The EMBO Journal vol.7 no.13 pp.4035-4044, 1988

Binding site requirements for pea nuclear protein factor GT-1 correlate with sequences required for light-dependent transcriptional activation of the *rbcS-3A* gene

Pamela J.Green[1], Mun-Heng Yong[2], Maria Cuozzo, Yuriko Kano-Murakami, Peter Silverstein[3] and Nam-Hai Chua

Laboratory of Plant Molecular Biology, The Rockefeller University, 1230 York Avenue, New York, NY 10021-6399, USA

[1]Present address: MSU-DOE Plant Research Laboratory and Department of Biochemistry, Michigan State University, East Lansing, MI 48824, USA
[2]Present address: Institute of Molecular and Cell Biology, National University of Singapore, Kent Ridge, 0511, Republic of Singapore
[3]Present address: Department of Biology, Indiana University, Bloomington, IN 47401, USA

Communicated by D.M.Lonsdale

Nuclear protein factor GT-1 binds to sequence boxes II, III, II* and III* upstream of the light-responsive pea *rbcS-3A* gene. We have shown previously that box II and box III are required for expression of *rbcS-3A* when redundant elements upstream of −170 (relative to the transcription start site) are removed. Here we present evidence that deletion and substitution mutations downstream of −170 which eliminate expression also decrease binding. Using a series of 2 bp substitution mutations we have defined a core of six residues (GGTTAA) within box II (GTGTGGTTAATATG) that are critical for binding. The most detrimental mutation for binding, which changes the double Gs to Cs, is sufficient to eliminate detectable expression *in vivo* when only 170 bp of 5′ flanking sequences are present. The simplest interpretation of these data is that GT-1 is an activator of *rbcS-3A* transcription. Footprinting experiments show that GT-1 from both light-grown and dark-adapted plants binds to the same sequences *in vitro*. Therefore, the lack of expression of *rbcS-3A* in the dark is not due to the absence of GT-1. In our analysis of the sequence elements upstream of −170, we have mapped two additional GT-1 sites (boxes II and III**) between −330 and −410. The similarities and differences among the GT-1 sites located upstream and downstream of −170 are discussed in terms of the different sequence requirements for *rbcS-3A* expression during development.**
Key words: trans-acting factors/light-responsive transcription/transcriptional activator/multiple binding sites

Introduction

Light-responsive transcription is perhaps the most interesting and well studied mode of gene regulation in plants. The large number of plant genes reported to be regulated by light attests to the importance of this process for normal plant growth and development (Tobin and Silverthorne, 1985). At the molecular level the most detailed work has focused on the nuclear gene families encoding two chloroplast proteins, the small subunit of ribulose bisphosphate carboxylase (*rbcS*) and the chlorophyll ab binding protein (*cab*) of the light harvesting complex (Kuhlemeier *et al.*, 1987b; Nagy *et al.*, 1988). The abundance of the transcripts for both of these proteins increases markedly following exposure to light primarily in cells which contain chloroplasts (Apel, 1979; Tobin, 1981; Coruzzi *et al.*, 1984; Fluhr *et al.*, 1986a,b). Nuclear run-on experiments (Gallagher and Ellis, 1982; Silverthorne and Tobin, 1984; Gallagher *et al.*, 1985; Mosinger *et al.*, 1985) and chimeric gene constructs expressed in transgenic plants (Fluhr *et al.*, 1986b; Nagy *et al.*, 1986; Simpson *et al.*, 1986a) have demonstrated that both gene families are regulated at the transcriptional level. *Cis*-acting DNA sequence elements upstream of *rbcS* and *cab* transcription start sites have been shown to enhance the expression of reporter genes in a light-dependent manner in transgenic tobacco leaves (Fluhr *et al.*, 1986a; Simpson *et al.*, 1986b; Nagy *et al.*, 1987). These *cis*-acting elements are thought to function by binding to *trans*-acting protein factors present in plant cell nuclei (Green *et al.*, 1987). Such factors could then interact directly or via other proteins with RNA polymerase II to modulate transcription.

In addition to light regulated genes, *cis*-element analysis in transgenic plants has been accomplished for a variety of other genes (see, e.g. Chen *et al.*, 1986; Colot *et al.*, 1987; Stockhaus *et al.*, 1987; Walker *et al.*, 1987). Several *trans*-acting factors have also been identified (Ferl and Nick, 1987; Green *et al.*, 1987; Jofuku *et al.*, 1987; Mikami *et al.*, 1987; Maier *et al.*, 1987; Jensen *et al.*, 1988). From these reports and numerous others from non-plant systems, *cis*-acting elements have been conventionally divided into two classes: (i) promoter elements and associated upstream activation sequences which generally reside within ~100 bp of the start of transcription, and (ii) enhancer elements located further upstream which function in a position independent manner (for reviews see Dynan and Tjian, 1985; Sassone-Corsi and Borelli, 1986). However, evidence has begun to mount which indicates these two classes of elements can be structurally and functionally related, and this has lessened the distinction between them (Maniatis *et al.*, 1987). One theme that has become apparent from the analysis of *trans*-acting factors is that multiple binding sites within an enhancer and/or promoter often exist for a given factor (Gidoni *et al.*, 1984; McKnight and Tjian, 1986). Conversely, multiple factors have been identified which bind to the same or overlapping sites (Zinn and Maniatis, 1986; Jones *et al.*, 1988).

From our studies with the *rbcS-3A* gene, the most highly expressed member of the pea *rbcS* gene family, it is clear that the *cis*-acting elements which contribute to light induction can be very complex. Analysis of 5′ deletion mutations in transgenic plants indicated that 166 bp of 5′ flanking sequences upstream of the transcription initiation site were sufficient to induce the expression of the *rbcS-3A* gene in a light-dependent manner (Kuhlemeier *et al.*, 1987a). Therefore, a light-responsive element (LRE) exists

Table I. Sequence comparison of GT-1 binding sites upstream of *rbcS-3A* gene

II		G	T	G	T	G	G	T	T	A	A	T	A	T	G
II*		G	T	G	A	G	G	T	A	A	T	A	T	C	C
II**	(rev)	T	A	T	G	G	G	T	A	A	C	A	T	T	T
III	(rev)	T	A	G	T	G	A	A	A	A	T	G	A	T	A
III*	(rev)	G	A	G	T	G	T	A	A	A	T	G	T	G	T
III**		T	T	G	T	G	A	A	G	T	A	A	C	A	G
	A	0	3	0	1	0	2	3	4	5	2	3	2	1	1
	G	3	0	5	1	6	3	0	1	0	0	2	0	1	2
	C	0	0	0	0	0	0	0	0	0	1	0	1	1	1
	T	3	3	1	4	0	1	3	1	1	3	1	3	3	2
Consensus		G/T	A/T	G	T	G	Pu	A/T	A	A	A/T	Pu	A/T	–	–

The critical core in box II deduced from the series of 2 bp substitution mutations is underlined. A tally of the bases in each position and a consensus are shown beneath the sequences. All sequences are written from 5′ to 3′. Those labeled (rev) are from the bottom strand of the sequence as in Figure 7C. All other are from the top strand.

downstream of −166 (relative to the *rbcS-3A* transcription initiation site). However, in more recent studies we have demonstrated that the sequences between −410 and −170, which were deleted in the −166 mutant, can act independently as a light-responsive enhancer-like element in transgenic plants (Kuhlemeier *et al.*, 1988a). Fluhr *et al.* (1986a) noted five sequence motifs (boxes I to V) which are conserved between the members of the pea *rbcS* gene family. In the upstream region of the *rbcS-3A* gene, certain of these sequences have homologous but not identical copies (see Table I). These short homologous sequences, present both upstream (boxes II* and III*) and downstream of −170 (boxes II and III), were predicted to play a role in the apparent functional redundancy. To test this hypothesis we introduced substitution mutations in either box II or box III in an *rbcS-3A* gene deleted to position −170. We found that disruption of either box eliminated expression indicating that boxes II and III were part of a positive transcription element (Kuhlemeier *et al.*, 1988a). Boxes II and III may play a dual role in LRE function because they can also decrease transcription in the dark when positioned between a constitutive enhancer and TATA box (Kuhlemeier *et al.*, 1987a).

We recently reported identifying a protein factor from pea nuclei that binds to *rbcS-3A* sequences both upstream and downstream of −170 (Green *et al.*, 1987). DNase I footprinting experiments and competition studies showed that this factor bound to boxes II, III, II* and III*, upstream of *rbcS-3A*. The factor was designated GT-1 because each of the boxes contains a GT-motif. We found that GT-1 binding could be inhibited by methylation of specific G residues within the boxes and suggested these Gs were critical for binding. Based on the results of gel-retardation assays we provided evidence that GT-1 was present in extracts from both light-grown and dark-adapted plants. Based on this observation two modes of GT-1 binding are possible: GT-1 from light-grown and dark-adapted plants could bind to different sites on the DNA or to the same sites.

In this manuscript we have investigated the interaction of GT-1 with *rbcS-3A* upstream sequences in considerably more detail. Using a series of 2 bp substitution mutants we have identified a 6−8 bp core sequence within box II that is

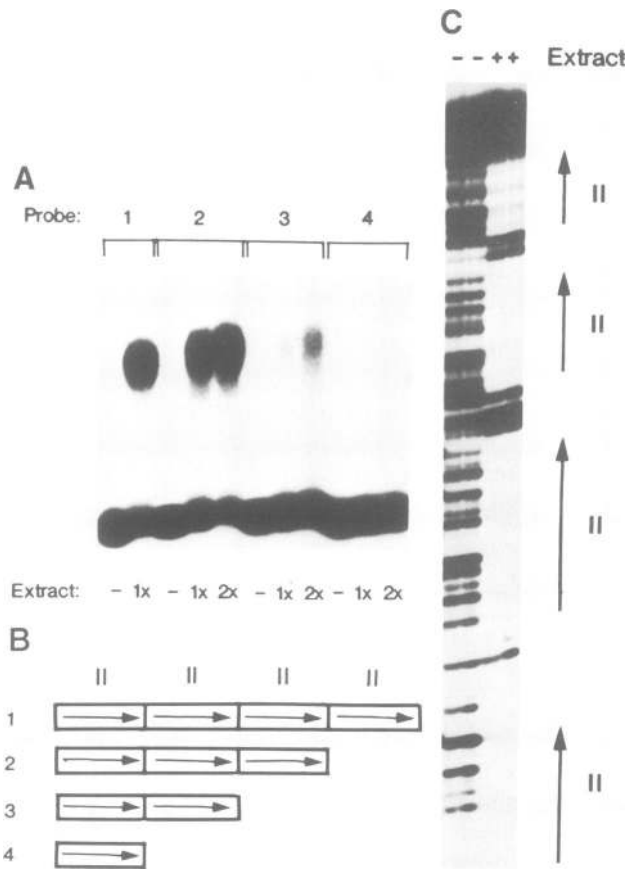

Fig. 1. Binding of GT-1 to cloned oligonucleotides containing one to four copies of box II. (**A**) Gel retardation assays contained 0.4 μg/μl (1×) or 0.8 μg/μl (2×) pea nuclear extract protein as indicated below the lanes. Probe designations above the lanes correspond to those indicated in (B). (**B**) The structure of binding probes containing four (probe 1), three (probe 2), two (probe 3) and one copy of box II (probe 4), used in (A) are shown. A plasmid containing a pair of 84mer oligonucleotides with the structures *Xho*I−box II−*Bam*HI−box II−*Nhe*I−box II−*Bgl*II−box II−*Hin*dIII, was used to generate all probes. It was 5′ end labeled at the *Xho*I site, divided into four samples, and each was cut with a different restriction enzyme before preparative gel electrophoresis and probe isolation. Binding reactions in (A) contained equal molar amounts normalized by radioactivity. (**C**) For DNase I footprinting, a probe containing four tandem copies of box II was prepared from the plasmid described for (B). It was end labeled at *Hin*dIII and cut at *Xho*I to generate the 84 bp probe. The presence (+) or absence (−) of pea nuclear protein extract (2.2 μg/μl) is indicated above the lanes. The positions of the box II sequences are shown to the right. For each 20 μl reaction, 0.2 ng of probe was used. For DNase I cleavage, samples were treated with 2 μl of 10−30 μg/ml DNase I for 1 min.

critical for tobacco and pea GT-1 binding. We have observed a clear correlation between mutations which decrease light-dependent transcriptional activation in transgenic plants and those which decrease binding of GT-1 *in vitro*. This correlation is evident with a 2 bp point mutation as well as with more extensive deletions and substitutions. We have extended our previous results in showing that the footprints obtained with extracts from light-grown and dark-adapted plants cover the same GT-1 sites. A model where GT-1 is an activator that is blocked from functioning in the dark could explain these results.

We have also identified two additional GT-1 binding sites located between −410 and −330. These sites resemble the sequence of those previously mapped and point to several

interesting features that are shared or different among the *rbcS-3A* sequences that interact with GT-1.

Results

GT-1 binds efficiently to cloned oligonucleotides containing three or four tandem copies of box II

Recent work has demonstrated that a 12 bp substitution mutation in box II eliminated detectable expression of an *rbcS-3A* gene with 170 bp of 5′ flanking sequences (Kuhlemeier *et al.*, 1988a). We have also reported that this same mutation decreases GT-1 binding. Our next goal was to define exactly which nucleotides within box II were critical for GT-1 binding. In order to determine the best simple context for efficient screening of box II point mutations, we investigated GT-1 binding to probes containing one to four copies of box II (see Figure 1A and B).

As shown in Figure 1A, GT-1 binds well to probes consisting of four or three tandem copies of box II. Fragments containing two or one copy bind less effectively, perhaps due to their decreased lengths. As shown in Figure 1C, we also used the four-copy box II probe for footprinting experiments. In addition to protection of the box II sequences from DNase I cleavage, GT-1 binding was associated with enhanced cleavage in the linker regions between the sites. These results show that GT-1 binds efficiently to each of the four tandem copies of box II. Based on these data we chose to analyze a series of box II mutations within the context of the four copy derivative.

Box II point mutations define a core sequence critical for GT-1 binding

A set of seven box II mutants was created using synthetic oligonucleotides. The mutants in this series, labeled 2–8 in Figure 2, contain a mutation in the context of the four copy box II derivative such that each of the tandem copies contains the given mutation. Figure 2A shows a gel retardation assay where each of the mutants has been used as a labeled binding probe. Mutations replacing either of the first two GTs have little effect on GT-1 binding, while replacing the double Gs results in a drastic decrease. The latter mutation is particularly significant because it was predicted by our previously reported methylation interference experiments (Green *et al.*, 1987) where methylation of the double Gs was found to inhibit GT-1 binding.

Beyond the double Gs, mutations in the 4–6 bp A/T-rich stretch also markedly decrease binding, while substitution of the last two bp in box II (TG) has little effect (Figure 2A). Binding experiments performed with a cloned synthetic version of the SV40 enhancer 'core' sequence (Weiher *et al.*, 1983) further support the importance of the long A/T-rich stretch for GT-1 binding. Box II is quite homologous to the SV40 core, particularly over the GTGG A/T A/T A/T region, as shown in Figure 2C. Yet, relative to box II very little GT-1 binding is detected using the SV40 probe (Figure 2A, probe 9). The SV40 fragment is also a very poor competitor for GT-1 binding (data not shown). In contrast to the type II and type III boxes, the SV40 sequence lacks the long A/T-rich stretch following the double Gs (Figure 2C).

The pea *rbcS-3A* gene is expressed and regulated normally in transgenic tobacco plants (Fluhr *et al.*, 1986a) and the sequences required for regulated expression have been

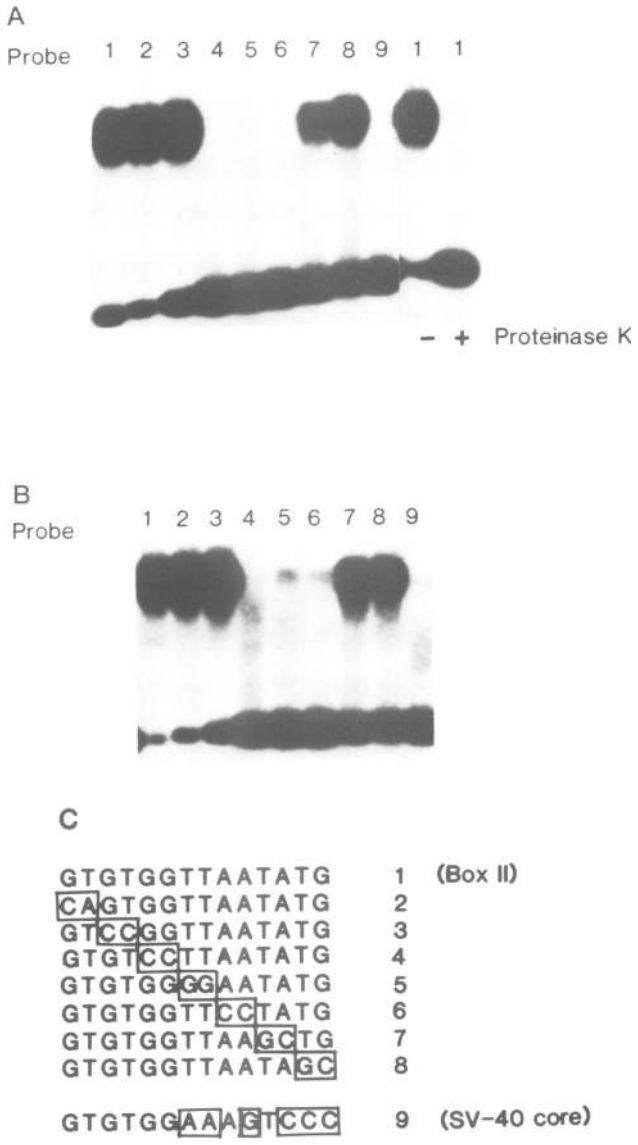

Fig. 2. Gel retardation assays with probes containing 2 bp substitution mutations in box II. (**A**) Assays contained 0.8 μg/μl pea nuclear extract protein and the probes indicated above each lane which correspond to those in (C). For the last two lanes, extracts were preincubated for 15 min at room temperature in the absence (–) or presence (+) of 0.1 μg/μl proteinase K prior to the binding reaction as indicated below the lanes. (**B**) Same as in (A) but with tobacco nuclear protein extract. (**C**) The sequences shown represent the wild-type box II (probe 1), the box II mutant derivatives containing 2 bp substitution mutations (probes 2–8) and the SV40 enhancer core (probe 9). These sequences are tandemly repeated four times in the corresponding 84 bp binding probes used in (A) and (B). The bp that differ from the wild-type box II sequence are boxed. Detailed sequences are available upon request.

studied in some detail. To correlate *in vivo* with *in vitro* data, it is important to look at GT-1 binding from tobacco. The binding assay shown in Figure 2B was performed with a nuclear extract from tobacco leaves. As with pea nuclear extracts, binding to the wild-type box II and the derivatives altered at either end occurred efficiently while substitution of the double Gs or the A/T-rich stretch clearly decreased binding when the tobacco extract was used. Thus the double Gs and the 4–6 A/T bp that follow define a core of residues that are critical for tobacco or pea GT-1 to bind box II in this context.

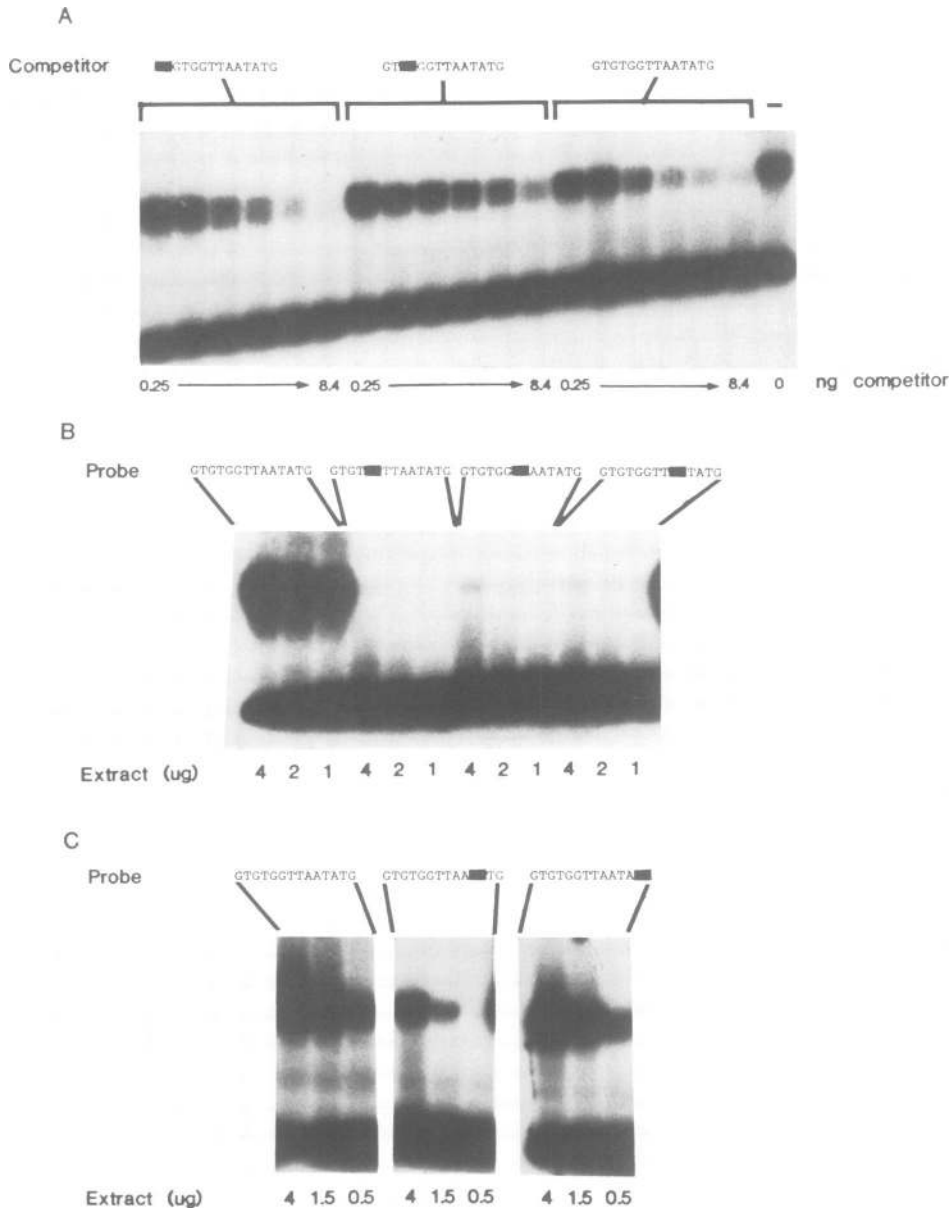

Fig. 3. Titration experiments comparing GT-1 binding to the box II mutants. (**A**) Gel retardation assays (10 μl) contained 1.6 μg nuclear extract protein, and the wild-type box II 84 bp fragment (0.084 ng) as probe. Except for the last lane, reactions also contained the 84 bp competitor (unlabeled) indicated above the lanes, in gradually increasing amounts as shown below the lanes. Each set of six lanes contained 0.25, 0.5, 1, 2, 3 and 8.4 ng competitor respectively. (**B**) and (**C**) Each set of three lanes corresponds to gel retardation assays performed with the probes indicated above the lanes. The amount of nuclear extract protein used in each reaction is shown below the lanes.

To investigate more subtle differences among the mutants we set up competition titration experiments with the box II derivatives, and an example of these is shown in Figure 3A. This experiment demonstrates that the wild-type box II and the box II mutated in the first GT pair have approximately the same competitive ability. Both are slightly better than the box II mutated in the second GT pair. Visual comparison of these results indicates that differences are of the order of ~2-fold. With such small differences (which become most apparent at the lowest binding levels), we have found other forms of quantitation to be less reproducible. We have also used extract titrations to compare the severity of the mutations. This type of assay is more sensitive to differences among box II mutants that bind (and compete) very poorly (see Figure 3B). These analyses indicate that the double G mutation decreases binding >10-fold while mutating the next TT or AA positions decreases binding ~5- to 8-fold (see Figure 3B). Towards the 3′ end of box II, substitution of the TA or the TG results in ~3- to 5-fold, or ~2- to 3-fold decrease in binding, respectively (Figure 3C).

A 2 bp mutation in box II, which eliminates expression in vivo, also eliminates the footprint over box II and decreases the footprint over box III in vitro

Our binding experiments with cloned synthetic oligonucleotides showed that the GG to CC mutation in box II was the most deleterious mutation tested. Recently we introduced this mutation into an *rbcS-3A* gene with 170 bp of 5′ flanking sequence (Kuhlemeier *et al.*, 1988a) by oligonucleotide-directed site-specific mutagenesis (Inouye and Inouye, 1986). In light-grown transgenic tobacco plants containing the GG to CC mutation, no activation of *rbcS-3A*

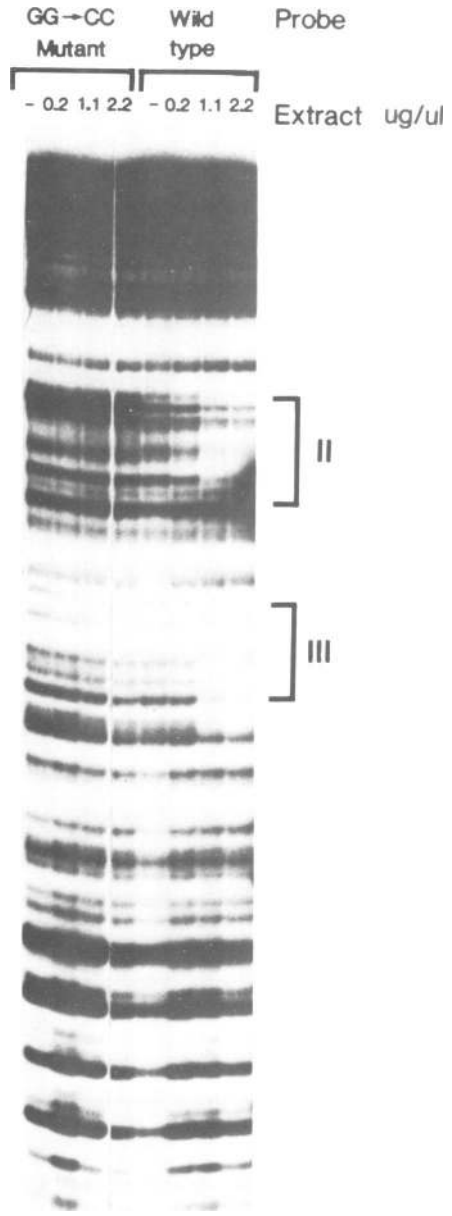

Fig. 4. Footprint titration experiment comparing wild-type and box II mutant probes. As indicated above the gel, the probe for the first four lanes contained a 2 bp substitution mutation in box II (GG changed to CC), while the probe for the next four lanes was wild-type. Both probes (−189 to −50) were 5′ end-labeled at −50 and 0.3 ng was used for each 20 μl reaction. The amount of extract used for each sample is also indicated above the lanes. For DNase I cleavage, samples were treated with 2 μl of a 10−14 μg/ml DNase I for 1 min. The positions of boxes II and III are indicated to the right of the lanes.

transcription could be observed. To see whether the GG to CC mutation also decreased binding in this natural *rbcS-3A* context, we performed a footprint titration experiment (see Figure 4). The probes for this experiment lacked the upstream GT-1 binding sites (boxes II* and III*) and were identical except that the mutant contained the GG to CC substitution in box II. With the wild-type probe, clear footprints over both boxes II and III are evident at the two highest extract concentrations. In contrast, little or no protection of box II is observed with the mutant. Some binding to the box III region is observed with the mutant probe but the protection is less extensive than with the

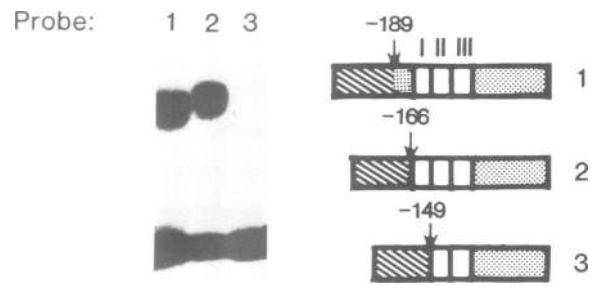

Fig. 5. Gel retardation assays with deletion mutants affected in a region critical for *in vivo* expression. The probes used for each gel retardation assay are indicated above the lanes and diagrammatically to the right of the gel. Reactions contained 0.4 μg/μl of nuclear extract protein. Each corresponds to an *Mst*II−*Bst*XI fragment from an *rbcS-3A* 5′ deletion mutant, whose expression has been characterized previously (Kuhlemeier *et al.*, 1987a) in transgenic plants. *rbcS-3A* upstream sequences extend from the *Bst*XI site at −46 to −189 (probe 1), −166 (probe 2) and −149 (probe 3). The slashed region corresponds to 70 bp of pMON200 DNA included to conserve the fusion junction present in the corresponding transgenic plants.

wild-type. The effect of the box II mutation on binding at box III indicates that in the wild-type configuration GT-1s bound at boxes II and III interact in some way, perhaps cooperatively. In the absence of redundant elements, both boxes are required for expression of *rbcS-3A* in transgenic plants.

Loss of GT-1 binding correlates with loss of transcriptional activation in 5′ deletion mutants

In mature green leaves of transgenic plants, *rbcS-3A* genes with either 189 or 166 bp of 5′-flanking sequences are expressed at the same level as a gene with 410 bp of upstream sequences. Moreover, each of these constructs is regulated normally in response to light and tissue type. In contrast, an *rbcS* gene deleted of all but 149 bp upstream of the transcription start is very poorly expressed (Kuhlemeier *et al.*, 1987a). A series of binding probes was isolated from these mutants in order to compare their interaction with GT-1. In transgenic plants the DNA immediately 5′ of the deletion endpoint is derived from the vector pMON200 (Rogers *et al.*, 1987). Each binding probe was isolated so as to include 70 bp of 5′ vector sequences, thereby preserving each deletion junction as it exists *in vivo* (see Figure 5). Both of the deletion mutants that were expressed in transgenic plants bind GT-1 quite effectively (Figure 5, lanes 1 and 2), whereas the poorly expressed mutant with only 149 bp of 5′-flanking sequences binds more weakly (lane 3). In competition experiments, the −149 deletion mutant also competes less effectively for GT-1 binding activity (data not shown). This deletion removes the first GT pair of box II, substitution of which had little effect on binding in the four copy box II derivative. The different effects of the two mutations could be due to the context of box II or the nucleotides substituted for the GT pair which differ between the two cases.

Similar footprints are observed with extracts from light-grown or dark-adapted plants

Thus far we have observed a very good correlation between mutations which decrease binding and those which decrease transcriptional activation. If GT-1 is an activator for *rbcS-3A* transcription it might be synthesized *de novo* in the light or it might be present under light and dark conditions.

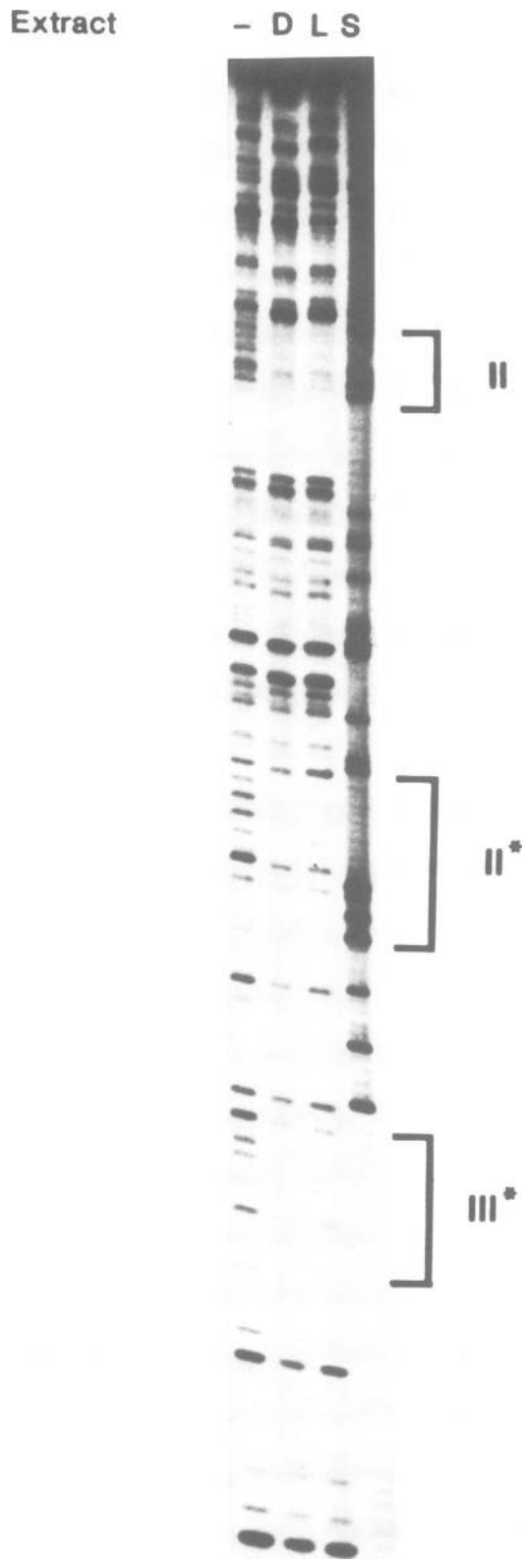

Fig. 6. Comparison of DNase I footprints with extracts from greenhouse-grown and dark-adapted plants. Footprinting was performed with a −330 to −50 fragment from *rbcS-3A* (5′ end-labeled at −330) as described previously (Green *et al.*, 1987). The extracts used are indicated above the lanes. The (**D**) reactions contained 9 μg/μl of nuclear extract protein from dark-adapted plants and the (**L**) reaction 3 μg/μl of nuclear extract protein from greenhouse-grown plants. (−) indicates the minus extract control, and (**S**) the position of Gs on the top strand. GT-1 binding sites corresponding to boxes II, II* and III* are indicated to the right of the lanes.

Therefore, in order to better understand the role of GT-1 it is of interest to characterize the activity of GT-1, in extracts from light-grown and dark-adapted plants. Our initial gel-retardation experiments (Green *et al.*, 1987) showed that extracts from dark-adapted pea leaves contained an activity with the binding and competition characteristics of GT-1. The DNase I footprint experiment shown in Figure 6 was designed to determine whether the GT-1 in extracts from light and dark-adapted plants binds to the same site. The results show that the regions protected by extracts from light-grown and dark-adapted leaves are very similar. This demonstrates that GT-1 is present in extracts from dark-adapted plants, and at least *in vitro*, is capable of binding to the same sites as GT-1 from light-grown plants.

GT-1 binds to a pair of GT-motifs upstream of −330

The *rbcS-3A* upstream region between −410 and −170 contains at least one light responsive enhancer-like element that can function independently of redundant elements located downstream of −170. We have previously described two GT-1 binding sites in the −230 region (boxes II* and III*) which are homologous to boxes II and III and therefore might be associated with the regulatory activity of the upstream element. All four of these GT-1 sites were detected with binding probes consisting of *rbcS-3A* sequences between −330 and −50. Using binding probes extending to position −410 we have obtained evidence for additional GT-1 binding upstream of −330. As shown in the gel retardation assay in Figure 7A, one bound species is observed with either the −410 to −330 probe (probe 3, Figure 7B) or the −330 to −170 probe (probe 2, Figure 7B). The longer fragment from −410 to −170 (probe 1, Figure 7B) forms two bound species as did the −330 to −50 fragment described previously (Green *et al.*, 1987). At present the exact nature of the upper bound species that forms using probes with two pairs of sites is unknown. It may correspond to probes bound at both pairs of boxes. However, if this is true we would expect that the −410 to −50 probe should form three bound species but we have only observed two with this probe as well (data not shown).

The footprint in Figure 7D indicates that the pea nuclear extract protects two sequences in the −410 to −330 fragment from DNase I digestion. The protected sequences (highlighted in Figure 7C) have considerable homology to the type II and type III boxes identified previously. The box at −380 has the double Gs shared by the type II boxes and the box at −350 has the GTG characteristic of the type III boxes. On this basis, we have designated the −380 element box II** and the −350 element box III** as shown in Table I. These protections are considered to be due to GT-1 binding because they are competed away entirely when a fragment containing four copies of box II is included in the footprint binding reaction (see Figure 7D). Binding is also effectively competed away in gel retardation assays with a wild-type, but not a mutant box II competitor (data not shown). This brings to six the number of GT-motifs in the *rbcS-3A* upstream region that interact with nuclear factor GT-1.

Discussion

Correlation of in vitro and in vivo data supports an activator role for GT-1

Here we have described a complete correlation between GT-1 binding and positive activity, i.e. mutations which

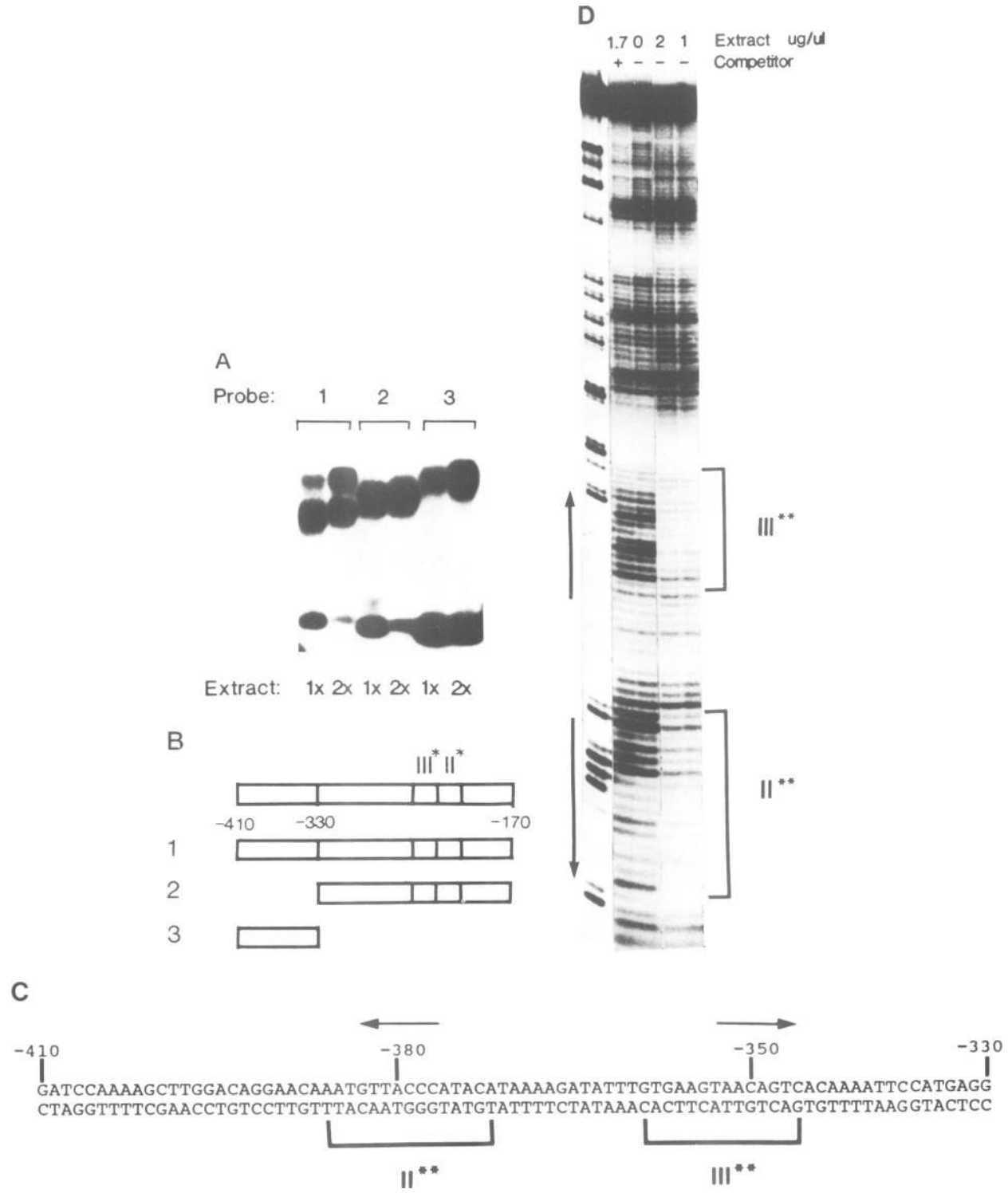

Fig. 7. GT-1 binding to sequences both upstream and downstream of −330. **(A)** Gel retardation assays were as described in Figure 1 except that reactions contained 0.4 μg/μl (1×) or 0.8 μg/μl (2×) nuclear extract protein. Probe designations correspond to those in (B). **(B)** A diagram of the −410 to −170 region of *rbcS-3A* is shown. 1, 2 and 3 correspond to the binding probes used for (A). **(C)** The *rbcS-3A* DNA sequence from −410 to −330 is shown. Lines under the sequence mark the new GT-1 sites, boxes II** and III**, detected in (D). Arrows illustrate the orientation of the GT-motifs in each box. **(D)** The fragment used to generate the GT-1 footprints was 3′ end-labeled at −410 and then cut at −275. The first lane shows the position of G residues and the next four lanes the DNase I cleavage products. The amount of pea nuclear extract protein and the presence (+) or absence (−) of box II competitor DNA (20 ng of an 84 bp fragment containing four tandem copies of box II) is indicated above the lanes. Positions of the footprints corresponding to boxes II** and III** are shown to the right of the lanes. Footprint reactions (20 μl) contained 0.1 ng of probe. For DNase I cleavage, samples were incubated with 1.5−3 μl of 15 μg/ml DNase I for 1 min.

inactivate transcription *in vivo* also decrease binding *in vitro*. This correlation can be made at three levels: deletion mutations, 8−12 bp substitution mutations, and a 2 bp substitution mutation. Our study has included a series of 5′ deletion mutants analogous to those assayed previously in transgenic plants. Both GT-1 binding *in vitro* and *rbcS-3A*

expression *in vivo* (Kuhlemeier *et al.*, 1987a) are decreased in a −149 5′ deletion mutant of *rbcS-3A*, relative to mutants deleted to −166 and −189. The −149 mutant has lost the 5′ end of box II. We have reported previously that 8−12 bp substitution mutations in box II and to a lesser degree those in box III decrease GT-1 binding (Green *et al.*, 1987). This result also correlates with expression requirements in transgenic plants, where both of these boxes are required to activate transcription of the *rbcS-3A* gene when only 170 bp of flanking DNA are present (Kuhlemeier *et al.*, 1988a). We have achieved the highest level of resolution with the 2 bp GG to CC substitution mutation in box II. The 2 bp mutation nearly eliminates binding to synthetic oligonucleotides or, in the natural context, in the absence of redundant elements, drastically decreases GT-1 binding to box II. Extract titrations at the footprint level indicate that the 2 bp mutation also decreases the extent of protection of box III. A model where GT-1 is required for activation of *rbcS-3A* is consistent with all of these observations.

We have shown that GT-1 binding sites can function as light-responsive elements in transgenic plants (Green *et al.*, 1987; Kuhlemeier *et al.*, 1987a). This association prompted us to investigate the nature of GT-1 binding using extracts from light-grown and dark-adapted plants. The footprints in Figure 6 show that both types of extracts protect the same GT-1 sites from DNase I digestion. These *in vitro* results demonstrate that GT-1 is present in both the light and the dark, but they do not prove GT-1 binds in a light independent manner *in vivo*. Relevant to this last point are the observations of Becker *et al.* (1987) which demonstrate that several mammalian factors which are clearly ubiquitous on the basis of *in vitro* footprints, bind in a cell specific manner *in vivo*.

Thus, GT-1 might be an activator that is present in the light and dark but only binds to boxes II and III in the light in plant cells. In the dark, GT-1 binding might be blocked by a repressor not yet detected *in vitro*. A result consistent with the existence of a repressor is the finding that boxes II and III are also capable of decreasing transcription of a constitutive gene in the dark (Kuhlemeier *et al.*, 1987a). This model does not require that GT-1 be specific for light-responsive genes because binding of the repressor could be the pivotal light switch. An analogous mechanism regulates induction of the beta interferon gene where binding of a repressor excludes binding of an activator at an overlapping site (Zinn and Maniatis, 1986).

With this as a working model, our data allow us to make one further prediction; GT-1 binding is required but insufficient for transcriptional activation. Although removal of GT-1 binding sites leads to dramatic decreases in expression *in vivo*, the opposite is not always true. One example of such non-productive binding is that of the cloned box II oligonucleotide (containing four tandem box IIs) discussed above. Like the three copy derivative we reported earlier (Kuhlemeier *et al.*, 1987a), this oligonucleotide fails to activate a reporter gene consisting of a truncated version of the 35S promoter fused to the bacterial CAT gene (M. Cuozzo *et al.*, unpublished). The idea that productive and non-productive binding can occur is not surprising if GT-1 must interact with the TATA factor, RNA polymerase or other activators to enhance transcription. We recently observed that light induction of *rbcS-3A* expression can be inhibited by cycloheximide, so one or more of the factors required for induction may be labile or be synthesized *de novo* in the light (E.Lam, P.J.Green and N.-H.Chua, unpublished results).

Alternative models for GT-1 function

A model ascribing activator function to GT-1 is only the simplest interpretation of our results. Alternative models where GT-1 is a repressor, a platform, or a bifunctional activator/repressor still cannot be ruled out due to the overlap of positive (Kuhlemeier *et al.*, 1988a) and negative (Kuhlemeier *et al.*, 1987a) *cis*-acting elements in the region of boxes II and III. If the overlap is not complete, it should be possible to find point mutations that only inactivate negative or positive function *in vivo*. In the present study we have made considerable progress in identifying the 6 nt core GGTTAA within box II which is critical for GT-1 binding. This information will be essential to interpret future *in vivo* studies designed to delineate the positive and negative elements and thereby further define the role of GT-1 and other *cis*- and *trans*-acting factors mediating *rbcS-3A* expression. A detailed knowledge of the binding site requirements of GT-1 should also be useful in the development of purification strategies based on differential affinity chromatography. Final proof of its role, however, will require the development of a suitable *in vitro* transcription system where the purified factor can be assayed for function.

Implications of multiple binding sites for GT-1

The region between −410 and −50 upstream of the *rbcS-3A* gene contains six sequences that serve as binding sites for nuclear factor GT-1 (see Table I). These sites are arranged in three pairs centered at ~ −150 (boxes II and III), −220 (boxes II* and III*) and −370 (boxes II** and III**). It is likely that each pair of boxes corresponds to two sites rather than a single site for GT-1 binding due to their symmetry. The box II and III GT-motifs face each other in a head to head arrangement, whereas for box II* and III*, and boxes II** and III**, they are arranged tail to tail. If GT-1 recognizes each box as a separate binding site, both arrangements are easy to accommodate. The proximity of the sites could then facilitate the interaction of GT-1 monomers and/or GT-1 binding as a dimer. Our data indicate that GT-1s bound at boxes II and III interact in some way, because the box II GG to CC mutation affects binding at boxes II and III in footprinting experiments. It is not yet known whether an interaction takes place at the other pairs of GT-1 sites. The tail to tail orientation of the starred box pairs differs from the box II/box III pairs and a different G is critical in box III* than in box III. This may suggest that GT-1 binding upstream of −170 differs in some subtle way compared to binding downstream of −170. One feature that the three pairs of GT-1 sites have in common is that for each pair the double Gs in one site and the GTG of the second site are exposed in the major groove on the same face of the double helix. A requirement for this arrangement would predict that one should be able to mimic the effects of a box II mutation by simply changing the spacing between boxes II and III so as to position one of them on the opposite side of the helix from the other.

The role of the *cis* elements upstream of the −170 may be more complicated than those located downstream. Like the sequences downstream of −170, the sequences between

−410 and −170 contain at least one LRE (Kuhlemeier *et al.*, 1987a, 1988a). However, sequences upstream of −166 are not required for maximal expression of *rbcS-3A* in mature leaves but they are required for maximal expression in young leaves (Kuhlemeier *et al.*, 1988a). The function of these sequences in young leaves could be the result of the four additional GT-1 sites, their arrangement or to additional transcription factors interacting with sequences upstream of −170. Now that the GT-1 footprints have been mapped in this region these possibilities can be investigated further.

GT-1 binding sites in other genes

We have observed that GT-1 binds to several other genes in addition to *rbcS-3A*. Strong binding has been observed using probes from other members of the pea *rbcS* gene family, *rbcS-E9* (Kuhlemeier *et al.*, 1988b), and *rbcS-3C*, and from the *Nicotiana plumbaginifolia atp2-1* (P.J.Green, unpublished results) gene encoding the beta subunit of the mitochondrial ATPase (Boutry and Chua, 1985). We have also detected a strong interaction with a fragment from the Cauliflower Mosaic Virus 35S promoter from −343 to +8 relative to the transcription start (F.Nagy and P.J.Green, unpublished). The site of interaction in the 35S fragment may be located upstream of −300 because GT-1 binds poorly to 35S sequences between −300 and +8 (Green *et al.*, 1987). Of the aforementioned genes, only the *rbcS* genes have been reported to be light regulated. The GT-1 binding sites in so-called 'constitutive genes' may correspond to LREs that are masked by other elements. Alternatively they may indicate that GT-1 is a general transcription factor that activates a large set of genes.

Materials and methods

Probe and competitor fragment preparation

DNA manipulations were performed essentially as described by Maniatis *et al.* (1982). Probes were 3′ labeled with Klenow enzyme or 5′ labeled with T4 polynucleotide kinase. Labeled probes and competitor fragments were isolated on 5% polyacrylamide gels in TBE followed by electroelution. For the fragments containing four tandem copies of box II or mutant derivatives, 84mer synthetic oligonucleotides were synthesized on an Applied Biosystems Model 380A DNA synthesizer. Full length products were purified on denaturing polyacrylamide gels, annealed, cloned into a derivative of pEMBL 8+ (Dente *et al.*, 1983), and sequenced. The 84-bp fragments were then excised from the plasmids for use in binding experiments. Specific activity of the probes and concentrations of competitor fragments were determined as described previously (Green *et al.*, 1987).

Nuclear extract preparation

Pea nuclear extracts from leaves of greenhouse-grown plants were used for binding experiments unless otherwise indicated. The preparation of pea nuclear extracts has been described (Green *et al.*, 1987). Tobacco nuclear extracts were prepared using the same methods from greenhouse-grown *Nicotiana tabacum* leaves.

Gel retardation assays

Gel retardation assays were performed essentially as described previously (Green *et al.*, 1987). The amount of nuclear extract for each experiment is indicated in the figure legends.

DNase I footprinting

Unless otherwise indicated in the figure legends, footprinting reactions contained 2 μl of probe mix [2 μg/μl poly(dIdC):poly(dIdC), 1.5 mM EDTA, and the probe as indicated in the figure legends] and 18 μl of pea nuclear extract (Green *et al.*, 1987) and/or EB (40 mM KCl, 0.2 mM EDTA, 25 mM Hepes, pH 6.8 or 7.0, 10% glycerol, 0.5 mM DTT, 0.8 mM PMSF). Following a 30-min incubation at room temperature, DNase I in 25 mM $MgCl_2$ was added as indicated in the figure legends, and incubation was continued for 1 min. Digestion was stopped by the addition of 10 μl of 0.1 M EDTA, 5 μg/μl proteinase K and incubation at 37°C for 20 min, followed by phenol : chloroform extraction and ethanol precipitation. Samples were denatured in formamide dye and resolved on 6% sequencing gels.

Acknowledgements

We thank Dr Cris Kuhlemeier for comments on the manuscript, Elisabeth Goyvaerts for technical assistance, and Wendy Roine for secretarial assistance. This work was supported by a grant from Monsanto Company. P.J.G. and M.C. were supported by NIH postdoctoral fellowships GM11797 and GM10853 respectively.

References

Apel,K. (1979) *Eur. J. Biochem.*, **97**, 183−188.
Becker,P.B., Ruppert,S. and Schutz,G. (1987) *Cell*, **51**, 435−443.
Boutry,M. and Chua,N.-H. (1985) *EMBO J.*, **4**, 2159−2165.
Chen,Z.-L., Schuler,M.A. and Beachy,R.N. (1986) *Proc. Natl. Acad. Sci. USA*, **83**, 8560−8564.
Colot,V., Robert,L.S., Kavanagh,T.A., Bevan,M.W. and Thompson,R.D. (1987) *EMBO J.*, **6**, 3559−3564.
Coruzzi,G., Broglie,R., Edwards,C. and Chua,N.-H. (1984) *EMBO J.*, **3**, 1671−1679.
Dente,L., Cesareni,G. and Cortese,R. (1983) *Nucleic Acids Res.*, **11**, 1645−1655.
Dynan,W.S. and Tjian,R. (1985) *Nature*, **316**, 774−778.
Ferl,R.J. and Nick,H.S. (1987) *J. Biol. Chem.*, **262**, 7947−7950.
Fluhr,R., Kuhlemeier,C., Nagy,F. and Chua,N.-H. (1986a) *Science*, **232**, 1106−1112.
Fluhr,R., Moses,P., Morelli,G., Coruzzi,G. and Chua,N.-H. (1986b) *EMBO J.*, **5**, 2063−2071.
Gallagher,T.F. and Ellis,R.J. (1982) *EMBO J.*, **1**, 1493−1498.
Gallagher,T.F., Jenkins,G.I. and Ellis,R.J. (1985) *FEBS Lett.*, **186**, 241−245.
Gidoni,D., Dynan,W.S. and Tjian,R. (1984) *Nature*, **312**, 409−413.
Green,P.J., Kay,S.A. and Chua,N.-H. (1987) *EMBO J.*, **6**, 2543−2549.
Inouye,S. and Inouye,M. (1986) In Narang,S. (ed.), *Synthesis of DNA, RNA and Their Application*. Academic Press, New York, p. 181.
Jensen,E.O., Marcker,K.A., Schell,J. and deBruijn,F.J. (1988) *EMBO J.*, **7**, 1265−1271.
Jofuku,K.D., Okamuro,J.K. and Goldberg,R.G. (1987) *Nature*, **328**, 734−737.
Jones,N.C., Rigby,P.W. and Ziff,E.B. (1988) *Genes Dev.*, **2**, 267−281.
Kuhlemeier,C., Fluhr,R., Green,P. and Chua,N.-H. (1987a) *Genes Dev.*, **1**, 247−255.
Kuhlemeier,C., Green,P. and Chua,N.-H. (1987b) *Annu. Rev. Plant Physiol.*, **38**, 221−257.
Kuhlemeier,C., Cuozzo,M., Green,P.J., 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.
Maier,U.-G., Brown,J.W.S., Toloczyki,C. and Feix,G. (1987) *EMBO J.*, **6**, 17−22.
Maniatis,T., Fritsch,E.F. and Sambrook,J. (1982) *Molecular Cloning. A Laboratory Manual*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
Maniatis,T., Goodbourn,S. and Fischer,J.A. (1987) *Science*, **236**, 1237−1244.
McKnight,S. and Tjian,R. (1986) *Cell*, **46**, 795−805.
Mikami,K., Tabata,T., Kawata,T., Nakayama,T. and Iwabuchi,M. (1987) *FEBS Lett.*, **223**, 273−278.
Mosinger,E., Batschauer,A., Schafer,E. and Apel,K. (1985) *Eur. J. Biochem.*, **147**, 137−142.
Nagy,F., Kay,S., 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.
Nagy,F., Kay,S.A. and Chua,N.-H. (1988) *Trends Genet.*, **4**, 37−42.
Rogers,S.G., Klee,H.J., Horsch,R.B. and Fraley,R.T. (1987) *Methods Enzymol.*, **153**, 253−277.
Sassone-Corsi,P. and Borelli,E. (1986) *Trends Genet.*, **2**, 215−219.
Silverthorne,J. and Tobin,E.M. (1984) *Proc. Natl. Acad. Sci. USA*, **81**, 1112−1116.
Simpson,J., Schell,J., Van Montagu,M. and Herrera-Estrella,L. (1986a) *Nature*, **323**, 551−554.

Simpson,J., Van Montagu,M. and Herrera-Estrella,L. (1986b) *Science*, **233**, 34–38.
Stockhaus,J., Eckes,P., Rocha-Sosa,M., Schell,J. and Willmitzer,L. (1987) *Proc. Natl. Acad. Sci. USA*, **84**, 7943–7947.
Tobin,E.M. (1981) *Plant Mol. Biol.*, **1**, 35–51.
Tobin,E.M. and Silverthorne,J. (1985) *Annu. Rev. Plant Physiol.*, **36**, 569–593.
Walker,J.C., Howard,E.A., Dennis,E.S. and Peacock,W.J. (1987) *Proc. Natl. Acad. Sci. USA*, **84**, 6624–6628.
Weiher,H., Konig,M. and Gruss,P. (1983) *Science*, **219**, 626–631.
Zinn,K. and Maniatis,T. (1986) *Cell*, **45**, 611–618.

Received on August 23, 1988