The Plant Cell, Vol. 6, 1805–1813, December 1994

Molecular Dissection of GT-1 from Arabidopsis

Kazuyuki Hiratsuka, Xiaodong Wu,[1] Hideya Fukuzawa,[2] and Nam-Hai Chua[3]

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

We isolated and characterized an Arabidopsis cDNA encoding the DNA binding protein GT-1. This protein factor, which contains 406 amino acids, is highly homologous to the previously described tobacco DNA binding protein GT-1a/B2F but is 26 amino acids longer. Recombinant Arabidopsis GT-1, which was obtained from in vitro translation, bound to probes consisting of four copies of pea small subunit of ribulose bisphosphate carboxylase *rbcS-3A* box II and required the same GGTTAA core binding site as the binding activity of an Arabidopsis nuclear protein preparation. However, unlike the truncated tobacco GT-1a prepared from *Escherichia coli* extracts, the full-length Arabidopsis GT-1 bound to pea *rbcS-3A* box III and Arabidopsis chlorophyll *a/b* binding protein *CAB2* light-responsive elements, both of which contain GATA motifs. Deletion and mutational analyses suggested that the predicted trihelix region of GT-1 is essential for DNA binding. Moreover, GT-1 binds to target DNA as a dimer, and its C-terminal region contains a putative dimerization domain that enhances the binding activity. Transient expression of the GT-1::β-glucuronidase fusion protein in onion cells revealed the presence of a nuclear localization signal(s) within the first 215 amino acids of GT-1.

INTRODUCTION

Studies on the function of plant promoters have demonstrated the presence of regulatory *cis*-acting elements that mediate developmental or environmental signals. Analyses of many light-responsive genes, such as the small subunit of ribulose-1,5-bisphosphate carboxylase/oxygenase (*rbcS*) and the chlorophyll *a/b* binding protein (*CAB*), showed that *cis*-acting elements responsible for light regulation are located within the 5′ upstream region (Gilmartin et al., 1990). Light-responsive expression of the pea *rbcS-3A* gene in transgenic tobacco has been used as a model system to define a complex array of *cis*-acting elements that interact with a number of different *trans* factors (Kuhlemeier et al., 1987, 1988; Gilmartin et al., 1990). The nuclear GT-1 activity that binds to six binding sites in the upstream region of *rbcS-3A* was identified as one such factor (Green et al., 1987, 1988). GT-1 binding sites were also found within the promoter regions of several other light-responsive genes and were shown to be involved in light-responsive transcription in vivo (Stockhaus et al., 1987; Manzara and Gruissem, 1988; Dean et al., 1989; Schindler and Cashmore, 1990). Gain-of-function experiments have confirmed the critical role of GT-1 binding sites in mediating light-responsive and tissue-specific gene expression. A synthetic tetramer of the pea *rbcS-3A* box II (GTGTGGTTAATATG), a GT-1 binding site, which was fused to the −90 deleted cauliflower mosaic virus 35S promoter linked to the β-glucuronidase (*GUS*) reporter gene, showed specific expression in chloroplast-containing cells when exposed to light (Lam and Chua, 1990).

A tobacco cDNA encoding the box II DNA binding protein, designated as GT-1a or B2F, has been cloned and partially characterized (Gilmartin et al., 1992; Perisic and Lam, 1992). The predicted structure of the encoded protein contains a novel trihelix motif that does not share significant homology with any other helix-containing DNA binding proteins, such as helix-turn-helix and homeodomain proteins. The relationship between GT-1a/ B2F and nuclear GT-1 was demonstrated by the similar 6-bp core DNA binding sequence requirement and immunological relationship. Notwithstanding, the recombinant GT-1a did not bind to a nuclear GT-1 binding site, *rbcS-3A* box III. In addition to this discrepancy, several important issues surrounding the properties of GT-1a remained unresolved. For example, the DNA binding domain of GT-1a was not defined, and it is not known whether GT-1a binds to target DNA as a monomer or a dimer. However, it was not possible to carry out detailed DNA binding experiments because of the difficulty in expressing recombinant GT-1a in either *Escherichia coli* or a reticulocyte lysate system in vitro. Our inability to express tobacco GT-1a in *E. coli* prompted us to clone GT-1 from another plant species with the hope of alleviating the apparent toxicity problem. We chose to use Arabidopsis because of the presence of GT-1 binding sites within the promoter sequence of several light-responsive genes of this model plant (Donald and Cashmore, 1990; Koncz et al., 1990; K. Hiratsuka and N.-H. Chua, unpublished data). Furthermore, recent advances in Arabidopsis genetic research may enable us to use light-responsive mutants in relation to GT-1 function. As a first step toward understanding the mode of interaction of GT-1 in

[1] Current address: Laboratory of Molecular Biophysics, The Rockefeller University, 1230 York Avenue, New York, NY 10021-6399.
[2] Current address: Department of Agricultural Chemistry, Faculty of Agriculture, Kyoto University, Kyoto 606-01, Japan.
[3] To whom correspondence should be addressed.

Arabidopsis, we conducted a series of experiments to define the functional domains of the GT-1 protein. In this study, we describe the cDNA cloning, expression, and characterization of Arabidopsis GT-1.

RESULTS

Isolation and Sequence Analysis of the Arabidopsis GT-1 cDNA

We screened an Arabidopsis cDNA library constructed in λ ZAPII with a DNA fragment derived from the tobacco GT-1a

```
   1 CTTTTTTTCATTTTGTTCTGAGGGGGCTTTTGTGTTGACGGTGGTGGGAACAAAGGAACA    60

  61 TCGTATCCATTCTGTCCGGGAAACACAGCAATGTTCATTTCCGACAAATCTCGTCCTACT   120
   1                                M  F  I  S  D  K  S  R  P  T    10

 121 GATTTCTACAAAGACGATCATCACAATTCCTCCACCACCAGCACCACACGCGATATGATG   180
  10  D  F  Y  K  D  D  H  H  N  S  S  T  T  S  T  T  R  D  M  M    30

 181 ATCGATGTACTCACCACTACCAACGAATCAGTAGATCTACAATCTCACCACCACCACAAT   240
  30  I  D  V  L  T  T  T  N  E  S  V  D  L  Q  S  H  H  H  H  N    50

 241 CACCACAATCATCATCTCCACCAATCTCAGCCACAACAACAGATTCTCCTCGGAGAAAGC   300
  50  H  H  N  H  H  L  H  Q  S  Q  P  Q  Q  Q  I  L  L  G  E  S    70

 301 AGTGGAGAAGATCACGAAGTTAAAGCACCAAAGAAACGAGCGGAGACATGGGTTCAAGAC   360
  70  S  G  E  D  H  E  V  K  A  P  K  K  R  A  E  T  W  V  Q  D    90

 361 GAAACTCGTAGCTTAATCATGTTCCGTAGAGGTATGGATGGTTTATTCAATACATCCAAA   420
  90  E  T  R  S  L  I  M  F  R  R  G  M  D  G  L  F  N  T  S  K   110

 421 TCTAATAAACATCTCTGGGAACAGATTTCGTCTAAGATGAGAGAAAAAGGGTTTGATCGA   480
 110  S  N  K  H  L  W  E  Q  I  S  S  K  M  R  E  K  G  F  D  R   130

 481 TCTCCGACTATGTGTACTGATAAATGGAGGAATCTGTTGAAAGAGTTTAAGAAAGCTAAG   540
 130  S  P  T  M  C  T  D  K  W  R  N  L  L  K  E  F  K  K  A  K   150

 541 CATCATGATAGAGGAAATGGATCGGCGAAGATGTCGTATTACAAAGAGATTGAAGATATT   600
 150  H  H  D  R  G  N  G  S  A  K  M  S  Y  Y  K  E  I  E  D  I   170

 601 CTTAGAGAGAGGAGCAAAAAAGTGACACCACCACAGTATAATAAGAGCCCTAATACACCA   660
 170  L  R  E  R  S  K  K  V  T  P  P  Q  Y  N  K  S  P  N  T  P   190

 661 CCTACATCAGCCAAAGTTGATTCCTTTATGCAATTTACTGATAAAGGTTTTGATGATACG   720
 190  P  T  S  A  K  V  D  S  F  M  Q  F  T  D  K  G  F  D  D  T   210

 721 AGCATTTCTTTTGGATCCGTTGAAGCTAATGGCAGGCCAGCCTTAAACCTTGAAAGGCGT   780
 210  S  I  S  F  G  S  V  E  A  N  G  R  P  A  L  N  L  E  R  R   230

 781 CTTGATCATGATGGTCATCCTCTTGCAATCACTACAGCAGTTGATGCTGTTGCAGCAAAT   840
 230  L  D  H  D  G  H  P  L  A  I  T  T  A  V  D  A  V  A  A  N   250

 841 GGAGTTACTCCTTGGAATTGGAGAGAGACTCCTGGAAACGGTGATGATAGTCATGGTCAG   900
 250  G  V  T  P  W  N  W  R  E  T  P  G  N  G  D  D  S  H  G  Q   270

 901 CCTTTTGGTGGTAGGGTCATAACAGTGAAATTTGGTGACTATACAAGAAGAATCGGTGTT   960
 270  P  F  G  G  R  V  I  T  V  K  F  G  D  Y  T  R  R  I  G  V   290

 961 GATGGTAGTGCAGAAGCAATCAAAGAGGTAATCAGATCTGCTTTTGGGTTAAGAACTCGA  1020
 290  D  G  S  A  E  A  I  K  E  V  I  R  S  A  F  G  L  R  T  R   310

1021 AGGGCTTTTTGGTTAGAAGATGAAGATCAGATTATTCGCTGTCTTGACCGAGACATGCCC  1080
 310  R  A  F  W  L  E  D  E  D  Q  I  I  R  C  L  D  R  D  M  P   330

1081 TTAGGGAACTATCTACTCCGTCTGGATGATGGACTGGCCATTAGGGTTTGCCATTATGAT  1140
 330  L  G  N  Y  L  L  R  L  D  D  G  L  A  I  R  V  C  H  Y  D   350

1141 GAATCCAACCAATTACCAGTCCATTCAGAAGAGAAAATCTTCTACACTGAAGAAGACTAC  1200
 350  E  S  N  Q  L  P  V  H  S  E  E  K  I  F  Y  T  E  E  D  Y   370

1201 CGCGAGTTTCTGGCTCTACAGGGATGGTCAAGCCTGCAAGTTGATGGTTTTAGGAACATA  1260
 370  R  E  F  L  A  L  Q  G  W  S  S  L  Q  V  D  G  F  R  N  I   390

1261 GAAAACATGGATGATCTTCAACCTGGTGCTGTGTATCGAGGTGTGAGATGAGGAAATGAG  1320
 390  E  N  M  D  D  L  Q  P  G  A  V  Y  R  G  V  R               406

1321 ATCACAAAACTTCTTCTCCAATTCTATCAACAGTTATCTGAACAGAAAACAGTCCCCTGT  1380

1381 AGATATGATCTCATTCTCTTTATACATTTCTTCTTTTTTTCATGTACTTGCTCAAATATG  1440

1441 AAATATCATAACTGGCATTTACCGTACAGACCAAGAAGACCTCAAAATAGATTTTTGGTC  1500

1501 TGTTACATTTGTAAGAACAGAACATAAGAGAAGTGAATAAATTGTTCACAATCAAAAAAA  1560

1561 AAAAAA                                                        1566
```

Figure 1. Nucleotide Sequence and Derived Amino Acid Sequence of Arabidopsis GT-1.

The predicted amino acid sequence derived from the continuous open reading frame is shown in the standard one-letter code below the nucleotide sequence. The nucleotide sequence data of Arabidopsis GT-1 are available in EMBL, GenBank, and DDBJ nucleotide sequence data bases as accession number L36806.

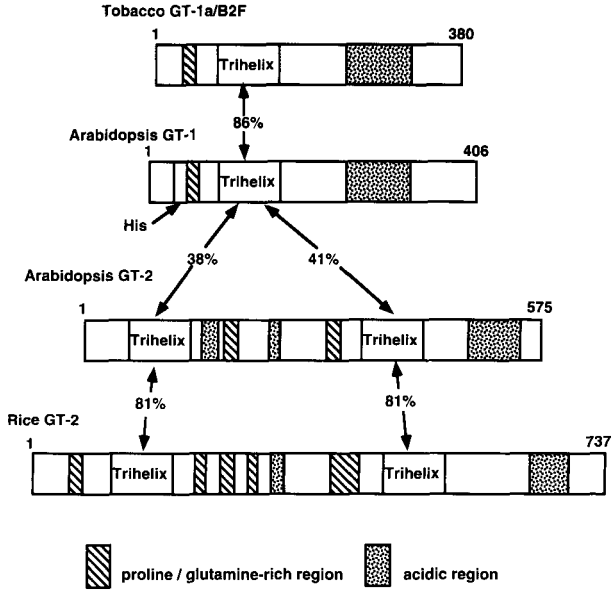


Figure 2. Predicted Structure of Arabidopsis GT-1 and Comparison of GT Box Binding Proteins.

Putative trihelices of tobacco GT-1a/B2F (amino acids 75 to 138; Gilmartin et al., 1992), Arabidopsis GT-1 (amino acids 87 to 150; this study), Arabidopsis GT-2 (amino acids 81 to 161 and amino acids 437 to 512; Kuhn et al., 1993), and rice GT-2 (amino acids 98 to 175 and amino acids 489 to 564; Dehesh et al., 1992) are indicated. The degree of homology between the trihelices is shown. Proline- and glutamine-rich regions and acidic regions are represented by diagonal striped and stippled boxes, respectively.

sequence (Gilmartin et al., 1992) under low-stringency hybridization conditions. Several positively hybridizing cDNA clones were isolated and shown to encode identical sequences. The longest 1.6-kb insert was excised from λ ZAPII, propagated in the plasmid pBluescript II SK+, and subjected to DNA sequence analysis. Figure 1 shows that the longest open reading frame encodes a protein of 406 amino acids, which is highly homologous to, but 26 amino acids longer than, tobacco GT-1a. The predicted secondary structure suggested the presence of a putative trihelix motif contained within the region from Trp-87 to Lys-150. This region showed the highest homology not only to the tobacco GT-1a/B2F putative trihelix region (Gilmartin et al., 1992; Perisic and Lam, 1992), but also to the putative trihelices of GT-2 from Arabidopsis (Kuhn et al., 1993) and rice (Dehesh et al., 1992) (Figure 2).

Previous studies on tobacco GT-1a demonstrated the presence of more than one GT-1a/B2F-related sequence in the tobacco genome. We therefore investigated the complexity of the GT-1 gene family in the Arabidopsis genome by DNA gel blot analysis. Figure 3 shows that digestion of Arabidopsis genomic DNA with four restriction enzymes resulted in a single hybridization band under standard conditions, suggesting that GT-1 is encoded by a single-copy gene.

RNA gel blot analysis showed that the GT-1 mRNA expression level was unaffected by light and displayed no tissue specificity (data not shown). Similar results were obtained previously with the tobacco GT-1a/B2F expression pattern (Gilmartin et al., 1992; Perisic and Lam, 1992).

DNA Binding Properties of Recombinant GT-1 Protein

To study the DNA binding properties of recombinant Arabidopsis GT-1, gel shift assays were performed with the full-length protein expressed via a T7 promoter in a rabbit reticulocyte lysate system. Figure 4 shows that recombinant Arabidopsis GT-1 binds to a probe consisting of four copies of pea *rbcS-3A* box II with high affinity (Green et al., 1988). Moreover, it showed the same requirement for the GGTTAA core sequence as the Arabidopsis nuclear GT-1 activity, although the binding affinity for each box II mutant probe was slightly different (data not shown). This observation is consistent with the previous finding that some differences existed between the binding specificity of tobacco nuclear GT-1 and the recombinant tobacco GT-1a (Gilmartin et al., 1992).

Previous studies demonstrated that tobacco nuclear GT-1 can interact with pea *rbcS-3A* box II and box III (Green et al., 1987) and the rice phytochrome *PHYA* GT2 box (Kay et al., 1989; Dehesh et al., 1990) although with different affinity. On the other hand, *E. coli* extracts containing a truncated tobacco GT-1a showed strong binding only to box II, with no detectable binding to box III (Gilmartin et al., 1992). This observation prompted us to investigate the binding activity of the recombinant Arabidopsis GT-1 to various GT-related motifs. Figure 4 shows the results of gel shifts using probes of tetramers of *rbcS-3A* box III (Sarokin and Chua, 1992) and the rice *PHYA* GT2 box and GT3 box (Dehesh et al., 1992). The full-length

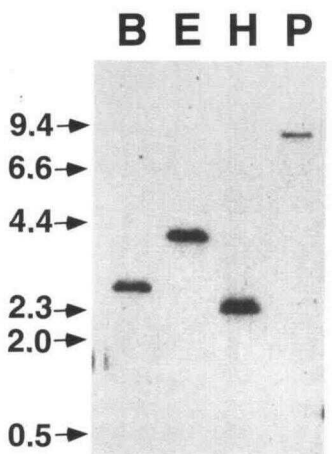


Figure 3. DNA Gel Blot Analysis of Arabidopsis Genomic DNA.

Arabidopsis DNA (1 μg) digested with BamHI, EcoRI, HindIII, and PstI was electrophoresed in lanes B, E, H, and P, respectively, and probed with a radiolabeled BglII- and EcoRI-digested Arabidopsis GT-1 cDNA fragment. Molecular length markers at left are in kilobases.

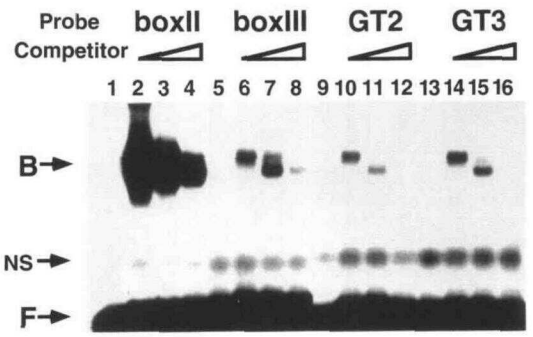


Figure 4. Sequence Specificity for Binding of Arabidopsis GT-1.

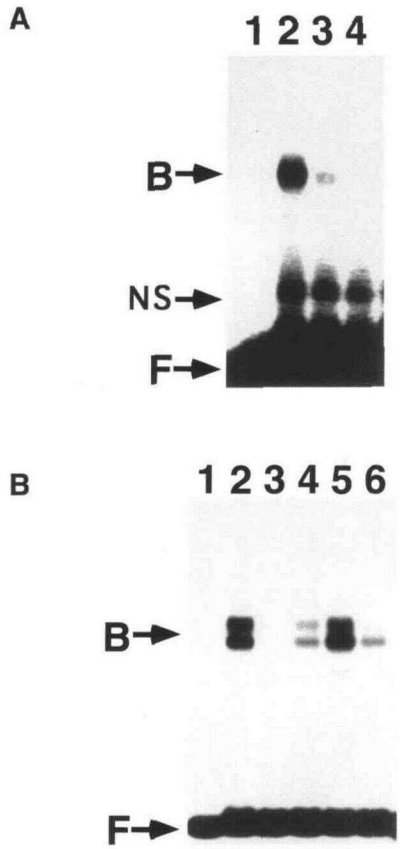
Gel shift assay with transcription–translation products of an Arabidopsis GT-1 cDNA clone in vitro using as probes tetrameric oligonucleotides (0.1 ng) of *rbcS-3A* box II (lanes 1 to 4); *rbcS-3A* box III, TAGTGAAAATGATA (lanes 5 to 8); rice GT2 box, GGCGGTAATTAAC (lanes 9 to 12); and rice GT3 box, CGAGGTAAATCCG (lanes 13 to 16). Either 0.5 ng (lanes 3, 7, 11, and 15) or 2.5 ng (lanes 4, 8, 12, and 16) of the box II competitor was added to the incubation mixtures. Lanes 1, 5, 9, and 13 contain the probe alone. The free probe (F), nonspecific complex (NS), and bound complex (B) are indicated. Two different binding complexes, a faster and a slower migrating band, observed in Figure 4, are due to the presence of multiple binding sites in each tetramer probe. These two bands are not seen in Figure 4, lanes 2 to 4, because of an overexposure of the autoradiogram. With short exposure times, however, two complexes similar to those in Figure 5B were obtained.

Arabidopsis GT-1 protein expressed in a rabbit reticulocyte lysate system formed a specific DNA–protein complex with each of these four probes, with the box II probe showing the highest affinity. Whereas the binding affinity of box III was ~50 times lower than that of box II, it was still higher (~5 to 10 times) than those of GT2 box and GT3 box probes, as demonstrated by competition experiments. Nevertheless, similar gel shift patterns of protein–DNA complexes were obtained with all the four probes, suggesting that the mode of interaction of recombinant Arabidopsis GT-1 with box III, GT2 box, and GT3 box is similar to box II.

The 78-bp fragment of the Arabidopsis *CAB2* promoter region from −111 to −33 was recently shown to confer both circadian- and phytochrome-mediated regulation on the −90 35S promoter (Anderson et al., 1994). Gel shift assays using recombinant Arabidopsis GT-1 protein and the radiolabeled 78-bp fragment showed the formation of a specific protein–DNA complex (Figure 5A). Moreover, competition of box II DNA binding activity using wild-type and mutant *CAB2* promoter sequences demonstrated that the binding activity of recombinant GT-1 is similar to that of CGF-1 (Anderson et al., 1994), which requires the repeated GATA motif for binding (Figure 5B).

Analyses of Deletion and Site-Specific Mutants of GT-1

Previous deletion studies suggested that the N-terminal region of tobacco GT-1a/B2F is needed for DNA binding



CGF-1 site GATAAAGATTACTTCAGATATAACAAACGTTAC

CGF123M site CCCAAACCCTACTTCACCCATAACAAACGTTAC

Figure 5. Arabidopsis GT-1 Binds to the *CAB2* Promoter Sequence.

(A) Gel shift assay with transcription–translation products of an Arabidopsis GT-1 cDNA clone in vitro using an Arabidopsis *CAB2* promoter fragment as a probe (0.1 ng). Lane 1 contains the probe alone; lane 2, recombinant GT-1 and the probe; lane 3, recombinant GT-1, the probe, and 2.5 ng of *rbcS-3A* box II competitor; lane 4, reticulocyte lysate (control) and the probe.

(B) Gel shift assay with transcription–translation products of an Arabidopsis GT-1 cDNA clone in vitro and *rbcS-3A* box II probe (0.05 ng) in the presence of different competitors. Lane 1 contains the probe alone; lane 2, recombinant GT-1 and the probe; lane 3, recombinant GT-1, the probe, and 5 ng of *rbcS-3A* box II competitor; lane 4, recombinant GT-1, the probe, and 10 ng of the CGF-1 binding site (Anderson et al., 1994); lane 5, recombinant GT-1, the probe, and 10 ng of the CGF123M mutant binding site (Anderson et al., 1994); lane 6, recombinant GT-1, the probe, and 10 ng of the Arabidopsis *CAB2* promoter fragment.

The free probe (F), nonspecific complex (NS), and bound complex (B) are indicated. Sequences of the CGF-1 and CGF123M binding sites are shown below. GATA motifs and mutated sequences are underlined. Two different binding complexes, a faster and a slower migrating band, observed in Figure 5B, are due to the presence of multiple binding sites in the tetramer probe.

(Gilmartin et al., 1992). Because only one truncated mutant was used in the previous work, more refined analysis was required to define the region for DNA binding. To this end, we generated mutants of Arabidopsis GT-1 using polymerase chain reaction (PCR) and site-directed mutagenesis. Arabidopsis GT-1 mutants were produced by transcription and translation of the appropriate templates in vitro. The primary structures of the various mutants are presented in schematic forms in Figure 6A. Figure 6B shows that the dN80 mutant of GT-1 can still bind to the box II probe with an affinity equivalent to that of the full-length protein. On the other hand, removal of an additional 10 amino acids from the N terminus totally abolished all DNA binding activity. These data suggest that the region between amino acids 80 and 90, which encodes the putative first helix, is essential for DNA binding.

The importance of the putative trihelix structure for DNA binding was confirmed by site-specific amino acid mutation designed to disrupt the tertiary structure of each of the helices. Figure 7 shows that binding activity to box II is decreased dramatically by the Gln-89 to Pro mutation in helix 1, whereas a similar mutation within helix 2 (Gln-118 to Pro) and helix 3 (Leu-143 to Pro) abolishes DNA binding activity completely.

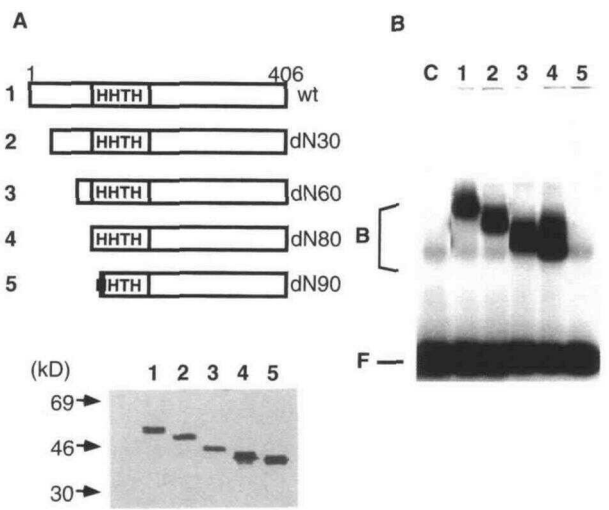


Figure 6. Analysis of N-Terminal Deletion Mutants of Arabidopsis GT-1.

(A) SDS-PAGE analysis of N-terminal deletion mutants of GT-1 proteins generated by transcription and translation in vitro. The primary structures of the mutants are illustrated in schematic form above the electrophoretogram. Numbers refer to the number of amino acid residues deleted from the N terminus. H represents predicted helices 1 to 3; T represents the putative turn region; wt, wild type. Protein molecular mass markers are indicated at left in kilodaltons.

(B) Gel shift analysis of N-terminal deletion mutants of GT-1 using a tetramer of *rbcS-3A* box II as a probe. Equivalent amounts of mutant and wild-type GT-1 proteins were used for SDS-PAGE and for gel shift analysis. Lane C contains the probe alone; lanes 1 to 5 contain translation products 1 to 5 shown in **(A)**, respectively. The free probe (F) and bound complex (B) are indicated. A short exposure of the autoradiogram revealed two closely migrating complexes in lanes 1 to 4.

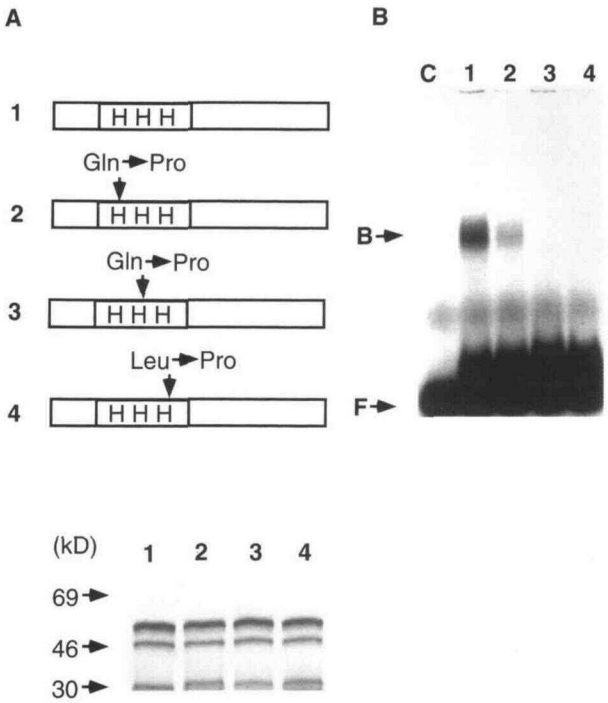


Figure 7. Mutational Analysis of the Putative Trihelix Region.

(A) SDS-urea-PAGE analysis of site-specific mutants of GT-1 proteins generated by transcription and translation in vitro. The primary structures of the mutants are illustrated in schematic form above the electrophoretogram. H represents predicted helices 1 to 3; mutated sites are indicated. Protein molecular mass markers are indicated at left in kilodaltons.

(B) Gel shift analysis of site-specific mutants of GT-1 using a tetramer of *rbcS-3A* box II as a probe. Lane C contains the probe only; lanes 1 to 4 contain translation products 1 to 4 from **(A)**, respectively. The free probe (F) and bound complex (B) are indicated.

Surprisingly, a short deletion (24 amino acids) of the C-terminal region of GT-1 decreased the DNA binding activity drastically and resulted in the formation of a complex with higher mobility. However, further C-terminal deletion to amino acid 146, which removes part of helix 3, abolished DNA binding activity completely (Figure 8). Collectively, these observations provided evidence that the trihelix region is essential for DNA binding and that the C-terminal region is required for high-affinity DNA binding.

GT-1 Binds to the Target Sequence as a Dimer

The result of C-terminal deletion studies of GT-1 prompted us to investigate the function of this region of the molecule. We speculated that the formation of a complex with higher mobility with C-terminal deletion mutants was due to the removal of a dimerization domain that enhances the DNA binding activity of GT-1. To examine this possibility, we used an N-terminal deletion mutant of GT-1, dN80, which retains the trihelix region. This mutant exhibited high DNA binding affinity, but because of its shorter length, it formed a slightly faster migrating complex compared with that of the full-length GT-1. Figure 9A shows that cotranslation of mRNA derived from dN80 and the wild-type GT-1 cDNA in a reticulocyte lysate system leads to the formation of a DNA–protein complex of an intermediate mobility, demonstrating that GT-1 binds as a dimer. However, the intermediate complex was not observed when the two proteins were mixed after translation. Cotranslation of mRNAs derived from the C-terminal deletion mutant dC382 with wild-type GT-1 did not produce any intermediate complex (data not shown). These results suggest that the C-terminal region of GT-1 mediates homodimer formation, which enhances the DNA binding activity.

The importance of the C-terminal region in mediating GT-1 homodimer formation was further confirmed by glutaraldehyde cross-linking experiments. Figure 9B shows that full-length GT-1 and dN80 form a homodimer complex that can be cross-

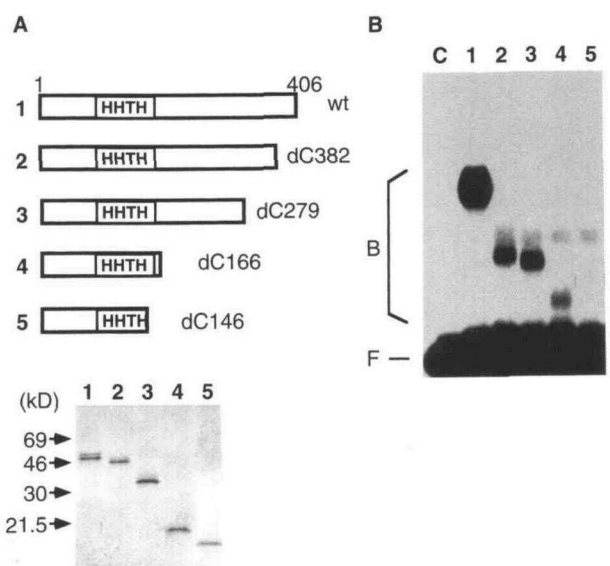


Figure 8. Analysis of C-Terminal Deletion Mutants of Arabidopsis GT-1.

(A) SDS-PAGE analysis of C-terminal deletion mutants of GT-1 generated by transcription and translation in vitro. The primary structures of the mutants are illustrated in schematic form above the electrophoretogram. For each mutant, the number refers to the last amino acid residue. H represents predicted helices 1 to 3; T represents the putative turn region; wt, wild type. Protein molecular mass markers are indicated at left in kilodaltons.

(B) Gel shift analysis of C-terminal deletion mutants of GT-1 using a tetramer of *rbcS-3A* box II as a probe. Equivalent amounts of mutant and wild-type GT-1 proteins were used for SDS-PAGE and gel shift analysis. Lane C contains only the probe; lanes 1 to 5 contain translation products 1 to 5 shown in **(A)**, respectively. A short exposure of the autoradiogram revealed two closely migrating complexes in lane 1. The free probe (F) and bound complex (B) are indicated.

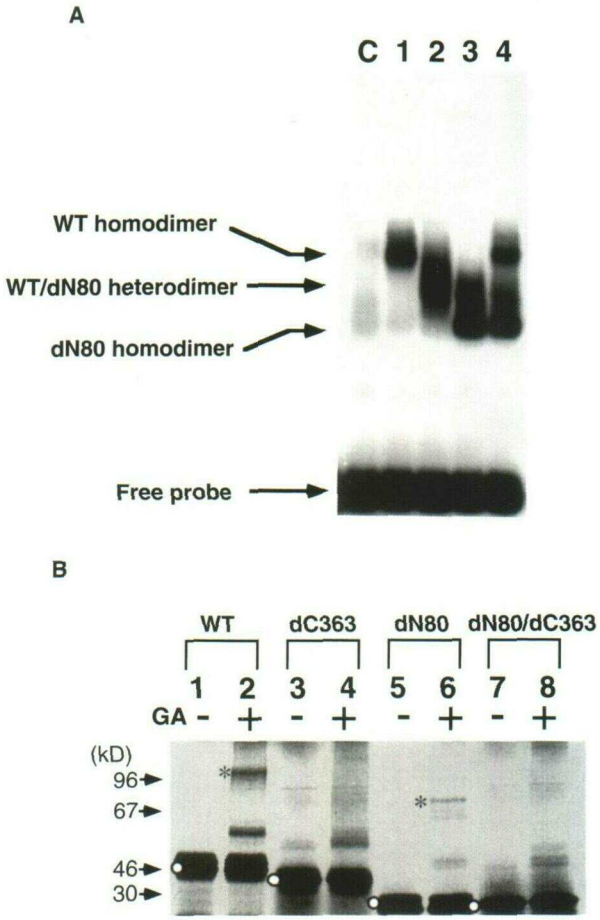


Figure 9. Arabidopsis GT-1 Binds to Target DNA as a Dimer.

(A) Gel shift analysis of the full-length GT-1 protein (lane 1), dN80 (lane 3), and cotranslated (lane 2) or post-translationally mixed full-length GT-1 and dN80 (lane 4) with a tetramer of the *rbcS-3A* box II probe. Lane C contains reticulocyte lysate (control) and the probe. WT, wild type.
(B) Glutaraldehyde cross-linking of Arabidopsis GT-1 deletion mutants. ^{35}S-labeled in vitro–translated proteins were incubated with (+) or without (−) 0.01% glutaraldehyde at 0°C for 60 min. Lanes 1 and 2 contain full-length GT-1 (WT); lanes 3 and 4, dC363; lanes 5 and 6, dN80; lanes 7 and 8, dN80/dC363. Molecular size markers are indicated at left in kilodaltons. The asterisks indicate the position of the dimer, and the open circles indicate the position of the monomer.

linked by glutaraldehyde treatment, whereas dC363 does not. These results also suggest that Arabidopsis GT-1 can form homodimers in solution even in the absence of target DNA.

GT-1 Contains a Nuclear Localization Signal(s)

To determine the intracellular localization of GT-1, a GT-1 cDNA fragment was fused in-frame to the *E. coli GUS* coding sequence. Because the N-terminal portion of GT-1 contains a basic domain within the trihelix motif, which may include putative nuclear localization signals, GT-1 cDNA corresponding to Met-1 to Gly-215 was used for this purpose. The fusion gene was cloned into the plant expression vector VIP26 as described previously (van der Krol and Chua, 1991). The GT-1::*GUS* fusion construct was introduced into onion skin cells by particle gun bombardment, and the fusion protein was expressed under the control of the cauliflower mosaic virus 35S promoter (Shieh et al., 1993). The intracellular localization of the GT-1–GUS activity was determined by histochemical staining with X-gluc. Figure 10A shows an example of the histochemical staining of onion monolayer cells bombarded with the GT-1::*GUS* construct. Note that the blue precipitate was localized within the nucleus. By contrast, GUS activity was observed mainly in the cytoplasm when the *GUS* gene alone was used (Figure 10B). These results demonstrate the presence of a nuclear localization signal(s) within the first 215 amino acids of the Arabidopsis GT-1 molecule.

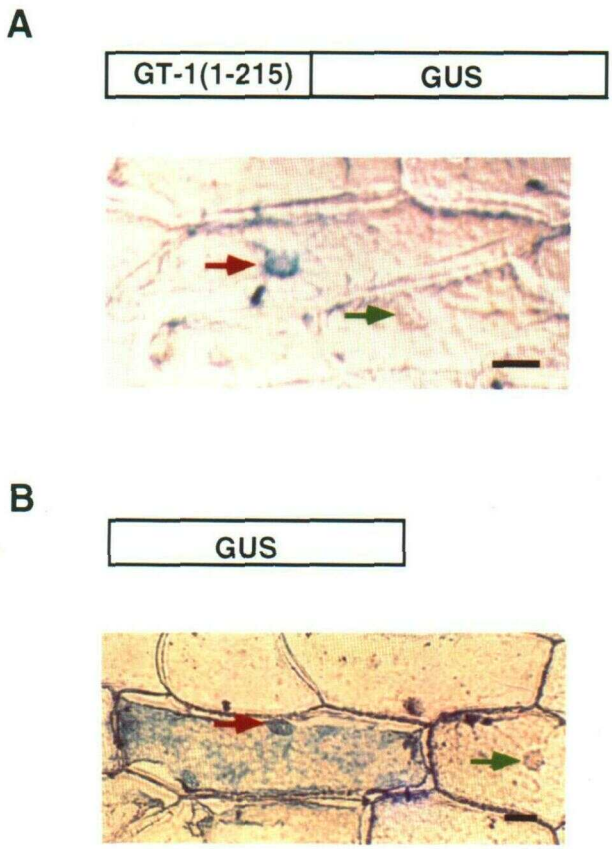


Figure 10. Histochemical Localization of GUS Activity in Onion Skin Cells.

(A) Transfected with 35S–GT-1::*GUS*.
(B) Transfected with 35S–*GUS*.
The nucleus of a transfected (red arrows) cell and that of a nontransfected (green arrows) cell are indicated. Bars = 20 mm.

DISCUSSION

Our study was undertaken in an effort to dissect the molecular structure of the Arabidopsis GT-1 protein because only a limited amount of information was obtained from the earlier studies with tobacco GT-1a/B2F.

The amino acid sequence of Arabidopsis GT-1 deduced from its cDNA nucleotide sequence shows a high degree of homology with tobacco GT-1a/B2F. Homology is especially high within the putative trihelix region, with an amino acid identity of ~86%. Together with the fact that GT-1 is encoded by a single-copy gene within the Arabidopsis genome, these findings lead us to conclude that the cDNA clone described in this report is an Arabidopsis homolog of tobacco GT-1a/B2F. Furthermore, this conclusion is reinforced by the identical DNA binding property of the two proteins with respect to the box II mutant series (data not shown) and the absence of another Arabidopsis gene that is homologous to tobacco GT-1a, as confirmed by DNA gel blot analysis. On the other hand, Arabidopsis GT-1 (406 amino acids) is significantly longer than tobacco GT-1a/B2F (380 amino acids) and contains an additional histidine-rich region ranging from His-46 to His-57. The function of this domain remains to be investigated.

As summarized in Figure 2, two previously reported DNA binding proteins, GT-2 from rice and Arabidopsis, exhibit some homology with Arabidopsis GT-1 and tobacco GT-1a/B2F. The putative trihelix region of Arabidopsis GT-1 shows homology with each of the Arabidopsis GT-2 trihelices: 38 and 41% for the N-terminal domain and the C-terminal domain, respectively. In addition to the homology between the DNA binding domains, GT-1 and GT-2 share specific domains, such as the proline/glutamine–rich and the acidic domains, both of which are presumably involved in the transcriptional regulatory function of the protein. The alignment of these specific amino acid clusters of GT-1 is similar to the C-terminal domain of rice GT-2. These structural similarities suggest a close relationship between two groups of DNA binding proteins.

RNA gel blot analysis of GT-1 transcripts showed no difference in expression level among the samples tested (data not shown), suggesting that GT-1–mediated transcriptional regulation is unlikely to be controlled by the expression level of its mRNA. This result is consistent with the previous observation that box II binding activity is present in nuclear extracts prepared from light-grown as well as dark-adapted plants (Green et al., 1987, 1988). An attractive hypothesis is that Arabidopsis GT-1 activity is regulated at the post-translational level, although much more biochemical characterization of the protein, such as reversible phosphorylation and/or protein–protein interaction, will be needed to elucidate the mechanisms involved.

The GT-2 protein has two separate trihelix DNA binding domains, one at the N terminus and the other at the C terminus, which bind preferentially to the rice *PHYA* GT3 box (GGTAAAT) and GT2 box (GGTAATT), respectively. Although the DNA binding domain of Arabidopsis GT-1 shows similarity with each of the twin trihelix motifs of GT-2, the GT1 box (GGTTAAT) of rice *PHYA*, which shares an identical 6-bp core sequence to the pea *rbcS-3A* box II, is bound only weakly by the N-terminal domain of rice GT-2 and not at all by the C-terminal domain (Dehesh et al., 1992). In this study, we detected only a weak binding activity of Arabidopsis GT-1 to rice GT2 and GT3 boxes, suggesting that GT-1 most likely does not interact with such sequences in vivo. We also found that Arabidopsis GT-1 could bind to the pea *rbcS-3A* box III probe, with a binding affinity clearly higher than those of the GT2 and GT3 boxes; however, the affinity was ~50 times lower as compared with box II. The latter result is consistent with the previous observation that box III has a lower transcription activity in vivo (Gilmartin and Chua, 1990). Previous experiments showed that the recombinant tobacco GT-1a did not show any detectable binding to the GT2 box and box III (Gilmartin et al., 1992). It is noteworthy that the tobacco GT-1 protein used for these experiments was a C-terminal deletion mutant missing the last 244 amino acids. It is possible that the difference in DNA binding activity between Arabidopsis GT-1 and tobacco GT-1a is due to the presence or absence of the C-terminal region that mediates dimerization and, consequently, enhances the DNA binding activity of GT-1.

In this study, we observed specific binding of recombinant Arabidopsis GT-1 to a light-responsive promoter region from the Arabidopsis *CAB2* gene, which contains GATA motifs. These motifs are likely involved in GT-1 interaction because mutations in these sequences blocked binding (Figure 4). Because GATA boxes are frequently found within the promoter region of light-responsive genes (Grob and Stuber, 1987; Gilmartin et al., 1990), GT-1 may be involved in GATA box–mediated transcriptional regulation as well.

The fact that Arabidopsis GT-1 shows a clear binding preference for box II and box III is important when considering the interaction of GT box binding proteins with DNA. The available evidence indicates that Arabidopsis GT-1 binds preferentially to *cis* elements of light-inducible genes (e.g., *rbcS-3A*), whereas the GT-2 protein prefers *cis* elements of genes that are negatively regulated by light (e.g., the GT2 box of rice *PHYA*).

Cotranslation experiments and gel shift assays with full-length GT-1 and the dN80 mutant indicated the formation of a homodimer. Although the dimerization domain was not localized precisely, a 24–amino acid truncation of the C terminus was enough to abolish high-affinity DNA binding to box II, resulting in a faster migrating complex. This result suggests that this C-terminal region is essential for dimer formation. Sequence comparison of the C-terminal region of GT-1 with the protein sequence data base showed no significant homology with previously reported proteins. The dimer formation of Arabidopsis GT-1 is consistent with the fact that two GT-1 binding sites are required for nuclear GT-1 binding in vitro and for light-dependent gene expression in vivo (Green et al., 1988; Gilmartin and Chua, 1990). It is also important because of the possible interaction of GT-1 with other molecules that may contain a compatible dimerization interface. Association of GT-1

with another molecule may generate new DNA binding specificities, leading to the interaction with different target sequences not recognized by the GT-1 homodimer.

Originally, GT-1 was identified as a nuclear protein that binds to light-responsive promoters, although no information was available concerning its intracellular localization. Because the GUS fusion protein strategy has been successfully applied to the study of nuclear localization of plant DNA binding proteins (Raikhel, 1992), we tested a GT-1::*GUS* chimeric construct by transient expression assays. Our results strongly suggest Arabidopsis GT-1 is localized in the nucleus in vivo. Although it is probable that GT-1 contains at least one nuclear localization signal, we could not identify any previously reported nuclear localization signal consensus sequence within the amino acid sequence of GT-1. Further molecular dissection will be necessary to define the amino acid sequence directly involved in nuclear targeting. In addition, because nuclear localization may be a regulatory step of GT-1–mediated transcriptional regulation, studies on the in vivo mechanism of GT-1 intracellular localization will also be interesting.

Taken together, our results provide strong evidence that the Arabidopsis GT-1 described in this report is functionally similar to the nuclear GT-1 that binds to light-responsive elements. The availability of a full-length cDNA clone encoding GT-1 would allow future experiments designed to manipulate the activity of this protein in vivo.

METHODS

General Molecular Biological Techniques

RNA and DNA gel blot analyses as well as other standard molecular cloning techniques were performed as described previously (Sambrook et al., 1989). DNA sequencing was carried out using a Sequenase DNA sequencing kit (U.S. Biochemical Corp.) according to the manufacturer's instructions. DNA and amino acid sequence data were processed using a Macintosh DNASIS (Hitach Software Engineering Co., San Bruno, CA) program software package.

Screening of an Arabidopsis cDNA Library

An Arabidopsis cDNA library in λ ZAPII (Stratagene) made from light-grown Arabidopsis (Columbia) mature plants was screened with tobacco GT-1a cDNA. The bacteriophage library was plated at a density of ~50,000 plaque-forming units per 150-mm-diameter Petri dish. Plaques were lifted onto GeneScreen Plus membranes (Du Pont) and prehybridized for 2 hr at 37°C in 20% formamide, 5 × SSC (1 × SSC is 0.15 M NaCl, 0.015 M sodium citrate), 50 mM Tris-HCl, pH 7.6, 1% SDS, 5 × Denhardt's solution (1 × Denhardt's solution is 0.02% Ficoll, 0.02% PVP, 0.02% BSA), and 50 mg/mL denatured salmon sperm DNA. The probe was prepared by random prime labeling of the isolated tobacco GT-1a EcoRI insert with the Klenow fragment of DNA polymerase I. The labeled probe was added to the prehybridization mix and incubated at 37°C for 12 hr. Filters were washed twice in 2 × SSC, 1% SDS at 60°C for 30 min and exposed to Kodak XAR films (Eastman Kodak Company, Rochester, NY) with intensifying screens.

Deletional and Mutational Studies

N-terminal deletions were constructed by polymerase chain reaction (PCR) using the following primers: dN30, 5′-GGGAAGCTTGCCGCCACCATGATGATCGATGTACTCACCACT-3′; dN60, 5′-GGGAAGCTTGCCGCCACCATGTCTCAGCCACAACAACAGATT-3′; dN80, 5′-GGGAAGCTTGCCGCCACCATGGCACCAAAGAAACGAGCGGAG-3′; dN90, 5′-GGGAAGCTTGCCGCCACCATGCAAGACGAAACTCGTAGCTTA-3′; and the reverse primer 5′-AACAGCTATGACCATG-3′.

C-terminal deletions were generated by exonuclease III digestion of the full-length clone and insertion of an NheI stop codon linker (New England Biolabs, Beverly, MA). dC146 was constructed by PCR using the primer dC146 (5′-CTAAAACTTAAACTCTTTCAACAGATTC-3′) and the M13 −20 primer. Internal mutations were introduced by site-directed mutagenesis using a transformer site-directed mutagenesis kit (Clontech Laboratories, Palo Alto, CA) according to the supplier's instructions.

Gel Shift Studies

Gel shift studies were performed as described previously (Green et al., 1988; Foster et al., 1992). All plasmid constructs were transcribed with T7 polymerase in vitro using an RNA transcription kit (Stratagene); RNAs were translated with a rabbit reticulocyte lysate translation kit (Promega) in the presence of ^{35}S-methionine, according to the manufacturer's instructions. Protein production in each reaction was verified by SDS-PAGE or SDS-urea-PAGE followed by autoradiography.

Transformation of Onion Cells and Histochemical Staining

pBI121 was digested with BamHI and SacI to release the β-glucuronidase (*GUS*) coding sequence. The BamHI-SacI fragment was inserted into a pBluescript II SK+ vector containing a full-length Arabidopsis GT-1 coding sequence to generate the GT-1::*GUS* fusion construct. Protein production of the GT-1::*GUS* construct was first verified by transcription and translation in vitro as previously described. Subsequently, the chimeric coding sequence was introduced into a plant expression vector (van der Krol and Chua, 1991). Onion peel cells were bombarded with 1.6-mm gold particles coated by GT-1::*GUS* or *GUS* DNA using a model PDS-1000/He particle delivery system (Bio-Rad). After incubation at 28°C in the dark for 48 hr, samples were treated with 50 mM sodium phosphate buffer containing 10 mg/mL X-gluc (5-bromo-4-chloro-3 indolyl-β-D-glucuronic acid) for 2 to 12 hr at 37°C (Jefferson et al., 1987). Intracellular localization of GUS activity was observed using a Zeiss Axioskop microscope (Carl Zeiss Instruments, Inc., Pelham, NY) and photographed with a Nikon FX-35A camera (Nikon Inc., Melville, NY).

ACKNOWLEDGMENTS

We thank Drs. Shawn Anderson and Steve Kay for sharing their unpublished results and the CGF-1 wild-type and mutant binding sites and Dr. Simon Barnes for critical reading of the manuscript. K.H. was supported by a fellowship from the Japanese Society for the Promotion of Science. This work was supported by National Institutes of Health Grant No. GM 44640.

Received August 15, 1994; accepted October 13, 1994.

REFERENCES

Anderson, S.L., Teakle, G.R., Martino-Catt, S.J., and Kay, S.A. (1994). Circadian clock- and phytochrome-regulated transcription is conferred by a 78 bp *cis*-acting domain of the *Arabidopsis CAB2* promoter. Plant J., **6,** 457–470.

Dean, C., Pichersky, E., and Dunsmuir, P. (1989). Structure, evolution and regulation of *rbcS* genes of higher plants. Annu. Rev. Plant Physiol. Plant Mol. Biol. **40,** 415–439.

Dehesh, K., Bruce, W.B., and Quail, P.H. (1990). A *trans*-acting factor that binds to a GT-motif in a phytochrome gene promoter. Science **250,** 1397–1399.

Dehesh, K., Hung, H., Tepperman, J.M., and Quail, P.H. (1992). GT-2: A transcription factor with twin autonomous DNA-binding domains of closely related but different target sequence specificity. EMBO J. **11,** 4131–4144.

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

Foster, R., Gasch, A., Kay, S., and Chua, N.-H. (1992). Analysis of protein/DNA interactions. In Methods in Arabidopsis Research, C. Koncz, N.-H. Chua, and J. Schell, eds (Singapore: World Scientific Publishing), pp. 378–392.

Gilmartin, P.M., and Chua, N.-H. (1990). Localization of a phytochrome responsive element within the upstream region of a pea *rbcS-3A*. Mol. Cell. Biol. **10,** 5565–5568.

Gilmartin, P.M., Sarokin, L., Memelink, J., and Chua, N.-H. (1990). Molecular light switches for plant genes. Plant Cell **2,** 369–378.

Gilmartin, P.M., Memelink, J., Hiratsuka, K., Kay, S.A., and Chua, N.-H. (1992). Characterization of a gene encoding a DNA binding protein with specificity for a light-responsive element. Plant Cell **4,** 839–849.

Green, P.J., Kay, S.A., and Chua, N.-H. (1987). Sequence-specific interactions of a pea nuclear factor with light-responsive elements upstream of the *rbcS-3A* gene. EMBO J. **6,** 2543–2549.

Green, P.J., Yong, M.-H., Cuozzo, M., Kano-Murakami, Y., Silverstein, P., and Chua, N.-H. (1988). Binding site requirements for pea nuclear protein factor GT-1 correlate with sequence required for light-dependent transcriptional activation of the *rbcS-3A* gene. EMBO J. **7,** 4035–4044.

Grob, U., and Stuber, K. (1987). Discrimination of phytochrome dependent light inducible from non-light inducible plant genes: Prediction of a common light-responsive element (LRE) in phytochrome dependent light inducible plant genes. Nucl. Acids Res. **15,** 9957–9973.

Jefferson, R.A., Kavanagh, T.A., and Bevan, M.W. (1987). GUS fusions: β-Glucuronidase as a sensitive and versatile gene fusion marker in higher plants. EMBO J. **6,** 3901–3907.

Kay, S.A., Keith, B., Shinozaki, K., Chye, M.-L., and Chua, N.-H. (1989). The rice phytochrome gene: Structure, autoregulated expression, and binding of GT-1 to a conserved site in the 5′ upstream region. Plant Cell **1,** 351–360.

Koncz, C., Mayerhofer, R., Koncz-Kalman, Z., Nawrath, C., Reiss, B., Redei, G.P., and Schell, J. (1990). Isolation of a gene encoding a novel chloroplast protein by T-DNA tagging in *Arabidopsis thaliana*. EMBO J. **9,** 1337–1346.

Kuhlemeier, C., Fluhr, R., Green, P.J., and Chua, N.-H. (1987). Sequences in the pea *rbcS-3A* gene have homology to constitutive mammalian enhancers but function as negative regulatory elements. Genes Dev. **1,** 247–255.

Kuhlemeier, C., Cuozzo, M., Green, P.J., Goyvaerts, E., Ward, K., and Chua, N.-H. (1988). Localization and conditional redundancy of regulatory elements in *rbcS-3A*, a pea gene encoding the small subunit of ribulose-bisphosphate carboxylase. Proc. Natl. Acad. Sci. USA **85,** 4662–4666.

Kuhn, R.M., Casper, T., Dehesh, K., and Quail, P.H. (1993). DNA binding factor GT-2 from *Arabidopsis*. Plant Mol. Biol. **23,** 337–348.

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

Manzara, T., and Gruissem, W. (1988). Organization and expression of the genes encoding ribulose-1,5-bisphosphate carboxylase in higher plants. Photosynth. Res. **16,** 117–139.

Perisic, O., and Lam, E. (1992). A tobacco DNA binding protein that interacts with a light-responsive box II element. Plant Cell **4,** 831–838.

Raikhel, N. (1992). Nuclear targeting in plants. Plant Physiol. **100,** 1627–1632.

Sambrook, J., Fritsch, E.F., and Maniatis, T. (1989). Molecular Cloning: A Laboratory Manual, 2nd ed. (Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press).

Sarokin, L.P., and Chua, N.-H. (1992). Binding sites for two novel phosphoproteins, 3AF5 and 3AF3, are required for *rbcS-3A* expression. Plant Cell **4,** 473–483.

Schindler, U., and Cashmore, A.R. (1990). Photoregulated gene expression may involve ubiquitous DNA binding proteins. EMBO J. **9,** 3415–3427.

Shieh, M.W., Wessler, S.R., and Raikhel, N.V. (1993). Nuclear targeting of the maize R protein requires two nuclear localization sequences. Plant Physiol. **101,** 353–361.

Stockhaus, J., Eckes, P., Rocha-Sosa, M., Schell, J., and Willmitzer, L. (1987). Analysis of *cis*-acting sequences involved in the leaf-specific expression of a potato gene in transgenic plants. Proc. Natl. Acad. Sci. USA **84,** 7943–7947.

van der Krol, A.R., and Chua, N.-H. (1991). The basic domain of plant B-ZIP proteins facilitates import of a reporter protein into plant nuclei. Plant Cell **3,** 667–675.