Molecular study of SRF-cofactor interactions

Alexia-Ileana Zaromytidou

This thesis is submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy from the University of London

2007

Transcription Laboratory
Cancer Research UK
44 Lincoln's Inn Fields
London
WC2A 3PX

University College London
Gower Street
London
WC1E 6BT

ProQuest LLC
789 East Eisenhower Parkway
P.O. Box 1346
Ann Arbor, MI 48106-1346

All experiments in this thesis were performed by myself with the exception of the sequence analysis presented in Figure 8.14 B and C.

Alexia-Ileana Zaromytidou

Abstract

Serum Response Factor regulates a large array of genes involved in diverse processes including cell proliferation, muscle differentiation and development, and cytoskeletal processes such as cell migration and adhesion. The specificity and versatility of the SRF responses is achieved by combinatorial interactions with accessory factors. SRF binds to the $CC(A/T)_2A(A/T)_3GG$ CArG box consensus sequence within the promoters of its target genes and acts as a docking platform for diverse signal regulated and cell-type specific cofactors to elicit their distinct responses. In fibroblasts two pathways signal through SRF in a mutually exclusive manner. MAP kinase signalling results in transcriptional activation of a subset of SRF target genes, via the interaction of SRF with members of the TCF family of Ets domain proteins. In contrast Rho-signalling induced changes in actin dynamics result in the association of SRF with members of the Myocardin-related family of SRF cofactors (MAL/MRTF-A/MKL1 and MAL16/MRTF-B/MKL2). The results described in this thesis characterise the molecular mechanism of MAL-SRF complex formation. MAL binds SRF as a dimer via a seven-residue core sequence within the MAL B1 region. Residues in the neighbouring Q-box enhance MAL-SRF complex formation, although these do not contact SRF directly. The MAL-SRF interaction displays the properties of a Rho-regulated cofactor. MAL competes with TCF for SRF binding due to the interaction of both cofactors with the same hydrophobic groove and pocket on SRF. In contrast to TCF, MAL-SRF complex formation depends on the intact N-terminus of the SRF DNA-binding domain. Mutations in the SRF αI-helix that reduce DNA bending also impair complex formation with MAL. These mutations however do not affect DNA distortion in the MAL-SRF complex. Efficient MAL-SRF complex formation requires that SRF be bound to its cognate DNA and that MAL directly contacts DNA on either side of the CArG box. My results support a model in which each MAL monomer adds a β-strand consisting of the core B1 sequence, to the β-sheet of the SRF DNA-binding domain in a similar way to TCF, while also making direct DNA contacts in the ternary complex facilitated by SRF-induced DNA distortion. My analysis of complex formation between MAL and SRF demonstrates that members of the MRTF and TCF families of SRF cofactors interact with SRF using related but distinct mechanisms, thus providing a molecular rationale for their mutually exclusive transcriptional responses and the specificity of signalling to SRF.

To my parents

Acknowledgements

I would like to thank Richard for giving me the opportunity to work in his lab, and for being such a good supervisor especially during the final stages of my studentship.

Cisco, without whom life in the lab would have been much harder. Thanks for sharing your knowledge with me, for teaching me so many things and for giving me good advice, even though I didn't always follow it. Above all thank you for being such a good friend.

Rob, another great friend, for always making time for me and never telling me off for interrupting his work with my questions.

Kasumi, for teaching me the most important technique in this thesis.

Cristina and Patrick for their friendship and advice. Also Sara, Gemma, John, Robert, Sebastian, Jane and all the other past and present members of the lab.

Special thanks to Fiona, whose help over the years has been invaluable.

Amanda, who has been a great friend through some tough times and who has provided so much support during the writing of this thesis.

I would also like to thank everybody in the equipment park, the peptide synthesis lab and the bioinformatics unit, for making life easier with their excellent assistance.

Finally a big big thank you to my family. My sister Christina, for being such an amazing presence in my life and for always knowing how to cheer me up.
And my mum and dad for their love, for always believing in me and for letting me hijack their lives these past 25 years.

PUBLICATIONS

Some of the results described in this thesis have been presented in the following publications:

Miralles, F., Posern, G., Zaromytidou, A. I., and Treisman, R. (2003). Actin dynamics control SRF activity by regulation of its coactivator MAL. Cell *113*, 329-342.

Zaromytidou, A. I., Miralles, F., and Treisman, R. (2006). MAL and ternary complex factor use different mechanisms to contact a common surface on the serum response factor DNA-binding domain. Mol Cell Biol *26*, 4134-4148.

Table of Contents

List of Figures

List of Tables

1 Introduction

The different cell types of multicellular organisms need to ensure the correct temporal and spatial patterns of expression of their genes in order to respond to diverse stimuli during growth, differentiation and development and perform their specialised functions. To this end control mechanisms are in place at different stages of gene expression, including RNA synthesis, transcript processing and translation. The process of transcriptional activation is a major point of regulation that ensures the correct execution of the transcriptional programme of each one of the vast number of eukaryotic genes (Orphanides, G. *et al.*, 2002).

1.1 Transcriptional control of gene expression

Transcription involves the simple principle of the localised action of an enzyme with the ability to catalyse the synthesis of RNA from a DNA template using a pool of the appropriate nucleotides. In practice however transcription is a complex process involving many proteins acting in concert within large multicomponent complexes (Lee, T. I. *et al.*, 2000; Lemon, B. *et al.*, 2000).

The initiation of mRNA synthesis requires the RNA polymerase II complex with its associated general transcription factors (GTFs), including TFIIF, TFIIB, TFIIH, TFIIE and TFIID, which together comprise the basal transcriptional machinery and are involved in recognition of the core proximal promoter elements (Woychik, N. A. *et al.*, 2002). Formation of this complex is sufficient for a basal level of transcription. Regulated transcription depends on a wide range of sequence specific DNA-binding proteins, which associate with regulatory promoter elements in a controlled manner to elicit specific transcriptional responses. Coregulators, such as the Mediator and GTF TFIID complexes, that bridge the sequence-specific factors with the basal machinery and transmit the regulatory signals (Malik, S. *et al.*, 2005; Naar, A. M. *et al.*, 2001), play an important role in this process. Additionally transcriptional control is critically linked with the state of chromatin, with many chromatin remodelling factors required to mobilise nucleosomes and expose DNA for transcription factor association (Narlikar, G. J. *et al.*, 2002), as well as a variety of enzymes that catalyse covalent modifications of

histones (including methylation, acetylation, phosphorylation, ubiquitination; (Khorasanizadeh, S., 2004)) that promote or inhibit transcription.

The correct ordered recruitment and concerted action of these factors in response to specific regulatory cues is necessary to ensure that only the genetic information needed for biological function and not the full complement of cellular genes, is expressed at any given time.

1.1.1 Sequence specific transcription factors

Sequence-specific DNA-binding transcription factors are a focal point of transcriptional regulation. These proteins bind to their cognate DNA elements that are present in the regulatory regions (e.g. enhancers) of target genes and in response to particular signals orchestrate transcriptional activation through the recruitment and/or stimulation of the activity of the basal transcriptional complex or repression by inhibition of this process (Kadonaga, J. T., 2004). Thus sequence specific factors (which hereafter will be referred to as transcription factors) function as a crucial interface between the regulatory information stored in the promoters of genes and the transcriptional machinery whose activity depends on the correct interpretation of this information.

Transcription factors are modular in structure, and usually contain a DNA-binding domain capable of sequence specific binding and discrete transcriptional activation and/or repression domains. In contrast to the extensively characterised DNA-binding domains, the transactivation domains of transcription factors remain structurally poorly defined, although acidic regions and polyglutamine and polyproline stretches have been found to be important for transcriptional activity ((Garvie, C. W. *et al.*, 2001) and references therein). Transactivation domains are thought to transduce transcriptional responses by recruiting co-regulatory complexes, which either physically link the DNA-specific transcription factors with the basal machinery complex via protein-protein interactions or function in chromatin regulation (Garvie, C. W. *et al.*, 2001).

Transcription factors may also contain additional regulatory modules, such as the ligand binding domains of nuclear receptors (Nagy, L. *et al.*, 2004) or the actin binding RPEL domain of the MRTFs (see Section 1.2.5.7.1) and dimerisation domains. As will be discussed in the following sections the modular character of transcription factors

does not preclude multifunctionality of their various domains, as seen in the cases for example of basic leucine-zipper (bZIP) factors which dimerise and bind DNA through the same domain and also the MADS box transcription factors (see Sections 1.1.1.1.2 and 1.2.3).

1.1.1.1 Protein-DNA Interactions

Transcription factors derive their ability to interact with specific DNA sequences from the fact that their DNA-binding domains contain surfaces that are chemically complementary to particular patterns of base pairs (Garvie, C. W. *et al.*, 2001). Thus the DNA-binding surfaces of transcription factors form multiple base-specific hydrogen bonds and van der Waals interactions with the functional groups of base pairs, which are exposed in the minor and major grooves of the DNA. DNA binding also involves non-specific contacts through hydrogen bonds or salt bridges with the highly negatively charged sugar-phosphate backbone of the double helix (Garvie, C. W. *et al.*, 2001). The backbone conformation which is also sequence dependent to a certain degree may in some cases affect the protein-DNA interactions, as exemplified by MADS box transcription factors and their recognition of narrow minor grooves created by A-T tracts (see Section 1.2.3).

Different transcription factors employ distinct modes of DNA recognition in which different combinations of structural elements provide the main DNA-binding determinants. The most common interactions include the binding of an α-helix in the major groove, although examples also exist of β-strands or loops binding in the minor or major groove (Garvie, C. W. *et al.*, 2001). A great variety of DNA-binding motifs have been described to date and major categories include the helix-turn-helix, zinc finger, basic region leucine-zipper and helix-loop-helix motifs (Garvie, C. W. *et al.*, 2001).

1.1.1.1.1 The helix-turn-helix motif

The helix-turn-helix motif is one of the most common DNA-binding motifs. The broad range of proteins belonging to this domain superfamily, contain different variations of a three-helix bundle in which the conserved second and third helices provide direct DNA binding (Garvie, C. W. *et al.*, 2001). The C-terminal recognition helix, mediates the majority of DNA contacts by binding the major groove, while the second helix is located

above the recognition helix at a fixed angle and provides additional DNA contacts. The homeodomain is a variant type of the HTH motif, which uses an N-terminal extension from the α1 helix to contact the minor groove (illustrated in the ternary complex of the MATα1-MATα2 homeodomain heterodimer; (Li, T. *et al.*, 1995)). Another variant is the winged-HTH motif found in ETS domain proteins, which contains two flexible loops on either side of the main HTH domain through which it makes additional DNA contacts (see also Section 1.2.5.6).

1.1.1.1.2 The basic region leucine-zipper motif

Basic region leucine-zipper (bZIP) proteins contain long α-helices which dimerise by a coiled-coil leucine-zipper motif formed by their the C-terminal regions, with the N-termini spread apart to enter the major groove of DNA, so that the two α-helices adopt a scissor-like conformation via which they dimerise and interact with DNA (Garvie, C. W. *et al.*, 2001). bZIP proteins can interact with DNA as homo- or heterodimers as seen with the Fos and Jun proteins, providing added functional versatility (Chinenov, Y. *et al.*, 2001).

1.1.1.1.3 The helix-loop-helix motif

The helix-loop-helix motif is similar to the bZIP since it also mediates DNA-binding and dimerisation. Each monomer consists of two α-helices connected by a loop. These α-helical elements interact forming a four-helix bundle in which the N-terminal basic regions are inserted in the major groove of DNA. Like bZIP factors bHLH proteins can homo- or heterodimerise (Garvie, C. W. *et al.*, 2001).

1.1.1.2 Regulation of transcription factor activity

Regulated transcription depends on the tight control of transcription factor activity, so that the correct genes are active at the right time. Several different ways exist in which cells modulate transcription factor activity and these include post-translational modifications such as phosphorylation, regulation of the subcellular localisation, control of transcription factor levels by expression or degradation, interaction with additional regulatory proteins or a combination of these (Calkhoven, C. F. *et al.*, 1996).

1.1.1.2.1 Regulation of transcription factor activity by phosphorylation

Regulation of transcription factor activity by phosphorylation is exemplified by the TCF transcription factors which are heavily phosphorylated in response to MAP-kinase signalling resulting in stimulation of transactivation activity via interactions with transcriptional coactivators (see Section 1.2.5.6). The bZIP protein CREB is also phosphorylated by multiple kinases resulting in the recruitment of the CBP/p300 co-activator (De Cesare, D. *et al.*, 1999).

1.1.1.2.2 Regulation of transcription factor activity by subcellular compartmentalisation

Nuclear translocation of transcriptional factors provides inducibility to transcriptional regulation in response to extracellular stimuli. For example, the transcription factor NFκB is retained in the cytoplasm by association with the IκB inhibitory protein. In response to extracellular cues IκB is targeted for phosphorylation and is subsequently degraded, resulting in the nuclear translocation of NFκB and transcriptional activation (DiDonato, J. A. *et al.*, 1997). Other transcriptional regulators, such as MRTF and FHL2 also accumulate in the nucleus as a result of signalling inputs ((Miralles, F. *et al.*, 2003; Muller, J. M. *et al.*, 2002); for the regulation of the MRTFs see also Section 1.2.5.7).

1.1.1.2.3 Regulation of transcription factor activity by degradation

Whereas degradation of an interacting partner can result in the activation of a transcription factor, as in the case of IκB-NFκB, direct degradation of the transcription factor itself can also provide a means for regulating transcription. The best understood example for this kind of regulation is the activation of Tcf/Lef-1 via protein stabilisation of its activator β-catenin upon Wnt-signalling (Molenaar, M. *et al.*, 1996). In the absence of Wnt-signalling cytoplasmic β-catenin forms a complex with glycogen synthase kinase 3 (GSK3) and adenomatous polyposis coli (APC) (Rubinfeld, B. *et al.*, 1996). This complex allows the phosphorylation of β-catenin by GSK3, which in turn tags β-catenin for ubiquitination and subsequent degradation (Aberle, H. *et al.*, 1997). Disruption of the APC-GSK3-β-catenin complex by Wnt-signalling prevents β-catenin phosphorylation by GSK3 and inhibits its degradation, resulting in β-catenin nuclear translocation and the activation of gene expression.

1.1.1.2.4 Regulation of transcription factor activity by acetylation

Transcription factor activity can be regulated by the post-translational acetylation of lysines, which may result in modulations in DNA binding affinity, transcriptional activity and protein stability (reviewed in (Glozak, M. A. *et al.*, 2005)). Acetylation of a non-histone protein was first described for the tumour suppressor p53 (Gu, W. *et al.*, 1997). p53 is acetylated at five lysines within its C-terminal regulatory domain, which stimulates its DNA-binding affinity and transcriptional activity. Increased p53-dependent gene expression by acetylation has been reported to occur in vivo in response to cellular stresses (Luo, J. *et al.*, 2004).

1.1.1.2.5 Regulation of transcription factor activity by differential expression

Transcription factors are key mediators of the genetic programmes that underlie development. The specific spatial and temporal expression pattern of a transcription factor will dictate the cell growth and differentiation of particular tissues and their morphogenesis into organs, as well as the maintenance of the differentiation state. There are numerous examples of transcription factors involved in determining developmental fate via their specific patterns of expression, such as MyoD-Myf5-MRF4 in myogenic determination ((Berkes, C. A. *et al.*, 2005) and references therein), GATA-1 in erythroid cell differentiation (Pevny, L. *et al.*, 1991), Pax6 in eye development (Hill, R. E. *et al.*, 1991), or the Drosophila Dorsal transcription factor in determining dorsal-ventral axis (Roth, S. *et al.*, 1989).

1.1.1.2.6 Regulation of transcription factor activity by multiple mechanisms

As inferred from the examples presented in the previous section, it is common for a combination of mechanisms to be in place for the regulation of transcription factor activity. One such example is the multiple events controlling the activity of the Smad protein mediators of TGFβ signalling (Massague, J. *et al.*, 2005). Under basal conditions the Smads are continuously shuttling between the nucleus and cytoplasm. Activation of the TGFβ pathway results in their phosphorylation and retention in the nucleus, where they interact with additional factors to activate transcription. A similar multi-level regulatory mechanism is at place for members of the MRTF family, which are also subject to nucleocytoplasmic shuttling, phosphorylation and interaction with the SRF transcription factor (see Section 1.2.5.7.3).

1.1.1.3 Combinatorial interactions – regulation of transcription by multicomponent transcription factor complexes

Unlike prokaryotic transcription factors that often function individually to modulate transcription, eukaryotic gene transcription is regulated by multiple interacting transcription factors. This involves interactions between sequence-specific transcription factors bound to adjacent or distant regulatory promoter elements as well as association of DNA-binding factors with cofactors devoid of specific DNA-recognition activity (Ogata, K. *et al.*, 2003).

The regulation of transcription by combinatorial interactions between diverse transcription factors contributes to the versatility of their transcriptional responses, since different protein combinations support different transcriptional outcomes. This is exemplified by the interactions of MADS box transcription factors with their various cofactors which confer functional specificity (analysed in detail in Sections 1.2.4 and 1.2.5). The NFAT transcription factors are also involved in selective interactions with Fos/Jun heterodimers and FoxP3 in T-cell activation (Chen, L. *et al.*, 1998; Wu, Y. *et al.*, 2006). These interactions are mutually exclusive and controlled by different signals. NFAT cofactor selectivity is thought to function in T-cell lineage determination, with binding to Fos/Jun resulting in effector T-cell differentiation, binding to FoxP3 facilitating regulatory T-cell differentiation and lack of signals promoting either interaction leading to T-cell anergy (Rudensky, A. Y. *et al.*, 2006; Wu, Y. *et al.*, 2006). Thus combinatorial interactions of transcription factors also function to integrate multiple signals at gene promoters allowing diverse modes of regulation of the same gene depending on different combinations of signalling inputs.

1.1.1.3.1 The enhanceosome

The significance of combinatorial interactions between transcription factors in the regulation of transcription is illustrated by the assembly of distinct transcription factor – enhancer complexes at gene regulatory regions called enhanceosomes in response to specific signals (reviewed in (Carey, M., 1998; Courey, A. J., 2001; Merika, M. *et al.*, 2001)).

Assembly of these multiprotein complexes depends on the cooperative binding of individual transcription factors to their enhancer regions, which requires the specific composition and spatial arrangement of their DNA-binding sites. The prototype enhanceosome model is that assembled at the promoter region of the interferon β (IFNβ) gene in response to virus infection (Thanos, D. *et al.*, 1995). Assembly of this enhanceosome involves the cooperative binding of NFκB, c-Jun, ATF-2, and IRF family members to a 60 bp DNA segment containing specific DNA-binding sites. Binding of these factors critically depends on the architectural protein HMG I (Y), which alters the curvature of the DNA and acts to stabilise the enhanceosome complex ((Thanos, D. *et al.*, 1995); reviewed in (Merika, M. *et al.*, 2001)).

Correct stepwise enhanceosome assembly results in synergistic transcriptional activation, by presenting a surface structure, which is complementary to components of the coregulator complexes and the basal transcriptional machinery (Carey, M., 1998). The absence of a single factor can greatly destabilise the complex, and thus the enhanceosome represents a transcriptional switch which is inactive until all the necessary components are present (Courey, A. J., 2001).

1.1.2 Summary

In summary transcriptional regulation depends on the concerted action of multiple diverse transcription factors. The critical characteristics of these factors are their abilities to bind DNA in a sequence-specific manner and to ultimately decode the regulatory information present in the promoters of genes and transmit it to the basal transcriptional machinery. Moreover combinatorial interactions between transcription factors greatly impact on the correct execution of the transcriptional programme of each eukaryotic gene at different developmental stages and in response to different environmental conditions. The following sections will focus on the MADS box family of transcription factors, the properties of which illustrate the principles of regulatory versatility of transcriptional responses.

1.2 The MADS box transcription factor family

The MADS box family of eukaryotic transcriptional regulators is characterised by the presence of a conserved 56 amino acid domain named the MADS box after the five

founding members of the family: MCM1 (Minichromosome maintenance 1, found in yeast, (Passmore, S. *et al.*, 1989)), AG (agamous, found in plants, (Yanofsky, M. F. *et al.*, 1990)) or ARG80 (found in yeast, (Dubois, E. *et al.*, 1987)), DEFA (Deficiens found in plants, (Sommer, H. *et al.*, 1990)) and SRF (Serum Response Factor, found in animals, (Norman, C. *et al.*, 1988)).

Thus MADS box containing proteins are found in fungi, plants and animals, including nematodes, arthropods, lower vertebrates and mammals. Four MADS proteins have been described in *Saccharomyces cerevisiae* (MCM1, ARG80, Rlm1 and Smp1), while SRF and the Myocyte Enhancer 2 factors (MEF2A, B, C and D) are found in metazoans. In plants the family is greatly expanded and over one hundred MADS box containing candidate genes have been identified by genomic sequencing in *Arabidopsis thaliana* and over 70 in *Oryza sativa* (De Bodt, S. *et al.*, 2003).

MADS box containing proteins are involved in diverse biological processes ranging from cell proliferation and muscle differentiation in animals, to mating cell type determination, stress response and metabolism in yeast, and reproductive and vegetative development in plants (reviewed in (Messenguy, F. *et al.*, 2003) and (Shore, P. *et al.*, 1995b)). The importance of the MADS encoding sequences in mediating these functions has imposed a functional constraint on their evolution, resulting in a high degree of conservation of the MADS box in all the three major eukaryotic kingdoms.

1.2.1 Classification of MADS box transcription factors

Phylogenetic analysis based on the conservation of the MADS box sequences reveal that family members can be assigned to two broad lineages, which have been designated Type I and Type II (Figure 1.1; (Alvarez-Buylla, E. R. *et al.*, 2000)). These groups can be further subdivided based on the conservation of their sequences C-terminal to the MADS box. The presence of fungal, animal and most likely also plant members in both Type I and Type II MADS lineages has led to the proposal that these arose by a gene duplication that took place in a common ancestor before the divergence of the three major eukaryotic kingdoms (Alvarez-Buylla, E. R. *et al.*, 2000).

Thus Type II proteins comprise the MEF2-like and MIKC subfamilies. The MEF2-like class is defined by the presence C-terminally to the MADS box of the MEF2 domain, a

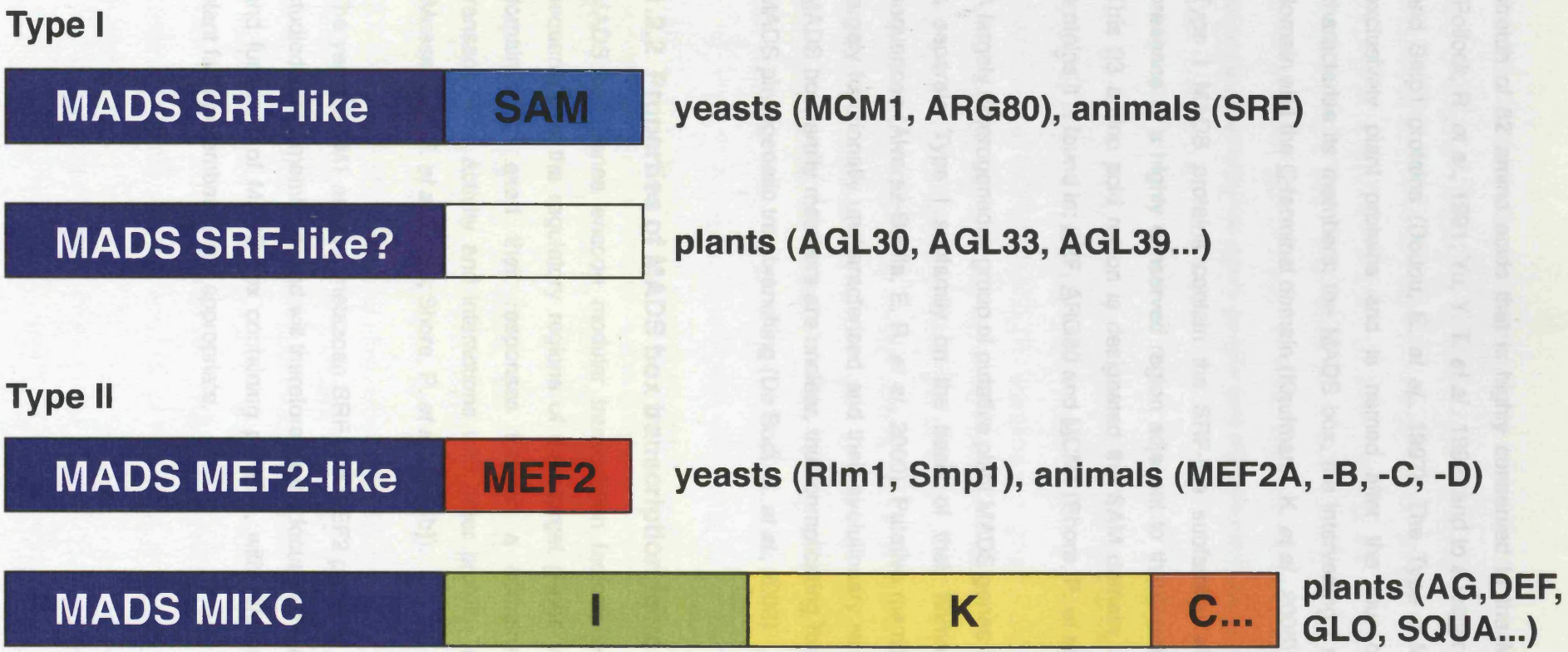


Figure 1.1. Schematic representation of the domains of plant, animal and fungal Type I and Type II MADS box proteins. In plant Type I proteins the "?" in the MADS box denotes the uncertainty surrounding the classification of their MADS domain and the white box C-terminal to the MADS domain indicates domains not well defined and of variable lengths. SRF-like SAM domain proteins and MEF2-like proteins also contain transactivation domains that are not pictured for simplicity. (Adapted from [Alvarez-Buylla et. al., 2000] and [Messenguy and Dubois, 2003].

stretch of 32 amino acids that is highly conserved in the MEF2 factors of animals (Pollock, R. *et al.*, 1991; Yu, Y. T. *et al.*, 1992) and to a lesser extent in the yeast Rlm1 and Smp1 proteins (Dodou, E. *et al.*, 1997). The Type II MIKC subfamily contains exclusively plant proteins and is named after the four conserved domains that characterise its members: the MADS box, the intervening I region, the keratin-like K domain and the C-terminal domain (Kaufmann, K. *et al.*, 2005).

Type I MADS proteins contain the SRF-like subfamily, which is grouped by the presence of a highly conserved region adjacent to the C-terminus of the MADS box. This 23 amino acid region is designated the SAM domain, after the initials of three proteins it is found in: SRF, ARG80 and MCM1 (Shore, P. *et al.*, 1995b).

A largely heterogeneous group of putative plant MADS genes has also been classed as a separate Type I subfamily on the basis of their homology in the MADS box sequences (Alvarez-Buylla, E. R. *et al.*, 2000). Putative members of this subfamily are largely functionally uncharacterised and their evolutionary relationships with the other MADS box family members are unclear, thus complicating the in depth analysis of the MADS phylogenetic tree branching (De Bodt, S. *et al.*, 2003).

1.2.2 Properties of MADS box transcription factors

MADS box genes encode modular transcription factors, which bind specific DNA sequences in the regulatory regions of their target genes via discrete DNA-binding domains and exert their responses through a combination of their separate transactivation activity and interactions with other proteins (see below; reviewed in (Messenguy, F. *et al.*, 2003; Shore, P. *et al.*, 1995b)).

The yeast MCM1 and the metazoan SRF and MEF2 proteins are the most extensively studied family members, and will therefore be the focus of the analysis of the properties and functions of MADS box containing proteins, with references to other yeast and plant family members were appropriate.

1.2.2.1 DNA binding site specificity of MADS box transcription factors

The DNA sites recognised by MADS box transcription factors are dyad symmetric sequences approximately 10 bp in length, which contain an A/T rich core flanked by conserved C-G base pairs. The DNA binding site specificities of various family members have been determined by in vitro binding site selection assays. Although these methods are designed to select high affinity sites, the consensus sequences determined in vitro for each protein usually correlate well with the naturally occurring sites found in target promoters.

The SRF consensus DNA binding sequence was determined by in vitro binding site selection as CC(A/T)TATA(A/T)GG (Pollock, R. *et al.*, 1990). This is a more stringent version of the $CC(A/T)_2A(A/T)_3GG$ consensus originally identified in the promoters of SRF controlled genes (Minty, A. *et al.*, 1986; Treisman, R., 1985; Treisman, R., 1986). Non-consensus changes in these positions in SRF binding sites decrease the DNA binding affinity of SRF by tenfold (Leung, S. *et al.*, 1989; Wynne, J. *et al.*, 1992). Nevertheless mismatched binding sites are frequently found in the regulatory regions of SRF target genes, with mismatches often conserved between species suggesting that they serve a specific role in vivo (SRF target genes analysed in section 1.2.5.4; (Miano, J. M., 2003; Selvaraj, A. *et al.*, 2004; Zhang, S. X. *et al.*, 2005)). The binding site selection experiments also discovered that the nucleotides flanking the core 10bp SRF consensus contribute to binding specificity and identified a degree of asymmetry in the flanking sequences selected by SRF (Pollock, R. *et al.*, 1990).

The same method identified the consensus binding site of MCM1 as CC(C/T)(A/T)(A/T)(A/T)NN(A/G)G (Wynne, J. *et al.*, 1992). In vivo binding site mutagenesis analysis determined a similar but symmetrical consensus (TTACCNAATTNGGTAA) and uncovered the importance of the three residues flanking the core 10bp site on either side (Acton, T. B. *et al.*, 1997). Thus the DNA specificity of MCM1 is related to that of SRF, but differs in significant ways, the most important being a lower stringency for the central A/T sequences. These differences explain the ability of SRF and MCM1 to bind subsets of each other's sites with variable efficiencies, since certain SRF sites represent varieties of the MCM1 consensus and vice versa (Wynne, J. *et al.*, 1992).

The CTA(A/T)$_4$TAG consensus binding sequence of the MEF2 proteins was also determined by binding site selection and differs significantly from those of its SRF-like relatives (Pollock, R. *et al.*, 1991). Sequences almost identical to the MEF2 consensus are recognised by the yeast MEF2-like proteins Rlm1 and Smp1 (Dodou, E. *et al.*, 1997). As a result of their distinct DNA specificities MEF2-like and SRF-like proteins fail to recognise each other's sites, reflecting crucial differences in the ways they contact DNA ((Pollock, R. *et al.*, 1991); analysed in Section 1.2.3.1).

The consensus binding sites of various plant proteins have also been determined by similar methods. Thus the *Arabidopsis thaliana* proteins AG and AGL1 recognise sequences resembling the MCM1 consensus, while AGL2 and AGL3 recognise sequences more closely related to MEF2 sites (Huang, H. *et al.*, 1993; Riechmann, J. L. *et al.*, 1996b). Another Arabidopsis protein, SQUA has dual specificity, since it binds SRF and MEF2 sites and a wide range of intermediate sites with varying affinities (West, A. G. *et al.*, 1998).

Thus despite their overall common characteristics, significant differences exist in the consensus DNA binding sites recognised by different MADS box proteins, which preclude or greatly reduce binding of one family member to the recognition site of another.

1.2.3 The DNA-binding domain of MADS box transcription factors

The minimal DNA-binding domains of many family members have been characterised, including those of SRF, MCM1, MEF2 and various MIKC plant proteins. These encompass the MADS box and the conserved sequences C-terminal to it (SAM, MEF2 or I domain) and are sufficient for dimerisation and high affinity specific DNA binding (Figure 1.2; (Mueller, C. G. *et al.*, 1991; Norman, C. *et al.*, 1988; Pollock, R. *et al.*, 1991; Riechmann, J. L. *et al.*, 1996a)). In this module the primary DNA and dimerisation contacts are mediated by the MADS box, with the C-terminal region contributing to dimerisation and high affinity DNA binding without directly contacting DNA (see below).

MADS box transcription factors commonly depend on combinatorial interactions with other proteins to fulfil their regulatory roles and achieve target gene specificity or

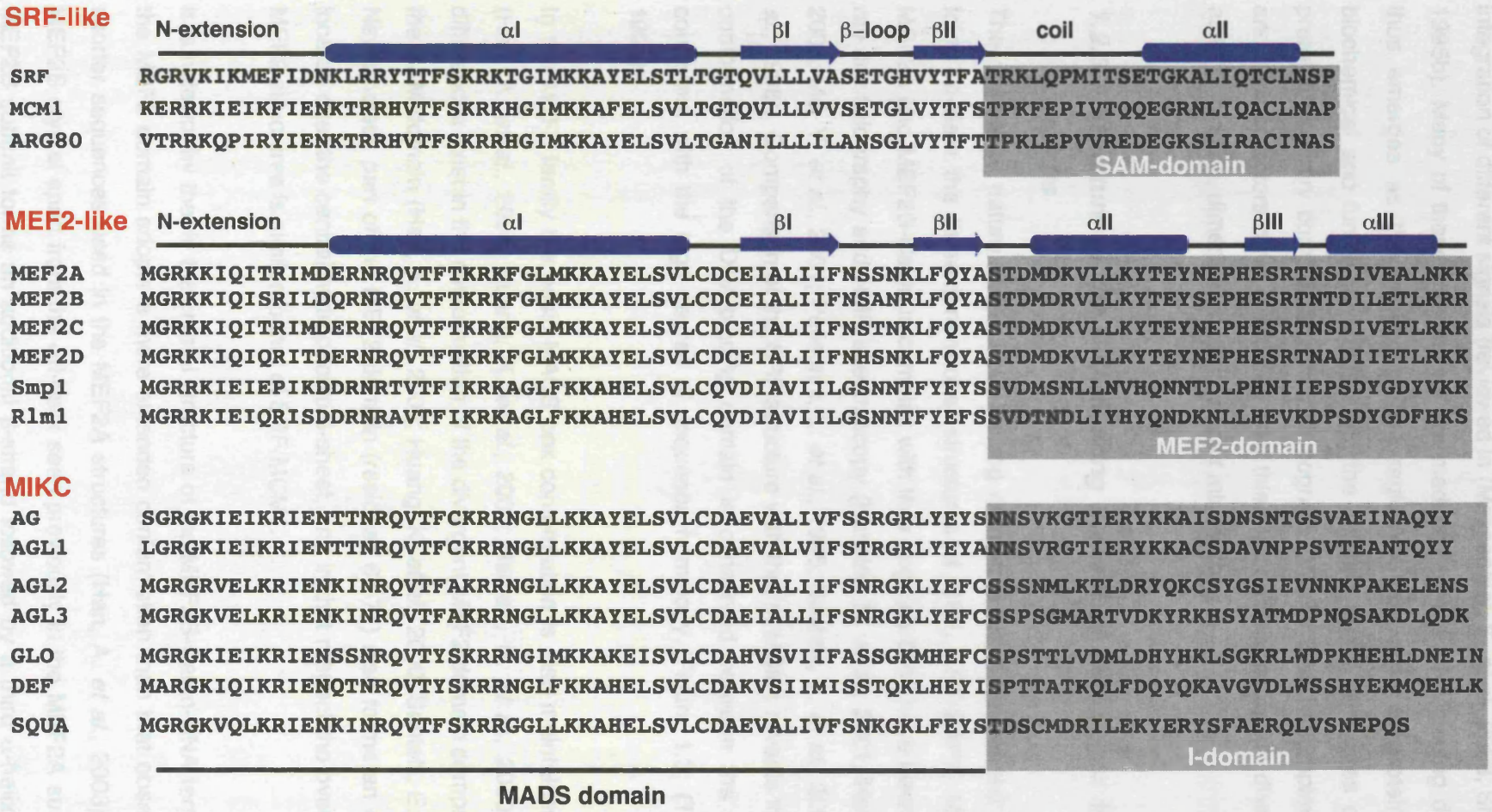


Figure 1.2. The DNA-binding domains of SRF-like, MEF2-like and MIKC MADS box proteins. The sequences of the MADS boxes from human (SRF, MEF2), yeast (MCM1, ARG80, Smp1, Rlm1) and selected plant MIKC proteins are shown, with the C-terminal SAM, MEF2 and I- domains indicated by grey boxes. The secondary structure shown for SRF- and MEF2-like proteins is based on the solution of their three-dimentional structures (Section 1.2.3). The highly homologous MADS boxes of the plant proteins indicate that they adopt similar conformations. The methionine at the beginning of the MEF2-like factors and some MIKC proteins is the initiation codon.

integration of different signals (reviewed in (Messenguy, F. *et al.*, 2003; Shore, P. *et al.*, 1995b). Many of these interactions are mediated by the DNA-binding domain, which thus emerges as the most important region of this class of proteins. Extensive biochemical and functional studies and the solution of the structures of MADS box proteins in binary complexes with their cognate DNA or ternary complexes with DNA and their cofactors have elucidated how this region integrates such diverse functions as DNA-binding, dimerisation and cofactor interaction.

1.2.3.1 Structure of the DNA-binding domain of MADS box transcription factors

The composite nature of the DNA-binding domain is largely explained by its unique fold. To date the three-dimensional structures of SRF, SRF-SAP1, MCM1-MATα2, MEF2A and MEF2B-Cabin in complex with their cognate DNA have been solved by X-ray crystallography and NMR spectroscopy (Hassler, M. *et al.*, 2001; Huang, K. *et al.*, 2000; Mo, Y. *et al.*, 2001; Pellegrini, L. *et al.*, 1995; Santelli, E. *et al.*, 2000; Tan, S. *et al.*, 1998). Comparison of the SRF structure with that of MCM1 reveals that the overall conformation of the DNA-binding domain is conserved between the two proteins, consistent with the high degree of sequence homology (Figure 1.3; (Tan, S. *et al.*, 1998)).

In the MEF2 family the basic MADS box conformation is also maintained (Figure 1.3; (Han, A. *et al.*, 2003; Huang, K. *et al.*, 2000; Santelli, E. *et al.*, 2000)). Significant differences exist in the conformation of the divergent MEF2 domain compared to that of the SAM domain (Han, A. *et al.*, 2003; Huang, K. *et al.*, 2000; Santelli, E. *et al.*, 2000). Nevertheless part of the MEF2 domain (residues 61-71) also forms an α-helix that is located over the central hydrophobic β-sheet, and in that respect the overall fold of the MEF2 structures is similar to that of SRF/MCM1.

It is noteworthy that in the crystal structure of the MEF2B-Cabin-DNA ternary complex, the MEF2 domain adopts a more extended conformation than that observed with the shorter sequences used in the MEF2A structures (Han, A. *et al.*, 2003). Thus in the MEF2B crystal apart from the αII-helix seen previously in the MEF2A structures, each MEF2B subunit forms an additional β-strand followed by a third α-helix (Figure 1.4). These elements reach over the opposing subunit so that the βIII-strands of one

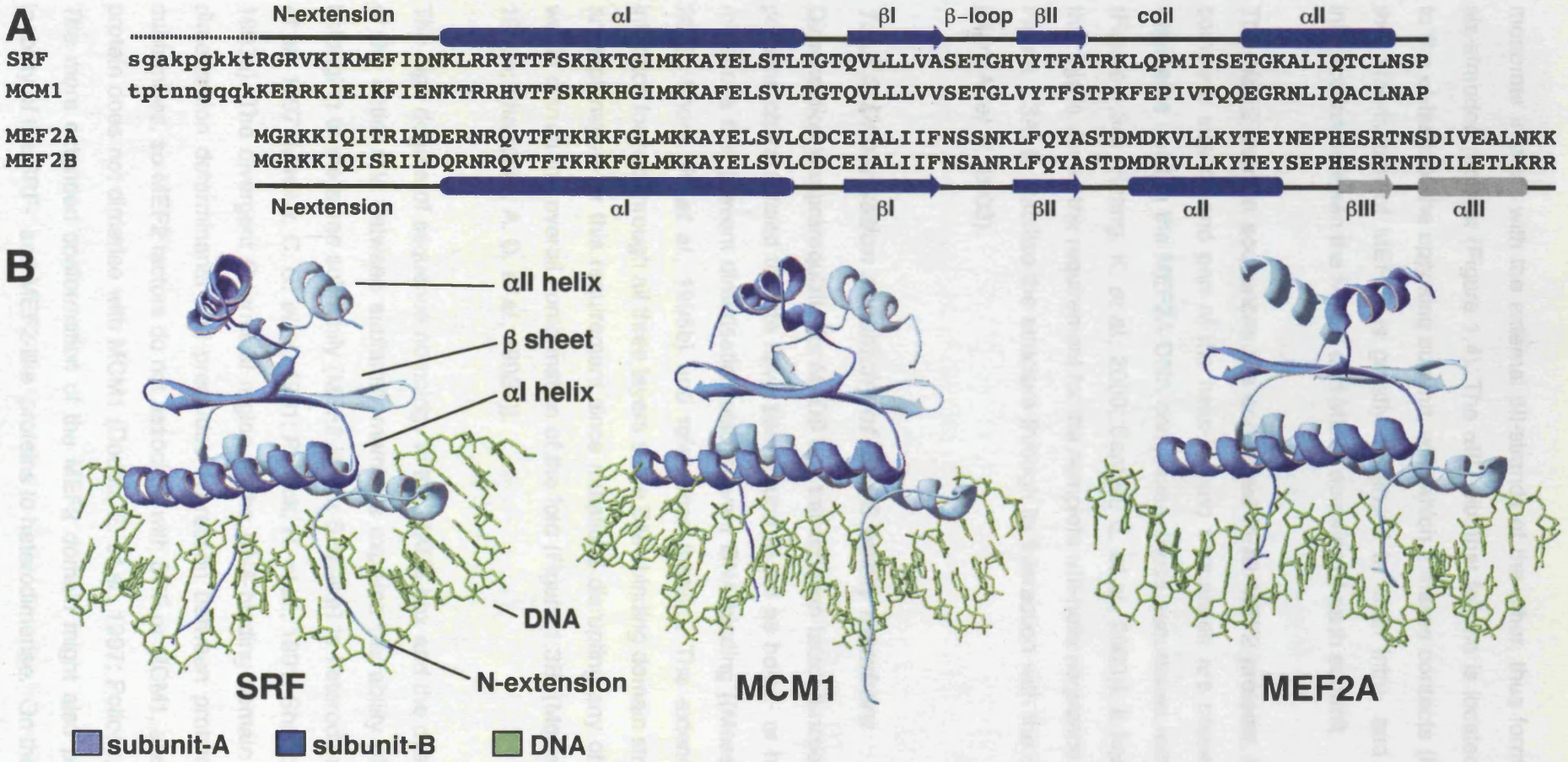


Figure 1.3. Structures of the DNA-binding domains of SRF, MCM1 and MEF2A on DNA. (A) Sequences and secondary structure of the DNA-binding domains (SRF residues 132-222, MCM1 residues 7-97, MEF2A and -B residues 1-91). The MADS box extends from the N-extension to the bII-strand and is followed by the SAM (in SRF and MCM1) or MEF2-domain (MEF2 proteins). The additional N-extension regions of SRF and MCM1 prior to the MADS box are unstructured and shown in lower case letters.The extended MEF2-domain sequence is shown with the secondary structure derived from the MEF2B structure in grey (see Section 1.2.3). (B) Three-dimensional structures of SRF, MCM1 and MEF2A bound to their DNA sites (produced from PDB files 1SRS, 1MNM and 1EGW). Only the aII-helix is ordered in the MEF2A structure (see text in Section 1.2.3).

monomer interact with the external βII-strands of the other, thus forming an extended six-stranded β-sheet (Figure 1.4). The αIII-helix that follows is located perpendicularly to the αI-helix of the opposing subunit, with which it makes contacts (Figure 1.4). Thus the dimerisation of MEF2B is greatly enhanced by both intra- and inter-subdomain interactions between the MADS and MEF2 domains of each subunit.

The MEF2 domain sequences are conserved in all MEF2 proteins, but although the complete strand- and part of the helix-forming sequences are present in the protein fragments used in the MEF2A-DNA complexes, these sequences remain unstructured (Figure 1.3A; (Huang, K. *et al.*, 2000; Santelli, E. *et al.*, 2000)). It has been proposed that this is due to the requirement for the complete αIII-helix sequence (residues 82-89, Figure 1.3A) to stabilise the structure through its interaction with the opposing αI-helix (Han, A. *et al.*, 2003).

1.2.3.2 Dimerisation specificity of MADS family members

Dimerisation is a prerequisite for MADS box transcription factor function. All MADS box proteins characterised to date have been found to act as homo- or heterodimers and mutations that prevent dimerisation also prevent DNA binding ((Messenguy, F. *et al.*, 2003; Shore, P. *et al.*, 1995b) and references therein). The extended dimerisation interface formed through all three layers of the DNA-binding domain structures provides an explanation for this requirement since mutations disrupting any of the three layers would disrupt the overall conformation of the fold (Figure 1.3B; (Molkentin, J. D. *et al.*, 1996a; Sharrocks, A. D. *et al.*, 1993a)).

The high degree of sequence homology in the MADS box and the overall conservation of the entire fold between subfamily members explains the ability of MADS proteins belonging to the same subfamily (MEF2-like or SRF-like) to heterodimerise (Dodou, E. *et al.*, 1997; Mueller, C. G. *et al.*, 1991; Pollock, R. *et al.*, 1991; Sharrocks, A. D. *et al.*, 1993a). The divergent C-terminal region of the DNA-binding domain provides further dimerisation determinants and precludes interaction between proteins from different subfamilies, so MEF2 factors do not associate with SRF or MCM1, and the yeast Rlm1 protein does not dimerise with MCM1 (Dodou, E. *et al.*, 1997; Pollock, R. *et al.*, 1991). The more extended conformation of the MEF2 domain might also play a role in the inability of the SRF- and MEF2-like proteins to heterodimerise. On the other hand the

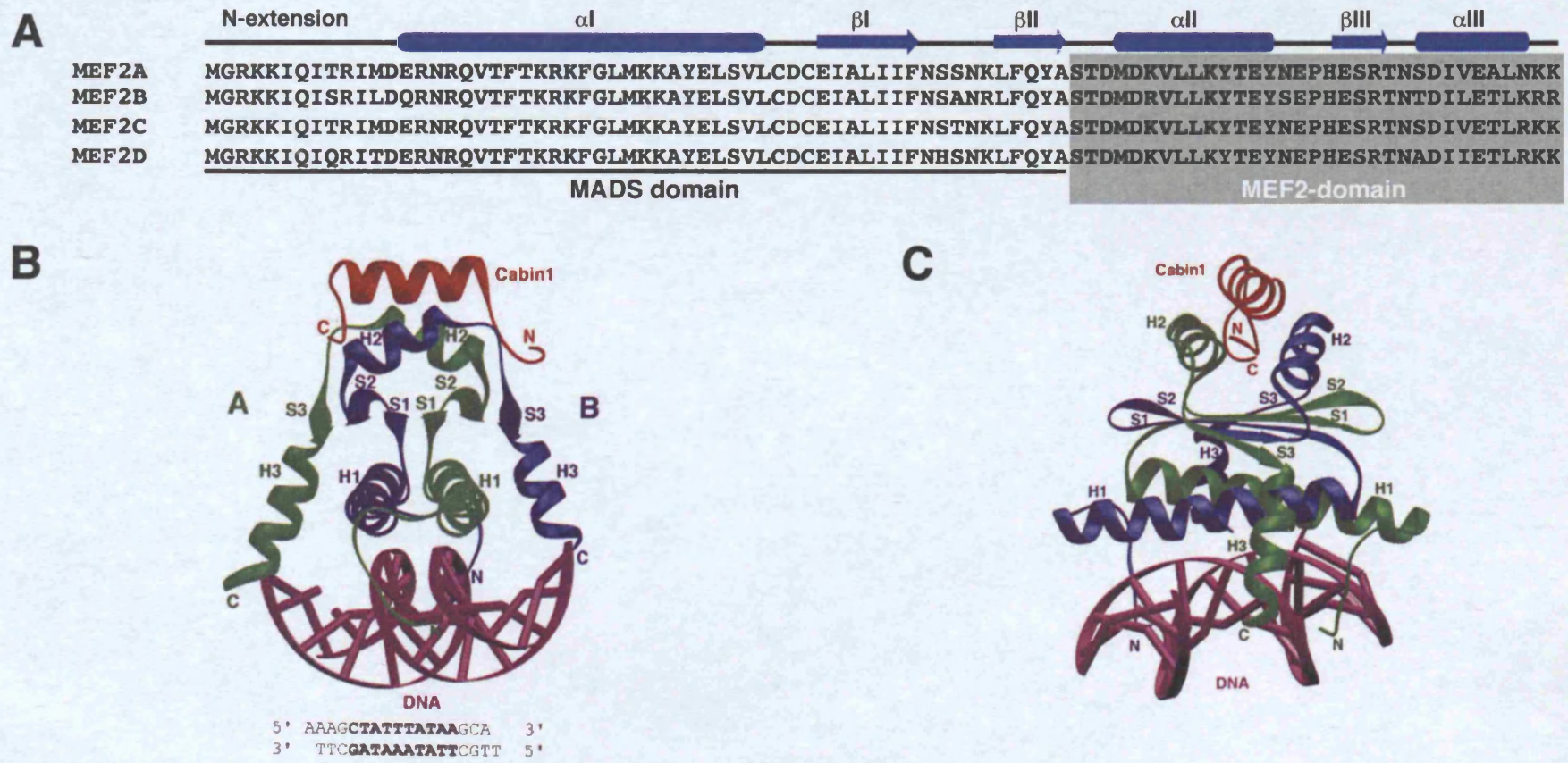


Figure 1.4. Structure of the Cabin1-MEF2B-DNA ternary complex. (A) The sequences of the DNA-binding domains of the mouse MEF2 proteins. The MADS box is indicated by a black line and the MEF2 domain by a grey box. The secondary structure elements are shown in each domain. (B) Ribbon representation of the MEF2B DNA-binding domain domains bound to DNA . The two MEF2B subunits are shown in blue and green the DNA in magenta.The helix of the Cabin1 repressor that interacts with MEF2B is also shown in red. The secondary structure elements are labeled in each MEF2 subunit (S1, S2 and S3 denote the βI, βII, and βIII strands and H1, H2, H3 the αI, αII and αIII helices respectively). The sequence of the DNA site of MEF2 is shown below. (C) The same complex viewed at 90° to B. Images were taken from (Han et. al., 2003).

overall conservation of the DNA-binding domain fold allows chimeric MADS proteins comprising N-terminal and C-terminal regions derived from different subfamily members to dimerise and bind DNA (Mueller, C. G. *et al.*, 1991; Pollock, R. *et al.*, 1991).

Although the structure of the plant MIKC proteins has not been solved their MADS domains are predicted to adopt similar conformations to those of yeast and metazoan proteins based on sequence conservation (Kaufmann, K. *et al.*, 2005). Biochemical analysis of MIKC subfamily members identified the I-region that follows the MADS domain as specifying dimerisation (Riechmann, J. L. *et al.*, 1996a). This region is predicted to adopt a helical conformation, and is thus thought to act like the MEF2 and SAM domains by dimerising and shielding the hydrophobic central β-sheet from solvent (Kaufmann, K. *et al.*, 2005). Certain members of the MIKC plant family such as the DEF/GLO heterodimers of *Antirrhinum majus* and the PI and AP3 proteins of *Arabidopsis thaliana* require the presence of their K-region in addition to the minimal DNA-binding domain for efficient dimerisation and subsequent DNA binding (Riechmann, J. L. *et al.*, 1996b; West, A. G. *et al.*, 1998). The K-region is predicted to mediate dimerisation via three α-helices that form extended coiled coils, and is also thought to be involved in the formation of higher order complexes between family members (reviewed in (Kaufmann, K. *et al.*, 2005)).

1.2.3.3 DNA binding and sequence recognition

The elucidation of the three-dimensional structures of the DNA-binding domains of SRF- and MEF2-like factors confirmed previous biochemical analyses that mapped the DNA-binding modules to the N-terminal part of the MADS domain, with conserved basic residues mediating crucial DNA contacts (Figure 1.5A and B; (Molkentin, J. D. *et al.*, 1996a; Nurrish, S. J. *et al.*, 1995; Pollock, R. *et al.*, 1991; Sharrocks, A. D. *et al.*, 1993a)).

The antiparallel coiled coil formed by the αl-helices of the MADS box of each monomer is the primary DNA binding unit in all the structures described in the previous section (Figure 1.5). This coiled coil lies over the minor groove in the middle of the dyad-symmetric binding site and interacts with the major groove on either DNA half-site, allowing the otherwise unstructured N-terminal extension to enter the minor groove and

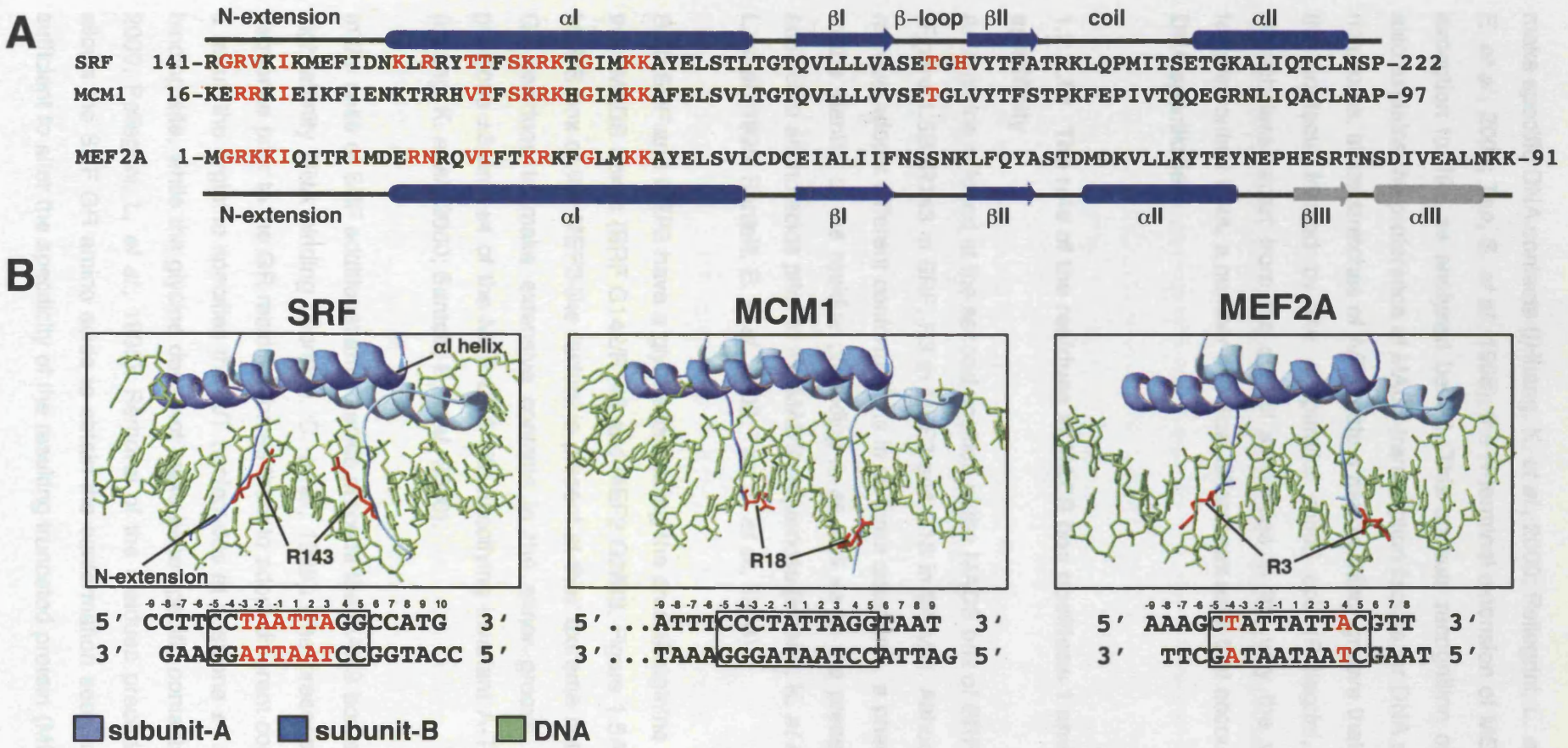


Figure 1.5. The interaction of the DNA-binding domains of SRF, MCM1 and MEF2A with their cognate DNA. (A) Sequences and secondary structures of the DNA-binding domains. Residues that make direct DNA contacts are shown in red. (B) The conserved N-terminal arginine of the MADS box adopts different conformations in different MADS proteins. Three-dimensional structures of SRF, MCM1 and MEF2A bound to their DNA sites (produced from PDB files 1SRS, 1MNM and 1EGW). Only the N-extension and al-helix of the DBDs are shown in ribbon representation. The side-chains of SRF R143, MCM1 R18 and MEF2A R3 are shown in red. The sequences DNA sites used in the crystals are shown below. Base-pairs contacted by the arginine residue are shown in red.

make specific DNA contacts ((Huang, K. *et al.*, 2000; Pellegrini, L. *et al.*, 1995; Santelli, E. *et al.*, 2000; Tan, S. *et al.*, 1998); the N-terminal extension of MCM1 is an important exception to this, as analysed below). This unusual recognition of the minor groove also explains the preference of MADS transcription factors for DNA sequences with A/T rich cores, since stretches of A/T create a narrow minor groove that can accommodate the contacts formed by the antiparallel coiled coil (Pellegrini, L. *et al.*, 1995). Nonetheless, apart from the overall similarities in the way the various MADS box factors contact DNA, a number of critical differences exist that account for their different DNA specificities.

1.2.3.3.1 The role of the residues at MADS box positions 1 and 2 in DNA binding specificity

An arginine is found at the second position of the MADS box of SRF, MEF2 and MCM1 (Figure 1.5B; R143 in SRF, R3 in MEF2 and R18 in MCM1). Although conserved, this residue adopts different conformations in all three structures, a phenomenon attributed to the identity of the residue preceding it, as well as to the presence or absence of additional amino acids prior to the MADS sequences (Huang, K. *et al.*, 2000; Pellegrini, L. *et al.*, 1995; Santelli, E. *et al.*, 2000; Tan, S. *et al.*, 1998).

Both SRF and MEF2 have a glycine preceding the crucial arginine at the beginning of their MADS boxes (SRF G142/R143 and MEF2 G2/R3, Figure 1.5A). The fact that the MADS box of the MEF2-like factors is present at their extreme N-terminus allows the GR residues to make extensive contacts in the minor groove with the bases at positions ±3 and ±4 of the MEF DNA site specifying invariant A-T at these positions (Huang, K. *et al.*, 2000; Santelli, E. *et al.*, 2000).

In the case of SRF additional amino acids precede the MADS box and are required for high affinity DNA binding (Norman, C. *et al.*, 1988). The presence of this extended sequence prior to the GR residues forces them to adopt different conformations and as a result the arginine specifies the A/T nucleotides at positions ±1 and ±2 of the SRF binding site, while the glycine does not make base specific contacts (Huang, K. *et al.*, 2000; Pellegrini, L. *et al.*, 1995). Removal of the residues preceding the MADS box allows the SRF GR amino acids to obtain the conformation seen in MEF2 and this is sufficient to alter the specificity of the resulting truncated protein ($METcore^{SRF}$) from the

wild-type SRF binding consensus to that of MEF2 (Nurrish, S. J. *et al.*, 1995; Sharrocks, A. D. *et al.*, 1993b). Thus the presence of a glycine prior to the conserved arginine specifies MEF2 DNA sites, unless the protein sequence is extended N-terminally to the MADS box in which case SRF-specific sites are recognised.

The arginine residue of MCM1 (R18) adopts a different conformation to those seen in MEF2 and SRF and does not enter the minor groove thus allowing greater flexibility in the central DNA sequences (Tan, S. *et al.*, 1998). This change in conformation is attributed to the presence of a glutamate (E17) in the position occupied by glycine in MEF2 and SRF (Huang, K. *et al.*, 2000; Santelli, E. *et al.*, 2000). This residue has been found to be crucial for MCM1 DNA specificity since a G143E substitution in SRF is sufficient to convert its DNA specificity to that of MCM1 (Nurrish, S. J. *et al.*, 1995). This property of the N-terminal MADS box sequences was successfully exploited to create altered DNA binding specificity SRF mutant derivatives that could bind MCM1 specific sites ((Hill, C. S. *et al.*, 1993); see also Chapter 2).

1.2.3.3.2 The role of the residues at MADS box positions 11-15 on DNA binding specificity

Another important DNA specificity determinant is the identity of the residues at positions 11-15 of the MADS box (SRF residues 152-156, MCM1 27-31 and MEF2A 12-16) (Nurrish, S. J. *et al.*, 1995). In SRF substitution of these sequences for those of MEF2A relaxes its DNA specificity so it can recognise both the SRF and MEF2 consensus sites. Additional removal of the sequences preceding the SRF MADS box completes the specificity switch from SRF to MEF sites exclusively (Nurrish, S. J. *et al.*, 1995). In the case of MCM1 combined substitutions at positions E17 and 27-31 with the MEF2 equivalents change its DNA specificity to that of MEF2 and vice versa (Nurrish, S. J. *et al.*, 1995). The effect of these residues on DNA specificity is not direct since none of them contacts the core 10bp binding site (Huang, K. *et al.*, 2000; Pellegrini, L. *et al.*, 1995; Santelli, E. *et al.*, 2000; Tan, S. *et al.*, 1998). At least in the case of SRF and MEF2 biochemical analysis of residues within this region has linked their effects on DNA binding specificity to their effects on DNA bending (analysed in the following section).

1.2.3.4 DNA bending by MADS box transcription factors

DNA bending is a highly conserved property of MADS box transcription factors through the plant, fungal and animal kingdoms. The ability of MADS proteins to distort their DNA sites to varying degrees has been well documented biochemically. SRF-like factors, including SRF, MCM1 and ARG80 induce substantial DNA bending, whereas MEF2-like factors including the MEF2 family and the yeast Rlm1 and Smp1 proteins have minimal DNA bending effects (Gustafson, T. A. *et al.*, 1989; Meierhans, D. *et al.*, 1997; West, A. G. *et al.*, 1998; West, A. G. *et al.*, 1999; West, A. G. *et al.*, 1997). Many plant MIKC MADS proteins have also been found to induce DNA bending including SQUA, AP1, AG, PLENA and also AP3/PI and DEF/GLO heterodimers (Riechmann, J. L. *et al.*, 1996b; West, A. G. *et al.*, 1998; West, A. G. *et al.*, 1999). It is clear however that different MADS proteins employ subtly different mechanisms to distort their DNA sites.

The difference between the DNA bending abilities of SRF-like and MEF2-like factors is apparent in the three-dimensional complexes of these proteins with their cognate DNA, in which SRF and MCM1 bend DNA by 72°, while MEF2A only by 15°-17° (Huang, K. *et al.*, 2000; Pellegrini, L. *et al.*, 1995; Santelli, E. *et al.*, 2000; Tan, S. *et al.*, 1998). In all three complexes there is a 15°-20° bend at the centre of the dyad axis of the DNA, and this is the only bend seen in the MEF2A-DNA complex (Huang, K. *et al.*, 2000; Santelli, E. *et al.*, 2000). SRF and MCM1 on the other hand induce further bends on either side of their core binding site, which add together resulting in significantly higher overall bend angles of the DNA in the complex (Pellegrini, L. *et al.*, 1995; Tan, S. *et al.*, 1998). These additional bends are attributed to the fact that both SRF and MCM1 contact the DNA flanking their core binding sites using residues from their αl-helices and also the β-loop located in their middle layer (Pellegrini, L. *et al.*, 1995; Tan, S. *et al.*, 1998). These interactions, which are thought to create a "pull" on the DNA, bending it around the N-terminus of the αl-helix of each subunit, are absent in the MEF2 structure and as a result the DNA is only bent in the middle of the dyad (Huang, K. *et al.*, 2000; Santelli, E. *et al.*, 2000).

In the case of SRF central to the interaction with the flanking DNA sequence is residue K154, which is located at position 14 of the MADS box within a sequence previously found to influence DNA-specificity (Figure 1.2; see also section 1.2.3.3.2;(Nurrish, S. J. *et al.*, 1995)). K154 acts in concert with other αl-helix residues to facilitate the binding

of β-loop residues to DNA. In MEF2 factors the position equivalent to K154 is occupied by a glutamate (glutamine in the case of MEF2B, see below), which does not contact DNA (Han, A. *et al.*, 2003; Huang, K. *et al.*, 2000; Santelli, E. *et al.*, 2000). A K154E substitution in the relaxed specificity METcore^{SRF} not only converts its specificity from SRF to MEF2 sites exclusively (Nurrish, S. J. *et al.*, 1995; Sharrocks, A. D. *et al.*, 1993b), but also reduces DNA bending (West, A. G. *et al.*, 1997). The reverse mutation E14K in MEF2A increases DNA bending and also relaxes its DNA specificity to include both MEF2 and SRF consensus sites (West, A. G. *et al.*, 1997). Although these results implicate this position in the MADS box in both DNA binding and DNA bending, they must be treated with caution since they use the relaxed specificity METcore^{SRF}.

MEF2B has a glutamine instead of glutamate at position 14, but retains the DNA specificity of the other MEF2 factors and also induces minimal DNA bending (Molkentin, J. D. *et al.*, 1996b; West, A. G. *et al.*, 1999). Thus a link emerges between the identity of the residue at SRF position 154 (MEF2 position 14) and the ability of the protein to bend DNA, and this correlation depends on whether the residue at this position is able to contact DNA. MEF2C, which also has a glutamate at position 14, appears to be an exception, since it has been reported to bend DNA significantly (Meierhans, D. *et al.*, 1997). This result is surprising considering that the sequence of the MEF2C MADS box is almost identical to that of MEF2A and is also very well conserved with the other MEF2 factors (Figure 1.2). It should be noted that DNA bending analysis in this study was not done in parallel with other MADS factors of known bending abilities, and it is thus possible that experimental conditions resulted in an unusually high apparent bend angle value.

The correlation between the presence of a DNA contact forming residue at position K154/E14 with the ability of the protein to induce DNA bending, appears to be context dependent, since although a lysine is found at the same position in MCM1 (K29), this residue does not contact DNA and has only a small effect on DNA bending (Lim, F. L. *et al.*, 2003). Instead MCM1 employs other αI-helix residues to contact the DNA bordering its core site, and aid the β-loop residue-DNA contacts. This is another example of SRF and MCM1 sharing overall similarities in the ways they contact DNA, while at the same time exhibiting subtle but crucial differences. Another such difference is that while DNA bending appears to be sequence independent in the case of SRF (West, A. G. *et al.*, 1997), the degree of bending induced by MCM1 depends on the

sequence of the DNA bordering its core site (Acton, T. B. *et al.*, 1997), with symmetrically placed T:A base pairs at positions ±7 from the centre of the binding site increasing the DNA bend (Lim, F. L. *et al.*, 2003).

The *Arabidopsis thaliana* SQUA protein, which binds DNA sites ranging between the SRF and MEF2 consensus, also induces DNA bending in a sequence dependent manner (West, A. G. *et al.*, 1999). SQUA has a lysine (K14) at the position equivalent to SRF K154, and this is also involved in bending but depends on residues in the β-loop to specify DNA binding (West, A. G. *et al.*, 1999).

1.2.3.4.1 The significance of DNA bending by MADS box transcription factors

Despite the wealth of biochemical information, the functional significance of DNA bending by MADS box transcription factors in vivo remains unclear. In the case of MCM1, mutations have been identified that impede DNA bending and exhibit slow-growth phenotypes, hinting at the importance of this property for MCM1 function in vivo (Acton, T. B. *et al.*, 2000; Lim, F. L. *et al.*, 2003).

The correlation between the abilities of SRF and MEF2 factors to recognise their consensus DNA site and induce DNA bending, combined with the fact that in general MEF2-like factors induce minimal DNA bending, while SRF-like factors bend their DNA sites significantly indicates that one of the functions of DNA bending in MADS box transcription factors is to provide a mechanism of "indirect readout" for these highly similar DNA-binding modules to distinguish between their DNA sites (Pellegrini, L. *et al.*, 1995; West, A. G. *et al.*, 1999).

Sequence specific DNA bending as seen for MCM1 and SQUA also provides extra specificity determinants for target gene selection, an important feature for proteins, which regulate multiple cell processes and control the activity of different subsets of genes. Furthermore at least for MCM1 and SRF the ability to distort their DNA sites influences cofactor interaction (see section 1.2.4 and Chapter 4; (Lim, F. L. *et al.*, 2003; Zaromytidou, A. I. *et al.*, 2006)), thus providing additional mechanisms by which to ensure specificity of their responses. In the case of the interaction of SRF with its cofactor MAL the induction of appropriate DNA distortion in the SRF-DNA complex is

critical for MAL binding and this is thought to facilitate direct MAL- DNA contacts (See Chapters 4 and 5; (Zaromytidou, A. I. *et al.*, 2006)).

MADS box transcription factors display a startling diversity in the transcriptional responses they elicit in different cell types and in response to different developmental and environmental stimuli (reviewed in (Messenguy, F. *et al.*, 2003; Shore, P. *et al.*, 1995b)). The versatility of their functions is achieved mainly by combinatorial interactions with accessory factors, which produce specific regulatory complexes on different target gene promoters. Thus MADS box proteins act as docking sites for a diverse array of cofactors through which distinct and sometimes opposing signalling cues are transduced. The following sections will discuss the MCM1 and SRF MADS box proteins focusing on the latter, and will analyse how combinatorial interactions contribute to the diverse functional roles of these transcription factors.

1.2.4 MCM1

Mcm1 was originally identified as a *Saccharomyces cerevisiae* gene necessary for the extrachromosomal maintenance of plasmids carrying specific autonomously replicating sequences (Minichromosome maintenance 1, ARS; (Maine, G. T. *et al.*, 1984)). It was subsequently discovered that this gene coded for a 286 amino acid transcription factor responsible for cell-type specific gene expression (Bender, A. *et al.*, 1987; Keleher, C. A. *et al.*, 1988; Passmore, S. *et al.*, 1988).

MCM1 contains an N-terminal DNA binding domain, which comprises the MADS and SAM domains (see Section 1.2.3) and is necessary and sufficient for cell viability (Bruhn, L. *et al.*, 1992; Primig, M. *et al.*, 1991). The MADS box is preceded by a 17 amino acid N-terminal extension which is phosphorylated in vivo at two major sites (Kuo, M. H. *et al.*, 1997). Deletion of the N-extension or mutation of the phosphorylation sites impairs growth on high salt. A stretch of 19 acidic residues located directly C-terminal to the SAM domain has been reported to affect α-cell type specific gene expression (Bruhn, L. *et al.*, 1992). MCM1 also contains a C-terminal transactivation domain, which is composed of 50% glutamine residues and is required for optimal transcriptional activation (Bruhn, L. *et al.*, 1992).

1.2.4.1 Biological roles and cofactor interactions of MCM1

MCM1 is expressed in all three yeast cell types (α, **a** and α/**a**, see later). This protein is involved in multiple functions, including cell type specification, the pheromone response, cell cycle control, minichromosome maintenance, osmotolerance and arginine metabolism. MCM1 elicits these diverse responses through interactions with different cofactors that are signal regulated, cell-type or promoter- specific (Table 1.1; reviewed in (Messenguy, F. *et al.*, 2003)). The following sections will describe some of the biological roles of MCM1 and the cofactor interactions that result in specificity of the transcriptional response.

Table 1.1 Combinatorial interactions involving MCM1 and its cofactors[1]

Cofactor	Cellular process	Experimental evidence
MATα1	Cell type specification: Activation of α- specific genes	EMSA, DNase I protection (Bender, A. *et al.*, 1987; Tan, S. *et al.*, 1988)
MATα2	Cell type specification: Repression of **a**-specific genes	EMSA, DNase I protection (Keleher, C. A. *et al.*, 1988; Keleher, C. A. *et al.*, 1989) Crystal structure (Tan, S. *et al.*, 1998)
STE12	Pheromone response Activation of α- and **a**- genes	EMSA (Errede, B. *et al.*, 1989)
Fkh2	Cell cycle control Activation of CLB2 cluster genes	EMSA (Kumar, R. *et al.*, 2000; Pic, A. *et al.*, 2000)
Yox1 and Yhp1	Cell cycle control Repression of ECB genes	Co-IP (Pramila, T. *et al.*, 2002)
ARG80	Arginine metabolism	EMSA (Dubois, E. *et al.*, 1991)
ARG81	Arginine metabolism	EMSA (Amar, N. *et al.*, 2000)
ARG82	Arginine metabolism	Yeast 2-hybrid, GST-pulldown (El Bakkoury, M. *et al.*, 2000)

[1] Table adapted from Messenguy, F., and Dubois, E. (2003). Role of MADS box proteins and their cofactors in combinatorial control of gene expression and cell development. Gene *316*, 1-21..

1.2.4.1.1 Cell type specification

Saccharomyces serevisiae is able to exist in either a haploid or a diploid state. It has two haploid cell types, α and **a**, which produce different cell surface receptors and pheromones. When one cell-type is exposed to secreted pheromones of the other, the α and **a** cells are able to fuse and form the diploid **a**/α cell (see Section 1.2.4.1.2). MCM1 is expressed in all three types and is required for establishing the different gene expression patterns that define cell type identity (reviewed in (Dolan, J. W. *et al.*, 1991)). Thus **a**- and α-cells express **a**- and α-specific genes respectively and also a set of haploid-specific genes, all of which are repressed in the diploid state. MCM1 is able to orchestrate the expression of different gene networks by engaging cell type-specific partners. Thus, in α-cells MCM1 binds the MATα2 repressor on the promoters of **a**-specific genes and actively blocks their transcription (Keleher, C. A. *et al.*, 1988). In contrast its association with the MATα1 activator on the promoters of α-specific genes allows their expression (Bender, A. *et al.*, 1987). Neither MATα1 nor MATα2 are present in **a**-cells, resulting in the α-specific genes being effectively repressed, while the **a**-specific genes are transcribed. In **a**/α cells both the **a**- and haploid specific genes are repressed by the interaction of MATα2 with MCM1 and the MAT**a**1 protein respectively, whereas α-specific genes are repressed due to the absence of the MATα1 activator.

1.2.4.1.1.1 The MCM1 - MATα2 interaction

Repression of **a**-specific genes in α- and **a**/α-cells is achieved by the cooperative binding of MCM1 and MATα2 at MCM1 consensus sites called P-sites, flanked by two α2-binding sites ((Keleher, C. A. *et al.*, 1989; Passmore, S. *et al.*, 1989) and reviewed in (Dolan, J. W. *et al.*, 1991)).

The 210 amino acid MATα2 protein contains a C-terminal homeodomain that is responsible for DNA-binding, preceded by a flexible linker sequence that is required for cooperative binding to MCM1. Transcriptional repression depends on the 100 amino acid N-terminal domain of MATα2, which homodimerises and recruits the Ssn6/Tup1 repressor complex ((Tan, S. *et al.*, 1998) and references therein).

Crystallographic analysis has elucidated the structural details of the interactions of the MATα2 linker and homeodomain fragment with the MCM1-DNA complex (Tan, S. *et al.*, 1998), which exhibit both striking similarities and contrasts to the TCF-SRF ternary complex. In this structure, the MATα2 homeodomain binds its DNA site on the 5' side of the MCM1 dimer (Figure 1.6A), with a second MATα2 homeodomain unexpectedly binding on the other side of MCM1 at the junction of two DNA molecules (not pictured). Both α2 homeodomains have the same fold and make similar DNA contacts and for clarification purposes are termed cis- (bound on the 5' side of MCM1) and trans- (bound at the DNA junction) monomers.

The cis- MATα2 monomer interacts with MCM1 via both its homeodomain and linker regions. Nevertheless as predicted by mutagenesis studies it is the linker-MCM1 contacts that define complex formation (Mead, J. *et al.*, 1996; Tan, S. *et al.*, 1998). The cis- α2 monomer uses an 8-residue sequence in its flexible linker to add a parallel β-strand to the central antiparallel β-sheet of MCM1 (Figure 1.6A). This sequence had been previously identified as necessary for MATα2-MCM1 mediated repression (Mead, J. *et al.*, 1996). This study determined that all residues within the β-strand are sensitive to alanine substitution, which abolishes MCM1 binding, in contrast to the interaction of TCF with SRF ((Hassler, M. *et al.*, 2001; Ling, Y. *et al.*, 1997; Mead, J. *et al.*, 1996); analysed in Section 1.2.5.6.3). The β-strand addition is achieved by the formation of a hydrophobic interface between the α2-residues and a hydrophobic groove on the surface of MCM1 (the MCM1 residues contacting MATα2 are summarised in Figure 1.7). Central to this interaction is the insertion of a phenylalanine residue in a hydrophobic pocket created by residues V69, V81, R87 and I90 in the β-sheet and αII-helices of MCM1 (Figure 1.6B; (Tan, S. *et al.*, 1998)). The importance of these contacts was confirmed by mutagenesis studies where removal of the α2 F116 sidechain or restricting the size of the pocket by substituting MCM1 residues with bulkier ones abolishes complex formation (Mead, J. *et al.*, 2002; Mead, J. *et al.*, 1996). The residues and interactions involving the hydrophobic pocket of MCM1 are conserved in the TCF-SRF complex (discussed in Section 1.2.5.6.3)

The significant degree of DNA bending induced by MCM1 was proposed to facilitate the interaction with the cis- MATα2 by bringing the two DNA binding sites closer (Tan, S. *et al.*, 1998). In vitro DNA bending experiments however do not support a dependence of MATα2 binding on DNA distortion (Lim, F. L. *et al.*, 2003).

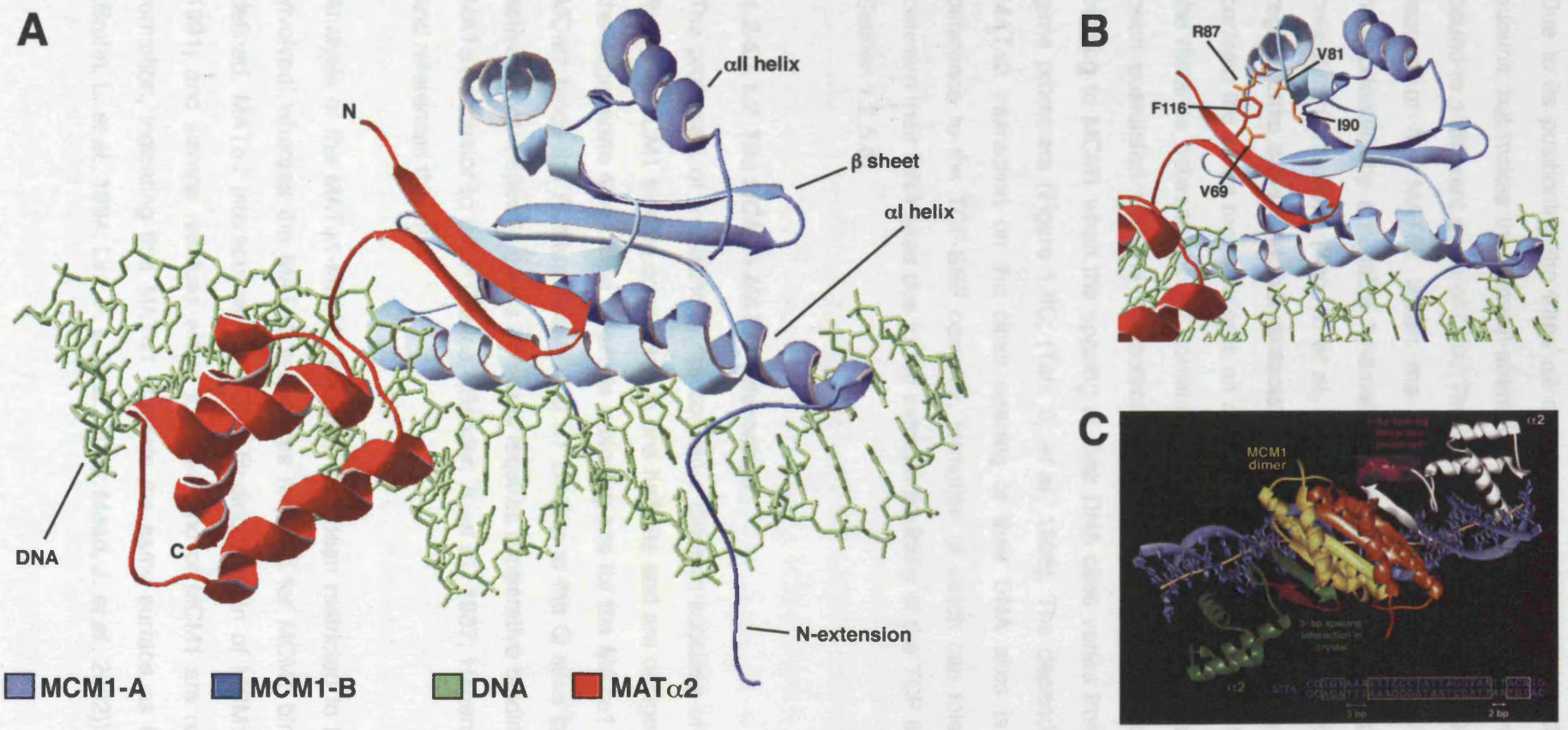


Figure 1.6. Structure of the MATα2 - MCM1 - DNA ternary complex. (A) Ribbon representation of the cis- MATα2 homeodomain and flexible linker sequence bound to DNA and MCM1 respectively. The two MCM1 subunits are shown in blue and the DNA in green. MATα2 is shown in red: the flexible linker forms a β-hairpin and adds a parallel β-strand to the β-sheet of MCM1. Image was produced from the 1mnm PDB file. (B) The interaction of the F116 MATα2 with the V69, V81, R87, I90 residues (sidechains shown in orange) that form the hydrophobic pocket of MCM1. (C) Model of the interaction of MATα2 and MCM1 on a composite DNA site where two α2-sites are spaced by 3 and 2bp from the MCM1 P-site . Structures are viewed from the top. The chameleon sequence of MATα2 is shown in red. Image taken from (Tan and Richmond, 1998).

Due to its positioning the trans α2 copy does not interact with the opposing MCM1 subunit, but makes the same β-strand contacts with a symmetry-related MCM1 dimer bound to a different piece of DNA. This fortuitous interaction uncovered an unexpected feature of the MATα2 linker: the MCM1 binding β-strand is connected to the homeodomain by a bistable "chameleon" region, which can adopt either β-strand or helical conformations (Tan, S. *et al.*, 1998). Thus in the cis-monomer this chameleon region forms a β-strand that contacts the MCM1-binding strand to form a β-hairpin. In contrast the same region adopts an α-helical conformation in the trans-copy to bridge the distance between the α2 homeodomain and the symmetry related MCM1. It has been postulated that the chameleon sequence functions to accommodate MATα2 binding to MCM1 when the spacing of their DNA sites varies from 2 to 3bp in target gene promoters (Figure 1.6C; (Tan, S. *et al.*, 1998). The dependence of the MCM1-MATα2 interaction on the close spacing of their DNA sites is another significant difference to the TCF-SRF complex, formation of which can tolerate long distances between their DNA sites due to the extreme flexibility of the TCF linker sequence (see Section 1.2.5.6).

1.2.4.1.1.2 The MCM1 - MATα1 Interaction

The promoters of α-specific genes contain a different subclass of MCM1 that deviate from the MCM1 binding consensus in one half-site and are designated P'. Adjacent to the degenerate side of the P' site is a binding site for the MATα1 activator named Q. MCM1 binds the P' sites weakly and α1 binding to the Q sites is undetectable, thus activation of α-specific genes in α-cells requires cooperative binding of MCM1 and the MATα1 activator to the P'Q sites ((Bender, A. *et al.*, 1987; Hagen, D. C. *et al.*, 1993) and references therein).

Analysis of the MATα1-MCM1 interaction has been restricted to the MCM1 residues involved, whereas the MATα1 sequences required for MCM1 binding remain poorly defined. MATα1 interacts with the DNA-binding domain of MCM1 (Primig, M. *et al.*, 1991) and several residues within the βII-strand of MCM1 are required for complex formation, indicating that MATα1 contacts the same surface as MATα2 (Figure 1.7; (Bruhn, L. *et al.*, 1994; Lim, F. L. *et al.*, 2003; Mead, J. *et al.*, 2002)).

In contrast to MATα2 however, MATα1 binding to MCM1 appears sensitive to mutations in the β-loop (at residues T66 and L68) and αI-helix (K40) that also affect the degree of DNA bending (Lim, F. L. *et al.*, 2003; Mead, J. *et al.*, 2002). Nevertheless some MCM1 derivatives that harbour mutations in the αI-helix and are defective for DNA bending are still able to interact with MATα1 (Carr, E. A. *et al.*, 2004). MATα1 binding to these mutant MCM1 proteins alleviates their bending defect and restores the overall bend angle in the ternary complex to that observed with the wild-type proteins. Furthermore proper DNA bending in the ternary complex correlates with transcriptional activity and it has thus been proposed that the interaction of MATα1 with MCM1 is required to impose a bend on the promoter necessary for transcriptional activation.

1.2.4.1.2 The pheromone response

As mentioned previously the α and **a** cell types secrete different pheromones or mating factors (α-factor is the pheromone produced by α-cells and **a**-factor the **a**-specific one). In addition each cell type expresses different cell-surface receptors, so that the α-cell recognises the **a**-secreted mating factor and vice versa. Exposure of one cell-type to secreted pheromones of the other induces the expression of genes required for mating, leading to cell cycle arrest and morphological changes that culminate in the fusion of the α- and **a**-cells and the formation of the **a**/α diploid cell (reviewed in (Dolan, J. W. *et al.*, 1991)).

Central to the pheromone response is the activation of a MAPK phosphorylation cascade that targets the STE12 transcription factor resulting in the expression of pheromone-responsive genes (reviewed in (Banuett, F., 1998)). In **a**-cells this depends on cooperative binding of STE12 and MCM1 at composite promoter elements consisting of an MCM1 P-site and a STE12 pheromone response element (PRE) (Errede, B. *et al.*, 1989). On the other hand the induction of pheromone-responsive genes in α-cells is thought to depend on the interaction of the MCM1-bound MATα1 activator with the STE12 transcription factor (Yuan, Y. O. *et al.*, 1993). It is noteworthy that MAPK signalling also regulates the activity of the mammalian SRF-TCF complex (see Section 1.2.5.6).

As with the other MCM1 cofactors, STE12 interacts with the DNA-binding domain of MCM1, although its MCM1 binding region is undefined (Mueller, C. G. *et al.*, 1991;

Primig, M. *et al.*, 1991). Mutagenesis studies revealed that the STE12 binding surface overlaps that of MATα1 and MATα2, although the contacts of the different cofactors are not completely conserved (Figure 1.7; (Bruhn, L. *et al.*, 1994; Mueller, C. G. *et al.*, 1991)).

1.2.4.1.3 Cell-cycle control

During the cell cycle MCM1 cooperates with the Fkh2 protein to control the expression of the CLB2 gene cluster, a group of about 30 genes required for the G2–M transition and then for mitotic progression and cytokinesis (Kumar, R. *et al.*, 2000; Pic, A. *et al.*, 2000; Zhu, G. *et al.*, 2000).

Fkh2 belongs to the Forkhead family of winged-helix proteins and contains a C-terminal DNA-binding domain (FKH domain) and an N-terminal Forkhead Associated domain (FHA) that is also found in the related cell cycle protein FKH1.

Fkh2 binds cooperatively with MCM1 to composite MCM1/FKh2 DNA consensus sites in the CLB2 gene promoters (Pic, A. *et al.*, 2000) and recruits the Ndd1 co-activator in a cell-cycle phosphorylation-dependent manner to activate transcription in the G2 and M stages (Darieva, Z. *et al.*, 2003; Koranda, M. *et al.*, 2000). This complex is also known as SFF (SWI Five Factor) since it was first identified as the activity that induces transcription of the SWI5 gene, one of the best studied CLB2 cluster members ((Pic, A. *et al.*, 2000) and references therein). It has been proposed that Fkh2 and MCM1 are constitutively bound to the CLB2 promoters but in the absence of Ndd1 act to repress transcription (Koranda, M. *et al.*, 2000; Pic, A. *et al.*, 2000).

Mutagenesis analysis of Fkh2 identified a stretch of 20 amino acids that are required for cooperative MCM1 binding (Boros, J. *et al.*, 2003). This sequence does not share obvious homology to the MATα2 β-strand, and is similar to the mammalian TCF B box (see later) in that it contains differently spaced hydrophobic residues required for complex formation (Boros, J. *et al.*, 2003; Hassler, M. *et al.*, 2001; Ling, Y. *et al.*, 1997; Mead, J. *et al.*, 1996; Tan, S. *et al.*, 1988). Mutation of a tyrosine residue within this stretch abolishes the interaction with MCM1, as does mutation of the MCM1 V69 residue that lines the hydrophobic pocket of MCM1, implicating the hydrophobic pocket of MCM1 in the Fkh2 interaction (Figure 1.7).

The spacing and orientation of the MCM1 and Fkh2 DNA sites are critical for the MCM1-Fkh2 interaction and thus the role of DNA bending in the formation of the MCM1-Fkh2 complex is thought to involve the correct juxtaposition of the two proteins on DNA (Boros, J. *et al.*, 2003; Lim, F. L. *et al.*, 2003). Mutations at MCM1 positions T66 and L68 that reduced the apparent bend angle correlated with reduced binding of Fkh2 (Lim, F. L. *et al.*, 2003). This correlation was not true for all mutations tested however, and the proximity of these residues to the proposed Fkh2 binding surface cannot exclude effects of certain substitutions on protein-protein interactions.

MCM1 also interacts with the Yox1 and Yhp1 homeodomain proteins to repress transcription of the early cell-cycle box (ECB) genes at the transition between the G1 and M phases of the cell cycle (Pramila, T. *et al.*, 2002). MCM1 is constitutively bound at the ECB promoters, and it is the periodic expression of the repressor that restricts gene expression. Although the two homeodomain proteins display different expression patterns they function redundantly and the molecular mechanism of their interaction with MCM1 has not been elucidated.

1.2.4.1.4 Arginine metabolism

MCM1 cooperates with three other proteins, the MADS box containing ARG80, and the unrelated ARG81 and ARG82 to control arginine metabolism by repressing biosynthetic genes and activating catabolic ones in the presence of arginine (reviewed in (Messenguy, F. *et al.*, 2003)). The interaction of these proteins is poorly understood. Although MCM1 and ARG80 can interact with both ARG82 and ARG81 the two latter proteins do not bind each other. ARG82 acts as a chaperone that stabilises the ARG80 and MCM1 MADS box proteins in the nucleus by binding their α1 helix (El Bakkoury, M. *et al.*, 2000). ARG81 is the sensor of arginine and forms a complex with the MCM1 and ARG80 MADS box proteins on the promoters or arginine regulated genes (Amar, N. *et al.*, 2000). Although MCM1 and ARG80 are both required for the control of arginine metabolism genes, the exact mechanism of their interaction is not known. Heterodimerisation of ARG80 and MCM1 has recently been proposed as a possible mechanism of arginine co-regulation (Messenguy, F. *et al.*, 2003), this however remains to be substantiated.

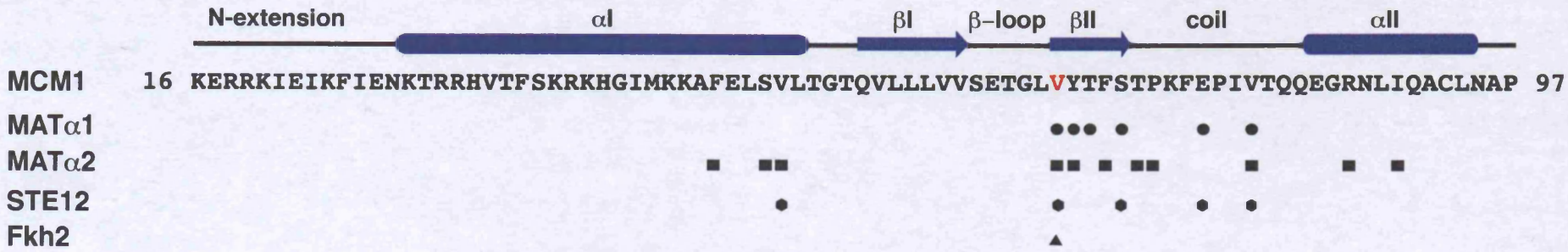


Figure 1.7. MCM1 cofactors contact overlapping surfaces on the MCM1 DNA-binding domain. The sequence and secondary structure of the MCM1 DNA-binding domain (amino acids 16-97) are shown, with residues involved in cofactor interactions indicated. Residue V69 which is located at the hydrophobic pocket of MCM1 and which correlates with SRF V194 is shown in red. Circles: residues affecting complex formation with MATα1 (Bruhn and Sprague, 1994); squares: residues contacting MATα2 in the crystal structure (Tan and Richmond, 1998); hexagons: residues affecting STE12 binding (Bruhn and Sprague, 1994); triangle: residue involved in Fkh2 binding (Boros et. al., 2003). Note that only those amino acids directly implicated in protein-protein contacts are shown, with residues affecting cofactor binding through effects on DNA bending omitted.

1.2.5 SRF

Serum Response Factor is a ubiquitously expressed 508-amino acid protein, which was initially identified as a critical component for the transcription of the c-fos proto-oncogene in response to serum stimulation (Gilman, M. Z. *et al.*, 1986; Norman, C. *et al.*, 1988; Prywes, R. *et al.*, 1986; Treisman, R., 1986). SRF is conserved throughout the animal kingdom and is found in nematodes, arthropods and chordates. It is constitutively expressed, although its expression can be upregulated in response to stimuli in various cell lines (Hela, NIH3T3 and 10T1/2 cells) (Hirschi, K. K. *et al.*, 2002; Norman, C. *et al.*, 1988; Sotiropoulos, A. *et al.*, 1999; Spencer, J. A. *et al.*, 1996).

SRF comprises a highly conserved N-terminal DNA-binding domain (Figure 1.8, see Section 1.2.3.1), and a C-terminal transactivation domain, which is not functionally conserved between vertebrates and invertebrates (Avila, S. *et al.*, 2002; Johansen, F. E. *et al.*, 1993).

1.2.5.1 SRF splice variants

Four SRF isoforms have been discovered to date: the originally identified full-length SRF protein coded by seven exons, SRF-M, which lacks exon 5, SRF-S which lacks exons 4 and 5 and SRF-I which lacks exons 3, 4 and 5 (Belaguli, N. S. *et al.*, 1999; Kemp, P. R. *et al.*, 2000). SRF-M is highly expressed in smooth and skeletal muscle cells, SRF-S expression is smooth muscle specific and SRF-I is restricted to embryonic tissues (Belaguli, N. S. *et al.*, 1999; Kemp, P. R. *et al.*, 2000). All three splice variants contain the complete DNA-binding domain and can therefore dimerise and bind DNA, but they lack regions of the transactivation domain. As a result they vary in their transactivation potential, with SRF-M activating SRF reporter genes weakly and SRF-I blocking transcription altogether (Belaguli, N. S. *et al.*, 1999; Kemp, P. R. *et al.*, 2000). This has led to the hypothesis that the biological role of the SRF splice variants is to control SRF target gene expression in different tissues by forming homo- or heterodimers of different activation potentials (Belaguli, N. S. *et al.*, 1999; Kemp, P. R. *et al.*, 2000; Miano, J. M., 2003). Nevertheless the mechanism regulating SRF alternative splicing and its functional significance in vivo remains unexplored.

1.2.5.2 SRF subcellular localisation

A nuclear localisation signal located N-terminal to its DNA-binding domain (residues 95-100, Figure 1.8) renders SRF constitutively nuclear in most cell lines (Gauthier-Rouviere, C. *et al.*, 1991). Recent studies have implicated altered subcellular localisation in the regulation of SRF gene expression. Partial redistribution of SRF to the cytoplasm has been reported in certain settings, such as upon differentiation of NIH3T3 cells to adipocytes due to loss of mitogenic responsiveness (Ding, W. *et al.*, 1999), and also after prolonged serum starvation of airway myocytes, due to downregulation of the RhoA pathway (Camoretti-Mercado, B. *et al.*, 2000; Liu, H. W. *et al.*, 2003). In both cases cytoplasmic redistribution of SRF is accompanied by downregulation of SRF dependent genes and has thus been proposed to be a control mechanism of SRF-dependent gene expression. The prevalence of the phenomenon and the mechanism by which SRF redistribution is controlled remains unclear. SRF was recently reported to be predominantly cytoplasmic in round embryonic mesenchymal cells, and to translocate to the nucleus following cell spreading and differentiation to smooth muscle (Beqaj, S. *et al.*, 2002), and it was proposed that the decrease of RhoA activity upon differentiation that drives SRF to the nucleus. In this case it was suggested that the low nuclear SRF levels in undifferentiated mesenchymal cells were sufficient for the mitogenic response of SRF, while the high nuclear levels observed after differentiation were required for smooth muscle specific gene expression. This hypothesis however is yet to be substantiated. Although regulation of SRF by alterations in subcellular distribution is an intriguing idea, further studies are required to explore the generality and impact of the observed phenomena described here.

1.2.5.3 SRF post-translational modifications

SRF is subject to various post-translational modifications including phosphorylation, glycosylation and sumoylation. Several sites of O-linked glycosylation have been mapped within the C-terminal activation domain, however the functional implications of these modifications remain unknown (Reason, A. J. *et al.*, 1992). SRF is modified by SUMO-1 at lysine 147 within its DNA-binding domain (Matsuzaki, K. *et al.*, 2003). This has been suggested to reduce its transcriptional activity in response to RhoA signalling, but this aspect of SRF regulation has not been further explored.

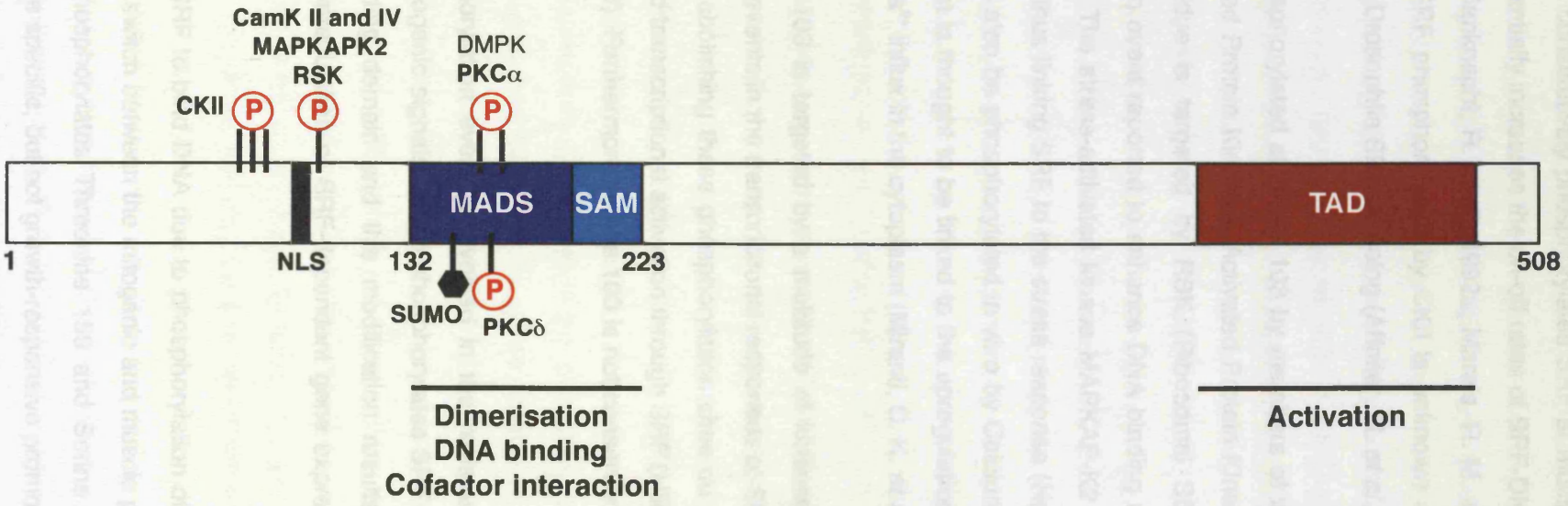


Figure 1.8. Schematic representation of Serum Response Factor. The MADS, SAM, TAD (Transcactivation domain) and NLS (nucler localisation signal) are shown as coloured boxes. Phosphorylation sites are indicated by vertical lines and the letter P in red and the kinases are shown: CK II phosphorylates residues 75-85; RSK, MAPKAPK2 and CamK II and IV phosphorylate Serine 103; PKCα and DMPK target Threonine 159 and Serine 162; PKCδ targets Threonine 160. The SUMO conjugation site at lysine 147 is indicated by a vertical line and black hexagon.

Casein Kinase II constitutively phosphorylates SRF at multiple sites between residues 77-85 and substantially increases the on-off rates of SRF-DNA binding without affecting binding affinity (Janknecht, R. *et al.*, 1992a; Marais, R. M. *et al.*, 1992). The functional significance of SRF phosphorylation by CKII is unknown and the CKII sites are not conserved in the Drosophila SRF homolog (Affolter, M. *et al.*, 1994)

SRF is also phosphorylated at serine 103 by members of the MAPKAP-Kinase family (Mitogen Activated Protein Kinase – Activated Protein Kinase). In response to growth factors this residue is targeted by RSK (Ribosomal S6 Kinase, also known as MAPKAP-K1), an event reported to enhance DNA binding rate and affinity (Rivera, V. M. *et al.*, 1993). The stress-activated kinase MAPKAP-K2 also targets serine 103 in vitro and in vivo thus linking SRF to the stress response (Heidenreich, O. *et al.*, 1999). This residue can also be phosphorylated in vitro by Calcium Calmodulin Kinase II and IV, and this effect is thought to be linked to the upregulation of SRF dependent genes in response to Ca^{2+} influx in the cytoplasm (Miranti, C. K. *et al.*, 1995).

Although serine 103 is targeted by a multitude of kinases the significance of these phosphorylation events in the transcriptional responses of SRF in vivo remains unclear. Point mutations abolishing these phosphorylation sites do not affect growth factor or calcium regulated transcriptional activation through SRF (Hill, C. S. *et al.*, 1994; Miranti, C. K. *et al.*, 1995). Furthermore Serine 103 is not conserved in Xenopus (Mohun, T. J. *et al.*, 1991).

An SRF phosphorylation event is involved in the resistance of cultured senescent fibroblasts to mitogenic signals: PKCδ phosphorylates SRF at T160 within the αI-helix of the DNA-binding domain and this modification results in decreased SRF-DNA binding and downregulation of SRF-dependent gene expression (Wheaton, K. *et al.*, 2004).

The inability of SRF to bind DNA due to phosphorylation of its αI-helix has also been implicated in the switch between the mitogenic and muscle programmes. In this case it is PKCα that phosphorylates Threonine 159 and Serine 162 abolishing SRF-DNA binding on muscle specific, but not growth-responsive promoters (Iyer, D. *et al.*, 2006).

The muscle/neuronal cell restricted Myotonic Dystrophy Protein Kinase (DMPK) has also been reported to phosphorylate Threonine 159 and Serine 162 within the αl coiled-coil of the DNA-Binding domain (Iyer, D. *et al.*, 2003). In this case however the phosphorylation event has been reported to activate SRF dependent transcription of reporter genes, a puzzling result considering that phosphorylation at these positions of the MADS box introduces negative charges close to DNA.

Thus the literature on SRF phosphorylation in response to signals is uncertain and sometimes contradicting. Despite the variety of phosphorylation events on SRF itself the best-characterised SRF-dependent transcriptional responses are elicited through interactions with accessory factors, which are themselves targets of signalling/phosphorylation events. The signal-induced responses of SRF will be discussed in more detail in the following sections.

1.2.5.4 SRF target genes and biological roles

The first SRF dependent gene discovered was the c-fos proto-oncogene. This was found to contain a dyad symmetric sequence in its promoter that was recognised by SRF and was required for c-fos expression in response to serum stimulation (Mohun, T. *et al.*, 1987; Treisman, R., 1985; Treisman, R., 1986). This element was therefore named the Serum Response Element (SRE). During the same period a 10bp promoter element was independently discovered in the promoters of α-cardiac actin genes of different species and also other muscle genes (Minty, A. *et al.*, 1986). Due to the conservation of the central A/T rich region and the flanking C-G base pairs this element was named the CArG box (CC A-rich GG;). It was soon discovered that the CArG box corresponded to the core SRF binding site in the centre of the SRE and CArG boxes were subsequently identified in the promoters of many muscle specific and growth-responsive genes (Miano, J. M., 2003; Treisman, R., 1995), including the c-fos and egr-1 immediate early genes and also the SRF gene itself (Spencer, J. A. *et al.*, 1996). Thus SRF is involved in the growth-response of proliferating cells, and also the regulation of the myogenic programme in post-replicative myocytes. In the latter case SRF is considered a central regulator of smooth muscle (Miano, J. M., 2003) and cardiac gene expression (Balza, R. O., Jr. *et al.*, 2006) and to a lesser extent skeletal muscle (see section 1.2.5.4.1; reviewed in (Pipes, G. C. *et al.*, 2006)).

Approximately 160 SRF target genes have been identified to date and over half of these have been validated (Philippar, U. *et al.*, 2004; Selvaraj, A. *et al.*, 2004; Sun, Q. *et al.*, 2006; Zhang, S. X. *et al.*, 2005). Although SRF target genes have traditionally been divided in growth-regulated and muscle specific, the latest studies show a clear bias for regulation of genes involved in the actin cytoskeleton (Philippar, U. *et al.*, 2004; Selvaraj, A. *et al.*, 2004; Sun, Q. *et al.*, 2006). Indeed *Srf*-null Embryonic Stem (ES) cells display severe cytoskeleton-related abnormalities ((Schratt, G. *et al.*, 2002); see following section). Therefore SRF emerges as a regulator of many cytoskeletal processes such as cytoskeletal architecture, cell migration and cell adhesion.

The fact that SRF controls such diverse subsets of genes raises questions as to the contribution of the CArG box sequence itself to the specificity of the response. Many natural CArG boxes deviate from the optimal $CC(A/T)_6GG$ consensus by one or more mismatched bases (Selvaraj, A. *et al.*, 2004; Sun, Q. *et al.*, 2006). These sites, are predicted to have lower affinities for SRF at least in vitro (Leung, S. *et al.*, 1989; Wynne, J. *et al.*, 1992) but are often evolutionarily conserved and are *bona fide* SRF binding sites (Miano, J. M., 2003; Selvaraj, A. *et al.*, 2004; Sun, Q. *et al.*, 2006). Many such elements are found in the promoters of muscle specific genes (Sun, Q. *et al.*, 2006), and in some cases substitution of the low affinity CArG with a high-affinity one results in more widespread expression (Hautmann, M. B. *et al.*, 1998). This lead to the hypothesis that low-affinity CArG boxes are a control mechanism of muscle-specific genes that contributes to the specificity of expression. Although intriguing this proposal remains largely unsubstantiated, especially since there is no strict correlation between perfect and imperfect CArG box sequences and muscle-specificity (Sun, Q. *et al.*, 2006) and also since many SRF muscle-specific genes are often also muscle cell-type restricted.

CArG boxes from growth-responsive and muscle-specific promoters are interchangeable (Taylor, M. *et al.*, 1989). Although the sequences flanking the CArG boxes of muscle specific genes have been previously implicated in muscle specificity in vivo, this proposal was based on substitution of the sequences flanking muscle-specific CArG boxes with those of the growth responsive c-*fos* CArG element (Chang, P. S. *et al.*, 2001; Santoro, I. M. *et al.*, 1991). These sequences also contain an Ets DNA binding site, recognised by members of the TCF family of SRF cofactors, which respond to growth signals (see Section 1.2.5.6). The widespread expression of the

muscle-specific genes observed upon these substitution experiments, can thus be attributed to sensitising their promoters to growth signals due to the interaction of the TCFs with SRF. As discussed in Section 1.2.2.1, in vitro binding site selection experiments identify further SRF-DNA binding determinants in the sequences bordering the CArG box (Pollock, R. *et al.*, 1990). The significance of these results and their possible implication in specificity of the SRF transcriptional response remains unknown.

1.2.5.4.1 SRF loss-of-function phenotypes

The importance of SRF in regulating cytoskeletal processes and myogenesis is apparent from SRF loss-of-function studies in mice and also *Drosophila melanogaster*.

Mutations of the Drosophila melanogaster SRF homologue (DSRF) (Affolter, M. *et al.*, 1994) reveal important roles of SRF in Drosophila development. The DSRF allele *pruned* is required for development of the tracheal system, the fruit-fly equivalent of respiratory and vascular systems of vertebrates (Guillemin, K. *et al.*, 1996). The *pruned* mutation is homozygous lethal with larvae exhibiting severely impaired terminal branching of the tracheal system due to inability of the cells to extend cytoplasmic outgrowths towards target tissues.

DSRF is also involved in wing formation where it is expressed in intervein tissue promoting its terminal differentiation, concurrently suppressing vein tissue formation (Montagne, J. *et al.*, 1996). Intervein differentiation involves special cytoskeletal arrangements, which facilitate the adherence of the two epithelial layers that will eventually form the wing. The loss-of-function SRF allele *blistered* prevents this process resulting in the blistered wing phenotype (Montagne, J. *et al.*, 1996). Both the pruned and blistered phenotypes involve alterations in cell morphology and the migratory and/or adherent properties of the cells, consistent with the role of SRF in the regulation of actin-mediated cytoskeletal processes in Drosophila development (see also Section 1.2.5.7.3.3).

Homozygous *Srf*-null mice display a severe gastrulation defect and die at embryonic day 7.5 (E7.5) (Arsenian, S. *et al.*, 1998). Srf (-/-) mice also fail to form mesodermal tissue in a non-cell autonomous manner, possibly due to the downregulation of a

signalling factor required to induce mesoderm (Arsenian, S. *et al.*, 1998; Weinhold, B. *et al.*, 2000). *Srf* (-/-) ES cells display defects in cell migration, adhesion and spreading (Schratt, G. *et al.*, 2002) highlighting the importance of SRF in these processes in vivo. Furthermore SRF has found to promote cell survival during ES cell differentiation by regulating the Bcl-2 antiapoptotic factor (Schratt, G. *et al.*, 2004), however the extent of the involvement of SRF in antiapoptotic gene programmes has not yet been explored. The importance of cytoskeletal organisation and cell migration in gastrulation raises the possibility that the phenotype observed in Srf-null mouse embryos reflects problems in such processes, although Bcl-2 regulation by SRF is likely to contribute in part.

Thus, although SRF was first identified due to its involvement in the growth-response, the SRF-regulated immediate-early gene programme appears dispensable for embryonic cellular proliferation since the SRF knock-out mice develop normally up to E6.5 and Srf-null ES cells do not display proliferative defects (Arsenian, S. *et al.*, 1998; Schratt, G. *et al.*, 2001).

The early embryonic lethality of SRF knockout mice precludes the use of this model to study the role of SRF in differentiation processes. Targeted inactivation of SRF has shown that it plays important roles in myogenesis and neurogenesis.

Inactivation in murine muscle tissues has been widely used to elucidate the central role of SRF in smooth, cardiac and also skeletal myogenesis. Conditional inactivation of SRF in the developing heart leads to severe cardiac defects and lethality between E10.5 and E13.5 due to downregulation of SRF-dependent muscle genes (Miano, J. M. *et al.*, 2004; Niu, Z. *et al.*, 2005; Parlakian, A. *et al.*, 2004). Deletion of SRF in embryonic smooth muscle cells severely reduces the smooth muscle cell population and causes severe cytoskeletal defects in the remaining ones (Miano, J. M. *et al.*, 2004). Selective inactivation of SRF in skeletal muscle resulted in hypoplastic skeletal muscles and perinatal lethality (Charvet, C. *et al.*, 2006; Li, S. *et al.*, 2005). Furthermore surviving mice displayed growth retardation due to skeletal muscle hypotrophy (Charvet, C. *et al.*, 2006).

Conditional inactivation of SRF in neuronal tissues also demonstrated the importance of SRF for normal neuronal development and function. SRF deletion in adult neuronal populations led to downregulation of immediate-early gene expression and showed that

SRF is essential for synaptic plasticity in response to neuronal activity-induced gene expression (Ramanan, N. *et al.*, 2005).

The significance of SRF in cytoskeletal processes and especially cell migration was highlighted by deletion of SRF in the developing mouse forebrain, which led to severely impaired neuronal migration in the rostral migratory stream and impairment of axon guidance, neuronal outgrowth and synapse formation in the hippocampus (Alberti, S. *et al.*, 2005; Knoll, B. *et al.*, 2006). These effects were linked to the downregulation of cytoskeletal SRF dependent genes.

Further insights into the biological functions of SRF have been gained by gene targeting studies of its cofactors and are described further below (Sections 1.2.5.7.3 and 1.2.5.6.6).

1.2.5.5 Specificity of SRF dependent gene expression

How is SRF activity differentially controlled according to cell-type or signalling pathway? SRF is central to many different gene expression programmes that are often opposing, the main example being cell proliferation and myogenesis since cells need to exit the cell cycle in order to differentiate to muscle cells. During the proliferative programme SRF is able to respond to different signals, transduced either through the MAP kinase phosphorylation cascades or through the effects of activated Rho GTPases on the actin cytoskeleton (analysed in sections 1.2.5.6.5 and 1.2.5.7.3). In contrast during myogenesis SRF elicits cell-type specific effects and controls smooth, cardiac and skeletal muscle genes.

Despite the various studies on regulation of SRF itself, involving cytoplasmic to nuclear translocation, alternative splicing, and posttranslational modifications (see sections 1.2.5.1 to 1.2.5.3) it is widely acknowledged that these effects do not constitute the crux of the differential context- and signal-dependent responses of SRF. SRF derives its versatility indirectly by physically associating with a range of signal-regulated or tissue-specific regulatory cofactors. Thus by engaging different partners SRF acts as a platform through which different incoming signals are interpreted.

This feature of SRF is best described in the fibroblast model of SRF-mediated serum response (Figure 1.9). Activation of the MAPK pathway in response to whole serum or mitogens such as TPA leads to ternary complex formation between SRF and members of the Ternary Complex Family of transcription factors and subsequent expression of immediate-early genes containing both SRF and TCF DNA binding sites such as c-fos and egr-1 (see Section1.2.5.6).

On the other hand, activation of the RhoA GTPase in response to whole serum or mitogens such as LPA induces actin polymerization and decreases the G-actin pool. These changes in actin dynamics culminate in the association of SRF with members of the Myocardin Related Family of SRF cofactors and activation of a different subset of genes that do not contain TCF binding sites (e.g. vinculin, actin and srf itself; see Section 1.2.5.7). Thus in this system the ability of SRF to form ternary complexes with either the TCFs or MRTFs dictates the MAPK- or Rho-responsiveness of SRF with the two pathways being mutually exclusive.

The following sections will describe the two major families of SRF cofactors, the Ternary Complex Factor family and the Myocardin Related Transcription Factor family, and briefly discuss other reported SRF partners.

1.2.5.6 The TCFs

The TCF family of SRF cofactors are one of the major links of SRF to the mitogenic response and the activation of immediate-early gene expression (Treisman, R., 1994). They interact with SRF on the promoters of their target genes and regulate their transcription dependent on activation of the Mitogen Activated Protein Kinase pathway.

TCFs form a distinct subgroup within the ETS-domain transcription factor superfamily (reviewed in (Sharrocks, A. D., 2001)). As denoted by their name members of this superfamily are characterised by the presence of an ETS DNA-binding domain (Donaldson, L. W. *et al.*, 1994). This common characteristic aside, the various ETS-domain proteins differ greatly in the ways they are regulated and the responses they elicit, allowing their subdivision into various subfamilies based on the conservation of the ETS domain and the presence of other functional regions.

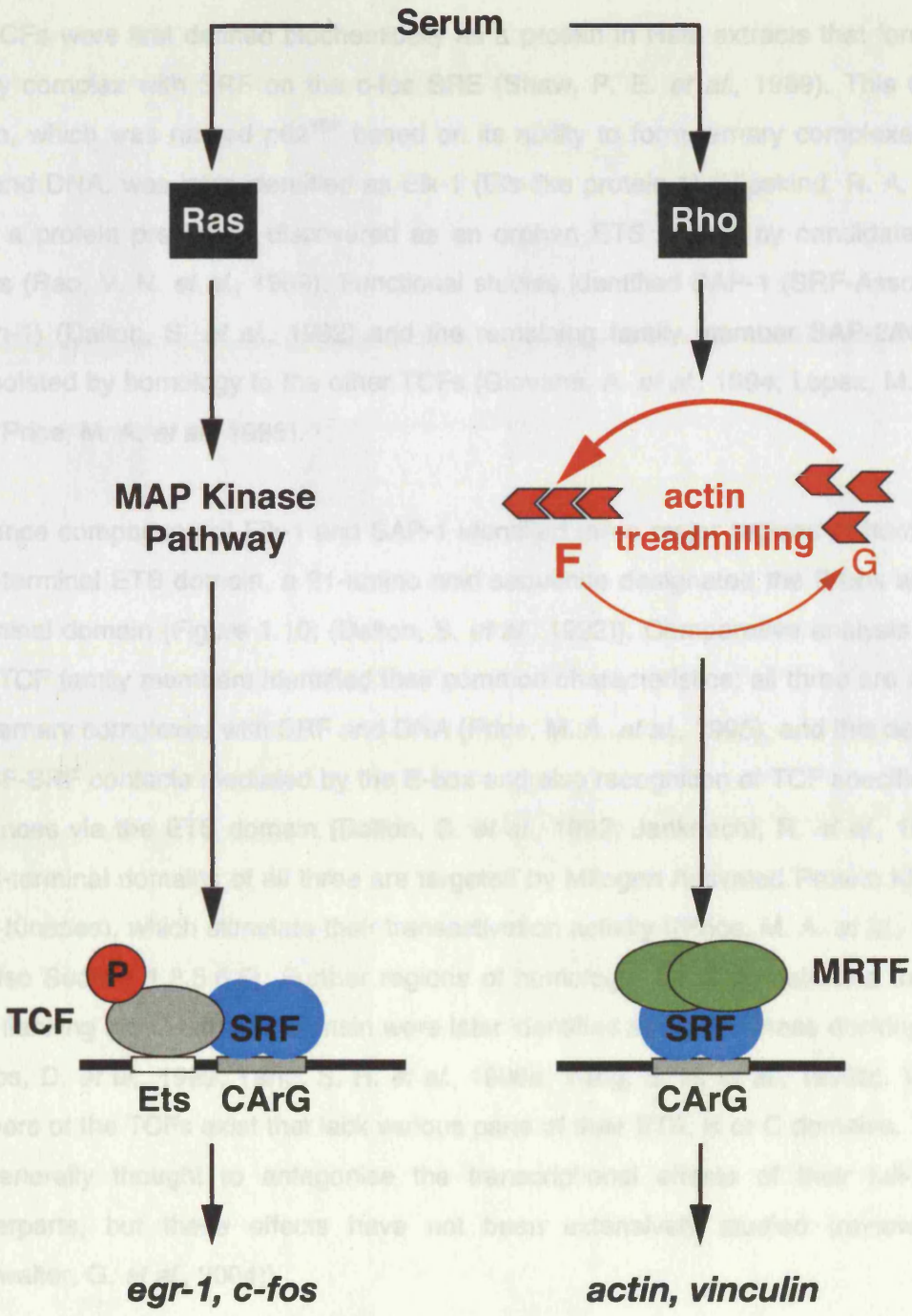


Figure 1.9. Signalling pathways converging at SRF. In the fibroblast model serum stimulation activates two distinct pathways that induce SRF-dependent gene expression by regulating different cofactors. Activation of MAP kinase signalling cascades results in the phoshorylation of members of the TCF family and transcription of genes containing TCF Ets binding sites such as egr-1 and c-fos. Activation of Rho GTPases induces changes in actin dynamics that activate members of the MRTF cofactor family, which bind SRF and activate transcription of a different subset of SRF dependent genes. GTPases are shown as black squares, actin is shown in red, SRF in blue, TCF in grey with a P in a red circle indicating phosphorylation. MAL is green and the DNA is shown as a black line with white and grey boxes indicating the Ets and CArG sites respectively.

The TCFs were first defined biochemically as a protein in Hela extracts that formed a ternary complex with SRF on the c-fos SRE (Shaw, P. E. *et al.*, 1989). This 62kDa protein, which was named $p62^{TCF}$ based on its ability to form ternary complexes with SRF and DNA, was later identified as Elk-1 (Ets-like protein-1) (Hipskind, R. A. *et al.*, 1991) a protein previously discovered as an orphan ETS protein by candidate gene studies (Rao, V. N. *et al.*, 1989). Functional studies identified SAP-1 (SRF-Associated protein-1) (Dalton, S. *et al.*, 1992) and the remaining family member SAP-2/Net/Erp was isolated by homology to the other TCFs (Giovane, A. *et al.*, 1994; Lopez, M. *et al.*, 1994; Price, M. A. *et al.*, 1995).

Sequence comparison of Elk-1 and SAP-1 identified three major regions of homology: the N-terminal ETS domain, a 21-amino acid sequence designated the B-box and the C-terminal domain (Figure 1.10; (Dalton, S. *et al.*, 1992)). Comparative analysis of the three TCF family members identified their common characteristics: all three are able to form ternary complexes with SRF and DNA (Price, M. A. *et al.*, 1995), and this depends on TCF-SRF contacts mediated by the B-box and also recognition of TCF specific DNA sequences via the ETS domain (Dalton, S. *et al.*, 1992; Janknecht, R. *et al.*, 1992b). The C-terminal domains of all three are targeted by Mitogen Activated Protein Kinases (MAP-Kinases), which stimulate their transactivation activity ((Price, M. A. *et al.*, 1995); see also Section 1.2.5.6.5). Further regions of homology, the D-domain and the FxF motif, flanking the C-terminal domain were later identified as MAP-Kinase docking sites (Jacobs, D. *et al.*, 1999; Yang, S. H. *et al.*, 1998a; Yang, S. H. *et al.*, 1998b). Variant members of the TCFs exist that lack various parts of their ETS, B or C domains. These are generally thought to antagonise the transcriptional effects of their full-length counterparts, but these effects have not been extensively studied (reviewed in (Buchwalter, G. *et al.*, 2004)).

The TCFs also contain repressive domains, the R domain of Elk-1 (Yang, S. H. *et al.*, 2002), the CID domain of Net (Criqui-Filipe, P. *et al.*, 1999) and the NID domains of SAP-1 and Net (Maira, S. M. *et al.*, 1996). These will be discussed in Section 1.2.5.6.5.

1.2.5.6.1 DNA binding by the TCFs

The ETS domain is a structural variant of the winged helix-turn-helix motif, the structure and DNA-binding properties of which have been well characterised for many ETS-

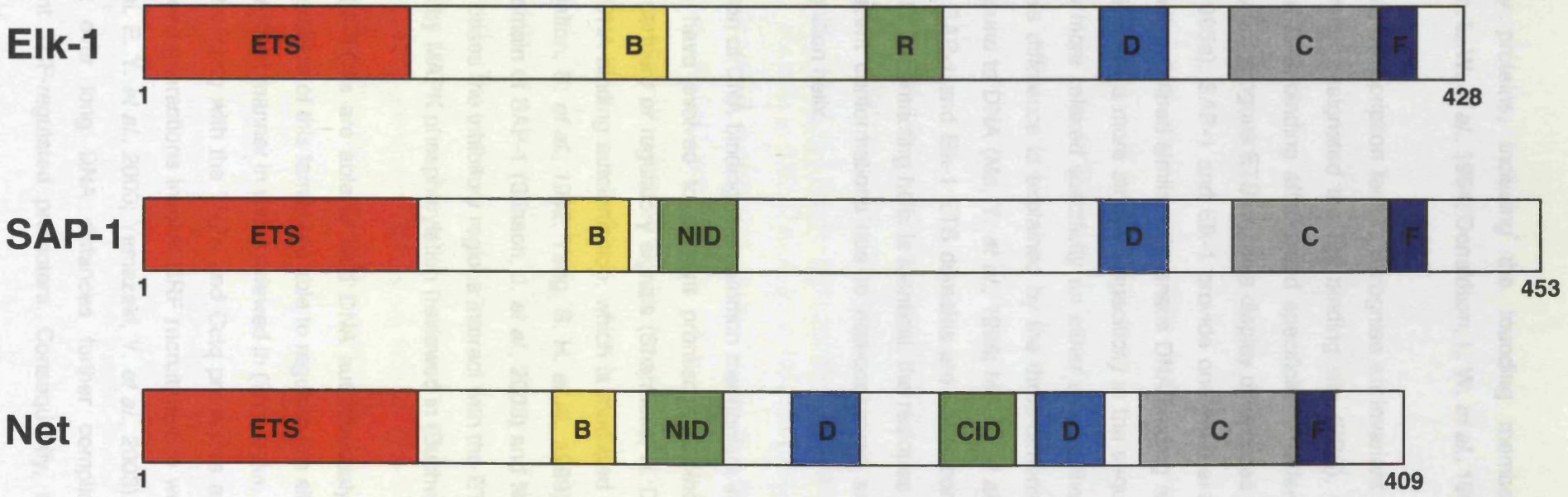


Figure 1.10. Schematic representation of the TCF family members. The ETS domain binds DNA, the B-box interacts with SRF, the C-domain has multiple MAPK phosphorylation sites and acts as a transactivation domain. The D and F (FxF) domains are MAPK docking sites. The R, NID and CID green boxes are transcriptional repression domains. (Adapted from Buchwalter et. al. 2004).

superfamily proteins, including the founding member Ets-1 (E-26 protein -1; (Donaldson, L. W. *et al.*, 1994; Donaldson, L. W. *et al.*, 1996)).

ETS domain transcription factors recognise an invariant GGA triad central to a more extended motif designated the Ets binding site (EBS), with surrounding nucleotides defining the DNA-binding affinity and specificity of different family members, so that even highly homologous ETS proteins display differences in their DNA contacts (Shore, P. *et al.*, 1995a). SAP-1 and Elk-1 provide one such example: although in vitro DNA binding studies defined similar consensus DNA binding sites (5'-AACCGGAAGT(A/G)-3'), Elk-1 displays a more stringent specificity in the sequences it recognizes whereas SAP-1 has more relaxed specificity on either side of the GGA triad (Shore, P. *et al.*, 1995a). This difference is explained by the three-dimensional structures of their ETS domains bound to DNA (Mo, Y. *et al.*, 1998; Mo, Y. *et al.*, 2000). Although the overall fold of the SAP-1 and Elk-1 ETS domains are ~80% conserved and the sequence of their main DNA-contacting helix is identical, the residues within this helix are forced to obtain different conformations due to non-conserved amino-acids C-terminal to the DNA-recognition helix.

Autoinhibition of DNA binding is a common mechanism in ETS domain proteins and is thought to have evolved to prevent promiscuous binding in the absence of co-regulatory partners or regulatory signals (Sharrocks, A. D., 2001). The TCFs are also subject to DNA-binding autoinhibition, which is mediated through the B and C regions of Elk-1 (Dalton, S. *et al.*, 1992; Yang, S. H. *et al.*, 1999), and the NID (Net Inhibitory Domain) domain of SAP-1 (Stinson, J. *et al.*, 2003) and Net (Maira, S. M. *et al.*, 1996). In all three cases the inhibitory regions interact with the ETS domain, and inhibition can be relieved by MAPK phosphorylation (reviewed in (Buchwalter, G. *et al.*, 2004)).

Although the TCFs are able to bind DNA autonomously in vitro, it remains unclear whether members of this family are able to regulate Ets site containing promoters in an SRF independent manner in vivo (reviewed in (Sharrocks, A. D., 2002)). Recent reports of Elk-1 interacting with the TNFα and Cctq promoters autonomously do not address whether these interactions involve SRF recruitment to weak or non-consensus CArG boxes (Tsai, E. Y. *et al.*, 2000; Yamazaki, Y. *et al.*, 2003). The ability of SRF and TCF to interact over long DNA distances further complicates the search for SRF independent TCF-regulated promoters. Consequently, the best characterised TCF

target genes to date contain CArG elements and belong to the immediate-early category (Gineitis, D. *et al.*, 2001).

1.2.5.6.2 Ternary complex formation between TCF and SRF

The Ets recognition sequences of various superfamily members display significant overlap. This raises the question of how specificity of DNA binding is achieved (reviewed in (Sharrocks, A. D., 2001)). It is possible that the interaction of TCFs with SRF serves as such a target gene specificity mechanism. Unlike many ETS proteins, which associate with their partners via their ETS domains, the TCF-SRF interaction is mediated by contacts external to the ETS domain. Thus ternary complex formation between TCF and SRF depends on both ETS-DNA and B box-SRF contacts (Dalton, S. *et al.*, 1992; Janknecht, R. *et al.*, 1992b).

The combination of protein-DNA and protein-protein interactions involved in the TCF-SRF ternary complex allows the recruitment of TCF to low affinity Ets sites as the one located on the 5' side of the CArG box on the c-fos promoter, which is not bound by TCFs in the absence of SRF (Dalton, S. *et al.*, 1992). The ability of SRF to recruit TCF to low affinity sites is illustrated by in vitro binding site selection experiments: when in the presence of SRF bound to a high affinity CArG box, Elk-1 selects the (G/A)(C/A)(C/A)GGA(A/T)(G/A)(T/C) consensus sequence, which represents a relaxed specificity version of that selected by autonomously binding Elk-1 (Treisman, R. *et al.*, 1992). Furthermore the sequences between the ETS and B regions of the TCFs act as a flexible linker between the DNA and SRF interaction domains. This linker allows great flexibility in the spacing and relative orientations of the Ets and CArG sites contributing to the ternary complex, as observed by the non-conservation of the Ets-CArG spacing in different promoters and confirmed by in vitro DNA-binding selection experiments ((Treisman, R. *et al.*, 1992) and references therein).

1.2.5.6.3 The TCF-SRF interaction

Extensive biochemical studies have elucidated the molecular mechanism of the TCF-SRF complex formation and the crystal structure of the SAP-1 ETS and B regions bound to SRF and DNA provides further insights in the mode of interaction of these proteins (Figure 1.11; (Hassler, M. *et al.*, 2001).

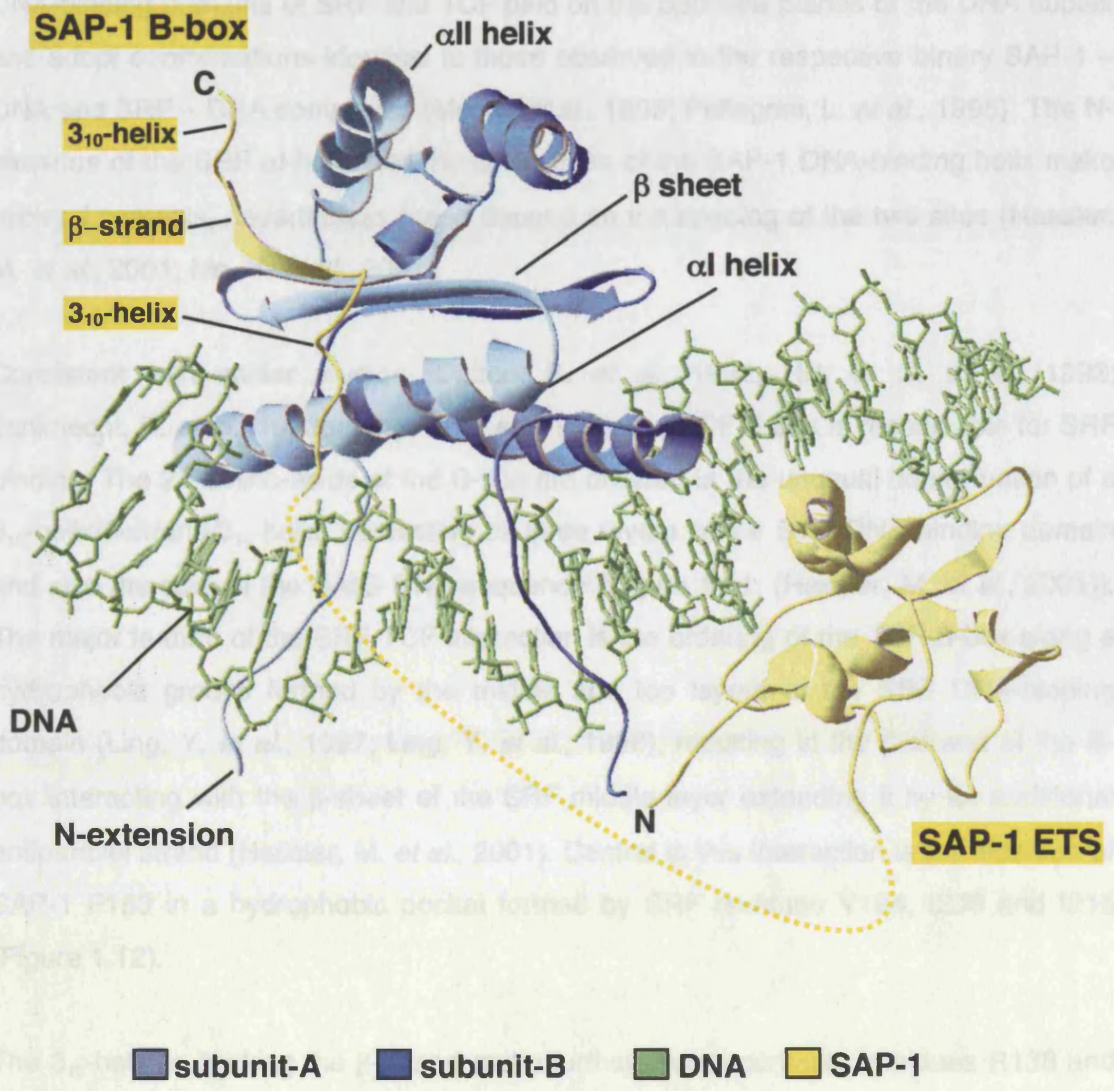


Figure 1.11. Structure of the SAP-1 - SRF - DNA ternary complex. Ribbon representation of the SAP-1 ETS and B domains bound to DNA and SRF repsectively. The two SRF subunits are shown in blue and the DNA in green. SAP-1 is shown in yellow: the B-box adds an antiparallel β-strand to the β-sheet of SRF, and is flanked by two 3_{10}-helices that contact the αII-helix and αI-helix and DNA. The dashed yellow line connecting the ETS and B domains represents the flexible linker between these two regions that is disordered in the crystal. The figure was made from the 1hbx PDB file (Hassler et. al. 2001).

In this structure the Ets binding site is located on the 5' side of the CArG box and the two high affinity sites are three nucleotides apart as seen in the c-fos promoter. The DNA-binding domains of SRF and TCF bind on the opposite planes of the DNA duplex and adopt conformations identical to those observed in the respective binary SAP-1 – DNA and SRF – DNA complexes (Mo, Y. *et al.*, 1998; Pellegrini, L. *et al.*, 1995). The N-terminus of the SRF αI-helix and the C-terminus of the SAP-1 DNA-binding helix make minimal contacts, nevertheless these depend on the spacing of the two sites (Hassler, M. *et al.*, 2001; Mo, Y. *et al.*, 2001).

Consistent with earlier studies (Dalton, S. *et al.*, 1992; Hill, C. S. *et al.*, 1993; Janknecht, R. *et al.*, 1992b; Ling, Y. *et al.*, 1997), the TCF B-box is responsible for SRF binding. The 21 amino-acids of the B-box are ordered in the unusual conformation of a 3_{10}-helix/β-strand/3_{10}-helix, contacting all three layers of the SRF DNA-binding domain and also the core of the CArG DNA sequence (Figure 1.11; (Hassler, M. *et al.*, 2001)). The major feature of the SRF-TCF interaction is the ordering of the TCF B-box along a hydrophobic groove formed by the middle and top layers of the SRF DNA-binding domain (Ling, Y. *et al.*, 1997; Ling, Y. *et al.*, 1998), resulting in the β-strand of the B-box interacting with the β-sheet of the SRF middle layer extending it by an additional antiparallel strand (Hassler, M. *et al.*, 2001). Central to this interaction is the insertion of SAP-1 F150 in a hydrophobic pocket formed by SRF residues V194, I206 and I215 (Figure 1.12).

The 3_{10}-helices flanking the β-strand make further crucial contacts: residues R138 and N139 of the N-terminal 3_{10}-helix make phosphate contacts with the core of the CArG box, while residue Y141 interacts with the SRF αI-helix. Moreover residues Y141 and I142 form a hydrophobic cluster with Y147 of the β-strand stabilising the overall conformation of the B box (Figure 1.12A). The C-terminal 3_{10}-helix contributes residues L152 and L155, which bind the SRF αII-helix.

The importance of many of these residues had been previously pinpointed by alanine scanning mutagenesis and in vitro binding studies of the highly homologous Elk-1 B box, which identified residues Y153, Y159, F162 and I164 (corresponding to SAP-1 Y141, Y147, F150 and L152, see Figure 1.12) as crucial for ternary complex formation (Ling, Y. *et al.*, 1997). Similar analyses had also mapped the amino acids forming the

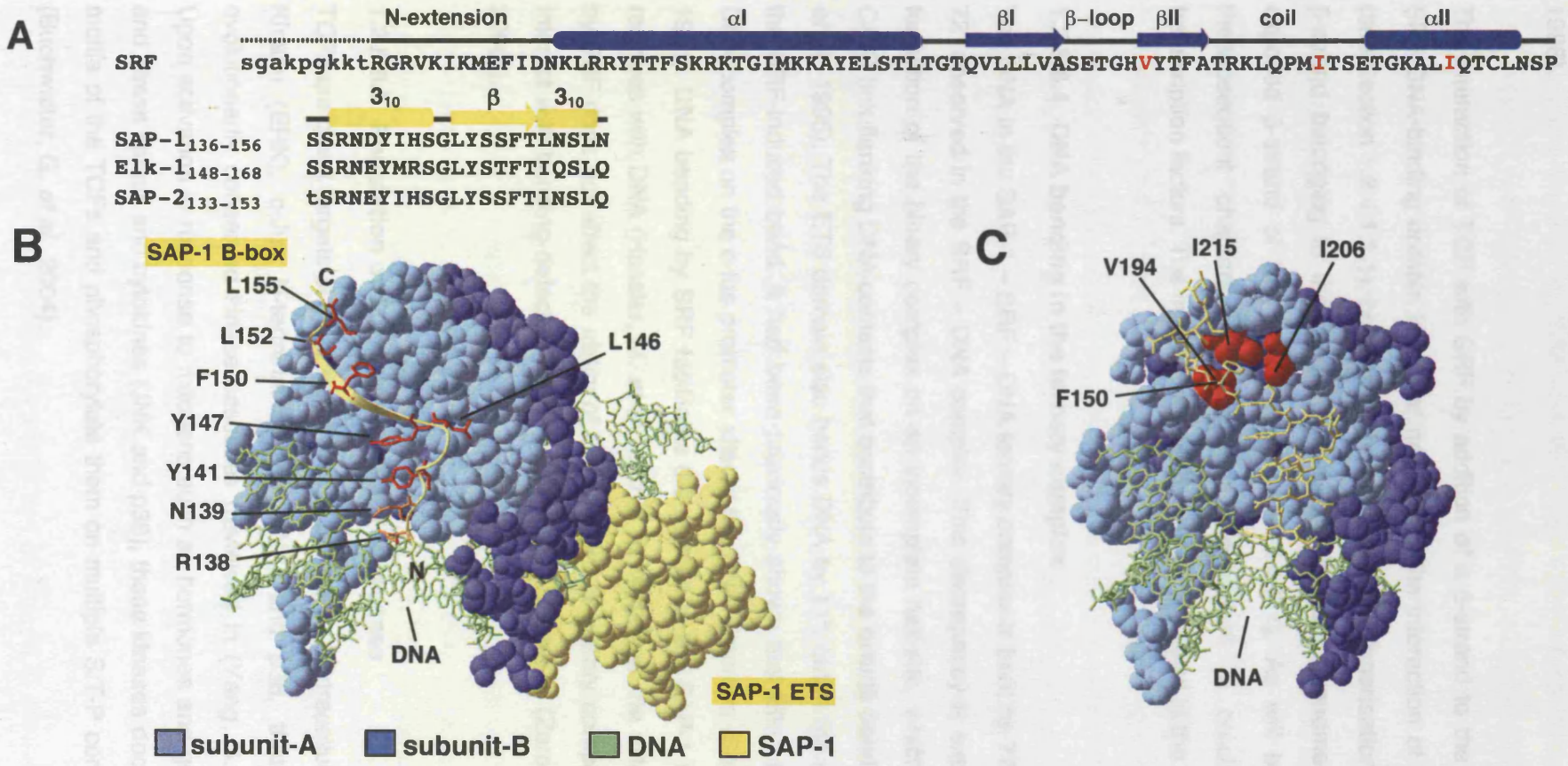


Figure 1.12. Interactions between the TCF B box and the SRF DNA-binding domain in the ternary complex. (A) Sequence and secondary structure elements of the SRF DNA-binding domain and the SAP-1 and Elk-1 B boxes. The SRF hydrophobic pocket residues are shown in red. (B)Van der Waals representation of the SRF DNA-binding domain and SAP-1 ETS domain bound to DNA. The SAP-1 B domain bound to SRF is showin as ribbon representation. The two SRF subunits are shown in blue and the DNA in green. SAP-1 is shown in yellow, with the side chains of the B-box contacting SRF in red, and those interacting with DNA in orange. (C) Van der Waals representation of the SRF DNA-binding domain complexed with the SAP-1 B box and DNA. Colour coding is as in B. The SRF residues lining the hydrophobic pocket are shown in red and the phenylalanine inserted into the pocket is labeled. Images were produced from the 1hbx PDB file (Hassler and Richmond, 2001).

hydrophobic interaction surface of SRF (see Chapter 4 for more details; (Ling, Y. *et al.*, 1998).

The interaction of TCF with SRF by addition of a β-strand to the middle layer of the SRF DNA-binding domain is similar not only to the interaction of MATα2 with MCM1 (see Section 1.2.4.1.1.1), but also to the mode of homodimerisation of MEF2B where a β-strand belonging to the top structural layer of each monomer interacts with the exposed β-strand of the other (see Section 1.2.3.1). As will become apparent in thesubsequent chapters this interaction mechanism is crucial for MADS box transcription factors. The implications of this will be analysed in the Discussion.

1.2.5.6.4 DNA bending in the ternary complex

The DNA in the SAP-1 – SRF – DNA ternary complex is bent by 77° compared with the 72° observed in the SRF – DNA complex. This discrepancy is explained partly by the formation of the binary complex on an incomplete half-site, which lacks some of the CArG box flanking DNA contacts that contribute to the overall bend angle (Pellegrini, L. *et al.*, 1995). The ETS domain also bends DNA by 11°, but in the opposite direction to the SRF-induced bend. It had been previously shown that Elk-1 binding to the SRF-DNA complex on the c-fos promoter altered the DNA distortion (Sharrocks, A. D. *et al.*, 1995). DNA bending by SRF facilitates the contact of the SAP-1 R138/N139 3_{10}-helix residues with DNA (Hassler, M. *et al.*, 2001). Nevertheless the extent of DNA bending by SRF does not affect the ability of Elk-1 to form a ternary complex, since it can still interact with bending-defective SRF mutants (see Chapter 4; (Zaromytidou, A. I. *et al.*, 2006)).

1.2.5.6.5 Regulation of TCF transcriptional responses

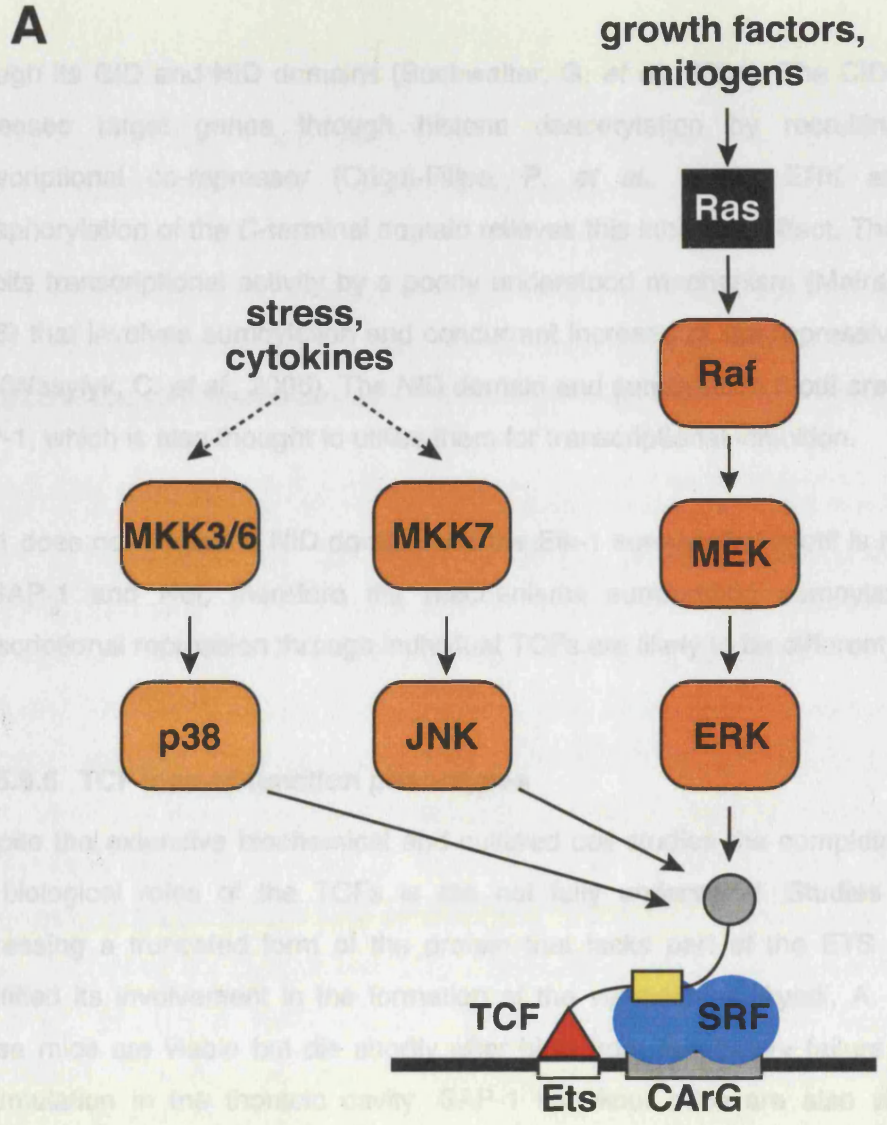
TCFs are direct targets of three major groups of MAPKs, Extracellular signal Regulated Kinase (ERK), c-Jun N-terminal Kinase (JNK) and p38, that are regulated by evolutionarily conserved kinase cascades (reviewed in (Yang, S. H. *et al.*, 2003b)). Upon activation in response to mitogens such as hormones and growth factors (ERK) and stress stimuli and cytokines (JNK and p38), these kinases dock on the D and FxF motifs of the TCFs and phosphorylate them on multiple S/T-P core consensus motifs (Buchwalter, G. *et al.*, 2004).

Elk-1 was one of the first transcription factors to be identified as a MAPK substrate and can be phosphorylated by ERK and JNK; SAP-1 can be phosphorylated by ERK and p38, whereas SAP-2 has been reported as an ERK, p38 and JNK target ((Sharrocks, A. D., 2002) and references therein; Figure 1.13). Thus despite the high sequence conservation of the TCF C-terminal domains the MAPK cascades target the TCFs differentially. Nevertheless the complexities of MAPK phosphorylation of TCFs in vivo remain unresolved. MAPK phosphorylation enhances DNA binding of the TCFs and stimulates their transcriptional activity by recruiting co-activators (the mechanisms of TCF activation are reviewed in (Buchwalter, G. *et al.*, 2004)).

In addition to their transactivation functions the TCFs are involved in transcriptional repression. Sumoylation of Elk-1 reduces its activity, as seen with many transcription factors (Verger, A. *et al.*, 2003). Sumoylation involves covalent linkage of a SUMO (Small Ubiquitin-like Modifier) molecule to lysines within sumo-acceptor motifs of target proteins. Although the exact mechanism through which this event attenuates transcription factor activity is still unclear, one of its functions appear to be the recruitment of transcriptional co-repressor complexes (Gill, G., 2005). The R domain of Elk-1 is a sumoylation motif and when SUMO-conjugated represses its target genes by recruiting the HDAC2 histone deacetylase complex (Yang, S. H. *et al.*, 2003a; Yang, S. H. *et al.*, 2004)

Studies of the role of phosphorylation and sumoylation in the regulation of Elk-1 activity have led to the following model of transcriptional regulation through Elk-1. In the absence of activating signals Elk-1 is SUMO-conjugated and bound to the promoters of its target genes in ternary complexes with SRF (Yang, S. H. *et al.*, 2003a). Sumoylation allows Elk-1 to recruit HDAC complexes that deacetylate the chromatin and inhibit transcription of the Elk-1 target genes (Yang, S. H. *et al.*, 2004). Activation of the MAPK pathway and subsequent phosphorylation of the C-terminal domain of Elk-1 leads to its rapid desumoylation (Yang, S. H. *et al.*, 2003a) and transcriptional activation by the recruitment of co-activators, such as the Sur2 subunit of the Mediator complex (Stevens, J. L. *et al.*, 2002). Finally transcription is terminated by the recruitment of histone-deacetylase complexes such as the mSin3A-HDAC complex to the N-terminal region of Elk-1 (Yang, S. H. *et al.*, 2001).

Net displays limited transactivation ability and is thought to act mainly as a repressor



B

ERK	→	Elk-1, SAP-1, Net
p38	→	SAP-1, Net
JNK	→	Elk-1, Net

Figure 1.13. Regulation of TCFs by MAP kinase signalling. (A) MAP kinase signalling cascades to TCF. The serum activated ERK pathway is shown in full. MAP kinases are shown as orange boxes and the Ras GTPase is shown in black. Dashed arrows indicate missing intervening pathaway components. The TCF-SRF-DNA ternary complex is shown: the TCF ETS domain (red triangle) is bound to the Ets site (white box) and the B box (yellow) binds SRF (blue) on the CArG site (grey box). Activation of the MAPK kinase cascades by signal culminates in the phosphorylation of the TCF C-terminal domain (grey circle) and subsequent transcriptional activation. MKK: MAP kinase kinase. [Adapted from (Posern and Treisman, 2006) and (Sharrocks, 2002)].(B) The TCF family members targeted by the three major MAP kinases.

through its CID and NID domains (Buchwalter, G. *et al.*, 2004). The CID domain also represses target genes through histone deacetylation by recruiting the CtBP transcriptional co-repressor (Criqui-Filipe, P. *et al.*, 1999). ERK activation and phosphorylation of the C-terminal domain relieves this inhibitory effect. The NID domain inhibits transcriptional activity by a poorly understood mechanism (Maira, S. M. *et al.*, 1996) that involves sumoylation and concurrent increase of the repressive potential of Net (Wasylyk, C. *et al.*, 2005). The NID domain and sumoylation motif are conserved in SAP-1, which is also thought to utilise them for transcriptional inhibition.

Elk-1 does not contain a NID domain and the Elk-1 sumoylation motif is not conserved in SAP-1 and Net, therefore the mechanisms surrounding sumoylation-mediated transcriptional repression through individual TCFs are likely to be different.

1.2.5.6.6 TCF loss-of-function phenotypes

Despite the extensive biochemical and cultured cell studies the complete spectrum of the biological roles of the TCFs is still not fully understood. Studies of Net mice expressing a truncated form of the protein that lacks part of the ETS domain have identified its involvement in the formation of the vasculature (Ayadi, A. *et al.*, 2001). These mice are viable but die shortly after birth from respiratory failure due to chyle accumulation in the thoracic cavity. SAP-1 knockout mice are also viable but are defective in single-positive T-cell selection and develop a form of Castleman's disease (Costello, P. S. *et al.*, 2004). Deletion of Elk-1 does not affect mouse development and viability and has no negative effect in immediate early gene expression apart from a downregulation of c-fos in certain parts of the brain (Cesari, F. *et al.*, 2004). The lack of severe phenotypes upon inactivation of individual TCFs in combination with their overlapping expression patterns (Price, M. A. *et al.*, 1995) and the overall structural similarities suggest a certain degree of functional redundancy. Nevertheless in light of their differences in DNA-sequence recognition, responsiveness to different branches of the MAPK pathway and differing abilities to act as transcriptional activators or repressors the lack of severe loss-of-function phenotypes is surprising. Double and triple knockout studies will help elucidate the functional significance of these proteins in vivo.

1.2.5.7 The MRTFs

The Myocardin-related transcription factor (MRTF) family of SRF cofactors consists of three members in mammals: Myocardin (Wang, D. *et al.*, 2001), MAL/MKL1/MRTF-A (Ma, Z. *et al.*, 2001; Mercher, T. *et al.*, 2001) and MAL16/MKL2/MRTF-B) (Selvaraj, A. *et al.*, 2003b; Wang, D. Z. *et al.*, 2002). The MRTFs are conserved through evolution in the animal kingdom and are present from arthropods to vertebrates, although no clear homolog is found in nematodes.

Myocardin is the founding member of the family and was discovered in a bioinformatics screen for murine cardiac specific genes (Wang, D. *et al.*, 2001). MAL had been identified previously as the product of a chromosomal translocation between human chromosomes 1 and 22 t(1;22)(p13;q13), that leads to acute megakaryoblastic leukaemia (AML:M7), a rare and aggressive form of childhood leukaemia, (Ma, Z. *et al.*, 2001; Mercher, T. *et al.*, 2001). This translocation product results in the in-frame fusion of almost the complete MAL reading frame C-terminal to a gene product from chromosome one that contains an RNA-binding motif. Since neither gene had been assigned a function at the time of discovery, the t(1:22) translocation product was named OTT-MAL (for One-Twenty-Two – Megakaryocytic Acute Leukaemia; (Mercher, T. *et al.*, 2001)) or RBM15-MKL-1 (RNA-Binding Motif 15 – Megakaryoblastic Leukemia-1; (Ma, Z. *et al.*, 2001)). A second MAL isoform termed MAL(BSAC) and containing distinct N-terminal sequences, was later discovered in a genetic screen for antiapoptotic factors (Sasazuki, T. *et al.*, 2002). The third family member, MAL16/MKL2/MRTF-B (Selvaraj, A. *et al.*, 2003b; Wang, D. Z. *et al.*, 2002), had been originally identified in a lacZ gene trap screen, due to the embryonic lethal effect of the gene trap insertion (Skarnes, W. C. *et al.*, 1992). Nevertheless it wasn't until the characterisation of Myocardin as an SRF cofactor (Wang, D. *et al.*, 2001), and the identification of MAL as such a cofactor (Miralles, F. *et al.*, 2003; Sasazuki, T. *et al.*, 2002; Wang, D. Z. *et al.*, 2002) that a function was assigned to MAL16. In contrast to Myocardin whose expression is restricted to the cardiac and smooth muscle lineages (Chen, J. *et al.*, 2002; Du, K. L. *et al.*, 2003; Wang, D. *et al.*, 2001), MAL and MAL16 are more widely expressed (Wang, D. Z. *et al.*, 2002). Studies in our lab demonstrated that MAL and MAL16 have the properties of the elusive cofactor mediating Rho signalling ((Miralles, F. *et al.*, 2003; Zaromytidou, A. I. *et al.*, 2006) and Cristina Perez-Sanchez, personal communication; analysed in detail in section 1.2.5.7.3.1).

1.2.5.7.1 Functional domains of the MRTFs

The three mammalian proteins share a degree of similarity in multiple regions (Figure 1.14; (Miralles, F. *et al.*, 2003; Wang, D. *et al.*, 2001; Wang, D. Z. *et al.*, 2002)).

1.2.5.7.1.1 The RPEL motifs

Their highly conserved N-termini contain three RPEL motifs, which have been named for their almost invariant central RPxxxEL sequence (Miralles, F. *et al.*, 2003). These repeats had not been assigned a function prior to their functional analysis in the MRTFs, which established that they represent a novel actin binding structure (Miralles, F. *et al.*, 2003; Posern, G. *et al.*, 2004). The ability of the MRTF RPEL domain to interact with actin confers Rho-inducibility to the MRTFs, since disruption of the MAL-actin complex in response to Rho-signalling activates MAL-SRF dependent gene expression (Miralles, F. *et al.*, 2003). Functional analysis of the MAL RPEL domain indicates that this region is necessary and sufficient for regulation of MAL nuclear accumulation in response to Rho activation (Miralles, F. *et al.*, 2003; Vartiainen, M. K. *et al.*, 2006) and that each RPEL motif binds three actin molecules ((Posern, G. *et al.*, 2004) and Sebastian Guettler, personal communication). MAL16 is also able to bind actin through the RPELs, whereas Myocardin whose RPEL sequences are more divergent has negligible affinity for actin and is refractory to Rho signalling (Francesc Miralles and Sebastian Guettler, personal communication). The significance of the interaction of MRTFs with actin in their regulation and the transcription of subsets of SRF target genes will be discussed in Section 1.2.5.7.3.

1.2.5.7.1.2 The B1 and Q regions

MRTFs share high homology through a region rich in basic residues designated B1 and a glutamine-rich region designated Q (Miralles, F. *et al.*, 2003; Wang, D. *et al.*, 2001; Wang, D. Z. *et al.*, 2002). As will be shown in this thesis, the B1 region mediates MRTF binding to SRF (Miralles, F. *et al.*, 2003; Wang, D. *et al.*, 2001). The Q region is not required for, but enhances MAL binding to SRF (Miralles, F. *et al.*, 2003). Although the Q box has been reported to be necessary for the interaction of Myocardin with SRF (Wang, D. *et al.*, 2001), and was proposed to bind SRF in a manner similar to the TCF B box (Wang, Z. *et al.*, 2004), it was later shown that this role is fulfilled by the B1 box in both MAL and Myocardin (Zaromytidou, A. I. *et al.*, 2006). The results presented in this thesis analyse the role of the B1 box and support a model in which the MRTFs use

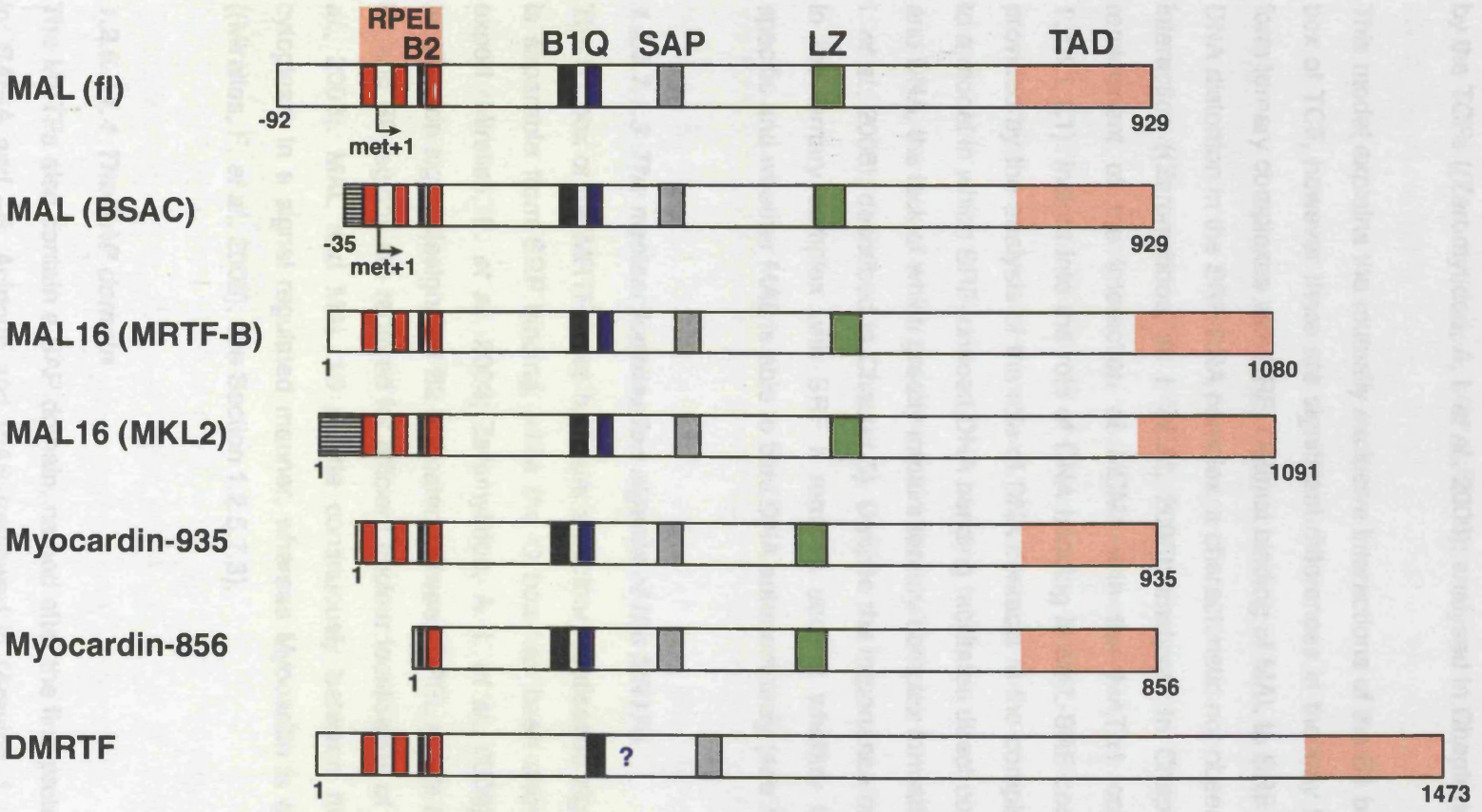


Figure 1.14. Schematic representation of the MRTF family members. The domain structure of the murine MAL, MAL16 and Myocardin isoforms are shown with the Drosophila DMRTF included. MAL(fl) is MKL1/MRTF-A and is translated from a leucine 92 residues prior to the first in-frame methionine. The divergent MAL(BSAC) sequences N-terminal to the RREL domain are shown as a striped box. The two MAL16 isoforms are identical apart from their different N-termini prior to the RPEL domain (shown as striped box). Myocardin-935 is identical to Myocardin A (not pictured) which contains an extra exon C-terminal to the LZ domain. Note that the original Myocardin cDNA was a predicted open reading frame of 807 amino acids that lacked the RPEL domain (Wang et. al., 2001).The question mark in DMRTF indicates the position of residues resembling the hydrophobic core of the mammalian Q box.

a conserved predominantly hydrophobic core sequence within their B1 box to contact the hydrophobic groove and pocket of SRF via a mechanism analogous to that utilised by the TCFs ((Zaromytidou, A. I. *et al.*, 2006); analysed in Chapters 3 and 4).

This model explains the mutually exclusive interactions of the B1 box of MAL and the B box of TCF, however there are significant differences in the way that the two proteins form ternary complexes with SRF. Optimal binding of MAL to SRF requires appropriate DNA distortion in the SRF-DNA complex, a characteristic not observed in the TCF-SRF interaction ((Zaromytidou, A. I. *et al.*, 2006); analysed in Chapter 4), but which is reminiscent of the interaction of MCM1 with the MATα1 cofactor (see Section 1.2.4.1.1.1). Insight into the role of DNA bending in MAL-SRF complex formation was provided by the analysis of the role of DNA contacts in the complex. This analysis led to a model in which SRF-induced DNA bending facilitates direct contacts between MAL and DNA, the lack of which greatly impairs ternary complex formation ((Zaromytidou, A. I. *et al.*, 2006); described in Chapter 5). Despite the importance of MAL-DNA contacts in the ternary complex with SRF it remains unclear whether these are sequence specific and whether MAL is able to bind DNA autonomously (see Appendix).

1.2.5.7.1.3 The nuclear localisation signals of the MRTFs

The B1 box of the MRTFs also harbours a nuclear localisation signal, but this function is separable from SRF binding, while the Q box has been shown to affect nuclear export (Miralles, F. *et al.*, 2003; Zaromytidou, A. I. *et al.*, 2006). A second nuclear localisation signal designated B2 is located between RPEL motifs 2 and 3 and both the B1 and B2 regions are required for efficient nuclear localisation of MAL (Miralles, F. *et al.*, 2003). MAL and MAL 16 shuttle continuously between the nucleus and the cytoplasm in a signal regulated manner, whereas Myocardin is constitutively nuclear ((Miralles, F. *et al.*, 2003); see Section 1.2.5.7.3).

1.2.5.7.1.4 The SAP domain

The MRTFs also contain a SAP domain, named after the first proteins it was identified in: SAF-A and –B, Acinous and PIAS (reviewed in (Aravind, L. *et al.*, 2000)). This domain is found in a variety of proteins involved in diverse processes like recognition of scaffold- or matrix attachment regions on chromatin, DNA repair, RNA processing, chromatin remodelling and apoptotic degradation of chromatin (Aravind, L. *et al.*,

2000). DNA binding appears to be the common characteristic between SAP containing proteins and in some cases, such as in SAF-A/B, Ku70 and PIAS1 this domain has been found to mediate recognition of A/T rich DNA regions (Kipp, M. *et al.*, 2000; Okubo, S. *et al.*, 2004).

The role of the SAP domain in MRTFs and its ability to bind DNA remain unclear. Deletion or disruption of this region does not affect ternary complex formation with SRF (Miralles, F. *et al.*, 2003; Wang, D. *et al.*, 2001). Disrupting the SAP domain of Myocardin affects target gene activation in a promoter dependent manner (Wang, D. *et al.*, 2001) and deleting it reduces the myogenic potential of the protein in cultured cells (Wang, Z. *et al.*, 2003). This functional requirement of the Myocardin SAP domain suggests a context dependent role either due to specific protein-protein or protein-DNA interactions. Moreover the involvement of SAP domain proteins in chromatin organisation, raises the possibility that the effects of this domain in the MRTFs are obscured by the experimental use of naked DNA as opposed to chromatin.

1.2.5.7.1.5 The leucine-zipper motif and transactivation domains

MRTFs contain a leucine-zipper motif that is responsible for homo- and heterodimerisation, with MAL existing as a stable dimer and Myocardin as a monomer in solution ((Miralles, F. *et al.*, 2003; Selvaraj, A. *et al.*, 2003b; Wang, D. *et al.*, 2001; Wang, Z. *et al.*, 2003); see Chapter 2). All three family members also contain C-terminal transactivation domains, which act as autonomous transcriptional units and which do not appear to be controlled by signalling (Miralles, F. *et al.*, 2003; Selvaraj, A. *et al.*, 2003a; Wang, D. *et al.*, 2001; Wang, D. Z. *et al.*, 2002).

1.2.5.7.2 MRTF isoforms

The original Myocardin cDNA is a 935-amino acid isoform that contains the complete RPEL domain (Wang, D. *et al.*, 2001; Wang, D. Z. *et al.*, 2002). It later became apparent that the Myocardin open reading frame contained four potential ATG start sites in total, the first of which gives rise to Myocardin-935 and is enriched in cardiac cells (Creemers, E. E. *et al.*, 2006). Alternative splicing introduces a premature stop codon at the N-terminus of Myocardin and translation starts at the third ATG giving rise to a shorter 856 residue isoform that contains only RPEL 3 and is abundant in smooth muscle cells (Figure 1.14; (Creemers, E. E. *et al.*, 2006). The Myocardin-935 isoform

interacts with both SRF and MEF2 proteins. The interaction with MEF2 is mediated via a short motif located within the N-terminal sequences and activates transcription of MEF2-dependent genes (Creemers, E. E. *et al.*, 2006). MEF2 binding is not observed with the Myocardin-856 isoform, which lacks the N-terminal motif. A third isoform is the 983 amino acid Myocardin A, which contains an extra exon in its C-terminal part, is also cardiac-enriched and is functionally identical to the original Myocardin-935 (Ueyama, T. *et al.*, 2003)*.

MAL is also present as two isoforms with different N-termini: MAL(fl) which is translated from a leucine 92 amino acids upstream of the first in-frame methionine (Miralles, F. *et al.*, 2003), and MAL(BSAC) which contains divergent sequences N-terminal to its first RPEL motif (Figure 1.14; (Sasazuki, T. *et al.*, 2002). Both isoforms are widely expressed (Sasazuki, T. *et al.*, 2002; Wang, D. Z. *et al.*, 2002) and the significance of their diverse N-termini is unclear. Short splice variants of MAL lacking the C-terminal transcactivation domains have also been detected in various tissues, but these have not been characterised ((Wang, D. Z. *et al.*, 2002) and Francesc Miralles, personal communication).

Alternative MAL16 isoforms analogous to MAL(fl) and MAL(BSAC) also exist and have been named MRTF-B (Wang, D. Z. *et al.*, 2002) and MKL2 ((Selvaraj, A. *et al.*, 2003b); Figure 1.14). Both isoforms contain all three RPEL domains however their N-termini prior to the first RPEL are distinct. Although both are widely expressed the MKL2 form is more abundant in skeletal muscle where the MRTF-B type is barely detectable (Selvaraj, A. *et al.*, 2003b; Wang, D. Z. *et al.*, 2002). Despite the differences in their expression pattern, the roles of their different N-termini remains unexplored.

1.2.5.7.3 Regulation of MRTF transcriptional activity and signal convergence at SRF

Myocardin was the first MRTF to be identified as an SRF cofactor (Wang, D. *et al.*, 2001). Although Myocardin can activate transcription of SRF target genes in a CArG-dependent manner, its expression is restricted to muscle lineages and its

* The Myocardin isoform used in this thesis is a 977 amino acid version of Myocardin, A which lacks six glutamine residues from the polyglutamine stretch of the Q box.

transcriptional activity appears to be constitutive (Kuwahara, K. *et al.*, 2005; Wang, D. *et al.*, 2001; Wang, Z. *et al.*, 2003). In contrast the discovery of the more widely expressed MAL and MAL16 created a breakthrough in the field of SRF dependent signal regulated transcription.

As mentioned in Section 1.2.5.5 SRF acts as a platform through which different signals can be transduced. Signal convergence at SRF has been extensively studied during the serum response of proliferating fibroblasts ((Miralles, F. *et al.*, 2003; Treisman, R., 1994) and references therein). In the fibroblast model serum stimulation induces the expression of many SRF-dependent genes via pathways initiated by different mitogens (Gineitis, D. *et al.*, 2001; Hill, C. S. *et al.*, 1995a). One such pathway is the ERK MAPK signalling cascade that results in the phosphorylation of TCF within the ternary complex with SRF and DNA (Section1.2.5.6).

Early in the dissection of the signalling events leading to activation of transcription by SRF it was discovered that the MAPK cascades were not the only pathway inducing SRF dependent gene expression (Graham, R. *et al.*, 1991; Hill, C. S. *et al.*, 1994). More specifically it was discovered that attenuation of TCF binding to SRF on the c-fos promoter did not abolish its responsiveness to serum indicating that a TCF independent pathway was at work (Hill, C. S. *et al.*, 1994; Johansen, F. E. *et al.*, 1994). Furthermore it was shown that the integrity of the SRF DNA-binding domain including the sequences implicated in TCF binding was instrumental for the responsiveness of SRF to serum in the absence of TCF. Furthermore the TCF-independent response required that the SRF DNA-binding domain be tethered to its cognate DNA site (Hill, C. S. *et al.*, 1994). This led to a model in which authentic SRF-DNA contact was required for the interaction of wild-type SRF with a putative "recognition factor", that mediated serum induced signalling independently of TCF (Hill, C. S. *et al.*, 1994). The TCF-independent pathway was later found to involve members of the Rho family of small GTPases ((Hill, C. S. *et al.*, 1995b); analysed in the following section).

Further studies revealed that the Rho and MAPK-TCF pathways are mutually exclusive in their abilities to activate SRF dependent genes (Gineitis, D. *et al.*, 2001; Sotiropoulos, A. *et al.*, 1999). The differential sensitivity of target genes to each pathway correlates with the presence or absence of a TCF binding site in their promoters and the physical interaction of the TCF B box with SRF on target promoters

is sufficient to inhibit Rho mediated activation, indicating that combinatorial interactions of SRF with different cofactors at target gene promoters are the root of pathway selectivity (Murai, K. *et al.*, 2002).

1.2.5.7.3.1 The TCF-independent pathway: Rho-actin signalling to SRF

Rho GTPases belong to the Ras GTPase superfamily and have major roles in the regulation of the actin cytoskeleton and diverse cellular processes including adhesion, migration, morphology, membrane trafficking and proliferation (reviewed in (Etienne-Manneville, S. *et al.*, 2002)). Activation of RhoA in response to whole serum or mitogens such as lysophosphatidic acid (LPA) is necessary and sufficient for activation of SRF dependent transcription in the absence of TCF (Hill, C. S. *et al.*, 1995a). The related Rac1 and Cdc42 GTPases have the same effect, however SRF activation by extracellular signals that act through these GTPases remain to be formally demonstrated (Hill, C. S. *et al.*, 1995b).

The function of Rho GTPases in TCF independent activation of SRF is also linked to their role as cytoskeletal regulators. This was initially revealed indirectly, by a mammalian screen for SRF activators, which recovered LIM kinase-1, a regulator of the actin treadmilling cycle (Sotiropoulos, A. *et al.*, 1999). LIMK stabilises actin filaments by phosphorylating and inactivating cofilin, a protein that promotes the dissociation of actin monomers from F-actin pointed ends and induces F-actin severing (Arber, S. *et al.*, 1998; Yang, N. *et al.*, 1998). Although the requirement of LIMK for serum-regulated activation of SRF appears to be cell-type specific (Geneste, O. *et al.*, 2002; Sotiropoulos, A. *et al.*, 1999), its discovery was crucial since it established a link between the actin cycle and transcription via SRF.

More specifically proteins that promote actin polymerisation, such as members of the Diaphanous and WASP families activate SRF in a RhoA dependent manner, while dominant negative forms of these proteins inhibit both F-actin assembly and SRF activation (Copeland, J. W. *et al.*, 2002; Grosse, R. *et al.*, 2003; Sotiropoulos, A. *et al.*, 1999). Actin binding drugs that inhibit polymerisation such as Latrunculin B, inhibit RhoA signalling to SRF, while reagents that promote actin polymerisation such as Jasplakinolide activate SRF (Sotiropoulos, A. *et al.*, 1999). Furthermore overexpression of actin itself or non-polymerisable actin mutants inhibits SRF activation, while

overexpression of actin mutants that stabilise F-actin has the opposite effect (Posern, G. *et al.*, 2002; Sotiropoulos, A. *et al.*, 1999). Thus RhoA-dependent activation of SRF was shown to be mediated by changes in actin dynamics that promote actin polymerisation and G-actin depletion, and it was the latter that emerged as the event leading to SRF activation (Posern, G. *et al.*, 2002; Sotiropoulos, A. *et al.*, 1999).

Taken together these data support a model according to which RhoA activation and ensuing changes in actin dynamics result in the depletion of the G-actin pool and subsequent activation of an actin-regulated SRF-binding cofactor that interacts with the same or overlapping surface on SRF as the TCFs.

In addition to defining certain regulatory behaviours of the putative cofactor, the functional studies discussed in the previous sections also make clear predictions about the abilities of this cofactor to interact with SRF. Using these criteria, soon after the discovery of the MRTFs MAL was identified as the elusive Rho-regulated SRF cofactor (Figure 1.15; (Miralles, F. *et al.*, 2003)). MAL cannot bind SRF efficiently in the absence of cognate SRE DNA, and cannot interact with altered-specificity SRF mutants that have heterologous N-terminal sequences in their DNA-binding domains ((Miralles, F. *et al.*, 2003; Zaromytidou, A. I. *et al.*, 2006); analysed in Chapters 2 and 5). Moreover MAL competes with the TCFs for binding the same surface on SRF as the TCFs ((Miralles, F. *et al.*, 2003); see Chapters 2 and 4).

MAL binding to actin renders it predominantly cytoplasmic and inactive under basal conditions and its nuclear localisation and transcriptional activity is promoted in response to Rho signalling ((Miralles, F. *et al.*, 2003); see below). Treatment with proteins/reagents that promote F-actin assembly or actin binding drugs that disrupt the actin-MAL complex, like Cytochalasin D has the same effect, whereas proteins that induce F-actin disassembly or drugs such as Latrunculin B that prevent actin polymerisation without affecting the actin-MAL interaction inhibit nuclear accumulation and transcriptional activation (Miralles, F. *et al.*, 2003).

In the fibroblast system MAL is continuously shuttling between the cytoplasm and the nucleus under basal conditions, but its high export rates result in its predominantly cytoplasmic appearance (Miralles, F. *et al.*, 2003) (Vartiainen, M. K. *et al.*, 2006). Activation of the Rho-actin pathway promotes MAL accumulation in the nucleus where

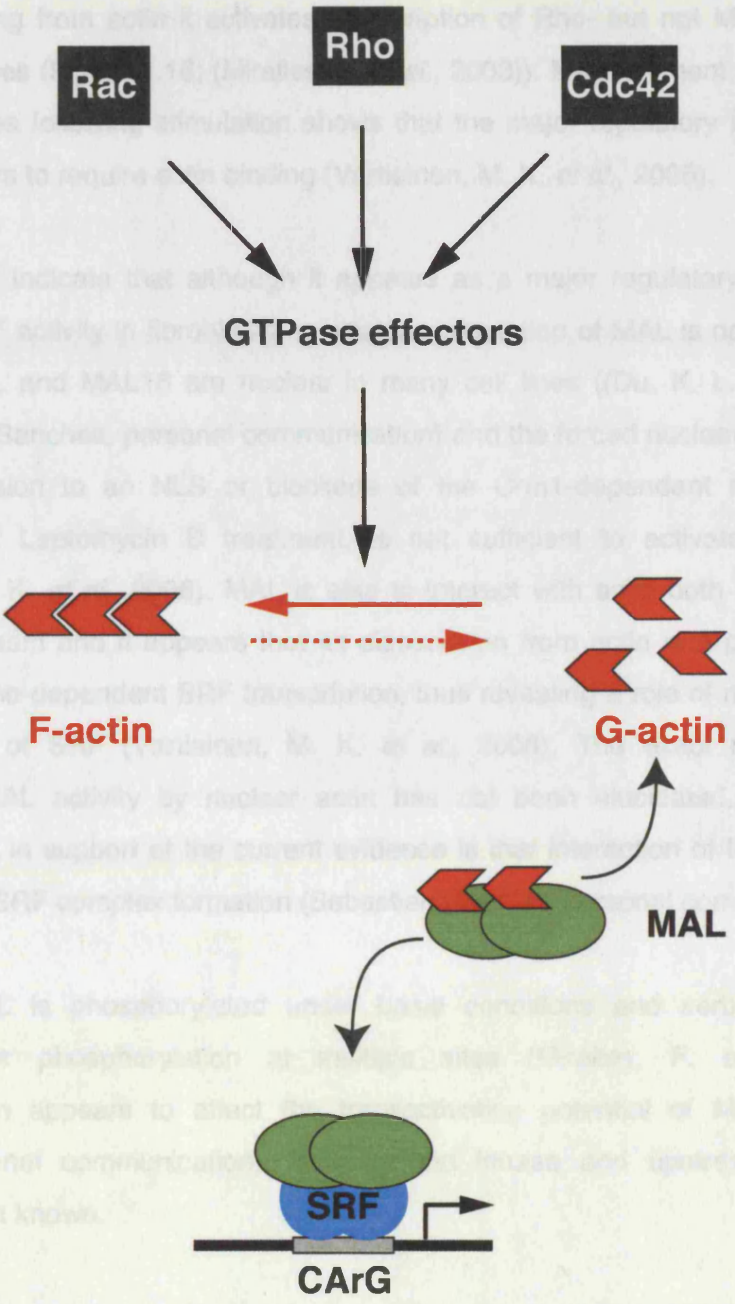


Figure 1.15. Rho-controlled actin dynamics regulate SRF activity through MAL. Signalling through Rho GTPases and their effectors (including Diaphanous and LIMK) induces actin polymerisation (shown as a solid red arrow). The depletion of the G-actin pool (dashed red arrow) results in the dissociation of actin from the SRF cofactor MAL, which then binds SRF and activates transcription. Rho GTPases are shown as black squares, MAL as a green oval and SRF as a blue circle. The CArG site on DNA is shown as a grey box on a black line. Adapted from (Posern and Treisman, 2006).

after dissociating from actin it activates transcription of Rho- but not MAPK-sensitive SRF target genes (Figure 1.16; (Miralles, F. *et al.*, 2003)). Measurement of MAL import and export rates following stimulation shows that the major regulatory step is export, and this appears to require actin binding (Vartiainen, M. K. *et al.*, 2006).

Recent studies indicate that although it appears as a major regulatory step in Rho-dependent SRF activity in fibroblasts, nuclear accumulation of MAL is not sufficient for activation. MAL and MAL16 are nuclear in many cell lines ((Du, K. L. *et al.*, 2004); Cristina Perez-Sanchez, personal communication) and the forced nuclear accumulation of MAL by fusion to an NLS or blockade of the Crm1-dependent nuclear export mechanism by Leptomycin B treatment, is not sufficient to activate transcription (Vartiainen, M. K. *et al.*, 2006). MAL is able to interact with actin both in the nucleus and the cytoplasm and it appears that its dissociation from actin is a prerequisite for activation of Rho-dependent SRF transcription, thus revealing a role of nuclear actin in the regulation of SRF (Vartiainen, M. K. *et al.*, 2006). The exact mechanism of inhibition of MAL activity by nuclear actin has not been elucidated, although the simplest model in support of the current evidence is that interaction of MAL with actin prevents MAL-SRF complex formation (Sebastian Guettler, personal communication).

Moreover, MAL is phosphorylated under basal conditions and serum stimulation induces further phosphorylation at multiple sites (Miralles, F. *et al.*, 2003). Phosphorylation appears to affect the transactivation potential of MAL (Francesc Miralles, personal communication), however the kinase and upstream regulators involved are not known.

In contrast to Myocardin, which does not respond to Rho signalling, MAL16 is regulated much like MAL in this system (Cristina Perez-Sanchez, Francesc Miralles, Sebastian Guettler, personal communication).

While the activation of Rho-regulated transcription via MAL-SRF is well understood the means by which MAL activity is down-regulated is less clear. MAL remains nuclear for hours after the initial transcriptional response and the purpose of this localisation is unknown (Miralles, F. *et al.*, 2003). MAL has recently been reported to be a target of sumoylation in response to Rho-signalling ((Nakagawa, K. *et al.*, 2005); Francesc Miralles, personal communication). This event follows the early MAL-induced

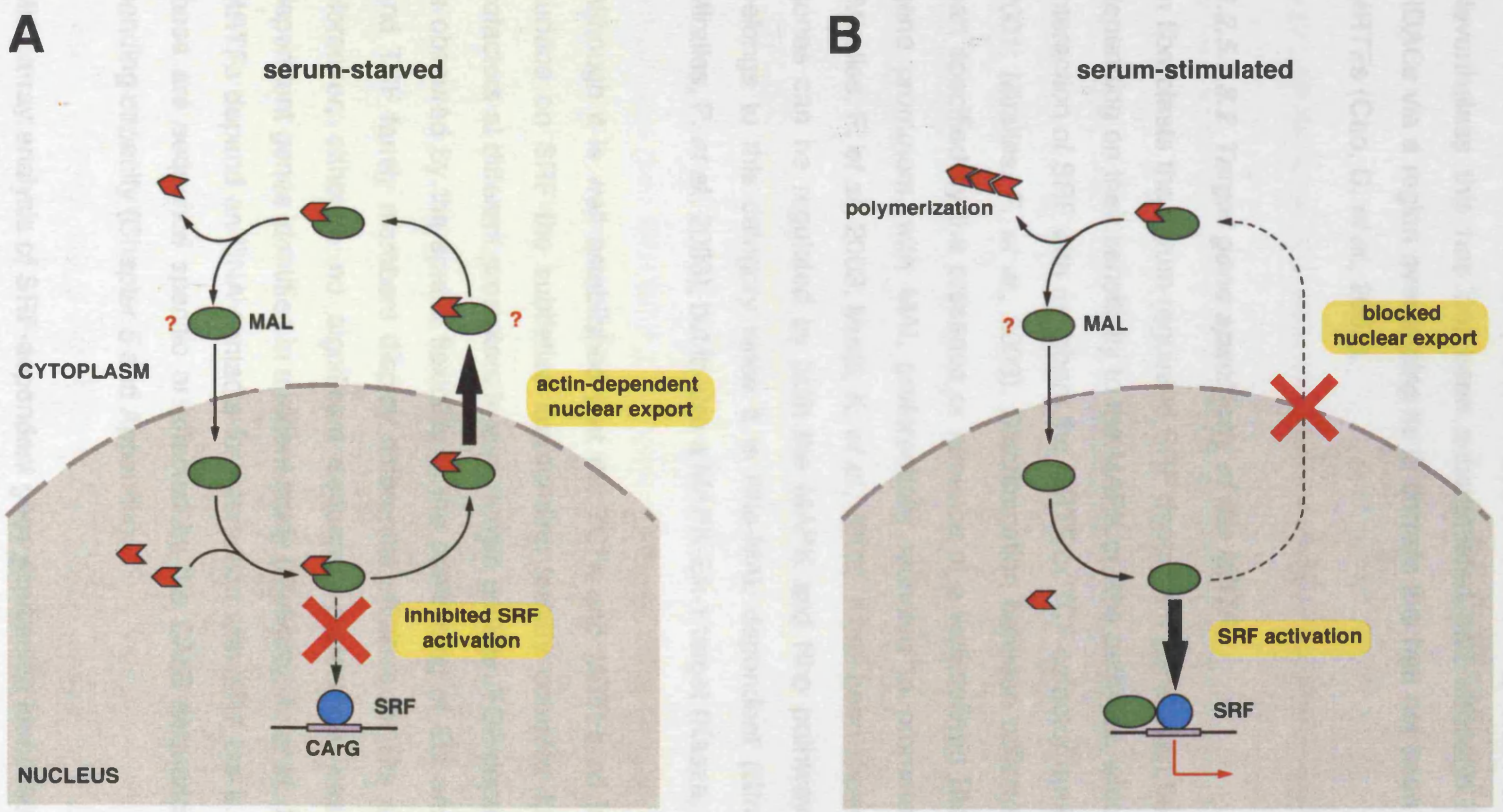


Figure 1.16. Regulation of MAL localisation and activity by actin in the fibroblast model. (A) Under basal conditions MAL is continuously shuttling between the nucleus and cytoplasm, but high export rates maintain it mainly in the cytoplasm and actin binding prevents SRF activation by the reduced nuclear population of MAL. (B) Serum stimulation activates Rho signalling and actin polymerisation resulting in the nuclear accumulation of MAL due to decreased export rates and activation of SRF-dependent transcription due to the dissociation from actin. The red "?" indicates the uncertainty surrounding the actin-binding state of MAL upon nuclear import and export. MAL is shown in green, SRF in blue and actin in red. Proteins are shown as monomers for simplicity. Adapted from (Vartiainen et. al., submitted).

transcriptional response and represses MAL activity. Thus sumoylation has been proposed as a potential mechanism for MAL-SRF transcription termination, possibly by recruitment of HDAC complexes as seen in other systems (see Section 1.2.5.6.5). Nevertheless this has not been substantiated and although Myocardin can recruit HDACs via a region overlapping its Q domain this has not been explored for the other MRTFs (Cao, D. *et al.*, 2005).

1.2.5.7.3.2 Target gene specificity of the MRTFs

In fibroblasts the serum-regulated SRF dependent genes can be divided in two groups depending on their sensitivity to the MAPK or Rho pathway, which results in differential interaction of SRF with members the MRTF or TCF cofactor families (Gineitis, D. *et al.*, 2001; Miralles, F. *et al.*, 2003). Discrimination between cofactors appears at least in part specified by the presence or absence of a well-defined Ets binding site at target gene promoters, with MAL preferentially recruited to promoters lacking such sites (Miralles, F. *et al.*, 2003; Murai, K. *et al.*, 2002). It has been shown however that certain genes can be regulated by both the MAPK and Rho pathways. The *srf* gene itself belongs to this category since it is Rho-MAL dependent (Gineitis, D. *et al.*, 2001; Miralles, F. *et al.*, 2003), but is also a MAPK-Elk-1 target (Kasza, A. *et al.*, 2005).

Although it is well established that the TCFs and MRTFs all interact with the same surface on SRF the subtleties surrounding ternary complex formation with different cofactors at different promoters in vivo remain unclear. TCF dependence of SRF genes is obscured by the spatial flexibility in the positioning of Ets and CArG sites on DNA and TCF family members display differential affinities for Ets sites (Section 1.2.5.6). Moreover, although no significant sequence preference was found in the MAL-dependent genes identified in a recent study ((Selvaraj, A. *et al.*, 2004); see below), the MRTFs depend on DNA contacts for interaction with SRF but it is not known whether these are sequence specific or affected by the CArG sequence itself or its intrinsic bending capacity (Chapter 5 and Appendix).

Microarray analysis of SRF-dependent gene expression identified 28 SRF target genes that are likely to be regulated by Rho-MAL signalling in response to serum, as defined by sensitivity to repression by a MAL mutant competent for SRF binding but not transcriptional activation (Selvaraj, A. *et al.*, 2004). Some of these code for

transcriptional regulators and a small number is involved in cytoskeletal processes. It should be noted however, that the genes identified in this study are unlikely to represent the complete spectrum of serum-induced MAL dependent SRF targets, since the dominant-negative MAL approach used may block both MRTF and TCF dependent targets.

Overexpression of a C-terminally truncated SRF-binding MAL derivative could perturb the interaction of SRF with other cofactors at different promoters, with currently unpredictable outcomes on their serum-inducibility. Thus identification of MRTF target genes requires more specific methods of cofactor inactivation such as siRNA-targeted knockdown. In addition, as will be discussed in the following sections, the overlapping expression patterns of MRTF family members in various tissues and their possible functional redundancy as well as the similar modes of regulation in the case of MAL and MAL16 also need to be taken into account when investigating the target genes of specific family members.

1.2.5.7.3.3 Biological roles of the MRTFs in cytoskeletal processes

The involvement of SRF in regulating genes involved in cytoskeletal processes is well established (see Section 1.2.5.4), and Rho signalling through the MRTFs has also been found to regulate expression of cytoskeletal genes such as vinculin, zyxin and actin itself (Gineitis, D. *et al.*, 2001; Miralles, F. *et al.*, 2003; Selvaraj, A. *et al.*, 2004). Indeed the cytoskeleton related effects of SRF in neurite outgrowth have been linked to Rho-actin signalling through MAL (Knoll, B. *et al.*, 2006).

Further evidence on the partnership of MAL and SRF in regulating cytoskeletal processes comes from studies in Drosophila melanogaster. There is one MAL isoform in fruit flies, named DMRTF/MAL-D (Figure 1.14; (Han, Z. *et al.*, 2004; Somogyi, K. *et al.*, 2004)). DMRTF shares homology with its mammalian counterparts in the N-terminal RPEL-, B1- and SAP domains and also contains a C-terminal transactivation domain. DMRTF does not contain a Q-box equivalent, although a stretch of hydrophobic residues resembling the hydrophobic core of the mammalian Q-box is located C-terminal to the B1 region and a glutamine-rich region exists in the C-terminal part of the protein (Han, Z. *et al.*, 2004; Somogyi, K. *et al.*, 2004). It is noteworthy that

this is the first SRF cofactor identified in Drosophila, since no TCF equivalent has been found in this organism.

DMRTF acts together with SRF during fly development and its deletion is lethal in homozygous larvae (Han, Z. *et al.*, 2004; Somogyi, K. *et al.*, 2004). DMRTF-null embryos display abnormalities in tracheal branching that are similar to the DSRF pruned phenotype, indicating that the two proteins cooperate during tracheal development ((Han, Z. *et al.*, 2004); and Section 1.2.5.4.1). Expression of dominant active or negative DMRTF derivatives in wing imaginal discs results in increase or decrease of intervein tissue correlating with the role of SRF in wing development ((Han, Z. *et al.*, 2004) and Section 1.2.5.4.1). Furthermore expression of dominant negative DMRTFs in Drosophila embryos affects the migration of mesodermal cells (Han, Z. *et al.*, 2004).

DMRTF activity is also required for cytoskeletal integrity and migration of border cells during Drosophila oogenesis (Somogyi, K. *et al.*, 2004). In these cells DMRTF is activated in response to perceived mechanical tension or cell deformation and translocates to the nucleus where it partners with SRF. Regulation of this system appears comparable to the mammalian Rho-actin pathway since an active form of the Drosophila Diaphanous protein could induce the translocation of DMRTF to the nucleus (Somogyi, K. *et al.*, 2004).

Thus the link of Rho-SRF regulation to the actin treadmilling cycle via the actin-binding MRTFs points to a signalling feedback mechanism through which cytoskeletal rearrangements autoregulate the expression of actin and other cytoskeletal components.

1.2.5.7.3.4 Biological roles of the MRTFs in myogenesis

All three MRTF family members are involved in multiple aspects of myogenesis. The founding member of the MRTF family, Myocardin is restricted to smooth and cardiac muscle and is one of the earliest markers of cardiac and smooth muscle cell lineages (Wang, D. *et al.*, 2001). The significance of Myocardin in cardiomyogenesis was demonstrated in Xenopus embryos were expression of a dominant negative form of the

protein and morpholino-mediated knockdown both inhibited heart development and cardiac gene expression (Small, E. M. *et al.*, 2005; Wang, D. *et al.*, 2001).

Surprisingly Myocardin-null mice do not display defects in cardiac development, but instead die in the embryonic stages due to the absence of vascular smooth muscle cells (Li, S. *et al.*, 2003). This severe phenotype combined with studies in which forced expression of Myocardin in cultured cells induces SRF dependent expression of smooth muscle genes (Chen, J. *et al.*, 2002; Du, K. L. *et al.*, 2003; Wang, Z. *et al.*, 2003; Yoshida, T. *et al.*, 2003) identified Myocardin is an important component of the smooth-muscle gene programme. Nevertheless this protein is not absolutely required for smooth muscle differentiation since Myocardin-null ES cells differentiate normally to smooth muscle cells in vitro (Pipes, G. C. *et al.*, 2005).

The high expression levels of MAL and MAL16 in muscle tissues (Selvaraj, A. *et al.*, 2003b; Wang, D. Z. *et al.*, 2002) combined with the fact that MAL and MAL16 are also able to induce smooth muscle gene expression when overexpressed in non-muscle cell lines including fibroblasts and ES cells (Du, K. L. *et al.*, 2004; Wang, D. Z. *et al.*, 2002) raise the possibility of functional compensation by other MRTFs in the absence of Myocardin.

In support of the involvement of the MRTFs in the smooth-muscle differentiation programme, inactivation of MAL16 results in severe lethal phenotypes related to cardiovascular abnormalities due to defects in the differentiation of neural crest derived smooth muscle cells (Li, J. *et al.*, 2005; Oh, J. *et al.*, 2005).

In contrast to the severe phenotype observed in MAL16-inactivated mice, MAL knockout mice are viable (Li, S. *et al.*, 2006; Sun, Y. *et al.*, 2006). Furthermore MAL appears dispensable for muscle development in mice, with knockouts only displaying a defect in the myoepithelial cells that are required for milk secretion from the mammary gland of lactating females. Myoepithelial cells are similar to smooth muscle cells in their structure, gene expression pattern, and contractile properties and the MAL-/- defect is related to downregulation of smooth muscle specific genes (Li, S. *et al.*, 2006; Sun, Y. *et al.*, 2006). MAL16 is also expressed in myoepithelial cells, but is not able to substitute for the function of MAL, and although this suggests that the two proteins have distinct roles it has also been suggested that they are indeed functionally

redundant and the inactivation of MAL results in the decrease of a threshold level of MRTF activity required for normal function (Li, S. *et al.*, 2006).

In one of the two studies of MAL knockout mice published to date approximately 40% of MAL-null embryos display cardiac defects and die due to cardiac cell necrosis, although the embryos that survive do not display any muscle-related phenotypes (Sun, Y. *et al.*, 2006). Although this effect connects MAL with aspects of cardiac development the reasons behind the incomplete penetrance of the phenotype and the variability between the two studies remain unknown.

Although the role of Myocardin is confined to cardiac and smooth muscle the other MRTFs have been implicated in skeletal muscle differentiation, since expression of dominant negative MAL derivatives or siRNA knockdown of MAL and MAL16 block skeletal muscle gene expression in cultured cells (Kuwahara, K. *et al.*, 2005; Selvaraj, A. *et al.*, 2003).

Despite the involvement of all three MRTF family members in myogenic differentiation the significant overlap in their expression patterns in muscle lineages (Selvaraj, A. *et al.*, 2003; Wang, D. *et al.*, 2001; Wang, D. Z. *et al.*, 2002) and the fact that they interact with SRF in the same manner (Zaromytidou, A. I. *et al.*, 2006), make functional redundancy between family members a central issue. Although the differences in the severities of the knockout phenotypes and the inability in some cases of one family member to replace the other point to at least some level of functional individuality in their roles, double or triple knockout studies are needed to unravel the specific contributions of each family member to cardiac, smooth and skeletal muscle development.

1.2.5.8 Interactions of SRF with other cofactors

Although TCF and MRTF family members are the best characterised SRF cofactors involved in well-defined complex formation with SRF, many other proteins have been reported to interact with SRF and to influence its transcriptional responses. Many of these interactions lack rigorous biochemical characterisation and it is unclear to what extent they represent bona fide SRF cofactors. Table 1.2 lists most of the SRF partners

reported to date. The present section will discuss some of these interactions and their effects on SRF dependent transcription.

Table 1.2 Interactions of SRF with other cofactors

Cofactor	Cofactor characteristics	Experimental evidence
Phox1	Homeodomain	EMSA kinetic effect (Grueneberg, D. A. *et al.*, 1992) (Simon, K. J. *et al.*, 1997)
Nkx2.5	Homeodomain Cardiac muscle specific	GST-pulldown (Chen, C. Y. *et al.*, 1996)
Nkx3.1	Homeodomain protein	Co-IP (Carson, J. A. *et al.*, 2000)
Nkx3.2	Homeodomain Smooth muscle specific	EMSA kinetic effect, GST-pulldown (Nishida, W. *et al.*, 2002)
Barx2b	Homeodomain	GST-pulldown (Herring, B. P. *et al.*, 2001)
HOP	Homeodomain protein, muscle specific	GST-pulldown, Co-IP, EMSA competition (Chen, F. *et al.*, 2002; Shin, C. H. *et al.*, 2002)
Fhl2	LIM-only FHL family	Co-IP, ChIP (Philippar, U. *et al.*, 2004)
CRP2 , CRP1	LIM-only CRP family	2-hybrid, GST-pulldown, Co-IP, EMSA (Chang, D. F. *et al.*, 2003)
TEF-1	TEA/ATTS transcription factor family, involved in muscle-specific gene expression	GST-pulldown, Co-IP, far western (Gupta, M. *et al.*, 2001)
Myogenin/E12 **MyoD/E12**	Basic-helix-loop-helix proteins involved in myogenesis	Yeast 2-hybrid, GST pulldown (Groisman, R. *et al.*, 1996)
GATA4	Zinc finger transcription factor, muscle	GST-pulldown, Co-IP

Cofactor	Cofactor characteristics	Experimental evidence
	specific	(Belaguli, N. S. *et al.*, 2000) EMSA, Co-IP (Morin, S. *et al.*, 2001)
GATA6	Zinc finger transcription factor	EMSA, GST-pulldown (Nishida, W. *et al.*, 2002)
Fli1	ETS domain	EMSA (Watson, D. K. *et al.*, 1997) (Dalgleish, P. *et al.*, 2000)
C/EBPβ	Ras-regulated activation of c-fos	Yeast 2-hybrid, Co-IP (Hanlon, M. *et al.*, 1999)
Smad3	Receptor activated Smad, TGFβ regulated cofactor	Co-IP (Qiu, P. *et al.*, 2003)
Smad7	Inhibitory-Smad, TGFβ signalling repressor	Co-IP (Camoretti-Mercado, B. *et al.*, 2006)
p65/NF-κB	Rel homology domain	GST-pulldown (Franzoso, G. *et al.*, 1996)
YY1	Zinc finger protein	EMSA (Natesan, S. *et al.*, 1995)

SRF has been shown to functionally co-operate with many homeodomain proteins, resulting in SRF-dependent gene activation. Most of these interactions are thought to involve binding of the homeodomain to the major groove of the CArG site on DNA (Phox1, Nkx2.5, Barx2b; (Chen, C. Y. *et al.*, 1996; Grueneberg, D. A. *et al.*, 1992; Herring, B. P. *et al.*, 2001)), or to sites adjacent to the CArG box (Nkx3.1; (Carson, J. A. *et al.*, 2000)) and result in increased SRF-DNA binding. Nevertheless despite clear enhancement of SRF-DNA binding in the presence of these proteins in in vitro binding assays, no direct interaction in a ternary complex with DNA has been shown to date.

Binding to the major groove of the SRF-occupied CArG box has also been proposed for YY1, a Zinc finger protein that confers significant bend on its DNA sites and can act as an activator or repressor depending on promoter context ((Natesan, S. *et al.*, 1995) and references therein). YY1 is able to form a ternary complex with SRF and enhances the

kinetics of SRF-DNA binding on the c-fos promoter, although the biological significance of this interaction remains unknown (Natesan, S. *et al.*, 1995).

In the case of the Phox1-SRF interaction a factor named SPIN (SRF-Phox1-Interacting protein) has been identified that is thought to bind both proteins forming an SRF/Phox1/SPIN complex (Grueneberg, D. A. *et al.*, 1997). The SPIN protein is identical to the basal transcription factor TFII-I and its interaction with SRF-Phox1 has been implicated in induction of SRF-dependent genes such as c-fos in reporter assays (Grueneberg, D. A. *et al.*, 1997).

A multicomponent complex has also been suggested to form between SRF and the heart specific Nkx2.5 and GATA4 factors to enhance transcription of cardiac genes (Belaguli, N. S. *et al.*, 2000; Chen, C. Y. *et al.*, 1996; Sepulveda, J. L. *et al.*, 2002), and between SRF and Nkx3.2 and GATA6 for smooth muscle specific gene expression (Nishida, W. *et al.*, 2002). Although this functional cooperation has been clearly demonstrated with coexpression of the proteins in reporter assays, the biochemical evidence of the interaction is weaker.

SRF has also been shown to synergise with GATA factors in complexes containing members of the LIM-only cysteine-rich protein family (CRP1 and CRP2; (Chang, D. F. *et al.*, 2003)). These proteins contain two LIM domains through which they are thought to mediate subcellular protein targeting and assembly of multiprotein complexes, for example cell-adhesion complexes by binding the actin-binding proteins α-actinin and zyxin ((Chang, D. F. *et al.*, 2003) and references therein). CRP1 and CRP2 however are nuclear in early embryonic stages and were shown by GST-pulldown to interact with SRF via their N-terminal LIM domain and GATA factors via the C-terminal one (Chang, D. F. *et al.*, 2003). Although no in vitro quaternary complex formation or indeed recruitment of these factors to promoters in vivo has been shown, the presence of CRP2 increases SRF binding to DNA in vitro and cotransfection of all three factors potentiates smooth muscle gene expression in transient transfections (Chang, D. F. *et al.*, 2003). Thus CRP1/2 have been proposed to act as bridging molecules between SRF and GATA factors that integrate their regulatory functions in target gene expression.

In addition to these positively acting muscle-restricted interactions, several repressors of SRF-dependent muscle gene expression have also been identified. HOP (Homeodomain Only Protein) is a cardiac specific protein, which unlike most homeodomain factors is incapable of specific DNA binding due to its divergent sequences in the third homeodomain helix (Chen, F. *et al.*, 2002; Shin, C. H. *et al.*, 2002). HOP interacts with SRF and prevents its binding to DNA thus repressing SRF-dependent transcription (Chen, F. *et al.*, 2002; Shin, C. H. *et al.*, 2002). HOP-null mice display an embryonic lethal phenotype due to cardiac defects and thus this protein is considered an important regulator of cardiomyogenesis.

FHL2 (Four and a half LIM domain protein 2) is another LIM domain containing inhibitor of SRF activation. FHL2 was identified by microarray analysis as an SRF target gene upregulated in response to RhoA activation and muscle differentiation (Philippar, U. *et al.*, 2004). This muscle-specific protein interacts with SRF in vitro and upon RhoA activation translocates to the nucleus where it is recruited to the promoters of SRF-dependent genes such as SM22 and αSM-actin, resulting in inhibition of MAL-induced transcriptional activation (Philippar, U. *et al.*, 2004). Furthermore FHL2 appears to compete with MAL for SRF binding in vitro, although this has not been analysed in detail. These observations have led to a model in which Rho-MAL signalling to SRF creates a negative feedback loop resulting in FHL2 production and downregulation of MAL-dependent gene expression.

Combinatorial interactions with other transcription factors are the crux of the transcriptional versatility of SRF. Identifying the ways these interactions are modulated temporally and spatially, as well as the potential integration of different signalling pathways at target gene promoters through multiprotein complexes including SRF and other transcription factors will provide necessary insights in the biological significance of SRF and the diversity of the roles it fulfils.

This thesis will describe the molecular mechanism of the SRF interaction with the downstream activators of the RhoA pathway the MRTFs, and compare it to the interaction of SRF with the MAPK-regulated TCFs in an attempt to elucidate how the physical association of SRF with members of each family is able to confer differential signal-sensitivity to target gene promoters.

2 The MAL-SRF complex

2.1 Aims

The identification of the Myocardin-related transcription factor family established a novel group of SRF coactivators (Ma, Z. *et al.*, 2001; Mercher, T. *et al.*, 2001; Wang, D. *et al.*, 2001; Wang, D. Z. *et al.*, 2002). Initial studies in our lab showed that in contrast to the constitutively active Myocardin, the transcriptional activity of MAL was serum-inducible and RhoA-dependent, thus making it a candidate for the Rho-actin pathway mediator. Functional studies had previously identified a number of properties expected of the Rho-regulated SRF cofactor. These included that binding of this cofactor to SRF and subsequent transcriptional activation required the intact N-terminal region of the SRF DNA-binding domain and also that the cofactor would compete with the TCFs for a common surface on the SRF DNA-binding domain (Hill, C. S. *et al.*, 1993; Hill, C. S. *et al.*, 1994; Murai, K. *et al.*, 2002). My initial aim was therefore to establish whether MAL interacts with SRF on DNA, and subsequently to investigate whether the SRF-binding properties of MAL correlate with those predicted for the Rho-pathway coactivator.

2.2 Formation of the MAL-SRF-DNA complex

The open reading frame of the predominant MAL mRNA present in NIH3T3 cells is predicted to produce MAL(met), a 929 amino-acid protein starting at the first ATG codon, N-terminal to the second RPEL motif (Figure 2.1). Indeed many studies on MAL and the MRTF protein family, describe MAL(met) as the full-length form of the protein (Cen, B. *et al.*, 2003; Sasazuki, T. *et al.*, 2002; Wang, D. Z. *et al.*, 2002). However, a MAL cDNA including the 5' UTR gives rise to a bigger protein, suggesting that translation begins upstream of the first in-frame methionine. Mutagenesis studies confirmed this and showed that translation of MAL begins at or just N-terminal to a leucine, 92 amino-acids upstream of the first methionine, producing MAL(fl), a protein that contains a third RPEL motif ((Miralles, F. *et al.*, 2003); Figure 2.1).

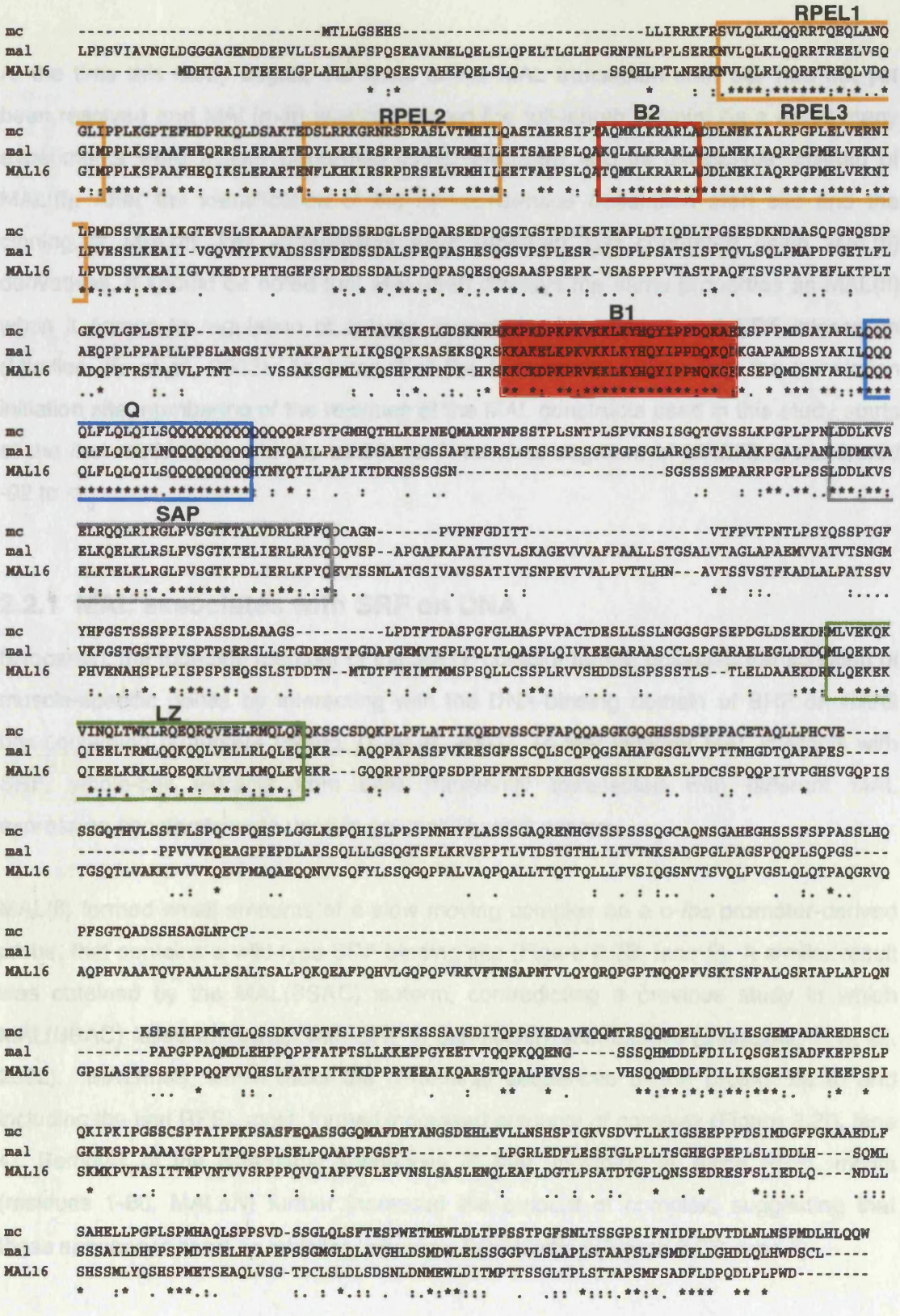


Figure 2.1. Sequence alignment of the mouse isoforms of the MRTF family members. Myocardin residues 1 to 976, MAL residues -92 to 929 and MAL16 residues 1 to 1091 were aligned using the CLUSTALW programme. For simplicity the N-terminal variant isoforms of MAL and MAL16 are not included in the alignment as their divergent sequences stop before RPEL1. Conserved motifs are labeled and indicated by coloured boxes. Identical residues are marked by asterisks and conservative replacements by colons and dots.

At the time this study began, the issue of the MAL translation start site had not yet been resolved and MAL(met) was considered the full-length protein. As a result many experiments were initially performed using MAL(met) and its derivatives, instead of MAL(fl). After the identification of the non-consensus translation start site and the cloning of MAL(fl), key experiments were repeated and confirmed using MAL(fl) derivatives. It should be noted that MAL(met) displays the same properties as MAL(fl) when it comes to regulation of activity, subcellular localisation and SRF interaction (Miralles, F. *et al.*, 2003). As a result of the uncertainty surrounding the translation initiation site, numbering of the residues of the MAL constructs used in this study starts at the first methionine and the additional N-terminal segment of MAL(fl) is numbered -92 to -1.

2.2.1 MAL associates with SRF on DNA

Myocardin, the founding member of the MRTF cofactor family, activates transcription of muscle-specific genes by interacting with the DNA-binding domain of SRF on CArG box containing promoters (Wang, D. *et al.*, 2001). To test whether MAL interacts with SRF, whole-cell extracts from cells transiently transfected with different MAL expression constructs were used in gel-mobility shift assays.

MAL(fl) formed small amounts of a slow moving complex on a c-*fos* promoter-derived probe, that contains a wild-type SRF binding site (Figure 2.2B, lane 2). A similar result was obtained by the MAL(BSAC) isoform, contradicting a previous study in which MAL(BSAC) failed to interact with SRF in gel-mobility shift assays (Sasazuki, T. *et al.*, 2002). MAL(met), which lacks the N-terminal sequences of the protein up to and including the first RPEL motif, formed increased amounts of complex (Figure 2.2B, lane 4). Removal of the N-terminal sequences of MAL including all three RPEL motifs (residues 1-80, MALΔN) further increased the amount of complex, suggesting that these sequences exert an inhibitory effect on SRF binding (Figure 2.2B, lane 5).

All complexes could be supershifted by anti-MAL and anti-SRF antibodies indicating that both proteins were present in the complex. Furthermore, no complexes formed on probe FOS.M, which has a mutated CArG box that cannot bind SRF, indicating that complex formation was dependent on the presence of endogenous SRF in the extracts (Figure 2.2B, lanes 16-20).

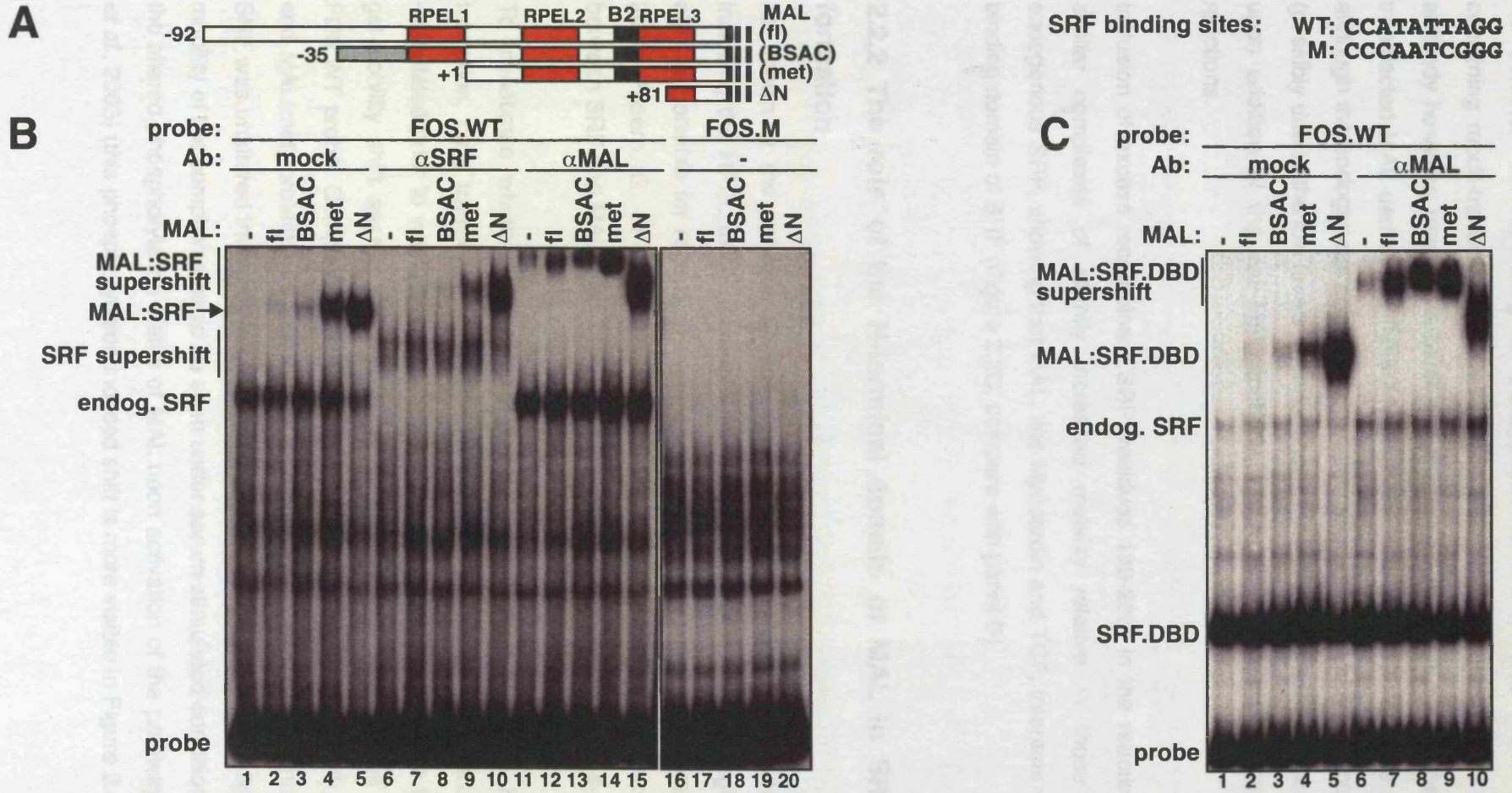


Figure 2.2. The MAL-SRF complex. (A) The different N-termini of MAL constructs are shown on the left. The sequences of the SRF-binding sites (WT and M) are shown on the right. (B) MAL complex formation with endogenous SRF. Gel mobility-shift assays contained the indicated MAL whole-cell extracts and probes, and anti-SRF or anti-MAL antibodies. The anti-MAL antibody was raised against MAL(met) residues 1-170. Positions of the MAL-SRF complex and the supershifted complexes are indicated. (C) MAL complex formation with the SRF DNA binding domain. Gel mobility-shift assays as in B, including recombinant SRF (residues 133-265).

No complex corresponding to endogenous MAL-SRF was visible in reactions containing mock-transfected extract (Figure 2.2B, lane 1). Inclusion of the anti-MAL antibody however, revealed a band at a position equivalent to that of the supershifted transfected MAL derivatives (lane 11, compare with lanes 12-14). This suggests that although the endogenous complex is undetectable under the bandshift conditions used (possibly due to the low levels of endogenous MAL in the extract), it becomes apparent upon addition of the anti-MAL antibody, which visibly enhances complexes in all reactions.

Inclusion of excess recombinant SRF (residues 133-265) in the reactions generated similar complexes of slightly increased mobility relative to those formed with endogenous SRF, showing that MAL, like Myocardin and TCF, interacts with the DNA-binding domain of SRF (Figure 2.2C, compare with panel B).

2.2.2 The role of the N-terminal domain of MAL in SRF-complex formation

As shown in the previous section, removal of the N-terminal sequences of MAL including the RPEL motifs increases MAL-SRF complex yield. This region of MAL is also responsible for regulation by RhoA and interaction with actin (Miralles, F. *et al.*, 2003; Posern, G. *et al.*, 2004), thus raising the possibility that complex formation between SRF and MAL is also regulated and requires stimulation of the system.

To investigate whether activation of the Rho pathway affects MAL-SRF complex formation, cells transiently transfected with MAL derivatives were serum-starved or -stimulated prior to whole-cell extract preparation. These extracts were then tested in gel-mobility shift assays for their ability to interact with endogenous SRF on the FOS.WT probe. Complex formation between the Rho-regulated MAL(fl), MAL(BSAC) and MAL(met) proteins, as well as the constitutively active MALΔN, and endogenous SRF was unaltered irrespective of serum stimulation (Figure 2.3B). A slight shift in the mobility of the complexes could be seen under serum-stimulated conditions, alluding to the altered phosphorylation status of MAL upon activation of the pathway (Miralles, F. *et al.*, 2003) (this phosphorylation-induced shift is more visible in Figure 2.4B).

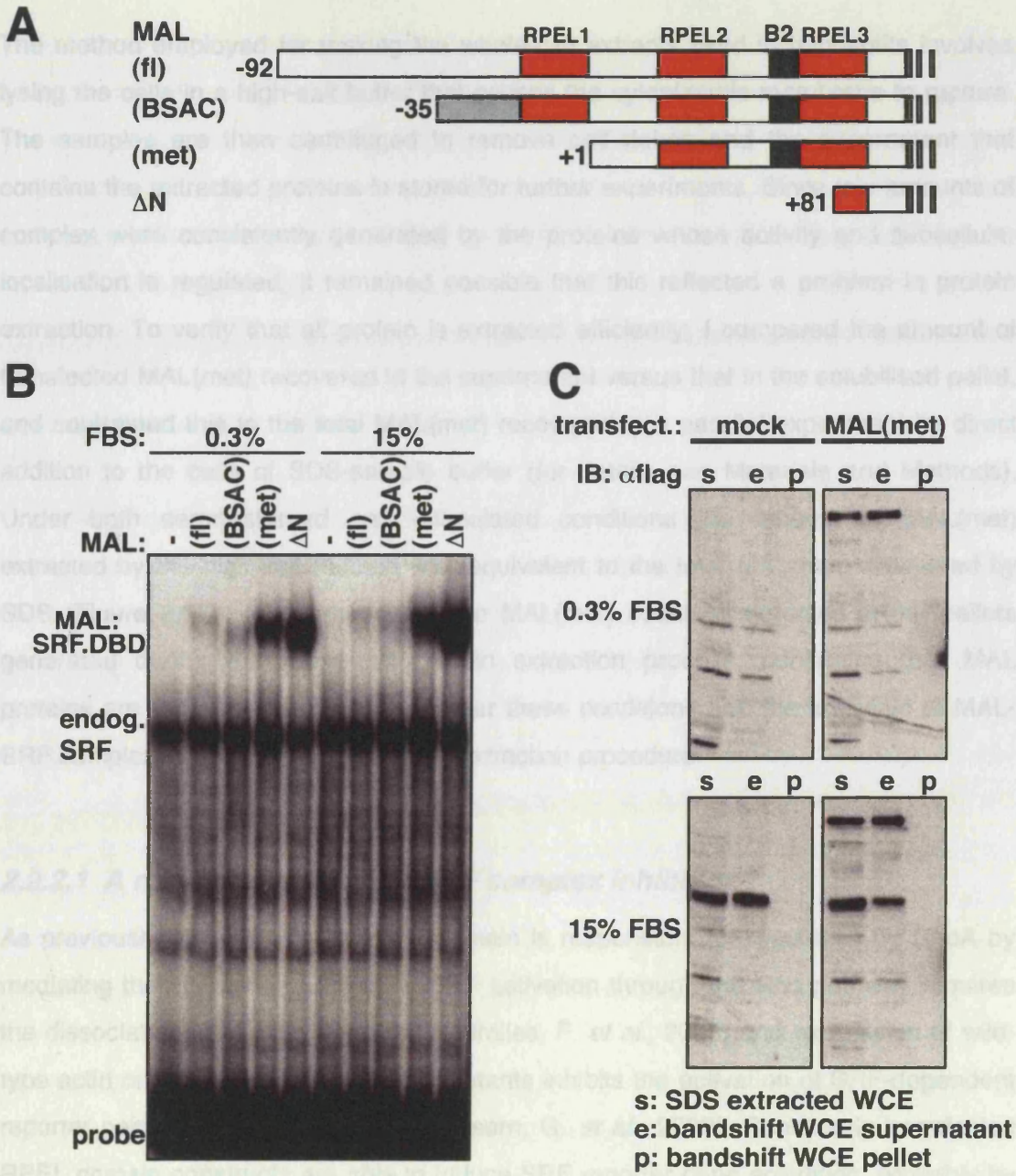


Figure 2.3. SRF complex formation by N-terminal MAL derivatives is not affected by serum stimulation. (A) Structures of the different N-termini of MAL constructs. (B) Complex formation of MAL N-terminal derivatives with endogenous SRF. Gel mobility-shift assays contained the indicated MAL whole-cell extracts and c-fos ΔTCF SRE probe. (C) Efficiency of protein extraction by the whole-cell extract technique used for bandshifts. Whole-cell extracts were made from Flag-tagged MAL(met) transfected cells and the recovery of MAL(met) protein in the supernatant and pellet was compared to the total MAL(met) extracted by addition to the cells of SDS sample buffer.

The method employed for making the whole-cell extracts used in bandshifts involves lysing the cells in a high-salt buffer that causes the cytoplasmic membrane to rupture. The samples are then centrifuged to remove cell debris and the supernatant that contains the extracted proteins is stored for further experiments. Since low amounts of complex were consistently generated by the proteins whose activity and subcellular localisation is regulated, it remained possible that this reflected a problem in protein extraction. To verify that all protein is extracted efficiently, I compared the amount of transfected MAL(met) recovered in the supernatant versus that in the solubilised pellet, and contrasted this to the total MAL(met) recovered in a parallel experiment by direct addition to the cells of SDS-sample buffer (for details see Materials and Methods). Under both serum-starved and -stimulated conditions the amount of MAL(met) extracted by the high-salt method was equivalent to the total MAL(met) recovered by SDS (Figure 2.3C). More importantly no MAL(met) could be detected in the pellets generated during the whole-cell protein extraction process, confirming that MAL proteins are recovered successfully under these conditions and the low yield of MAL-SRF complexes is not an artifact of the extraction procedure.

2.2.2.1 A role for actin in MAL-SRF complex inhibition?

As previously mentioned, the RPEL domain is responsible for regulation by RhoA by mediating the interaction with actin. SRF activation through the Rho pathway requires the dissociation of G-actin from MAL (Miralles, F. *et al.*, 2003) and expression of wild-type actin or non-polymerisable actin mutants inhibits the activation of SRF-dependent reporter genes in luciferase assays (Posern, G. *et al.*, 2002). Conversely transfected RPEL domain constructs are able to induce SRF reporter gene activation, probably by competing with MAL for actin binding (Sebastian Guettler, personal communication). The exact manner however, in which these events affect recruitment of MAL to SRF bound promoters and subsequently transcriptional activation remains unclear.

Taking into account that the bandshift experiments presented so far were performed with whole-cell extracts, it remained possible that the actin present in the extracts was inhibiting MAL-SRF complex formation by binding to the RPEL domain and preventing MAL from interacting with SRF. This possibility cast doubt on the suitability of the gel mobility-shift assay for investigating the interactions of full-length MAL with SRF. If actin-bound MAL is unable to interact with SRF, it is conceivable that the complexes

attributed to RPEL-containing MAL proteins and SRF are in fact formed by MAL derivatives that lack all or part of their RPEL sequences (perhaps due to protein translation from an internal start site, or N-terminal degradation) and are therefore unable to bind actin. To investigate whether the N-termini of the MAL proteins complexed with SRF in bandshifts were intact I performed antibody supershift assays using extracts expressing 5'-Flag-tagged MAL derivatives (Figure 2.4). All MAL-SRF complexes were efficiently supershifted by an anti-Flag antibody, demonstrating that the MAL derivatives complexed with SRF contained their complete N-terminal sequences.

I next employed a binding competition approach to investigate the possibility of actin-induced inhibition of the MAL-SRF complex: if the actin present in the extracts prevents MAL-SRF complex formation by binding to the RPEL domain of MAL, then inclusion of excess MAL-RPEL domain in the reactions would be expected to relieve this inhibition by competing with the full-length MAL protein for interaction with actin. Early attempts to test this by titrating GST-RPEL fusion proteins or coexpressing RPEL-domain and MAL constructs in cell extracts, did not affect MAL-SRF complex yield (data not shown). It remained unclear however whether this was a genuine negative result: actin is very abundant in whole-cell extracts and it is conceivable that the concentration range of titrated RPEL domain or the expression levels of the transfected RPEL construct were not sufficient to relieve an actin-induced inhibition of SRF binding.

Furthermore, attempts to titrate purified actin into MAL-SRF reactions in order to compete for complex formation were also unsuccessful, since the salt conditions used in the bandshift assays induce actin polymerisation. This was clearly visible in the samples containing high actin concentrations, were the polymerised actin remained in the wells and did not enter the gel (data not shown).

2.2.2.2 The effect of mutations in the RPEL motifs on the MAL-SRF complex

Having established that authentic full-length MAL-SRF complexes can be detected by the bandshift assay, and since addressing the role of actin in complex formation directly was prevented by technical difficulties, I attempted to further explore the role of the RPEL domain in SRF-binding. The RPEL motifs of MAL were identified by their

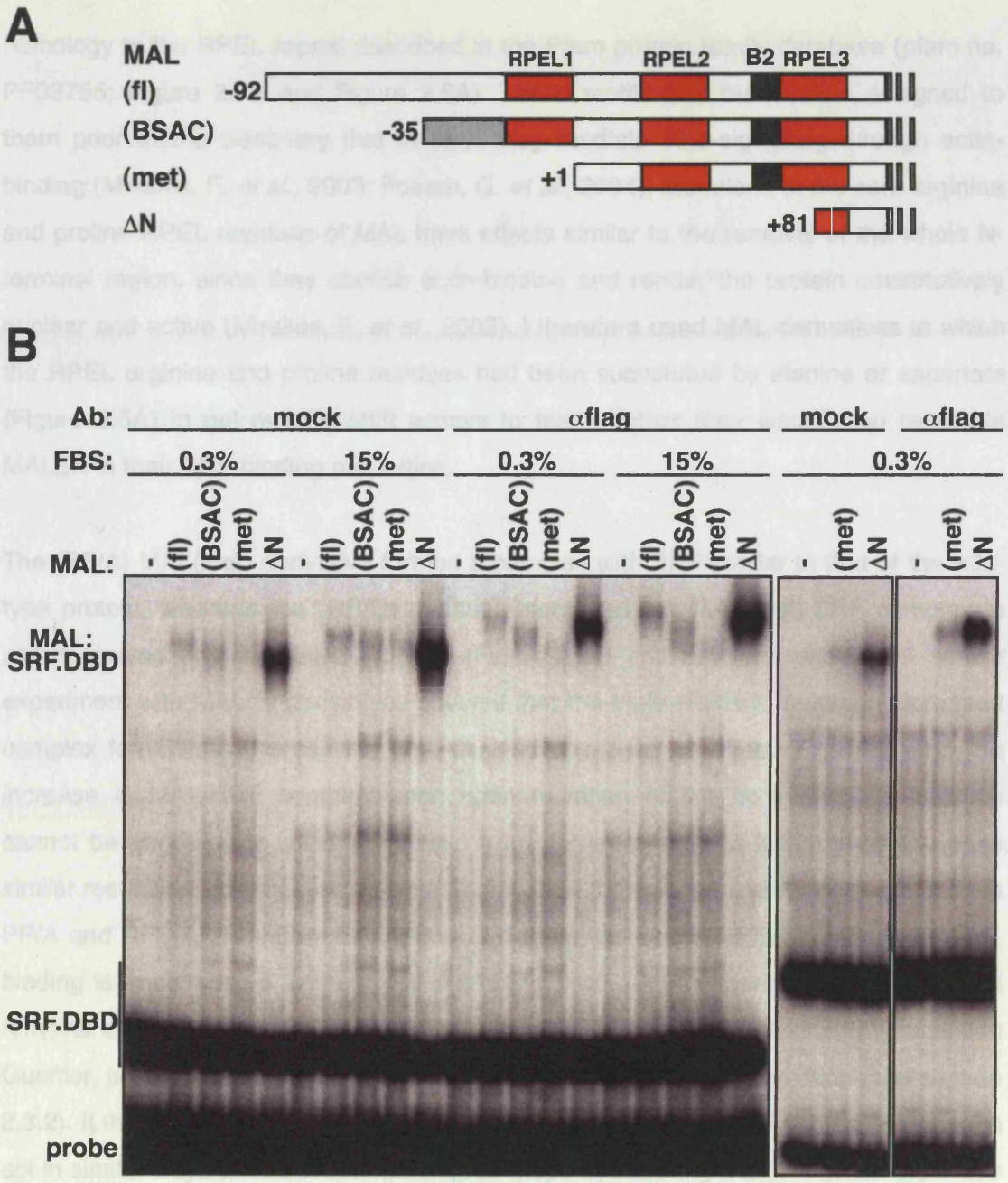


Figure 2.4. N-terminal derivatives of MAL complexed with SRF, are not N-terminally degraded. (A) Structures of the different N-termini of MAL constructs. (B) Supershift of complexes containing MAL N-terminal derivatives and SRF. Gel mobility-shift assays contained the indicated N-terminally Flag-tagged MAL whole-cell extracts, SRF.DBD (residues 132-223 and 133-265 in the left and right panels respectively), c-fos ΔTCF SRE probe and anti-Flag antibody as indicated. The supershift of the MAL(met)/SRF complex is shown as an extra panel on the right for clarity.

homology to the RPEL repeat described in the Pfam protein family database (pfam no. PF02755; Figure 2.1A and Figure 2.5A). These motifs had no function assigned to them prior to the discovery that in MAL they mediate Rho-signalling through actin-binding (Miralles, F. *et al.*, 2003; Posern, G. *et al.*, 2004). Mutations at the core arginine and proline RPEL residues of MAL have effects similar to the removal of the whole N-terminal region, since they abolish actin-binding and render the protein constitutively nuclear and active (Miralles, F. *et al.*, 2003). I therefore used MAL derivatives in which the RPEL arginine and proline residues had been substituted by alanine or aspartate (Figure 2.5A) in gel mobility-shift assays to test whether they would also resemble MALΔN in their SRF binding properties.

The (PP/A) MAL(met) derivative formed a complex with SRF similar to that of the wild-type protein, whereas the (RR/D) mutation increased the MAL(met)-SRF complex in serum-starved or –stimulated extracts (Figure 2.5B and data not shown). A similar experiment with MAL(fl) derivatives showed that the triple (RRR/A) mutation increased complex formation, whereas the (RPP/A) derivative had no effect (Figure 2.5B). The increase in MAL-SRF complex seen upon mutation of the core arginine residues cannot be attributed to removal of actin from the protein, since if that were the case similar results would be expected upon mutation of the core proline residues (mutations PP/A and RPP/A). It appears therefore, that the effect of the RPEL mutations in SRF-binding is unconnected to their role in binding actin. Later experiments revealed that removal of the N-terminus of Myocardin, whose affinity for actin is minimal (Sebastian Guettler, personal communication), also increases SRF complex formation (see section 2.3.2). It therefore remains possible that the N-terminal domains of MAL and Myocardin act in similar ways to inhibit SRF binding, perhaps by masking the SRF-binding surface or imposing structural constraints on the MAL-SRF interaction (see Discussion).

2.3 The role of the conserved motifs of MAL and Myocardin in SRF-binding

MRTF family members share high homology in multiple regions (Figure 2.1). In order to identify the regions of MAL mediating the interaction with SRF, MAL constructs were created, in which areas of significant homology to Myocardin were removed. These MAL domain-deletion derivatives were then tested for their ability to bind SRF in gel mobility shift assays. The amounts of complex formed by MAL(fl) and SRF are very low

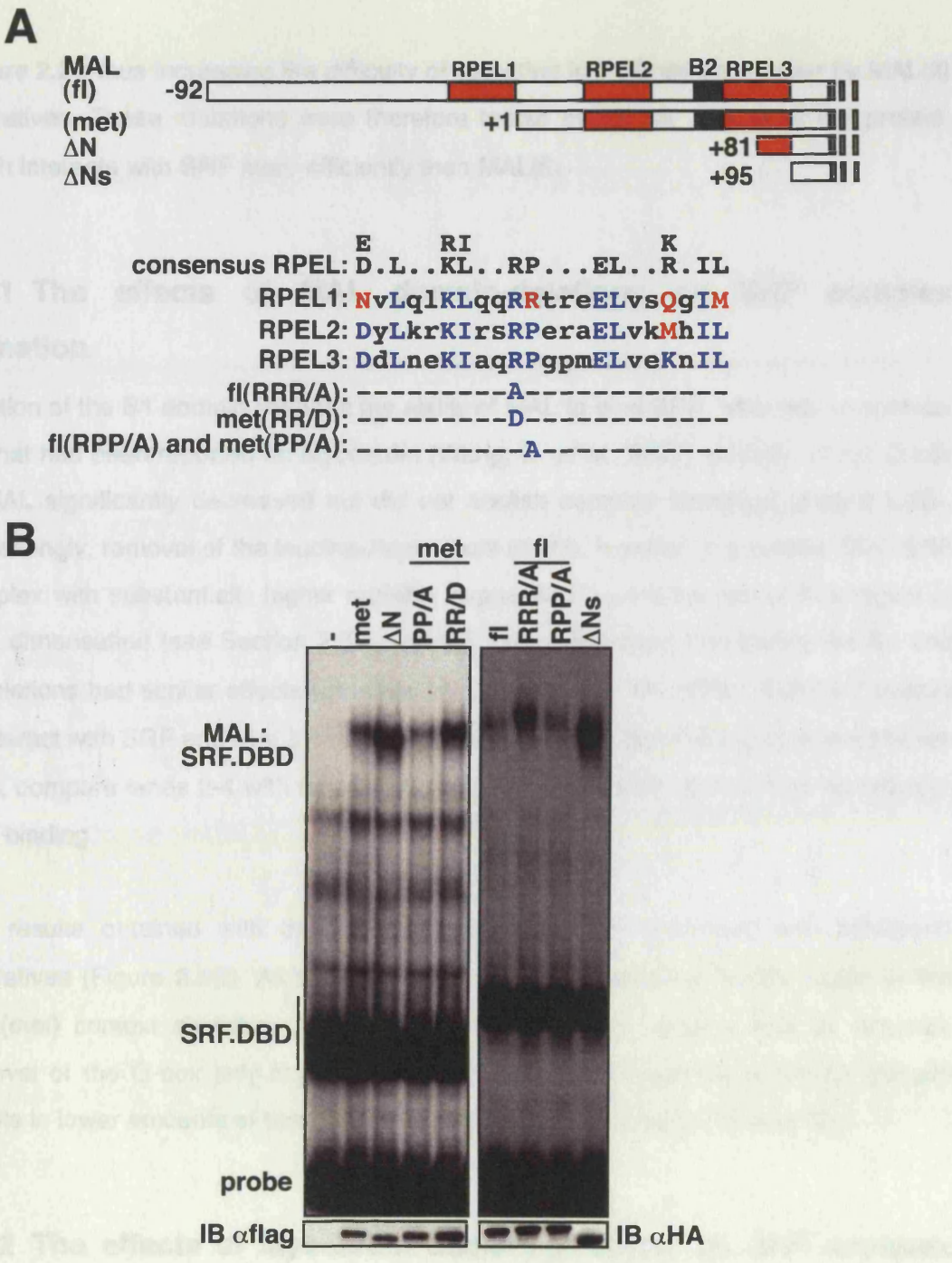


Figure 2.5. Mutations in the RPEL motifs affect the interaction of MAL and SRF. (A) Structures of the different N-termini of MAL constructs. RPEL motifs are shown as red bars. MALΔN includes part of RPEL3 whereas MALΔNs starts directly C-terminal to it. The sequence of the consensus RPEL repeat is shown in black with the sequences of the three RPELs also shown below. Framework residues are in blue and residues divergent from the consensus in red. The point mutations at the framework R and P residues of the motifs are also shown. (B) Complex formation of MAL RPEL mutants with the SRF.DBD. Gel mobility-shift assays contained the indicated N-terminally tagged MAL whole-cell extracts, SRF.DBD (residues 132-223 and 120-265 in the left and right panels respectively), and c-fos ΔTCF SRE probe. Expression levels of the MAL derivatives were confirmed by immunoblotting.

(Figure 2.2), thus increasing the difficulty of detecting low-affinity complexes by MAL(fl) derivatives. These mutations were therefore tested in the ΔN context of the protein, which interacts with SRF more efficiently than MAL(fl).

2.3.1 The effects of MAL domain-deletions on SRF complex formation

Deletion of the B1 domain inhibited the ability of MAL to bind SRF, whereas, in contrast to what had been reported on Myocardin (Wang, D. *et al.*, 2001), deletion of the Q box of MAL significantly decreased but did not abolish complex formation (Figure 2.6B). Interestingly, removal of the leucine-zipper motif of MAL resulted in a weaker MAL-SRF complex with substantially higher mobility, suggesting a possible role of this region in MAL dimerisation (see Section 2.3.4). ΔNΔLZ MAL derivatives harbouring the B1 and Q deletions had similar effects with their ΔN counterparts, with MALΔNΔB1ΔLZ unable to interact with SRF and MALΔNΔQΔLZ forming a weaker, fast-moving complex (Figure 2.6B, compare lanes 2-4 with lanes 7-9). Removal of the SAP domain had no effect in SRF binding.

The results obtained with the MALΔN constructs were confirmed with MAL(met) derivatives (Figure 2.6C). As observed with MALΔN, removal of the B1 region in the MAL(met) context abolishes complex formation (compare lanes 4 and 5), whereas removal of the Q-box only impairs it (lanes 6 and 7) and deletion of the LZ domain results in lower amounts of complex with increased mobility (lanes 11 and 12).

2.3.2 The effects of Myocardin domain-deletions on SRF complex formation

Analysis of the equivalent Myocardin derivatives revealed significant differences in the SRF-binding properties of the two proteins. Although the sizes of the MAL and Myocardin ORFs are comparable, the Myocardin-SRF complex displayed an inherently higher mobility compared to that of MAL-SRF and appeared analogous to the MALΔNΔLZ-SRF complex (Figure 2.7B, compare lane 1 with lanes 6 and 7). In contrast to the MALΔLZ derivatives, deletion of the LZ domain of Myocardin did not change the mobility of the complex (Figure 2.7B, lane 4), suggesting that the properties of the MAL LZ region are not shared by that of Myocardin. Deletion of either the B1 and Q domains

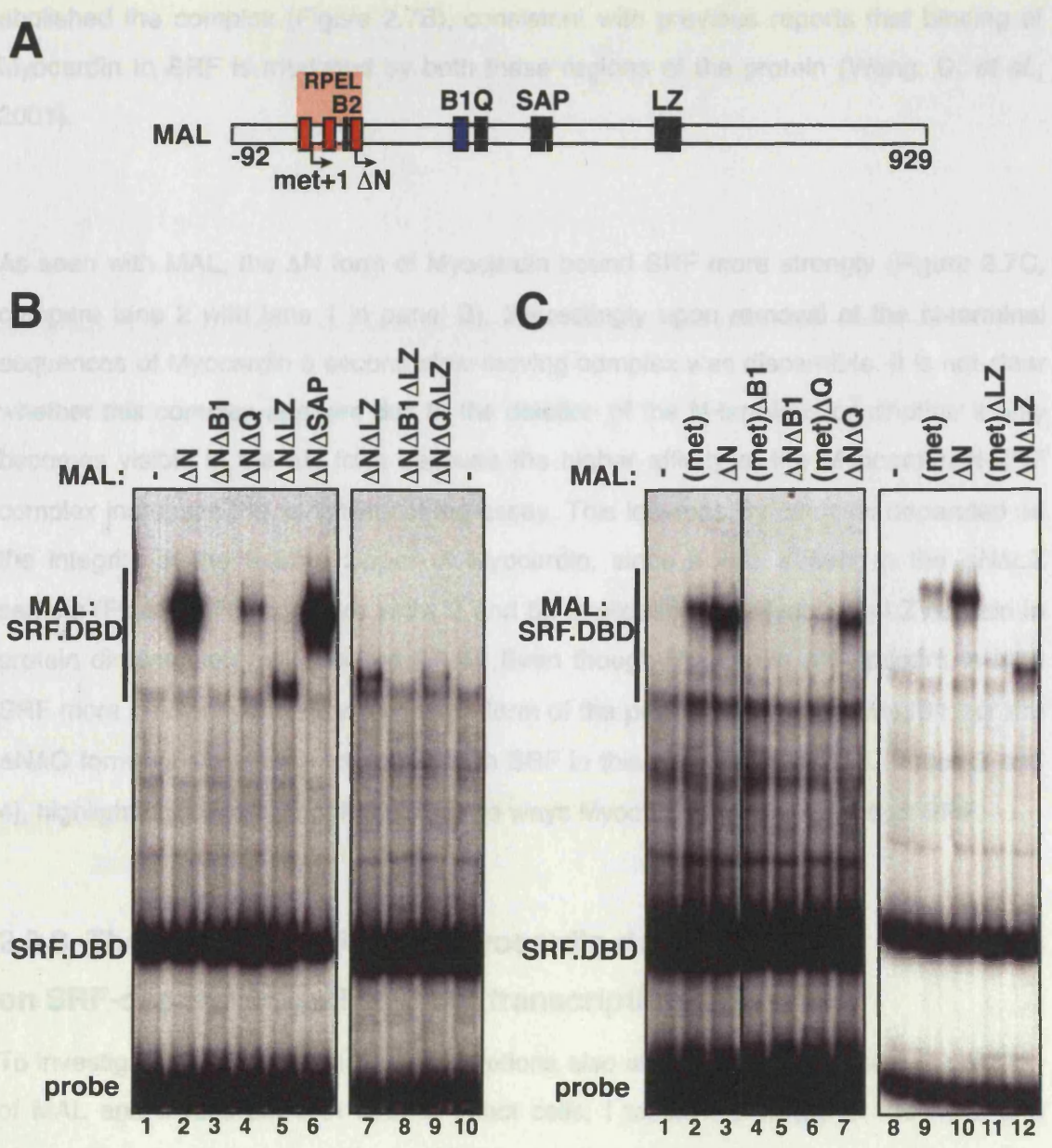


Figure 2.6. The role of the MAL conserved domains in SRF complex formation. (A) Diagram of MAL. Conserved motifs are shown as boxes and the MAL truncations to positions +1 (met) and +81 (ΔN) are indicated. (B) Complex formation between MALΔN deletion mutants and SRF. Gel mobility-shift assays contained the indicated MAL whole-cell extracts, SRF.DBD (133-265) and c-fos ΔTCF probe. (C) Complex formation between MAL(met) deletion mutants and SRF. Gel mobility-shift assays as in B.

abolished the complex (Figure 2.7B), consistent with previous reports that binding of Myocardin to SRF is mediated by both these regions of the protein (Wang, D. *et al.*, 2001).

As seen with MAL, the ΔN form of Myocardin bound SRF more strongly (Figure 2.7C, compare lane 2 with lane 1 in panel B). Interestingly upon removal of the N-terminal sequences of Myocardin a second slow-moving complex was discernible. It is not clear whether this complex appears due to the deletion of the N-terminus or whether it only becomes visible in the ΔN form because the higher affinity of the MyocardinΔN-SRF complex increases the sensitivity of the assay. This low-mobility complex depended on the integrity of the leucine zipper of Myocardin, since it was absent in the ΔNΔLZ sample (Figure 2.7C, compare lanes 2 and 5), implicating the Myocardin LZ domain in protein dimerisation (see Section 2.3.4). Even though Myocardin ΔN appears to bind SRF more efficiently than the full-length form of the protein, neither the ΔNΔB1 nor the ΔNΔQ forms of Myocardin interacted with SRF in this assay (Figure 2.7C, lanes 3 and 4), highlighting a possible difference in the ways Myocardin and MAL contact SRF.

2.3.3 The effects of MAL and Myocardin domain deletion derivatives on SRF-dependent activation of transcription

To investigate whether these domain deletions also affected the functional cooperation of MAL and Myocardin with SRF in intact cells, I tested the ability of the MAL and Myocardin constructs to activate SRF-dependent transcription in a reporter gene assay. Use of the MALΔN constructs allowed direct evaluation of their contribution to transcriptional activation without the interference of endogenous MAL, since MALΔN is constitutively nuclear and active (Miralles, F. *et al.*, 2003).

Deletion of the MAL B1 region failed to activate the reporter and removal of the LZ domain impaired reporter activity in accordance with the effects of these constructs on complex formation (Figure 2.8B). Removal of the Q-box however, did not have a significant effect on reporter gene activation. Similar results were described by Cen *et. al.* (Cen, B. *et al.*, 2003).

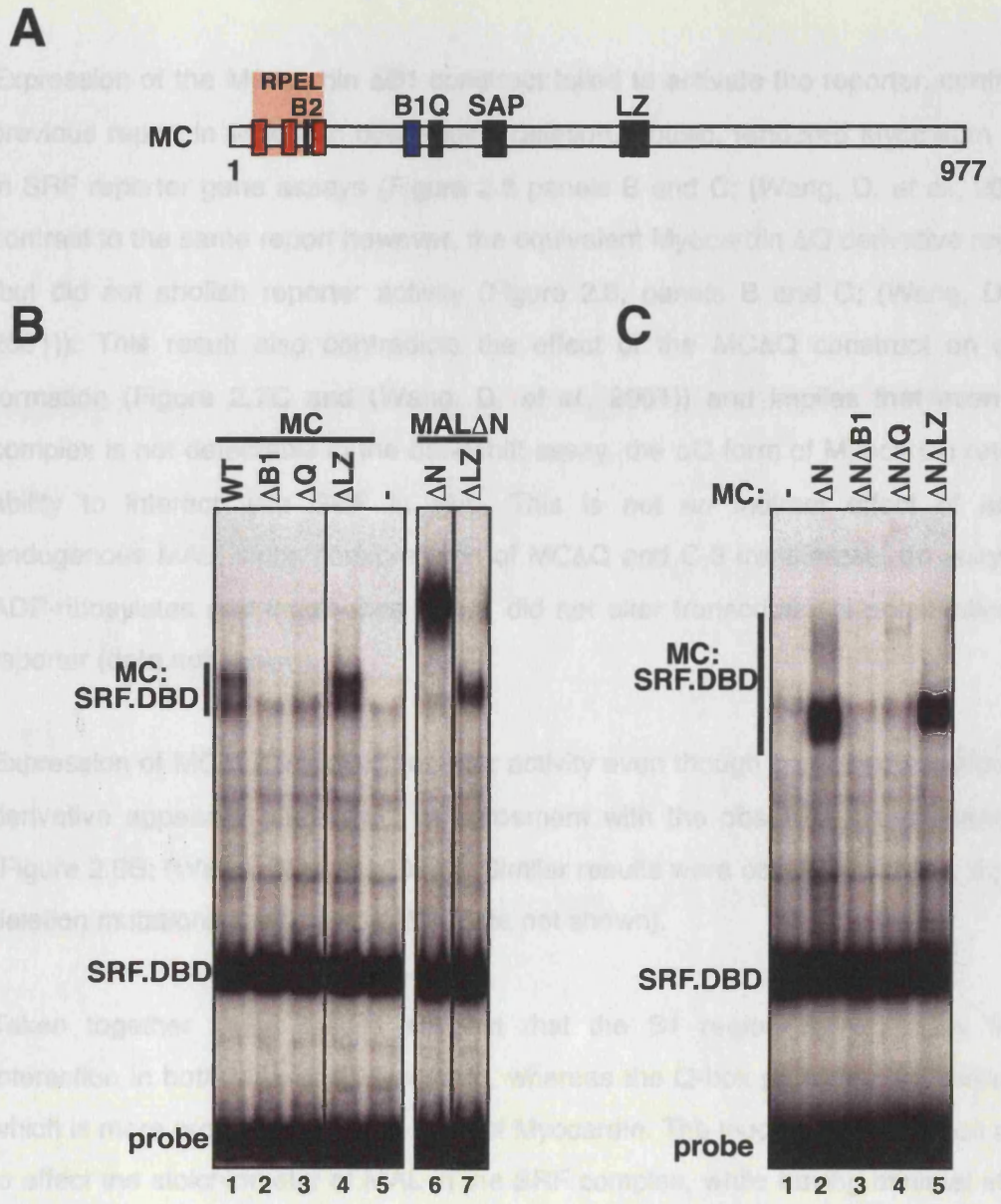


Figure 2.7. The role of the Myocardin conserved domains in SRF complex formation. (A) Diagram of Myocardin. Conserved motifs are shown as boxes. (B) Complex formation between Myocardin deletion mutants and SRF. Gel mobility-shift assays contained the indicated MAL whole-cell extracts, SRF.DBD (133-265) and c-fos ΔTCF probe. The complexes formed by MALΔN and MALΔNΔLZ are also shown for comparison. (C) Complex formation between MyocardinΔN deletion mutants and SRF. Gel mobility-shift assays as in B.

Expression of the Myocardin ΔB1 construct failed to activate the reporter, confirming a previous report in which an overlapping deletion, Δbasic, rendered Myocardin inactive in SRF reporter gene assays (Figure 2.8 panels B and C; (Wang, D. *et al.*, 2001)). In contrast to the same report however, the equivalent Myocardin ΔQ derivative reduced but did not abolish reporter activity (Figure 2.8, panels B and C; (Wang, D. *et al.*, 2001)). This result also contradicts the effect of the MCΔQ construct on complex formation (Figure 2.7C and (Wang, D. *et al.*, 2001)) and implies that even though complex is not detectable in the bandshift assay, the ΔQ form of Myocardin retains the ability to interact with SRF in vivo. This is not an indirect effect of activating endogenous MAL, since coexpression of MCΔQ and C-3 transferase, an enzyme that ADP-ribosylates and inactivates RhoA, did not alter transcriptional potentiation of the reporter (data not shown).

Expression of MCΔLZ impaired reporter activity even though complex formation by this derivative appeared unaffected, in agreement with the observations of Wang *et. al.* (Figure 2.8B; (Wang, Z. *et al.*, 2003)). Similar results were obtained with the equivalent deletion mutations in Myocardin ΔN (data not shown).

Taken together these results suggest that the B1 region is necessary for SRF interaction in both MAL and Myocardin, whereas the Q-box performs an auxiliary role, which is more pronounced in the case of Myocardin. The leucine-zipper region appears to affect the stoichiometry of MAL in the SRF complex, while having minimal effects in Myocardin (see also section 1.3.4), but influences reporter gene activation by both proteins.

2.3.4 The role of the leucine-zipper domain in MAL and Myocardin

The results described in the previous section implicate the leucine zipper domains of MAL and possibly Myocardin in protein dimerisation. Protein association through leucine-zipper domains relies on the interaction of two amphipathic α-helices to form a hydrophobic interface. This is achieved by a slight overtwisting of the helices, so that the structure repeats itself after two helical turns or 7 amino-acids (Alber, T., 1992; Baxevanis, A. D. *et al.*, 1993). The residues of each heptad repeat are designated ***a,b,c,d,e,f,g*** (Figure 2.9A). Residues ***a*** and ***d*** form the dimerisation interface of the zipper and are usually hydrophobic, with leucines most commonly found at position ***d***.

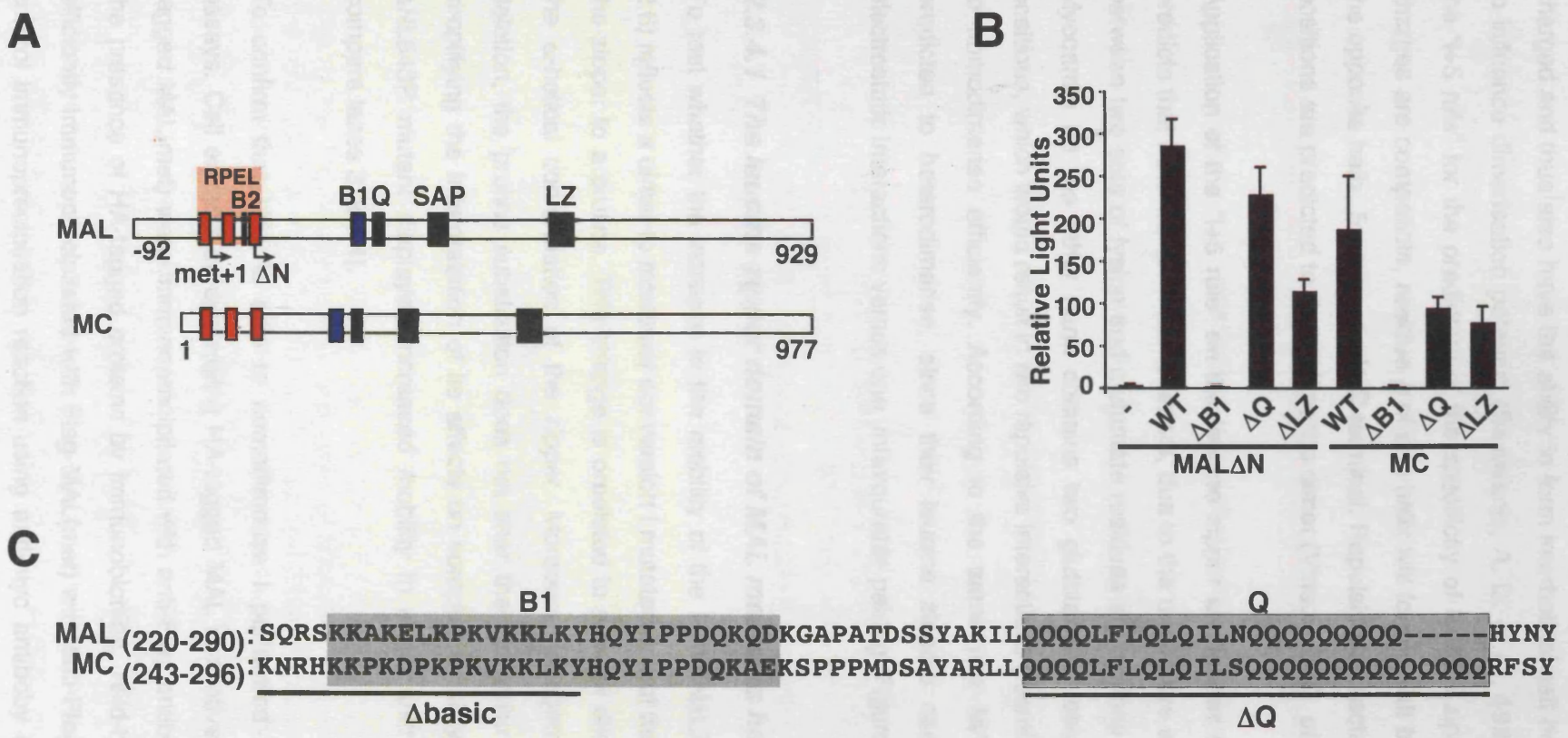


Figure 2.8. Effects of MAL and Myocardin deletion mutations on transcriptional activation through SRF. (A) Diagrams of MAL and Myocardin. Conserved motifs are shown as boxes and the MALΔN truncation to position +81 is indicated. (B) Activation of the 3DA.luc SRF reporter gene by MALΔN and Myocardin deletion mutants. NIH3T3 cells were transfected with 50ng of the indicated plasmids and maintained in 0.3% FBS until assayed for luciferase activity. (C) The B1 and Q regions of MAL and Myocardin. The sequences removed in the ΔB1 and ΔQ deletion mutations are boxed in grey. The black lines indicate the Myocardin Δbasic and ΔQ deletions used by Wang. et. al. (Wang et. al. 2001).

Residues ***e*** and ***g*** flank the hydrophobic core of the zipper. These residues are often charged and thus also have the ability to form interhelical salt bridges, which is thought to influence dimerisation potential (Baxevanis, A. D. *et al.*, 1993). This is the basis of the "i+5 rule" for the prediction of the specificity of leucine zipper partners: when the charges are compatible, residue ***g*** of one helix will form a salt bridge with residue ***e*** on the opposite helix, 5 amino-acids C-terminal. Repulsive interactions between these two positions are predicted to destabilise the dimer (Vinson, C. R. *et al.*, 1993).

Application of the "i+5 rule" on the leucine zipper sequences of MAL and Myocardin predicts that MAL will form homodimers, due to the productive electrostatic interactions between two sets of lysine and glutamate residues at positions ***g*** and ***e*** (Figure 2.9B). Myocardin on the other hand contains two glutatamate residues at key specificity positions, which would result in two repulsive interactions potentially rendering it unable to homodimerise efficiently. According to the same rule MAL and Myocardin are predicted to heterodimerise, since their leucine zippers can form two productive electrostatic interactions versus one unfavourable pairing (Figure 2.9B).

2.3.4.1 The leucine zipper domain of MAL mediates homodimerisation

To test whether the increase in the mobility of the MALΔNΔLZ-SRF complex (Figure 2.6) reflects a dimer to monomer conversion I mutated one of the conserved leucines of the zipper to a proline. This change is predicted to prevent dimerisation by disrupting the α-helical conformation of the zipper. Moreover, in contrast to the LZ domain deletion, the proline substitution does not alter the molecular weight of the protein, simplifying the interpretation of its effects on complex mobility. Like MALΔNΔLZ, the ΔNL543P mutant displayed increased mobility in a gel shift assay (Figure 2.10B, compare lanes 3 and 4).

To confirm that MAL is able to homodimerise I performed co-immunoprecipitation assays. Cell extracts co-expressing HA-tagged MAL derivatives with or without Flag-tagged MAL(met) were immunoprecipitated with anti-Flag antibodies and analysed for the presence of HA-tagged proteins by immunoblotting. Wild-type HA-MAL(met) was efficiently immunoprecipitated with Flag-MAL(met) with anti-Flag antibody, but not in a control immunoprecipitation reaction using anti-myc antiboby (Figure 2.10C, lanes 2 and 3). This interaction was severely decreased when HA-MAL(met)ΔLZ was used,

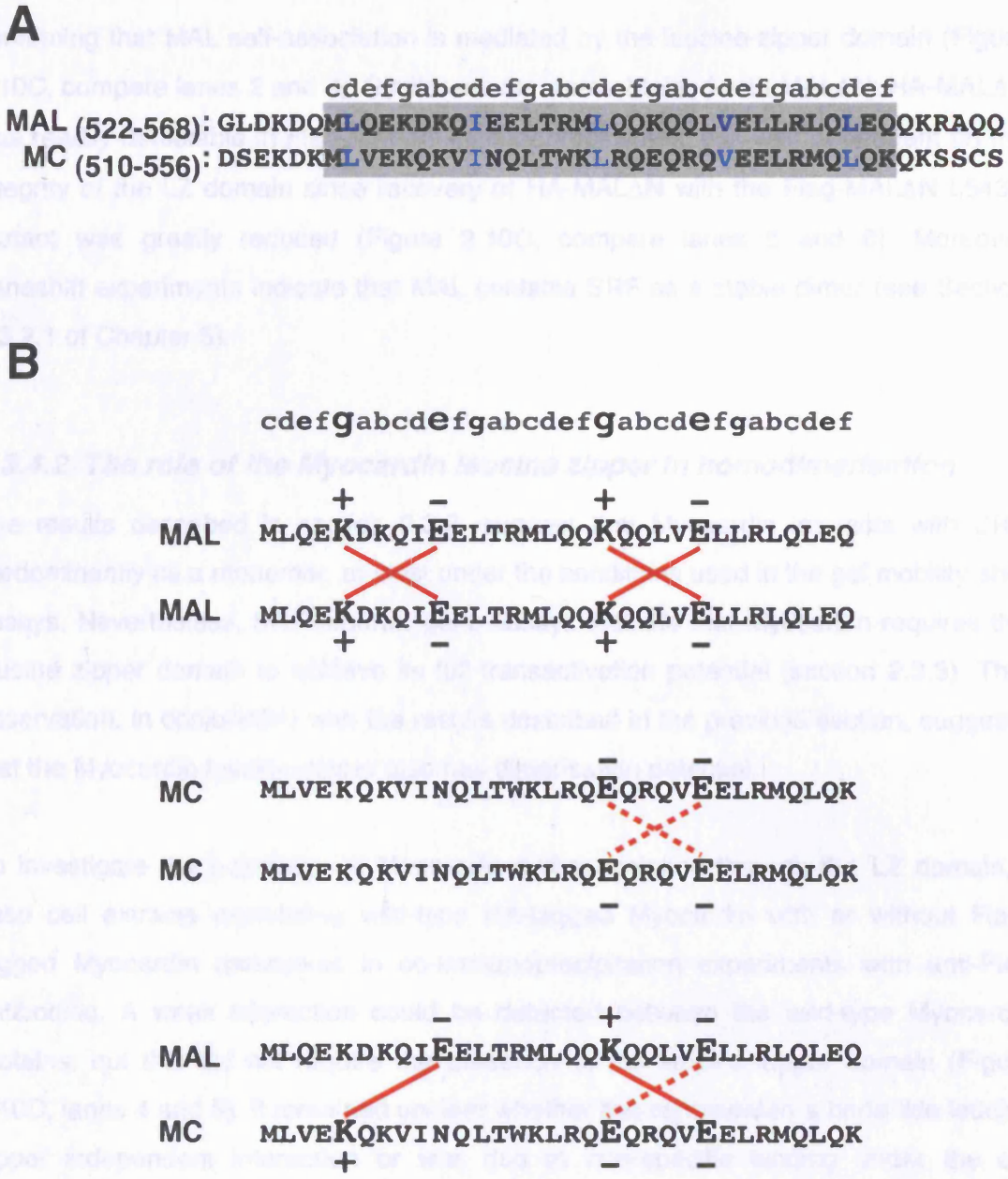


Figure 2.9. The leucine-zipper domains of MAL and Myocardin. (A) Sequences of the MAL and Myocardin LZ motifs. The top line shows the standard LZ nomenclature. Residues at position ***d*** predicted to contribute to the hydrophobic dimerisation interface are shown in blue. (B) Application of the "i+5" rule of leucine zipper dimerisation specificity on the sequences of MAL and Myocardin. Residues at positions ***e*** and ***g*** that could be participating in interhelical electrostatic interactions are shown in large characters and their charge is indicated. Putative productive and repulsive electrostatic interactions between these residues are indicated by solid and dashed red lines respectively. According to this rule MAL is predicted to form homodimers, and also heterodimers with Myocardin, whereas Myocardin homodimerision is not favoured.

confirming that MAL self-association is mediated by the leucine-zipper domain (Figure 2.10C, compare lanes 2 and 4). Similar results were obtained with MALΔN: HA-MALΔN was readily detectable in Flag-MALΔN immunoprecipitates; this was dependent on the integrity of the LZ domain since recovery of HA-MALΔN with the Flag-MALΔN L543P mutant was greatly reduced (Figure 2.10C, compare lanes 5 and 6). Moreover bandshift experiments indicate that MAL contacts SRF as a stable dimer (see Section 5.3.2.1 of Chapter 5).

2.3.4.2 The role of the Myocardin leucine zipper in homodimerisation

The results described in section 2.3.2, suggest that Myocardin interacts with SRF predominantly as a monomer, at least under the conditions used in the gel mobility-shift assays. Nevertheless, SRF reporter gene assays indicate that Myocardin requires the leucine zipper domain to achieve its full transactivation potential (section 2.3.3). This observation, in conjunction with the results described in the previous section, suggests that the Myocardin leucine-zipper also has dimerisation potential.

To investigate the possibility of Myocardin self-associating through the LZ domain, I used cell extracts expressing wild-type HA-tagged Myocardin with or without Flag-tagged Myocardin derivatives in co-immunoprecipitation experiments with anti-Flag antibodies. A weak interaction could be detected between the wild-type Myocardin proteins, but this did not require the presence of the leucine zipper domain (Figure 2.10D, lanes 4 and 5). It remained unclear whether this represented a bona fide leucine zipper independent interaction or was due to non-specific binding under the co-immunoprecipitation conditions used. Hence, these experiments failed to show Myocardin associating with itself through the leucine zipper domain, in contrast to another report in which Myocardin could be weakly immunoprecipitated with itself in a leucine zipper dependent manner (Wang, Z. *et al.*, 2003). The discrepancy between these results could be due to the sensitivity/stringency of the methods used: it is conceivable that the self-association of Myocardin is too weak or unstable to be easily detectable under the co-immunoprecipitation conditions used.

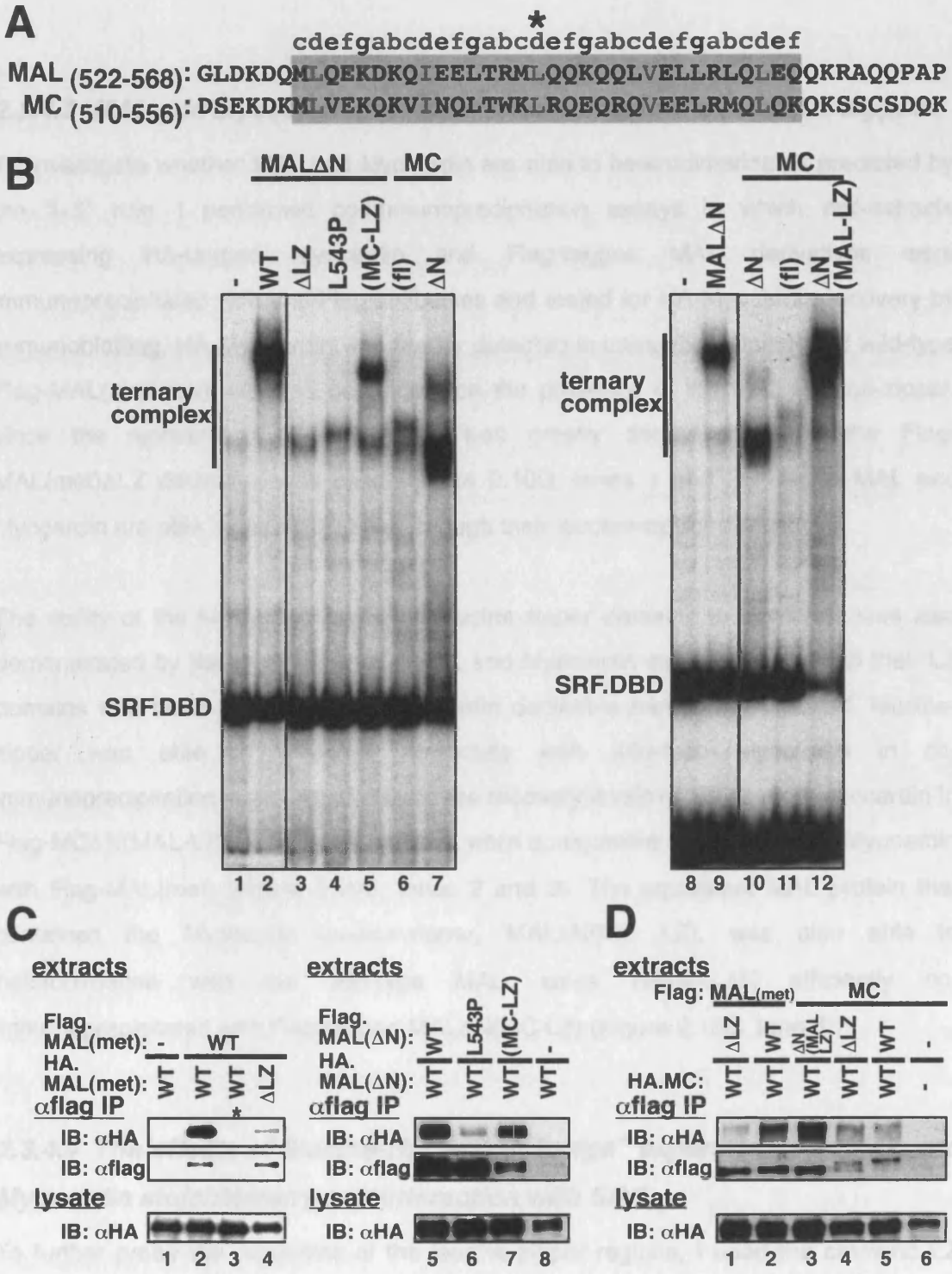


Figure 2.10. The role of the leucine-zipper domain of MAL and Myocardin. (A) The sequences of the MAL and Myocardin LZ motifs are boxed in grey. The top line shows the standard LZ nomenclature. Residues at position ***d*** predicted to contribute to the hydrophobic dimerisation interface are shown in blue. An asterisk indicates the position of the MAL L543P mutation. (B) Effects of MAL and Myocardin LZ mutations in complex formation with SRF. Gel mobility-shift assays contained the indicated MAL and Myocardin whole-cell extracts, SRF.DBD (133-265) and c-*fos* ΔTCF probe. (C) MAL dimerises through the LZ domain. Extracts expressing Flag- and HA-MAL derivatives were immunoprecipitated with anti-Flag beads (anti-myc in lane 3 as a negative control, indicated by asterisk) and analysed by immunoblotting. (D) Myocardin can heterodimerise with MAL. Extracts expressing HA-tagged Myocardin and Flag-MAL and -Myocardin derivatives were immunoprecipitated with anti-Flag beads and analysed by immunoblotting.

2.3.4.3 MAL and Myocardin heterodimerise through their leucine zippers

To investigate whether MAL and Myocardin are able to heterodimerise as predicted by the "i+5" rule, I performed co-immunoprecipitation assays in which cell-extracts expressing HA-tagged Myocardin and Flag-tagged MAL derivatives were immunoprecipitated with anti-Flag antibodies and tested for HA-Myocardin recovery by immunoblotting. HA-Myocardin was readily detected in immunoprecipitates of wild-type Flag-MAL(met) and this was dependent on the presence of the MAL leucine-zipper, since the recovery of HA-Myocardin was greatly decreased when the Flag-MAL(met)ΔLZ derivative was used (Figure 2.10D, lanes 1 and 2). Hence MAL and Myocardin are able to heterodimerise through their leucine-zipper domains.

The ability of the MAL and Myocardin leucine-zipper domains to associate was also demonstrated by the use of chimeric MAL and Myocardin constructs in which their LZ domains had been exchanged. A Myocardin derivative harbouring the MAL leucine-zipper was able to efficiently associate with wild-type Myocardin in co-immunoprecipitation experiments, where the recovery levels of HA-tagged Myocardin in Flag-MCΔN(MAL-LZ) immunoprecipitates, were comparable to those of HA-Myocardin with Flag-MAL(met) (Figure 2.10D, lanes 2 and 3). The equivalent MAL protein that contained the Myocardin leucine-zipper, MALΔN(MC LZ), was also able to heterodimerise with the wild-type MAL, since HA-MALΔN efficiently co-immunoprecipitated with Flag-tagged MALΔN(MC LZ) (Figure 2.10C, lane 6).

2.3.4.4 The effects of leucine-zipper "exchange" experiments in MAL and Myocardin stoichiometry and interaction with SRF

To further probe the properties of the leucine-zipper regions, I used the chimeric LZ MAL and Myocardin derivatives in gel-mobility shift assays with SRF. Whole-cell extracts expressing MCΔN(MAL-LZ) formed high amounts of a slow-moving complex with SRF (Figure 2.10B, lane 12). This complex comigrated with MALΔN-SRF instead of the MCΔN-SRF complex (Figure 2.10B, compare lane 12 with lanes 9 and 10), indicating that the exchange of the Myocardin leucine-zipper for that of MAL enhances the ability of Myocardin to dimerise.

The MAL leucine-zipper chimera had a surprising effect on SRF-complex formation: when used in gel mobility-shift assays, extracts expressing MALΔN(MC-LZ) formed two complexes with SRF, one corresponding to the wild-type MAL-SRF complex and one comparable to that formed by the ΔLZ mutant (Figure 2.10B, compare lane 5 with lanes 2 and 3). This effect is similar to the two complexes formed by MCΔN with SRF, with the important difference that whereas MCΔN forms higher amounts of the fast-moving complex, in the case of MALΔN(MC LZ) formation of the low-mobility complex is more efficient (Figure 2.10B, compare lanes 5 and 7). Thus this result implies that the ability of the Myocardin leucine-zipper to self-associate is influenced by the protein context (MAL or Myocardin). Since the complexes detected in the bandshift assays depend on both the dimerisation properties of the MRTFs and their interactions with SRF itself, this could be reflecting fundamental differences in the ways MAL and Myocardin interact with SRF or the presence of other sequences in these proteins able to promote or inhibit dimerisation.

Taken together, the data described in this section broadly agree with the "i+5" rule predictions on the dimerisation specificities of the MAL and Myocardin leucine zippers. MAL contacts SRF as a dimer and this property is dependent on the integrity of its leucine zipper domain. In contrast Myocardin appears unable to form stable homodimers and contacts SRF as a monomer at least as measured by bandshift experiments. Nevertheless several lines of evidence point to a limited homodimerisation ability of the leucine zipper of this protein. Finally MAL and Myocardin are able to heterodimerise through their leucine zipper domains.

2.4 MAL has the SRF-DNA binding properties of the Rho-controlled SRF cofactor

The mechanism by which Rho-signalling regulates SRF was not unravelled until the discovery of the actin-MAL link (Miralles, F. *et al.*, 2003; Sotiropoulos, A. *et al.*, 1999) and even at present many aspects of the regulation of this system remain unclear. Nevertheless, a wealth of information was compiled over time as to what is required for SRF to respond to this pathway and a model emerged in which Rho-induced changes in actin dynamics regulate an SRF accessory factor that binds the SRF DNA-binding domain and activates transcription of a subset of SRF target genes ((Gineitis, D. *et al.*, 2001; Hill, C. S. *et al.*, 1994; Hill, C. S. *et al.*, 1995b; Sotiropoulos, A. *et al.*, 1999); see

also Introduction). Moreover a set of predictions were made as to what is required for this cofactor to interact with SRF and activate transcription (see below). Having established that MAL associates with SRF, I sought to examine whether the SRF-binding properties of MAL correlate with those expected of the Rho-actin pathway coactivator.

2.4.1 The N-terminus of the SRF DNA-binding domain is required for binding and activation by MAL

The different DNA-binding specificities conferred on the various MADS box proteins by the N-terminal sequences of their core domains had been previously exploited to create an altered DNA-specificity SRF form, SRF.M2 that contains the N-terminal extension and αl helix of the SRF-related yeast protein MCMI (Figure 2.11A; (Hill, C. S. *et al.*, 1993)). As a result SRF.M2 recognises MCMI-specific DNA sequences, while retaining weak affinity for the wild-type SRF CArG box (Figure 2.11B). Functional studies showed that SRF.M2 does not respond to the Rho pathway when bound to its cognate DNA site (FOS.LM; (Hill, C. S. *et al.*, 1993; Hill, C. S. *et al.*, 1994)), and overexpression of this altered-specificity mutant also fails to restore activity to a weak SRE site that is inducible in vivo by overexpression of the wild-type SRF (Hill, C. S. *et al.*, 1994). Furthermore SRF does not respond to serum stimulation when tethered to DNA via a heterologous DNA-binding domain (Hill, C. S. *et al.*, 1994). Thus the prediction was made that transcriptional activation through the Rho-pathway depended on an SRF-binding cofactor, which required the intact N-terminal region of the SRF DNA-binding domain bound to the authentic SRE site in order to interact with SRF. This cofactor would thus be unable to bind SRF.M2 on its cognate site.

To investigate whether this was true for MAL I tested its ability to bind SRF.M2 in gel mobility-shift assays. The DNA probes used were constructed from sequences of the c-*fos* promoter: FOS.WT contained the wild-type c-*fos* CArG box, while the FOS.M probe had the CArG box sequences mutated to convert its specificity from wild-type SRF to SRF.M2 (Figure 2.11B and (Hill, C. S. *et al.*, 1993)).

Whole-cell extracts expressing MALΔN formed a discrete complex with endogenous SRF on the FOS.WT probe, but not on the mutated FOS.M site, which is not recognised by wild-type SRF (Figure 2.12A, lanes 1 and 2, see also Figure 2.2A).

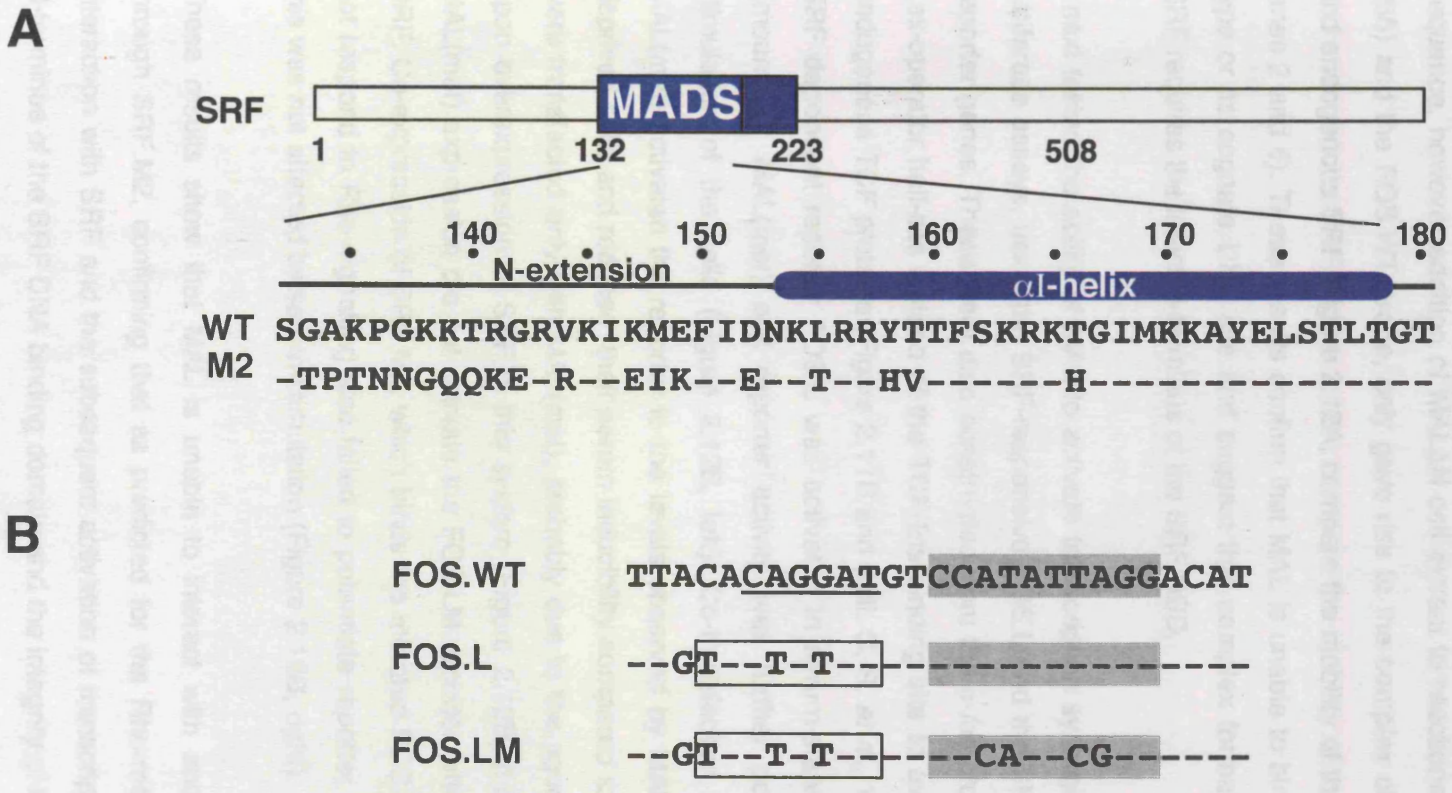


Figure 2.11 The altered DNA-specificity SRF.M2 and c-fos derived SRE derivatives. (A) Schematic representation of SRF, with the MADS box containing DNA-binding domain shown in blue. The sequence and secondary structure elements of the N-terminal half of the DBD and the substitutions in the SRF.M2 derivative are shown below. (B) Sequences of the intact *c-fos* SRE (FOS.WT) and its derivatives. In FOS.L the LexA half-site (boxed) replaces the Ets motif (underlined). The SRF CArG box is shaded. The base changes in FOS.LM that convert it to a high affinity MCMI site are shown.

Addition of excess SRF(1-265), gave rise to a similar complex of increased mobility on the FOS.WT site, but not on the mutated FOS.M site (Figure 2.12A lanes 3,4 and 7,8). Furthermore MALΔN failed to interact with SRF.M2 on its cognate DNA site FOS.M (Figure 2.12A, lanes 9,10). SRF.M2 has a weak affinity for the wild-type CArG box sequence, however addition of MALΔN cell extract to reactions containing SRF.M2(1-265) and the FOS.WT probe, only gave rise to the complex characteristic of MALΔN and endogenous SRF (Figure 2.12A, compare the mobility of the MAL-SRF complex in lanes 2 and 6). These results confirm that MAL is unable to bind SRF.M2 on the wild-type or its cognate DNA site and suggest that complex formation between MAL and SRF requires the intact N-terminus of the SRF.DBD.

I next tested the ability of MAL to activate transcription synergistically with SRF.M2 in luciferase assays, using the SRF-responsive FOS.L and the SRF.M2-specific FOS.LM reporter genes. These were also constructed from the c-*fos* promoter and contained a Lex-operator half-site instead of the TCF Ets binding site to avoid interference by the endogenous TCF proteins (Figure 2.11B and (Hill, C. S. *et al.*, 1993)). As expected the SRF dependent reporter FOS.L was activated in serum-starved cells by increasing amounts of MAL(met) and reporter activity was further potentiated upon serum stimulation of the cells (Figure 2.12B, left). Co-transfection of wild-type SRF with MAL(met) activated the reporter to the levels observed by MAL(met) alone in serum-deprived cells and reduced their serum inducibility compared to their counterparts that were transfected only with MAL(met), probably due to the squelching effect observed upon overexpression of SRF in this system (Figure 2.12B, (Hill, C. S. *et al.*, 1993)). MAL(met) expression did not activate the FOS.LM reporter which cannot interact with SRF. Co-expression of SRF.M2, which binds the mutated FOS.LM sequence but does not respond to Rho-signalling also failed to potentiate reporter activation by MAL and this was not affected by serum-stimulation (Figure 2.12B, right).

These results show that MAL is unable to interact with and activate transcription through SRF.M2, confirming that as predicted for the Rho-responsive cofactor MAL interaction with SRF and the subsequent activation of transcription requires the intact N-terminus of the SRF DNA binding domain and the integrity of the authentic SRF DNA binding site.

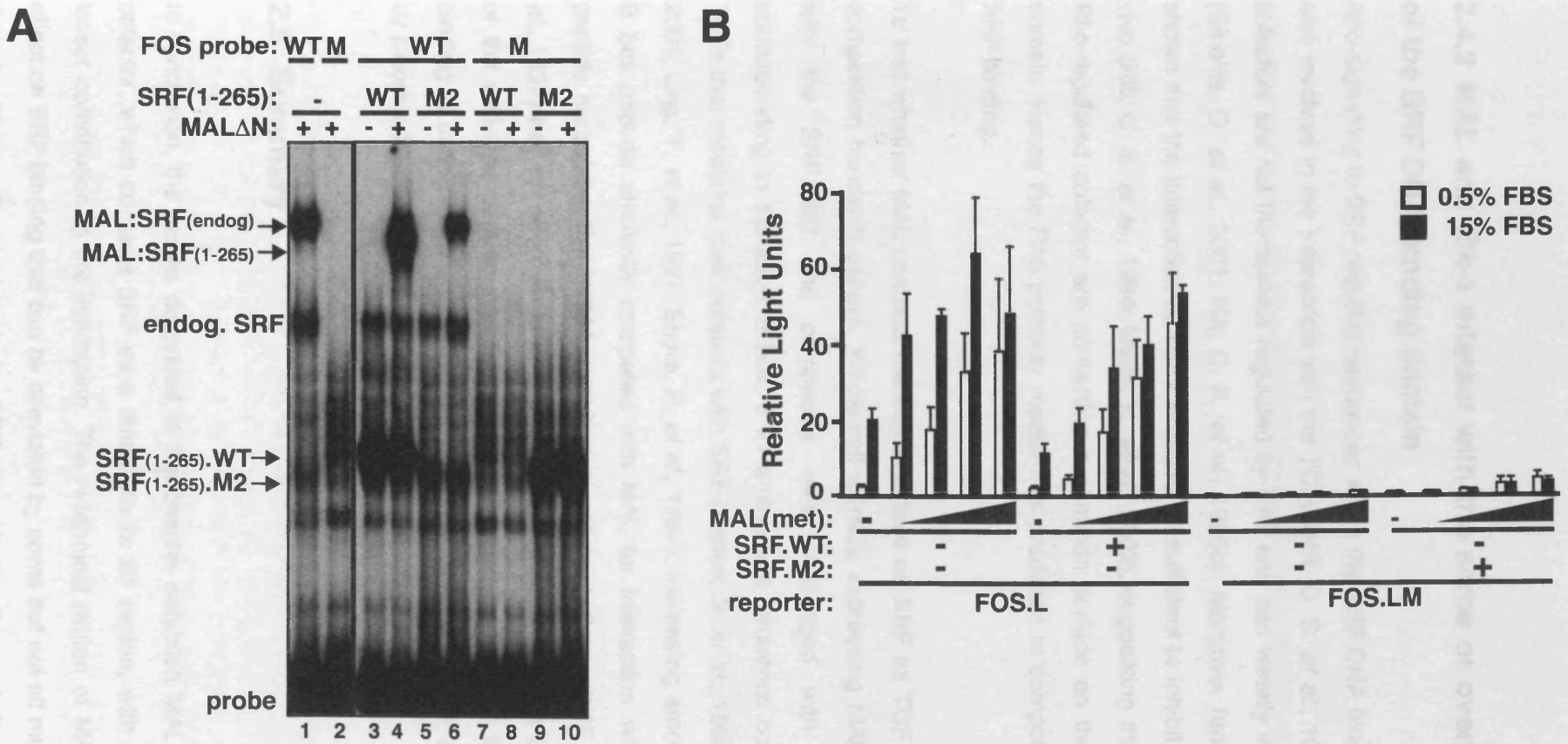


Figure 2.12. Activation of SRF through MAL requires the intact N-terminal sequences of the SRF.DBD. (A) MAL does not bind the altered-specificity SRF.M2. Gel mobility-shift assays contained MALΔN and wildtype SRF(residues 1-265) or SRF.M2(1-265) whole-cell extracts, and c-*fos* ΔTCF wildtype or mutated M probe as indicated. The complexes formed between MALΔN and SRF(1-265) or endogenous SRF are indicated. (B) MAL does not activate the altered-specificity SRF.M2 on its cognate site. NIH3T3 cells were transfected with the FOS.L or FOS.LM reporter plasmids and MAL(met) (0, 5, 50, 100 or 250ng) and SRF.WT or SRF.M2 expression plasmids as indicated. Cells were maintained in 0.5% FBS or stimulated with 15% FBS prior to being assayed for luciferase activity.

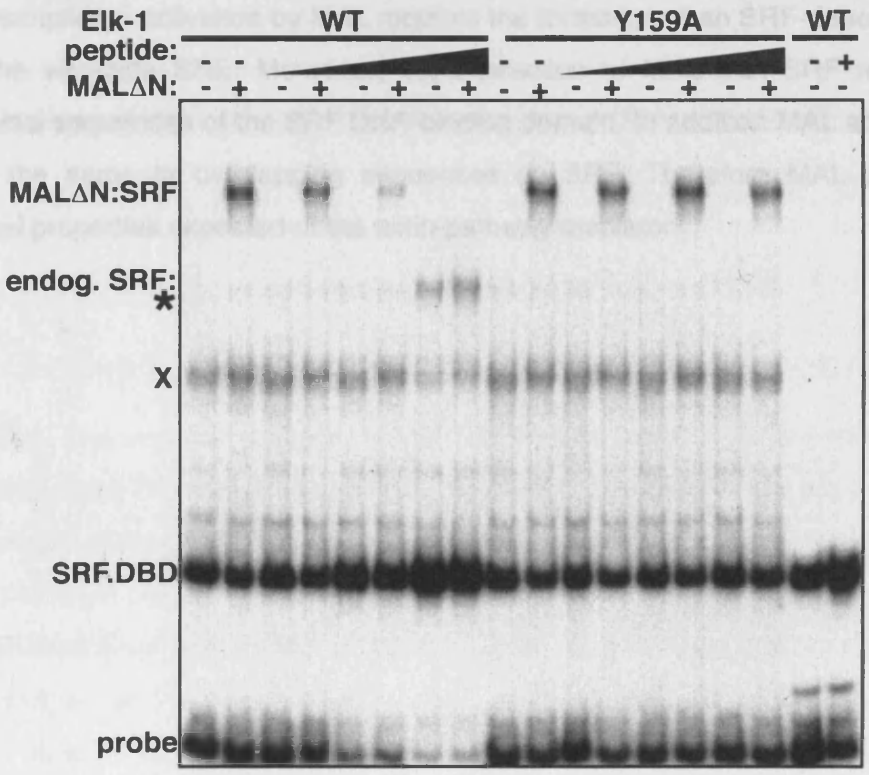
2.4.2 MAL and Elk-1 interact with the same or overlapping regions of the SRF DNA-binding domain

Rho-signalling to SRF requires sequences within the SRF DNA binding domain that are also involved in the interaction with the TCFs (Hill, C. S. *et al.*, 1993), however these cofactors are not themselves regulated by Rho and can weakly interact with SRF.M2 (Gineitis, D. *et al.*, 2001; Hill, C. S. *et al.*, 1995b). Moreover functional studies have shown that the interaction of TCFs with SRF is sufficient to inhibit the Rho pathway in vivo (Hill, C. S. *et al.*, 1994; Murai, K. *et al.*, 2002), suggesting that the TCFs and the Rho-regulated cofactor are contacting a common surface on the SRF DNA-binding domain. Hence the Rho pathway mediator is predicted to compete with the TCFs for SRF binding.

To test whether MAL contacts the same surface on SRF as TCF I performed binding competition bandshift assays. Whole-cell extracts expressing MALΔN were incubated with the SRF.DBD and complexes were challenged with synthetic peptides corresponding to the Elk-1 B box, a 21 amino-acid sequence conserved in all three TCFs that mediates their contacts with SRF (Dalton, S. *et al.*, 1992; Hassler, M. *et al.*, 2001; Ling, Y. *et al.*, 1997; Shore, P. *et al.*, 1994). Increasing amounts of the wild-type B box peptide efficiently competed with MAL for interaction with SRF, whereas a peptide harbouring the Y159A substitution, which abolishes SRF binding (Ling, Y. *et al.*, 1997) had no effect on MAL-SRF complex formation (Figure 2.13). The specificity of the wild-type peptide was further demonstrated by the fact that it increased the binding of endogenous and core SRF to the probe (Figure 2.13, complexes indicated by asterisks).

2.5 Summary

In conclusion, the results described in this chapter establish MAL as an SRF-binding cofactor, which contacts SRF as a dimer via its B1 region, with the Q-box having a lesser contribution to the interaction. The N-terminal region of MAL has an inhibitory effect on SRF binding that can be alleviated by some but not all mutations in the RPEL domain. Myocardin also contacts SRF through the B1 and Q regions, in this case



Elk-1 WT: **SSRNEYMRSGLYSTFTIQSLQ**
Elk-1 Y159A: **----------A---------**

Figure 2.13. MAL and Elk-1 compete for the same binding surface on SRF. Peptide competition assay of the MAL-SRF complex with Elk-1 B-box peptides. Gel mobility-shift assays contained MALΔN whole-cell extract, SRF.DBD (residues 132-265), c-*fos* probe and 0, 0.8, 4 and 20μM wildtype or mutant Elk-1 peptides as indicated. Peptide sequences are shown below. An asterisk indicates the specific SRF-peptide interaction revealed by the increased binding of endogenous SRF and of SRF(132-265). X indicates the position of a complex that is depleted by the wild-type peptide. It is not clear whether this represents the SRF.DBD bound to endogenous TCF or whether it is a non-specific effect seen due to the limiting amounts of probe at high peptide concentrations.

however the Q-box appears to have a more prominent role. The N-terminus of Myocardin also limits SRF interaction, and Myocardin appears unable to form stable dimers. Nevertheless, its leucine-zipper domain is likely to retain limited homodimerisation ability, as judged by its effects on reporter gene activation. Transcriptional activation by MAL requires the formation of an SRF-dependent complex on the wild-type SRE. Moreover, the interaction of MAL with SRF requires the N-terminal sequences of the SRF DNA-binding domain. In addition MAL and TCF interact with the same or overlapping sequences on SRF. Therefore MAL possesses two critical properties expected of the actin-pathway mediator.

3 The SRF-binding surface of MAL

3.1 Aims

Having established that MAL is an SRF cofactor that utilises its B1 and Q regions to interact with SRF, I next sought to characterise the SRF-binding surface of MAL. In this chapter I describe the detailed analysis of the B1 and Q regions of MAL and their respective contributions to SRF complex formation.

3.2 Analysis of the B1 box requirement for SRF-binding

I initially focused my analysis on the B1 region, since it is essential for complex formation (see Chapter 2, section 2.3). Small deletions were introduced into the B1 box of MALΔN (Figure 3.1A) and the ability of the resulting MAL derivatives to interact with the SRF DNA-binding domain was assessed in gel mobility-shift assays. Removing the basic N-terminal sequences of the B1 region slightly increased complex formation, whereas a small internal deletion (Δ230-235) reduced but did not abolish it (Figure 3.1B, lanes 4 and 5; this will be futher analysed in Chapter 5 and the Discussion). Deletion of the C-terminal sequences had the same effect as removing the entire region (MALΔB1) and completely abolished complex formation (Figure 3.1B, lane 6), implying that the main SRF-contacting residues were present in the C-terminal half of B1.

3.2.1 Alanine-scanning mutagenesis of the B1 box

In order to investigate which residues of the B1 box form the SRF binding surface I employed an alanine-scanning mutagenesis approach, which had been previously successfully applied on both Elk-1 and SRF, in order to map their respective interacting surfaces (Ling, Y. *et al.*, 1997; Ling, Y. *et al.*, 1998).

Alanine point-mutations were introduced individually at every position of the B1 region, thus removing side chains that potentially participate in intermolecular interactions, while at the same time avoiding major structural perturbations. Gel mobility shift assays of the point mutants identified a short stretch of predominantly hydrophobic residues

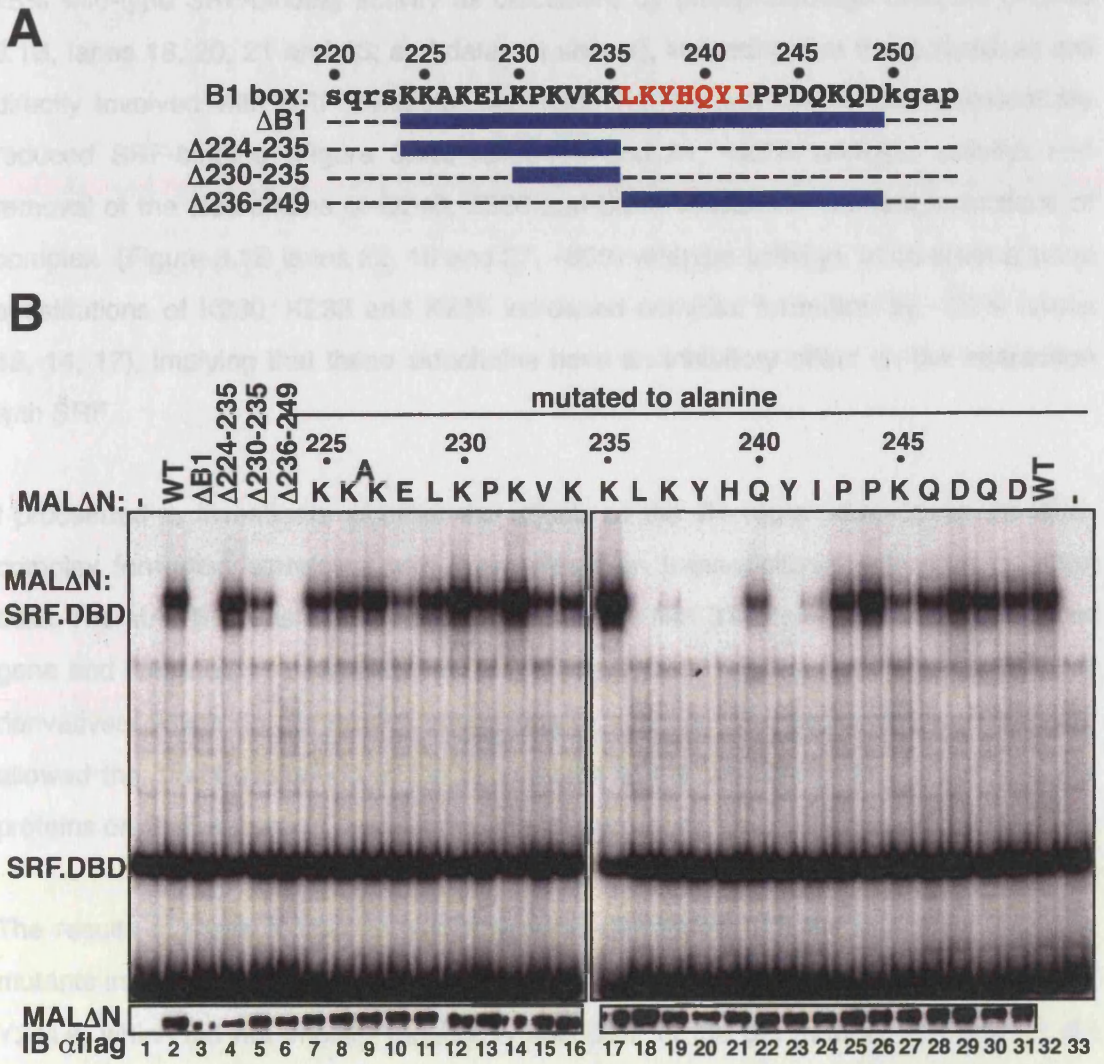


Figure 3.1. Identification of MAL B1 region residues critical for SRF binding. (A) The sequence of the B1 region is shown in capital letters. The 7-residue critical sequence identified by the alanine scan is highlighted in red. Deletion mutations are indicated by blue bars. (B) Deletion and alanine-scanning mutations in the MAL B1 region affect complex formation. Gel mobility-shift assays contained Flag-tagged MALΔN B1 region derivatives, recombinant SRF.DBD (residues 133-265), and the c-*fos* ΔTCF SRE probe. Bottom, quantitation of MALΔN construct expression by immunoblotting.

that are required for interaction with SRF. Alanine substitutions at positions L236, Y238, H239 and Y241, abolished complex formation with SRF, since they displayed <5% wild-type SRF-binding activity as calculated by phosphorimage analysis (Figure 3.1B, lanes 18, 20, 21 and 23, and data not shown), indicating that these residues are directly involved with SRF contacts. Mutations K237A and I242A also substantially reduced SRF-binding (Figure 3.1B, lanes 19 and 24, ~25% wildtype activity) and removal of the side chains of Q240, E228 and D245 resulted in modest reductions of complex (Figure 3.1B lanes 22, 10 and 27, ~50% wildtype activity). In contrast alanine substitutions of K230, K232 and K235 increased complex formation by ~30% (lanes 12, 14, 17), implying that these sidechains have an inhibitory effect on the interaction with SRF.

I proceeded to investigate whether the effects of the B1 region derivatives on SRF-complex formation correlated with their effects on transcriptional activation in intact cells. The MAL B1 constructs were cotransfected in NIH 3T3 cells with an SRF reporter gene and their ability to stimulate luciferase activity was assessed. The use of MALΔN derivatives, which do not require stimulation to activate SRF-dependent transcription, allowed the direct evaluation of the contribution of the transiently expressed MAL B1 proteins on SRF activation, without the interference of the endogenous MAL.

The results of these functional studies broadly correlated with the effects of the MAL mutants in SRF complex formation. The MALΔN derivatives L236A, Y238A, H239A and Y241A, which did not interact detectably with SRF in the gel mobility-shift assay, did not activate an SRF-controlled reporter gene (Figure 3.2A). The Q240A and I242A substitutions, which reduced SRF complex formation, were able to activate transcription to almost wild-type levels. The effects of these mutations were more pronounced at lower plasmid inputs as shown by titration experiments (Figure 3.2B), implying that although under the expression conditions used, MAL-SRF binding is not limiting for reporter activation, the assay can become easily saturated with higher plasmid amounts. Furthermore the assay also suffers from the squelching effect observed widely in the SRF activation system at high protein expression levels (see Chapter 2; (Hill, C. S. *et al.*, 1993)), obscuring the subtler effects of certain mutants such as I242A. A more marked decrease of SRF activation at lower expression levels was also seen with the K237A substitution (Figure 3.2B), which affects both complex formation with SRF and also MAL nuclear localisation (Figure 3.1B and Section 3.4.1).

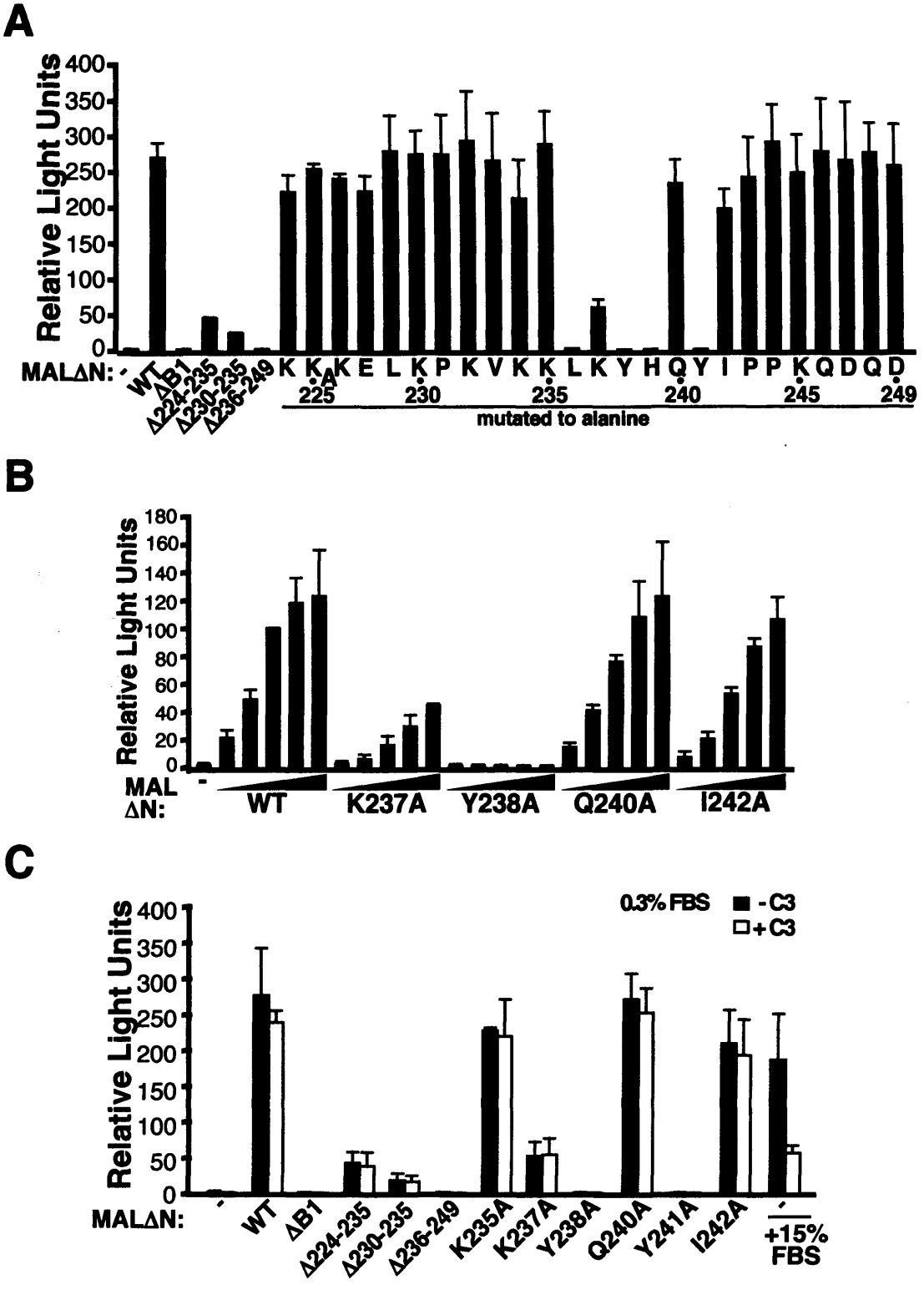


Figure 3.2. Effects of MAL B1 region mutations on SRF reporter gene activation. (A) Activation of the SRF-dependent reporter 3DA.luc by MALΔN B1 region derivatives. NIH3T3 cells were transfected with 50ng of the indicated plasmids and maintained in 0.3% FBS until assayed for luciferase activity. Results were normalised to activity by SRF.VP16 set to 100 and are presented as ±SEM of three independent experiments. (B) Titration of MALΔN B1 derivatives (6ng, 18ng, 55ng, 166ng, and 500ng MALΔN plasmid inputs). Reporter assays were as in A, with results normalised to 55ng input of MALΔN set to 100. (C) Effects of C3-transferase coexpression on SRF-reporter activation by MALΔN B1 region derivatives. Reporter assays were as in A, with 25ng C3-transferase plasmids cotransfected as indicated.

Mutations K234A and K235A also displayed lower SRF reporter gene activation at lower plasmid inputs (data not shown). These derivatives do not reduce SRF complex formation but are defective for nuclear import (data not shown; for nuclear import see Section 3.4.1). The effects of the K237A, K234A and K235A derivatives on nuclear localisation are obscured by high expression levels, which render the localisation of all MAL derivatives pancellular. Thus increased expression of K234A, K235A and K237A results in higher SRF reporter gene activation levels, than those expected for nuclear import defective derivatives. E228A and D245A also potentiated transactivation to wild-type levels, as did the proteins that formed increased amounts of complex with SRF (K230 and K232), and this can probably also be attributed to saturation of the system (Figure 3.2A). Reporter gene activation by deletions Δ224-235, Δ230-235, was reduced to a disproportionately large degree compared to their results on complex formation (compare Figure 3.2A with Figure 3.1B). This however is probably due to the pronounced negative effects these changes have on nuclear accumulation (see section 3.4.1).

To demonstrate that the effects of the MALΔN B1 derivatives on reporter gene activation reflect their specific roles on SRF interaction and do not involve indirect activation of endogenous MAL, they were co-expressed with C3-transferase in luciferase assays (Figure 3.2C). As mentioned previously this enzyme inactivates Rho and inhibits activation of the system by endogenous MAL. C3-transferase expression reduced transactivation of the reporter upon serum stimulation, but had no effect on reporter-gene activation by key MALΔN B1 region derivatives, indicating that these do not act indirectly through the endogenous protein.

3.2.2 Mutagenesis of key B1 residues

Having identified a stretch of residues in the B1 region that abolish or greatly reduce SRF binding, I proceeded to investigate the side chain requirements for MAL-SRF complex formation (Figure 3.3A).

The experiments presented in Chapter 2, demonstrate that MAL and TCF must contact the same or overlapping surfaces on SRF. Furthermore, alanine scanning mutagenesis of the TCF B box established the essential role of an aromatic residue (F162 in Elk-1, and F150 in SAP-1) in SRF-binding, since its bulky aromatic side-chain fits in a deep

hydrophobic pocket on the surface of SRF ((Hassler, M. *et al.*, 2001; Ling, Y. *et al.*, 1997); see also Introduction and Chapter 4).

To test the possibility that one of the tyrosines in the MAL B1-box might play an analogous role, different hydrophobic residues were inserted at positions Y238 and Y241. Substitutions with phenylalanine were expected to increase the binding efficiency of SRF binding or leave it unaffected, since they would allow optimal docking in the pocket (Hassler, M. *et al.*, 2001), whereas non-aromatic substitutions were predicted to inhibit the interaction. Indeed a Y238F substitution increased complex formation considerably, while Y238I and Y238L substantially reduced but did not abolish it, apparently implicating this residue in the hydrophobic pocket interaction (Figure 3.3B, lanes 4-6). Nevertheless, the Y241F mutation also left complex formation unaffected, whereas other mutations at that position had severe effects on SRF binding (Figure 3.3B, lanes 8-10). Thus although these observations establish the importance of aromaticity at positions 238 and 241, they preclude a direct correlation of either residue with the TCF aromatic position (see Discussion).

At position H239 all substitutions had significant effects on SRF interaction, with conversion to phenylalanine reducing it, and H239L, H239T, or H239K virtually abolishing it (Figure 3.3B, lanes 12-15). Hence the planar character of this residue appears to be important for SRF binding. A lysine to arginine substitution at position 237 did not affect complex formation, suggesting that at this position a basic residue is important (Figure 3.3B, lanes 18-19).

I proceeded to investigate the effects of these MALΔN B1 region derivatives in SRF transcriptional activity, by testing their ability to activate an SRF reporter gene. The substitutions at residues Y238, H239, Y241 and K237 affected reporter activity largely in accordance to their effects on complex formation, although for unclear reasons Y238F impaired rather than enhanced reporter activation (Figure 3.3C).

The seven-residue sequence that is essential for SRF binding is identical in all MRTF family members in vertebrates. In contrast a related sequence is found in MAL orthologs in arthropods. To test whether this sequence could mediate SRF binding in the context of the mammalian proteins I transplanted the Drosophila heptamer in the mouse MAL protein (Figure 3.3A). The MAL (D.B1box) derivative bound SRF very

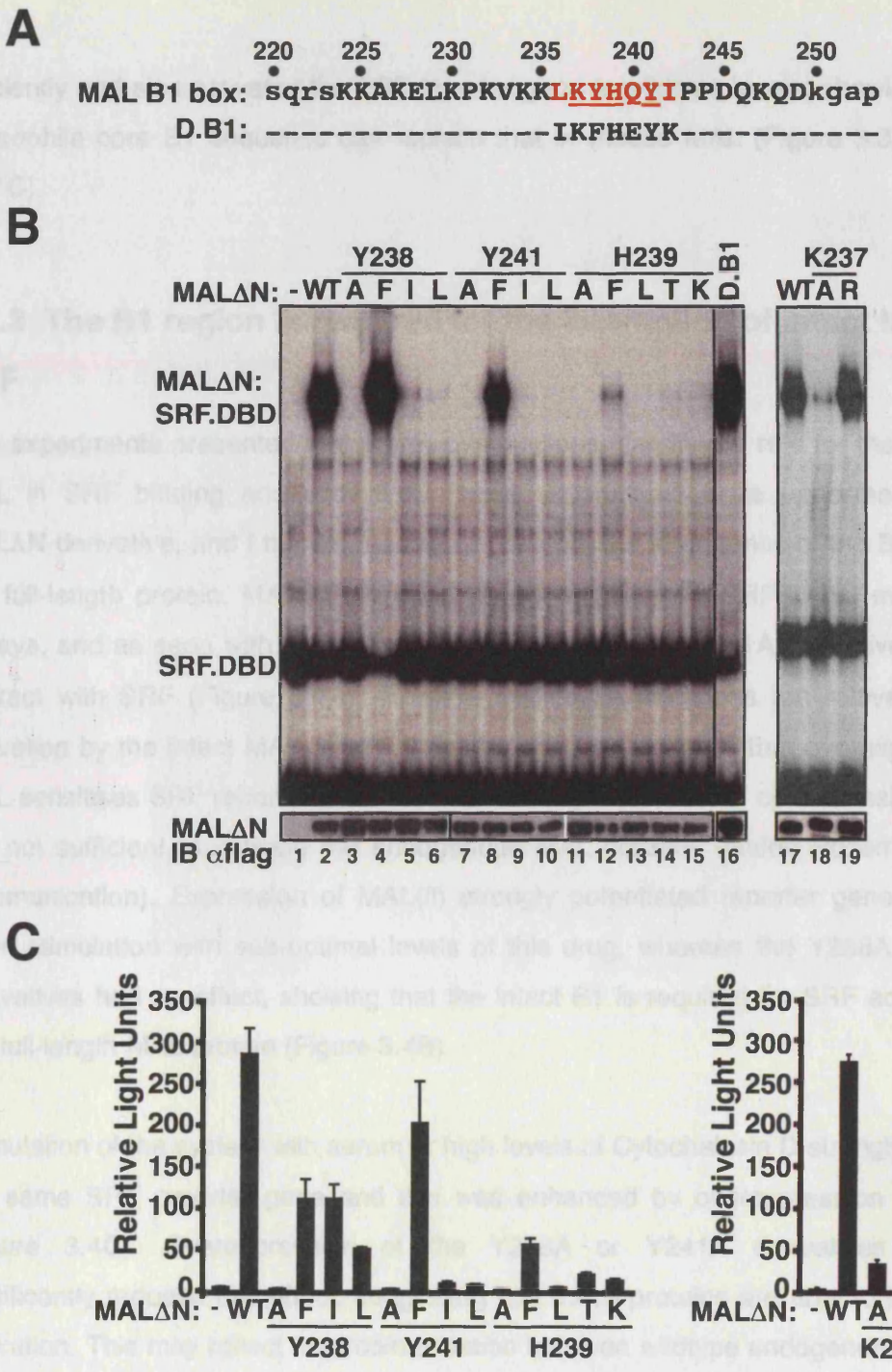


Figure 3.3. Analysis of interactions by residues K237, Y238, H239 and Y241. (A) The sequence of the B1 region is shown in capital letters. The 7-residue critical sequence is highlighted in red with mutated residues underlined. The sequence of the DMAL B1 construct (D.B1) is shown. (B) Gel mobility-shift assays contained Flag-tagged MALΔN B1 region derivatives, recombinant SRF.DBD (residues133-265) (lanes 17-19 contain *in vitro* translated SRF.DBD residues 120-265), and the c-*fos* ΔTCF SRE probe. Bottom, quantitation of MALΔN construct expression by immunoblotting. (C) Activation of the SRF-dependent reporter 3DA.luc by B1 region derivatives. NIH3T3 cells were transfected with 50ng of the indicated plasmids and maintained in 0.3% FBS untill assayed for luciferase activity.

efficiently and also activated the SRF-reporter gene to wildtype levels, showing that the Drosophila core B1 sequence can replace that of mouse MAL (Figure 3.3, panels B and C).

3.2.3 The B1 region is required for the interaction of intact MAL with SRF

The experiments presented in the previous sections identified a role for the B1 box of MAL in SRF binding and activation. These experiments were performed with the MALΔN derivative, and I therefore sought to confirm the importance of the B1 region in the full-length protein. MAL(fl) formed a weak complex with SRF in gel mobility-shift assays, and as seen with MALΔN, the MAL(fl) Y238A and Y241A derivatives failed to interact with SRF (Figure 3.4A). To show that these mutations are relevant in SRF activation by the intact MAL *in vivo*, I exploited the observation that overexpression of MAL sensitises SRF reporter genes to activation by low levels of cytochalasin D, which are not sufficient to activate the endogenous MAL proteins (Guido Posern, personal communication). Expression of MAL(fl) strongly potentiated reporter gene activation upon stimulation with sub-optimal levels of this drug, whereas the Y238A or Y241A derivatives had no effect, showing that the intact B1 is required for SRF activation by the full-length MAL protein (Figure 3.4B).

Stimulation of the system with serum or high levels of Cytochalasin D strongly activated the same SRF reporter gene and this was enhanced by overexpression of MAL(fl) (Figure 3.4C). Overexpression of the Y238A or Y241A derivatives however, significantly reduced this effect, suggesting that these proteins are able to inhibit SRF activation. This may reflect heterodimerisation between wildtype endogenous MAL and the overexpressed mutant derivatives, resulting in weak or unstable MAL-SRF complexes and hence in inefficient transcriptional activation (see also Discussion).

Considered together, these results demonstrate that the intact B1 region is required for potent and efficient activation of SRF by MAL.

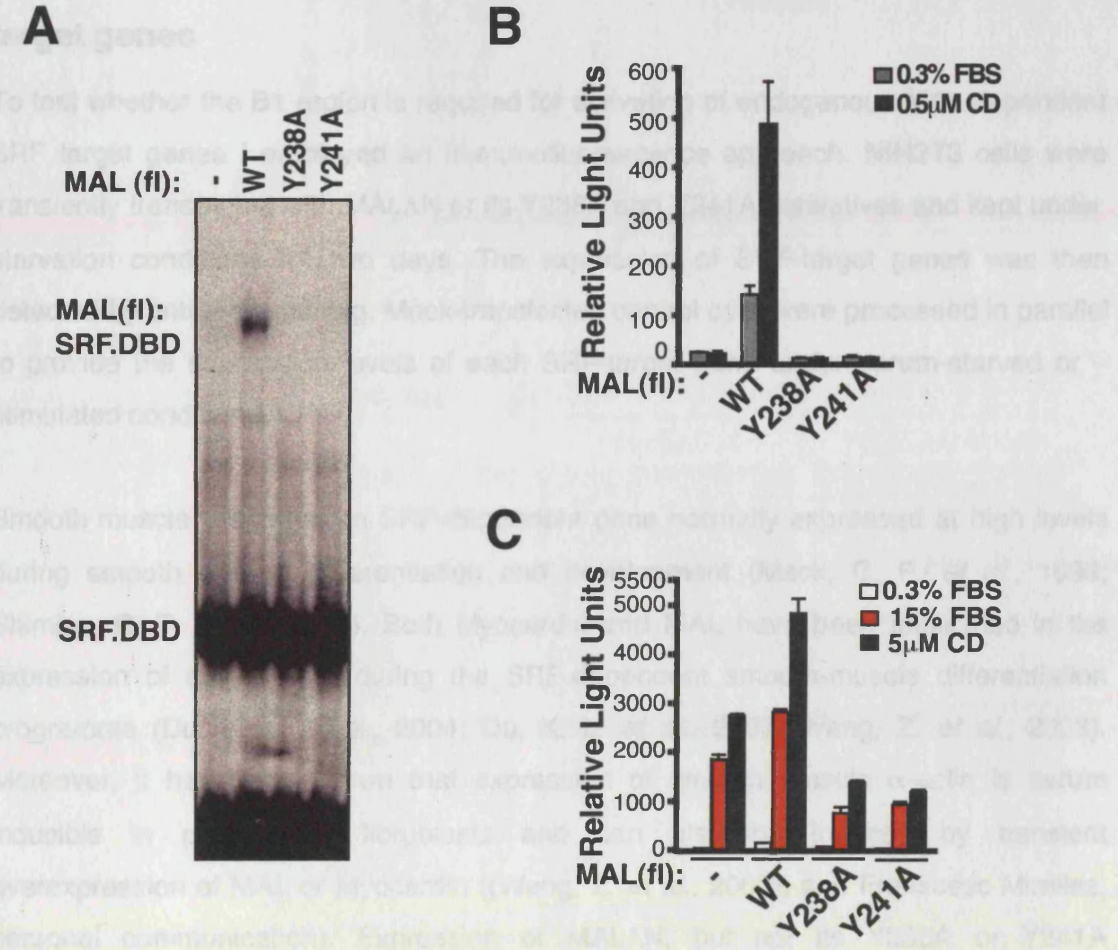


Figure 3.4. The B1 region is necesary for SRF activation by MAL. (A) SRF complex formation by the MAL(fl) Y238A and Y241A derivatives. Reactions contained whole-cell extracts transiently transfected with Flag-tagged MAL(fl) derivatives, *in vitro* translated SRF.DBD (residues 120-265), and the c-*fos* ΔTCF SRE probe. (B) Potentiation of SRF reporter activity induced by low levels of cytochalasin D requires the integrity of the MAL B1 region. NIH3T3 cells were transfected with the SRF-dependent 3DA.luc reporter and 50ng of the indicated plasmids. After 18 hours of serum starvation cells were treated with 0.3% FBS or 0.5μM cytochalasin D for 7 hours as indicated. Results are presented as the average ±SEM of three independent experiments. (C) B1 region mutations that abolish SRF complex formation, inhibit activation of the SRF-dependent reporter 3DA.luc. Reporter assays were as in B.

3.2.4 The B1 region is required for expression of endogenous SRF target genes

To test whether the B1 region is required for activation of endogenous MAL-dependent SRF target genes I employed an immunofluorescence approach. NIH3T3 cells were transiently transfected with MALΔN or its Y238A and Y241A derivatives and kept under starvation conditions for two days. The expression of SRF-target genes was then detected by antibody staining. Mock-transfected control cells were processed in parallel to provide the expression levels of each SRF-target gene under serum-starved or –stimulated conditions.

Smooth-muscle α-actin is an SRF-dependent gene normally expressed at high levels during smooth muscle differentiation and development (Mack, C. P. *et al.*, 1999; Shimizu, R. T. *et al.*, 1995). Both Myocardin and MAL have been implicated in the expression of this marker during the SRF-dependent smooth-muscle differentiation programme (Du, K. L. *et al.*, 2004; Du, K. L. *et al.*, 2003; Wang, Z. *et al.*, 2003). Moreover, it has been shown that expression of smooth muscle α-actin is serum inducible in proliferating fibroblasts and can also be induced by transient overexpression of MAL or Myocardin ((Wang, Z. *et al.*, 2003), and Franscesc Miralles, personal communication). Expression of MALΔN, but not its Y238A or Y241A derivatives, induced high levels of smooth-muscle α-actin expression in serum-deprived cells (Figure 3.5, panels A and B). In contrast MALΔN remained unable to activate the TCF-controlled *egr-1* and c-*fos* genes, which are insensitive to Rho-actin signalling (Figure 3.5, panels A and B; (Gineitis, D. *et al.*, 2001)), indicating that the specificity of target gene expression is maintained under MAL overexpression conditions.

I also tested the effects of MAL overexpression in the SRE.FOS.HA stable cell line. This NIH3T3-derived cell-line harbours an epitope-tagged c-*fos* transcript (HA.FOS), controlled by the synthetic 3D.A promoter, (Mohun, T. *et al.*, 1987) (Alberts, A. S. *et al.*, 1998). This promoter lacks TCF binding sites and this renders HA.FOS expression in this cell-line solely dependent on Rho-signalling. Overexpression of MALΔN in the SRE.FOS.HA cell-line efficiently activated HA.FOS expression, as did serum-stimulation, whereas the Y238A and Y241A mutants had no effect (Figure 3.5, panels A and B). Taken together these data show that the integrity of the B1 region of MAL is

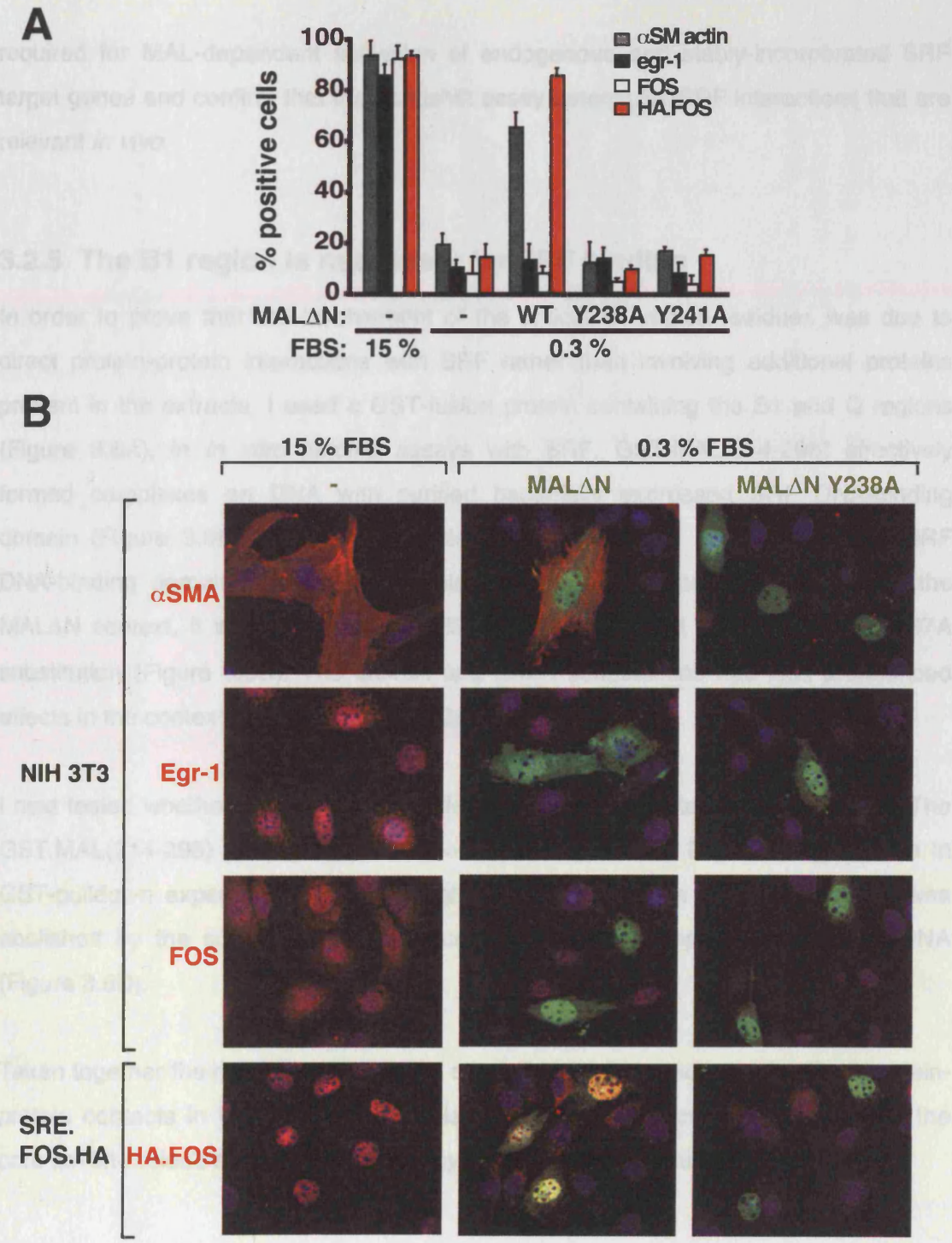


Figure 3.5. Activation of expression of endogenous and stably transfected SRF-dependent genes by MALΔN derivatives. (A) NIH3T3 and NIH3T3(SRE.FOS.HA) cells were transfected with 150ng of the indicated MALΔN plasmids and maintained in 0.3% serum for two days. Where indicated, stimulation with 15 % FBS was for 30 minutes prior to sample processing. Gene expression was scored by immunofluorescence in 100-200 transfected cells in two independent experiments. (B) Expression of SRF-dependent genes induced by serum and MALΔN derivatives. Assays were as in A. Endogenous smooth muscle α-actin, Egr-1 and FOS, and stably transfected HA.FOS (red) were visualised by staining with the appropriate antibodies, MALΔN (green) was visualised with anti-Flag antibody and DNA (blue) with DAPI staining.

required for MAL-dependent activation of endogenous and stably-incorporated SRF target genes and confirm that the bandshift assay detects B1-SRF interactions that are relevant *in vivo*.

3.2.5 The B1 region is necessary for SRF binding

In order to prove that the involvement of the critical B1 region residues was due to direct protein-protein interactions with SRF rather than involving additional proteins present in the extracts, I used a GST-fusion protein containing the B1 and Q regions (Figure 3.6A), in *in vitro* binding assays with SRF. GST.MAL(214-298) effectively formed complexes on DNA with purified bacterially expressed SRF DNA-binding domain (Figure 3.6B). The same complex was formed with *in vitro* translated SRF DNA-binding domain (Figure 3.6C). This interaction was specific, since, as in the MALΔN context, it was abolished by L236A and H239A and reduced by the K237A substitution (Figure 3.6C). The Q240A and I242A substitutions had less pronounced effects in the context of GST.MAL(214-298).

I next tested whether these mutations affect protein-protein interactions with SRF. The GST.MAL(214-298) protein could recover the wildtype SRF DNA-binding domain in GST-pulldown experiments. Even though SRF recovery was very inefficient, it was abolished by the substitutions at residues required for complex formation on DNA (Figure 3.6D).

Taken together the results in this section show that the B1 region makes direct protein-protein contacts in the MAL-SRF complex, and these contacts are mediated by the core seven-residue sequence identified by alanine-scanning mutagenesis.

3.2.6 The B1 region is sufficient for SRF binding

The results presented in the preceding sections are consistent with the view that residues L236 to I242 of the B1 region represent the primary interaction surface of MAL with SRF. To test this directly I used synthetic B1 peptides in binding competition gel mobility-shift assays with the MALΔN-SRF.DBD complex. For that purpose I designed a 21-residue peptide comprising the hydrophobic core region flanked by seven N- and C-terminal residues (Figure 3.7A, peptide **A**; residues 229-249). This peptide was able

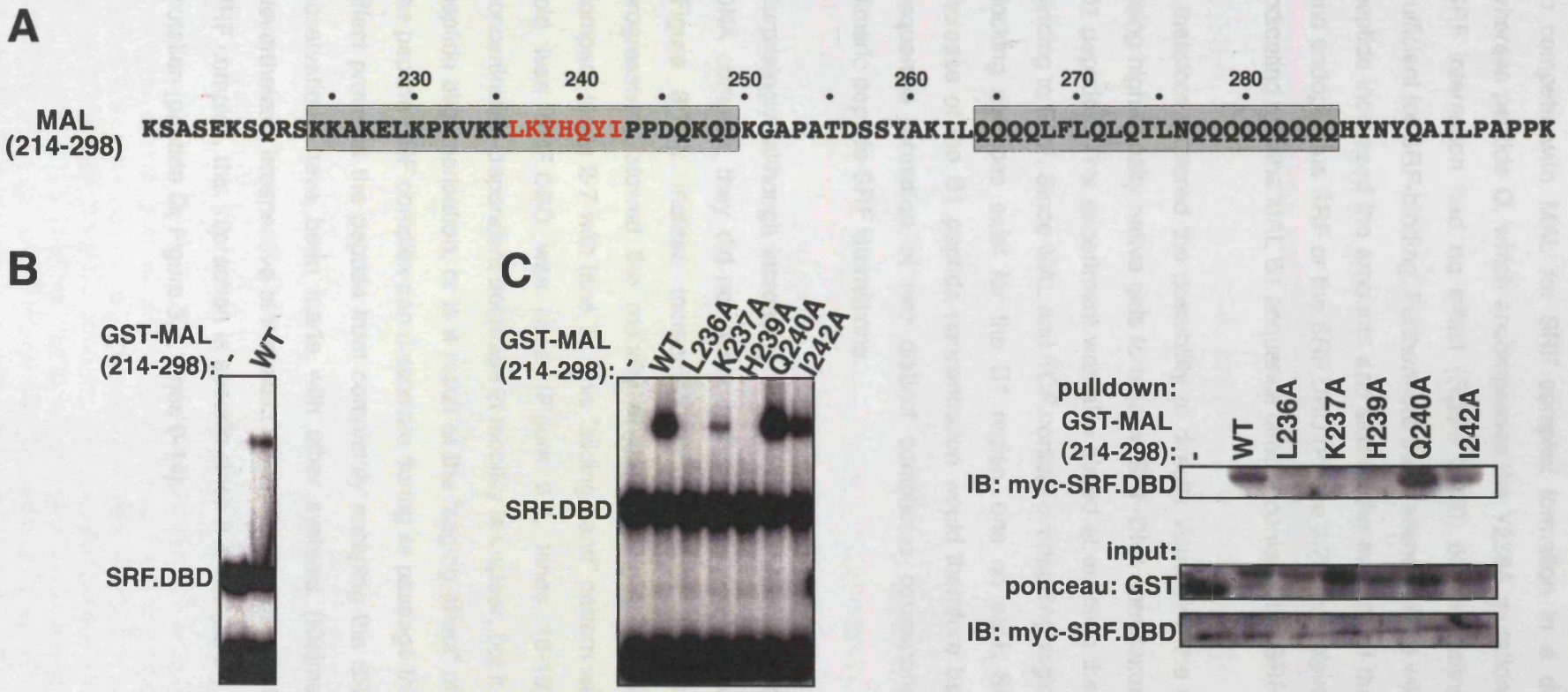


Figure 3.6. The B1 region mediates protein-protein interactions with SRF. (A) The sequence of the MAL(214-298) GST-fusion protein. The B1 and Q regions are boxed in grey. The critical B1 residues are shown in red. (B) Complex formation between purified GST.MAL(214-298) and recombinant SRF.DBD (residues 133-222). (C) Mutations in GST.MAL(214-298) have similar effects on complex formation with *in vitro* translated SRF.DBD (residues 120-265) in gel mobility-shift assays (left) and in pulldown assays (right).

to compete with MAL for SRF complex formation in a dose dependent manner, whereas peptide **Q**, which encompasses the Y238A mutation that abolishes the MAL-SRF interaction had no effect (Figure 3.7B), demonstrating that the B1 region is sufficient for SRF-binding. Furthermore the presence of the wild-type but not the mutant peptide increased the amounts and slowed the mobility of the complex between DNA and endogenous SRF or the SRF.DBD (Figure 3.7B, complexes indicated by asterisks) indicating that the MAL B1 sequence directly contacts the SRF DNA binding domain.

I therefore explored the possibility of directly visualising the peptide-SRF complex by using high-density native gels to resolve SRF-DNA complexes in the presence of MAL B1 peptides. This experiment was also aimed at exacting the stoichiometry of peptide binding to SRF. Since MAL and TCF contact overlapping regions on SRF, two possible docking surfaces exist for the B1 region, one on each SRF subunit. Progressive increase of the B1 peptide concentration would therefore be expected to induce the sequential formation of two distinct complexes corresponding to monomeric and dimeric peptide-SRF interactions.

Surprisingly, although increasing amounts of the peptide **A** enhanced the SRF.DBD-DNA complex, they did not result in the formation of discrete complexes with SRF (Figure 3.7C). Instead increasing concentrations of the wild-type B1 peptide progressively slowed the mobility of the existing SRF-DNA complex (Figure 3.7, compare lanes 2-7 with lane 1). This "sliding-band" pattern was unchanged when five-fold less SRF.DBD was used (Figure 3.7, lanes 15-19). The reason for this concentration-dependent decrease in mobility is unclear, but it is possible that it reflects peptide oligomerisation, or is a result of the "caging effect" of the gel matrix: although the peptide-SRF complex can dissociate during its passage through the gel, the caging effect prevents the peptide from completely escaping the SRF-DNA complex. Similar observations have been made with other systems (Klejman, M. P. *et al.*, 2004). Nevertheless, irrespective of the exact reason for the unusual behaviour of the peptide-SRF complex, this interaction is specific since it is completely abolished by the Y238A mutation (peptide **Q**; Figure 3.7, lanes 9-14).

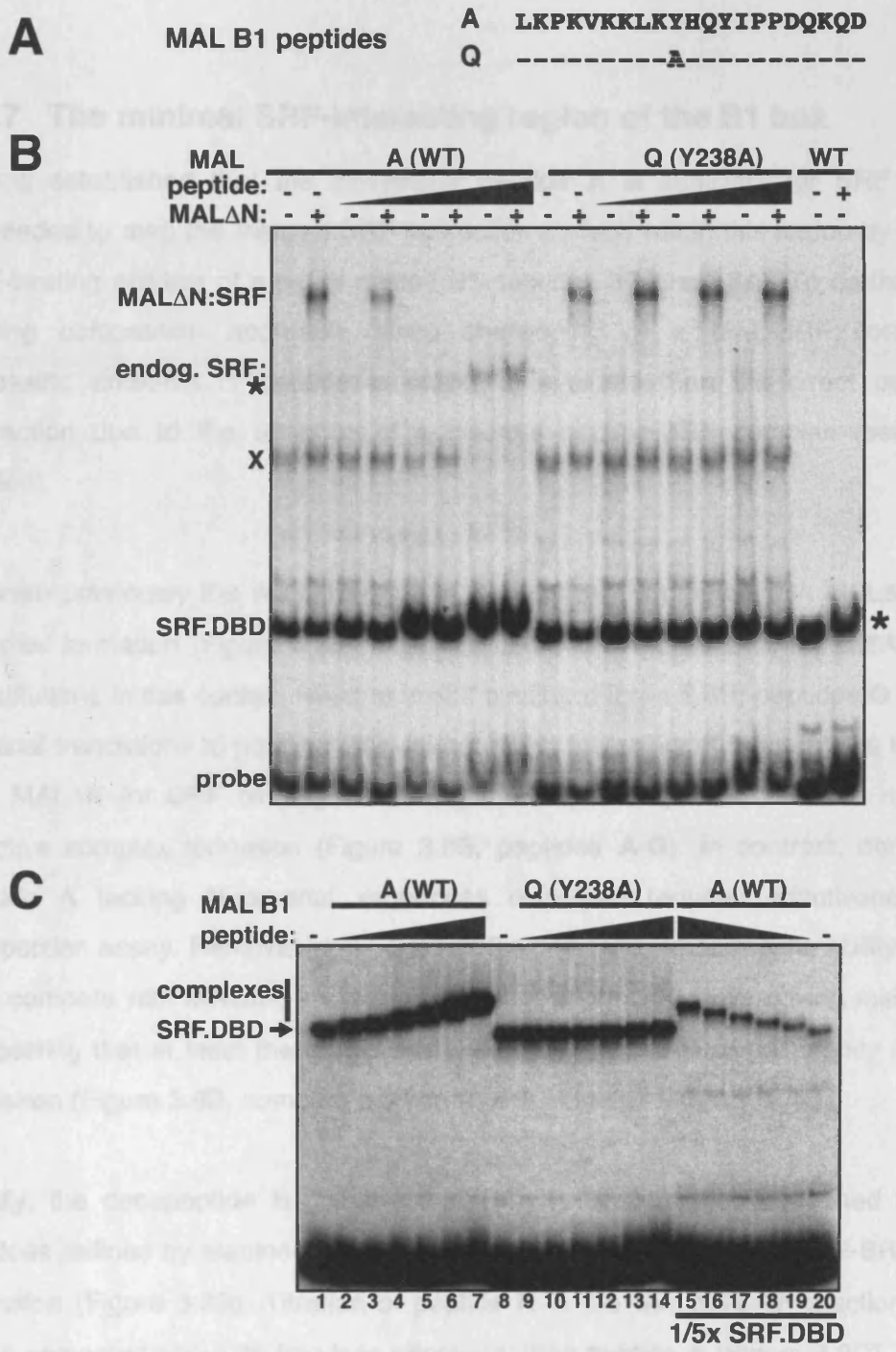


Figure 3.7. The B1 region is sufficient for interaction with SRF. (A) Sequences of the MAL B1 peptides used (**A**: wild-type; **Q**: Y238A mutant). (B) Peptide competition assay of the MAL-SRF complex with MAL B1-box peptides. Gel mobility-shift assays contained MALΔN whole-cell extract, SRF.DBD (residues 132-265), c-*fos* probe and 0, 0.8, 4 and 20μM wildtype or mutant MAL peptides as indicated. An asterisk indicates the specific SRF-peptide interaction revealed by the increased binding of endogenous SRF and of SRF(132-265). X indicates the position of a complex that is depleted by the wild-type peptide. It is not clear whether this represents the SRF.DBD bound to endogenous TCF or whether it is a non-specific effect seen due to the limiting amounts of probe at high peptide concentrations. (C) Complex formation between MAL B1 peptides and the SRF.DBD. Binding reactions contained *in vitro* translated SRF.DBD (residues 120-265), and c-*fos* ΔTCF probe with B1 peptides (0.16, 0.32, 0.63, 1.25, 2.5 and 5 μM). Lanes 15-20 contain probe with five times less SRF.DBD and 5, 2.5, 1.25, 0.63 and 0.32 μM peptide **A**. Complexes were resolved in 8% native gels.

3.2.7 The minimal SRF-interacting region of the B1 box

Having established that the 21-residue peptide **A** is sufficient for SRF binding, I proceeded to map the minimal SRF interaction surface within this region by testing the SRF-binding abilities of a set of nested B1-peptides (Figure 3.8A). To do that I used a binding competition approach, since challenging of a MAL-SRF complex with increasing amounts of peptide is easier to evaluate than the direct peptide-SRF interaction due to the absence of a discrete peptide-SRF complex (see previous section).

As seen previously the wildtype peptide **A** effectively competed with MALΔN for SRF complex formation (Figure 3.8B), whereas introduction of either the Y238A or Y241A substitutions in this context failed to inhibit binding (Figure 3.8B; peptides **Q** and **R**). C-terminal truncations to position 243 did not affect the ability of the peptides to compete with MALΔN for SRF binding, indicating that these sequences are not required for effective complex formation (Figure 3.8B, peptides **A-G**). In contrast, derivatives of peptide **A** lacking N-terminal sequences displayed reduced effectiveness in the competition assay. Removal of residue K230 significantly impaired the ability of peptide **N** to compete with MALΔN, and this was further decreased by removing residue K232, suggesting that at least these two basic residues contribute to the affinity of complex formation (Figure 3.8B, compare peptide **A** with peptides **M/N** and **J/K**).

Finally, the decapeptide **H** (N234-KKLKYHQYIP-C243), which contained the critical residues defined by alanine-scanning, also weakly competed for MALΔN-SRF complex formation (Figure 3.8B). Titration of peptide **H** in the MALΔN-SRF reactions showed that it competed about 30-fold less effectively than peptide **A** (Figure 3.8C), confirming the contribution of the basic B1 sequences in the affinity of the interaction, but indicating that they are not essential for complex formation. Introduction of the Y238A substitution in the context of peptide **H** failed to compete for MAL-SRF complex formation, indicating the interaction was specific (Figure 3.8C, compare peptides **H** and **S**). Hence, this decapeptide encompasses the minimal B1 surface required for SRF interaction.

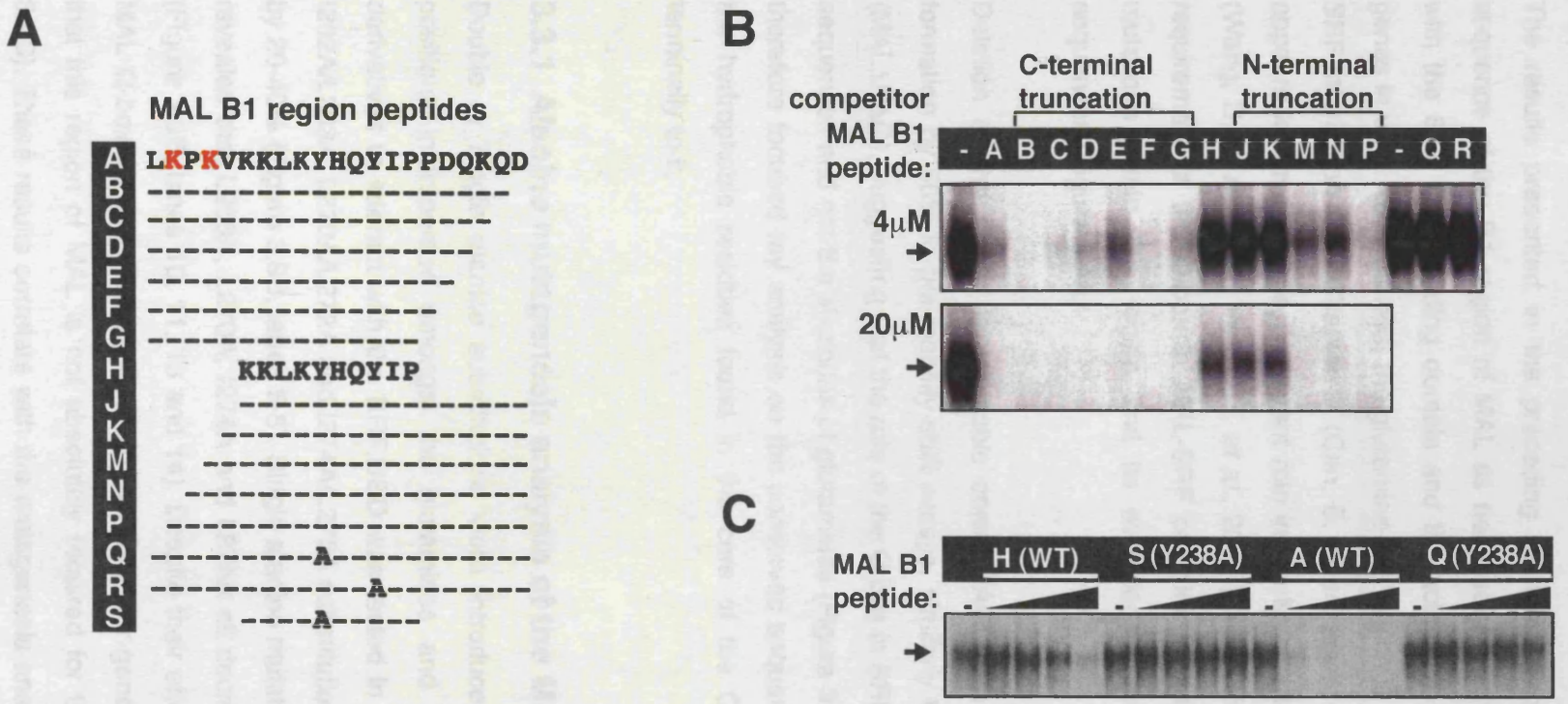


Figure 3.8. Identification of the minimal B1 region required for complex formation with SRF. (A) MAL B1 peptide sequences. Lysines K230 and K232 are shown in red. (B) Peptide competition studies. Gel mobility-shift binding reactions contained MALΔN extract, SRF.DBD (residues 132-223) and c-*fos* wildtype SRE probe with 4μM or 20μM peptide competitor as indicated. C-terminally and N-terminally truncated peptides are indicated by brackets. Only the SRF.DBD-MAL complex band is shown. (C) The hydrophobic core of the B1 region is sufficient for SRF complex formation. Binding reactions were as in B, with 1.3, 3.8, 11.3 and 34μM of B1 peptide titrated.

3.3 The role of the Q-box in SRF-binding

The results presented in the preceding section establish a minimal seven-residue sequence of the B1 region of MAL as necessary for the formation of direct contacts with the SRF DNA-binding domain and the activation of transcription of SRF target genes in vivo. Nevertheless, the glutamine-rich region is also required for optimal MAL-SRF interaction (see Chapter 2; (Cen, B. *et al.*, 2003; Miralles, F. *et al.*, 2003)) and appears to have a more prominent role in the Myocardin-SRF interaction (Chapter 2; (Wang, D. *et al.*, 2001; Wang, Z. *et al.*, 2004)). I therefore sought to characterise the requirement of the Q-box in MAL-SRF complex formation, by testing the effects of mutations within this region and its evolutionarily conserved N-terminal flanking sequences (Figure 3.9A).

Deletion of the Q-box hydrophobic core (MALΔNΔΦ) reduced MAL-SRF complex formation by 50% in gel mobility-shift assays, similarly to removal of the entire Q-box (MALΔNΔQ), suggesting that the role of the Q box in SRF binding is mediated by these sequences and not the stretches of glutamines (Figure 3.9B, compare lanes 3 and 4). I therefore focused my analysis on the conserved sequences in the Q box, particularly the hydrophobic residues found in the core of the Q-box and the sequences N-terminally to it.

3.3.1 Alanine mutagenesis analysis of the MAL Q box

Double or single alanine substitutions were introduced at the mainly hydrophobic positions interspersed amongst the glutamines and the ability of these MALΔN derivatives to interact with the SRF.DBD was tested in gel mobility-shift assays. The I262A/L263A, L270A/L272A and I274A/L275A substitutions reduced complex formation by 20-40% (Figure 3.9B, lanes 5-8). Single alanine mutations at each of these residues revealed that L263A, L270A, I274A and L275A all decrease the interaction with SRF (Figure 3.9B, lanes 10, 11, 13 and 14). Despite their effects on complex formation, all MAL Q-box derivatives activated the SRF reporter gene to wildtype levels, indicating that this region of MAL is not absolutely required for SRF activation *in vivo* (Figure 3.9C). These results correlate with the mutagenesis analysis of the Myocardin Q-box, in which residues L286, L293 and I297 (equivalent to MAL L263, L270 and I274) but not L298 (L275 in MAL) were necessary for SRF binding (Wang, Z. *et al.*, 2004). As

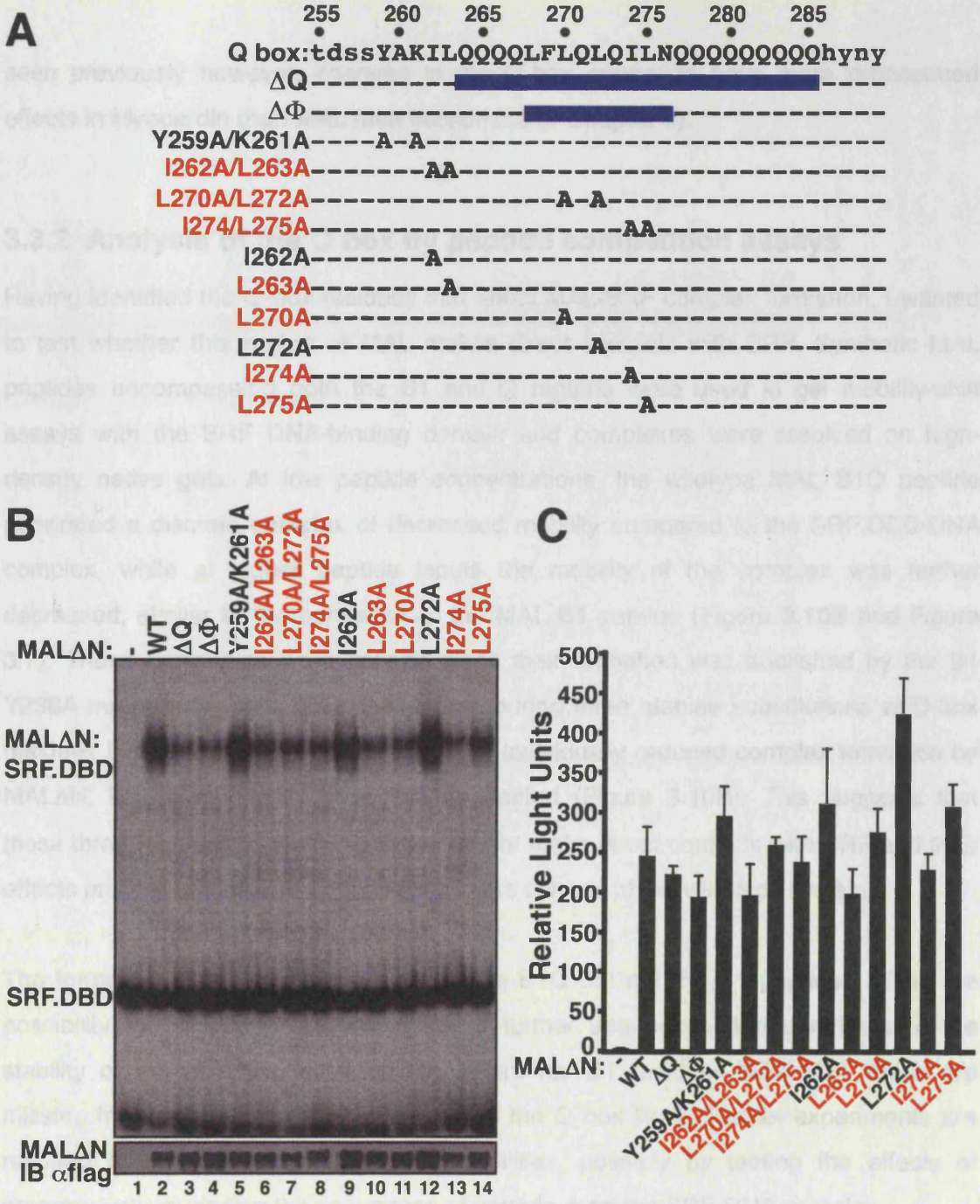


Figure 3.9. Effects of the Q-box mutations in the MAL-SRF interaction. (A) The Q-box sequence and mutations are shown. Deletion mutations are indicated by blue bars and point mutations that affect SRF binding are shown in red. (B) Mutations in the Q-box decrease but do not abolish MAL-SRF complex formation. Gel mobility-shift assays were performed using Flag-tagged MALΔN Q-box derivatives, SRF.DBD (residues 133-265), and the *c-fos* ΔTCF SRE probe. Expression levels of the MAL derivatives were confirmed by immunoblotting. MAL mutations that affect SRF binding are shown in red. (C) Q-box mutations do not affect activation of the 3DA.luc SRF reporter. NIH3T3 cells were transfected with 50ng of the indicated plasmids and maintained in 0.3% FBS until assayed for luciferase activity.

seen previously however, changes in the Q-box appear to have more pronounced effects in Myocardin than MAL (see section 2.3 of Chapter 2).

3.3.2 Analysis of the Q box by peptide competition assays

Having identified the Q-box residues that affect MAL-SRF complex formation, I wanted to test whether this region of MAL makes direct contacts with SRF. Synthetic MAL peptides encompassing both the B1 and Q regions were used in gel mobility-shift assays with the SRF DNA-binding domain and complexes were resolved on high-density native gels. At low peptide concentrations, the wildtype MAL B1Q peptide generated a discrete complex of decreased mobility compared to the SRF.DBD-DNA complex, while at higher peptide inputs the mobility of the complex was further decreased, similar to the behaviour of the MAL B1 peptide (Figure 3.10B and Figure 3.7). These complexes were specific since their formation was abolished by the B1 Y238A mutation. A MAL B1Q peptide harbouring three alanine substitutions at Q-box residues L263, L270 and I274, all of which individually reduced complex formation by MALΔN, left SRF-complex formation unaffected (Figure 3.10B). This suggests that these three residues of the MAL Q-box do not make direct contacts with SRF and their effects on SRF binding are only obvious in the context of the wild-type protein.

The formation of discrete complexes by the B1Q but not the B1 peptides raises the possibility that the B1Q peptide contains further sequences that contribute to the stability of the complex, such as the N-terminal B1 residues (224-228) which are missing from peptide **A**, or the presence of the Q box itself. Further experiments are required in order to explore these possibilities, possibly by testing the effects of progressively extending the sequences of peptide **A** on the SRF-DNA complex.

As mentioned previously, Wang and coworkers reported that Myocardin residues L286, L293 and I297 (corresponding to MAL positions L263, L270 and I274), were necessary for SRF binding by Myocardin (Wang, Z. *et al.*, 2004). I therefore tested the analogous Myocardin B1Q peptides in the SRF-binding assay (Figure 3.10A). As seen with MAL, the wildtype Myocardin B1Q peptide formed a complex with the SRF.DBD, and at higher peptide inputs a second complex could be discerned, suggesting that individual peptides can interact with each SRF monomer (Figure 3.10B). These complexes were abolished by the Y261A mutation, which is equivalent to MAL Y238A. The Myocardin

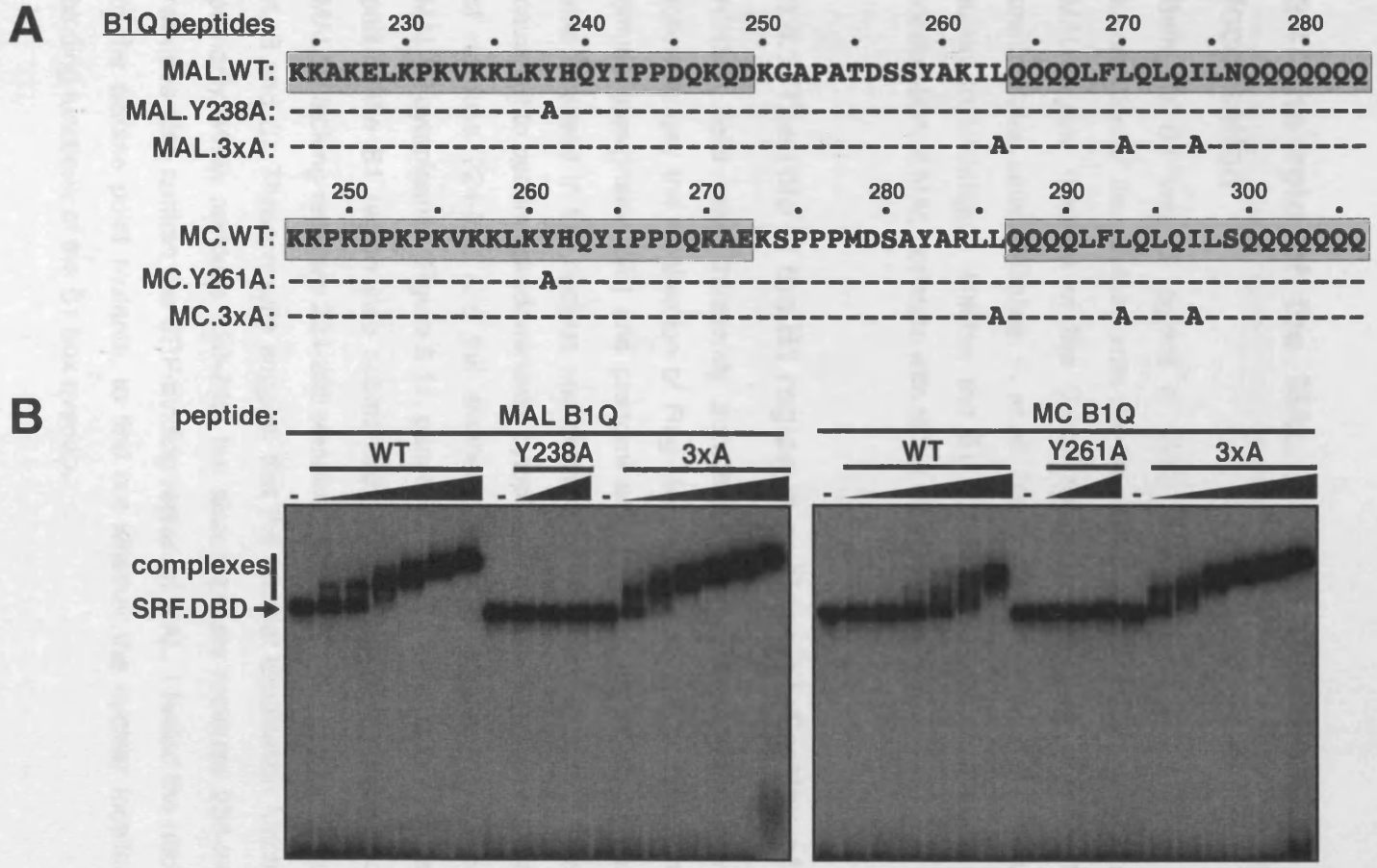


Figure 3.10. The B1 but not the Q region is necessary for SRF complex formation by both MAL and Myocardin. (A) The sequences of the MAL and Myocardin B1Q peptides. The B1 and Q regions are boxed in grey and the positions of point mutations are shown. (B) Complex formation between MAL and Myocardin B1Q peptides and SRF.DBD. Binding reactions contained *in vitro* translated SRF.DBD (residues 120-265), and c-*fos* ΔTCF probe with B1Q peptides (wildtype or 3xA peptide: 0.014, 0.04, 0.12, 0.37, 1.11, and 3.33 μM; Y238A or Y261A peptide: 0.37, 1.11, and 3.33 μM).

3xA peptide that contains alanines at positions L286, L293 and I297 was also able to interact with SRF, suggesting that also in the case of Myocardin these residues do not mediate direct protein-protein interactions with SRF.

3.4 The role of the MAL B1 and Q regions in subcellular localisation

Both the B1 and Q regions of MAL have been shown to affect the subcellular localisation of the protein, with the B1 region necessary for nuclear accumulation of MALΔN and deletion of the Q-box resulting in nuclear accumulation of MAL in unstimulated cells (Miralles, F. *et al.*, 2003). I therefore used the immunofluorescence assay to investigate whether the B1 and Q sequences involved in the subcellular-localisation of MAL correlate with those affecting SRF-interaction.

3.4.1 The role of the B1 region in nuclear localisation of MALΔN

NIH3T3 cells were transiently transfected with different MAL derivatives and the following day the localisation of Flag-MAL was scored as predominantly cytoplasmic, diffuse (designated c/n) and predominantly nuclear. As previously reported, MALΔN was localised in the nucleus and deletion of the entire B1 region (residues 224-249) caused it to become predominantly cytoplasmic (Figure 3.11, panels B and D). Deletion of residues 224-235 and the smaller deletion of residues 230-235 also rendered MALΔN cytoplasmic (Figure 3.11, panels A, B and C). Removal of the entire C-terminal part of the B1 region also substantially reduced MALΔN nuclear localisation, while MALΔN lacking residues 224-229 remained predominantly nuclear (Figure 3.11, panels A, B and C). These results suggest that the nuclear localisation function is contained primarily within residues 230-235, but also implicate residues 236-249. Since these residues also contain the SRF-binding region of MAL, I tested the nuclear localisation of the alanine point mutants, to find out whether the nuclear localisation and SRF-binding functions of the B1 box overlap.

Alanine substitutions K234A and K235A greatly reduced nuclear localisation of MALΔN, with a K234/235A double mutant having as severe an effect as deletion of the entire B1-box (Figure 3.11 panels C and D). Mutation K237A also reduced nuclear localisation (Figure 3.11, panels C and D), but all other substitutions, including those

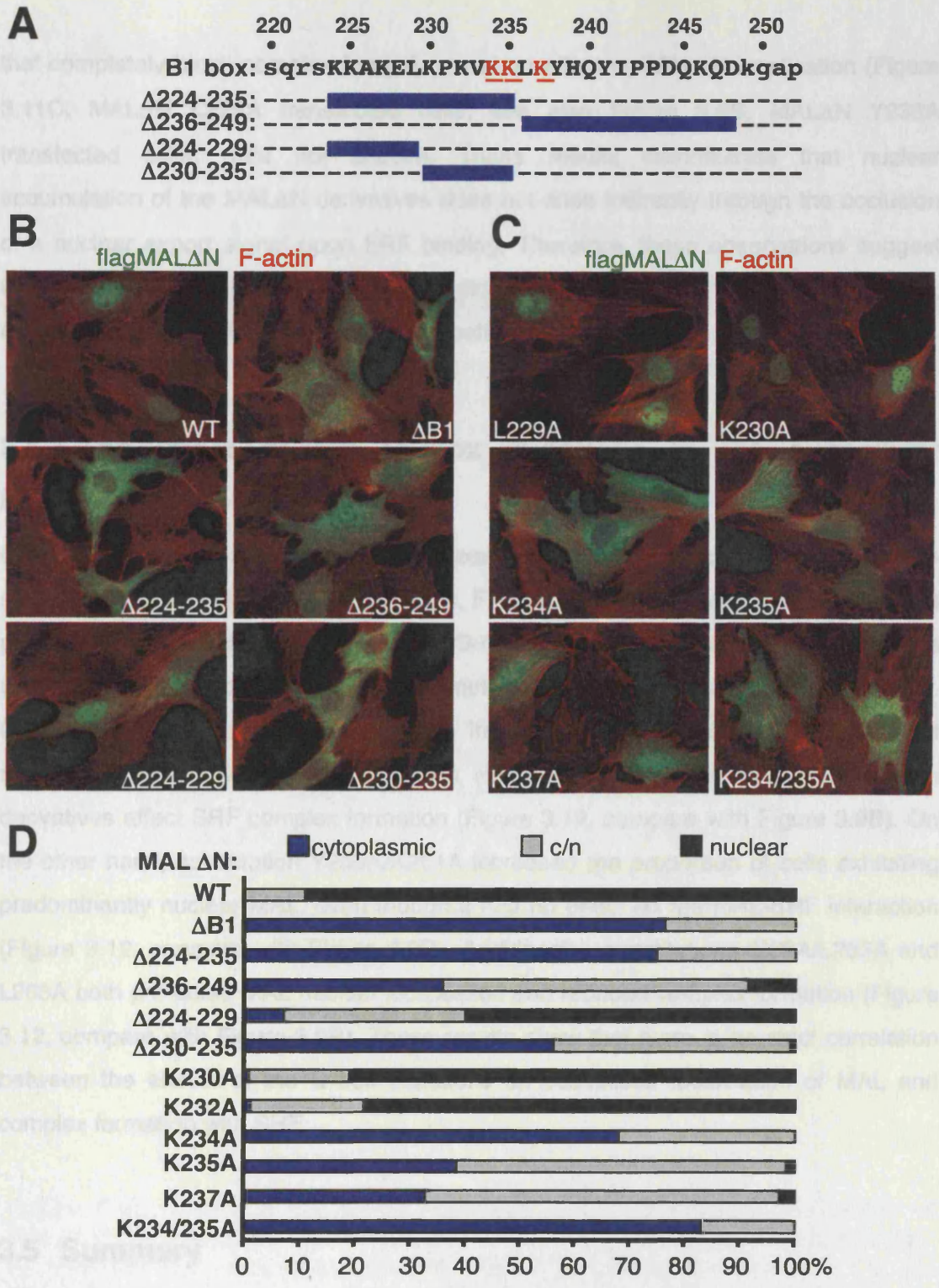


Figure 3.11. Identification of MAL B1 region sequences involved in nuclear import. (A) B1 region sequence, with deletions in blue and residues K234, K235 and K237 highlighted in red. (B) Subcellular localisation of the MALΔN B1 region deletion derivatives. NIH3T3 cells were transfected with 50ng of the indicated MALΔN plasmids and stained for F-actin (red) and Flag-tagged MALΔN (green). (C) Subcellular localisation of MALΔN B1 region point-mutation derivatives, as in B. (D) MALΔN B1 mutations affect nuclear import. Subcellular localisation was scored as predominantly cytoplasmic, evenly distributed or nuclear. Results are average of two independent experiments (100-200 cells each).

that completely block complex formation, had no effect on MALΔN localisation (Figure 3.11C, MALΔN L229A transfected cells; see also Figure 3.5B, MALΔN Y238A transfected cells; data not shown). These results demonstrate that nuclear accumulation of the MALΔN derivatives does not arise indirectly through the occlusion of a nuclear export signal upon SRF binding. Therefore, these observations suggest that the nuclear localisation and SRF-binding functions of the B1 box are separable, even though residue K237 plays a role in both activities.

3.4.2 The effect of the Q box mutations in the subcellular localisation of MAL(met)

Deletion of the Q-box promotes the nuclear localisation of intact MAL under serum-starved conditions (Figure 3.12B, (Miralles, F. *et al.*, 2003)). To investigate whether this property correlates with the effects of the Q-box mutations on SRF-binding I tested the subcellular localisation of various MAL(met) Q derivatives by immunofluorescence. Deletion of the Q-box hydrophobic core increased MAL nuclear accumulation, but mutation I274A/L275A had no effect on subcellular localisation even though both derivatives affect SRF complex formation (Figure 3.12, compare with Figure 3.9B). On the other hand, substitution Y259A/K261A increased the proportion of cells exhibiting predominantly nuclear MAL, even though it had no effect on the MAL-SRF interaction (Figure 3.12, compare with Figure 3.9B). Additionally, substitutions I262A/L263A and L263A both promoted MAL nuclear localisation and reduced complex formation (Figure 3.12, compare with Figure 3.9B). These results show that there is no strict correlation between the effects of the Q-box mutations on subcellular localisation of MAL and complex formation with SRF.

3.5 Summary

The results described in this chapter characterise the B1 box as the SRF-binding surface of MAL and identify the residues within this surface required for efficient SRF interaction and activity. The integrity of the B1 region is required for SRF activation by the full-length MAL protein and is also necessary for expression of endogenous MAL-dependent SRF-target genes, such as smooth muscle α-actin.

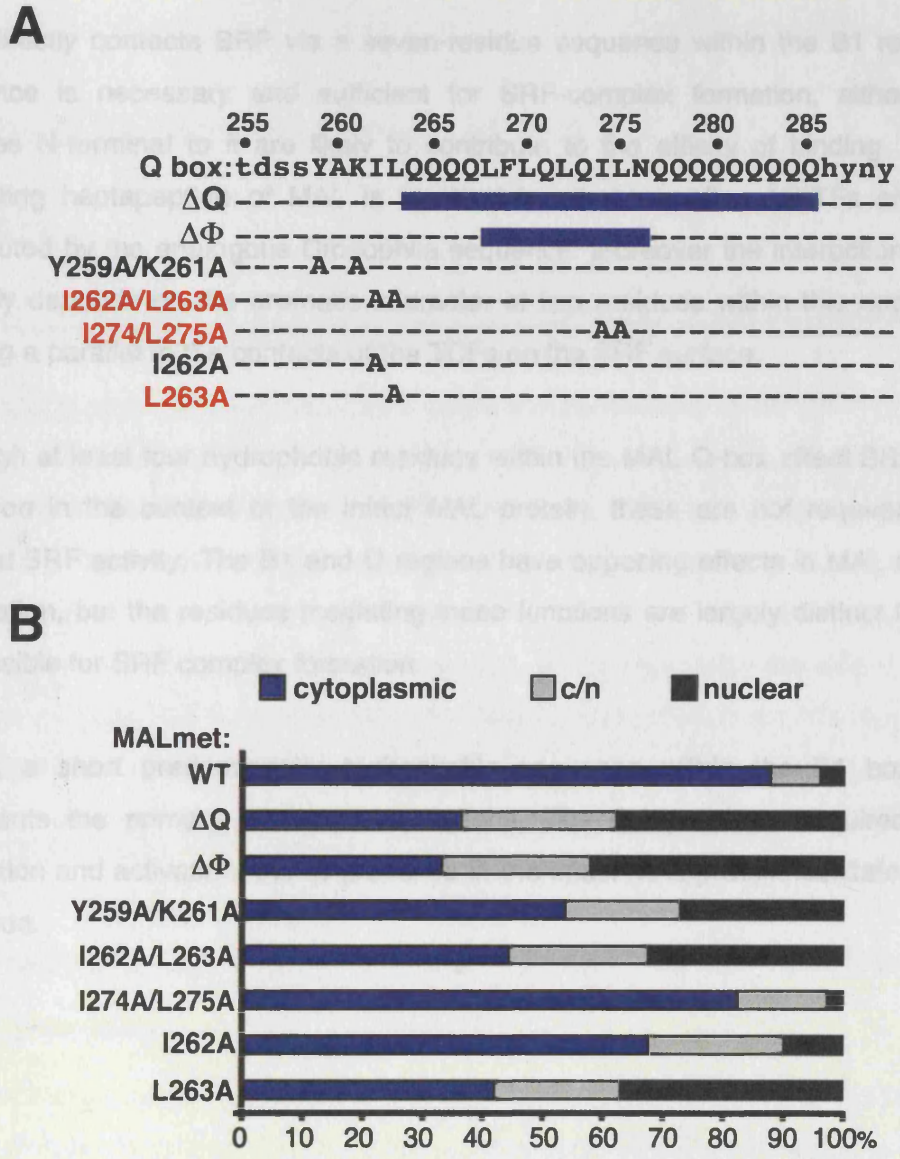


Figure 3.12. Effects of the Q-box mutations in the subcellular localisation of MAL(met). (A) The Q-box sequence and mutations. Deletion mutations are indicated by blue bars and point mutations that affect SRF binding are shown in red. (B) MAL(met) Q-box mutations affect nuclear import. Subcellular localisation of the MAL(met) Q-box region derivatives. NIH3T3 cells were transfected with 50ng of the indicated MAL(met) plasmids and stained for F-actin and Flag-tagged MAL(met). MAL subcellular localisation was scored as predominantly cytoplasmic, evenly distributed (c/n) or nuclear. Results are average of two independent experiments (100-200 cells each).

MAL directly contacts SRF via a seven-residue sequence within the B1 region. This sequence is necessary and sufficient for SRF-complex formation, although basic residues N-terminal to it are likely to contribute to the affinity of binding. The SRF-contacting heptapeptide of MAL is identical in all mammalian MRTFs and can be substituted by the analogous Drosophila sequence. Moreover the interaction with SRF critically depends on the aromatic character of two residues within this heptapeptide, drawing a parallel to the contacts of the TCFs on the SRF surface.

Although at least four hydrophobic residues within the MAL Q-box affect SRF complex formation in the context of the intact MAL protein, these are not required for MAL induced SRF activity. The B1 and Q regions have opposing effects in MAL subcellular localisation, but the residues mediating these functions are largely distinct from those responsible for SRF complex formation.

Hence, a short predominantly hydrophobic sequence within the B1 box of MAL represents the primary SRF contact surface. The Q-box is not required for SRF interaction and activation, but its presence in the intact MAL protein facilitates complex formation.

4 The MAL-binding surface of SRF

4.1 Aims

The results presented in the previous chapters established important similarities in the properties that govern the interaction of SRF with the MRTF and the TCF cofactor families. Although MAL and the TCFs physically compete for SRF binding, indicating that they contact a common surface on the SRF DNA-binding domain, MAL unlike the TCFs cannot interact with altered DNA-specificity SRF derivatives, which harbour mutations in the N-terminal region of the DNA-binding domain. The aim of the present chapter was to further analyse the requirements of the MAL-SRF interaction by mapping the surface of SRF that contacts MAL and investigating the extent to which it overlaps with the TCF-binding surface. Furthermore, I wished to explore the role of the N-terminal part of the SRF DNA-binding domain in MAL complex formation and determine whether it directly contributes to the MAL-binding surface of SRF.

4.2 Mapping of the MAL-binding surface of SRF by point mutagenesis

Detailed biochemical and structural studies identified the TCF-binding surface of SRF as a hydrophobic groove formed by the βII-strand, proceeding coil region and αII-helix of the DNA-binding domain and showed that TCF interacts with this surface by adding a β-strand to the β-sheet of the SRF middle layer (see Introduction; (Hassler, M. *et al.*, 2001; Ling, Y. *et al.*, 1998)). These studies pinpointed the key SRF DNA-binding domain residues, including residues within this hydrophobic groove, involved in TCF contacts Figure 4.1A and B). Given that MAL and TCF physically compete for SRF binding, I sought to map the MAL-binding surface of SRF by first analysing the role of the hydrophobic groove residues in the MAL-SRF complex.

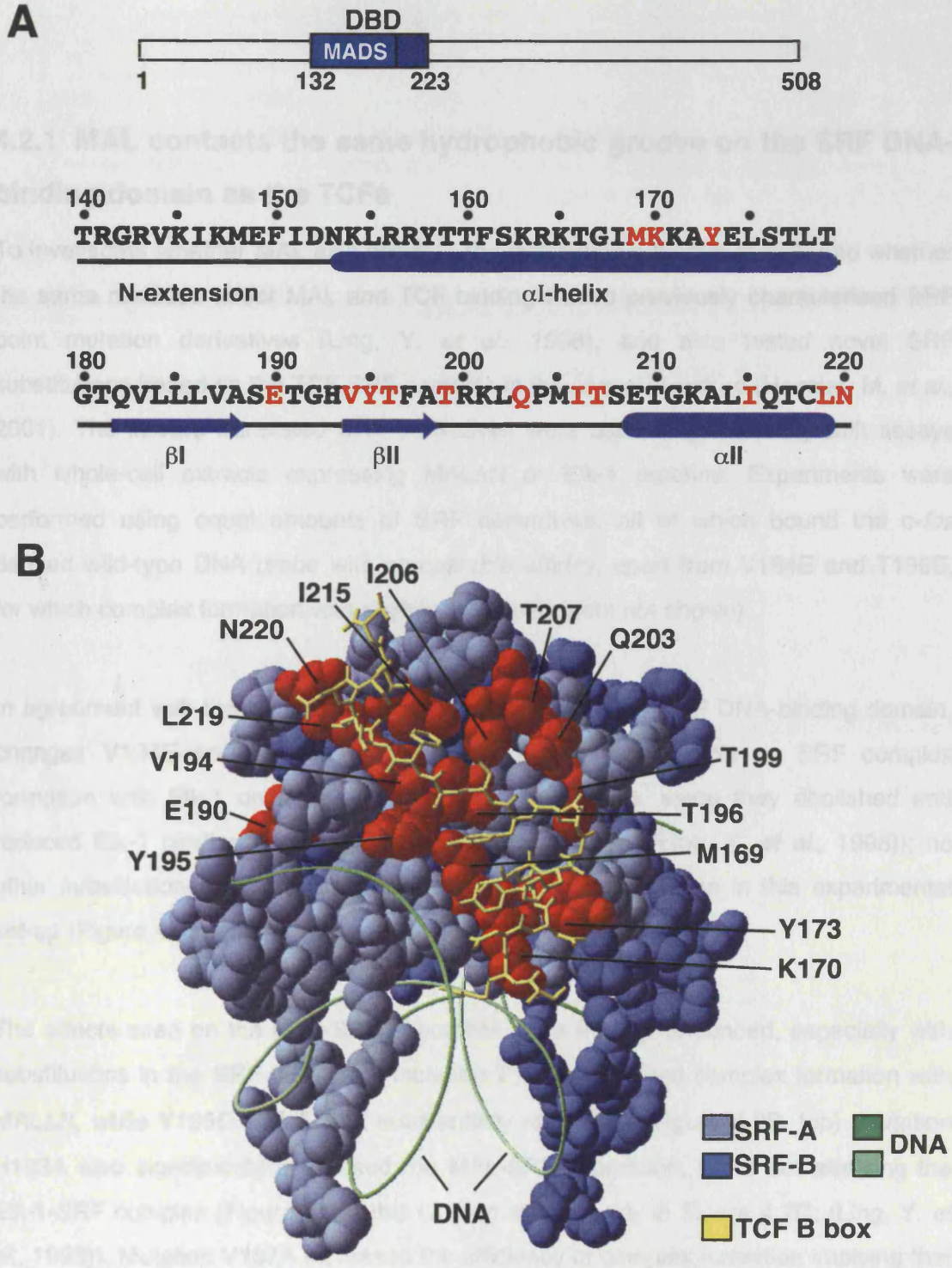


Figure 4.1. Residues in the SRF DNA-binding domain involved in TCF binding. (A) Schematic representation of SRF. Secondary structure elements in the SRF.DBD (residues 140- 220) are shown below. Residues involved in interaction with the TCF B box are highlighted in red. (B) Van der Waals representation of the DNA-binding domain of SRF (blue) complexed with DNA (green, shown as ribbon for simplicity) and the SAP1a B box (yellow, backbone and side-chain representation). Red SRF residues are involved in the TCF interaction. E190, V194, Y195, T196, T199, Q203, T207 were originally identified by mutagenesis analysis (Ling et. al. 1998) and M169, K170, Y173, I206, I215, L219, N220 were later identified by crystallographic analysis (Hassler and Richmond, 2001).

4.2.1 MAL contacts the same hydrophobic groove on the SRF DNA-binding domain as the TCFs

To investigate whether MAL also contacts the hydrophobic groove of SRF and whether the same residues affect MAL and TCF binding I used previously characterised SRF point mutation derivatives (Ling, Y. *et al.*, 1998), and also tested novel SRF substitutions based on the TCF-SRF contacts in the crystal structure (Hassler, M. *et al.*, 2001). The in vitro translated SRF derivatives were used in gel mobility-shift assays with whole-cell extracts expressing MALΔN or Elk-1 proteins. Experiments were performed using equal amounts of SRF derivatives, all of which bound the c-*fos* derived wild-type DNA probe with comparable affinity, apart from V194E and T196E, for which complex formation was slightly increased (data not shown).

In agreement with the original mutagenesis analysis of the SRF DNA-binding domain, changes V194E and T196E had the most pronounced effects on SRF complex formation with Elk-1 on a wild-type c-*fos* derived probe, since they abolished and reduced Elk-1 binding respectively (Figure 4.2B, bottom; (Ling, Y. *et al.*, 1998)); no other substitutions had detectable effects on complex formation in this experimental set-up (Figure 4.2, bottom).

The effects seen on the MAL-SRF interaction were more pronounced, especially with substitutions in the SRF βII-strand: mutation V194E abolished complex formation with MALΔN, while Y195D and T196E substantially reduced it (Figure 4.2B, top). Mutation H193A also significantly decreased the MAL-SRF interaction, while not affecting the Elk-1-SRF complex (Figure 4.2B, this is seen more clearly in Figure 4.7C; (Ling, Y. *et al.*, 1998)). Mutation V187A increased the efficiency of complex formation implying that this side-chain might interfere with MAL-SRF complex formation. Change Q203E had the opposite effect on MAL binding than that seen with TCF, since it increased MAL-SRF complex formation (Figure 4.2B and Figure 4.1B; (Ling, Y. *et al.*, 1998)), suggesting that the negative charge at position 203 might productively interact with MAL in the complex. The other substitutions did not have significant effects.Taken together these results suggest that MAL contacts the β-sheet in the middle layer of the SRF.DBD.

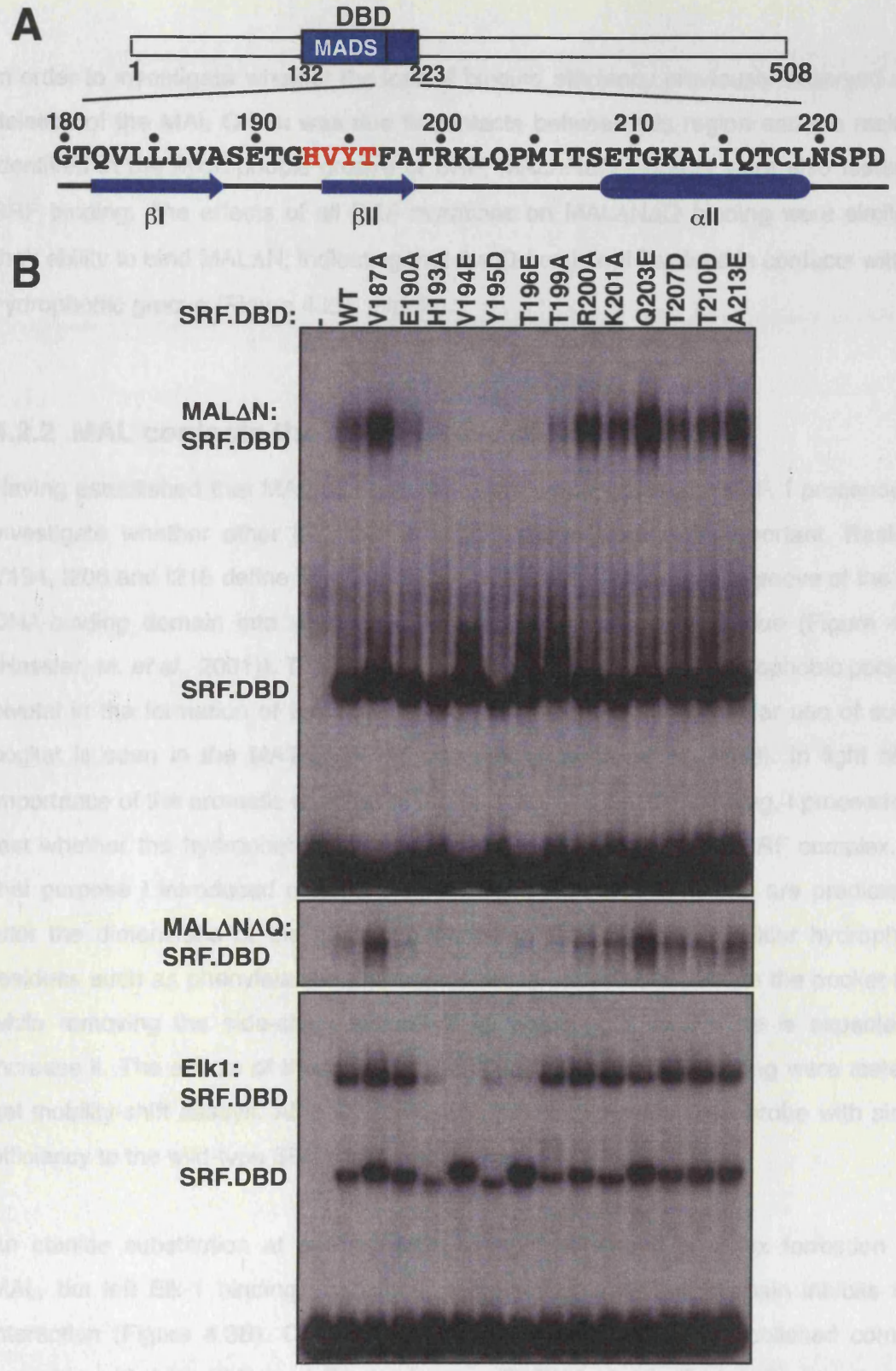


Figure 4.2. MAL and TCF contact the same hydrophobic groove on the SRF DNA-binding domain. (A) Schematic representation of SRF, with the DBD in blue. The sequence and secondary structure elements of the C-terminal half of the DBD are shown below, with hydrophobic groove residues in red. (B) SRF hydrophobic groove mutations affect complex formation. MALΔN (top), MALΔNΔQ (middle) or Elk-1 (bottom) whole-cell extracts were used in gel mobility-shift assays with SRF.DBD (residues 120-265), and either ΔTCF SRE probe (top and middle) or WT SRE probe (bottom).

In order to investigate whether the loss of binding efficiency previously observed upon deletion of the MAL Q-box was due to contacts between this region and the residues identified in the hydrophobic groove of SRF, MALΔNΔQ extracts were also tested for SRF binding. The effects of all SRF mutations on MALΔNΔQ binding were similar to their ability to bind MALΔN, indicating that the Q-box is not involved in contacts with the hydrophobic groove (Figure 4.2B, middle).

4.2.2 MAL contacts the hydrophobic pocket of SRF

Having established that MAL contacts the hydrophobic groove of SRF, I proceeded to investigate whether other interactions in this region were also important. Residues V194, I206 and I215 define a hydrophobic pocket in the hydrophobic groove of the SRF DNA-binding domain into which the TCFs insert an aromatic residue (Figure 4.1B; (Hassler, M. *et al.*, 2001)). The interaction of this residue with the hydrophobic pocket is pivotal in the formation of the TCF-SRF complex. Furhtermore, similar use of such a pocket is seen in the MATα2-MCM1 complex (Tan, S. *et al.*, 1998). In light of the importance of the aromatic residues in the MAL B1 box for SRF binding, I proceeded to test whether the hydrophobic pocket was also central in the MAL-SRF complex. For that purpose I introduced mutations at positions I206 and I215 that are predicted to alter the dimensions of the pocket. Exchanging isoleucine with bulkier hydrophobic residues such as phenylalanine and tryptophan is expected to reduce the pocket size, while removing the side-chain altogether by changing it to alanine is expected to increase it. The effects of these substitutions on MAL and TCF binding were tested in gel mobility-shift assays. All SRF derivatives bound the c-*fos* DNA probe with similar efficiency to the wild-type SRF (data not shown).

An alanine substitution at position I206 strongly enhanced complex formation with MAL, but left Elk-1 binding unaffected, suggesting that this side-chain inhibits MAL interaction (Figure 4.3B). Conversely a change to phenylalanine abolished complex formation with both MAL and Elk-1, whereas I206W had no effect, indicating that this change can be tolerated in the interaction of both MAL and Elk-1 with SRF (Figure 4.3B). Substitutions I215A and I215W impaired complex formation with both MAL and Elk-1 whereas I215F decreased complex formation by both cofactors to a lesser extent, indicating that the aliphatic side-chain is important at this position. Similar results were obtained with a MAL derivative lacking the Q-box (Figure 4.3B, middle).

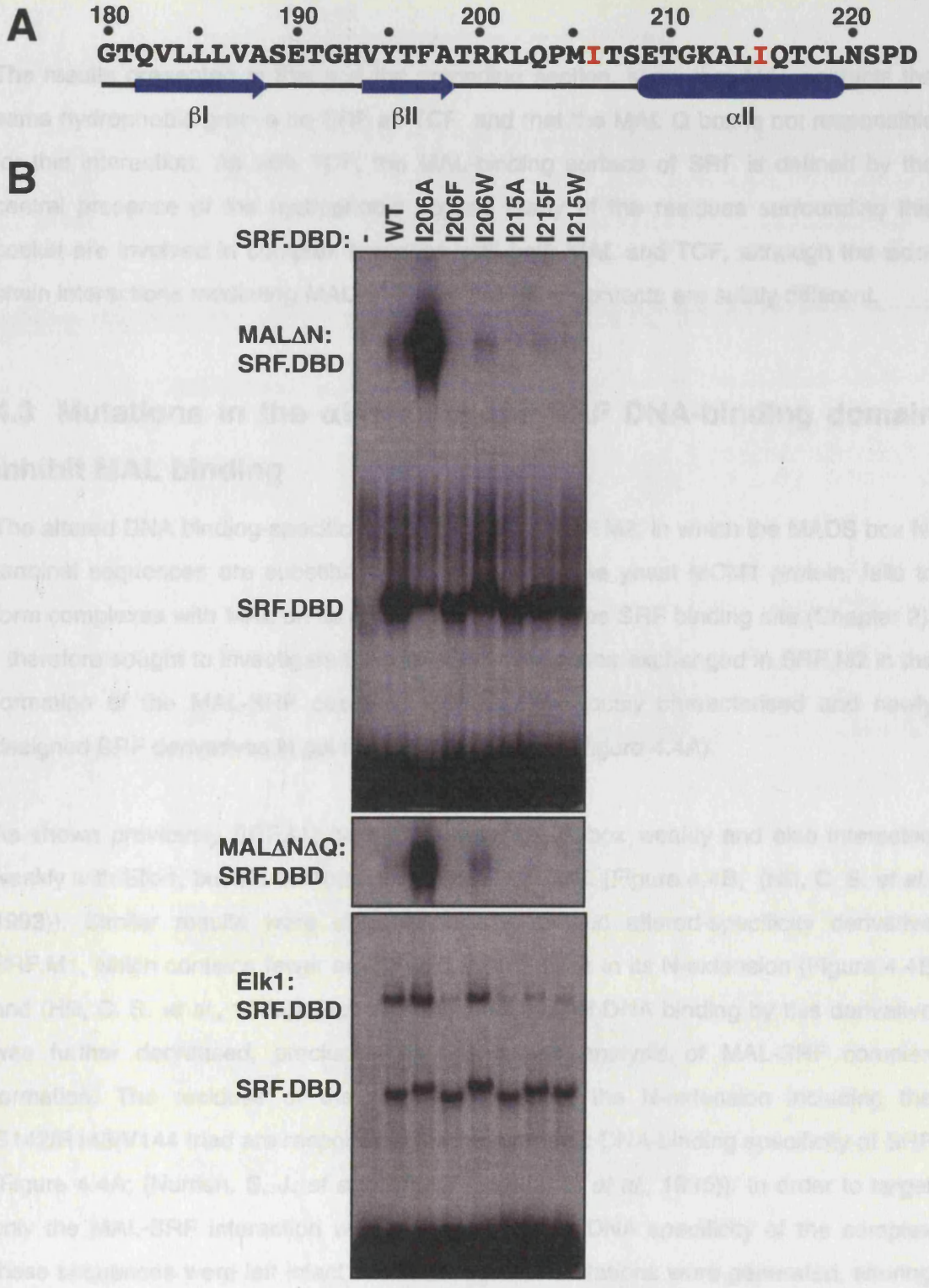


Figure 4.3. Mutations in the SRF hydrophobic pocket differentially affect MAL-SRF and Elk-1-SRF interactions. (A) The sequence and secondary structure elements of the C-terminal half of the DBD are shown, with mutated hydrophobic pocket residues in red. (B) Mutations predicted to alter the dimensions of the SRF hydrophobic pocket have different efffects in the interaction of MAL and Elk-1 with SRF. MALΔN (top), MALΔNΔQ (middle) or Elk-1 (bottom) whole-cell extracts were used in gel mobility-shift assays with SRF.DBD (residues 120-265), and either ΔTCF SRE probe (top and middle) or WT SRE probe (bottom).

The results presented in this and the preceding section, show that MAL contacts the same hydrophobic groove on SRF as TCF, and that the MAL Q box is not responsible for this interaction. As with TCF, the MAL-binding surface of SRF is defined by the central presence of the hydrophobic pocket. Many of the residues surrounding this pocket are involved in complex formation with both MAL and TCF, although the side-chain interactions mediating MAL-SRF and TCF-SRF contacts are subtly different.

4.3 Mutations in the αI-helix of the SRF DNA-binding domain inhibit MAL binding

The altered DNA binding-specificity SRF derivative SRF.M2, in which the MADS box N-terminal sequences are substituted with those from the yeast MCM1 protein, fails to form complexes with MAL on its cognate or the wild-type SRF binding site (Chapter 2). I therefore sought to investigate the role of the sequences exchanged in SRF.M2 in the formation of the MAL-SRF complex, by using previously characterised and newly designed SRF derivatives in gel mobility-shift assays (Figure 4.4A).

As shown previously SRF.M2 bound the c-*fos* CArG box weakly and also interacted weakly with Elk-1, but did not form complexes with MAL (Figure 4.4B; (Hill, C. S. *et al.*, 1993)). Similar results were obtained with a second altered-specificity derivative SRF.M1, which contains fewer amino acid substitutions in its N-extension (Figure 4.4B and (Hill, C. S. *et al.*, 1993)). However the efficiency of DNA binding by this derivative was further decreased, precluding its use in the analysis of MAL-SRF complex formation. The residues of the N-terminal half of the N-extension including the G142/R143/V144 triad are responsible for the authentic DNA-binding specificity of SRF (Figure 4.4A; (Nurrish, S. J. *et al.*, 1995; Pellegrini, L. *et al.*, 1995)). In order to target only the MAL-SRF interaction without perturbing the DNA specificity of the complex these sequences were left intact and two clustered mutations were generated, altering the C-terminal part of the N-extension and the αI-helix respectively.

The resulting SRF derivatives SRF(N-extension) and SRF(αI-helix) bound the wild-type c-*fos* CArG box with efficiencies similar to the wild-type SRF (Figure 4.4A and data not shown). The four mutations in the SRF N-extension derivative had no effect on either MAL or Elk-1 binding to SRF (Figure 4.4B). The same result was seen with individual substitutions K147E and M148I, while changes E149K and D152E caused small

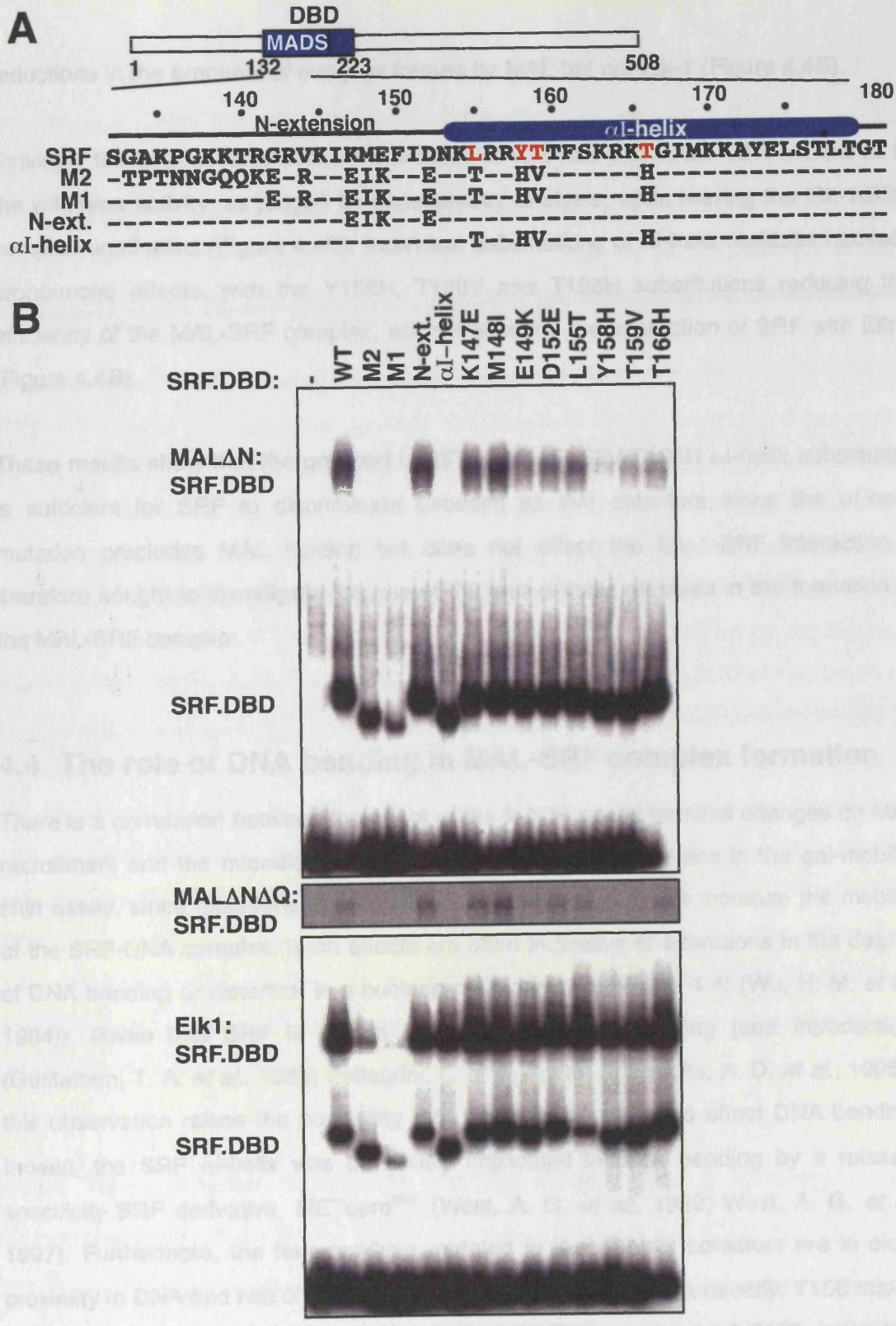


Figure 4.4. Mutations in the SRF.DBD αI-helix inhibit MAL-SRF complex formation. (A) Secondary structure elements and mutations in the N-terminal half of the SRF DBD. Critical αI-helix residues are highlighted in red. (B) SRF αI-helix mutations disrupt MAL-SRF complex formation. Whole-cell extracts from cells expressing MALΔN (top), MALΔNΔQ (middle) or Elk-1 (bottom) were used in gel mobility-shift assays with the indicated SRF.DBD derivatives (residues 120-265), and either ΔTCF c-*fos* SRE probe (top and middle) or WT c-fos SRE probe (bottom).

reductions in the amounts of complex formed by MAL but not Elk-1 (Figure 4.4B).

Strikingly the SRF αI-helix derivative reduced MAL binding to below 10% compared to the wild-type activity, as judged by densitometry analysis, while leaving the Elk-1-SRF complex unaffected (Figure 4.4B). Individual substitutions of αI-helix residues had less pronounced effects, with the Y158H, T159V and T166H substitutions reducing the efficiency of the MAL-SRF complex, without affecting the interaction of SRF with Elk-1 (Figure 4.4B).

These results show that the grouped L155T/Y158H/T159V/T166H αI-helix substitution is sufficient for SRF to discriminate between its two cofactors since the αI-helix mutation precludes MAL binding but does not affect the Elk-1-SRF interaction. I therefore sought to investigate the role of the four αI-helix residues in the formation of the MAL-SRF complex.

4.4 The role of DNA bending in MAL-SRF complex formation

There is a correlation between the effect of the MADS box N-terminal changes on MAL recruitment and the migration properties of the SRF-DNA complex in the gel-mobility shift assay, since mutations that impair the MAL interaction also increase the mobility of the SRF-DNA complex. Such effects are often indicative of alterations in the degree of DNA bending or distortion in a nucleoprotein complex (Figure 4.4; (Wu, H. M. *et al.*, 1984)). Given that SRF is known to bend DNA upon binding (see Introduction; (Gustafson, T. A. *et al.*, 1989; Pellegrini, L. *et al.*, 1995; Sharrocks, A. D. *et al.*, 1995)), this observation raises the possibility that the αI-helix mutations affect DNA bending. Indeed, the SRF αI-helix was previously implicated in DNA bending by a relaxed-specificity SRF derivative, METcore^{SRF} (West, A. G. *et al.*, 1999; West, A. G. *et al.*, 1997). Furthermore, the four residues mutated in the αI-helix construct are in close proximity to DNA and two of them are known to interact with DNA directly: Y158 makes a phosphate contact with DNA in the SAP1-SRF-DNA crystal, and T159 contacts a thymine on the 5' side of the CArG box in the binary SRF-DNA complex (Hassler, M. *et al.*, 2001; Pellegrini, L. *et al.*, 1995). This suggests that the role of these substitutions in MAL-SRF complex formation might be indirect, through their effects in the way SRF interacts with DNA.

To compare DNA bending by the different SRF derivatives I performed circular permutation analysis (Wu, H. M. *et al.*, 1984). This technique is based on the fact that DNA distortion is a determinant of the electrophoretic mobility of a DNA fragment and employs the gel-mobility shift assay to calculate the degree of DNA distortion induced in a nucleoprotein complex upon protein binding. This is achieved by generating a set of DNA probes of equal lengths, that are circular permutations of the same sequence, resulting in the protein binding site being located at different positions along the DNA fragment (Figure 4.5). Electrophoresis of the protein of interest bound to these probes generates complexes of varying mobilities: when the DNA distortion such as a bend is present in the middle of the DNA fragment, the nucleoprotein complex migrates through the native gel more slowly than an equivalent complex in which the bend is present at the end of the DNA fragment. Calculation of the relative mobilities of these complexes and fitting of the data to a cosine function allows the determination of the minimum (bend in the middle of the fragment) and maximum (bend at the end of the fragment) migration points (Figure 4.5). These are then used to estimate the apparent bend angle α of the complex from the empirical equation $[R_f^{(\text{middle binding site})}/R_f^{(\text{end binding site})} = \cos(\alpha/2)]$ (Thompson, J. F. *et al.*, 1988). This value is termed "apparent bend angle" because the circular permutation technique cannot unequivocally identify directional DNA bending as opposed to other DNA distortions such as locations of increased DNA flexibility (Kerppola, T. K. *et al.*, 1991). Moreover, bend angles calculated with this method are not absolute values, since their magnitude depends on experimental conditions including gel density, fragment length, electrical field strength and temperature (Kerppola, T. K. *et al.*, 1991).

4.4.1 The αI-helix mutations inhibit MAL binding through their effects on DNA bending

Circular permutation analysis of the SRF αI-helix derivatives was performed using a set of circularly permuted DNA probes, generated from the c-*fos* promoter. The wild-type SRF DNA-binding domain induced an apparent bend angle of 55.5° ± 1.07° in this assay (Figure 4.6B). This value is lower than the 72° observed in crystallographic and previous circular permutation studies, presumably owing to the different experimental conditions used. The SRF.M2 mutation, which abolished interaction with MAL, reduced the apparent bend angle in the binary SRF-DNA complex to 34. 7° ± 0.66°. A similar

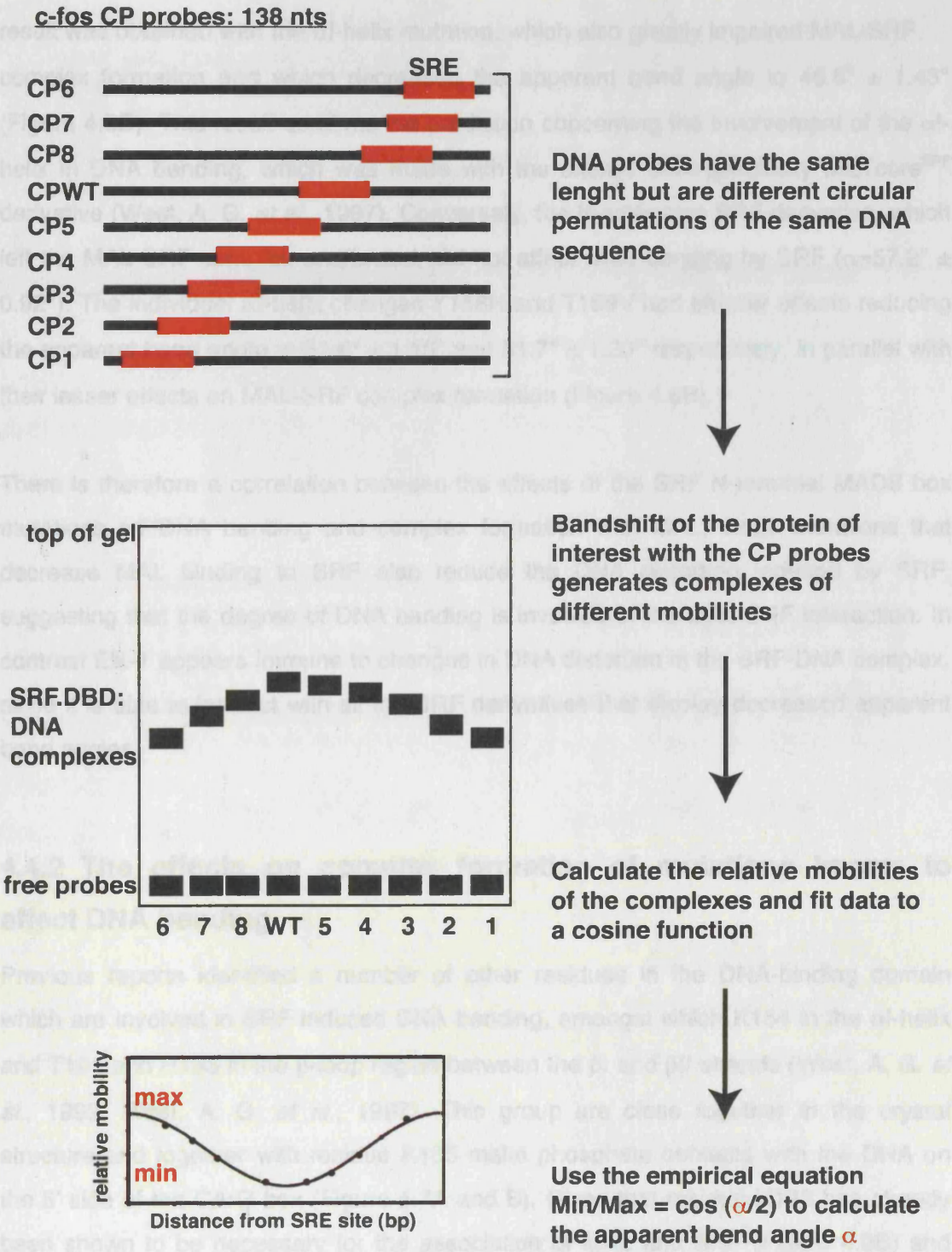


Figure 4.5. Experimental design of Circular Permutation analysis of the SRF.DBD derivatives with c-fos derived DNA probes. Circular permutation probes were generated by PCR with the c-fos promoter as template. Details of probe generation and programmes used for data manipulation can be found in Materials and Methods.

result was obtained with the αI-helix mutation, which also greatly impaired MAL-SRF complex formation and which decreased the apparent bend angle to 46.8° ± 1.43° (Figure 4.6B). This result confirms the prediction concerning the involvement of the αI-helix in DNA bending, which was made with the altered DNA-specificity METcore^{SRF} derivative (West, A. G. *et al.*, 1997). Conversely, the N-extension SRF derivative, which left the MAL-SRF complex unaffected, did not affect DNA bending by SRF (α=57.2° ± 0.92°). The individual αI-helix changes Y158H and T159V had smaller effects reducing the apparent bend angle to 51.6° ± 1.19° and 51.7° ± 1.20° respectively, in parallel with their lesser effects on MAL-SRF complex formation (Figure 4.6B).

There is therefore a correlation between the effects of the SRF N-terminal MADS box mutations on DNA bending and complex formation with MAL, since mutations that decrease MAL binding to SRF also reduce the DNA distortion induced by SRF, suggesting that the degree of DNA bending is involved in the MAL-SRF interaction. In contrast Elk-1 appears immune to changes in DNA distortion in the SRF-DNA complex, since it is able to interact with all the SRF derivatives that display decreased apparent bend angles.

4.4.2 The effects on complex formation of mutations known to affect DNA bending

Previous reports identified a number of other residues in the DNA-binding domain which are involved in SRF induced DNA bending, amongst which K154 in the αI-helix and T191 and H193 in the β-loop region between the βI and βII strands (West, A. G. *et al.*, 1999; West, A. G. *et al.*, 1997). This group are close together in the crystal structure and together with residue K165 make phosphate contacts with the DNA on the 5' side of the CArG box (Figure 4.7A and B). Given that residue H193 has already been shown to be necessary for the association of MAL and SRF (Figure 4.2B) and since the results described in the preceding section implicate DNA bending by SRF in complex formation with MAL, I next investigated whether there is a link between the effects of these residues in DNA bending and MAL binding.

Experiments with the altered DNA specificity SRF mutant METcore^{SRF} previously showed that negatively charged amino acid substitutions at these positions greatly reduce the magnitude of the SRF induced DNA bend, while changes to alanine affect

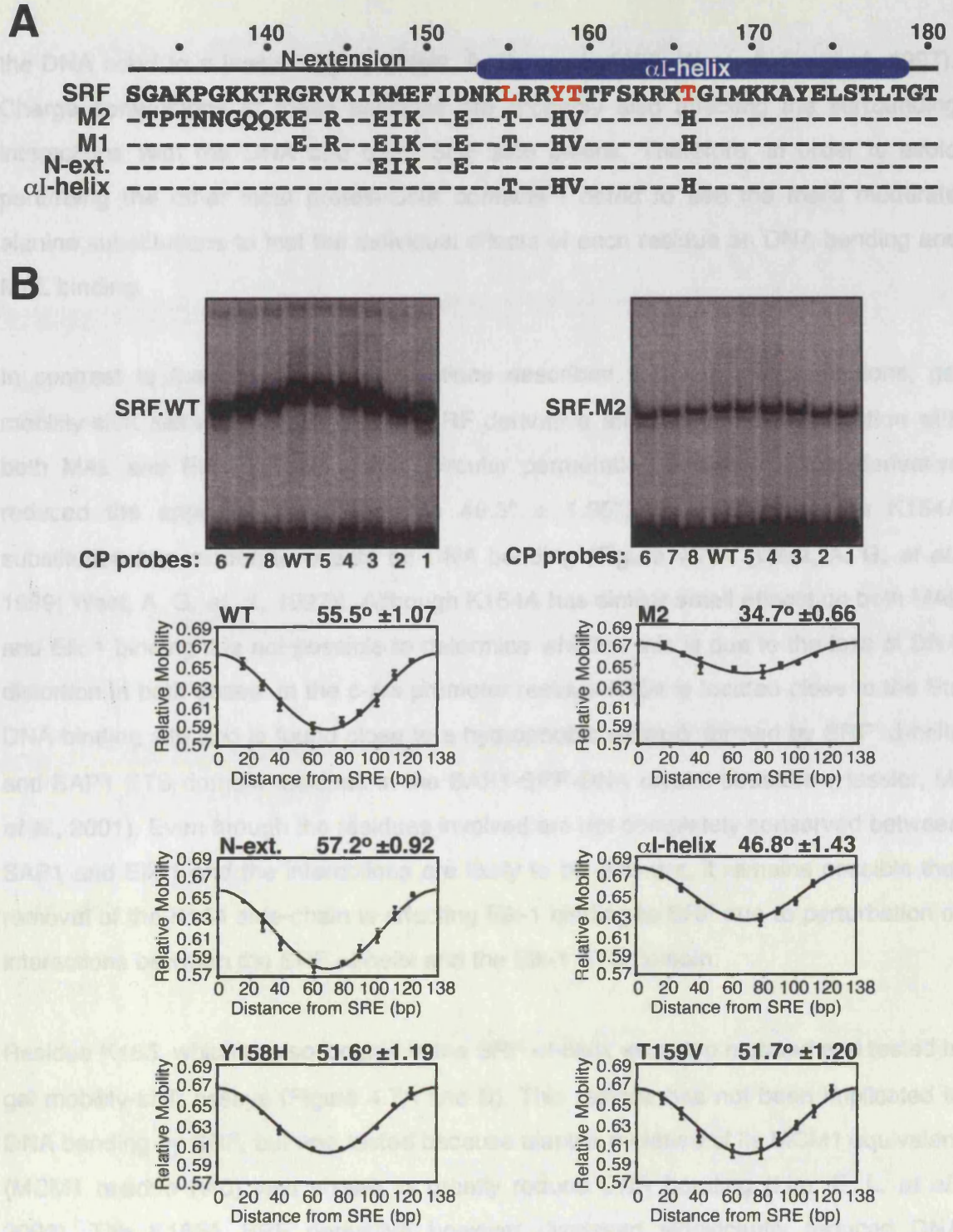


Figure 4.6. Mutations in the SRF αI-helix inhibit MAL-SRF complex formation through their effect on DNA bending. (A) Secondary structure elements, sequence and mutations in the N-terminal half of the SRF DBD. Critical αI-helix residues are highlighted in red. (B) Circular permutation analysis of SRF αI-helix derivatives. Gel mobility-shift assays contained SRF.DBD (residues 120-265) and the indicated CP probes. Plots represent the relative mobilities of the SRF-DNA complexes versus the distance of the centre of the SRE site from the probe end. Data were fitted to a cosine function to calculate the apparent bend angles (shown for each graph) and are presented as mean ±SEM of three independent experiments.

the DNA bend to a lesser degree (West, A. G. *et al.*, 1999; West, A. G. *et al.*, 1997). Charge conversions at these positions are probably also affecting the surrounding interactions with the DNA and other SRF side chains. Therefore, in order to avoid perturbing the other local protein-DNA contacts I opted to use the more moderate alanine substitutions to test the individual effects of each residue on DNA bending and MAL binding.

In contrast to the other αI-helix mutations described in the previous sections, gel mobility-shift assays with the K154A SRF derivative affected complex formation with both MAL and Elk-1 (Figure 4.7B). Circular permutation analysis of this derivative reduced the apparent bend angle to 49.3° ± 1.95°, confirming that the K154A substitution has moderate effects on DNA bending (Figure 4.7C; (West, A. G. *et al.*, 1999; West, A. G. *et al.*, 1997)). Although K154A has similar small effects on both MAL and Elk-1 binding it is not possible to determine whether this is due to the loss of DNA distortion in both cases. In the c-*fos* promoter residue K154 is located close to the Ets DNA binding site and is found close to a hydrophobic network formed by SRF αI-helix and SAP1 ETS domain residues in the SAP1-SRF-DNA crystal structure (Hassler, M. *et al.*, 2001). Even though the residues involved are not completely conserved between SAP1 and Elk-1 and the interactions are likely to be different, it remains possible that removal of the K154 side-chain is affecting Elk-1 binding to SRF due to perturbation of interactions between the SRF αI-helix and the Elk-1 ETS domain.

Residue K165, which is also located in the SRF αI-helix was also mutated and tested in gel mobility-shift assays (Figure 4.7A and B). This residue has not been implicated in DNA bending by SRF, but was tested because alanine mutation of its MCM1 equivalent (MCM1 residue K40) was shown to greatly reduce DNA bending (Lim, F. L. *et al.*, 2003). The K165A SRF derivative however displayed significantly reduced DNA binding, probably because it affects both DNA binding and SRF dimerisation (Pellegrini, L. *et al.*, 1995) and was therefore not included in the circular permutation experiments (Figure 4.7C).

In the SRF β-loop region, individual changes T191A and H193A, both substantially reduced complex formation with MAL but not Elk-1, but had only marginal effects on bending, decreasing the apparent bend angle to 52.2° ± 0.09° and 52.8° ± 1.01° respectively (Figure 4.7B and C). A double substitution at these positions, T191/H193A

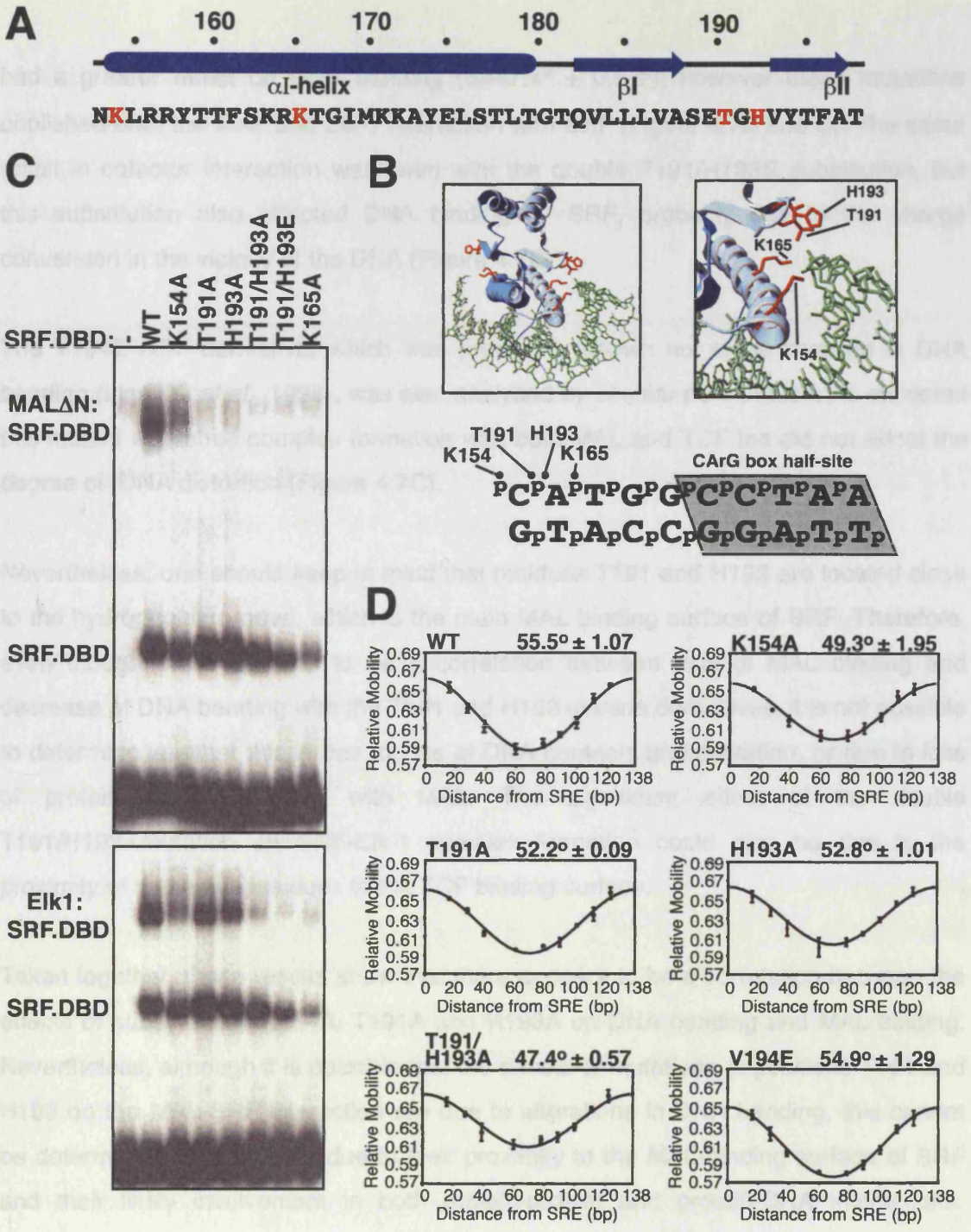


Figure 4.7. Effects of SRF.DBD mutations known to affect DNA bending on complex formation with MAL. (A) Secondary structure elements and sequence of the SRF DBD. Residues reported to affect DNA bending are shown in red. (B) The DNA contacts of residues K154, K165, T191 and H193. The SRF-DNA complex is shown (blue ribbon representation left, enlarged section on the right), with the DNA in green and the four side-chains in red shown in both SRF subunits. The phosphate DNA contacts are shown below on the sequence of the SRE site used in the crystal structure. Image adapted from the 1SRS PDB file (Pelegrini and Richmond 1995). (C) Complex formation of SRF.DBD DNA-bending derivatives with MAL and Elk-1. Gel mobility-shift assays contained SRF.DBD (residues 120-265) derivatives and MALΔN (top) or Elk-1 (bottom) extracts with c-fos SRE probe. (D) Circular permutation analysis. Data are presented as described in the legend to Figure 4.6.

had a greater effect on DNA bending (α=47.4° ± 0.57°); however these mutations abolished both the MAL and Elk-1 interaction with SRF (Figure 4.7B and C). The same result in cofactor interaction was seen with the double T191/H193E substitution, but this substitution also affected DNA binding by SRF, probably due to the charge conversion in the vicinity of the DNA (Figure 4.7B).

The V194E SRF derivative, which was previously shown not to be involved in DNA bending (Ling, Y. *et al.*, 1998), was also analysed by circular permutation. As expected this mutant abolished complex formation with both MAL and TCF but did not affect the degree of DNA distortion (Figure 4.7C).

Nevertheless, one should keep in mind that residues T191 and H193 are located close to the hydrophobic groove, which is the main MAL binding surface of SRF. Therefore, even though there appears to be a correlation between loss of MAL binding and decrease of DNA bending with the T191 and H193 alanine derivatives, it is not possible to determine whether this is due to loss of DNA contacts and distortion, or due to loss of protein-protein contacts with MAL. The significant effect of the double T191/H193Amutation on SRF-Elk-1 complex formation could also be due to the proximity of these two residues to the TCF binding surface.

Taken together, these results show that there appears to be a correlation between the effects of substitutions K154A, T191A and H193A on DNA bending and MAL binding. Nevertheless, although it is possible that the effects of mutations at positions T191 and H193 on the MAL-SRF interaction are due to alterations in DNA bending, this cannot be determined with certainty due to their proximity to the MAL-binding surface of SRF and their likely involvement in both protein-protein and protein-DNA interactions. Attempts to resolve this issue by testing the ability of the SRF mutants to interact with MAL in co-precipitation assays were not fruitful due to the low recovery of MAL-SRF complexes in the absence of DNA (see Chapter 5 and Discussion).

4.4.3 DNA bending in the MAL-SRF complex

Having established that the clustered αI-helix substitution decreases the degree of DNA distortion in the binary SRF-DNA complex, I next investigated whether it also affects DNA bending in the ternary MAL-SRF-DNA complex by performing circular

permutation analysis. The different protein composition between the ternary MAL-SRF-DNA complex and the binary SRF-DNA complex prohibits a direct quantitative comparison of their apparent DNA bend angles. Nevertheless, the assay can be used to compare DNA bending by complexes of equivalent protein content, which include different SRF derivatives.

The location of the MAL-SRF complex near the top of the native gels in bandshift assays precludes the use of the full-length protein for circular permutation analysis, since the loss of resolution at the top of the gel prevents the calculation of the differences in the relative mobilities of the complexes. I therefore used a small GST fusion protein of the MAL B1Q domains, GST.MAL(214-298), complexed with SRF.DBD(120-265) and the c-*fos* circular permutation DNA probes.

GST.MAL(214-298) formed complexes efficiently with SRF on all DNA probes, apart from the two probes (CP1 and CP6) in which the SRF binding site was centred 16 nucleotides from the end of the fragment (Figure 4.8; see Materials and Methods, section for CP probe details). This was also observed when full-length MAL was used (data not shown; analysed in detail in Chapter 5). For this reason the apparent bend angles of the complexes in this experiment were calculated using the data from the MAL-SRF complexes formed on the seven remaining probes. This had only marginal effects on the magnitudes of the apparent bend angles of the binary SRF-DNA complexes (the apparent bend angle from seven versus nine probes was 54.1° ±0.31° versus 55. 5° ±1.07 for wild-type SRF, and 45.39° ±3.01° versus 46.8o ±1.43° for the αI-helix derivative).

Complexes between the wild-type SRF DNA binding domain and GST.MAL(214-298) distorted DNA with an apparent bend of 55.6° ± 0.25° (Figure 4.8). The SRF αI-helix mutation left DNA distortion in the MAL-SRF-DNA complex unaffected, at 57.5° ± 1.25° (Figure 4.8, top), in contrast to its effect on DNA bending in the binary SRF-DNA complex, where a reduction of the apparent bend angle is observed (Figure 4.6B). This suggests that the interaction of MAL with the SRF(αI-helix)-DNA complex increases the extent of DNA distortion, either through direct MAL-DNA contacts, or through conformational changes in the complex induced by MAL binding. Thus a simple model in which MAL binding to SRF requires appropriate DNA distortion in the SRF-DNA complex could explain the inability of MAL to efficiently interact with the αI-helix mutant

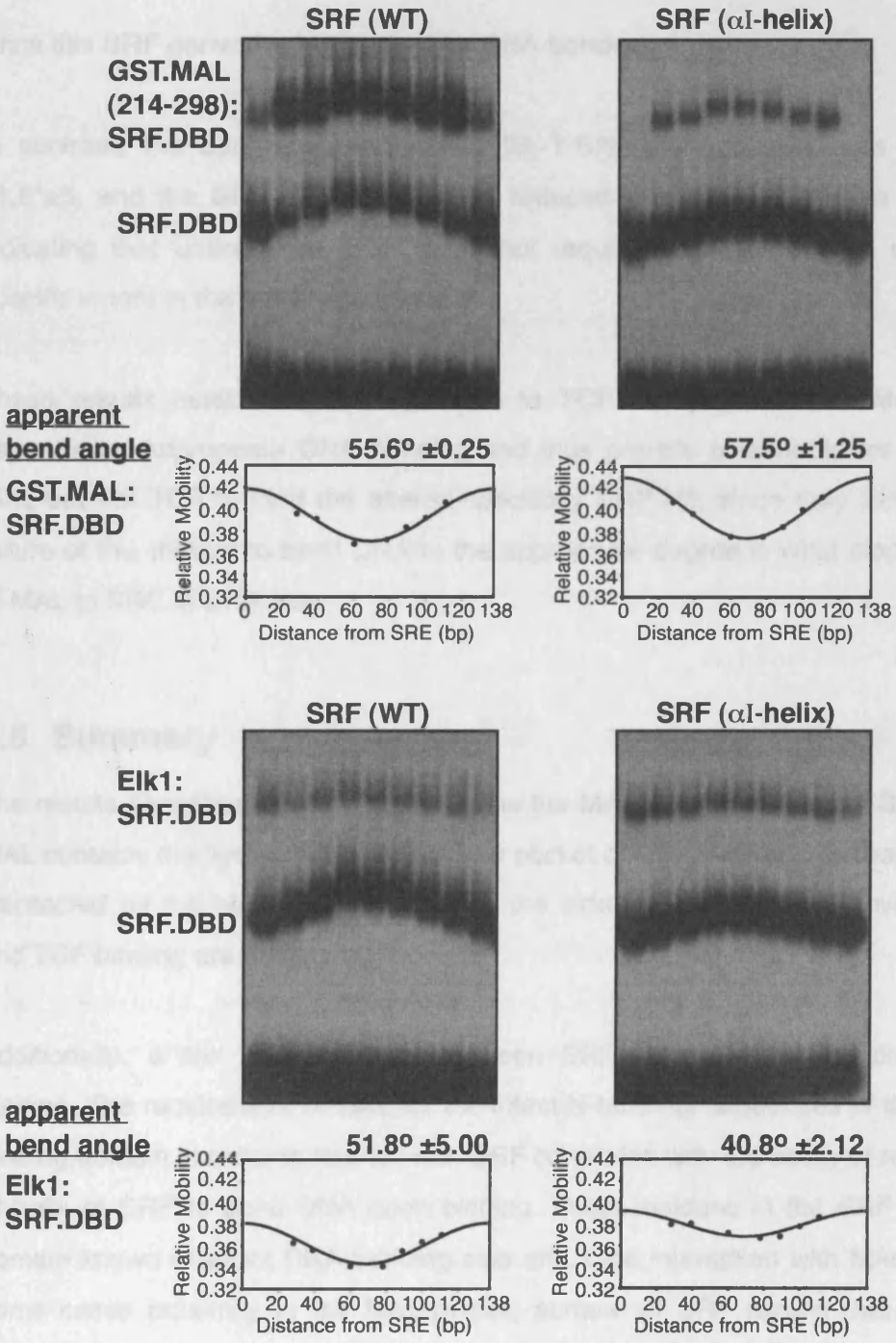


Figure 4.8. Optimal MAL-SRF interaction requires the induction of an appropriate DNA bend in the SRF-DNA complex. Circular permutation assays were performed using wildtype SRF.DBD or its αI-helix derivative (residues 120-265) and either GST.MAL(214-298) or transiently expressed Elk-1 as indicated. The data was analysed as described in the legend to Figure 4.5. Apparent bend angles for the ternary complexes and representative plots are shown below each panel. Note that MAL complex formation is specifically impaired on probes in which the SRF binding site is centred 16 basepairs from the fragment end.

since this SRF derivative is impaired for DNA bending.

In contrast, the apparent bend in the Elk-1-SRF-DNA complex was estimated at 51.8°±5, and the SRF αl-helix mutation reduced it to 40.8°±2 (Figure 4.8, bottom), indicating that unlike MAL, Elk-1 does not require that the DNA be distorted to a specific extent in the ternary complex.

These results establish that in contrast to TCF the interaction of MAL with SRF depends on appropriate DNA bending and thus provide a rationale for the failure of MAL but not TCF to bind the altered-specificity SRF.M2, since they indicate that the failure of this mutant to bend DNA to the appropriate degree is what stops recruitment of MAL to SRE.M DNA site.

4.5 Summary

The results described in this chapter define the MAL binding surface of SRF. Like TCF MAL contacts the hydrophobic groove and pocket of SRF. However, although the areas contacted by the two cofactors overlap, the side-chain interactions involved in MAL and TCF binding are subtly different.

Additionally, a link is established between SRF induced DNA bending and MAL binding. The requirement of MAL for the intact N-terminal sequences of the SRF DNA-binding domain in order to interact with SRF correlates with the ability of residues in the αl-helix of SRF to bend DNA upon binding. Other residues in the SRF DNA-binding domain known to affect DNA bending also affect the interaction with MAL, although in some cases proximity to the MAL-binding surface of SRF means that their precise effect on MAL binding is unclear.

Furthermore, appropriate distortion of the DNA in the SRF-DNA binary complex is a prerequisite for MAL binding, whereas the ability of SRF to bend DNA leaves the interaction with TCF largely unaffected, establishing a major difference in the ways the two cofactors bind SRF.

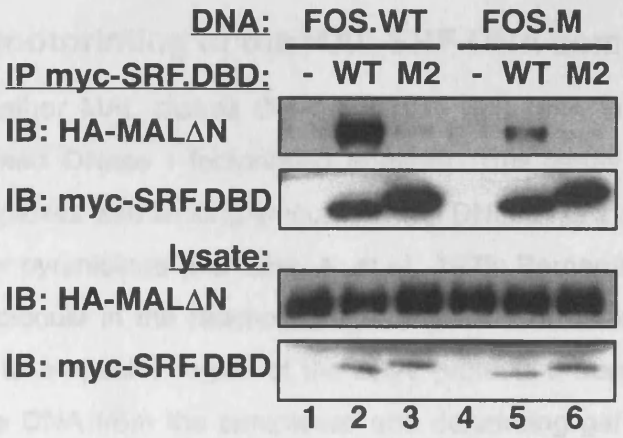
5 The role of DNA in the MAL-SRF interaction

5.1 Aims

The results presented in the previous chapter demonstrate that appropriate DNA distortion in the SRF-DNA complex is a prerequisite for MAL binding. A simple model to explain this requirement is that MAL contacts DNA directly, or that appropriate distortion induces a conformational change on SRF necessary for MAL complex formation. The aim of the present chapter was therefore to analyse the role of the DNA in the MAL-SRF interaction, by investigating how the presence of DNA affects MAL binding to SRF and studying whether the formation of the MAL-SRF-DNA complex involves direct MAL-DNA interactions.

5.2 Cognate DNA enhances the interaction of MAL and SRF

To study the requirement for DNA in MAL-SRF binding in detail I tested whether binding of SRF to its cognate DNA facilitates the MAL-SRF interaction by a co-immunoprecipitation approach. In vitro translated myc-tagged SRF.WT and SRF.M2 derivatives were combined with whole-cell extracts expressing HA-tagged MALΔN in the presence of either FOS.WT DNA or the mutated FOS.M DNA. Reactions were immunoprecipitated using myc-beads and immunoblotted for HA-MALΔN recovery. MALΔN was efficiently immunoprecipitated with wild-type SRF in the presence of the FOS.WT DNA that contains the wild-type SRF binding site, while its recovery was greatly reduced in the presence of the mutated SRF binding site FOS.M, which cannot bind wild-type SRF (Figure 5.1, compare lanes 2 and 5; (Hill, C. S. *et al.*, 1993)). These results demonstrate that MAL binding to SRF is facilitated when SRF is bound to its cognate DNA. The increased efficiency of MAL-SRF complex formation in the presence of DNA also provides a potential explanation for the consistently low recovery of MAL-SRF complexes in co-immunoprecipitation and GST-pulldown experiments. MALΔN failed to immunoprecipitate with SRF.M2 in the presence of either the wild-type or mutant c-fos DNA (Figure 5.1, lanes 3 and 6), confirming that MAL does not interact with SRF.M2 even when its cognate DNA is present, consistent with the gel mobility-shift assay results (Chapter 2).



FOS.WT TTACACAGGATGTCCATATTAGGACAT

FOS.M ----------CA--CG------

Figure 5.1. Cognate DNA enhances the interaction of MAL and SRF. Extracts containing HA-tagged MALΔN were incubated with the wild-type or SRF.M2 forms of myc-tagged in vitro translated SRF.DBD (120-265) as indicated, in the presence or absence of either the wild-type c-fos DNA probe (FOS.WT) or its mutated FOS.M derivative generated by PCR. Following immunprecipitation with anti-myc beads, MAL ΔN recovery was detected by HA immunoblotting. The sequences of the SRF binding sites are shown below, with the SRF binding site boxed in grey and the Ets binding site underlined.

5.3 MAL makes DNA contacts in the MAL-SRF-DNA complex

5.3.1 DNase I footprinting of the MAL-SRF-DNA complex

To investigate whether MAL makes direct contacts with DNA in the MAL-SRF-DNA complex, I performed DNase I footprinting analysis. This assay involves incubating nucleoprotein complexes with limiting amounts of the DNase I enzyme, which cuts DNA predominantly after pyrimidines (Bernardi, A. *et al.*, 1975; Bernardi, G. *et al.*, 1973), so that each DNA molecule in the reaction is cut only once on average. Binding of the protein of interest to a specific region of the DNA, protects it from DNase I cleavage. After elution of the DNA from the complexes and denaturing gel electrophoresis this becomes apparent as a gap in the DNA ladder, creating a characteristic protein "footprint". Protein binding to the DNA may also change its conformation exposing a neighbouring DNA region and making it more susceptible to DNase I digestion. This appears in the gel as a DNase I hypersensitive cleavage site.

To identify protein-DNA interactions in the MAL-SRF-DNA complex, c-fos promoter probes were radioactively labelled on the 5' end of their top strand or the 3' end of the bottom strand, and were then combined with recombinant SRF DNA-binding domain (residues 132-223), either alone or with increasing amounts of GST.MAL(214-298) derivatives. After incubation with limiting amounts of DNase I the DNA was eluted and resolved on denaturing gels alongside chemical degradation products of the probes that served as sequence markers. The SRF DNA-binding domain protected the CArG box and its flanking DNA sequences symmetrically to positions ±11 on the 5', and ±14 on the 3' side of each strand, consistent with the original characterisation of the classical SRE element (Figure 5.2; (Treisman, R., 1986)).

Inclusion in the reactions of increasing amounts of wild-type GST.MAL(214-298) induced additional changes in the DNase I digestion pattern symmetrically around the SRE dyad (Figure 5.2; summarised in Figure 5.3).

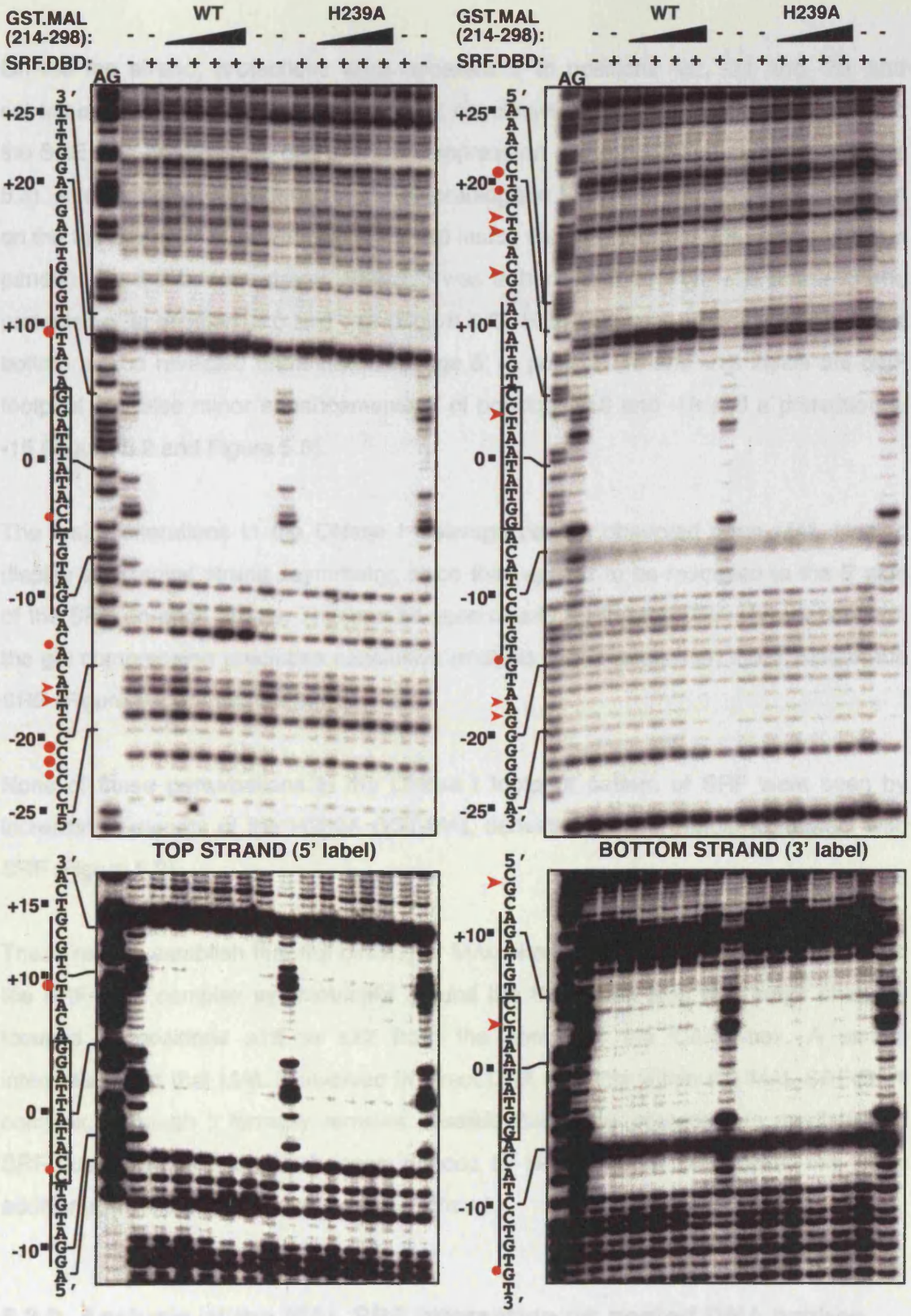


Figure 5.2. DNase I footprinting analysis of the MAL-SRF complex. Reactions contained c-fos DNA, SRF.DBD (residues 132-223) as indicated and increasing amounts (0.8, 2.5, 7.6 and 22.8 ng) of wild-type or H293A GST.MAL(214-298) proteins. The DNA sequences are aligned with the AG marker ladders, the line indicates the classical SRE and the CArG sequence is boxed. Red dots represent protections from and arrowheads enhancements of DNase I cleavage. Prolonged exposure autoradiograms of the footprints are shown below the main experiments.

On the top strand, protections were apparent 5' to positions -22, -21 and -20, and enhancements 5' to positions -17 and -16; conclusive interpretation of the pattern 3' to the SRE was not possible due to a gel compression (Figure 5.2; summarised in Figure 5.3). After prolonged exposure of the autoradiogram further protections were apparent on the top strand 5' to positions -4 and +10 inside the SRF footprint (Figure 5.2, bottom panels). On the bottom strand cleavage was enhanced 5' to +17, +16 and +13 and protected 5' to position +20 and +19 (Figure 5.2). High exposure autoradiograms of the bottom strand revealed enhanced cleavage 5' to position +3 and +13 inside the SRF footprint and also minor enhancements 5' of positions -18 and -19 and a protection at -15 (Figure 5.2 and Figure 5.3).

The major alterations in the DNase I cleavage pattern observed upon MAL binding display substantial strand asymmetry, since they appear to be restricted to the 5' side of the SRE on each strand. This can be seen clearly for the top DNA strand, although the gel compression precludes conclusive analysis of the pattern on the 3' side of the SRE (Figure 5.2 and Figure 5.3).

None of these perturbations in the DNase I footprint pattern of SRF were seen by increasing amounts of the H239A GST.MAL derivative that is unable to interact with SRF (Figure 5.2).

These results establish that the binding of MAL alters the DNase I cleavage pattern in the SRF-DNA complex symmetrically around the SRE dyad axis with major changes focused at positions ±16 to ±22 from the centre of the CArG box. A simple interpretation is that MAL is involved in direct DNA contacts within the MAL-SRF-DNA complex, although it formally remains possible that these contacts are mediated by SRF due to conformational changes induced by MAL binding. To address this issue additional DNA binding studies were performed.

5.3.2 Analysis of the MAL-SRF interaction on nested DNA probes

The DNase I footprinting analysis suggested that sequences outside the classical SRE are required for formation of the MAL-SRF complex. In order to directly investigate the role of these sequences in the interaction of MAL with SRF I performed further gel mobility-shift assays. MAL was previously shown to be unable to interact with SRF on

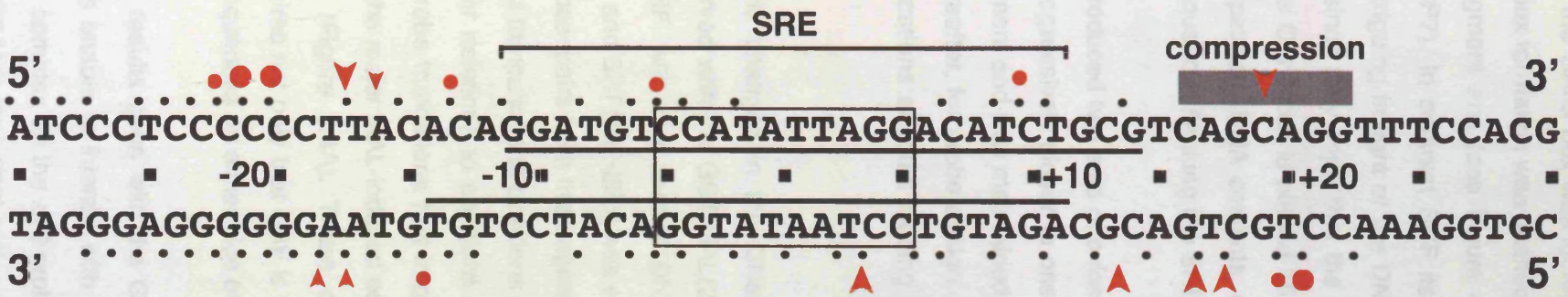


Figure 5.3. Summary of the DNase I footprinting data. The c-fos SRE region is shown, with black dots representing phosphodiester bonds where DNase I cleavage was readily detectable on naked DNA. The footprint of SRF alone is indicated by solid lines for simplicity and the core CArG sequence is boxed. The changes induced by MAL binding are shown by red symbols: circles represent proctection from DNase I cleavage and arrowheads cleavage enhancement. The grey bar with red arrowhead indicates the location of a gel compression that precludes interpretation of the changes on the 3' side of the SRE.

DNA probes where the CArG box was located 16 bp from either the 5' or 3' fragment end, whereas complex formation was unaffected when the CArG box was centred 27 or 28 bp from the fragment end (see Figure 4.8, compare probes CP1 and CP6 with probes CP2 and CP7). In contrast SRF itself bound all probes equally well. These observations are intriguing in light of the DNase I footprinting results described in the preceding section, since they imply that the inability of MAL to interact with SRF-DNA complexes when the CArG box is placed 16 nucleotides from the probe end is due to the elimination of upstream DNA contacts, consistent with the idea that MAL makes contacts with the sequences flanking the SRE.

To test this idea I produced two sets of c-*fos* derived probes, in which the SRF binding site was brought progressively closer to one fragment end in 3 base pair increments, while the other fragment end was maintained at the same position (Figure 5.4A; termed "nested probes" hereafter, for probe design details see Materials and Methods). None of these probe truncations affected binding of SRF to the DNA, as expected (Figure 5.4B).

Since the symmetric changes in the DNase I accessibility in the MAL-SRF-DNA complex were observed with the GST.MAL(214-298) protein, I tested the ability of this protein to bind SRF with the nested DNA probes. Complex formation between GST.MAL(214-298) and SRF(132-223) was efficient on probes in which the binding site was centred 25-28 basepairs from the fragment end (Figure 5.4B, probes -25, -27, +25 and +28). In contrast truncations to positions ±22 greatly decreased MAL-SRF complex formation and further deletions to positions ±19 and ±16 almost abolished it (Figure 5.4A and B). The probe truncations that impair MAL-SRF complex formation coincide with the location of the major MAL-induced perturbations in DNase I accessibility to the SRF-DNA complex (Figure 5.4A). These data provide further evidence that MAL binding to SRF requires not only that SRF is bound to the SRE but also that MAL itself contacts the DNA sequences on either side of this site.

Consistent with the results seen with the GST.MAL(214-298) protein on the nested probes, MALΔN was unable to interact with the ±16 probes and formed substantially reduced amounts of complex on the ±19 probes (Figure 5.4B). This demonstrates that the inability of the GST.MAL(214-298) fusion protein to bind SRF on the short probes is a bona fide property of the MAL protein.

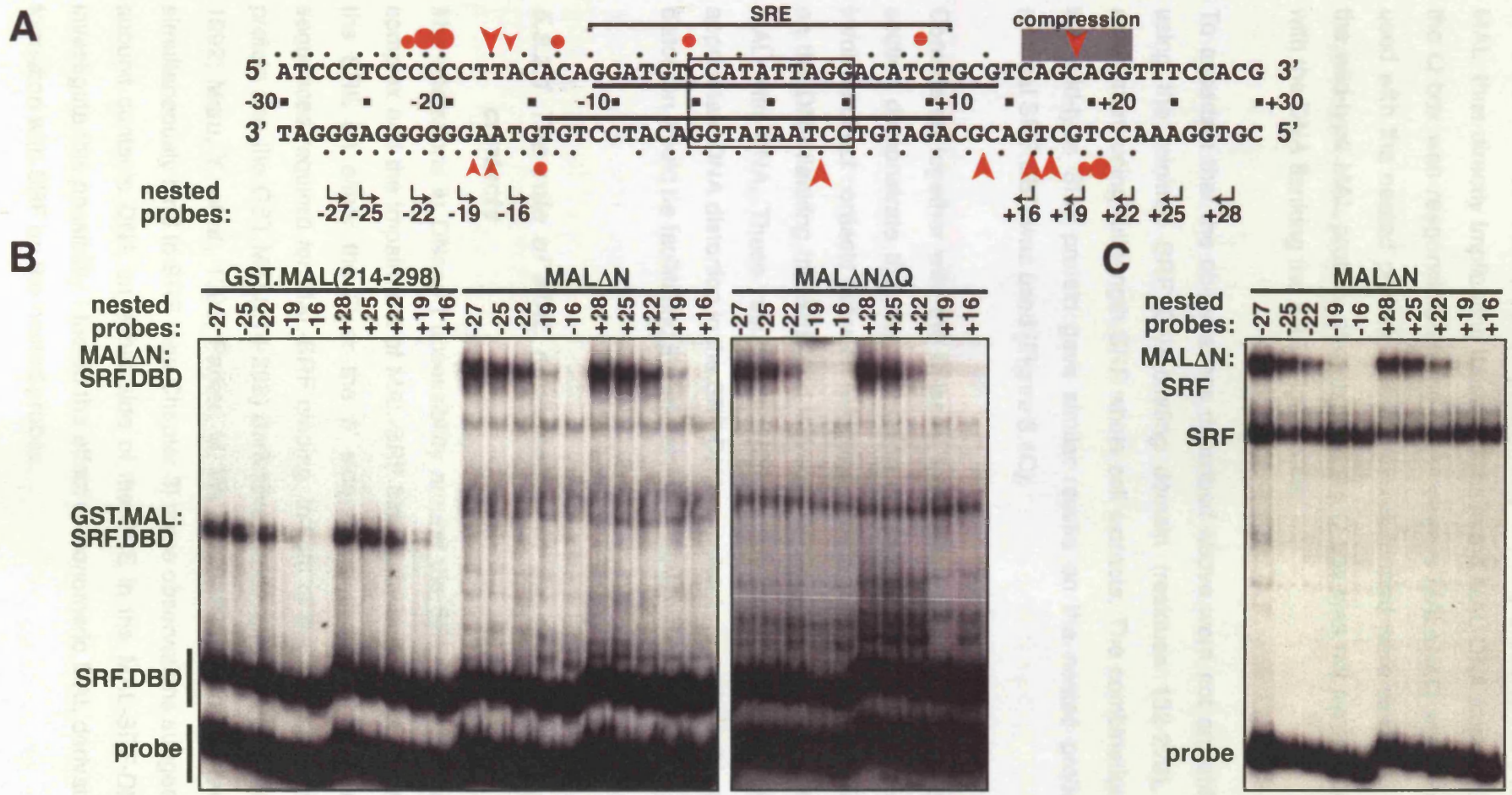


Figure 5.4. MAL-SRF complex formation on nested DNA probes. (A) The sequence of the c-fos DNA site. The SRE is shown with the CArG sequence boxed. Solid lines indicate the footprint of SRF alone, and red symbols represent DNase I cleavage enhancements (arrowheads) and protections (circles). The start positions and direction of the nested bandshift probes in relation to the SRE dyad are shown. (B) Complex formation between MAL derivatives and the SRF DNA-binding domain on the nested probes. Bandshift reactions contained purified SRF.DBD (132-223) and either GST.MAL(214-298) or MALΔN or MALΔNΔQ whole-cell extracts and the indicated probes. (C) Complex formation between MALΔN and full-length SRF on the c-fos nested probes. Gel mobility-shift assays contained whole-cell extracts expressing MALΔN and SRF and the indicated probes.

The GST.MAL(214-298) fusion protein only encompasses the B1 and Q regions of MAL, thus directly implicating these regions in the MAL-DNA contacts. To test whether the Q box was responsible for the DNA contacts MALΔNΔQ whole-cell extracts were used with the nested probes. The complexes formed were similar to those seen with the wild-type MAL protein, showing that the Q box was not responsible for interacting with the DNA flanking the SRE (Figure 5.4B).

To ascertain that the observations described above were not an artifact produced by using the minimal SRF DNA-binding domain (residues 132-223), I repeated the experiment using full-length SRF whole cell extracts. The combination of MALΔN with the wild-type SRF protein gave similar results on the nested probes as when the minimal SRF.DBD was used (Figure 5.4C).

Considered together with the DNase I footprinting data, the results presented in this section demonstrate that the interaction of MAL with the binary SRF-DNA complex involves direct contacts between MAL and DNA in the vicinity of positions ± 16 to ± 22 on the DNA flanking the SRE and indirectly implicate the B1 box in the interaction of MAL with DNA. These results also provide an explanation for the requirement of appropriate DNA distortion in the SRF-DNA complex in order for MAL to bind, since this distortion would be facilitating the interaction of MAL with DNA.

5.3.2.1 The role of MAL dimerisation in the formation of the MAL-DNA contacts

The alterations in DNase I accessibility around the SRE dyad in the MAL-SRF-DNA complex and the impairment of MAL-SRF binding by truncations of the DNA bordering the SRE on either the 5' or the 3' side reveal a striking symmetry in the DNA sequences required for MAL-SRF binding. In light of the fact that both wild-type MAL protein and the GST.MAL(214-298) derivative are dimeric (Chapter 2 and (Ji, X. *et al.*, 1992; Maru, Y. *et al.*, 1996; Parker, M. W. *et al.*, 1990)) and that two B1 boxes can simultaneously bind to SRF (see Chapter 3) these observations suggest that each MAL subunit contacts DNA on either side of the SRE in the MAL-SRF-DNA complex. To investigate this possibility I tested the effect of monomeric MAL derivatives on complex formation with SRF on the nested probes.

Whole cell-extracts expressing the monomeric MALΔNΔLZ (see Chapter 2), bound SRF efficiently on probes -27, +28, -25 and +25 (Figure 5.5A). In contrast to what was seen with MALΔN, further truncations of the DNA probes to positions ±22, ±19 and ±16 gradually reduced but did not abolish complex formation between MALΔNΔLZ and SRF (Figure 5.5A), consistent with each monomer making contacts at only one side of the SRE.

Since Myocardin is predominantly monomeric in complex with SRF (see Chapter 2, Fig. 2.4), I proceeded to test its ability to interact with SRF on the nested probes. Transiently expressed full-length Myocardin bound the nested probes with reduced efficiency, and complex formation was gradually decreased but not completely impaired, as observed with the MALΔLZ derivative (Figure 5.5B).

I next tested the ΔN Myocardin derivative, which in gel mobility-shift assays forms two SRF interacting species, one monomeric co-migrating with the full-length Myocardin-SRF complex and one dimeric migrating at roughly the position of the MAL-SRF complex (see Chapter 2, Fig. 2.4). Higher amounts of MCΔN whole-cell extract were used in this experiment to adequately visualise both SRF-Myocardin complexes.

Myocardin ΔN interacted strongly with SRF on the longer probes, and formed a smear that covered both the dimer and monomer positions (Figure 5.5B, compare with MAL-SRF complexes in Figure 5.5A). This was probably due to a combination of the high amounts of Myocardin in the extracts and the instability of the interaction between the dimeric Myocardin and SRF. Sequential truncations of the probes to ±25 and then ±22 nucleotides from the centre of the SRE decreased the amount of MCΔN bound to SRF and induced the formation of a discrete MCΔN-SRF complex corresponding to the monomeric MC position (Figure 5.5B, compare complexes formed on probes ±25 and ±22). Further deletions of the probes to positions ±19 and ±16 almost abolished the smear corresponding to the dimeric MCΔN-SRF interaction, but only slightly decreased the amount of the monomeric MCΔN-SRF complex (Figure 5.5B).

Thus, the monomeric forms of MAL and Myocardin can tolerate the loss of DNA sequences bordering on one side of the SRE. This implies that when complexed with the longer nested probes the MAL or Myocardin monomer is able to exchange SRF subunit partners in the complex and interact with DNA on either side of the SRE.

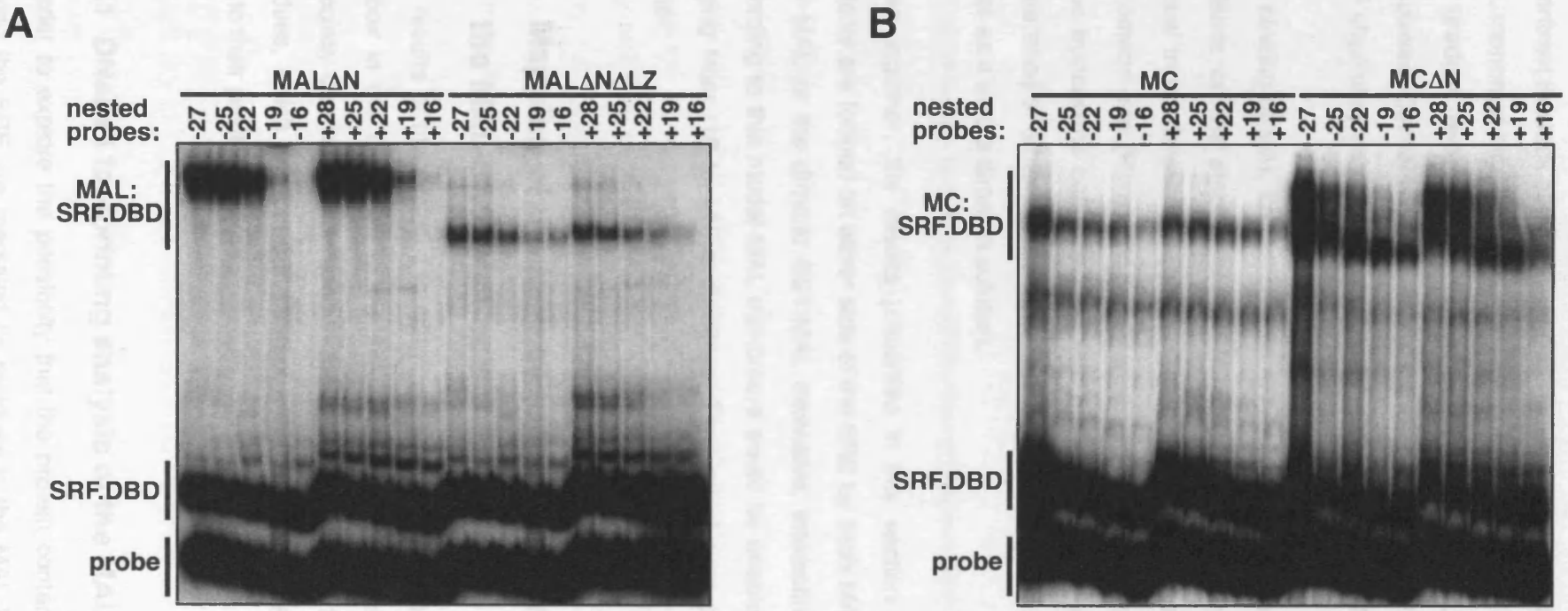


Figure 5.5. Complex formation by MALΔLZ and Myocardin derivatives on the nested probes. (A) Gel mobility-shift assays contained the indicated MALΔN whole-cell extracts and recombinant SRF (residues 132-223), with the indicated c-fos SRE nested probes. (B) Gel mobility-shift assays contained the indicated Myocardin extracts and SRF.DBD (132-223) with the c-fos nested probes. Twice the normal amount of Myocardin extract was used in this experiment.

In contrast the lack of the critical DNA sequences in the shorter probes would constrain MAL monomer binding to only one side of the SRF-DNA complex, thus accounting for the gradual reduction in the amounts of MALΔNΔLZ-SRF and Myocardin-SRF complexes. Conversely these results show that binding of dimeric MAL or Myocardin to SRF absolutely requires that both subunits interact with SRF and DNA in the complex.

The binding of MAL and Myocardin ΔN derivatives on the nested probes also provides evidence on the stability of MRTF homodimerisation. In the case of Myocardin ΔN gradual truncations of the probes result in the loss of the dimeric complex but not the monomeric one (Figure 5.5B), showing that homodimerisation is unstable. In contrast probe truncations completely abolish the interaction of MALΔN and SRF without the appearance of a monomeric MAL-SRF complex (Figure 5.5A), demonstrating that MAL exists as a stable dimer in solution.

Taken together, the results presented in this section indicate that the MAL-DNA contacts are formed on either side of the SRE by each MAL subunit of the dimeric wild-type MAL or the dimeric GST.MAL derivative, interacting with each SRF monomer. According to this model MAL monomers would be unable to bind to very short probes ranging from -16 to +16 nucleotides. Time constraints precluded pursuing this point further.

5.4 Mapping of the MAL sequences mediating DNA contacts in the MAL-SRF-DNA complex

The results of the DNase I footprinting and nested probe analyses implicate the MAL B1 box in the formation of the MAL-DNA contacts. This is an intriguing possibility especially due to the presence in the N-terminal part of the B1 box of multiple lysine residues, which are found in contact with the DNA in many nucleoprotein complexes due to their positive charge.

5.4.1 DNase I footprinting analysis of the MAL B1 peptides

In order to explore the possibility that the protein contacts with the DNA sequences flanking the SRE are mediated by residues in the MAL B1 box I performed DNase I

footprinting experiments using different MAL peptides (Chapter 3; Figure 5.6A) in complex with the SRF DNA-binding domain and c-fos DNA.

The MAL B1Q peptide (residues 224-283), which encompasses the intact B1 region, generated a footprint essentially identical to that produced by GST.MAL(214-298) (Figure 5.6B). In contrast, use of peptide **A**, which lacks the N-terminal residues of the B1 region, resulted in alterations in the pattern of DNase I cleavage, the most striking of which was an additional enhanced cleavage site on the bottom strand 5' to position -16 (Figure 5.6B, summarised in panel C; note that cleavage 5' to -15 was also observed after prolonged autoradiogram exposure, data not shown). Symmetrical changes in the DNase I cleavage pattern of the top strand could not be interpreted due to a gel compression (Figure 5.6B).

Further changes in the MAL-SRF footprint included the loss of the protections seen with GST.MAL and the B1Q peptide on the bottom strand, while only two of the previously identified DNase I hypersensitive sites were seen 5' to positions +17 and +3 (Figure 5.6B). On the top strand, the peptide **A**-SRF footprint lacked protections 5' to positions -22, -21, -20 and -13, but retained the enhanced cleavage sites 5' to positions -17 and -16 (Figure 5.6B). Prolonged autoradiogram exposure revealed a loss of the protections 5' to positions -4 and +12 inside the SRF footprint (data not shown). Footprinting of peptides **M** and **J**, which are further N-terminally truncated, generated DNase I digestion patterns identical to the one seen with peptide **A**, indicating that the removed residues, including K230 and K232 which were previously seen to affect the affinity of the MAL-SRF interaction in peptide competition experiments (Chapter 3) do not participate in direct MAL-DNA contacts detectable by DNase I footprinting (Figure 5.6A and B). These results imply that residues closer to the core B1 seven-residue sequence, such as K234 or K235 (see following section), are responsible for the perturbed footprint.

The changes in the DNase I footprinting pattern seen with the B1 peptides that lack the N-terminal part of the basic region suggest that MAL residues 224-228 are required for authentic MAL-DNA contacts. These data however do not address whether these sequences are themselves responsible for contacting DNA, or whether their absence simply perturbs DNA contacts made by other B1 residues (see Discussion).

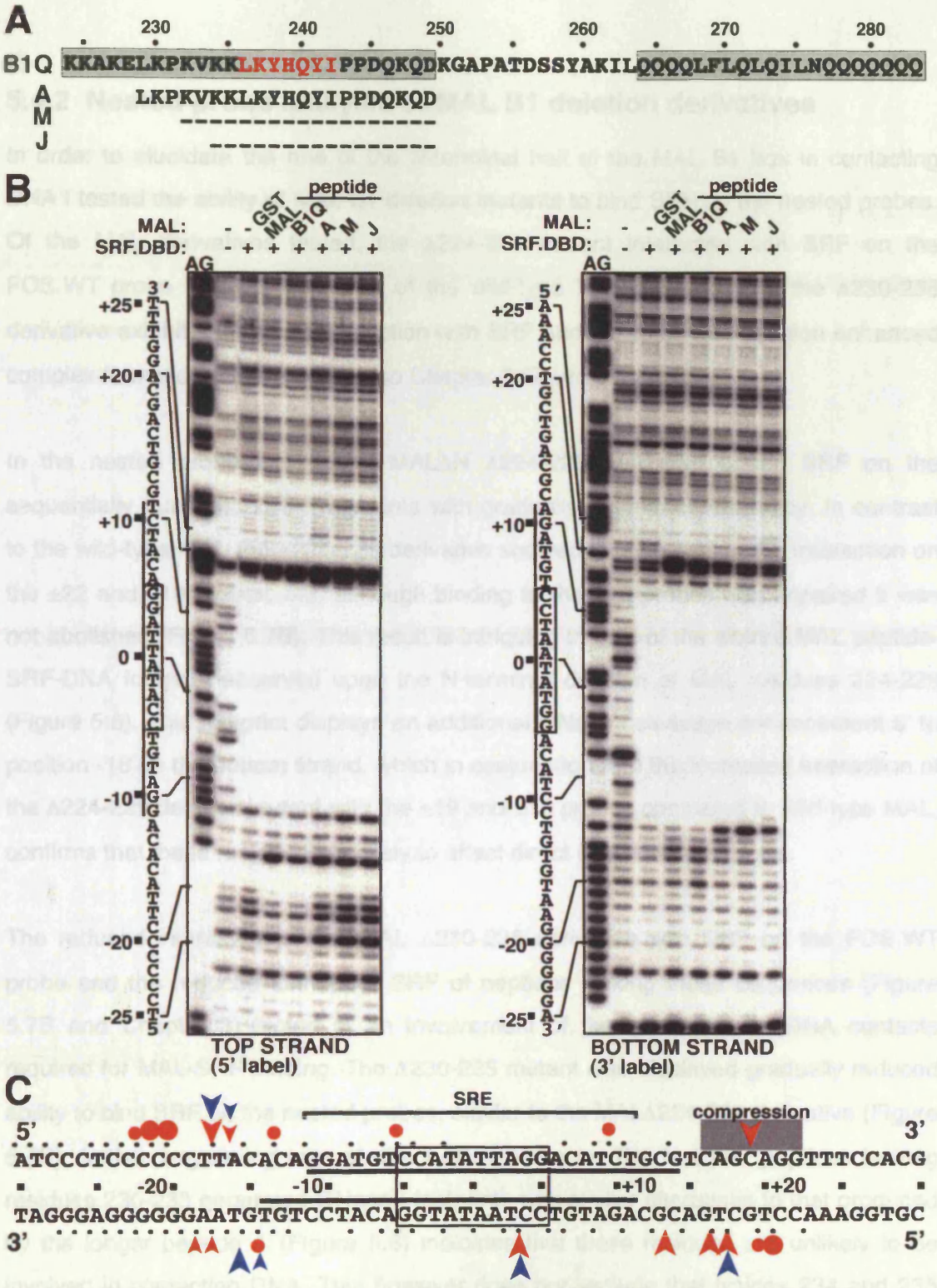


Figure 5.6. DNase I footprinting analysis of the MAL B1 region peptides in complex with SRF. (A) Sequences of the MAL peptides. The B1 and Q regions are boxed in grey and the critical B1 residues are red. (B) Footprinting reactions contained c-fos DNA and SRF.DBD (residues 132-223) with 7.6ng GST.MAL(214-298), 1.7μM peptide B1Q, 8.5μM peptides A and M and 85μM peptide J as indicated. The DNA sequences are aligned with the AG marker ladders. (C) Summary of the DNase I footprinting data.The c-fos SRE is shown with the CArG sequence boxed. The SRF footprint is shown with a line and the MAL induced changes in DNase I cleavage patterns with circles (protections) and arrowheads (enhancements); red symbols, GST.MAL (214-298) and B1Q peptide; blue symbols, peptides A, M and J.

5.4.2 Nested probe analysis of MAL B1 deletion derivatives

In order to elucidate the role of the N-terminal half of the MAL B1 box in contacting DNA I tested the ability of MAL B1 deletion mutants to bind SRF on the nested probes. Of the MAL derivatives tested, the Δ224-229 mutant interacted with SRF on the FOS.WT probe with the efficiency of the wild-type MAL protein, while the Δ230-235 derivative exhibited reduced interaction with SRF and the Δ224-235 deletion enhanced complex formation (Figure 5.7B, also Chapter 3 Figure 3.1).

In the nested probe assay, the MALΔN Δ224-229 derivative bound SRF on the sequentially truncated DNA fragments with gradually decreased efficiency. In contrast to the wild-type MAL the Δ224-229 derivative showed little change in the interaction on the ±22 and ±19 probes, and although binding to the ±16 probes was impaired it was not abolished (Figure 5.7B). This result is intriguing in light of the altered MAL peptide-SRF-DNA footprint observed upon the N-terminal deletion of MAL residues 224-228 (Figure 5.6). This footprint displays an additional DNase I cleavage enhancement 5' to position -16 on the bottom strand, which in conjunction with the increased interaction of the Δ224-229 deletion mutant with the ±19 and ±16 probes compared to wild-type MAL, confirms that these residues are likely to affect direct MAL-DNA contacts.

The reduced interaction of the MAL Δ230-235 derivative with SRF on the FOS.WT probe and the reduced affinity for SRF of peptides lacking these sequences (Figure 5.7B and Chapter 3) hinted at an involvement of these residues in DNA contacts required for MAL-SRF binding. The Δ230-235 mutant also displayed gradually reduced ability to bind SRF on the nested probes, similar to the MALΔ224-229 derivative (Figure 5.7B), further suggesting loss of direct DNA contacts. The fact that peptides lacking residues 230-233 generated DNase I footprints with similar alterations to that produced by the longer peptide A (Figure 5.6) indicates that these residues are unlikely to be involved in contacting DNA. This however does not exclude that lysines 234 and 235 which are also deleted in the Δ230-235 derivative are involved in DNA interactions.

I next tested the MALΔ224-235 derivative, which lacks the complete N-terminal half of the B1 box. This MAL derivative formed increased amounts of complex with SRF (Chapter 3, Figure 3.1 and Figure 5.7B). Use of MAL Δ224-235 in the nested probe assay generated surprising results, since this derivative was able to interact efficiently

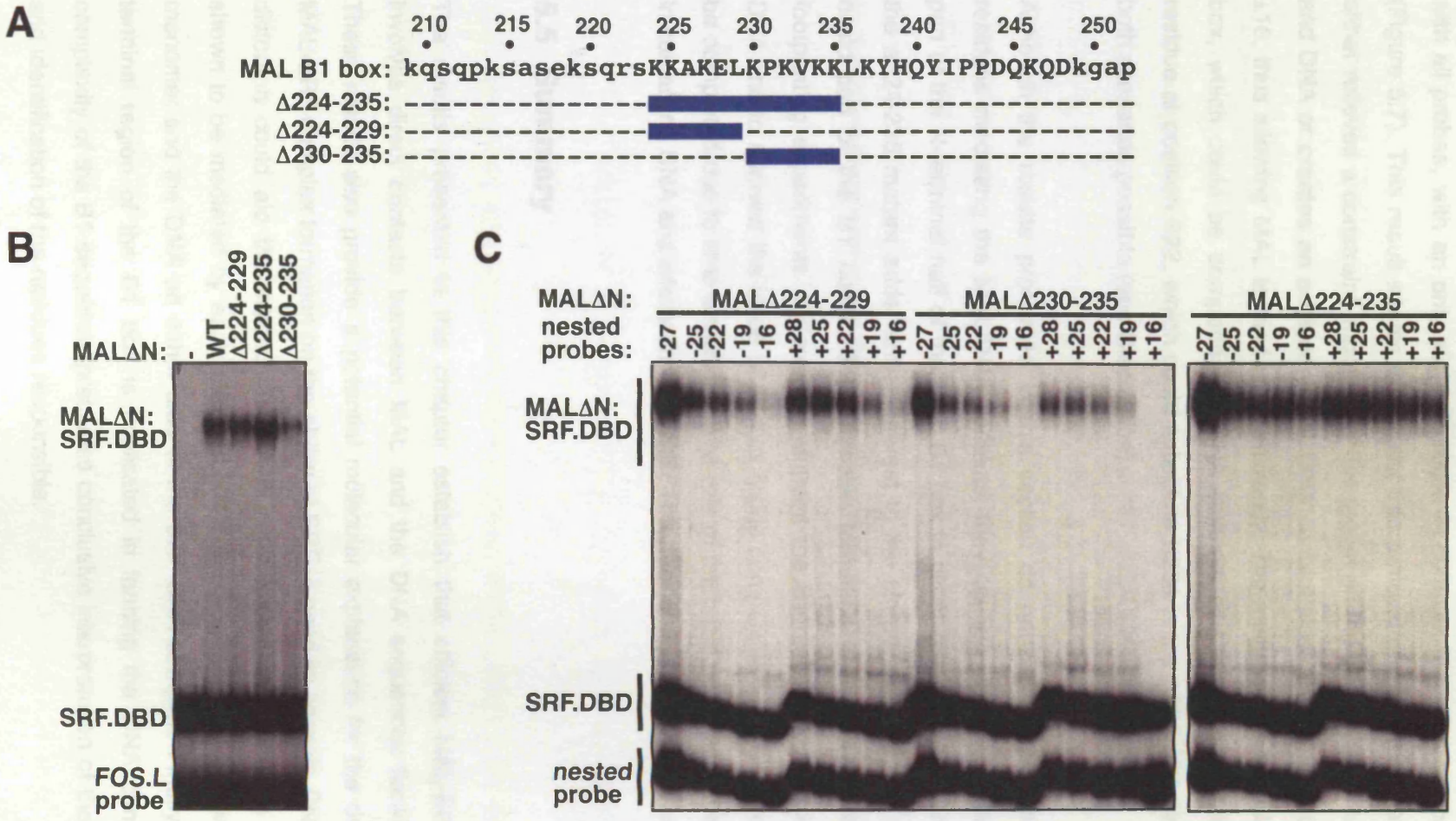


Figure 5.7. Complex formation by MAL B1 deletion derivatives on the nested probes. (A) The sequence of the B1 region is shown in capital letters. Deletion mutations are indicated by blue bars. (B) Interaction of the MAL B1 region. Gel mobility-shift assays contained the indicated MALΔN whole-cell extracts, recombinant SRF(132-223) and the ΔTCF c-fos probe (FOS.L). (C) Gel mobility-shift assays contained the indicated MALΔN extracts and recombinant SRF (residues 132-223), with the indicated c-fos SRE nested probes.

with all probes, with an only slight decrease in complex formation on the ±16 probes (Figure 5.7). This result seems to suggest that removal of the N-terminal B1 residues either relieves a constraint imposed by this region on the interaction of MAL with SRF and DNA or creates an additional MAL-DNA contact closer to the SRE than positions ±16, thus allowing MAL to bind more efficiently. The multiple lysine residues in the B1 box, which could be competing for DNA contacts and the presence of an arginine residue at position 222, which could substitute K234 in the Δ224-235 derivative make both scenarios possible (see Discussion).

Although the results presented in this section do not provide evidence as to the residues mediating the MAL-DNA contacts, they confirm the involvement of at least part of the N-terminal half of the MAL B1 box in direct DNA contacts. The analysis of the Δ224-235 mutant adds further interest to the already contradicting effects of the basic part of the B1 box in SRF complex formation. Gel mobility-shift assay and footprinting experiments designed to pinpoint the B1 residue or residues contacting DNA and to address the issue of B1 box lysine competition for DNA binding could not be completed due to time limitations. The role of the basic residues of the MAL B1 box in contacting DNA and interacting with SRF is further analysed in the Discussion.

5.5 Summary

The results presented in this chapter establish that efficient MAL-SRF interaction involves direct contacts between MAL and the DNA sequences flanking the SRE. These results also provide a potential molecular explanation for the dependence of MAL-SRF complex formation on the ability of SRF to bend its cognate DNA, since such distortion could aid the interaction of MAL with DNA. The MAL-DNA contacts are shown to be mediated by each MAL subunit interacting symmetrically with each SRF monomer and the DNA on either side of the SRF-DNA complex. Finally the basic N-terminal region of the B1 box is implicated in forming the DNA contacts but the complexity of the B1 sequence precludes conclusive interpretation of the present data and identification of the residues responsible.

6 Discussion

6.1 Summary

In the present thesis I have analysed the interaction between the SRF transcription factor and its cofactor MAL, a member of the Myocardin related transcription factor family. The results presented here show that the MAL-SRF complex has the predicted properties of the Rho-actin regulated SRF cofactor and extend these findings. The molecular mechanism of MAL-SRF complex formation has been characterised in detail through the analysis of the MAL and SRF interaction surfaces and of the role played by DNA contact and distortion in complex formation has been analysed.

In the following sections the different aspects of MAL-SRF complex formation will be discussed and a model for the interaction of SRF with the MRTFs will be presented. Additionally the importance of cofactor competition and exchange in the execution of the diverse transcriptional programmes regulated by SRF will be addressed.

6.1.1 The effect of different MRTF domains in the interaction with SRF

Both the MAL(fl) and MAL(BSAC) isoforms form SRF-dependent complexes on DNA fragments containing wild-type SRF binding sites. Efficient formation of these complexes depends on the presence of the B1 and Q regions and GST-fusion proteins encompassing these regions can form specific complexes with recombinant SRF (Chapter 3), showing that MAL directly contacts SRF. This contradicts a previous report where the inability to detect MAL(BSAC)-SRF complexes lead to the proposal that an unknown molecule other than SRF was targeting MAL to CArG box containing DNA (Sasazuki, T. *et al.*, 2002). It is likely that the lack of MAL(BSAC)-SRF complexes in this study reflected a combination of the different experimental conditions used in the gel mobility-shift assay and the low amounts of complex formed by full-length MAL proteins.

6.1.1.1 Inhibition of complex formation by the RPEL domain

Different MAL forms that include the N-terminal RPEL motifs only form small amounts of complex with SRF. Removal of the extreme N-terminal sequences including RPEL1 increases SRF binding and this effect is further enhanced by deletion of the complete N-terminus of the protein.

How could the N-terminal region of MAL affect complex formation with SRF? One possibility is that inhibition of SRF binding is due to the interaction of the three RPEL motifs with actin present in the extracts. Dissociation of actin from MAL is a prerequisite for transcriptional activity of the protein, and preliminary data indicate that actin cannot interact with the MAL-SRF complex (Sebastian Guettler, personal communication).

The different binding competition strategies employed to test this possibility were not fruitful. Titration of RPEL domain in order to compete with full-length MAL for actin binding had no effect on the MAL-SRF complex. It is conceivable however that due to the abundance of actin in the whole-cell extracts higher concentrations of RPEL domain are required to alleviate its negative effect on the MAL-SRF interaction. It is noteworthy that titration of Cytochalasin D in the bandshift reactions in an attempt to disrupt actin-MAL binding and increase complex formation with SRF also had no effect. The high actin protein levels in the extracts could also have influenced the outcome of this experiment.

Use of purified actin in an attempt to compete with SRF for MAL binding was also unsuccessful due to technical difficulties raised by the fact that the actin-polymerisation and bandshift conditions coincide. This experiment is currently being repeated with unpolymerisable actin forms that are still able to interact with MAL, such as the R62D actin mutant (Posern, G. *et al.*, 2002) or actin purified in the presence of Latrunculin B, a drug that sequesters actin monomers and blocks the Rho pathway (Sotiropoulos, A. *et al.*, 1999). Latrunculin B treated actin is also unable to polymerise but retains the ability to bind MAL. The issue of actin interfering with MAL-SRF binding could be unequivocally solved by the use of purified wild-type and N-terminally truncated MAL derivatives with recombinant SRF, in gel mobility-shift assays. If actin is responsible for complex inhibition, the purified MAL proteins would be expected to form equal amounts of complex with SRF, irrespective of the presence of the RPEL motifs. Titration of unpolymerisable actin would then be predicted to compete with SRF for MAL-binding.

Although the competition binding experiments could neither prove nor oppose the possibility of an actin-imposed inhibition of MAL-SRF complex formation, two lines of evidence make the involvement of actin unlikely. MAL derivatives containing mutations in the RPEL motifs have variable effects on SRF complex formation, despite being equally compromised for actin binding. Furthermore N-terminal truncations of Myocardin, whose affinity for actin is minimal (Sebastian Guettler, personal communication), also exhibit enhanced interaction with SRF. It therefore appears that the N-termini of MRTF family members act in a similar way to inhibit SRF binding independently of their ability to bind actin.

One possibility is that the N-terminal domain is involved in intraprotein interactions that mask the SRF-binding surface of MAL. It should be noted however that attempts to investigate potential interactions between the RPEL domain and other regions of MAL by GST-pulldown assays have not produced any positive results (Sebastian Guettler, personal communication). Thus although the data so far show that the structural integrity of the N-terminal domain is required for the inhibition of MAL binding to SRF, they do not establish a mechanism for this effect.

6.1.1.2 The B1 box

Complex formation between SRF and MAL or Myocardin depends on the B1 region, which is required for direct protein-protein interactions with the SRF DNA-binding domain (Chapters 2 and 3). Alanine scanning mutagenesis of the MAL B1 box identified a seven-residue sequence that is critical for interaction with SRF. Furthermore the aromatic and planar properties of residues at key positions within this sequence are crucial for SRF binding (see also Section 6.1.2.1). This seven-amino acid stretch is conserved between MRTF family members and the analogous DMRTF sequence is able to substitute for that of MAL in the context of the murine MAL protein, indicating that the mode of interaction of different MRTFs with SRF is identical. Short peptides encompassing this core sequence are sufficient for specific SRF binding, demonstrating that this conserved predominantly hydrophobic seven residue sequence constitutes the minimal SRF-interaction surface of the MRTFs.

Peptide competition studies indicate that basic B1 residues N-terminal to this core sequence contribute to the affinity of complex formation. Nevertheless, the alanine scanning mutagenesis of the B1 box shows that individual basic residues in this region are dispensable for specific interaction with SRF. Indeed alanine substitution of many of these residues increases rather than decreases complex formation (discussed further in Section 6.1.3).

In addition to its role in SRF complex formation, the B1 region affects nuclear accumulation of MALΔN. Three B1 residues, K234, K235 and K237 are required for this function. These functions are separable since the ability of different B1 mutant derivatives to bind SRF does not correlate with nuclear accumulation, with only residue K237 affecting both functions. These observations indicate that the B1 sequence acts as a nuclear localisation signal, as opposed to a nuclear export signal occluded upon SRF binding or a sequence promoting MAL nuclear retention through SRF interaction.

Functional studies demonstrate that residues L236, Y238, H239 and Y241, which are absolutely required for in vitro complex formation with SRF, are also critical for the interaction in intact cells, since alanine mutants fail to activate SRF-dependent transcription. Substitutions, which cause small reductions on complex formation in the bandshift assay, do not exhibit an absolute correlation in their effects on SRF activation. The fact that their impact on SRF activation is more noticeable at lower plasmid inputs implies that the assay is easily saturated by high MAL expression levels.

The integrity of the B1 region is also required for transcription of chromosomal Rho-regulated SRF-dependent genes. This was demonstrated using an immunofluoresence approach to visualise SRF-dependent protein expression in intact cells. Although this approach involved MAL overexpression it succeeded in retaining pathway specificity. Thus wild-type MAL efficiently induced expression of the Rho-dependent αSM-actin, whereas the expression of MAPK-controlled immediate early genes such as c-*fos* and *egr-1* was unchanged.

Despite the successful use of chromatin immunoprecipitation to detect the recruitment of endogenous MAL to SRF target gene promoters (Miralles, F. *et al.*, 2003), attempts to employ this technique to demonstrate the significance of the MAL B1 region in SRF-

binding in vivo using transiently transfected epitope-tagged MAL derivatives were unsuccessful. Although different antibodies, epitope tags and transfection techniques were tested, the cause of this is unclear since other studies have reported the successful chromatin immunoprecipitation of transfected Flag-Myocardin with SRF using similar conditions (Cao, D. *et al.*, 2005). A further approach would be to evaluate cell-lines stably expressing tagged-MAL derivatives in order to ensure high MAL protein levels.

6.1.1.3 The Q box

Gel mobility-shift assays of MAL domain-deletion mutants show that the lack of the Q box decreases but does not abolish the MAL-SRF complex. This is corroborated by the reporter-gene activation data, where deletion of the Q-box has a minimal effect on transactivation potential in agreement with another report (Cen, B. *et al.*, 2003).

In their original Myocardin paper, Wang *et. al.* showed that both the B1 and Q regions of Myocardin were necessary for complex formation with SRF and also reported that Myocardin ΔB1 or ΔQ derivatives were unable to activate SRF-reporter genes (Wang, D. *et al.*, 2001). The results described here partially confirm this report, since neither ΔB1 nor ΔQ forms of Myocardin interact with SRF in complex formation assays. However my results show that the Myocardin ΔQ derivative can still activate an SRF-dependent reporter genes, independently of the Rho-actin pathway. The failure to visualise MCΔQ-SRF complexes despite the ability of these proteins to interact functionally could be a result of the Q box deletion rendering the Myocardin-SRF interaction too weak or unstable to be detectable by the sensitivity levels of the bandshift assay.

Further analysis of the MAL Q box identified a number of conserved hydrophobic residues within the Q box and N-terminally to it that contribute to complex formation. In support of the non-essential role of the Q box in SRF binding removal of these sidechains reduces complex formation but has minimal effects on transcriptional activation.

A previous study identified the corresponding Myocardin residues as absolutely required for SRF binding and proposed that the role of the Myocardin Q region is

equivalent to that of the TCF B box (Wang, Z. *et al.*, 2004). Several lines of evidence presented here contest this proposal. The Q box mutations shown to affect SRF binding in the context of the full-length MAL or Myocardin have no effect on the interaction of B1Q peptides with SRF. In addition the integrity of the Q box is not sufficient to mediate peptide binding to SRF when the B1 region harbours mutations that abolish the complex. Mutations in the SRF DNA-binding domain that affect the interaction with MAL do so irrespective of the presence of the Q-box, indicating that this region of MAL does not directly contact any surface on SRF. Moreover the Q box is not involved in contacting DNA within the ternary complex with SRF, since efficient interaction of MAL derivatives with SRF remains dependent on the DNA sequences flanking the SRE irrespective of the presence of the Q region. Despite the presence of hydrophobic residues in the regions roughly corresponding to the Q box in arthropods, this region is poorly conserved through evolution in contrast to B1 (Figure 6.1), and the Q residues important for complex formation cannot be satisfactorily aligned with those mediating the interaction of the TCF B box with SRF (Hassler, M. *et al.*, 2001; Ling, Y. *et al.*, 1997).

What could be the role of the Q box in the MRTF-SRF interaction if it does not involve contacts with SRF or DNA? One possibility is that the Q box is required to stabilise the interaction of the B1 box with SRF, perhaps by making intraprotein contacts with this region. Such stabilising interactions are seen with the MATα2 repressor, which contacts the MCM1 β-sheet via a β-hairpin structure (see Introduction; (Tan, S. *et al.*, 1998)).

The Q box also inhibits the nuclear localisation of MAL, since in the MAL(met) context the absence of this region results in nuclear accumulation under basal conditions (Miralles, F. *et al.*, 2003). Although there is no strict correlation between the Q-residues contributing to the SRF interaction and those affecting the subcellular localisation of the protein, one possibility is that the Q box interacts with the B1 region in uncomplexed MAL resulting in the occlusion of the nuclear import signal. In this model Q box mutations would disrupt this interaction and expose B1 residues critical for nuclear import resulting in MAL nuclear accumulation.

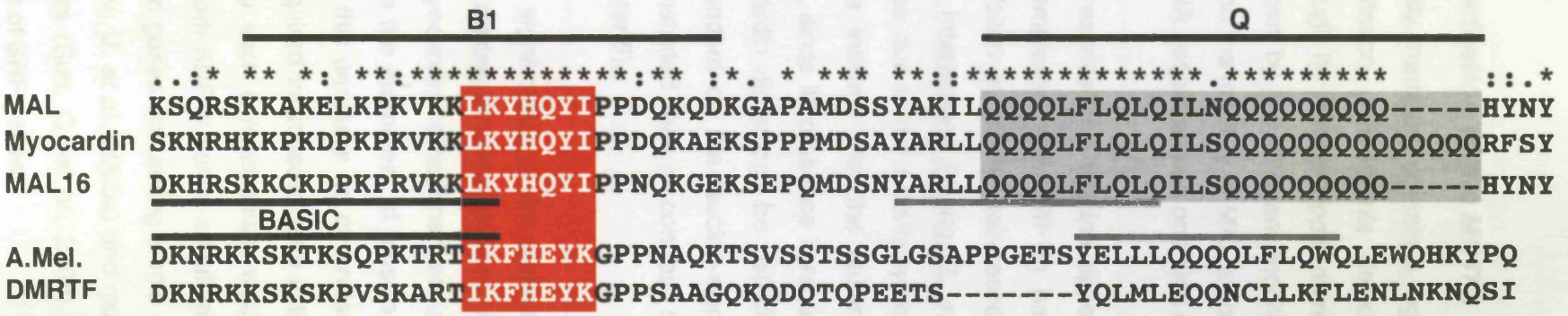


Figure 6.1 Sequence conservation in the B1 and Q regions of the MRTFs. Alignment of the B1 and Q regions of mouse MAL, Myocardin and MAL16. The B1 and Q boxes as defined by homology between the MAL and Myocardin regions are indicated by solid lines. Identical residues are indicated by asterisks and conservative replacements by colons and dots above the sequence. The critical seven residue B1 sequence that mediates SRF binding is boxed in red. The Q box is grey. The conservation of the critical seven-residue MAL sequence (red) and Q box-related sequence (indicated by grey lines) in bee (*Apis Melifera*) and fruitfly MRTFs are shown below. The basic region of the B1 box and the extended homology on the 5' side of B1 are indicated. Note that all B1 lysines are conserved in the B1 regions of the murine MRTFs.

6.1.1.4 Dimerisation of the MRTFs

All MRTF family members contain a region resembling a leucine-zipper domain. Co-immunoprecipitation experiments demonstrate that MAL has the ability to self-associate through its leucine-zipper domain. Moreover MAL contacts SRF as a stable dimer, as shown by the different mobilities of MAL LZ mutants in gel mobility-shift assays and also the inability of MAL to form monomeric SRF complexes on probes that lack critical DNA sequences on one side of the SRE (see later).

Despite the presence of a leucine-zipper-like region in Myocardin, this protein forms monomeric complexes with SRF in bandshift experiments. Co-immunoprecipitation experiments failed to show self-association of Myocardin, although others have reported such interactions (Wang, Z. *et al.*, 2003). The discrepancy between these results could be due to the sensitivity/stringency of the methods used. Reporter gene activation data indicate that the leucine-zipper of Myocardin has the potential to homodimerise, since its presence is required for full transactivation potential. Moreover a weak Myocardin dimer can be detected in the bandshift assays, the presence of which is dependent on the leucine-zipper motif. Experiments using truncated DNA probes demonstrated that in contrast to MAL, Myocardin is an unstable dimer in solution (Chapter 5).

The biological significance of the preferentially monomeric state of Myocardin despite its ability to dimerise remains unclear. Wang *et. al.* have proposed that the normally monomeric Myocardin homodimerises upon contacting SRF dimers on neighbouring CArG boxes in the promoters of muscle-specific genes (Wang, Z. *et al.*, 2003). They proposed that this unmasks an otherwise cryptic activation domain of the protein and this step is required for muscle-gene activation, but this has yet to be substantiated. It is unclear why such a mechanism would be in place since the MRTFs are able to interact with both SRF subunits simultaneously (Chapters 3 and 5). Moreover smooth muscle specific genes containing single CArG boxes are still responsive to Myocardin activation (Zhou, J. *et al.*, 2005a) and multiple CArG boxes are not restricted to muscle gene promoters (Sun, Q. *et al.*, 2006) and hence they do not define a Myocardin-specific subset of SRF genes.

Another possibility is that such a mechanism provides Myocardin with a wider range of gene targets, since in contrast to the stable MAL dimer the interaction of the monomeric Myocardin with SRF does not depend on direct DNA contacts on both sides of the CArG box (see Section 6.1.3). It is also conceivable that it reflects the interaction of Myocardin with a transcription factor other than SRF, which is not possible by a dimeric protein. MEF2 has recently been identified as a partner of the Myocardin cardiac specific monomer (Creemers, E. E. *et al.*, 2006). This protein does not provide an MRTF interaction surface analogous to that presented by SRF (see Section 6.1.2 and the Introduction), and it would be of interest to investigate the interaction of Myocardin dimers with MEF2 and the possibility of Myocardin multimerisation on MEF2 targets.

Leucine-zipper domains are described in terms of a heptad repeat in which residues are designated ***abcdefg***. In this repeating pattern residues ***a*** and ***d*** form the dimerisation interface and residues ***e*** and ***g*** are thought to contribute to the stability of the interaction by forming interprotein salt-bridges (Baxevanis, A. D. *et al.*, 1993). The charge compatibility of positions ***e*** and ***g*** in different leucine zippers defines the "i+5 rule" for the prediction of the specificity of leucine zipper partners (Vinson, C. R. *et al.*, 1993). According to this rule the MAL leucine-zipper is expected to homodimerise efficiently, in contrast to that of Myocardin. Furthermore the two proteins are predicted to heterodimerise. The results presented here broadly agree with the "i+5" predictions for MRTF dimerisation. Furthermore MAL16 has also been shown to homodimerise and heterodimerise with MAL, which also fits the "i+5" rule ((Selvaraj, A. *et al.*, 2003a); Cristina Perez-Sanchez, personal communication).

The dimerisation potentials of MAL and Myocardin cannot be interpreted only on the basis of the "i+5" predictions for leucine-zipper interaction as shown by the use of a chimeric MAL construct in which the authentic MAL LZ was swapped for that of Myocardin. The MAL(MC LZ) derivative formed both high- and low-mobility complexes with SRF, implying that rather than being a self-contained dimerisation unit, the leucine-zipper of Myocardin displays different abilities to self-associate depending on the protein context. So far, no other region in MAL has been shown to affect intraprotein interactions although the roles of conserved regions such as the SAP domain and Q box still remain obscure.

It should also be kept in mind that the potential interactions between residues ***g*** and ***e*** are not the only specificity determinants for leucine-zipper interaction. Position ***a*** of the leucine zipper has also been implicated in dimerisation-partner selectivity (Baxevanis, A. D. *et al.*, 1993; Lavigne, P. *et al.*, 1995). Polar or charged residues are often found at these positions and are thought to provide further stability and specificity determinants (Alber, T., 1992; Baxevanis, A. D. *et al.*, 1993; Lavigne, P. *et al.*, 1995). A closer inspection of the MAL, MAL16 and Myocardin leucine-zippers reveals that these positions are largely occupied by polar/charged residues. As yet, it remains unknown what, if any, significance these residues have in the ability of the MRTFs to partner with themselves and each other.

6.1.2 The MAL binding surface of SRF

6.1.2.1 The Interaction of cofactors with the hydrophobic groove of SRF

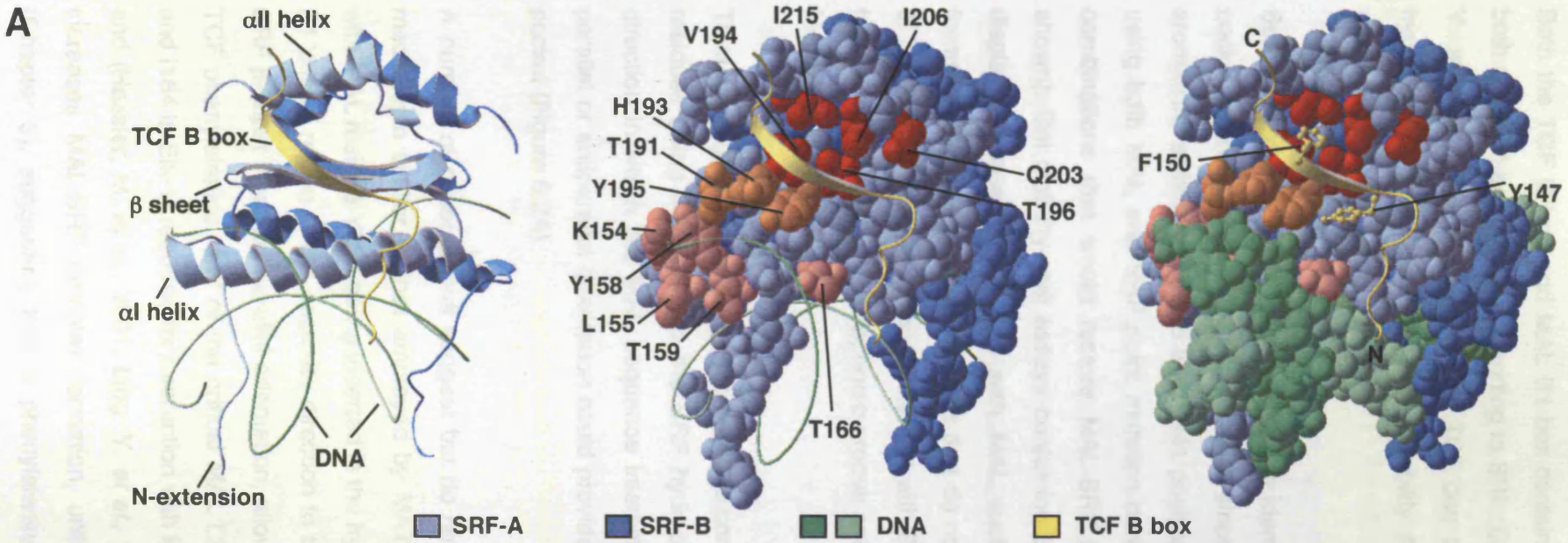
Functional studies have previously shown that physical interaction between the TCF B box and SRF inhibits the activity of the Rho-actin pathway (Murai, K. *et al.*, 2002). Thus the prediction was formulated that the Rho-regulated cofactor would compete with the TCFs for SRF complex formation. This was confirmed by peptide competition experiments, which demonstrated that the TCF B box and MAL contact the same surface on the SRF DNA-binding domain (Chapter 2).

The TCF-binding surface of SRF corresponds to a hydrophobic groove and pocket formed by the βII-strand and αII-helix of each SRF monomer ((Hassler, M. *et al.*, 2001; Ling, Y. *et al.*, 1998); see also Introduction and Chapter 4). The MAL-binding surface of SRF was also mapped along this groove using mutated SRF derivatives. Mutations H193A, V194E, Y195D and T196E disrupt MAL-SRF complex formation on DNA fragments derived from the c-*fos* promoter. In contrast in the Elk-1-SRF complex only the V194E and T196E changes detectably impair interaction, consistent with the work of Ling *et. al* (Ling, Y. *et al.*, 1998). In this study further changes that decrease Elk-1-SRF complex formation (E190A, Y195D, T199A, Q203E, and T207D), were identified by the use of a high affinity Ets DNA site to bind Elk1 and subsequently recruit SRF to a weak CArG box, which presumably renders the Elk-1 – SRF interaction more sensitive to individual amino acid substitutions.

TCF interacts with the hydrophobic groove of SRF by adding an antiparallel β-strand to the central β-sheet of the SRF DNA-binding domain (Hassler, M. *et al.*, 2001). Although TCF B box residues N- and C-terminal to the β-strand sequence are also required for SRF binding due to their interactions with the SRF αI-helix and DNA, and αII-helix respectively, it is noteworthy that MATα2 complex formation with MCM1 also depends on the addition of a β-strand to the MCM1 DNA-binding domain (Hassler, M. *et al.*, 2001; Ling, Y. *et al.*, 1997; Tan, S. *et al.*, 1998); see also Introduction). Moreover alanine substitution of any of the eight residues forming the β-strand sequence of MATα2 results in loss of interaction with MCM1 in functional assays (Mead, J. *et al.*, 1996).

Although the TCF and MATα2 β-strands are added to SRF and MCM1 respectively in opposite orientations, the interactions involved are highly conserved (Hassler, M. *et al.*, 2001; Tan, S. *et al.*, 1998). Central to the interaction of SRF with the TCFs is the insertion of an aromatic side-chain from the β-strand in a deep hydrophobic pocket on the SRF DNA-binding domain, defined by residues V194, I206 and I215 (Figure 6.2A; see also Chapter 4; (Hassler, M. *et al.*, 2001)). The aromatic character of the TCF residue inserted in the pocket is crucial for complex formation (Ling, Y. *et al.*, 1998). Such an interaction is also seen with MCM1 and MATα2, and other MCM1 interacting proteins have been implicated in interactions with residues surrounding the hydrophobic pocket (see Introduction), suggesting that this represents a common cofactor-binding mechanism for Type-I MADS box transcription factors.

MAL binding to SRF also involves interactions with the hydrophobic pocket as demonstrated by substitutions at positions I206 and I215 that are predicted to alter the pocket dimensions. These substitutions affect complex formation with both MAL and Elk-1, however their effects on the MAL-SRF and TCF-SRF interaction are not identical, indicating that the interactions of each cofactor with the SRF pocket are subtly different. These observations raise the possibility that MAL also binds the SRF DNA-binding domain by adding a β-strand to the central β-sheet of the structure. If β-strand addition is indeed the mechanism of MAL binding to SRF, is this strand added to the SRF β-sheet in an antiparallel orientation as seen with TCF or in parallel as seen with MATα2 and MCM1?



B

MAL	N→C	236LKYHQYIP243	MAL	N→C	236LKYHQYIP243
SAP-1	C←N	152LTFSSYLG145	SAP-1	N→C	145GLYSSFTL152
SAP-2	C←N	149ITFSSYLG142	SAP-2	N→C	142GLYSSFTI149
Elk-1	C←N	164ITFTSYLG157	Elk-1	N→C	157GLYSTFTI164
MATα2	N→C	114LVFNVVTQ121	MATα2	C←N	121QTVVNFVL114

Figure 6.2 The MAL binding surface of SRF and the "β-strand addition" cofactor binding model. (A) The SAP-1-SRF ternary complex (Hassler and Richmond, 2001) is shown: blue, SRF; green, DNA; yellow, SAP-1 B-box. Left, ribbon model, with secondary structure elements indicated. Centre, SRF and SAP-1 are shown in van der Waals and ribbon representations respectively. SRF residues implicated in MAL interaction are red (hydrophobic pocket), orange (βII-strand residues also contacting DNA), and pink (critical αI-helix residues). Right, as centre but with DNA shown as van der Waals representation and SAP-1 aromatic side chains Y147 and F150 in backbone representation. Images were produced from the 1hbx PDB file. (B) Potential sequence relationships between the critical MAL B1 region sequence and the β-strand segments of the TCF B-box, and MATα2. The aromatic residues are highlighted in yellow. Note that the TCF β-strand is added to the SRF β-sheet in antiparallel orientation, whereas that of MATα2 is added to the MCM1 β-sheet in paralell. Both cofactors insert a phenylalanine in the hydrophobic pocket of their MADS box partner protein.

Both the TCF B box and MAL B1 box contain two equally spaced aromatic residues, both of which are critical for binding to SRF (Chapter 3; (Hassler, M. *et al.*, 2001; Ling, Y. *et al.*, 1998)). In the case of TCF one of these aromatics is inserted into the hydrophobic pocket, raising the possibility that one of the MAL Y238 and Y241 residues fulfils the same role.

Substitutions at these positions aimed to identify the MAL residue responsible for the pocket interaction were not successful since SRF complex formation depends on aromaticity at both the Y238 and Y241 positions (Chapter 3). Additionally experiments using both MAL and SRF point mutation derivatives in an attempt to identify mutant combinations that would restore MAL-SRF binding were not successful (data not shown). Gel mobility-shift assays containing SRF hydrophobic pocket mutants, which display increased interaction with MAL such as I206A, failed to restore complex formation with the MAL B1 mutants that do not interact with SRF. Similarly bandshifts of MAL mutants, which strongly interact with SRF such as Y238F, did not compensate for the inability of SRF hydrophobic groove and pocket mutants to form a complex with MAL.

Thus although the data presented here are consistent with one of the MAL B1 aromatic residues being inserted into the SRF hydrophobic pocket, they do not address the direction in which the MAL sequence interacts with the SRF β-sheet, since either a parallel or antiparallel orientation could provide an aromatic residue for insertion in the pocket (Figure 6.2A).

A number of observations suggest but do not prove that a parallel β-strand addition mechanism similar to that employed by MATα2 operates in the MAL-SRF complex, with MAL residue Y238 being inserted in the hydrophobic pocket. Alignment of the MAL B1 critical region in the opposite direction to the TCF β-strand, which is added to the SRF β-sheet in an antiparallel orientation, allows alignment of both MAL Y238 with the TCF phenylalanine, and of the critical MAL L236 with a TCF residue (L152 in SAP-1 and I164 in Elk-1) required for interaction with the SRF αII-helix (Figure 6.2A; Chapter 3 and (Hassler, M. *et al.*, 2001; Ling, Y. *et al.*, 1997). Moreover the Y238F substitution increases MAL-SRF complex formation, unlike Y241F, which leaves it unaffected (Chapter 3), suggesting that a phenylalanine at position 238 facilitates MAL-SRF

interaction. This is intriguing in light of the fact that the residue inserted in the SRF hydrophobic pocket is a conserved phenylalanine in all three TCFs and a phenylalanine also interacts with the pocket in the MATα2-MCM1 complex (Hassler, M. *et al.*, 2001; Ling, Y. *et al.*, 1997; Tan, S. *et al.*, 1998). Furthermore the DMRTF residue corresponding to MAL Y238 is a phenylalanine (Figure 6.1).

The interaction of the MAL B1 region with SRF via a β-strand mechanism could also provide an explanation for the differential effect of the I206A substitution at the SRF hydrophobic pocket, on the MAL and TCF interaction. This change barely affects complex formation with TCF, presumably because the B box residue in the vicinity is a serine (S149 in SAP-1), whose small size does not interfere with the long isoleucine sidechain. In contrast the I206A change causes a striking increase of MAL binding, suggesting that the isoleucine sidechain clashes with a MAL residue. Assuming that one of the MAL tyrosines is inserted in the hydrophobic pocket, the residue closer to I206 would be H239 or Q240 depending on the orientation of β-strand addition. Both these amino acids have bulkier sidechains than serine, which could possibly interfere with I206. Such a model would suggest that removal of the interfering MAL sidechain should also increase SRF binding. This is not the case however since mutation H239A abolishes binding, and Q240A decreases it, indicating that these residues are also involved in important interactions with the SRF DNA-binding domain.

6.1.3 The role of DNA in the MAL-SRF interaction

6.1.3.1 The role of DNA bending

Formation of the MAL-SRF complex involves the N-terminal sequences of the SRF DNA-binding domain, since MAL does not interact with SRF.M2, an altered-specificity mutant that contains heterologous N-terminal sequences. This requirement seems to be independent of the DNA sequence contacted by SRF, since MAL and SRF.M2 fail to form a complex on both the wild-type SRE and the SRF.M2 specific SRE.M. These results confirm and extend previous findings that proposed the integrity of the N-terminal sequences of the SRF.DBD as a prerequisite for activation through the Rho pathway (Hill, C. S. *et al.*, 1994).

Mutagenesis analysis based on the SRF.M2 substitutions that do not affect DNA specificity identified a group of four residues within the SRF αI-helix whose mutation to the analogous residues from MCM1 abolished MAL binding. In contrast the αI-helix sequences do not affect the TCF-SRF complex, which is also weakly formed on SRF.M2.

The critical αI-residues are located close to DNA in the binary SRF-DNA complex and two of them are known to interact with DNA directly (Hassler, M. *et al.*, 2001; Pellegrini, L. *et al.*, 1995). Furthermore, the cumulative effect of all four substituted residues is required to inhibit MAL binding, since no individual substitution has similar effects on MAL-SRF complex formation. These observations suggest that the inhibition of MAL binding by the αI-helix derivative does not involve direct protein-protein contacts. Instead the effects of the SRF.M2 and αI-helix mutations on MAL binding correlate with their effects on SRF-induced DNA bending.

The ability to bend their DNA sequences is a well characterised property of Type-I SRF-like MADS box transcription factors (see Introduction). Residues within the αI-helix of SRF have been previously implicated in DNA bending in the context of the relaxed-specificity METcore^{SRF} construct and residues within the analogous region of MCM1 have been shown to mediate DNA bending (Acton, T. B. *et al.*, 1997; Lim, F. L. *et al.*, 2003; West, A. G. *et al.*, 1997). The results presented here confirm the involvement of the SRF αI-helix in DNA bending and implicate this property of SRF in cofactor selectivity, since MAL but not TCF is sensitive to mutations that decrease the degree of DNA bending.

The identification of DNA bending as a possible determinant of cofactor interaction led me to test other residues located in the αI-helix and β-loop of the SRF DNA binding domain previously reported to affect DNA bending through phosphate mediated effects on DNA binding (West, A. G. *et al.*, 1999; West, A. G. *et al.*, 1997). The results presented here show no easily interpreted correlation between the effects of these residues on MAL binding and DNA bending. The alanine substitution of residue K154, previously reported to be a major bending determinant in the context of METcore^{SRF} has modest effects on DNA bending, but reduces SRF complex formation with both MAL and Elk-1. It is possible that the effect on Elk-1 complexes reflects altered interactions between SRF and the TCF ETS binding domain arising due to the spacing

of the CArG and Ets sites in the c-*fos* promoter probe used (Chapter 4; see also (Hassler, M. *et al.*, 2001)).

Substitutions of the β-loop residues T191 and H193 had striking results on MAL binding, which were accompanied by small effects on DNA bending. Although this is consistent with an effect on complex formation mediated by DNA bending, it should be kept in mind that these residues are located close to the hydrophobic groove, which is the main MAL binding surface of SRF. It is therefore not possible to determine whether the impairment of MAL binding is due to loss of DNA contacts and distortion, or due to disruption of protein-protein contacts with MAL. The significant effect of the double T191/H193A mutation on SRF-Elk-1 complex formation could also be due to the proximity of these two residues to the TCF binding surface. The reservations on the connection between the effects of the β-loop residues on MAL binding and DNA bending could in principle be dispelled by testing the ability of the mutant derivatives to bind MAL in the absence of DNA. However as will be discussed later such an analysis was precluded by the required presence of DNA for efficient MAL-SRF interaction.

6.1.3.1.1 DNA bending in the MAL-SRF complex

In contrast to its effect on DNA bending in the binary SRF-DNA complex, the αI-helix mutation does not alter the magnitude of DNA bending in the context of the MAL-SRF-DNA ternary complex. This suggests that formation of the MAL-SRF complex depends on the induction of an appropriate DNA bend. According to this view, SRF mutations compromised for DNA bending are accompanied by the impairment of MAL-SRF complex formation since they require MAL to expend binding energy to further distort the DNA in the ternary complex.

Why would DNA bending be necessary for MAL binding to SRF? A simple model is that DNA bending or distortion facilitates direct contacts between MAL and DNA in the complex (see below) or that it induces an SRF conformation required for MAL binding. It is noteworthy that appropriate DNA bending has been proposed to be required for correctly juxtaposing MCM1 and some of its cofactors, such as Fkh2 and MATα1, to achieve optimal complex formation (Lim, F. L. *et al.*, 2003; Mead, J. *et al.*, 2002). Another possibility that is not incompatible with the previous explanations is that DNA bending also provides a specificity determinant for MAL target gene selection, based

on the inherent ability of a DNA sequence to bend. This could be a feature of the CArG box itself, since the central A-T rich sequence has the intrinsic tendency to bend, or sequences adjacent to it, as seen with MCM1 and the SQUA protein of *Antirrhinum majus* (Acton, T. B. *et al.*, 1997; West, A. G. *et al.*, 1999). Binding site selection experiments designed to investigate this possibility were not successful due to technical issues (see Appendix for a detailed analysis).

It is important to note that although useful in identifying overall effects in DNA distortion within a nucleoprotein complex, circular permutation experiments do not provide information on whether binding of a protein induces a DNA bend or simply an increase in local DNA flexibility (Kerppola, T. K. *et al.*, 1991). Furthermore, such experiments cannot determine the direction of the DNA bend. In the case of the SRF-DNA binary complex, extensive biochemical and crystallographic analyses have determined that SRF induces directional bending upon DNA binding (Hassler, M. *et al.*, 2001; Pellegrini, L. *et al.*, 1995; Sharrocks, A. D. *et al.*, 1995). The results presented here however do not address the effect on this bend of MAL binding to the SRF-DNA complex. The present analysis involves the qualitative comparison of the effects of the αI-helix mutation on DNA distortion in the MAL-SRF-DNA complex, and thus the directionality of the DNA bend within the MAL-SRF-DNA complex was not a critical point.

Structural analysis of the SRF-MAL complex could resolve this issue, but an additional way to tackle this would be to perform phasing analysis experiments (Crothers, D. M. *et al.*, 1991; Kerppola, T. K. *et al.*, 1991). This technique involves placing the binding site of the protein of interest at different locations in respect with an intrinsic DNA bend induced by a poly A:T-tract, such that the spacing between the two bending loci is phased over a helical DNA turn. The poly A:T tract will bend DNA towards the minor groove, and if the binding of the protein of interest induces an in-phase bend to the same direction the bends will cooperate to induce a maximum DNA bend slowing the migration of the nucleoprotein complex through a native gel. If the protein induced bend is in the opposite orientation the two bends will counteract each other resulting in a fast moving complex. Calculation of the relative mobilities of the complexed versus free DNA as a function of the distance between the intrinsic and protein induced bend centres allows the determination of the orientation of the protein-induced bend.

6.1.3.2 The Interaction of MAL with DNA in the ternary complex

The mutagenesis and DNA bending experiments discussed in the previous sections strongly suggest that DNA interactions are important in the MAL-SRF complex formation. Three lines of evidence show that this is indeed the case. First, MAL binding to SRF results in symmetric DNA contacts at positions ±16 through ±21 with respect to the SRE dyad, as demonstrated by DNase I footprinting. Second, these interactions are required for ternary complex formation with SRF, since truncated probes lacking the critical sequences flanking the SRE do not mediate MAL-SRF interaction. Third, DNA is required for efficient recovery of SRF with MAL in coimmunoprecipitation experiments. The data presented here do not formally exclude the possibility of the additional DNA contacts being mediated by SRF itself due to a conformational change induced by MAL binding. The solution of the MAL-SRF-DNA structure will unequivocally resolve this issue. Nevertheless the simplest interpretation of the data points to a model in which MAL binding to SRF is accompanied by symmetric MAL-DNA contacts around the SRE.

The dimeric nature of MAL is central in contacting DNA on either side of the SRE monomeric forms of MAL and Myocardin show a decreased dependence on the presence of the critical DNA regions. This result in combination with the presence of two MAL binding surfaces on the SRF dimer and the ability of MAL B1 region peptides to occupy both surfaces at the same time (Chapters 3 and 4), indicate that the formation of the MAL-SRF complex involves the interaction of the two B1 boxes present in the MAL dimer with the hydrophobic grooves of the two SRF subunits, with each MAL subunit also contacting DNA. This model also provides an explanation for the interference on SRF activation by the endogenous MAL protein under stimulated conditions observed with MAL mutants that do not themselves bind SRF (Chapter 3). According to the model for MAL-SRF binding delineated above heterodimerisation between the wild-type endogenous MAL and overexpressed mutated derivatives such as Y238A and Y241A would be expected to destabilise the MAL-SRF-DNA complex due to the inability of one B1 box in the MAL dimer to bind SRF, resulting in inefficient transcriptional activation.

6.1.3.2.1 MAL-DNA contacts through the B1 region

The nested probe analysis of MAL-SRF complex formation reveals that the B1 region is most likely responsible for the MAL-DNA contacts, since GST-fusion MAL proteins encompassing only the B1 and Q regions require the appropriate DNA fragment length for efficient SRF-complex formation, and this is not affected by deletion of the Q box. Although the possibility that the region linking the B1 and Q boxes is somehow involved has not been formally excluded, several lines of evidence support the involvement of the N-terminal part of the B1 sequence in formation of the MAL-DNA contacts. First, different deletions of the basic B1 sequences show differences in their abilities to interact with the short nested probes compared to the wild-type MAL protein. Second, N-terminal truncations of the peptides encompassing the B1 box exhibit a drop in SRF binding affinity in competition assays. Third, B1 peptides lacking the extreme N-terminal B1 residues give rise to an altered DNase I footprint, directly implicating the missing sequences in affecting MAL-DNA contacts.

Which B1 residue or residues could be responsible for DNA binding? Peptides lacking residues 224-228 from the N-terminal part of the B1 region give rise to an altered footprint and removal of sequences 224-229 in the MALΔN context increases the ability of MAL to interact with the shorter nested probes, implying that these sequences affect MAL-DNA contacts (Chapter 5). One explanation for the effect of these sequences on MAL-SRF complex formation is that their involvement in the MAL-DNA contacts is indirect, and that deletion of these sequences perturbs the authentic contacts between neighbouring B1 residues and the DNA. On the other hand it is also possible that residues 224-229 are themselves involved in direct DNA contacts and their deletion from the B1 peptides results in the loss of these contacts, but also induces the formation of new or perturbation of pre-existing contacts by the remaining B1 residues. The results presented here do not exclude either scenario.

Deletion of residues 230-235 causes a decrease in the overall efficiency of MAL-SRF binding, suggesting that it reflects the loss of MAL-DNA interactions (Chapter 5). However despite removal of residues K230 and K232 from the MAL B1 peptides resulting in marked drops in SRF binding affinity, these peptides produce footprints apparently identical to the longer peptide **A**, indicating that these lysines do not directly interact with DNA (Chapters 3 and 5). This does not exclude the possibility of other residues such as K234 and K235 being responsible for DNA contacts. For example

substitution of K234 with alanine in the context of MALΔN decreases complex formation with SRF (Chapter 2), raising the possibility that this lysine interacts with DNA. Thus the results presented here cannot conclusively identify the residues involved, since amino acids within either the 224-229 or 230-235 B1 sequences could be involved in DNA binding.

Removal of the complete N-terminal half of the B1 region relieves MAL of its dependence on the DNA sequences flanking the SRE and the Δ224-235 derivative is able to bind all nested probes efficiently (Chapter 5). This result is surprising since the opposite effect would be expected if residues within the deleted region were required for direct DNA contacts as indicated by the results described above. One possible explanation is that the 224-235 deletion brings a residue close to the SRF-binding region that is able to make a DNA contact closer to the core SRF-CArG box complex, thus sustaining binding. Interestingly the 224-235 deletion results in an arginine residue being brought to the position occupied by K234 in the wild-type protein.

A second possibility, is that the N-terminal part of the B1 box occludes the seven-residue sequence mediating binding to SRF, and the MAL-DNA contacts act to accommodate the basic region and unmask the SRF-binding stretch. Deletion of residues 224-235 relieves this masking effect resulting in the increased ability of MAL to bind SRF. Although this scenario cannot be ruled out, it appears unlikely since the smaller B1 deletions would be expected to also disrupt such an autoinhibitory mechanism. As already discussed however these small deletions do not allow binding of MAL to the shorter nested probes.

A third and more attractive explanation is that the multiple lysines in the B1 region compete with each other for DNA binding. According to this model deletion of the whole region relieves this competition and increases the ability of MAL to interact with SRF. This suggestion is also supported by the effects of alanine point mutations of the B1 box lysines, where substitutions K227A, K230A, K232A and K235A increase MAL-SRF complex formation in vitro, implying that these lysine side-chains somehow obstruct MAL-SRF binding (Chapter 3).

The data presented here do not address the possibility of sequence specificity in the MAL-DNA contacts. The fact that these contacts are symmetric around the SRE dyad

in the DNase I footprint, in combination with the possible involvement of the B1 region lysines suggests that these contacts represent non-specific interactions with the DNA phosphate backbone. However base-specific contacts within these regions cannot be excluded. Attempts to investigate this possibility with DNA binding site selection experiments using MAL and SRF were not brought to completion due to time limitations. Nevertheless preliminary results indicate that although MAL binding may bias the sequence identity of the bases adjacent to the CArG box, there appears to be no specific sequence patterns enriched in the ±16 to ±22 regions contacted by MAL in the DNase I footprinting experiments (see Appendix).

The data discussed in this section and the presence of multiple lysines in the N-terminal half of the B1 box that are conserved between MRTF family members (Figure 6.1) make it tempting to speculate that more than one of these residues are responsible for interacting with DNA, possibly by making phosphate contacts with the DNA backbone. Further experiments are required to investigate this and to determine the MAL-DNA binding mechanism. The use of MAL derivatives or long B1Q peptides harbouring single or multiple lysine to alanine substitutions will potentially clarify the puzzling and sometimes contradicting effects of the B1 region on MAL-SRF complex formation. MAL GST-fusion derivatives containing such changes were constructed to address these issues, however preliminary DNase I footprinting and nested probe bandshift experiments were inconclusive and could not be completed due to time constraints.

6.1.4 Conclusions: A model for the interaction of MAL and SRF

My analysis of the mechanism mediating complex formation between MAL and SRF demonstrates that members of the MRTF and TCF families of SRF cofactors interact with SRF using related but distinct mechanisms. MAL contacts the same hydrophobic groove on the SRF DNA-binding domain as members of the TCF family of SRF cofactors. Interactions with a hydrophobic pocket on this surface of SRF are central for both MAL and TCF binding. Formation of the MAL-SRF complex is mediated by a short predominantly hydrophobic core sequence within the conserved MAL B1 region. Although this sequence is sufficient for specific binding, basic residues N-terminal to it are thought to contribute to the affinity of the interaction. Hydrophobic residues in the neighbouring Q-box also enhance MAL-SRF complex formation but these are unlikely

to contact SRF directly. MAL contacts SRF as a dimer and in contrast to the TCFs, mutations in the SRF αl-helix that reduce SRF-induced DNA bending also impair complex formation with MAL. These mutations however do not affect DNA distortion in the MAL-SRF complex. Efficient MAL-SRF binding requires that SRF be bound to DNA and that MAL directly contacts DNA on either side of the CArG box.

My results support a model in which each MAL monomer adds a β-strand consisting of the core hydrophobic B1 sequence, to the β-sheet region in the middle layer of the SRF DNA-binding domain, while simultaneously making direct DNA contacts with the DNA flanking the CArG box. These DNA contacts are probably facilitated by the appropriate DNA distortion in the SRF-DNA complex and mediated by basic residues in the B1 region.

6.1.4.1 Implications of cofactor competition

"β-strand addition" is a common interaction mechanism utilised by MADS box factors, and is employed both for cofactor binding as exemplified by SRF and MCM1 (Hassler, M. *et al.*, 2001; Tan, S. *et al.*, 1998), and also for dimerisation interactions as seen in the extended contacts between the MADS and MEF2 domains of MEF2 factors (Han, A. *et al.*, 2003). SRF utilises the overlapping surfaces to interact with the MRTFs and TCFs, and the same is observed in the interactions of MCM1 with its cofactors (see Introduction for analysis). Although the β-strand addition mechanism is not available for cofactor binding in MEF2 factors, these proteins also use overlapping surfaces in their MEF2 domains to interact with the Cabin repressor and HDACs (Han, A. *et al.*, 2003). The ability to recruit many different interacting proteins via the same binding mechanism expands the regulatory potential of MADS box transcription factors and the mutually exclusive interactions with their cofactors ensure specificity of their responses.

One such example is the competition between MRTFs and TCFs for SRF complex formation presented in this thesis. MRTF-TCF competition has been shown to be relevant in vivo, since stimulation of the MAPK pathway through PDGF in smooth muscle cells results in Elk-1 activation and displacement of Myocardin from target gene promoters (Wang, Z. *et al.*, 2004). Thus cofactor competition and exchange creates a binary switch that controls SRF responses in muscle differentiation and cell proliferation.

Although the influence of SRF interaction with its cofactors by signalling inputs to the cofactors themselves is well established (Hill, C. S. *et al.*, 1993; Miralles, F. *et al.*, 2003; Murai, K. *et al.*, 2002; Wang, Z. *et al.*, 2004), it remains unclear whether signalling to SRF itself affects cofactor selectivity. Phosphorylation of S162 in the αI-helix of the SRF DNA-binding domain has been recently implicated in selective inhibition of complex formation with the MRTFs (Iyer, D. *et al.*, 2006). S162 phosphorylation by PKCα blocks efficient SRF-DNA binding resulting in inhibition of Myocardin-SRF complex formation. TCF-SRF complex formation remains unaffected since TCF binding to adjacent Ets sites facilitates recruitment of the phosphorylated SRF to the CArG box (Iyer, D. *et al.*, 2006). The modulation of cofactor binding by this phosphorylation event has been implicated in the control of the myogenic versus proliferative response of SRF, and such phenomena involving cofactor exchange via signalling to SRF will be interesting topics for future investigation.

6.1.4.2 Cofactor specificity of SRF target genes

An additional unresolved issue in regulation of transcription via the combinatorial interactions of SRF with its cofactors is the cofactor specificity of SRF target genes. In the case of the TCFs it is clear that the presence of an Ets DNA site confers MAPK-sensitivity on SRF-dependent genes (Gineitis, D. *et al.*, 2001; Murai, K. *et al.*, 2002; Zhou, J. *et al.*, 2005b). Nevertheless it is not possible to determine the TCF-responsiveness of SRF target genes based on the DNA sequence alone due to the flexibility observed in the spacing and orientation of the Ets site with respect to the CArG box (Treisman, R. *et al.*, 1992). The MRTFs also contact DNA in the ternary complex with SRF, but the sequence-dependence of these interactions and the influence of the DNA sequence of the CArG box and flanking regions on MRTF target gene selection remain unresolved.

6.1.4.2.1 Target gene specificity and functional diversity of the MRTFs

A related outstanding question is what determines target gene specificity between MRTF family members. The results presented in this thesis and the sequence conservation of the MRTFs in the B1 and Q regions (Figure 6.1) indicate that different family members employ the same mechanism to interact with SRF. Moreover the

MRTFs exhibit distinct but overlapping expression patterns in mammals and all are implicated in muscle-specific gene expression (reviewed in (Pipes, G. C. *et al.*, 2006); see Introduction for detailed analysis). Despite these observations and indications of functional redundancy, in vivo data suggest that different family members also have discrete effects ((Li, S. *et al.*, 2006; Li, S. *et al.*, 2003; Oh, J. *et al.*, 2005; Sun, Y. *et al.*, 2006); for analysis of the in vivo effects of different MRTFs see Introduction).

Different expression levels in diverse cell types and developmental stages are likely to affect recruitment of MRTFs to target genes by competition for SRF binding. It is also possible that the dimerisation state of the MRTFs plays a role in target gene selectivity. The requirement of dimeric MRTFs to contact DNA flanking both sides of the CArG box could limit accessibility to certain SRF genes due to competition with factors binding to DNA elements close to the CArG box, as indicated by the inhibition of Rho-MAL signalling by the Ets and AP1/ATF sites flanking the CArG box on the c-fos promoter (Murai, K. *et al.*, 2002). Fewer restrictions of this type would be expected for preferentially monomeric MRTFs like Myocardin, since these only require SRF and DNA contacts on one side of the SRF-DNA complex.

Another as yet unexplored possibility is that additional specificity determinants are imposed on MRTF family members by their interactions with proteins other than SRF. Synergistic effects of SRF with myogenic factors such as GATA and MyoD have been widely reported (Belaguli, N. S. *et al.*, 2000; Groisman, R. *et al.*, 1996; Morin, S. *et al.*, 2001; Nishida, W. *et al.*, 2002). Recent studies suggest that at least some of the GATA-mediated effects involve Myocardin (Yin, F. *et al.*, 2005), although the promoter context is highly significant for the transcriptional outcome of the Myocardin-GATA interactions (Oh, J. *et al.*, 2004; Yin, F. *et al.*, 2005). The extent to which these effects are restricted to Myocardin or involve other MRTF family members remains unknown.

The differential functional roles of the MRTFs could also be due to SRF-independent transcription factor interactions, as seen with the cardiac isoform of Myocardin, which interacts with MEF2 and activates MEF2-dependent gene targets (Creemers, E. E. *et al.*, 2006). Thus at least one extra level of specificity exists in the case of Myocardin, and it remains possible that the other MRTFs also interact with as yet unidentified partners.

6.1.4.3 Future prospects

Several outstanding questions remain on how the formation and modulation of the interactions of SRF with its cofactors determines the specificity of its transcriptional responses depending on signalling inputs and cell type. The structural analysis of the MRTF-SRF complex, the investigation of the DNA sequence dependence of complex formation and the effects of cofactor specific SRF derivatives in vivo will provide important insights into the mechanisms via which SRF fulfils its diverse biological functions via combinatorial interactions with cofactors.

7 Materials and Methods

MATERIALS

7.1 Chemicals and reagents

This is a general list of reagents. Chemicals were purchased from Sigma, Merck and Roche unless otherwise stated. Those reagents used specifically for one method are described in the relevant section.

Acrylamide/Bisacrylamide 37.5:1 and 19:1 solution	AMRESCO
Agarose	Gibco BRL
Ammonium persulphate	Sigma
Ampicillin	Sigma
Aprotinin	Sigma
Benzamidine	Sigma
Bromophenol Blue	Biorad
BSA (acetylated)	Sigma
Chloramphenicol	Boehringer Mannheim
Complete protease inhibitor cocktail tablets	Roche
Coomassie Brilliant Blue	Biorad
Cytochalasin D	Calbiochem
Dithiothreitol (DTT)	Calbiochem
Dimethyldichlorosilane solution	BDH
Ethidium bromide	Boehringer Mannheim
Glycogen	Boehringer Mannheim
Kanamycin	Sigma
Leupeptin	Sigma
Linear Acrylamide	Ambion
β-mercaptoethanol	Sigma
milk powder	Marvel

dNTPs	Pharmacia
Okadaic acid	Calbiochem
Orange G	Sigma
Pepstatin	Sigma
Phenylmethyl-sulfonyl fluoride (PMSF)	Sigma
Poly(dIdC)•poly(dIdC)	Amersham
Protein assay reagent	BioRad
Revidue 32P dNTPs	Amersham
Spermidine	Sigma
TEMED	Sigma
Trizma-base	Sigma
Triton X-100	Sigma
Xylene cyanol	Biorad
3MM paper	Whatman

Restriction enzymes were purchased from New England Biolabs (NEB) and were used with the recommended supplied NEB buffers and BSA solution. Additional enzymes used were purchased from the following companies:

Biotaq Red DNA polymerase	Bioline
Calf intestinal alkaline phosphatase	NEB
DNase I	Sigma
Klenow DNA polymerase	NEB
Pfu turbo DNA polymerase	Stratagene
Proofstart DNA polymerase	Qiagen
Proteinase K	Gibco BRL
RNase A	Sigma
RNase inhibitor	Boehringer Mannheim
T4 Polynucleotide kinase	NEB

Where applicable, enzymes were used with the buffers supplied by the manufacturer.

7.2 Buffers and solutions

All buffers and solutions were made with deionised water (Milli-Q plus system, Millipore) and, where appropriate sterilised by filtration on a 0.2 μm vaccum-driven filtration system (Stericup). A list of the most commonly used solutions follows. Solutions used specifically for one technique are listed in the relevant section.

PBS	0.17 mM NaCl 3 mM KCl 1 mM Na_2HPO_4 1.8 mM KH_2PO_4 pH 7.4
TBE	89 mM Tris Base 89 mM Boric acid 2 mM EDTA
TE	10 mM Tris pH 7.5, 1 mM EDTA pH 7.5
TEN	10 mM Tris pH 7.5, 1 mM EDTA pH 7.5 100 mM NaCl

7.3 Plasmids and oligonucleotides

7.3.1 Expression vectors

Protein expression in mammalian cells

MAL, SRF and Elk-1 cDNAs were expressed from the following vectors:

MLV.plink	described in (Dalton, S. *et al.*, 1992).
pEF.Flag	derived from EF.plink (Hill, C. S. *et al.*, 1995b), contains an N-terminal Flag epitope tag

pEF.HA	derived from EF.plink (Hill, C. S. *et al.*, 1995b), contains two N-terminal HA epitope tags

Protein expression in reticulocyte lysate in vitro translation systems
SRF cDNAs were expressed from the following vector:
pFTX5 derived from T7Plink (Howell, M. *et al.*, 1997) has an N-terminal Myc epitope tag

Protein expression in E. coli
MAL cDNAs were expressed as GST-fusion proteins from the following vector:
pET41a Novagen

7.3.1.1 Oligonucleotides

Oligonucleotide primers were synthesised by the in-house Cancer Research UK oligonucleotide synthesis service or from Sigma. All oligonucleotides were purified by reverse phase chromatography (RP1) unless otherwise stated. Lyophilised oligonucleotides were dissolved in water to final concentrations of 10 or 100 μM. Details of oligonucleotide design for MAL and SRF derivative construction can be found in section 7.7.4.4. Oligonucleotides used to generate bandshift probes are listed in section 7.9.1.7.

7.4 Peptides

All peptides were synthesised by the Cancer Research UK Peptide Synthesis Laboratory. Peptides were dissolved in water, or when they were too hydrophobic in 50 % acetonitrile/water solution. After centrifuging for 10 minutes at maximum speed and room temperature to remove undissolved material a 1/10 dilution of peptide solution was used to measure the absorbance of the peptide bond at 215 nm in a fluorimeter using Quartz cells. The absorbance measurement was used to calculate the peptide concentration as follows:

(Absorbance x peptide dilution) ÷ (number of peptide bonds) = A*

where A* =Absorbance units per peptide bond

The peptide concentration was then worked out from the Beer-Lambert law:

$$A^* = \varepsilon c l$$

Where ε = molar extinction coefficient (1000 l/M/cm for each peptide bond)

And l = cell path length, (3mm cells were used)

The pH of the peptide solutions was checked with pH indicator test strips and pH was brought to 7-8 with Tris-HCl pH 7.9.

The peptides were then buffered to a final concentration of 10 mM Tris-HCl pH 7.9, 100 mM NaCl and 25% acetonitrile. Peptide stock solutions were stored at -20°C and -80°C.

Peptide Name	**Peptide Sequence**	**Residues**
Elk-1 B box WT	SSRNEYMRSGLYSTFTIQSLQ	Elk-1 148 - 168
Elk-1 B box Y159A	SSRNEYMRSGLASTFTIQSLQ	Elk-1 148 - 168
MAL B1 A	LKPKVKKLKYHQYIPPDQKQD	MAL 229 - 249
MAL B1 B	LKPKVKKLKYHQYIPPDQKQ	MAL 229 - 248
MAL B1 C	LKPKVKKLKYHQYIPPDQK	MAL 229 - 247
MAL B1 D	LKPKVKKLKYHQYIPPDQ	MAL 229 - 246
MAL B1 E	LKPKVKKLKYHQYIPPD	MAL 229 - 245
MAL B1 F	LKPKVKKLKYHQYIPP	MAL 229 - 244
MAL B1 G	LKPKVKKLKYHQYIP	MAL 229 - 243
MAL B1 H	KKLKYHQYIP	MAL 234 - 243
MAL B1 J	KKLKYHQYIPPDQKQD	MAL 234 - 249
MAL B1 K	VKKLKYHQYIPPDQKQD	MAL 233 - 249
MAL B1 M	KVKKLKYHQYIPPDQKQD	MAL 232 - 249
MAL B1 N	PKVKKLKYHQYIPPDQKQD	MAL 231 - 249
MAL B1 P	KPKVKKLKYHQYIPPDQKQD	MAL 230 - 249
MAL B1 Q	LKPKVKKLKAHQYIPPDQKQD	MAL 229 - 249
MAL B1 R	LKPKVKKLKYHQAIPPDQKQD	MAL 229 - 249
MAL B1 S	KKLKAHQYIP	MAL 234 - 243
MAL B1Q	KKAKELKPKVKKLKYHQYIPPDQKQDKGA PATDSSYAKILQQQQLFLQLQILNQQQQQ QQ	MAL 224 - 283
MAL B1Q Y238A	KKAKELKPKVKKLKAHQYIPPDQKQDKGA	MAL 224 - 283

Peptide Name	Peptide Sequence	Residues
	PATDSSYAKILQQQQLFLQLQILNQQQQQ QQ	
MAL B1Q 3xA	KKAKELKPKVKKLKYHQYIPPDQKQDKGA PATDSSYAKIAQQQQLFAQLQALNQQQQ QQQ	MAL 224 - 283
MC B1Q	KKPKDPKPKVKKLKYHQYIPPDQKAEKSP PPMDSAYARLLQQQQLFLQLQILSQQQQ QQQ	Myocardin 247 - 306
MC B1Q Y261A	KKPKDPKPKVKKLKAHQYIPPDQKAEKSP PPMDSAYARLLQQQQLFLQLQILSQQQQ QQQ	Myocardin 247 - 306
MC B1Q 3xA	KKPKDPKPKVKKLKYHQYIPPDQKAEKSP PPMDSAYARLAQQQQLFAQLQALSQQQQ QQQ	Myocardin 247 - 306

METHODS

All water used was deionised on a Milli-Q plus system (Millipore) and, where appropriate, solutions and culture media were sterilised by filtration on a 0.2 μm vaccum-driven filtration system (Stericup).

7.5 Bacterial Techniques

7.5.1 Bacterial media and plates

LB media	1% w/v Bacto-tryptone, 0.5% w/v Bacto-yeast extract, 1% w/v NaCl
LB agar	1% w/v Bacto-tryptone, 0.5% w/v Bacto-yeast extract, 1% w/v NaCl, 1.5% w/v Bacto-agar

Depending on plasmid antibiotic resistance 100 μg/ml Ampicillin, 30 μg/ml Kanamycin, or 34 μg/ml Chloramphenicol were added to the media and agar plates for plasmid selection.

7.5.2 Bacterial strains

DH5α	**Invitrogen; used for all cloning manipulations**
TOP10	**Invitrogen; used for all cloning manipulations.**
JM110	**Stratagene; used for production of unmethylated DNA**
Rosetta DE3 pLysS	**Novagen; used for expression of GST-fusion proteins**

7.5.3 Transformation of *E. coli*

7.5.3.1 Preparation of electrocompetent E. coli

A single colony of electrocompetent E. coli previously grown on an agar plate, was inoculated in a sterile flask containing 10 ml of LB media and grown overnight at 37°C at 200-220 rpm. The saturated culture was divided between two 1l flasks containing 500 ml of LB medium. Cells were incubated at 37°C with shaking until they reached $OD_{600} = 0.6$.

The cultures were then chilled on ice for 20 min and the cells were collected by centrifugation at 1200xg (4000 rpm) for 10 min at 4°C. The pellet was resuspended in ice-cold 10% glycerol made in highly purified water and filtered (resuspension volume was equal to the original culture volume) and incubated on ice for 20 min. Cells were then pelleted by centrifugation at 4000 rpm at 4°C for 10 min. The wash and centrifugation steps were repeated twice under the same conditions, reducing the pellet resuspension volume first to half and then to one quarter of the original culture volume. After the third wash and centrifugation step, as much supernatant as possible was removed and cells were gently resuspended in 1/500 volume (routinely 2ml per 1l of original bacterial culture) of ice-cold 10% glycerol. This bacterial suspension was

divided into 40-200 µl aliquots in prechilled eppendorf tubes and snap-frozen on dry ice for 15 min before storing at -80°C.

7.5.3.2 Electroporation of DNA into E. coli

40 µl of electrocompetent bacteria were thawed on ice and mixed with 1µl of purified DNA dissolved in water or TE, or 1-2.5 µl of ligation mix. The suspension was placed in an ice cold electroporation cuvette (BioRad 0.2 cm separation) and cells were subjected to an electric pulse of 2.5 kV (capacitance setting 25 µF and resistance setting 200 Ω) using a BioRad Gene pulser with a pulse controller. 1 ml of LB media was immediately added to the cell suspension which was incubated at 37°C for 1 hr. Cells were then plated on LB plates containing the appropriate concentration of antibiotic, depending on plasmid antibiotic resistance. For propagation of purified DNA 100 µl were plated. For ligation samples, all the electroporated cell culture was plated: cells were pelleted by flash-spin centrifugation to 9000 rpm for 5 s and after removing all but 100-200 µl of media, cells were resuspended and plated.

7.5.3.3 Chemical Transformation of E. coli

Chemically competent E. coli JM110, were transformed according to the manufacturer's instructions (Stratagene).

7.6 Mammalian Cell culture

7.6.1 Cell culture media

E4	equivalent to DMEM CRUK media production
Foetal Bovine Serum (FBS)	Gibco BRL
OPTIMEM-1	Reduced Serum Medium, Gibco BRL
5 x Trypsin/Versene	2.5 g trypsin, 8 g NaCl, 1.15g Na_2HPO_4,200 µg KH_2PO_4, 1 g versene (EDTA), 1.5 ml 1 % (w/v) phenol red CRUK media production

7.6.2 Cell lines

NIH3T3	A mouse fibroblast cell line (Treisman Laboratory, CRUK)
SRE.FOS.HA	An NIH3T3 derived cell line stably expressing the SRE.FOS.HA construct. Described in (Alberts, A. S. *et al.*, 1998).

7.6.3 General culture conditions

Cell lines were cultured on plastic dishes of tissue culture grade (Corning Incorporated and Beckton Dickinson) in an incubator (Innova Co-170, New Brunswick Scientific) at 37 °C and 10 % CO_2. Cells were cultured in E4 supplemented with 10 % foetal bovine serum. All media were warmed to 37 °C in a water bath before adding to the cells. To remove cells from the flask, the cells were washed once with trypsin:versene and then incubated with trypsin:versene solution for 1-2 minutes at 37°C.

7.6.4 Transient cell transfection with Lipofectamine reagent

Transient cell transfections using the Lipofectamine reagent were performed according to the manufacturer's recommendations (see Table). Transfections for luciferase assays and coverslip seeding for immunofluorescence assays were performed in 6-well plates, transfections for whole-cell extracts were in 60mm plates and transfections for co-immunoprecipitations in 10cm plates. Cells were seeded (see Table) 18-24 hr prior to transfection, in E4 medium supplemented with 10% FBS. Transfections were performed the following day as follows: the DNA and Lipofectamine Reagent (see Table) were mixed separately with Optimem-1 Reduced Serum Medium and left at room temperature for 10 minutes. After this the two mixes were combined, vortexed vigorously and allowed to incubate at room temperature for 30 minutes. Cells were washed twice in Optimem-1, and the appropriate volume of Optimem-1 was added to each dish (see Table). The transfection mix was added to the cells and they were incubated at 37°C and 10% CO_2 for 5 hr. After 4-5 hr the transfection solution was

replaced with starvation media (E4 + 0.3% FCS). Stimulations of cells with the appropriate stimuli were performed 18 - 24 hr later and prior to sample processing.

Plate size	Number of cells per plate or well	[DNA]	Volume of Lipofectamine	Volume of Optimem	Volume of Optimem on cells
6-well	1.5×10^5	1 μg	6 μl	200 μl	1500 μl
6 cm	3×10^5	2 μg	12 μl	400μl	1500 μl
10 cm	1×10^6	2-4 μg	20 μl	1000 μl	4000 μl

7.6.5 Luciferase assay

Firefly luciferase was measured using luciferase reporter assay system according manufacturer's recommendations (Promega). Transfections for luciferase assays were performed using Lipofectamine reagent as described in section 7.6.4. Routinely 40 ng 3D.A.luc or 50 ng of the SRE.L2 or SRE.LM2 reporter plasmids were transfected with MAL or Myocardin expression plasmids (standard amounts were 50 ng). 150 ng of the MLV.lacZ reporter gene was co-transfected as an internal control for luciferase activity normalisation. Each experiment included a sample transfected with 50 ng of the MLV.SRF.VP16, for normalisation of SRF activity. Where necessary, cells were stimulated for 7 hr with the appropriate stimuli. Cells were washed twice with ice-cold PBS, prior to scraping with a rubber policeman in 150 μl 1 x Reporter Lysis Buffer (Promega). After centrifugation at 13,000 rmp for 5 min to remove cell debris 20 μl of the supernatant was used to determine firefly luciferase activity. 45 μl Luciferase Assay Reagent (Promega) were added to the cell lysates in a Microtiter plate (Dynex Technologies, Inc) and the light produced was measured on a Microtiter Plate Luminometer (Dynex) using Revelation Version 3.2 software. The activity was normalised to βgal activity or to total protein measured with the Biorad protein assay reagent (see section 7.8.2.1).

7.6.5.1 β-GAL assay

The β-Galactosidase assays for luciferase activity normalisation were performed in a 96-well plate and contained 20 μl cell lysate,180 μl LacZ buffer (60 mM $Na_2HPO_4.7H_20$,

40 mM NaH_2PO_4, 10 mM KCl, 1 mM $MgSO_4$, 2.7 ml/l ß-mercaptoethanol) and 20 µl of 4 mg/ml O-Nitro-Phenyl-ß-D-galactoside (ONPG). Reactions were incubated at 37°C for 30 min, or until the colour was changed from light to dark yellow. βgal activity was quantitated spectrophotometrically at 595 nm using a SpectraMax Plus spectrophotometer (Molecular Devices).

LacZ buffer	60 mM $Na_2HPO_4.7H_20$ 40 mM NaH_2PO_4 10 mM KCl 1 mM $MgSO_4$ 2.7 ml/l β-mercaptoethanol

7.6.5.2 Mammalian reporter plasmids

3D.A.LUC	a derivative of 3D.ACAT (Mohun, T. *et al.*, 1987) with firefly luciferase in place of the CAT sequence. Constructed by O. Geneste (Geneste, O. *et al.*, 2002).
SREL2.LUC	a derivative of SRE.L2 (Hill, C. S. *et al.*, 1994), with firefly luciferase in place of the CAT sequence. Constructed by Gemma Smith.
SRELM2.LUC	a derivative of SRE.LM2 (Hill, C. S. *et al.*, 1994), with firefly luciferase in place of the CAT sequence. Constructed by inserting the SRE.LM2 sequence in the pGL3 (Clontech) luciferase reporter vector after digestion with Xho1.
MLVLacZ	described in (Marais, R. *et al.*, 1993).

7.6.6 Immunofluorescence assay

Cells were seeded in 6-well plates, each well containing 2-3 coverslips and transfected as described in section 7.6.4, with 50 ng of MAL expression plasmids. Approximately 24 hours after transfection cells were stimulated with the appropriate stimuli for 30 minutes prior to processing.

Cells were fixed in fresh 4% paraformaldehyde or formaldehyde in PBS, prewarmed to 37°C, for 15 min. After washing three times with PBS cells were permeabilised with 0.2% Triton X-100 in PBS for 10 minutes. Coverslips were blocked with Gelatin blocking solution for 30 min – 1 hr at room temperature and washed three times with 0.2% Triton X-100 in PBS. A piece of whatman paper was placed in a plastic tissue culture dish and after wetting it with water a piece of parafilm was placed on top. Primary antibodies were diluted in Gelatin blocking solution and approximately 20 μl drops were placed on the parafilm sheet. Each coverslip was then placed on the antibody solution and incubated for 1hr at room temperature or 30 min at 37°C. After incubation coverslips were washed three times with 0.2% Triton X-100 in PBS before being incubated with the secondary antibody (and other staining reagents such as phalloidin and DAPI) following the procedure described for the primary antibodies. Coverslips were washed three times with 0.2% Triton X-100 in PBS, followed by a final wash in water prior to mounting on glass slides. Slides were dried at 37°C for 1 hour or overnight at room temperature. Images were obtained using a Zeiss Axiovert microscope and Smart Capture system (Vysis UK) and processed as PICT files using Adobe photoshop CS2.

Endogenous gene expression after MAL overexpression was performed in essentially the same way, with the exception that 150 ng of MALΔN constructs were transfected, and cells were kept under starvation conditions (1 mg/ml BSA in E4 media) for 48 hr prior to processing.

Immunofluorescence solutions	
Gelatin blocking solution	1 % (v/v) Gelatin 0.2 % (v/v) Triton X-100 10 % FBS 10 % (w/v) Milk in PBS
mounting fluid	1 % (w/v) Mowiol 4-88 (Calbiochem) 2.5 % (w/v) 1,4-diazabicyclo(2.2.2) octane antifade (Sigma) 25 % Glycerol 100 mM Tris-HCl pH 8.5

Reagents used for immunofluorescence analysis:

Primary antibodies			
Antibody	**Specificity**	**Species**	**Concentration**
Flag (Sigma)	Flag	Rabbit	1/300
M2 Flag (Sigma)	Flag	Mouse	1/200
SM α-actin (Sigma)	SM α-actin	Mouse	1/200
Egr-1 (Santa Cruz)	Egr-1	Rabbit	1/50
Fos (Santa Cruz)	Fos	Rabbit	1/50

Secondary Antibodies			
Antigen	**Conjugate**	**Species**	**Concentration**
Rabbit IgG (Jackson Labs)	FITC	Goat	1/200
Mouse IgG (Jackson Labs)	Cy2	Donkey	1/200
Rabbit IgG (Jackson Labs)	Cy2	Donkey	1/200
Mouse IgG (Jackson Labs)	Cy3	Donkey	1/500
Rabbit IgG (Jackson Labs)	Cy3	Donkey	1/500

Cell staining reagents		
Reagent	**Cell component**	**Concentration**
Phalloidin toxin TRITC-conjugated (Molecular Probes)	F-actin	1/100
4',6-diamidino-2-phenylindole stain (DAPI)	DNA	1/10000

7.7 Nucleic Acid Manipulations

7.7.1 Purification of plasmid DNA

To prepare plasmid DNA, a single antibiotic-resistant colony was picked and used to inoculate either 4 ml (small-scale preparation, miniprep) or 200 ml (large-scale preparation, maxiprep) of LB containing the appropriate plasmid selection antibiotic. The culture was then shaken overnight at 37°C. For minipreps, the DNA was isolated

using QIAprep spin miniprep kit (Qiagen) according to the Manufacturers' instructions or by the CRUK Equipment Park Miniprep service using the Qiagen Biorobot 9600. For maxipreps, the DNA was isolated using a Plasmid Maxi Kit (Qiagen) according to the Manufacturers' instructions.

7.7.2 Quantitation of Nucleic Acid concentration

The concentration of purified double stranded DNA was quantified by measuring the OD at 260nm of a DNA sample diluted by 1/100 in water, on an LKB Biochrom Ultraspec II spectrophotometer. 1 absorbance unit at 260 nm corresponds to a 50 μg/ml double stranded DNA solution, therefore the following equation was used to obtain the concentration of the DNA in μg/μl:

(OD reading × dilution factor × 50 μg/ml) ÷ 1000

The purity of the sample was estimated by measuring the OD at 280. A ratio (OD_{260}/OD_{280}) of ≥ 1.8 but ≤ 2.0 is an indication of pure DNA.

7.7.3 Agarose gel electrophoresis

Agarose gels (0.8-2.5%) were prepared in 1xTBE with 0.5-2 μg/ml ethidium bromide. The molten gel was poured into a sealed gel tray with the appropriate gel comb(s) and allowed to set at room temperature. All gels were run in 1xTBE with a 1kb DNA ladder (New England Biolabs; NEB) at 120-150 volts for 30min-1hr. Visualisation and photography of gels was by ultraviolet illumination using a UVP 2UV Transilluminator.

Agarose gel loading buffer	0.01 % w/v OrangeG 5 % Glycerol in TBE buffer

7.7.4 Cloning Techniques

For cDNA subcloning, the appropriate DNA fragments were generated by restriction enzyme digestion from the original constructs when possible and ligated into appropriately R.E. digested vectors. Otherwise for other cloning procedures including generation of mutants by site-directed mutagenesis, the desired DNA fragments were produced by Polymerase Chain Reaction (PCR) with oligonucleotides containing appropriate restriction enzyme sites, and ligated into the chosen vectors. All PCR-

generated fragments were verified by DNA sequencing. Descriptions of the main methods used in all cloning procedures follow.

7.7.4.1 Purification of DNA fragments from enzymatic reactions

After enzymatic reactions, such as PCR, restriction enzyme digestion and DNA dephosphorylation, DNA fragments were purified using either the QIAquick™ or MinElute PCR purification Kit (Qiagen) according to the manufacturer's recommendations.

7.7.4.2 Gel extraction of DNA fragments

Reactions that generated more than one DNA species were electrophoresed on agarose gels and the appropriate band excised from the gel with a scalpel using a low power UV lamp (365 nm). The DNA was extracted from the gel slice using either the QIAquick™ or MinElute Gel Extraction Kit (Qiagen) according to the manufacturer's recommendations.

7.7.4.3 Polymerase Chain Reaction

PCR was used to amplify cDNA for subcloning and also for the production of mutants by site-directed mutagenesis (see below). Oligonucleotide primers were designed to be 21-36 mers depending on their purpose. Additionally, some primers were designed to incorporate restriction sites or the desired changes for introducing mutations.

Standard 50 µl PCR reactions contained:

1 µl of 100 ng/µl template DNA

1 µl each of 10 pmol forward and reverse primers

4 µl of 2.5 mM dNTPS

2.5 units DNA polymerase (Pfu or Proofstart)

in 1xPCR buffer (10 mM KCl, 20 mM Tris-HCl pH8.8, 2 mM $MgSO_4$, 10 mM $(NH_4)_2SO_4$,100 µg/ml BSA, 0.1 % Triton X-100).

PCRs were carried out in a Biometra TRIO-thermoblock thermal cycler.

Standard amplification conditions were as follows:

94°C	5 min	
94°C	1 min	⎫
47-55°C	30 s - 1 min	⎬ x 30 cycles
72°C	1 min	⎭
72°C	10 min	
4°C	∞	

After PCR, 1/10th of the reaction were run on an agarose gel to verify PCR product generation. Depending on the expreriment the rest of the reaction was purified directly or gel extracted after agarose gel electrophoresis

7.7.4.4 Site directed mutagenesis

Primers were designed on both strands of DNA that overlapped the region to be mutated. The mutation was centrally located within a primer, and was flanked either side by 15-18 bases of wild type sequence. Primers were also designed at locations 5' and 3' of the mutation which included unique restriction enzyme sites. Two PCR reactions were set up as described above, in both cases 1 primer containing the mutation and the corresponding primer containing the restriction enzyme site. The products from the first round of PCRs were analysed by agarose gel electrophoresis, and purified in 10 µl final volume. 1 µl of each first round PCR was used as template for a second round of PCR with the two outer primers flanking the mutation site. This product was again analysed on an agarose gel and the correct DNA fragment was purified. After restriction enzyme digestion, the DNA fragment was purified again and ligated into the appropriate restriction enzyme digested and phosphatase treated construct to generate a mutation within a wild type cDNA.

7.7.4.5 Restriction enzyme digestion

Restriction enzyme digestions were performed for 1-2 hr in the appropriate NEB buffer, at 37°C, unless recommended otherwise by the manufacturer. 10 µl of reaction mixture and 1 unit of enzyme were used per microgram of DNA. Cleavage was monitored by agarose gel electrophoresis. If the DNA fragment was required for further subcloning,

the reaction was either purified directly or electrophoresed and the correct fragment subsequently extracted from an agarose gel.

7.7.4.6 Dephosphorylation of DNA fragment ends

5' phosphate groups were removed from digested vectors that were going to be used for subcloning in order to prevent religation using calf intestinal alkaline phosphatase (CIP). The enzyme was added to the DNA directly after restriction enzyme digestion in one of the NEB buffers recommended by the manufacturer and incubated for 30 min – 1 hr at 37 °C prior to purification.

7.7.4.7 DNA Ligation

10 μl ligation reactions were carried out using approximately 250 - 500 ng of vector DNA, cut and treated with phosphatase, with a 3 fold molar excess of insert DNA and 200 units of T4 DNA ligase in 1 x T4 DNA ligase buffer (NEB) at 16°C overnight or at room temperature for 30 min. 1.5-2.5 μl of the ligation mixture were transformed and plated on agar plates (see section 7.5.3.2). Colonies formed were cultured the following day for miniprep DNA purification. Succesfuly ligated constructs were identified by restriction enzyme digestion and/or sequencing of miniprep DNA.

7.7.4.8 Sequencing

Sequencing reactions were carried out according to the ABI PRISM Dye terminator cycle sequencing kit. 150-200 ng of plasmid was mixed with 3.2 pmol of primer and 8 μl of Perkin Elmer Dirhodamine big dye terminator cycle mix (BDT versions 1.1 or 3.1) in a 20 μl reaction.

Thermal cycling was as follows:

96°C	5 min		
96°C	30 s	⎫	
50°C	15 s	⎬	x 25 cycles
72°C	4 min	⎭	
4°C	∞		

Unincorporated dye terminators were removed from sequencing reactions by ethanol precipitation: 2 µl of 0.125 mM EDTA and 2 µl of 3M NaOAc pH 5.2 and 50 µl 96% Ethanol were added to the PCR reactions and products were precipitated at room temperature for 15 min. Samples were centrifuged at 4 °C for 30 min, washed in 50 µl 70% ethanol and after a second centrifugation step the pellet was dried. Sequencing of precipitated DNA products was performed by the sequencing service of the Cancer Research UK Equipment Park. DNA sequences were analysed using an ABI PRISM 377 DNA sequencer. Sequence analysis was conducted with ABI Sequence Navigator and 4Peaks software.

7.8 Protein Manipulations

7.8.1 Protein production

7.8.1.1 Preparation of whole-cell extracts for gel mobility-shift assay

Transfections for whole-cell extracts used in bandshifts were performed using Lipofectamine reagent as described in section 7.6.3. Routinely 1-2 µg of expression plasmids were transfected. Where required stimulations of cells were performed for 30 minutes prior to extract preparation. Cells were washed twice with ice-cold PBS, ensuring all PBS is removed after the second wash before adding 80-100 µl of D0.4 lysis buffer to each plate. Cells were scraped into 1.5 ml micro-centrifuge tubes, and centrifuged at 13 000 rpm at 4°C for 5 min to remove cell debris. The supernatant was transferred to a cold microcentrifuge tube and stored at -80°C. Total protein concentration was determined using the Bradford protein assay (see section 7.8.2.1). Concentrations were in the range of 1.5 – 3 µg/µl protein solution. Approximately 2 µg/µl were used for electrophoretic mobility shift analysis. The expression levels of the proteins of interest were verified by using 5-10 µl of extract for western blot analysis (section 7.8.2.3).

D0.4 lysis buffer	20 mM HEPES pH 7.9
	10% Glycerol
	0.4 M KCl
	0.4% tritonX-100
	10 mM EGTA

	5 mM EDTA
	5 mM NaF
	1 mM DTT
Phosphatase and protease inhibitors and DTT were added fresh before use.	
Protease/phosphatase inhibitors used were:	
Pepstatin, Leupeptin,	
Aprotinin, Okadaic acid	10 mg/ml stock in water, used at 1/1000 dilution
PMSF	0.5mM stock in ethanol, used at 1/200 dilution

7.8.1.2 Preparation of whole-cell extracts for western blot analysis

Transfections for whole-cell extracts used in western blot analysis were performed using Lipofectamine reagent as described in section 7.6.3. Routinely 1-2 μg of expression plasmids were transfected. Cells were washed twice using ice-cold PBS, 80-100 μl of 1x SDS sample buffer was added and the cells were scraped off the dish using a 'rubber policeman'. 5-10 μl of extract was loaded per lane.

7.8.1.3 GST-fusion protein production

A fresh transformant of the GST fusion protein was used to inoculate a 10 ml preculture of LB media containing the appropriate antibiotics (chloramphenicol and kanamycin for MAL GST-fusions) and 1 % glucose to avoid premature induction. The culture was grown overnight in a shaking incubator at 37 °C and was diluted 1/200 the following day in the same media. This large-scale culture was grown to an OD_{600} of 0.6 before induction with 0.5 mM IPTG for 5 hr at 25°C. The culture was then centrifuged at 4000xg for 15 minutes at 4 °C, and the pellet was lysed in a suitable volume of ice cold Lysis Buffer (routinely 1/10 of the original culture volume) and was sonicated three times for 1 minute on ice and at maximum energy, on a SANYO Soniprep 150 MSE sonicator. The sonicated culture was centrifuged at 30,000xg for 30 min and the supernatant was respun under the same conditions for a further 15 min to remove all cell debris. The supernatant was added to glutathione-agarose bead 1:1 slurry (1 ml per litre of culture) pre-equilibrated by washing twice with Lysis buffer and incubated on

a rotor for 4 hours at 4 °C. The beads were centrifuged at 1500 rpm for 1 minute at 4 °C and washed twice with 10-15 ml of Lysis Buffer followed by a third wash with room temperature ATP Buffer to remove the bacterial chaperone proteins. The beads were washed again with Lysis Buffer and 500 µl of Glutathione Elution buffer were added to the beads which were incubated at 4°C on a rotor for 1 hour. The beads were centrifuged at 3500 rpm for 1 minute at 4 °C, the supernatant was transferred to an eppendorf, and the elution procedure was repeated. The eluates were pooled and dialysed against Dialysis buffer (1 litre of Dialysis buffer per 1 ml of protein solution) using Slide-A-Lyzer dialysis cassettes (Pierce) pre-equilibrated in water. Dialysis was overnight at 4°C with one buffer change and further dialysis for 4 hours the following day. Proteins were then visualised by Coomassie staining after SDS electorphoresis. Protein concentration was determined by comparison to a set of BSA standards run alongside, using in parallel the Bradford assay (section 7.8.2.1). Purified proteins were stored at -80°C.

GST Lysis Buffer	50 mM Tris-HCl pH 7.9
	100 mM NaCl
	1 % Triton X-100
	1 mM EDTA
	1 mM PMSF
	1 mM DTT
	15 µg/ml Benzamidine
ATP wash Buffer	50 mM Tris-HCl pH 7.5
	50 mM KCl
	20 mM $MgCl_2$
	5 mM ATP
	1 mM DTT
GST Elution Buffer	50 mM Tris-HCl pH 7.9
	100mM NaCl
	1 % Triton X-100
	1 mM EDTA
	1 mM PMSF
	1 mM DTT

	15 μg/ml Benzamidine
Dialysis Buffer	15 mM Tris-HCl pH 8.8 100mM NaCl 10 mM Glutathione (30 mg/10 ml) 1 mM DTT
Benzamidine	at 10 mg/ml stock in water
ATP	0.2 M stock (in 200 mM NaOH and 5 mM Tris pH 8.8)

7.8.1.4 In vitro protein translation with Reticulocyte lysate systems

Protein was prepared using the TNT Coupled Reticulocyte Lysate System (Promega) according to the manufacturer's instructions. After in vitro translation Ribonuclease A (0.2 mg/ml final concentration) was added to the 50 μl reticulocyte lysate reactions, and samples were incubated at 30°C for 5 minutes. A sample of each reaction (3 μl) was boiled in 2x SDS sample buffer and separated on a 16 % SDS gel (see section 7.8.2.2). Protein yield and integrity was analysed by western blotting (section 7.8.2.3). In vitro translated protein samples were stored at -80°C. 0.5-2 μl of sample were used in gel mobility-shift assays, while co-immunoprecipitation and pulldown assays contained whole 50 μl reactions.

7.8.2 Protein Analysis

7.8.2.1 Protein concentration quantitation (Bradford assay)

Protein concentration was measured using 200 μl of a 1 in 5 dilution of the Biorad Protein Assay reagent. A standard curve was prepared by making dilutions of a stock BSA solution (NEB). 1 μl of the sample whose concentration was to be determined was used in the assay, which was performed according to the manufacturer's instructions. The OD was measured at 595 nm using a SpectraMax Plus spectrophotometer (Molecular Devices). The protein concentration of the sample was then determined from the BSA standard curve.

7.8.2.2 SDS Polyacrylamide Gel Electrophoresis (SDS-PAGE)

Proteins were separated according to their size by SDS-PAGE using a minigel apparatus (ATTA). 6 ml of resolving gel (9 or 10 % gels were used for MAL protein separation and 16% for SRF DNA binding domain derivatives; see table for composition) was poured between two glass plates and the surface overlaid with methanol to ensure a level, air-free polymerisation interface. Following polymerisation, the methanol was removed, the stacking gel mix (see table for composition) was poured on top of the resolving gel and a plastic comb was inserted between the plates to produce the wells.

	Resolving gel			**Stacking gel**
Percentage	9 %	10 %	16 %	
40% acrylamide/ bisacrylamide (37.5:1)	3.38 ml	3.75 ml	6.0 ml	1.27 ml
1 M Tris-HCl (pH 8.8)	5.65 ml	5.65 ml	5.65 ml	-
1 M Tris-HCl (pH 6.8)	-	-	-	1.25 ml
10 % (w/v) SDS	150 µl	150 µl	150 µl	100 µl
Water	5.66 ml	5.29 ml	3.05 ml	6.37 ml
50 % Glycerol	-	-	-	0.9 ml
10 % (w/v) APS	150 µl	150 µl	150 µl	100 µl
TEMED	15 µl	15 µl	15 µl	10 µl

After polymerisation the comb and plastic seals were removed and the wells were washed with water and 1 x SDS running buffer. Samples were boiled for 5 min in SDS sample buffer to denature proteins before loading. Protein molecular weight markers (Rainbow markers: Amersham) were included in all experiments. Gels were run at 120-180 V in SDS running buffer until the dye front was run off the gel. After SDS-PAGE, gels were either stained with Coomassie blue stain for 30 minutes and then destained (multiple washes with destain solution on a shaker) to visualise the proteins or used for Western blot analysis of the proteins (section7.8.2.3).

SDS-PAGE buffers and solutions	
1 x SDS running buffer	192 mM glycine 25 mM Tris base pH 8.3 0.1 % SDS
2 x SDS sample buffer	125 mM Tris-HCl pH 6.8 4 % SDS 20 % glycerol 2% β-mercaptoethanol 0.01 % w/v bromophenol blue
Coomassie stain	0.5 % (w/v) Coomassie blue dye 40 % (v/v) methanol 7 % (v/v) acetic acid
Destain solution	40 % (v/v) methanol 7 % (v/v) acetic acid

7.8.2.3 Western Blot analysis

Following electrophoresis, proteins were transferred onto PVDF membranes (pre-soaked in methanol; Amersham) or nitrocellulose membranes (Whattman) sandwiched between Whatmann 3MM paper. Protein transfer was carried out at room temperature for 1 hr at 250-300 mA using a Mini Trans-Blot Cell (Biorad) with western blotting transfer buffer kept cool with an ice pack.

Following transfer the membrane was incubated in 5 % milk-blocking solution for 30 minutes at room temperature to block non-specific binding of antibody to filter. Next the membrane was incubated with the primary antibody in 2 % milk-blocking solution for 1 hour at room temperature or overnight at 4°C on a rocking platform. The membrane was then washed twice for 15 minutes in 1 % milk-blocking solution prior to incubation with a secondary antibody conjugated to horseradish peroxidase (HRP) in 2 % milk-blocking solution for 45 min - 1 hr. Finally the membrane was washed again twice for

15 minutes in 1 % milk-blocking solution, and then twice for 15 min in PBS and the HRP was detected with ECL Western blotting detection reagents (Amersham Pharmacia Biotech). ECL solutions A and B were mixed at 1:1 ratio and 5 ml were added to the membranes for 1 min. Excess liquid was removed, and the filter was exposed to Amersham ECL Hyperfilm.

For the HRP-conjugated primary antibodies (anti-HA-HRP and anti-Flag-HRP), after blocking the membrane was incubated with the appropriate dilution of antibody for 1 hr at room temperature or overnight at 4°C before washing twice for 15 minutes in 1 % milk-blocking solution, and detection with ECL Western blotting reagents was as described above.

Western blotting buffers and solutions	
Western blotting transfer buffer	192 mM glycine 25 mM Tris base, 20 % methanol
5 % milk-blocking solution	5 % (w/v) milk powder (Marvel) 0.5 % (v/v) Triton X-100 in PBS

Antibodies used for western blotting analysis:

Primary antibodies			
Antibody	**Specificity**	**Species**	**Concentration**
9E10 (Cancer Research UK)	Myc	Mouse	1/1000
M2 Flag (Sigma)	Flag	Mouse	1/2000
Flag-HRP (Sigma)	Flag	Mouse	1/500
HA-HRP 3F10 (Roche)	HA	Rat	1/500

Secondary Antibodies			
Antigen	**Conjugate**	**Species**	**Concentration**
Rabbit IgG (DAKO)	HRP	Goat	1/2000
Mouse IgG (DAKO)	HRP	Goat	1/2000

7.8.2.4 Membrane stripping

To strip membranes of antibodies, they were incubated in glycine stripping buffer (25 mM glycine, pH 2.5, 0.1 % SDS) for 30 minutes at room temperature with gentle agitation. The membrane was then washed twice for 15 minutes with PBS and blocked for 30 minutes with 5 % milk blocking solution, before proceeding normally with the primary antibody.

7.8.2.5 Protein co-immunoprecipitation assay

Cells in 10 cm dishes were lysed in 300 - 1000 µl RIPA buffer containing freshly added protease inhibitors (Complete protease inhibitor cocktail – Roche). Lysates were centrifuged at 13,000 rpm for 20 minutes at 4 °C. The supernatant was transferred to a sterile pre-chilled eppendorf and diluted 1/2 with 1 % TX buffer. 40 µl samples were retained for use as controls. The supernatant was then incubated with 20 µl Anti-Flag M2 or Anti-myc Affinity Gel beads (Sigma), pre-equilibrated in 1% TX buffer, for 2-4 hours at 4 °C with rotation. The beads were washed three times with 1 % TX buffer, resuspended in 2 x SDS sample buffer and Western blotted to visualise associated proteins. MAL-SRF co-immunoprecipitations in the presence of DNA included non-radioactive c-fos promoter PCR products produced as described in section 7.9.1.4 and purified as described in section 7.7.4.1.

Co-immunoprecipitation buffers	
RIPA buffer	20 mM Tris-HCl pH 7.9
	150 mM NaCl
	1 mM EDTA
	5 % Glycerol
	1 % Triton X-100
	0.5 % Deoxycholate
	0.1 % SDS
1 % TX buffer	20 mM Tris-HCl pH 7.9
	150 mM NaCl
	1 mM EDTA
	5 % Glycerol
	1 % Triton X-100

7.8.2.6 GST-protein pulldown assay

10μg of purified GST-MAL proteins (section 7.8.1.3) were incubated with 100μl in vitro translated SRF.DBD (section 7.8.1.4) and 30μl Glutathione Sepharose beads (washed once in pulldown buffer) in 500μl pulldown buffer for 4hr at 4°C with rotation. After washing twice with 500 μl of pulldown buffer and discarding the supernatant the beads were resuspended in 20 μl 2x SDS sample buffer and proteins were fractionated on 16% gels as described in section 7.8.2.2. GST-MAL input was visualised by Ponceau S staining, and recovery of SRF.DBD by immunoblotting (section 7.8.2.3).

Pulldown buffer	50 mM Tris-HCl pH 7.9 150 mM NaCl 0.1 % Triton X-100 protease inhibitors were added fresh
Ponceau S solution	2 % (w/v) Ponceau S 30 % (v/v) trichloroacetic acid

7.9 Protein-DNA interaction analysis

7.9.1 Gel mobility-shift assay

Bandshift assays (Electrophoretic Mobility Shift Assays – EMSA) were used to investigate protein–DNA interactions. Two glass plates (19 cm x 20 cm) one of which had previously been siliconised with Dimethyldichlorosilane solution were assembled with 1.5 mm plastic spacers and sealed with electrical tape. Non-denaturing acrylamide gel mix (50 ml) was then poured between the plates and a 1.5 mm plastic comb was inserted between the plates to produce the wells. 4% gels were routinely used for MAL whole-cell extract bandshifts, circular permutation analysis was performed with 5 % gels and 8 % gels were used for MAL peptide – SRF.DBD bandshifts.

Non-denaturing acrylamide gel composition			
Percentage	4 %	5 %	8 %
40% (w/v) acrylamide	6 ml	7.5 ml	12 ml
1 x TBE	3 ml	3 ml	3 ml
Water	50 ml	49.5 ml	44 ml
10 % (w/v) APS	58 µl	58 µl	58 µl
TEMED	700 µl	700 µl	700 µl

After gel polymerisation the comb and tape were removed and the gels were mounted on the gel electrophoresis tanks (Cambridge). The wells were rinsed with 0.5 x TBE to removed unpolymerised gel mix. Bandshift gels were pre-run at 170 Volts for 1-2 hours in 0.5 x TBE at room temperature prior to sample loading.

Bandshift binding reactions containing different combinations of whole-cell extract (up to 2.5 µl cell extract; see section 7.8.1.1), recombinant SRF DNA binding domain (5 – 10 ng/µl) and in vitro translated proteins (usually 1 µl of reticulocyte lysate) were mixed with 5 µl 2 x DBB to a final volume of 9 µl. Recombinant SRF DNA-binding domain was either the SRF (residues 133-265) fragment (previously produced in baculovirus and purified in the Treisman lab) or the SRF(132-223) fragment ((Pellegrini, L. *et al.*, 1995) produced in E. coli from the pET3a vector and purified by the CRUK protein production lab). Reactions were incubated on ice for 5 min before addition of 0.25 - 1 ng/µl radiolabelled DNA probe and further incubation for 15 min at room temperature. Bandshift loading buffer was then added, and samples were loaded onto the non-denaturing polyacrylamide gel. Gels were run for the appropriate length of time at 170 Volts before being transferred onto 3MM Whatman paper and dried on a gel drier at 80 °C for 45 minutes. The bandshift was visualised by overnight exposure on KODAK Biomax Film at -80°C.

Bandshift buffers and solutions	
2xDBB	2 mM EDTA 20 mM Tris-Hcl pH 7.9 100 mM NaCl, 1 mM DTT

	100 ng/ml BSA
	50-200 ng/µl poly(dI-dC)•poly(dI-dC)
Bandshift loading buffer	60 % Glycerol
	10 mM EDTA
	0.01 % w/v Xylene cyanol
	0.01 % w/v bromophenol blue
poly(dI-dC)•poly(dI-dC)	1 mg/ml stock made in TEN
The solution was heated to 72°C for 5-10 minutes to anneal and was allowed to cool at room temperature for 30 min before sonicating twice for 15 seconds at 3/4 maximum energy to ensure fragment sizes between 200 – 400 bp	

7.9.1.1 Supershift assays

Supershift assays were performed as described in section 7.9.1. Appropriate antibodies were added to bandshift reactions and incubated with the protein for 10 minutes on ice prior to addition of the DNA probe. The following antibodies were used: anti-Flag M2 (Sigma), anti-SRF (Santa Cruz), anti-MAL (raised against MAL residues 1 - 170 by the CRUK antibody production facility (Miralles, F. *et al.*, 2003)). 0.2 – 0.4 ng of antibody was used per reaction.

7.9.1.2 Peptide competition bandshifts

Peptide competition bandshifts were performed as described in section 7.9.1. Peptides were added to bandshift reactions containing the proteins of interest and reactions were incubated for a further 5 minutes on ice prior to addition of the DNA probe. For a description of the composition of all peptides used in bandshift reactions, see section 7.4.

7.9.1.3 MAL peptide – SRF.DBD bandshifts

Peptide bandshifts were performed as described in section 7.9.1. Peptides (concentrations ranging from 0.014 - 20 µM) were added to bandshift reactions containing in vitro translated SRF.DBD and reactions were incubated for 5 minutes on

ice prior to addition of the DNA probe. Complexes were resolved on 8% native gels. For a description of the composition of all peptides used in bandshift reactions, see section 7.4.

7.9.1.4 Generation of Bandshift DNA probes

Probes were labelled with [α^{32}P] dCTP and generated by PCR.

20 µl PCR reactions contained:
2 µl 10 x Biotaq buffer (Bioline)
2 µl $MgCl_2$ (15mM)
2 µl d(A, G and T)TP (500µM)
2 µl dCTP (200µM)
4 µl [α^{32}P] dCTP (3000µCi)
2 µl Forward oligonucleotide (7pmol/µl)
2 µl Reverse oligonucleotide (7pmol/µl)
1 µl DNA template (50 ng/µl)
0.5 µl BioTaq polymerase (1 u/µl)
2.5 µl H_2O

Thermal cycling was:			
	95 °C	2 min	
	95 °C	1 min	]
	65 °C	1 min	} x 30cycles
	72 °C	1 min	]
	72 °C	5 min	
	4 °C	∞	

The PCR product was precipitated by incubation with an equal volume of 5M NH_4OAc and 8 volumes of 95 % ethanol at -20°C for at least 1 hr. The precipitant was washed before loading on a 5 - 7% non-denaturing polyacrylamide gel in TBE buffer. The gel was run at 170 V for approximately 2 hours and then subjected to autoradiography with MXB Film (Kodak) for 1 – 2 minutes to identify the location of the probe band on the gel. Identified probe bands were excised from the gel, and probe was extracted by

overnight by incubation with 450 µl probe extraction buffer at 37°C. The probe was precipitated with 2.5 volumes of ethanol, washed with 70 % ethanol, air dried and then dissolved in TEN. The amount of hot probe was quantitated by Cerenkov counting of 1 µl of the product. 0.5 – 1 ng/µl was used per electrophoretic mobility shift sample.

Bandshift probe loading buffer	60 % Glycerol 10 mM EDTA 0.01 % w/v Xylene cyanol 0.01 % w/v bromophenol blue
Probe Extraction buffer	0.5 M NH_4OAc 1 mM EDTA

7.9.1.5 Probe Quantitation

The amount of PCR product was quantitated by calculating the Decay Factor of the ^{32}P dCTP stock: Decay Factor $^{32}P = \exp^{(-0.0485 \times t)}$. This was then used to calculate the specific activity of the PCR product in cpm/pg using the following equation:

$$\text{Specific Activity} = \frac{40\ (\mu Ci) \times 10^6\ (\text{Cerenkov cpm}/\mu Ci) \times \text{Decay Factor}}{20\ (\mu l) \times 40\ (\text{pmol}/\mu l) \times 4\ (\text{dNTPs})}$$

1 µl of the resuspended probe was counted in a scintillation counter and the cpm value was divided by the specific activity to give the amount of probe in pg/µl.

7.9.1.6 c-fos plasmids used for bandshift DNA probe construction

FOS.WT	pF711, described in (Treisman, R., 1985).
FOS.M	pF711 derivative, described in (Hill, C. S. *et al.*, 1995b).
FOS.L	pF711 derivative, described in (Hill, C. S. *et al.*, 1995b). Also called FOS. ΔTCF

7.9.1.7 Oligonucleotides used for bandshift DNA probe construction

The table that follows lists the oligonucleotides used to generate bandshift probes, including their sequences and probe names. Oligonucleotide pairs used for circular permutation probes are numbered and designated F and R (forward and reverse). The wild-type FOS probe generated with oligos P10 and P11 was also used in circular permutation analysis. Generation of nested probes used forward oligos CP1 F, CP1.3 F, CP1.6 F, CP1.9 F and CP2 F with reverse oligo CP5 R for one set and forward oligo P10 with reverse oligos CP6 R, CP6.3 R, CP 6.6 R, CP6.9 R and CP7 R for the other set. Nested probes are named after the number of oligonucleotides between the centre of the SRF CArG box and the closest fragment end.

Probe	**Oligo pair**	**Oligonucleotide sequence**
FOS.(WT, L, M and LM)	P10 (forward)	CGCACTGCACCCTCGGTGTTGGCTGC
	P11(reverse)	ATGGCTCCCCCCAGGGCTACAGGGAAA
CP1	CP1 F	ACACAGGATGTCCATATTAGGACAT
	CP1 R	GAGCATTTCGCAGTTCCTGTCTCAG
CP2	CP2 F	CCTCCCCCCTTACACAGGATGTCCA
	CP2 R	AGTTCCTGTCTCAGAGGTCTCGTGG
CP3	CP3 F	TCCCGTCAATCCCTCCCCCCTTACA
	CP3 R	CAGAGGTCTCGTGGGCCCCCCAAGA
CP4	CP4 F	CCGCGAGCAGTTCCCGTCAATCCCT
	CP4 R	TGGGCCCCCCAAGATGAGGGGTTTC
CP5	CP5 F	TTGGCTGCAGCCCGCGAGCAGTTCC
	CP5 R	AGATGAGGGGTTTCGGGGATGGCTC
CP6	CP6 F	CCGTTCCCGCCTCCCCTCCCCCAGC
	CP6 R	GACGCAGATGTCCTAATATGGACAT
CP7	CP7 F	CCCCTCCCCCAGCCGCGGCCCCCGC
	CP7 R	TGGAAACCTGCTGACGCAGATGTCC
CP8	CP8 F	GCCGCGGCCCCCGCCTCCCCCCGCA
	CP8 R	GGGAAAGGCCGTGGAAACCTGCTGA
(-16)	CP1 F	ACACAGGATGTCCATATTAGGACAT
	CP5 R	AGATGAGGGGTTTCGGGGATGGCTC

Probe	Oligo pair	Oligonucleotide sequence
(-19)	CP1.3 F	CTTACACAGGATGTCCATATTA
	CP5 R	AGATGAGGGGTTTCGGGGATGGCTC
(-22)	CP1.6 F	CCCCTTACACAGGATGTCCATA
	CP5 R	AGATGAGGGGTTTCGGGGATGGCTC
(-25)	CP1.9 F	TCCCCCCTTACACAGGATGTCC
	CP5 R	AGATGAGGGGTTTCGGGGATGGCTC
(-27)	CP2 F	CCTCCCCCCTTACACAGGATGTCCA
	CP5 R	AGATGAGGGGTTTCGGGGATGGCTC
(+16)	P10 (forward)	CGCACTGCACCCTCGGTGTTGGCTGC
	CP6 R	GACGCAGATGTCCTAATATGGACAT
(+19)	P10 (forward)	CGCACTGCACCCTCGGTGTTGGCTGC
	CP6.3 R	GCTGACGCAGATGTCCTAATAT
(+22)	P10 (forward)	CGCACTGCACCCTCGGTGTTGGCTGC
	CP6.6 R	CCTGCTGACGCAGATGTCCTAA
(+25)	P10 (forward)	CGCACTGCACCCTCGGTGTTGGCTGC
	CP6.9 R	AAACCTGCTGACGCAGATGTCC
(+28)	P10 (forward)	CGCACTGCACCCTCGGTGTTGGCTGC
	CP7 R	TGGAAACCTGCTGACGCAGATGTCC

7.9.2 DNase I footprinting assay

DNase I footprint reactions contained 10,000cpm 5' or 3' radiolabelled probe, 0.85ng recombinant SRF(132-223) and 0.85-22.8ng GSTMAL(214-298) derivatives or 1.7 – 85 μM MAL peptides in 1x DNase I buffer with 2.5ng poly(dI-dC)•poly(dI-dC) and 3 mM spermidine. After incubating at room temperature for 30min, 0.25 units DNase I were added for 5min on ice. Reactions were stopped with Stop buffer at 50°C for 1hr. The DNA was precipitated with 10 μl of 1M LiCl and 3 volumes 96% ethanol on dry ice for 30 minutes. After washing the DNA was pelleted and resuspended in 6μl of formamide loading buffer. Reactions were counted in a scintillation counter and equal cpm were loaded on a 8% polyacrylamide/8M Urea sequencing gel, which had been pre-run for 30 minutes. Samples were electrophoresed at 16mAmps for approximately 2 hr. Gels were transferred onto 3MM Whatman paper and dried on a gel drier at 80°C for 45 minutes. The footprint was visualised by overnight exposure on KODAK Biomax Film at -80°C or by exposure to phosphorimager for 1hr.

DNase I footprint buffers and solutions	
10x DNase I buffer	200 mM Tris-HCl pH 7.5 500 mM NaCl, 30 mM $MgCl_2$ 10 mM $CaCl_2$ 20 mM DTT 10 μM $ZnCl_2$ 20 % glycerol 1 mg/ml BSA protease inhibitors added fresh
DNase I stock solution	1 mg/ml (2 units/μl) in 10 mM Tris-HCl pH 7.5 and 50 % glycerol kept at -20°C
Stop buffer	20mM Tris-HCl pH 7.5 50 mM EDTA 2 % SDS 0.25 mg.ml linear acrylamide 0.2mg Proteinase K
Formamide loading buffer	95% formamide 20 mM EDTA 0.01 % w/v Xylene cyanol 0.01 % w/v bromophenol blue
8% denaturing gel solution	25ml 40% acrylamide/bisacrylamide (19:1) solution 12.5 ml 10x TBE 62.5g Urea 42 ml H_2O The mix was heated to 37°C, filtered and stored at 4°C until use.

7.9.2.1 Generation of DNase I footprinting probes

Probes were generated by EcoRI digestion of the Δ(-363) pF711 c-fos promoter construct (Treisman, R., 1986) and labelling of either the 5' end with γ ^{32}P-ATP and T4 Polynucleotide Kinase or the 3' end by filling in with α ^{32}P-dATP and Klenow Polymerase. The DNA was then digested with NotI to generate radiolabelled probes of 245bp (5'label) and 251bp (3'label). Probes were purified on native polyacrylamide gels and Cerenkov counted in a scintillation counter. 10,000cpm of probe were used per DNase I footprinting reaction.

5' end (top strand) labelling with T4 Polynucleotide Kinase

10 μg of DNA were digested with EcoRI as described in section 7.7.4.5. An equal volume of phenol:chloroform solution was added to the reaction and after vortexing centrifugation was at maximum speed and room temperature for 10 minutes. The aqueous phase removed and DNA was ethanol precipitated with 1/10 volume 3M NaOAc pH 5.2 and 2.5 volumes 96% ethanol. The DNA pellet was air dried and resuspended in 10 μl water.

5' end phosphorylation reactions contained:
2 μl 10 x PNK buffer (NEB)
5 μl [γ^{32}P] ATP (10 μCi/μl)
10 μl EcoRI digested DNA
1 μl Polynucleotide kinase
2 μl H_2O

Reactions were incubated for 30 minutes at 37°C, followed by addition of 0.5 μl of PNK and incubation for a further 30 minutes at 37°C. PNK was heat-inactivated for 20 minutes at 65°C and 20 μl TE were added to neutralise the reaction, which was ethanol precipitated as described above in the presence of 10-20 μg/ml linear acrylamide. The pellet was resuspended in 20 μl water and the radiolabelled DNA was digested with NotI as described in section 7.7.4.5 to release the radiolabelled EcoRI-NotI probe fragment. The reaction was ethanol precipitated and resuspended in 10 μl TE followed by DNA probe purification by native polyacrylamide gel electrophoresis as described in section 7.9.1.4.

3' end (bottom strand) labelling with α ^{32}P-dATP and Klenow Polymerase

10 μg of DNA were digested with EcoRI as described in section 7.7.4.5. The DNA was purified with the QIAQUICK reaction purification kit (section 7.7.4.1) and eluted in 30 μl water.

3' end labelling reactions contained:
5 μl 10 x NEB buffer 2
5 μl [α^{32}P] dATP (10μCi/μl)
1 μl d(C, G and T)TP (1mM)
30 μl EcoRI digested DNA
4 μl Klenow polymerase
5 μl H_2O

Reactions were incubated for 30 minutes at room temperature, followed by addition of 1 μl of 10mM dNTPs (A, G, C, T) and incubation for a further 15 minutes at room temperature. The reaction, was ethanol precipitated as described above in the presence of 10-20 μg/ml linear acrylamide. The pellet was resuspended in 20 μl water and the radiolabelled DNA was digested with NotI as described in section 7.7.4.5 to release the radiolabelled EcoRI-NotI probe fragment. The reaction was ethanol precipitated and resuspended in 10 μl TE followed by DNA probe purification by native polyacrylamide gel electrophoresis as described in section 7.9.1.4.

7.9.2.2 Preparation of AG marker for footprinting analysis

Maxam-Gilbert AG sequencing reactions of the DNA probes were produced by partial depurination of DNA with formic acid and subsequent DNA backbone cleavage with piperidine, and were used as DNA marker ladders. 50,000 – 100,000 cpm of radiolabelled DNA probe were mixed with 1μg Sonicated Salmon Sperm DNA (Stratagene) in TE buffer in a 10 μl final volume. After addition of 1μl of 4% Formic Acid the mix was incubated at 37°C for 25 minutes and then placed on ice. 150 μl of 1M piperidine solution were added and the reaction was initially incubated at 90°C for 30 minutes and subsequently placed on ice for 5 minutes. 1 ml n-butanol was added and the mix was vortexed vigorously and centrifuged for 2 minutes at maximum speed and at room temperature to pellet the DNA. After removal of the supernatant 150 μl of 1% SDS solution and 1 ml of n-butanol were added to the pellet and the vortexing and

centrifugation steps were repeated. The supernatant was removed and the pellet was washed twice with 0.5 ml of n-butanol and dried under vacuum for 10 minutes. 10 µl of formamide loading dye were added and the recovery of DNA marker was measured with a scintillation counter before storing at -20°C. Approximately 5,000 cpm of AG marker were loaded per experiment.

7.9.3 In vitro selection of transcription factor DNA binding sites

Binding site selection was carried out essentially as described in (Pollock, R. *et al.*, 1990).

7.9.3.1 Oligonucleotides used in Binding Site Selection

Binding site selection oligos were designed to contain a random oligonucleotide sequence flanked by primers that included restriction enzyme sites for cloning and analysis of selected sites. Oligo SS1 32N contained 32 random nucleotides, whereas oligo SS2 conSRE included a partially set SRE (CCWWAWWWGG) flanked by 21 random nucleotides on either side, thus constraining SRF binding to the middle of the sequence.

Oligonucleotide name	**Oligonucleotide sequence**
SS1 32N	CAGGTCAGTTCAGCGAATTCTGTCG$(N)_{32}$GAGGC AAGCTTAGTGCAACTGCAGC
SS1 32N EcoRI F	CAGGTCAGTTCAGCGAATTCTGTCG
SS1 32N HindIII R	GCTGCAGTTGCACTAAGCTTGCCTC
SS2 conSRE	CAGTTCAGCGAATTCTGATG $(N)_{21}CC(W)_2A(W)_3GG\ (N)_{21}$ CAGCGAAGCTTAGTGCTACT
SS2 conSRE EcoRI F	CAGTTCAGCGAATTCTGATG
SS2 conSRE HindIII R	AGTAGCACTAAGCTTCGCTG

7.9.3.2 Preparation of double stranded oligonucleotide randomer

The oligonucleotide randomer was rendered double stranded and labelled with [α^{32}P] dCTP by PCR.

20 µl PCR reactions contained:
2 µl 10 x Biotaq buffer (Bioline)
2 µl $MgCl_2$ (15mM)
2 µl d(A, G and T)TP (500µM)
2 µl dCTP (40µM)
5 µl [α^{32}P] dCTP (10 µCi/µl)
2 µl Reverse primer SS1 R or SS2 R (80ng/µl)
2 µl oligonucleotide randomer (50 ng/µl)
1 µl BioTaq polymerase (1 u/µl)
2 µl H_2O

The following PCR cycle was performed once:	
94 °C	1 min
62 °C	3 min
72 °C	10 min
2 µl of 0.5 mM cold dCTP were added and reactions were incubated at	
72 °C	10 min

The product was purified on an 8% native polyacrylamide gel as described in section 7.9.1.4. The final DNA pellet was resuspended in 20 µl water and after Cerenkov counting and amount quantitation (see section 7.9.3.5) the ds oligo was diluted to 0.1 ng/µl in water.

7.9.3.3 Binding and Recovery of oligonucleotides

Reaction buffers and conditions including protein amounts were kept as similar as possible to the known MAL-SRF bandshift conditions (section 7.9.1). The binding site selection using SS1 32N as the starting oligo pool was performed by GST-pulldown of GST.MAL(214-298) and recombinant SRF(132-223). Binding reactions contained 6 ng/µl GST fusion proteins, 1.36 ng/µl SRF(132-223) (both diluted in buffer D0.4 described in section 7.8.1.1). The binding site selection using SS2 conSRE as the starting oligo pool was performed by immunoprecipitation of whole-cell extracts expressing flag-tagged MALΔN and MALΔNΔB1 with untagged SRF whole-cell extract. Mock-transfected extracts combined with untagged SRF were used as negative control and flag-tagged SRF extract was used as a positive control.

Binding reactions were set up in 0.5 ml tubes and included 0.2 ng/µl of radiolabelled oligonucleotide (0.4 ng/µl in the first round of selection) in 1x SS Binding Buffer. Reactions were incubated for 30 minutes at room temperature before addition of 20 µl GST-sepharose bead slurry (washed once in SS wash buffer). Samples were incubated on a rotor at 4°C for 2.5 hours. The supernatant was removed and beads were washed twice with SS wash buffer. Selected oligonucleotides were eluted by addition of 200 µl elution buffer and incubation at 45 °C for 1 hour. After extraction with an equal volume of phenol:chloroform solution the selected oligos were ethanol precipitated. Pellets were dissolved in 10 µl water and recovery was quantitated by Cerenkov counting to work out the concentration of selected DNA (section 7.9.3.5).

Binding site selection buffers

2x SS Binding Buffer	40 mM Hepes pH 7.9
	200 mM NaCl
	0.4 mM EDTA
	0.4 mM EGTA
	20 % glycerol
	100 ng/ml BSA
	16 ng/µl poly(dI-dC)•poly(dI-dC)
	0.2 % Triton X-100
	2 mM DTT
	6 mM spermidine
	protease inhibitor cocktail (Roche)
SS wash buffer	20 mM Hepes pH 7.9
	100 mM NaCl
	0.2 mM EDTA
	0.2 mM EGTA
	0.1 % Triton X-100
	1 mM DTT
	protease inhibitor cocktail (Roche)

Buffers were stored at 4°C. poly(dI-dC)•poly(dI-dC), spermidine, DTT and protease inhibitors were added fresh prior to use.

SS Elution solution	50 mM Tris pH 7.9
	100 mM NaOAcl
	5 mM EDTA
	0.5 % SDS

7.9.3.4 Amplification of selected oligonucleotides

1 pg of selected DNA was amplified for use in subsequent rounds of selection.

20 µl PCR reactions contained:

2 µl 10 x Biotaq buffer (Bioline)
2 µl $MgCl_2$ (15mM)
3.2 µl d(A, G and T)TP (500µM)
2 µl dCTP (40µM)
1 µl [α^{32}P] dCTP (10 µCi/µl)
2 µl Forward primer SS1 F or SS2 F (80ng/µl)
2 µl Reverse primer SS1 R or SS2 R (80ng/µl)
1 µl selected oligonucleotide (1 pg/µl)
0.5 µl BioTaq polymerase (1 u/µl)
2 µl H_2O

The following PCR programme was used:

94 °C	30 sec	]
62 °C	30 sec	} x 17 cycles
72 °C	30 sec	]
72 °C	2 min	

The product was purified on an 8% native polyacrylamide gel as described in section 7.9.1.4. The final DNA pellet was resuspended in 10 µl water and after Cerenkov counting and amount quantitation (see section 7.9.3.5) the selected DNA was diluted to 0.1 ng/µl in water.

7.9.3.5 Quantitation of oligonucleotides

The amount of double stranded oligonucleotide material produced for use as the original sequence pool was determined as follows:

1μl of the radiolabelled DNA was Cerenkov counted in a scintillation counter. The ratio of hot/total dCTP that can be incorporated by the double stranding reaction was calculated for each oligonucleotide, to allow an estimation of the original specific activity of the sample.

The specific activity of the synthesised DNA was then calculated using the radioactive decay equation: $N/No = e^{-0.0485xt}$

where No= the original specific activity of the sample, and N= the specific activity after time t.

Assuming that $1x10^6$ cpm is approximately equal to 1μCi, the Specific Activity can be converted into moles and using the MW of each oligonucleotide this can be converted into grammes, thus allowing the estimation of the concentration of the 1μl sample counted originally.

The same method was used to quantitate the amouns of amplified selected DNA, with the exception that the hot/total dCTP ratio was different in this case.

7.9.3.6 Recovery of selected DNA by gel mobility-shift assay and analysis by cloning and sequencing

After four rounds of selection the selected probes were used in gel mobility-shift assays (section 7.9.1) with the selecting proteins and appropriate controls. After drying and the gel was exposed to film to visualise. Specific bands formed between the protein of interest and the DNA were excised and soaked in 70-100 μl water at 37°C overnight.

10-20 μl of this were used in a 30 μl PCR reaction:

- 3 μl 10 x Biotaq buffer (Bioline)
- 3 μl $MgCl_2$ (15mM)
- 2.5 μl d(A, G and T)TP (1 mM)
- 1 μl dCTP (120μM)
- 1.5 μl [α^{32}P] dCTP (10 μCi/μl)
- 3 μl Forward primer SS1 F or SS2 F (80ng/μl)
- 3 μl Reverse primer SS1 R or SS2 R (80ng/μl)

10 µl selected oligonucleotide (1 pg/µl)
0.5 µl BioTaq polymerase (1 u/µl)
2.5 µl H_2O

The following PCR programme was used:		
94 °C	1 min	⎫
62 °C	1 min	⎬ x 17 cycles
72 °C	1 min	⎭
72 °C	5 min	

The PCR product was purified as described in section 7.9.1.4, including a phenol extraction step prior to the final ethanol precipitation. The DNA pellet was resuspended in 20 µl water and digested with the appropriate restriction enzymes (EcoRI and HindIII) as described in section 7.7.4.5. After phenol extraction with an equal volume of phenol:chloroform solution and subsequent ethanol precipitation the pellet was dissolved in water and ligated (see section 7.7.4.7) into the pBSKS+ vector (Stratagene), which had previously been digested with EcoRI and HindIII and treated with alkaline phosphatase (sections 7.7.4.5 and 7.7.4.6). E. coli were subsequently transformed with ligated DNA and individual colonies were grown into miniprep cultures (sections 7.5.3 and 7.7.1). Purified DNA was sequenced as described in section 7.7.4.8.

7.9.3.7 Analysis of selected oligonucleotides

Selected oligonucleotides were analysed manually for the SS1 32N experiments. Details are explained in the text of Chapter 8/Appendix. The sequences derived from the SS conSRE experiments were analysed by Mike Mitchell of the CRUK Cgal facility, using the Improbizer programme, a tool specialised in searching sets of sequences for motifs. Improbizer randomly samples the input sequences many times searching for over-represented sub-sequences, scoring the identified matrix according to conservation. Improbizer will start with motifs of the specified size and allow them to grow if the conservation extends outside of the initial high scoring region. The sequence motif images for the SS2 conSRE selected sequences were made by Mike Mitchell using the Weblogo programme.

8 Appendix

8.1 Binding site selection aims

The results described in the preceding chapters highlight the central role of DNA in the interaction of MAL with SRF and raise the question of the significance of the CArG sequence itself for the formation of the ternary MAL-SRF-DNA complex. Additionally, the discovery that the SRF-induced DNA distortion serves to facilitate direct interactions between MAL and the DNA sequences flanking the SRE raises the issue of specificity in the formation of these MAL-DNA contacts. In the present chapter I describe my preliminary attempts to investigate the SRE specificity of the MAL-SRF complex and also the possible sequence dependence of the MAL-DNA contacts within the MAL-SRF-DNA complex.

8.2 Site selection 1: selection of MAL-SRF specific DNA binding sites

To explore the possibility that the MAL-SRF complex recognises DNA sites deviating from the SRF consensus I employed an in vitro binding site selection technique (Pollock, R. *et al.*, 1990; Thiesen, H. J. *et al.*, 1990). This technique involves using a pool of random sequence oligonucleotides as the source of potential binding sites. To facilitate amplification the random sequence is flanked on either side by specific primer sequences, which contain the desired restriction enzyme sites for cloning and analysis of the selected sequences. After incubation of the double stranded oligonucleotides with the protein of interest the formed nucleoprotein complexes are isolated by precipitation on affinity beads. The selected DNA is then recovered and PCR amplified for use in subsequent rounds of selection. After a few selection rounds the recovered DNA is used in a gel mobility-shift assay with the selecting protein and specific bands are excised in order to extract selected protein-bound DNA. This DNA is then cloned into an appropriate vector and analysed by sequencing. This technique has been used successfully in the past to identify the consensus binding sites of many proteins including SRF, MCM1 and of Elk-1 in complex with SRF ((Pollock, R. *et al.*, 1990; Treisman, R. *et al.*, 1992; Wynne, J. *et al.*, 1992); see also Section 1.2.2.1 of the Introduction).

The experimental design of the MAL-SRF binding site selection experiment is outlined in Figure 8.1. SS1 32N, the oligonucleotide designed for the experiment included a random 32 nucleotide sequence. MAL-SRF specific DNA sites were selected with purified proteins to avoid non-specific binding effects from other proteins in cell extracts and were isolated using a GST-pulldown assay. Binding conditions and protein concentrations were identical to those used in the standard MAL-SRF gel mobility-shift assay.

To identify DNA sites specified by the minimal MAL and SRF regions required for complex formation I used the GST.MAL(214-298) fusion protein that contains the B1 and Q regions and the recombinant SRF DNA-binding domain SRF(132-223). A sample of GST.MAL(214-298) alone was also included to test whether the B1-Q region of MAL has the ability to bind DNA specifically in the absence of other proteins. Samples containing empty GST and empty GST combined with SRF(132-223) served as negative controls. Given the well characterised SRF binding consensus my initial experiments did not include GST.SRF(132-223) fusion protein as a positive control. The initial oligonucleotide pool and the DNA samples recovered after each amplification step were radiolabelled to facilitate analysis by EMSA.

After four rounds of selection the recovered DNA from each sample was tested in gel mobility shift assays for the formation of specifically enriched complexes with the selecting proteins (Figure 8.2, the red line under each bandshift specifies the reactions where specifically enriched complexes might be found).

Selection with GST.MAL(214-298) in the absence of SRF did not enrich for DNA that generated specific complexes. Thus the B1-Q regions of MAL are not competent for autonomous DNA binding under these binding conditions (Figure 8.2, lanes 1-5), although this does not exclude the possibility that they might make specific DNA contacts within a ternary complex with SRF. A non-specific band was present in most reactions including those containing starting DNA material (Round 0).

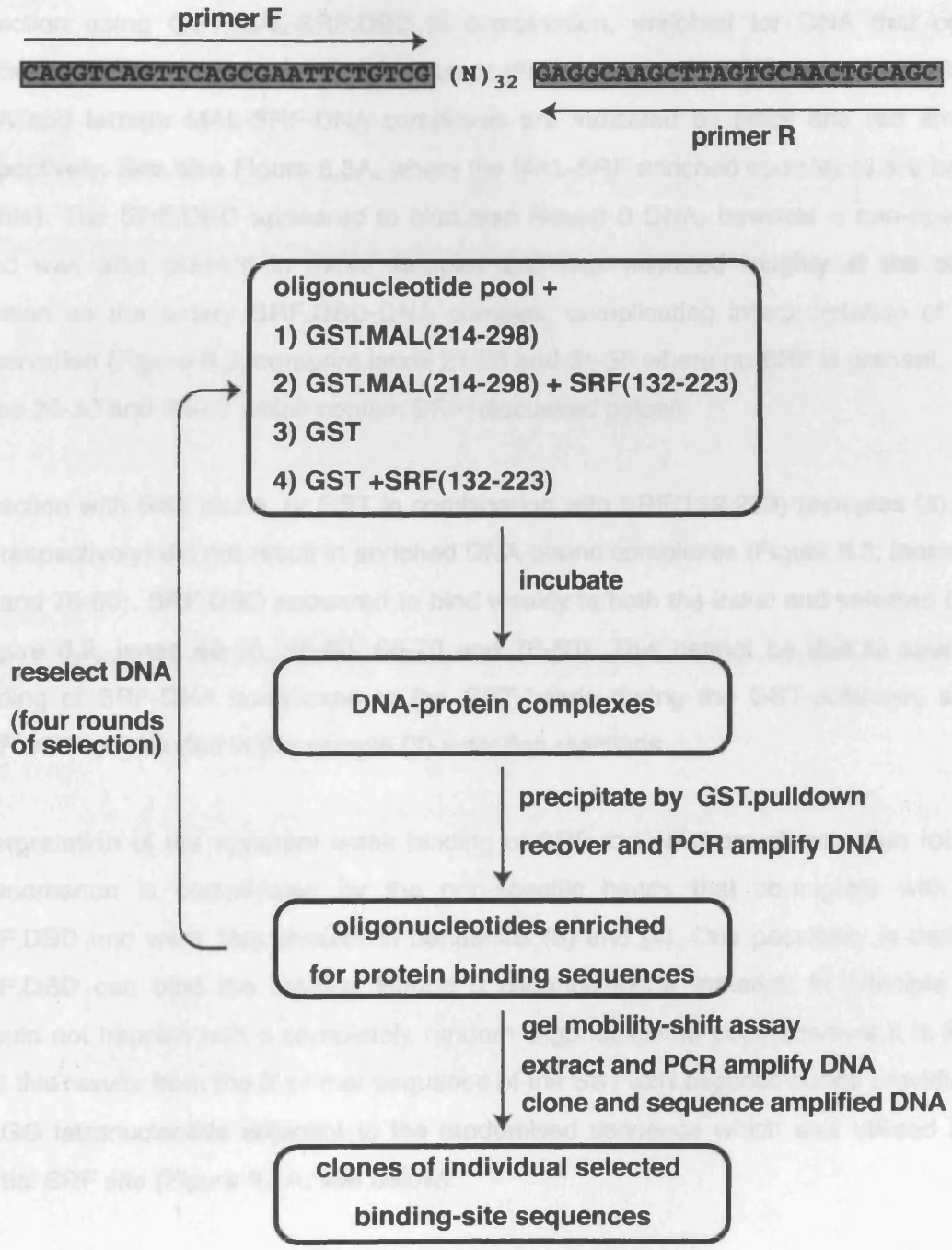


Figure 8.1. Experimental design of DNA binding site selection using with the SS1 32N random oligo pool. The sequence of the SS1 32N oligo which contains 32 random oligonucleotides (designated N) is shown. Binding site selection reactions contained GST.MAL(214-298) with or without recombinant SRF(132-223) (samples 1 and 2) or empty GST protein with or without recombinant SRF(132-223) (samples 3 and 4).

Selection using GST.MAL-SRF.DBD in combination, enriched for DNA that could efficiently be bound by MAL-SRF complexes (Figure 8.2, lanes 26-30; the binary SRF-DNA and ternary MAL-SRF-DNA complexes are indicated by black and red arrows respectively. See also Figure 8.3A, where the MAL-SRF enriched complexes are better visible). The SRF.DBD appeared to bind also Round 0 DNA, however a non-specific band was also present in these samples and was migrated roughly at the same position as the binary SRF.DBD-DNA complex, complicating interpretetation of this observation (Figure 8.2, compare lanes 21-25 and 31-35 where no SRF is present, with lanes 26-30 and 36-40 which contain SRF; discussed below).

Selection with GST alone, or GST in combination with SRF(132-223) (samples (3) and (4) respectively) did not result in enriched DNA-bound complexes (Figure 8.2, lanes 51-55 and 76-80). SRF.DBD appeared to bind weakly to both the initial and selected DNA (Figure 8.2, lanes 46-50, 56-60, 66-70 and 76-80). This cannot be due to spurious binding of SRF-DNA complexes to the GST-beads during the GST-pulldown, since SRF was not included in the sample (3) selection reactions.

Interpretation of the apparent weak binding of SRF to DNA from all selection rounds phenomenon is complicated by the non-specific bands that co-migrate with the SRF.DBD and were also present in bandshifts (3) and (4). One possibility is that the SRF.DBD can bind the starting Round 0 oligonucleotide material. In principle this should not happen with a completely random oligonucleotide pool, however it is likely that this results from the 3' primer sequence of the SS1 32N oligonucleotide providing a GAGG tetranucleotide adjacent to the randomised sequence which was utilised as a partial SRF site (Figure 8.1A; see below).

8.2.1 Analysis of the MAL-SRF selected binding sites from the SS1 32N oligonucleotide pool

To analyse MAL-SRF specific DNA binding sites the DNA present in MAL-SRF complexes formed with Round 4 DNA was recovered, cloned and sequenced (Figure 8.3A, the band is indicated by an arrow).

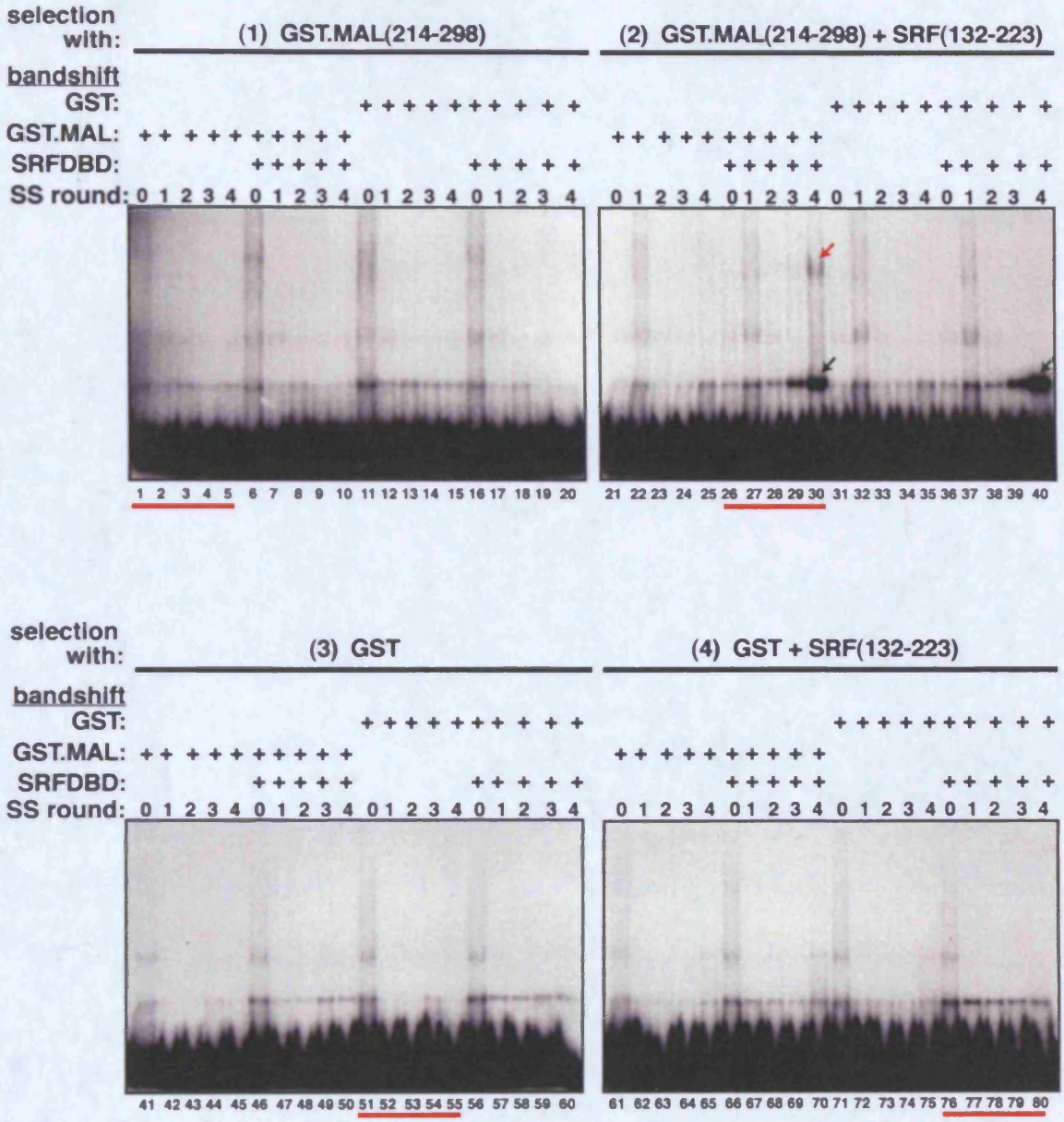


Figure 8.2. Gel mobility-shift assays of the selected SS1 32N DNA from four selection rounds. The selecting proteins for each enriched DNA sample are indicated over each gel. Each gel mobility-shift assay included reactions with the starting DNA (SS round 0), and enriched DNA samples (SS rounds 1-4), assayed for complex formation with GST-MAL(214-298) or empty GST, with or without recombinant SRF (residues 132-223). The reactions assaying selected DNA with the selecting proteins, where specifically enriched complexes might arise are indicated by red lines. The specific complexes of GST.MAL(214-298) and SRF(132-223) in gel (2) are indicated by arrows: red arrow GST.MAL-SRF-DNA complex, black arrow SRF-DNA complex. Note that non-specific bands are seen in all gels at the same position as SRF(132-223), which may represent degraded probe DNA or contamination of the DNA samples with smaller fragments during the DNA purification step.

Of the 182 cloned oligonucleotides, 142 represented unique sequences, with the remaining 40 sequences representing oligonucleotides recovered more than once (results summarised in Figure 8.3B). The identification of multiple identical sequences shows that after four rounds of selection the complexity of the original randomised oligonucleotide pool has been significantly decreased, indicating that the experiment had effectively been taken to completion.

Only 8 of the selected sequences matched the $CC(A/T)_2A(A/T)_3GG$ CArG consensus found in natural SRF sites and of these only one matched the more stringent CC(A/T)TATA(A/T)GG high affinity consensus derived in the original SRF binding site selection study (Figure 8.4; (Pollock, R. *et al.*, 1990) and also Section 1.2.2.1 of the Introduction). The majority of selected sequences contained sites with one or more mismatches in the CArG consensus (Figure 8.4; summarised in Figure 8.3). For discussion purposes hereafter sites that match the $CC(A/T)_2A(A/T)_3GG$ SRF binding consensus will be classed as CArG sequences, while those that contain mismatches in the 10bp SRF-binding sequence will be referred to as CArG-like.

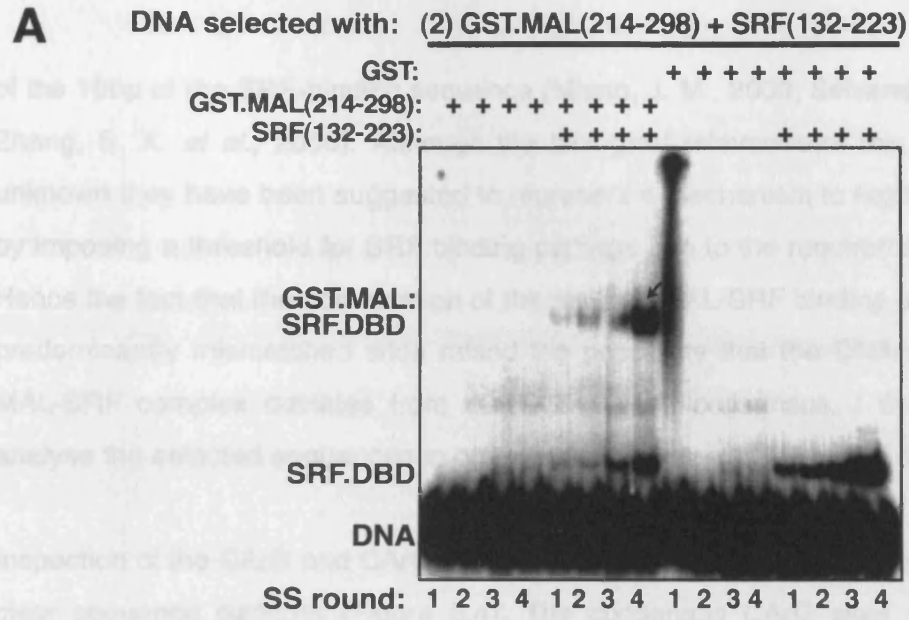
50 of the selected oligonucleotides contained sites overlapping the 3' GAGG primer sequence, and half of these contained further mismatches within the CArG-like site (Figure 8.8; see Section 8.2.1.2). 21 of the selected oligonucleotides contained two 10bp sequences resembling CArG sites and were thus classed separately (Figure 8.9). Five oligonucleotides did not contain a sequence resembling a CArG box.

8.2.1.1 Analysis of the MAL-SRF selected CArG-like sites located completely in the random oligonucleotide sequences

The sequential rounds of oligonucleotide recovery and amplification are set to select high affinity sites. Thus the identification of predominantly mismatched CArG like sequences was surprising, since even single mismatches reduce the affinity of SRF binding by tenfold and CArG sites containing two or more mismatches are predicted to have negligible affinity for SRF (Leung, S. *et al.*, 1989; Wynne, J. *et al.*, 1992).

Nevertheless high affinity binding does not always correlate with functional significance. Many CArG boxes found in natural promoters contain mismatches in one

A



B

Round 4 sites selected with GST.MAL(214-298) + SRF(132-223)	
unique sequences	142
1 CArG-like sequence	117
CArG (consensus)	8
CArG-like (1 mismatch)	35
CArG-like (>1 mismatch)	24
CArG with GAGG primer overlap	50
>1 CArG-like sequence	21
without primer overlap	1
with GAGG primer overlap	16
with other primer overlap	4
no obvious CArG	5

CAGGTCAGTTCAGCGAATTCTGTCG $(N)_{32}$ GAGGCAAGCTTAGTGCAACTGCAGC

Figure 8.3. [GST.MAL+SRF.DBD] selected DNA sites from the SS1 32N random oligo pool. (A) Gel mobility-shift assay of the DNA selected with GST.MAL(214-298) and SRF(132-223). Bandshift reactions contained GST-MAL(214-298) or empty GST, with or without recombinant SRF(132-223), and DNA from the indicated selection rounds. The band corresponding to the GST.MAL-SRF.DBD complex with Round 4 DNA (indicated by arrow) was excised and DNA was amplified and cloned for sequence analysis. (B) Summary of the binding site selection results after sequence analysis. CArG (consensus) sequences match the $CC(A/T)_3A(A/T)_2GG$ consensus. CArG-like sequences contain one or more non-consensus change. CArGs with primer overlap utilise the the 3' primer sequence GAGG. The SS1 32N oligo sequence is shown below with the 3' GAGG primer sequence underlined. Oligonucleotides containing more than one CArG-like sequences are listed separately.

of the 10bp of the SRF-binding sequence (Miano, J. M., 2003; Selvaraj, A. *et al.*, 2004; Zhang, S. X. *et al.*, 2005). Although the biological relevance of this phenomenon is unknown they have been suggested to represent a mechanism to regulate SRF activity by imposing a threshold for SRF binding perhaps due to the requirement of a cofactor. Hence the fact that the combination of the minimal MAL-SRF binding domains selected predominantly mismatched sites raised the possibility that the DNA specificity of the MAL-SRF complex deviates from the $CC(A/T)_6GG$ consensus. I thus proceeded to analyse the selected sequences in order to define potential MAL-SRF specific patterns.

Inspection of the CArG and CArG-like sites with one or more mismatches did not reveal clear sequence patterns (Figure 8.4). The consensus CArG sites were too few to attempt sequence analysis, whereas the CArG-like sites with more than one mismatch could not be divided into groups since many different combinations of mismatch position and identity were observed.

Alignment of the single mismatched sites at the CArG box so that non-consensus bases were contained on the right CArG halfsite, revealed no clear grouping of the mismatched sites (Figure 8.4, middle group), since they contained G/C changes in the central A/T tract and also mismatches in the flanking C-G residues with not obvious base or position bias.

Similarly grouping of the CArG boxes containing one or no mismatches according to their conserved halfsite sequence (e.g. TTAGG versus TATGG) did not identify any bias in the position or base identity of the non-consensus changes, or patterns in the sequences flanking the 10bp core site (Figure 8.5).

The original SRF binding site selection which derived the CArG box consensus with the central invariant A/T tract flanked by C-G residues, had also revealed that the sequence specificity of the prefferred SRF binding sites extended further from the 10bp core to the flanking sequences on either side (Pollock, R. *et al.*, 1990). This study established that SRF selects asymmetric sites with a well-conserved ATGC tetranucleotide directly adjacent to the 5' side of the CArG box and a less conserved sequence on the 3' side and determined the SRF binding consensus halfsite as 5'-ATG(not G)CC(A/T)TA-3'.

CArG box consensus: $CC(A/T)_2A(A/T)_3GG$

CArG consensus

```
92-          gaattctgtcgGATTATAACGGGCCATATAAGGTTTTCTTTGGgaggcaagctt
23-        aagcttgcctcGGAAGGGCACACAAGCCATTTATGGTAACCAAcgacagaattc
13-              gaattctgtcgTGGGTACCATTTATGGTATCCAATCACTGCCTgaggcaagctt
86-               aagcttgcctcCCACACCATTTAAGGCTGGCAGTGTATTCCAGcgacagaattc
54-              gaattctgtcgGGAACCCCATATTTGGTCTTCGCGGGGGCTCCcgacagaattc
78-            gaattctgtcgATGCTTGTCCAATTTTGGGCTTATGAGTACGCgaggcaagctt
47-         gaattctgtcgGGGGTAGAATCTTTCCAAATTTGGTGGCCATCgaggcaagctt
21-       gaattctgtcgCCGGCAGACGGAATATACCTTTTTTGGTCATCgaggcaagctt
```

CArG-like with one non-consensus change

```
13b-           gaattctgtcgGAGCTAATGGCTTATAAGGTCTCGATGTCGGGgaggcaagctt
13c-                gaattctgtcgGGGGGCTTAAAAGGTTTTCGGTAGGATACTTAgaggcaagctt
27-                 aagcttgcctcAAGGGCATAAATGGCAACCTAGTTCACGGGGTcgacagaattc
12-      aagcttgcctcTCAGCCACACCGCCCACTGTCATATAAGGCTGCgacagaattc
5-         gaattctgtcgCACGGGTGCCTGCCTGTCTTTATAGGGAGGTgaggcaagctt
18-                  aagcttgcctcGGGACATATAAGGGCCCGACGACGAAAAACGCCgacagaattc
36-                     gaattctgtcgACATATTTGGTGTATACGCGGGGGTGGGCCCCgaggcaagctt
31-               aagcttgcctcCAGACAACTATATAAGGGTACGGGACTCGCCAcgacagaattc
28-             gaattctgtcgACAGCATCAGCTATATAAGGTGTTGTCGCCCGgaggcaagctt
59-           gaattctgtcgGCCGTTTCATAACTAAATATGGAGAGGTTCTTgaggcaagctt
35-             aagcttgcctcGAGATAATACTATAAAAGGCTCAAAAAGACTTcgacagaattc
84-          gaattctgtcgAGAGTGGGCGGGGCTATAAATGGTTATATTGCgaggcaagctt
33-       gaattctgtcgTTCGGATGGCGAGGCTTGTCTTAATTTGGTCCgaggcaagctt
19b-              gaattctgtcgTCTGTGGCCGTTTAAGGCGCTTATTAGCTTAGgaggcaagctt
17-         gaattctgtcgGTGAATCGTTTGTATGCCGTAAAAGGCATGGGgaggcaagctt
23c-         gaattctgtcgGTATGTTGTTTTCATCCGAAAATGGTAACTAAgaggcaagctt
42-            aagcttgcctcCTGTCAGCTACCCTTTATGGTTATCAGTATGTcgacagaattc
23b-      aagcttgcctcTACGACCAGCTGAGACCGACCCTAAAAGGCCCcgacagaattc
19-               aagcttgcctcACGTACCCTAAAAGGCATCCATCAGCACCAAcgacagaattc
11-              aagcttgcctcCTACCCCCACATAAGGTAAAAACGTCTTATAAcgacagaattc
66-                  gaattctgtcgCCACATAAGGTTCGATTTGATGCTACTCAGTgaggcaagctt
4-     aagcttgcctcATAATCACACCAATTACCAGCACCACTTATGGcgacagaattc
21b-      aagcttgcctcAAAGCACTCCACGACATACCTCTTATGGATACcgacagaattc
49-               aagcttgcctcAATCACTTCCTCAAAAGGAAATGCTGCTcgacagaattc
10-           aagcttgcctcCTGAGCTTCAATCACCTTGTAAGGCATTCCAAcgacagaattc
83-                aagcttgcctcGACTAGCCTTGTAAGGCACTAGCAATAGACCAcgacagaattc
21c-               gaattctgtcgTATGGCCTAGTTAGGGTAGATTGCCGTTTAGGgaggcaagctt
7-          gaattctgtcgTGGGATATTCACCTAGACCATGTATGGTTCCTgaggcaagctt
34-                  gaattctgtcgCCTTGTAAGGTGGTATCCGAGGTGCCGTCGCGgaggcaagctt
8-              gaattctgtcgCATCAGGTCCATGTTTGGTTTGATGTGTGAGGgaggcaagctt
56-                 gaattctgtcgTGTGGCCTAGTTAGGGTAGATTGCCGTTTAGGgaggcaagctt
30-           gaattctgtcgTGAGTCGGAGTACCTTCTAAGGGTTGGGACTgaggcaagctt
30b-          gaattctgtcgAATGGCGCGATGCACCCTTCTAAGGGCGGAACgaggcaagctt
2-            aagcttgcctcACCCGGTATACCTTACCTTCCAATGGACACCAAcgacagaattc
27-              gaattctgtcgGGGGGTGTGTCCTTCAAAGGGTTGGGGAGGCTgaggcaagctt
```

CArG-like with more than one non-consensus change

```
79-          gaattctgtcgTCGGTATAGATTTTTGACTATATATGTCGCCTgaggcaagctt
20-        gaattctgtcgGAAGTGGATCAAAACTGGACTCATATGGGCGTgaggcaagctt
90-              aagcttgcctcACCACGGCGACCTATATGGACACTAATCTCGCcgacagaattc
95-                   gaattctgtcgAGACTGTATTGGTCTTGTTAATCTCTCGCTTCgaggcaagctt
88.1-  aagcttgcctcAGGAGGTCCCGAGACACTATAGCTACAAAAGGcgacagaattc
34c-          gaattctgtcgTCGCAGCATTGTGCCTATGTTTGGGTGTTTGTgaggcaagctt
42b-          gaattctgtcgGGGGGCTCGTAAGTCTAATCATGGTGTGCGGTgaggcaagctt
45-             gaattctgtcgATGTACGCCTTTATACGGGTGTAGTCGGGGCgaggcaagctt
38-          gaattctgtcgGGATGGATGTAATGACTGTATAAGGTCACTGTgaggcaagctt
1-           gaattctgtcgCTTATGGTGATTCTTCTGTTTATGGTTAGAAGgaggcaagctt
89-                gaattctgtcgGGGAGGGACGATTAAGGTTATAGTAGTCAGTGgaggcaagctt
11c-            gaattctgtcgTGCAAGGCACCGTAGTTGGTAGTCGTGTTGGGgaggcaagctt
55-          gaattctgtcgAGCGATATGTGTCGGGCTATGTAGGGGCCTCTgaggcaagctt
35b-                  gaattctgtcgTCATAGATGGTCACGGACTGCTTTGGGTATGGgaggcaagctt
82.2-               aagcttgcctcGAGACCCTACAAGGAAATAAAGTAGGACACAGcgacagaattc
36b-          aagcttgcctcCCTGACAACTTAAACCCTACAAGGTCAAAGAAcgacagaattc
26-             aagcttgcctcACTCAAAGATAACCCTACTTGGCTCCATCAGAcgacagaattc
34b-              aagcttgcctcACAAACACCCAAACATAGGCACAATGCTGCGAcgacagaattc
62-                   gaattctgtcgCCTCATCTGGGGTTTTCTATGTTTAAATAAGgaggcaagctt
51-                gaattctgtcgTATTGTCCACTTTCGGGCCGAGTTTGTAGCGTgaggcaagctt
10-b                  gaattctgtcgGTGCATATAGGGATGCATCGGCGTGGTTTCGgaggcaagctt
82.1-            gaattctgtcgTATGCCTTAAGTACGAATTTGGGCATGTTTCAgaggcaagctt
14-                   gaattctgtcgATTGCCCTTTAGGGTGATATTTACGTCATTCTgaggcaagctt
11b-               gaattctgtcgGAGACGACCGTGTAGGGGTGCGGGGCTCACTGgaggcaagctt
```

Figure 8.4. [GST.MAL(214-298) + SRF(132-223)] selected binding sites in which the CArG box is within the random SS1 32N sequence. The sequences of 65 binding sites containing one CArG-like site completely within the random sequence and derived by four rounds of selection are shown below the consensus CArG site. Sequences were aligned at the CArG box, so that CArG-like sequences with one mismatch contained non-consensus bases in the same half-site. CArG-like sequences with more than one mismatches cannot be divided in clear subclasses. Blue: bases that match the CArG consensus; red: non-consensus bases; primer sequences are shown in lowercase letters.

CArG halfsite

TAAGG

```
92-          gaattctgtcgGATTATAACGGGCCATATAAGGTTTTCTTTGGgaggcaagctt
12-    aagcttgcctcTCAGCCACACCGCCCACTGTCATATAAGGCTGcgacagaattc
18-                  aagcttgcctcGGGACATATAAGGGCCCGACGACGAAAAACGCcgacagaattc
31-               aagcttgcctcCAGACAACTATATAAGGGTACGGGACTCGCCAcgacagaattc
28-            gaattctgtcgACAGCATCAGCTATATAAGGTGTTGTCGCCCGgaggcaagctt
11-               aagcttgcctcCTACCCCCACATAAGGTAAAAACGTCTTATAAcgacagaattc
66-                     gaattctgtcgCCACATAAGGTTCGATTTGATGCTACTCAGTgaggcaagctt
86-               aagcttgcctcCCACACCATTTAAGGCTGGCAGTGTATTCCAGcgacagaattc
13b-           gaattctgtcgGAGCTAATGGCTTATAAGGTCTCGATGTCGGGgaggcaagctt
10-         aagcttgcctcCTGAGCTTCAATCACCTTGTAAGGCATTCCAAcgacagaattc
83-               aagcttgcctcGACTAGCCTTGTAAGGCACTAGCAATAGACCAcgacagaattc
30-          gaattctgtcgTGAGTCGGAGTACCTTCTAAGGGTTGGGACTgaggcaagctt
30b-       gaattctgtcgAATGGCGCGATGCACCCTTCTAAGGGCGGAACgaggcaagctt
34-                  gaattctgtcgCCTTGTAAGGTGGTATCCGAGGTGCCGTCGCGgaggcaagctt
19b-              gaattctgtcgTCTGTGGCCGTTTAAGGCGCTTATTAGCTTAGgaggcaagctt
```

TATGG

```
4-       aagcttgcctcATAATCACACCAATTACCAGCACCACTTATGGcgacagaattc
23-           aagcttgcctcGGAAGGGCACACAAGCCATTTATGGTAACCAAcgacagaattc
13-                gaattctgtcgTGGGTACCATTTATGGTATCCAATCACTGCCTgaggcaagctt
7-          gaattctgtcgTGGGATATTCACCTAGACCATGTATGGTTCCTgaggcaagctt
42-               aagcttgcctcCTGTCAGCTACCCTTTATGGTTATCAGTATGTcgacagaattc
21b-      aagcttgcctcAAAGCACTCCACGACATACCTCTTATGGATACcgacagaattc
59-            gaattctgtcgGCCGTTTCATAACTAAATATGGAGAGGTTCTTgaggcaagctt
```

AAAGG

```
49-                aagcttgcctcAATCACTTCCTCAAAAGGAAATGCTGCTcgacagaattc
35-               aagcttgcctcGAGATAATACTATAAAAGGCTCAAAAAGACTTcgacagaattc
13c-                 gaattctgtcgGGGGGCTTAAAAGGTTTTCGGTAGGATACTTAgaggcaagctt
17-          gaattctgtcgGTGAATCGTTTGTATGCCGTAAAAGGCATGGGgaggcaagctt
23b-       aagcttgcctcTACGACCAGCTGAGACCGACCCTAAAAGGCCCcgacagaattc
19-                aagcttgcctcACGTACCCTAAAAGGCATCCATCAGCACCAAcgacagaattc
27-              gaattctgtcgGGGGGTGTGTCCTTCAAAGGGTTGGGGAGGCTgaggcaagctt
```

TTTGG

```
54-                   gaattctgtcgGGAACCCCATATTTGGTCTTCGCGGGGGCTCCCgacagaattc
36-                        gaattctgtcgACATATTTGGTGTATACGCGGGGGTGGGCCCCgaggcaagctt
8-                  gaattctgtcgCATCAGGTCCATGTTTGGTTTGATGTGTGAGGgaggcaagctt
21-           gaattctgtcgCGGCAGACGGAATATACCTTTTTTGGTCATCgaggcaagctt
78-                 gaattctgtcgATGCTTGTCCAATTTTGGGCTTATGAGTACGCgaggcaagctt
47-            gaattctgtcgGGGGTAGAATCTTTCCAAATTTGGTGGCCATCgaggcaagctt
33-         gaattctgtcgTTCGGATGGCGAGGCTTGTCTTAATTTGGTCCgaggcaagctt
```

AATGG

```
2-           aagcttgcctcACCCGGTATACCTTACCTTCAATGGACACCAAcgacagaattc
27-                  aagcttgcctcAAGGGCATAAATGGCAACCTAGTTCACGGGGTcgacagaattc
84-             gaattctgtcgAGAGTGGGCGGGGCTATAAATGGTTATATTGCgaggcaagctt
23c-           gaattctgtcgGTATGTTGTTTTCATCCGAAAATGGTAACTAAgaggcaagctt
```

TTAGG

```
21c-                  gaattctgtcgTATGGCCTAGTTAGGGTAGATTGCCGTTTAGGgaggcaagctt
56-                   gaattctgtcgTGTGGCCTAGTTAGGGTAGATTGCCGTTTAGGgaggcaagctt
```

ATAGG

```
5-          gaattctgtcgCACGGGTGCCTGCCTGTCTTTATAGGGAGGTgaggcaagctt
```

Figure 8.5. The CArG-like sequences selected by [GST.MAL(214-298)+SRF(132-223)] do not belong to distinct subclasses depending on CArG halfsite identity. The sequences of 43 binding sites containing one CArG-like site with no or a single mismatch were grouped according to CArG half-site identity. Sequences were aligned at the CArG box so that non-consensus bases were on the right halfsite. Light blue: bases that follow the perfect CArG consensus sequence; red: non-consensus bases; primer sequences are shown in lowercase letters.

Despite no such obvious 5' flanking sequence specificity in the MAL-SRF selected sites I attempted to align them according to the best match to the flanking SRF halfsite consensus (ATG-not G) in order to discover the degree of variability between the originally derived SRF consensus site and that of the GST.MAL(214-298)-SRF(132-223) selected sites. Sequences were aligned so that the maximum matches would be on the 5' side of the CArG box (Figure 8.6). Oligonucleotides where the central 10 and flanking 4 bp on either side impinged on primer sequences were ignored in order to derive an extended consensus binding sequence.

Although this consensus broadly agreed with the one derived in the original SRF binding site selection, significant differences were apparent. The MAL-SRF derived consensus contained mismatches in all 10 positions of the CArG box, with the highest number of changes at positions ±1 and ±3 (Figure 8.6), in contrast to the original SRF consensus where the flanking C-G and central A-T bases are invariant.

Furthermore although in the extended MAL-SRF derived consensus there appears to be a preference for ATG at positions -9 to -7 and an overall asymmetry in the flanking sequence selection, when the data are combined into the halfsite consensus positions ±9 to ± 6 do not display any strong specificity bias for any sequence pattern (Figure 8.6). Thus it would appear that under the conditions used the minimal MAL-SRF complex displays more relaxed specificity in the DNA sequence it recognises. Nevertheless, the possibility that the relaxed specificity observed is a property of the minimal SRF DNA-binding domain cannot be excluded, due to the lack of a positive control dataset. The previous SRF DNA binding site selections were performed with full-length SRF (Pollock, R. *et al.*, 1990), and although SRF(132-223) has been widely used in biochemical and crystallographic analyses its wild-type DNA specificity has not been confirmed.

To investigate whether the mismatched CArG like sequences represent genuine MAL-SRF specific sites, I tested their ability to interact with SRF and form a MAL-SRF complex was tested. Radiolabelled probes were produced from four cloned oligos, one containing a single G mismatch in the A-T core and three containing two mismatches at different position of the CArG-like box (Figure 8.7A). These were used in gel mobility-shift assays with MAL and SRF derivatives, in parallel with a c-fos promoter derived DNA probe, which contains a high affinity SRF site.

extended SRF halfsite consensus: A(T/A)G(notG)CC(A/T)TA
(Pollock and Treisman 1990)

```
21-                    aagcttgcctcGATGACCAAAAAAGGTATATTCCGTCTGCCGGcgacagaattc
54-              aagcttgcctcGGAGCCCCCGCGAAGACCAAATATGGGGTTCCcgacagaattc
78-                aagcttgcctcGCGTACTCATAAGCCCAAAATTGGACAAGCATcgacagaattc
92-                   aagcttgcctcCCAAAGAAAACCTTATATGGCCCGTTATAATCcgacagaattc
23-                      gaattctgtcgTTGGTTACCATAAATGGCTTGTGTGCCCTTCCgaggcaagctt
13-              aagcttgcctcAGGCAGTGATTGGATACCATAAATGGTACCCAcgacagaattc
86-             gaattctgtcgCTGGAATACACTGCCAGCCTTAAATGGTGTGGgaggcaagctt
47-               gaattctgtcgGGGGTAGAATCTTTCCAAATTTGGTGGCCATCgaggcaagctt

10-                      gaattctgtcgTTGGAATGCCTTACAAGGTGATTGAAGCTCAGgaggcaagctt
13b-                    gaattctgtcgGAGCTAATGGCTTATAAGGTCTCGATGTCGGGgaggcaagctt
35-                     aagcttgcctcGAGATAATACTATAAAAGGCTCAAAAAGACTTcgacagaattc
27-                        aagcttgcctcAAGGGCATAAATGGCAACCTAGTTCACGGGGTcgacagaattc
13c-            aagcttgcctcTAAGTATCCTACCGAAAACCTTTTAAGCCCCCcgacagaattc
28-                   AAGCTTGCCTCCGGGCGACAACACCTTATATAGCTGATGCTGTCGACAGAATTC
21c-                     gaattctgtcgTATGGCCTAGTTAGGGTAGATTGCCGTTTAGGgaggcaagctt
49-                 gaattctgtcgCGAGCAGCATTTCCTTTTGAGGAAGTGATTgaggcaagctt
84-                   AAGCTTGCCTCGCAATATAACCATTTATAGCCCCGCCCACTCTCGACAGAATTC
59-                 gaattctgtcgGCCGTTTCATAACTAAATATGGAGAGGTTCTTgaggcaagctt
42-                 gaattctgtcgACATACTGATAACCATAAAGGGTAGCTGACAGgaggcaagctt
8-                     gaattctgtcgCATCAGGTCCATGTTTGGTTTGATGTGTGAGGgaggcaagctt
30-                 gaattctgtcgTGAGTCGGAGTACCTTCTAAGGGTTGGGACTgaggcaagctt
83-               gaattctgtcgTGGTCTATTGCTAGTGCCTTACAAGGCTAGTCgaggcaagctt
31-                     aagcttgcctcCAGACAACTATATAAGGGTACGGGACTCGCCAcgacagaattc
2-                      gaattctgtcgTTGGTGTCCATTGAAGGTAAGGTATACCGGGTgaggcaagctt
19b-                    gaattctgtcgTCTGTGGCCGTTTAAGGCGCTTATTAGCTTAGgaggcaagctt
27-                  gaattctgtcgGGGGGTGTGTCCTTCAAAGGGTTGGGGAGGCTgaggcaagctt
56-                     gaattctgtcgTGTGGCCTAGTTAGGGTAGATTGCCGTTTAGGgaggcaagctt
23c-                    aagcttgcctcTTAGTTACCATTTTCGGATGAAAACAACATACcgacagaattc
19-               gaattctgtcgTTGGTGCTGATGGATGCCTTTTAGGGTACGTgaggcaagctt
30b-                gaattctgtcgAATGGCGCGATGCACCCTTCTAAGGGCGGAACgaggcaagctt
18-            GAATTCTGTCGGCGTTTTTCGTCGTCGGGCCCTTATATGTCCCGAGGCAAGCTT
5-                gaattctgtcgCACGGGTGCCTGCCTGTCTTTATAGGGAGGTgaggcaagctt
21b-            aagcttgcctcAAAGCACTCCACGACATACCTCTTATGGATACcgacagaattc
17-                     aagcttgcctcCCCATGCCTTTTACGGCATACAAACGATTCACcgacagaattc
7-                gaattctgtcgTGGGATATTCACCTAGACCATGTATGGTTCCTgaggcaagctt
11-               gaattctgtcgTTATAAGACGTTTTTACCTTATGTGGGGGTAGgaggcaagctt

14-                         gaattctgtcgATTGCCCTTTAGGGTGATATTTACGTCATTCTgaggcaagctt
82.2-           gaattctgtcgCTGTGTCCTACTTTATTTCCTTGTAGGGTCTCgaggcaagctt
38-               gaattctgtcgGGATGGATGTAATGACTGTATAAGGTCACTGTgaggcaagctt
26-                  aagcttgcctcACTCAAAGATAACCCTACTTGGCTCCATCAGAcgacagaattc
11c-                aagcttgcctcCCCAACACGACTACCAACTACGGTGCCTTGCAcgacagaattc
11b-                   gaattctgtcgGAGACGACCGTGTAGGGGTGCGGGGCTCACTGgaggcaagctt
45-                  gaattctgtcgATGTACGCCTTTATACGGGTGTAGTCGGGGCgaggcaagctt
42b-               gaattctgtcgGGGGGCTCGTAAGTCTAATCATGGTGTGCGGTgaggcaagctt
89-                   gaattctgtcgGGGAGGGACGATTAAGGTTATAGTAGTCAGTGgaggcaagctt
82.1-               gaattctgtcgTATGCCTTAAGTACGAATTTGGGCATGTTTCAgaggcaagctt
36b-                aagcttgcctcCCTGACAACTTAAACCCTACAAGGTCAAAGAAcgacagaattc
1-                GAATTCTGTCGCTTATGGTGATTCTTCTGTTTATGGTTAGAAGGAGGCAAGCTT
55-               gaattctgtcgAGCGATATGTGTCGGGCTATGTAGGGGCCTCTgaggcaagctt
51-                     gaattctgtcgTATTGTCCACTTTCGGGCCGAGTTTGTAGCGTgaggcaagctt
79-               gaattctgtcgTCGGTATAGATTTTTGACTATATATGTCGCCTgaggcaagctt
90-                     aagcttgcctcACCACGGCGACCTATATGGACACTAATCTCGCcgacagaattc
34c-               GAATTCTGTCGTCGCAGCATTGTGCCTATGTTTGGGTGTTTGTGAGGCAAGCTT
34b-                    aagcttgcctcACAAACACCCAAACATAGGCACAATGCTGCGAcgacagaattc
20-            gaattctgtcgGAAGTGGATCAAAACTGGACTCATATGGGCGTgaggcaagctt
```

CArG position	-9	-8	-7	-6	-5	-4	-3	-2	-1	+1	+2	+3	+4	+5	+6	+7	+8	+9
A	29	17	10	23	4	1	22	14	25	12	39	22	2	-	6	11	16	8
G	12	7	21	17	3	1	5	-	7	2	2	6	52	53	15	11	13	19
C	7	9	3	6	47	45	4	3	5	5	-	5	-	1	14	14	14	18
T	7	22	21	9	1	8	24	38	18	36	14	22	1	1	20	19	12	10
consensus	A	T/A	G/T	T/G	C	C	T/A	T A	A/T	T	A T	A/T	G	G	T G/C	N	N	C/G

HALFSITE	±9	±8	±7	±6	±5	±4	±3	±2	±1
A	39	29	29	43	5	2	44	28	61
G	30	21	35	31	4	1	10	-	12
C	26	22	14	21	100	97	10	5	7
T	15	38	32	15	1	10	46	77	30
consensus	(T)	N	(C)	(T)	C	C	T/A	T A	A T

Figure 8.6. Derivation of a consensus DNA binding sequence for [MAL(214-298)+SRF(132-223)]. The 55 sequences listed contain the CArG box and four flanking nucleotides on either side completely within the randomised sequence and were used to derive an extended DNA binding consensus sequence. CArG boxes were aligned for maximum match to the extended SRF halfsite consensus [shown at the top; (Pollock and Treisman,1990)] was on the 5' side of the CArG box. Blue: CArG consensus bases; red: non-consensus bases; lowercase letters: primer sequences. The derived [MAL(214-298)+SRF(132-223)] consensus is shown at the bottom. Bases separated by forward slash are interchangeable; bases shown one over the other denote preference for the top base; brackets denote least favoured bases.

The probes were first tested for their interaction with different sized SRF derivatives (Figure 8.7B). The efficiency of complex formation between each probe and the different SRF forms is not directly comparable, due to the different source of the proteins: SRF(132-223) was purified from of E. coli, SRF(120-265) was produced in reticulocyte lysates, whereas SRF(fl) was expressed in whole-cell extracts. Nevertheless the ability of each SRF derivative to bind the different probes can be correlated to its high affinity interaction with the c-fos probe (Figure 8.7B, lanes 1, 11 and 16).

All three SRF derivatives tested bound the mismatched sites with lower efficiency compared to their interaction with the c-fos probe (Figure 8.7B), and phosphorimager analysis revealed that irrespective of their length, the different SRF derivatives displayed 5 to 10-fold lower affinities for the mismatched probes (data not shown), thus excluding the possibility of SRF(132-223) recognising non-consensus sites.

Inclusion of MALΔN in the reactions containing the c-fos probe gave rise to substantial amounts of complex with all three SRF forms (Figure 8.7C). In contrast complex formation between MALΔN and either full-length SRF or SRF(120-265) with probes A, B, C and D was greatly reduced (Figure 8.7C, compare lanes 12-15 with lane 11 and lanes 7-10 with lane 6) and was shown to correspond to 10-20% of the c-fos probe binding activity by phosphorimager analysis (data not shown). Surprisingly MALΔN was more competent in binding the mismatched probes when complexed with SRF(132-223) (Figure 8.7C, lanes 2-4). In this case, the complex formation efficiency was in the range of 20-40% compared to the wild-type complex formed with the c-fos probe.

Thus the altered DNA specificity observed in the MAL-SRF selected sequences is not a genuine property of the MAL-SRF complex, since in that case MAL should have recruited all SRF forms into high affinity ternary complexes. In contrast this result appears to imply that MAL has more relaxed DNA binding specificity when bound to the minimal SRF DNA-binding domain compared to when it interacts with longer forms of SRF. SRF(132-223) encompasses only the MADS and SAM domains required for DNA binding and dimerisation. Although the sequences preceding the N-terminal extension are not thought to affect DNA binding it is conceivable that their absence renders the MAL-SRF more flexible conformationally increasing the ability of SRF to interact with non-consensus sites.

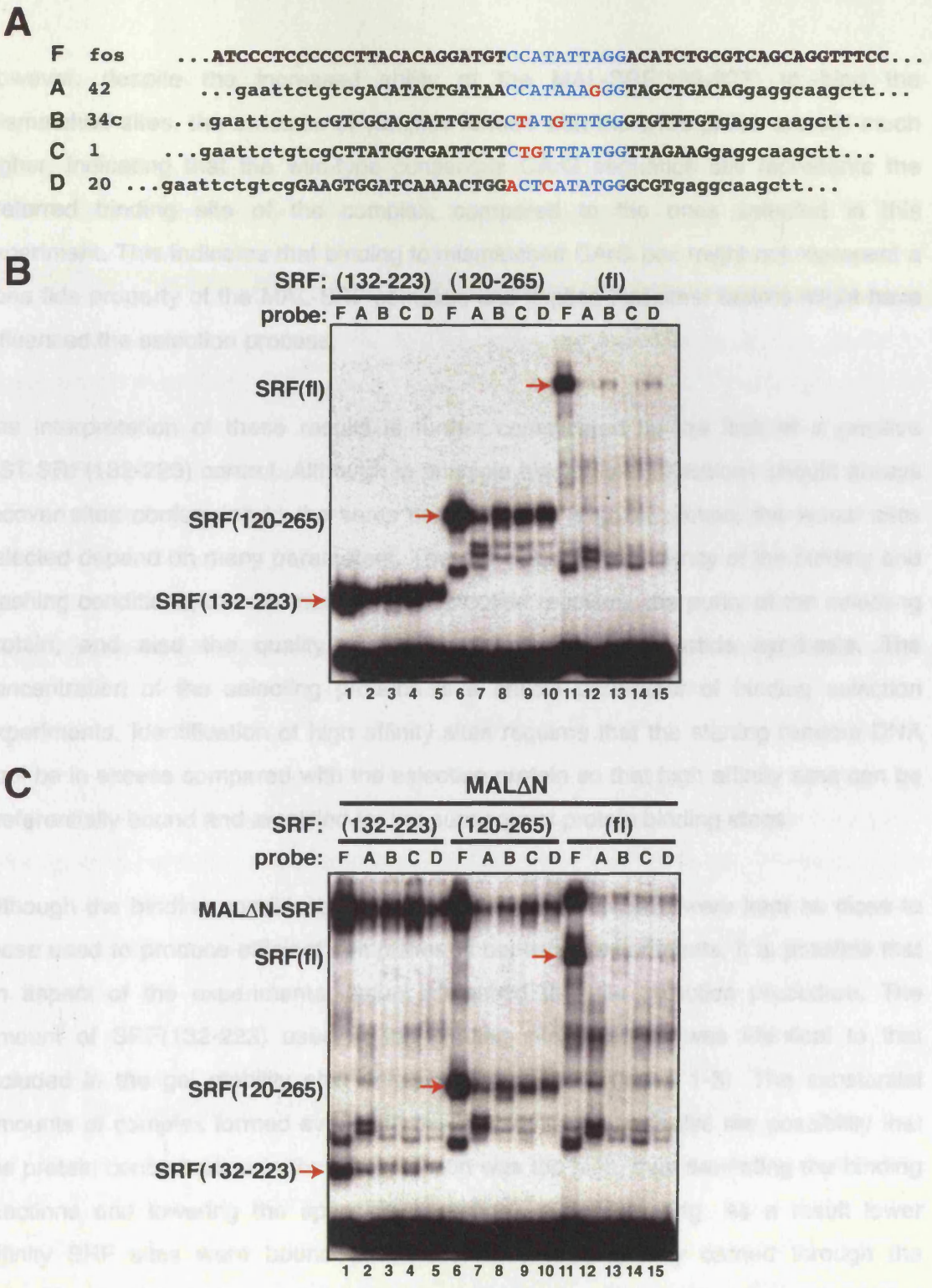


Figure 8.7. Characterisation of [MAL(214-298)+SRF(132-223)] non-consensus CArG selected sites for SRF binding and MAL-SRF complex formation. (A) Sequences of the cloned sites that were used as probes in bandshift assays. Probe F is the wild-type c-fos CArG sequence. Blue: bases that follow the CArG consensus sequence; red: non-consensus bases; primer sequences are shown in lowercase letters. Only part of the sequences is shown. (B) Gel mobility-shift assays contained purified SRF(132-223) or in vitro translated SRF(120-265) or whole-cell extract expressing full-length SRF, combined with the indicated probes. The red arrows indicate the SRF-DNA complexes formed by the three different size SRF derivatives. (C) Gel-mobility shift assays were as in B, with reactions also including MALΔN whole-cell extract.

However, despite the increased ability of the MAL-SRF(132-223) to bind the mismatched sites, the amounts of complex formed with the c-fos probe are still much higher, indicating that the wild-type consensus CArG sequence still represents the preferred binding site of the complex, compared to the ones selected in this experiment. This indicates that binding to mismatched CArG box might not represent a bona fide property of the MAL-SRF complex, and implies that other factors might have influenced the selection process.

The interpretation of these results is further complicated by the lack of a positive GST.SRF(132-223) control. Although in principle binding site selections should always recover sites conforming to the same consensus for a given protein, the actual sites selected depend on many parameters. These include the stringency of the binding and washing conditions, the manner of oligonucleotide recovery, the purity of the selecting protein, and also the quality of the randomised oligonucleotide synthesis. The concentration of the selecting proteins is a critical parameter of binding selection experiments. Identification of high affinity sites requires that the starting random DNA pool be in excess compared with the selecting protein so that high affinity sites can be preferentially bound and amplified for the subsequent protein binding steps.

Although the binding conditions of the MAL-SRF site selection were kept as close to those used to produce efficient complexes in bandshift experiments, it is possible that an aspect of the experimental setup influenced the site selection procedure. The amount of SRF(132-223) used in the binding site selection was identical to that included in the gel mobility shift assay in Figure 8.7B (lanes 1-5). The substantial amounts of complex formed even with the low affinity probes raise the possibility that the protein concentration in the site selection was too high, thus saturating the binding reactions and lowering the specificity threshold of SRF binding. As a result lower affinity SRF sites were bound and these were subsequently carried through the selection rounds.

8.2.1.2 The CArG-like sites utilising the primer GAGG sequences

In light of the results presented in the previous section it is perhaps not surprising that the combination of purified MAL(214-298) and SRF(132-223) recognised the GAGG primer sequence as a CArG-like halfsite (Figure 8.8). It is noteworthy that the utilisation

of a GAGG primer sequence as a half-site had been previously observed in binding site selection studies with MCM1 and MCM1 hybrids containing the SRF SAM domain (Wynne, J. *et al.*, 1992), but not with wild-type SRF (Pollock, R. *et al.*, 1990). The MCM1 consensus is more relaxed within the A-T rich tract compared to that of SRF due to the unconstrained conformation of residue R18, which allows the presence of G-C base pairs (see Section 1.2.3.3.1 of the Introduction). Thus in selecting the GAGG-halfsite sequences, the minimal MAL-SRF complex displays a DNA specificity closer to that of MCM1 than SRF. Although such a marked change in the DNA specificity of SRF when complexed with MAL is an intriguing idea, the experimental weaknesses of this binding site selection discussed in the previous sections indicate that recognition of the GAGG halfsite resulted from inordinately high amounts of SRF in the binding reactions.

The fortuitous utilisation of the GAGG primer tetranucleotide resulted in constraining the CArG-like sites at one end of the random 32-nucleotide sequence. Since MAL itself contacts DNA between positions ±13 and ±22, the -13 and -22 regions of the selected oligos were inspected for sequence specific patterns (Figure 8.8), however no sequence specificity was apparent. The inability to identify sequence specific motifs on the 5' region of the GAGG halfsite sequences does not unequivocally preclude MAL-DNA specific contacts. It is possible for instance that the result was affected by the relaxed specificity of the sites selected by the minimal MAL-SRF complex or the fact that the MAL-DNA contacts were constrained to the primer sequences on the 3' side of the CArG box.

8.2.1.3 The oligonucleotides containing two CArG-like sites

The double CArG-like selected sites represent 15% of the total unique sequences identified. Although the majority of these included more than one mismatches in the core 10bp, most of the sequences recovered multiple times belonged in this category, implying that possibly both sites could be recognised by the MAL-SRF complexes (Figure 8.9).

Multiple functional CArG sites are often found in SRF controlled promoters, however these are well spaced from each other, and some times are located hundreds of base pairs apart (Miano, J. M. *et al.*, 2004; Selvaraj, A. *et al.*, 2004; Sun, Q. *et al.*, 2006). The footprint of the minimal SRF DNA-binding domain covers the area to nucleotides

general CArG box consensus: CC(A/T)$_6$GG

perfect CArG consensus: CC(A/T)TATA(A/T)GG

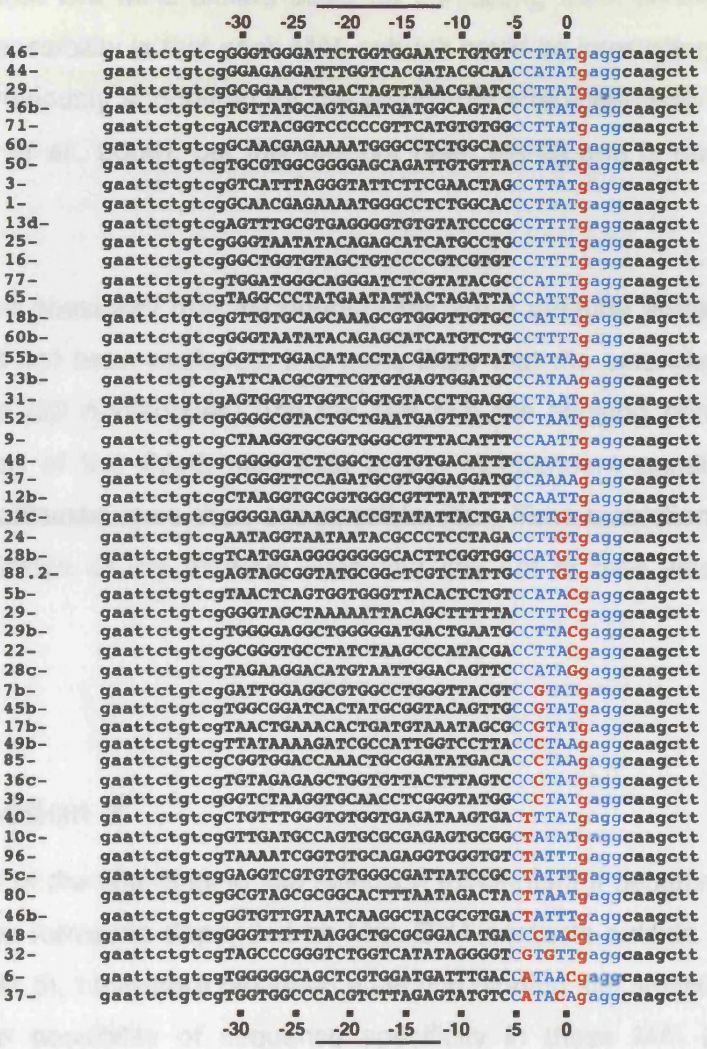


Figure 8.8.Selected binding sites for GST.MAL(214-298) and SRF(132-223) in which the CArG box overlaps the 3' primer sequence. The sequences of 50 oligonucleotides containing one CArG-like site that utilises the 3' GAGG primer sequence are shown below the consensus CArG site. Sites were recovered after four rounds of selection and are aligned on the fixed GAGG sequence. Blue: bases that follow the perfect CArG consensus sequence; red: non-consensus bases; primer sequences are shown in lowercase letters. Nucleotides in the random sequence are numbered from the centre of the CArG box. The solid lines represent the area of MAL-DNA contacts as defined by DNase I footprinting and nested probe analysis.

±14 from the centre of the SRE (Chapter 5; Figures 5.2 and 5.3). Hence it is formally possible for some of the selected sites that two SRF dimers were bound on the same DNA molecule at the same time (Figure 8.9). The DNA-binding requirements of MAL make it unlikely that two MAL dimers could be contacting them simultaneously (see Chapter 5). One possibility is that each MAL subunit could be interacting with a distinct SRF dimer, as previously suggested for the Myocardin interaction with multiple CArG boxes (Wang, Z. *et al.*, 2004), but this has not been tested and is not supported by concrete data.

Thus, although the possibility that the double sites are a genuine feature of the MAL-SRF complex has not been excluded, it is more likely that the extended length of the random sequence (32 nucleotides) and the fact that the flanking primer sequences were used as part of the CArG site, lead to the independent parallel selection of sequences that contained more than one possible sites. Time restrictions did not allow testing of the affinities of the different sites and this set of data was not analysed further.

8.3 Site selection 2

During the course of the first binding site selection experiment it became apparent that MAL-SRF complex formation depended on MAL-DNA contacts outside the core CArG sequence (Chapter 5). I therefore designed a second binding site selection experiment to investigate the possibility of sequence specificity in these MAL-DNA contacts. DNase I footprinting and nested probe analysis located the DNA regions contacted by MAL to positions ±12 to ±22 from the centre of the CArG box, requiring a long random oligonucleotide sequence to identify the extensive sequence covered by the MAL and SRF complex. To avoid complications by CArG boxes located at varying positions within the randomer and to avoid selecting multiple CArG boxes by chance the oligonucleotide was designed to contain a partially set CArG box, where the flanking C and G bases and a central adenine were invariant, and the rest of the core was constrained for A or T (Figure 8.10). This partially set CArG box was flanked by 21 random nucleotides, thus constraining SRF binding to the middle of the sequence and including long enough randomers on either side for MAL binding. Care was taken to avoid primer sequences that could be recognised as halfsites.

Figure 8.9. [GST.MAL+SRF.DBD] selected oligonucleotides after four rounds of selection containing more than one potential CArG site. Potential CArG sequences are shown in blue with mismatches in red. The red numbers next to some sequences indicate the number of times these were recovered during the sequencing process.

Due to the complications encountered using purified short forms of MAL and SRF and isolating oligonucleotides by GST-pulldown (see previous section) I opted to use whole cell extracts expressing transfected full-length proteins and recover sequences by immunoprecipitation (Figure 8.11). Four selection conditions were set: 1) mock-transfected empty Flag vector with untagged SRF, 2) Flag-tagged MALΔN with untagged SRF, 3) Flag-tagged MALΔNΔB1 with untagged SRF as a negative control, 4) Flag-tagged SRF as a positive control.

After four rounds of selection recovered DNA samples were assayed alongside the starting DNA material in bandshifts for specific complex formation with the selecting extracts (Figure 8.11, the red line under each bandshift specifies the reactions where enriched complexes might arise).

Selection with empty Flag.vector and SRF did not enrich DNA able to generate specific complexes (Figure 8.11, lanes 1-5). Although both SRF and MAL-SRF complexes could bind DNA from all selection rounds due to the presence of the set CArG box, the DNA in the complexes was not enriched.

In contrast sequential rounds of selection with Flag-tagged MALΔN and SRF enriched MAL-SRF specific sites as can be seen by the progressive increase in the amounts of SRF-DNA and MAL-SRF-DNA complexes (Figure 8.11, lanes 26-30). Similar enrichment was seen in the Flag.SRF recovered DNA (Figure 8.11, lanes 61-65). As expected this was not observed with the DNA selected in the presence of the MALΔNΔB1 derivative that cannot bind SRF (Figure 8.11, lanes 51-55).

8.3.1 Analysis of the MAL-SRF and SRF selected sequences from the SS2 (conSRE) oligonucleotide pool

To analyse MAL-SRF specific DNA sites, DNA from round four bound by the MAL-SRF complex, was recovered, cloned and sequenced (Figure 8.11, lanes 30 band indicated by red arrow). A complex formed by SRF and round four specific DNA was also processed to provide a control SRF specific dataset (Figure 8.11, lane 65).

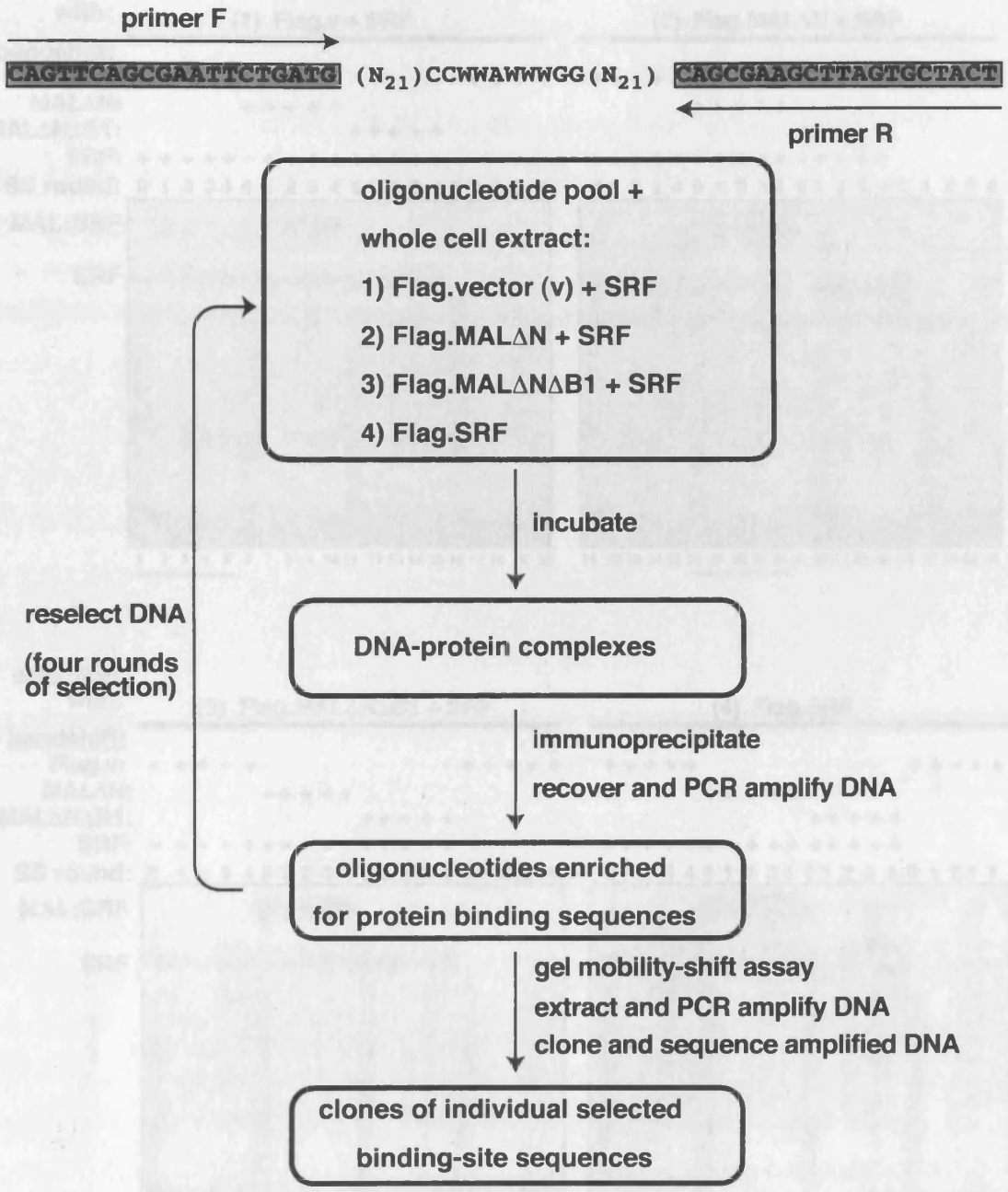


Figure 8.10. Binding site selection experimental design with the SS2 (conSRE) random oligo pool. (A) The sequence of the SS2 (conSRE) oligo is shown (N denotes any nucleotide and W denotes A or T). The double stranded oligonucleotide pool was incubated with whole-cell extracts expressing untagged SRF with Flag-tagged MALΔN wild-type (sample 2) or Flag-tagged MALΔNΔB1 (sample 3) . Extract (v) was mock transfected with empty Flag.vector and combined with untagged SRF was used as a negative control (sample 1). Flag-tagged SRF was used as a positive control (sample 4).

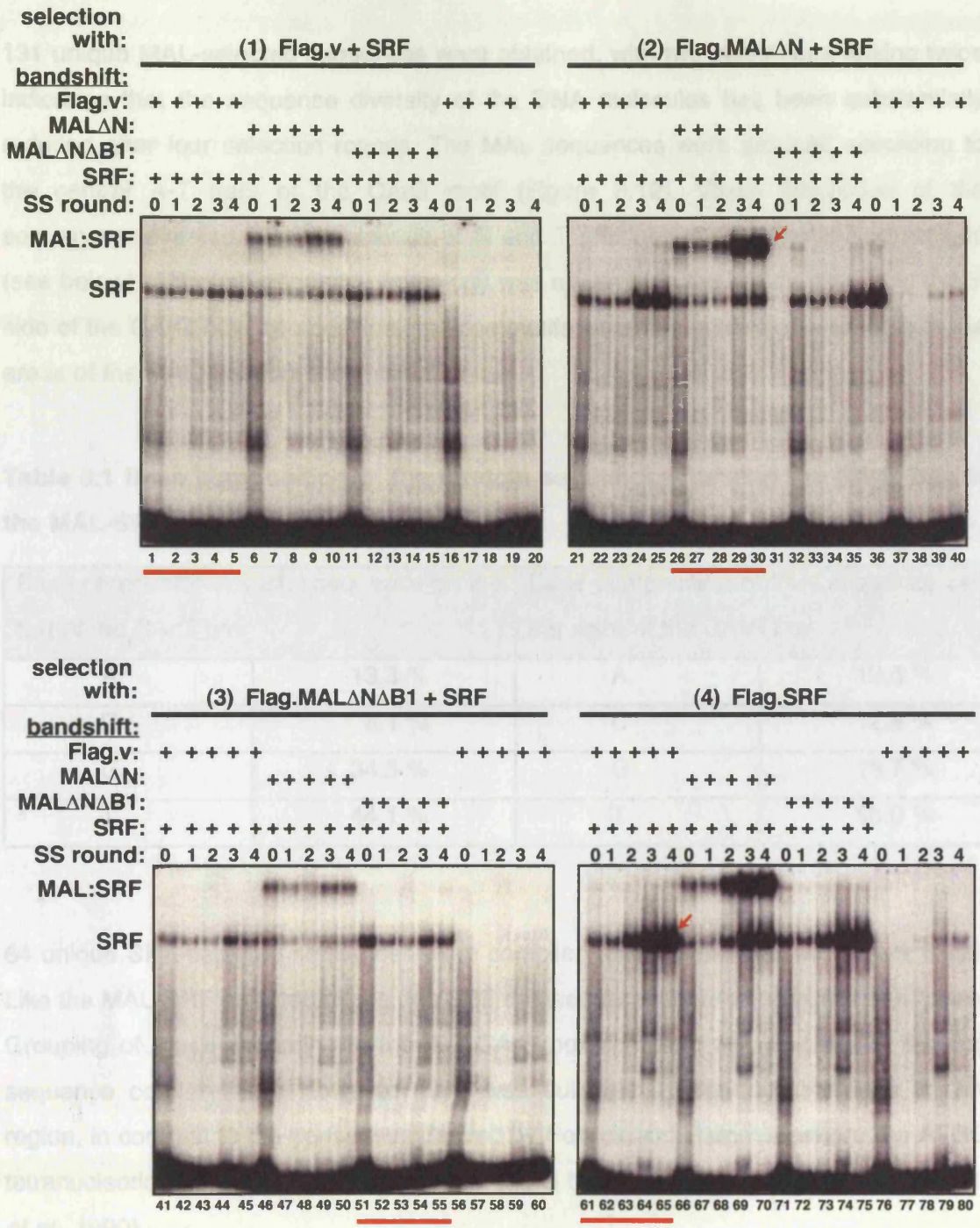


Figure 8.11. Bandshifts of the selected SS2 (conSRE) DNA after four rounds. The selecting proteins for each enriched DNA sample are indicated over each gel. Each gel mobility-shift assay included reactions with the starting DNA (SS round 0), and enriched DNA samples (SS rounds 1-4), assayed for complex formation with the following whole-cell extract combinations: [mock-transfected Flag.vector and SRF], [MALΔN and SRF], [MALΔNΔB1 and SRF], [mock-transfected Flag.vector].The reactions assaying selected DNA with the selecting proteins, where specifically enriched complexes might arise are indicated by red lines. The specific complexes of [MALΔN and SRF] in gel (2) and SRF alone in gel (4) that were excised for further analysis are indicated by red arrows. (The empty Flag vector is designated Flag.v)

131 unique MAL-selected sequences were obtained, with two sequences arising twice indicating that the sequence diversity of the DNA molecules has been substantially reduced after four selection rounds. The MAL sequences were grouped according to the central A-T tract of the CArG motif (Figure 8.12). Visual inspection of the sequences revealed a predominance of G and T residues in the randomised regions (see below). Although sequence homology was apparent in the region bordering the 5' side of the CArG box, no specific sequence motifs were immediately discernable in the areas of the MAL-DNA contacts.

Table 8.1 Base composition of the random sequences flanking the CArG box in the MAL-SRF selected sites

Base composition of 21N sequence on the **left** of the CArG box		Base composition of 21N sequence on the **right** of the CArG box	
A	13.3 %	A	12.5 %
C	8.1 %	C	12.8 %
G	34.5 %	G	18.7 %
T	44.1 %	T	56.0 %

64 unique SRF-selected sequences were compiled as a control dataset (Figure 8.13). Like the MAL-SRF selected oligos, the SRF dataset also contained significant G-T bias. Grouping of these according to the A-T CArG box core also revealed the 5' flanking sequence conservation. However there was substantial sequence variation in this region, in contrast to the consensus derived by Pollock and Treisman where the ATGC tetranucleotide bordering the 5' side of the CArG box was almost invariant (Pollock, R. *et al.*, 1990).

The G-T bias observed in the control SRF selected sequences is unexpected since SRF has not been shown to have such a preference in the flanking sequences it recognises. It is thus possible that this represents an error in oligonucleotide synthesis. Ideally selected oligonucleotides from earlier rounds and samples from the starting oligo pool would have been sequenced to resolve this issue, however this was not possible due to time constraints. It therefore remains possible that the motifs identified from these datasets do not represent *bona fide* consensus sequences but are

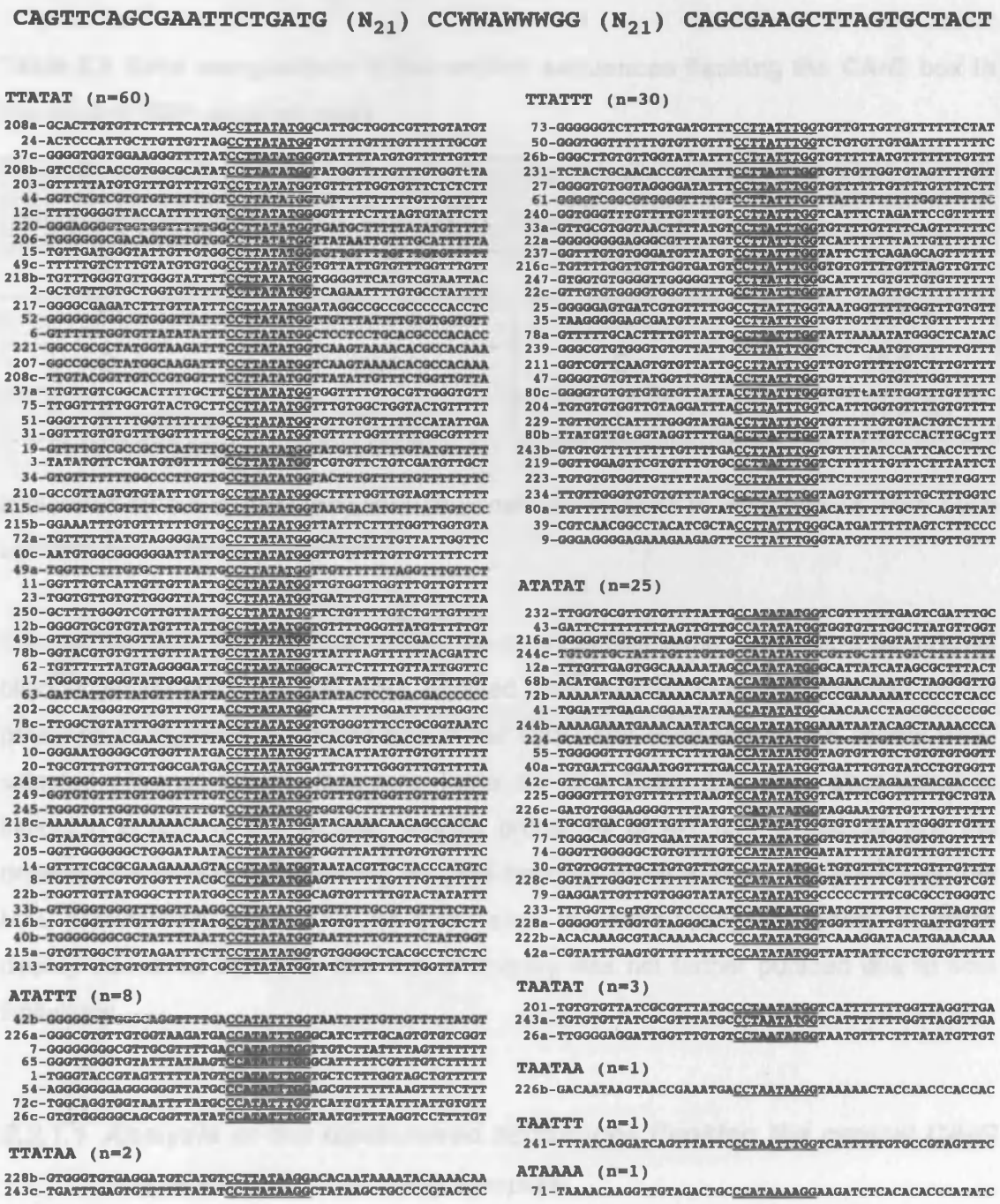


Figure 8.12. MAL-SRF selected oligonucleotides from the SS2(conSRE) oligo pool after four rounds of selection. The 131 unique oligonucleotides are grouped depending on the core A-T tract of the CArG box. Sequences have been aligned so that maximum match is on the 5' flanking region of the CArG box. Only the random sequences with the central partially set CArG box are shown. The CArG box is underlined. The SS2(conSRE) oligonucleotide is shown at the top (the letter N denotes any nucleotide and W denotes A or T).

Table 8.2 Base composition of the random sequences flanking the CArG box in the control SRF selected sites

Base composition of 21N sequence on the **left** of the CArG box		Base composition of 21N sequence on the **right** of the CArG box	
A	19.0 %	A	14.5 %
C	12.5 %	C	25.5 %
G	35.4 %	G	18.8 %
T	33.1 %	T	41.2 %

influenced by the excess presence of certain nucleotides in what should have been completely randomised sequences.

It should be noted however that despite the possible bias in the synthesis of the original oligonucleotide pool, the MAL-SRF selected DNA appears to further enriched for the presence of thymines on either side of the CArG box compared to the control SRF-selected oligonucleotides (compare Table 8.1 and Table 8.2). This observation is intriguing in light of the intrinsic bending properties of A-T rich sequences and the dependence of MAL-SRF binding on DNA bending, since it raises the possibility that MAL preferentially binds SRF on CArG sites the surrounding sequences of which display increased flexibility. This line of enquiry was not further pursued due to time limitations.

8.3.1.1 Analysis of the randomised sequences flanking the central CArG box of the SS2 (con SRE) samples

Despite these concerns the two datasets were used to identify MAL-specific sequence preferences. The Improbizer motif recognition programme was used to analyse the sequence motif content of each dataset, however this approach only identified the partially set CArG box with confidence. Smaller patterns isolated in the randomised regions were barely above the cut-off for nucleotide combinations arising by chance (data not shown).

Attempts to refine the search for MAL-specific motifs in the randomised regions using the SRF selected oligos or human and mouse genomic sequences as background

CAGTTCAGCGAATTCTGATG (N_{21}) CCWWAWWWGG (N_{21}) CAGCGAAGCTTAGTGCTACT

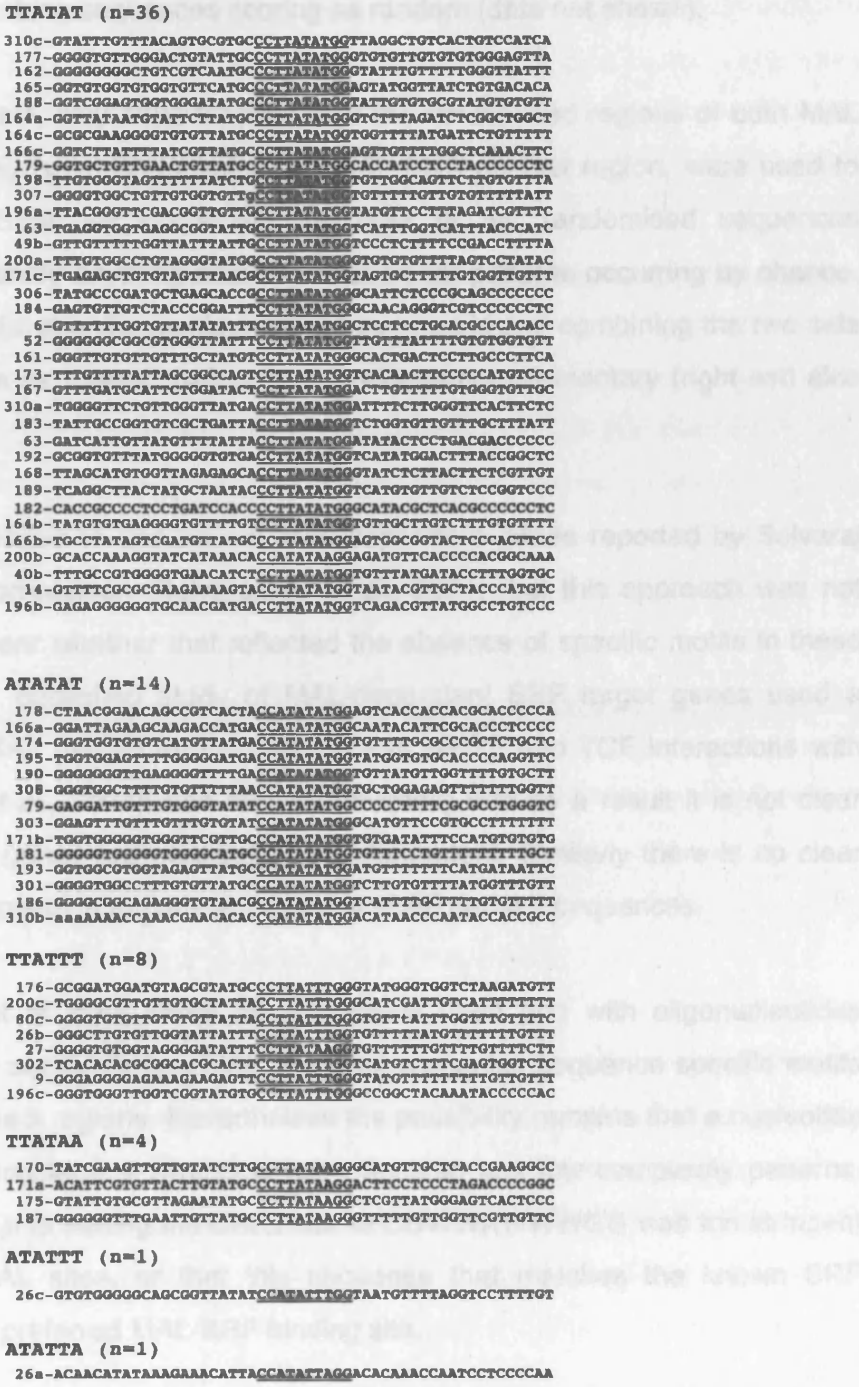


Figure 8.13. SRF selected oligonucleotides from the SS2(conSRE) oligo pool after four rounds of selection. The 64 unique oligonucleotides are grouped depending on the core A-T tract of the CArG box. Sequences have been aligned so that maximum match is on the 5' flanking region of the CArG box. Only the random sequences with the central partially set CArG box are shown. The CArG box is underlined. The SS2(conSRE) oligonucleotide is shown at the top (the letter W denotes A or T).

populations not thought to contain MAL-specific motifs, also only recognised the central CArG box with the flanking sequences scoring as random (data not shown).

To simplify the analysis, only 16bp fragments of the randomised regions of both MAL and SRF selected oligos, which included the MAL-DNA contact region, were used to search for motifs. Searches containing the right or left randomised sequences separately did not identify anything above the cut-off for patterns occurring by chance. Assuming that potential specific motifs might be symmetric and combining the two sets of random sequences as forward (left set) and reverse complementary (right set) also produced no results.

Attempts were also made to use known MAL-dependent genes reported by Selvaraj and coworkers for comparison (Selvaraj, A. *et al.*, 2004), but this approach was not informative. It is unclear whether that reflected the absence of specific motifs in these promoters. The only published study of MAL-dependent SRF target genes used a dominant negative MAL that indiscriminately blocks MRTF and TCF interactions with SRF ((Selvaraj, A. *et al.*, 2004); see also Introduction) and as a result it is not clear whether all reported genes are genuinely MAL dependent. Similarly there is no clear MAL-independent population to compare the MAL-SRF derived sequences.

Thus it appears that at least under the conditions used and with oligonucleotides containing a partially set CArG box MAL does not recognise sequence specific motifs in the MAL-DNA contact regions. Nevertheless the possibility remains that a nucleotide bias during the oligonucleotide synthesis has obscured any low complexity patterns. Another possibility is that setting the CArG box to CCWWAWWWGG was too stringent to select optimal MAL sites, or that this sequence that matches the known SRF consensus is not the preferred MAL-SRF binding site.

8.3.1.2 Investigation of the differences in the central CArG sequences selected by SRF versus MAL-SRF

As discussed previously the original SRF binding site selection derived an extended SRE consensus that included the CArG box halfsite and the sequences immediately adjacent to it ((Pollock, R. *et al.*, 1990); see Section 8.2.1.1). Although the ATG(notG) consensus flanking the CArG box identified in that study does not appear to be

conserved in either the SRF or the SRF-MAL datasets, there appears to be sequence bias in this region (Figure 8.12 and Figure 8.13). I therefore proceeded to investigate the sequence preference of the MAL-SRF and SRF selected oligos in the nucleotides immediately adjacent to CArG box. The fact that the CArG box has been partially set limits the potential observable differences that might be observed within the core 10bp of the site.

The Improbizer programme was used to search for sequence preferences in the regions bordering the CArG box and a plot of nucleotide frequency was produced. The extended motif derived from the SRF dataset broadly matched that identified by Pollock and Treisman ((Pollock, R. *et al.*, 1990); Figure 8.14, compare panels A and B), although in contrast to the invariance of the ATGC tetranucleotide on the 5' side in the original study, the base usage in the SRF motif derived here varied greatly.

The process was repeated for the MAL-SRF dataset and a nucleotide frequency plot was produced (Figure 8.14C). This differed from that of SRF at positions -7 and -6. At position -7 MAL selected predominantly T followed by G, while SRF selected G and T with roughly equal frequencies. At position -6 SRF showed no nucleotide bias whereas MAL showed preference for T or G. Differences were also seen at position +7 where MAL showed no preference, while SRF displayed a small bias for C or T.

In order to identify the consensus halfsites for each dataset, the oligonucleotides were divided in half and the right side was combined with the reverse complementary version of the left side. The halfsite motifs identified by the Improbizer programme for the SRF and MAL-SRF selected sequences varied only in position ±7 where SRF selected predominantly G, while MAL selected mainly T. It should be noted however that the selection frequencies of the predominant bases in either case are quite low. Thus although small differences can be identified in the CArG box flanking sequences selected by SRF and the MAL-SRF complex, these are not striking. This does not preclude a possible significance of these differences in the formation of MAL-SRF-DNA complexes. For instance DNA bending and cofactor selection by MCM1 depends to an extent on the sequences directly flanking the MCM1 site (Acton, T. B. *et al.*, 1997; Lim, F. L. *et al.*, 2003). The importance of these differences was not investigated experimentally due to time limitations.

Figure 8.14. Consensus motifs derived from the MAL-SRF and SRF selected SS2.(conSRE) oligonucleotides. (A) The extended consensus and halfsite consensus derived by (Pollock and Treisman, 1990). Bases shown one over the other denote preference for the top base. (B) Extended and halfsite consensus motif derived from the SRF dataset. The size of the letters represents their frequency at each position. The preset CArG sequence within the random oligonucleotide is shown in black under the motif. The letter W designates A or T. The extended consensus motifs were derived with the Improbizer programme, searching for sequence bias extending on either side of the 10bp CArG motif. The halfsites were derived in the same manner by dividing the oligos in half and combining the forward (left side) and reverse complement (right side) sequences. Motif images were made with the weblogo programme. (C) Extended and halfsite consensus motif derived from the MAL-SRF dataset. Motifs were derived as in B.

8.4 Evaluation of the site selection experiments and future directions

The two binding site selection experiments presented in this section do not conclusively address the sequence dependence of the MAL-SRF complex formation. Attempts to select MAL-SRF specific sites from a completely random oligonucleotide pool in order to investigate whether the sequence of the CArG box itself affected MAL-SRF complex formation, were unsuccessful due to technical problems. The subsequent discovery that MAL contacts DNA flanking the main SRF-CArG complex, and that these contacts are required for efficient MAL-SRF complex formation increased the length of the sequences that would have to be examined in order to identify authentic MAL-SRF DNA specificity. The second binding site selection approach involved including a partially set CArG box in the middle of the random sequence in order to explore the sequence specificity of the MAL-DNA interactions. The results of this experiment were inconclusive, since time constraints did not allow completion of the data analysis. Analysis of the preliminary data indicates that despite the possibility of a nucleotide bias in the original oligonucleotide pool, the MAL-SRF selected sequences exhibit a preference for T-tracts on either side of the CArG box. These sequences did not reveal specific patterns, but this could be due to the biased oligonucleotide synthesis. It would be interesting to pursue this observation further in order to investigate whether MAL selects CArG sites located within intrinsically more flexible sequences. A starting point to achieve this would be using a similar binding site selection approach. Control procedures should be added to ensure the unbiased synthesis of the original random oligonucleotide material. This could be easily achieved by sequencing samples of the original double-stranded DNA. This experiment could also be modified to explore the sequence specificity of MAL-SRF selected CArG boxes, by constraining fewer nucleotides in the central random sequence to the CArG box consensus.

References

Aberle, H., Bauer, A., Stappert, J., Kispert, A., and Kemler, R. (1997). beta-catenin is a target for the ubiquitin-proteasome pathway. Embo J *16*, 3797-3804.

Acton, T. B., Mead, J., Steiner, A. M., and Vershon, A. K. (2000). Scanning mutagenesis of Mcm1: residues required for DNA binding, DNA bending, and transcriptional activation by a MADS-box protein. Mol Cell Biol *20*, 1-11.

Acton, T. B., Zhong, H., and Vershon, A. K. (1997). DNA-binding specificity of Mcm1: operator mutations that alter DNA-bending and transcriptional activities by a MADS box protein. Mol Cell Biol *17*, 1881-1889.

Affolter, M., Montagne, J., Walldorf, U., Groppe, J., Kloter, U., LaRosa, M., and Gehring, W. J. (1994). The Drosophila SRF homolog is expressed in a subset of tracheal cells and maps within a genomic region required for tracheal development. Development *120*, 743-753.

Alber, T. (1992). Structure of the leucine zipper. Curr Opin Genet Dev *2*, 205-210.

Alberti, S., Krause, S. M., Kretz, O., Philippar, U., Lemberger, T., Casanova, E., Wiebel, F. F., Schwarz, H., Frotscher, M., Schutz, G., and Nordheim, A. (2005). Neuronal migration in the murine rostral migratory stream requires serum response factor. Proc Natl Acad Sci U S A *102*, 6148-6153.

Alberts, A. S., Geneste, O., and Treisman, R. (1998). Activation of SRF-regulated chromosomal templates by Rho-family GTPases requires a signal that also induces H4 hyperacetylation. Cell *92*, 475-487.

Alvarez-Buylla, E. R., Pelaz, S., Liljegren, S. J., Gold, S. E., Burgeff, C., Ditta, G. S., Ribas de Pouplana, L., Martinez-Castilla, L., and Yanofsky, M. F. (2000). An ancestral MADS-box gene duplication occurred before the divergence of plants and animals. Proc Natl Acad Sci U S A *97*, 5328-5333.

Amar, N., Messenguy, F., El Bakkoury, M., and Dubois, E. (2000). ArgRII, a component of the ArgR-Mcm1 complex involved in the control of arginine metabolism in Saccharomyces cerevisiae, is the sensor of arginine. Mol Cell Biol *20*, 2087-2097.

Aravind, L., and Koonin, E. V. (2000). SAP - a putative DNA-binding motif involved in chromosomal organization. Trends Biochem Sci *25*, 112-114.

Arber, S., Barbayannis, F. A., Hanser, H., Schneider, C., Stanyon, C. A., Bernard, O., and Caroni, P. (1998). Regulation of actin dynamics through phosphorylation of cofilin by LIM-kinase. Nature *393*, 805-809.

Arsenian, S., Weinhold, B., Oelgeschlager, M., Ruther, U., and Nordheim, A. (1998). Serum response factor is essential for mesoderm formation during mouse embryogenesis. EMBO J *17*, 6289-6299.

Avila, S., Casero, M. C., Fernandez-Canton, R., and Sastre, L. (2002). Transactivation domains are not functionally conserved between vertebrate and invertebrate serum response factors. Eur J Biochem *269*, 3669-3677.

Ayadi, A., Zheng, H., Sobieszczuk, P., Buchwalter, G., Moerman, P., Alitalo, K., and Wasylyk, B. (2001). Net-targeted mutant mice develop a vascular phenotype and up-regulate egr-1. Embo J *20*, 5139-5152.

Balza, R. O., Jr., and Misra, R. P. (2006). Role of the serum response factor in regulating contractile apparatus gene expression and sarcomeric integrity in cardiomyocytes. J Biol Chem *281*, 6498-6510.

Banuett, F. (1998). Signalling in the yeasts: an informational cascade with links to the filamentous fungi. Microbiol Mol Biol Rev *62*, 249-274.

Baxevanis, A. D., and Vinson, C. R. (1993). Interactions of coiled coils in transcription factors: where is the specificity? Curr Opin Genet Dev *3*, 278-285.

Belaguli, N. S., Sepulveda, J. L., Nigam, V., Charron, F., Nemer, M., and Schwartz, R. J. (2000). Cardiac tissue enriched factors serum response factor and GATA-4 are mutual coregulators. Mol Cell Biol *20*, 7550-7558.

Belaguli, N. S., Zhou, W., Trinh, T. H., Majesky, M. W., and Schwartz, R. J. (1999). Dominant negative murine serum response factor: alternative splicing within the activation domain inhibits transactivation of serum response factor binding targets. Mol Cell Biol *19*, 4582-4591.

Bender, A., and Sprague, G. F., Jr. (1987). MAT alpha 1 protein, a yeast transcription activator, binds synergistically with a second protein to a set of cell-type-specific genes. Cell *50*, 681-691.

Beqaj, S., Jakkaraju, S., Mattingly, R. R., Pan, D., and Schuger, L. (2002). High RhoA activity maintains the undifferentiated mesenchymal cell phenotype, whereas RhoA down-regulation by laminin-2 induces smooth muscle myogenesis. J Cell Biol *156*, 893-903.

Berkes, C. A., and Tapscott, S. J. (2005). MyoD and the transcriptional control of myogenesis. Semin Cell Dev Biol *16*, 585-595.

Bernardi, A., Gaillard, C., and Bernardi, G. (1975). The specificity of five DNAases as studied by the analysis of 5'-terminal doublets. Eur J Biochem *52*, 451-457.

Bernardi, G., Ehrlich, S. D., and Thiery, J. P. (1973). The specificity of deoxyribonucleases and their use in nucleotide sequence studies. Nat New Biol *246*, 36-40.

Boros, J., Lim, F. L., Darieva, Z., Pic-Taylor, A., Harman, R., Morgan, B. A., and Sharrocks, A. D. (2003). Molecular determinants of the cell-cycle regulated Mcm1p-Fkh2p transcription factor complex. Nucleic Acids Res *31*, 2279-2288.

Bruhn, L., Hwang-Shum, J. J., and Sprague, G. F., Jr. (1992). The N-terminal 96 residues of MCM1, a regulator of cell type-specific genes in Saccharomyces cerevisiae, are sufficient for DNA binding, transcription activation, and interaction with alpha 1. Mol Cell Biol *12*, 3563-3572.

Bruhn, L., and Sprague, G. F., Jr. (1994). MCM1 point mutants deficient in expression of alpha-specific genes: residues important for interaction with alpha 1. Mol Cell Biol *14*, 2534-2544.

Buchwalter, G., Gross, C., and Wasylyk, B. (2004). Ets ternary complex transcription factors. Gene *324*, 1-14.

Calkhoven, C. F., and Ab, G. (1996). Multiple steps in the regulation of transcription-factor level and activity. Biochem J *317 (Pt 2)*, 329-342.

Camoretti-Mercado, B., Fernandes, D. J., Dewundara, S., Churchill, J., Ma, L., Kogut, P. C., McConville, J. F., Parmacek, M. S., and Solway, J. (2006). Inhibition of transforming growth factor beta-enhanced serum response factor-dependent transcription by SMAD7. J Biol Chem *281*, 20383-20392.

Camoretti-Mercado, B., Liu, H. W., Halayko, A. J., Forsythe, S. M., Kyle, J. W., Li, B., Fu, Y., McConville, J., Kogut, P., Vieira, J. E., *et al.* (2000). Physiological control of smooth muscle-specific gene expression through regulated nuclear translocation of serum response factor. J Biol Chem *275*, 30387-30393.

Cao, D., Wang, Z., Zhang, C. L., Oh, J., Xing, W., Li, S., Richardson, J. A., Wang, D. Z., and Olson, E. N. (2005). Modulation of smooth muscle gene expression by association of histone acetyltransferases and deacetylases with myocardin. Mol Cell Biol *25*, 364-376.

Carey, M. (1998). The enhanceosome and transcriptional synergy. Cell *92*, 5-8.

Carr, E. A., Mead, J., and Vershon, A. K. (2004). Alpha1-induced DNA bending is required for transcriptional activation by the Mcm1-alpha1 complex. Nucleic Acids Res *32*, 2298-2305. Print 2004.

Carson, J. A., Fillmore, R. A., Schwartz, R. J., and Zimmer, W. E. (2000). The smooth muscle gamma-actin gene promoter is a molecular target for the mouse bagpipe homologue, mNkx3-1, and serum response factor. J Biol Chem *275*, 39061-39072.

Cen, B., Selvaraj, A., Burgess, R. C., Hitzler, J. K., Ma, Z., Morris, S. W., and Prywes, R. (2003). Megakaryoblastic leukemia 1, a potent transcriptional coactivator for serum response factor (SRF), is required for serum induction of SRF target genes. Mol Cell Biol *23*, 6597-6608.

Cesari, F., Brecht, S., Vintersten, K., Vuong, L. G., Hofmann, M., Klingel, K., Schnorr, J. J., Arsenian, S., Schild, H., Herdegen, T., *et al.* (2004). Mice deficient for the ets transcription factor elk-1 show normal immune responses and mildly impaired neuronal gene activation. Mol Cell Biol *24*, 294-305.

Chang, D. F., Belaguli, N. S., Iyer, D., Roberts, W. B., Wu, S. P., Dong, X. R., Marx, J. G., Moore, M. S., Beckerle, M. C., Majesky, M. W., and Schwartz, R. J. (2003). Cysteine-rich LIM-only proteins CRP1 and CRP2 are potent smooth muscle differentiation cofactors. Dev Cell *4*, 107-118.

Chang, P. S., Li, L., McAnally, J., and Olson, E. N. (2001). Muscle specificity encoded by specific serum response factor-binding sites. J Biol Chem *276*, 17206-17212.

Charvet, C., Houbron, C., Parlakian, A., Giordani, J., Lahoute, C., Bertrand, A., Sotiropoulos, A., Renou, L., Schmitt, A., Melki, J., *et al.* (2006). New role for serum response factor in postnatal skeletal muscle growth and regeneration via the interleukin 4 and insulin-like growth factor 1 pathways. Mol Cell Biol *26*, 6664-6674.

Chen, C. Y., Croissant, J., Majesky, M., Topouzis, S., McQuinn, T., Frankovsky, M. J., and Schwartz, R. J. (1996). Activation of the cardiac alpha-actin promoter depends upon serum response factor, Tinman homologue, Nkx-2.5, and intact serum response elements. Dev Genet *19*, 119-130.

Chen, F., Kook, H., Milewski, R., Gitler, A. D., Lu, M. M., Li, J., Nazarian, R., Schnepp, R., Jen, K., Biben, C., *et al.* (2002). Hop is an unusual homeobox gene that modulates cardiac development. Cell *110*, 713-723.

Chen, J., Kitchen, C. M., Streb, J. W., and Miano, J. M. (2002). Myocardin: a component of a molecular switch for smooth muscle differentiation. J Mol Cell Cardiol *34*, 1345-1356.

Chen, L., Glover, J. N., Hogan, P. G., Rao, A., and Harrison, S. C. (1998). Structure of the DNA-binding domains from NFAT, Fos and Jun bound specifically to DNA. Nature *392*, 42-48.

Chinenov, Y., and Kerppola, T. K. (2001). Close encounters of many kinds: Fos-Jun interactions that mediate transcription regulatory specificity. Oncogene *20*, 2438-2452.

Copeland, J. W., and Treisman, R. (2002). The diaphanous-related formin mDia1 controls serum response factor activity through its effects on actin polymerization. Mol Biol Cell *13*, 4088-4099.

Costello, P. S., Nicolas, R. H., Watanabe, Y., Rosewell, I., and Treisman, R. (2004). Ternary complex factor SAP-1 is required for Erk-mediated thymocyte positive selection. Nat Immunol *5*, 289-298.

Courey, A. J. (2001). Cooperativity in transcriptional control. Curr Biol *11*, R250-252.

Creemers, E. E., Sutherland, L. B., Oh, J., Barbosa, A. C., and Olson, E. N. (2006). Coactivation of MEF2 by the SAP domain proteins myocardin and MASTR. Mol Cell *23*, 83-96.

Criqui-Filipe, P., Ducret, C., Maira, S. M., and Wasylyk, B. (1999). Net, a negative Ras-switchable TCF, contains a second inhibition domain, the CID, that mediates repression through interactions with CtBP and de-acetylation. Embo J *18*, 3392-3403.

Crothers, D. M., Gartenberg, M. R., and Shrader, T. E. (1991). DNA bending in protein-DNA complexes. Methods Enzymol *208*, 118-146.

Dalgleish, P., and Sharrocks, A. D. (2000). The mechanism of complex formation between Fli-1 and SRF transcription factors. Nucleic Acids Res *28*, 560-569.

Dalton, S., and Treisman, R. (1992). Characterization of SAP-1, a protein recruited by serum response factor to the c-fos serum response element. Cell *68*, 597-612.

Darieva, Z., Pic-Taylor, A., Boros, J., Spanos, A., Geymonat, M., Reece, R. J., Sedgwick, S. G., Sharrocks, A. D., and Morgan, B. A. (2003). Cell cycle-regulated transcription through the FHA domain of Fkh2p and the coactivator Ndd1p. Curr Biol *13*, 1740-1745.

De Bodt, S., Raes, J., Van de Peer, Y., and Theissen, G. (2003). And then there were many: MADS goes genomic. Trends Plant Sci *8*, 475-483.

De Cesare, D., Fimia, G. M., and Sassone-Corsi, P. (1999). Signaling routes to CREM and CREB: plasticity in transcriptional activation. Trends Biochem Sci *24*, 281-285.

DiDonato, J. A., Hayakawa, M., Rothwarf, D. M., Zandi, E., and Karin, M. (1997). A cytokine-responsive IkappaB kinase that activates the transcription factor NF-kappaB. Nature *388*, 548-554.

Ding, W., Witte, M. M., and Scott, R. E. (1999). Transformation blocks differentiation-induced inhibition of serum response factor interactions with serum response elements. Cancer Res *59*, 3795-3802.

Dodou, E., and Treisman, R. (1997). The Saccharomyces cerevisiae MADS-box transcription factor Rlm1 is a target for the Mpk1 mitogen-activated protein kinase pathway. Mol Cell Biol *17*, 1848-1859.

Dolan, J. W., and Fields, S. (1991). Cell-type-specific transcription in yeast. Biochim Biophys Acta *1088*, 155-169.

Donaldson, L. W., Petersen, J. M., Graves, B. J., and McIntosh, L. P. (1994). Secondary structure of the ETS domain places murine Ets-1 in the superfamily of winged helix-turn-helix DNA-binding proteins. Biochemistry *33*, 13509-13516.

Donaldson, L. W., Petersen, J. M., Graves, B. J., and McIntosh, L. P. (1996). Solution structure of the ETS domain from murine Ets-1: a winged helix-turn-helix DNA binding motif. Embo J *15*, 125-134.

Du, K. L., Chen, M., Li, J., Lepore, J. J., Mericko, P., and Parmacek, M. S. (2004). Megakaryoblastic leukemia factor-1 transduces cytoskeletal signals and induces smooth muscle cell differentiation from undifferentiated embryonic stem cells. J Biol Chem *279*, 17578-17586. Epub 12004 Feb 17517.

Du, K. L., Ip, H. S., Li, J., Chen, M., Dandre, F., Yu, W., Lu, M. M., Owens, G. K., and Parmacek, M. S. (2003). Myocardin is a critical serum response factor cofactor in the transcriptional program regulating smooth muscle cell differentiation. Mol Cell Biol *23*, 2425-2437.

Dubois, E., Bercy, J., and Messenguy, F. (1987). Characterization of two genes, ARGRI and ARGRIII required for specific regulation of arginine metabolism in yeast. Mol Gen Genet *207*, 142-148.

Dubois, E., and Messenguy, F. (1991). In vitro studies of the binding of the ARGR proteins to the ARG5,6 promoter. Mol Cell Biol *11*, 2162-2168.

El Bakkoury, M., Dubois, E., and Messenguy, F. (2000). Recruitment of the yeast MADS-box proteins, ArgRI and Mcm1 by the pleiotropic factor ArgRIII is required for their stability. Mol Microbiol *35*, 15-31.

Errede, B., and Ammerer, G. (1989). STE12, a protein involved in cell-type-specific transcription and signal transduction in yeast, is part of protein-DNA complexes. Genes Dev *3*, 1349-1361.

Etienne-Manneville, S., and Hall, A. (2002). Rho GTPases in cell biology. Nature *420*, 629-635.

Franzoso, G., Carlson, L., Brown, K., Daucher, M. B., Bressler, P., and Siebenlist, U. (1996). Activation of the serum response factor by p65/NF-kappaB. Embo J *15*, 3403-3412.

Garvie, C. W., and Wolberger, C. (2001). Recognition of specific DNA sequences. Mol Cell *8*, 937-946.

Gauthier-Rouviere, C., Cavadore, J. C., Blanchard, J. M., Lamb, N. J., and Fernandez, A. (1991). p67SRF is a constitutive nuclear protein implicated in the modulation of genes required throughout the G1 period. Cell Regul *2*, 575-588.

Geneste, O., Copeland, J. W., and Treisman, R. (2002). LIM kinase and Diaphanous cooperate to regulate serum response factor and actin dynamics. J Cell Biol *157*, 831-838.

Gill, G. (2005). Something about SUMO inhibits transcription. Curr Opin Genet Dev *15*, 536-541.

Gilman, M. Z., Wilson, R. N., and Weinberg, R. A. (1986). Multiple protein-binding sites in the 5'-flanking region regulate c-fos expression. Mol Cell Biol *6*, 4305-4316.

Gineitis, D., and Treisman, R. (2001). Differential usage of signal transduction pathways defines two types of serum response factor target gene. J Biol Chem *276*, 24531-24539.

Giovane, A., Pintzas, A., Maira, S. M., Sobieszczuk, P., and Wasylyk, B. (1994). Net, a new ets transcription factor that is activated by Ras. Genes Dev *8*, 1502-1513.

Glozak, M. A., Sengupta, N., Zhang, X., and Seto, E. (2005). Acetylation and deacetylation of non-histone proteins. Gene *363*, 15-23.

Graham, R., and Gilman, M. (1991). Distinct protein targets for signals acting at the c-fos serum response element. Science *251*, 189-192.

Groisman, R., Masutani, H., Leibovitch, M. P., Robin, P., Soudant, I., Trouche, D., and Harel-Bellan, A. (1996). Physical interaction between the mitogen-responsive serum response factor and myogenic basic-helix-loop-helix proteins. J Biol Chem *271*, 5258-5264.

Grosse, R., Copeland, J. W., Newsome, T. P., Way, M., and Treisman, R. (2003). A role for VASP in RhoA-Diaphanous signalling to actin dynamics and SRF activity. Embo J *22*, 3050-3061.

Grueneberg, D. A., Henry, R. W., Brauer, A., Novina, C. D., Cheriyath, V., Roy, A. L., and Gilman, M. (1997). A multifunctional DNA-binding protein that promotes the formation of serum response factor/homeodomain complexes: identity to TFII-I. Genes Dev *11*, 2482-2493.

Grueneberg, D. A., Natesan, S., Alexandre, C., and Gilman, M. Z. (1992). Human and Drosophila homeodomain proteins that enhance the DNA-binding activity of serum response factor. Science *257*, 1089-1095.

Gu, W., and Roeder, R. G. (1997). Activation of p53 sequence-specific DNA binding by acetylation of the p53 C-terminal domain. Cell *90*, 595-606.

Guillemin, K., Groppe, J., Ducker, K., Treisman, R., Hafen, E., Affolter, M., and Krasnow, M. A. (1996). The pruned gene encodes the Drosophila serum response factor and regulates cytoplasmic outgrowth during terminal branching of the tracheal system. Development *122*, 1353-1362.

Gupta, M., Kogut, P., Davis, F. J., Belaguli, N. S., Schwartz, R. J., and Gupta, M. P. (2001). Physical interaction between the MADS box of serum response factor and the TEA/ATTS DNA-binding domain of transcription enhancer factor-1. J Biol Chem *276*, 10413-10422.

Gustafson, T. A., Taylor, A., and Kedes, L. (1989). DNA bending is induced by a transcription factor that interacts with the human c-FOS and alpha-actin promoters. Proc Natl Acad Sci U S A *86*, 2162-2166.

Hagen, D. C., Bruhn, L., Westby, C. A., and Sprague, G. F., Jr. (1993). Transcription of alpha-specific genes in Saccharomyces cerevisiae: DNA sequence requirements for activity of the coregulator alpha 1. Mol Cell Biol *13*, 6866-6875.

Han, A., Pan, F., Stroud, J. C., Youn, H. D., Liu, J. O., and Chen, L. (2003). Sequence-specific recruitment of transcriptional co-repressor Cabin1 by myocyte enhancer factor-2. Nature *422*, 730-734.

Han, Z., Li, X., Wu, J., and Olson, E. N. (2004). A myocardin-related transcription factor regulates activity of serum response factor in Drosophila. Proc Natl Acad Sci U S A *101*, 12567-12572.

Hanlon, M., and Sealy, L. (1999). Ras regulates the association of serum response factor and CCAAT/enhancer-binding protein beta. J Biol Chem *274*, 14224-14228.

Hassler, M., and Richmond, T. J. (2001). The B-box dominates SAP-1-SRF interactions in the structure of the ternary complex. Embo J *20*, 3018-3028.

Hautmann, M. B., Madsen, C. S., Mack, C. P., and Owens, G. K. (1998). Substitution of the degenerate smooth muscle (SM) alpha-actin CC(A/T-rich)6GG elements with c-fos serum response elements results in increased basal expression but relaxed SM cell specificity and reduced angiotensin II inducibility. J Biol Chem *273*, 8398-8406.

Heidenreich, O., Neininger, A., Schratt, G., Zinck, R., Cahill, M. A., Engel, K., Kotlyarov, A., Kraft, R., Kostka, S., Gaestel, M., and Nordheim, A. (1999). MAPKAP kinase 2 phosphorylates serum response factor in vitro and in vivo. J Biol Chem *274*, 14434-14443.

Herring, B. P., Kriegel, A. M., and Hoggatt, A. M. (2001). Identification of Barx2b, a serum response factor-associated homeodomain protein. J Biol Chem *276*, 14482-14489.

Hill, C. S., Marais, R., John, S., Wynne, J., Dalton, S., and Treisman, R. (1993). Functional analysis of a growth factor-responsive transcription factor complex. Cell *73*, 395-406.

Hill, C. S., and Treisman, R. (1995a). Differential activation of c-fos promoter elements by serum, lysophosphatidic acid, G proteins and polypeptide growth factors. Embo J *14*, 5037-5047.

Hill, C. S., Wynne, J., and Treisman, R. (1994). Serum-regulated transcription by serum response factor (SRF): a novel role for the DNA binding domain. Embo J *13*, 5421-5432.

Hill, C. S., Wynne, J. K., and Treisman, R. H. (1995b). The Rho family GTPases RhoA, Rac1 and CDC42hs regulate transcriptional activation by SRF. Cell *81*, 1159-1170.

Hill, R. E., Favor, J., Hogan, B. L., Ton, C. C., Saunders, G. F., Hanson, I. M., Prosser, J., Jordan, T., Hastie, N. D., and van Heyningen, V. (1991). Mouse small eye results from mutations in a paired-like homeobox-containing gene. Nature *354*, 522-525.

Hipskind, R. A., Rao, V. N., Mueller, C. G., Reddy, E. S., and Nordheim, A. (1991). Ets-related protein Elk-1 is homologous to the c-fos regulatory factor p62TCF. Nature *354*, 531-534.

Hirschi, K. K., Lai, L., Belaguli, N. S., Dean, D. A., Schwartz, R. J., and Zimmer, W. E. (2002). Transforming growth factor-beta induction of smooth muscle cell phenotpye requires transcriptional and post-transcriptional control of serum response factor. J Biol Chem *277*, 6287-6295.

Howell, M., and Hill, C. S. (1997). XSmad2 directly activates the activin-inducible, dorsal mesoderm gene XFKH1 in Xenopus embryos. Embo J *16*, 7411-7421.

Huang, H., Mizukami, Y., Hu, Y., and Ma, H. (1993). Isolation and characterization of the binding sequences for the product of the Arabidopsis floral homeotic gene AGAMOUS. Nucleic Acids Res *21*, 4769-4776.

Huang, K., Louis, J. M., Donaldson, L., Lim, F. L., Sharrocks, A. D., and Clore, G. M. (2000). Solution structure of the MEF2A-DNA complex: structural basis for the modulation of DNA bending and specificity by MADS-box transcription factors. Embo J *19*, 2615-2628.

Iyer, D., Belaguli, N., Fluck, M., Rowan, B. G., Wei, L., Weigel, N. L., Booth, F. W., Epstein, H. F., Schwartz, R. J., and Balasubramanyam, A. (2003). Novel phosphorylation target in the serum response factor MADS box regulates alpha-actin transcription. Biochemistry *42*, 7477-7486.

Iyer, D., Chang, D., Marx, J., Wei, L., Olson, E. N., Parmacek, M. S., Balasubramanyam, A., and Schwartz, R. J. (2006). Serum response factor MADS box serine-162 phosphorylation switches proliferation and myogenic gene programs. Proc Natl Acad Sci U S A *103*, 4516-4521.

Jacobs, D., Glossip, D., Xing, H., Muslin, A. J., and Kornfeld, K. (1999). Multiple docking sites on substrate proteins form a modular system that mediates recognition by ERK MAP kinase. Genes Dev *13*, 163-175.

Janknecht, R., Hipskind, R. A., Houthaeve, T., Nordheim, A., and Stunnenberg, H. G. (1992a). Identification of multiple SRF N-terminal phosphorylation sites affecting DNA binding properties. Embo J *11*, 1045-1054.

Janknecht, R., and Nordheim, A. (1992b). Elk-1 protein domains required for direct and SRF-assisted DNA-binding. Nucleic Acids Res *20*, 3317-3324.

Ji, X., Zhang, P., Armstrong, R. N., and Gilliland, G. L. (1992). The three-dimensional structure of a glutathione S-transferase from the mu gene class. Structural analysis of the binary complex of isoenzyme 3-3 and glutathione at 2.2-A resolution. Biochemistry *31*, 10169-10184.

Johansen, F. E., and Prywes, R. (1993). Identification of transcriptional activation and inhibitory domains in serum response factor (SRF) by using GAL4-SRF constructs. Mol Cell Biol *13*, 4640-4647.

Johansen, F. E., and Prywes, R. (1994). Two pathways for serum regulation of the c-fos serum response element require specific sequence elements and a minimal domain of serum response factor. Mol Cell Biol *14*, 5920-5928.

Kadonaga, J. T. (2004). Regulation of RNA polymerase II transcription by sequence-specific DNA binding factors. Cell *116*, 247-257.

Kasza, A., O'Donnell, A., Gascoigne, K., Zeef, L. A., Hayes, A., and Sharrocks, A. D. (2005). The ETS domain transcription factor Elk-1 regulates the expression of its partner protein, SRF. J Biol Chem *280*, 1149-1155.

Kaufmann, K., Melzer, R., and Theissen, G. (2005). MIKC-type MADS-domain proteins: structural modularity, protein interactions and network evolution in land plants. Gene *347*, 183-198.

Keleher, C. A., Goutte, C., and Johnson, A. D. (1988). The yeast cell-type-specific repressor alpha 2 acts cooperatively with a non-cell-type-specific protein. Cell *53*, 927-936.

Keleher, C. A., Passmore, S., and Johnson, A. D. (1989). Yeast repressor alpha 2 binds to its operator cooperatively with yeast protein Mcm1. Mol Cell Biol *9*, 5228-5230.

Kemp, P. R., and Metcalfe, J. C. (2000). Four isoforms of serum response factor that increase or inhibit smooth-muscle-specific promoter activity. Biochem J *345 Pt 3*, 445-451.

Kerppola, T. K., and Curran, T. (1991). DNA bending by Fos and Jun: the flexible hinge model. Science *254*, 1210-1214.

Khorasanizadeh, S. (2004). The nucleosome: from genomic organization to genomic regulation. Cell *116*, 259-272.

Kipp, M., Gohring, F., Ostendorp, T., van Drunen, C. M., van Driel, R., Przybylski, M., and Fackelmayer, F. O. (2000). SAF-Box, a conserved protein domain that specifically recognizes scaffold attachment region DNA. Mol Cell Biol *20*, 7480-7489.

Kiejman, M. P., Pereira, L. A., van Zeeburg, H. J., Gilfillan, S., Meisterernst, M., and Timmers, H. T. (2004). NC2alpha interacts with BTAF1 and stimulates its ATP-dependent association with TATA-binding protein. Mol Cell Biol *24*, 10072-10082.

Knoll, B., Kretz, O., Fiedler, C., Alberti, S., Schutz, G., Frotscher, M., and Nordheim, A. (2006). Serum response factor controls neuronal circuit assembly in the hippocampus. Nat Neurosci *9*, 195-204.

Koranda, M., Schleiffer, A., Endler, L., and Ammerer, G. (2000). Forkhead-like transcription factors recruit Ndd1 to the chromatin of G2/M-specific promoters. Nature *406*, 94-98.

Kumar, R., Reynolds, D. M., Shevchenko, A., Shevchenko, A., Goldstone, S. D., and Dalton, S. (2000). Forkhead transcription factors, Fkh1p and Fkh2p, collaborate with Mcm1p to control transcription required for M-phase. Curr Biol *10*, 896-906.

Kuo, M. H., Nadeau, E. T., and Grayhack, E. J. (1997). Multiple phosphorylated forms of the Saccharomyces cerevisiae Mcm1 protein include an isoform induced in response to high salt concentrations. Mol Cell Biol *17*, 819-832.

Kuwahara, K., Barrientos, T., Pipes, G. C., Li, S., and Olson, E. N. (2005). Muscle-specific signaling mechanism that links actin dynamics to serum response factor. Mol Cell Biol *25*, 3173-3181.

Lavigne, P., Kondejewski, L. H., Houston, M. E., Jr., Sonnichsen, F. D., Lix, B., Skyes, B. D., Hodges, R. S., and Kay, C. M. (1995). Preferential heterodimeric parallel coiled-coil formation by synthetic Max and c-Myc leucine zippers: a description of putative electrostatic interactions responsible for the specificity of heterodimerization. J Mol Biol *254*, 505-520.

Lee, T. I., and Young, R. A. (2000). Transcription of eukaryotic protein-coding genes. Annu Rev Genet *34*, 77-137.

Lemon, B., and Tjian, R. (2000). Orchestrated response: a symphony of transcription factors for gene control. Genes Dev *14*, 2551-2569.

Leung, S., and Miyamoto, N. G. (1989). Point mutational analysis of the human c-fos serum response factor binding site. Nucleic Acids Res *17*, 1177-1195.

Li, J., Zhu, X., Chen, M., Cheng, L., Zhou, D., Lu, M. M., Du, K., Epstein, J. A., and Parmacek, M. S. (2005). Myocardin-related transcription factor B is required in cardiac neural crest for smooth muscle differentiation and cardiovascular development. Proc Natl Acad Sci U S A *102*, 8916-8921.

Li, S., Chang, S., Qi, X., Richardson, J. A., and Olson, E. N. (2006). Requirement of a myocardin-related transcription factor for development of mammary myoepithelial cells. Mol Cell Biol *26*, 5797-5808.

Li, S., Czubryt, M. P., McAnally, J., Bassel-Duby, R., Richardson, J. A., Wiebel, F. F., Nordheim, A., and Olson, E. N. (2005). Requirement for serum response factor for skeletal muscle growth and maturation revealed by tissue-specific gene deletion in mice. Proc Natl Acad Sci U S A *102*, 1082-1087.

Li, S., Wang, D. Z., Wang, Z., Richardson, J. A., and Olson, E. N. (2003). The serum response factor coactivator myocardin is required for vascular smooth muscle development. Proc Natl Acad Sci U S A *100*, 9366-9370. Epub 2003 Jul 9316.

Li, T., Stark, M. R., Johnson, A. D., and Wolberger, C. (1995). Crystal structure of the MATa1/MAT alpha 2 homeodomain heterodimer bound to DNA. Science *270*, 262-269.

Lim, F. L., Hayes, A., West, A. G., Pic-Taylor, A., Darieva, Z., Morgan, B. A., Oliver, S. G., and Sharrocks, A. D. (2003). Mcm1p-induced DNA bending regulates the formation of ternary transcription factor complexes. Mol Cell Biol *23*, 450-461.

Ling, Y., Lakey, J. H., Roberts, C. E., and Sharrocks, A. D. (1997). Molecular characterization of the B-box protein-protein interaction motif of the ETS-domain transcription factor Elk-1. EMBO J *16*, 2431-2440.

Ling, Y., West, A. G., Roberts, E. C., Lakey, J. H., and Sharrocks, A. D. (1998). Interaction of transcription factors with serum response factor. Identification of the Elk-1 binding surface. J Biol Chem *273*, 10506-10514.

Liu, H. W., Halayko, A. J., Fernandes, D. J., Harmon, G. S., McCauley, J. A., Kocieniewski, P., McConville, J., Fu, Y., Forsythe, S. M., Kogut, P., *et al.* (2003). The RhoA/Rho kinase pathway regulates nuclear localization of serum response factor. Am J Respir Cell Mol Biol *29*, 39-47.

Lopez, M., Oettgen, P., Akbarali, Y., Dendorfer, U., and Libermann, T. A. (1994). ERP, a new member of the ets transcription factor/oncoprotein family: cloning, characterization, and differential expression during B-lymphocyte development. Mol Cell Biol *14*, 3292-3309.

Luo, J., Li, M., Tang, Y., Laszkowska, M., Roeder, R. G., and Gu, W. (2004). Acetylation of p53 augments its site-specific DNA binding both in vitro and in vivo. Proc Natl Acad Sci U S A *101*, 2259-2264.

Ma, Z., Morris, S. W., Valentine, V., Li, M., Herbrick, J. A., Cui, X., Bouman, D., Li, Y., Mehta, P. K., Nizetic, D., *et al.* (2001). Fusion of two novel genes, RBM15 and MKL1, in the t(1;22)(p13;q13) of acute megakaryoblastic leukemia. Nat Genet *28*, 220-221.

Mack, C. P., and Owens, G. K. (1999). Regulation of smooth muscle alpha-actin expression in vivo is dependent on CArG elements within the 5' and first intron promoter regions. Circ Res *84*, 852-861.

Maine, G. T., Sinha, P., and Tye, B. K. (1984). Mutants of S. cerevisiae defective in the maintenance of minichromosomes. Genetics *106*, 365-385.

Maira, S. M., Wurtz, J. M., and Wasylyk, B. (1996). Net (ERP/SAP2) one of the Ras-inducible TCFs, has a novel inhibitory domain with resemblance to the helix-loop-helix motif. Embo J *15*, 5849-5865.

Malik, S., and Roeder, R. G. (2005). Dynamic regulation of pol II transcription by the mammalian Mediator complex. Trends Biochem Sci *30*, 256-263.

Marais, R., Wynne, J., and Treisman, R. (1993). The SRF accessory protein Elk-1 contains a growth factor-regulated transcriptional activation domain. Cell *73*, 381-393.

Marais, R. M., Hsuan, J. J., McGuigan, C., Wynne, J., and Treisman, R. (1992). Casein kinase II phosphorylation increases the rate of serum response factor-binding site exchange. Embo J *11*, 97-105.

Maru, Y., Afar, D. E., Witte, O. N., and Shibuya, M. (1996). The dimerization property of glutathione S-transferase partially reactivates Bcr-Abl lacking the oligomerization domain. J Biol Chem *271*, 15353-15357.

Massague, J., Seoane, J., and Wotton, D. (2005). Smad transcription factors. Genes Dev *19*, 2783-2810.

Matsuzaki, K., Minami, T., Tojo, M., Honda, Y., Uchimura, Y., Saitoh, H., Yasuda, H., Nagahiro, S., Saya, H., and Nakao, M. (2003). Serum response factor is modulated by the SUMO-1 conjugation system. Biochem Biophys Res Commun *306*, 32-38.

Mead, J., Bruning, A. R., Gill, M. K., Steiner, A. M., Acton, T. B., and Vershon, A. K. (2002). Interactions of the Mcm1 MADS box protein with cofactors that regulate mating in yeast. Mol Cell Biol *22*, 4607-4621.

Mead, J., Zhong, H., Acton, T. B., and Vershon, A. K. (1996). The yeast alpha2 and Mcm1 proteins interact through a region similar to a motif found in homeodomain proteins of higher eukaryotes. Mol Cell Biol *16*, 2135-2143.

Meierhans, D., Sieber, M., and Allemann, R. K. (1997). High affinity binding of MEF-2C correlates with DNA bending. Nucleic Acids Res *25*, 4537-4544.

Mercher, T., Coniat, M. B., Monni, R., Mauchauffe, M., Khac, F. N., Gressin, L., Mugneret, F., Leblanc, T., Dastugue, N., Berger, R., and Bernard, O. A. (2001). Involvement of a human gene related to the Drosophila spen gene in the recurrent t(1;22) translocation of acute megakaryocytic leukemia. Proc Natl Acad Sci U S A *98*, 5776-5779.

Merika, M., and Thanos, D. (2001). Enhanceosomes. Curr Opin Genet Dev *11*, 205-208.

Messenguy, F., and Dubois, E. (2003). Role of MADS box proteins and their cofactors in combinatorial control of gene expression and cell development. Gene *316*, 1-21.

Miano, J. M. (2003). Serum response factor: toggling between disparate programs of gene expression. J Mol Cell Cardiol *35*, 577-593.

Miano, J. M., Ramanan, N., Georger, M. A., de Mesy Bentley, K. L., Emerson, R. L., Balza, R. O., Jr., Xiao, Q., Weiler, H., Ginty, D. D., and Misra, R. P. (2004). Restricted inactivation of serum response factor to the cardiovascular system. Proc Natl Acad Sci U S A *101*, 17132-17137.

Minty, A., and Kedes, L. (1986). Upstream regions of the human cardiac actin gene that modulate its transcription in muscle cells: presence of an evolutionarily conserved repeated motif. Mol Cell Biol *6*, 2125-2136.

Miralles, F., Posern, G., Zaromytidou, A. I., and Treisman, R. (2003). Actin dynamics control SRF activity by regulation of its coactivator MAL. Cell *113*, 329-342.

Miranti, C. K., Ginty, D. D., Huang, G., Chatila, T., and Greenberg, M. E. (1995). Calcium activates serum response factor-dependent transcription by a Ras- and Elk-1-independent mechanism that involves a Ca2+/calmodulin-dependent kinase. Mol Cell Biol *15*, 3672-3684.

Mo, Y., Ho, W., Johnston, K., and Marmorstein, R. (2001). Crystal structure of a ternary SAP-1/SRF/c-fos SRE DNA complex. J Mol Biol *314*, 495-506.

Mo, Y., Vaessen, B., Johnston, K., and Marmorstein, R. (1998). Structures of SAP-1 bound to DNA targets from the E74 and c-fos promoters: insights into DNA sequence discrimination by Ets proteins. Mol Cell *2*, 201-212.

Mo, Y., Vaessen, B., Johnston, K., and Marmorstein, R. (2000). Structure of the elk-1-DNA complex reveals how DNA-distal residues affect ETS domain recognition of DNA. Nat Struct Biol *7*, 292-297.

Mohun, T., Garrett, N., and Treisman, R. (1987). Xenopus cytoskeletal actin and human c-fos gene promoters share a conserved protein-binding site. Embo J *6*, 667-673.

Mohun, T. J., Chambers, A. E., Towers, N., and Taylor, M. V. (1991). Expression of genes encoding the transcription factor SRF during early development of Xenopus laevis: identification of a CArG box-binding activity as SRF. Embo J *10*, 933-940.

Molenaar, M., van de Wetering, M., Oosterwegel, M., Peterson-Maduro, J., Godsave, S., Korinek, V., Roose, J., Destree, O., and Clevers, H. (1996). XTcf-3 transcription factor mediates beta-catenin-induced axis formation in Xenopus embryos. Cell *86*, 391-399.

Molkentin, J. D., Black, B. L., Martin, J. F., and Olson, E. N. (1996a). Mutational analysis of the DNA binding, dimerization, and transcriptional activation domains of MEF2C. Mol Cell Biol *16*, 2627-2636.

Molkentin, J. D., Firulli, A. B., Black, B. L., Martin, J. F., Hustad, C. M., Copeland, N., Jenkins, N., Lyons, G., and Olson, E. N. (1996b). MEF2B is a potent transactivator expressed in early myogenic lineages. Mol Cell Biol *16*, 3814-3824.

Montagne, J., Groppe, J., Guillemin, K., Krasnow, M. A., Gehring, W. J., and Affolter, M. (1996). The Drosophila Serum Response Factor gene is required for the formation of intervein tissue of the wing and is allelic to blistered. Development *122*, 2589-2597.

Morin, S., Paradis, P., Aries, A., and Nemer, M. (2001). Serum response factor-GATA ternary complex required for nuclear signaling by a G-protein-coupled receptor. Mol Cell Biol *21*, 1036-1044.

Mueller, C. G., and Nordheim, A. (1991). A protein domain conserved between yeast MCM1 and human SRF directs ternary complex formation. Embo J *10*, 4219-4229.

Muller, J. M., Metzger, E., Greschik, H., Bosserhoff, A. K., Mercep, L., Buettner, R., and Schule, R. (2002). The transcriptional coactivator FHL2 transmits Rho signals from the cell membrane into the nucleus. Embo J *21*, 736-748.

Murai, K., and Treisman, R. (2002). Interaction of Serum Response Factor (SRF) with the Elk-1 B-Box inhibits RhoA-actin signalling to SRF and potentiates transcriptional activation by Elk-1. Mol Cell Biol *22*, 7083-7092.

Naar, A. M., Lemon, B. D., and Tjian, R. (2001). Transcriptional coactivator complexes. Annu Rev Biochem *70*, 475-501.

Nagy, L., and Schwabe, J. W. (2004). Mechanism of the nuclear receptor molecular switch. Trends Biochem Sci *29*, 317-324.

Nakagawa, K., and Kuzumaki, N. (2005). Transcriptional activity of megakaryoblastic leukemia 1 (MKL1) is repressed by SUMO modification. Genes Cells *10*, 835-850.

Narlikar, G. J., Fan, H. Y., and Kingston, R. E. (2002). Cooperation between complexes that regulate chromatin structure and transcription. Cell *108*, 475-487.

Natesan, S., and Gilman, M. (1995). YY1 facilitates the association of serum response factor with the c-fos serum response element. Mol Cell Biol *15*, 5975-5982.

Nishida, W., Nakamura, M., Mori, S., Takahashi, M., Ohkawa, Y., Tadokoro, S., Yoshida, K., Hiwada, K., Hayashi, K., and Sobue, K. (2002). A triad of serum response factor and the GATA and NK families governs the transcription of smooth and cardiac muscle genes. J Biol Chem *277*, 7308-7317.

Niu, Z., Yu, W., Zhang, S. X., Barron, M., Belaguli, N. S., Schneider, M. D., Parmacek, M., Nordheim, A., and Schwartz, R. J. (2005). Conditional mutagenesis of the murine serum response factor gene blocks cardiogenesis and the transcription of downstream gene targets. J Biol Chem *280*, 32531-32538.

Norman, C., Runswick, M., Pollock, R., and Treisman, R. (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.

Nurrish, S. J., and Treisman, R. (1995). DNA binding specificity determinants in MADS-box transcription factors. Mol Cell Biol *15*, 4076-4085.

Ogata, K., Sato, K., and Tahirov, T. H. (2003). Eukaryotic transcriptional regulatory complexes: cooperativity from near and afar. Curr Opin Struct Biol *13*, 40-48.

Oh, J., Richardson, J. A., and Olson, E. N. (2005). Requirement of myocardin-related transcription factor-B for remodeling of branchial arch arteries and smooth muscle differentiation. Proc Natl Acad Sci U S A *102*, 15122-15127.

Oh, J., Wang, Z., Wang, D. Z., Lien, C. L., Xing, W., and Olson, E. N. (2004). Target gene-specific modulation of myocardin activity by GATA transcription factors. Mol Cell Biol *24*, 8519-8528.

Okubo, S., Hara, F., Tsuchida, Y., Shimotakahara, S., Suzuki, S., Hatanaka, H., Yokoyama, S., Tanaka, H., Yasuda, H., and Shindo, H. (2004). NMR structure of the N-terminal domain of SUMO ligase PIAS1 and its interaction with tumor suppressor p53 and A/T-rich DNA oligomers. J Biol Chem *279*, 31455-31461.

Orphanides, G., and Reinberg, D. (2002). A unified theory of gene expression. Cell *108*, 439-451.

Parker, M. W., Lo Bello, M., and Federici, G. (1990). Crystallization of glutathione S-transferase from human placenta. J Mol Biol *213*, 221-222.

Parlakian, A., Tuil, D., Hamard, G., Tavernier, G., Hentzen, D., Concordet, J. P., Paulin, D., Li, Z., and Daegelen, D. (2004). Targeted inactivation of serum response factor in the developing heart results in myocardial defects and embryonic lethality. Mol Cell Biol *24*, 5281-5289.

Passmore, S., Elble, R., and Tye, B. K. (1989). A protein involved in minichromosome maintenance in yeast binds a transcriptional enhancer conserved in eukaryotes. Genes Dev *3*, 921-935.

Passmore, S., Maine, G. T., Elble, R., Christ, C., and Tye, B. K. (1988). Saccharomyces cerevisiae protein involved in plasmid maintenance is necessary for mating of MAT alpha cells. J Mol Biol *204*, 593-606.

Pellegrini, L., Tan, S., and Richmond, T. J. (1995). Structure of serum response factor core bound to DNA. Nature *376*, 490-498.

Pevny, L., Simon, M. C., Robertson, E., Klein, W. H., Tsai, S. F., D'Agati, V., Orkin, S. H., and Costantini, F. (1991). Erythroid differentiation in chimaeric mice blocked by a targeted mutation in the gene for transcription factor GATA-1. Nature *349*, 257-260.

Philippar, U., Schratt, G., Dieterich, C., Muller, J. M., Galgoczy, P., Engel, F. B., Keating, M. T., Gertler, F., Schule, R., Vingron, M., and Nordheim, A. (2004). The SRF target gene Fhl2 antagonizes RhoA/MAL-dependent activation of SRF. Mol Cell *16*, 867-880.

Pic, A., Lim, F. L., Ross, S. J., Veal, E. A., Johnson, A. L., Sultan, M. R., West, A. G., Johnston, L. H., Sharrocks, A. D., and Morgan, B. A. (2000). The forkhead protein Fkh2 is a component of the yeast cell cycle transcription factor SFF. Embo J *19*, 3750-3761.

Pipes, G. C., Creemers, E. E., and Olson, E. N. (2006). The myocardin family of transcriptional coactivators: versatile regulators of cell growth, migration, and myogenesis. Genes Dev *20*, 1545-1556.

Pipes, G. C., Sinha, S., Qi, X., Zhu, C. H., Gallardo, T. D., Shelton, J., Creemers, E. E., Sutherland, L., Richardson, J. A., Garry, D. J., *et al.* (2005). Stem cells and their derivatives can bypass the requirement of myocardin for smooth muscle gene expression. Dev Biol *288*, 502-513.

Pollock, R., and Treisman, R. (1990). A sensitive method for the determination of protein-DNA binding specificities. Nucleic Acids Res *18*, 6197-6204.

Pollock, R., and Treisman, R. (1991). Human SRF-related proteins: DNA-binding properties and potential regulatory targets. Genes Dev *5*, 2327-2341.

Posern, G., Miralles, F., Guettler, S., and Treisman, R. (2004). Mutant actins that stabilise F-actin use distinct mechanisms to activate the SRF coactivator MAL. Embo J *23*, 3973-3983.

Posern, G., Sotiropoulos, A., and Treisman, R. (2002). Mutant actins reveal a role for unpolymerised actin in control of transcription by Serum Response Factor. Mol Biol Cell *13*, 4167-4178.

Pramila, T., Miles, S., GuhaThakurta, D., Jemiolo, D., and Breeden, L. L. (2002). Conserved homeodomain proteins interact with MADS box protein Mcm1 to

restrict ECB-dependent transcription to the M/G1 phase of the cell cycle. Genes Dev *16*, 3034-3045.

Price, M. A., Rogers, A. E., and Treisman, R. (1995). Comparative analysis of the ternary complex factors Elk-1, SAP-1a and SAP-2 (ERP/NET). Embo J *14*, 2589-2601.

Primig, M., Winkler, H., and Ammerer, G. (1991). The DNA binding and oligomerization domain of MCM1 is sufficient for its interaction with other regulatory proteins. Embo J *10*, 4209-4218.

Prywes, R., and Roeder, R. G. (1986). Inducible binding of a factor to the c-fos enhancer. Cell *47*, 777-784.

Qiu, P., Feng, X. H., and Li, L. (2003). Interaction of Smad3 and SRF-associated complex mediates TGF-beta1 signals to regulate SM22 transcription during myofibroblast differentiation. J Mol Cell Cardiol *35*, 1407-1420.

Ramanan, N., Shen, Y., Sarsfield, S., Lemberger, T., Schutz, G., Linden, D. J., and Ginty, D. D. (2005). SRF mediates activity-induced gene expression and synaptic plasticity but not neuronal viability. Nat Neurosci *8*, 759-767.

Rao, V. N., Huebner, K., Isobe, M., ar-Rushdi, A., Croce, C. M., and Reddy, E. S. (1989). elk, tissue-specific ets-related genes on chromosomes X and 14 near translocation breakpoints. Science *244*, 66-70.

Reason, A. J., Morris, H. R., Panico, M., Marais, R., Treisman, R. H., Haltiwanger, R. S., Hart, G. W., Kelly, W. G., and Dell, A. (1992). Localization of O-GlcNAc modification on the serum response transcription factor. J Biol Chem *267*, 16911-16921.

Riechmann, J. L., Krizek, B. A., and Meyerowitz, E. M. (1996a). Dimerization specificity of Arabidopsis MADS domain homeotic proteins APETALA1, APETALA3, PISTILLATA, and AGAMOUS. Proc Natl Acad Sci U S A *93*, 4793-4798.

Riechmann, J. L., Wang, M., and Meyerowitz, E. M. (1996b). DNA-binding properties of Arabidopsis MADS domain homeotic proteins APETALA1, APETALA3, PISTILLATA and AGAMOUS. Nucleic Acids Res *24*, 3134-3141.

Rivera, V. M., Miranti, C. K., Misra, R. P., Ginty, D. D., Chen, R. H., Blenis, J., and Greenberg, M. E. (1993). A growth factor-induced kinase phosphorylates the serum response factor at a site that regulates its DNA-binding activity. Mol Cell Biol *13*, 6260-6273.

Roth, S., Stein, D., and Nusslein-Volhard, C. (1989). A gradient of nuclear localization of the dorsal protein determines dorsoventral pattern in the Drosophila embryo. Cell *59*, 1189-1202.

Rubinfeld, B., Albert, I., Porfiri, E., Fiol, C., Munemitsu, S., and Polakis, P. (1996). Binding of GSK3beta to the APC-beta-catenin complex and regulation of complex assembly. Science *272*, 1023-1026.

Rudensky, A. Y., Gavin, M., and Zheng, Y. (2006). FOXP3 and NFAT: partners in tolerance. Cell *126*, 253-256.

Santelli, E., and Richmond, T. J. (2000). Crystal structure of MEF2A core bound to DNA at 1.5 A resolution. J Mol Biol *297*, 437-449.

Santoro, I. M., and Walsh, K. (1991). Natural and synthetic DNA elements with the CArG motif differ in expression and protein-binding properties. Mol Cell Biol *11*, 6296-6305.

Sasazuki, T., Sawada, T., Sakon, S., Kitamura, T., Kishi, T., Okazaki, T., Katano, M., Tanaka, M., Watanabe, M., Yagita, H., *et al.* (2002). Identification of a novel transcriptional activator, BSAC, by a functional cloning to inhibit tumor necrosis factor-induced cell death. J Biol Chem *277*, 28853-28860. Epub 22002 May 28817.

Schratt, G., Philippar, U., Berger, J., Schwarz, H., Heidenreich, O., and Nordheim, A. (2002). Serum response factor is crucial for actin cytoskeletal organization and focal adhesion assembly in embryonic stem cells. J Cell Biol *156*, 737-750.

Schratt, G., Philippar, U., Hockemeyer, D., Schwarz, H., Alberti, S., and Nordheim, A. (2004). SRF regulates Bcl-2 expression and promotes cell survival during murine embryonic development. Embo J *23*, 1834-1844.

Schratt, G., Weinhold, B., Lundberg, A. S., Schuck, S., Berger, J., Schwarz, H., Weinberg, R. A., Ruther, U., and Nordheim, A. (2001). Serum response factor is required for immediate-early gene activation yet is dispensable for proliferation of embryonic stem cells. Mol Cell Biol *21*, 2933-2943.

Selvaraj, A., and Prywes, R. (2003a). Megakaryoblastic leukemia-1/2, a transcriptional co-activator of serum response factor, is required for skeletal myogenic differentiation. J Biol Chem *278*, 41977-41987.

Selvaraj, A., and Prywes, R. (2003b). Megakaryoblastic leukemia-1/2, a transcriptional co-activator of serum response factor, is required for skeletal myogenic differentiation. J Biol Chem *278*, 41977-41987. Epub 42003 Aug 41914.

Selvaraj, A., and Prywes, R. (2004). Expression profiling of serum inducible genes identifies a subset of SRF target genes that are MKL dependent. BMC Mol Biol *5*, 13.

Sepulveda, J. L., Vlahopoulos, S., Iyer, D., Belaguli, N., and Schwartz, R. J. (2002). Combinatorial expression of GATA4, Nkx2-5, and serum response factor directs early cardiac gene activity. J Biol Chem *277*, 25775-25782.

Sharrocks, A. D. (2001). The ETS-domain transcription factor family. Nat Rev Mol Cell Biol *2*, 827-837.

Sharrocks, A. D. (2002). Complexities in ETS-domain transcription factor function and regulation: lessons from the TCF (ternary complex factor) subfamily. The Colworth Medal Lecture. Biochem Soc Trans *30*, 1-9.

Sharrocks, A. D., Gille, H., and Shaw, P. E. (1993a). Identification of amino acids essential for DNA binding and dimerization in p67SRF: implications for a novel DNA-binding motif. Mol Cell Biol *13*, 123-132.

Sharrocks, A. D., and Shore, P. (1995). DNA bending in the ternary nucleoprotein complex at the c-fos promoter. Nucleic Acids Res *23*, 2442-2449.

Sharrocks, A. D., von Hesler, F., and Shaw, P. E. (1993b). The identification of elements determining the different DNA binding specificities of the MADS box proteins p67SRF and RSRFC4. Nucleic Acids Res *21*, 215-221.

Shaw, P. E., Schroter, H., and Nordheim, A. (1989). The ability of a ternary complex to form over the serum response element correlates with serum inducibility of the human c-fos promoter. Cell *56*, 563-572.

Shimizu, R. T., Blank, R. S., Jervis, R., Lawrenz-Smith, S. C., and Owens, G. K. (1995). The smooth muscle alpha-actin gene promoter is differentially regulated in smooth muscle versus non-smooth muscle cells. J Biol Chem *270*, 7631-7643.

Shin, C. H., Liu, Z. P., Passier, R., Zhang, C. L., Wang, D. Z., Harris, T. M., Yamagishi, H., Richardson, J. A., Childs, G., and Olson, E. N. (2002). Modulation of cardiac growth and development by HOP, an unusual homeodomain protein. Cell *110*, 725-735.

Shore, P., and Sharrocks, A. D. (1994). The transcription factors Elk-1 and serum response factor interact by direct protein-protein contacts mediated by a short region of Elk-1. Mol Cell Biol *14*, 3283-3291.

Shore, P., and Sharrocks, A. D. (1995a). The ETS-domain transcription factors Elk-1 and SAP-1 exhibit differential DNA binding specificities. Nucleic Acids Res *23*, 4698-4706.

Shore, P., and Sharrocks, A. D. (1995b). The MADS-box family of transcription factors. Eur J Biochem *229*, 1-13.

Simon, K. J., Grueneberg, D. A., and Gilman, M. (1997). Protein and DNA contact surfaces that mediate the selective action of the Phox1 homeodomain at the c-fos serum response element. Mol Cell Biol *17*, 6653-6662.

Skarnes, W. C., Auerbach, B. A., and Joyner, A. L. (1992). A gene trap approach in mouse embryonic stem cells: the lacZ reported is activated by splicing, reflects endogenous gene expression, and is mutagenic in mice. Genes Dev *6*, 903-918.

Small, E. M., Warkman, A. S., Wang, D. Z., Sutherland, L. B., Olson, E. N., and Krieg, P. A. (2005). Myocardin is sufficient and necessary for cardiac gene expression in Xenopus. Development *132*, 987-997.

Sommer, H., Beltran, J. P., Huijser, P., Pape, H., Lonnig, W. E., Saedler, H., and Schwarz-Sommer, Z. (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.

Somogyi, K., and Rorth, P. (2004). Evidence for tension-based regulation of Drosophila MAL and SRF during invasive cell migration. Dev Cell *7*, 85-93.

Sotiropoulos, A., Gineitis, D., Copeland, J., and Treisman, R. (1999). Signal-regulated activation of serum response factor is mediated by changes in actin dynamics. Cell *98*, 159-169.

Spencer, J. A., and Misra, R. P. (1996). Expression of the serum response factor gene is regulated by serum response factor binding sites. J Biol Chem *271*, 16535-16543.

Stevens, J. L., Cantin, G. T., Wang, G., Shevchenko, A., Shevchenko, A., and Berk, A. J. (2002). Transcription control by E1A and MAP kinase pathway via Sur2 mediator subunit. Science *296*, 755-758.

Stinson, J., Inoue, T., Yates, P., Clancy, A., Norton, J. D., and Sharrocks, A. D. (2003). Regulation of TCF ETS-domain transcription factors by helix-loop-helix motifs. Nucleic Acids Res *31*, 4717-4728.

Sun, Q., Chen, G., Streb, J. W., Long, X., Yang, Y., Stoeckert, C. J., Jr., and Miano, J. M. (2006). Defining the mammalian CArGome. Genome Res *16*, 197-207.

Sun, Y., Boyd, K., Xu, W., Ma, J., Jackson, C. W., Fu, A., Shillingford, J. M., Robinson, G. W., Hennighausen, L., Hitzler, J. K., *et al.* (2006). Acute myeloid leukemia-associated Mkl1 (Mrtf-a) is a key regulator of mammary gland function. Mol Cell Biol *26*, 5809-5826.

Tan, S., Ammerer, G., and Richmond, T. J. (1988). Interactions of purified transcription factors: binding of yeast MAT alpha 1 and PRTF to cell type-specific, upstream activating sequences. Embo J *7*, 4255-4264.

Tan, S., and Richmond, T. J. (1998). Crystal structure of the yeast MATalpha2/MCM1/DNA ternary complex. Nature *391*, 660-666.

Taylor, M., Treisman, R., Garrett, N., and Mohun, T. (1989). Muscle-specific (CArG) and serum-responsive (SRE) promoter elements are functionally interchangeable in Xenopus embryos and mouse fibroblasts. Development *106*, 67-78.

Thanos, D., and Maniatis, T. (1995). Virus induction of human IFN beta gene expression requires the assembly of an enhanceosome. Cell *83*, 1091-1100.

Thiesen, H. J., and Bach, C. (1990). Target Detection Assay (TDA): a versatile procedure to determine DNA binding sites as demonstrated on SP1 protein. Nucleic Acids Res *18*, 3203-3209.

Thompson, J. F., and Landy, A. (1988). Empirical estimation of protein-induced DNA bending angles: applications to lambda site-specific recombination complexes. Nucleic Acids Res *16*, 9687-9705.

Treisman, R. (1985). Transient accumulation of c-fos RNA following serum stimulation requires a conserved 5' element and c-fos 3' sequences. Cell *42*, 889-902.

Treisman, R. (1986). Identification of a protein-binding site that mediates transcriptional response of the c-fos gene to serum factors. Cell *46*, 567-574.

Treisman, R. (1994). Ternary complex factors: growth factor regulated transcriptional activators. Curr Opin Genet Dev *4*, 96-101.

Treisman, R. (1995). Journey to the surface of the cell: Fos regulation and the SRE. Embo J *14*, 4905-4913.

Treisman, R., Marais, R., and Wynne, J. (1992). Spatial flexibility in ternary complexes between SRF and its accessory proteins. Embo J *11*, 4631-4640.

Tsai, E. Y., Falvo, J. V., Tsytsykova, A. V., Barczak, A. K., Reimold, A. M., Glimcher, L. H., Fenton, M. J., Gordon, D. C., Dunn, I. F., and Goldfeld, A. E. (2000). A lipopolysaccharide-specific enhancer complex involving Ets, Elk-1, Sp1, and CREB binding protein and p300 is recruited to the tumor necrosis factor alpha promoter in vivo. Mol Cell Biol *20*, 6084-6094.

Ueyama, T., Kasahara, H., Ishiwata, T., Nie, Q., and Izumo, S. (2003). Myocardin expression is regulated by Nkx2.5, and its function is required for cardiomyogenesis. Mol Cell Biol *23*, 9222-9232.

Vartiainen, M. K., Guettler, S., Larijani, B., and Treisman, R. (2006). Nuclear actin regulates dynamic subcellular localization and activity of the SRF cofactor MAL. submitted.

Verger, A., Perdomo, J., and Crossley, M. (2003). Modification with SUMO. A role in transcriptional regulation. EMBO Rep *4*, 137-142.

Vinson, C. R., Hai, T., and Boyd, S. M. (1993). Dimerization specificity of the leucine zipper-containing bZIP motif on DNA binding: prediction and rational design. Genes Dev *7*, 1047-1058.

Wang, D., Chang, P. S., Wang, Z., Sutherland, L., Richardson, J. A., Small, E., Krieg, P. A., and Olson, E. N. (2001). Activation of cardiac gene expression by myocardin, a transcriptional cofactor for serum response factor. Cell *105*, 851-862.

Wang, D. Z., Li, S., Hockemeyer, D., Sutherland, L., Wang, Z., Schratt, G., Richardson, J. A., Nordheim, A., and Olson, E. N. (2002). Potentiation of serum response factor activity by a family of myocardin-related transcription factors. Proc Natl Acad Sci U S A *99*, 14855-14860. Epub 12002 Oct 14823.

Wang, Z., Wang, D. Z., Hockemeyer, D., McAnally, J., Nordheim, A., and Olson, E. N. (2004). Myocardin and ternary complex factors compete for SRF to control smooth muscle gene expression. Nature *428*, 185-189.

Wang, Z., Wang, D. Z., Pipes, G. C., and Olson, E. N. (2003). Myocardin is a master regulator of smooth muscle gene expression. Proc Natl Acad Sci U S A *100*, 7129-7134. Epub 2003 May 7119.

Wasylyk, C., Criqui-Filipe, P., and Wasylyk, B. (2005). Sumoylation of the net inhibitory domain (NID) is stimulated by PIAS1 and has a negative effect on the transcriptional activity of Net. Oncogene *24*, 820-828.

Watson, D. K., Robinson, L., Hodge, D. R., Kola, I., Papas, T. S., and Seth, A. (1997). FLI1 and EWS-FLI1 function as ternary complex factors and ELK1 and SAP1a function as ternary and quaternary complex factors on the Egr1 promoter serum response elements. Oncogene *14*, 213-221.

Weinhold, B., Schratt, G., Arsenian, S., Berger, J., Kamino, K., Schwarz, H., Ruther, U., and Nordheim, A. (2000). Srf(-/-) ES cells display non-cell-autonomous impairment in mesodermal differentiation. Embo J *19*, 5835-5844.

West, A. G., Causier, B. E., Davies, B., and Sharrocks, A. D. (1998). DNA binding and dimerisation determinants of Antirrhinum majus MADS-box transcription factors. Nucleic Acids Res *26*, 5277-5287.

West, A. G., and Sharrocks, A. D. (1999). MADS-box transcription factors adopt alternative mechanisms for bending DNA. J Mol Biol *286*, 1311-1323.

West, A. G., Shore, P., and Sharrocks, A. D. (1997). DNA binding by MADS-box transcription factors: a molecular mechanism for differential DNA bending. Mol Cell Biol *17*, 2876-2887.

Wheaton, K., and Riabowol, K. (2004). Protein kinase C delta blocks immediate-early gene expression in senescent cells by inactivating serum response factor. Mol Cell Biol *24*, 7298-7311.

Woychik, N. A., and Hampsey, M. (2002). The RNA polymerase II machinery: structure illuminates function. Cell *108*, 453-463.

Wu, H. M., and Crothers, D. M. (1984). The locus of sequence-directed and protein-induced DNA bending. Nature *308*, 509-513.

Wu, Y., Borde, M., Heissmeyer, V., Feuerer, M., Lapan, A. D., Stroud, J. C., Bates, D. L., Guo, L., Han, A., Ziegler, S. F., *et al.* (2006). FOXP3 controls regulatory T cell function through cooperation with NFAT. Cell *126*, 375-387.

Wynne, J., and Treisman, R. (1992). SRF and MCM1 have related but distinct DNA binding specificities. Nucleic Acids Res *20*, 3297-3303.

Yamazaki, Y., Kubota, H., Nozaki, M., and Nagata, K. (2003). Transcriptional regulation of the cytosolic chaperonin theta subunit gene, Cctq, by Ets domain transcription factors Elk-1, Sap-1a, and Net in the absence of serum response factor. J Biol Chem *278*, 30642-30651.

Yang, N., Higuchi, O., Ohashi, K., Nagata, K., Wada, A., Kangawa, K., Nishida, E., and Mizuno, K. (1998). Cofilin phosphorylation by LIM-kinase 1 and its role in Rac-mediated actin reorganization. Nature *393*, 809-812.

Yang, S. H., Bumpass, D. C., Perkins, N. D., and Sharrocks, A. D. (2002). The ETS domain transcription factor Elk-1 contains a novel class of repression domain. Mol Cell Biol *22*, 5036-5046.

Yang, S. H., Jaffray, E., Hay, R. T., and Sharrocks, A. D. (2003a). Dynamic interplay of the SUMO and ERK pathways in regulating Elk-1 transcriptional activity. Mol Cell *12*, 63-74.

Yang, S. H., and Sharrocks, A. D. (2004). SUMO promotes HDAC-mediated transcriptional repression. Mol Cell *13*, 611-617.

Yang, S. H., Sharrocks, A. D., and Whitmarsh, A. J. (2003b). Transcriptional regulation by the MAP kinase signaling cascades. Gene *320*, 3-21.

Yang, S. H., Shore, P., Willingham, N., Lakey, J. H., and Sharrocks, A. D. (1999). The mechanism of phosphorylation-inducible activation of the ETS-domain transcription factor Elk-1. EMBO J *18*, 5666-5774.

Yang, S. H., Vickers, E., Brehm, A., Kouzarides, T., and Sharrocks, A. D. (2001). Temporal recruitment of the mSin3A-histone deacetylase corepressor complex to the ETS domain transcription factor Elk-1. Mol Cell Biol *21*, 2802-2814.

Yang, S. H., Whitmarsh, A. J., Davis, R. J., and Sharrocks, A. D. (1998a). Differential targeting of MAP kinases to the ETS-domain transcription factor Elk-1. Embo J *17*, 1740-1749.

Yang, S. H., Yates, P. R., Whitmarsh, A. J., Davis, R. J., and Sharrocks, A. D. (1998b). The Elk-1 ETS-domain transcription factor contains a mitogen-activated protein kinase targeting motif. Mol Cell Biol *18*, 710-720.

Yanofsky, M. F., Ma, H., Bowman, J. L., Drews, G. N., Feldmann, K. A., and Meyerowitz, E. M. (1990). The protein encoded by the Arabidopsis homeotic gene agamous resembles transcription factors. Nature *346*, 35-39.

Yin, F., and Herring, B. P. (2005). GATA-6 can act as a positive or negative regulator of smooth muscle-specific gene expression. J Biol Chem *280*, 4745-4752.

Yoshida, T., Sinha, S., Dandre, F., Wamhoff, B. R., Hoofnagle, M. H., Kremer, B. E., Wang, D. Z., Olson, E. N., and Owens, G. K. (2003). Myocardin is a key regulator of CArG-dependent transcription of multiple smooth muscle marker genes. Circ Res *92*, 856-864. Epub 2003 Mar 2027.

Yu, Y. T., Breitbart, R. E., Smoot, L. B., Lee, Y., Mahdavi, V., and Nadal-Ginard, B. (1992). Human myocyte-specific enhancer factor 2 comprises a group of tissue-restricted MADS box transcription factors. Genes Dev *6*, 1783-1798.

Yuan, Y. O., Stroke, I. L., and Fields, S. (1993). Coupling of cell identity to signal response in yeast: interaction between the alpha 1 and STE12 proteins. Genes Dev *7*, 1584-1597.

Zaromytidou, A. I., Miralles, F., and Treisman, R. (2006). MAL and ternary complex factor use different mechanisms to contact a common surface on the serum response factor DNA-binding domain. Mol Cell Biol *26*, 4134-4148.

Zhang, S. X., Garcia-Gras, E., Wycuff, D. R., Marriot, S. J., Kadeer, N., Yu, W., Olson, E. N., Garry, D. J., Parmacek, M. S., and Schwartz, R. J. (2005). Identification of direct serum-response factor gene targets during Me2SO-induced P19 cardiac cell differentiation. J Biol Chem *280*, 19115-19126.

Zhou, J., and Herring, B. P. (2005a). Mechanisms responsible for the promoter-specific effects of myocardin. J Biol Chem *280*, 10861-10869.

Zhou, J., Hu, G., and Herring, B. P. (2005b). Smooth muscle-specific genes are differentially sensitive to inhibition by Elk-1. Mol Cell Biol *25*, 9874-9885.

Zhu, G., Spellman, P. T., Volpe, T., Brown, P. O., Botstein, D., Davis, T. N., and Futcher, B. (2000). Two yeast forkhead genes regulate the cell cycle and pseudohyphal growth. Nature *406*, 90-94.

Yuan, Y. O., Stroke, I. L., and Fields, S. (1993). Coupling of cell identity to signal response in yeast: interaction between the alpha 1 and STE12 proteins. Genes Dev *7*, 1584-1597.

Zaromytidou, A. I., Miralles, F., and Treisman, R. (2006). MAL and ternary complex factor use different mechanisms to contact a common surface on the serum response factor DNA-binding domain. Mol Cell Biol *26*, 4134-4148.

Zhang, S. X., Garcia-Gras, E., Wycuff, D. R., Marriot, S. J., Kadeer, N., Yu, W., Olson, E. N., Garry, D. J., Parmacek, M. S., and Schwartz, R. J. (2005). Identification of direct serum-response factor gene targets during Me2SO-induced P19 cardiac cell differentiation. J Biol Chem *280*, 19115-19126.

Zhou, J., and Herring, B. P. (2005a). Mechanisms responsible for the promoter-specific effects of myocardin. J Biol Chem *280*, 10861-10869.

Zhou, J., Hu, G., and Herring, B. P. (2005b). Smooth muscle-specific genes are differentially sensitive to inhibition by Elk-1. Mol Cell Biol *25*, 9874-9885.

Zhu, G., Spellman, P. T., Volpe, T., Brown, P. O., Botstein, D., Davis, T. N., and Futcher, B. (2000). Two yeast forkhead genes regulate the cell cycle and pseudohyphal growth. Nature *406*, 90-94.