© 1993 Oxford University Press

Nucleic Acids Research, 1993, Vol. 21, No. 20 4769–4776

Isolation and characterization of the binding sequences for the product of the *Arabidopsis* floral homeotic gene *AGAMOUS*

Hai Huang, Yukiko Mizukami, Yi Hu and Hong Ma*
Cold Spring Harbor Laboratory, Cold Spring Harbor, NY 11724-2212, USA

Received July 2, 1993; Revised and Accepted August 30, 1993

ABSTRACT

The *Arabidopsis* floral homeotic gene *AGAMOUS* (*AG*) is required for normal flower development. The deduced AG protein contains a region which shares substantial sequence similarity with the DNA-binding domains of known transcription factors, SRF (human) and MCM1 (yeast). Therefore, it is likely that AG is also a DNA-binding protein regulating transcription of floral genes. We describe here several experiments to characterize AG-DNA binding *in vitro*. We show that AG indeed binds a DNA sequence matching the consensus of SRF targets. Further, we have selected the AG-binding sequences from a pool of random oligonucleotides, and deduced an AG-binding consensus sequence of TT(A/T)CC(A/T)(A/t)$_2$(T/A)NNGG(-G)(A/t)$_2$. We have demonstrated that AG binds to the consensus region of three of the oligonucleotides by footprinting analysis. Finally, we have examined AG's relative binding affinity for different sequences, as compared to SRF, by gel mobility shift analysis. Our results indicate that AG is a sequence-specific DNA-binding protein, and that the AG-binding consensus sequence is similar to those of MCM1 and SRF.

INTRODUCTION

Genetic studies in *Arabidopsis thaliana* have identified mutations affecting flower development. Several genes controlling floral organ identity have been extensively studied genetically and at the molecular level (1, 2). These homeotic genes include *APETALA2*, *APETALA3*, *PISTILLATA* and *AGAMOUS* (3–8). Mutations in the *AGAMOUS* (*AG*) gene cause homeotic conversion of stamens to petals, and replace the gyneocium (of fused carpels) with a secondary *ag* flower in which third whorl organs are again petals and a tertiary *ag* flower occupies the center (4, 6, 7). Therefore, normal *AG* function is required for the proper morphogenesis of stamens and the gyneocium.

The *AG* DNA sequence (6) indicates that a region near the N-terminus of the deduced AG protein shares striking sequence similarity with the human serum response factor (SRF, ref.9), and a regulator of yeast cell-type specific genes, MCM1 (10). Another yeast protein, ARG80, also has the same type of conserved sequence motif (11). Furthermore, the *Antirrhinum majus* floral homeotic gene *DEFICIENS* (*DEF A*, a homolog of *APETALA3*) also encodes a protein with a similarly conserved motif (12). The conserved motif found in all these gene products has been designated the MADS-box (for MCM1, AG and ARG80, DEF A, and SRF; ref.13). Additional floral MADS-box genes have been isolated from *Arabidopsis* (8, 14, 15) and *Antirrhinum* (16–18), suggesting that a family of MADS-box genes encode transcriptional regulators during flower development.

Both SRF and MCM1 have been studied extensively *in vivo* and *in vitro* (9, 19–21). SRF binds *in vitro* to the SRE (serum response element) in the promoter of the c-*fos* proto-oncogene, and mediates a transient transcriptional activation by growth factors in cultured cells (22, 23). In addition to the *c-fos* gene, the SRF target sites (SREs) are found in actin genes (24). MCM1 is essential for regulating cell-type specific genes (25). In **a** cells, MCM1 activates **a**-specific genes, whereas in α cells, MCM1 cooperates with MATα1 to activate α-specific genes, and with MATα2 to repress **a**-specific genes. Natural MCM1 target sites are found in the upstream regulatory regions of both **a**-specific and α-specific genes (20, 26, 27). The SRF and MCM1 target sequences have been shown to be very similar, and contain a characteristic consensus CC(A/T)$_6$GG (the CArG box, refs.20, 27, 28). MCM1 can bind SREs (20, 27), and SRF can bind to many of the MCM1 target sequences as well. Further analysis using pools of random oligonucleotides indicates that both SRF and the MCM1 proteins show preferences *in vitro* for more specific sequences. The SRF-binding sequence consensus is ATGCCC**(A/t)TATA(T/a)GG**(T/A)NNT (ref.28; see Table II), while the consensus of MCM1 binding sequence is (T/a)T-(-C)**CC(T/C)AA(T/A)NNGG**TAA (ref.29; see Table II). The region in bold face corresponds to the CArG box, which we will refer to as the core region.

To understand the function of *AG* at the molecular level, it is important to address the following questions: is AG a DNA-binding protein, and what is the binding specificity of the AG protein? We have performed a series of experiments to answer these questions. First, our results indicate that AG is indeed a

* To whom correspondence should be addressed

DNA-binding protein, and that it binds CArG-box containing sequences. We have also selected a large number of oligonucleotides with a wide spectrum of different sequences to which AG binds *in vitro*. Furthermore, we show that AG and SRF have different relative binding affinities for several sequences. Our results indicate that AG is a sequence-specific DNA-binding protein, and provides strong evidence that AG is most likely a transcription factor controlling *Arabidopsis* flower development.

MATERIALS AND METHODS

Plasmid construction and expression of the AG protein

For the *in vitro* studies the AG protein was produced in *E.coli* cells by placing the AG cDNA under the control of a bacteriophage T7 promoter (30). Because the longest *AG* cDNA clone lacks an ATG initiation codon (6), we used an *AG* cDNA, ATG-34 (31), with an engineered ATG codon at a position matching the translational start site of the *AG* homologs from *Brassica napus* (*BAG1*, ref.32), tomato and tobacco (M. Yanofsky, pers. comm.). ATG-34 lacks the first 33 amino acid residues found in the AG protein predicted from the longest *AG* cDNA clone, and probably lacks a few additional residues potentially present in the native AG protein. ATG-34 was inserted into the *Nco* I–*Bam* HI sites of the pET-9d vector (30). Cells of the *E.coli* strain BL21 were transformed with the *AG* construct and the expression of the AG protein was induced as previously described (30). Crude *E.coli* lysate was shown to contain the AG protein near the expected size, as well as some degradation products, by western blot analysis (data not shown) using an antiserum recognizing the AG protein (C.Flanagan and H.Ma, unpublished work). Fractionation studies indicate that most of the AG protein was in the insoluble protein fraction. For DNA-binding studies, a 50-ml culture of the cells carrying the *AG* construct was induced. After the cells were harvested, they were resuspended in 1.0 ml TE and lysed by sonication on ice (at scale 4 on the VirSonic 300 from Virtis Co. Inc.), 3 times of 10 second each, with 30 second incubations on ice in between. The lysate was diluted to 20 ml with ice-chilled 2.5 M NaCl in TE, and the AG-containing fraction were then pelleted by centrifugation. The pellet was washed three times with ice-chilled TE, and then solubilized in 1.5 ml of buffer A (buffer A contains 10 mM Tris–HCl, pH 7.8, 50 mM NaCl, 1 mM EDTA, 6M urea and 1 mM PMSF). The cell debris was removed by centrifugation in a Sorvall SA-600 rotor at 12,000 rpm for 30 minutes at 4°C. The protein concentration of the AG-containing supernatant was determined by the Bio-Rad protein assay using BSA as a standard, and the partially purified AG protein was stored as 50 μl aliquots at −70°C until use.

In vitro DNA-binding of AG and SRF

Gel-mobility shift assays were done essentially according to Garcia *et al.* (33). For AG binding, the urea-solubilized partially purified AG protein was diluted 5-fold with buffer B (buffer B is the same as buffer A except that it contains 20% glycerol and no urea). Protein–DNA incubations were performed in binding reaction buffer (10 mM Tris–HCl, pH7.5, 50 mM NaCl, 1 mM DTT, 1 mM EDTA, 5% glycerol, 50 μg/ml poly(dI-dC)·poly(dI-dC), 100 μg/ml BSA) with about 1×10^4 cpm of radioactive probe (labeled by polymerase chain reaction [PCR] with end-labeled primer pMC069, see below) and 0.5 to 3 μg of the partially purified AG protein. The reaction mix was incubated at room temperature for 15 minutes, and free- and bound-DNAs were separated on a 5% polyacrylamide gel (80:1 acrylamide:bis) in 0.25×TBE. After the gel was dried onto 3M blotting filter paper, the radiolabeled DNAs were visualized by autoradiography. For SRF binding to AG-binding sequences, *E.coli* lysate containing the SRF protein (T7 expression system; gift of Dr D.Grueneberg) was used to bind several oligonucleotide probes using both the above AG binding conditions and standard SRF binding conditions (same as the AG binding conditions except 75 mM NaCl, 20 μg/ml poly(dI-dC)·poly(dI-dC), 3 mM spermidine, no EDTA, no BSA and no urea, D. Grueneberg pers. comm.). Two sequences (oligo A and oligo B) were used for initial DNA binding analysis with AG. These sequences (Fig. 1B) were chemically synthesized and cloned into the *Hin*dIII site of pGEM7Zf(+) (Promega). Primers (pMC068 and pMC069, see below) matching polylinker sequences flanking the *Hin*dIII site were used to generate probes by PCR.

Selection of AG binding sequences from random oligonucleotides

A pool of single-stranded random oligonucleotides (oMC070) with constant sequences of 25 bases [derived from the polylinker region of pGEM7Zf(+) which lacks any AG binding sequences (see Results)] at both ends and a variable sequence of 26 bases in the center was chemically synthesized: 5′-ACTCGAG-GAATTCGGTACCCCGGGT-N_{26}TGGATCCGGAGAGC-TCCCAACGCGT-3′. The underlined *Sma* I and *Bam* HI sites were used later to clone the selected AG-binding sites into pGEM7Zf(+). Two additional oligonucleotides were synthesized to serve as primers for PCR: oMC068, 5′-ACTCGAGGAATT-CGGTACCCCGGGT-3′ (the same as the left end of oMC070) and oMC069, 5′-ACGCGTTGGGAGCTCTCCGGATCCA-3′ (complement of the right end of oMC070). Primer oMC069 was used to convert the single-stranded oMC070 into double-stranded DNA using *E.coli* DNA polymerase I large fragment (Klenow). When a radioactive probe was required, primer oMC069 was end-labeled with γ-^{32}P ATP and T4 polynucleotide kinase. *In vitro* selections of AG binding sequences were performed essentially according to a procedure (E.Grotewold and T.Peterson, pers. comm.) which is similar in principle to a previously described method (28). Instead of using immunoprecipitation to purify the protein–DNA complexes (28), in the Grotewold and Peterson procedure, the protein–DNA complexes are purified by running the binding reaction on a polyacrylamide gel and excising the region of the gel corresponding to the protein–DNA complexes for each round starting with the second round. We have made two additional changes for use with the AG protein: 1) the DNA binding reaction was done as described in the previous section; and 2) the purification of the AG-DNA complex in the first round was also performed using a gel-mobility shift assay with the position of the complexes (not detectable) estimated from that of the AG-oligo A complexes (similar in size) on the same gel. The DNA was eluted from the gel piece by incubating in 1 ml TE at 4°C overnight, extracted with phenol/chloroform and chloroform, precipitated with ethanol, and resuspended in 20 μl H_2O. Five microliters of the DNA solution were used as a template for PCR (annealing temperature of 60°C) with radiolabeled oMC068 and oMC069 as primers. After 20 cycles of amplification, the amplified DNA was purified using PAGE (8% 40:1 acrylamide: bis, 1×TBE, visualized by autoradiography) followed by elution from the gel piece as described above. The purified radiolabeled DNAs were used for the next round of binding reaction and

complex purification. After five rounds of AG-binding reaction and gel purification, the selected AG-bound DNA was amplified with PCR for 40 cycles, digested with *Eco* RI and *Bam* HI restriction enzymes and cloned into the *Eco* RI and *Bam* HI sites of pGEM7Zf(+), restoring the flanking non-random sequences. Sequence analysis of double-stranded DNA was performed using the Sequenase kit (US Biochemicals) according to the protocol provided by the supplier.

Footprinting analysis

The probes for footprinting analysis were prepared according to an unpublished protocol (S. Bell, pers. comm.). Briefly, to generate ends for labeling, plasmids which carry AG-binding sequences were digested with *Sph* I (for top strand) or *Nsi* I (for bottom strand), which cut near the ends of the pGEM7Zf(+) polylinker. Following the digestion, one unit of calf intestinal phosphatase (CIP, from Boehringer Mannheim) was added. The reaction mix was incubated at 37°C for 30 minutes and then heated at 68°C for 20 minutes. After two phenol/chloroform extractions, the DNA was precipitated with ethanol. One microgram of CIP-treated DNA was rephosphorylated in a 20 μl reaction mixture containing 2 μl 10×kinase buffer (NEB), 9.6 μCi of γ-^{32}p ATP (ICN), and 15 units of T4 polynucleotide kinase (NEB), and incubated for 30 minutes on ice, followed by inactivation of the kinase at 68°C for 15 minutes. The labeled DNA was digested with *Nsi* I (for top strand) or *Sph* I (for bottom strand) to allow the removal of the vector sequences (also labeled), and the probes were separated from unincorporated γ-^{32}P ATP using an 8% acrylamide (40:1 acrylamide to bis-acrylamide) non-denaturing gel in TBE buffer. The labeled probes were excised from gel, eluted by incubating in 1 ml TE at 4°C overnight, and precipitated with ethanol.

For the footprinting analysis, a binding reaction essentially as described for the gel-mobility shift assay was performed by incubating labeled probes (about 0.01 pmol) with or without AG protein for 15 minutes at room temperature in a total volume of 100 μl. Subsequently, 2 μl of 0.1 M $MgCl_2$ and 500 ng of DNase I were added to the DNA−protein mixture, and incubated at room temperature for 8 minutes. The DNase I treatment was terminated by adding an equal volume of stop solution containing 20 mM EDTA, 200 mM NaCl, 1% SDS and 0.2 mg/ml of glycogen (Boehringer Mannheim). The samples were extracted twice with phenol/chloroform and precipitated with ethanol. The resulting DNA fragments were dissolved in the sequencing stop solution and separated on an 8% polyacrylamide-8 M urea sequencing gel in 1×TBE buffer. The A>C sequencing reaction of the same DNA fragments was performed by the Maxam-Gilbert sequencing method (34).

RESULTS

AG binds a SRF consensus sequence

To study AG DNA-binding *in vitro*, we expressed the AG protein in *E.coli* cells using a construct (Fig. 1A) based on the T7 expression system (see Materials and Methods for a description of the construct). It is known that SRF and MCM1 have overlapping DNA binding sequence specificities; therefore, AG might also bind to a sequence to which both SRF and MCM1 bind. As a first step towards understanding the DNA binding activity of the AG protein, we tested the ability of the *E.coli* expressed AG protein to bind a synthetic oligonucleotide (oligo A; Fig. 1B) containing a SRF consensus CArG box (underlined

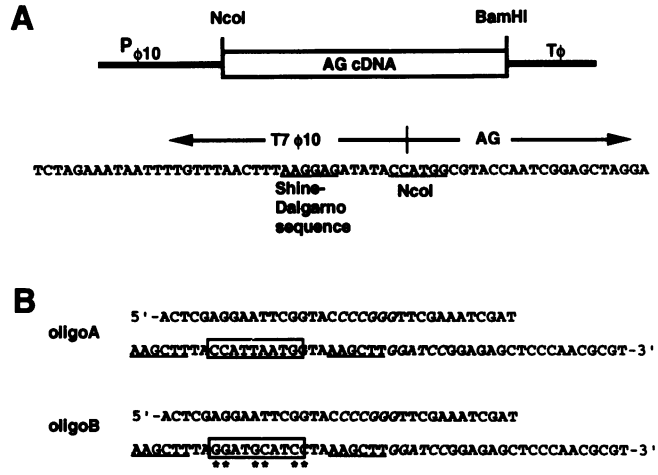

Figure 1. *AG* expression construct and synthetic oligonucleotides for AG−DNA binding studies. **A.** Schematic representation of the bacteriophage T7 gene 10 promoter and translational leader sequence ($P_{\phi 10}$), fused to the *AG* cDNA, with the transcriptional terminator sequence (Tφ) downstream of the *AG* cDNA. The sequence of the fusion junction is shown below the construct, with the Shine−Dalgarno sequence and the fusion *Nco* I site underlined. **B.** The sequences of oligo A, which contains an SRE CArG box (underlined) and oligo B in which 6 bases (with asterisks below) of the CArG box were altered. Each oligo sequence is shown in two lines, and only the upper strands of the double-stranded DNAs are shown. The *Hin*dIII sites (underlined) mark the junctions between pGEM7Zf(+) polylinker and the synthetic oligonucleotides, oligo A and oligo B. Later, the oligonucleotides selected from the random pool (sequences shown in Table I) were cloned into the *Sma*I and *Bam*HI sites (italicized in the oligo A sequence shown here) of the pGEM7Zf(+) vector (see Materials and Methods).

in Fig. 1B). As shown in Fig. 2, the oligo A probe was shifted when an *E.coli* lysate containing the AG protein (Lane 6), but not a lysate from the same host strain with only the vector (Lane 4), was used in the binding reaction. This AG-dependent mobility shift requires the CArG box present in oligo A, since a mutant oligonucleotide (oligo B, Fig. 1B) containing six nucleotide changes (those with asterisks below, Fig. 1B) cannot be shifted by the *E.coli* lysate containing AG (Fig. 2, Lane 5). Furthermore, the AG-oligo A binding can be competed away by non-radioactive oligo A (Fig. 2, Lanes 7 and 8), but not by the mutant oligo B (Fig. 2, Lane 9). These results indicate that AG is a sequence-specific DNA-binding protein.

Determination of an AG consensus sequence *in vitro*

Because no natural target gene for AG has been reported, we sought to determine the spectrum of sequences to which AG binds by selecting binding sequences from a pool of random oligonucleotides, using a modified procedure of a previously described *in vitro* method (see Materials and Methods). Briefly, AG-binding sequences were selected from random oligonucleotides by using multiple rounds of a binding reaction followed by PCR amplification of the DNA bound to AG. Fig. 3 shows the enrichment of AG binding DNA sequences after each round of selection and amplification. The additional faint bands are probably due to complexes with *E.coli* proteins (band I) or AG degradation fragments (band II). After five rounds, the oligonucleotides were cloned, and their sequences determined. As shown in Table I, the selected sequences contain regions resembling the CArG box. A random subset of these sequences (those with a $ sign next to the sequence number) was individually tested for binding, and all of them bind AG (data not shown).

Table I. Sequences of AG binding nucleotides selected from random oligodeoxynucleotides

Site Clone	Sequence
CC(A/T)$_6$GG	
108*	TCCGGATCCAAAAATGGCAATATAATCGATTCCCATAC
4*	TCCGGATCCAAAATAGGCAAAGTTTGTAGCGTTTAAAC
128	GGTATTACCAAATTAGGTAATCCCCATGATG
133	GTATTTACCAAATATGGTAAGAGGAGCTTG
106	GTAATCCTAACCAAATATGGAAAGAATCTG
112	GTAATAATTTACCAAATATGGTAACACCTG
65*	CATTTGTTTCCAATTAAGGCTGTCTTTCAC
116	GTGTCTAACCAATTATGGTAACTTTACGTG
62*$	CAGTGACCTTTCCATATTAGGCTGCCCCAC
85$#	GTGGGCAATACATTCCATATTTGGCAGGTGG
73	GTTCTACCGCGTTTTTTCCATTTTTGGATGGATC
118	GGGTTTACCTAAAAAGGAAAGAGGTGGTTCTG
10.1*	CATATATTAAAGATTACCTAAAATGGCAACCCG
38$	GTTTTAGCCTAAATTGGCAATTTGGTAATG
12*	CAAGTACGATTTCCTAATTTGGCAGTGAAC
52.1	CGGGTTACCTAATTAGGCAAAGTATTTGCCTTG
94*	CACACAACTTTCCTAATTAGGCGTTGCAC
95$	CGGGTTGCCTAATTAGGGAATAATTCCAAGCTG
58$	GTTCTTCAGCTTTCCTTAATAGGTATAGTG
72$	GTATGGAATTACCTTAAATGGATCTCACTG
CC(A/T)$_4$NNGG	
48*$	TCCGGATCCAAAAACGGTAATAAATTGTGGCTGTAAAC
56$	GGTCCTTCCAAAAGAGGAAACCTGGCTTATG
129*	CATAATGAACTTGCCAAAAGCGGTAAGTAC
18	GGTAGTTCCAAAACGGGAATATTGAGCTATG
119*	TCCGGATCCAAAACGGGTAACATTCTGCGGGTGCGCAC
49.2*	CATTAAAACTTTCCAAATAGGGTAAGATAC
54*$	CAGAACTATTAGCCAAATGGGGTAAAGCAC
96*	ATCCATGCCAAATGGGGTAAGATTGTTTACCAC
44*$	CATATTTATTACCATAATCGGCAATTAGAC
13*	CAAAACATACCATAACGGGTAAACACGTAC
60*$	CCAGTAACCATATTGGGTAAATTTACCAGAC
120$	GTTAATCATACCATATCAGGCATATGCAGT
51*	CATTAGAGTTTCCATATCAGGCATTTAAAC
89*$	CAAACTCAAGCCATTTTGGGAAATATAGAC
53.1	GTAATTTCCATTTCAGGCAATAGTAATTTG
50*$	CACATTATTTCCTAAACCGGCAGTGCTAC
15	GTCTGTACCTAAAGCGGTAAGTTATATATG
100*$	CAGAATACTTTCCTAATACGGTAACTTAAC
64*$	CAAGTTTGCCTAATTCGGGTATATACTAAC
97*	ATCCATGCCTAATTCGGATACTTACATCTCAAC
122$	GTGTCGATTACCTAATTCGGTTATTTATGCTG
16	GTGAGCGGTACTTACCTAATTGGGATGTTGG
135*	CAATTGCCGGTATTGCCTAATTGGGCACGACC
124	GTACTTTCCTAATCGGGCTCAGTAGACCTG
52.2*	CAGCTTCTTACCTAATCTGGATTGATTGAC
42$	GTTTCCCCTTACCTAATCGGGCAAATCCTG
130*	TCCATTACCTAATCGGGTAATAGGGGACTCAC
121	GTGCTAATTACCTTATGAGGTAATACGGTG
CC(A/T)$_4$NN(A/T)G or CC(A/T)$_4$NNG(A/T)	
49.1*	ATCCATTCCAAAAAGAGCAAGAATTAGATTGAC
125*	CATATGATCTTTCCATAAATAGAAATAAC
71*	CCAGTTACCATAATTAGTATGCTATGAATAC
9*	TCCACTACCTAAAATAGAAAAACTCCTTAAAC
11	CCGGGTACCTAATAAAGGAAAGAATGTGAGACTG
111$#	GTAAACCGGGGACATTACCTATTTTAGGTGGATC
10*	CAATCGTCTCCACTTTCCTTAAGTAGGGACCCG
134*	CACACTTGCCTATTATTGGGTGATTATGAC
91*	CACGTTCACTACCTAATCGTGTTATAGCAC
67$	GGGTATACCAAAAGGGAAAGAATAGACTATTG
109*	CAATTGGCGCCTTTCCCAATTTAGAATGACCC
99*	CAAGAAGTTTACCCAATAAAGTGAAATTAC
55	GTCACAATCCCAAAAATGTTACCCAAAATG
39*$#	CATTCTATACCCAAATGAGTAACGTAGAAC
113*	CATGTAAAAAGTTCCCAATTTAGTATAAACC
7	GTAGGTGATCTTTACCATAATAGTACTCTGG
115	GTCCCGCTTTACCTAAAATGTGCTAATTTG
127*	CATCGTCGACATTTCCTAATTTGTACATACC
40	GTCATTATTTCCTAATCAGTCGATAGCATG
Others	
19*	ATCCATACCAAAGGCGGAAATAGTCTGCCTTAC
88$	GTCTAGGCGCAGGTTTCCTAACTCGGAATGGAT
114	GTTATTTTATTTCCTAAGCGGGCAACCCTG
90	GTATCTACATAACTTACCTTTCTAGGCATGGAT
69*	CATTTATACCACAATCGGAAATTCTAAAAC
43	CGGGTTTCCAGATCAGGCATACTTTTTATGCTG
83*$#	TCCGGATCCACTTCCGGTAACTACAGCGAGTTGTAGAC
92	GTCGAAAATTACCCAATTGGGAACTACATG
110	GTAACATATTCCGAAAGGGGAAAGTTACTG
41$#	GTATACTTTACCGAATGGGGTTAGACTATG
104*$	CCAGTTGCCCTTTCGGGTAATATTCCACGAC
131*	CATATGACCCGATTGGGGTGAAATCTCAAC
78*$	TCCGGATCCACTAAAGACGCATTTCCGTAAAGAGAAAC
3	GTACCACATATACCTTTCCACTTTTAGATGGATC
132*	CAAATTATTCCACTTTGAGGGACTTTCCAC
45*$	CACATTGTCCGAGAATAGAAACTTCTAGAC
57*	GATCCACCCGTAAGTAGAAAGTAGTTACTATTAC
102$	GTTTCTTACCGAAAGTAGAAAGACTCGATG
101$#	GTCGATTTACATTACTTAAAATAGCCCGTGG
117	CCCGGGTCCAAACATTGCTCGATTAGGCATTAATG
61*	ATCCAAAACCCAAAGGGTAAAATTGAAATGTAC
75	GTACGCTCTGATTCCCCATATAGTAATTTGG
77	GTGTACATATATTACCCGTTTCAGTTAATGGA
Twin sites	
53.2*	TCCATTCCCGAAGAAGTAATTTTAACGAGTAC
68*	TCCACTACCGTAACGAGTAATTAACCGCGAAC
123*$	TCCACTACCAGAACAAGTAAAGTATAAGCGAC
6	GTTATTACTACATTAGTACATTATCGATTG
14	GTAATTACTATATCTGAACATAATTGTGTG

The numbers are the clone number of the AG binding sequences; the asterisks indicate that the sequence has been inverted to produce a better fit with the consensus; the dollar ($) signs denote the sequences which have been tested for AG binding individually; and the number (#) signs indicate the sequences which were analyzed for relative binding affinities with AG or SRF.
The bold-faced nucleotides correspond to the core region, similar to the SRF and MCM1 consensus; the underlined ones are from the constant ends; and the italicized ones show possible secondary sites in the oligonucleotides.

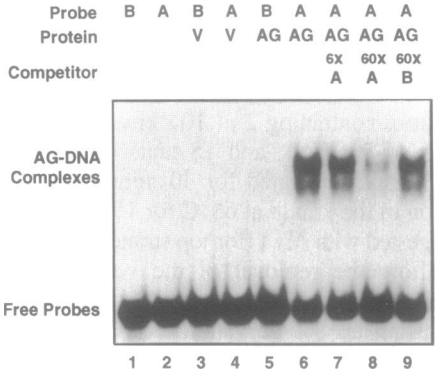

Figure 2. AG binds a specific sequence. The gel mobility shift assay was used to test AG binding to oligo A (A), with oligo B (B) as a negative control. An extract from *E.coli* cells containing only the vector (V) was used as a control for binding of proteins other than AG. Competition was done with two concentrations of non-radioactive oligo A (6×A and 60×A) and one of oligo B (60×B).

The sequence information is summarized in Table II; and the consensus for AG binding sequences is TT(A/T)**CC(A/T)-(A/t)A(T/A)NNGG**(-G)$(A/t)_2$, where the upper case letters represent the most common nucleotide(s) and the lower case letters represent a relatively frequent nucleotide; the nucleotides in bold face correspond to the CArG box, also found in SRF and MCM1 consensus sequences (28, 29). The consensus of AG binding sequences is very similar to that of MCM1 (29), and partly similar to that of SRF (Table II; ref.28).

AG protein protects a region spanning the consensus

Because the selected oligonucleotides contain sequences in addition to the AG binding consensus, we examined three different sequences for the ability of AG to protect the consensus region against DNase I treatment by footprinting analysis. As shown in Fig. 4, in all three sequences AG protects a region including the deduced consensus core region and additional flanking nucleotides. The results show protection of the consensus, which includes not only the 10-nucleotide core region, but also three nucleotides on either side which show significant preferences for particular bases. Taking into account the possible steric hindrance effect of AG on DNase I activity, the footprinting

Table II. Comparison of AG-binding sequence consensus with those of SRF and MCM1

	-4	-3	-2	-1	1	2	3	4	5	6	7	8	9	10	11	12	13	14
AG consensus (n=90, all except twin sites)																		
A	23	14	18	36	0	0	44	63	70	35	24	26	13	3	29	62	56	21
G	18	11	2	11	0	0	7	1	1	2	16	24	74	76	9	7	11	25
C	29	4	0	9	90	89	4	5	0	3	18	13	0	0	26	4	9	15
T	20	61	70	34	0	1	35	21	19	50	32	27	3	11	26	17	14	29
	N	T	T/a	A/T	C	C	A/T	A/t	A/t	T/A	N	N	G	G	-G	A/t	A	N
AG consensus (n=66, totally random [a])																		
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/t	A/t	N
SRF consensus [b] (n=33, totally random)																		
A	28	1	1	7	0	0	23	2	33	0	31	10	0	0	12	13	11	1
G	2	0	31	0	0	0	0	0	0	0	0	0	33	33	3	8	7	8
C	1	0	0	25	33	33	0	0	0	0	0	0	0	0	4	9	6	4
T	2	32	1	1	0	0	10	31	0	33	2	23	0	0	14	3	9	20
	A	T	G	C	C	C	A/t	T	A	T	A	T/a	G	G	T/A	N	N	T
MCM1 consensus [b] (n=23, totally random)																		
A		8	3	6	0	0	0	18	19	11	5	8	5	1	6	14	15	
G		2	4	5	0	0	0	1	1	0	4	6	18	21	0	2	2	
C		3	4	0	23	23	10	1	0	0	10	4	0	0	2	6	0	
T		10	12	12	0	0	13	3	3	12	4	5	0	1	15	1	6	
		T/a	T	-C	C	C	T/C	A	A	T/A	N	N	G	G	T	A	A	

[a]The sequences with no overlap between the constant ends and the region from positions −3 to 13.
[b]From Pollock *et al.* 1990 (SRF, ref.28); and Wynne *et al.* 1992 (MCM1, ref.29).
The CArG boxes are underlined for each consensus sequence.

results support the AG-binding consensus deduced from selected sequences. Interestingly, one or two nucleotides near the CC residues at the beginning of the core sequence on each strand are not protected but rather are hypersensitive to the DNase I treatment. This has also been seen with MCM1 to a lesser extent (20), but not with SRF. The presence of hypersensitive sites suggests that the DNA conformation is altered upon AG binding.

AG binds to different sequences with high affinity

Because many different sequences were selected from the random oligonucleotides, we were interested to learn whether they have different binding affinities for AG. Several oligonucleotides were selected to represent the different types of CArG boxes shown in Table I. These sequences were labeled along with oligo A using the same end-labeled primer, so all the probes have the same specific radioactivity. An identical amount of radioactivity for each probe was used in a gel mobility shift experiment, and as shown in Fig. 5A, oligo #83 (lane 8) has lower affinity for AG than oligo A (lane 2), while the others all have slightly higher affinity for AG than oligo A. The sequences [see Table I, indicated with a (#) sign] tested here have different numbers of mismatches to the AG binding consensus sequence, or matches with the lower case choices. Although it is hard to generalize based on the limited amount of information, it seems that nucleotides within the core region may influence the binding affinity more than the flanking nucleotides.

Several of these sequences contain a core region (corresponding to the CArG box) that is different from human SRF binding sequences. We therefore tested the binding of SRF to these sequences. *E.coli* lysate containing SRF (a gift from Dr D.Grueneberg) was used to bind the same probes shown in Fig. 5A in a gel mobility shift assay, and the results are shown

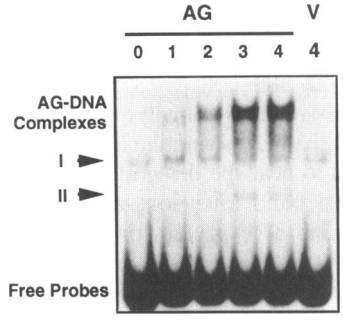

Figure 3. Enrichment of AG binding sequences. Gel mobility shift assay with DNAs after successive rounds of selection. AG protein was used in all lanes except the rightmost one, which contains the control extract from the vector-only cells (V). The numbers above the lanes indicate the number of cycles of the AG binding reaction and gel purification which the probes had gone through. Band (I) is present in all lanes, and it does not seem to be enriched by the AG-binding cycles; therefore, it is likely due to binding of *E.coli* protein(s). Band (II) is AG-specific and seems to be enriched by AG-binding cycles, and may be due to binding by an AG degradation fragment. Only the top bands were excised for amplication and the next round of selection.

in Fig. 5B. Because the standard SRF-binding conditions are slightly different from our AG-binding conditions, we tested SRF binding using both SRF- (no urea, lanes 1−7) and AG- (with urea, lane 8−14) binding conditions (see Materials and Methods). The results show that the relative SRF-binding affinities for different probes are consistent under these two binding conditions. As expected, SRF binds very well to oligo A and oligo #85, which matches the SRF-binding consensus sequence very well (only one or two changes between A and T in the central portion of the CArG box). Oligo #111 has an 'A' instead of 'G' at the

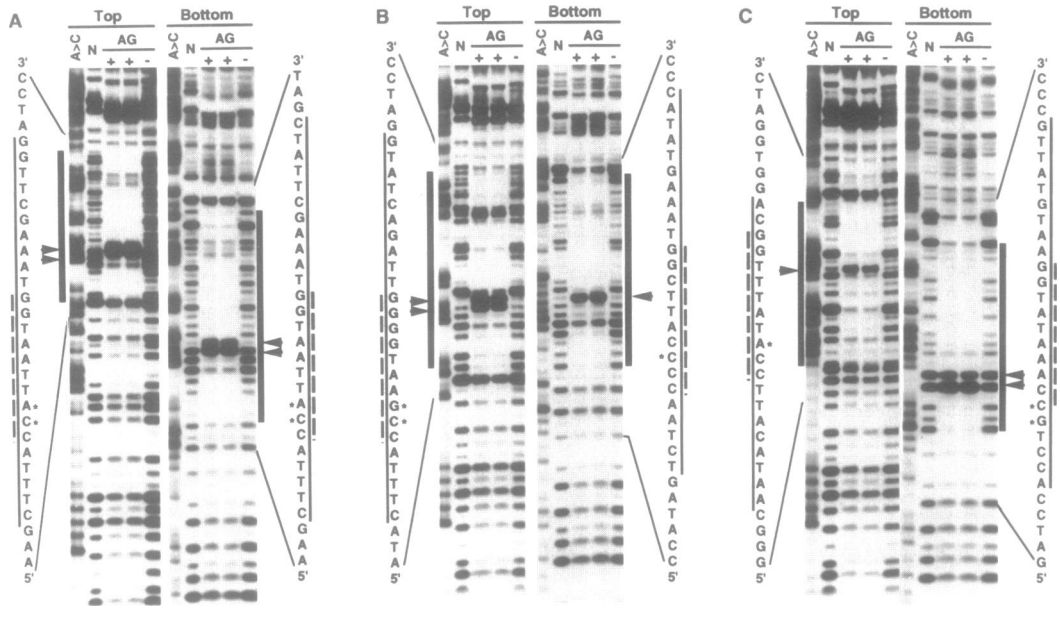

Figure 4. Footprinting analysis with AG. DNase I protection analysis was performed for AG binding to oligo A (**A**), oligo # 41 (**B**) and oligo # 85 (**C**). The protected regions on the autoradiographs are indicated with solid bars, and hypersensitive sites with arrowheads. The protected sequences are indicated with thin lines, and the core region of the consensus is highlighted by dashed lines, and hypersensitive nucleotides with asterisks. Top, top strand; bottom, bottom strand. Lanes: A > C, chemical sequencing reaction; N, no protein; AG(+), extract from cells carrying the *AG* construct; AG(−), extract from cells carrying the vector.

second to last position of the core region (and three changes between A and T), and it has a slightly reduced affinity with SRF. Oligos # 39 and # 101 have two A/T to G/C (or G/C to A/T) differences from the SRF consensus core, and their binding to SRF is dramatically reduced. Finally, oligos # 41 and # 83 have three A/T to G/C changes within the SRF core, and the binding is further reduced to near the detection limit of the assay (# 83 has a faint shifted band after longer exposure, not shown). Because the AG and SRF protein preparations are both from *E.coli* lysates and the AG and SRF protein concentrations are not known, it is not possible to determine the relative binding affinity of AG and SRF to a particular sequence.

DISCUSSION

Genetic studies have identified a number of genes which regulate normal flower development. However, relatively little is known about the molecular mechanisms of their action. Many of these genes potentially encode transcription factors, including a large number of the MADS-box gene family members. Our results presented here show that the product of the floral homeotic gene *AG* indeed binds DNA in a sequence-specific fashion, in agreement with previous findings using an *in vitro* translated AG peptide containing primarily the AG MADS-box domain (aa 41−131, ref.35). Our results support the hypothesis that AG functions by regulating transcription of other genes during flower development, although additional experiments are needed to demonstrate that AG is indeed a transcription factor.

Our results from the analysis of AG-binding oligonucleotides selected from random sequences indicate that AG binds to a remarkably wide range of sequences. We have tested a random subset of 30 sequences, and found that all of them bind to AG. However, a more quantitative analysis of the relative affinities

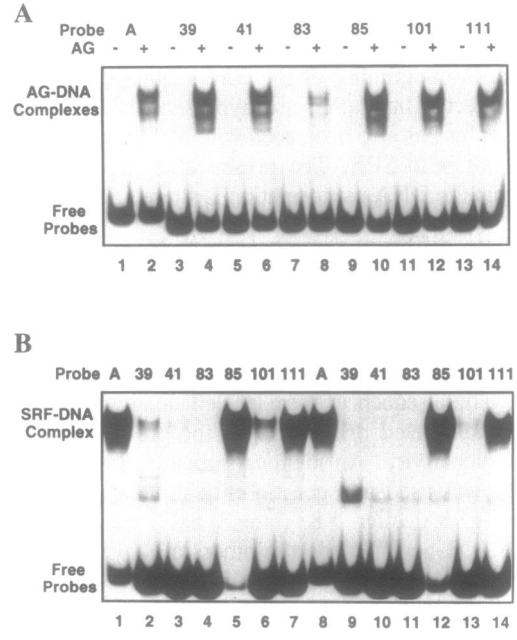

Figure 5. Relative binding affinities of AG and SRF to AG-binding sequences. **A**. AG binding with oligo A and some of the selected AG-binding sequences, with the probe numbers referring to the designations as shown in Table I. Both the vector-only extract (−) and the AG-containing extract (+) were used for binding reactions. **B**. SRF binding with oligo A and selected AG-binding sequences. The probe numbers are the same as in A. Lanes 1−7 were reactions performed using standard SRF binding conditions, and lanes 8−14 were from those using AG binding conditions. The lower bands have also been seen in binding reactions using other sequences and E.coli extract containing SRF, and may be non-specific binding of E.coli proteins or binding by SRF monomer (44).

Table III. Arabidopsis sequences found in the GenBank database which match the AG binding consensus sequence

Locus [a]	Region [b]	Sequence (range [c])	Reference
Exact match (some match the inverted sequence)			
ATHBG1A	5' NTL	TTTCCTTTTTAGGCAA (10-25)	(37)
ATHCDC2A	TL	TTTCCCTAAATGGAAA (2259-2274)	(38)
ATHCHL1A	TL	TTACCAATTTGGGCAA (1111-1126)	(39)
ATTS0377		TTTCCGCAAATGGAAA (35-50)	*
One mismatch (not a complete list)			
ATHACS	5' NTC	CTACCAAATTCGGATA (2068-2083)	(36)
ATHB3MR	5' NTL	TTACCAAAAGAAGCAA (28-43)	(40)
ATHMERI5G	3' NTL	TGTCCAATAAAGGAAA (1482-1497)	(41)
ATHMGCOA	TL	TTACCTAATCCGGAAC (539-554)	(42)
ATTS0500		TTACCAAATTCGTAAA (103-118)	**
ATTS0885		CTTCCAAAACTGGCAA (196-211)	***
T04709		TTACCATTGAAGGATA (120-135)	$
T04448		TTATCAAAATGGGAAA (249-264)	$
T04688		TTACCTTTGAGGGATT (436-451)	$
T04341		TTGCCAAATCTGGAAA (53-68)	$
T04428		TTGCCATTTTGGGAAT (213-228)	$
T04241		TTTCCAAAAAGGGACA (337-352)	$
T04660		TTGCCAAAAAGGGATT (128-143)	$
T04280		TTGCCAAATCTGGAAA (120-135)	$
A centromere associated sequence			
ATT27C1C		TTGCCTTAAGAGTATT (2469-2484)	(43)

[a]The locus designations are as in the Genbank database.
[b]NTC, non-transcribed; NTL non-translated; NL, translated.
[c]The ranges shown are from the Genbank database, and do not necessarily indicate the distnace to transcribed or translated regions.
*Expressed sequence tag (EST) submitted by Parmentier, Y., Criqui, M. C., Durr, A., and Fleck, J.
**EST submitted by Philipps, G, and Gigot, C.
***EST submitted by Berthomieu, P., Guerrier, D, and Giraudat, J.
$EST submitted by Newman, T.

of several sequences indicates that some of these have similar affinities while others have different affinities for AG. Therefore, the *in vitro* procedure is able to select different sequences with a range of affinities. It is worth noting that none of our selected oligonucleotides has the same sequence as any other, indicating that the complexity of AG binding sequences could be even greater.

The consensus sequence of AG binding was determined to be TT(A/T)CC(A/T)(A/t)$_2$(T/A)NNGG(-G)(A/t)$_2$. Since this consensus is derived from *in vitro* binding studies, it is not known whether it reflects the *in vivo* DNA-binding properties of AG. For example, the AG-binding sequences with the highest affinities *in vitro* might not be the natural target sites of AG, rather those sequences with moderate affinities might resemble the true AG targets. Nevertheless, this approach has been shown to be successful in deriving SRF and MCM1 binding sequence consensus (28, 29), which are very similar to their native targets [refs.20, 27; for example, the natural MCM1 targets contain the core sequence of CCTAAT(T/A)(A/G)GG]. Therefore, the results described here probably reflect properties of the AG protein. Since the haploid *Arabidopsis* genome is about 70,000 kb with an (AT) frequency of about 65%, the number of an exact match to the consensus is about 230 per haploid genome {or once every $[0.325^2 \times 0.65^7 \times 0.175^4 \times 0.675]^{-1} = 3.05 \times 10^5$ bp}. A search among the more than 2,000 entries of *Arabidopsis* sequences in the GenBank database found several precise matches to the consensus sequence (Table III); in addition, numerous sequences were found to have one mismatch with the consensus, some of which are in a noncoding region of a gene (Table III). In particular, the *ACC2* gene (locus ATHACS) 5' non-transcribed region contains a site with one mismatch to the AG binding consensus (36), and it is expressed in flowers and seedpods where *AG* is expressed. We have observed that sequences with one or two mismatches to the consensus can bind to AG *in vitro*, suggesting that they might serve as targets of AG *in vivo*. That most of the sequences with matches to the AG binding consensus are in the coding region is probably a reflection of the fact that most of the entries in the GenBank database is from coding regions. Additional potential sites may be located away from transcribed regions and have not been cloned and sequenced (see Table III for an example); these sites are not likely to serve as target sites for AG.

Footprinting studies demonstrate that AG binds to the region spanning the deduced AG-binding sites, supporting the AG-binding consensus. Furthermore, one or two nucleotides near the CC residues were found to be hypersensitive to DNase I treatment. This was also observed to a lesser extent with MCM1 (20). The hypersensitivity sites suggest that the binding of AG results in a conformational change rendering the DNA more sensitive to DNase I. It is interesting that the hypersensitive sites in individual sequences have slightly different positions, suggesting that the conformational change varies somewhat depending on the DNA sequence or the AG–DNA interaction.

The AG-binding consensus sequence is overall similar to those of SRF and MCM1, particularly to that of MCM1 (28, 29). Our comparison of relative binding affinities of AG and SRF to the same set of oligonucleotides showed that SRF binds poorly to some of the sequences to which AG binds very well, suggesting that AG and SRF have similar but distinct sequence specifities. Although side-by-side comparison of AG and MCM1 was not carried out, the *in vitro* binding consensus sequences for these two proteins are remarkably similar, suggesting that AG may have binding specificities very similar to those of MCM1. Both SRF and MCM1 are known to bind DNA as homodimers, and their binding consensus sequences have diad symmetry. That the consensus sequence for AG also shows diad symmetry suggests that AG may also bind as a homodimer. We have also found that one half of the imperfect palindromic AG binding consensus sequence is more conserved than the other, suggesting that if AG binds as a dimer, one AG subunit could bind to the more highly conserved half site and cooperatively stimulate the binding of the other subunit to the less conserved half site. That the two half sites of the binding consensus are conserved to different extents has also been observed with MCM1, but not with SRF (28, 29). These findings suggest that AG and MCM1 subunits may have a stronger cooperativity than SRF. Our results also indicate that AG shows a preference for specific nucleotides in areas flanking the CArG box, similar to SRF and MCM1.

Our finding that AG has a similar binding consensus sequence to those of SRF and MCM1 indicates that the general aspects of protein–DNA interaction have remained intact through evolution. Because there are a number of differences in the amino acid sequence of the MADS-boxes of these proteins, the differences between the consensus sequences are likely to reflect the divergence in specific nucleotide–amino acid contacts. In *Arabidopsis*, there are at least 12 different MADS-box genes, including *AG*, *APETALA1* and *APETALA3*, which have different functions (6, 8, 14, 15). It is possible, given the similarity of the AG-binding consensus sequence to those of SRF and MCM1, that several of the *Arabidopsis* MADS-box proteins bind to similar target sequences. The specificity of gene function may be

mediated by subtle differences in the binding sequences, by different binding affinities of the proteins to the same sequence, or by differential interaction with other proteins. AG is known to be required for both stamen and carpel development, which involve very different morphogenesis programs. AG may control these developmental fates by regulating different sets of genes, possibly through interaction with other proteins, as does MCM1 (25).

CONCLUSION

We have shown that AG is a sequence-specific DNA-binding protein, have presented a wide spectrum of AG binding sequences, and have shown that the consensus for AG binding sequences is related to those of SRF and MCM1. Our results support the hypothesis that AG is a transcription factor regulating flower development. The function of the AG-binding consensus sequence will await the identification and characterization of AG target gene(s).

ACKNOWLEDGEMENTS

We would like to thank E.Grotewold for helpful advice and discussion on DNA-binding experiments, C.Flanagan for the antiserum used in the western analysis of AG, D.Grueneberg for the gift of SRF and advice, S.Bell for advice on footprinting analysis, M.Tudor for help with initial experiments, and E.Grotewold, D.Grueneberg, C.Flanagan, P.Rubinelli, C.Weiss and the anonymous reviewers for helpful comments on this manuscript. This work was supported by funds from the Cold Spring Harbor Laboratory Robertson Fund and by a grant from the National Science Foundation (DCB-9105260) to H.M.

REFERENCES

1. Coen, E.S. and Meyerowitz, E.M. (1991). *Nature*, **353**, 31–37.
2. Meyerowitz, E.M., Bowman, J.L., Brockman, L.L., Drews, G.N., Jack, T., Sieburth, L.E. and Weigel, D. (1991). *Development*, **Supplement 1**, 157–167.
3. Komaki, M.K., Okada, K., Nishino, E. and Shimura, Y. (1988). *Development*, **104**, 195–203.
4. Bowman, J.L., Smyth, D.R. and Meyerowitz, E.M. (1989). *Plant Cell*, **1**, 37–52.
5. Kunst, L., Klenz, J.E., Martinez-Zapater, J. and Haughn, G.W. (1989). *Plant Cell*, **1**, 1131–1135.
6. Yanofsky, M.F., Ma, H., Bowman, J.L., Drews, G.N., Feldmann, K.A. and Meyerowitz, E.M. (1990). *Nature*, **346**, 35–39.
7. Bowman, J.L., Smyth, D.R. and Meyerowitz, E.M. (1991b). *Development*, **112**, 1–20.
8. Jack, T., Brockman, L.L. and Meyerowitz, E.M. (1992). *Cell*, **68**, 683–697.
9. Norman, C., Runswick, M., Pollock, R. and Treisman, R. (1988). *Cell*, **55**, 989–1003.
10. Passmore, S., Maine, G.T., Elble, R., Christ, C. and Tye, B.K. (1988). *J. Mol. Biol.*, **204**, 593–606.
11. Dubois, E., Bercy, J. and Messenguy, F. (1987). *Mol. Gen. Genet.*, **207**, 142–148.
12. Sommer, H., Beltran, J.-P., Huijser, P., Pape, H., Lonnig, W.-E., Saedler, H. and Schwarz-Sommer, Z. (1990). *EMBO J.*, **9**, 605–613.
13. Schwarz-Sommer, Z., Huijser, P., Nacken, W., Saedler, H. and Sommer, H. (1990). *Science*, **250**, 931–936.
14. Ma, H., Yanofsky, M.F. and Meyerowitz, E.M. (1991). *Genes & Dev.*, **5**, 484–495.
15. Mandel, M.A., Gustafson-Brown, C., Savidge, B. and Yanofsky, M.F. (1992). *Nature*, **360**, 273–277.
16. Huijser, P., Klein, J., Lönnig, W.-E., Meijer, H., Saedler, H. and Sommer, H. (1992). *EMBO J.*, **11**, 1239–1249.
17. Tröbner, W., Ramirez, L., Motte, P., Hue, I. and Huijser (1992). *EMBO J.*, **11**, 4693–4704.
18. Bradley, D., Carpenter, R., Sommer, H., Hartley, N. and Coen, E.S. (1993). *Cell*, **72**, 85–95.
19. Jarvis, E.E., Clark, K.L. and Sprague, G.F., Jr. (1989). *Genes & Dev.*, **3**, 936–945.
20. Passmore, S., Elble, R. and Tye, B.K. (1989). *Genes & Dev.*, **3**, 921–935.
21. Ryan, W.A., Jr.,, Franza, J., B.R. and Gilman, M.Z. (1989). *EMBO J.*, **8**, 1785–1792.
22. Treisman, R.H. (1986). *Cell*, **46**, 567–574.
23. Treisman, R.H. (1987). *EMBO J.*, **6**, 2711–2717.
24. Boxer, L.M., Prywes, R., Roeder, R.G. and Kedes, L. (1989). *Mol. Cell. Biol.*, **9**, 515–522.
25. Herskowitz, I. (1990). *Nature*, **342**, 749–757.
26. Bender, A. and Sprague, G.F., Jr. (1987). *Cell*, **50**, 681–691.
27. Hayes, T.E., Sengupta, P. and Cochran, B.H. (1988). *Genes & Dev.*, **2**, 1713–1722.
28. Pollock, R. and Treisman, R. (1990). *Nuc. Acids Res.*, **18**, 6197–6204.
29. Wynne, J. and Treisman, R. (1992). *Nuc. Acids Res.*, **20**, 3297–3303.
30. Studier, W., Rosenberg, A.H., Dunn, J.J. and Dubendorff, J.W. (1990). *Methods Enzymol.*, **185**, 60–89.
31. Mizukami, Y. and Ma, H. (1992). *Cell*, **71**, 119–131.
32. Mandel, M.A., Bowman, J.L., Kempin, S.A., Ma, H., Meyerowitz, E.M. and Yanofsky, M.F. (1992). *Cell*, **71**, 133–143.
33. Garcia, A., LaMontagne, K., Reavis, D., Stober-Grässer, U. and Lipsick, J.S. (1991). *Oncogene*, **6**, 265–273.
34. Sambrook, J., Fritsch, E.F. and Maniatis, T. (1989). Molecular Cloning: A Laboratory Manual, 2nd Edt. Cold Spring Harbor: Cold Spring Harbor Laboratory Press.
35. Mueller, C.G. and Nordheim, A. (1991). *EMBO J.*, **10**, 4219–4229.
36. Liang, X., Keller, J.A., Abel, S., Shen, N.F. and Theologis, A. (1992). *Proc. Natl. Acad. Sci. USA*, **89**, 11046–11050.
37. Dong, X., Mindrinos, M., Davis, K.R. and Ausubel, F.M. (1991). *Plant Cell*, **3**, 61–72.
38. Imajuku, Y., Hirayama, T., Endoh, H. and Oka, A. (1992). *FEBS Lett.*, **304**, 73–77.
39. Tsay, Y.-F., Schroeder, J.I., Feldmann, K.A. and Crawford, N.M. (1993). *Cell*, **72**, 705–713.
40. Mattsson, J., Soederman, E., Svenson, M., Borkird, C. and Engstroem, P. (1992). *Plant Mol. Biol.*, **18**, 1019–1022.
41. Medford, J., Elmer, J. and Klee, H.J. (1991). *Plant Cell*, **3**, 359–370.
42. Caelles, C., Ferrer, A., Balcells, L., Hegardt, F.G. and Boronat, A. (1989). *Plant Mol. Biol.*, **13**, 627–638.
43. Richards, E.J., Goodman, H.M. and Ausubel, F.M. (1991). *Nucleic Acids Res.*, **19**, 3351–3357.
44. Grueneberg, D.A., Natesan, S., Alexandre, C. and Gilman, M.Z. (1992). *Science*, **257**, 1089–1095.