Characterization of a Gene Encoding a DNA Binding Protein with Specificity for a Light-Responsive Element

Philip M. Gilmartin,1 Johan Memelink,2 Kazuyuki Hıratsuka, Steve A. Kay,3 and Nam-Hai Chua4
Laboratory of Plant Molecular Biology, The Rockefeller University, 1230 York Avenue, New York, New York 10021-6399

The sequence element of box II (GTGTGGTTAATATG) is a regulatory component of a light-responsive element present within the upstream region of pea rbcS-3A. The nuclear protein GT-1 was defined previously as a DNA binding activity that interacts with box II. Here, we describe the isolation and characterization of cDNA sequences that encode a DNA binding protein with specificity for this element. The recombinant protein, tobacco GT-1a, shows similar sequence requirements for DNA binding to nuclear GT-1, as assayed by its ability to interact with previously defined 2-bp scanning mutations of box II, and is shown to be immunologically related to nuclear GT-1. The predicted structure of the 43-kD protein derived from the cDNA sequence suggests the presence of a novel helix-helix-turn-helix (HHTH) motif. Comparison between the predicted protein sequence encoded by the tobacco GT-1a cDNA and that of another GT binding protein, rice GT-2, reveals strong amino acid conservation over the HHTH region; this motif appears to be involved in the interaction between the recombinant protein and box II. Genomic DNA gel blot analysis indicated the presence of a small gene family of related sequences within the tobacco nuclear genome. RNA gel blot analysis of tobacco mRNA using the isolated cDNA as a probe showed that transcripts are present in several tissues, including both light-grown and dark-adapted leaves.

INTRODUCTION

Light plays a critical role in regulating plant growth and development through the modulation of transcription levels of light-responsive genes. Following photoperception, by specific photoreceptors, the signal is subsequently transduced through as yet undefined pathways to mediate the transcriptional response (Kuhlemeier et al., 1987a; Silverthorne and Tobin, 1987; Jenkins, 1988; Nagy et al., 1988). Our studies on light-responsive transcription have focused on the pea ribulose bisphosphate carboxylase small subunit 3A (rbcS-3A) gene. Light-responsive expression of rbcS-3A in tobacco is mediated by a complex array of cis-acting elements (Kuhlemeier et al., 1987b, 1988, 1989; Davis et al., 1990; Gilmartin et al., 1991). Several nuclear protein factors that interact with these elements have been identified (see Gilmartin et al., 1990). One of these factors, GT-1, binds to six binding sites present in the upstream region of rbcS-3A (Green et al., 1987, 1988a). Similar sequence motifs are also present within several other light-responsive genes (Stockhaus et al., 1987; Manzara and Gruissem, 1988; Dean et al., 1989; Elliot et al., 1989; Kay et al., 1989; Dehesh et al., 1990; Schindler et al., 1990; Kay, 1991; Lawton et al., 1991). The –166 deleted rbcS-3A promoter retains light responsiveness (Kuhlemeier et al., 1987); the light-responsive element (LRE) that mediates this response is located between positions –166 and –55 (Kuhlemeier et al., 1989; Davis et al., 1990). This LRE contains two binding sites for GT-1, box II (–151GTGTGGTTAATATG–138) and box III (–128ATCATTTTCACT–114) (Green et al., 1987); both of these boxes are essential for the phytochrome-responsive activity conferred by this element (Kuhlemeier et al., 1988; Gilmartin and Chua, 1990a, 1990b). There is a strong correlation between the affinity of GT-1 for box II and box III in vitro and the level of transcriptional activity conferred by these elements in vivo (Gilmartin and Chua, 1990a). It has been shown that box III is a weak binding site for GT-1 as compared to box II. Replacement of box II by box III within the –166 promoter results in a 95% reduction in transcriptional activity as compared with wild-type levels (Gilmartin and Chua, 1990a). Interactions between box III and GT-1 may play only a minor role in the transcriptional activity conferred by the LRE.

The critical role of box II as a regulatory sequence, as opposed to a purely quantitative element, was established by gain-of-function experiments in which a synthetic tetramer of box II was fused to the –90 deleted cauliflower mosaic virus 35S promoter. In this context, the box II tetramer was able to confer light responsiveness upon the heterologous light-insensitive promoter (Lam and Chua, 1990). Further studies demonstrated a requirement for an interaction between the
box II tetramer and an element present between positions -90 and -48 of the cauliflower mosaic virus 35S promoter, most likely as T (Davis et al., 1990). These studies illustrate that GT-1 binding is required but not necessarily sufficient for light-responsive transcription. In addition, they establish a regulatory role for the GT-1 box II binding site in the transcriptional light response.

With the long-term aim of elucidating the signal transduction pathway that links photoperception to the box II-mediated transcriptional response of rbcS-3A outlined above, we have isolated cDNA sequences encoding a DNA binding protein with specificity for box II. Here, we describe the characterization of the cDNA sequences and the encoded protein.

RESULTS

Isolation of a cDNA Encoding a DNA Binding Protein with Specificity for rbcS-3A Box II

Definition of the critical role of rbcS-3A box II as a regulatory component of the LRE prompted us to attempt to isolate cDNA sequences that encode the nuclear proteins through which box II-mediated light-responsive transcription is modulated. Based on the observation that GT-1 binding activity can be detected in nuclear extracts prepared from plants grown in both the light and the dark, we expected GT-1 transcripts to be present in RNA samples prepared from both light- and dark-grown tissue. Because light-inducible transcripts, for example, rbcS and chlorophyll a/b binding (cab) transcripts, are reduced in abundance in dark-grown tissue, low-abundance mRNA sequences present in both light-grown and dark-adapted plants would be relatively more highly represented in RNA prepared from dark-grown tissue.

Following a slightly modified version of the screening protocol for the isolation of phase encoding DNA binding proteins (Singh et al., 1989; Katagiri et al., 1990), we screened 6 x 10^6 phages of an etiolated tobacco cDNA library constructed in ZAP with a box II tetramer (Gilbert et al., 1991). From this screen, we obtained a single recombinant phage that encoded a DNA binding protein with specificity for the box II tetramer. The 1.4-kb insert was excised from λ ZAP, propagated in the plasmid pBluescript II SK+ and subjected to DNA sequence analysis; this sequence is referred to as clone A. This cDNA insert was used as a probe to screen 1 x 10^6 phages of a second cDNA expression library prepared from light-grown tobacco leaf mRNA (Stratagene). A recombinant phage was isolated from this screen that also contains a 1.4-kb insert; this sequence is referred to as clone B. Comparison of the two cDNA sequences revealed that they are almost identical with the following exceptions. The second sequence, clone B, lacks 27 bp at the 5’ end as compared to clone A but contains an additional 27 bp including a poly(A) tract at the 3’ end. These differences can be accounted for by the extent of cDNA synthesis at the 5’ end and by the presence of an EcoRI restriction.

Figure 1. Nucleotide Sequence of Tobacco GT-1a Derived from Two Independent cDNA Clones and Partial Genomic Sequence Data.

The sequence is numbered from the first nucleotide of the EcoRI linker (a regulatory role for the GT-1 box II binding site in the transcriptional response of rbcS-3A outlined above, we have isolated cDNA sequences encoding a DNA binding protein with specificity for box II. Here, we describe the characterization of the cDNA sequences and the encoded protein.

RESULTS

Isolation of a cDNA Encoding a DNA Binding Protein with Specificity for rbcS-3A Box II

Definition of the critical role of rbcS-3A box II as a regulatory component of the LRE prompted us to attempt to isolate cDNA sequences that encode the nuclear proteins through which box II-mediated light-responsive transcription is modulated. Based on the observation that GT-1 binding activity can be detected in nuclear extracts prepared from plants grown in both the light and the dark, we expected GT-1 transcripts to be present in RNA samples prepared from both light- and dark-grown tissue. Because light-inducible transcripts, for example, rbcS and chlorophyll a/b binding (cab) transcripts, are reduced in abundance in dark-grown tissue, low-abundance mRNA sequences present in both light-grown and dark-adapted plants would be relatively more highly represented in RNA prepared from dark-grown tissue.

Following a slightly modified version of the screening protocol for the isolation of phase encoding DNA binding proteins (Singh et al., 1989; Katagiri et al., 1990), we screened 6 x 10^6 phages of an etiolated tobacco cDNA library constructed in ZAP with a box II tetramer (Gilbert et al., 1991). From this screen, we obtained a single recombinant phage that encoded a DNA binding protein with specificity for the box II tetramer. The 1.4-kb insert was excised from λ ZAP, propagated in the plasmid pBluescript II SK+ and subjected to DNA sequence analysis; this sequence is referred to as clone A. This cDNA insert was used as a probe to screen 1 x 10^6 phages of a second cDNA expression library prepared from light-grown tobacco leaf mRNA (Stratagene). A recombinant phage was isolated from this screen that also contains a 1.4-kb insert; this sequence is referred to as clone B. Comparison of the two cDNA sequences revealed that they are almost identical with the following exceptions. The second sequence, clone B, lacks 27 bp at the 5’ end as compared to clone A but contains an additional 27 bp including a poly(A) tract at the 3’ end. These differences can be accounted for by the extent of cDNA synthesis at the 5’ end and by the presence of an EcoRI restriction.
site at the 3' end that was probably cleaved in clone A during construction of the first library. The 5' and 3' untranslated regions of clones A and B are otherwise identical, suggesting that they are derived from the same gene.

One further difference between these two cDNA sequences lies within the open reading frame. In a comparison of clone A to clone B, 2 bp (positions 466 and 467) are missing from clone A, causing a frameshift mutation. In addition, comparing clone B with clone A, 1 bp at position 349 is missing from clone B, causing a frameshift mutation; clone B also has a single base pair deletion at position 855. The discrepancy between these two cDNA sequences was resolved by polymerase chain reaction amplification of the corresponding region from tobacco genomic DNA. The amplified fragments were ligated into pBluescript II SK+, and several independent recombinant clones were subjected to sequence analysis. The genomic sequence spanning this region contains the single base pair deletion at position 855. The discrepancy between the two sequences and demonstrated the presence of a continuous open reading frame within the basic region. In addition, a putative, proline residue-containing turn region is predicted between helix 2 and helix 3. The theoretical structure of the protein including this predicted helix-helix-turn-helix (HHTH) motif is shown in Figure 2A.

Comparison of the amino acid sequence of tobacco GT-1a with that of rice GT-2, which binds to a motif within the rice phytochrome promoter containing the sequence GGTAAT, reveals a 48% amino acid sequence (27 of 64) identity over the predicted HHTH region (Figure 2B). The remainder of the two proteins do not show any sequence homology. The HHTH region of tobacco GT-1a does not have any striking amino acid homology to any other helix-containing DNA binding proteins such as helix-loop-helix (HLH) and homeodomain proteins.

Figure 2. Predicted Secondary Structure of Tobacco GT-1a.

(A) The acidic domains at the extreme 5' and 3' termini of the protein are indicated as is the basic domain that extends throughout the three indicated helices. The putative turn region between helix 2 and helix 3 is indicated by a T. The putative DNA binding/dimerization and activation domains are shown.

(B) Amino acid sequence comparisons between the conserved region of tobacco GT-1a and rice GT-2. The standard one-letter code is used. The tobacco GT-1a sequence between amino acids 75 and 380 is shown; the GT-2 sequence (Dehesh et al., 1990) between amino acids 87 and 150 is shown. Helices 1, 2, and 3 of tobacco GT-1a are indicated as are helices 1 and 2 of rice GT-2, as defined by Dehesh et al. (1990).

Amino acid identities between tobacco GT-1a and rice GT-2 are indicated by an asterisk.

Binding Site Specificity of Tobacco GT-1a

To define the binding site specificity of tobacco GT-1a, DNA-protein filter binding assays were performed with wild-type and mutant binding sites for the two nuclear proteins activation sequence factor 2 (ASF-2) (Lam and Chua, 1989) and GA Factor 1 (GAF-1) (J. Memelink, P. M. Gilmartin, and N.-H. Chua, manuscript in preparation) in comparison with the GT-1 box II binding site and its mutant box II<sup>mut</sup> derivative. Figure 3A shows these results as well as the sequence of the wild-type and mutant binding sites. These data demonstrate that tobacco GT-1a shows specificity for box II but is unable to bind to its mutant derivative, box II<sup>mut</sup>, or to any of the other four sequences tested.

From our analyses of rbcS-3A, as well as studies of GT-1 binding sites present within the upstream regions of the rice phytochrome A (phyA) gene (Kay et al., 1989; Kay, 1991) and similar sequences within other regulatory elements (Schindler et al., 1990; Lawton et al., 1991), it is apparent that nuclear GT-1 can interact with several distinct, yet closely related sequence motifs. Dehesh et al. (1990) have demonstrated that recombinant rice GT-2, which binds to the rice phyA GT motif GCCGTAATT, interacts only weakly with both the rice phyA TAGGTAAAT motif and the similar rbcS-3A box II element. This observation prompted us to assay the specificity of tobacco GT-1a for different nuclear GT-1 binding sites.

These studies were performed by assaying phage containing the tobacco GT-1a cDNA clone A for the ability of the
encoded protein to bind to the previously defined \textit{rbcS-3A} nuclear GT-1 binding sites box II, its mutant derivative box II$'^m'$, and box III. In addition, the rice GT-2 binding site was assayed. This last sequence was originally identified as a binding site for nuclear GT-1 (Kay et al., 1989) and subsequently shown also to be the specific target sequence of the recombinant rice DNA binding protein GT-2 (Dehesh et al., 1990). The results presented in Figure 3B demonstrate that tobacco GT-1a shows dramatically reduced binding to box II$'^m'$ and to box III as well as to the GT-2 binding site from the rice phytochrome promoter. The observation that tobacco GT-1a does not interact strongly with \textit{rbcS-3A} box III was confirmed by gel shift studies using \textit{E. coli} extracts containing tobacco GT-1a (Figure 3C). Thus, to summarize, nuclear GT-1 can interact with box II, box III (Figure 3C; Green et al., 1987), and the rice GT-2 binding site (Kay et al., 1989; Kay, 1991); tobacco GT-1a shows strong binding only to box II.

**Binding Specificity of Nuclear and Tobacco GT-1a for Box II**

Previous studies of 2-bp scanning mutations through \textit{rbcS-3A} box II (GTGTGGTTAATATG) demonstrated a critical core of GGTTAA with some sequence requirements for the following TA nucleotides (Green et al., 1988a). Having demonstrated that tobacco GT-1a is a box II–specific binding protein, we wished to define the critical nucleotides within box II required for binding. Figure 4 shows the results obtained using the wild-type and seven mutant sequences described previously (Green et al., 1988a). As a comparison, the binding specificity of nuclear GT-1 for the eight probes was assayed by gel shift studies (Figure 4A). As shown previously (Green et al., 1988a), these data reveal the 6-bp GGTTAA core region to be critical for binding of nuclear GT-1.

To determine the binding specificity of tobacco GT-1a clone A, a phage containing the cDNA sequence was screened for the ability of the encoded protein to interact with each of the same eight probe preparations used in Figure 4A. As shown in Figure 4B, it is clear that tobacco GT-1a shares the same GGTTAA core binding requirements as nuclear GT-1. The specificity of tobacco GT-1a was confirmed by gel shift studies, again using the same eight probe preparations with \textit{E. coli} extracts prepared from cells expressing tobacco GT-1a. Figure 4C shows that tobacco GT-1a shares the same GGTTAA core binding site as nuclear GT-1 when assayed by gel shift analysis. In combination, these data demonstrate the similarity of the core binding site requirements for tobacco GT-1a and nuclear GT-1. Some differences are seen, however, between the binding specificity of nuclear GT-1 and tobacco GT-1a, and these are discussed below.

**Nuclear GT-1 and Tobacco GT-1a Are Antigenically Related**

Having demonstrated that the isolated cDNA encodes a protein with similar binding specificity to nuclear GT-1, we wanted to determine whether tobacco GT-1a and nuclear GT-1 are antigenically related. Oligopeptides were synthesized based on the predicted amino acid sequence of tobacco GT-1a. These oligopeptides—Pep-1, Lys-65 to Ile-84; Pep-2, Lys-98 to Asp-117;
Characterization of Tobacco GT-1a

Pep-3. Lys-259 to Phe-278 (Figure 1)—were synthesized with two lysine residues at both the amino and carboxy termini to facilitate subsequent conjugation to BSA carrier protein. Antibodies were raised to the conjugated oligopeptides in rabbits. The specificity of the resulting antisera was confirmed by protein gel blotting. The antisera prepared following immunization with Pep-1, Pep-2, and Pep-3 contained antibodies with specificity for these synthetic peptides. These antisera were assayed for their ability to interact with nuclear GT-1 in gel shift studies. Figure 5 shows that the antibodies raised against Pep-1 cross-react with nuclear GT-1 in such an assay. Incubation of 2 μL of antisera raised against Pep-1 in a gel shift reaction with the tetramer of box II as a probe and tobacco nuclear extract resulted in a super shift of the retarded fragment (lane 5) as compared to that seen when either no antisera (lanes 2 and 3) or preimmune sera (lane 4) were added. Antisera raised to Pep-2 and Pep-3 showed weak interactions with the nuclear protein (data not shown). Data presented in Figure 5 demonstrate that the protein encoded by the isolated cDNA sequences is antigenically related to nuclear GT-1.

Genomic Organization of Tobacco GT-1a Genes

The demonstration that tobacco GT-1a does not interact strongly with other GT motifs and the identification of the related rice GT-2 protein (Dehesh et al., 1990) prompted us to examine the

Figure 4. Sequence Specificity for Binding of Nuclear GT-1 and Tobacco GT-1a to Box II.

(A) Gel shift study with nuclear GT-1 using tetramer oligonucleotides of box II and its mutant derivatives. WT, GTGTTGGAATATG; GT, CAGTGGGAATATG; GG, GTGTCCCTAATATG; TT, GTGTTGGAATATG; AA, GTGTTGGAATATG; TA, GTGTTGGAATATG.
The free probe and bound complex are indicated.

(B) Filter binding assay using tobacco GT-1a clone A in comparison with a nonspecific DNA binding protein with the same eight box II derivative probes as above.

(C) Gel shift assay using tobacco GT-1a prepared from E. coli infected with phage carrying GT-1a clone A. The probes are as described above. The free probe and bound complex are indicated as are nonspecific complexes due presumably to E. coli proteins.

Figure 5. Antibodies Raised to an Oligopeptide from Tobacco GT-1a Interact with Nuclear GT-1.

Radiolabeled box II tetramer was used as a gel shift probe with tobacco leaf nuclear extract (0.25 μg/μL). Lane 1, probe alone; lanes 2, 4, and 5, 0.5 μg nuclear extract; lane 3, 1 μg nuclear extract. Either 2 μL of undiluted preimmune sera (lane 4) or 2 μL of undiluted serum from Pep-1 (lane 5) was added to the incubations. Free probe, bound complex, and super shift are indicated.
organization of sequences related to GT-1a within the tobacco genome. Figure 6 shows these results. DNA gel blot analysis of tobacco nuclear DNA revealed the presence of a small number of restriction fragments that show homology to the cDNA sequence. These data suggest the presence of more than one GT-1a–related gene within the tobacco genome. It is possible that these sequences could indicate the presence of a tobacco equivalent of rice GT-2 as well as other GT binding proteins. On the other hand, it is also possible that they could be accounted for in part by the amphidiploid nature of tobacco.

Domain Definition of Tobacco GT-1a

From comparisons between tobacco GT-1a and rice GT-2 (Figure 2B), it is apparent that these two DNA binding proteins share homology over the HHTH region. The similarity between the core binding sites of tobacco GT-1a and GT-2, GGTTAA and GGTAAT, respectively, suggests the possibility that the conserved regions of tobacco GT-1a and GT-2 play a role in interacting with the related target DNA elements. This possibility is consistent with the DNA binding function of helix-containing structures of other DNA binding proteins (Desplan et al., 1988; Murre et al., 1989).

To localize the DNA binding domain of tobacco GT-1a, we compared the ability of the truncated polypeptides encoded by clones A and B to bind to box II. The longest possible translation product predicted from the cDNA sequences is shown in both Figures 1 and 7A. The protein sequence shown starts with the first methionine of the open reading frame, which is presumably the translation start site in the plant. However, both cDNA sequences are in the same reading frame within the polylinker of λ ZAP, yet out of frame with the β-galactosidase coding region. The recombinant protein from E. coli is therefore likely not produced as a fusion protein but probably results from internal initiation of translation from Met-1, Met-18, or Met-25. The most likely initiation point in E. coli is Met-18 because of the presence of a Shine-Delgarno (AGGAG) sequence directly 5’ to this ATG.

The initial GT-1a cDNA, clone A, was isolated by the ability of the encoded protein to bind box II. The encoded polypeptide must therefore necessarily contain the DNA binding domain of tobacco GT-1a. The second cDNA sequence, clone B, was isolated by DNA sequence homology to clone A. Comparison of the two cDNA sequences, assuming translation from Met-1, reveals that the frameshift mutation in clone A results in a predicted polypeptide of 136 amino acids. The frameshift mutation in clone B results in a predicted polypeptide of only 108 amino acids, 27 amino acids shorter than that encoded by clone A (Figure 7A). By comparison of the two sequences, it is apparent that the shorter protein encoded by clone B lacks the second and third helices of the HHTH region (Figure 7B). The ability of the polypeptide encoded by clone B to bind to box II was assayed by both DNA-protein screening of a phage containing this cDNA (Figure 7C) and gel shift analysis of E. coli extracts prepared from cells expressing tobacco GT-1a (Figure 7D). It is of interest to note that the truncated polypeptide gives rise to a retarded band of a similar mobility to nuclear GT-1. From these data, it is apparent that polypeptide B that lacks helix 2 and helix 3 is unable to bind box II. This is in striking contrast to polypeptide A that contains all three predicted helices. These observations therefore suggest that the second and third helices of the HHTH motif are involved in the interaction between tobacco GT-1a and box II.

Expression of Genes Encoding Tobacco GT-1a

Previous studies on both tobacco and pea nuclear GT-1 demonstrated that GT-1 activity is present in extracts prepared from both light-grown and dark-adapted plants. We wished, therefore, to analyze the expression pattern of the gene from which the tobacco GT-1a cDNA was derived. From our screens of the two cDNA libraries, one derived from RNA isolated from etiolated tobacco seedlings and the other derived from RNA isolated from light-grown tobacco leaves, it is clear that tobacco GT-1a mRNA is present in both populations. These observations were confirmed by RNA gel blot analyses of poly(A)+ RNA isolated from various tissue samples. These data are presented in Figure 8. GT-1a transcripts are present in both
Characterization of Tobacco GT-1a

DISCUSSION

Tobacco Nuclear GT-1 and Recombinant GT-1a Are Related

With the aim of dissecting the signal transduction pathway that links photoperception by phytochrome to transcriptional activation of pea rbcS-3A, we isolated cDNA sequences that encode a DNA binding protein with specificity for the rbcS-3A GT-1 box II binding site. The protein encoded by the isolated cDNA sequences is related to nuclear GT-1. This is demonstrated by two lines of evidence. First, antibodies raised to synthetic peptides derived from the predicted amino acid sequence of tobacco GT-1a can interact with the nuclear protein. Second, both nuclear and tobacco GT-1a have similar DNA binding site requirements for the box II sequence element. It is also evident that tobacco GT-1a shares some sequence homology to rice GT-2 (Dehesh et al., 1990); however, the relationship between tobacco GT-1a and the purified DNA binding activity of the nuclear silencer binding factor 1 (SBF-1), which also requires a GGTTAA core consensus (Harrison et al., 1991), remains to be determined.

Binding Specificity of Tobacco GT-1a

It is clear from the binding specificity data that tobacco GT-1a binds specifically to box II (GTGTGGTTAATATG) with sequence requirements similar to those of nuclear GT-1. A core of GGTTAA was previously defined as critical for binding of pea and tobacco nuclear GT-1 to box II in vitro (Green et al., 1988a) and transcriptional activity conferred by this element in transgenic tobacco plants (Kuhlemeier et al., 1988; Sarokin and Chua, 1992). Tobacco GT-1a does, however, show some differences in its binding specificity from nuclear GT-1. Nuclear GT-1 shows some requirement for the TA dinucleotide following the 6-bp core sequence (Figure 4; Green et al., 1988a). However, mutation of this TA dinucleotide does not appear to affect binding of tobacco GT-1a, as demonstrated by both filter binding assays and gel shift studies. Mutation of the first GT dinucleotide of box II results in reduced binding of tobacco GT-1a. This same

Figure 7. DNA Binding Domain Definition of Tobacco GT-1a.

(A) Comparison between the predicted amino acid sequences derived from the two independent GT-1a cDNA clones. Clone A is shown above clone B. The polypeptide encoded by cDNA clone A is 136 amino acid residues; the polypeptide encoded by clone B is 108 amino acid residues. The amino acids present within the polypeptides as a consequence of the frameshift mutations in clone A and B are indicated in bold. The predicted helix 1, 2, and 3 regions are underlined; helix 2 and helix 3 are absent from the polypeptide encoded by clone B. (B) Predicted structure of GT-1a clone A and clone B. The predicted acidic, basic, helix, and turn (T) regions are indicated. (C) Filter binding assay with phage carrying GT-1a clone A and GT-1a clone B with a tetramer of box II (GTGTGGTTAATATG) as probe. (D) Gel shift analysis of GT-1a clone A (A) and GT-1a clone B (B) as compared with nuclear GT-1 (N) using a tetramer of box II as probe. The free probe and bound complex are indicated.

light-grown (lane 1) and dark-adapted (lane 2) leaf tissue. In addition, stem (lane 3), root (lane 4), and etiolated seedling (lane 5) tissues also contain GT-1a mRNA. The expression level of the gene encoding the GT-1a mRNA sequence is low, requiring the use of poly(A)+ RNA and precluding quantitative comparisons from these studies.

Figure 8. RNA Gel Blot Analysis of Tobacco GT-1a mRNA.

Tobacco poly(A)+ RNA (3 μg) from light-grown leaf tissue (lane 1), 3-day dark-adapted leaf tissue (lane 2), stem tissue (lane 3), root tissue (lane 4), and etiolated seedlings (lane 5) was probed with the radiolabeled GT-1a cDNA clone A.
mutation does not dramatically affect binding of nuclear GT-1 (Green et al., 1988a); however, a different two-base substitution of this GT dinucleotide does affect binding of nuclear GT-1 (Green et al., 1988a), demonstrating a requirement for these nucleotides under certain conditions.

The inability of tobacco GT-1a to interact with box II (ACTT-TATCATTTTCACTACTCT) or with the rice phyA 3' GT motif (TTGGCGGTAATTAAC) is in contrast to the broader binding specificity of nuclear GT-1, which interacts with all three sequences (Green et al., 1987; Kay et al., 1989). The tobacco protein GT-1a has a subset of binding activities shown by tobacco nuclear GT-1. Three possible explanations for these differences are considered. First, the recombinant protein is produced as a truncated polypeptide and is composed of only the amino terminal 136 amino acids of the full-length protein, which may affect the specificity as determined by sequences flanking the core binding site. Efforts to assay the binding specificity of the full-length polypeptide have proven unsuccessful because of our inability to recover full-length GT-1a cDNA sequences. Second, the E. coli–produced protein may not be faithfully modified as compared with nuclear GT-1. Third, it is possible that, whereas tobacco GT-1a alone can bind to box II, additional proteins are required to stabilize such interactions with box II–related sequences such as box III and the rice phyA 3' GT motif. The presence of such proteins in nuclear extracts and their absence in E. coli extracts containing only tobacco GT-1a could account for the more stringent sequence specificity of the recombinant protein compared with nuclear GT-1. It is clear, however, that both the recombinant and nuclear proteins share a requirement for the 6-bp core sequence of box II. Earlier studies demonstrated that box III is a much weaker binding site for nuclear GT-1 as compared with box II (Gilmartin et al., 1990a). The difference in affinities of nuclear GT-1 for these two sequences may be heightened when the recombinant protein, as opposed to the nuclear protein, is used in such studies.

A Family of GT Binding Proteins

It is as yet unclear whether tobacco GT-1a can interact with other box II–like elements present upstream of rbcS-3A (Green et al., 1987, 1988a) and other light-responsive genes (Stockhaus et al., 1987; Manzara and Grunissem, 1988; Dean et al., 1989; Elliot et al., 1989; Schindler et al., 1990; Lawton et al., 1991) or whether additional GT binding proteins exist. The characterization of rice GT-2 (Dehesh et al., 1990) and the purification of SBF-1 (Harrison et al., 1991) suggest that GT binding proteins probably comprise a small family of related factors of which tobacco GT-1a is one.

The observation that tobacco GT-1a does not bind to the rice phyA GT-2 binding site is complementary to the observations of Dehesh et al. (1990). These authors showed that recombinant rice GT-2 binds to the 3' rice GT motif (GGCGGTAATTA) with a relative affinity two orders of magnitude higher than it does to the pea rbcS-3A box II element (GTGTGTTAATG). In addition, they note that binding of recombinant GT-2 to the rice phyA 5' GT motif (TAGGTTAATTA) is severely reduced. Our combined findings demonstrate reciprocal binding specificities for tobacco GT-1a and rice GT-2 with reference to the binding sites containing the core motifs GGTAA and GGTAT, respectively.

Structure of Tobacco GT-1a

The predicted structure of tobacco GT-1a, derived from cDNA sequences and a polymerase chain reaction–amplified genomic DNA fragment, indicates the presence of an HHTH region. Sequence comparison between tobacco GT-1a and rice GT-2 shows that this region is highly conserved between the two proteins. Comparisons between the ability of clone A and clone B to interact with box II suggest a role for this region in such an interaction. Similar helical structures are present within HLH and homeodomain DNA binding proteins (cf. Johnson and McKnight, 1989) and have been shown to be involved in DNA binding (Desplan et al., 1988; Murre et al., 1989). However, the absence of any striking amino acid homology between tobacco GT-1a and such proteins suggests that tobacco GT-1a and rice GT-2 may contain a novel DNA binding motif that shares predicted structures similar to those of HLH and homeodomain proteins.

We have used the designation HHTH as opposed to HLH for the following reasons. First, sequence structure predictions for this region of tobacco GT-1a suggest the presence of three helices as opposed to two. The second and third helices are separated by a short sequence that contains a potentially turn-inducing proline residue. Second, the term HHTH discriminates between the predicted structure of tobacco GT-1a and other previously described helix-containing proteins. The predicted structure of tobacco GT-1a is of interest in that few plant DNA binding proteins have been isolated from cDNA expression libraries that do not contain a basic leucine zipper (bZIP) motif (Katagiri and Chua, 1992).

Comparisons of the DNA binding abilities of the proteins encoded by the two independent cDNA clones for tobacco GT-1a show that the second and third helices are required in the DNA binding domain of the recombinant protein. Because this region is highly conserved between tobacco GT-1a and rice GT-2, it is likely that this motif may also play a similar role in binding of rice GT-2 to its target sequence. The similarities between the two proteins over this region may explain their affinities for such similar DNA sequence elements. The lack of homology between other regions of tobacco GT-1a and rice GT-2 likely reflects their proposed reciprocal roles in mediating positive regulation of pea rbcS-3A in the light and rice phyA in the dark, respectively. By analogy with other DNA binding proteins, the acidic domains within tobacco GT-1a may comprise part of the activation domain of the protein. The availability of the cDNA sequences will enable this question to be addressed.
A Model for the Role of GT-1

We have previously proposed three possible models to reconcile the presence of nuclear GT-1 in the light and the dark with the ability of the GT-1 binding sites to activate transcription specifically in the light (Lam and Chua, 1990). Our observations from RNA gel blot data that genes encoding tobacco GT-1a show constitutive expression are consistent with the presence of GT-1 binding activity in light- and dark-grown plants (Green et al., 1988). Our present findings are consistent with these models, which are presented in Figure 9. Two of the models predict the presence of an additional protein, either as a dark repressor (Figure 9, model 1) or as an activator in the light (Figure 9, model 2). The third possibility is that GT-1 undergoes a photoreversible modification that leads to its activation specifically in the light. In addition, the presence of GT-1a transcripts in stem and root tissue raises the question of how the box II tetramer confers activity specifically in chloroplast-containing leaf cells (Lam and Chua, 1990).

The isolation of cDNA sequences that encode a protein which binds specifically to the regulatory element box II and is antigenically related to nuclear GT-1 will enable further studies to address these questions and may provide access to the signal transduction chain that links photoperception to a transcriptional response.

METHODS

Isolation of Sequence-Specific DNA Binding Proteins

Approximately 600,000 plaque-forming units (pfu) of phage (20,000 pfu per 150-cm-diameter Petri dish) of a tobacco seedling cDNA library were screened with the box II tetramer (4GTGTGGTTAATATG) as a probe using the procedure of Singh et al. (1989) and Katagiri et al. (1990) and modified as described previously (Gilmartin et al., 1991). The library was made from mRNA isolated from etiolated tobacco seedlings by oligo(dT) priming. The cDNA was methylated with EcoRI methylase, and EcoRI linkers were added and ligated into λ ZAP (Stratagene).

DNA-Protein Filter Hybridization Assays

These were performed essentially as described by Gilmartin et al. (1991). The recombinant phage was plated at a density of approximately 200 pfu per 75-cm-diameter Petri dish. Following growth of the plaques on nitrocellulose, the filters were cut into sectors and each one treated identically with separate probes. Oligonucleotide probes for the binding sites were excised from plasmids, gel purified, and end labeled with the Klenow fragment of DNA polymerase I with the following exception: GT-2 binding site monomer oligonucleotides (10 ng each) were heated to 100°C in 20 μL of 1 × kinase buffer and allowed to cool to room temperature. ATP, polynucleotide kinase, and T4 DNA ligase (Stratagene) were added, and the reaction was incubated at room temperature overnight. The ligated oligonucleotide mixture was random prime-labeled using T7 DNA polymerase (Stratagene) and passed once over a pushcolumn (Stratagene).

Screening of Tobacco cDNA Library

Following the growth and isopropyl β-thiogalactoside induction of approximately 1 × 10^6 pfu of a λ ZAP tobacco leaf library (Stratagene) as outlined above, plaques were lifted and UV cross-linked in a Stratalinker (Stratagene) onto Amersham HyBond N nylon membranes. Filters were prehybridized for 1 hr at 42°C in 50% formamide, 5 × SSPE (1 × SSPE is 0.15 M NaCl, 10 mM sodium phosphate, 1 mM EDTA, pH 7.4), 5 × SDS. The probe was prepared by random prime labeling the isolated cDNA insert with a T7 DNA polymerase kit (Stratagene). The probe was added to the prehybridization mix, and incubation at 42°C continued for 16 hr. Filters were washed twice for 15 min in 0.1 × SDS, 0.5 × SSC (1 × SSC is 0.15 M NaCl, 0.015 M sodium citrate) at 65°C and exposed to Kodak X-Omat film. Positive plaques were purified by subsequent rounds of screening.

Gel Shift Studies

Tobacco nuclear extracts were prepared and gel shift studies performed as described previously (Green et al., 1988b). Escherichia coli extracts containing tobacco GT-1a were prepared by infecting 100 μL of competent E. coli XL-1 blue cells (Hanahan, 1983) with 10^6 to 10^7 pfu of phage. Following 30 min on ice and a subsequent 5-min incubation at 37°C, 800 μL of SOC medium (Sambrook et al., 1989) and 100 μL of 100 mM isopropyl β-thiogalactoside were added. The mixture was incubated at 37°C for 5 hr, and the cells were pelleted subsequently in a microcentrifuge for 2 min. The pellet was resuspended in 30 μL of nuclear extract buffer (Green et al., 1987) containing 5 mM DTT and 5 μg/mL each antipain and leupeptin; the suspension was lysed by freezing and thawing.
Antibody Production

One milliliter of synthetic peptide (5 mg/mL) was mixed with 1 mL of BSA (10 mg/mL). Two milliliters of 0.2% glutaraldehyde was added stepwise, and the mixture was incubated at room temperature for 1 hr. One milliliter of 1 M glycine in PBS was added to give a final concentration of 200 mM. The mixture was incubated with rocking at room temperature for 1 hr. The peptide-protein conjugate was separated from free peptide by dialysis against PBS (Harlow and Lane, 1988). Antibodies were raised in New Zealand white rabbits against the peptide-protein conjugate (Harlow and Lane, 1988).

General Molecular Biological Techniques

RNA and DNA gel blot analyses as well as other standard molecular biology procedures were performed as described previously (Sambrook et al., 1989). DNA sequencing was carried out using a United States Biochemicals Sequenase kit according to the manufacturer’s instructions. DNA and protein sequence data were processed using the DNASIS and PROSIS packages from Hitachi (Brisbane, CA).

ACKNOWLEDGMENTS

We are most grateful to Dr. Maria Deak for providing the etiolated tobacco seedling cDNA library, Dr. Steve McKnight, Carnegie Institute of Washington, for providing clone B9, Dr. Laura Sarokin for the box III probe, and Wen-Shu Lin, Institute of Molecular and Cellular Biology, Singapore, for the synthetic peptides. We also thank the members of the Laboratory of Plant Molecular Biology for helpful discussions and comments during this work. R.M.G., J.M., and K.H. were supported by fellowships from the Winston Foundation, the North Atlantic Treaty Organization, and the Japanese Society for the Promotion of Science, respectively. This work was supported by National Institutes of Health Grant No. GM44640.

Received March 18, 1992; accepted May 11, 1992.

REFERENCES


Characterization of a gene encoding a DNA binding protein with specificity for a light-responsive element.

P M Gilmartin, J Memelink, K Hiratsuka, S A Kay and N H Chua

Plant Cell 1992;4:839-849

DOI 10.1105/tpc.4.7.839

This information is current as of January 8, 2018