DNA fragment (D2, panel D and D3, panel E) and the second the binding reaction in the presence of 1 μg pea extract (50% ammonium sulfate precipitate) and 2 μg poly(dI.dC). For lanes 3−6, binding assays were performed as described for lane 2, but 0.1 pmol of competitor plasmids, carrying different *cab-E* promoter fragments, defined in panel A and Figure 5A, were included in the reactions; pUC18 DNA (0.1 pmol) was included as a negative control (lane 7).

Seven GT-motifs reside within the cab-E promoter

Initial binding studies with tobacco nuclear extract showed a pronounced DNA−protein complex when fragment A1, C1 or E1 (Figure 5A) was used as a radiolabeled probe (data not shown). For further characterization, subfragments were obtained and subjected to mobility shift assays. The results demonstrated that the DNA recognition sites reside within region A7, A17, C4 and E3 (Figure 5A). Competition studies indicated that these fragments were bound by the same protein(s), although with different binding affinities (data not shown). Furthermore, binding activity was observed in significantly larger amounts when extracts were prepared from pea nuclei.

To delineate these binding sites more precisely, methylation interference experiments were performed on fragments A7, A17, C4 and E3. As shown in Figure 5B and C, protein binding to fragment A7 was strongly inhibited when the G residue at position −146 (box 2r) was methylated; a less pronounced effect was observed on the complementary strand at positions −108 and −109 (box 1). Using the fragment A17 as a probe, the G residues critical for protein binding were located at positions −366, −368 and −369 (Figure 5D, box 3r). A weak binding interference was observed at position −340. Three distinct protein binding sites, box 4 (G residues −831 and −832), box 5 (G residues −866 and −867) and box 6 (G residues −887 and −888) were identified within fragment C4 by methylation interference assays (Figure 5E).

In an electrophoretic mobility shift assay with fragment E3, three distinct protein−DNA complexes, designated b1, b2 and b3, were resolved (Figure 5F, lane 2). These binding activities were separated on a Q-Sepharose column, indicating that they were mediated by distinct proteins (Figure 5F). In a methylation interference experiment with either the crude or partially purified nuclear extract, the formation of only the slower migrating complex (b3) was inhibited by the methylation of G residue −1429 (Figure 5G, box 7). DNase I footprinting experiments and electrophoretic mobility shift assays with 3′ deletion clones defined the box 7 binding site between −1433 and −1415 (data not shown).

The GT-motifs in the cab-E promoter are bound by GT-1

All seven GT-motifs that were characterized by methylation interference experiments share some sequence homology (Figure 5H), in keeping with the likelihood that they are bound by the same protein. A comparison with known DNA binding activities already characterized for plant nuclear extracts revealed a similarity to the core binding sequence of GT-1 (Figure 5H), a pea nuclear protein interacting with several promoter elements within the pea *rbcS-3A* promoter (Green *et al.*, 1987). Green and coworkers (Green *et al.*, 1988) demonstrated that GT-1 is only able to form a stable protein−DNA complex when at least two DNA recognition sites are present; consistent with this observation it is noted that fragments A7 and C4 contain two or three GT-motifs respectively (Figure 5A). Furthermore, a stable interaction

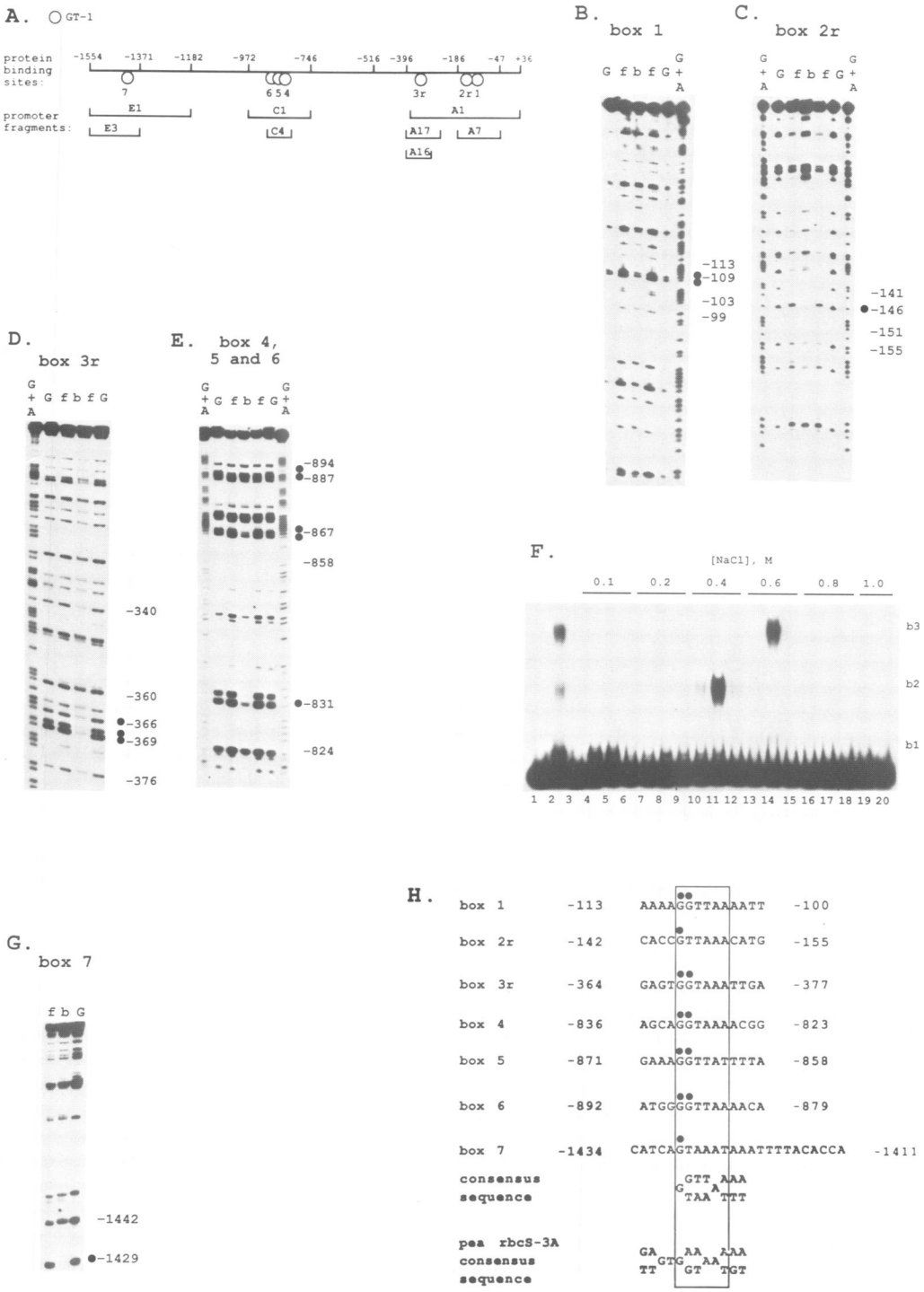

Fig. 5. Characterization of the GT-1 binding sites. **A**. Schematic representation of the *cab-E* promoter showing the location of the GT-1 binding sites and the fragments used for the binding studies. The GT-1 recognition sites were designated as boxes 1–7, r denotes those sites which reside on the complementary DNA strand. **B–E** and **G**. Methylation interference experiments. Filled circles indicate the G residues which when methylated impair protein binding. The position of the guanines (G) and adenines (G+A) were determined according to Maxam and Gilbert (1980). The experiments were performed as described in Materials and methods and the cleavage products were analyzed on a 6 or 8% sequencing gel; f and b represent the free and bound fractions, respectively. B and C; fragment A7 (−186 to −47) was cloned into the *Eco*RV site of pBluescript in both orientations and 3′ end-labeled at the *Eco*RI site of the polylinker. D; fragment 17 (−396 to −252) was excised from the *in vitro* mutagenized clone mp3 with *Xho*I (−252) and *Eco*RI and 3′ end-labeled at the *Eco*RI site. E; fragment C4 (903 to −823), subcloned into the *Sma*I and *Eco*RI sites of pBluescript, was 3′ end-labeled at the *Eco*RI site. F; fragment E3 (−1554 to −1371) was cloned into the *Eco*RI and *Sph*I sites of pUC18 and 3′ end-labeled at the *Hin*dIII site. **F**. Separation of nuclear proteins interacting with promoter fragment E3. Crude pea nuclear extract was applied to a Q-Sepharose column. The column was washed with buffer containing 0.1 M NaCl and then eluted with increasing salt concentrations. Fractions were dialyzed and assayed for DNA binding activity in a mobility shift assay with fragment E3 as a radiolabeled probe. Lane 1 represents the free DNA fragment. Lane 2 is the binding reaction in the presence of 0.5 μg crude pea nuclear extract and 2 μg poly(dI.dC). Lanes 3–20 are binding reactions performed with aliquots of column fractions eluted with buffer containing the indicated salt concentrations; b1, b2 and b3 refer to the DNA–protein complexes obtained with fragment E3. **H**. Summary of the GT-1 binding sites identified in the *cab-E* promoter. The consensus sequence derived from our results is given below and compared with the consensus sequence derived from the GT motifs present in the pea *rbcS-3A* promoter (Green *et al.*, 1987).

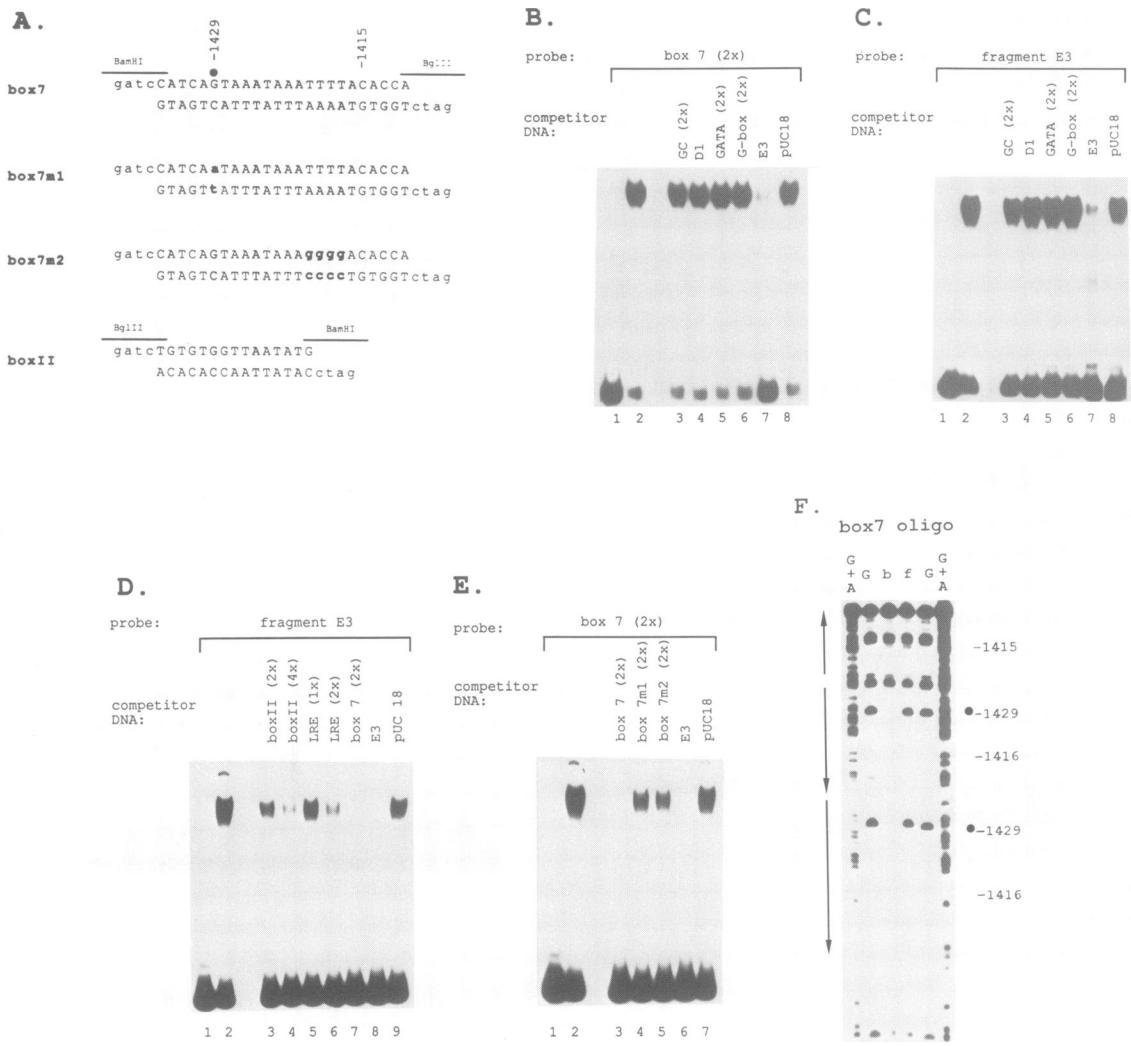

Fig. 6. GT-1 interacts specifically with the *cab-E* and the *Arabidopsis rbcS-1A* promoter. **A**. DNA sequence of the wild-type box 7 oligonucleotide derived from the *cab-E* promoter, as well as the mutant derivatives box 7m1 and box 7m2. The mutations are indicated in bold face and small letters. The boxII oligonucleotide is derived from the pea *rbcS-3A* promoter (Green *et al.*, 1988). **B** and **C**. Competitive binding assay with the radiolabeled fragment E3 (panel B) and the dimer oligonucleotide box 7(2×) (panel C). In each set of eight lanes the probes were incubated in the absence (lane 1) or presence (lanes 2–8) of 0.5 μg pea nuclear extract and 2 μg poly(dI.dC). Lanes 3–7, specific competitor plasmids (0.05 pmol each), carrying the indicated inserts were included in the binding reaction. The competitor DNAs are defined in the legends to Figures 1, 2, 3 and 4; pUC18 DNA was included as a negative control (lane 8). **D**. BoxII of the pea *rbcS-3A* and the LRE fragment of the *Arabisopsis rbcS-1A* promoter compete for GT-1. The binding assay with the radiolabeled *cab-E* promoter fragment E3 was performed as described for panels B and C. The plasmids used as competitor DNAs (0.05 pmol each) contained two and four copies of the *rbcS-3A* boxII oligonucleotide (lanes 3 and 4), one and two copies of the *Arabidopsis rbcS-1A* LRE fragment (−125 to −320, lanes 5 and 6), the dimerized box 7 oligonucleotide (lane 7) and fragment E3 (lane 8). **E**. Mutant box 7 oligonucleotides do not compete for GT-1. The binding reactions were performed as described above, lane 1 represents the free probe box 7(2×). The sequence of the oligonucleotides used as specific competitor DNAs (0.05 pmol each), indicated above each lane, is given in panel A; pUC18 DNA was included as a negative control (lane 7). For all experiments the competitor DNAs were added as linearized plasmids. **F**. Methylation interference experiment. The filled circles indicate the positions critical for protein binding. f and b represent the free and bound fractions, respectively. G and G+A refer to sequencing reactions (Maxam and Gilbert, 1980). The trimer box 7 (the orientation of the individual oligonucleotide is indicated by the arrows), was cloned into a pBluescript derivative (see Materials and methods), end-labeled at the *Bam*HI site and released with *Hin*dIII. The experiment was performed as described in Materials and methods.

with the box 3r was only obtained when the low affinity binding site at positions −347 to −340 is present. This latter sequence (GTCAAATA) contains one important mismatch in comparison to the core recognition sequence (Figure 5H). The DNA fragment A16 (Figure 5A), extending from −396 to −350, contains box 3r but the −347 to −340 sequence has been deleted; this fragment was incapable of binding GT-1. An interesting exception to the apparent requirement for two or more GT-motifs for complex formation is box 7, which is the only GT-motif within fragment E3 (Figure 5A). This box is characterized by a palindromic structure which we presume contributes to its strong binding affinity.

Oligonucleotides corresponding to box 7 and the pea *rbcS-3A* box II sequence were synthesized and used for competitive binding studies in order to determine their relative affinity for GT-1 activity (Figure 6A). The single box 7 oligonucleotide was bound weakly in an electrophoretic mobility shift assay; however a dimer [box 7(2×)] showed the same strong binding characteristics as the original E3 fragment (Figure 6B and C). Various DNA fragments containing binding sites for other proteins failed to compete for binding with box 7(2×) (Figure 6B, lanes 3–6) or fragment E3 (Figure 6C, lanes 3–6); in contrast, E3 was, in both cases, an effective competitor (Figure 6B and C, lane

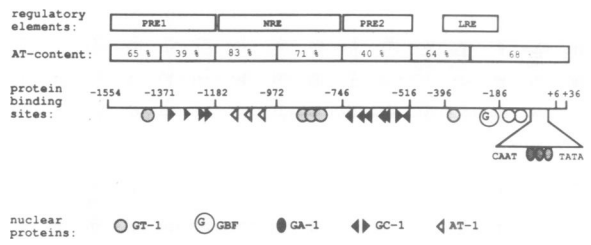

Fig. 7. Multiple binding sites within the *cab-E* promoter. Binding sites of the various factors are shown, as are the different regulatory elements and their AT content. Each binding site has been defined either by competitive binding studies and/or by methylation interference studies, as described in the text. The exceptions to this are the GC-1 binding sites; one of these motifs was defined as described in Figure 3, and the other sites are predicted on the basis of sequence homology (Castresane *et al.*, 1988).

7). We also observed that two copies of an oligonucleotide derived from the pea *rbcS-3A* box II sequence (Figure 6A) competed weakly with E3 for complex formation (Figure 6D, lane 3); four copies of the same sequence were more effective (lane 4). Fragment E3 itself (lane 8), or the oligonucleotide box 7(2×) representing a dimer of box 7, always showed the highest binding affinity (lane 7).

In order to investigate whether both half-sites of the palindrome in box 7 were responsible for the strong binding interaction observed for this sequence, two mutant oligonucleotides were designed (Figure 6A) and the effect of these mutations was analyzed in competitive binding experiments. Box 7m1 contains a single base pair exchange in which the critical G residue (−1429) is substituted by an adenine. The four T and A residues following this guanine were shown to be essential for GT-1 binding (Green *et al.*, 1988); in box 7m2 we therefore substituted the four distal A residues, residing within the other half of the palindrome, by cytidines. Competitive binding studies with the mutant dimerized oligonucleotides showed in both cases that they could not compete for binding with the dimerized wild-type box 7 oligonucleotide (Figure 6E, lanes 4 and 5); competition was observed with fragment E3 or the wild-type dimer itself (lanes 3 and 6). From these experiments we conclude that the strong affinity of GT-1 for box 7 is at least in part due to the long AT stretch and the palindromic nature of the binding site. We note that the G residues within the second half of the palindrome (−1415, −1413 or −1412) do not seem to be critical for binding, since no methylation interference was obtained on the complementary strand of E3 (data not shown) or when a trimer oligonucleotide was used as radiolabeled probe (Figure 6F).

An Arabidopsis rbcS-1A promoter fragment containing the conserved L-box sequence also binds GT-1

We have previously described conserved L-, I- and G-box sequences for the promoters of tobacco, tomato and *Arabidopsis* RBCS genes (Giuliano *et al.*, 1988a; Donald and Cashmore, 1990). From an examination of the *Arabidopsis rbcS-1A* promoter containing these conserved elements, we identified potential GT-1 recognition motifs. This fragment, designated LRE, contains *Arabidopsis rbcS-1A* promoter sequences from −320 to −125 and confers light regulated expression on a heterologous alcohol dehydrogenase reporter gene (Donald and Cashmore, 1990). Putative GT-1 recognition sites were identified within this fragment by comparison with previously characterized GT-motifs (Green *et al.*, 1987, 1988; this manuscript). One of these sites overlaps the L-box sequences, raising the possibility that the conserved L-box sequences are variants of a GT-1 binding site. We explored this possibility by including the *rbcS-1A* LRE fragment in a competitive binding assay. Figure 6D shows that a plasmid containing two copies of the LRE region (lane 6) competes for the retarded complex formed with fragment E3 to the same extent as a plasmid containing four copies of the pea *rbcS-3A* box II motif (lane 4). We conclude that GT-1 also interacts with this *Arabidopsis rbcS-1A* promoter fragment and from the sequence comparison it seems likely that this interaction involves binding to the conserved L-box sequence.

Discussion

Multiple factors bind to the cab-E promoter

The experiments reported in this paper demonstrate the complex binding pattern of five different nuclear proteins interacting with different regulatory elements within the *cab-E* promoter (Castresana *et al.*, 1988). A schematic representation of this promoter and the location of the protein binding sites is shown in Figure 7. Our analyses imply that the complex regulation of CAB gene expression in response to several different stimuli (Fluhr and Chua, 1986; Fluhr *et al.*, 1986; Giuliano *et al.*, 1988b; Nagy *et al.*, 1988; Piechulla and Gruissem, 1987; Simpson *et al.*, 1985 and 1986; Tobin and Silverthorne, 1985) involves the specific interaction of multiple and distinct nuclear factors.

We note that the five factors in question do not represent an exhaustive list; we have ample evidence that there are additional proteins interacting with this promoter. In this respect, our estimate of the complexity of this promoter and its interacting factors is certainly an underestimate. However, we also note that whereas we commonly refer to the promoter in question as the *cab-E* promoter, some of the regulatory elements we have analyzed within this 1.5 kb promoter fragment are likely to be involved in the regulation of the *cab-F* gene; *cab-E* and *cab-F* are divergently transcribed genes separated by ~2 kb of non-coding sequences (Castresana *et al.*, 1987). In this respect it could be argued that the picture we present overestimates the complexity of the *cab-E* regulatory elements and their cognate factors, as some of the far-upstream elements may be more correctly assigned to the *cab-F* promoter. However, conversely, the likely possibility of intergenic interactions, where some of the regulatory elements might mediate the expression of both genes, adds an additional level of complexity that eventually needs to be considered.

In this study the five different factors we focused on either interacted with transcriptionally important DNA motifs (P.Pringmann and A.R.Cashmore, in preparation; Donald and Cashmore, 1990; Gidoni *et al.*, 1989; Kuhlemeier *et al.*, 1989) or showed very pronounced DNA binding activity. Although most assays were performed with crude nuclear extracts, preliminary chromatographic and renaturation experiments demonstrated that the five identified binding activities corresponded to different proteins.

Many of the factors binding to the cab-E promoter appear to be ubiquitous

We have recently characterized a nuclear factor GBF (Giuliano *et al.*, 1988a) which interacts with the G-box motif

in RBCS genes. In this present study we have characterized GBF binding to a G-box-like sequence in the *cab-E* promoter. The methylation interference pattern that we observe for both elements is very similar (Figure 1C and D; Donald *et al.*, 1990). Within the *cab-E* promoter, the G-box sequence is localized in a region known to be indispensable for light regulation (Figure 7, and Castresana *et al.*, 1988). Furthermore, recent site-directed mutagenesis and *in vivo* expression studies indicate that mutation of this site, in the context of the entire *cab-E* promoter, strongly reduces the expression of a reporter gene in transgenic plants (P.Pringmann and A.R.Cashmore, in preparation). These studies imply that there is an *in vivo* interaction between GBF and its cognate binding site within the *cab-E* promoter.

CG-1 interacts with the chalcone synthase promoter of *Antirrhinum majus* (Staiger *et al.*, 1989) and requires the sequence CACGTG, identical to the G-box motif, for *in vitro* binding. The relationship between GBF and this second G-box binding protein, CG-1, is still unclear. Although complex formation with CG-1 was competed by an oligonucleotide derived from several RBCS G-box-containing elements, no competition was obtained with an oligonucleotide representing the *cab-E* G-box motif (Staiger *et al.*, 1989), implying, that the factor in question would not tolerate the mismatches within the *cab-E* G-box. There are two explanations for these results that somewhat contradict our experiments. CG-1 may be the same protein as GBF and the slightly reduced binding affinity which we also observed for the *cab-E* element (Figure 1C) may have been missed in the earlier study. Alternatively, CG-1 and GBF could be different proteins, interacting with similar sequences as described for several mammalian and yeast transcription factors (see later for a discussion on this point).

G-box-like sequences are known to occur in promoters of non-light regulated plant genes including both maize and *Arabidopsis adh*, where a factor has been demonstrated to bind *in vivo* to a G-box promoter sequence (Ferl and Nick, 1987; McKendree *et al.*, 1990). Similarly in yeast (Vogel *et al.*, 1989), *Xenopus* (Hall and Taylor, 1989; Scotto *et al.*, 1989) and in mammals (Beckmann *et al.*, 1990; Chodosh *et al.*, 1987; Sawadogo and Roeder, 1985; Sawadogo *et al.*, 1988), G-box-like promoter sequences and corresponding factors have been characterized and in yeast we have identified a factor that mediates plant G-box-dependent expression (Donald *et al.*, 1990).

The second factor that we have characterized, GA-1, binds to the GATA-box sequences present in the *cab-E* and other CAB promoters. These sequences also appear to have their counterparts in RBCS genes; we refer to these latter motifs as I-boxes and they often occur as the sequence GATAAG in the −150 to −300 region of these genes (Giuliano *et al.*, 1988; Donald and Cashmore, 1990). In addition, a related complementary sequence (CCTTATCAT) was identified as a uniquely conserved sequence commonly residing 5′ proximal to the TATA box of RBCS genes (Grob and Stueber, 1987). By competitive *in vitro* binding studies, we have demonstrated that both the CAB GATA-box and RBCS I-box sequences bind the factor GA-1. Mutation of either the GATA element within two different CAB promoters (Bringmann *et al.*, in preparation; Gidoni *et al.*, 1989) or the I-boxes of the *Arabidopsis rbcS-1A* promoter (Donald and Cashmore, 1990) strongly decreases the expression of a reporter gene in transgenic plants, suggesting that the binding we observe may be functionally significant. Recently Lam and Chua (1989) described a nuclear protein, ASF-2, interacting with the as-2 site of the CaMV 35S promoter. We have demonstrated that *in vitro* the as-2 site is also bound by GA-1, however the affinity for this motif is significantly reduced in comparison to the *cab-E* GATA element (Figure 2C). It remains to be determined whether GA-1 and ASF-2 are the same factor. We note that proteins interacting with GATA motifs are not restricted to the plant kingdom. The erythroid specific transcription factor GF-1 (Eryf1/NFE-1/EF-1/EFγ) recognizes a similar GATA-motif within several globin genes (Perkin *et al.*, 1989; Trainor *et al.*, 1990; Tsai *et al.*, 1989). Furthermore, the zinc-finger motif of GF-1 is homologous to a protein sequence derived from the *areA* gene of *Aspergillus nidulans*, encoding a regulatory protein involved in nitrogen acquisition (Arst *et al.*, 1989). However, the DNA binding site of the *areA* gene product is unknown.

Three AT-1 boxes reside within the negative regulatory element of the *cab-E* promoter and are recognized by the third protein we have characterized, AT-1. DNA binding proteins interacting with AT-rich sequences have been reported for a variety of plant promoters (de Brujin *et al.*, 1989; Deikman and Fischer, 1988; Holdsworth and Laties, 1989a and b; Jensen *et al.*, 1988; Jofuku *et al.*, 1987; Metz *et al.*, 1988; Stougaard *et al.*, 1987). With the exception of our own studies (Datta and Cashmore, 1989), these reports all concern non-photoregulated genes. The relationship between these various factors and the role they play in mediating specific gene expression is unknown.

Multiple copies of the GC-motif are clustered within the PRE1 and PRE2 positive regulatory elements of the *cab-E* promoter (Figure 3A). These GC-elements act as strong activators of expression in transgenic tobacco plants when fused to a truncated *cab-E* promoter (Alonso,E., Castresana,C. and Cashmore,A.R., unpublished) and thus it seems likely that they are at least partly responsible for the activity mediated by PRE1 and PRE2. Based on sequence comparison and competitive binding studies we argue that they are bound by GC-1, the fourth factor described in this paper. An almost identical sequence occurs twice within a DNA region of the MSV promoter (Mullineaux *et al.*, 1984) which was demonstrated to be essential for transient expression and occupied by a maize nuclear protein (Fenoll *et al.*, 1988). Although the MSV fragment does compete for GC-1 binding *in vitro*, the relationship between these two factors is unknown.

GT-motifs, the recognition sites for the fifth factor, are distributed throughout the entire *cab-E* promoter region, residing within both positive and negative regulatory elements and in the close vicinity of the light regulatory element (Figure 7). Expression studies with the pea *rbcS-3A* promoter imply that both positive and negative elements overlap at the GT-motifs (Kuhlemeier *et al.*, 1987b and 1988a; Strittmatter and Chua, 1987). A 116 bp fragment from the pea *rbcS-3A* promoter containing two GT-1 binding sites is able to confer light-regulated expression on a truncated 35S promoter, as determined by the dark-adaptation assay (Kuhlemeier *et al.*, 1989). However, two points should be noted. Firstly; the physiological relevance of the dark-adaptation assay, although commonly performed in many laboratories using transgenic plants to study light regulation (including our own), remains to be demonstrated

and secondly; GT-1 binding sites occur in several promoters (Green *et al.*, 1988), some of which, such as the 35S promoter, are non-photoregulated.

The large number of GT-1 binding sites associated with the *cab-E* promoter is quite striking. It is possible that differential affinity for boxes 1–7 influences the function exerted by GT-1 although no good correlation was observed between binding affinity and promoter strength (Kuhlemeier *et al.*, 1988b). Another possibility is that the distance between the binding motif and the transcriptional start site is functionally important or that adjacent DNA binding proteins have modulating effects, influencing the regulatory function of GT-1, as demonstrated for several mammalian genes (Calame, 1989; Chui *et al.*, 1989; Jones *et al.*, 1988). Site-directed mutagenesis studies are necessary to confirm the importance of these GT-motifs within the *cab-E* promoter.

What is the role of these apparently ubiquitous DNA binding proteins in photoregulated gene expression?

Our results demonstrate that the photoregulated *cab-E* promoter is occupied by several nuclear proteins, most of them interacting with multiple sites. We were not able to demonstrate any pronounced differences in DNA binding activity between extracts prepared from light-grown and dark-adapted plants for any of these factors. Furthermore, DNA sequences homologous to the identified DNA binding sites are also present in non-photoregulated genes and proteins with similar binding characteristics have been decribed for other organisms. What is the relationship between these various factors and do ubiquitous factors play a role in mediating photoregulated gene expression? In reference to this question we discuss the following points.

We first note that whereas factors binding similar DNA sequences associated with different promoters may be the same factor, there is ample evidence indicating that this need not be the case. Hence in many cases, factors that appear ubiquitous may in fact be unique to the promoters in question (Johnson and McKnight, 1989; Maity *et al.*, 1988). In mammalian cells for example, both USF and TFE3 bind to a similar G-box like sequence (Beckmann *et al.*, 1990; Sawadogo and Roeder, 1985; Sawadogo *et al.*, 1988). Similarly in yeast, where we have demonstrated *Arabidopsis rbcS-3A* G-box-dependent expression (Donald *et al.*, 1990), the pho4 gene product binds to G-box-like sequences (Vogel *et al.*, 1989) but it is not the factor that mediates *rbcS-3A* G-box-dependent expression, as this expression is observed in a yeast pho4 mutant. Thus, in yeast, as in mammalian cells, there are multiple factors that bind to G-box-like sequences. We conclude that the relationship between the factors that interact *in vivo* with G-boxes associated with RBCS, CAB, ADH (Ferl and Nick, 1987; McKendree *et al.*, 1990) and chalcone synthase promoters (Schulze-Lefert *et al.*, 1989; Staiger *et al.*, 1989) is at this stage quite unknown. We note that in many cases the factors interacting with these sequences *in vitro* may well be the same, but that does not allow one to conclude that the factors mediating expression *in vivo* are identical. Similar uncertainties exist, for example, concerning the identity of factors that bind to GT-motif and GATA-box sequences in RBCS, CAB and CaMV promoters.

A second point is that in many situations a general transcription factor may indeed be involved in mediating the expression of genes with quite diverse expression characteristics. In the simplest situation, different expression characteristics would simply reflect, in an additive way, the distinct transcription factors or expression modules associated with the particular genes.

The final point we wish to discuss reflects the complexity that we have characterized with respect to both the number of factors and the binding sites associated with the *cab-E* promoter. This complexity, at least in theory, could result in quite distinct expression characteristics, reflecting, both qualitatively and quantitatively, something substantially more than a simple additive effect of the properties of the individual factors. In this model, the activity of a particular promoter would be strongly influenced by protein–protein interactions. The likelihood of such interactions influencing promoter activity has been well recognized and was recently elaborated on in a review by Johnson and McKnight (1989). The significance of these interactions is illustrated by the yeast mating type regulatory proteins PRTF, which either stimulates or represses expression depending on the promoter context (Johnson and McKnight, 1989). A point we wish to emphasize here concerns the enormous structural complexity that could be generated, simply via the utilization of a relatively small number of general transcription factors. This complexity would reflect not only the order and spacing of factors along a promoter, but would also be expected to be influenced by nucleotide sequences residing between protein binding elements and affecting DNA conformation (reviewed by Adhya, 1989 and Travers, 1989). This structural complexity might be expected to result in a similar state of complexity at the level of transcriptional regulation.

As an example of the above considerations we note that it is possible that a general transcription factor might directly mediate light regulation, that is the factor may reside at the end of a photoregulatory signal transduction pathway, if positioned in an appropriate promoter context. In contrast, in a different promoter setting, the factor may be quite unaffected by the same signal transduction pathway. In raising this possibility we are influenced by the fact that we have evidence that the profile of proteins associated with the LRE is substantially more complex than that depicted in Figure 7. In making this proposal we are certainly not excluding the obvious alternative: that light regulation is mediated by a factor unique to photoregulated genes; this factor could be one of those we have described or it could be as yet uncharacterized.

We note that such complex promoters in which the overall activity is clearly something quite distinct from the simple sum of their parts, will be difficult to analyze, both by genetic means and by the methods of molecular biology and biochemistry. For example, the analyses of individual *cis*-acting elements will not necessarily shed light on the activity of the same elements in the context of a complex promoter. Similarly, genetic characterization of 'unique' *trans*-acting factors will be difficult when 'uniqueness' is defined in the context of a promoter and not simply by the amino acid sequence of the factor, as mutations would generally be expected to have pleiotropic, if not lethal affects.

In keeping with our suggestion that certain transcription factors may behave in unique ways reflecting their particular promoter setting, are the observations demonstrating gene-specific defects resulting from mutations in RNA polymerase subunits (Scafe *et al.*, 1990 and references cited therein). In concluding on an optimistic theme, we note that at least

in these cases, genetic studies have been fruitful in illuminating unique interactions of general transcription factors.

Materials and methods

Recombinant DNA techniques and preparation of radiolabeled DNA probes

All promoter fragments were subcloned into pUC18 or pBluescript-SK(+) using general cloning procedures (Maniatis *et al.*, 1982). Radiolabeled probes for mobility shift assays were prepared by filling 5'-overhangs with [α-^{32}P]dATP or [α-^{32}P]dGTP and DNA polymerase (Klenow, Promega) or [γ-^{32}P]ATP and T4 polynucleotide kinase (Promega). Probes for methylation interference experiments were generated in the same way, followed by a second restriction digestion. The fragments were electrophoretically separated from the vector and unincorporated nucleotides. Oligonucleotides were synthesized using an Applied Biosystems 380B DNA synthesizer, annealed and cloned into the *Bgl*II and *Bam*HI sites of pBglII, a pBluescript-SK(+) derivative, where a *Bgl*II linker was inserted into the *Eco*RV site. The correct basepair composition was confirmed by dideoxy sequencing (Chen and Seeburg, 1985).

Preparation of nuclear extracts

Nuclear extracts were either prepared from 3-week-old tomato or 8-week-old *N.plumbaginifolia* seedlings grown under greenhouse conditions or from 6 day old pea plumuli grown in the dark.

Proteins were isolated as described by Maier *et al.* (1987) except that 2.5% (w/v) dextran T40 was used in buffer A; the sonication step was omitted. Instead the extract was incubated on ice for 30 min with occasional shaking. The protein concentrations were determined using the Bio-Rad Assay System. The extract was either used directly for DNA binding assays, ammonium sulfate precipitated or applied onto a Q-Sepharose column (Pharmacia) equilibrated with column buffer (20 mM HEPES, pH 7.5, 1 mM $MgCl_2$, 0.2 mM EDTA, 20% (v/v) glycerol, 0.5 mM DTT, 0.5 mM PMSF). The proteins were then step-eluted in the same buffer supplemented with increasing NaCl concentrations and dialyzed for 4 h at 4°C against column buffer supplemented with 100 mM KCl.

Electrophoretic mobility shift assay

A typical binding reaction contained 10 mM Tris–HCl, pH 7.5, 40 mM NaCl, 1 mM EDTA, 1 mM 2-mercaptoethanol, 4% glycerol, 2–10 fmol (10 000 c.p.m.) radioactively labeled DNA, poly(dI.dC), nuclear extract and specific competitor DNA. The concentration of the competitor DNAs was determined spectrophotometrically. Linearized plasmids, isolated DNA fragments or annealed oligonucleotides were included in the binding reactions, as indicated in the figure legends. After incubation for 30 min at room temperature, 1 μl loading buffer (30% glycerol, 0.025% bromphenol blue, 0.025% xylene cyanol) was added and the reactions were loaded onto a 4 or 5% polyacrylamide gel in 25 mM Tris–HCl, 190 mM glycine, 1 mM EDTA, pH 8.3 or 0.25× TBE (Maniatis *et al.*, 1982), which was prerun for 1–2 h at 6 V/cm. Electrophoresis was carried out at the same voltage. Gels were dried and exposed overnight to X-ray films.

Methylation interference assay

Methylation interference experiments were performed as described by Donald *et al.* (1990). G and G+A reactions were carried out according to Maxam and Gilbert (1980). The individual promoter fragments used for the assays are indicated in the figure legend.

Acknowledgements

We thank Dr S.H.Howell (Cornell University) for providing the clone CF208, containing the MSV promoter fragment, Drs C.Castresana and E.Alonso for the plasmids pA.1, pB.1, pD.1 and pGC(2×), as well as Dr P.Bringmann for the *in vitro* mutagenized clone mp3. We are also grateful to members in the Plant Science Institute for critical reading of this manuscript. This work was supported by grants to A.R.C. from the DOE (DE-FG02-87ER13680) and NIH (GM 38409). U.S. was supported in part by a grant from the Deutsche Academische Austauschdienst (DAAD), Sonderprogramm Gentechnology.

References

Adhya,S. (1989) *Annu. Rev. Genet.*, **23**, 227–250.
Arst,H.N.J., Kudla,A., Martinez-Rossi,N., Cadick,M.X., Sibley,S. and Davies,R.W. (1989) *Trends Genet.*, **5**, 291.
Beckmann,H., Su,L.-K. and Kadesch,T. (1990) *Genes Dev.*, **4**, 167–179.
Benfey,P.N. and Chua,N.-H. (1989) *Science*, **244**, 174–181.
Calame,K.L. (1989) *Trends Genet.*, **5**, 395–399.
Castresana,C., Staneloni,R., Malik,V.S. and Cashmore,A.R. (1987) *Plant Mol. Biol.*, **10**, 117–126.
Castresana,C., Garcia-Luque,I., Alonso,E., Malik,V.S. and Cashmore,A.R. (1988) *EMBO J.*, **7**, 1929–1936.
Chen,E,Y. and Seeburg,P.H. (1985) *DNA*, **4**, 165–70.
Chodosh,L.A., Carthew,R.W., Morgan,J.G., Crabtree,G.R. and Sharp,P.A. (1987) *Science*, **238**, 684–688.
Chui,R., Angel,P. and Karin,M. (1989) *Cell*, **59**, 979–986.
Datta,N. and Cashmore,A.R. (1989) *Plant Cell*, **1**, 1069–1077.
de Brujin,F.J., Felix,G., Grunenberg,B., Hoffmann,H.J., Metz,B., Ratet,P., Simons-Schreier,A., Szabados,L., Welters,P. and Schell,J. (1989) *Plant Mol. Biol.*, **13**, 319–325.
Deikman,J. and Fischer,R.L. (1988) *EMBO J.*, **7**, 3315–3320.
Donald,R.G.K. and Cashmore,A.R. (1990) *EMBO J.*, **9**, 1717–1726.
Donald,R.G.K., Schindler,U., Batschauer,A. and Cashmore,A.R. (1990) *EMBO J.*, **9**, 1727–1735.
Dynan,W.S. (1989) *Cell*, **58**, 1–4.
Fenoll,C., Black,D.M. and Howell,S.H. (1988) *EMBO J.*, **7**, 1589–1596.
Ferl,R.J. and Nick,H.S. (1987) *J. Biol. Chem.*, **262**, 7947–7950.
Fluhr,R. and Chua,N.-H. (1986) *Proc. Natl. Acad. Sci. USA*, **232**, 1106–1112.
Fried,M. and Crother,D.M. (1981) *Nucleic Acids Res.*, **9**, 6505–6525.
Garner,M.M. and Revzin,A. (1981) *Nucleic Acids Res.*, **9**, 3047–3060.;
Gidoni,D., Dynan,W.S. and Tjian,R. (1984) *Nature*, **312**, 409–413.
Gidoni,D., Brosio,P., Bond-Nutter,D., Bedbrook,J. and Dunsmuir,P. (1989) *Mol. Gen. Genet.*, **215**, 337–344.
Giuliano,G., Pichersky,E., Malik,V.S., Timko,M.P., Scolnik,P.A. and Cashmore,A.R. (1988a) *Proc. Natl. Acad. Sci. USA*, **85**, 7089–7093.
Giuliano,G., Hoffman,N.E., Ko,K., Scolnik,P.A. and Cashmore,A.R. (1988b) *EMBO J.*, **7**, 3635–3642.
Green,P.J., Kay,S.A. and Chua,N.-H. (1987) *EMBO J.*, **6**, 2543–2549.
Green,P.J. Yong,M.-H., Cuozzo,M., Kano-Murakami,Y., Silverstein,P. and Chua,N.-H. (1988) *EMBO J.*, **7**, 4035–4044.
Grob,U. and Stueber,K. (1987) *Nucleic Acids Res.*, **15**, 9957–5011.
Holdsworth,M.J. and Laties,G.G. (1989a) *Planta*, **180**, 74–81.
Holdsworth,M.J. and Laties,G.G. (1989b) *Planta*, **179**, 17–23.
Jensen,E.O., Marcker,K.A., Schell,J. and de Bruijn,F.J. (1988) *EMBO J.*, **7**, 1265–1271.
Jofuku,K.D., Okamuro,J.K. and Goldberg,R.B. (1987) *Nature*, **328**, 734–737.
Johnson,P.F. and McKnight,S.L. (1989) *Annu. Rev. Biochem.*, **58**, 799–839.
Jones,N.C., Rigby,P.W.J. and Ziff,E.B. (1988) *Genes Dev.*, **2**, 267–281.
Kuhlemeier,C., Green,P. and Chua,N.-H. (1987a) *Annu. Rev. Plant Physiol.*, **38**, 221–257.
Kuhlemeier,C., Fluhr,R., Green,P.J. and Chua,N.-H. (1987b) *Genes Dev.*, **1**, 247–255.
Kuhlemeier,C., Cuozzo,M., Green,P., Goyvaerts,E., Ward,K. and Chua,N.-H. (1988a) *Proc. Natl. Acad. Sci. USA*, **85**, 4662–4666.
Kuhlemeier,C., Fluhr,R. and Chua,N.-H. (1988b) *Mol. Gen. Genet.*, **212**, 405–411.
Kuhlemeier,C., Strittmatter,G., Ward,K. and Chua,N.-H. (1989) *Plant Cell*, **1**, 471–478.
Lam,E. and Chua,N.-H. (1989) *Plant Cell*, **1**, 1147–1156.
Lee,W., Haslinger,A., Karin,M. and Tjian,R. (1987) *Nature*, **325**, 368–372.
Maier,U.-G., Brown,J.W.S., Toloczyki,C. and Feix,G. (1987) *EMBO J.*, **6**, 17–22.
Maity,S.N., Golumbek,P.T., Karsenty,G. and De Crombrugghe,B. (1988) *Science*, **241**, 582–585.
Maniatis,T., Fritsch,E.F. and Sambrook,J. (1982) *Molecular Cloning: A Laboratory Manual.* Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
Manzara,T. and Gruissem,W. (1988) *Photosynth. Res.*, **16**, 117–130.
Maxam,A.M. and Gilbert,W. (1980) *Methods Enzymol.*, **65**, 499–560.
McKendree,W.L., Paul,A.-L., DeLisle,A.J. and Ferl,R.J. (1990) *Plant Cell*, **2**, 207–21.
Metz,B.A., Welters,P., Hoffmann,H.J., Jensen,O., Schell,J. and de Bruijn,F.J. (1988) *Mol. Gen. Genet.*, **214**, 181–191.
Mitchell,P.J. and Tjian,R. (1989) *Science*, **245**, 371–378.
Mullineaux,P.M., Donson,J., Morris-Krsinich,B.A.M., Boulton,M.I. and Davies,J.W. (1984) *EMBO J.*, **3**, 3036–3068.
Nagy,F., Kay,S.A. and Chua,N.-H. (1988) *Genes Dev.*, **2**, 376–382.
Perkin,N.D., Nicolas,R.H., Plumb,M.A. and Goodwin,G.H. (1989) *Nucleic Acids Res.*, **17**, 1299–1315.

Piechulla,B. and Gruissem,W. (1987) *EMBO J.*, **6**, 3593–3599.
Sawadogo,M. and Roeder,R.G. (1985) *Cell*, **43**, 165–175.
Sawadogo,M., Van Dyke,M.W., Gregor,P.D. and Roeder,R.G. (1988) *J. Biol. Chem.*, **263**, 11985–11993.
Scafe,C., Nonet,M. and Young,R.A. (1990) *Mol. Cell. Biol.*, **10**, 1010–1016.
Schell,J.S. (1987) *Science*, **237**, 1176–1182.
Schulze-Lefert,P., Dangl,J.L., Becker-Andre,M., Hahlbrock,K. and Schulz,W. (1989) *EMBO J.*, **8**, 651–656.
Scotto,K.W., Kaulen,H. and Roeder,R.G. (1989) *Genes Dev.*, **3**, 651–662.
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.
Staiger,D., Kaulen,H. and Schell,J. (1989) *Proc. Natl. Acad. Sci. USA*, **86**, 6930–6934.
Stougaard,J., Sandal,N.N., Gron,A., Kuehle,A. and Marcker,K.A. (1987) *EMBO J.*, **6**, 3565–3569.
Strittmatter,G. and Chua,N.-H. (1987) *Proc. Natl. Acad. Sci. USA*, **84**, 8986–8990.
Tobin,E.M. and Silverthorne,J. (1985) *Annu. Rev. Plant Physiol.*, **36**, 569–594.
Trainor,C.D., Evans,T., Felsenfeld,G. and Boguski,M.S. (1990) *Nature*, **343**, 92–96.
Travers,A.A. (1989) *Annu. Rev. Biochem.*, **23**, 427–452.
Tsai,S.-F., Martin,D.I.K., Zon,L.I., D'Andrea,A.D., Wong,G.G. and Orkin,S.H. (1989) *Nature*, **339**, 446–451.
Vogel,K., Hoerz,W. and Hinnen,A. (1989) *Mol. Cell. Biol.*, **9**, 2050–2057.
Willmitzer,L. (1988) *Trends Genet.*, **4**, 13–18.

Received on April 23, 1990; revised on July 19, 1990