Trihelix DNA-binding Protein with Specificities for Two Distinct cis-Elements

BOTH IMPORTANT FOR LIGHT DOWN-REGULATED AND DARK-INDUCIBLE GENE EXPRESSION IN HIGHER PLANTS

Yukio Nagano‡, Takehito Inaba§, Hirofumi Furuhashi, and Yukiko Sasaki
From the Graduate School of Agricultural Sciences, Nagoya University, Nagoya 464-8601, Japan

Received for publication, March 20, 2001
Published, JBC Papers in Press, April 4, 2001, DOI 10.1074/jbc.M102474200

The DE1 sequence is a cis-regulatory element necessary and sufficient for light down-regulated and dark-inducible expression of the pea GTPase pra2 gene. This sequence does not show any sequence similarity to the previously reported ones involved in light-regulated gene expression. A one-hybrid screen isolated a cDNA encoding a DNA-binding protein, named DF1, with specificity for the DE1 sequence 5′-TACAGT. DF1 has domains similar to the trihelix DNA-binding domain found in the GT-1 and GT-2 proteins, which are plant transcription factors. The DE1-binding domain of DF1 is most similar to the carboxyl-terminal trihelix domain of the rice GT-2 protein with specificity for the GT2 sequence 5′-GGTAATT, which is also necessary for dark-inducible expression of the rice phyA gene. An electrophoretic mobility shift assay showed that this DNA-binding domain specifically binds to two types of DNA sequences, DE1 and GT2. Additionally, using DF1/GT-1 chimeras, we show that the second and third helices of the trihelix DNA-binding domain of DF1 are responsible for this dual DNA binding specificity. Our results show that DF1 has specificity for the two distinct cis-regulatory elements, both important for light down-regulated and dark-inducible gene expression in higher plants.

Light is one of the most important signals affecting plant gene expression (1). Light signals regulate the transcription of a number of genes in positive and negative manners. The mechanisms of light-enhanced transcription have been extensively studied (1). Various cis-regulatory elements and trans-acting factors responsible for light-enhanced gene expression have been characterized (1). In contrast, the mechanism of light down-regulated transcription is less clear. Only a few cis-regulatory elements and trans-acting factors responsible for down-regulation have been reported. For example, the GT2 (5′-GGTAATT) sequence in the light down-regulated rice phyA gene is necessary for dark-activation of this gene (2). Affinity screening using GT2 sequences has been used to isolate the DNA-binding protein, named GT-2 (2). The GT-2 protein is a member of the trihelix DNA-binding proteins (3), which are considered to be specific to plants and have domain(s) containing three α-helices involved in DNA binding. Although the GT-2 protein is one of the well characterized plant DNA-binding proteins, the function of its binding site, the GT2 sequence, is less clear because there has been no report concerning a gain-of-function analysis using the GT2 sequence fused to the minimal promoter. Therefore, combinatorial interaction between the GT2 element and the other element may be required for dark activation. Except for the GT2 protein, we do not know of a DNA-binding protein with specificity for the cis-regulatory element involved in light down-regulated gene expression.

Previously, we reported that a small GTPase gene in pea, pra2, which belongs to the YPT/rab family, is one of the genes whose expression is down-regulated by photoreceptor phytochrome (4). The pra2 gene is mainly expressed in the growing zone of etiolated epicotyls, and its expression is repressed when the plant is illuminated (5). The DE1 sequence (5′-GGATTTTACAGT) in the pra2 gene is another cis-regulatory element necessary for light down-regulated and dark-enhanced expression in plants (6). Gain-of-function analysis has shown that the DE1 sequence is sufficient to confer dark-inducible and light down-regulated expression to a minimal promoter (7). This sequence does not show any sequence similarity to the previously reported ones involved in light-regulated gene expression. In addition, to our knowledge, DE1 is the first discovered cis-regulatory element that is both necessary and sufficient for light regulation. Physiological and genetic analyses have shown that the DE1 sequence receives the signals from phytochrome A, phytochrome B, and blue-light photoreceptors (7). Therefore, the DE1 sequence should be a useful tool for studying the molecular mechanism involved in light down-regulated gene expression in plants. In this study, we identified a cDNA whose product binds to the DE1 sequence. We found that this protein is a member of the trihelix DNA-binding proteins and binds to two distinct cis-regulatory elements, DE1 and GT2, both important for light down-regulated and dark-inducible expression in plants.

EXPERIMENTAL PROCEDURES

Reporter Constructs for the Library Screen—The methods for the creation of reporter constructs were described previously (7). For the first screening, the pHiSI reporter plasmid (CLONTECH, Palo Alto, CA) containing five tandem copies of the 18-bp sequence, which con-
Fig. 1. Schematic diagram of one-hybrid strategy.

Preparation of the cDNA Library—Total RNA was prepared from the growing zones of pea (*Pisum sativum* cv. Alaska) stems (between 0 and 1 cm from the top of the hook) that had been grown in the dark for 7 days at 25 °C as described elsewhere (8). Poly(A)^+^-RNA was extracted by the PolyATtract mRNA isolation system (Promega, Madison, WI). Complementary DNA was synthesized with the HybriZAP-2.1 two-hybrid cDNA synthesis kit (Stratagene, La Jolla, CA), cloned into the EcoRI and XhoI sites of HybriZAP-2.1 or λZapII, and packaged using the MaxPlax packaging extract (Epitect Technologies, Madison, WI). The plasmid library was constructed by excising from the HybriZAP library by infection with the helper phage ExAssist (Stratagene). The resulting HybriZAP cDNA library contained ~8 × 10^7 independent cDNAs.

Screening of the cDNA Library—The histidine yeast reporter strain was transformed with a pea cDNA library by the LiAc/polyethylene glycol method. Approximately 1.5 × 10^7 cDNA plasmids were transformed with a pea cDNA library by the LiAc/polyethylene glycol method. Approximately 1.5 × 10^7 cDNA plasmids were transformed with a pea cDNA library by the LiAc/polyethylene glycol method. Approximately 1.5 × 10^7 cDNA plasmids were transformed with a pea cDNA library by the LiAc/polyethylene glycol method. Approximately 1.5 × 10^7 cDNA plasmids were transformed with a pea cDNA library by the LiAc/polyethylene glycol method. Approximately 1.5 × 10^7 cDNA plasmids were transformed with a pea cDNA library by the LiAc/polyethylene glycol method. Approximately 1.5 × 10^7 cDNA plasmids were transformed with a pea cDNA library by the LiAc/polyethylene glycol method.

The yeast one-hybrid strategy (10) was used to screen for pea cDNA encoding for protein that binds to the DE1 sequence. Tandem copies of the 18-bp sequence containing the 12-bp DE1 sequence with a neighboring 3 bp at each end were subcloned into the upstream region of *HIS3* and *lacZ* reporter genes. About 500 histidine-positive clones were selected from 1.5 × 10^7 transformants (Fig. 1A). We rescreened the plasmids recovered from histidine-positive clones by transforming the *lacZ*-reporter strain and obtained three plasmids showing *lacZ*-positive phenotypes (Fig. 1B). To exclude the cDNAs whose products bind to the joint sequence between the 18 bp sequences, we rescreened the plasmids using the reporter constructs having tandem copies of the 20-bp sequence containing the DE1 sequence with a neighboring 4 bp at both ends. Finally, one plasmid showed a *lacZ*-positive phenotype (Fig. 1C). To isolate longer cDNAs, this cDNA was used to screen a λZAPII cDNA library. We further characterized the longest cDNA of the isolated clones.

The gene, designated as *DF1*, contains an open reading frame of 682 amino acid residues. A sequence similarity search revealed that *DF1* has domains similar to the trihelix DNA-binding domain (3) found in GT-1 (11, 12), GT-2 (2, 13), and GTL1 (14) proteins. GT-1 and GT-2 are DNA-binding proteins that recognize light-responsive cis-regulatory elements and are homologous within trihelix DNA-binding domains, which contain three α helices involved in DNA binding. Fig. 2A shows the schematic representation of *DF1* and other related proteins. Among trihelix DNA-binding proteins, *DF1* is the most similar to proteins having twin trihelix DNA-binding domains, such as GT-2 and GTL-1 proteins. To date, these twin trihelix DNA-binding proteins differ structurally from GT-1 because GT-1 has only one trihelix DNA-binding domain. The *DF1* protein also has twin trihelix domains. In addition to having two conserved DNA-binding domains, GT-2, GTL-1, and *DF1* proteins each have an additional conserved domain, a central domain, which is located between two DNA-binding domains. However, except for the three conserved domains, the amino acid sequences of these twin trihelix DNA-binding proteins are not conserved. Fig. 2B shows the multiple alignments among several trihelix DNA-binding proteins. As recent *in silico* analysis has postulated that the trihelix domain is distantly related to the Myb DNA-binding domain (15), the alignment includes the c-Myb DNA-binding domains. All these proteins have well conserved amino acid residues. Among Arabidopsis proteins, the amino acid sequence of *DF1* is most similar to that of the protein predicted from the BAC clone F7O12 (Fig. 2C). Thus, *DF1* is not orthologous to GT-2 and GTL-1 proteins. However, proteins having twin trihelix DNA-binding domains, such as DF1, GT-2, and GTL-1, may have overlapping functions because they share a high degree of similarity. The orthologous Arabidopsis gene is located adjacent to the GT-2 gene, in the opposite orientation, on chromosome 1.
DF1 Is a DNA-binding Protein with Specificity for the DE1 Element—To analyze the DNA-binding specificity of the DF1 protein, we conducted an EMSA using a 31-bp synthetic DNA probe containing the single DE1 sequence (Fig. 3A, WT). We focused only on the carboxyl-terminal trihelix domain of DF1 because a one-hybrid screen first isolated this DNA-binding domain, and studies of both domains would have been complicated. We expressed the carboxyl-terminal trihelix DNA-binding domain of DF1 in *E. coli* as a His10 tag or a GST-fused protein. The addition of the His10 tag DF1 protein to the binding reaction mixture showed the retarded band(s) of the DNA-protein complexes (Fig. 3B, lane 2). The addition of the GST-fused DF1 protein to the binding reaction mixture also showed the retarded bands of DNA-protein complexes (Fig. 3B, lane 3). Because it is easy to produce sufficient amounts of recombinant GST-fused protein and it is difficult to dissolve the His10 tag DF1 protein in solution, we used the GST-fused DF1 protein in the following experiments. The results were essentially the same as those of the experiments using the His10 tag protein (data not shown).

To test whether the observed binding was specific to the DE1 sequence, we carried out competition experiments. The competitor sequences are shown in Fig. 3A. The addition of a 400-fold wild-type competitor diminished the retarded band (Fig. 3B, lane 4, WT). The MT competitor contained the three nucleotide changes, which could not bind to the nuclear factors with specificity for the DE1 sequence (6). A gain-of-function analysis using the 3-bp mutated constructs fused to the mini-promoter has shown that this construct does not have the ability for light down-regulation (7). The EMSA showed that the MT competitor did not diminish the retarded band (Fig. 3B, lane 5). These data indicate that the DNA binding specificity is consistent with the results of the EMSA using nuclear extracts (6) and the gain-of-function analysis (7).

To determine the core-binding site of the DF1 protein, we carried out competition experiments using the 6-bp mutated competitors (Fig. 3A, LS2-LS4). In our previous work (6), we identified the DE1 element using the 6-bp mutated constructs by transient assay analysis. The LS3 construct (Fig. 3A, lane 1) did not show red-light down-regulation. The MT competitor contained the three nucleotide changes, which could not bind to the nuclear factors with specificity for the DE1 sequence (6). A gain-of-function analysis using the 3-bp mutated constructs fused to the minimal promoter has shown that this construct does not have the ability for light down-regulation (7). The EMSA showed that the MT competitor did not diminish the retarded band (Fig. 3B, lane 5). These data indicate that the 12-bp sequence, designated DE1 (5'-GGATTTTACAGT), mediates phytochrome down-regulation and that the 6-bp sequence (5'-TACAGT) is the core region for down-regulation. The addition of a 400-fold LS3 competitor did not diminish the retarded band (Fig. 3B, lane 7). The addition of an LS2 competitor diminished the band slightly (Fig. 3B, lane 6), whereas
the addition of an LS4 competitor diminished the band greatly (Fig. 3B, lane 8). These results show that recombinant DF1 binds the DE1 sequence and the core-binding site is the sequence 5'-TACAGT. Thus, these results are very consistent with those of the transient assay analysis (6).

To determine the DF1-binding site more precisely, we carried out competition experiments using the pairwise-mutated competitors (Fig. 3A, MT1-MT9). The addition of MT4, MT5, and MT6 competitors, which correspond to the core region of the DE1 element, did not diminish the retarded band (Fig. 3B, lanes 12–14). The addition of the MT2 and MT3 competitors diminished the band slightly (Fig. 3B, lanes 10 and 11), and the addition of other competitors diminished the band greatly (Fig. 3B, lane 9 and lanes 15–17). These results show that the sequence 5'-TACAGT is the core-binding site of the DF1 protein, and its surrounding sequence is slightly involved in the binding of the DF1 protein.

DF1 Is a DNA-binding Protein with Specificity for the GT2 Element—The DE1-binding domain shows sequence similarities to many trihelix DNA-binding domains. Among these, the binding sequences of GT-1 (11, 12) and GT-2 (2, 13) proteins have been characterized. The GT1 element (5'-GTGGTTGTATATG) of pea rbcS-3A is the binding site of tobacco GT-1. (The core-binding sequence is underlined.) The GT2 (5'-TGCGCCGTAATTAAC) and the GT3 (5'-TGCGGTAATACCGG) sequences of rice phyA are the binding sites of carboxyl-terminal and amino-terminal trihelix DNA-binding domains, respectively, of the rice GT-2 protein (13). Thus, the DF1 binding sequence (5'-TACAGT) is clearly different from the binding sequences of GT-1 and GT-2 proteins. The DE1-binding domain of the DF1 protein is most similar to the GT2 element-binding domain of the rice GT-2 protein. To determine whether DF1 has the ability to bind to the binding sites of these trihelix DNA-binding proteins, we carried out competition experiments using competitors containing the above binding sequences (Fig. 3A, GT1-GT3). The addition of GT1 and GT3 competitors did not diminish the retarded band (Fig. 3B, lanes 18 and 20). However, the addition of a DF1 competitor diminished the band (Fig. 3B, lane 19). Thus, these results suggest that DF1 has the ability to bind to the GT2 element.

The complementary sequence of the GT2 probe (5'-TGCGCCGTAATTAAAC) is 5'-GTAAATTAACCCGCA. In this complementary sequence, the sequence 5'-TACCCCG is partly similar to the core DE1 sequence 5'-TACAGT. Therefore, the DF1 protein might bind to this sequence and not to the GT2 core sequence. To determine the DNA binding specificity to the GT2 sequence, we also conducted the EMSA using a 31-bp synthetic DNA probe containing the GT2 sequence (Fig. 4A, GT2) and the pairwise-mutated competitors (Fig. 4A). The addition of the GST-fused DF1 protein to the binding reaction mixture showed the retarded bands of the DNA-protein complex (Fig. 4B, lane 2). The addition of MT12, MT13, MT 14, and MT15 competitors, which correspond to the core region of the GT2 element, did not diminish the retarded bands (Fig. 4B, lanes 6–9). The addition of the MT11 competitors diminished the bands slightly (Fig. 4B, lane 5), and the addition of the GT2, MT10, MT16, and WT (containing the DE1 sequence) competitors diminished the bands greatly (Fig. 4B, lanes 3–4 and lanes 10–11). This analysis showed that the core-binding site of DF1 is 5'-GTAAATTA. This recognition specificity of the DF1 protein is very similar to that of the GT-2 protein (2). Thus, these results again show that DF1 has the ability to bind specifically to the GT2 sequence, which is clearly different from the core DE1 sequence.

**Fig. 4. Determination of the DNA sequence bound to the DF1 protein using the GT2 probe.** A, nucleotide sequences of various competitors used to define the DF1-binding site. The positions of the mutated nucleotide are shaded. The core-binding site of the DF1 protein is underlined. B, EMSA using various competitors. The competitors WT and MT are shown in Fig. 3.

Second and Third Helices of the DNA-binding Domain Are Responsible for the Dual DNA Binding Specificity—We are interested in knowing which helices of the trihelix DNA-binding domain of DF1 are responsible for the dual DNA binding specificity. In addition, it has not been determined which helices of the trihelix DNA-binding domain of the other related protein can recognize DNA sequences, although deletion and mutational analyses have shown that the trihelix DNA-binding domain of Arabidopsis GT-1 is essential for DNA binding (16). To determine the mode of action for recognizing two distinct DNA sequences, we conducted the EMSA using DF1/GT-1 chimeras (Fig. 5). Fig. 5A shows the recombinant proteins used in this study. Both DF1 and GT-1 helix 1-DF1-helices 2–3 proteins were able to bind to two types of sequences, the GT2 and DE1 sequences (Fig. 5B, lanes 2, 3, 7, and 8), although the DF1 protein bound to these sequences more strongly. Each of two types of proteins could have similar affinities to the GT2 and DE1 sequences. In contrast, the DF1-helices 1-2-GT1-helix 3 protein bound to neither of the two types of sequences (Fig. 5B, lanes 4 and 9). The GT1 protein bound to both sequences only weakly (Fig. 5B, lanes 5 and 10). Thus, the second and third helices of the trihelix domain are responsible for the binding to the DE1 and GT2 sequences.

The creation of the chimera protein may affect the DNA binding specificity. To determine the DNA binding specificity of the DF1-helices 1-2-GT1-helix 3 protein to both sequences, we conducted the EMSA using the pairwise-mutated competitors. The
addition of MT12, MT13, MT14, and MT15 competitors to the binding reaction mixture containing the labeled GT2 probe did not diminish the retarded bands (Fig. 5C, lanes 5–8). On the other hand, the addition of GT2, MT11, and MT16 competitors diminished the bands slightly (Fig. 5C, lanes 3, 4, and 9). Thus, although the specificity to the sequence was reduced markedly, the core GT2 sequence, 5′-GGTTAATT, could be the core-binding site of the DF1helix 1–2–GT-1helix 3 protein. Similarly, the addition of MT4, MT5, and MT6 competitors to the binding reaction mixture containing the labeled DE1 probe did not diminish the retarded bands (Fig. 5C, lanes 14–16). On the other hand, the addition of WT, MT3, and MT7 competitors diminished the bands slightly (Fig. 5C, lanes 12, 13, and 17). These results show that the DE1 sequence 5′-TAGCAGT is the core-binding site of the DF1helix 1–2–GT-1helix 3 protein. Thus, this chimera protein can recognize two distinct DNA sequences, DE1 and GT2, indicating that second and third helices of the DNA-binding domain are responsible for the dual DNA binding specificity.

### DISCUSSION

The carboxyl-terminal DNA-binding domain of DF1 can recognize two distinct cis-regulatory elements, DE1 and GT2, both important for light down-regulated and dark-inducible gene expression in higher plants. This is an especially exquisite mechanism as DF1 protein transduces information from the light signal to the two effectors, DE1 and GT2. Plants must have acquired this mechanism during the course of evolution. One possible explanation for this evolved mechanism is that ancestral DF1 protein could bind to only one recognition sequence and the subsequent discontinuous change(s) in the DNA sequence of the DNA-binding domain resulted in the dual DNA binding specificity. Another possibility is that continual changes in both the DNA-binding domain and its recognition sequences eventually resulted in the dual DNA binding specificity.

In the current study, we have described a DNA-binding protein having a domain that recognizes two distinct DNA sequences, although both are cis-regulatory elements having similar functions. Interestingly, two recognition sequences show no or little sequence similarities. To date, cases in which one DNA-binding protein has the specificity for distinct, although related, sequences have been reported. For example, the DNA-binding protein MYB.Ph3, a member of the MYB proteins in *Petunia*, binds to two types of sequences: 5′-A/aD/a/D/aG/C/GTTA (where a/D is A, G, or T, A as the preferred base) and 5′-AGTTAGTTA (17). In contrast, murine c-MYB only binds to the former sequence. A single residue substitution in MYB.Ph3 caused a switch from the dual DNA binding specificities to the c-MYB specificity (17). In addition, c-MYB with the reciprocal substitution gained MYB.Ph3 specificity (17). Interestingly, the trihelix DNA-binding domain may be distantly related to the Myb DNA-binding domain (15). Single or multiple residue substitutions in the trihelix DNA-binding domain from the ancestral protein might cause dual DNA binding specificity for two distinct and unrelated cis-regulatory elements.

The dual DNA binding specificity might be the common feature of the trihelix DNA-binding protein. *Arabidopsis* GT-1 binds not only to the GT1 sequence 5′-GGTTA but also to the repeated GATA sequences (16). However, the precise characteristics of the protein have not been determined because the precise specificity to the GATA sequence has not been determined and these two recognition sequences may be related.

Affinity screening using GT-1 binding sites has isolated the GT1 protein (11, 12). These studies have postulated that three putative a helices (trihelix domain) in GT-1 might be involved in DNA binding. Subsequent deletion analyses of GT-1 have shown that the trihelix DNA-binding domain of GT-1 is certainly involved in DNA binding (16). However, it has not been determined which helices are responsible for DNA binding.

Using DF1/GT-1 chimeras, the current study showed that second and third helices of the trihelix DNA-binding domain of DF1 are responsible for specific DNA binding. This mode of action is similar to those of the other DNA-binding proteins containing three helices, such as the Myb DNA-binding domain and the homeodomain. A recent *in silico* study has postulated that the trihelix DNA-binding domain is distantly related to the Myb DNA-binding domain (15). Thus, this postulation is consistent with our results using DF1/GT-1 chimeras.

In the current study, we added a characterized member of...
trihelix DNA-binding proteins. The carboxyl-terminal DNA-binding domain of DF1 can bind to two distinct cis-regulatory elements, DE1 and GT2, both responsible for light down-regulated and dark-inducible gene expression in higher plants. The rice GT-2 protein binds to the GT2 sequence, which is necessary for dark-inducible expression of the rice phyA gene. As the GT-2 protein shares a high degree of similarity to DF1, the GT-2 protein might also bind to the DE1 sequence. Another trihelix DNA-binding protein, GT-1, can bind to the GT1 sequence, which is necessary but not sufficient for light-induced expression of the pea rbcS-3A gene (11, 12). Thus, the trihelix DNA-binding proteins could be generally important for light-regulated gene expression. In addition, the current study emphasizes the importance of twin trihelix DNA-binding proteins for light down-regulated and dark-inducible gene expression in plants. Arabidopsis, the genome of which was almost completely sequenced (18), contains more than 10 trihelix DNA-binding proteins (several of which are listed in Refs. 4 and 16). Further characterization of the trihelix DNA-binding proteins, especially of twin trihelix DNA-binding proteins including DF1 protein, should be an important task for the research of light-regulated gene expression in plants.

Acknowledgment—We thank Dr. K. Hiratsuka for useful suggestions.
Trihelix DNA-binding Protein with Specificities for Two Distinct cis-Elements: BOTH IMPORTANT FOR LIGHT DOWN-REGULATED AND DARK-INDUCIBLE GENE EXPRESSION IN HIGHER PLANTS

Yukio Nagano, Takehito Inaba, Hirofumi Furuhashi and Yukiko Sasaki

J. Biol. Chem. 2001, 276:22238-22243.
doi: 10.1074/jbc.M102474200 originally published online April 11, 2001

Access the most updated version of this article at doi: 10.1074/jbc.M102474200

Alerts:
• When this article is cited
• When a correction for this article is posted

Click here to choose from all of JBC's e-mail alerts

This article cites 19 references, 11 of which can be accessed free at
http://www.jbc.org/content/276/25/22238.full.html#ref-list-1