Binding Site Selection for the Plant MADS Domain Protein AGL15
AN IN VITRO AND IN VIVO STUDY

Weining Tang and Sharyn E. Perry‡
From the Department of Agronomy, University of Kentucky, Lexington, Kentucky 40546-0312

AGL15 (for AGAMOUS-like 15) is currently the only reported member of the plant MADS domain family of transcriptional regulators that preferentially accumulates during embryo development. Additionally, AGL15 is one of the more divergent members of the MADS domain family, including within the DNA-binding domain. Previous studies have shown that MADS domain proteins bind to DNA sequences with an overall consensus of CC(A/T)6GG (called a CArG motif). Nonetheless, different MADS domain proteins exhibit similar yet distinct binding site preferences that may be critical for differential gene regulation. To determine the consensus sequence preferentially bound by AGL15 in vitro, PCR-assisted binding site selection assays were performed. AGL15 was observed to prefer a CArG motif with a longer A/T-rich core and is to date the only plant MADS domain protein having such a preference. Next, the Arabidopsis genome data base was searched for genes containing AGL15 binding sites as candidates for direct regulation by AGL15. One gene, DTA4 (for Downstream Target of AGL15-4), was identified by this method, and then confirmed as a direct target of AGL15 in vivo.

MADS domain proteins are a family of regulatory factors found in yeast, animals, and plants, containing a conserved ∼60-amino acid domain that mediates DNA binding and dimerization (for a review of the MADS domain, see Ref. 1). MADS domain proteins generally bind to a consensus DNA sequence called a CArG motif† (CC(A/T)6GG; canonical sequence); however, more detailed studies revealed that different MADS factors have overlapping but distinct binding site preferences (reviewed in Ref. 2). In plants, MADS box genes are expressed throughout the life cycle, and many members have been shown to be involved in homeotic regulation of programs in flower development (reviewed in Ref. 1). MADS box genes are more numerous in plants (e.g. over 80 in Arabidopsis thaliana) (3) than in organisms from other kingdoms. The reason for this greater diversity in plants is not clear, but it raises an important question concerning how these factors specifically regulate their downstream target genes. One possible mechanism to achieve such specificity is through differential binding to target genes.

AGL15 (for AGAMOUS-like 15) is especially intriguing because it is currently the only reported MADS domain protein that preferentially accumulates in developing plant embryos (4, 5). AGL15 has been found in nuclei of tissue developing in an embryonic mode in a wide variety of situations (5, 6). Although AGL15 appears to be expressed at its highest level during plant embryo development, it is also expressed at lower levels after germination (7). AGL15 is also one of the more divergent members of the MADS domain family in terms of primary structure, including within the MADS domain (8, 9). Therefore, it would be particularly interesting to determine whether this relatively divergent member of the family preferentially binds DNA sequences distinct from the general CArG motif. In this study, the consensus binding sequence of AGL15 was determined using PCR-assisted in vitro binding site selection assay. AGL15 was found to prefer a type of CArG motif with a longer central A/T-rich sequence, similar to the N-10 site of the MEF2A binding consensus (10). This preference is thus far unique for plant MADS domain proteins and may contribute to functional specificity of AGL15.

To understand the regulatory roles AGL15 plays, it is important to have a comprehensive knowledge of the downstream target genes it regulates in vivo. The in vitro information obtained from the binding site experiments was applied to identify potential in vivo targets of AGL15. The Arabidopsis genome data base was searched for genes containing AGL15 binding sites. One such identified gene, DTA4 (for Downstream Target of AGL15-4), was further tested to verify that it represents a direct in vivo target of AGL15. Such an approach may be useful for identifying downstream target genes of other DNA-binding regulatory proteins.

EXPERIMENTAL PROCEDURES

PCR-assisted in Vitro Binding Site Selection—The approach for binding site selection was modified from Huang et al. (11). DNA sequences encoding full-length AGL15, AGL15 lacking the DNA-binding MADS domain (AGL15ΔM), and AG were modified by PCR to add a T7 tag sequence to the C-terminal ends of the recombinant proteins, cloned into the bacterial expression vector pET15b (Novagen, Madison, WI), and transformed into Escherichia coli BL21 (DE3) cells. Expression and isolation of inclusion body proteins were performed as described previously (4). The protein samples were further purified using anti-T7 antibody-conjugated agarose beads (Novagen) according to the manufacturer’s instructions. The purity and integrity of the proteins were assessed by SDS-PAGE and by Western analysis using anti-T7 antibodies (Novagen), purified proteins were dialyzed against 1× binding buffer (10 mM Tris-HCl, pH 7.5, 50 mM NaCl, 1 mM EDTA, 1 mM dithiothreitol).

To produce a pool of double-stranded degenerate oligonucleotides, a 76-mer template was synthesized containing a core of 26 randomized positions, hand-mixed during synthesis to ensure equal representation of each base (IDT Inc., Coralville, IA). ACTCGAGGAAATTCGGTACCC-GGGTNN_TGGATCCGGAGGCTCCAAACCGT. Two additional primers corresponding to the nondegenerate arms of the 76-mer were

*This work was supported by National Science Foundation Grant IBN-9884274 (to S. E. P.) and by the University of Kentucky. This paper (number 02-06-199) is published with the approval of the Director of the Kentucky Agricultural Experiment Station. The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked “advertisement” in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.
†To whom correspondence should be addressed: Dept. of Agronomy, University of Kentucky, 1405 Veterans Rd., 307 Plant Science Building, Lexington, KY 40546-0312. Tel.: 859-257-5020 (ext. 80732); Fax: 859-257-7125; E-mail: sperr2@uky.edu.
‡The abbreviations used are: CArG motif, C(A/T)-rich G sequence; EMSA, electrophoretic mobility shift assay; ChIP, chromatin immunoprecipitation; SRF, serum response factor.
used to convert and amplify the 76-mer to double strand, using KlenTaq DNA polymerase (AB Peptide, St. Louis, MO). The double-stranded 76-mer population was labeled with [α-32P]dCTP (Amersham Biosciences) using PCR and purified on an 8% polyacrylamide gel (0.5× TBE), as described by Ausubel et al. (12).

For electrophoretic mobility shift assays (EMSAs), ~100 ng of full-length AGL15, AGL15-3M, or AG was added to 1× binding buffer supplemented with 5% glycerol, 0.5 ng/μl poly(dI-dC), 100 μg/ml bovine serum albumin. In the AG binding reaction, 1.2 μM was included (11). After preincubation at room temperature for 15 min, 102 to 103 cpm of 32P-labeled 76-mer probe was added. After a further 15 min at room temperature, the entire 20-μl binding reaction was resolved on a 5% polyacrylamide gel (0.5× TBE, 5% glycerol) to separate free probe and protein-DNA complexes. The gel was dried and exposed to a phosphorImager screen (Amersham Biosciences). The regions corresponding to shifted bands were excised, electroeluted in 0.1× TBE buffer, and precipitated. An aliquot was amplified and radiolabeled with PCR, gel-purified, and used as probe for the next round of binding and EMSA. This procedure was repeated for 3–5 rounds to enrich the oligonucleotide bound by AGL15 and AG.

**Analysis of Selected Sequences**—The selected oligonucleotides were cloned into pGEM-T vector (Promega, Madison, WI) and sequenced on an ABI Prism 310 Genetic Analyzer (Applied Biosystems, Foster City, CA). To reveal the consensus binding sequences, both strands of the random core and part of the flanking linker region were aligned using ClustalW and Jalview programs (available on the World Wide Web at www.ebi.ac.uk/clustalw), followed by manual inspection and adjustment.

**Competitive EMSA**—Probes were labeled by PCR with [α-32P]dCTP as above. To ensure equal specific activity of the probes, the labeled to unlabeled dCTP ratio was adjusted for each PCR according to the base composition of the templates. Equal cpm (102 to 103) of probes were used in EMSA. Unlabeled competitor concentrations were verified on an 8% polyacrylamide gel. EMSAs were performed under the same conditions as for binding site selection, except various amounts of unlabeled competitors were incubated with proteins for 15 min at room temperature prior to the addition of probes.

**Expression Assay**—Eight tandem repeats including a high affinity AGL15 binding site (TTATGTATATATAGA) or a mutated binding site (TTATGTATATATATA) were placed before the minimal 35S CaMV promoter (TATA-box region) (13) and an intron-containing GUS (β-glucuronidase) reporter gene (14). The reporter constructs were cloned into the binary vector pH101 (Clontech). The effector encoded an activated form of AGL15 containing a viral VP16 transcription activation domain (15) at the C-terminal end (constructed for constitutive expression. All constructs were transformed into Agrobacterium tumefaciens strain GV3850.

The transient expression assay was performed using a petunia infiltration system (a formal detailed procedure will be reported by Drs. M. Schoenbeck and J. Chappell, University of Kentucky). Briefly outlined here, overnight agrobacteria cultures were washed once and resuspended to A600 of 0.5 in 10% sucrose. Cells harboring reporter and effector constructs were mixed at various ratios. The mixture was infiltrated into the abaxial side of leaves of 4–6 week old Petunia hybrida plants grown under standard greenhouse conditions. The leaves were then left in a moist chamber at room temperature for 3–5 days. GUS activity was measured by 4-MUG assay (16) using a Hoefer DynaQuant 200 fluorometer (Hoefer Scientific Instruments, San Francisco, CA), following the manufacturer’s instructions.

**Enrichment Tests**—To determine whether AGL15 binds in vivo to DNA sequences identified in vitro, enrichment assays were performed on populations of DNA fragments isolated by chromatin immunoprecipitation (ChIP) from an embryonic culture tissue using AGL15-specific antibody. The AGL15 antisem is has been previously described and detailed (17). The detailed ChIP protocol is described elsewhere (17). Potential in vivo targets of AGL15 were identified by searching the data base (available on the World Wide Web at www.arabidopsis.org) with the preferred in vitro binding sites. Oligonucleotide primers were synthesized to amplify the target DNA fragments containing these sites, as well as the coding regions of β-tubulin (18), and elongation factor 1α (19, 21). Oligonucleotide primers were designed for both bring primer pairs were used in multiplex PCR. A series of dilutions of total input chromatin DNA (before ChIP), AGL15-antisem ChIP-selected DNA, and preimmune serum ChIP-selected DNA were used as templates. PCRs were resolved on 2% agarose gels.

**Expression Pattern of Target Gene**—A. thaliana silique RNA was isolated using the hot borate method (20). cDNAs were synthesized using oligo(dT) and Moloney murine leukemia virus reverse transcriptase (Promega), according to the manufacturer’s instructions. Gene expression level was determined semiquantitatively using PCR. The coding region of elongation factor 1α was amplified as a normalization control.

**RESULTS**

**Selection of AGL15 Binding Sites in Vitro**—PCR-assisted in vitro binding site selection was used to isolate DNA sequences preferentially bound by AGL15 and AG from a random population of oligonucleotides. Binding site selection has previously been performed for AG (11, 21) and served as a positive control in our experiments. AGL15 and AG proteins were isolated as inclusion bodies from E. coli and further purified via a C-terminal T7 tag. The integrity and purity of recombinant proteins were confirmed by Coomassie staining and Western analysis (not shown). Oligonucleotides bound by AGL15 or AG were separated from free DNA by EMSA and amplified by PCR. After 3–5 rounds of selection, shifted bands were clearly visible, indicating enrichment of AGL15 and AG binding sites (not shown). Binding site selection was also performed using a form of AGL15 that lacked the DNA-binding MADS domain (AGL15-3M). This form of AGL15 does not bind to a CArG motif from the yeast STE6 promoter, whereas full-length AGL15 is able to bind to this site in EMSA (5). Therefore, AGL15-3M served to monitor the background binding site selection. No DNA was recovered when AGL15-3M was used (not shown).

Selected binding site sequences were aligned and consensus sequences were calculated. Summaries of base preferences of AGL15 and AG at each position of the binding site are shown in Fig. 1, A and B, respectively. Selected sequences were categorized into three groups, where the number refers to the length of the A/T-rich core: CC-6-GG represents a canonical CArG motif with the sequence 5′-CC-(A/T-rich)-6-GG-3′; C-8-G contains a longer A/T-rich core (i.e. 5′-C-(A/T-rich)-6-GG-3′); and C-7-GG/CC-7-G is an intermediate form. The calculated consensus sequence from alignment indicated that the preferred binding site of AGL15 contained a longer A/T-rich core. In the AGL15-selected population, the canonical CC-6-GG motifs constituted less than 25% of the selected sites (Fig. 1C). The majority of the binding sites were of the C-8-G and C-7-GG/CC-7-G forms, with the former predominating. The opposite trend was observed in the AG-selected population. The preferred AG binding sites had the canonical form, as previously reported (11, 21). The preference of AGL15 for a C-8-G type of CArG motif is unique among plant MADS domain protein binding site sequences reported thus far (11, 21–23). Even among MADS domain proteins from different kingdoms, only MEF2A has similar binding site preference (10).

**AGL15 Has Higher Affinity for CArG Motifs with a Longer A/T-rich Core**—To test whether AGL15 binds to the C-8-G CArG motif with higher affinity, several individual CArG sequences were selected from the populations for EMSAs. CarG1 and CarG2 were AG-selected CC-6-GG type motifs. CarG3 was a C-8-G type CArG repeatedly selected by AGL15. A representative EMSA is shown in Fig. 2A. For AG, all three individual CArGs had similar binding affinity. This was not an artifact of probe or protein saturation, because the same pattern was observed over a wide range of probe/protein ratios (not shown).

AGL15, however, showed clear differential affinity among the three CArGs. CarG3, containing the longer A/T core, showed a strong interaction with AGL15 in EMSA, whereas binding to CarG1 and CarG2 (CC-6-GG type) was considerably weaker. Reciprocal competition assays were used to confirm the preference of AGL15 for the C-8-G CArG motif. Unlabeled CArG sequences were used as competitors against themselves and
AGL15 could bind under site selection assay identified an array of DNA sequences that although the situation preference for binding to an unusual C-8-G type CArG motif. Specific competitor than CC-6-GG type CArG, and, therefore, AGL15 had a particular form of CArG motif, although CC-6-GG type CArG motifs competed better than CC-6-GG type CArGs, and, therefore, AGL15 had higher binding affinity to this form of CArG sequence. Competitive EMSAs revealed a lack of preference of AG for any partner, but, for simplicity, only the top protein-DNA complexes are shown.

Taken together, these data suggest that for AGL15, a C-8-G type CArG3 with a longer A/T core was a much stronger competitor than CC-6-GG type CArG3 also competed better (Fig. 2B, compare lanes 2–4, lanes 5–7 and lanes 8–10 with lanes 2–4; note the difference in -fold level of unlabeled competitor). When CArG1 or CArG2 was used as radiolabeled probe, unlabeled CArG3 could effectively compete with the probe (Fig. 2B, lanes 2–4). In the presence of a 750-fold excess of CArG3 competitor, shifted bands were virtually undetectable (not shown). In contrast, unlabeled CArG1 or CArG2 competitors had little effect in the competition with radiolabeled CArG3 probe, even at a much higher competitor excess (Fig. 2B, compare lanes 5–7 and lanes 8–10 with lanes 2–4; note the difference in -fold level of unlabeled competitor). When CArG1 or CArG2 was used as radiolabeled probe, unlabeled CArG3 also competed better (Fig. 2B, compare lanes 15–17 with lanes 12–14, and compare lanes 22–24 with lanes 19–21). Taken together, these data suggest that for AGL15, a C-8-G type CArG3 with a longer A/T core was a much stronger competitor than CC-6-GG type CArGs, and, therefore, AGL15 had higher binding affinity to this form of CArG sequence. Competitive EMSAs revealed a lack of preference of AG for any particular form of CArG motif, although CC-6-GG type CArG motifs competed slightly better than the C-8-G motif (Fig. 2C).

Identifying in Vivo Targets of AGL15—PCR-assisted binding site selection assay identified an array of DNA sequences that AGL15 could bind under in vitro conditions and revealed a preference for binding to an unusual C-8-G type CArG motif. Although the situation in vivo is likely to be very different, correlations have been reported between in vitro binding affinity and functional in vivo binding (24). In the context of the chromatin environment, with the presence of interacting proteins, it is very likely that AGL15 binds DNA sequences that it is unable to bind in vitro (17), just as the animal MADS domain protein SRF binds sites in vivo that it is unable to bind in vitro (reviewed in Ref. 2). Conversely, due to chromatin structure, it is also likely that DNA sequences that correspond to preferred sites in vitro are not bound in vivo (17, 25). Therefore, it is important to test whether the information obtained through a strictly in vitro study could be used to identify in vivo targets of AGL15.

Among all of the selected AGL15 binding sites, one particular sequence (TTACTATATACTAGAA) was especially intriguing. AGL15 showed high affinity for oligonucleotides containing this perfect 16-bp palindromic sequence (e.g. CArG3 in Fig. 2B). More interestingly, this CArG motif was highly represented in the AGL15-selected populations. About 30% (12 of 44) of the oligonucleotides after three rounds of selection contained this site. The percentage rose to 50% after the fourth round of selection and almost 100% after the fifth round (not shown). In addition, the consensus derived from alignment of nonredundant binding sites selected by AGL15 matched this particular motif (Fig. 1A). Does AGL15 bind to this type of CArG motif in vivo?

To address this question, this site was first assessed for function in vivo using a transient reporter assay. Eight tandem repeats containing this site were placed in front of a minimal 3S CaMV promoter (TATA-box), which in turn drives the expression of a GUS reporter gene. As a control, a similar construct containing tandem repeats of a mutated version of the binding site (TTAGTATATACTACT) was also prepared. AGL15 cannot bind to this mutated sequence in vitro, as tested.
AGL15 could bind to this DNA sequence in vivo, the question remains whether this naturally occurring site in the chromatin context was also associated with AGL15. To test this, enrichment tests in conjunction with chromatin immunoprecipitation (ChIP) assays were performed. ChIP isolates in vivo AGL15-DNA complexes and has been shown to be an effective approach to select direct downstream targets of AGL15 (17). For the enrichment tests, primers were designed to amplify the genomic DNA region containing the potential AGL15 binding site within DTA4 as well as primers to amplify coding regions of various genes that are not expected to be bound by AGL15 as controls. Multiplex PCR was performed using both target and control primer pairs on ChIP-selected DNA populations (Fig. 4, ChIP) from embryonic culture tissue as well as DNA populations selected with preimmune serum (PI-ChIP) DNA, using the control amplification product as a reference. An in vivo AGL15-bound DNA fragment should be represented at higher abundance in the population after ChIP selection than in total DNA (i.e. it is enriched by ChIP). A representative gel image of an enrichment test is shown in Fig. 4. Input chromatin DNA was diluted so that the β-2-tubulin control was amplified to a comparable amount as in the ChIP and PI-ChIP populations. The intensity of the DTA4 band in ChIP was significantly stronger than observed in input or PI-ChIP, indicating strong enrichment of the DTA4 fragment by ChIP. Reproducible results were obtained from enrichment assays performed on several independently isolated ChIP populations. The DTA4 fragment was also not enriched in control ChIP experiments using immune serum depleted for anti-AGL15 IgGs, or in ChIP populations from wild type flower buds that accumulate many AGL15 transcripts (7). RNA was isolated from wild type and transgenic plant materials (DTA4) and control (β-2-tubulin) levels in each sample were determined by semiquantitative reverse transcriptase-PCR (RT-PCR). For comparison, the template amount in each PCR was adjusted empirically so that β-2-tubulin was amplified to a similar extent in all samples. Input, total input chromatin DNA; ChIP, DNA selected using AGL15-specific antisera; PI-ChIP, DNA selected using preimmune serum.

**DISCUSSION**

Many DNA-binding regulatory factors are members of large families, generally grouped by homology to a DNA-binding domain. Members of these families often recognize similar or identical DNA sequences. Although cross-regulation could be...
involved in extending the developmental roles of a transcription factor, discrimination is likely to be an important aspect to execute specific developmental programs. How do individual members achieve different developmental consequences? It is possible that there is discrimination in binding sites, leading to differential gene regulation. Or perhaps binding site recognition overlaps extensively, and different developmental programs are realized through other mechanisms such as interactions with other proteins or chromatin architecture. Most likely, all of these mechanisms contribute to functional specificity.

The problem of functional specificity is especially obvious for plant MADS domain proteins, which number over 80 in *A. thaliana*, and play diverse, critical roles in development (3). Several members of the family have been found to bind related DNA binding sites (CArG motifs), although with distinct preferences in some cases. Despite the fact that AG, AP1, AP3, and PI, members of the *Arabidopsis* MADS domain family involved in flower development, can bind the same or very similar CArG motifs (28), they have different developmental roles. In some cases, domain swapping experiments have localized biological specificity within the plant MADS and adjoining linker domains (27), but other experiments suggest that DNA-binding site recognition is not the determining factor for execution of specific developmental programs (27, 28).

In animals, the MADS domain proteins SRF and MEF2A have more stringent in vitro binding site preferences. SRF binds to a serum response element of the form CC(A/T)\(_n\)GG *in vitro*, but not to the N-10 site, which has a consensus sequence of CTA(A/T)\(_n\)TAG. Conversely, MEF2A has been reported to bind to the N-10 site but not a serum response element *in vitro* (29). Also important is the degree of DNA bending induced by binding of these proteins (−73° for SRF; −19° for MEF2A). Bending has been shown to be important for gene regulation for DNA-binding proteins in general as well as for members of the MADS family (30, 31). The key determinant in DNA site recognition and bending within the MADS domain of SRF and MEF2A is a single amino acid residue that is not strongly conserved among MADS box genes (29). In contrast, the degree of bending by the plant MADS domain protein SQUAMOSA was found to be intrinsic to the DNA-binding site rather than determined by the protein (32). Also, for the yeast MADS domain protein MCM1, the DNA sequence greatly influences the extent of bending (31). A single base change within the A/T-rich sequences flanking the CArG motif does not affect binding significantly but has a dramatic effect on degree of bending and on activation and repression of downstream targets of MCM1 (31).

Recent work has indicated that the DNA-binding site is not a passive player in the DNA-protein interaction. Several studies have found that a given transcription factor can exhibit different activities depending on DNA-binding site (reviewed in Ref. 33). Binding to different DNA sequences may result in changes in protein conformation and/or interactions with other proteins, with consequences ranging from different degrees of transcriptional activation to activation at some sites but repression at others. Differences in binding affinity to different sites may also have developmental consequences. In a microarray study to identify targets of the *C. elegans* FOXA protein, PHA-4, it was found that the relative affinity, as assessed *in vitro* by EMSA, correlated with the onset of gene expression during development. Genes with higher affinity *in vitro* binding sites are expressed earlier in development, whereas genes with lower affinity sites are expressed later (34). Therefore, an understanding of preferred DNA sequences bound *in vitro* can be very relevant to understanding how a DNA-binding protein functions *in vivo*.

**AGL15 Has a Unique Binding Site Preference**—AGL15 offers a distinct opportunity to examine binding site preference because AGL15 is one of the most diverse members of the MADS domain family in higher plants, including within the DNA-binding MADS domain (8, 9). PCR-assisted binding site selection revealed that AGL15 binds with highest affinity to a variant CArG of form C-S-G with strong preference for AT-flanking sequences. Although other plant MADS domain proteins will bind to this type of sequence, to our knowledge, preference for this type of CArG is unique to AGL15 among plant MADS domain binding sites studied to date. MEF2A, a MADS domain protein involved in mammalian muscle-specific gene regulation, specifically recognizes DNA sites of form C-S-G (10). In particular, AGL15 recognized a palindromic sequence of TTACTATATATAGTTA that was preferentially selected in two entirely independent experiments. EMSAs with individual selected CArG sequences (Fig. 2A) and competition experiments (Fig. 2B) confirmed that AGL15 preferentially binds to CArG motifs of the C-S-G form *in vitro*. AGL15 was also able to bend DNA (−90°, not shown), as determined by circular permutation assay (35, 36) using a truncated form of AGL15. Interestingly, AGL15 also bent a CArG motif of the form CC-6-GG to a similar extent (not shown), indicating that bending may be a function of the protein for AGL15, as found for the animal MADS domain proteins, rather than of the DNA binding site, as found for SQUAMOSA.

**The Value of in Vitro Study: Identifying in Vivo Targets**—We demonstrate in this report that in *vitro* binding site information for AGL15 is relevant for *in vivo* function and can be used directly to identify previously unknown or unsuspected target genes. Binding sites identified for AGL15 *in vitro* are able to function *in vivo*. A transient reporter system demonstrated that a high affinity *in vitro* site (TTACTATATATAGTTA) could lead to a high level of GUS reporter expression upon AGL15 binding *in vivo*. The mutated site was not active, indicating sequence-specific interaction. The *Arabidopsis* genome was searched with the sequence corresponding to the highest affinity binding site. Only one exact match was found, corresponding to sequence in the single intron of *DTA4*, a gene encoding a hypothetical protein. This *cis*-element was evaluated for association with AGL15 in the chromatin context by performing ChiP to isolate DNA fragments that bind AGL15 *in vitro* and testing whether a DNA fragment containing the *DTA4* putative CArG motif was preferentially selected compared with DNA fragments not expected to be bound by AGL15. As shown in Fig. 4, the DNA fragment containing the palindromic CArG motif was enriched in the ChiP population relative to total input DNA or preimmune serum-selected DNA, indicating selection in ChiP *via in vivo* bound AGL15. In addition, expression of *DTA4* responds to the level of AGL15. In the presence of increased levels of AGL15 as found in 35S::AGL15 transgenic plants, expression of *DTA4* is increased. Therefore, AGL15 binds and regulates gene expression *in vivo* via a *cis*-element.
corresponding to a DNA sequence identified in vitro.

It has been suggested that for a gene to qualify as a direct downstream target of a transcription regulator, three criteria must be satisfied (37, 38): 1) the regulatory protein should bind to the cis-element in vitro; 2) expression of the target gene should respond to changes in the level of the regulator; and 3) the binding site for the regulator in the target gene should confer regulation by the regulatory protein in vitro. DTA4, a gene identified based on in vitro binding site information, passes all three of these tests for demonstrating direct regulation by AGL15. Further characterization of this gene and its role in development is currently underway.

Although PCR-assisted binding site selection can lead to identification of in vivo targets, it is important to recognize that the information obtained is strictly in vitro. In the chromatin environment, consensus sites cannot be assumed to be occupied by a DNA-binding factor. Several other CArG motifs that contained CTATATATAG with A/T-rich flanking sequences were identified in the 5' regulatory regions of genes by data base searches. Of three tested, two appeared to bind AGL15 in vivo (not shown). Selective binding to a subset of potential DNA-binding sites is a theme emerging from genome-wide studies mapping the comprehensive in vivo binding sites of transcription factors (25). Additionally, binding may occur to nonconsensus sites in vivo. Nevertheless, in vitro binding site selection provides valuable information. It is the most convenient method to select preferentially bound sequences from a pool containing virtually all possibilities, and binding affinity, as determined in vitro, has been correlated with in vivo activity (24, 34).

Several DTA sequences identified by direct sequencing of a ChIP population were reported in a previous study (17). Whereas DTA2 contains a canonical CArG motif of form CC(A/T)6CG in the 5' regulatory region (17), other DNA fragments isolated by ChIP do not contain canonical binding sites for MADS domain proteins. Information obtained from the binding site selection experiments reported here has assisted in identification of potential binding sites in DNA fragments isolated by ChIP. The regulatory regions of DTA1 and DTA3, for example, do not contain any canonical CArG motifs. The putative binding sites for AGL15 in the regulatory region of DTA1 contain a C-S-G motif. Recent work indicates that this site is involved in the expression of DTA1 in response to AGL15 in vivo.2 The DNA fragment isolated by ChIP from DTA3 is strongly enriched in the ChIP populations, but the most likely binding site for AGL15 has a form of CC(A/T)6CG, which is not of a form isolated by binding site selection assays.3 It seems likely that co-factors are involved in binding of AGL15 to DTA3 regulatory regions in vivo. Other DNA fragments isolated by ChIP have a variety of potential binding sites, including CC-6-GG, C-7-GG/CC-7-G, and C-S-G forms. In addition, the information obtained from the experiments described in this report allowed identification of in vivo binding sites for AGL15 within its own regulatory regions.3 These sites of form C-S-G were previously ignored because they did not fit the canonical type CArG motif. Therefore, information obtained by binding site selection experiments has been useful for discovery of novel direct targets (DTA4), determination of binding sites for AGL15 in DNA fragments selected by ChIP (DTA1 and DTA2), and identification of binding motifs in suspected targets (AGL15).

Acknowledgments—We thank Dr. H. Wang for providing AGL15-VP16 effector and help with the reporter constructs; Drs. M. Schoenbeck and J. Chappell for the petunia system; and Drs. S. Adhya for the bending vector pBEND2, D. Falcone for pMMIC, D. Fernandez for AGL15-T7, and H. Ma for AG constructs. We thank Drs. J. Chappell, R. Dinkins, L. Dirk, and H. Wang for helpful comments on the manuscript.

REFERENCES
1. Riechmann, J. L., and Meyerowitz, E. M. (1997) Biol. Chem. 378, 1079–1101
2. Shore, P., and Sharrocks, A. D. (1995) Eur. J. Biochem. 229, 1–13
3. Jack, T. (2001) Plant Mol. Biol. 46, 515–529
4. Heck, G. R., Perry, S. E., Nichols, K. W., and Fernandez, D. E. (1995) Plant Cell 7, 1271–1282
5. Perry, S. E., Nichols, K. W., and Fernandez, D. E. (1996) Plant Cell 8, 1977–1989
6. Perry, S. E., Lehti, M. D., and Fernandez, D. E. (1999) Plant Physiol. 120, 121–129
7. Fernandez, D. E., Heck, G. R., Perry, S. E., Patterson, S. E., Bleeker, A. B., and Fang, S.-C. (2000) Plant Cell 12, 183–197
8. Theissen, G., Kim, J. T., and Saedler, H. (1999) J. Mol. Biol. 343, 484–516
9. Arias-Buylla, E. R., Pelas, S., Liégeuen, S. J., Gould, S. E., Burgeff, C., Ditta, G. S., de Pouplana, L. R., Martinez-Castilla, L., and Yanofsky, M. F. (2000) Proc. Natl. Acad. Sci. U. S. A. 97, 5328–5333
10. Pollock, R., and Treisman, R. (1991) Genes Dev. 5, 2327–2341
11. Huang, H., Mizukami, Y., Hu, Y., and Ma, H. (1993) Nucleic Acids Res. 21, 4769–4776
12. Ausubel, F. M., Brent, R., Kingston, R. E., Moore, D. D., Seidman, J. G., Smith, J. A., and Struhl, K., eds. (1998) Current Protocols in Molecular Biology, pp. 12.11.1–12.11.8, John Wiley & Sons, Inc., New York
13. Campisi, L., Yang, Y. T., Yi, H., Herman, E., Herman, B., Cassista, A. J., Allen, D. W., Xiang, H. J., and Jack, T. (1999) Plant J. 17, 699–707
14. Vancanneyt, G., Schmidt, R., Ooncensantes, A., Willmitzer, L., and Rochassa, M. (1990) Mol. Gen. Genet. 220, 245–250
15. Dalymply, M. A., McGeoch, D. J., Davison, A. J., and Preston, C. M. (1985) Nucleic Acids Res. 13, 7865–7879
16. Yang, Y. N., Li, R. G., and Qi, M. (2000) Plant J. 22, 543–551
17. Wang, H., Tang, W., Zhu, C., and Perry, S. (2003) Plant J. 32, 831–843
18. Snustad, D. P., Haas, N. A., Kopczak, S. D., and Silflow, C. D. (1992) Plant Physiol. 100, 7865–7879
19. Axelos, M., Bardet, C., Liboz, T., Thai, A. L. V., Curie, C., and Lescure, B. (1999) Mol. Gen. Genet. 261, 196–211
20. Wilkins, T. A., and Smith, M. B. (1996) A Laboratory Guide to DNA Isolation, Analysis, and Synthesis (Krieg, P. A., ed) pp. 21–41, Wiley-Liss, Inc., New York
21. Shirahashi, H., Okada, K., and Shimura, Y. (1993) Plant J. 4, 385–398
22. Huang, H., Tudor, M., Su, T., Zhang, Y., Hu, Y., and Ma, H. (1996) Plant Cell 8, 81–94
23. West, A. G., Causier, B. E., Davies, B., and Sharrocks, A. D. (1998) Nucleic Acids Res. 26, 5277–5287
24. Carr, A., and Buggin, M. D. (1999) EMBO J. 18, 1508–1608
25. Wyrick, J. J., and Young, H. A. (2002) Curr. Opin. Genet. Dev. 12, 130–136
26. Riechmann, J. L., Krizek, B. A., and Meyerowitz, E. M. (1996) Proc. Natl. Acad. Sci. U. S. A. 93, 4793–4798
27. Krizek, B. A., and Meyerowitz, E. M. (1996) Proc. Natl. Acad. Sci. U. S. A. 93, 4063–4070
28. Krizek, B. A., Riechmann, J. L., and Meyerowitz, E. M. (1999) Sex. Plant Breed. 12, 14–26
29. West, A. G., Shore, P., and Sharrocks, A. D. (1997) Mol. Cell. Biol. 17, 2876–2887
30. Murphy, E. C., Zharkov, V. B., Louis, J. M., Corniscule, G., and Clare, G. M. (2001) J. Mol. Biol. 312, 481–499
31. Acton, T. B., Zhong, H., and Verashon, A. K. (1997) Mol. Cell. Biol. 17, 1881–1889
32. West, A. G., and Sharrocks, A. D. (1999) J. Mol. Biol. 286, 1311–1323
33. Latchman, D. S. (2001) Trends Biochem. Sci. 26, 211–213
34. Gaudet, J., and Mango, S. E. (2002) Science 295, 821–825
35. Thompson, J. F., and Landy, A. (1988) Nucleic Acids Res. 16, 9687–9705
36. Kim, J., Zwieb, C., Wu, C., and Adhya, S. (1989) Gene (Amst.) 85, 15–23
37. Andrew, D. J., and Scott, M. P. (1992) New Biol. 4, 5–15
38. Carey, M., and Smale, S. T. (1999) Transcriptional Regulation in Eukaryotes: Concepts, Strategies, and Techniques, pp. 291–317, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY

2 H. Wang and S. Perry, unpublished observation.
3 C. Zhu and S. Perry, unpublished observation.
