Nucleic Acids Research, 1993, Vol. 21, No. 2 215–221

The identification of elements determining the different DNA binding specificities of the MADS box proteins p67^{SRF} and RSRFC4

Andrew D.Sharrocks*, Friedrich von Hesler and Peter E.Shaw
Max-Planck-Institut für Immunbiologie, Spemann Laboratories, Postfach 1169, Stübeweg 51, 7800 Freiburg, Germany

Received October 26, 1992; Revised and Accepted December 7, 1992

ABSTRACT

The human transcription factors p67^{SRF} and RSRFC4 recognise similar but distinct binding sites which are found in the promoters of both muscle-specific and 'immediate early' genes. Both proteins share a common basic DNA-binding domain, which is defined by the MADS box homology region. The DNA-binding specificity of a truncated form of p67^{SRF} (core^{SRF}) can be converted to that of RSRFC4. Removal of residues immediately N-terminal to the MADS box relaxes the specificity of core^{SRF} for its cognate sequence (CC(A/T)$_6$GG) as it improves binding to the RSRFC4 site (CTA(A/T)$_4$TAG). Moreover, the introduction of a single, additional mutation, K154E, into the N-terminal truncated derivative completes the change in specificity to the RSRFC4 binding site. It also influences the salt dependence of DNA binding and ternary complex formation with p62^{TCF}. However, residues at this position do not appear to be involved in direct base-pair recognition. These results indicate that although the DNA binding specificity of p67^{SRF} can be converted to that of RSRFC4, the two proteins may bind DNA in different ways. Furthermore, they suggest that binding site specificity can be determined by an indirect mechanism involving residues which are not directly involved in base recognition.

INTRODUCTION

The eukaryotic transcription factor p67^{SRF} has been implicated in the transcriptional control of the both muscle-specific genes and growth factor-induced 'immediate-early' genes (1,2). In the case of the well studied immediate-early gene c-*fos*, p67^{SRF} forms a ternary complex with a second protein, p62^{TCF}, and the serum response element (SRE) (3,4). It appears that through this ternary complex the transcriptional induction of the c-*fos* gene in response to both serum growth factors and mitogens is elicited (3). Consistent with this interpretation is the recent demonstration that binding of p62^{TCF} is stimulated upon direct phosphorylation by MAP kinases in response to extracellular growth factor stimulation (5). Moreover, the binding site for p67^{SRF} has been implicated in the rapid shut-off and subsequent repression of the c-*fos* promoter following mitogenic stimulation (6,7, 8,9,10,11,12).

p67^{SRF} contains a distinct basic DNA-binding motif within its DNA binding domain which, in common with many DNA-binding proteins, appears to utilise an α-helix for interacting with DNA (13). This motif is part of the MADS box region whose sequence is conserved among a family of related DNA-binding proteins including MCM1, Arg80/AG, DEFA, and p67^{SRF} (14). Residues that are essential for dimerisation also map within the MADS box (13). Several MADS box proteins have been shown to bind to sequences related to the p67^{SRF} binding site CC(A/T)$_6$GG (also referred to as CArG box) (4,13,15,16,17). However, two human DNA-binding proteins from the same family, RSRFC4 and RSRFR2, possess an altered DNA-binding specificity and bind to the sequence CTA(A/T)$_4$TAG (18). Binding sites for these proteins are also found in the promoters of muscle-specific and immediate-early genes, suggesting a similar, but distinct, *in vivo* role to that of p67^{SRF} (18). In this study we have examined the mechanism that underlies the different DNA-binding specificities exhibited by p67^{SRF} and the RSRF proteins.

Amino acids at the N-terminus of the DNA-binding domains contribute to sequence specificity in the case of p67^{SRF}, and to DNA-binding affinity in the case of RSRFC4. In addition, a single amino acid, which maps to the putative recognition helix within the basic region of p67^{SRF} (13), is involved in sequence specificity determination, albeit without apparent participation in DNA contacts. Alteration of this residue also influences the salt dependence of DNA binding and ternary complex formation with p62^{TCF}. These observations support the notion that, despite their high degree of amino acid similarity, p67^{SRF} and the RSRFs may bind DNA in distinct ways.

* To whom correspondence should be addressed at Department of Biochemistry and Genetics, The Medical School, The University, Newcastle-upon-Tyne NE2 4HH, UK

MATERIALS AND METHODS

Plasmid constructions and mutagenesis

The plasmid pBOC3st encodes p67SRF amino acids 132–222 (coreSRF) (4,13). pAS37 (encoding METcoreSRF; amino acids 142–222 from p67SRF) was constructed by PCR amplification of coreSRF sequences with the oligonucleotides MET1(5′-GGGTAAGCTTACCATGGGCCGCGTGAAGATC) and FOR (GTAAAACGACGGCCAGT) followed by cleavage with HindIII and BamHI and subsequent ligation into BamHI/HindIII-cut pBOC3st. pAS54 (encoding C4coreSRF; amino acids 1–30 from RSRFC4 fused to amino acids 170–222 from p67SRF) was constructed by ligating the NcoI/PstI fragment from pT7C4/SRF (a gift from R. Treisman, 18) into pAS37 cut by the same enzymes.

The plasmid pAS63 was constructed as follows. A ClaI site was introduced into pBOC3st by 'single primer' PCR (19) to produce pBOC4st. A 'cassette' (encoding RSRFC4 amino acids 13–30) was synthesised by primer extension directed by the oligonucleotides (template: 5′-GCTCATCGATGACGAGAGG AACCGCCAGGTGACGTTCACCAAGAGGAAGTT-CGGCATCATGAAGAAGG and primer: 5′-CCTTCTT CATGATGCC). The double-stranded product was cleaved with ClaI (underlined) and inserted into pBOC4st cleaved with ClaI and StuI. After amplification with PCR primers MET1 and FOR and restriction with NcoI and PstI the resulting fragment was inserted into pAS37 cut with the same enzymes.

Plasmids pAS38-49 and pAS51-53 (encoding the respective AS proteins) were constructed and the mutations K154E and K154A were introduced into coreSRF and METcoreSRF by 'single-primer' PCR as described previously (13,19). Details of the oligonucleotides used and plasmid constructions can be obtained upon request. The sequences of all plasmids encoding mutant proteins were verified by dideoxy sequencing with Sequenase™ according to the manufacturer's instructions.

In vitro transcription and translation

Plasmids encoding mutant proteins were linearised with BamHI and used as templates for transcription by T3 RNA polymerase. RNA thus produced was translated in the presence of ^{35}S-methionine in rabbit reticulocyte lysates according to the manufacturer's instructions (Promega). Translations were verified by electrophoresis of an aliquot through SDS-polyacrylamide gels followed by autoradiography.

Gel retardation analysis

Gel retardation analysis of mutant proteins generated by in vitro translation was carried out as described previously (13). End-labelled DNA fragments were prepared from the following oligonucleotide pairs: N10A (5′-TCGAAGGAAAA CTATTTATAGATCAAAT) & N10B (5′-TCGAATTTGAT CTATAAATAGTTTTCCT) (12), M1A (5′-TCGAAGG AAAACCATTTATGGATCAAAT) & M1B (5′-TCGAATTT GATCCATAAATGGTTTTCCT) (12) and SRN10A (5′-GG CCACACAGGATGTCTATTTATAGATCAAAT) & SRN 10B (5′-GGCCATTTGATCTATAAATAGACATCCTGTGT). Residues defining 'core' consensus binding sites are underlined. Equivalent amounts of protein were used in each binding reaction unless otherwise indicated. DNA-protein complexes were quantified by scintillation counting of bands excised from dried gels. The enriched fraction of p62TCF was prepared from glycerol gradient fractions depleted of p67SRF by wheatgerm aglutinin affinity chromatography as previously described (3,20).

RESULTS

Deletion of the N-terminus of coreSRF changes its DNA-binding specificity

It has previously been shown that a 90 amino acid moiety of p67SRF (coreSRF) is capable of dimerising and binding specifically to the CArG box. It is also able to recruit p62TCF into a ternary complex with the c-*fos* SRE (4,16,20,21). A mutagenic study on coreSRF identified several amino acids essential for DNA-binding (13), which lie on the face of a putative α-helix (Fig. 1A).

RSRFC4 recognises a different binding site (CTA(A/T)$_4$TA-G) to that of p67SRF. In this study we utilised the RSRFC4 N10 binding site (18) and a mutant binding site M1 (18), which shows only two changes from the N10 site; a T-C transition at position −4 and a A-G transition at position +4 to recreate a CArG box (fig. 1B). p67SRF binds to the M1 site with a 10-fold lower affinity than to the c-*fos* SRE, but with a 10 fold higher affinity than to the N10 site (18). In accordance with this, coreSRF binds to the M1 site with a consistently greater affinity than to the N10 site over a wide range of salt concentrations (fig. 2). Under optimal binding conditions, more than a 20-fold difference in affinity is observed (fig. 2B).

An obvious characteristic of the RSRF proteins that distinguishes them from other members of the MADS box family is the lack of amino acids N-terminal to the MADS box. Thus a 10 amino acid truncation of coreSRF was carried out to produce METcoreSRF, which has an N-terminus analogous to that of RSRFC4 (fig. 2C). METcoreSRF has a drastically reduced affinity for the M1 binding site, but a concomitantly increased affinity for the N10 binding site (fig. 2A). It should be noted though, that this binding to N10 follows a salt-dependent profile similar to that of coreSRF binding to the M1 oligonucleotide (fig. 2A and B). This indicates that one or more residues in the N-terminal 10 amino acids of coreSRF strongly influence the ability of coreSRF to discriminate between A:T and G:C pairs at positions ±4 in its binding site. As methylation/carboxymethylation interference experiments indicate that coreSRF is in close contact with guanine at these positions (1,22) and RSRFC4 is in contact with the adenine at the analogous positions (18), it is probably purines at positions ±4 that are important for determining the respective DNA-binding

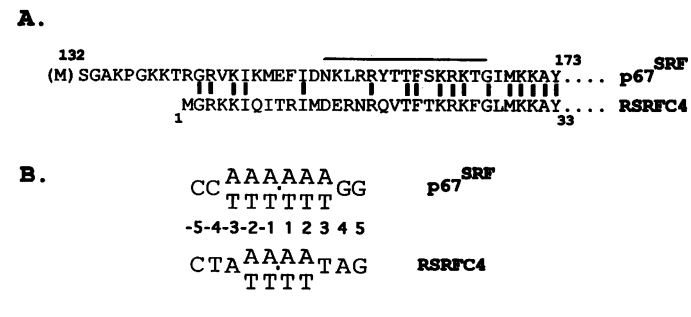

Figure 1. The DNA-binding domains and binding sites of p67SRF and RSRFC4. **(A)** Residues known to be essential for determining the DNA-binding specificity of p67SRF (20,21) and RSRFC4 (18) are shown. Vertical lines indicate residues conserved between the two proteins. The bracketed methionine is an artificial initiator residue to allow *in vitro* expression of coreSRF. Residues that constitute the putative DNA-binding helix (13) are overlined. **(B)** The 'core' consensus DNA-binding sites of p67SRF and RSRFC4 as deduced by selection from pools of random oligonucleotides in vitro (15,18) are shown.

specificities, although a role for the two pyrimidines cannot be excluded.

In common with METcore^{SRF}, the hybrid protein C4core^{SRF}, which contains the RSRFC4 DNA-binding domain linked to the core^{SRF} dimerisation domain (fig. 2C), binds preferentially to the N10 site over a similar range of salt concentrations (fig. 5A). In this case, however, the binding to the N10 site is absolute with no detectable binding to the M1 site. This sequence specificity is consistent with results obtained with a larger fusion protein (18). In addition, C4core^{SRF} has a distinct salt-dependent DNA-binding profile (fig. 5B), with maximal binding observed at 41mM KCl in contrast to that exhibited by core^{SRF} and METcore^{SRF} (166mM). This suggests that despite their primary sequence similarity, METcore^{SRF} and C4core^{SRF} might bind to DNA in distinct ways.

The sequence specificity determinant maps to the N-terminal end of the putative recognition helix

RSRFC4 clearly has an absolute requirement for adenine at position ±4 in its binding site throughout a wide range of salt concentrations. In contrast, at low salt concentrations, both core^{SRF} and METcore^{SRF} are less able to discriminate between purines at this position. We took advantage of this fact in order to locate the residues which determine the absolute specificity of C4core^{SRF} for the N10 binding site. To this end, a series of chimaeric METcore^{SRF}/RSRF proteins was created (fig. 3A). The mutant protein AS38, which has a four amino acid substitution between RSRFC4 and METcore^{SRF}, shows a phenotype that is characteristic of RSRFC4, i.e. exclusive binding to the N10 site (fig. 3B). In contrast, the alteration of five other N-terminal amino acids (fig. 3B) to either those found in RSRFR2 (AS39) or RSRFC4 (AS41) does not produce a protein that can bind the N10 site (fig. 3B). In addition, neither of these proteins can bind to the M1 site. This may be due to incorrect folding of the protein upon the introduction of multiple amino acid changes. In contrast, AS38 clearly does fold in a manner compatible with strong binding to the N10 site. The results indicate that the residues determining the sequence specificity of C4core^{SRF} reside in a four amino acid stretch which maps to the N-terminus of the putative DNA-binding a-helix (13). The simultaneous incorporation of all nine RSRF changes into METcore^{SRF} (AS40 + AS42) abrogates binding to both the N10 site and the M1 site. However, the lack of DNA-binding shown here is again possibly the result of incorrect folding of the proteins (see below).

It is conceivable that the same residues are involved in DNA-binding by core^{SRF} and RSRFC4, but contact different bases in each case. This would require a re-orientation of the DNA-binding interface in the major groove. Such a change could be governed by residues flanking the 'hinge' glycine residue located between the DNA-binding and dimerisation domains (13). The respective RSRFC4 amino acids flanking this glycine were therefore inserted into METcore^{SRF} and various derivatives. However, no change of specificity was obtained and only AS44

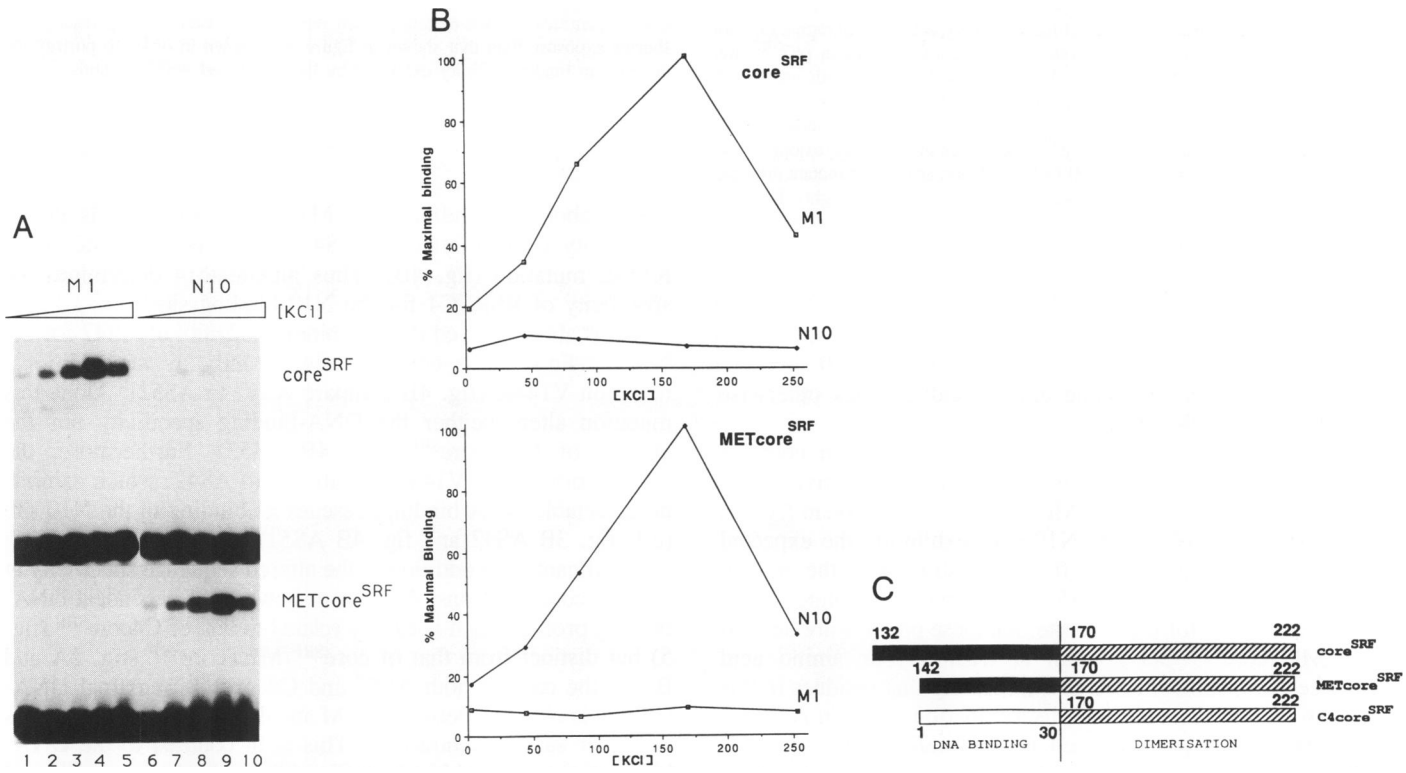

Figure 2. The unique N-terminus of core^{SRF} is a major determinant of its DNA-binding specificity **(A)** Gel retardation analysis of core^{SRF} (top) and METcore^{SRF} (bottom) binding to the M1 oligonucleotide (lanes 1–5) or N10 oligonucleotide (lanes 6–10). KCl concentrations in the binding reactions are; OmM (lanes 1 & 6), 41mM (lanes 2 & 7), 83mM (lanes 3 & 8), 166mM (lanes 4 & 9), 250mM (lanes 5 & 10). **(B)** The quantity of DNA bound by core^{SRF} (top) and METcore^{SRF} (bottom) was determined at each salt concentration in relation to the maximal binding observed. Open symbols represent values for binding to the M1 oligonucleotide, closed symbols represent binding to the N10 oligonucleotide. **(C)** Schematic representation of the structure of core^{SRF} and chimaeric proteins used in this study. The portion of core^{SRF} encoding sequence specificity is indicated by a solid box whereas that of RSRFC4 is indicated by an open box. The core^{SRF} dimerisation domain is indicated by a hatched box. Coordinates of p67^{SRF}-derived residues are shown above whereas those from RSRFC4 are shown below the boxes.

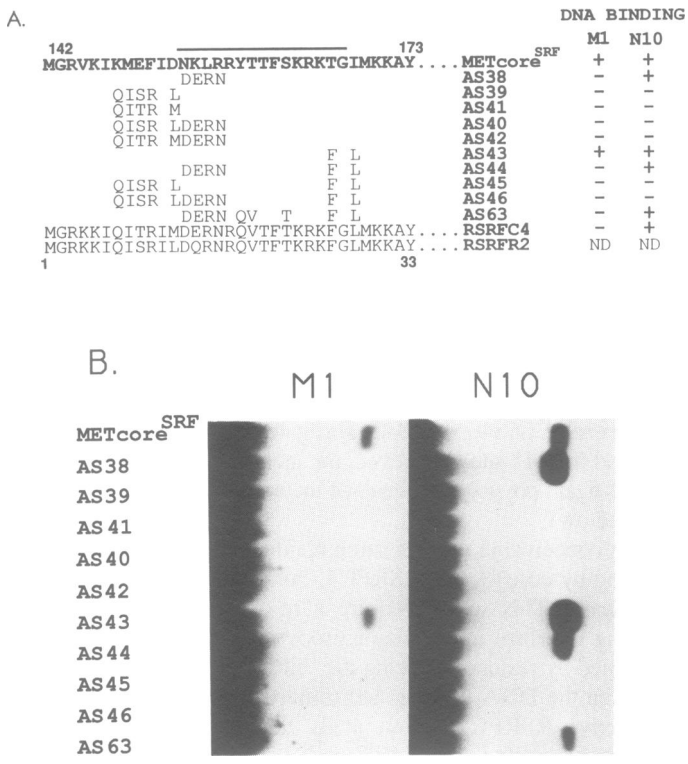

Figure 3. Residue(s) determining the absolute specificity of RSRFC4 reside within a four amino acid block. **(A)** METcore^{SRF}/C4core^{SRF} chimaeric proteins. Residues within the RSRFC4 and RSRFR2 specificity-determining region (amino acids M1−Y33) are shown at the bottom. Analogous residues in p67^{SRF} are shown at the top in bold (amino acids G142−Y173). RSRF amino acids introduced into METcore^{SRF} in mutant derivatives are indicated between these sequences. Detectable DNA-binding as judged by gel retardation analysis is indicated on the right. The line above the METcore^{SRF} sequences indicates the extent of the putative DNA-binding α-helix (13). **(B)** Gel retardation analysis of mutant proteins with the M1 oligonucleotide (left) and the N10 oligonucleotide (right). Binding reactions were carried out at 30mM KCl. Equal quantities of each mutant protein were used in each binding reaction.

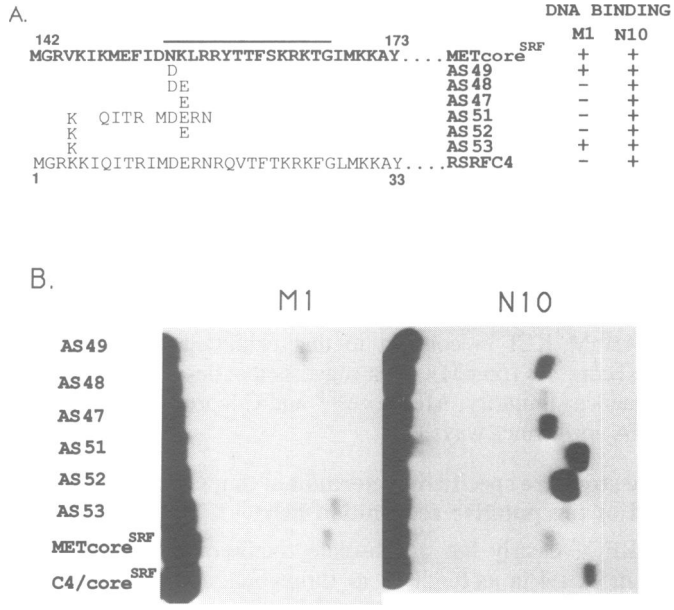

Figure 4. The absolute specificity of RSRFC4 is determined by a single amino acid residue. **(A)** The sequence of METcore^{SRF}/C4core^{SRF} chimaeric proteins is shown. Residues within the RSRFC4 specificity-determining region are shown at the bottom (amino acids M1−Y33). Analogous residues in p67^{SRF} are shown at the top in bold (amino acids G142−Y173). RSRF amino acids introduced into METcore^{SRF} in mutant derivatives are indicated between these sequences. **(B)** Gel retardation analysis of mutant proteins with the M1 oligonucleotide (left) and the N10 oligonucleotide (right). Binding reactions were carried out at 30mM KCl. Equal quantities of each mutant protein were used in each binding reaction. A shorter exposure than that shown in figure 3 is shown in order to portray the increase in binding affinity exhibited by the AS51 and AS52 proteins.

retained the ability to interact uniquely with the N10 site (fig. 3B). AS44 contains the same amino acid changes otherwise present in AS38 (fig. 3A).

Finally, a hybrid protein was constructed which contains RSRFC4 amino acids 13−30 (encompassing the putative DNA-binding α-helix) inserted into METcore^{SRF}. This protein (AS63) again bound uniquely to the N10 site, exhibiting the expected RSRFC4 DNA-binding specificity. In summary, the mutant proteins AS38, AS44 and AS63 all exhibit an absolute DNA-binding specificity for the N10 site. All these proteins are derived from METcore^{SRF} but contain at least a four amino acid sequence derived from RSRFC4, indicating that residues in this block are involved in efficient discrimination between A:T and G:C base-pairs at positions ±4 in the binding site.

Glutamate14 determines the sequence specificity of RSRFC4

In order to precisely identify the amino acids that determine the specificity of DNA recognition, single point mutations were introduced into the four amino acid sequence identified above as critical for this function (fig. 4A). The mutation N153D does not affect the sequence specificity of METcore^{SRF} (AS49; fig. 4B). In contrast, the double mutation, N153D/K154E (AS48) clearly abolishes binding to the M1 site. Moreover, this altered specificity is manifested by AS47, which contains the single K154E mutation (fig. 4B). Thus glutamate14 determines the specificity of RSRFC4 for the N10 binding site.

It was also observed that the binding affinity of AS47 for the N10 binding site is enhanced significantly by the additional mutation V144K (fig. 4B compare AS47 to AS52). Alone this mutation alters neither the DNA-binding specificity nor the affinity of METcore^{SRF} (fig. 4B AS53). Furthermore, the introduction of the V144K mutation into AS42, which exhibits no detectable DNA binding, rescues its binding to the N10 site (c.f. fig. 3B AS42 and fig. 4B AS51).

Significantly, in addition to the altered sequence specificity at low salt concentrations, AS52 also shows a salt-dependent DNA-binding profile which is clearly related to that of C4core^{SRF} (fig. 5) but distinct from that of core^{SRF}/METcore^{SRF} (fig. 2A and B). In the case of both AS52 and C4core^{SRF}, maximal DNA-binding is obtained between 0mM and 83mM KCl and decreases at higher salt concentrations. This is in contrast to the DNA-binding maxima at 166mM KCl exhibited by both core^{SRF} and METcore^{SRF}. These observations are consistent with a role for lysine at this position in stabilising the RSRF protein structure and thus increasing the affinity of binding. Interestingly, all complexes formed with proteins containing both the K154E and V144K mutations exhibit lower mobility (fig. 4B; AS51, AS52, AS53 and C4core^{SRF}). This may be indicative of an altered structure of the complexes.

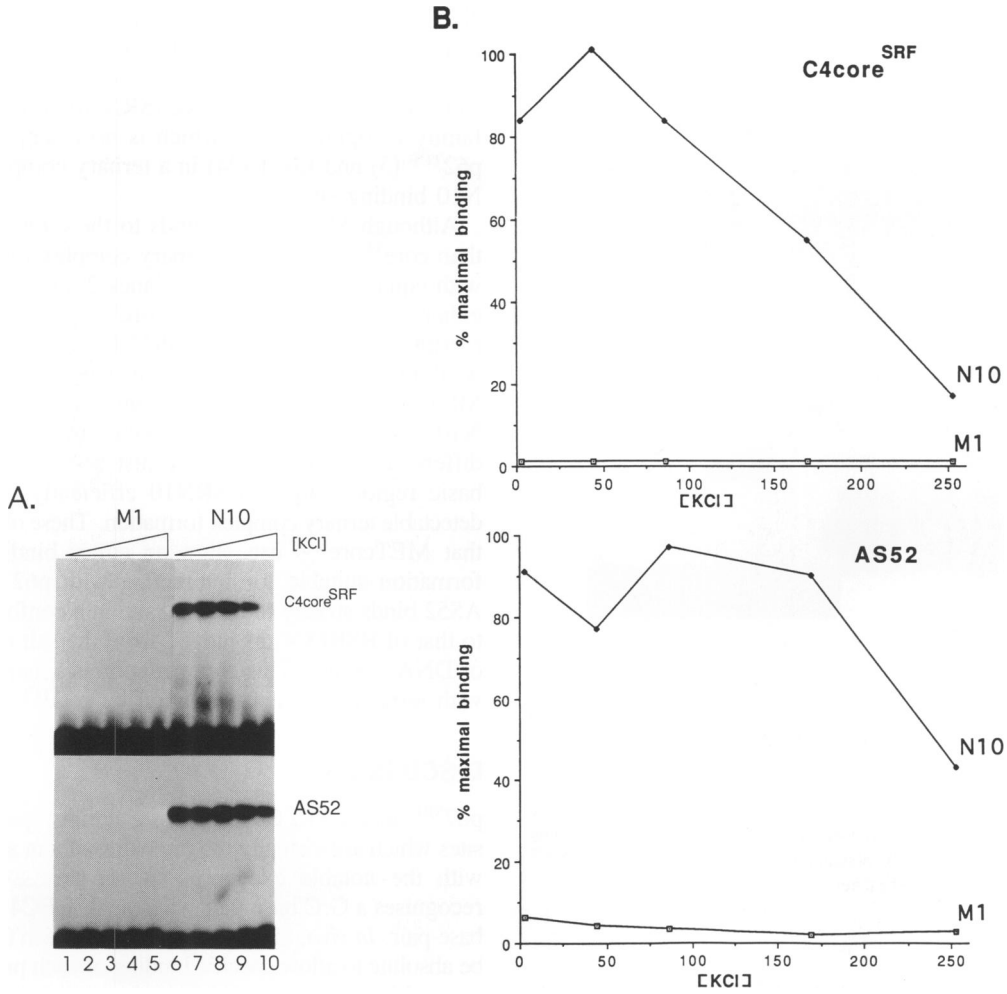

Figure 5. The mutation K154E in METcore^{SRF} recreates the DNA-binding specificity of RSRFC4. **(A)** Gel retardation analysis of C4core^{SRF} (top) and AS52 (bottom) binding to the M1 oligonucleotide (lanes 1–5) or N10 oligonucleotide (lanes 6–10). KC1 concentrations in the binding reactions are; 0mM (lanes 1 & 6), 41mM (lanes 2 & 7), 83mM (lanes 3 & 8), 166mM (lanes 4 & 9), 250mM (lanes 5 & 10). **(B)** The quantity of DNA bound by C4core^{SRF} (top) and AS52 (bottom) was determined at each salt concentration in relation to the maximal binding observed. Open symbols represent values for binding to the M1 oligonucleotide, closed symbols binding to the N10 oligonucleotide.

In summary, the introduction of the mutation K154E into METcore^{SRF} causes an alteration in DNA-binding specificity. The resulting protein, AS47, binds with absolute specificity to the N10 oligonucleotide, exactly as RSRFC4 itself. Moreover, the salt-sensitivity of DNA-binding is very similar to that of C4core^{SRF}. A second mutation located 10 amino acids N-terminally appears to stabilise proteins that contain the K154E mutation.

Neither lysine154 in p67^{SRF} nor glutamate14 in RSRFC4 directly determine the recognition of positions ±4 in the DNA-binding site

The substitution of lysine154 with glutamate abolishes the binding of METcore^{SRF} to the M1 site. A simple explanation of this result would be that lysine154 directly determines the recognition of G:C base-pairs at positions ±4. This cannot be the case however, as the introduction of the same mutation into core^{SRF} does not impair binding to the M1 site (fig. 6, lane 5). Moreover, the introduction of an alanine residue, a 'loss of contact' mutation (23), in place of lysine154 does not alter binding to the M1 site,

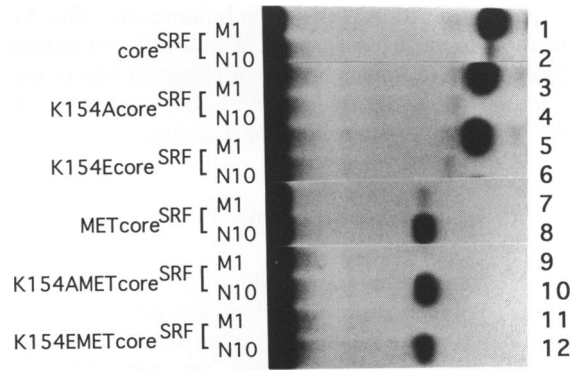

Figure 6. Lysine154 does not specify a base-pair contact. Gel retardation analysis of wild-type and mutant core^{SRF} and METcore^{SRF} derivatives. DNA-binding reactions were set up with the indicated mutant proteins and either the M1 oligonucleotide (lanes 1,3,5,7,9,11) or N10 oligonucleotide (lanes 2,4,6,8,10,12). Reactions were carried out at 41mM KCl. All reactions contained equal amounts of protein except lanes 1,2,7 & 8, where more protein was added to achieve levels of binding equivalent to those observed with mutant proteins.

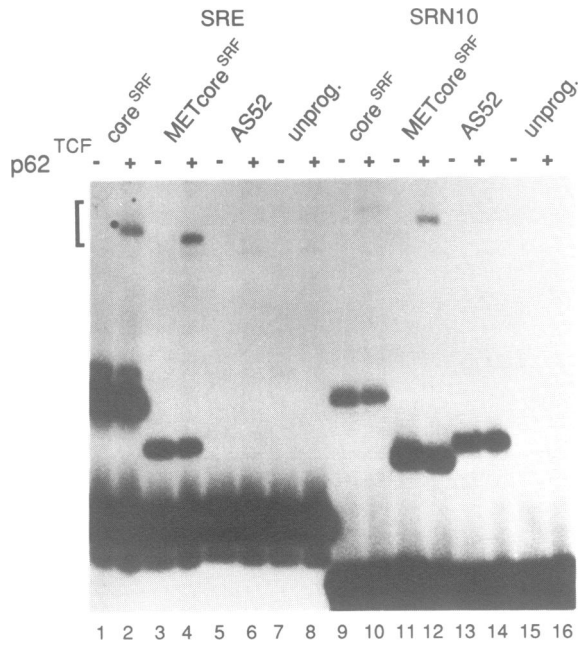

Figure 7. Ternary complex formation with coreSRF derivatives and an N10 binding site. core^{SRF} derivatives, prepared by cell-free translation in reticulocyte lysates, were incubated with either a SRE (lanes 1–8) or a N10 (lanes 9–16) DNA-binding probe (SRN10, see Materials and Methods) as indicated, in the absence (−) or presence (+) of an enriched fraction of p62^{TCF}. 'Unprog.' refers to a control translation performed without the addition of exogenous RNA. The positions of ternary complexes formed between the core^{SRF} derivatives, p62^{TCF} and the DNA are indicated by the bracket.

emphasising that this position does not specify a base-pair contact. Significantly however, residual binding to the N10 site by core^{SRF} proteins is abolished by mutations at this position (K154A and K154E: fig. 6, lanes 4 and 6). A similar effect of these mutations is observed in the context of METcore^{SRF}. In this case, the residual binding to the M1 site is abolished (fig. 6, lanes 9 and 11).

In summary, lysine154 in core^{SRF} allows binding to non-consensus sites and therefore less discrimination between the A:T and G:C pairs at positions ±4 in the binding site. This loss of discrimination is most pronounced at low salt concentrations. Substitution of lysine154 with alanine or glutamate blocks residual binding to the N10 site by core^{SRF} and to the M1 site by METcore^{SRF} thereby increasing their ability to discriminate between the two sites.

METcore^{SRF} forms a ternary complex with p62^{TCF} and a RSRFC4 binding site

METcore^{SRF}, which binds ambivalently to the N10 and M1 binding sites, is identical to core^{SRF} except for the deletion of the N-terminal 10 amino acids. All other residues are derived from core^{SRF} including those which are necessary for the recruitment of p62^{TCF} into a ternary complex (4,16). AS52 differs from METcore^{SRF} by two point mutations in the basic region that abolish binding to CArG box sites (eg. M1) and allow high affinity binding to the N10 binding site, a sequence recognised by RSRFC4. As RSRFC4 cannot recruit p62^{TCF} to the N10 binding site (18) it was of interest to see if p62^{TCF} could form a ternary complex with either METcore^{SRF} or AS52 at the N10 binding site. To this end an oligonucleotide duplex incorporating the N10 site into the c-*fos* SRE in place of the C-ArG box was synthesised and used as a probe for complex formation. This modified SRE (SRN10) provides an ETS protein family recognition site, which is necessary for the binding of p62^{TCF} (3) and Elk-1 (24) in a ternary complex, adjacent to an N10 binding site.

Although METcore^{SRF} binds to the c-*fos* SRE more weakly than core^{SRF}, it supports ternary complex formation by p62^{TCF} with equal efficiency (fig. 7, lanes 2 and 4). AS52 fails to do either (lanes 5 and 6). METcore^{SRF} is also able to bind and recruit p62^{TCF} efficiently to SRN10, and even core^{SRF} supports weak ternary complex formation (lanes 10 and 12). Similarly, METcore^{SRF} can recruit Elk-1 into a ternary complex with the N10 binding site (data not shown). In contrast, AS52, which differs from METcore^{SRF} by just two point mutations in the basic region, binds to SRN10 efficiently but exhibits barely detectable ternary complex formation. These observations indicate that METcore^{SRF} can align on either binding site in a conformation suitable for interacting with p62^{TCF}/Elk-1 whereas AS52 binds strictly to the N10 site in a conformation analogous to that of RSRFC4, as judged from the salt-dependence profile of DNA-binding. This conformation is apparently incompatible with ternary complex formation by p62^{TCF}.

DISCUSSION

p67^{SRF} and RSRFC4 recognise similar, palindromic binding sites which are virtually the same throughout a central 10bp core, with the notable exception that at the ±4 position p67^{SRF} recognises a G:C base-pair whereas RSRFC4 recognises an A:T base-pair. *In vivo*, the discrimination between these two sites must be absolute to allow specific binding of each protein to its cognate recognition sequence. Here we have shown that minimal core peptides containing the p67^{SRF} or RSRFC4 DNA-binding domains retain the ability to discriminate between these two sites. This discrimination is brought about by two complementary determinants that map to the unique N-terminal amino acids of core^{SRF} and to a single amino acid located in the putative DNA-binding α-helix (13).

In the case of core^{SRF}, the deletion of the N-terminal 10 amino acids causes a dramatic relaxation of DNA-binding specificity. The resulting protein, METcore^{SRF}, recognises adenine in place of guanine at the ±4 position, essentially the DNA-binding specificity of RSRFC4. This suggests that the N-terminal 10 amino acids play an inhibitory role in core^{SRF} by blocking the recognition of an A:T base-pair and simultaneously stipulating the recognition of a G:C base-pair. This could take place either by impeding residues that determine the 'A' contact or, alternatively, by repositioning residues so that only a 'G' contact can be made. In either case, the removal of the N-terminus would release residue(s) to interact with the A:T base-pair.

In combination with the N-terminal deletion the mutation K154E further defines the DNA-binding specificity of RSRFC4. This single exchange cancels the relaxed specificity observed with METcore^{SRF} in favour of specific binding to the N10 site. Furthermore, the salt dependence of DNA-binding changes with this mutation from a profile characteristic of core^{SRF} to one resembling that of C4core^{SRF}. However, the role of this residue in direct base-pair recognition can be ruled out. The K154E mutation in core^{SRF} does not perturb binding to sites containing G:C base-pairs at the ±4 positons and the reciprocal mutation

in C4core^{SRF} does not allow binding to a CArG box (data not shown). Moreover, the mutation K154A in the context of either core^{SRF} or METcore^{SRF} reduces neither DNA-binding affinity nor sequence specificity. On the contrary, DNA-binding specificity is increased significantly: in the case of core^{SRF} towards recognising a G:C base-pair and in the case of METcore^{SRF} towards recognising an A:T base-pair at the ±4 positions. This suggests a modulatory role for lysine154, as it allows some flexibility in DNA-binding specificity. Substitution of this residue cancels the relaxation of specificity. In RSRFC4 it appears that the N-terminal amino acids also play an essential role in DNA-binding. In this case, the presence of lysine144 increases the DNA-binding affinity of METcore^{SRF} derivatives containing the K154E mutation. Taken together with the fact that the N-terminal 10 amino acids in core^{SRF} are involved in sequence specificity determination, these results imply a 'looping-back' of the N-terminal residues in core^{SRF} and RSRFC4 in the 3-D structure of the proteins. This would bring N-terminal residues into close proximity with both lysine154/Glutamate14 and residues that are proposed to lie on the recognition face of the DNA binding helix (13).

Previous results have indicated that lysine154 is located at the N-terminus of a putative DNA-binding helix of p67^{SRF} (fig.1)(13). The results presented here are consistent with this hypothesis, with lysine154 lying at the edge of the proposed DNA-binding face of this helix. Such a position would allow close association both with residues forming the DNA-binding interface and residues in the N-terminus of the protein. Similarly, lysine4 in RSRFC4 would be in close proximity to glutamate14 in this model. In view of their opposite charges, it is tempting to speculate that lysine4 and glutamate14 of RSRFC4 form a salt bridge. This would provide the increased stability proposed to explain the observed higher binding affinities of proteins containing these two residues and the observed effects of salt on DNA binding.

It is interesting to note that several members of the MADS box family have the same N-terminus as RSRFC4 (25). The DNA-binding specificity of these proteins has not been determined but our results predict that these proteins recognise an A:T base-pair rather than a G:C base-pair at the ±4 position. Several proteins that possess an N-terminal extension all bind strongly to CArG boxes as predicted by our results (4,16,17,20). We have shown *in vitro* that the core^{SRF} derivative METcore^{SRF} possesses a relaxed DNA-binding specificity and has the ability to form a ternary complex with p62^{TCF} and a N10 binding site. The loss of this ability to support ternary complex formation seen with the mutant AS52 parallels the cancellation of relaxed DNA-binding specificity and the altered salt-dependent DNA-binding profile. These differences suggest that the complexes formed between p67^{SRF}, RSRFC4 and their cognate binding sites are structurally dissimilar.

The observation of a ternary complex with p62^{TCF}, METcore^{SRF} and a binding site divergent from the functionally characterised SRE points to a means whereby the study of p67^{SRF} function may be possible *in vivo* without interference from the ubiquitous, endogenous p67^{SRF}. Such a system would be invaluable for delineating the role of p67^{SRF} and mutant derivatives thereof in gene regulation. We are currently testing the feasibility of this strategy.

It is tempting to speculate that ternary complex factor binding takes place at the N10 site in its natural location within the N10 gene promoter (26). Although neither p62^{TCF} nor Elk-1 form ternary complexes with RSRFC4 at the N10 site (18), it is conceivable that the presence of the correct partners for RSRF proteins would allow ternary complexes to form. In summary, we have shown that the difference in DNA-binding specificities of the human MADS box proteins p67^{SRF} and RSRFC4 is determined by a combinatorial mechanism. In both cases, N-terminal residues in the DNA-binding domain contribute to DNA-binding. The unique N-terminus of core^{SRF} is a major determinant of the G:C specificity at the ±4 position. The deletion of this N-terminus in combination with the K154E mutation creates the DNA-binding specificity of RSRFC4. Both determinants act by an indirect mechanism in specifying base-pair recognition. Although the generality of this phenomenon is not clear, it may be widespread and thus have important consequences when DNA-binding specificities of transcription factors are determined using 'core domains' that lack potential modulatory C-terminal and N-terminal sequences.

ACKNOWLEDGEMENTS

We thank Richard Treisman for the plasmids pT7C4SRF, Katja Stingl for excellent technical assistance, Veronika Person for the oligonucleotides and Lore Lay for the photographic work. This work was supported by the Max-Planck-Gesellschaft.

REFERENCES

1. Treisman,R. (1990). *Sem. Cancer Biol.* **1**, 47–58.
2. Tuil,D., Clergue,N., Montarras,D., Pinset,C., Khan,A. and Phan-Dinh-Tuy,F. (1990). *J. Mol. Biol.* **213**, 677–686.
3. Shaw,P.E., Schröter,P.E. and Nordheim,A. (1989). *Cell* **56**, 563–572.
4. Shaw,P.E. (1992). *EMBO J.* **11**, 3011–3019.
5. Gille,H., Sharrocks,A.D. and Shaw,P.E. (1992). *Nature* **358**, 414–417.
6. Sassone-Corsi,P. and Verma,I.M. (1987). *Nature* **326**, 507–510.
7. Wilson,T. and Treisman,R. (1988). *EMBO J.* **7**, 4193–4202.
8. König,H., Ponta,H., Rahmsdorf,U., Büscher,M., SchÖnthal,A., Rahmsdorf,H.J. and Herrlich,P. (1989). *EMBO J.* **8**, 2559–2566.
9. Lucibello,F.C., Lowag,C., Neuberg,M. and Müller,R. (1989). *Cell* **59**, 999–1007.
10. Schönthal,A., Büscher,M., Angel,P., Rahmsdorf,H.J., Ponta,H., Hattori,K., Chiu,R, Karin,M. and Herrlich,P. (1989). *Oncogene* **4**, 629–636.
11. Shaw,P.E., Frasch,S. and Nordheim,A. (1989). *EMBO J.* **8**, 2567–2574.
12. Ofir,R., Dwarki,V.J., Rashid,D. and Verma, I.M. (1990). *Nature* **348**, 80–82.
13. Sharrocks,A.D., Gille,H. and Shaw,P.E. (1992). *Mol. Cell. Biol.* In press.
14. Schwarz-Sommer,Z., Huijser,P., Nacken,W., Saedler,H. and Sommer,H. (1990). *Science* **250**, 931–936.
15. Pollock,R. and Treisman,R. (1990). *Nucl. Acids Res.* **18**, 6197–6204.
16. Mueller,C.G.F. and Nordheim,A. (1991). *EMBO J.* **10**, 4219–4229.
17. Hayes,T.E., Sengupta,P. and Cochran,B.H. (1988). *Genes & Dev.* **2**, 1713–1722.
18. Pollock,R. and Treisman,R. (1991). *Genes & Dev.* **5**, 2327–2341.
19. Sharrocks,A.D. and Shaw,P.E. (1992). *Nucl. Acids Res.* **20**, 1147.
20. Schröter,H., Mueller,C.G.F., Meese,K. and Nordheim,A. (1990). *EMBO J.* **9**, 1123–1130.
21. Norman,C., Runswick,M., Pollock,R. and Treisman,R. (1988). *Cell* **55**, 989–1003.
22. Prywes,R. and Roeder,R.G. (1987). *Mol. Cell. Biol.* 7, 3482–3489.
23. Ebright,R.H. (1991). *Methods in Enzymol.* **208**, 620–640.
24. Hipskind,R.A., Rao,V.N., Mueller,C.G.F., Reddy,E.S.P. and Nordheim,A. (1991). *Nature* **354**, 531–534.
25. Ma,H., Yanofsky,M.F. and Meyerowitz,E.M. (1991). *Genes & Dev.* **5**, 484–495.
26. Ryseck,R.P., Macdonald-Bravo,H., Mattei,M.G., Ruppert,S. and Bravo,R. (1989) *EMBO J.* **8**, 3327–3335.