MOLECULAR AND CELLULAR BIOLOGY, Aug. 1995, p. 4076–4085 Vol. 15, No. 8
0270-7306/95/$04.00+0

DNA Binding Specificity Determinants in MADS-Box Transcription Factors

STEPHEN J. NURRISH† AND RICHARD TREISMAN*

Transcription Laboratory, Imperial Cancer Research Fund Laboratories, London WC2A 3PX, United Kingdom

Received 3 March 1995/Returned for modification 13 April 1995/Accepted 26 April 1995

The MADS box is a conserved sequence motif found in the DNA binding domain of a family of transcription factors which possess related but distinct DNA binding specificities. We investigated the basis of differential sequence recognition by the MADS-box proteins serum response factor (SRF), MCM1, and MEF2A, using chimeric proteins and site-directed mutants in conjunction with gel mobility shift and binding site selection assays. Deletion of sequences immediately N terminal to the SRF MADS box alters its preferred binding site to that of MEF2A, although the resulting protein still weakly binds SRF-specific sites: exclusive binding to MEF2 sites requires further mutations, at MADS-box residues 11 to 15. In contrast to SRF, the sequence specificity of MCM1 (and of MEF2A) is determined entirely by sequences within its MADS box, and mutation of only SRF MADS-box residue 1 is sufficient to alter its binding specificity to that of MCM1. However, changes at both MADS-box positions 1 and 11 to 15 are necessary and sufficient to alter the specificity of the MCM1 MADS box to that of MEF2, and vice versa. The role of SRF MADS-box residues which differ from those present in the other proteins was investigated by selection of functional SRF variants in yeast cells. SRF MADS-box position 1 was always a glycine in the variants, but many different sequences at the other nonconserved MADS-box residues were compatible with efficient DNA binding. We discuss potential mechanisms of DNA recognition by MADS-box proteins.

The MADS (MCM1-agamous-ARG80-deficiens-SRF)-box transcription factors constitute a family of proteins that includes the metazoan transcription factors serum response factor (SRF) and MEF2, the yeast transcription factors MCM1 and ARG80, and a large number of plant homeotic gene products (1, 10, 14, 15, 17, 23, 28, 29). Both the SRF and MCM1 proteins act to recruit accessory factors to DNA via protein interactions mediated by their DNA binding domains (26). The MADS-box motif (20) is a 56-amino-acid region which comprises a highly conserved basic N-terminal region followed by a less well conserved, relatively hydrophobic C-terminal segment. It forms part of the DNA binding domain, which is generally located at the extreme N terminus of the protein. The MADS box is not sufficient for high-affinity DNA binding, however, since it encodes only the N-terminal part of the dimerization surface (1, 14, 17, 21). Efficient dimerization requires some 25 additional residues, located immediately C terminal to the MADS box, which are conserved only among subgroups of MADS proteins such as SRF/MCM1/ARG80 and the four MEF2 proteins (14, 17).

Although different MADS-box family members generally recognize A/T-rich consensus sequences, they possess distinct DNA binding specificities; for example, SRF binds the core consensus decanucleotide $CC(A/T)_2A(A/T)_3GG$, MCM1 binds $CC(C/T)(A/T)_3NNGG$, and MEF2 binds $(C/T)TA(T/A)_4TA(G/A)$ (13, 16, 17, 21, 27). Little is known, however, about the structural basis of this differential sequence recognition. Chimeric proteins in which the N-terminal basic region of the MADS box, together with any sequences N terminal to it, are joined to the SRF dimerization region and C-terminal sequences have been investigated by in vitro binding-site selection experiments (17, 27). The chimeric proteins possess exactly the same binding specificity as the parental proteins. Thus, the dimerization region makes no contribution to binding specificity, which is solely a function of the MADS-box N-terminal basic region and its flanking sequences. In the case of SRF, MADS-box residues 3 and 13 were reported to play an important role in determination of binding specificity (22) (note that we define MADS-box residue 1 as the conserved glycine preceding the original MADS-box sequence string [20]). However, the SRF MADS box does not contain sufficient information to completely determine binding specificity, since deletion of the sequences N terminal to it relaxes DNA binding specificity (22) in addition to reducing binding affinity (14).

In this work, we have used site-directed mutagenesis and a yeast screening protocol to investigate the basis of differential sequence-specific DNA binding by the MADS-box proteins MCM1, MEF2, and SRF. We show that MADS-box position 1 plays a crucial role in the determination of the binding specificity of all three proteins, while both positions 11 to 15 and the sequences immediately N terminal to the MADS box can place additional constraints on the types of sequences that can be recognized.

MATERIALS AND METHODS

Plasmids. For cell-free translation, derivatives of either T7ΔATG or T7Δ723 (14) were used. Plasmids were constructed by standard techniques, and sequences were verified where appropriate; full details are available on request. The coordinates of the MADS box used here are as indicated in each figure; note that we have redefined MADS-box residue 1 as a highly conserved glyine residue one residue N terminal to the MADS-box consensus string originally defined (20).

(i) SRF derivatives. T7SRFΔ122b encodes SRF residues MGLE-(SRF codons 122 to 508); it is identical to T7Δ723 (14) but contains mutations that introduce four unique restriction sites as follows: *Not*I, G-781→C; *Eco*RI, G-805→A; *Cla*I, C-814→T, and *Bst*BI, CAGC (nucleotides 841 to 844)→TTCG. T7SRFΔ137, T7SRFΔ141, and T7SRFΔ142 encode N-terminally truncated SRF derivatives encoding M-(SRF sequences including and C terminal to the indicated codon).

* Corresponding author. Mailing address: Transcription Laboratories, Imperial Cancer Research Fund Laboratories, 44 Lincoln's Inn Fields, London WC2A 3PX, United Kingdom.

† Present address: Molecular Biology, Massachusetts General Hospital, Boston, MA 02114.

Other SRF derivatives encode sequences as specified in the figures; in the exchange mutant names, numbers in parentheses indicate substituted MADS-box residues (M = MCM1, F = MEF2A).

(ii) MCM1 derivatives. MCM* (previously described as pT7MCM1[1–41]/SRF[167–508] [27]) comprises MCM1 codons 1 to 41 joined to SRF codons 167 to 508. MCM*Δ17 encodes M-(MCM1 codons 17 to 41)-(SRF codons 167 to 508). Other MCM* derivatives encode sequences as specified in the figures; in the exchange mutant names, numbers in parentheses indicate substituted MADS-box residues (S = SRF, F = MEF2A).

(iii) MEF2* derivatives. MEF2* (previously described as pT7C4/SRF [17]) encodes MEF2A codons 1 to 28 joined to SRF codons 169 to 508). Other MEF2* derivatives encode sequences as specified in the figures; in the exchange mutant names, numbers in parentheses indicate substituted MADS-box residues (M = MCM1, S = SRF).

(iv) Yeast expression plasmids. Plasmid SD08.3, which encodes the SRF-VP16 fusion protein MTGSGLE-(SRF codons 122 to 412)-(VP16 codons 410 to 490) was derived from SD08 (5) as follows. (i) The unique SD08 *Eco*RI and *Cla*I restriction sites were removed by digestion and end repair; (ii) SRF codons 1 to 245 (*Xho-Bgl*II) were replaced by SRF codons 122 to 245 from T7SRFΔ122b (see above). This removes SRF N-terminal sequences, thereby facilitating PCR, and introduces *Not*I, *Eco*RI, *Cla*I, and *Bst*BI sites for manipulation of the SRF MADS box. Plasmids SD08.3b and SD08.3c are derivatives of SD08.3 in which the SRF MADS-box *Eco*RI and *Cla*I sites, respectively, have been destroyed by digestion and end repair.

Construction of randomized libraries. Plasmids SD08.3b and SD08.3c, which encode nonfunctional SRF derivatives and therefore score white in the colony color assay, were used for insertion of randomized cassette oligonucleotides. Only library plasmids that contain the mutagenized sequences, thereby restoring the open reading frame through the SRF MADS box, can score positive in the *lacZ* colony color assay. Library DNA was purified following electrotransformation of *Escherichia coli* (5); 3,000, 30,000, and 91,000 independent clones were obtained for libraries A, B, and C, respectively.

Library A randomized SRF MADS-box positions 1 and 3. A PCR was carried out with T7Δ122b as the template and oligonucleotides T7 (AATACGACTC ACTATAG) and 43981 (CCCCCCATCGATGAATTCCATCTTGATCTTN NNGCGNNNGCGGGTCTTCTTACC) as primers. The product was digested with *Xho*I and *Cla*I, and the resulting fragment was substituted for the corresponding fragment of SD08.3b.

Library B randomized SRF MADS-box positions 3, 6, 8, and 9. The T7 and 15985 (CTTATCGATNNNNNNCATNNNGATCTTNNNGCGGCCGCCTAT AGTGAGTCGTATT) oligonucleotides were annealed and rendered double stranded by using reverse transcriptase. The product was digested with *Not*I and *Cla*I, and the resulting 30-bp fragment was used to replace the *Not*I-*Cla*I fragment of SD08.3b.

Library C randomized SRF MADS-box positions 11, 13, 14, and 15. The T7 and 18705 (CCTCTTCGAAAAGGTCGTGTAGCGNNNNNNNNNGTYNNN GATGAATTCTATAGTGAGTCGTATT) oligonucleotides were annealed and rendered double stranded by using reverse transcriptase. The product was digested with *Eco*RI and *Bst*BI, and the resulting 32-bp fragment was used to replace the corresponding fragment of SD08.3c.

Screening the libraries of SRF variants. The yeast reporter strain 62L was used for screening the library DNAs by previously published methods (5). Five micrograms of library A plasmid generated ~1,500 independent colonies, of which 7% scored as blue in the color assay. Of these, approximately one-half represented contaminating wild-type SD08.3 DNA. Sixty micrograms of library B and C plasmid generated 30,000 and 24,000 independent yeast colonies, of which 4.6 and 12.5%, respectively, scored positive. DNA recovered from positive colonies was retransformed into 62L cells to confirm the phenotype and sequenced. For in vitro binding studies, the *Xho*I-*Stu*I fragment (SRF codons 122 to 171) containing the variant sequences was substituted for the corresponding fragment of T7Δ723 and used for in vitro translation. Proteins and gel mobility shift assays of the SRF variants were quantitated with a PhosphorImager (Molecular Dynamics).

Gel mobility shift assay. Gel mobility shift assays were performed as described previously (11) but with only 0.2 μg of poly(dI-dC) poly(dI-dC). Cell-free translation was performed with the TNT coupled transcription-translation rabbit reticulocyte lysate system (Promega) as directed by the manufacturer. Protein expression was assessed by [^{35}S]methionine incorporation; generally binding reaction mixtures included 0.5 to 2 μl of lysate adjusted to a constant volume with unprogrammed lysate.

Reaction mixtures contained 0.2 ng (3 fmol) of radiolabelled probe. The core consensus sequences of the c-*fos* serum response element (SRE), MCM1, and MEF2 probes are shown in Fig. 1B. Probes were generated by PCR using the following templates: for c-*fos* SRE, SRE-TKCAT (7); for MCM1, SRE.M-TK CAT (7); for MEF2, pN10.WTx1 (17) (all using the M13 and pBLCAT2R primers [2]); and for ACT.L, pACT.L (25), using the M13F and M13R primers. Binding competition assay mixtures also contained 20 ng (1.2 nmol; ~350-fold excess over probe) of double-stranded unlabelled competitor oligonucleotide. Oligonucleotides were previously described: S and mS were identical to FOS and FOS* (16); F was N10 (17); and M was SRE.M (7).

Other methods. Binding-site selection was performed as previously described (16). DNA manipulations, in vitro transcription, and sodium dodecyl sulfate-polyacrylamide gel electrophoresis were performed by standard methods. Yeast transformation, plating on nylon filters, and processing for the colony color assay were performed as described previously (5).

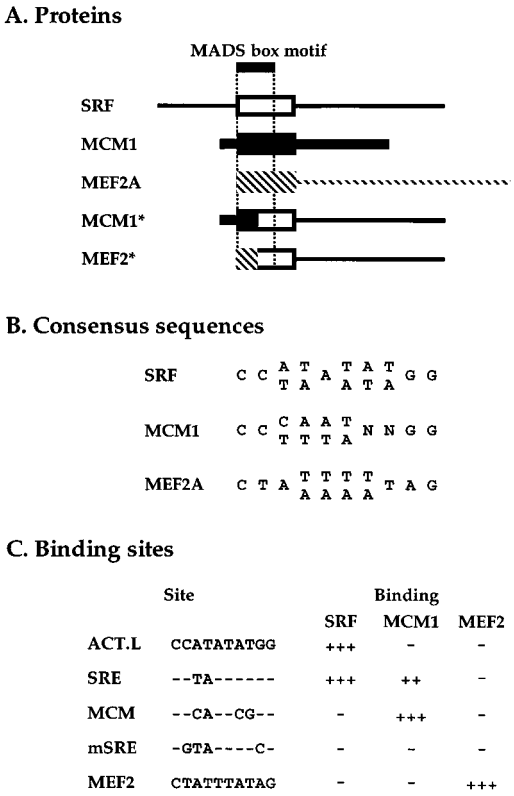


FIG. 1. MADS-box proteins and their binding specificities. (A) Schematic representation of the MADS-box proteins SRF, MCM1, and MEF2A and the derivatives MCM* (27) and MEF2* (17) used in this work. Conserved regions within DNA binding domains are shown as boxes. SRF sequences are shown as thin lines and a white box, MCM1 sequences are shown as thick lines and a black box, and MEF2 sequences are shown as hatched lines and a hatched box. The location of the MADS-box motif is shown at the top. The fusion proteins MCM* and MEF2* have been described elsewhere (17, 27). (B) Sequences of the decamer cores of the consensus binding sites of SRF, MCM1, and MEF2A. N represents any base. (C) Binding properties of probe sequences used in this work. The relationship of the SRF and MCM1 probe sequences is indicated by using dashes to show identity with the SRF-specific probe ACT.L. Note that the c-*fos* SRF site is inverted with respect to its orientation in the c-*fos* SRE; its mutated derivative mSRE (16) contains mutations at conserved residues that prevent SRF binding (7, 25). The MEF2 sequence is shown in its entirety. For details, see references 16, 17, 25, and 27.

RESULTS

Strategy for in vitro binding studies. We used derivatives of either SRF itself or the fusion proteins MCM* and MEF2* for in vitro DNA binding studies. In the fusion proteins, sequences N terminal to the MADS box, and the N-terminal basic sequences of the MADS box itself, are derived from MCM1 or MEF2A, while the dimerization domain and C-terminal sequences are from SRF (Fig. 1A). The use of a common dimerization region was designed to eliminate any effects of the different dimerization regions on binding specificity: the binding specificities of the fusion proteins are identical to those of the parent proteins, indicating that all specificity-determining protein-DNA contacts are maintained in the chimeras (17, 27).

Each protein was produced by cell-free translation and DNA binding assessed by the gel mobility shift assay using the DNA probes shown in Fig. 1B. For SRF, we used either the c-*fos*

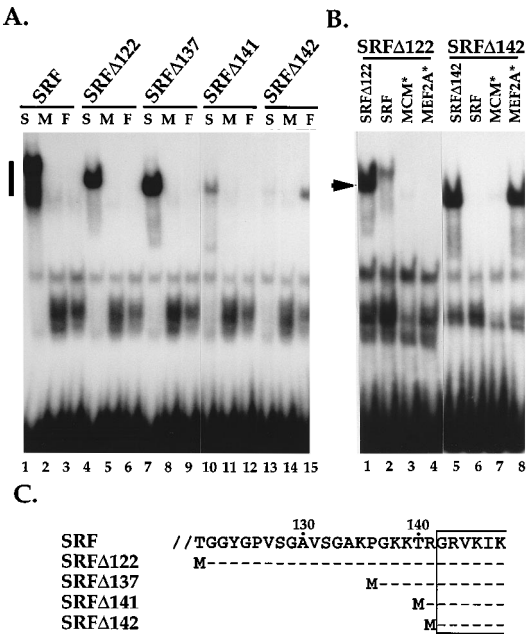


FIG. 2. Residues N terminal to the SRF MADS box are required for SRF binding specificity and affinity. (A) DNA binding by SRF N-terminal truncation mutants. Equal amounts of the different SRF derivatives shown in panel C were tested in gel mobility shift assays using the probe indicated above each lane. S, c-*fos* SRE; M, MCM site; F, MEF2 site. Specific SRF complexes are indicated by a line. Note that each probe binds lysate proteins with various degrees of efficiency; these interactions do not appear specific as judged by binding competition and have not been investigated further. (B) Binding-site selection analysis of SRFΔ122 and SRFΔ142. SRFΔ122 (lanes 1 to 4) and SRFΔ142 (lanes 5 to 8) were used in the binding-site selection assay. Populations of radiolabelled oligonucleotides present after five rounds of selection were used as probes in the gel mobility shift assay with SRFΔ122 (lane 1), SRFΔ142 (lane 5), SRF (lanes 2 and 6), MCM* (lanes 3 and 7), and MEF2* (lanes 4 and 8). For sequences of representative oligonucleotides, see Fig. 3. (C) Schematic representation of the SRF mutants tested. The wild-type SRF sequence is shown at the top, with the N-terminal boundary of the MADS box indicated; sequence identities are indicated by dashes. Numbers refer to the SRF sequence.

SRE (CCTAATATGG), which can also bind MCM1, or the synthetic SRF-specific site ACT.L (CCATATATGG) (25); neither of these sites binds MEF2. For MCM1, we used the synthetic MCM1-specific site M (CCCAATCGGG), which cannot be bound by SRF or MEF2 (7, 27); for MEF2, we used the MEF2-specific N10 probe (CTATTTATAG), which cannot be bound by SRF or MCM1 (17). To enhance the detection of low-affinity interactions, we also performed binding-site competition experiments in which a 350-fold excess of unlabelled competitor DNA was included in the DNA binding reaction mixtures. In some cases, the optimal sequence preference of a protein was determined by using it to select oligonucleotides containing binding sites from a pool of random sequence oligonucleotides (16).

Sequences N terminal to the SRF MADS box are required for sequence specificity. We previously found that sequences N terminal to the SRF MADS box are required for high-affinity DNA binding (14). To define the N terminus of the SRF DNA binding domain at high resolution, we analyzed a set of N-terminal truncations in a gel mobility shift assay. Deletion of sequences N terminal to residues 122 and 137 had no substantial effect on binding affinity or specificity, whereas removal of further sequences to position 141 resulted in a marked reduction in binding affinity but no loss of sequence specificity (Fig. 2A, lanes 1 to 12). In contrast, SRFΔ142, a deletion mutant which lacks all sequences N terminal to the MADS box, bound only very weakly to the c-*fos* SRE but more efficiently to the MEF2 probe; the complexes formed with the MEF2 probe exhibited slightly greater mobility than those formed on the c-*fos* SRE (Fig. 2; compare lanes 13 and 15). The SRFΔ142 protein also bound weakly to the SRF-specific ACT.L site but could not detectably bind the MCM1 site (Fig. 2, lane 14, and data not shown). Consistent with these data, in a binding competition experiment, binding of SRFΔ142 to the MEF2 probe could be efficiently diminished by a 350-fold excess of the c-*fos* SRE or MEF2 oligonucleotide but not by the MCM1 or mutant c-*fos* SRE (CGTAATATCG) oligonucleotide (data not shown; see Fig. 6).

Although truncation of SRF clearly alters its binding specificity, it remained possible that the optimal binding site for SRFΔ142 is distinct from those tested. The SRFΔ122 and SRFΔ142 proteins were therefore compared in a binding-site selection experiment. Following five rounds of selection, the radiolabelled oligonucleotide population was used as a probe in gel mobility shift assays with SRF, MEF2*, and MCM*. DNA selected by SRFΔ122 was bound only by wild-type SRF in the gel mobility shift assay (Fig. 2B, lanes 1 to 4), while DNA selected by SRFΔ142 was efficiently bound only by MEF2*, suggesting that it contains MEF2 consensus sequences (Fig. 2B, lanes 5 to 8). Sequence analysis of oligonucleotides selected by SRFΔ142 showed that each contained a perfect match to the MEF2 consensus binding site CTA(T/A)$_4$TAG, with the same range of sequence variations that occur in bona fide MEF2 binding sites (Fig. 3A; cf. Fig. 1C).

We used site-directed mutagenesis of SRFΔ137 to investigate the role of sequences in the region of SRF immediately N terminal to its MADS box in DNA binding specificity and affinity. No mutant bound the MEF2 site, although the mutations KK138/139AA and T140Q caused significant reductions in binding affinity (Fig. 4, proteins C to G). Although the presence of a basic residue immediately N terminal to the SRF MADS box is sufficient to restrict binding to SRF sites, albeit with low affinity, in the context of SRFΔ137, substitution of this residue by alanine had little effect on affinity or specificity (Fig. 4, protein D). Moreover, substitution of the sequences flanking the SRF MADS box by the corresponding sequences from MCM1 did not affect specificity (Fig. 4, protein H). Since both the SRF and MCM1 N-terminal sequences prevented the SRF MADS box from binding the MEF2 site, we tested whether their presence N terminal to the MEF2* MADS box would also prevent binding. This was indeed the case (Fig. 4, proteins J and K).

Taken together, the results show that the SRF MADS box possesses a dual specificity: it preferentially recognizes the same consensus sequences as MEF2 but retains the ability both to bind the c-*fos* SRE and to discriminate between SRF- and MCM-specific sequences. The sequences N terminal to the SRF MADS box both prevent binding to MEF2 sites and increase DNA binding affinity.

MADS-box position 1 is important for MCM1 binding specificity. We showed above that sequences N terminal to the SRF MADS box are essential for correct DNA binding specificity. To test whether this is also true for other MADS box proteins containing sequences N terminal to the MADS box, we analyzed DNA binding by the MCM1 derivative MCM*. We first compared the binding properties of MCM* and a derivative which lacks all residues N terminal to the MADS box, MCM*Δ17. Both proteins efficiently bound to the MCM1 site and c-*fos* SRE but not to the MEF2 site, indicating that the MCM* N-terminal sequences do not influence its binding specificity (Fig. 5A, lanes 1 to 6). Consistent with this finding,

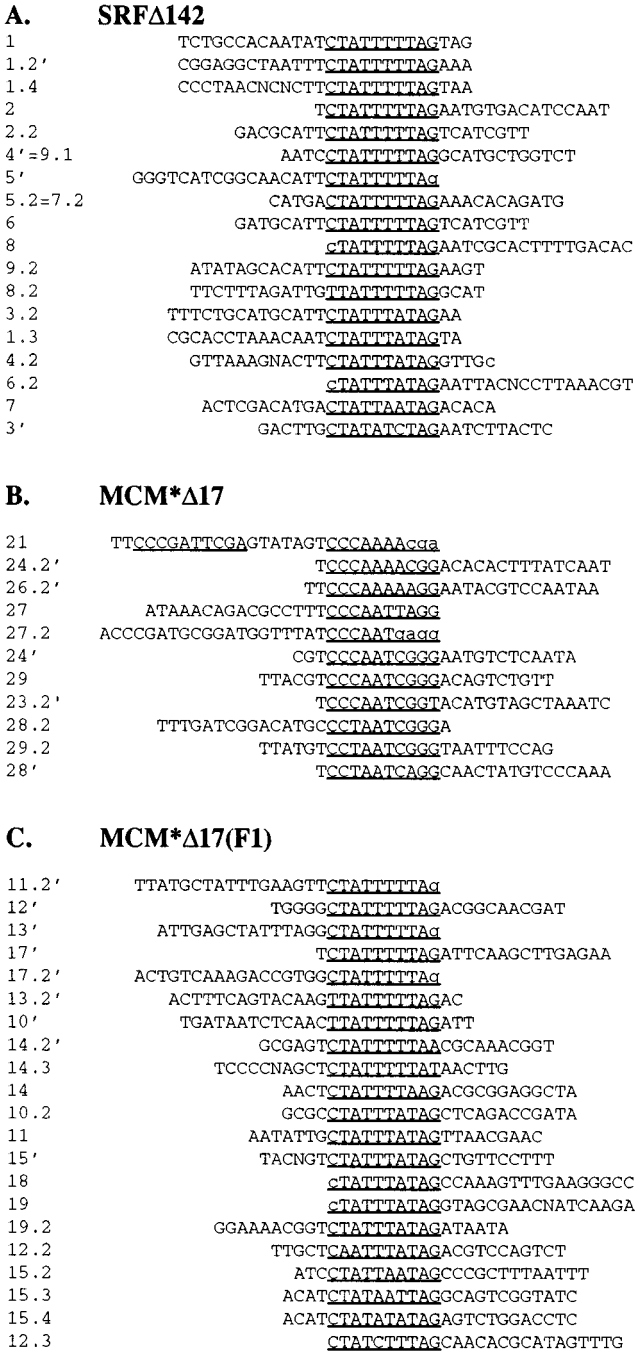


FIG. 3. Sequences of selected oligonucleotides recovered after five rounds of selection, aligned to maximize match to the underlined consensus sequence. Lowercase letters denote sequences derived from the primer oligonucleotides.

in in vitro binding-site selection analysis, MCM*Δ17 selected sequences that contained the MCM1 consensus sequence CCY(A/T)(A/T)(T/A)NNGG (Fig. 3B). Thus, in contrast to SRF, all of the residues that define sequence-specific DNA binding by MCM* lie within its MADS box.

Since the MCM1 MADS box has a specificity different from that of SRF, we used SRF-MCM chimeras to map the residues within the MCM* MADS box important for binding specificity. We constructed mutant derivatives of the SRFΔ142 protein in which SRF residues are replaced by progressively smaller segments of the MCM1 MADS box. A mutant of SRFΔ142 in which MADS-box positions 1 to 15 are substituted by the corresponding residues from MCM1 bound the SRF and MCM1 sites with reduced affinity but did not detectably bind the MEF2 site (Fig. 5A, lanes 10 to 12). Similar results were obtained with a mutant in which MCM1 MADS-box positions 1 to 9 were substituted for those of SRF, although binding was weaker (data not shown). Surprisingly, SRFΔ142(M1), in which only SRF MADS-box residue 1 is replaced by the corresponding residue from MCM1, also bound DNA with MCM1-like specificity, albeit very weakly (Fig. 5A, lanes 13 to 15). This result was confirmed in a binding competition experiment in which excess c-*fos* SRE and MCM1, but not MEF2, oligonucleotides competed for binding of SRFΔ142(M1) to a radiolabelled MCM1 probe (Fig. 5B). We next tested whether N-terminal MADS-box residues were sufficient to alter the binding specificity of SRFΔ122, which binds DNA with SRF-type specificity. Substitution of MADS-box residue 1, but not residue 3, was sufficient to alter specificity (Fig. 5A, lanes 16 to 21). We conclude that MADS-box residue 1 is an important determinant of SRF DNA binding specificity.

MADS-box positions 11 to 15 can restrict binding specificity. The experiments described above highlight the importance of MADS-box residue 1 and sequences N terminal to it for sequence-specific binding by MCM1 and SRF, respectively. However, a previous report demonstrated a role for MADS-box residue 13 in sequence-specific binding of another MADS-box protein, MEF2A (22). We therefore compared the relative roles of these regions in sequence-specific binding by MEF2*. In MEF2*, as in MCM*, the basic portion of the MADS box must be sufficient for MEF2-specific DNA binding, since MEF2* contains no sequences N terminal to the MADS box (17).

To identify which MEF2* MADS-box residues restrict binding to MEF2 sites alone, we substituted residues in the SRFΔ142 MADS box, which can bind both SRF and MEF2 sites, with the corresponding residues from MEF2A. Since SRFΔ142 exhibits only weak binding to the c-*fos* SRE probe, we analyzed binding in a competition assay in which we tested the abilities of different sequences to compete for binding of each protein to a radiolabelled MEF2 site probe. In this assay, excess unlabelled c-*fos* SRE or MEF2 oligonucleotide, but not oligonucleotides containing MCM1 sites or a mutant c-*fos* SRE, efficiently compete for binding of SRFΔ142 to a radiolabelled MEF2 site probe (Fig. 6A). In contrast, only the MEF2 site oligonucleotide affects binding of MEF2* to the MEF2 probe (Fig. 6, protein B). An SRFΔ142 derivative containing MEF2 MADS-box residues 1 to 15 also binds with correct MEF2 specificity (Fig. 6, protein C), whereas a derivative containing MEF2 MADS-box residues 1 to 9 maintains the dual binding specificity of SRFΔ142 (Fig. 6, protein D). However, substitution of SRFΔ142 MADS-box residues 11 to 15 by the corresponding MEF2 sequences restricts its specificity to the MEF2 site (Fig. 6, protein E).

We next tested whether MADS-box residues 11 to 15 are important for specificity of DNA binding in the context of the MEF2* MADS box. Substitution of MEF2* MADS-box residues 11 to 15 by the corresponding residues of SRF broadened binding specificity such that the SRF site can be bound, although very weakly (Fig. 6, protein F). Substitution by MCM1 MADS-box residues 11 to 15 had the same effect, rather than allowing recognition of an MCM1 site (Fig. 6G). Taken together with the previous data, these results show that MEF2A MADS-box residues 11 to 15 are associated with exclusive binding to MEF2 consensus sequences, while the corresponding residues of MCM1 and SRF are compatible with binding to

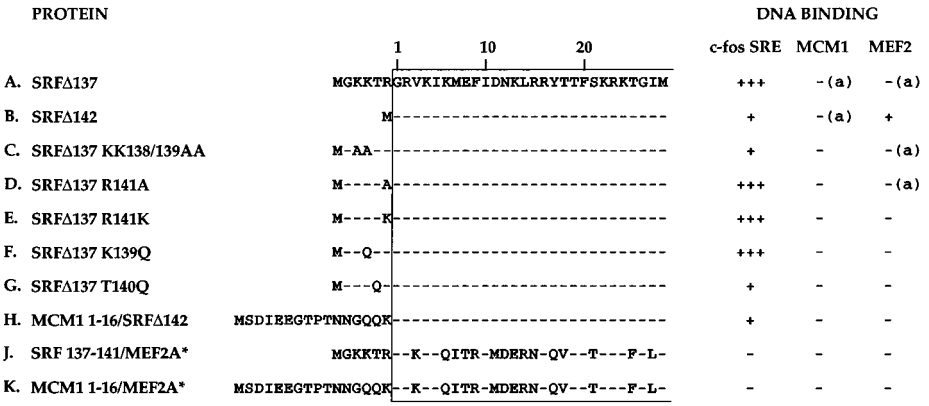


FIG. 4. Sequence requirements for function of the SRF sequences immediately N terminal to the MADS box. The mutant SRF derivatives shown were constructed and tested for binding to the c-*fos* SRE, MCM1, and MEF2 sites by gel mobility shift assay. Dashes indicate identity to SRF; the MADS-box boundary is outlined, and MADS-box residues are numbered. Binding to the various probe sequences is shown schematically: +++, binding comparable to that of wild-type SRF in the gel mobility shift assay; +, binding less than or equal to 20% the level of wild-type SRF. (a) indicates that the results were confirmed by binding competition studies in which a 350-fold molar excess of an unlabelled oligonucleotide did not compete for binding to the c-*fos* SRE (proteins A, C, and D) or the MEF2 site (protein B).

the MCM1, SRF, and MEF2 sites according to their protein contexts.

Alteration of MCM1 specificity to that of MEF2A. The experiments described so far pinpoint MADS-box positions 1 and 11 to 15 as important DNA binding specificity determinants for binding to MCM1 and MEF2 sites, respectively. We therefore used MCM*-MEF2* chimeras to investigate whether changes at either or both of these regions are sufficient to alter the DNA binding specificity of the MCM* MADS box to that of MEF2*. The results are shown in Fig. 7.

We first replaced MCM*Δ17 MADS-box position 1 by the corresponding residue from MEF2, an E→G change. This mutant protein, MCM*Δ17(F1), did not bind the MCM1 site detectably but very weakly bound the c-*fos* SRE and MEF2 sites (Fig. 7A; compare lanes 4 to 6 with lanes 1 to 3). Binding competition analysis confirmed this result: the c-*fos* SRE and MEF2 oligonucleotides, but not an MCM1 oligonucleotide, efficiently competed for binding of MCM*Δ17(F1) to a radiolabelled MEF2 site (Fig. 7B, top panel). We determined the optimal binding site for MCM*Δ17(F1) by binding site selection. Oligonucleotides selected by MCM*Δ17(F1) bound the MEF2* protein efficiently but not the SRF or MCM* protein (data not shown). Sequence analysis of 21 oligonucleotides selected by MCM*Δ17(F1) showed that each contained a good match to the MEF2 consensus sequence (Fig. 3C). The binding specificity of MCM*Δ17(F1) is thus identical to that of

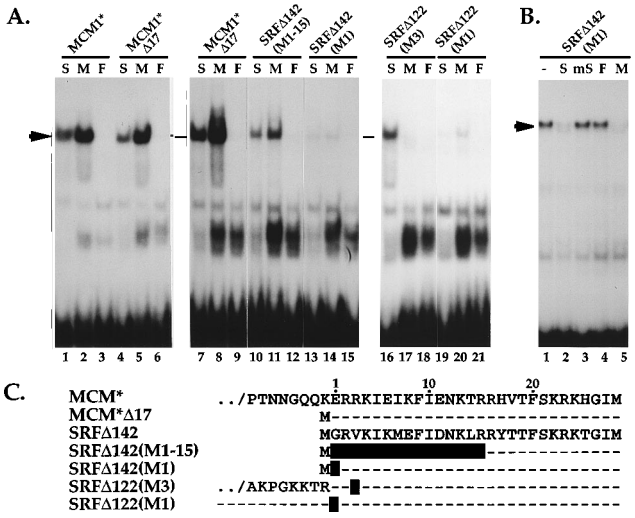


FIG. 5. MADS-box residue 1 is sufficient for MCM1-type DNA binding specificity. (A) DNA binding by MCM* derivatives. Equal amounts of the different MCM1 derivatives shown in panel C were tested in gel mobility shift assays, using the probe indicated above each lane. S, c-*fos* SRE; M, MCM site; F, MEF2 site. (B) Binding competition analysis of SRFΔ142(M1). Binding reaction mixtures contained radiolabelled MCM1 probe and a 350-fold excess of the following unlabelled competitor oligonucleotides: c-*fos* SRE (S), mutant c-*fos* SRE (mS), MEF2 site (F), and MCM1 site (M). (C) Schematic representation of the MCM* derivatives. The sequences of the N termini of MCM* and MCM*Δ17 are shown at the top, with identities indicated by dashes. Below are shown the sequences of SRFΔ142 and SRFΔ122 and their derivatives carrying different extents of MCM1 sequence, shown as black boxes; sequence identities are indicated by dashes. Numbers indicate MADS-box positions.

FIG. 6. MADS-box residues are important but not sufficient to determine binding specificity. Binding competition analysis was used to evaluate the sequence specificities of various SRFΔ142 and MEF2A derivatives. Sequences of each protein are shown at the left; identities with SRF are indicated by dashes, and MADS box positions are numbered. Segments of MEF2 sequence substituted for SRF sequences are shown as shaded boxes; in protein MEF2*(M11–15), the MCM1 sequences are shown in full. Binding reaction mixtures contained radiolabelled MEF2 probe and where indicated a 350-fold excess of the following unlabelled competitor oligonucleotides: c-*fos* SRE S, mutant c-*fos* SRE (mSRE), MEF2 site, and MCM1 site.

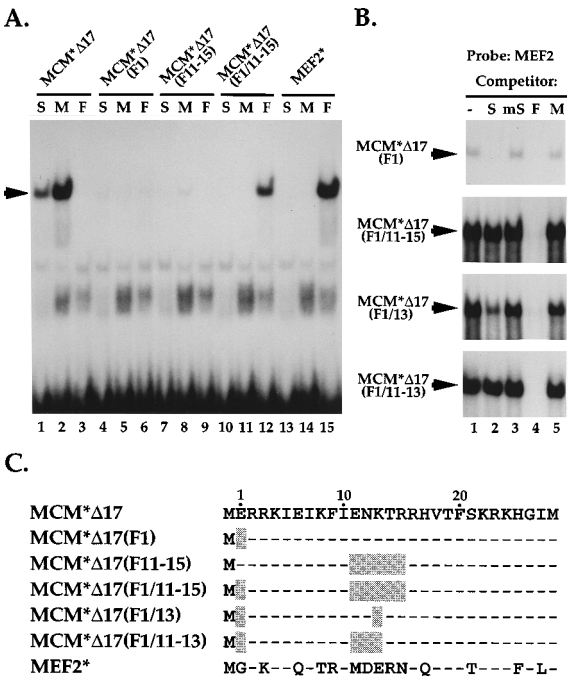


FIG. 7. Conversion of the specificity of MCM* to that of MEF2*. (A) DNA binding by MCM*Δ17 exchange mutants. Equal amounts of the different MCM*Δ17 or MEF2A derivatives shown in panel C were tested in gel mobility shift assays using the probe indicated above each lane. S, c-*fos* SRE; M, MCM site; F, MEF2 site. (B) Binding competition analysis of MCM*Δ17 exchange mutants. Binding reaction mixtures contained radiolabelled MEF2 probe and a 350-fold excess of the following unlabelled competitor oligonucleotides: c-*fos* SRE (S), mutant c-*fos* SRE (mS), MEF2 site (F), and MCM1 site (M). (C) Schematic representation of the MCM*Δ17 exchange mutants. The N-terminal sequences of MCM*Δ17 are shown at the top, and those of MEF2* are shown at the bottom; identities are indicated by dashes. In the exchange mutants, MEF2A sequences substituted for those of MCM1 are shaded. Numbers indicate MADS-box positions.

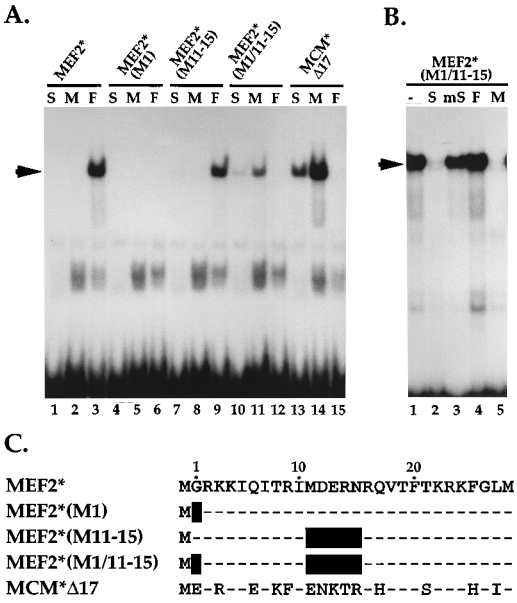
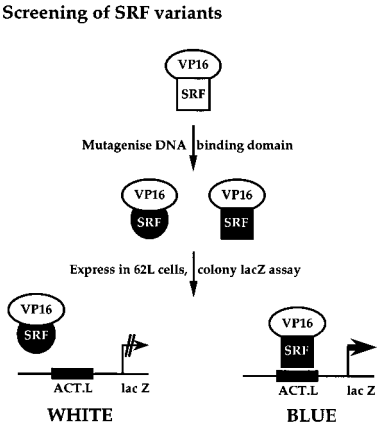


FIG. 8. Conversion of the specificity of MEF2* to that of MCM*. (A) DNA binding by MEF2A exchange mutants. Equal amounts of the different MCM*Δ17 or MEF2A derivatives shown in panel C were tested in gel mobility shift assays using the probe indicated above each lane. S, c-*fos* SRE; M, MCM site; F, MEF2 site. (B) Binding competition analysis of MEF2*(M1/11–15). Binding reaction mixtures contained a radiolabelled MCM site probe and a 350-fold excess of the following unlabelled competitor oligonucleotides: c-*fos* SRE (S), mutant c-*fos* SRE (mS), MEF2 site (F), and MCM1 site (M). (C) Schematic representation of the MEF2* exchange mutants. The N-terminal sequences of MEF2* are shown at the top, and those of MCM*Δ17 are shown at the bottom; sequence identities are indicated by dashes. In the exchange mutants, MCM1 sequences substituted for those of MEF2* are shown in black. Numbers indicate MADS-box positions.

SRFΔ142: the proteins recognize the same spectrum of MEF2 consensus sequences, weakly bind sequences matching the SRF consensus, and fail to bind MCM1-specific sequences. These data confirm the importance of MADS-box residue 1 as a DNA binding specificity determinant in the context of the MCM1 MADS-box sequences.

We next examined the role of MCM*Δ17 MADS-box positions 11 to 15 by substituting them with the corresponding residues from MEF2. This protein, MCM*Δ17(F11–15), exhibited reduced binding to the MCM1 site but did not bind to the MEF2 site (Fig. 7A, lanes 7 to 9). However, when these changes were combined with the MADS-box position 1 exchange, the resulting protein, MCM*Δ17(F1/11–15), bound with high affinity to the MEF2 site but not to the MCM1 or SRF site (Fig. 7A; compare lanes 1 to 3 with lanes 10 to 12). Moreover, in a binding competition analysis, binding of MCM*Δ17(F1/11–15) to the MEF2 site probe was unaffected by a 350-fold excess of unlabelled MCM1 site oligonucleotide (Fig. 7B, second panel). A previous report suggested that MADS-box residue 13 is a crucial specificity determinant (22). We therefore tested whether an exchange of MADS-box residues 1 and 13 is sufficient to achieve a complete specificity switch. Although this mutant, MCM*Δ17(F1/13), bound the MEF2 probe efficiently, it retained the ability to bind the c-*fos* SRE weakly, as assessed by a binding competition experiment (Fig. 7B; compare middle panels). However, substitution of MCM1 MADS-box positions 1 and 11 to 13 by the corresponding MEF2 residues was sufficient to alter the specificity of MCM*Δ17 to that of MEF2* (Fig. 7B, bottom panel). We found that no single MEF2 residue in the region from residues 11 to 15 was sufficient to alter specificity in combination with MADS-box residue 1 (data not shown). Thus, binding specificity is not a function of a single residue in the MADS-box region from residues 11 to 15.

Alteration of MEF2A binding specificity to that of MCM1. We next wished to determine whether the exchange of both MADS-box residues 1 and 11 to 15 is also necessary and sufficient to alter the specificity of the MEF2* MADS box to that of MCM1*. Substitution of MEF2* MADS-box position 1 by the corresponding residue from MCM*, a G→E change, abolished sequence-specific binding, as judged by the gel mobility shift assay (Fig. 8A; compare lanes 4 to 6 with lanes 1 to 3). Consistent with this result, the protein failed to select specific oligonucleotides through five rounds of binding-site selection (data not shown). Substitution of MEF2* MADS-box positions 11 to 15 with the corresponding MCM1 residues had essentially no effect on binding affinity or specificity in the gel mobility shift assay (Fig. 8A; compare lanes 7 to 9 with lanes 1 to 3); however, as already shown, this exchange allowed detectable binding to the c-*fos* SRE in the binding competition assay (see Fig. 6G). However, substitution of both MEF2* MADS-box positions 1 and 11 to 15 by the corresponding MCM1 residues generated a protein with a specificity strikingly similiar to that of MCM1, albeit of lower affinity. This protein, MEF2*(M1/11–15), bound the MCM1 site and the c-*fos* SRE, but not the MEF2 site, in the gel mobility shift assay (Fig. 8A, lanes 10 to 12). Moreover, in the binding competition assay,



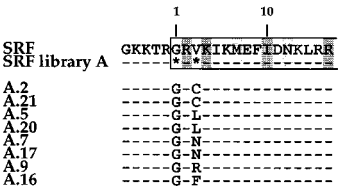


FIG. 9. SRF MADS-box position 1 must be glycine for efficient binding in vivo. (A) In vivo screen for SRF variants that can bind the SRF-specific ACT.L SRE. At the top, the SRF-VP16 fusion protein is shown schematically with the DNA binding domain as an open square and the VP16 activation domain as an oval. Mutagenized SRF DNA binding domains are filled (squares, binders; circles, nonbinders). The structure of the indicator gene is indicated. Steps in the screening procedure are outlined. (B) Part of the sequence of the wild-type SRF sequence is shown, with the MADS box outlined. Dark shading, invariant residues; light shading, conserved residues. Residues varied in library A are indicated by asterisks. Below are the sequences of eight independently derived library A SRF variants. Five micrograms of library A plasmid DNA generated ~1,500 independent colonies, of which 7% scored as blue in the color assay (of these, approximately one-half represented contaminating wild-type SD08.3 DNA).

complex formation between MEF2*(M1/11–15) and a radiolabelled MCM1 oligonucleotide was unaffected by the presence of a 350-fold excess of unlabelled MEF2 oligonucleotide but was abolished by a similar excess of unlabelled MCM1 or c-*fos* SRE oligonucleotide (Fig. 8B).

Selection of SRE-binding SRF variants in vivo. The experiments described above establish that the residues responsible for the differential binding specificities of MCM* and MEF2* are MADS-box positions 1 and 11 to 15. We next investigated the roles of these and other MADS-box residues in SRF DNA binding in vivo. We used a previously described yeast genetic screen to identify SRF variants capable of binding the SRF-specific ACT.L site in vivo (5) (Fig. 9A), expecting strong selection for particular residues at any MADS-box positions directly or indirectly involved in specificity-determining DNA contacts. The indicator yeast strain, 62L, carries a *lacZ* reporter gene under the control of the SRF-specific binding site ACT.L; endogenous MCM1 protein cannot bind the ACT.L sequence, and 62L cells are therefore white in a colony color assay (5, 27). In contrast, an SRF-VP16 fusion protein can bind ACT.L efficiently and activate *lacZ* expression, and cells expressing this protein are therefore blue in the assay. We therefore subjected a VP16-tagged derivative of SRFΔ122, which has wild-type DNA binding properties, to cassette mutagenesis and screened for functional SRF variants by using the colony color assay. We constructed three libraries of mutants in which nonconserved MADS-box residues were varied as follows: (i) MADS-box positions 1 and 3 randomized; (ii) MADS-box positions 3, 5, 8, and 9 randomized; and (iii) MADS-box positions 11, 13, 14, and 15 randomized, position 12 = D or N. Each library generated blue colonies at high frequency, between 7 and 12.5%; plasmid DNA was recovered from these colonies, and the sequence through the mutagenized region was determined.

Sequence analysis of eight positive clones derived from library A showed absolute preference for the glycine at MADS-box position 1, confirming that this residue is essential for binding the SRF-specific ACT.L site; in contrast, position 3 was variable (Fig. 9B). The high frequency with which positive SRF variants were recovered from libraries B and C was unexpected considering the large number of different possible sequence variants present in these libraries. Sequence analysis of the variants suggested that many different sequences at the mutagenized positions are compatible with binding to the ACT.L SRF site: SRF variants from library B showed no obvious sequence preferences apart from MADS-box position 10, at which large hydophobic residues are preferred, while variants from library C showed sequence preference for acid or amide side chains at MADS-box position 11 and for basic side chains at position 13; no discrimination between D and N was apparent at position 12 (Table 1). To verify that the library B and C SRF variants can bind the ACT.L SRE in vitro, the variants were tested in the gel mobility shift assay. All library B mutants bound a radiolabelled ACT.L probe comparably to the wild-type protein (Table 1). Binding of the library C mutants was more variable: in general, variants containing a basic residue at position 13 bound comparably to wild-type SRF (Table 1). All of the variants formed ternary complexes with the Elk-1 protein with comparable efficiencies on an appropriate probe containing the ACT.L SRF site (data not shown).

These results confirm the importance of SRF MADS-box position 1 for DNA binding and show that a huge amount of sequence variation can be tolerated at other positions in the SRF MADS box with little effect on DNA binding. Nevertheless, the large number of mutants that do not bind detectably in the colony color assay indicates that many sequences are not tolerated at SRF MADS-box positions 11 to 15.

DISCUSSION

MADS-box DNA binding specificity determinants. We have investigated the basis of differential DNA sequence recognition by the three MADS-box transcription factors SRF, MCM1, and MEF2A. Our results are summarized in Fig. 10. We found that MADS-box positions 1 and 11 to 15 act as specificity determinants; in addition, the presence of sequences immediately N terminal to the MADS box can also affect binding specificity but only in the context of particular MADS-box sequences. Strikingly, these different determinants act in an essentially combinatorial fashion. For example, at position 1, the presence of glycine is compatible with binding to SRF or MEF2 consensus sequences, while glutamate is compatible with binding to the MCM1 consensus. At positions 11 to 15, the MEF2 sequence MDERN is compatible only with binding to MEF2 consensus sites. In contrast, according to their context, the corresponding MCM1 and SRF sequences are compatible with binding to SRF-, MCM1-, or MEF2-specific consensus sequences. Sequences immediately N terminal to the MADS box appear incompatible with binding to the MEF2 consensus sequence. Our results were obtained by using chimeric proteins in which the N-terminal basic part of the MADS

TABLE 1. Sequences and DNA binding properties of SRF variants selected in vivo

Variant	Sequence[a]		Binding activity[b]
	MADS-box positions 2–10	MADS-box positions 10–16	
Wild type	RVKIKMEFI	IDNKLRR	100
Library B[c]	R*KI*M**I	IDNKLRR	
B.4	-L--R-EL-	-------	130
B.11	-V--P-TI-	-------	121
B.10	-Q--K-QL-	-------	112
B.6	-V--Q-RH-	-------	105
B.7	-M--G-HY-	-------	98
B.5	-R--P-QH-	-------	94
B.3	-R--P-LL-	-------	93
B.2	-C--S-VR-	-------	81
B.9	-R--E-AL-	-------	60
B.12	-C--R-VF-	-------	58
B.1	-R--H-LF-	-------	48
B.8	-R--R-QW-	-------	39
Library C[d]	RVKIKMEFI	I*D/N***R	
C.16	---------	-NNRMS-	111
C.24	---------	-DDQTR-	92
C.14	---------	-QDRHG-	85
C.5	---------	-RNRHD-	84
C.20	---------	-QNTQS-	71
C.1	---------	-LNRSH-	70
C.13	---------	-RDSNR-	60
C.7	---------	-NNKET-	55
C.9	---------	-ADRRL-	52
C.22	---------	-SNVTK-	49
C.25	---------	-VDRIK-	48
C.19	---------	-NNKSL-	36
C.8	---------	-ENKGL-	36
C.4	---------	-QNRYG-	33
C.18	---------	-SNSLR-	32
C.6	---------	-KNSSN-	26
C.10	---------	-DDATK-	25
C.12	---------	-DDRLT-	21
C.23	---------	-RNCHT-	21
C.3	---------	-IDVKR-	19
C.21	---------	-RDARK-	15
C.17	---------	-SDKKE-	12
C.11	---------	-PDPRV-	11
C.2	---------	-QDGRQ-	9

[a] Sequences of the variants are compared with the wild-type SRF sequence; residues varied in each library are indicated by asterisks, and identities to the wild-type SRF sequence are indicated by dashes.

[b] DNA binding activity was determined by gel mobility shift assay (results are averages of two experiments).

[c] Sixty micrograms of plasmid DNA from library B generated 30,000 independent yeast colonies, of which 4.6% scored positive in the *lacZ* colony color assay.

[d] Sixty micrograms of plasmid DNA from library C generated 24,000 independent yeast colonies, of which 12.5% scored positive in the *lacZ* colony color assay.

box is joined to the dimerization region of SRF. However, all specificity-determining protein-DNA contacts must be maintained in these chimeras, since their binding specificities are identical to those of the parent proteins (17, 27).

In alternative approach, we used a yeast screening protocol to identify SRF variants capable of binding the SRF-specific ACT.L SRE in vivo from libraries of SRF mutants in which nonconserved MADS-box residues are randomized. Similar approaches have been used to study the roles of conserved residues in GCN4 (18), Tet repressor (3, 4), and 434 repressor (9). The results obtained in the analysis strongly supported the conclusions from the in vitro studies. Strikingly, the only SRF MADS-box position strongly selected in this analysis is position 1, which was invariably glycine; in contrast, the two other regions examined, MADS-box positions 3, 6, 8, and 9 and positions 11 to 15, showed that extreme variability in sequence can be tolerated with little effect on binding affinity for an SRF-specific site. These data strongly suggest that at least in SRF, these MADS-box residues do not make specificity-determining base contacts; however, it should be noted that a similar analysis of the yeast GCN4 protein showed that considerable variation can be tolerated at three positions involved in direct DNA contacts (18).

Although our experiments clearly demonstrate the importance of these elements for DNA binding specificity, other residues may be important in other MADS family members. For example, sequences bound by the *Arabidopsis* agamous protein, which possesses a glycine at MADS-box position 1, are similar to those bound by SRF but can tolerate significant mismatch at one side of the dyad (8); it will be interesting to analyze the roles of the elements identified here in agamous sequence specificity.

Role of MADS-box N-terminal flanking sequences. We found that SRFΔ142, an SRF truncation that lacks the sequences N terminal to the MADS box, exhibits an altered DNA binding specificity in that it can also bind the N10 MEF2 consensus site, in agreement with a previous report (22). Recognition of the MEF2 site is not a fortuitous consequence of a simple relaxation of binding specificity: in binding-site selection experiments, SRFΔ142 selects sequences that exactly match the MEF2A consensus (17). However, binding competition experiments show that the protein also retains the ability to bind the c-*fos* SRE and can still discriminate between SRF- and MCM1-specific sites. A single arginine residue N terminal to the MADS box is sufficient to restrict binding to SRF consensus sequences; however, this residue is not absolutely required, since in the context of the intact SRF N-terminal flanking sequences it can be changed to lysine or alanine without reducing specificity. In contrast, although sequences N terminal to the MCM1 MADS box are not required for its correct DNA binding specificity, these sequences can functionally substitute for those of SRF. This ability may arise from the conservation of a basic residue in the MCM1 flanking sequences, but it remains possible that the presence of N-terminal flanking sequences per se is sufficient to restrict the binding specificity of the SRF MADS box.

A simple interpretation of these observations is that the SRF MADS box contains a structure which can interact with DNA in two different ways, each compatible with a different spectrum of DNA sequences. Perhaps the SRF sequences immediately N terminal to the MADS box affect its structure such that it is rendered incapable of interaction with MEF2 consensus sequences. An alternative possibility is that the N-terminal sequences make additional, not necessarily sequence-specific, DNA contacts outside the consensus decamer and that this interaction is incompatible with binding to MEF2 consensus sites. Two observations are consistent with the N-terminal flanking region making DNA contacts. First, mutation of basic residues within the region causes a loss in DNA binding affinity but does not affect specificity. Second, in binding-site selection experiments, SRF, but not MCM1 or MEF2A, selects sites that contain additional sequence preferences to one side of the decamer consensus motif, as might be expected if SRF makes additional DNA contacts (16). The difference in mobility of SRFΔ142 complexes formed on MEF2 and c-*fos* SRE probes is consistent with the notion that the protein may adopt different conformations at the different sites.

Role of MADS-box position 1. Our data show that MADS-box residue 1, generally conserved as a glycine, is a major determinant of DNA binding specificity in MADS-box tran-

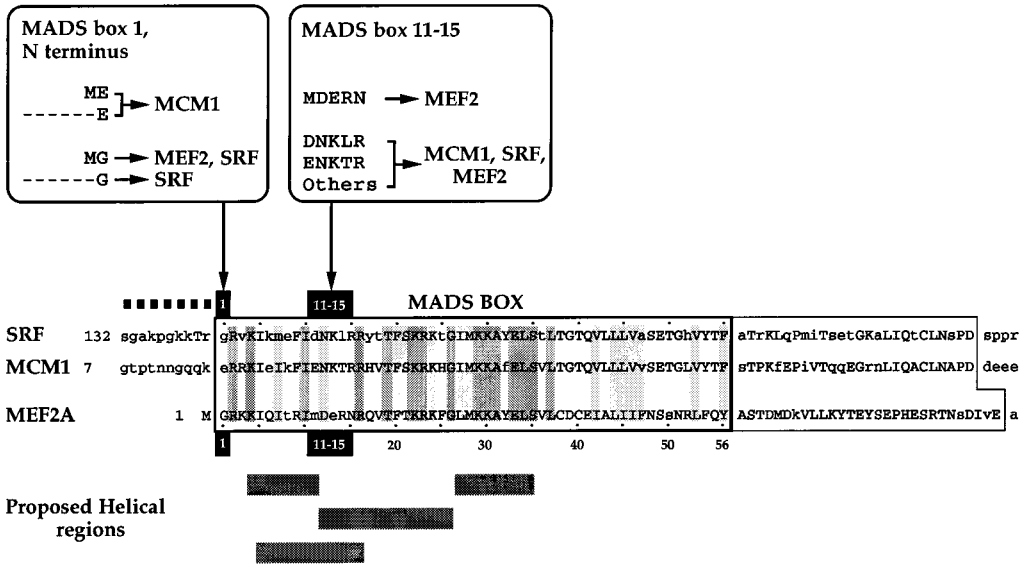


FIG. 10. A summary of results is shown at the top. The protein sequences are shown at the left of each panel, and the binding sites with which they are compatible are shown at the right. Results for the N-terminal flanking sequences and MADS-box position 1 are summarized together. The locations of the DNA binding specificity determinants in the SRF, MCM1, and MEF2A DNA binding domains are indicated on the sequences below by black bars (MADS-box positions 1 and 11 to 15) or by dots (N-terminal flanking sequences). Below are shown the sequences of the SRF, MCM1, and MEF2A DNA binding domains. The MADS box is surrounded by a thick line. A thin line encloses the C-terminal flanking residues, conserved among subgroups of MADS-box proteins, which are necessary for high-affinity DNA binding. Invariant MADS-box residues are darkly shaded, while conservatively substituted residues are lightly shaded. Uppercase letters indicate the residues conserved between members of the SRF/MCM1/ARG80 and MEF2A/B/C/D family subgroups. Below the sequences, proposals for helical secondary structures in the specificity-determining region are summarized (see references 17, 21, and 24 [top to bottom]). For a discussion, see the text.

scription factors. Mutation of SRF MADS-box position 1 from glycine to glutamate is sufficient to alter the binding specificity of both intact and truncated SRF to that of MCM1; in MEF2, MADS-box positions 1 and 11 to 15 together determine sequence specificity. Moreover, a glycine is invariably recovered at MADS-box position 1 when SRF variants are selected in vivo for the ability to bind the SRF-specific ACT.L SRE. The striking effect of mutations at MADS-box position 1 is consistent with this residue either affecting secondary structure in the basic N-terminal part of the MADS box or making direct DNA contacts. For example, if the N-terminal region of the MADS box were helical in character (see below), the glutamate at MCM1 MADS-box position 1 might stabilize an N-terminal extension of such a helix via a salt bridge with the lysine at MADS-box position 4; conversely, the increased main chain flexibility associated with a glycine at this position could destabilize a helix. A secondary function of the glycine present at SRF MADS-box position 1 may be to allow sequences N terminal to it to interact with either DNA or the MADS box itself, as proposed above. These interactions would presumably be precluded by the presence of glutamate at this position, since neither the affinity nor the specificity of MCM1 binding is affected by the presence of sequences N terminal to the MADS box.

The presence of glutamate at MADS-box position 1 is essential for MCM1-type binding specificity. We have observed that aspartate at this position also generates MCM1-type specificity (unpublished observations), but it remains possible that other residues at this position generate further specificities. It would be interesting to evaluate this possibility by selection of MCM1 variants in vivo. Two other MADS-box proteins, ARG80 and deficiens, lack a glycine at position 1, containing instead threonine and alanine, respectively. Little is known about their binding specificity, although ARG80 may bind with an MCM1-related specificity distinct from SRF (12).

Role of MADS-box positions 11 to 15. We found that MADS-box residues 11 to 15 are required as a crucial specificity determinant in MADS-box proteins; in the case of MEF2A, residues 11 to 13 appear sufficient for specificity. However, these sequences are not necessarily sufficient to determine sequence specificity, but rather they act in concert with the N-terminal and MADS-box position 1 sequences. Using MCM-MEF2 chimeras, we found that no single residue among MADS-box residues 11 to 15 was sufficient to alter specificity: substitution of at least residues 11 to 13 was required for high-affinity specific binding. A previous report suggested that MADS-box residue 13 alone might be sufficient to determine specificity in the context of the SRF MADS box (22); however, it is unlikely that the relatively insensitive direct binding assay used in that study would detect weak binding to the c-*fos* SRE, which in our hands was detectable only by the indirect binding competition assay.

Selection of SRF variants in vivo demonstrated that a large amount of sequence variation at MADS-box positions 11 to 15 is compatible with efficient binding to an SRF-specific site. We favor the idea that MADS-box positions 11 to 15, like the SRF N-terminal flanking residues, are involved in orienting a recognition surface with respect to DNA. According to this view, both MCM1 and SRF MADS-box positions 11 to 15 would allow flexibility in orientation such that MCM1-, SRF-, or MEF2-specific consensus sequences could be recognized. In MEF2A, however, MADS-box positions 11 to 15 would not allow orientation of the recognition surface in the manner required for binding to the SRF or MCM site. Consistent with this, substitution of SRF or MCM1 MADS-box positions 11 to 15 by the corresponding MEF2A sequences prevents sequence-specific high-affinity DNA binding. Different members of the MEF2 family differ at these positions (19, 31). However, since the sequence of MADS-box positions 11 to 15 appears to have a more restrictive role in DNA binding specificity in

MEF2A than in SRF, it might be expected that fewer sequences are compatible with correct DNA binding by MEF2, and it will be interesting to test this possibility.

What is the DNA recognition structure? A number of proposals concerning potential secondary-structure elements within the MADS box have been advanced (Fig. 10). According to the combined secondary-structure prediction algorithm of Eliopoulos et al. (6), MADS-box residues 4 to 11, 27 to 35, and possibly 37 to 45 generally have helical character (14, 17). However, the yeast ARG80 protein contains a proline at MADS-box residue 6, and our data show that the SRF N-terminal region can tolerate the presence of helix-breaking glycine or proline residues. In contrast, structural predictions from the EMBL-PHD program (19) suggests that MADS-box residues 3 to 10 are mostly β sheet. Whatever the structure present between MADS-box residues 4 and 11, it is interesting that it is flanked by the specificity determinants mapped here, in agreement with the idea that these determinants may play a role in its orientation. Two alternative structural proposals arise from consideration of MADS-box sequence conservation. The first, prompted by the (i, i + 3/4) spacing of invariant MADS box residues and ability of MADS family members to bind $CC(A/T)_6GG$-related consensus sequences, is that MADS-box residues 12 to 26 form a helix (21). A second model, prompted by comparison of MADS boxes with DNA recognition motifs of known structure, is that MADS-box residues 5 to 16 form a fixed probe helix in which positions 5, 8, 9, and 12 make base-specific contacts (24). This model is not supported by our data concerning the identity of these positions in our SRF variants, nor does it suggest a role for MADS-box residue 1 as a specificity determinant.

In conclusion, our work shows that the differential DNA binding specificity of MADS-box transcription factors is governed by three determinants, which act to some extent combinatorially. Many residues which differ between MADS-box family members do not appear to contribute substantially to differential sequence specificity. We propose that the specificity determinants are involved in the structure and orientation of a DNA recognition surface located at the N terminus of the MADS box. Our analysis does not address the nature of any MADS box-DNA contacts that are conserved between family members, a likely possibility given the generally A/T-rich character of different MADS-box factor consensus sequences, and it remains possible other regions play a role in such contacts. Indeed, the dependence of SRF DNA binding on the basic residues at MADS-box positions 16, 23, and 24, which are highly conserved through the MADS-box family, may reflect such conserved contacts (21). It will be interesting to see how our observations relate to the structure of individual MADS-box protein–DNA complexes.

ACKNOWLEDGMENTS

This work was supported by the Imperial Cancer Research Fund and the Howard Hughes Medical Institute. R.T. is an HHMI International Research Scholar.

We thank Steve Dalton for plasmids and advice on yeast techniques and Ian Goldsmith and his team for oligonucleotide synthesis. We thank Justin Blau, Paul Freemont, Steve Goodbourn, Nic Jones, and members of the laboratory for technical advice, helpful discussions, and comments on the manuscript.

REFERENCES

1. **Ammerer, G.** 1990. Identification, purification, and cloning of a polypeptide (PRTF/GRM) that binds to mating-specific promoter elements in yeast. Genes Dev. **4:**299–312.
2. **Angel, P., I. Baumann, B. Stein, H. Delius, H. J. Rahmsdorf, and P. Herrlich.** 1987. 12-*O*-Tetradecanoyl-phorbol-13-acetate induction of the human collagenase gene is mediated by an inducible enhancer element located in the 5′-flanking region. Mol. Cell. Biol. **7:**2256–2266.
3. **Baumeister, R., V. Helbl, and W. Hillen.** 1992. Contacts between Tet repressor and tet operator revealed by new recognition specificities of single amino acid replacement mutants. J. Mol. Biol. **226:**1257–1270.
4. **Baumeister, R., G. Muller, B. Hecht, and W. Hillen.** 1992. Functional roles of amino acid residues involved in forming the alpha-helix-turn-alpha-helix operator DNA binding motif of Tet repressor from Tn10. Proteins **14:**168–177.
5. **Dalton, S., and R. Treisman.** 1992. Characterization of SAP-1, a protein recruited by serum response factor to the c-fos serum response element. Cell **68:**597–612.
6. **Eliopoulos, E. E., A. J. Geddes, M. Brett, D. J. C. Pappin, and J. B. C. Findley.** 1982. A structural model of the chromophore binding domain of ovine rhodopsin. Int. J. Biol. Macromol. **4:**263–268.
7. **Hill, C. S., R. Marais, S. John, J. Wynne, S. Dalton, and R. Treisman.** 1993. Functional analysis of a growth factor-responsive transcription factor complex. Cell **73:**395–406.
8. **Huang, H., Y. H. Mizukami, and M. Hong.** 1993. Isolation and characterization of the binding sequences for the product of the *Arabidopsis* floral homeotic gene AGAMOUS. Nucleic Acids Res. **21:**4769–4776.
9. **Huang, L., T. Sera, and P. G. Schultz.** 1994. A permutational approach toward protein-DNA recognition. Proc. Natl. Acad. Sci. USA **91:**3969–3973.
10. **Jarvis, E. E., K. L. Clark, and G. F. Sprague, Jr.,** 1989. The yeast transcription activator PRTF, a homolog of the mammalian serum response factor, is encoded by the MCM1 gene. Genes Dev. **3:**936–945.
11. **Marais, R., J. Wynne, and R. Treisman.** 1993. The SRF accessory protein Elk-1 contains a growth factor-regulated transcriptional activation domain. Cell **73:**381–393.
12. **Messenguy, F., and E. Dubois.** 1993. Genetic evidence for a role for MCM1 in the regulation of arginine metabolism in *Saccharomyces cerevisiae*. Mol. Cell. Biol. **13:**2586–2592.
13. **Mueller, C. G., and A. Nordheim.** 1991. A protein domain conserved between yeast MCM1 and human SRF directs ternary complex formation. EMBO J. **10:**4219–4229.
14. **Norman, C., M. Runswick, R. Pollock, and R. Treisman.** 1988. Isolation and properties of cDNA clones encoding SRF, a transcription factor that binds to the c-*fos* serum response element. Cell **55:**989–1003.
15. **Passmore, S., G. T. Maine, R. Elble, C. Christ, and B. K. Tye.** 1988. Saccharomyces cerevisiae protein involved in plasmid maintenance is necessary for mating of MAT alpha cells. J. Mol. Biol. **204:**593–606.
16. **Pollock, R., and R. Treisman.** 1990. A sensitive method for the determination of protein-DNA binding specificities. Nucleic Acids Res. **18:**6197–6204.
17. **Pollock, R., and R. Treisman.** 1991. Human SRF-related proteins: DNA-binding properties and potential regulatory targets. Genes Dev. **5:**2327–2341.
18. **Pu, W. T., and K. Struhl.** 1991. Highly conserved residues in the bZIP domain of yeast GCN4 are not essential for DNA binding. Mol. Cell. Biol. **11:**4918–4926.
19. **Rost, B., and C. Sander.** 1993. Prediction of protein structure at better than 70% accuracy. J. Mol. Biol. **232:**584–599.
20. **Schwarz-Sommer, Z., P. Huijser, W. Nacken, H. Saedler, and H. Sommer.** 1990. Genetic control of flower development by homeotic genes in *Antirrhinum majus*. Science **250:**931–936.
21. **Sharrocks, A. D., H. Gille, and P. E. Shaw.** 1993. Identification of amino acids essential for DNA binding and dimerization in $p62^{SRF}$: implications for a novel DNA-binding motif. Mol. Cell. Biol. **13:**123–132.
22. **Sharrocks, A. D., F. von Hesler, and P. E. Shaw.** 1993. The identification of elements determining the different DNA binding specificities of the MADS box proteins $p67^{SRF}$ and RSRFC4. Nucleic Acids Res. **21:**215–221.
23. **Sommer, H., J. P. Beltran, P. Huijser, H. Pape, W. E. Lonnig, H. Saedler, and S. Z. Schwarz-Sommer.** 1990. Deficiens, a homeotic gene involved in the control of flower morphogenesis in Antirrhinum majus: the protein shows homology to transcription factors. EMBO J. **9:**605–613.
24. **Suzuki, S.** 1993. Common features in DNA recognition helices of eukaryotic transcription factors. EMBO J. **12:**3221–3226.
25. **Treisman, R.** 1987. Identification and purification of a polypeptide that binds to the c-fos serum response element. EMBO J. **6:**2711–2717.
26. **Treisman, R. H., and G. Ammerer.** 1991. The SRF and MCM1 transcription factors. Curr. Opin. Genet. Dev. **2:**221–226.
27. **Wynne, J., and R. Treisman.** 1992. SRF and MCM1 have related but distinct DNA binding specificities. Nucleic Acids Res. **20:**3297–3303.
28. **Yanofsky, M. F., H. Ma, J. L. Bowman, G. N. Drews, K. A. Feldmann, and E. M. Meyerowitz.** 1990. The protein encoded by the Arabidopsis homeotic gene agamous resembles transcription factors. Nature (London) **346:**35–39.
29. **Yu, Y.-T., R. E. Breitbart, L. B. Smoot, L. Youngsook, V. Mahdavi, and B. Nadal-Ginard.** 1992. Human myocyte-specific enhancer factor 2 comprises a group of tissue-restricted MADS box transcription factors. Genes Dev. **6:**1783–1798.