3134–3141 *Nucleic Acids Research, 1996, Vol. 24, No. 16*

DNA-binding properties of *Arabidopsis* MADS domain homeotic proteins APETALA1, APETALA3, PISTILLATA and AGAMOUS

José Luis Riechmann, Minqin Wang and Elliot M. Meyerowitz*

Division of Biology 156-29, California Institute of Technology, Pasadena, CA 91125, USA

Received May 28, 1996; Accepted July 3, 1996

ABSTRACT

The MADS domain proteins APETALA1 (AP1), APETALA3 (AP3), PISTILLATA (PI), and AGAMOUS (AG) specify the identity of *Arabidopsis* floral organs. AP1 and AG homocomplexes and AP3–PI heterocomplexes bind to CArG-box sequences. The DNA-binding properties of these complexes were investigated. We find that AP1, AG and AP3–PI are all capable of recognizing the same DNA-binding sites, although with somewhat different affinities. In addition, the three complexes induce similar conformational changes on a CArG-box sequence. Phasing analysis reveals that the induced distortion is DNA bending, oriented toward the minor groove. The molecular dissection of AP1, AP3, PI and AG indicates that the boundaries of the dimerization domains of these proteins vary. The regions required to form a DNA-binding complex include, in addition to the MADS box, the entire L region (which follows the MADS box) and the first putative amphipathic helix of the K box in the case of AP3–PI, while for AP1 and AG only a part of the L region is needed. The similarity of the DNA-binding properties of AP1, AP3–PI and AG is discussed with regard to the biological specificity that these proteins exhibit.

INTRODUCTION

According to a well-established genetic model (1–3), the identities of the organs of an *Arabidopsis* flower are specified by the action of at least five homeotic genes: *APETALA1* (*AP1*), *APETALA2* (*AP2*), *APETALA3* (*AP3*), *PISTILLATA* (*PI*) and *AGAMOUS* (*AG*) (4–8). While these genes have been extensively characterized at the genetic level, little is known about the molecular mechanisms by which the organ-identity proteins act. AP1, AP3, PI and AG are all MADS domain proteins (4,6–8). The MADS domain is a conserved DNA-binding/dimerization region present in a variety of transcription factors from different organisms (SRF, serum response factor; MCM1; the MEF2 family) (9,10). Within the family of MADS domain proteins, a particular characteristic of the plant proteins is that the vast majority of them contain another conserved region, the K box (11,12). This region has similarity to the coiled-coil segment of keratin, suggesting that the K box may form amphipathic alpha helices, perhaps involved in protein–protein interactions (11,13). SRF and MCM1 recognize CArG-box sequences (consensus $CC(A/T)_6GG$) (14,15), and *in vitro* experiments have shown that AG (16–20), AP1 and AP3–PI (20) complexes bind to such sites. These four proteins exhibit partner specificity for the formation of DNA-binding complexes: neither AP3 nor PI have been found to bind DNA by themselves or in combination with either AP1 or AG (20).

Since these four related proteins act to specify the development of different organ types in the *Arabidopsis* flower, we were interested in comparing the DNA binding properties of the AP1, AP3–PI and AG complexes, which are dimers, in an effort to understand how the biological specificity of these (presumed) transcription factors is achieved. We show that the DNA-binding specificities of AP1, AP3–PI and AG dimers are very similar, since they recognize the same DNA-binding sites, although differences in affinities were detected. The three complexes are also similar in the distortion that they induce on the DNA, that is (at least in part) DNA bending toward the minor groove. In addition, the molecular dissection of AP1, AP3, PI and AG has revealed differences in the regions that are required for dimerization among these four proteins, which correlate with the partner specificity that they exhibit.

MATERIALS AND METHODS

Plasmids for *in vitro* transcription/translation

pSPUTK (Stratagene)-derived plasmids to produce AP1, AP3, PI and AG in *in vitro* transcription/translation reactions have been described previously (20). Several derivatives of *AP3*, *PI* and *AG* sequences were synthesized by PCR in order to make N- and C-terminal truncated proteins, as listed below, and cloned into pSPUTK. Throughout this article, the N- (N-terminal extension that precedes the AG MADS-box), M- (MADS domain), L- (linker between the MADS domain and the K box), K- (K box) and C- (C-terminal) regions of AP1, AP3, PI and AG, as well as the corresponding amino acid numbering, are as shown in Figure 1 of ref. 20. AP3ΔMLCK: AP3 protein lacking the first 26 aa of the MADS box (Asn residue at position 26 is changed into the initiation Met). AP3ML: truncated AP3 protein comprising only the MADS box and the L region (a stop codon was

* To whom correspondence should be addressed

introduced after the Gln residue at position 88). PIML: truncated PI protein comprising only the MADS box and the L region. AGΔMLCK: AG protein lacking the N region and the first half of the MADS box (Asn residue is changed into the initiation Met). AGLKC: the AGLKC protein lacks the N region and the entire MADS box except the two last aa (it therefore starts with the sequence Met-Glu-Tyr-Ser...). AGNM: truncated AG protein comprising only the N region and the MADS box. AGNML: truncated AG protein comprising the N, MADS and L regions. AGNMLK: truncated AG protein lacking the C-terminal region.

DNA-binding site probes

Seven different probes (A–G) were used. Probes A and B are derived from the promoters of the *Arabidopsis AP3* and *SUPERMAN* (*SUP*) genes, respectively, and have been described previously (20); probe D is derived from the *Arabidopsis AGL5* promoter (19); sites C and E were obtained in sequence-selection experiments performed with AG (17; site clones #85 and #41, respectively); sites F and G were obtained in sequence-selection experiments performed with AGL3 (21; site clones #3 and #103, respectively). All binding sites were cloned into pGEM vectors. Probe A, 5′-*ggatcc*TCACTTAGTTTTCATCAACTTCTGAACTTACCTTTCATGGATTAGGCAATACTTTCCATTTTTAGTAACT*aagctt*-3′, (an additional CArG-box like sequence is also present in this probe, but it was determined by site-directed mutagenesis that it is not recognized by any of the proteins used in this study); probe B, 5′-*ggatcc*TAAGAAAAATGGGAGAAAGGAACATCCACTTTTCCATTTTTGGTATAAAACTTTTGATATAATATGTCCTTTTGCT*aagctt*-3′; probe C, 5′-*aagctt*gcatgcctgcaggtcgactctagaggatccacagcAATACATTCCATATTTGGCAGGTGGCtccg*gaattc*-3′; probe D, 5′-*ggatcC*AATAAAAAGAAAAGGAGAATAAAAAGGGATTACCAAAAAAGGAAAGTTTCCAAAAGGTGATTCTGATG*Aagctt*-3′; probe E, 5′-*tctaga*ctcaggaattcggtaccccgggtATACTTTACCGAATGGGGTTAGACTA*tggatcc*-3′; probe F, 5′-*tctaga*ctcgaggaattcggtaccccgggtTCAACCCCATTTATAGCCACGTCAGT*tggatcc*-3′; probe G, 5′-*tctaga*ctcgaggaattcggtaccccgggtACGCATGCACCACATATAGTAACGTG*tggatcc*-3′; (the CArG-boxes are underlined and the plasmid-derived linker sequences are in lower case). Binding probes were prepared as described (20).

In vitro transcription and translation

Proteins were synthesized using the TNT coupled transcription/translation reticulocyte lysate system (Promega). Labeled ([^{35}S]methionine) *in vitro* translation reactions demonstrated that the proteins were produced in similar amounts. Some of the C-terminal deletion derivatives of AP1, AP3, PI and AG proteins were obtained by digesting the plasmids encoding the full-length proteins with internal restriction sites prior to the *in vitro* transcription reaction. AP1M–2, AP1M+3, AP1M+15, AP1M+29, AP1M+33, AP1ML+6, and AP1ML+34 proteins were obtained from RNAs synthesized from pSPUTK-AP1 linearized with *Bst*BI, *Hin*fI, *Rsa*I, *Ple*I, *Bsr*I, *Afl*II and *Alw*NI, respectively. To obtain AP3ML+12, AP3ML+31 and AP3ML+42, pSPUTK-AP3 was linearized with *Ple*I, *Alw*NI and *Fok*I, respectively. The PI open reading frame was linearized at *Ecl*136II and *Bpm*I sites to generate PIML+16 and PIML+20, respectively. AGNM+22 and AGNM+28 were obtained after digestion of pSPUTK-AG with *Ase*I and *Msp*I. The RNAs were purified by agarose gel electrophoresis prior to their use in *in vitro* translation reactions performed with standard reticulocyte lysate (Promega).

DNA-binding assays and immunoprecipitation experiments

In vitro translated proteins were tested for DNA-binding activity by electrophoretic mobility shift assay (EMSA). Binding reactions were performed as described previously (20). Gels for resolving protein–DNA complexes were 5% (except when indicated otherwise) polyacrylamide:bisacrylamide (60:1) in 1× TBE. Immunoprecipitation experiments were carried out as described previously (20).

Apparent K_d values in DNA binding

Saturation-binding assays to determine the dissociation constants (K_d) were carried out by incubating a fixed amount of *in vitro* translated proteins (2 µl of the translation reaction) with increasing amounts of probes A or B under the standard conditions (the incubation time after addition of the probe was extended to 90 min to allow the binding reactions to reach equilibrium, as determined in pilot experiments). Probes were used at concentrations between 1 and 80 nM, the concentration range depending on the protein/probe combination. After gel electrophoresis, bound and free probe were quantitated with a phosphorimager (Molecular Dynamics). The production of both the full-length and a truncated AP1 protein in the *in vitro* translation reactions resulted in the formation of three different AP1 DNA-binding complexes. The amount of probe bound by all of them was quantitated, and the values obtained were used for the calculations as the total amount of bound probe. DNA-binding reactions with AG also showed band shifts originated by truncated AG proteins, but the amount of probe that was bound in the AG reactions is very low. This ensured that the concentration of free probe at equilibrium was approximately equal to the concentration of total probe, and therefore that the values obtained for the probe bound only by full-length AG could be used to calculate the apparent K_ds. K_ds were estimated by the method of Scatchard and calculation of the least-square fit line of the primary data, wherein $K_d = -1/\text{slope}$ (22).

Circular permutation and phasing analyses

For circular permutation analysis, two annealed complementary oligonucleotides containing the site A CArG-box, 5′-CTAGAGCAATACTTTCCATTTTTAGTAACTCAAGTC-3′ and 5′-TCGAGACTTGAGTTACTAAAAATGGAAAGTATTGCT-3′, were cloned into the *Xba*I/*Sal*I sites of pBend2 (23), generating plasmid pBendA. Probes were prepared by digestion of pBendA with the appropriate restriction enzymes and labeling of the purified fragments with ^{32}P using T4 polynucleotide kinase. The magnitude of apparent DNA bending was calculated using the formula $\mu_{min}/\mu_{max} = \cos(\alpha_D/2)$ (24), where α_D is the distortion angle, and μ_{min} and μ_{max} are the relative mobilities of the slowest and fastest migrating species. The values of μ_{min} and μ_{max} were calculated from the curve produced after fitting the data using a computer function (Cricket Graph III, Cricket Software).

For the phasing analysis, sequences containing the site A CArG-box separated by a linker of variable length from an A tract (intrinsically bent toward the minor groove by approximately 54°; 25) were cloned into the *Xba*I/*Sal*I sites of pBend2 (23). The distance between the center of the CArG-box and the center of the

A tract was 21, 23, 26, 28 or 30 bases. Sequences were as follows: 5′-ctcagaTTTCCATTTTTAGTATAAAAACGGGCAAAAACGGGCAAAAACGgtcgac-3′; 5′-ctcagaTTTCCATTTTTAGTAAGCAAAAACGGGCAAAAACGGGCAAAAACGgtcgac-3′; 5′-ctcagaTTTCCATTTTTAGTAACTGGCAAAAACGGGCAAAAACGGGCAAAAACGgtcgac-3′; 5′-ctcagaTTTCCATTTTTAGTAACTGTAGCAAAAACGGGCAAAAACGGGCAAAAACGgtcgac-3′; and 5′-ctcagaTTTCCATTTTTAGTAACTGTACTGCAAAAACGGGCAAAAACGGGCAAAAACGgtcgac-3′ (the CArG-box and the A tract are underlined). Probes of 164–173 bp in length were prepared by digestion of the resulting plasmids with *Pvu*II, and labeling of the purified DNA fragments with ^{32}P using T4 polynucleotide kinase.

RESULTS

Comparison of DNA-binding by AP1, AP3–PI and AG complexes

The DNA-binding capabilities of AP1, AP3–PI and AG complexes were compared using several CArG-box containing sequences as binding sites. *In vitro* translated AP1, AP3, PI and AG were incubated with probes A, B, C and D, and the protein–DNA complexes analyzed by electrophoretic mobility shift assays (EMSA) (Fig. 1). Probes A, B and D contain CArG sequences that are found in the promoters of three *Arabidopsis* genes (see Materials and Methods), while probe C is based on a synthetic AG-binding site identified in sequence-selection experiments (17). The probes were labeled to the same specific activity, allowing direct comparison between the reactions containing the same protein. The shifted bands present in the reactions with AP1 and AG correspond to protein–DNA complexes formed by the full-length proteins as well as by truncated proteins also produced in the translation reactions. AP1 showed the strongest binding to probes A and D, recognizing the probes in the order A~D>B>C (Fig. 1). A similar behavior was observed for AP3–PI, while the affinities of AG for probes A, B and D were comparable and higher than that for probe C (Fig. 1). The binding of AG to probe C is revealed in a longer exposure of the autoradiogram (Fig. 1, lane 13).

Apparent dissociation constants (K_ds) were estimated by Scatchard analyses of saturating binding assays (22) in which a constant amount of *in vitro* translated protein was titrated with increasing amounts of the A and B probes (Fig. 2). AP1 showed a higher affinity for probe A (K_d = 4.6 nM, Fig. 2) than for probe B (K_d = 43.4 nM; r^2 value for the least square fit line was 0.893). The AP3–PI complex showed a similar behavior, although with somewhat lower affinities: K_d for probe A was 12.5 nM (r^2 = 0.942), while that for probe B could not be estimated because saturation was not reached in the range of probe concentration used. AG has comparable affinities for probes A and B: K_ds were 3.8 (r^2 = 0.744) and 2.7 (r^2 = 0.820), respectively. Similar results were obtained in duplicated experiments. The fact that the intensities of the bands due to AG were substantially weaker than those of the bands produced by AP1 or AP3–PI complexes (Fig. 1) indicates that only a very minor fraction of the AG protein synthesized was active in DNA-binding, since the amounts of AP1, AP3, PI and AG that are produced by the *in vitro* translation reactions are comparable (not shown) and AG binds with high affinity to probes A and B.

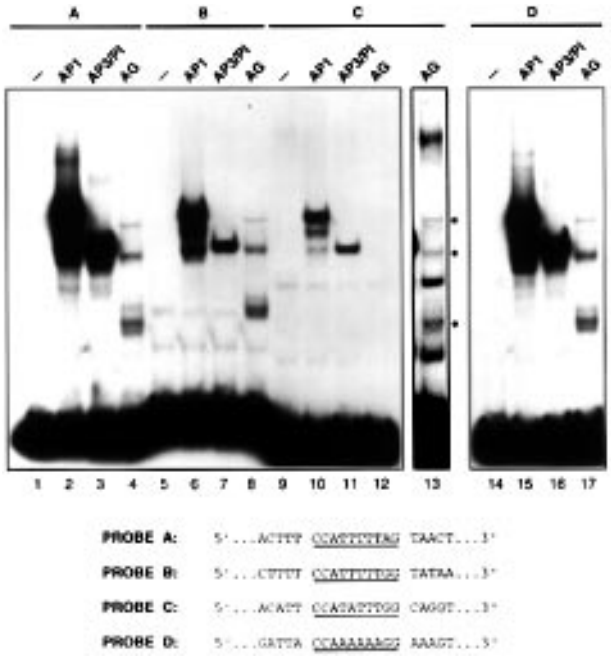

Figure 1. AP1, AP3–PI, and AG complexes bind to CArG-box containing sequences. *In vitro* translated AP1, AP3 and PI (cotranslated), and AG were assayed for DNA-binding activity with probes A, B, C and D. Controls with unprogrammed lysate for each of the probes are included (lanes 1, 5, 9 and 14). Lane 13 shows a longer exposure of lane 12. The shifted bands visible on lane 13 that are specific to the presence of AG in the reaction are indicated by asterisks; additional shifted bands are non-specific and originated by the reticulocyte lysate.

AP1, AP3–PI and AG DNA-binding activity was assayed with three additional CArG-box containing sequences, probes E, F and G, derived from sequence-selection experiments performed with either AG or AGL3 (17,21). AP1, AP3–PI and AG bind to all three probes (a single exception being AP1 and probe E, a combination for which no binding could be detected in the experimental conditions used) (data not shown). AP1, AP3–PI and AG bound probes E, F and G with much lower affinities than probes A, B, C and D, consistent with the fact that the former vary more from the canonical CC(A/T)$_6$GG site and the consensus sequence [5′-(T/a)(T/a)(A/T/g)CC(A/T)$_4$(A/T/g/c)$_2$(G/a)(G/t)(A/T/C)(A/t)(A/t/g/c)-3′] deduced from AG-sequence-selection experiments (17,18) (data not shown).

DNA bending by AP1, AP3–PI and AG complexes

Circular permutation analysis was used to determine whether AP1, AP3–PI and AG complexes induce conformational changes on the DNA upon binding to a CArG-box sequence. This assay is based on the position-dependent effects of DNA distortion on the electrophoretic mobility of DNA fragments of the same length (26). A series of probes were prepared in which the position of the site A CArG-box varies with respect to the ends of the fragments, that are otherwise of identical sequence (Fig. 3A). These circularly permutated probes were used in EMSAs with AP1, AP3–PI and AGNML. In all cases, protein–DNA complexes in which the CArG-box sequence is localized toward the center of the DNA fragment (probes 3, 4 and 5) showed lower mobility than those in which the CArG-box is located near either end

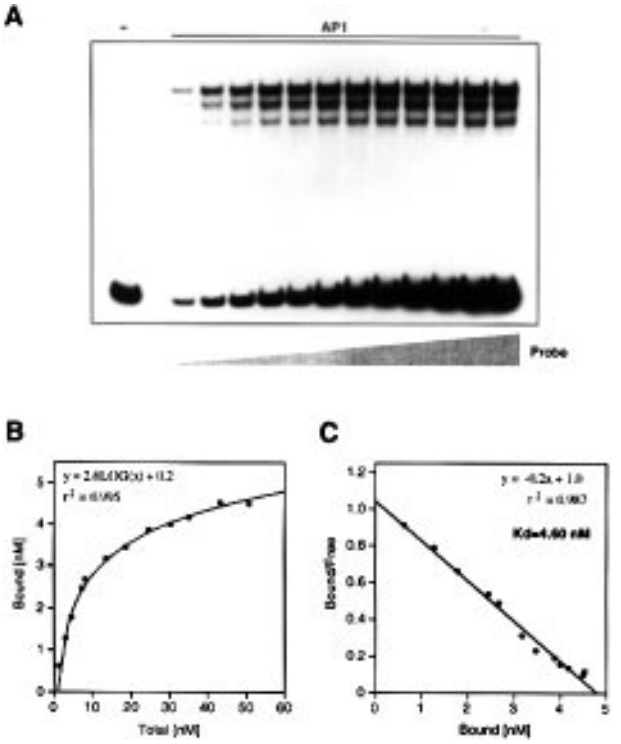

Figure 2. Measurement of DNA-binding affinity of AP1 for probe A. (**A**) Saturation binding assay with increasing concentrations of probe A (from 1 to 50 nM) and a constant amount of *in vitro* translated AP1. The production of a truncated AP1 protein as a minor product of the *in vitro* translation reaction results in the formation of three different protein–DNA complexes. Free and bound (by all the AP1 complexes) probe were quantitated. A control with unprogrammed lysate and probe A at 12 nM is included (left lane). (**B**) The amount of bound probe is plotted as a function of total input. (**C**) Scatchard plot (ratio of bound and free versus bound) of the saturation curve shown in (B). A linear correlation was observed between the two variables, allowing calculation of the dissociation constant ($K_d = -1$/slope). This value is an average of the K_ds for the various AP1 protein complexes with DNA.

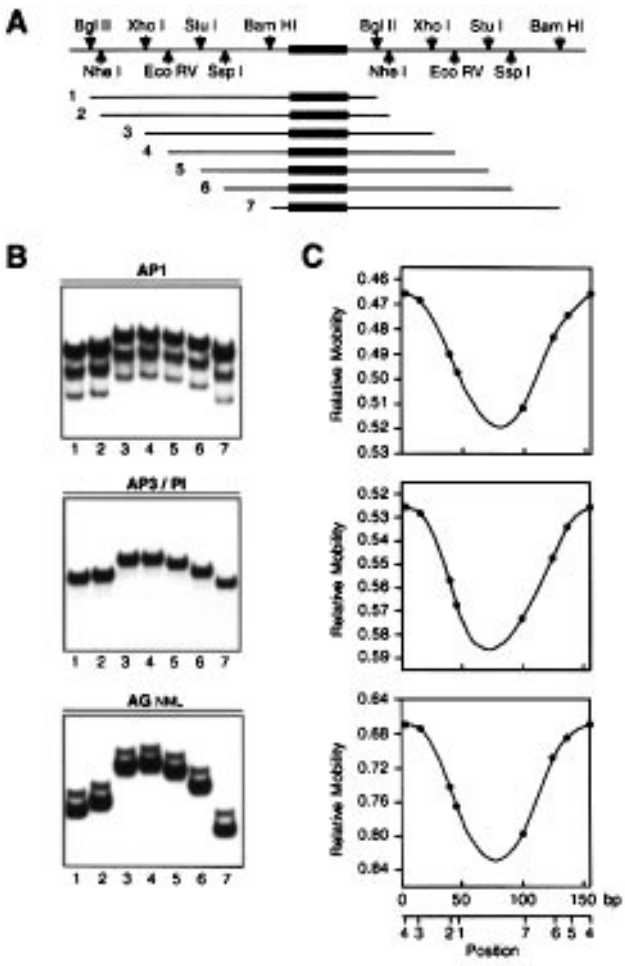

Figure 3. Circular permutation analysis of DNA distortions induced by AP1, AP3–PI and AGNML complexes. (**A**) The probes used for circular permutation analysis were generated by restriction endonuclease cleavage of pBendA, containing the site A CArG-box sequence (black box) flanked by two tandem polylinker sequences, with the seven enzymes shown (1–7). (**B**) Electrophoretic mobility shift analysis of AP1, AP3–PI and AGNML bound to circularly permutated probes. The faster migrating complexes visible in the reactions with AP1 are due to the presence of truncated proteins also produced in the translation reactions. Reactions with AGNML showed an additional band of lower mobility. (**C**) The relative mobilities of the AP1–, AP3–PI– and AGnml–DNA complexes were normalized for slight differences in probe mobilities and plotted as a function of the distance between the center of the CArG-box and the center of the probes.

(probes 1, 2, 6 and 7) (Fig. 3B), indicating that AP1, AP3–PI and AGNML induce DNA conformational changes. The unbound probes possessed similar mobilities, regardless of the position of the CArG-box, suggesting that they do not contain significant intrinsic DNA bends (data not shown). The distortion angles were calculated from the data obtained in the circular permutation analysis (Fig. 3C), and estimated to be 53° (AP1 and AP3–PI), and 73° (AGNML). The full-length AG protein was also used in EMSAs with the circularly permutated probes, and its induced apparent bend angle was estimated to be 70° (data not shown).

The DNA distortions induced by AP1, AP3–PI and AGNML were further investigated using phasing analysis (27–29), which determines the direction of the protein-induced bend with respect to an intrinsic DNA bend. A series of DNA probes were prepared such that the site A CArG-box sequence is separated by a linker of variable length from a 25 bp sequence that contains an A tract intrinsically bent toward the minor groove (25). In this set of probes, the distance between the center of the CArG-box and the center of the A tract is varied from 21 to 30 bp, almost a helical turn, to place the CArG-box on different faces of the DNA relative to the intrinsic bend (Fig. 4A). If AP1, AP3–PI or AGNML complexes bend the DNA at the CArG-box in the same orientation as the A tract, the two bends cooperate to increase the overall extent of bending, resulting in a slow-moving complex in the mobility shift assays. If, on the contrary, the protein induced bend and the intrinsic bend counteract each other, a faster-moving complex will be formed. The DNA–protein complexes formed between the phasing probes and AP1, AP3–PI and AGNML showed variations in electrophoretic mobility that depended on the spacing between the CArG-box and the intrinsic DNA bend, confirming that these MADS-domain proteins induce directed DNA bends (Fig. 4B and C). In all three cases, binding to probe 21, in which the centers of the two bends are separated by ~2 helical turns, resulted in the complex with the slowest mobility (Fig. 4B and C). Therefore, since the two bends cooperate when their centers are in phase, the net orientation of DNA bending

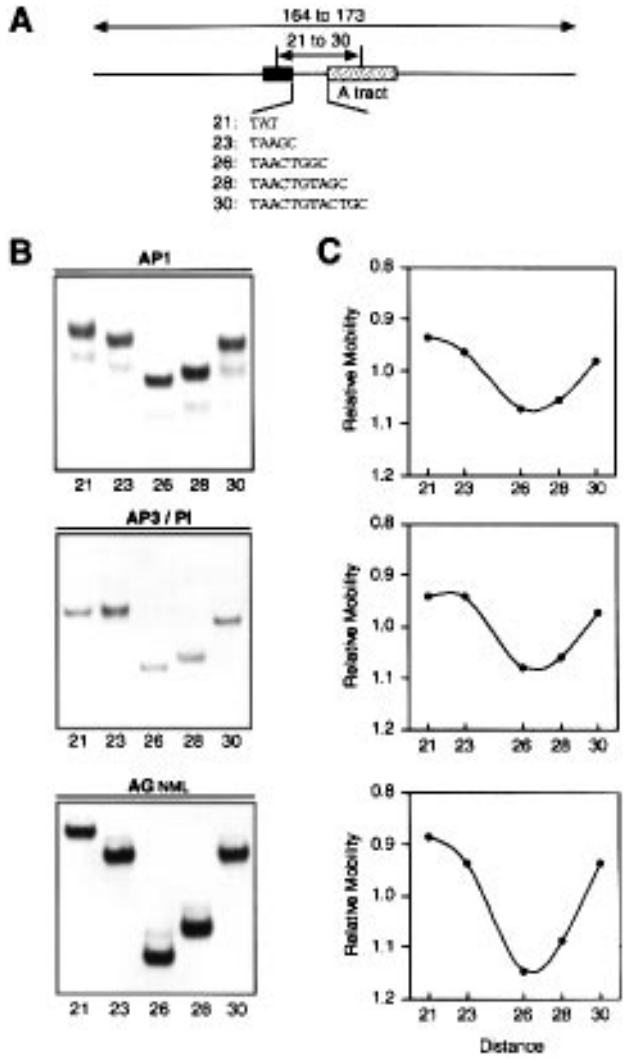

Figure 4. Phasing analysis of DNA bending by AP1, AP3–P1 and AGNML complexes. (**A**) The probes used for phasing analysis contained the site A CArG-box sequence (black box) separated by a linker of variable length from an A tract sequence (dotted box) that is bent intrinsically toward the minor groove. The distance between the center of the CArG-box and the center of the A tract in the different probes was 21, 23, 26, 28 or 30 bp. (**B**) Electrophoretic mobility shift analysis of AP1, AP3–PI and AGNML bound to phasing analysis probes. (**C**) Relative mobilities of the AP1–, AP3–PI– and AGNML–DNA complexes plotted as a function of the distance between the center of the CArG-box and the A tract. In each case, mobility of each protein–probe complex was normalized to the average mobility of all the complexes.

induced by these proteins with respect to the center of the CArG-box is toward the minor groove. The amplitude of the phasing curve obtained with AGNML was larger than those from AP1 and AP3–PI (Fig. 4C), which is indicative of a more pronounced bend (29), as was suggested by the circular permutation analysis.

The observation that AP1, AP3–PI and AGNML induce DNA bending toward the minor groove is in agreement with recent results obtained using circular permutation and phasing analyses that indicated that SRF induces bending in the same orientation (30), and with the crystal structure of core SRF bound to DNA that showed the DNA bent around the protein by 72° (9).

DNA-binding domains of AP1, AP3, PI and AG proteins

To investigate the role of the K box and other regions that are C-terminal to the MADS box in DNA-binding complex formation, a series of C-terminal deletion derivatives of AP1, AP3, PI and AG were produced by *in vitro* transcription/translation. Regions C-terminal to the AP1 MADS box are required to form a DNA-binding complex, since neither AP1M-2 nor AP1M+3 derivatives have such activity (Fig. 5A; AP1M–2 and AP1M+3 truncated proteins lack the last two amino acids of the MADS box or contain the first three amino acids of the L region, respectively). The K box is not required for DNA binding, as AP1M+15, AP1M+29 and AP1M+33 truncated proteins were capable of DNA binding (Fig. 5A). Therefore, the 'core' AP1 protein (minimal DNA-binding domain) consists of the MADS box and part of the L region. AP1M+15 binds to DNA, but at much reduced levels compared with AP1M+29 (Fig. 5A; comparable amounts of the truncated proteins were produced in the *in vitro* translation reactions). AP1 truncated proteins were also used to show that the DNA-binding complex is a protein dimer. The presence of both AP1M+29 and AP1ML+34 in the DNA-binding reaction leads to the formation of a single additional complex of intermediate mobility, corresponding to a heterodimer of both protein forms (Fig. 5A, lanes 10–12).

Similar results were obtained for the AG protein: AGNM+22 bound DNA, whereas AGNM did not (Fig. 5B). Thus, the minimal DNA-binding domain of both AP1 and AG encompasses the MADS box and part of the L region. Curiously, in those reactions with AG truncated proteins that do not include the K box (AGNM+22, AGNM+28, and AGNML) the amount of shifted probe was greater than that when assaying AGNMLK or AG (Fig. 5B). This did not result from substantial differences in the amounts of protein that were produced in the translation reactions (data not shown) or from differences in the DNA-binding affinities (K_ds for AGNML and probes A and B were determined and found to be in the same range as those of AG; data not shown). It may be that the full-length protein has more difficulty in folding properly in the *in vitro* translation. Reactions with AGNM+22, AGNM+28, and AGNML showed an additional retarded band, of weaker intensity, that could be due to a different conformation or shape of the protein–DNA complexes. AG truncated proteins were also used to show that the DNA-binding complex is a protein dimer (Fig. 5B, lanes 8–12).

In contrast with the results obtained for AP1 and AG, AP3ML truncated protein did not show DNA-binding activity when assayed together with PI or PIML (Fig. 5C, lanes 2 and 3), and neither did the PIML protein with several AP3 derivatives (Fig. 5C, lanes 3, 6, 11, 16 and 21). The first 12 amino acids of the AP3 K box were not enough to restore DNA-binding complex formation (AP3ML+12 variant; Fig. 5C, lanes 4–8), while AP3ML+31 and AP3ML+42 were functional when combined with an appropriate PI derivative or with full-length PI (Fig. 5C, lanes 9–18). PIML+16 and PIML+20 truncated proteins could form DNA-binding complexes together with AP3ML+31 (Fig. 5C, lanes 12 and 13) and AP3ML+42 (Fig. 5C, lanes 17 and 18) but, curiously, not with the full-length AP3 protein (Fig. 5C, lanes 22 and 23). In summary, the minimal DNA-binding domains of AP3 and PI differ from those of AP1 and AG. AP3 and PI proteins require amino acids in the K box, in addition to the MADS domain and the L region, to form a DNA-binding complex.

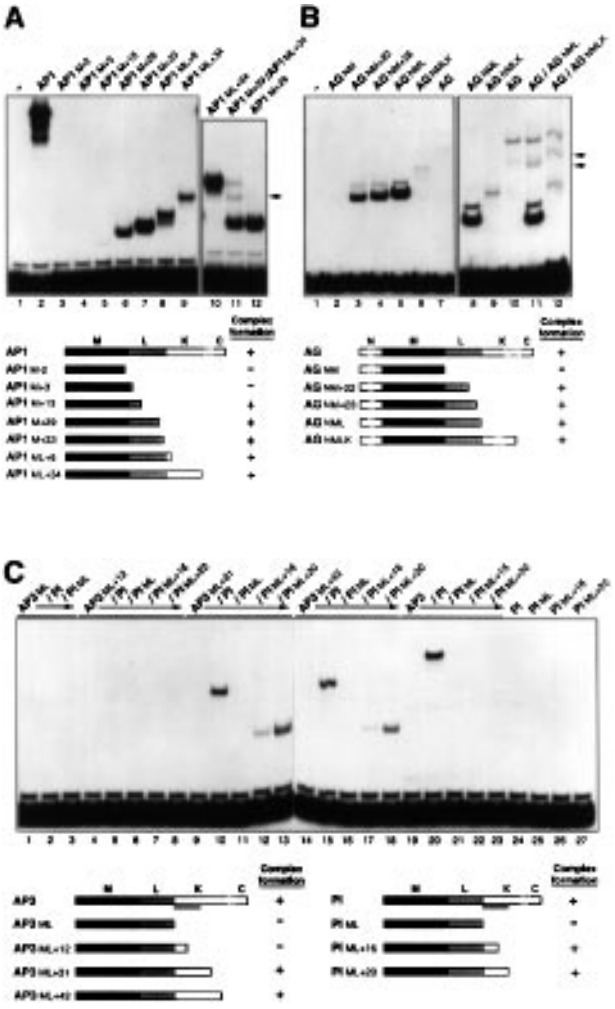

Figure 5. Analyses of C-terminal deletion mutants of AP1, AG, AP3 and PI. (**A**) Electrophoretic mobility shift analysis using C-terminal deletion derivatives of AP1, assayed with probe A. The structure of the different protein variants is represented in a schematic form below the panel, with the M, L, K and C regions indicated. The ability to form a complex with DNA is indicated with a + sign. That AP1 binds to DNA as a dimer is also shown (lanes 10–12; a 8% gel was used): when cotranslated AP1M+29 and AP1ML+34 (lane 11) were assayed, only one new band appeared, corresponding to a heterodimer of both protein forms (indicated by an arrow). (**B**) C-terminal deletion derivatives of AG tested with probe B. Protein–DNA complexes were separated from the free probe on a 7% gel. AG binds to DNA as a dimer (lanes 8–12; a 5% gel was used): AG was mixed with AGNML (lane 11) and with AGNMLK (lane 12), and in both cases only one new band appeared, corresponding to a heterodimer between the full-length and the truncated protein (indicated by an arrow). (**C**) Different combinations of AP3 and PI derivatives tested for DNA-binding with probe A. Protein–DNA complexes were separated from the free probe on 7% gels. The solid line in the diagram represents the first putative amphipathic helix of the K box. The ability to form a heterodimer complexed with DNA is indicated with a + sign.

The ability of N-terminally truncated proteins to dimerize was investigated by immunoprecipitation experiments. AGΔmlkc and AP3Δmlkc (which start at amino acid 26 of the MADS box) were still capable of interacting with AG and PI, respectively, although these interactions are reduced in comparison with those of the full-length AG and AP3 proteins (Fig. 6A). Complete removal of the AG MADS box (AGlkc protein) resulted in a

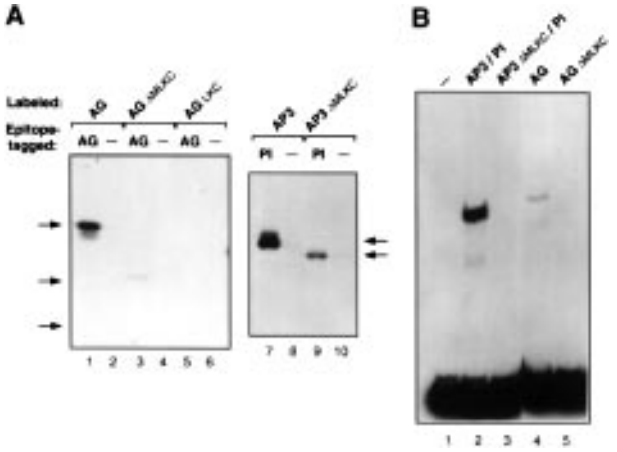

Figure 6. N-terminal deletion analysis of AG and AP3. (**A**) Similar amounts of [^{35}S]methionine-labeled *in vitro* translated AG, AGΔMLKC, and AGLKC were coimmunoprecipitated with epitope-tagged AG; labeled AP3 and AP3ΔMLKC were coimmunoprecipitated with epitope-tagged PI. Reactions with unprogrammed lysate (as control for non-specific precipitation) were included (even-numbered lanes). Arrows indicate the positions to which the different proteins migrate (**B**) Gel shift analysis of AGΔMLKC and AP3ΔMLKC assayed with probe A.

protein incapable of interacting with AG (Fig. 6A), in agreement with previous data showing that the MADS box was required for the interaction between AP3 and PI (7). DNA-binding experiments showed that neither AGΔMLKC nor AP3ΔMLKC–PI complexes could bind to probe A (Fig. 6B), indicating that a dimeric MADS protein complex requires the MADS-domains of both monomers to bind DNA. These results are in agreement with the recently determined crystal structure of SRF bound to DNA, which shows that residues in the N-terminal α-helix of the MADS box are involved in both DNA binding and forming part of the dimerization interface, while residues in the C-terminal half of the MADS box are critical for dimer formation (9).

DISCUSSION

Similarity of the DNA-binding properties of AP1, AP3–PI and AG complexes

AP1, AP3–PI and AG dimers were tested for DNA-binding with seven different CArG-box containing sequences, and of the resulting 21 different protein–DNA combinations only one failed to show DNA-binding, that between AP1 and probe E. Some of the probes used were synthetic binding sites identified in random sequence-selection experiments performed with either AG or AGL3 (17,21), but were nevertheless also bound by AP1 and AP3–PI. These results indicate that the sets of sequences recognized by AP1, AP3–PI and AG dimers are largely overlapping. Moreover, *AGL5* has been proposed to be regulated by AG (19); however, the CArG-box (probe D in this study) that might mediate such regulation is also very efficiently bound by AP1 and AP3–PI. Similarly, a CArG-box present in the *AP3* promoter (probe A), that might be involved in the autoregulation of *AP3* expression by AP3–PI (7,31,32), is also bound by AP1 and AG; and the three complexes recognized the probe derived from the *SUP* promoter (probe B). It is noteworthy that the three probes that are derived from the *Arabidopsis* genome were bound

with much higher affinities than those obtained from sequence-selection experiments, showing that the sequence-selection experiments did not unequivocally identify the highest affinity binding sites, and questioning the biological significance of the consensus sequences that are defined in those experiments. The similarity (or identity) of the sequences recognized by AP1, AP3–PI and AG implies that it would not be feasible to try to identify downstream genes of each particular MADS box protein complex by scanning *Arabidopsis* genomic sequences for CArG motifs. In addition, and most importantly, this similarity raises a question about the *Arabidopsis* MADS domain homeotic proteins that has been asked previously for other transcription factors: how do proteins that recognize the same or very similar sets of binding sites regulate the expression of different groups of downstream genes?

Although AP1, AP3–PI and AG recognize similar sets of target sites, their intrinsic DNA-binding specificities are not identical: differences in the *in vitro* DNA-binding affinities are detected. It is possible that these differences contribute to the biological specificity of these proteins. However, if subtle differences in DNA-binding affinities are, by themselves, the main determinants of the functional specificity of these four homeotic proteins, their concentrations in the cell should be critical and thus finely regulated. The available data, on the contrary, have not revealed a tight link between protein concentration and developmental outcome. First, none of the *ap1*, *ap3*, *pi* or *ag* alleles studied to date has been shown to be a haplo-insufficient mutation with respect to organ identity. In addition, *AG*, *AP3* and *PI* have been ectopically expressed under the control of the constitutive 35S promoter and shown to produce the expected organ identity changes (31–33). These data indicate that the levels of expression of AP1, AP3, PI and AG can be varied within a certain range without affecting their control of organ identity (it remains an open possibility that the level of protein of each gene is also regulated posttranscriptionally). Certain thresholds of homeotic protein concentration or function likely exist: the phenotype that is conferred by the ectopic expression of *AG* or *AP3* can vary in its severity between different transgenic lines, presumably owing to different levels of transgene expression (31,33). Nonetheless, the only functions identified in the ectopic expression experiments are those that are particular to the wild-type expression of each of those genes, and no new or different functions are shown by these proteins in the different transgenic lines. Therefore, the thresholds of protein concentration or function could in part be related to the DNA-binding activity of each of these proteins, but they do not indicate that the specific functions of each protein can be changed by under- or overexpression, as would be expected if subtle DNA affinity differences were responsible for specific functions.

AP1, AP3–PI and AG dimers were found to induce similar degrees of DNA bending toward the minor groove. It is noteworthy that a truncated core AG protein, AG$_{NML}$, induced the same DNA distortion as the full-length AG protein, suggesting that the results obtained in the circular permutation and phasing analysis experiments were not affected by a possible extended shape of the proteins (as has been described in other cases; 34). In addition, the crystal structure of core SRF bound to DNA has recently been determined and showed the DNA bent around the protein (9). The similarity of the conformational changes induced by AP1, AP3–PI and AG dimers suggests that the different regulatory specificities of these three complexes do not arise through the generation of different DNA structures that could direct the formation of transcripton complexes with distinct functional properties. It therefore seems at least possible that the biological specificity of AP1, AP3, PI and AG cannot be explained on the basis of their intrinsic DNA-binding properties alone. Consistent with this interpretation, *in vivo* analyses of the activity of chimeric genes formed by swapping regions between AP1, AP3, PI and AG have shown that, at least in some cases, the MADS domains can be interchanged without them determining the specific functions of the resulting chimeric proteins (35). In addition, we have recently found that the DNA-binding specificity of AP1, AP3, PI and AG can be altered without affecting their functions *in vivo* (J. L. Riechmann and E. M. Meyerowitz, unpublished results). Another possible mechanism by which the MADS domain homeotic proteins could direct the development of different organs is that they may act in conjunction with cofactors that modulate their ability to regulate the transcription of downstream genes. This could be a process in which DNA bending by the plant MADS-domain proteins might be involved, through determining DNA topology in nucleoprotein complexes, allowing interactions with other proteins that may bind to adjacent DNA sites, or facilitating the recognition by accessory proteins of their respective target sites, as has been suggested for SRF (9).

This situation of diverse and highly specific *in vivo* functions by related proteins with similar DNA-binding properties is reminiscent of that encountered for the *Drosophila* homeotic selector proteins. Homeodomain proteins also show very similar intrinsic DNA-binding specificities *in vitro* (with affinities on the order of K_d = 10^{-8}–10^{-9} M) (36). Some differences in the DNA-binding specificities are also detected, which might contribute in part to the functional specificity of the proteins (37). However, the analyses of different mutant and chimeric proteins in ectopic expression experiments have shown that the specificity of action of the homeodomain proteins *in vivo* also depends on protein–protein interactions (38). Examples of direct interactions between the MADS box proteins of animals and fungi and additional cofactors are already abundant. Some of these interactions result in modulation of the MADS box protein activity and a concomitant cell-specific differential gene expression, eventually leading to cell specialization or to different developmental pathways. The yeast MADS domain protein MCM1 is required for transcription in the three yeast cell types, but through interactions with different cofactors (α1 protein, α2 homeodomain protein) it regulates the transcription of cell-type specific genes. Thus the regulatory activities of MCM1 are determined by the availability of accessory proteins in conjunction with the sequence context of the MCM1 binding sites (10,39). The MADS domain protein MEF2A physically interacts with muscle bHLH transcription factors to control the cascade of myogenic development through cooperative activation of muscle gene expression (40,41).

Organization of AP1, AP3, PI and AG proteins

As expected from the high degree of sequence similarity, the organization of the AG, AP3 and PI MADS boxes is similar to that of SRF: the basic N-terminal half is essential for DNA-binding and the C-terminal half is required for dimerization (9,42). Since the MADS box proteins bind to DNA as dimers, the minimal DNA-binding domain includes the conserved 56 aa MADS box, and an additional C-terminal extension, whose

sequence is not conserved throughout the family but is necessary for dimerization. This extension is of ~24 aa in SRF and MCM1 (9,16,42). In addition to the MADS box, the minimal DNA-binding domains of AP1 and AG include extensions of ~20 amino acids (part of the L region), and similar results have been obtained recently with the *Arabidopsis* AGL2 protein, whose core includes the MADS domain and the first 21 aa of the L region (43). On the other hand, core AP3 and PI proteins comprise the entire L region and part of the K box (a total C-terminal extension to the MADS domain of ~50 aa). The involvement of the first amino acids of the K-box in dimerization has also been recently shown for the *Antirrhinum* homologous proteins of AP3 and PI: DEF and GLO, respectively (44). The difference in the size of the core proteins, AG and AP1 on one hand, and AP3 and PI on the other, correlates with the partner specificity that these proteins possess: AG and AP1 form DNA-binding homodimers but not DNA-binding heterodimers with AP3 or PI, which form a DNA-binding AP3–PI heterodimer (20).

Based on the presumptive coiled-coil structure of the K box, and by analogy to leucine zipper proteins, it has been suggested that this region could be involved in promoting dimerization (12). The analysis of C-terminal deletion mutants described here shows that the entire K box (in the case of AP1 and AG), or a substantial part of it (in the case of AP3 and PI), is dispensable for the formation of DNA-binding dimers. It is possible that the K box plays a role in dimer stabilization, but might not be required in the mild conditions used in the DNA-binding experiments. Consistent with this notion, it has been shown that deletion of part of the K box of an epitope-tagged PI protein (a deletion that did not include the region shown here as forming part of the core protein) reduced, but did not abolish, the immunoprecipitation of labeled AP3 protein (7). Alternatively, the K-box could be involved in interactions with additional (unknown) cofactors of the plant MADS box proteins.

In summary, the finding of differences in the organization of the AP1, AG and AP3 and PI proteins, and its correlation with the partner specificity that these proteins exhibit for the formation of DNA-binding dimers (20), support the idea that selective dimerization is part of the mechanism by which these proteins achieve their functional specificity. On the other hand, the DNA-binding activities of these dimers (AP1, AP3–PI and AG) are very similar, suggesting that the biological specificity that these proteins possess may not be explained on the basis of their intrinsic DNA-binding specificity alone. It is likely that at least part of their biological specificity is achieved through selective interactions with additional transcription factors, a mechanism that appears to be a common theme for the MADS box proteins of animals and fungi.

ACKNOWLEDGEMENTS

We are grateful to Hong Ma for providing binding site probes, to Sankar Adhya for pBend2, and to members of the laboratory for valuable comments on the manuscript. This work was supported by US National Science Foundation grant MCB-9204839 to E. M. M.; J. L. R. was supported by a fellowship from Ministerio de Educación y Ciencia (Spain) and was formerly an EMBO Postdoctoral Fellow.

REFERENCES

1 Coen,E.S. and Meyerowitz,E.M. (1991) *Nature*, **353**, 31–37.
2 Ma,H. (1994) *Genes Dev.*, **8**, 745–756.
3 Weigel,D. and Meyerowitz,E.M. (1994) *Cell*, **78**, 203–209.
4 Mandel,M.A., Gustafson-Brown,C., Savidge,B. and Yanofsky,M.F. (1992) *Nature*, **360**, 273–277.
5 Jofuku,K.D., den Boer,B.G.W., van Montagu,M. and Okamuro,J.K. (1994) *Plant Cell*, **6**, 1211–1225.
6 Jack,T., Brockman,L.L. and Meyerowitz,E.M. (1992) *Cell*, **68**, 683–697.
7 Goto,K. and Meyerowitz,E.M. (1994) *Genes Dev.*, **8**, 1548–1560.
8 Yanofsky,M.F., Ma,H., Bowman,J.L., Drews,G.N., Feldmann,K.A. and Meyerowitz,E.M. (1990) *Nature*, **346**, 35–39.
9 Pellegrini,L., Tan,S. and Richmond,T.J. (1995) *Nature*, **376**, 490–498.
10 Shore,P. and Sharrocks,A.D. (1995) *Eur. J. Biochem.*, **229**, 1–13.
11 Ma,H., Yanofsky,M.F. and Meyerowitz,E.M. (1991) *Genes Dev.*, **5**, 484–495.
12 Davies,B. and Schwarz-Sommer,Z. (1994) In Nover,L. (ed.), *Plant Promoters and Transcription Factors*. Springer-Verlag, Berlin, pp. 235–258.
13 Pnueli,L., Abu-Abeid,M., Zamir,D., Nacken,W., Schwarz-Sommer,Z. and Lifschitz,E. (1991) *Plant J.*, **1**, 255–266.
14 Pollock,R. and Treisman,R. (1990) *Nucleic Acids Res.*, **18**, 6197–6204.
15 Wynne,J. and Treisman,R. (1992) *Nucleic Acids Res.*, **20**, 3297–3303.
16 Mueller,C.G.F. and Nordheim,A. (1991) *EMBO J.*, **10**, 4219–4229.
17 Huang,H., Mizukami,Y., Hu,Y. and Ma,H. (1993) *Nucleic Acids Res.*, **21**, 4769–4776.
18 Shiraishi,H., Okada,K. and Shimura,Y. (1993) *Plant J.*, **4**, 385–398.
19 Savidge,B., Rounsley,S.D. and Yanofsky,M.F. (1995) *Plant Cell*, **7**, 721–733.
20 Riechmann,J.L., Krizek,B.A. and Meyerowitz,E.M. (1996) *Proc. Natl. Acad. Sci. USA*, **93**, 4793–4798.
21 Huang,H., Tudor,M., Weiss,C.A., Hu,Y. and Ma,H. (1995) *Plant Mol. Biol.*, **28**, 549–567.
22 Cao,Z., Umek,R.M. and McKnight,S.L. (1991) *Genes Dev.*, **5**, 1538–1552.
23 Kim,J., Zwieb,C., Wu,C., and Adhya,S. (1989) *Gene*, **85**, 15–23.
24 Thompson,J.F. and Landy,A. (1988) *Nucleic Acids Res.*, **16**, 9687–9707.
25 Koo,H.S., Drak,J., Rice,J.A. and Crothers,D.M. (1990) *Biochemistry*, **29**, 4227–4234
26 Wu,H.-M. and Crothers,D.M. (1984) *Nature*, **308**, 509–513.
27 Zinkel,S.S. and Crothers,D.M. (1987) *Nature*, **328**, 178–181.
28 Kerppola,T.K. and Curran,T. (1991) *Cell*, **66**, 317–326.
29 Kerppola,T.K. and Curran,T. (1993) *Mol. Cell. Biol.*, **13**, 5479–5489.
30 Sharrocks,A.D. and Shore,P. (1995) *Nucleic Acids Res.*, **23**, 2442–2449.
31 Jack,T., Fox,G.L. and Meyerowitz,E.M. (1994) *Cell*, **76**, 703–716.
32 Krizek,B.A. and Meyerowitz,E.M. (1996) *Development*, **122**, 11–22.
33 Mizukami,Y. and Ma,H. (1992) *Cell*, **71**, 119–131.
34 Sitlani,A. and Crothers,D.M. (1996) *Proc. Natl. Acad. Sci. USA*, **93**, 3248–3252.
35 Krizek,B.A. and Meyerowitz,E.M. (1996) *Proc. Natl. Acad. Sci. USA*, **93**, 4063–4070.
36 Gehring,W.J., Affolter,M. and Bürglin,T. (1994) *Annu. Rev. Biochem.*, **63**, 487–526.
37 Dessain,S., Gross,C.T., Kuziora,M.A. and McGinnis,W.M. (1992) *EMBO J.*, **11**, 991–1002.
38 Mann,R.S. (1995) *Bioessays*, **17**, 855–863.
39 Johnson,A.D. (1995) *Curr. Opin. Genet. Dev.*, **5**, 552–558.
40 Kaushal,S., Schneider,J.W., Nadal-Ginard,B. and Mahdavi,V. (1994) *Science*, **266**, 1236–1240.
41 Molkentin,J.D., Black,B.L., Martin,J.F. and Olson,E.N. (1995) *Cell*, **83**, 1125–1136.
42 Norman,C., Runswick,M., Pollock,R. and Treisman, R. (1988) *Cell*, **55**, 989–1003.
43 Huang,H., Tudor,M., Su,T., Zhang,Y., Hu,Y. and Ma,H. (1996) *Plant Cell*, **8**, 81–94.
44 Zachgo,S., de Andrade Silva,E., Motte,P., Tröbner,W., Saedler,H. and Schwarz-Sommer,Z. (1995) *Development*, **121**, 2861–2875.