DNA binding and dimerisation determinants of Antirrhinum majus MADS-box transcription factors

Adam G. West*, Barry E. Causier¹, Brendan Davies¹ and Andrew D. Sharrocks*

Department of Biochemistry and Genetics, The Medical School, University of Newcastle upon Tyne, Newcastle upon Tyne NE2 4HH, UK and 1 Leeds Institute for Plant Biotechnology and Agriculture, Centre for Plant Sciences, University of Leeds, Leeds LS2 9JT, UK

Received August 31, 1998; Revised and Accepted October 20, 1998

ABSTRACT

Members of the MADS-box family of transcription factors are found in eukaryotes ranging from yeast to humans. In plants, MADS-box proteins regulate several developmental processes including flower, fruit and root development. We have investigated the DNA-binding mechanisms used by four such proteins in Antirrhinum majus, SQUA, PLE, DEF and GLO. SQUA differs from the characterised mammalian and yeast MADS-box proteins as it can efficiently bind two different classes of DNA-binding site. SQUA induces bending of these binding sites and the sequence of the site plays a role in determining the magnitude of these bends. Similarly, PLE and DEF/GLO induce DNA bending although the direction of the resulting bends differ. Finally, we demonstrate that the MADS-box and I-domains are sufficient for homodimer formation by SQUA. However, the K-box in SQUA can also act as an oligomerisation motif and in the full-length protein, the K-box plays a different role in mediating dimerisation in the context of SQUA homodimers or heterodimers with PLE. Together these results contribute significantly to our understanding of the function of SQUA and other plant MADS-box proteins at the molecular level.

INTRODUCTION

MADS-box transcription factors have important roles in regulating diverse cellular processes in organisms ranging from yeast to humans (reviewed in 1). In plants, MADS-box proteins are involved in regulating a range of developmental events including floral organ and fruit development (reviewed in 2) and lateral root initiation in response to environmental conditions (3). In Antirrhinum majus, the MADS-box proteins SQUA, PLE, DEF and GLO are involved in determining meristem and floral organ identity. Mutations of genes encoding these proteins (SQUAMOSA, PLENA, DEFICENS and GLOBOSA, respectively) lead to homoeotic transformations such as the conversion of floral meristems into inflorescence meristems in SQUA mutants (4). Similarly, in Arabidopsis thaliana, homologues of these proteins exist, API (SQUA), AG (PLE), AP3 (DEF) and PI (GLO), which share significant amino acid identity and perform similar developmental functions (reviewed in 2).

To date, the analysis of the function of MADS-box proteins at the molecular level has focused mainly on the human and yeast family members (reviewed in 1). The recent elucidation of the structure of the DNA-binding domains of SRF and Mcm1 bound to DNA has further illuminated the molecular functions of this family of transcription factors (5,6). The minimal DNA-binding domains of the yeast and human MADS-box proteins are composed of the C-terminal MADS-box and a 25–30 amino acid C-terminal extension. Together these motifs are sufficient to direct dimerisation and sequence-specific DNA binding (7; reviewed in 1). MADS-box proteins bind to one of two classes of binding site based on the central consensus motifs 5′-CC(A/T)₆GG-3′ (SRE-like) and 5′-CTA(A/T)₆TAG-3′ (N10-like) (reviewed in 1). For example, SRF binds efficiently to SRE-like sites but only weakly to N10-like sites (8) whereas MEF2A binds in a reciprocal manner and exhibits negligible binding to SRE-like sites (9). Furthermore, although proteins such as SRF and Mcm1 are able to greatly distort their binding sites (5,6,10–12), other proteins such as MEF2A, Rml1 and Smp1 exhibit a much reduced propensity for inducing DNA bending (7,13).

Currently, little is known about in vivo target sites for plant MADS-box proteins (reviewed in 2). However, DNA-binding site-selection studies for the A.thaliana protein AG suggest that it binds sequences related to the SRE-like class of sites (14). Similarly, several Antirrhinum MADS-box proteins have been shown to bind to this class of sites by using various SRE-like binding sites (15–17). Currently, there is no evidence of plant MADS-box proteins which bind exclusively to the N10-like class of binding site. However, AGL2 exhibits a wider binding specificity than other Arabidopsis MADS-box proteins such as AG and AGL1 (18). Protein-induced DNA bending has not been demonstrated by Antirrhinum MADS-box proteins although the Arabidopsis proteins API, AG and AP3/PI all cause directional DNA bending (19).

The identification of a large number of MADS-box proteins in plant species, coupled with their propensity for homo- and heterodimer formation, suggests a high potential for the establishment of complex regulatory circuits. It is currently unclear how homodimer and heterodimer formation are mediated in all cases although the MADS-box and a short, C-terminal extension (I-domain) are sufficient for homodimer formation by the Arabidopsis proteins

*To whom correspondence should be addressed. Tel: +44 191 222 8800; Fax: +44 191 222 7424; Email: a.d.sharrocks@ncl.ac.uk

*Present address: Laboratory of Molecular Biology, NIDDK, 5-218, NIH, 9000 Rockville Pike, Bethesda, MD 20892, USA
API and AG (20). In other cases, such as heterodimer formation between AP3/AG (20), an additional conserved domain known as the K-box is required (20). In Antirrhinum, the situation is less clear although the K-box has been implicated in mediating interactions amongst MADS-box proteins (17,21).

In this study, we have investigated the molecular mechanisms by which Antirrhinum MADS-box proteins achieve sequence-specific DNA binding. In particular we have focused on SQUA and its DNA binding and dimerisation mechanisms in comparison to other MADS-box proteins. In contrast to mammalian and yeast MADS-box proteins, SQUA exhibits a broad DNA-binding specificity and can bind to two distinct classes of binding site. SQUA and PLE homodimers and DEF/GLO heterodimers distort DNA although the direction of the resulting bend differs. Dimers of these proteins also differ and the K-box appears to play a different role in individual MADS-box proteins. Together our results further our understanding of how plant MADS-box proteins function at the molecular level.

MATERIALS AND METHODS
Plasmid constructions and mutagenesis
The pBluescript-KS+-derived plasmids, pAS68 (encoding coreMEF2A; amino acids 1–86) (22), pAS1 (encoding coreSRF; amino acids 132–222), pAS37 (encoding METcoreSRF; amino acids 142–222) (23) and pAS701 (encoding coreSQUA/SQUAM1; amino acids 1–92) (7) have been described previously. pTZSQUA, pTZPLE, pTZDEF and pTZGLO encode full-length SQUA, PLE, DEF and GLO proteins, respectively (17). pAS702, pAS756, pAS757, pAS758 and pAS759 (encoding SQUAMIK, amino acids 1–170; DEFMI, amino acids 1–92; GLOMI, amino acids 1–92; PLEMI, amino acids 1–105; and PLEMIK, amino acids 1–170, respectively) were produced by ligating HindIII and EcoRI digested PCR fragments (primer pairs ADS186/ADS425, ADS186/ADS503, ADS186/ADS504, ADS186/ADS505 and ADS186/ADS506 using the templates pTZSQUA, pTZPLE, pTZGLO and pTZPLE) into pAS37 cleaved with the same enzymes.

The phasing vectors pAS525–529 containing the N10 consensus DNA sequence (5′-CCATATTAGG-3′) (7), pAS76 (c-fos SRE, central sequence; 5′-CCATATTAGG-3′) (7) and pAS152 (N10 site, central sequence; 5′-CTATTATTAG-3′) (13) have been described previously. The phasing vectors pAS525–529 containing the N10 site phased over one helical turn from intrinsically bent poly A:T tracts have been described previously (7). These vectors contain four phased A:T tracts with distances between the centre of the last A:T tract and the centre of the N10-binding site of 30, 32, 35, 37 and 40 bp.

Details of PCR primers and mutagenic oligonucleotides can be supplied upon request. The sequences of all plasmids encoding mutant proteins and PCR-derived sequences were confirmed by automated or manual dideoxy sequencing.

Protein production and immunoprecipitation analysis
Wild-type and mutant MADS-box proteins were produced by sequential in vitro transcription and translation and subsequently analysed and quantified as described previously (13). Immunoprecipitations were carried out using anti-Flag antibody (Kodak) coupled to protein-G beads (0.5 µl antibody and 20 µl 50% v/v protein-G beads per assay) in a total volume of 100 µl 1× PBS. Binding of 35S-labelled in vitro translated proteins was carried out at room temperature for 4 h followed by three washes in 500 µl 1× PBS. Remaining bound proteins were eluted from the beads and analysed by SDS–PAGE.

Gel retardation, circular permutation and phasing analysis
Gel retardation assays were carried out essentially as described previously (11,13) on the c-fos SRE (22), the N10 site (23) or the QPPAL sites (7). Binding reactions were allowed to reach equilibrium by incubating with DNA-binding sites for between 30 and 45 min. Relative DNA-binding affinities were calculated by PhosphorImager analysis of DNA–protein complexes (FUJI BAS1500; TINA 2.08e software). Experiments were carried out to achieve ≤50% of total DNA binding in protein–DNA complexes. Under these conditions, relative binding affinities within an experiment can be compared by direct quantification of DNA–protein complexes.

For circular permutation analysis, DNA fragments were produced by appropriate restriction enzyme digestion of PCR products derived from the vectors pAS76 (containing the c-fos SRE), pAS152 (N10) and pAS724 (QPPAL) followed by gel purification as described previously (11). Circular permutation analysis was carried out on 5% polyacrylamide gels cast in 1× TBE. Curve fitting and apparent DNA bend angles were calculated as described previously (13). Bend angles are not absolute but are comparative under the conditions used and are referred to as apparent bend angles. Bend angles are quoted as the average of three independent experiments. Standard deviations (n = 1) of bend angles are in the range 0.5–1.6°.

For phasing analysis, DNA fragments were produced by PCR amplification of the inserts from pAS525–529 with the primer pair ADS262/ADS346. These vectors were designed to optimise parameters based on the results of a recent study (25) to allow accurate detection of DNA bends (7). Relative magnitudes of the DNA bends are proportional to the amplitude of the phasing function (7,26). Binding sites were purified and phasing experiments were carried out as described for the circular permutation experiments. Data are shown fitted to a cosine function as described previously (7).
All figures were generated electronically from either Phosphor-Imager files or scanned images of autoradiographic images using Picture Publisher (Micrografx) and PowerPoint (Microsoft) software. Final images are representative of the original autoradiographic images.

DNA-binding site selection procedure

DNA-binding sites were selected from a pool of random double-stranded oligonucleotides (containing 26 random bases) essentially as detailed previously (27). Flag epitope-tagged SQUAMIK was synthesised by in vitro transcription/translation using the template pAST798. An aliquot of 3 µl of this protein was used in binding reactions to the original and subsequently amplified DNA pools as described previously (27). DNA binding and washing were carried out under the same buffer conditions as gel retardation analysis except that poly[dI-dC] was omitted. Bound sites were recovered by immunoprecipitation with an anti-Flag antibody as described above. Following four rounds of selection, bound DNA pools were amplified and ligated into EcoRV-cleaved pBluescript (Stratagene).

RESULTS

Mapping the minimal DNA-binding domains of Antirrhinum MADS-box proteins

High affinity, sequence-specific DNA binding by MADS-box proteins is usually mediated by a domain containing the MADS-box and a C-terminal extension (reviewed in 1). Structural studies on SRF and Mcm1 indicate that residues from within both the MADS-box and the C-terminal extension contribute to the dimerisation interface (5,6). In plants, an additional C-terminal domain known as the K-box is believed to contribute to this process (reviewed in 2). C-terminal truncations were made of the Antirrhinum MADS-box protein SQUA in order to map its minimal DNA-binding domain. Proteins were tested for binding to three different types of MADS-box protein binding site (Fig. 1D). Minimal DNA-binding domain. Proteins were tested for binding MADS-box protein SQUA in order to map its

Figure 1. DNA binding by Antirrhinum MADS-box proteins. (A) Schematic illustration of the domain structure of the proteins used in (B) and (C). Numbers indicate the positions of residues with respect to the full-length proteins. (B) Gel retardation analysis of wild-type and truncated SQUA proteins complexed with the N10 (lanes 1–3), c-fos SRE (lanes 4–6) or palindromic PAL (lanes 7–9) binding sites. Equal molar quantities of SQUAMI (lanes 1, 4 and 7), SQUAMIK (lanes 2, 5 and 8) and full-length SQUA (lanes 3, 6 and 9) are used on each site to demonstrate relative binding affinities. (C) Gel retardation analysis of C-terminally truncated DEF, GLO and PLE proteins. In vitro co-translated DEFMI/GLOMI (lane 1) and in vitro translated PLEMI (lane 2) and PLEMIK (lane 3) were complexed with the palindromic PAL binding site. (D) Sequences of the central 18 bp motifs of the c-fos SRE, N10 and PAL binding sites. Base residue numbers follow the nomenclature of Pellegrini et al. (5). The 10 bp CATG-box motif is boxed.

full-length proteins efficiently form DNA-bound heterodimers (e.g. Fig. 3C). With the exception of SQUA, efficient binding to the N10 site by various full-length and truncated homo- and heterodimer combinations was not detectable (data not shown), indicating that amongst these proteins, SQUA may possess a distinct binding specificity.

Together these results demonstrate that the core DNA-binding domains of SQUA and PLE are sufficient to permit sequence-specific DNA binding by these proteins.

SQUA possesses ‘dual site’ DNA-binding specificity

Generally, MADS-box proteins bind to one of two types of site related to either the c-fos SRE or N10 site (1). However, SQUA can bind to both of these sites with high efficiency (Fig. 1). We therefore carried out a DNA-binding site selection experiment...
Figure 2. Identification of SQUA-binding sites. (A) Alignment of the DNA-binding sites selected by SQUAMIK. Sites are grouped according to the presence of the consensus sequences 5′-CTA(A/T)4TAG-3′ (‘MEF2-like’) and 5′-CC(A/T)6GG-3′ (‘SRF-like’). Sites conforming to neither of these consensus sequences were classified as ‘intermediate’ and further subclassified according to the presence of an exclusively A/T-rich 6 bp central region (class I) or with single G-C base pairs within this region (class II). The central 10 bp of the binding sites are shown in bold and boxed. Upper and lower case letters represent bases derived from the random and constant parts, respectively, of the random binding site oligonucleotides. A further 10 sequences were analysed which did not contain obvious binding sites. Binding of SQUA to a selection of these sites could not be detected (data not shown). The summary and consensus represent compilations of all classes of binding sites. Nucleotides present in >50% of sites at a given position are shown as upper case letters. A second lower case letter indicates that this nucleotide is present in >80% of the remaining sites. Asterisks indicate that these positions are strongly influenced by the non-random primer sequences. (B) Gel retardation analysis of SQUAMIK (lanes 1–7), core-MEF2A (lanes 8–14) and core-SRF (lanes 15–21) binding to a series of the selected sites. Equal molar quantities of each in vitro translated protein and DNA site were included in the binding reactions. The identities of the binding sites used are indicated above each lane. (C) Comparison of the relative binding efficiencies of SQUAMI and SQUAMIK to a series of the selected sites. Data are presented relative to the binding of each protein to site 12 (taken as 100%).

with SQUA from a random double-stranded oligonucleotide pool to identify the spectrum of sites to which it can bind. The sequences of the sites selected after four rounds of selection can be categorised into three major groups (Fig. 2A); ‘MEF2-like’ [5′-CTA(A/T)4TAG-3′], ‘SRF-like’ [5′-CC(A/T)6GG-3′] and ‘intermediate’ sites which do not obviously conform to these two consensus sites. The latter class can be further subdivided into those which contain either an exclusively A/T-rich central 6 bp core (class I) or those with single G-C base pairs within this region (class II). Binding of SQUAMIK and the minimal DNA-binding domains of MEF2A and SRF to a series of the selected sites was compared (Fig. 2B). SQUAMIK binds to all of the sites in this series, although its affinity for each site differs (Fig. 2B, lanes 1–7). In contrast, MEF2 and SRF only bind efficiently to a subset of these sites with maximal binding in both cases being observed to sites which conform to their respective consensus binding motifs (Fig. 2B, lanes 8–21). In order to investigate a possible role for the K-box in modifying the DNA-binding specificity of SQUA, the binding of SQUAMI and SQUAMIK to the same series of sites was compared (Fig. 3C). Generally, SQUAMI bound to all the sites with lower efficiency but large differences in their relative affinities for individual sites are observed. For example, site 25 shows an 80% reduction whereas site 8 shows only a 15% reduction in SQUA binding upon deletion of the K-box, although these sites are bound with comparable efficiency by SQUAMIK. This suggests a novel role for the K-box in relaxing the DNA-binding specificity of SQUA.

These results therefore indicate that SQUA exhibits a wider DNA-binding specificity than most other MADS-box proteins and binds to sites which are not efficiently bound by MEF2A and SRF.

DNA binding by Antirrhinum MADS-box proteins

DNA binding is often accompanied by protein-induced DNA distortion. Indeed, the human MADS-box protein SRF induces considerable bending in its binding site (5,10,11). However, in comparison, other family members such as MEF2A induce minimal DNA distortion (13). We therefore investigated whether
Antirrhinum MADS-box proteins also induce DNA bending. Circular permutation analysis demonstrates that SQUA homodimers induce considerable DNA distortion of the N10 site (57°; Fig. 3C). In the case of SRF, the minimal 92 amino acid core DNA-binding domain is sufficient for inducing DNA bending (5, 11). Similarly, a truncated SQUA protein (SQUAMIK, Fig. 3D) and the minimal core DNA-binding domain of SQUA (SQUAMI; 7) are sufficient to induce increased DNA flexibility.

SRF has been demonstrated to bend DNA in a sequence-independent manner (13), whereas the magnitude of bends induced by Mcm1 are influenced by the sequence of its binding site (7, 30). As SQUA binds efficiently to a wide range of binding sites (Figs 1 and 2), its ability to bend sites containing different sequences was tested (Fig. 3D). Interestingly, the magnitude of induced bends differed according to the binding site used. The magnitude of the DNA bends induced by SQUA at these sites was c-fos SRE < N10 < PAL (Fig. 3D), thereby demonstrating a role for the sequence of the DNA-binding site in determining the ability of SQUA to bend DNA.

The results of circular permutation assays can be influenced by factors other than DNA bending such as protein shape (reviewed in 31) and do not provide information regarding the direction of DNA bending. However, in the case of other MADS-box proteins, this assay has been demonstrated to provide reliable estimates of the relative magnitude of bending induced by these proteins (13). In order to confirm that SQUA homodimers induce directional DNA bending, phasing analysis was used. This assay detects protein-induced bends by their effects on electrophoretic mobility (cooperative or antagonistic) as the protein binding site is moved throughout one helical turn from an intrinsic bend generated by a poly A:T tract (reviewed in 32). Full-length and C-terminally truncated SQUA proteins induce bending of the N10 binding site. In all cases, the overall direction of bending is virtually identical and is towards the minor groove (Fig. 4B). In comparison to the full-length protein, the amplitude of the phasing function is larger for the truncated proteins. This may reflect that the value for the full-length protein is an underestimation due to poor resolution of the larger complexes. Alternatively,

Figure 3. Circular permutation analysis of DNA bending by the Antirrhinum MADS-box proteins SQUA, PLE and DEF/GLO. (A) Schematic illustration of the domain structure of the proteins used in (C) and (D). Numbers indicate the size of each full-length protein. (B) Diagrammatic representation of DNA fragments generated by restriction digestion of pBEND2-derived vectors. Protein binding sites are represented by open boxes. Probes were generated by digestion with MluI (Ml), BglII (Bg), XhoI (Xh), EcoRV (EV), SmalI (Sm), StulI (St), RsaI (Rs) or BamHI (BH). (C) Gel-retardation analysis of full-length SQUA (lanes 1–7), C-terminally truncated PLE (PLEMI, lanes 8–15) and full-length DEF/GLO heterodimers (lanes 16–22) complexed with circularly permuted probes containing the N10 (lanes 1–7), PAL (lanes 8–15) or c-fos SRE (lanes 16–22) binding sites respectively. The following probes were used: Ml (lanes 1, 8 and 16), BgI (lanes 2, 9 and 17), XhI (lanes 3, 10 and 18), EV (lanes 4, 11 and 19), SmI (lanes 5, 12 and 20), StI (lane 13), Rs (lanes 6, 14 and 21) and BH (lanes 7, 15 and 22). The data from the circular permutation experiments are shown graphically beneath each set of primary data. The relative mobilities of protein–DNA complexes were normalised for differences in probe mobilities and plotted as a function of the position of the centre of the N10, PAL or c-fos SRE binding site from the 5′-end of the probe. The points are connected by a curve of the best fit to a cosine circular permutation function. Error bars represent standard deviations calculated from at least three independent experiments. The correlations (R²) between relative mobility data and the model circular permutation function calculated from curve fitting analyses were >0.95. The apparent bend angles calculated from these data are shown within each graph. (D) Gel retardation analysis of the truncated SQUA protein, SQUAMIK, complexed with each of the circularly permuted probes containing either the c-fos SRE (lanes 1–8), N10 (lanes 9–16) or palindromic PAL binding sites (lanes 17–24). The data from each circular permutation experiment are shown graphically beneath each set of primary data as described above. The following probes were used: Ml (lanes 1, 9 and 17), BgI (lanes 2, 10 and 18), XhI (lanes 3, 11 and 19), EV (lanes 4, 12 and 20), SmI (lanes 5, 13 and 21), StI (lanes 6, 14 and 22), Rs (lanes 7, 15 and 23) and BH (lanes 8, 16 and 24). The apparent bend angles calculated from these data are shown within each graph.
additional domains in the C-terminus of SQUA may influence the magnitude of DNA bending as observed with some bZIP proteins (33). However, it is clear that both full-length and truncated versions of SQUA induce directional DNA bending.

The ability of other Antirrhinum MADS-box proteins to bind DNA was also tested. Circular permutation analysis demonstrates that PLE homodimers and DEF/GLO heterodimers both induce substantial DNA distortion (82° and 47°; Fig. 3C). As observed with SQUA, the minimal DNA-binding domain of PLE (PLEMI, Fig. 3C) is sufficient to induce DNA bending. Phasing analysis demonstrates that DEF/GLO heterodimers induce directional DNA bending (Fig. 4C). However, when compared to SQUA, maximal cooperativity of the bends is induced with a longer spacer length between the intrinsic and protein-induced bends. Maximal cooperativity is displaced by 9 bp indicating that the direction of bending is displaced by ∼51° when viewed down the helical DNA axis. Together, these results demonstrate that considerable DNA distortion is induced in SQUA, PLE and DEF/GLO complexes. In the cases of SQUA and DEF/GLO, this distortion results in a directional DNA bend.

**Dimerisation determinants of Antirrhinum MADS-box proteins**

The minimal DNA-binding domain of SQUA (SQUAMI) is sufficient to permit high affinity sequence-specific binding (Fig. 1) and induce directional DNA bending (Fig. 4). As dimerisation is a prerequisite for DNA binding by MADS-box proteins, this result implies that the MADS-box and a short C-terminal extension (I-region) of SQUA are sufficient to mediate dimerisation by SQUA. The K-domain of plant MADS-box proteins has been proposed to represent a dimerisation motif (reviewed in 1,2). In order to assess the roles of these domains in homodimerisation by SQUA, DNA-binding assays were carried out with various combinations of full-length and truncated SQUA derivatives (Fig. 5). In this assay, the ability of two different sized proteins to form dimers is detectable by the production of either a new DNA-bound complex of intermediate mobility or a reduction in the intensity of existing complexes due to the production of a dimer which is incapable of DNA binding. SQUAMIK, SQUAMI and full-length SQUA all efficiently bind DNA and hence form homodimers (Figs 1B and 5B). When SQUAMI is mixed, or co-translated, with the full-length protein or the truncated SQUAMI derivative, no change in either the intensity or mobilities of the complexes is observed (Fig. 5B, lanes 1–4). In contrast, when SQUAMIK and the full-length protein were co-translated, a novel complex of intermediate mobility is generated (Fig. 5B, lane 5). Mixing of the proteins after translation is not sufficient to generate this complex (Fig. 5B, lane 6). These results therefore indicate that two K-boxes, one from each monomer, are required to allow homodimerisation of the full-length protein. The presence of a single K-box in one monomer precludes binding with a second SQUA monomer which lacks this motif.

The role of SQUA domains in directing heterodimer formation with other MADS-box proteins was subsequently analysed. Initially, heterodimerisation between the truncated SQUA derivative SQUAMIK and full-length DEF, GLO and PLE proteins was analysed (Fig. 6B). Efficient formation of DNA-binding homodimers by full-length PLE and DEF occurs on the PAL site although both bind with greatly reduced affinity on other sites such as the c-fos SRE (data not shown). Of these proteins, only heterodimers between PLE and SQUA could be detected (Fig. 6B, lane 7). As the K-box was shown to be required for homodimerisation with either SQUAMIK or full-length SQUA, the ability of SQUAMI (which lacks the K-box; Fig. 5A) to form heterodimers with PLE was tested. SQUAMI forms DNA-bound heterodimers with PLE (Fig. 6B, lane 1) but is unable to form heterodimers with DEF or GLO (data not shown). The K-box is therefore dispensable for dictating heterodimer formation with PLE and the MADS-box and I-region are sufficient to dictate heterodimer formation. This is in direct contrast to the requirement of the K-box for SQUA homodimerisation. These results therefore suggest a different requirement for the SQUA K-box in homodimerisation of SQUA (where it is required) and heterodimerisation with PLE (where it is not required).

The ability of the K-box from SQUA to act as a dimerisation motif was therefore analysed in a heterologous context and its interactions compared with the K-box from PLE. Chimeric proteins were created in which either the K-box alone or the
Figure 5. Dimerisation determinants of SQUA. (A) Schematic illustration of the domain structure of the SQUA proteins used in (B). Numbers indicate the positions of residues with respect to the full-length protein. (B) Gel retardation analysis of complexes formed between short- and long-form SQUA proteins and the PAL binding site. Proteins were either co-translated (c, lanes 1, 3 and 5) or mixed after translation (m, lanes 2, 4 and 6). The ratio of short- to long-form SQUA proteins is 2:1 (lanes 1 and 2) and 1:1 (lanes 3–6). Equal molar amounts of in vitro translated proteins were added in each pair of lanes (lanes 1 and 2, 3 and 4 and 5 and 6). The open, shaded and black arrows indicate complexes formed by full-length SQUA, MIK and MI homodimers, respectively. The position of complexes containing heterodimers is indicated by an asterisk.

K-boxes and the C-terminal region of SQUA or PLE were fused to the MADS-box of MEF2A (Fig. 7A). DNA binding was undetectable by all of these chimeric proteins (Fig. 7B; data not shown), indicating that the K-box is not sufficient to promote the formation of dimers which are competent in DNA binding. This is in contrast to the inclusion of a leucine zipper as an alternative dimerisation motif for MEF2A which promotes the formation of DNA-binding competent dimers (13).

The ability of the chimeric MEF2 proteins to bind to full-length SQUA was subsequently analysed. Full-length SQUA was either co-translated or mixed after translation with the chimeric MEF proteins. As these chimeric MEF2 proteins are unable to bind to DNA, the loss of DNA binding by SQUA homodimers is taken as indicative of the formation of heterodimers which are incapable of binding DNA (13,34). Upon co-translation of equimolar amounts of the MEF chimeras and full-length SQUA, a reduction in the amount of DNA-bound SQUA homodimers is detected (data not shown). When a 10-fold molar excess of the MEF chimeras was used, a substantial reduction in SQUA binding was observed in all cases (Fig. 7B, lanes 2, 4, 6 and 8). These results indicate that the K-boxes from SQUA and PLE are sufficient to permit dimer formation with SQUA but are insufficient to direct DNA binding in an analogous manner to the I-regions of these proteins or other classical dimerisation motifs like the leucine zipper.

In order to directly demonstrate interactions between chimeric MEF2/K-box proteins and SQUA, immunoprecipitation experiments were carried out using epitope-tagged SQUAMIK (Fig. 7C). Proteins were either translated individually (lanes 7–9), co-translated (lanes 3 and 4) or mixed following translation (lanes 5 and 6). When translated individually or mixed with epitope-tagged SQUAMIK following translation, only background levels of MEF:SQUAK and MEF:PLEK were precipitated (Fig. 7C, lanes 5–8). In contrast, both MEF:SQUAK and MEF:PLEK were efficiently precipitated when co-translated with epitope-tagged SQUAMIK (Fig. 7C, lanes 3 and 4). Together, these results therefore demonstrate that the K-boxes from SQUA and PLE are sufficient to mediate dimer formation with SQUA.

DISCUSSION

The Antirrhinum MADS-box transcription factors SQUA, PLE, DEF and GLO play important roles in regulating floral development (reviewed in 2). In this study, we have focused on SQUA and investigated its DNA-binding properties in comparison to the other proteins in order to further our understanding of their function at the molecular level (summarised in Table 1). Our results demonstrate differences in the DNA-binding specificity, the mechanisms of DNA bending and modes of dimerisation amongst this group of MADS-box proteins.
Figure 7. The K-boxes of SQUA and PLE act as dimerisation motifs. (A) Schematic illustration of the domain structure of the proteins used in (B) and (C). Numbers indicate the positions of residues with respect to the full-length proteins. (B) Heterodimerisation of chimeric MEF:SQUA and MEF:PLE proteins and full-length SQUA. Gel-retardation analysis of dimeric complexes binding to the palindromic PAL binding site. Proteins were either mixed after translation (m; lanes 1, 3, 5 and 7) or co-translated (c; lanes 2, 4, 6 and 8). The ratio of chimeric to full-length SQUA proteins is 10:1 (lanes 1–8). Equal molar amounts of in vitro translated proteins were added in each pair of lanes. The position of complexes containing SQUA homodimers is indicated by an arrow. (C) Immunoprecipitation analysis of FLAG epitope-tagged SQUAMIK and chimeric MEF:SQUA and MEF:PLE proteins. Proteins were either co-translated (c; lanes 3 and 4), mixed after translation (m; lanes 5 and 6) or translated individually (lanes 7–9). 20% of the input protein is shown in lanes 1 and 2. The immunoprecipitated epitope-tagged SQUAMIK and MEF:PLE/SQUA chimeras are indicated by open and closed arrows, respectively.

Table 1. Summary of the DNA binding and dimerisation properties of Antirrhinum MADS-box proteins determined in this study

<table>
<thead>
<tr>
<th>Protein</th>
<th>DNA-binding sites</th>
<th>Induction of DNA bending</th>
<th>K-box requirement</th>
</tr>
</thead>
<tbody>
<tr>
<td>SQUA</td>
<td>N10, c-fos SRE</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>PLE</td>
<td>PAL, [PLE]</td>
<td>Yes</td>
<td>Yes [PLE]</td>
</tr>
<tr>
<td>DEF</td>
<td>PAL, c-fos SRE</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>GLO</td>
<td>None</td>
<td>ND</td>
<td>ND</td>
</tr>
</tbody>
</table>

*The DNA-binding sites tested were: N10, c-fos SRE and PAL sites.

All combinations of binding sites and possible partner proteins were not tested. For SQUA/PLE, only binding to the PAL site was tested and for DEF/GLO, binding to the N10 (data not shown) and c-fos SRE sites were tested.

The direction of DNA bending for DEF/GLO heterodimers differs from that observed with SQUA homodimers.

ND, not determined. Heterodimer partners tested are shown in brackets.

Minimal DNA-binding domains of Antirrhinum MADS-box proteins

The minimal ‘core’ DNA-binding domains of the human and yeast MADS-box proteins have been mapped to a region encompassing the MADS-box and a short ~25–30 amino acid C-terminal extension (reviewed in 1,7). In Arabidopsis, similar constructs containing the MADS-box and C-terminal I-region constitute the minimal DNA-binding domains of AP1 (19) and
AG (19,35). Minimal core DNA-binding domains are also present in the related Antirrhinum proteins SQUA and PLE (Fig. 1). In contrast, the core DNA-binding domains of GLO and DEF (Fig. 1) and their Arabidopsis homologues PI and AP3 (19) are insufficient for DNA binding either as homodimers or heterodimers. Extra C-terminal amino acids encompassing the K-box are required to permit DNA binding by DEF/GLO heterodimers (see below).

DNA-binding specificities of MADS-box transcription factors

MADS-box proteins bind two classes of binding site based on the central consensus motifs 5′-CC(A/T)6GG-3′ (SRE-like) and 5′-CTA(A/T)4TAG-3′ (N10-like) (reviewed in 1). Proteins like SRF and Mcm1 bind efficiently to SRE-like sites but weakly to N10-like sites. The reciprocal situation occurs with MEF2, Rlm1 and Smp1 which only efficiently bind to N10-like sites (reviewed in 1; 35). Binding of the Antirrhinum MADS-box proteins PLE and DEF/GLO can only be detected on SRE-like sites and in the case of DEF homodimers, efficient DNA binding can only be observed on the palindromic PAL site (Fig. 1; data not shown). Similarly, the PLE homologue in Arabidopsis, AG, selects binding sites in the SRE-like class (14). In contrast, SQUA efficiently binds to both SRE-like and N10-like sites (Figs 1 and 2). Moreover, SQUA binds to sites which do not represent binding sites for SRF and MEF2A, indicating that it also recognises novel sites. SQUA therefore differs from other mammalian and yeast MADS-box proteins as it possesses ‘dual’ DNA-binding specificity and indicates that it has the potential to bind to a wide variety of promoter sites in vivo. Similarly, a DNA-binding site-selection study on the Arabidopsis protein AGL2 indicates that this protein can also bind to a wide spectra of sites, including SRF-like and MEF2-like sites (18). However, to date, no in vivo binding sites for SQUA have been identified. Our determination of the binding specificity of SQUA will facilitate the identification of SQUA-regulated promoters. Interestingly, SQUA preferentially selects symmetrically oriented T-A base pairs located outside the central 10 bp of the binding site at the ±7 position (Fig. 2A). Rlm1, Smp1, Mcm1 and AG also show a preference for T-A base pairs at this position (14,36,37) whereas SRF and MEF2A do not (8,9). The presence of a T-A base pair at this position also correlates with the ability of SQUA (7; see further discussion below) and Mcm1 (30) to bend DNA. This might reflect a link between DNA binding specificity determination and the propensity of a site to be distorted. A similar hypothesis has been proposed for the establishment of the different binding specificities of SRF and MEF2A but in this case, it is the sequence of the proteins rather than the DNA which determines the magnitude of DNA bending (13).

Previous studies indicate that a combination of N-terminal extensions to the MADS-box and the identity of residue 14 in the MADS-box play significant roles in determining the different DNA-binding specificities of MADS-box proteins (22,38). In SRF, deletion of the N-terminal MADS-box extension relaxes its DNA-binding specificity, allowing it to efficiently recognise N10-like sites. The introduction of the mutation K154E at position 14 further alters the selectivity of SRF towards N10-like sites. SQUA lacks an N-terminal extension and contains a lysine residue at position 14 and therefore resembles the N-terminally truncated SRF protein. Similarly, AGL2 also shares these two features which may account, at least in part, for the wider binding specificity of SQUA and AGL2 in comparison to other MADS-box proteins. The K-box appears to be an additional component which contributes to the DNA-binding specificity of SQUA. Inclusion of the K-box increases the efficiency of SQUA binding to a wider spectrum of sites than by its core DNA-binding domain (Fig. 2C). This is consistent with the observation that sequences outside the core DNA-binding domains can affect the DNA-binding specificities of chimeric proteins created between human and Arabidopsis MADS-box proteins (39).

A recent study used chimeric proteins made with the MADS-box of either SRF or MEF2A and the Arabidopsis proteins AP1, AG, AP3 and PI to analyse the role of DNA binding specificity in vivo (39). In most cases, overexpression of these chimeric proteins caused similar homeotic transformations or rescued mutant phenotypes. This suggests that the DNA-binding specificity of these proteins is unimportant. However, potential caveats with this study are the overexpression of the chimeric proteins and the ability of several of these to participate in heterodimers with endogenous MADS-box proteins. Indeed, the presence of endogenous heterodimerisation partners with ‘relaxed’ DNA-binding specificities would allow dimerisation with more stringent binding specificities to bind to a wider spectrum of sites. The demonstration that plant MADS-box proteins such as SQUA (this study) and AGL2 (36), possess such a wider binding specificity strengthens this hypothesis.

DNA bending by Antirrhinum MADS-box transcription factors

MADS-box proteins differ in their ability to bend DNA (13). The Arabidopsis proteins AP1, AG and AP3/PI all induce directional DNA bending (19). The related Antirrhinum proteins SQUA, PLE and DEF/GLO also induce DNA bending (Figs 3 and 4). However, two further significant observations were made in the present study. Firstly, the direction of bending induced by DEF/GLO heterodimers differs from that induced by either PLE or SQUA homodimers. Secondly, the sequence of the DNA binding sites plays a role in dictating the magnitude of DNA bends induced by SQUA. Sequences located outside the conserved central portion of the binding site are at least in part responsible for this differential DNA bending (7). Together these observations indicate that the local promoter architecture will differ depending on the type of MADS-box protein(s) bound and at least in the case of SQUA, the sequence of the binding sites themselves. Such differences in architecture are likely to lead to alterations in transcription factor function.

Dimerisation determinants of plant MADS-box proteins

In plants, multiple MADS-box proteins exist with the potential to form a complex regulatory network by dimerisation or higher order complex formation. Dimerisation of the human MADS-box proteins SRF (5; reviewed in 1) and the yeast protein Mcm1 (6) is mediated by motifs from within the MADS-box and a 25–30 amino acid C-terminal extension known as the SAM-domain. Similarly, in human MEF2 proteins, both the MADS-box and C-terminal extension (MEF2 domain) contribute to dimerisation (9,13,40). In plants, the role of the analogous C-terminal extensions from the MADS-box, known as the I-domains, in dimerisation appears to differ amongst different dimer combinations. An additional domain, the K-domain, is also thought to play a role in mediating interactions amongst plant MADS-box proteins (17,20; reviewed in 1,2). In Arabidopsis, the MADS-box and
I-domain are sufficient for permitting DNA-bound dimer formation by AP1 and AG homodimers but not AP3/PI heterodimers (20). Similarly, the MADS-box and I-domain in the Antirrhinum proteins is sufficient to mediate DNA-bound homodimer formation by SQUA and PLE but not heterodimers between DEF and GLO. The inclusion of the K-domain permits heterodimer formation between DEF and GLO (17) and the homologous proteins AP3 and PI (20). The K-domain also plays an important role in dimerisation of full-length SQUA (Fig. 5) but is not however required for heterodimer formation with full-length PLE (Fig. 6). This suggests a role for the K-domain in modulating the ability of SQUA to form complexes with itself and other MADS-box proteins. In the case of DEF and GLO, the K-box is the major dimerisation motif and probably acts to stabilise heterodimers between these two proteins. The K-box therefore appears to play alternative roles in different MADS-box proteins. Interestingly, the PLE and SQUA K-boxes are insufficient to permit DNA-bound dimer formation when fused to the MEF2A MADS-box (Fig. 7). This is in contrast to the ability of a leucine zipper dimerisation motif to direct DNA binding in similar chimeras (13). However, these K-boxes are sufficient to permit homo- and heterodimer formation with SQUA in a non-DNA-bound state (Fig. 7). Similarly, yeast two-hybrid analysis demonstrates that the K-box from SQUA is sufficient for binding to DEFH72 (B.C. and B.D., unpublished data). A model therefore emerges in which the I-domain allows the correct alignment of the DNA-binding motifs from within the MADS-box in an analogous manner to the SAM- and MEF2-domains in human and yeast proteins. The K-box is not sufficient to substitute for this function but can either act to stabilise or modulate homo- and heterodimer formation. One intriguing possibility is that in addition to a role in dimerisation the K-box may permit higher order complex formation and hence further combinatorial interactions with other MADS-box proteins bound elsewhere in the promoter or in the absence of DNA binding by these additional partners. The ABC model for the determination of floral organ identity predicts some form of functional interaction between these three gene classes, but no evidence has so far been obtained to suggest that the interaction is direct (17). The formation of higher order complexes would provide one possible mechanism for this interaction to take place.

**Comparison of biochemical properties of SQUA and AP-1**

Genetic evidence suggests that the Antirrhinum MADS-box protein SQUA is functionally similar to the Arabidopsis MADS-box protein AP-1 (reviewed in 2). A comparison of our biochemical data with previous studies on AP-1 indicate several similarities and differences in their function. Both SQUA and AP-1 bend DNA (Fig. 3; 19). SQUA exhibits a wide DNA-binding specificity (Figs 1 and 2) and AP-1 also binds to several different sites but in contrast appears more restricted than other Arabidopsis MADS-box proteins on the limited number of sites tested (19). The minimal core DNA-binding domains of SQUA (Fig. 1) and AP-1 (19) are sufficient for DNA binding, indicating that the K-box is not required for dimerisation. However, neither a possible regulatory role for the K-box in homodimer formation nor its requirement for heterodimer formation with the PLE-related protein AG, were tested. Thus, biochemically, the two proteins differ in several aspects although it remains to be determined whether these differences reflect different functions of SQUA and AP-1 in vivo.

In summary, our results provide important insights on the mechanisms of DNA-binding by Antirrhinum MADS-box proteins. Differences in DNA binding specificity and mechanisms of DNA bending are uncovered and further subtleties in the mechanisms of dimerisation are detected. As the number of plant MADS-box proteins and potential in vivo targets identified increases, such molecular information will contribute significantly to our understanding of the complex interactions amongst these key regulatory proteins.

**ACKNOWLEDGEMENTS**

We would like to thank Margaret Bell for excellent technical assistance and Katherine Stewart for secretarial assistance. We would also like to thank Bob Liddell for oligonucleotide synthesis and DNA sequencing. We are grateful to Fei-Ling Lim for comments on the manuscript and members of our laboratory for helpful discussions. We are grateful to Jason Kahn for reagents. We would also like to thank Tom Kerppola for helpful advice. This work was supported by the North of England Cancer Research Campaign, the Wellcome Trust and a Medical Research Council Studentship to A.G.W. B.E.C. is supported by a grant from the Leverhulme Trust. A.D.S. is a Research Fellow of the Lister Institute of Preventative Medicine.

**REFERENCES**