The Plant Cell, Vol. 8, 81–94, January 1996

DNA Binding Properties of Two Arabidopsis MADS Domain Proteins: Binding Consensus and Dimer Formation

Hai Huang,[a] Matthew Tudor,[a,1] Thomas Su,[a,2] Yue Zhang,[a,b] Yi Hu,[a] and Hong Ma[a,3]

[a] Cold Spring Harbor Laboratory, 1 Bungtown Road, Cold Spring Harbor, New York 11724-2212

[b] Molecular and Cellular Biology Graduate Program, State University of New York, Stony Brook, New York 11794-5215

MADS domain proteins are members of a highly conserved family found in all eukaryotes. Genetic studies clearly indicate that many plant MADS domain proteins have different regulatory functions in flower development, yet they share a highly conserved DNA binding domain and can bind to very similar sequences. How, then, can these MADS box genes confer their specific functions? Here, we describe results from DNA binding studies of AGL1 and AGL2 (for AGAMOUS-like), two Arabidopsis MADS domain proteins that are preferentially expressed in flowers. We demonstrate that both proteins are sequence-specific DNA binding proteins and show that each binding consensus has distinct features, suggesting a mechanism for specificity. In addition, we show that the proteins with more similar amino acid sequences have more similar binding sequences. We also found that AGL2 binds to DNA in vitro as a dimer and determined the region of AGL2 that is sufficient for DNA binding and dimerization. Finally, we show that several plant MADS domain proteins can bind to DNA either as homodimers or as heterodimers, suggesting that the number of different regulators could be much greater than the number of MADS box genes.

INTRODUCTION

Transcriptional regulation is a universal mechanism that controls development and differentiation. In animals, many developmental regulatory genes have been found to encode well-known transcription factors, such as homeodomain proteins, basic leucine zipper proteins, and helix-loop-helix proteins (for review, see Gehring, 1987; Biggins and Tjian, 1989; Scott et al., 1989; Struhl, 1989). The MADS box gene family is also a widely conserved group of genes that includes genes encoding transcription factors from humans (serum response factor [SRF]) and yeast (MCM1) (Norman et al., 1988; Passmore et al., 1988) and the floral homeotic genes *AGAMOUS* (*AG*) from Arabidopsis and *DEFICIENS* (*DEF*) from Antirrhinum (Sommer et al., 1990; Yanofsky et al., 1990). Hence, the conserved region was named MADS box, for MCM1, AG, DEF, and SRF (Schwarz-Sommer et al., 1990). In recent years, many MADS box genes have been isolated from a variety of plants, including both dicots and monocots (Ma, 1994; Yanofsky, 1995).

Many of the plant MADS box genes are expressed in the flower, and several have been shown to be important regulators of flower development by genetic and molecular studies (Ma, 1994). In Arabidopsis and Antirrhinum, genetic analysis has revealed that different MADS box genes have distinct roles, from controlling meristem transition to specifying various floral organ identities (Bowman et al., 1989; Irish and Sussex, 1990; Sommer et al., 1990; Yanofsky et al., 1990; Huijser et al., 1992; Jack et al., 1992; Mandel et al., 1992; Tröbner et al., 1992; Bradley et al., 1993; Goto and Meyerowitz, 1994). Many additional MADS box genes are expressed in the flower with distinct patterns and may play different roles in regulating flower development (Ma et al., 1991; Yanofsky, 1995). How do these genes carry out their distinct functions? It has been shown that AG and DEF/GLOBOSA (GLO) bind to DNA in vitro (Mueller and Nordheim, 1991; Schwarz-Sommer et al., 1992; Tröbner et al., 1992; Huang et al., 1993; Shiraishi et al., 1993), suggesting that they are indeed transcription factors; therefore, one obvious answer is that different MADS domain factors regulate different sets of target genes. The question then becomes how each MADS domain protein recognizes its own set of target genes. DNA binding is essential for sequence-specific transcription factors to function; the transcription factors SRF and MCM1 are known to bind to sequences with a consensus called the CArG box (CC[A/T]$_6$GG) (Pollock and Treisman, 1990; Wynne and Treisman, 1992), and AG can also bind to similar sequences (Huang et al., 1993; Shiraishi et al., 1993). Do other plant MADS domain proteins also bind to sequences with a CArG box? To address this question, we characterized the DNA binding properties of two Arabidopsis MADS domain proteins: AGL1 and AGL2 (for AG-like).

AG, *AGL1*, and *AGL2* are all preferentially expressed in the flower, but they have quite different expression patterns within the flower (Yanofsky et al., 1990; Bowman et al., 1991; Drews

[1] Current address: National Institute of Neurological Disorders and Stroke, National Institutes of Health, Building 10, Room 3D-04, Bethesda, MD 20892.

[2] Current address: Department of Biology, University of California, Los Angeles, CA 90024.

[3] To whom correspondence should be addressed.

et al., 1991; Ma et al., 1991; Flanagan and Ma, 1994). The wild-type Arabidopsis flower has four types of organs arranged in concentric rings or whorls—sepals, petals, stamens, and carpels—and its development has been divided into 12 stages up to anthesis (Smyth et al., 1990). Among these three genes, *AGL2* expression appears first, at stage 2 throughout the spherical floral meristem; it continues in all four types of floral organ primordia and diminishes as organs mature. *AG* expression begins slightly later than *AGL2* in the central region of the floral meristem at stage 3, when sepal primordia appear; *AG* expression continues in the reproductive organ primordia and later becomes localized in specific cells. *AGL1* expression comes on last, beginning at approximately stage 8, and is confined to the carpel wall and to developing ovules. The degrees of amino acid sequence similarity are also different among AG, AGL1, and AGL2. AGL1 is very similar to AG, with 95 and 52% identity within and outside the MADS domain, respectively. Although AGL1 and AGL2 are 82% identical in the MADS domain, they have only ~19% identity outside the MADS domain (Ma et al., 1991). Therefore, both expression patterns and amino acid sequences suggest that AGL1 and AGL2 are likely to have different regulatory functions.

Here, we describe results from in vitro DNA binding studies of AGL1 and AGL2. We found that they both bind to DNA in a sequence-dependent fashion, and we determined their binding consensus sequences. We note that the similarity between the binding consensus of the plant MADS domain proteins correlates with the amino acid sequence similarity of the DNA binding domains. We also found that MADS domain proteins have different affinities for different sequences. In addition, we demonstrate that AGL2 binds to DNA as a dimer, and we describe the regions involved in DNA binding and subunit interaction. Finally, we show that several plant MADS domain proteins can bind to DNA as homodimers and heterodimers. These results suggest mechanisms by which a limited set of individual MADS domain proteins can specifically regulate a diverse group of target genes.

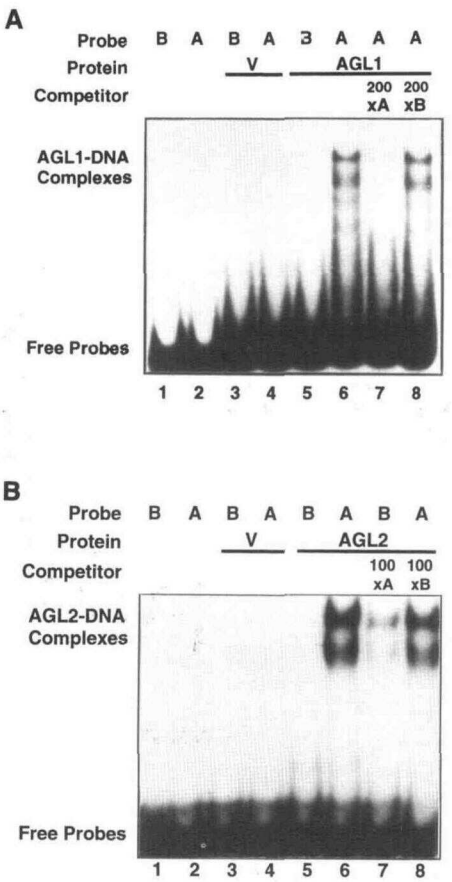

Figure 1. AGL1 and AGL2 Bind to a Specific Sequence.

(A) AGL1 gel mobility shift experiment.

(B) AGL2 gel mobility shift experiment.

The following probes were used: A, oligo(A); B, oligo(B). No protein was added to lanes 1 and 2, and extracts from cells with vector only (V) were added to lanes 3 and 4. Excess nonradioactive oligo(A) (xA, lane 7) or oligo(B) (xB, lane 8) was used as the competitor. More than one band was observed for AGL1– and AGL2–DNA complexes, most likely due to partial degradation of the proteins because gel shifts with truncated proteins had only one band (see Figure 6).

RESULTS

AGL1 and AGL2 Bind to Specific Sequences

To understand the AGL1 and AGL2 proteins as transcription factors, we characterized these proteins in vitro using DNA binding assays. They were expressed in *Escherichia coli* cells, using the T7 expression systems (Studier et al., 1990), and *E. coli* cell extracts containing the proteins were used throughout this study. First, we tested DNA binding with a known sequence (oligo(A), containing a CArG box, CCATTAATGG) that was shown to bind AG (Huang et al., 1993). Our results, as shown in Figure 1, indicate that both AGL1 (Figure 1A) and AGL2 (Figure 1B) bind to oligo(A) (lane 6) but not to a mutant oligonucleotide (lane 5; oligo(B), with CArG mutated to GGATGCATCC; Huang et al., 1993) and that the binding of the labeled probe was reduced by excess nonradioactive oligo(A) but not by oligo(B). This indicates that AGL1 and AGL2 are both sequence-specific DNA binding proteins.

To determine the spectrum of AGL1 and AGL2 binding sequences, we selected oligonucleotides from a pool of random sequences. DNA bound to the protein was purified, amplified by polymerase chain reaction (PCR), and used for the next round of selection. We found that the sequences capable of binding to AGL1 or AGL2 were being enriched through successive selections (Figures 2A and 2B). After four rounds of selection and PCR, we cloned the oligonucleotides and analyzed their sequences. Based on the isolated binding sequences, we derived the AGL1 and AGL2 binding consensus sequences (Table 1). These consensus sequences resemble

those of SRF, MCM1, and AG. Nevertheless, the AGL2 binding sequences show some differences from those of AG and AGL1; in particular, the CArG box of many AGL2 binding sequences allows an A residue instead of a G residue at position 9 and, to a lesser degree, a T instead of a C residue at position 2 (Table 1). In addition, AGL2 is much less tolerant of G or C residues at positions 7 and 8 than are AG and AGL1.

As an additional test for AGL2 binding to the consensus region of the selected sequences, we performed DNase I protection assays of four high-affinity oligonucleotides with AGL2; the results are shown in Figure 3. In each case, AGL2 protects a region of the oligonucleotide that includes both the consensus region and some flanking sequences. In addition, hypersensitive sites were detected in the presence of the protein, suggesting DNA conformational changes due to the binding of AGL2.

Because the AGL1 binding consensus is quite similar to that of AG but the AGL2 consensus is slightly different from those of AG and AGL1, we investigated this difference further by testing AGL2 or AG binding to a number of different sequences. Figure 4 shows the results from gel mobility shift experiments with AGL2 (Figure 4A) or AG (Figure 4B). Because the precise amount of active protein could not be determined, the amount of protein–DNA complex reflects only the relative affinity of a particular sequence for that protein. Two aspects of relative binding affinity are revealed by these experiments. First, for the same protein, different sequences may have different affinities. Here, we compare sequences in three pairs (Figure 4C): (1) oligo(A) is compared with oligonucleotide 31; (2) oligonucleotide 74 with oligonucleotide 115; and (3) oligonucleotide 125 with oligonucleotide 147. Oligo(A) has a standard CArG box with C and G residues at positions 2 and 9, respectively, but oligonucleotide 31 has a CArG box with T and A residues at these positions; the oligonucleotide 31 CArG box fits the AGL2 consensus but not the AG or AGL1 consensus. Accordingly, we found that AGL2 binds well to both sequences (Figure 4A), but AG binds well only to oligo(A) (Figure 4B). For the other two pairs, the CArG boxes are similar between the two within a pair; the only important difference is a substitution of a C or G for a T normally found in the internal positions (7 and 8, respectively). The results show that AGL2 binds much less well to the sequences with an internal C or G residue than it does to sequences with an internal T (Figure 4A, lane 4 versus 5, and lane 6 versus 7). On the other hand, AG binds equally well to both sequences of a pair (Figure 4B, lanes 4 versus 5, and lane 6 versus 7). These results further support the idea that AGL2 favors A and T residues at the central six positions of the CArG box and allows CT and AG at the ends. However, AG binds equally well to sequences with G/C or A/T at positions 7 and 8 and prefers C and G at positions 2 and 9, respectively. This is consistent with the AG binding sequence consensus.

These results also reveal another aspect of relative affinity. AGL2 and AG may have different affinities for the same sequence relative to the same standard. For example, using oligo(A) as a reference, oligonucleotide 147 binds more strongly than oligo(A) to AGL2 but more weakly than oligo(A) to AG. In addition, oligonucleotide 31 binds to AGL2 nearly as well as oligo(A), but it has a much lower affinity for AG than oligo(A). On the other hand, oligonucleotide 125 binds AGL2 extremely weakly, yet it binds AG more strongly. The relative affinities of these sequences suggest that both AG and AGL2 have different affinities for different nucleotide sequences and that the binding selections were able to recover sequences with a wide range of affinities.

DNA Binding Domain of AGL2

In AGL2, as in most plant MADS domain proteins, sequence comparisons define four structural domains, as shown in Figure 5: the MADS domain, which is the most conserved region;

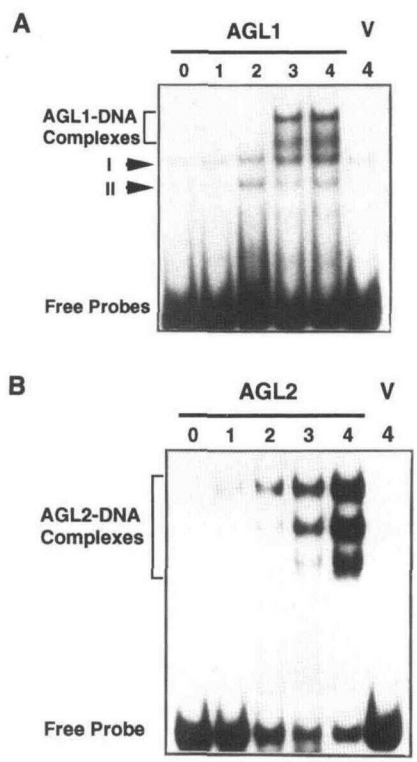

Figure 2. Selection of AGL1 and AGL2 Binding Sequences.

(A) Binding sequence selection with AGL1.

(B) Binding sequence selection with AGL2.

AGL1 or AGL2 was used in all lanes except for the rightmost lane. This lane contains the control extract from the vector-only cells (V). The numbers above the lanes indicate the number of cycles of the AGL1 or AGL2 binding reaction and gel purification that the probes had undergone. For AGL1, the top two bands represent AGL1–DNA complexes; DNA was purified from the top band. The lower bands (I and II) probably represent nonspecific binding because they are also present in the vector lane. For AGL2, all three bands represent AGL2–DNA complexes, but the top band was not reproducibly observed. Therefore, the DNA was purified from the middle band.

Table 1. AGL1 and AGL2 Binding Sequence Consensus[a]

	−4	−3	−2	−1	1	2	3	4	5	6	7	8	9	10	11	12	13	14
AGL1																		
Consensus																		
(*n* = 99)																		
A	25	19	8	29	0	1	45	47	67	37	21	23	18	7	42	71	60	21
G	26	13	10	27	0	1	9	9	1	3	13	14	79	89	7	8	11	33
C	26	9	1	9	99	97	8	10	2	5	25	10	0	0	6	14	14	31
T	22	58	80	34	0	0	37	33	29	54	40	22	2	3	44	6	14	14
	N	T	T	−C	C	C	A/T	A/T	A/t	T/A	N	N	G	G	T/A	A	A	N
AGL1																		
Consensus																		
(*n* = 65, totally random[b])																		
A	21	12	4	21	0	1	30	31	42	26	14	15	7	4	26	48	39	18
G	15	9	9	22	0	1	7	8	1	2	9	28	56	56	5	4	8	19
C	11	6	1	5	65	63	6	4	1	3	14	8	0	0	6	10	9	15
T	18	38	51	17	0	0	22	22	21	34	28	14	2	5	28	3	9	13
	N	T	T	−C	C	C	A/T	A/T	A/t	T/A	N	N	G	G	T/A	A	A	N
AGL2																		
Consensus																		
(*n* = 92, all except twin sites)																		
A	19	21	28	34	0	0	71	34	67	49	58	8	45	0	42	62	49	15
G	29	23	6	14	0	0	4	6	0	0	7	7	47	92	9	2	14	31
C	27	12	5	10	92	86	7	2	0	0	7	12	0	0	12	6	15	22
T	17	36	53	29	0	6	10	50	25	43	20	65	0	0	29	22	14	24
	N	N	T/a	N	C	C	A	T/A	A/t	A/T	A/t	T	G/A	G	A/T	A/t	A	N
AGL2																		
Consensus																		
(*n* = 51, totally random[b])																		
A	14	14	11	19	0	0	39	20	37	25	28	4	26	0	23	40	33	6
G	7	10	6	9	0	0	1	4	0	0	4	3	25	51	3	0	5	17
C	15	6	4	7	51	46	5	2	0	0	4	6	0	0	11	3	5	8
T	15	21	30	16	0	5	6	25	14	26	15	38	0	0	14	8	8	20
	N	N	T/a	N	C	C	A	T/A	A/t	T/A	A/T	T	A/G	G	A/t	A	A	N
AG																		
Consensus[c]																		
(*n* = 66, totally random)																		
A	21	9	10	29	0	0	31	47	52	25	17	19	7	2	22	45	40	15
G	6	3	1	8	0	0	6	0	1	1	11	20	57	54	5	5	9	16
C	20	3	0	8	66	65	3	2	0	0	15	8	0	0	17	4	6	10
T	19	51	55	21	0	1	26	17	13	40	23	19	2	10	22	12	11	25
	N	T	T	A/T	C	C	A/T	A/t	A/t	T/A	N	N	G	G	−G	A	A	N
AGL3																		
Consensus[d]																		
(*n* = 42, totally random)																		
A	18	8	5	14	2	0	34	18	28	19	28	3	29	0	12	25	23	9
G	6	8	3	6	0	0	0	3	0	0	1	0	11	40	2	6	7	10
C	11	5	5	9	40	36	1	2	0	0	1	3	0	0	9	2	6	8
T	7	21	29	13	0	6	7	19	14	23	12	36	2	2	19	9	6	15
	N	T	T	N	C	C	A	T/A	A/T	T/A	A/t	T	A/g	G	−G	A/t	A	N

[a] The consensus sequences are derived from selected binding sequences, which will be provided upon request, according to the following rules based on percentage of occurrence: (1) If A (for example, same for each base) is >50% and none of the others is >20%, then A is used; (2) if (A + T) is >70% and A is >T is >30%, then A/T is used; (3) if (A + T) is >70% and A is > (2 × T) is >20%, then A/t is used; (4) if C is <10% and each of the others is >20%, then −C is used; (5) if none is >50%, no two combined are >70%, and each is >10%, then N is used. When the situation is between two of those described above, then the one closest is used. The CArG boxes contain residues 1 to 10.

[b] These are the sequences with no overlap between the constant ends of the oligonucleotide and the region from positions −3 to 13.

[c] From Huang et al. (1993).

[d] From Huang et al. (1995).

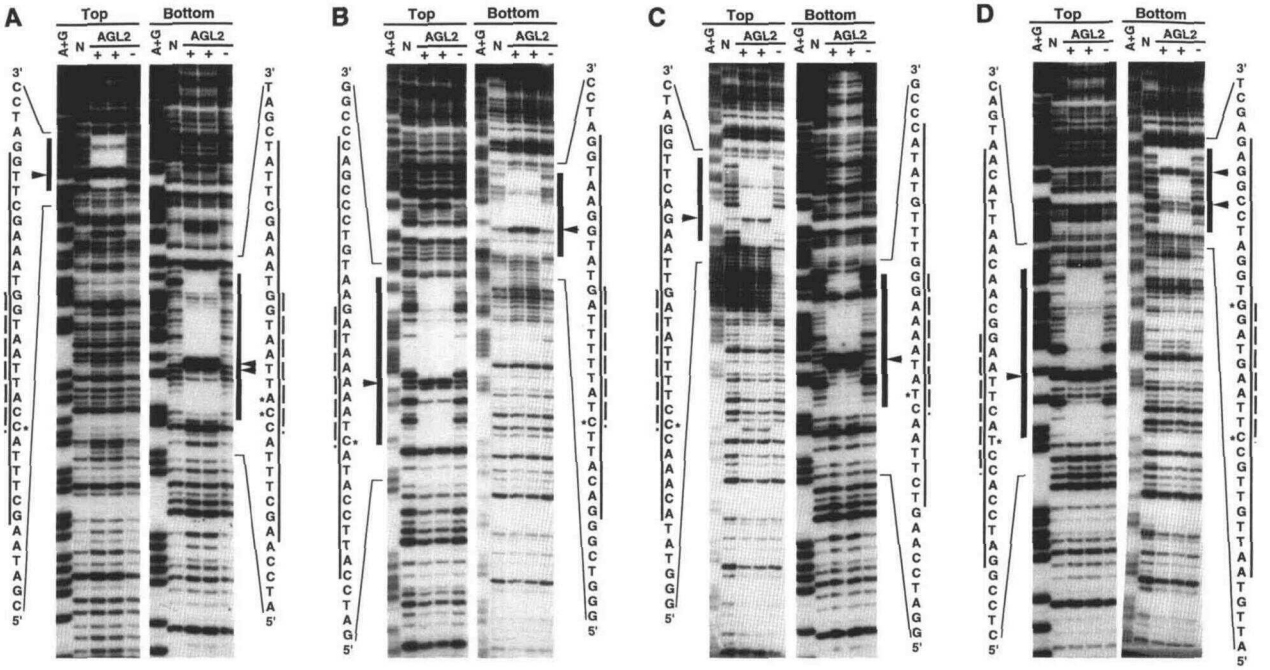

Figure 3. DNase I Protection Footprinting of AGL2.

(A) Footprint with oligo(A).
(B) Footprint with oligonucleotide 31.
(C) Footprint with oligonucleotide 76.
(D) Footprint with oligonucleotide 111.
Solid bars adjacent to the autoradiographs indicate the protected regions, and arrowheads point to hypersensitive sites. The sequences with thin lines are the protected sequences, the dashed lines highlight the core region of the consensus, and asterisks are adjacent to the hypersensitive nucleotides. Top, top strand (the sequences used for derivation of consensus); Bottom, bottom strand. Lanes: A+G, chemical sequencing reaction; N, no protein; AGL2(+), extract from cells carrying the *AGL2* construct; AGL2(−), extract from cells carrying the vector.

the K domain, a region of moderately conserved sequence and structural similarity with keratins; a divergent I domain between the two conserved domains; and a C-terminal nonconserved domain. To determine the domains required for AGL2 DNA binding, we generated several truncated AGL2 proteins: AGL2Δ158 lacks only the C-terminal nonconserved domain; AGL2Δ98 is missing the C-terminal nonconserved domain and most of the K domain; AGL2Δ77 has the MADS domain and a little over half of the I domain; and AGL2Δ70 has only the MADS domain and less than half of the I domain (Figure 5). These truncated AGL2 proteins were expressed in *E. coli* (data not shown) and tested for DNA binding, using gel mobility shift assays. We found that all of the truncated AGL2 proteins can bind to DNA (Figure 5 and Figure 6A, lanes 1 to 4), although the smallest one, AGL2Δ70, binds much more weakly than expected for the amount of protein added (Figure 6A, lane 1; also, data not shown). When the AGL2Δ70 protein was present at a lower level, no band shift was detected (data not shown). As controls, none of the extracts binds to oligo(B) (data not shown). The weak binding of AGL2Δ70 indicates that the last seven amino acid residues of AGL2Δ77 are important for DNA binding albeit not absolutely required.

Our results indicate that the K and C-terminal domains are not required for DNA binding, and neither is the C-terminal portion of the I domain. The MADS domain and the first 21 residues of the I domain are sufficient for DNA binding, and removal of the last seven of these residues considerably weakens the binding. The AGL2 DNA binding domain of 77 residues is the smallest among the characterized MADS domain proteins: SRF has an N-terminal extension, and both SRF and MCM1 DNA binding domains have longer C-terminal extensions (Norman et al., 1988; Nurrish and Treisman, 1995).

AGL2 Binds to DNA as a Dimer

It is known that SRF and MCM1 bind to DNA as homodimers (Norman et al., 1988; Passmore et al., 1989; Mueller and Nordheim, 1991). To test whether AGL2 also binds DNA as a dimer, we performed DNA binding experiments with mixtures

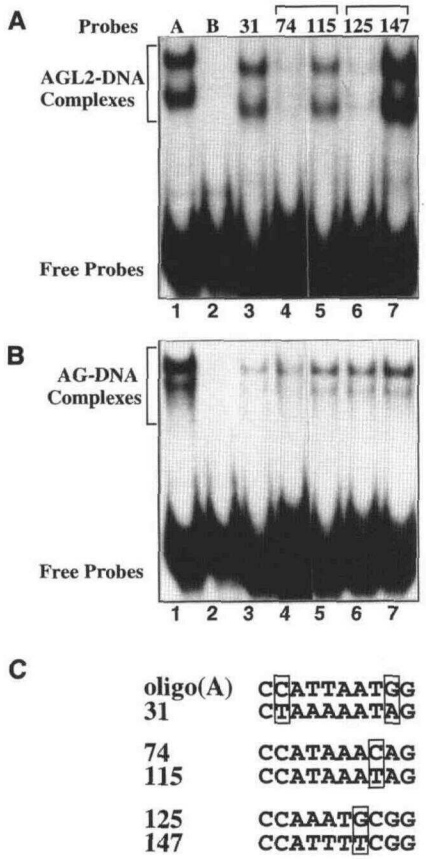

Figure 4. AGL2 Binds Different Sequences with Different Affinities.

(A) DNA binding by AGL2.

(B) DNA binding by AG.

(C) Sequences of the probes used in **(A)** and **(B)**. The G/C to A/T changes are boxed.

For controls, oligo(A) and oligo(B) (lanes 1 and 2) were used. The *AGL2 oligonucleotide* numbers are the clone numbers. The extract from cells with vector only was used as a negative control, and none of the oligonucleotides had any shifted bands (data not shown). The oligonucleotides were labeled with a radioactive primer using PCR, so they have the same specific activity; the same amount of radioactivity was used in each lane.

of different truncated AGL2 proteins. If each truncated AGL2 protein binds to DNA as a homodimer, then a heterodimer of the two differently sized truncated proteins will result in a gel mobility shift intermediate between those of the two truncated proteins alone. As shown in Figure 6A, for the three pairs of truncated proteins (AGL2Δ77/AGL2Δ98 [lane 6], AGL2Δ77/AGL2Δ158 [lane 8], and AGL2Δ98/AGL2Δ158 [lane 9]), there is an intermediate band in addition to the two bands representing binding of the individual truncated proteins (compare lane 6 with lanes 2 and 3, lane 8 with lanes 2 and 4, and lane 9 with lanes 3 and 4). This indicates that heterodimers were formed between any two of AGL2Δ77, AGL2Δ98, and AGL2Δ158 and that the heterodimers can bind to DNA (Figure 5). This result and the fact that AGL2Δ77 can bind to DNA as a homodimer (lane 2) indicate that AGL2Δ77 contains sufficient residues to form dimers in vitro. In addition, in the presence of longer proteins, the shortest one, AGL2Δ70, binds to DNA as a heterodimer with AGL2Δ98 (lane 5) and AGL2Δ158 (lane 7) much more readily than as a homodimer, even when AGL2Δ70 was present at a lower level (data not shown).

These results suggest that the interaction between two AGL2Δ70 subunits is weaker than those between AGL2Δ70 and longer proteins or those between any longer ones; this weak dimerization resulted in lower amounts of shifted probes even when AGL2Δ70 was alone and at a high concentration. AGL2Δ70/AGL2Δ77 heterodimer formation could not be tested because the mobilities of the dimers were not resolvable in our gel mobility shift assays. These results indicate that AGL2Δ77 contains amino acid residues sufficient for AGL2 dimerization and that AGL2Δ70 lacks several residues that are important, but not essential, for the dimerization.

DNA Binding by Heterodimers between MADS Domain Proteins

Several MADS box genes are known to be expressed in the flower, and many of these have overlapping expression patterns. For example, at stage 3, *AG* is expressed at the center of the floral meristem, and *AGL2* is expressed throughout the same meristem (Yanofsky et al., 1990; Drews et al., 1991; Flanagan and Ma, 1994); therefore, both *AG* and *AGL2* are expressed at the center. This coexpression provides an opportunity for the formation of heterodimers between different MADS domain proteins. We sought to test for formation of such heterodimers in vitro and for DNA binding by the heterodimers. As shown in Figure 6B, all three AGL2 truncated proteins, when mixed with a truncated AG protein lacking the C-terminal nonconserved domain (34M-I-K; see Methods), resulted in a band (lanes 5 to 7) intermediate between the bands from individual proteins (compare with lanes 1 to 4). We have also tested heterodimer formation of AGL2Δ77 or AGL2Δ70 with a truncated AG protein lacking both the K and C-terminal domains (34M-I; see Methods) and found that both of the truncated AGL2 proteins can form dimers with the shorter AG protein (data not shown). These results indicate that AGL2 can form heterodimers with AG in vitro and that AGL2Δ70 interacts with AG amino acid residues N-terminal to the AG K domain.

Two other Arabidopsis MADS box genes, *AGL1* and *AGL3*, are also expressed in flowers (Ma et al., 1991; Huang et al., 1995). *AGL1* is expressed during late stages of flower development in the carpel wall and ovules. Because *AG* and *AGL2* are both expressed in the ovules late in flower development, any pair of two proteins from among AG, AGL1, and AGL2 may be expressed in the same cells. Because *AGL2* is expressed in all floral organs and *AGL3* is expressed in flowers, it is likely that *AGL3* expression overlaps with that of *AGL2* and possibly

with expression of other MADS box genes. Therefore, we tested for heterodimer formation and DNA binding, using all possible pairings between these four proteins. The results shown in Figure 6C indicate that all of the pairs of these four proteins can bind to DNA as a heterodimer. For example, AGL1 can bind to DNA as a homodimer (Figure 6C, lane 1) or as a heterodimer with AGL2 (lane 2). Therefore, each of AG, AGL1, AGL2, and AGL3 can bind to DNA as a homodimer or as a heterodimer with one of the remaining three.

DISCUSSION

AGL1 and AGL2 Have Different DNA Binding Consensus Sequences

We show here that AGL1 and AGL2 are sequence-specific DNA binding proteins and that they bind to many related, yet distinct, sequences. In addition, AGL2 has different affinities for different binding sequences, as do AG and AGL3, and the relative binding affinities for different sequences correlate with the degree of similarity of the particular sequence to the AGL2 binding consensus. When two AGL2 binding sequences were compared, the one that deviates from the consensus sequence binds to AGL2 much less well than the other. There are exceptions, however, because oligonucleotide 147 has a C residue at position 8, which does not fit the consensus, yet it binds to AGL2 better than oligo(A) does. We also noted that AGL2 exhibits preferences for certain nucleotides at positions flanking the CArG box.

The sequences that we recovered have a range of binding affinities; thus, they probably represent a broad spectrum of AGL1 and AGL2 binding sequences. Similar sequences are found in upstream regions of cloned Arabidopsis genes (Huang et al., 1993, 1995); therefore, they could serve as binding sites in target genes of the MADS domain proteins. Furthermore, the fact that the binding affinity of a protein for different sequences varies in vitro suggests that such affinity may also vary in vivo. Because each MADS domain protein may regulate a number of target genes, binding affinity provides a mechanism for differential regulation of individual target genes by the same protein. This suggests that changes in the level of a MADS domain protein may affect some of its functions more than others, as supported by studies with antisense *AG* RNA in transgenic plants (Mizukami and Ma, 1995).

Arabidopsis has many MADS domain proteins with different functions, and some have overlapping regions of expression (Drews et al., 1991; Ma et al., 1991; Jack et al., 1992; Mandel et al., 1992; Flanagan and Ma, 1994; Goto and Meyerowitz, 1994; Kempin et al., 1995; Savidge et al., 1995). Therefore, a particular MADS domain protein must be able to distinguish its target genes from those of other MADS domain proteins. Our results with AG and AGL3 (Huang et al., 1993, 1995) and with AGL1 and AGL2 indicate that different MADS domain proteins can bind to distinct sequences with different affinities in vitro. This suggests that distinct binding sequences and different binding affinities may be important for binding site selection in vivo, contributing to the functional specificities of MADS domain proteins.

Correlation between the Amino Acid Sequences and DNA Binding Consensus Sequences of MADS Domain Proteins

We noted that among the four proteins AG, AGL1, AGL2, and AGL3, the proteins that have more similar DNA binding domains

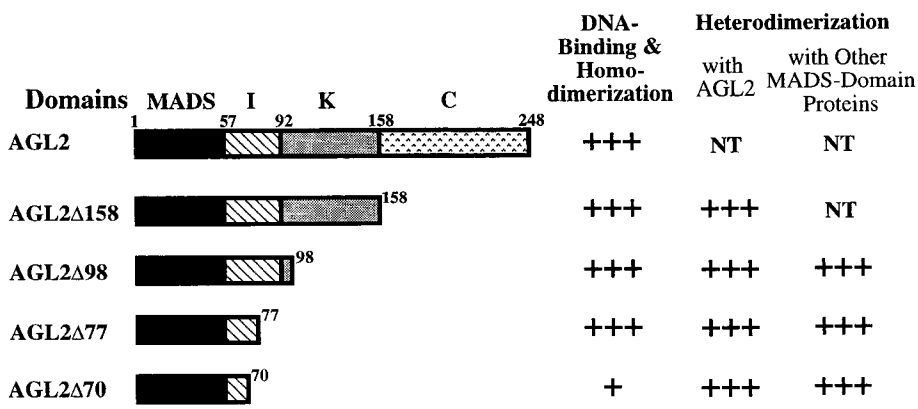

Figure 5. DNA Binding and Dimerization Properties of Truncated AGL2 Proteins.

Shown are schematic representations of the AGL2 domains—MADS, I domain (I), K domain (K), and the C-terminal domain (C)—as well as four truncated AGL2 proteins. The numbers indicate the amino acid number at the boundary of domains. The DNA binding and dimerization properties of the truncated proteins are indicated to the right. +++ and +, strong and weak bands, respectively, in gel mobility shift experiments; see Figure 6. NT, not tested.

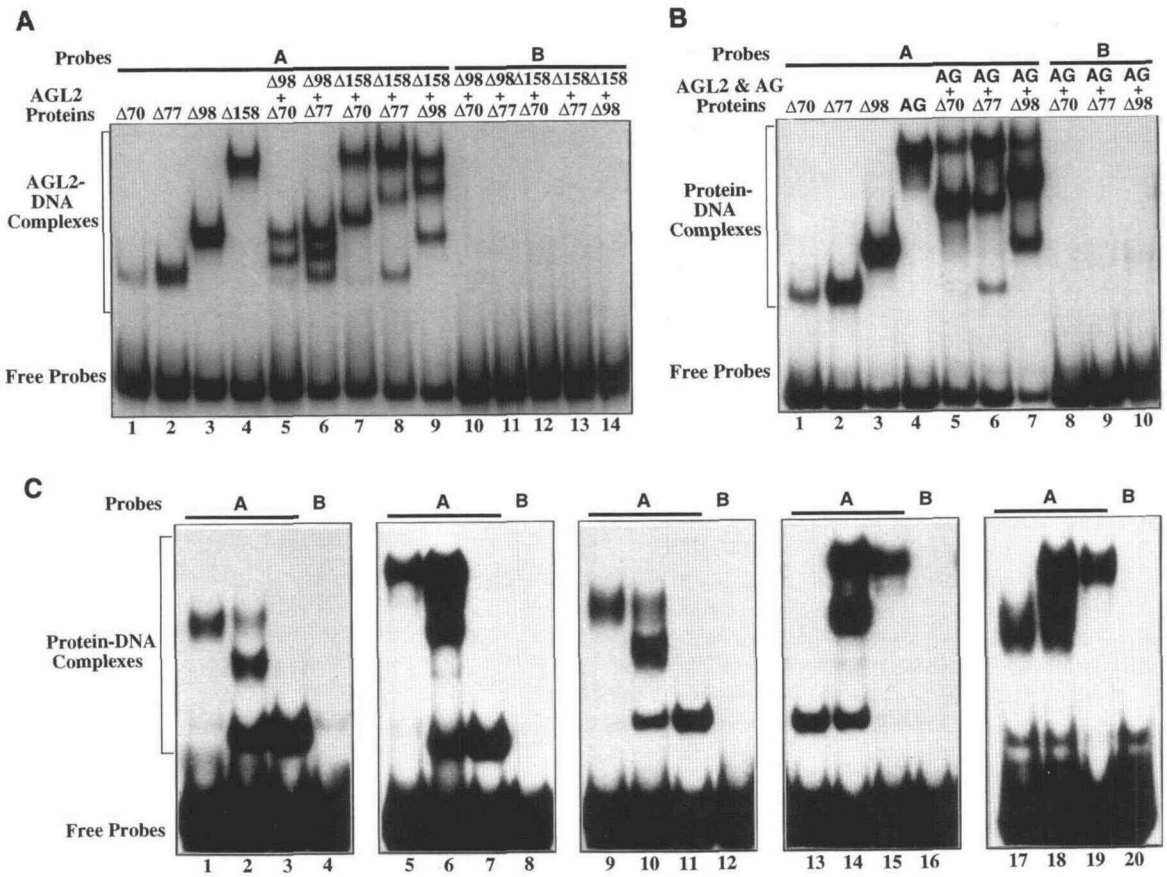

Figure 6. DNA Binding by AGL2 Dimers and Heterodimers.

(A) Heterodimers between truncated AGL2 proteins. Complexes of the AGL2Δ70 homodimer with DNA were observed in lanes 5 and 7 after a longer exposure (not shown). The amount of AGL2Δ70 used was at least four times as much as the others.

(B) Heterodimers between truncated AGL2 proteins and AG (34M-I-K). The faint bands in lanes 8 to 10 are nonspecific.

(C) Heterodimers among AG (M-I), AGL1 (lacking the C domain), AGL2 (Δ77), and AGL3 (full length). The odd-numbered lanes contain individual proteins (the left one being first in the order given below), and the even-numbered lanes contain both of the proteins: AGL1 + AGL2Δ77; AGL3 + AGL2Δ77; AGL1 + AG (M-I); AG (M-I) + AGL3; AGL1 and AGL3. The probes were A, oligo(A), and B, oligo(B); bars indicate that the same probe was used.

(MADS and I domains) also bind to more similar DNA sequences (Figures 7A and 7B). We compared the amino acid sequences of the DNA binding domains of these four plant MADS domain proteins in the context of the recent structure of the SRF–DNA cocrystal (Pellegrini et al., 1995). One way amino acid differences could contribute to binding sequence specificity is through the interaction between individual divergent amino acids and specific nucleotide bases. However, two SRF residues (Arg-143 and Lys-163) important for contacting the bases of the binding sequence (Pellegrini et al., 1995) are invariant in the four plant proteins (Figure 7C); therefore, most of the amino acids divergent between AG/AGL1 and AGL2/AGL3 are not responsible for differences in direct recognition of bases.

Nevertheless, there is one feature in the plant MADS domain protein sequences that could contribute to binding sequence specificity. In mammals, two SRF-related proteins, RSRFC4 and RSRFR2, have the methionine initiation residue just upstream of the MADS domain and bind to a CArG box ($CTA[A/T]_4TAG$; Pollock and Treisman, 1991) that is different from that of SRF ($CC[A/T]_6GG$). Pellegrini et al. (1995) proposed that the methionine residue just before the MADS domain of RSRFR proteins could contribute to binding sequence specificity. AG and AGL1, both with an N-terminal extension before the MADS domain, bind to sequences with a G residue at position 9 (similar to SRF), whereas the proteins AGL2 and AGL3, both predicted to begin with the MADS domain, bind to sequences with either a G or A residue at this

position (similar to RSRF). Therefore, the presence or absence of a methionine could explain the binding sequence specificity between AGL2/AGL3 and AG/AGL1.

Because most of the amino acid–base interactions are conserved in all of the plant MADS domain proteins, other factors must influence their binding sequence specificities. Pellegrini et al. (1995) concluded that the intrinsic and induced DNA conformation is the principal DNA feature recognized by SRF. First, the CArG box DNA sequence contains an AT-rich center that is similar to sequences known to bend (Haran et al., 1994), and the binding of SRF causes further bending (Gustafson et al., 1989). We have shown that AG, AGL2, or AGL3 binding to DNA caused DNase I hypersensitive sites (Huang et al., 1993, 1995), suggesting that binding by these proteins also induces further DNA bending. Second, most of the SRF residues contacting DNA interact with the phosphate or deoxyribose; therefore, although these interactions do not contribute to base recognition directly, they may recognize different degrees of intrinsic DNA bending resulting from different CArG sequences. As shown in Figure 7C, several of the residues that differ between AG/AGL1 and AGL2/AGL3 correspond to those in SRF that are involved in contacting the DNA backbone; therefore, AG and AGL1 could have different interactions with DNA than those of AGL2 and AGL3.

Another consequence of different amino acid sequences is differing protein conformation. In particular, the compactness of the dimer of MADS domain proteins is related to the strength of the interactions between the two subunits. Among the MADS domain residues that are conserved between AG and AGL1 and between AGL2 and AGL3, but not between AG/AGL1 and AGL2/AGL3, a few correspond to residues of SRF identified as being important for dimerization (Figure 7C). Furthermore, the I domain, which is required for dimerization, contains many residues that are conserved within each of the AG/AGL1 and AGL2/AGL3 pairs but divergent between the pairs (Figure 7C). Therefore, the dimeric conformations of AG and AGL1 may be different from those of AGL2 and AGL3, resulting in preferential binding to one type of DNA sequences (with a characteristic intrinsic conformation) instead of another. Because replacing AT with GC residues at positions 7 and 8 of the CArG box can alter the degree of bending of the DNA and AG and AGL1 can tolerate both AT and GC at these positions, it is possible that the conformations of AG and AGL1 allow the binding of sequences with a wider range of bending than those of AGL2 and AGL3. In short, the differences in both DNA backbone contacts and protein conformation may contribute to the different DNA binding sequence preferences for AG/AGL1 versus AGL2/AGL3. It is worth pointing out that because the I domain is required for dimerization, which may indirectly affect binding sequence specificity, differences in the I domain, in addition to those in the MADS domain, could contribute to target gene selection.

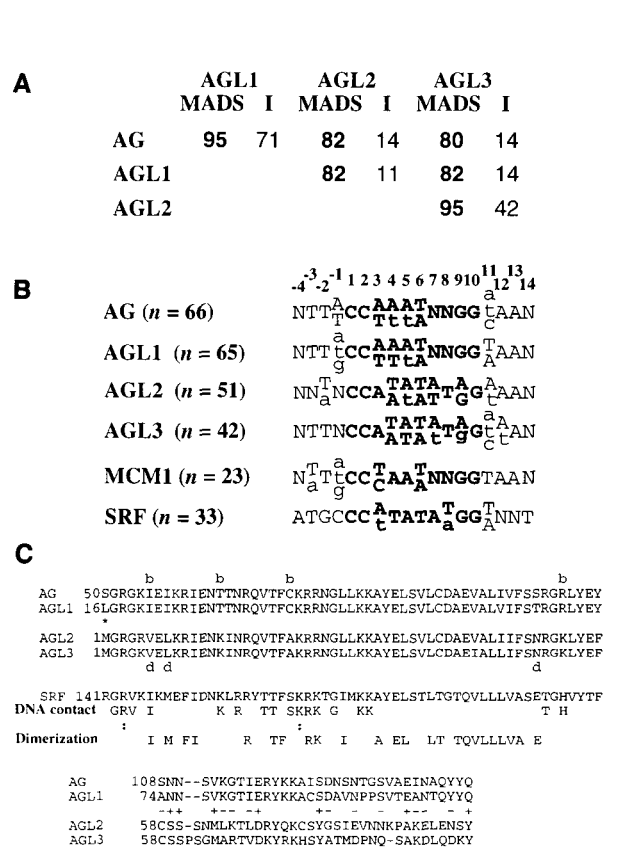

Figure 7. Amino Acid Sequences of MADS Domain Proteins and Their Binding Consensus Sequences.

(A) Comparison of MADS and I domain amino acid sequences. Numbers indicate percentage of identity.

(B) DNA consensus binding sequences: those of SRF (Pollock and Treisman, 1990) and MCM1 (Wynne and Treisman, 1992) are also shown. Uppercase and lowercase letters reflect relative frequencies; see Table 1 for explanations. N, any of the four nucleotides.

(C) MADS domain (top) and I domain (bottom) sequences. The MADS domain residues that are divergent between AG/AGL1 and AGL2/AGL3 and that correspond to those of SRF contacting the DNA are indicated with the letter (b) above the AG sequence; those that are divergent and correspond to the SRF residues involved in dimerization are indicated with the letter (d) below the AGL3 sequence. The asterisk indicates the methionine residue present in AGL2 and AGL3. The SRF residues involved in DNA interaction or dimerization are shown below the SRF sequence, and the two highly conserved base-contacting residues, Arg-143 and Lys-163, are indicated with a colon. The I domain residues that are divergent between subfamilies are indicated with either a plus sign for those that are identical within either pair of sequences or a minus sign for those that are identical in one of the pairs and similar in the other.

AGL2 Dimerization Involves More than One Region of Interaction

We show here that the dimeric AGL2 DNA binding domain includes both the MADS domain and a part of the adjacent C-terminal region (I domain), as do the DNA binding domains of SRF and MCM1 (Norman et al., 1988; Passmore et al., 1989;

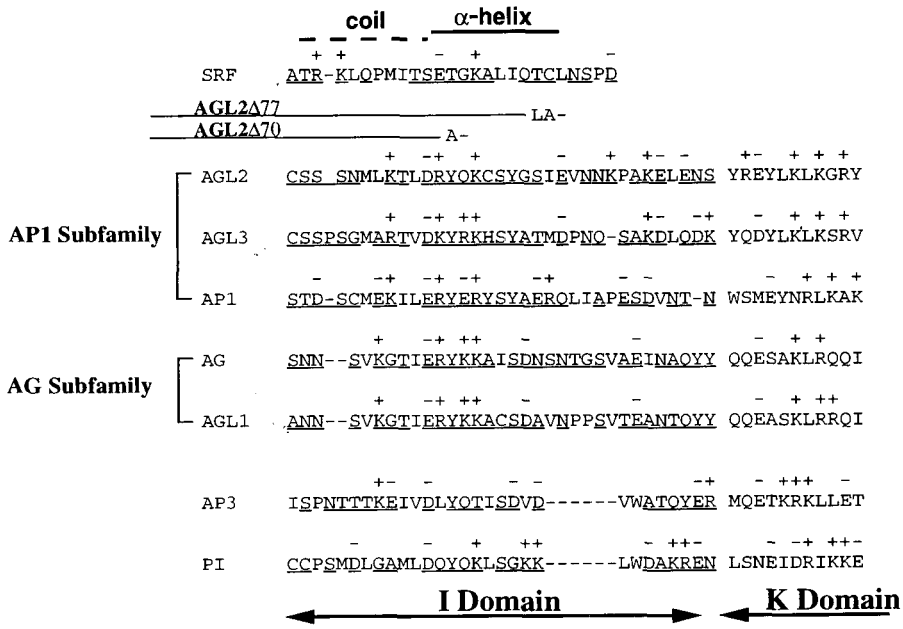

Figure 8. I Domain of MADS Domain Proteins.

Sequences of the I domains and the N termini of the K domains are presented. Shown are AG (Yanofsky et al., 1990); AGL1 and AGL2 (Ma et al., 1991); AGL3 (Huang et al., 1995); AP1 (Mandel et al., 1992); AP3 (Jack et al., 1992); and PI (Goto and Meyerowitz, 1994). For comparison, the coil and α helix of SRF (Pellegrini et al., 1995) are also shown. The charged residues are highlighted by the plus or minus signs above the sequences, and hydrophilic residues are underlined. The C termini of the AGL2Δ70 and AGL2Δ77 proteins are shown, with the (–) sign indicating the carboxyl group.

Mueller and Nordheim, 1991). Therefore, although the primary sequence of the AGL2 I domain is not similar to the corresponding regions in SRF and MCM1, they are all needed for DNA binding. In addition, the fact that the AGL2 K domain is not required for dimer formation in vitro suggests that it may have other functions, possibly interacting with other proteins.

We found that if two truncated AGL2 proteins were mixed under nondenaturing conditions and then used in gel mobility shift assays, the DNA complexes with the two homodimers, but not that with the heterodimer, could be detected (data not shown), indicating that subunit exchange did not occur when proteins were not denatured. The heterodimers formed only when denatured proteins were mixed and then renatured. The heterodimers between different plant MADS domain proteins were also observed after similar treatment. This is similar to the formation of DEF/GLO and SRF/MCM1 heterodimers (Mueller and Nordheim, 1991; Schwarz-Sommer et al., 1992). These observations suggest that MADS domain proteins exist as stable dimers separable only by denaturation. The three-dimensional structure of SRF indicates that there are two α helices and two β strands that are involved in dimerization (Pellegrini et al., 1995). The MADS domain contains the first long α helix and the two β strands, and these structures of both SRF subunits form an antiparallel coiled coil and a four-stranded antiparallel β sheet, respectively (Pellegrini et al., 1995). Both the coiled coil and the β sheet have many hydrophobic interactions, providing forces for dimerization that is stable in solution.

We also found that an AGL2 truncated protein containing only the first 70 amino acid residues (AGL2Δ70) can bind to DNA as a homodimer in vitro only at a high protein concentration. Based on the SRF crystal structure, the region C-terminal to the MADS domain is involved only in dimerization (Pellegrini et al., 1995). Therefore, the reduced amount of DNA binding by AGL2Δ70 suggests that the active AGL2Δ70 homodimer is less stable, or at a lower level, than those of longer proteins. On the other hand, heterodimers of AGL2Δ70 with the longer proteins AGL2Δ98 or AGL2Δ158 can be readily detected, suggesting that the extra sequences in the longer proteins allow the formation of more stable, or a higher level of, dimers with AGL2Δ70. The SRF crystal structure indicates that the region C-terminal to the MADS domain contains a second α helix that is oriented antiparallel to that of the other subunit (Pellegrini et al., 1995). Even though the AGL2 I domain has little sequence similarity with this region of SRF, the two sequences both have many hydrophilic residues (Figure 8). It is possible that the AGL2 I domain also has a second helix at a position similar to that in SRF; this would mean that AGL2Δ70 contains only the N terminus of the presumptive helix (Figure 8). Therefore, the residues still present in AGL2Δ70 would have

to interact with the C-terminal half of the helix in the other, longer subunit, explaining why AGL2Δ70 formed weaker homodimers than heterodimers with longer AGL2 proteins. It is significant that this dimeric interaction seems to be conserved between AG and AGL2, because AGL2Δ70 can form heterodimers with truncated AG proteins.

MADS Domain Protein Heterodimers Could Potentially Increase the Repertoire of Regulators

Heterodimers between similar proteins are important components of the eukaryotic transcriptional regulatory machinery. Because many plant MADS box genes have overlapping regions of expression, there are opportunities for heterodimer formation. The plant MADS box genes have been grouped into four subfamilies, represented by *AG*, *APETALA1* (*AP1*), *APETALA3* (*AP3*), and *PISTILLATA* (*PI*), with a few additional genes that do not belong to any of these subfamilies (Purugganan et al., 1995). Of the four proteins tested here, AG and AGL1 are in one subfamily and AGL2 and AGL3 are in the AP1 subfamily; members of these two subfamilies constitute the majority of plant MADS domain proteins. We found that these four MADS domain proteins can bind to DNA in vitro as either homodimers or heterodimers. This indicates that heterodimers can form when the subunits are either from the same subfamily or from different subfamilies. The Antirrhinum MADS domain proteins DEF and GLO can bind to DNA as a heterodimer but not as homodimers (Schwarz-Sommer et al., 1992). AP3 and PI, the Arabidopsis orthologs of DEF and GLO, respectively, behave in similar ways (J.L. Riechmann, B.A. Krizek, and E.M. Meyerowitz, personal communication). It is worth noting that DEF (AP3) and GLO (PI) are divergent enough to be placed in separate subfamilies, although they are more similar to each other than to members of other subfamilies (Purugganan et al., 1995).

The ability of many MADS domain proteins to form both homodimers and heterodimers, as well as that of AP3/PI and DEF/GLO to form heterodimers but not homodimers, could be explained at least in part by the characteristics of the I domains of the MADS domain proteins (Figure 8). The I domains of plant MADS domain proteins have many hydrophilic residues, including charged ones, as in SRF (Figure 8). In particular, the I domains of members of AG and AP1 subfamilies have at least four positive and three negative residues; some of these are within the region of each of these proteins corresponding to the SRF coil and the second α helix. The spacing of the charged residues is such that these proteins would be able to form ionic interactions between subunits of either a homodimer or a heterodimer (Figure 8). Because the I domains of the AG and AP1 subfamily members are of similar lengths, they may all be able to form heterodimers, as supported by results here. If the MADS domain proteins behave similarly in vivo, then, depending on the number of genes expressed in the same cells, the potential number of different MADS domain dimers could be significantly more than that of subunit genes expressed, dramatically increasing the complexity of the transcriptional machinery encoded by MADS box genes.

The AP3 and PI I domains are quite different from those of AG and AP1 subfamily members (Figure 8). AP3 has only two positive residues at positions far apart, with five negative residues in between; although the N-terminal two-thirds of the PI I domain do contain three positive and two negative residues, the same-charged residues are clustered. Therefore, ionic interactions can be easily formed between AP3 and PI but not between two of the same protein. It is noteworthy that the N-terminal end of the K domain of both AP3 (DEF) and PI (GLO) has several charged residues (Sommer et al., 1990; Jack et al., 1992; Tröbner et al., 1992; Goto and Meyerowitz, 1994), allowing possible additional interactions. This may explain why a point mutation in *DEF* that deletes a lysine at the N-terminal end of the K domain results in the inability of the mutant DEF protein to bind DNA with GLO (Schwarz-Sommer et al., 1992). The fact that both AP3 and PI have shorter I domains than the other plant MADS domain proteins may explain why AP3 and PI can bind to DNA as a heterodimer with each other but not with AG or AP1 (J.L. Riechmann, B.A. Krizek, and E.M. Meyerowitz, personal communication).

Arabidopsis and other plants have a large number of MADS box genes, many of which are preferentially expressed in flowers, with overlapping expression patterns. Therefore, plants use MADS domain proteins extensively to control development, particularly flower development. The fact that functionally different MADS domain proteins can exist in the same cells raises questions about the mechanisms for functional specificity and about potential interactions between MADS domain proteins. We have shown here that individual MADS domain proteins can bind in vitro to distinct DNA sequences with different affinities, providing a means for regulating distinct sets of target genes by different proteins. In addition, we found that several MADS domain proteins can form heterodimers as well as homodimers, suggesting that potentially a large number of MADS domain regulators could be encoded by a much smaller number of genes. The combination of the binding sequence specificity and heterodimer formation, with modulation by levels of expression, covalent modifications, and interaction with accessory factors, would allow differential regulation by MADS domain proteins of many target genes that are necessary to control complex processes such as flower development.

METHODS

Plasmid Constructs

All of the proteins used here were produced in BL21 (DE3) *Escherichia coli* cells (Studier et al., 1990). To express AGL1 (for AGAMOUS-like) protein, two constructs were made. A construct for a truncated AGL1 was first generated as follows. A polymerase chain reaction (PCR)

product encoding AGL1 from the N terminus to the C-terminal end of the K domain was generated using primers with XhoI sites. The primers are as follows: oMC048 (N terminus), 5′-GACTCGAG**AAGGTGGGAGTAGTCAC**-3′; and oMC049 (C-terminal end of K domain), 5′-CGCTCGAG**TTTCCCTCTTCTGCATA**-3′ (XhoI sites are underlined; the *AGL1* sequence is in boldface). The PCR product was digested with XhoI and ligated with XhoI-cut pGEM7Zf(+) (from Promega) to yield pMC772; the AGL1 insert was verified by sequence analysis. To allow in-frame fusion of AGL1 with an epitope tag present in the expression vector, pMC772 was modified by filling in the XhoI site near the AGL1 N terminus, producing pMC776; the *AGL1* cDNA was then released from pMC776 by cutting at flanking XbaI and BamHI sites in the pGEM7Zf(+) polylinker and ligated to the XbaI- and BamHI-digested expression vector pMC804, which was modified from pETN-Npou-mc (provided by M. Tanaka, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY) by filling in an upstream XbaI site. The resulting AGL1 expression construct (pMC805) carried an epitope tag at the N terminus of AGL1, with the following fusion junction: TCTAGACTCGATCGA**GAAGGTGGGAGT** (*AGL1* sequence in boldface). To obtain an expression construct for full-length AGL1, the C-terminal half of the *AGL1* cDNA was excised from pMC935 (Y. Mizukami and H. Ma, unpublished data) with ScaI (internal) and XhoI (in the polylinker downstream of *AGL1* cDNA) and ligated with ScaI- and XhoI-cut pMC805, producing pMC812, which restores the AGL1 reading frame. The full-length AGL1 protein was used for all DNA binding experiments, with the exception of the heterodimer experiments, in which the truncated form was used.

To generate an expression construct of the full-length AGL2 protein, we started with the plasmid pMC414 (Y. Mizukami and H. Ma, unpublished data), which had an artificial NcoI site at the AGL2 initiation ATG codon. The *AGL2* sequence was released from pMC414 with digests by NcoI and BamHI, which cut in the downstream polylinker, and ligated with the NcoI- and BamHI-digested vector pET9d (Studier et al., 1990) to yield pMC811.

Several constructs were generated to express truncated AGL2 proteins (see Figure 5 for their designations). For AGL2Δ158, we used a previously generated PCR product from the beginning of the AGL2 MADS domain to the end of the K domain. The primers used in the PCR are as follows: oMC050, 5′-GACTCGAG**GAAGAGGAAGAGTAGAG**-3′; and oMC051, 5′-CGCTCGAG**GCTCTTTATTTTGAAGATC**-3′ (the XhoI sites are underlined; the *AGL1* sequence is in boldface). This PCR fragment was inserted into the XhoI site in pGEM7Zf(+), resulting in pMC447, which has an XbaI site just upstream of the XhoI site and *AGL2* sequence. The initiation codon with an NcoI site was introduced into the construct by replacing an XbaI-BamHI (internal) fragment of pMC447 with the XbaI-BamHI fragment from pMC414 (see the previous paragraph). The resulting plasmid was (1) digested with XhoI, (2) treated with mung bean nuclease to create blunt ends, and (3) further digested with NcoI. The resulting *AGL2* cDNA was ligated with pET9d that had been (1) digested with NheI, (2) blunt-ended with the Klenow fragment of DNA polymerase I, and (3) digested with NcoI. The final construct (pMC1254) has a C-terminal junction sequence of **AAAGAG**:CCCTAG (*AGL2* in boldface): the AGL2 protein expressed from pMC1254 begins at the initiation ATG codon and ends after the last residue (glutamate) of the K domain with an additional proline. For the other AGL2 constructs, the *AGL2* cDNA was released from pMC414 by digesting with an enzyme (see below) that cuts within the AGL2 coding region, filling in with Klenow fragment or treating with mung bean nuclease (for AGL2Δ70), and digesting with NcoI; the *AGL2* cDNA was inserted into pET9d as was done for pMC1254. The restriction enzymes used for the deletions, the *AGL2* sequences near the corresponding sites (underlined), and the final junction sequences (*AGL2* in boldface) are as follows: for AGL2Δ98 (pMC1253), HindIII (CTGAAGCTT) and **CTGAAGCT**CTAG; for AGL2Δ77 (pMC1252), BamHI (TATGGATCC) and **TATGGA**CTAGCATGA; and for AGL2Δ70 (pMC1251), KpnI (CTTGATCGGTACC) and **CTTGATCG**AGCATGA.

For comparison of relative DNA binding affinity with AGL2, a previously described AG protein was used (Huang et al., 1993). For heterodimer experiments, truncated AG proteins, as briefly described below, and the full-length AGL3 protein (pMC810, which is pET9d with the *AGL3* coding region; Huang et al., 1995) were used. For the interaction between AG and AGL2, truncated AG proteins 34M-I and 34M-I-K from residue 34 in the N-terminal region to the ends of the I and K domains, respectively (pMC1003 and pMC1004; Y. Mizukami, H. Huang, M. Tudor, and H. Ma, unpublished data) were used; for interactions with the other proteins, a truncated AG protein (M-I) with only the MADS domain and the I region (pMC1037; Y. Mizukami, H. Huang, M. Tudor, and H. Ma, unpublished data) was used.

Protein Extraction, DNA Binding, and Other Analyses

E. coli BL21 (DE3) cells carrying either vector or one of the expression constructs were grown to mid-to-late exponential phase (OD_{600} from 0.5 to 1.0) and induced for 3 hr by the addition of isopropyl β-D-thiogalactoside to a final concentration of 0.5 mM. Cells were harvested and resuspended in one-tenth volume of buffer B (10 mM Tris-HCl, pH 7.8, 50 mM NaCl, 1 mM EDTA, 1 mM phenylmethylsulfonyl fluoride, and 20% glycerol; Huang et al., 1993, 1995), aliquoted, and centrifuged; the cell pellets were stored at −70°C. Cell lysates were prepared by lysing cells in buffer A (same as buffer B except that 20% glycerol was replaced with 6 M urea), followed by centrifugation to remove cell debris; the lysates were then diluted with buffer B as described previously (Huang et al., 1995). For heterodimer experiments, the cells from two strains, each for a different protein, were mixed first; then the extracts were obtained by lysis in buffer A, centrifugation, and diluting with buffer B. The diluted cell lysates were used in DNA binding experiments (Huang et al., 1995). Gel shift experiments, binding sequence selections, and DNase I protection footprinting experiments were performed with radiolabeled probes as described previously (Huang et al., 1993, 1995).

ACKNOWLEDGMENTS

We thank Yukiko Mizukami for the *AGL1* and *AGL2* full-length cDNAs; Masafumi Tanaka for the gift of the pETN-Npou-mc vector; Jose Luis Riechmann, Beth A. Krizek, and Elliot M. Meyerowitz for sharing results before publication; and Catherine A. Flanagan, Erich Grotewold, and Chong Huang for helpful comments. T.S. was a participant of the Cold Spring Harbor Laboratory Undergraduate Research Program. This work is supported by grants from the U.S. Department of Agriculture (No. 92-01551 and No. 94-00655) and a Junior Faculty Research Award (JFRA-520) from the American Cancer Society to H.M. and by the Cold Spring Harbor Laboratory Robertson Fund.

Received September 7, 1995; accepted November 9, 1995.

REFERENCES

Biggins, M.D., and Tjian, R. (1989). Transcription factors and the control of *Drosophila* development. Trends Genet. **5**, 377–383.

Bowman, J.L., Smyth, D.R., and Meyerowitz, E.M. (1989). Genes directing flower development in *Arabidopsis*. Plant Cell **1,** 37–52.

Bowman, J.L., Drews, G.N., and Meyerowitz, E.M. (1991). Expression of the Arabidopsis floral homeotic gene *AGAMOUS* is restricted to specific cell types late in flower development. Plant Cell **3,** 749–758.

Bradley, D., Carpenter, R., Sommer, H., Hartley, N., and Coen, E.S. (1993). Complementary floral homeotic phenotypes result from opposite orientations of a transposon at the *plena* locus of Antirrhinum. Cell **72,** 85–95.

Drews, G.N., Bowman, J.L., and Meyerowitz, E.M. (1991). Negative regulation of the Arabidopsis homeotic gene *AGAMOUS* by the *APETALA2* product. Cell **65,** 991–1002.

Flanagan, C.A., and Ma, H. (1994). Spatially and temporally regulated expression of the MADS-box gene *AGL2* in wild-type and mutant *Arabidopsis* flowers. Plant Mol. Biol. **26,** 581–595.

Gehring, W.J. (1987). Homeo boxes in the study of development. Science **236,** 1245–1252.

Goto, K., and Meyerowitz, E. (1994). Function and regulation of the *Arabidopsis* floral homeotic gene *PISTILLATA*. Genes Dev. **8,** 1548–1560.

Gustafson, T.A., Taylor, A., and Kedes, L. (1989). DNA bending is induced by a transcriptional factor that interacts with the human c-*fos* and α-actin promoters. Proc. Natl. Acad. Sci. USA **86,** 2162–2166.

Haran, T.E., Kahn, J.D., and Crothers, D.M. (1994). Sequence elements responsible for DNA curvature. J. Mol. Biol. **244,** 135–143.

Huang, H., Mizukami, Y., Hu, Y., and Ma, H. (1993). Isolation and characterization of the binding sequences for the product of the *Arabidopsis* floral homeotic gene *AGAMOUS*. Nucleic Acids Res. **21,** 4769–4776.

Huang, H., Tudor, M., Weiss, C.A., Hu, Y., and Ma, H. (1995). The *Arabidopsis* MADS-box gene *AGL3* is widely expressed and encodes a sequence-specific DNA-binding protein. Plant Mol. Biol. **28,** 549–567.

Huijser, P., Klein, J., Lönnig, W.-E., Meijer, H., Saedler, H., and Sommer, H. (1992). Bracteomania, an inflorescence anomaly, is caused by the loss of function of the MADS-box gene *squamosa* in *Antirrhinum majus*. EMBO J. **11,** 1239–1249.

Irish, V.F., and Sussex, I.M. (1990). Function of the *apetala-1* gene during *Arabidopsis* floral development. Plant Cell **2,** 741–753.

Jack, T., Brockman, L.L., and Meyerowitz, E.M. (1992). The homeotic gene *APETALA3* of Arabidopsis thaliana encodes a MADS box and is expressed in petals and stamens. Cell **68,** 683–697.

Kempin, S.A., Savidge, B., and Yanofsky, M.F. (1995). Molecular basis of the *cauliflower* phenotype in *Arabidopsis*. Science **267,** 522–525.

Ma, H. (1994). The unfolding drama of flower development: Recent results from genetic and molecular analyses. Genes Dev. **8,** 745–756.

Ma, H., Yanofsky, M.F., and Meyerowitz, E.M. (1991). *AGL1–AGL6,* an *Arabidopsis* gene family with similarity to floral homeotic and transcription factor genes. Genes Dev. **5,** 484–495.

Mandel, M.A., Gustafson-Brown, C., Savidge, B., and Yanofsky, M.F. (1992). Molecular characterization of the *Arabidopsis* floral homeotic gene *APETALA1*. Nature **360,** 273–277.

Mizukami, Y., and Ma, H. (1995). Separation of *AG* function in floral meristem determinacy from that in reproductive organ identity by expressing antisense *AG* RNA. Plant Mol. Biol. **28,** 767–784.

Mueller, C.G., and Nordheim, A. (1991). A protein domain conserved between yeast MCM1 and human SRF directs ternary complex formation. EMBO J. **10,** 4219–4229.

Norman, C., Runswick, M., Pollock, R., and Treisman, R. (1988). Isolation and properties of cDNA clones encoding SRF, a transcription factor that binds to the c-*fos* serum response element. Cell **55,** 989–1003.

Nurrish, S.J., and Treisman, R. (1995). DNA binding specificity determinants in MADS-box transcription factors. Mol. Cell. Biol. **15,** 4076–4085.

Passmore, S., Maine, G.T., Elble, R., Christ, C., and Tye, B.K. (1988). A *Saccharomyces cerevisiae* protein involved in plasmid maintenance is necessary for mating of MATα cells. J. Mol. Biol. **204,** 593–606.

Passmore, S., Elble, R., and Tye, B.K. (1989). A protein involved in minichromosome maintenance in yeast binds a transcriptional enhancer conserved in eukaryotes. Genes Dev. **3,** 921–935.

Pellegrini, L., Tan, S., and Richmond, T.J. (1995). Structure of serum response factor core bound to DNA. Nature **376,** 490–498.

Pollock, R., and Treisman, R. (1990). A sensitive method for the determination of protein–DNA binding specificities. Nucleic Acids Res. **18,** 6197–6204.

Pollock, R., and Treisman, R. (1991). Human SRF-related proteins: DNA-binding properties and potential regulatory targets. Genes Dev. **5,** 2327–2341.

Purugganan, M.D., Rounsley, S.J., Schmidt, R.J., and Yanofsky, M.F. (1995). Molecular evolution of flower development: Diversification of the plant MADS-box regulatory gene family. Genetics **140,** 345–356.

Savidge, B., Rounsley, S.D., and Yanofsky, M.F. (1995). Temporal relationships between the transcription of two Arabidopsis MADS-box genes and the floral organ identity genes. Plant Cell **7,** 721–733.

Schwarz-Sommer, Z., Huijser, P., Nacken, W., Saedler, H., and Sommer, H. (1990). Genetic control of flower development: Homeotic genes in *Antirrhinum majus*. Science **250,** 931–936.

Schwarz-Sommer, Z., Hue, I., Huijser, P., Flor, P.J., Hansen, R., Tetens, F., Lönnig, W.-E., Saedler, H., and Sommer, H. (1992). Characterization of the *Antirrhinum* homeotic MADS-box gene *deficiens*: Evidence for DNA binding and autoregulation of its persistent expression throughout flower development. EMBO J. **11,** 251–263.

Scott, M.P., Tamkun, J.W., and Hartzell, G.W.I. (1989). The structure and function of the homeodomain. Biochim. Biophys. Acta **989,** 25–48.

Shiraishi, H., Okada, K., and Shimura, Y. (1993). Nucleotide sequences recognized by the *AGAMOUS* MADS domain of *Arabidopsis thaliana in vitro*. Plant J. **4,** 385–398.

Smyth, D.R., Bowman, J.L., and Meyerowitz, E.M. (1990). Early flower development in *Arabidopsis*. Plant Cell **2,** 755–767.

Sommer, H., Beltran, J.-P., Huijser, P., Pape, H., Lonnig, W.-E., Saedler, H., and Schwarz-Sommer, Z. (1990). *Deficiens*, a homeotic gene involved in the control of flower morphogenesis in *Antirrhinum majus*: The protein shows homology to transcription factors. EMBO J. **9,** 605–613.

Struhl, K. (1989). Helix-turn-helix, zinc-finger, and leucine-zipper motifs for eukaryotic transcriptional regulatory proteins. Trends Biol. Sci. **14,** 137–140.

Studier, W., Rosenberg, A.H., Dunn, J.J., and Dubendorff, J.W. (1990). Use of T7 RNA polymerase to direct expression of cloned genes. Methods Enzymol. **185,** 60–89.

Tröbner, W., Ramirez, L., Motte, P., Hue, I., Huijser, P., Lönnig, W.-E., Saedler, H., Sommer, H., and Schwarz-Sommer, Z. (1992). *GLOBOSA*: A homeotic gene which interacts with *DEFICIENS* in the control of *Antirrhinum* floral organogenesis. EMBO J. **11,** 4693–4704.

Wynne, J., and Treisman, R. (1992). SRF and MCM1 have related but distinct DNA binding specificities. Nucleic Acids Res. **20,** 3297–3303.

Yanofsky, M.F. (1995). Floral meristems to floral organs: Genes controlling early events in *Arabidopsis* flower development. Annu. Rev. Plant Physiol. Plant Mol. Biol. **46,** 167–188.

Yanofsky, M.F., Ma, H., Bowman, J.L., Drews, G.N., Feldmann, K.A., and Meyerowitz, E.M. (1990). The protein encoded by the *Arabidopsis* homeotic gene *agamous* resembles transcription factors. Nature **346,** 35–39.