DNA-Binding Study Identifies C-Box and Hybrid C/G-Box or C/A-Box Motifs as High-Affinity Binding Sites for STF1 and LONG HYPOCOTYL5 Proteins

Young Hun Song, Cheol Min Yoo, An Pio Hong, Seong Hee Kim, Hee Jeong Jeong, Su Young Shin, Hye Jin Kim, Dae-Jin Yun, Chae Oh Lim, Jeong Dong Bahk, Sang Yeol Lee, Ron T. Nagao, Joe L. Key, and Jong Chan Hong*

Division of Applied Life Science (BK21 Program), Plant Molecular Biology and Biotechnology Research Center, Environmental Biotechnology Research Center, National Core Research Center, Gyeongsang National University, Jinju 660–701, Korea (Y.H.S., C.M.Y., A.P.H., S.H.K., H.J.J., S.Y.S., H.J.K., D.-J.Y., C.O.L., J.D.B., S.Y.L., J.C.H.); Division of Plant Sciences, University of Missouri, Columbia, Missouri 65211 (J.C.H.); and Department of Plant Biology, University of Georgia, Athens, Georgia 30602 (R.T.N., J.L.K.)

LONG HYPOCOTYL5 (HY5) is a bZIP (basic leucine zipper) transcription factor that activates photomorphogenesis and root development in Arabidopsis (Arabidopsis thaliana). Previously, STF1 (soybean [Glycine max] TGACG-motif binding factor 1), a homologous legume protein with a RING-finger motif and a bZIP domain, was reported in soybean. To investigate the role of STF1, the phenotypes of transgenic Arabidopsis plants overexpressing STF1 and HY5 were compared. In addition, the DNA-binding properties of STF1 and HY5 were extensively studied using random binding site selection and electrophoretic mobility shift assay. Overexpression of STF1 in the hy5 mutant of Arabidopsis restored wild-type photomorphogenic and root development phenotypes of short hypocotyl, accumulation of chlorophyll, and root gravitropism with partial restoration of anthocyanin accumulation. This supports that STF1 is a homolog of HY5 with a role in light and hormone signaling.

The bZIP (basic Leu zipper) proteins are a class of transcription factors involved in many plant growth and development processes, including photomorphogenic development and hormone signaling (Jakoby et al., 2002; Cluis et al., 2004). One of the best characterized bZIP factors thought to play a role in photomorphogenic seedling development and hormone signaling in Arabidopsis (Arabidopsis thaliana) is LONG HYPOCOTYL5 (HY5) (Oyama et al., 1997; Ang et al., 1998; Holm et al., 2002; Cluis et al., 2004). The function of HY5 in photomorphogenesis is well illustrated in hy5 mutant seedlings, which have defects in light inhibition of hypocotyl elongation, in light-induced chlorophyll, and in anthocyanin accumulation (Oyama et al., 1997; Sibout et al., 2006; Shin et al., 2007). The model for light-signaling pathways in photomorphogenic development includes the photoreceptor phytochromes, the ubiquitin ligase CONSTITUTIVE PHOTOMORPHOGENIC1 (COP1), and the positive signaling component HY5 (Deng et al., 1992; Ang and Deng, 1994; Holm et al., 2002). One regulatory circuit involving COP1 is nuclear degradation of HY5. Although degradation of the HY5 protein is mediated by nuclear COP1 in the dark, light exposure results in the extrusion of COP1 from the nucleus into the cytosol, a process that allows accumulated HY5 to interact with DNA and to activate a light-regulated gene (von Arnim and Deng, 1994; Osterlund et al., 2000).
The role of HY5 in hormone signaling has been implicated in the hy5 null allele that shows altered root morphology (Oyama et al., 1997). The hy5 mutation affects several aspects of root morphogenesis, resulting in an elevated number of lateral roots, less responsiveness to gravitropic stimulus and touching, and longer root hairs in hy5 seedlings than in wild type. The hy5 mutant traits are partly the result of an altered balance in the signaling of auxin (Sibout et al., 2006). Microarray analyses have shown that many auxin-responsive and auxin-signaling genes are misexpressed in hy5 mutants, an indication that the genes encoding auxin-signaling components are one group of the HY5 downstream genes (Cluis et al., 2004; Sibout et al., 2006). HY5 is also involved in cytokinin signaling (Vandenbussche et al., 2007). Cytokinin treatment results in similar growth responses to blue light, such as the development of leaves and chloroplasts, stimulation of anthocyanin production, and the inhibition of hypocotyl growth elongation (Chory et al., 1994). It has been proposed that cytokinins increase the level of HY5 by reducing the degradation mediated by COP1 (Vandenbussche et al., 2007).

Previously, STF1, a homologous bZIP protein that acts as a potential regulatory factor for hypocotyl elongation, was reported for soybean (Glycine max; Cheong et al., 1998). STF1 has the unusual feature of having two unrelated structural domains with high-sequence homology to the N-terminal RING-finger domain found in RADIODERMIES WOLOGYEN (RSW1): the cellulose synthase catalytic subunit and the C-terminal HY5-like bZIP domain. The bZIP proteins show a similar structural feature found in other legume bZIP proteins, including broad bean (Vicia faba) VF6BPZIF and Lotus japonicus LjBZF (Cheong et al., 1998; Nishimura et al., 2002). The role of LjBZF, a gene product of ASTRAY, was predicted in astray (Lyj&mbox;7), a root mutant that develops an increased number of nodules compared with the wild type. The astray mutant also shows photomorphogenic mutant phenotypes similar to those observed in hy5 mutants (Nishimura et al., 2002). However, the role of STF1 has not been studied in detail.

This article presents the findings of: (1) an analysis of the DNA-protein interactions of STF1 using both random binding site selection (RBSS) and gel mobility shift assay, and (2) a comparison of the DNA-binding properties and biological functions of STF1 with HY5 using a transgenic plant analysis. An in vitro binding analysis of STF1 and HY5 demonstrate that these bZIP proteins preferentially recognize C-1, hybrid C/G-1, and C/A-box motifs over G-box motifs. The in vitro analysis corresponds with earlier in vivo and functional analyses that identified the predicted locations of the HY5 binding sites in the promoters of anthocyanin biosynthetic genes (Hartmann et al., 2005; Shin et al., 2007). It also helps explain the abundance (approximately 3,900) of in vivo HY5 targets in the Arabidopsis genome identified by coupled chromatin immunoprecipitation and DNA chip hybridization (ChIP-chip; Lee et al., 2007). By identifying the HY5/STF1 recognition elements and the parameters associated with target genes, this study extends our understanding of the roles of HY5 and related bZIP proteins in the regulation of gene expression during plant development.

RESULTS

STF1 Can Replace HY5 in Photomorphogenesis and Hormone Signaling

STF1 is a bZIP factor of soybean that is homologous (71.8%) to the C-terminal half of the HY5 protein in Arabidopsis. It contains conserved amino acid motifs to the casein kinase II phosphorylation site and the COP1 interaction right before the bZIP domain (Fig. 1A; Hardtke et al., 2000). To test whether STF1 and HY5 play similar roles in photomorphogenesis and hormonal signaling, a complementation test was performed using the hy5 mutant. The coding region of STF1 was constructed under the control of the 35S cauliflower mosaic virus promoter and was stably introduced into Arabidopsis. STF1OX, the STF1 overexpression line, was compared with wild type and with HY5OX, the HY5 overexpression line. Figure 1 shows that expression of both STF1 and HY5 in the hy5 mutants restored normal levels of hypocotyl growth inhibition as well as chlorophyll accumulation in the light-grown seedlings. The accumulation of anthocyanin was partially restored in the STF1OX line. In L. japonicus, a mutation in LjBZF, the STF1 related gene, resulted in the reduction of anthocyanin accumulation (Nishimura et al., 2002). These results suggest that STF1 plays the same role as HY5 in photomorphogenesis (Fig. 1B).

We then compared the other aspect of the hy5 mutant phenotype (i.e. gravitropic response, waving growth of root, lateral root formation, and root hair elongation) that reflects normal auxin signaling (Oyama et al., 1997; Cluis et al., 2004). The hy5 mutant grown on agar plates showed widely spread lateral roots that were directed nearly horizontal rather than downward (Fig. 2B). The main roots of the hy5 mutants also showed reduced gravitropism with a slight slant to the left. In addition, hy5 mutants exhibited defects in the touch response; they fail to display the normal wavy pattern of root growth when grown in the agar plates set at an angle of 45° (Fig. 2F). The gravitropic and touching responses of the roots were restored in the STF1OX line (Fig. 2, D and H). The enhanced lateral-root formation and the longer root hairs observed in hy5 mutants were also complemented by overexpression of HY5 and STF1 (Fig. 2, K and L). Altogether, the transgenic plant analysis provides further support that STF1 and HY5 have the same role in photomorphogenesis and hormone signaling.

Binding Properties of STF1 and HY5 to ACGT-Containing Elements

A previous analysis of the recombinant STF1 protein revealed the C-box (nGACGTCn) to be a high-affinity
binding site (Cheong et al., 1998). The HY5 protein interacts with both the G- (CACGTG) and Z- (ATACGTGT) boxes of the light-regulated promoter of \( RbcS1A \) (ribulose bisphosphate carboxylase small subunit) and the \( CHS \) (chalcone synthase) genes (Ang et al., 1998; Chattopadhyay et al., 1998; Yadav et al., 2002). To test whether STF1 and HY5 have similar DNA-binding properties, the binding properties of each were compared with eight different DNA sequences that represent G-, C-, and C/G-box motifs (Fig. 3A). C-box sequences carrying the mammalian cAMP responsive element (CRE; TGACGTCA) motif and the Hex sequence (TGACGTGCC), a hybrid C/G-box (Cheong et al., 1998), were high-affinity binding sites for both proteins (Fig. 3B). No binding or limited binding was observed to as-1 (Lam et al., 1989), nos-1 (Lam et al., 1990), or the AP-1 site (TGACTCA; Kim et al., 1993). Binding to the palindromic G-box (PA G-box, GCCACGTGGC) was moderate. However, binding activity to the G-box of the light-responsive unit 1 (U1) region of the parsley (\( Petroselinum crispum \)) CHS promoter (CHS-U1: TCCACGTGGC; Schulze-Lefert et al., 1989) or the G-box of \( GmAux28 \) (TGACTGTC) was much weaker than to the PA G-box (Fig. 3, B and C). Gradual increases in protein concentrations resulted in detectable binding to very weak binding sites such as CHS-U1 and as-1 sequences (Fig. 3B).

The N terminus of STF1 contains structurally unrelated domains with high sequence homology to the N-terminal RING-finger domain found in RSW1 (Fig. 1;
Cheong et al., 1998). When STF1 and HY5 are compared, the full-length STF1 possesses weaker binding activity than HY5 to both the PA G- and the CHS-U1 G-boxes (Fig. 3C). However, the binding affinities of both bZIP proteins were similar to CRE sequence with flanking adenine and thymine (A/T) at positions −4 and +4. Deletion of the STF1-specific N-terminal domain resulted in enhanced binding to the G-box to a level comparable with HY5 (Fig. 3D). These results indicate that the bZIP domains of both STF1 and HY5 have similar binding properties for recognizing ACGT-containing elements (ACEs).

The in vitro binding experiments presented in this study show that, although the G-box is a known target site for the HY5 protein, the C-box sequences are the preferred binding sites for both STF1 and HY5.

STF1 Exhibits a Distinct DNA-Binding Property and Requires a Larger Recognition Sequence Than Does SGBF1 or STGA1

Three soybean bZIP proteins from different families have been described: SGBF1 (soybean G-box binding factor 1; Hong et al., 1995), STGA1 (soybean TGA1; Cheong et al., 1998), and STF1. To differentiate STF1 binding to ACEs, the electrophoretic mobility shift assay (EMSA) patterns for STGA1, STF1, and SGBF1 were compared using the same sets of binding site probes (Fig. 4A). SGBF1 interacts equally well with the two G-box sequences (PA G-box and CHS-U1). Hex, CRE, and AP-1 sequences are also well recognized. STGA1 bound strongly to the sequences containing TGACG and recognized, with high affinity, most of the sequences selected by STF1. Although some preference for the flanking sequence has been reported for these bZIP proteins, the same kind of flanking base preference was not observed for STF1.

STF1, STGA1, and SGBF1 exhibit distinct DNA-binding properties; however, each binds Hex oligonucleotide (Cheong et al., 1994; Hong et al., 1995; Cheong et al., 1998; Fig. 4A). Thus, the protein-DNA contacts mediated by STF1, STGA1, and SGBF1 were compared in more detail using the Hex sequence. Methylation interference experiments were performed to determine whether methylation of G residues modified binding to these proteins. The data in Figure 4B show that STF1 binding to the Hex sequence requires distinct and additional contacts (12–13 bp) than those required by STGA1 and SGBF1. STF1 binding to the Hex oligonucleotide was inhibited when the G residues at positions −5, −4, −2, +0, +2, and +3 (“upper” strand in the image) and −0, +4, and +5 (“lower” strand in the image) were methylated. Binding patterns of STGA1 and SGBF1 to Hex are similar to TGA1 and GBF1 of Arabidopsis (Schindler et al., 1992b).

Binding Site Selection from Random Oligonucleotides Defines the C-Box and the Hybrid ACEs C/G-Box and C/A-Box as High-Affinity Binding Sites for STF1

In its usual form, EMSA cannot be used to identify the wide spectrum of binding sites recognized by a DNA-binding protein. Thus to determine the DNA-binding site requirement and consensus sequence of STF1, RBSS was used (Oliphant et al., 1989). Random oligonucleotides were synthesized and allowed to bind with bacterially produced, purified recombinant STF1 at two different salt concentrations (50 and 150 mM KCl), which represent moderate- and high-stringency conditions, respectively. A total of 150 plasmids that contain DNA-binding sites selected from both conditions were isolated and sequenced. The sequences are shown in Figure 5A, where they are arranged according to the reference nucleotide at position +2. Ninety-five percent of the sequences contain the intact TGACGT motif. The consensus binding site for STF1 is thus 5′-RRTGACGTVDNNTGACGT-3′ [(G/A)G/A]TGACGT [(G/A)(G/A)/(A/T)/(G/A)]TGACGT (C/G/A)(A/T)/(G/A)-3′ (Fig. 5B).

When analyzed by type of ACE, these sequences can be grouped into four subclasses (Fig. 5C): C-box, where the C residue comes at the +2 position; a hybrid C/G-box (C/G-box), with G at the +2 position; C/A-box, with A at the +2 position; and C/T-box, with T at the +2 position. The C-box subclass contains the largest

Figure 2. Root morphologies of the wild type, the hy5 mutant, and the STF1 and HY5 overexpression lines. Wild-type Arabidopsis [Wassilewskija ecotype; A, E, and I], hy5-Ks50 (hy5; B, F, and J), 35S:HYS/hy5-Ks50 (HY5OX; C, G, and K), and 35S:STF1/hy5-Ks50 (STF1OX; D, H, and L) are shown. Plants were grown, vertically positioned, on MS agar plates supplemented with 2% Suc for 21 d (A–D). Arrowheads indicate the positions of root tips at the time of the position change. Root hairs of each plant are shown (I–L). The hy5 mutant displays elongated root hairs (J).
number of selected binding sites for STF1 (38% at 50 mM KCl and 48% at 150 mM), followed by the C/G- (25.3%) and the C/A-boxes (26%). Only a small number of C/T-boxes (4/100) and non-TGACGT sequences (4/100) were selected. Further arrangement of each subclass identified the significance of the base at position 1 (top strand): STF1 shows a strong preference for A at position 1 when it interacts with the C-box (TGACGTC); however, T is preferred when STF1 interacts with the C/G-box (TGACGTG). No C residue was observed at position 1. In addition, most selected sequences contain a base preference for purines (G and A) at positions −5 and −4 (top strand). At higher salt concentrations, T at position +5 and H (A, T, C) at +5 are preferred (Fig. 5, B and C).

**EMSA Confirms RBSS Analysis and Identifies the Significance of Combination of Half-Site**

To confirm that the selected sequences represent the binding affinity of STF1 as well as HY5, a few selected sequences representing each group of ACEs were analyzed using EMSA (Fig. 6). High-affinity binding was observed for all C-box sequences containing the (−3 TGACGTC +3) motif with flanking purines at positions −4 and −5 (CREA/T, no. 16-17, no. 4-27, no. 4-38, no. 4-47, no. 12-78), followed by the C/G-box (Hex, no. 4-21) and the C/A-box (no. 4-46), which is consistent with the RBSS data. Since purine bases are strongly favored at positions −5 and −4, the effect of base substitution at these positions was tested using CREA/T as a reference sequence. STF1 binding to the C-box was profoundly reduced when GA at positions −5 and −4 was converted to either GT (no. 16-15) or TG (no. 1-2). Furthermore, a mutation at position −4 reduced the binding of STF1 to the C-box more than the mutation at position −5 (see no. 16-5 versus no. 1-2). When flanking bases satisfied the preference for purines at positions −5 and −4, only a slight preference for pyrimidine at +4 and +5 was observed. There was no difference in the preference for G or A at positions −5 and −4 (no. 16-17, no. 4-27, no. 4-38, and
The lack of binding, or very weak binding, with number 4-33 (AGTGACGTATT), number 12-86 (GGTGACGCGGAG), and number 4-34 (ACTGACGCGGAG) further confirms that the C/T-box and TGACG-motifs without the ACGT-core are not the preferred binding sites for STF1. The significant reduction in binding to number 4-36 (GATGACGTTTA) is also consistent with the RBSS data that show T is rarely found at position +3 of the C-box.

The HY5 also showed a binding preference very similar to that of STF1 (Fig. 6A). This characteristically strong preference for flanking sequence to the TGACGT motif was not observed for STGA1 or SGBF1 (Fig. 6B). The DNA-binding parameters of STF1 and HY5 are summarized in Figure 7. The ACEs comprise two half-sites, either of which can be symmetric or dependent (Izawa et al., 1993; Niu and Guiltinan, 1994; Fig. 7A).

The degree to which STF1 binds to the ACEs depends on the combination of half-sites. The symmetric half-site can be bound by STF1 when made symmetric (i.e. C-box), and the dependent half-site cannot be bound or weakly bound when made symmetric but can be bound when combined with a symmetric half-site. The RBSS and EMSA analyses show that the dependent half-sites to STF1 are defined as G- and A-boxes. The hybrid ACEs that contain both symmetric and dependent half-sites are good binding sites for both STF1 and HY5, whereas the ACEs comprising two dependent half-sites (i.e. G-box, A-box, and G/Á-box [Z-box]) are weak binding sites (Fig. 7B).

Analysis of Target Genes of STF1 and HY5

Detailed analysis of the DNA-binding properties of STF1 identified a set of binding sites recognized by
Figure 5. DNA-binding sequences selected by STF1. A. Compilation of STF1 binding sites identified by RBSS in 50 and 150 mM KCl. Binding sites were selected from a pool of oligonucleotides carrying random 13 bp flanked by a defined sequence of 26 bp on either side. DNA sequences that bound to the GST-STF1 fusion protein were selected and analyzed using DNA sequencing. Binding sites were aligned according to the consensus sequence. Nucleotides corresponding to the flanking sequences on either side of the random 13 bp are underlined. An asterisk before a clone number indicates a selected sequence containing more or
both STF1 and HY5. The pleiotropic response observed in the hy5 mutant suggests that there are many genes regulated by HY5. This is in agreement with the finding of a large number (approximately 3,900) of in vivo HY5 targets in the Arabidopsis genome (Lee et al., 2007). From the EMSA, the high-affinity binding site for homodimeric STF1 and HY5 was defined as $5^\prime$-(G/A)(G/A)TGACGT(C/G/A)(A/T/G)$-3^\prime$.

As G-box sequences and hybrid ACEs are targets for the HY5 protein, the basal binding motif can be described as HBACGTVD, which includes C-box, C/G-box, C/A-box, and G- and G/A-boxes. The Z-box (ATACGTGT) is the G/A-box, a hybrid ACE recognized by HY5 (Eng et al., 1998). A pattern matching analysis of the upstream region of the whole Arabidopsis genome found that 48.4% (15,516 out of 32,041 sequences) of all genes have the basic consensus sequence HBACGTVD motif in the 1-kb 5' upstream regions of their translation start. The same motif was found in 72% (2,800 out of 3,894) of the genes selected by in vivo HY5 target sequences (Lee et al., 2007). It was 1.48 times higher in the target genes than in the whole genome.

The RRTGACGTVD motif defined as the high-affinity binding motif for HY5/STF1 is found in 516 (13.25%) in vivo HY5 targets, whereas 1,707 genes (5.3%) in the whole genome contain this motif. This represents 2.87 times enrichment in target genes than in the whole genome. This indicates that the high-affinity consensus motif is more likely to be observed less than 13 random nucleotides. In the image, "rev" refers to the reverse orientation of the individual sequence. Selected binding sites are grouped according to the bases at position -2/+2 and divided further by the base at position +3. B, Consensus sequences of STF1 binding sites derived from the base frequencies of the selected sequences at 50 and 150 mM KCl. Numbers indicate the base frequencies from the sequence data. R denotes either base G or A. V denotes bases G, A, and C. D denotes bases G, A, and T. C. The selected binding sites are arranged by C-box, C/G-box, C/A-box, and C/T-box sequences. The number in parentheses indicates the occurrence of the group.

Figure 5. (Continued.)

Figure 6. EMSA analysis of selected binding sites. A, EMSA analysis of purified STF1 and HY5 proteins using various ACEs. Purified STF1 (0.1 μM) and HY5 protein was incubated with 20,000 cpm of P32-labeled binding-site probes. The 24 binding sites used as probes for the EMSA analysis are listed on the right. These sequences are arranged to allow a comparison of the binding properties of each type of ACE. The bases, which are different from the high-affinity consensus binding site, are boxed. B, EMSA analysis of three soybean bZIP proteins: STF1, STGA1, and SGBF1. Three soybean bZIP proteins were interacted with selected C-box sequences to compare the binding properties of each bZIP protein against different flanking sequences to the ACGT-motif.
in the proposed HY5-target genes than in the whole genome (Supplemental Tables S1 and S2).

HY5 regulates a wide range of genes involved in photosynthesis and hormone signaling (Cluis et al., 2004; Sibout et al., 2006; Lee et al., 2007). First, we compared 3,103 genes differentially regulated by light (Ma et al., 2005). Of these genes, 1,681 (54%) contain the HBACGTVD motif, and 228 (7.3%) contain the RRTGACGTVD motif (Supplemental Table S3). We then compared auxin-responsive genes whose mRNA levels are affected by the hy5 mutation (Cluis et al., 2004; Sibout et al., 2006). Of the 246 auxin-responsive genes, 142 (57.2%) contain HBACGTVD, and eight (3.2%) contain the high-affinity binding site (Supplemental Table S4). These values indicate that high-affinity binding sites are underrepresented in both light-regulated and auxin-responsive genes. The identification of the HBACGTVD motif and the in vivo HY5 binding sites in many genes involved in auxin and cytokinin signaling provides further support for the role of HY5 in hormone signaling (Table I). Genes containing high-affinity binding sites in the promoter region encode proteins with diverse functions (Supplemental Table S2). Many regulatory genes such as transcription factors, protein kinases, and ubiquitin ligases contain the binding motif and are actually bound by the HY5 protein (Lee et al., 2007).

**The Promoters of Anthocyanin Biosynthetic Genes Are Targets of HY5 and Contain ACEs That Fit the Criteria of STF1 and HY5 Binding**

The anthocyanin biosynthetic genes were shown to be light regulated and targets of HY5 (Hartmann et al., 2005; Shin et al., 2007). We examined the effect of HY5 expression on the accumulation of mRNA among seven genes encoding phenylpropanoid biosynthetic enzymes: CHS, CHI, F3'H, F3'H, DFR, FLS, and LDOX. Expression of the seven genes showed drastic reduction and induction by the hy5 mutation and HY5 overexpression line (Fig. 8, B and C). The STF1 overexpression line showed partial restoration of the target gene expression (data not shown). Limited induction of LDOX and F3'H was observed by STF1 overexpression, which is consistent with the partial complementation of anthocyanin accumulation in the transgenic lines. The lack of STF1 mutant soybean plants makes it difficult to address the role of STF1 in anthocyanin biosynthesis. However, the reduction of anthocyanin accumulation in the L. japonium astray mutant supports a role for STF1 in anthocyanin biosynthesis (Nishimura et al., 2002).

Promoters of all seven genes contain many ACEs. The F3H promoter was extensively analyzed to compare in vitro and in vivo binding to HY5 (Lee et al., 2007; Shin et al., 2007). Among the five ACEs that were bound to HY5 in vitro, two ACEs—G-box at −246 and C/G-box at −429—fit the criteria of the HY5 binding motif. These two ACEs are in vivo HY5 target sites (Fig. 8D). The predicted locations of the HY5 binding sites in the seven genes are shown in Figure 8E. These sequences were confirmed as HY5 targets using in vivo binding and functional analyses (Hartmann et al., 2005; Lee et al., 2007; Shin et al., 2007). Although these ACEs are classified as low-affinity binding sequences, the half-site of ACE complies well with the binding requirements of HY5 and STF1. All together, these findings strongly support that HY5 is the bZIP protein involved in expression of anthocyanin biosynthetic genes. This study also helps to predict precisely the binding sites in the HY5 target promoters.

**DISCUSSION**

This study shows that STF1 is a bZIP protein that has roles in photomorphogenic development and hormone signaling, much like HY5 of Arabidopsis. Analyses of the DNA-binding properties and the binding site selectivity of STF1 provide ample information about the spectra of binding sites recognized by this protein as well as the effects of flanking sequences around the ACGT-core motif, which contributes to the binding specificity and affinity of the two related proteins, STF1 and HY5.
Table 1. The putative HY5 binding sites predicted by RBSS from genes involved in hormone signaling

| Gene Name * | Gene Identification | HBACGTVD | In Vivo HY5 Target | Sequence and Location | Classification |
|-------------|---------------------|----------|-------------------|----------------------|----------------|
| **Auxin-related genes** | | | | | |
| **Auxin carriers** | | | | | |
| AUX1 | At1g77690 | Yes | 1 | 134 AGAGCTGCTG 145 | G-box |
| EIR1/PIN2 | At5g37090 | Yes | 1 | 66 AACTGCTGCG 77 | G-box |
| PID | At2g34650 | Yes | 2 | 790 TCAAGCTGCTA 801 | G-box |
| PIN1 | At1g73590 | Yes | Yes | 806 TCTATGCTATC 817 | C/A-box |
| REH1/PIN3 | At1g70940 | Yes | Yes | 813 TTACAGTTAAC 824 | G/A-box |
| PIN4 | At1g01420 | Yes | Yes | 949 GCTGCTATTT 960 | G/A-box |
| **Auxin response factors** | | | | | |
| ARF4 | At5g60450 | Yes | 2 | 605 GCTAGCTCAAT 616 | C/A-box |
| ARF6 | At1g30330 | Yes | Yes | 716 TTTGAGCTATTT 727 | C/A-box |
| ARF18 | At1g23750 | Yes | Yes | 94 AAACACG 105 | G/A-box |
| ARF11 | At2g46530 | Yes | Yes | 934 CAAGCTGCTA 945 | G-box |
| ARF19 | At1g19220 | Yes | Yes | 826 TCAAGCTGAAA 837 | C/G-box |
| MP/ARF5 | At1g19850 | Yes | Yes | 605 GCTAGCTGOC 616 | G/A-box |
| **IAA genes** | | | | | |
| IAA3/SHY2 | At1g02420 | Yes | Yes | 54 TACATGCTAA 55 | G/A-box |
| IAA5 | At1g15580 | Yes | 4 | 201 AAGCTAATAC 212 | C/A-box |
| IAA6/SHY1 | At1g52830 | Yes | 1 | 436 ATACATGCT 445 | G/A-box |
| IAA7/AXR2 | At3g23050 | Yes | 3 | 530 TGGCTGCTGGC 541 | G/A-box |
| IAA8 | At2g22670 | Yes | Yes | 586 ATATGCTGATAT 597 | A-box |
| IAA9 | At5g65670 | Yes | 1 | 389 GCTAGCTGATG 400 | A-box |
| IAA10 | At1g04100 | Yes | Yes | 504 TGGCTGCTGAT 515 | G/A-box |
| IAA14/SLR | At4g14550 | Yes | 2 | 567 TACATGCTGCTC 576 | G-box |
| IAA15 | At1g80390 | Yes | Yes | 586 ATATGCTGATAT 597 | A-box |
| IAA17/AXR3 | At1g02420 | Yes | Yes | 434 GCTAGCTGCTC 445 | G/C-box |
| IAA19/MSG2 | At1g15580 | Yes | Yes | 527 TACATGCTAAC 538 | G/A-box |
| IAA20 | At1g20490 | Yes | Yes | 914 GCTAGCTGCTA 923 | G/A-box |
| IAA27/PAP2 | At4g29080 | Yes | 1 | 343 ATACATGCTGTC 354 | G/A-box |
| IAA28 | At5g25890 | Yes | 2 | 784 GCAGCTGCTTG 795 | G-box |
| IAA30 | At3g62100 | Yes | 4 | 863 ATACATGCTGTC 874 | G/A-box |
| **Other auxin signaling genes** | | | | | |
| AXR6/CUL1 | At1g02570 | Yes | Yes | 90 AGCTAGCTGAT 101 | C/A-box |
| NAC1 | At1g36010 | Yes | Yes | 149 GCTAGCTGAC 160 | G/A-box |
| SAUR-AC1-like | At1g06690 | Yes | Yes | 230 TACATGCTGAT 249 | G/A-box |
| SAUR-AC1-like | At3g61900 | Yes | Yes | 794 TACATGCTGAT 803 | G-box |

(Table continues on following page.)
Table I. (Continued from previous page.)

| Gene Name | Gene Identification | HBACGTVD | In Vivo HY5 Target | Sequence and Location | Classification |
|------------|---------------------|----------|--------------------|-----------------------|----------------|
| TIR1       | At1g12820           | Yes      | 2                  | 424 ATCATAATTT 435    | A-box          |
|            |                     |          |                    | 614 TTATACATGC 625    | A-box          |

Cytokinin-related genes

Cytokinin biosynthesis

| Gene Name | Gene Identification | HBACGTVD | In Vivo HY5 Target | Sequence and Location | Classification |
|-----------|---------------------|----------|--------------------|-----------------------|----------------|
| ACS1      | At3g61510           | Yes      | 1                  | 115 TTTTACAATGATT 126 | G-box          |
| ACS2      | At1g01480           | Yes      | 3                  | 30 AAATACATAGAG 41    | G/A-box        |
| ACS6      | At4g11280           | Yes      | 1                  | 346 TTTCATCTAAA 357   | A-box          |
| ACS9      | At3g49700           | Yes      | 1                  | 802 AGTATACAGACG 813  | G/A-box        |
| ANPT1     | At1g86460           | Yes      | 2                  | 461 TCTTACATGGCA 472  | G-box          |
| ANPT2     | At2g27760           | Yes      | 1                  | 818 TTTCATCTAT 829    | C/G-box        |
| ANPT7     | At3g23630           | Yes      | 2                  | 44 TAATACATGCTC 55    | A-box          |
|           |                     |          |                    | 230 ACACATACGTT 241   | G-box          |

Cytokinin oxidases

| Gene Name | Gene Identification | HBACGTVD | In Vivo HY5 Target | Sequence and Location | Classification |
|-----------|---------------------|----------|--------------------|-----------------------|----------------|
| CKX2      | At2g19500           | Yes      | 1                  | 403 ACATACATGATAT 414 | A-box          |
| CKX5/6    | At1g75450           | Yes      | Yes                | 8                    | 216 CACATACATGAC 227 |
|           |                     |          |                    | 303 ACACATACAGG 314   | G-box          |
|           |                     |          |                    | 336 TTACATACAGTC 325  | G-box          |
|           |                     |          |                    | 337 ATACACATACAG 348  | G-box          |
|           |                     |          |                    | 703 CTACATACATAC 714  | C/A-box        |
|           |                     |          |                    | 779 AGACACATGT 790    | G/A-box        |
|           |                     |          |                    | 859 TACACATGAT 870    | G/A-box        |
|           |                     |          |                    | 923 TACACATGAT 934    | C/G-box        |

Cytokinin + ethylene receptor

| Gene Name | Gene Identification | HBACGTVD | In Vivo HY5 Target | Sequence and Location | Classification |
|-----------|---------------------|----------|--------------------|-----------------------|----------------|
| ERS1      | At2g40940           | Yes      | 1                  | 560 TACACATGAAAT 571  | G/A-box        |
| ERS2      | At1g04310           | Yes      | 1                  | 780 TTCACATGAGT 791   | G/A-box        |
| ETR1      | At1g66340           | Yes      | 2                  | 845 CAATACATGATT 856  | C/A-box        |
|           |                     |          |                    | 963 ATACATACGGAA 974  | C-box          |

Phosphorelay proteins

| Gene Name | Gene Identification | HBACGTVD | In Vivo HY5 Target | Sequence and Location | Classification |
|-----------|---------------------|----------|--------------------|-----------------------|----------------|
| AHK1      | At3g17820           | Yes      | 1                  | 217 GATACATACAGA 228  | C/A-box        |
| AHP1      | At3g21510           | Yes      | 1                  | 95 TTTACCATAAA 106    | A-box          |
| AHP3      | At5g39340           | Yes      | 1                  | 177 GAAAATACATGAG 188  | C/G-box        |

Response regulators

| Gene Name | Gene Identification | HBACGTVD | In Vivo HY5 Target | Sequence and Location | Classification |
|-----------|---------------------|----------|--------------------|-----------------------|----------------|
| ARR5      | At3g48100           | Yes      | 1                  | 118 TTCTACATGCG 129   | C-box          |
| ARR6      | At5g62920           | Yes      | 1                  | 555 TTCTACATGAT 566   | A-box          |
| ARR7      | At1g19050           | Yes      | 2                  | 338 ATACATACAAAG 349  | C/A-box        |
| ARR8      | At2g41310           | Yes      | Yes                | 1                    | 294 AAACATACCA 305 |
| ARR9      | At3g57040           | Yes      | Yes                | 4                    | 104 TTCTACATGA 115 |
|           |                     |          |                    | 127 AGACACATATCG 138  | G/A-box        |
|           |                     |          |                    | 316 ATTACATATT 305    | A-box          |
|           |                     |          |                    | 387 AGTACATACAT 398   | C-box          |
|           |                     |          |                    | 878 AGATACATGAC 889   | C-box          |

Other (putative) cytokinin signaling genes

| Gene Name | Gene Identification | HBACGTVD | In Vivo HY5 Target | Sequence and Location | Classification |
|-----------|---------------------|----------|--------------------|-----------------------|----------------|
| SOB2/DRNL | At1g24590           | Yes      | 2                  | 44 ATCCATACGT 55      | G-box          |
|           |                     |          |                    | 896 AAATACATGATC 907  | A-box          |
|           |                     |          |                    | 426 GAACATACATAGA 437 | C/A-box        |
| AteRF1    | At1g28160           | Yes      | 1                  | 655 ATACATGCTAC 666   | G-box          |
| AteRF2    | At1g28370           | Yes      | 2                  | 893 ATACATACATGAA 904 | G/A-box        |
|           |                     |          |                    | 154 ATACATGCTAC 165   | C/A-box        |
|           |                     |          |                    | 718 AGACACATGAAT 729  | G/A-box        |
|           |                     |          |                    | 881 TTACATGATAG 870   | A-box          |

**PEP**

| Gene Name | Gene Identification | HBACGTVD | In Vivo HY5 Target | Sequence and Location | Classification |
|-----------|---------------------|----------|--------------------|-----------------------|----------------|
|           | At5g10480           | Yes      | 1                  | 141 ATACATACGTAAG 152 | C/G-box        |

*Gene name.  1Arabidopsis Genome Initiative identification number.  2The presence of the HY5 binding motif (HBACGTVD) in the −1,000 bp upstream of first amino acid (Met).  3The presence of in vivo HY5 target site according to Lee et al. (2007).  4Number of HY5 binding ACEs in the promoter region (−1,000 bp upstream).  5Locations and sequences of the HY5 binding sites in the reported genes. Number denotes base from the first ATG.  6ACE types are classified according to the half-site nomenclature. The half-site is named according to the base at +2.

RBSS identified a consensus site for high-affinity binding: 5′-RRTGACGTVDn-3′. An interesting feature of the STF1 protein-DNA interaction to emerge from the RBSS and the EMSA analyses is that STF1 has a strict requirement for binding sites that have large sequences and certain combinations of flanking se-
quences at positions −5, −4, +2, and +3. This feature has also been observed in two Antirrhinum bZIP proteins that preferentially bind to a 12-bp hybrid C/G-box motif [−5 G(A/G)TGACGTGG(C/A) +5; Martinez-Garcia et al., 1998]. The highest binding affinity was observed for PA C-boxes of the 12-bp sequence 5′-(G/A)TGACGTGG(C/A)-3′, which contains a purine at positions −5 and −4 and a pyrimidine at positions +4 and +5 (Figs. 3 and 4). Although STF1 has a narrower binding site spectra than soybean’s TGA1,
which is a C-box binding protein (Cheong et al., 1994), the binding of STF1 to the G-box, coupled with its ability to form heterodimers with GBFs, makes it a novel member of the bZIP protein family because of its capability to recognize both C-box and G-box motifs. Many researchers have also reported the importance of the flanking sequences for DNA-binding of bZIP proteins (Schindler et al., 1992b; Williams et al., 1992; Izawa et al., 1993, 1994; Niu and Guiltinan, 1994; Hong et al., 1995; Martinez-Garcia et al., 1998). However, this study stresses the importance of a single base at position −4 for STF1 binding, which is in contrast to other bZIP proteins that interact with these C-box sequences in an overlapping manner (Fig. 4B).

Detailed analyses of the selected binding sites and the EMSA data identified an optimal combination of flanking sequences at positions +2 and +3 after TGACGT. For C-boxes, the preferred bases are CA (TGACGTCA) and CG (TGACGTGC). The CC (TGACGTC) or CT (TGACGTCT) combinations are rarely selected, a finding confirmed by the EMSA using the TGACGTCT (no. 4-36; compare no. 4-36: GATGACGTCT versus no. 12-78 GATGACGTCAT and no. 11-14 GATGACGTCT) sequences (Fig. 4). Other preferred combinations of flanking sequences are GG (Hex, TGACGTGG; C/G-box), GT, and no. 36 GG (Hb, TGACGTGG; C/g-box), and AT (no. 4-36, TGACGTAT; C/A-box). The flanking sequences GG, GT, and CA at positions +3 and +4 were also observed in the high-affinity binding site of group 1 factors (GBFs, EmBP1, HBP1a) for which the optimum binding site is CCACGTGG (G-box; Schindler et al., 1992a, 1992b; Niu and Guiltinan, 1994). For hybrid ACEs, the correct base combination should be considered in predicting the target site.

Since C-box motifs and hybrid C/G- or C/A-boxes are the preferred binding sites for HY5 as well as for STF1, both HY5 and STF1 may regulate a wide range of genes in addition to G-box containing genes. The observation that the most dramatic morphological defects in the hy5 mutant are found in the hypocotyls, stems, and roots supports this conjecture (Ang and Deng, 1994; Oyama et al., 1997; Ang et al., 1998). Recently, LjBzf, a gene highly homologous to STF1, was isolated from L. japonicus (Nishimura et al., 2002). Mutation of LjBzf results in the mutant astray (Ljspm77), a root mutant with a higher number of nodules than that of the wild type. The astray mutant shows similar greening, hypocotyl, and root morphology as the hy5 mutant with reduction in anthocyanin accumulation (Nishimura et al., 2002). The ASTRAY protein has a structure highly similar to that of other legume bZIP proteins: STF1 of soybean; VFBIPZF of broad bean, which contains an N-terminal RING-finger domain found in RSW1; the cellulose synthase catalytic subunit; and the C-terminal HY5-like bZIP domain. Given these similarities, detailed analyses of STF1 DNA-binding properties may enhance our understanding of the roles of related legume bZIP proteins in plant development.

The HY5 target was reported to over 3,000 genes (Lee et al., 2007). The sequence analysis revealed that these genes carry several different types of ACEs. The ACEs are classified into C-box, G-box, and hybrid C/G, C/A, and G (Z-box) boxes. The analysis of HY5 target genes suggests that these bZIP proteins could regulate expression of many regulatory genes, such as those encoding transcription factors, kinases, or genes required for cell proliferation and elongation (Supplemental Tables S1 and S2). This is strongly indicative of HY5 having a role in a high hierarchical position (Lee et al., 2007). The identification of many signal transduction related genes and transcription factors involved in auxin and cytokinin signaling complies well with a role for HY5 in hormone signaling (Table 1). The HY5 binding sites in light-regulated genes fit the role of HY5 in photomorphogenesis.

Overall, the observation that STF1 and HY5 have similar binding properties and physiological roles and the identification of their binding criteria further facilitate our understanding of how these two bZIP proteins function in complex plant developmental processes, such as cell elongation, root development, and photosynthesis.

**Materials and Methods**

**Plant Materials and Growth Conditions**

Standard molecular biology techniques were used according to Sambrook et al. (1989). Arabidopsis (Arabidopsis thaliana) plants were cultivated in a growth chamber with a 16-h light/8-h dark cycle at 22°C under a combination of cool-white fluorescent and incandescent lights at 70 to 100 μmol m⁻² s⁻¹. For generation of the STFI overexpression (STFIox) transgenic line, the full-length coding region of STFI (Cheong et al., 1998) was amplified and cloned into the BanHI site of pCAMBIA1300 (CAMBIA, Canberra, Australia). For transformation of the hy5 mutant, Arabidopsis hy5-k50 (provided by Dr. Kiyotaka Okada, Kyoto University) was used. The 35S::HY5 overexpression line (Hy5ox) was obtained from Dr. K. Okada. Fifteen transgenic lines were obtained, and homozygous lines, which express higher levels of the STF1 gene, were established for phenotypic analysis. All plants used in these experiments were in the Wassilewskija background and were grown on half-strength Murashige and Skoog (MS) agar medium containing 2% Suc, except for measurement of hypocotyl length.

**RNA Extraction and Gene Expression Analysis**

Total RNA was extracted from 5-d-old seedlings grown on MS agar medium under continuous light (LL) using the Trizol method (Invitrogen) according to the manufacturer’s instructions. Gene expression was analyzed by semiquantitative reverse transcription (RT)-PCR (Kim et al., 2005). For synthesis of the complementary DNA (cDNA), 4 μg of total RNA were primed using oligo(dT)12-18 (Invitrogen) as the primer. The cDNA was diluted to 200 μL with water, and 5 μL of the diluted cDNA were used for PCR amplification. The primer sequences and cycles used to amplify each gene are as follows: HY5, 25 cycles, 5'-GAGGAGAACCTGCAGGAAA-3' and 5'-CTTCTTTGCTTCCACCTCGCTCA-3'; STFI, 25 cycles, 5'-CCGGAGATGGAGGGTAAGGAC-3' and 5'-CTTCTTCCTCCTCCCTGCTT3'; CHI, 23 cycles, 5'-CTTCTGCTTCTTCTCTGGT-3' and 5'-CATTGGTTGGTTGCTTCCAG-3'; CHI, 25 cycles, 5'-CTCTCCCTACATTTACACTTCCT-3' and 5'-CTTGGTGCTCCTTCCCACTGTTGG-3'; F3H, 25 cycles, 5'-GAATACAGGGAGGTGAATGAAGA-3' and 5'-GCAATGGCTAAGAACTCAATGG-3'; ACTIN2 (At2), 20 cycles, 5'-AAAACAACCTATCAGTGCTTCGG-3' and 5'-CTTGGGTTCCCTTGAC-3'; 5'-CCATGGCTAAGAATCCACTAAGG-3'; 5'-ATGCTTGCACTAAGACTCAACT-3'; 5'-GACGGATCCTATCCGCTCCAT-3'.

---

*Song et al.*

1874

Plant Physiol. Vol. 146, 2008
Analysis of Hypocotyl Length, Anthocyanin Levels, and Chlorophyll Content

To measure hypocotyl length, sterilized seeds were placed on half-strength MS agar medium without Suc and stratified in the dark at 4°C for 3 d. The seeds were exposed to white light (100 μmol m⁻² s⁻¹) for 1 h, returned to the darkness at 22°C for 23 h, and then placed into either continuous white light or the dark for 6 d. The hypocotyl lengths of 50 seedlings were measured using SCION Image software (Scion).

To measure anthocyanin accumulation, 5-d-old light-grown seedlings were used. Fifty seedlings per sample were incubated overnight in 300 μL of extraction buffer (methanol containing 1% (v/v) HCl) in the dark. After extraction, 200 μL of distilled water and 200 μL of chloroform were added to each sample, and absorbances were read at 530 and 657 nm. The quantity of anthocyanin was determined by spectrophotometric measurement of the aqueous phase (A₅₃₀/A₆₅₇) and normalized to the total fresh weight of tissues used in each sample.

Relative chlorophyll levels were determined from the same samples used for measuring the quantity of anthocyanin. Chlorophyll was extracted into 1 mL of 80% acetone by shaking the chloroform fractions overnight in the dark. Chlorophyll levels were measured spectrophotometrically, and the amount was calculated using MacKinney’s coefficients and the equation (chlorophyll, μg cm⁻² = 7.15 × OD₅₄₆ nm + 18.71 × OD₆₄₇ nm) described by Holm et al. (2002).

EMSA

For the gel, Hex, and CRE probes, the same DNA fragments were used as previously described (Hong et al., 1995). The oligonucleotides 5'-ACTCGCATCTTGGCGTCGCTCTCCGCCCACTAACCTTGCTTCCCTCA-3' and 5'-GAGTCTGAGGAGTTCTGGAGGAGCGGCCACCCGGATCCGTCGTTGAAATAGC-3' yielded CHS1-U1, which contains the light-responsive U1 element from the parsley (Petroselinum crispum) chalcone synthase gene promoter (Schulze-Leifer et al., 1989); 5'-GATCCATCCGCTGAGGAGGACGACAAGCCAAAACTTCACTACCCCTCTCA-3' and 5'-AGTGGTGTGCTGTCGTGCTCCCTACCTAGCTG-3' yielded nos-1 (Lam et al., 1989); and 5'-GATCCCTAAATAGCTGCACTAGGCAA-3' and 5'-AGGCTTCTGCTAGCCG-3' yielded a 150-base-pair (bp) HindIII-HindIII DNA fragment from the barley (Hordeum vulgare) TRS (Trichome-HI and II) tandem repeats.

Overexpression and Purification of the Proteins

Expression constructs of the pGEX vector (Pharmacia) were used to express the glutathione S-transferase (GST)-fusion proteins (STF1 and HY5) in E. coli BL21(DE3)/pLyS. Purification of the GST-fusion proteins was carried out according to Smith and Johnson (1988). E. coli transformants selected with ampicillin were grown to log phase at 37°C and induced with 1 mM isopropyl-β-D-thiogalactopyranoside (Promega). Three and one-half hours after induction, bacteria were centrifuged, and the pellets were washed with cold STE (0.1% NaCl; 10 mM Tris-HCl, pH 8.0; 1 mM EDTA; pH 8.0) and resuspended in a cold phosphate-buffered saline (PBS; 16 mM NaH₂PO₄; 4 mM NaCl, pH 7.3; 150 mM NaCl) that contained 0.25 mM phenylmethanesulfonyl fluoride, 0.25 μg/mL leupeptin, 2.5 μg/mL aprotinin, 2.5 μg/mL antipain, 0.25 μg/mL pepstatin A, and 1 mM DTT. After sonication, 1% Triton X-100 was added to the lysates, which were then clarified by centrifugation at 12,000g for 10 min at 4°C, and the supernatant was mixed with glutathione-linked agarose beads (Sigma). After a 30-min incubation with gentle shaking at 4°C, the beads were washed three times in a cold PBS. For RBSS, the fusion protein was eluted with 10 mM glutathione (Sigma) in PBS. For EMSA, the glutathione agarose beads containing the GST-fusion proteins were equilibrated with a thrombin cleavage buffer (2.5 mM CaCl₂, 150 mM Tris-HCl, pH 8.0; 150 mM NaCl, 0.1% β-mercaptoethanol), and the proteins were eluted with a thrombin cleavage buffer containing thrombin (0.024 units/mL, Boehringer Mannheim). After purification, all the proteins were dialyzed in an extraction buffer (20 mM HEPES, pH 7.9; 50 mM β-mercaptoethanol; 0.2 mM phenylmethanesulfonyl fluoride; 1 mM DTT; 1 mM EDTA; 100 mM NaCl; 10% glycerol) at 4°C, aliquoted, and frozen at −70°C. Protein concentration was determined using a combination of the Bradford reagent assay (Bio-Rad) and staining with Coomassie blue after SDS-PAGE.

BB5

The oligonucleotide synthesized for use in the binding selection was TN64 (5'-CGGAGACTTGCAGGAAATACAGTTAGTTAAGGCAAACTCACA-3'), which was derived from Schindler et al. (1992b). TN64 contains a randomized sequence of 13 nucleotides flanked by 26 nucleotides at the 5'-end containing a HindIII site and 25 nucleotides at the 3'-end containing a BamHI site. For PCR amplification, the oligonucleotides TN20 (5'-CGGAGACTTGCAGGAAATACAGTTAGTTAAGGCAAACTCACA-3') and TN20 (5'-GCTGTCGTGAGGCAGCCGCTC-3') served as forward and reverse primers, respectively. To produce double-stranded random binding DNA sequences, TN64 (400 pmol) and TN20 (800 pmol) were annealed, and the second strand was synthesized after adding the four deoxyribonucleotide triphosphates (final concentration of 0.25 mM each). Extension was performed using the Klenow fragment of the DNA polymerase. The GST-STF1 fusion protein was preincubated with 6 μg polyclonal antibodies, as a non-specific competitor, for 10 min at room temperature in a reaction mixture (20 mM HEPES, pH 7.9; 50 mM β-mercaptoethanol; 0.2 mM phenylmethanesulfonyl fluoride; 1 mM DTT; 15% glycerol) prior to addition of the double-stranded synthetic random oligonucleotide, producing a final volume of 30 μL. After an additional 15-min incubation, the DNA-protein complex was separated on a 7% polyacrylamide (29% acrylamide/bis) gel containing 0.5× Tris-borate/EDTA and 3% glycerol for 1.5 to 2 h at 15 to 20 V/cm. The DNA-protein complex, which comigrated with the GST-STF1 and the α-1 labeled Hex (5'-GCTGTCGTGAGGCAGCCGCTC-3') probe complex, was positioned, excised from the gel, and eluted by electrodialysis into a dialysis bag. DNA was extracted with phenol/chloroform and precipitated by ethanol. Recovered DNA was resuspended in approximately 20 volumes of deionized water. PCR was carried out in a final volume of 100 μL containing 80 pmol of each primer, 20 nmol of the four deoxyribonucleotide triphosphates, 10 mM Tris-HCl (pH 8.3 at 25°C), 1.5 mM MgCl₂, 50 mM KCl, and 0.5% (v/v) Taq polymerase. After initial denaturation for 30 s at 92°C, the selected binding sequences were amplified for 30 s at 92°C, 1 min at 55°C, and 30 s at 72°C for 30 cycles in a thermal cycler (Pharmacia LKB). The amplified DNA of 64 nucleotides was purified on a 2% agarose gel, eluted as described above, and served as the template for the following round of selection. Five selection cycles (binding/gel-shift/elution/PCR) were carried out using a binding buffer containing 50 mM KCl. The final pool of oligonucleotides was digested with BamHI and HindIII, ligated into pHuescript II (Stratagene) SK+, and transformed into XLI-Blue. A total of

DNA-Binding Properties of STF1 and LONG HYPOCOTYL5

Methylation Interference Analysis

Methylation interference analysis was performed as described previously with a minor modification (Schindler et al., 1992b). The plasmid containing the Hex sequence (5'-gggTACGTTGcaca-3') was 5'-end labeled on either strand with Klenow and [α-32P]dATP, at purified using PAGE. EMSA was then performed as previously described (Hong et al., 1995). Proteins were preincubated in a reaction buffer (20 mM HEPES, pH 7.9; 0.2 μg/μL poly(dI-dC); 0.5 mM dithiothreitol; 0.1 mM EDTA; 50 mM KCl) for 10 min at room temperature and incubated with 2 × 10⁷ cpm (0.5 ng) of end-labeled probe DNA for 15 min. The resulting protein-DNA complexes were analyzed by electrophoresis on non-denatured 5% PAGE using a 0.5× Tris-borate/EDTA electrophoresis buffer. Following electrophoresis, the gels were dried and subjected to autoradiography with intensifying screens at −70°C.
100 clones containing the selected binding site were randomly chosen. The sequences of the inserted DNA were determined by using the 7-deaza-dGTP sequencing kit (USB) with the T3 primer and [α-32P]dATP (Amersham).

**Sequence Analysis**

The potential targets for STF1/HY5 in the Arabidopsis genome database were searched using the Pattern Matching program on the Arabidopsis Information Resource Web site (http://www.arabidopsis.org). Duplicate entries between the two data sets were removed using Duplicate Remover, a biostatistical tool available on The Bio-Array Resource for Arabidopsis Functional Genomics Web site (http://bbc.botany.utoronto.ca).

Sequence data from this article can be found in the GenBank/EMBL data libraries under accession numbers BAA21327 (HY5), L28003 (STF1), L28005 (STGA1), and L01447 (SGBF1).

**Supplemental Data**

The following materials are available in the online version of this article.

**Supplemental Table S1.** Genes that carry high-affinity binding sites based on RBSs analysis and that were identified as HY5 targets by ChIP-chip (Lee et al., 2007).

**Supplemental Table S2.** The 1,705 genes that contain the high-affinity HY5 binding site (RRTGACGTVD); the number of HY5-binding motifs and the position of 12 nt sequence (start and end) are shown with sequences.

**Supplemental Table S3.** List of light-regulated genes that contain the HBACGTVD (1,681 genes) or RRTGACGTVD (228 genes) sequences (Ma et al., 2005) and their identification as HY5 target genes or not (Lee et al., 2007).

**Supplemental Table S4.** Auxin-responsive genes that contain HY5 binding sites (Sibout et al., 2006).

**ACKNOWLEDGMENTS**

We thank Kiyotaka Okada for the gift of hs5-Ks50 and 35S::HY5/hs5-Ks50 seeds.

Received November 16, 2007; accepted February 13, 2008; published February 20, 2008.

**LITERATURE CITED**

Ang LH, Chattopadhyay S, Wei N, Oyama T, Okada K, Batschauer A, Deng XW (1998) Molecular interaction between COP1 and HY5 defines a regulatory switch for light control of Arabidopsis development. Mol Cell 1: 213–222

Ang LH, Deng XW (1994) Regulatory hierarchy of photomorphogenic loci: allele-specific and light-dependent interaction between the HY5 and COP1 loci. Plant Cell 6: 613–628

Chattopadhyay S, Ang LH, Puente P, Deng XW, Wei N (1998) Arabidopsis bZIP protein HY5 directly interacts with light-responsive promoters in mediating light control of gene expression. Plant Cell 10: 673–683

Cheong YH, Park JM, Yoo CM, Bahk JD, Cho MJ, Hong JC (1994) Isolation and characterization of STGA1, a member of the TGA1 family of bZIP transcription factors from soybean. Mol Cells 4: 405–412

Cheong YH, Yoo CM, Park JM, Ryu GR, Goekjian VH, Nagao RT, Key JL, Cho MJ, Hong JC (1998) STF1 is a novel TGAACG-binding factor with a zinc-finger motif and a bZIP domain which heterodimerizes with GBF proteins. Plant J 15: 199–209

Chory J, Reinecke D, Sim S, Washburn T, Brenner M (1994) A role for cytokinins in de-etiolation in Arabidopsis (det mutants have an altered response to cytokinins). Plant Physiol 104: 339–347

Cluis CP, Mouchet C, Hardtke CS (2004) The Arabidopsis transcription factor HY5 integrates light and hormone signaling pathways. Plant J 38: 332–347

Deng XW, Matsu M, Wei N, Wagner D, Chu AM, Feldmann KA, Quail PH (1992) COP1, an Arabidopsis regulatory gene, encodes a protein with both a zinc-binding motif and a G beta homologous domain. Cell 71: 791–801

Hardtke CS, Gohda K, Osterlund MT, Oyama T, Okada K, Deng XW (2000) HY5 stability and activity in Arabidopsis is regulated by phosphorylation in its COP1 binding domain. EMBO J 19: 4997–5006

Hartmann U, Sagasser M, Mehrtens F, Stracke R, Weisshaar B (2005) Differential combinatorial interactions of cis-acting elements recognized by R2R3-MYB, BZIP, and BHLH factors control light-responsive and tissue-specific activation of phenylpropanoid biosynthesis genes. Plant Mol Biol 57: 155–171

Holm M, Ma LG, Qu LJ, Deng XW (2002) Two interacting bZIP proteins are direct targets of COP1-mediated control of light-dependent gene expression in Arabidopsis. Genes Dev 16: 1247–1259

Hong JC, Cheong YH, Nagao RT, Bahk JD, Key JL, Cho MJ (1995) Isolation of two soybean G-box binding factors which interact with a G-box sequence of an auxin-responsive gene. Plant J 8: 199–211

Izawa T, Foster R, Chua NH (1993) Plant bZIP protein DNA binding specificity. J Mol Biol 230: 1113–1144

Izawa T, Foster R, Nakajima M, Shimamoto K, Chua NH (1994) The rice bZIP transcriptional activator RTA-1 is highly expressed during seed development. Plant Cell 17: 1277–1287

Jakoby M, Weisshaar B, Droge-Laser W, Vicente-Carbajosa J, Tiedemann J, Kroj T, Parcy F (2002) bZIP transcription factors in Arabidopsis. Trends Plant Sci 7: 106–111

Kim J, Tazamias D, Ellenberger T, Harrison SC, Struhl K (1993) Adaptability of the protein-DNA interface is an important aspect of sequence recognition by bZIP proteins. Proc Natl Acad Sci USA 90: 4513–4517

Kim WY, Hicks KA, Somers DE (2005) Independent roles for EARLY FLOWERING 3 and ZEITLUPE in the control of circadian timing, hypocotyl length, and flowering time. Plant Physiol 139: 1557–1569

Lam E, Benfey PN, Gilmartin PM, Fang RX, Chua NH (1989) Site-specific mutations alter in vitro factor binding and change promoter expression pattern in transgenic plants. Proc Natl Acad Sci USA 86: 7890–7894

Lam E, Katagiri F, Chua NH (1990) Plant nuclear factor ASF-1 binds to an essential region of the nopaline synthase promoter. J Biol Chem 265: 9909–9913

Lee J, He K, Stolc V, Lee H, Figueroa P, Gao Y, Tongprasit W, Zhao H, Lee I, Deng XW (2007) Analysis of transcription factor HY5 genomic binding sites revealed its hierarchical role in light regulation of development. Plant Cell 19: 731–749

Ma L, Sun N, Liu X, Jiao Y, Zhao H, Deng XW (2005) Organ-specific expression of Arabidopsis genome during development. Plant Physiol 138: 80–91

Martinez-Garcia JF, Moyano E, Alcero MJ, Martin C (1998) Two bZIP proteins from Antirrhinum flowers preferentially bind a hybrid C-box/G-box motif and help to define a new sub-family of bZIP transcription factors. Plant J 13: 489–505

Maxam AM, Gilbert W (1980) Sequencing end-labeled DNA with base-specific chemical cleavages. Methods Enzymol 65: 499–560

Nishimura R, Ohmori M, Fujita H, Kawaguchi M (2002) A Lotus basic leucine zipper protein with a RING-finger motif negatively regulates the developmental program of nodulation. Proc Natl Acad Sci USA 99: 15206–15210

Niu X, Guilfoin MN (1994) DNA binding specificity of the wheat bZIP protein EmBP-1. Nucleic Acids Res 22: 4969–4979

Olipant AR, Brandle CJ, Struhl K (1989) Defining the sequence specificity of DNA-binding proteins by selecting binding sites from random-sequence oligonucleotides: analysis of yeast GCN4 protein. Mol Cell Biol 9: 2944–2949

Osterlund MT, Hardtke CS, Wei N, Deng XW (2000) Targeted destabilization of HY5 during light-regulated development of Arabidopsis. Nature 405: 462–466

Oyama T, Shimura Y, Okada K (1997) The Arabidopsis HY5 gene encodes a bZIP protein that regulates stimulus-induced development of root and hypocotyl. Genes Dev 11: 2983–2995

Sambrook J, Fritsch EF, Maniatis T (1989) Molecular Cloning: A Laboratory Manual, Ed 2. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY

Schindler U, Beckmann H, Cashmore AR (1992a) TGA1 and G-box binding factors: two distinct classes of Arabidopsis leucine zipper proteins compete for the G-box-like element TGACGTTC. Plant Cell 4: 1309–1319

1876 Plant Physiol. Vol. 146, 2008
DNA-Binding Properties of STF1 and LONG HYPOCOTYL5

Expressed in Escherichia coli as fusions with glutathione S-transferase. Gene 67: 31–40

Vandenbussche F, Habricot Y, Condiff AS, Maldiney R, Van der Straeten D, Ahmad M (2007) HY5 is a point of convergence between cryptochrome and cytokinin signalling pathways in Arabidopsis thaliana. Plant J 49: 428–441

von Arnim AG, Deng XW (1994) Light inactivation of Arabidopsis photomorphogenic repressor COP1 involves a cell-specific regulation of its nucleocytoplasmic partitioning. Cell 79: 1035–1045

Williams ME, Foster R, Chua NH (1992) Sequences flanking the hexameric G-box core CACGTG affect the specificity of protein binding. Plant Cell 4: 485–496

Yadav V, Kundu S, Chattopadhyay D, Negi P, Wei N, Deng XW, Chattopadhyay S (2002) Light regulated modulation of Z-box containing promoters by photoreceptors and downstream regulatory components, COP1 and HY5, in Arabidopsis. Plant J 31: 741–753

Schindler U, Terzaghi W, Beckmann H, Kadesch T, Cashmore AR (1992b) DNA binding site preferences and transcriptional activation properties of the Arabidopsis transcription factor GBF1. EMBO J 11: 1275–1289

Schulze-Lefert P, Becker-Andre M, Schulz W, Halbrock K, Dangl JL (1989) Functional architecture of the light-responsive chalcone synthase promoter from parsley. Plant Cell 1: 707–714

Shin J, Park E, Choi G (2007) PIF3 regulates anthocyanin biosynthesis in an HY5-dependent manner with both factors directly binding anthocyanin biosynthetic gene promoters in Arabidopsis. Plant J 49: 981–994

Sibout R, Sukumar P, Hettiarachchi C, Holm M, Muday GK, Hardtke CS (2006) Opposite root growth phenotypes of hy5 versus hy5 hyh mutants correlate with increased constitutive auxin signaling. PLoS Genet 2: e202

Smith DB, Johnson KS (1988) Single-step purification of polypeptides expressed in Escherichia coli as fusions with glutathione S-transferase. Gene 67: 31–40

Vandenbussche F, Habricot Y, Condiff AS, Maldiney R, Van der Straeten D, Ahmad M (2007) HY5 is a point of convergence between cryptochrome and cytokinin signalling pathways in Arabidopsis thaliana. Plant J 49: 428–441

von Arnim AG, Deng XW (1994) Light inactivation of Arabidopsis photomorphogenic repressor COP1 involves a cell-specific regulation of its nucleocytoplasmic partitioning. Cell 79: 1035–1045

Williams ME, Foster R, Chua NH (1992) Sequences flanking the hexameric G-box core CACGTG affect the specificity of protein binding. Plant Cell 4: 485–496

Yadav V, Kundu S, Chattopadhyay D, Negi P, Wei N, Deng XW, Chattopadhyay S (2002) Light regulated modulation of Z-box containing promoters by photoreceptors and downstream regulatory components, COP1 and HY5, in Arabidopsis. Plant J 31: 741–753