© 1992 Oxford University Press
Nucleic Acids Research, Vol. 20, No. 13 3297–3303

# SRF and MCM1 have related but distinct DNA binding specificities

Judy Wynne and Richard Treisman*
Imperial Cancer Research Fund, PO Box 123, Lincoln's Inn Fields, London WC2A 3PX, UK

Received May 6, 1992; Revised and Accepted June 3, 1992

## ABSTRACT

**The mammalian transcription factor SRF and the yeast regulatory protein MCM1 contain DNA binding domains that are 70% identical; moreover, both proteins can bind the serum response element in the human c-fos promoter. Here we present an analysis of MCM1 sequence specificity by selection of sites from random sequence oligonucleotides. In this assay the MCM1 DNA binding domain selects binding sites containing the consensus (NotC)CCY(A/T)(A/T)(T/A)NN(A/G)G, distinct from the SRF binding consensus CC(A/T)$_6$GG. Carboxylethylation interference analysis of a set of selected sites suggests that MCM1 contacts DNA in its major groove throughout one helical turn. These differences in specificity are largely due to sequence differences between the N terminal basic parts of the SRF and MCM1 DNA binding domains. Comparison of the relative binding affinities of MCM1 and SRF for a panel of representative binding sites showed that many high affinity MCM1 sites have negligible affinity for SRF and *vice versa*. Thus MCM1 and SRF have significantly different sequence specificities.**

## INTRODUCTION

Recent advances in the study of eukaryotic DNA binding proteins have led to the identifcation of several new protein sequence motifs associated with sequence specific DNA binding. One such motif is found in Serum Response Factor (SRF), a transcription factor involved in growth factor-regulated transcription (1; for review see 2). The SRF DNA binding domain is 70% identical to the DNA binding domains of two yeast regulatory proteins, MCM1 and ARG80 (3, 4). All three DNA binding domains include a conserved 56 amino acid sequence motif, the 'MADS box' which is conserved in more distantly related proteins from both plant and animal kingdoms (5–8, see Figure 1A). The N terminal half of the MADS box includes highly basic sequences involved in sequence-specific DNA binding (1, 8), while its C terminal half forms part of the dimerisation region. Where tested, sequences required for high affinity DNA binding extend some 30 residues to the C terminal side of the MADS box motif (1, 8, 9), a region that also specifies the recruitment of accessory factors by SRF and MCM1 (10, 11).

The extensive sequence identity between SRF and MCM1 is reflected by the similarity between the SRF binding consensus CC(A/T)$_6$GG and the sequences of naturally occuring MCM1 binding sites in celltype-specific yeast UASs, shown in Figure 1B (for references, see 9, 12–15). Indeed, MCM1 can bind the human c-fos SRF binding site both in vitro and in yeast cells in vivo (3, 15, 16). In spite of these similarities, however, the sequences and properties of naturally occuring MCM1 binding sites suggest that the binding specificities of the proteins are in fact different. For example, binding studies in vitro have shown that the affinity of SRF for CC(A/T)$_6$GG elements can be reduced up to tenfold by mutation of the conserved CG basepairs at positions 1,2,9 and10 (for convenience, the first C of the CC(A/T)$_6$GG consensus is designated position 1); by the introduction of deletions within the central AT core at positions 3–8; or by substitution of the AT core with CG basepairs. However, many naturally occuring MCM1 sites contain precisely such features (Figure 1; see 14, 15). In addition, although some SRF binding sites function efficiently as UAS sequences in yeast, others have only minimal UAS activity, suggesting that they are not efficiently bound by MCM1 (17).

To investigate differences in specificity between SRF and MCM1 in more detail, we have used a binding site selection technique to select MCM1 binding sites from a pool of random sequence oligonucleotides (13). We show that the sequence specificities of SRF and MCM1 are indeed significantly different, and that sites which exclusively bind one protein or the other can be recovered.

## MATERIALS AND METHODS

### Plasmids and proteins

Recombinant SRF was produced using a baculovirus vector and purified as previously described (18). MCM1 and its derivatives were produced by in vitro translation of appropriate cRNAs in reticulocyte lysates using as expression vectors T7βSal (1) or its derivative pT7βTAG (13). Plasmids were constructed by standard techniques and were as follows:

pT7MCM1, encodes MCM1 amino acids 1–286. It was derived from pGA1761 (9) and contains MCM1 sequences extending from the initiation codon to a ClaI site 6 basepairs 3′ to the termination codon, inserted between the NcoI and PstI sites

* To whom correspondence should be addressed

of T7βSal. For transcription the plasmid was linearised with BamHI.

pT7MCM1[1−112]T encodes MCM1 amino acids 1−112, followed by the c-myc 9E10 epitope (19). PCR with primer 5′ CCTCCATGGCTTCTTCCTCATCATC 3′ was used to introduce an NcoI site into the MCM1 coding sequence at residue 113, and standard techniques were used to insert MCM1 codons 1−112 into the NcoI site of pT7βTAG. For transcription the plasmid was linearised with EcoRI.

pT7MCM1[1−41]/SRF[167−508] was constructed using standard techniques by substitution of MCM1 DNA extending from the initiation codon to BspHI (codon 44) for SRF DNA extending from the NcoI to BspHI (codon 169) sites of pT7Δ2.9 (1). It was linearised for transcription with EcoRI. The last MCM1 derived amino acid in the hybrid coding region is MCM1 residue 41 (His).

### Site selection and DNA binding studies

Sites selections and DNA binding assays were performed exactly as described (13), except that the final KCl concentration in all binding reactions was raised to 200mM. Generation of probes for DNA binding studies by PCR from individual subcloned oligonucleotides, and mobility-shift gels were as described (13). DEPC interference analysis was done as previously described (20). Besides the MCM1-selected oligonucleotides and three oligonuceotides from a previous study of SRF binding (13), probes were generated from plasmids comprising the following oligonucleotides inserted at the EcoRI site of pUC12:

Fos, aattGGATGTCCATATTAGGACATC (21);
1/2 site, aattTCCTAATTTCCT;
STE3, aattGTGACACTAATTAGGAAACT;
ACT.L, aattAGATGCCCATATTTGGCGATCT (22);

## RESULTS

### Selection of MCM1 binding sites

Preliminary experiments to produce MCM1 protein by in vitro translation of cRNA in the rabbit reticulocyte lysate showed that production of full length polypeptide was inefficent, due to premature termination of translation just C terminal to the DNA binding domain (data not shown). However, the DNA binding properties of a subfragment of the protein containing the DNA binding domain alone are similar to those of the intact protein (9, 11; data not shown), so we therefore used the isolated DNA binding domain for site selection. A truncated MCM1 derivative, MCM1[1−112]T, was constructed comprising MCM1 amino acids 1−112, fused at its C terminus to sequences encoding the human c-myc 9E10 epitope (Figure 1A; 19). The protein was produced by in vitro translation, bound to random sequence oligonucleotides, and protein-DNA complexes were purified by immunoprecipitation using the 9E10 antibody (13). The associated DNA was recovered, amplified using the PCR and used for further rounds of selection. During the selection process, the oligonucleotide pool became enriched for MCM1 binding sites, as judged by the proportion of input DNA recovered or the amount of DNA complexed when the recovered DNA was used as a probe in gel mobility-shift assays with the MCM1 protein. No DNA was selected by unprogrammed lysate (data not shown; see 13).

After four rounds of selection, oligonucleotides containing MCM1 binding sites were recovered from MCM1[1−112]T-DNA complexes resolved on a mobility-shift gel, subcloned, and

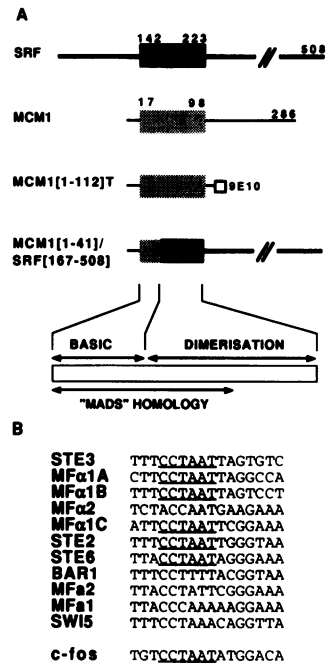



**Figure 1. A.** Proteins studied in this paper. The SRF and MCM1 polypeptides are shown schematically as thick and thin lines respectively; the homologous regions encompassing their DNA binding domains are shown as black and gray boxes respectively. The MCM1[1−112]T and the chimeric MCM1[1−41]/SRF[167−508] proteins are also shown below, with the 9E10 epitope tag is indicated by an open square. Below, the conserved region of the DNA binding domains is shown schematically as an open box, with the extents of the N terminal basic region and the C terminal dimerisation region indicated by arrows. The extent of a conserved sequence motif, the 'MADS box' shared by all proteins containing SRF type DNA binding domains is indicated at the bottom. **B.** Naturally occuring MCM1 binding sites. P box sequences from several MCM1 dependent UASs (for references see 14, 29, 30) are shown compared to the c-fos SRE. Sites STE3 through MFα1C are from a-cell specific UASs, sites STE2 through MFa1 are from a-cell specific UASs, and SWI5 is a cell-cycle regulated UAS. The conserved CCTAAT motif (15) is underlined.

sequenced. The sequences of 57 different oligonucleotides are shown in Table 1. Several oligonucleotides were recovered twice, indicating that the complexity of the original oligonucleotide pool was substantially reduced during the selection procedure. Each oligonucleotide sequenced contained a good match to consensus sequence (NotC)CCY(A/T)(A/T)(T/A)NN(A/G)G. For further discussion, the first C of this consensus will be designated position 1; positions 1−6 are underlined in Table 1. Thirteen of the different oligonucleotides contained consensus matches overlapping the EcoRI primer sequence (Table 1); sequences of this type were never recovered in site selection experiments with SRF (13). For derivation of an unbiassed consensus sequence, oligonucleotides in which primer sequences were located closer to the core motif than position −3 or +13 were omitted from the database (Table 1, second consensus).

The presumptive MCM1 consensus binding sites differ from the SRF consensus site CC(A/T)$_6$GG (13) in three major ways. First, the outer GG dinucleotide at positions 9 and 10 is not invariant, with AG and less commonly GT occuring at this position; second, a pyrimidine rather than A or T is conserved at position 3; and third, GC basepairs are commonly found at positions 3−8, especially at positions 6,7 and 8. The core consensus is embedded in AT rich DNA, and there is a marked discrimination against C at position −1 and G at position 11.

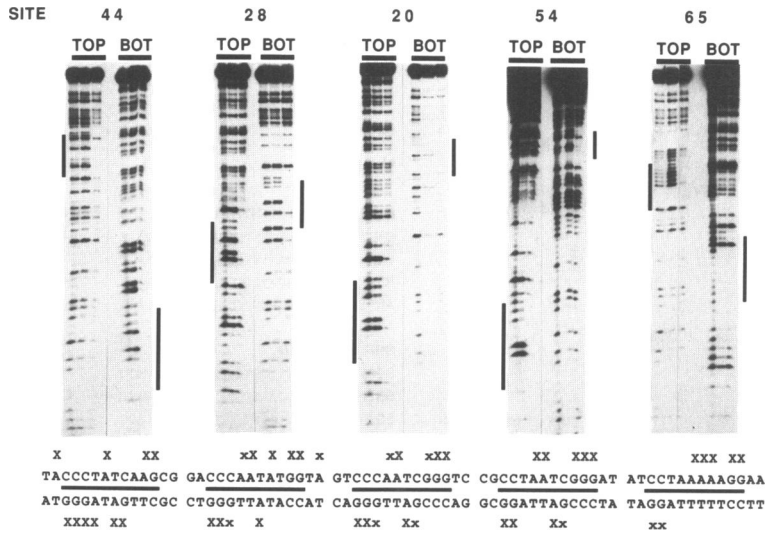



**Figure 2.** Carboxyethylation interference analysis of MCM1[1−112]T selected sites. The interaction of MCM1[1−112]T with sites M44, M28, M20, M54, and M65 was analysed by DEPC interference. Each panel displays the cleavage patterns using probes labelled on the 'top' and 'bottom' strands of the DNA. Each set of three lanes shows results from DEPC treated probe alone, and from free DNA and complexed DNA purified from a mobility shift gel. The patterns of interference are indicated at the bottom. The location of the consensus motif on the cleavage ladder is indicated by the thick bar next to each gel.

Many of the oligonucleotides selected by MCM1 thus contain multiple features that would adversely affect SRF binding.

## The consensus motif defines an MCM1 binding site

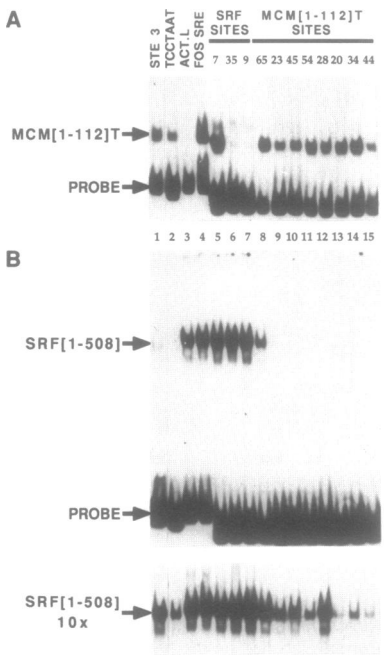
To prove that the conserved (NotC)CCY(A/T)(A/T)(T/A)NN(A/G)G motif in the oligonucleotides selected by MCM1 are actually binding sites for the protein, we performed carboxyethylation interference studies. Previous methylation interference studies of MCM1 binding at naturally occuring sites showed that positions 1,2,9,10 are close contact points for the protein (15, 23). We particularly wished to confirm that MCM1 is in close contact with the DNA throughout the observed consensus sequence, since many of the selected sites resemble halfsites of the SRF binding site. For analysis we therefore chose sites M20,M28,M44,M54, and M65 which contain from as few as 3 AT basepairs (site M20) to 6 AT basepairs (site M65) in their central region. (We shall show below that binding affinity does not correlate with the number of central AT basepairs.)

In order to allow protein-DNA contacts in the DNA major groove at each basepair to be monitored, we chose to modify probe DNA by diethylpyrocarbonate (DEPC) treatment, which carboxyethylates the N7 of A and G bases. Complexes between MCM1[1−112]T and DEPC-treated DNA were purified by gel electrophoresis, the DNA was eluted, and subsequently cleaved by piperidine treatment at positions of modification. The results, shown in Figure 2, indicate that the MCM1 DNA binding domain makes major groove DNA contacts throughout one complete helical turn of the DNA. The degree of interference varies with the position in the consensus site. Carboxyethylation of purines at positions 1,2,9, and 10 always interferes strongly with binding, indicating that these positions are always in close proximity to protein (Figure 2). Modification of purines at positions 3 to 8 also interferes with MCM1[1−112]T binding, although at these positions the strength of the effect is very variable (see site M65, Figure 2).

## The basic region determines binding specificity

The significant differences between the sites selected by MCM1 protein from those selected by SRF might be due either to sequence changes in the parts of the protein that contact DNA or to changes in the orientation of the DNA contact surface caused by sequence changes in the dimerisation domain. MCM1 was previously found to adopt different conformations at different binding sites (24). To investigate the MCM1 sequences responsible for the difference in specificity with SRF we constructed an MCM1/SRF hybrid protein, MCM1[1−41]/SRF[167−508] in which the N terminal sequences of MCM1, including the basic region of the DNA binding domain, are joined to the dimerisation region and C terminal sequences of SRF (Figure 1A). The hybrid protein was produced by cell free translation and the binding site selection process repeated, using an anti-SRF antiserum to recover protein-DNA complexes. Enrichment for binding sites was again observed with successive rounds of selection. After four rounds of selection, the MCM-SRF protein-DNA complexes were purified by gel electrophoresis and the associated DNA isolated, subcloned and sequenced.

The sequences of 44 different selected oligonucleotides are shown in Table 2. The consensus site established in this analysis is very similar to that established for the intact MCM1 DNA binding domain. Again, the sites differ from the SRF binding site consensus: the flanking CC and GG dinucleotides at positions 1,2 and 9,10 are not invariant; a pyrimidine is conserved at position 3; and the majority of the sites contain some GC basepairs between positions 3 and 8. Minor differences were observed between sites selected by the MCM1-SRF chimera and MCM1[1−112]T: several sites contained the dinucleotide AC rather than CC at positions 1,2; the frequency of GG dinucleotides at positions 9,10 was higher; and three sites contained a C at position −1. However, this experiment suggests that the principal determinant of sequence specificity in MCM1 is located in the N terminal basic part of its DNA binding domain.




C

| SITE | | BINDING | |
|---|---|---|---|
| NATURAL and SYNTHETIC | | MCM1 | SRF |
| STE3 | TAGTTTCCTAATTAGTGTCAC | ++ | ++ |
| 1/2 site | GAATTTCCTAATTTCCTAATT | + | + |
| FOS | AGATGTCCTAATATGGACATC | +++ | +++ |
| ACT.L | AGATGCCCATATATGGGCATC | +/- | +++ |
| SRF SELECTED | | | |
| S2-7' | tcgCTTCCTAATATGGGCATC | +++ | +++ |
| S35 | GCAAGTCCATATATGGGCATC | +/- | +++ |
| S9 | ctcTGTCCTTATATGGGCATG | +/- | +++ |
| MCM1 SELECTED | | | |
| M65 | AAGCATCCTAAAAAGGAAACC | +++ | ++ |
| M23' | gCGCGTCCTAAACAGGACATC | ++ | + |
| M45 | ATTTATCCTAATGTGGGTATC | ++ | + |
| M54' | ctgtcgCCTAATCGGGATACT | ++ | + |
| M28' | AAGTGACCCAATATGGTATGa | +++ | ++ |
| M20' | tcgAGTCCCAATCGGGTCATG | +++ | +/- |
| M34 | ATCTTGCCCAAAAGGGACACA | +++ | + |
| M44 | CTATTACCCTATCAAGcgaca | + | +/- |

**Figure 3.** Comparison of (A) MCM1[1−112]T and (B) SRF binding to a panel of binding sites (C). A panel of sites, including natural and synthetic MCM1 and SRF binding sites, and sites isolated by site selection using SRF and MCM1[1−112]T were analysed for binding to MCM1[1−112]T (panel A) or SRF (panel B). **A.** Binding reactions contained 1μl reticulocyte lysate programmed with MCM1[1−112]T together with probes from the indicated sites. The identity of the upper complex formed on the SRF selected sites is unclear. **B.** Binding reactions contained either 5ng (upper section) or 50ng (lower section) purified recombinant SRF expressed in insect cells (18), with probes as in A. Note that only the complexes are indicated in the lower section. **C.** Summary of binding data. Full sequences of the sites tested are shown, and their relative affinities indicated. The absolute affinities of MCM1 and SRF are not stictly comparable in this assay. MCM1: +++, high affinity; ++ comparable to STE3 UAS; +, less than STE3 UAS; ±, barely detectable in gel mobility shift. SRF: +++, comparable to c-fos SRE; ++, about 10 fold less than c-fos SRE; + 100 fold less than c-fos SRE; ±, binding barely detectable.

## Direct comparison of MCM1 and SRF binding

The differences between the MCM1 consensus binding site determined here and the SRF binding consensus detemined previously (13) suggest that many MCM1 binding sites may be poor SRF binding sites. For example, in SRF binding sites, nucleotides other than CC at positions 1 and 2 and GG at 9 and 10, or GC basepairs at positions 3−8 are known to lower significantly the affinity for SRF (12). Conversely, the presence of a conserved pyrimidine at position 3 of many sites selected by MCM1 contrasts with the requirement for A or T at this position in SRF binding sites (13). We therefore tested whether sites could be found that bound MCM1 but not SRF and vice versa. Probes of equal specific acitvity were prepared from the MCM1 and SRF sites shown in Figure 3C, and tested for SRF and MCM1 binding by gel mobility-shift assay. The eight MCM1-selected sites included sites with either C or T at position 3, and sites with between three and six AT basepairs at positions 3−8. In addition we tested the naturally occuring MCM1 binding site from the yeast STE3 gene UAS, which closely resembles the SRF consensus but has a T rather than a G at position 10 (see Figure 3C). The last MCM1 site tested is a synthetic site, TCCTAATTTCCT, previously shown to bind MCM1 weakly in vitro (15), which differs from an SRF consensus in the presence of CC in place of GG at positions 9 and 10. As SRF sites, we chose three SRF-selected sites from a previous analysis (13), which differ only in the sequence of AT basepairs in the central $CC(A/T)_6GG$ core. All of these sites contain an (unfavourable) G at position 11: however, at positions 1−6, site S2-7 contains a good consensus match to the MCM1 consensus element (CCTAAT), while site S9 contains a poor match (CCTTAT), and site S35 contains a mismatch (CCATAT; see Figure 3C). In addition we tested the c-fos SRE, which contains a good match to the consensus and was previously shown to bind MCM1 (15, 16); and ACT.L, a symmetric derivative of the X. laevis γ actin SRE (22), which contain an AT basepairs at the highly conserved position 3, disrupting the MCM1 consensus.

We first examined MCM1[1−112]T binding to the panel of sequences. As expected, the eight MCM1[1−112]T-selected sites were bound efficiently by this protein under our assay conditions, with affinities comparable to that of the STE3 MCM1 site (Figure 3A, compare lanes 8−15 with lane 1). Efficient binding of MCM1[1−112]T to the TCCTAATTTCCT site was also observed in this experiment (Figure 3A, lane 2). Similar results were obtained with intact MCM1 protein (data not shown). However, the affinity of MCM1[1−112]T for the SRF binding sites tested was highly variable. The c-fos SRE and the in vitro SRF-selected site S2-7, which both contain a good match to our MCM1 consensus at positions 1−6, bound MCM1[1−112]T with an affinity comparable to the MCM1[1−112]T-selected sequences (Figure 3A compare lanes 4,5 with 8−15). In contrast, the SRF-selected sites S35 and S9, which differ from site S2-7 only in the sequence of AT basepairs in the central region, exhibited greatly reduced binding affinity for MCM1[1−112]T (Figure 3A, compare lanes 5 with 6,7). We presume that the failure of site S9 to bind efficiently in spite of its match to the consensus at positions 1−6 (CCTTAT) probably arises from its divergence both from the preferred A at position 4 and the unfavoured G at position 11.

We obtained significantly different results when we tested these

**Table 1.** Sites selected by MCM1[1−112]T. DNA recovered after four rounds of selection by MCM1[1−112]T was used as a probe in gel mobility shift assay. Complexed oligonucleotides were recovered, subcloned, and sequenced. The sites are aligned around the central consensus motif of which the conserved CCY(A/T)$_3$ sequence is underlined. Sites that apparently overlap the flanking PCR primer sequences are grouped separately.

```
All except primer overlap (n=42)
A                  11 5  14 -  -  -  34 35 18 9  10 16 1  10 18 23     AG - 15
G                  3  8  8  -  -  -  2  1  -  8  18 26 39 5  9  9      AA - 1
C                  7  7  -  42 42 19 1  -  -  17 5  -  -  10 11 2      GT - 2
T                  21 22 20 -  -  23 5  6  24 8  9  -  2  17 4  8      GG - 24

-3 to +13 totally from random (n=23)
A                  8  3  6  -  -  -  18 19 11 5  8  5  1  6  14 15     AG - 4
G                  2  4  5  -  -  -  1  1  -  4  6  18 21 -  2  2      AA - 1
C                  3  4  -  23 23 10 1  -  -  10 4  -  -  2  6  -      GT - 1
T                  10 12 12 -  -  13 3  3  12 4  5  -  1  15 1  6      GG - 17

Position           -3 -2 -1 1  2  3  4  5  6  7  8  9  10 11 12 13     9-10

CCT SITES
M65                      gaattcgcctcAAGCATCCTAAAAAGGAAACCTTTG cgacaggatcc
M30              gaattcgcctcTAAGATGGGAGTAGGCCTAAAAGAGGcgacaggatcc
M52              gaattcgcctcTACCAAAGAGCATTACCTAAATTGGCcgacaggatcc
M62                   gaattcgcctcGACATTCATGCCTAAACAGGTAGTCAcgacaggatcc
M23'                     ggatcctgtcgCGCGTCCTAAACAGGACATCGGTATAgaggcgaattc
M11                    gaattcgcctcTATATCGCCTAAATAAGTAAGCCTTCcgacaggatcc
M17                gaattcgcctcTAGCGGACAAATTTCCTAAAGTAGAA cgacaggatcc
M31                gaattcgcctcATATAGCAGTACTCCTAAATAGGCTAcgacaggatcc
M47                   gaattcgcctcCAAGCAATTGCCTAAAGCGGTAATTGcgacaggatcc
M15              gaattcgcctcTATAATAAGACTTCTACCTAATCGGTcgacaggatcc
M22               gaattcgcctcGGCCACTATTTATACCTAATCGAGTTcgacaggatcc
M9'                   ggatcctgtcgTCAGTAGTTTCCTAATCGGTTGTTTCgaggcgaattc
M45                      gaattcgcctcATTTATTCCTAATGTGGTATGCATTCCcgacaggatcc
M54'                        ggatcctgtcgCCTAATCGGGATACTTGTTTTGTCCCgaggcgaattc
M43              gaattcgcctcAGAGAACCATTTATATCCTAATCGAGcgacaggatcc
M63              gaattcgcctcTACTTAGACCAAACGTCCTATTTAAGcgacaggatcc
M2'                         ggatcctgtcgCCTATTGTGGTAATAACGCGCAGTCCgaggcgaattc
M25'        ggatcctgtcgGGACGGAGTGCGCACATTTCCTATTCgaggcgaattc
M29                 gaattcgcctcATCAAGTAACTCCTATATAAATCGCCcgacaggatcc
M13'/M59'                ggatcctgtcgTTTTTCCTAGACCGGTCATAGCCCT gaggcgaattc
M46                gaattcgcctcATATAGCTTATGCCTTAACGGGTAATcgacaggatcc
M21'               ggatcctgtcgGCCGGTAATACCTTATGTGGTATATGgaggcgaattc
M49'               ggatcctgtcgCCATTAGTTACCTCATCCAGCCTCGTgaggcgaattc

CCC SITES
M34                gaattcgcctcAGACATCTTGCCCAAAAGGGACACAAcgacaggatcc
M35'                     ggatcctgtcgTCCCAAAAGGGAATACACTATGTCCTgaggcgaattc
M50           gaattcgcctcATATACCGACCAGATACCCAAAAGGGcgacaggatcc
M40'         ggatcctgtcgGCTTACTCCATTGGGCTCCCAAATGAgaggcgaattc
M10/M57      gaattcgcctcTTGTTGTACAACATTACCCAAATTGGcgacaggatcc
M27                    gaattcgcctcGAATACCCAAAGTAGAAATGTGGCTCcgacaggatcc
M26                gaattcgcctcACAATAAGTTCCCAATCAGGAGATTGcgacaggatcc
M53                      gaattcgcctcCCACCCAATCCAGTAATACGGTACACcgacaggatcc
M20'                     ggatcctgtcgAGTCCCAATCGGGTCATGTCCTAAGTgaggcgaattc
M51'/M61'                  ggatcctgtcgACCCAATGCAGCTTGTTTATTTCAACgaggcgaattc
M28'           ggatcctgtcgACAACGGAAGTGACCCAATATGGTATgaggcgaattc
M33'       ggatcctgtcgAAAAATGTGATTCCAGTTTCCCAATGgaggcgaattc
M3'                      ggatcctgtcgTATCCCATTAAGGTAAAGTGGGGCTGgaggcgaattc
M16'             ggatcctgtcgGCGGTATGTAAGTCCCATTAGGGTAAgaggcgaattc
M32               gaattcgcctcCTGCAACGAGTTCCCTATCGGGTAACcgacaggatcc
M44           gaattcgcctcACGCAGATTGCTATTACCCTATCAAGcgacaggatcc
M55'          ggatcctgtcgACTCCGACCTTAATTACCCTATCGGGgaggcgaattc
M12/M58             gaattcgcctcTGTGCATTACCCGATTTGGAATCACAcgacaggatcc
M19'       ggatcctgtcgAGTCAGGGGTCGAGTATATCCCGATAgaggcgaattc


PRIMER OVERLAP
M6'        ggatcctgtcgGACATCTTTTGGTTAAGTGACCTAATGaggcgaattc
M42'       ggatcctgtcgGGCTCTTTGGGTGTCGTGTCCCTAATGaggcgaattc
M48'       ggatcctgtcgTACATGACGCGGATAATTTTCCTAATGaggcgaattc
M8'        ggatcctgtcgCTGCTCGTTTAGTGCCGTTTCCTTTTGaggcgaattc
M5'        ggatcctgtcgGGTAATAGATATTTGGAGATCCCAAAGaggcgaattc
M7'        ggatcctgtcgACTAGTTGGGGTGTCTTTAACCCAATGaggcgaattc
M14'       ggatcctgtcgATATCACTAGCACATCATATCCCAATGaggcgaattc
M18'/M60'  ggatcctgtcgATATAACGCGGAAATAATGGCCCAATGaggcgaattc
M37'       ggatcctgtcgGGACAATGTTAAAACGATTTCCCAATGaggcgaattc
M39'       ggatcctgtcgACTCGCAAGTACGGATATTACCCTATGaggcgaattc
M64'       ggatcctgtcgTCGCCATTATTGGTAACTTTCCCTATGaggcgaattc
M36'       ggatcctgtcgGCCATTCAGGCGGTTAATCTCCCTTTGaggcgaattc
M38'       ggatcctgtcgATTAACGCAGCATGTGTTTACCCTTTGaggcgaattc

NO OBVIOUS SITES
M4'        ggatcctgtcgACGTAAATTGTGTCTTATTACCGGATgaggcgaattc
```

**Table 2.** Sites selected by the MCM1[1−41]/SRF[167−508] hybrid protein. DNA recovered after four rounds of selection by MCM1[1−41]/SRF[167−508] was used as a probe in gel mobility shift assay. Complexed oligonucleotides were recovered, subcloned, and sequenced. The sites are aligned around the central consensus motif of which the conserved CCY(A/T)$_3$ sequence is underlined. Sites that apparently overlap the flanking PCR primer sequences are grouped separately.

```
All except primer overlap (n=31)
A                        5  8  9  5  -  -  26 24 15 13 7  2  4  11 12 13
G                        2  8  8  -  -  -  3  2  -  5  13 29 27 4  5  8
C                        8  3  3  26 31 14 -  -  -  11 8  -  -  8  12 4
T                        16 12 11 -  -  17 2  5  16 2  3  -  -  8  2  6

-3 TO +13 all from random (n=18)
A                        2  6  7  2  -  -  14 13 8  9  6  1  2  8  10 10
G                        1  4  2  -  -  -  2  1  -  3  7  17 16 2  2  2
C                        4  -  1  16 16 8  -  -  -  5  4  -  -  4  6  2
T                        11 8  8  -  -  10 2  4  10 1  1  -  -  4  -  4

CCT SITES
19.3               gaattcgcctcAATTGGCAATAGCCTAAAAGGGCTAcgacaggatcc
24.2                 ggatcctgtcgTCACCGTATACCTAAAAGGGAGTGTTgaggcgaattc
27                         gaattcgcctcCCTAAAAGGGTGTAGATGGGCTCACTcgacaggatcc
1.2'                     gaattcgcctcCCAAACCTAAAAGGGAAATGTTAACAcgacaggatcc
39                       gaattcgcctcAACAACCTAAAGGGGACGTGCGCTATcgacaggatcc
40'                        ggatcctgtcgTTGCCCTAAACGGGTCACGGCACTCAgaggcgaattc
19.1             gaattcgcctcACGTGTCACGTTACCTAATTGGGACTcgacaggatcc
4.3                ggatcctgtcgGGGATAAACTGCCTAATACGGAAATTgaggcgaattc
11'          ggatcctgtcgACCCAAAGGGACGCGGAGTCCTAATCgaggcgaattc
12'                ggatcctgtcgGATTTCTTGTCCTATTAAGGACATGCgaggcgaattc
35'                  ggatcctgtcgACCATTTCCTATTCAGGTAAAGACAGgaggcgaattc
18'                        ggatcctgtcgCCTAGTCTGGACCGAGTTGTCTCTAAgaggcgaattc
9                gaattcgcctcTAAAAGGCTAGGTTTCCTGATCGGGTcgacaggatcc

CCC SITES
4                gaattcgcctcTCTTGGGTAAAATAACCCAAAAGGGCcgacaggatcc
4.2                        ggatcctgtcgCCCAAAGGGGCTACGCTTCCGGGTCTgaggcgaattc
5                gaattcgcctcAATGAGTCTCGCTAGCCCAAACGGGAcgacaggatcc
21                         gaattcgcctCCCAAAGCGGCGGGCTGACCGGTTAGGcgacaggatcc
30.2               ggatcctgtcgCGTGGTCGTATCCCAAAGTGGAAAAAgaggcgaattc
38                       gaattcgcctcATTACCCAATAGGGCGAACGAAAAT cgacaggatcc
32'          ggatcctgtcgAATGCGGTCCCAACGATACCCAATCAgaggcgaattc
37'                        ggatcctgtcgCCCCAATCCGATTACGAAATCTATAAgaggcgaattc
8                      gaattcgcctcTTGTTCCCAGAAAGGGCAACGCTTCCcgacaggatcc
17                     gaattcgcctcTTGTATCCCGATGAAGCACATCGTATCgacaggatcc
26'                  ggatcctgtcgTCTGCTTTATCCCGAAACGGTACATGgaggcgaattc
15'                  ggatcctgtcgACGATCAATTACCCTTTCGGACATGTgaggcgaattc
10'                        ggatcctgtcgTCGACCCTTTCAGACAATGCCTGTAAgaggcgaattc

ACT/ACC SITES
14'                              ggatcctgtcgGGACTAAAACGGTATGAATGCCAAATgaggcgaattc
29                 gaattcgcctcCGTGTCGAGACAATGACTAAAACGGAcgacaggatcc
34                   gaattcgcctcCTTAGCGACGCTTACTAATACGGTAGcgacaggatcc
7'                         ggatcctgtcgGACTATTTTGGCAAGAGAGTAATCATgaggcgaattc
1.3                      gaattcgcctcACTGTACCAATCCGGAAAATGCGCT cgacaggatcc

PRIMER OVERLAP
19.2'      ggatcctgtcgAACACGACGGGGTCTTATGGCCTAATGaggcgaattc
1.1'       ggatcctgtcgTTACTCTTTTCGGAGTTGCAACTAATGaggcgaattc
14'        ggatcctgtcgGGACTAAAACGGTATGAATGCCAAATGaggcgaattc
23'        ggatcctgtcgTTACTGCATTTGGCAATGGTCCATAAGaggcgaattc
24.1'      ggatcctgtcgATTGCCTATTGTGAACATAACCAATAGaggcgaattc
28'        ggatcctgtcgTCTGCTTAAGGGAATTTTTACCCATTGaggcgaattc
36'        ggatcctgtcgTGAGCATGCAGTGNGTATGCACCAATGaggcgaattc

NO OBVIOUS SITES
3          gaattcgcctcTAAACGGGAACATGTCTGTTTTAATCcgacaggatcc
6          gaattcgcctcTTCCTCCACCCGTTCTGGTAAGGAATCgacaggatcc
16         gaattcgcctcTTTCGGGTACTAATCCCCCTGTTTGTcgacaggatcc
30.1       gaattcgcctcAAATGAGACAGCTCGACTAATGCGGAcgacaggatcc
30.3       gaattcgcctcTTGCGGAAAACTGATAACGCTTTGAGcgacaggatcc
31         gaattcgcctcGAAACGTGATGCCGAATCCGGTCACAcgacaggatcc
```

various sites for SRF binding. For convenience we used recombinant SRF produced by a baculovirus vector: this has DNA binding properties identical to those of HeLa cell and in vitro translated SRF (18). Similar results were obtained with in vitro translated SRF (data not shown; C.Hill and R.T. unpublished data). Each site was tested in binding reactions containing 1ng or 10ng SRF to allow an approximate evaluation

of relative affinity (Figure 3B, upper and lower panels respectively). The five naturally occuring and SRF-selected sites bound the protein equally well even though these sites exhibited widely different affinities for MCM1 (Figure 3B, lanes 3−7). As expected, the STE3 UAS, which differs from the SRF consensus in that it contains a TA basepair at position 10, exhibits approximately 10-fold lower affinity for SRF than the c-fos SRE (Figure 3B, compare lanes 1,4), while the affinity of SRF for the TCCTAAT halfsite, which contains CG basepairs at positions 9 and 10, is even lower (Figure 3B compare lanes 2,4). By contrast, none of the sites selected by MCM1[1−112]T bound SRF efficiently in these assays. Of these sites only site M65 contains six AT basepairs in the core region; however, this site has AA at positions 5 and 6 in contrast to the majority of both naturally occuring and in vitro selected SRF binding sites (see 2, 13). In general, it appears that the fewer the number AT basepairs at positions 3 to 8, the lower the affinity for SRF, as borne out by the examples of sites M65 and M23 (Figure 3B, lanes 8,9) and sites M28 and M20 (Figure 3B, lanes 12,13).

## DISCUSSION

In this work we used a site selection method and in vitro DNA binding studies to compare the DNA binding specificities of the highly related yeast MCM1 and mammalian SRF DNA binding domains. Binding sites selected by MCM1 in vitro, which had affinities comparable to or greater than the naturally occuring STE3 UAS, all contained the consensus motif (NotC)CCY (A/T)(A/T)(T/A)NNGG. This consensus differs significantly from the SRF consensus $CC(A/T)_6GG$ determined by the same method (13). The subtle differences in specificity even between MCM1 and SRF underscore the importance of a rigorous assessment of the binding specificity of the other family members. In particular, although several plant SRF-related DNA binding domains can bind the c-fos SRE (10, 25), the RSRFs, a set of mammalian SRF-related proteins bind the distinct consensus sequence $YTA(A/T)_4TAR$ (8). Thus even closely related members of the SRF family of related DNA binding domains may have significantly different sequence specificities.

Our results at first appear surprising given the resemblance between the sequences of naturally occuring MCM1 and SRF binding sites, and the observation that in several cases SRF and MCM1 have been shown to bind each other's sites (10, 15, 16). However, the different MCM1 consensus determined here provides a rationale for the observation that many natural MCM1 binding sites contain deviations from the SRF $CC(A/T)_6GG$ consensus known to impair SRF binding. For example, analysis of SRF binding to mutant derivatives of the $CC(A/T)_6GG$ sequence has shown that its affinity is reduced up to tenfold by mutations at the conserved outer CG basepairs of the $CC(A/T)_6GG$ sequence, by substitution of the central AT basepairs with GC basepairs, or by deletion of AT basepairs from the central core (12, 13, 26). Thus deviations of natural MCM1 sites from the established SRF binding consensus probably reflect the subtly different binding specificity of the MCM1 protein.

Previous studies of the role of MCM1 in yeast celltype specific gene expression have demonstrated that the protein binds the P box motif, a degenerate dyad symmetry element found in α- and a-cell specific UASs (27, 28; for examples see Figure 1B). In vitro DNA binding studies with both cell extracts and recombinant MCM1 proteins suggested that the primary sequence recognition determinant for MCM1 within the P box is the halfsite sequence TCCTAAT (14, 15). Our results are in broad agreement with this halfsite assigment, and the more relaxed consensus suggested by our study is matched by the majority of natural MCM1 sites from both celltype-specific and -nonspecific promoters (see Figure 1). However, in contrast to our results, it was previously proposed that the optimal MCM1 binding site is symmetric, based on the binding properties and UAS activity of the synthetic site P(PAL), and the structure of a-cell specific P boxes (14, 15). Perhaps the greater apparent symmetry of a-cell specific MCM1 sites actually reflects their ability to bind MATα2 protein (see Figure 1B). Other proteins with SRF type DNA binding domains also select asymmetric DNA sequences in site selection assays, but in these cases the asymmetry is less pronounced (8, 13). Notwithstanding the asymmetric nature of many of these binding sites, interference analysis indicates that the protein makes close contact with DNA throughout a complete helical turn.

The difference in binding specificity between SRF and MCM1 cannot be due to changes in the relative orientations of the MCM1 subunits compared with those of SRF caused by their differing dimerisation domains: an MCM1-SRF chimera containing the dimerisation region of SRF in place of that of MCM1 binds DNA with essentially the same sequence specificity as MCM1 itself. This result also indicates that sequence differences between the N terminal basic parts of the two DNA binding domains must determine the sequence specificity of binding, as previously found in the case of the mammalian RSRF proteins (8). The availability of closely related proteins with distinct sequence specificities should allow the identification of residues involved in base-specific interactions with DNA, by study of the DNA binding properties of chimeric proteins derived from different SRF family members. We are currently pursuing this strategy.

In yeast we found that the c-fos SRE functions weakly as a UAS, in agreement with previous reports (15, 16), but that the ACT.L SRE has negligible UAS activity, consistent with its reduced affinity for MCM1 in vitro (17). The activity of ACT.L-controlled reporter genes in yeast is thus totally dependent on exogenously expressed SRF, and strains containing such reporters can be used for the study of the interaction of SRF with its accessory proteins and with DNA in the absence of interference due to MCM1 binding (17). We have not as yet examined whether the in vitro-selected MCM1 binding sites described here represent efficient UAS sequences in yeast, and whether such activity is celltype-specific. This question is of interest in the light of current models for celltype-specific gene regulation in yeast. One model proposes that MCM1 sites that bind the protein efficiently in the absence of cooperating MATα1 protein should act as celltype-independent UASs (14): according to this model our sites should function as MATα1 independent UASs. An alternative model is that most MCM1 sites bind the protein with comparable affinities but that it is only at a-specific UASs that MCM1 can adopt a conformation which allows it to activate transcription (23, 24). Failure of our high affinity binding sites to activate independently of cell type might therefore provide support for the conformation model for transcriptional activation by MCM1, and it will be interesting to test this idea.

The data presented here establish that it is possible to design binding sites for MCM1 which in vitro are not recognised by SRF, and vice versa. Are such differences observed when the proteins are expressed in mammalian cells? As yet, we have not extensively characterised the properties of these sites in mammalian cells in vivo. Our preliminary results indicate that an MCM1 binding site that cannot bind SRF in vitro exhibits

neither constitutive nor growth-factor inducible activity when linked to suitable reporter gene and tested in a mammalian cell transfection assay. However, the regulated activity of such a site can be partially restored by overexpression of either wildtype SRF protein itself or more effectively by mutant SRF proteins that can efficiently bind such sites (C. Hill and R.T., unpublished data). Our observations may thus allow the development of systems for the study of SRF and mutants *in vivo*.

## ACKNOWLEDGEMENTS

We thank Roy Pollock for advice on site selections, and members of the laboroatory for helpful discussions and comments on the manuscript.

## REFERENCES

1. Norman, C., Runswick, M., Pollock, R. and Treisman, R. (1988) *Cell* **55**, 989–1003.
2. Treisman, R. (1990) *Seminars in Cancer Biol.* **1**, 47–58.
3. Passmore, S., Maine, G.T., Elble, R., Christ, C. and Tye, B.K. (1988) *J Mol Biol* **204**, 593–606.
4. Dubois, E., Bercy, J., Descamps, F. and Messenguy, F. (1987) *Gene* **55**, 265–75.
5. Yanofsky, M.F., Ma, H., Bowman, J.L., Drews, G.N., Feldmann, K.A. and Meyerowitz, E.M. (1990) *Nature* **346**, 35–9.
6. Sommer, H., Beltran, J.P., Huijser, P., Pape, H., Lonnig, W.E., Saedler, H. and Schwarz-Sommer, S.Z. (1990) *EMBO J* **9**, 605–13.
7. Ma, H., Yanofsky, M.F. and Meyerowitz, E.M. (1991) *Genes Dev* **5**, 484–495.
8. Pollock, R. and Treisman, R. (1991) *Genes Dev* **5**, 2327–41.
9. Ammerer, G. (1990) *Genes Dev* **4**, 299–312.
10. Mueller, C.G.F. and Nordheim, A. (1991) *EMBO J.* **10**, 4219–4229.
11. Primig, M., Winkler, H. and Ammerer, G. (1991) *EMBO J.* **10**, 4209–4218.
12. Leung, S. and Miyamoto, N.G. (1989) *Nucleic Acids Res* **17**, 1177–95.
13. Pollock, R. and Treisman, R. (1990) *Nucleic Acids Res* **18**, 6197–204.
14. Bender, A.and Sprague, G.J. (1987) *Cell* **50**, 681–91.
15. Passmore, S., Elble, R. and Tye, B.K. (1989) *Genes Dev* **3**, 921–35.
16. Hayes, T.E., Sengupta, P. and Cochran, B.H. (1988) *Genes Dev* **2**, 1713–22.
17. Dalton, S. and Treisman, R. (1992) *Cell* **68**, 597–612.
18. Marais, R.M., Hsuan, J.J., McGuigan, C.M., Wynne, J. and Treisman, R. (1992) *EMBO J.* **11**, 97–105.
19. Evan, G.I., Lewis, G.K., Ramsay, G. and Bishop, J.M. (1985) *Mol Cell Biol* **5**, 3610–6.
20. Sturm, R., Baumruker, T., Franza, B.J. and Herr, W. (1987) *Genes Dev* **1**, 1147–60.
21. Treisman, R. (1986) *Cell* **46**, 567–74.
22. Treisman, R. (1987) *EMBO J* **6**, 2711–7.
23. Tan, S., Ammerer, G. and Richmond, T.J. (1988) *EMBO J* **7**, 4255–64.
24. Tan, S. and Richmond, T.J. (1990) *Cell* **62**, 367–77.
25. Schwarz-Sommer, Z., Hue, I., Huijser, P., Flor, P., Hansen, R., Tetens, F., Lonnig, W-E., Saedler, H., and Sommer, H. (1992) *EMBO J.* **11**, 251–264.
26. Christy, B. and Nathans, D. (1989) *Mol Cell Biol* **9**, 4889–95.
27. Dolan, J.W. and Fields, S. (1991) *Biochim Biophys Acta* **1088**, 155–69.
28. Sprague, G.F. (1990) *Adv. Gen.* **27**, 33–62.
29. Kronstad, J.W., Holly, J.A. and MacKay, V.L. (1987) *Cell* **50**, 369–77.
30. Lydall, D., Ammerer, G. and Nasmyth, K. (1991) *Genes Dev* **5**, 2405–19.