4698–4706 Nucleic Acids Research, 1995, Vol. 23, No. 22 

# The ETS-domain transcription factors Elk-1 and SAP-1 exhibit differential DNA binding specificities

**Paul Shore and Andrew D. Sharrocks***

Department of Biochemistry and Genetics, The Medical School, University of Newcastle upon Tyne, Newcastle upon Tyne, UK

Received July 10, 1995; Revised and Accepted October 17, 1995

## ABSTRACT

**The ETS DNA-binding domain is conserved amongst many eukaryotic transcription factors. ETS-domains bind differentially to specific DNA sites containing a central GGA trinucleotide motif. The nucleotides flanking this motif define the binding specificity of individual proteins. In this study we have investigated binding specificity of the ETS-domains from two members of the ternary complex factor (TCF) subfamily, Elk-1 and SAP-1. The ETS DNA-binding domains of Elk-1 (Elk-93) and SAP-1 (SAP-92) select similar sites from random pools of double stranded oligonucleotides based on the consensus sequence ACCGGAAGTR. However, SAP-92 shows a more relaxed binding site selectivity and binds efficiently to a greater spectrum of sites than does Elk-93. This more relaxed DNA binding site selectivity is most pronounced in nucleotides located on the 3′ side of the GGA motif. This differential DNA-binding specificity is also exhibited by longer TCF derivatives and, indeed by the full-length proteins. Our results suggest that the range of potential *in vivo* target sites for SAP-1 is likely to be greater than for Elk-1. We discuss our results in relation to other similar studies carried out with more divergent ETS-domains.**

## INTRODUCTION

The ETS-domain family of transcription factors is currently comprised of >30 members from organisms as diverse as sponges and humans (1,2). Members of this family were originally identified on the basis of a region of primary sequence homology with the protein product of the *ets-1* proto-oncogene (3). The conserved region, termed the ETS-domain, corresponds to the DNA-binding domain of these proteins (reviewed in refs 4,5). Minimal polypeptides (~86 amino acids) that encompass the ETS-domain bind DNA with both high affinity and specificity (reviewed in refs 4,5). It has been demonstrated that the ETS-domain of three representative members of this family consists of a mixture of α-helical and β-strand structural elements (6–8). Moreover, the 3-D structure of the ETS-domain from Fli-1 has been recently solved by NMR and shown to be a variation of the helix-turn-helix motif (9).

The vast majority of ETS-domain proteins bind to DNA sequences that contain a GGA trinucleotide motif. However specific residues flanking this motif are also required for high affinity sequence specific binding by individual family members (reviewed in refs 4,5). Indeed, a series of studies using pools of random double-stranded oligonucleotides to select high affinity binding sites have identified consensus sequences for Ets-1 (10–12); E74A (13) ER81, ER71, GABPα (14) and Fli-1 (15) which all contain a GGA motif. Furthermore, footprinting analyses indicate that Ets-1 makes DNA contacts over a 20 base pair region centred on the GGA motif. Major groove contacts are inferred between Ets-1 and the two guanine residues in this motif (12).

A sub-group of the ETS-domain family has been defined due to the N-terminal position and high primary sequence similarity exhibited by the ETS-domains of its members. This sub-group is referred to as the ternary complex factors (TCFs) due to the ability of its members to form complexes with SRF and derivatives of the *c-fos* SRE (16; reviewed in ref. 17). There are currently three known human TCFs, Elk-1 (18), SAP-1 (19) and SAP2 (20). Ternary complex factor homologues also exist in mice (16,21). The TCFs share three regions of primary sequence similarity; the N-terminal ETS DNA-binding domain (22,23), the B-box which mediates direct protein–protein interactions with SRF (24) and a C-terminal domain which is the target for MAP kinase signal transduction pathways (20,25–28). MAP kinase phosphorylation of the C-terminal domain is required for transcriptional activation by TCFs (20,27–31) and stimulation of their DNA-binding activity (28,32–34).

Deletion of the C-terminal phosphorylation domain allows TCFs to bind constitutively to SRF:SRE complexes (16,19–22,24). Further deletion of the B-box region blocks the ability of TCFs to form ternary complexes but does not compromise their ability to bind autonomously to DNA (19,22). Indeed, in the case of SAP-1 such a deletion removes a region which inhibits its autonomous binding to certain DNA sequences such as the *c-fos* SRE (19). In contrast, similar derivatives of Elk-1, in which residues located C-terminally to the ETS-domain are deleted, cannot bind efficiently to the *c-fos* SRE (22). However, these same, truncated derivatives of SAP-1 and Elk-1 can bind to other 'high affinity' ets-sites such as the *Drosophila* E74 site (22,24,31). The ETS-domain of SAP-1 can therefore recognise and efficiently bind sites that Elk-1 cannot, thereby

* To whom correspondence should be addressed

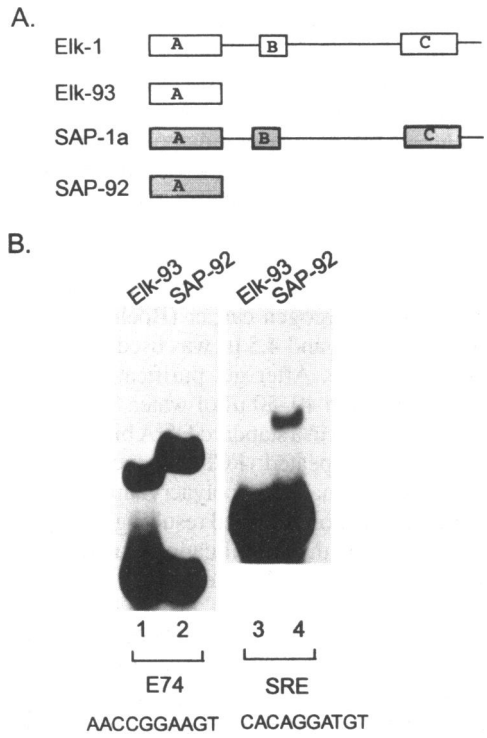


**Figure 1.** Elk-93 and SAP-92 show differential binding to the *c-fos* SRE. (**A**) Schematic representation of the domain structure of Elk-1 (open boxes) and SAP-1 (shaded boxes). The A (ETS DNA-binding domain), B (SRF interaction domain) and C (MAP kinase phosphorylation domain) boxes are shown. The regions encompassed by Elk-93 and SAP-92 are shown below the full-length proteins. (**B**) Gel retardation analysis of *in vitro* translated Elk-93 (lanes 1 and 3) and SAP-92 (lanes 2 and 4) binding to the *Drosophila* E74 site (lanes 1 and 2) or the *c-fos* SRE (lanes 3 and 4) equivalent amounts of each protein were used in each lane. The sequences flanking the central GGA motif in each binding site are shown below each set of lanes.

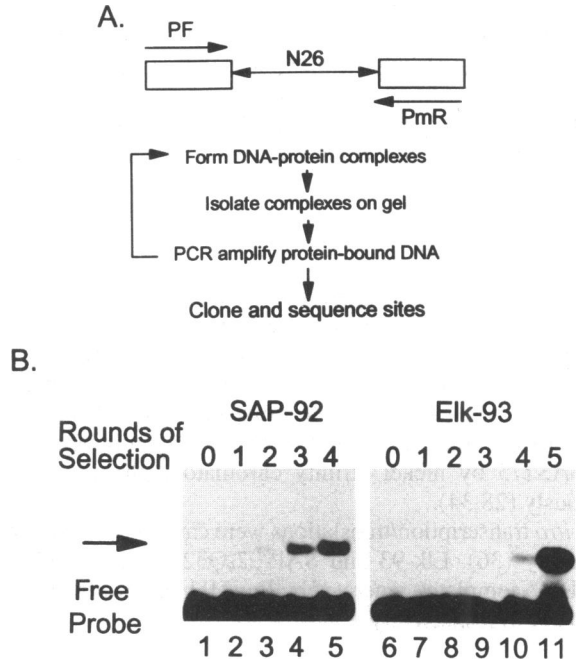


**Figure 2.** (**A**) Strategy for site selection. The structure of the degenerate oligonucleotide is shown. Flanking PCR primers PF and PmR are indicated by boxes. N26 indicates that the primers were separated by 26 random nucleotides. The steps utilised in the selection of binding sites are indicated below the oligonucleotide. (**B**) Gel retardation analysis of the selected pools of binding sites by bacterially expressed SAP-92 (lanes 2–5) and Elk-93 (lanes 7–11). The starting oligonucleotide pool is shown in lanes 1 and 6. The DNA probe represents the oligonucleotide pool after 0 (lanes 2 and 7), 1 (lanes 3 and 8), 2 (lanes 7 and 9), 3 (lanes 5 and 10) and 4 (lane 11) rounds of selection. The position of the protein–DNA complexes are indicated by an arrow. The DNA from the complexes in lane 4 (SAP-92 selected) or lanes 9 and 10 (Elk-93 selected) were amplified and cloned for sequence analysis.

suggesting that the two proteins have subtly different binding specificities.

In this study we have investigated the difference in the DNA-binding specificities of defined minimal ETS-domains of Elk-1 and SAP-1. Using a PCR-based site-selection procedure from random double-stranded oligonucleotide pools, we have derived consensus binding sequences for the Elk-1 and SAP-1 ETS-domains. The consensus binding sites selected by the ETS-domains of SAP-1 and Elk-1 are similar. However, the ETS-domain of SAP-1 shows a more relaxed binding specificity and binds efficiently to a greater spectrum of sequences than the corresponding ETS-domain of Elk-1. This differential DNA binding specificity is also exhibited by the full-length proteins. We discuss our results in relation to similar studies on the DNA-binding specificity of other ETS-domain proteins.

## MATERIALS AND METHODS

### Plasmids

*pAS74 and pAS72.* The plasmids pAS74 and pAS72 encode the ETS-domain of Elk-1 (Elk-93; amino acids 1–93) for expression in *E.coli,* as a fusion with GST or for synthesis by *in vitro* transcription/translation respectively and have been described previously (24).

*pAS183.* The sequence encoding the ETS-domain of SAP-1a (SAP-92; amino acids 1–92) was amplified from pT7 SAP-1a (19) using PCR with the primers ADS143, 5′-CGACGGATC-CGAATTCATCAGTTCAAAATCTCTGG-3′ and ADS167, 5′-CGACGGATCCATGGACAGTGCTATCAC-3′. The product was cleaved with *Bam*HI and *Eco*RI and ligated into the vector pGEX-2T (Pharmacia) cut with the same enzymes. The resultant plasmid encodes the N-terminal 92 amino acids of SAP-1 fused to glutathione S-transferase (GST) under the transcriptional control of the *tac* promoter for expression in *E.coli.*

*pAS136.* The plasmid pAS136 encodes the ETS-domain of SAP-1 (SAP-92 ; amino acids 1–92 with the mutation, Q32E) for synthesis by *in vitro* transcription/translation and has been described previously (35).

*pAS96.* The plasmid pAS96 encodes amino acids 1–168 from Elk-1 (Elk1-168) which encompass the ETS-domain and the B-box region (24).

*pAS168.* The sequence encoding the ETS-domain and B-box of SAP-1 (SAP1-156; amino acids 1–156) was amplified from pT7 SAP1a (19) using PCR with the primers ADS167, 5′-CGACG-

GATCCATGGACAGTGCTATCAC-3′ and ADS169, 5′-GCA-TAGATCTGAATTCATCAGTTCAAAGAGTTGAGAG-3′. The product was cleaved with *Nco*I and *Bgl*II and ligated into pAS37 (36) cut with *Nco*I and *Bam*HI.

pQE6/16Elk and pAS275 encode His-tagged full-length Elk-1 and SAP-1a under the control of the T5 and T7 promoters respectively (28,32,34).

### Protein purification

The ETS-domain polypeptides Elk-93 and SAP-92 were expressed in *E.coli* as fusion with GST and purified essentially as described previously for Elk-93 (8) with two modifications; the 750 mM NaCl wash and the final heparin affinity column steps were omitted. Full-length hexahistidine tagged Elk-1 and SAP-1a were purified from *E.coli* harboring the plasmids pQE6/16Elk and pAS275 by nickel affinity chromatography as described previously (28,34).

*In vitro* transcription/translations were carried out as described previously (36). Elk-93 and SAP-92(Q32E) were synthesised from DNA templates generated by *Bam*HI linearisation of pAS72 and pAS136 respectively.

### DNA-binding assays

Gel retardation assays were performed as described previously (37). Oligonucleotides used to construct the E74 and SRE binding sites have been described elsewhere (24). Radio-labelled DNA-binding sites for individual selected sites were synthesised using PCR essentially as described below except that ~10 ng of plasmid template containing the desired site was used in the PCR and 20–25 cycles were typically carried out. In standard reactions isolated ETS-domains were incubated with $^{32}$P-labelled DNA-binding sites in a total volume of 12 or 15 µl containing 42 or 125 mM KCl. Reactions were allowed to proceed for 20 min at room temperature prior to loading on a 5% non-denaturing polyacrylamide gel. Gels were fixed, dried and visualised by autoradiography or phosphorimaging (42S Phosphorimager, Molecular Dynamics). Quantification of DNA:protein complexes was achieved by analysis of phosphorimage data (ImageQuant software, Molecular Dynamics).

### DNA-binding site selection

*Binding sites.* Synthesis of a random pool of DNA-binding sites was carried out using PCR with the oligonucleotide R76 as a template and the primers PF (38) and PmR 5′-CAGGTCAGTT-CAGCGCATGCTGTCG-3′ in a 10 µl total reaction volume. R76 contains 26 random basepairs flanked by sequences complementary to PF and PR (38). PmR is identical to PR except that the *Bam*HI site has been destroyed. The reaction contained 0.5 ng of R76 oligonucleotide as template, 18 pmol of each of the primers PF and PmR, 5 µCi of [$^{32}$P]dCTP (3000 Ci/mmol), 200 µM dCTP, 500 µM each of dATP, dTTP, dGTP, 0.04% NP40, 1.5 mM $MgCl_2$, 1 µl 10 × Taq buffer (Promega) and 2.5 U of Taq polymerase (Promega). Amplification was performed for 15 cycles using the following conditions: 93°C, 1 min; 45°C, 1 min; 72°C, 3 min. The reaction mix was precipitated, resuspended in 20 µl of water and loaded onto a 10% polyacrylamide gel. Radio-labelled DNA-binding sites were purified from the gel by overnight elution into 400 µl 1 × TE followed by ethanol precipitation and resuspension in 40 µl of water.

*Selection procedure.* The first round of site selection was carried out with 2 µl of double-stranded $^{32}$P-labelled R76 in a standard DNA binding reaction with purified bacterially expressed Elk-93 or SAP-92. DNA–protein complexes were separated from unbound DNA on a 5% non-denaturing polyacrylamide gel. Gels were dried without fixing prior to being exposed to film. The position of the DNA:protein complexes was estimated and the corresponding area excised from the gel. The DNA from the complexes was then extracted by vigorous shaking in 400 µl of 1 × TE buffer for ~18 h. The DNA was ethanol precipitated in the presence of 10 µg of glycogen carrier (Boehringer) and resuspended in 10 µl of water and 4.5 µl was used as a template in a PCR as described above. After gel purification, the amplified DNA was resuspended in 40–50 µl of water. Two microlitres of this sample was then used in a standard DNA binding reaction and the selection process repeated. PCR products were routinely purified on 10% non-denaturing polyacrylamide gels. In some instances overamplification occurred resulting in single-stranded DNA. To overcome this the amplification was repeated using a lower number of cycles (10–12 cycles).

*Cloning selected sites.* One microlitre of DNA from the desired round of selection was taken directly from the PCR and ligated into pT7blue T-vector (Novagen). The ligation mixture was used to transform *E.coli* JM101. After incubation with Xgal/IPTG (39) white colonies were selected for plasmid isolation. Plasmids were analysed by restriction analysis and those that contained inserts of the desired length were selected for DNA sequence analysis. DNA sequencing was performed using the dideoxy chain-termination method.

## RESULTS

### The minimal ETS DNA-binding domains of Elk-1 and SAP-1 bind selectively to different sites

Truncated versions of SAP-1 which retain an intact ETS DNA-binding domain but lack the B-box domain, bind efficiently to the *c-fos* SRE (19) whereas similar deletion constructs of Elk-1 do not (22,24). Differences in the DNA sequences of the ets-motifs within the *c-fos* SRE and the E74 site (Fig. 1B) must determine this differential binding. In order to accurately assess the roles of the ETS-domains of SAP-1 and Elk-1 in mediating this differential DNA-binding selectivity, analogous regions of Elk-1 (amino acids 1–93; Elk-93) and SAP-1 (amino acids 1–92; SAP-92) (Fig. 1A) were expressed by either *in vitro* transcription/translation or as GST fusion proteins in *E.coli*. The GST fusion proteins were subsequently purified and the ETS-domains, Elk-93 and SAP-92, were cleaved from the GST moiety. *In vitro* translated Elk-93 and SAP-92 bound with similar affinities to the E74 site (Fig. 1B lanes 1 and 2). However, only SAP-92 bound efficiently to the *c-fos* SRE (Fig. 1B lane 4). No detectable Elk-93 binding to the *c-fos* SRE was observed (Fig. 1B lane 3). Similar results were obtained using the bacterially expressed ETS-domains (data not shown). These defined minimal ETS-domains therefore re-capitulate the DNA-binding selectivity exhibited by the various truncated derivatives previously examined (19,22).

### Selection of Elk-93 and SAP-92 binding sites

The DNA sequence requirements for the binding of Elk-93 and SAP-92 were investigated using a PCR based site-selection

**Table 1.** Sites selected by Elk-93: the sequences of DNA-binding sites selected by Elk-93 after four (A) or three (B) rounds of selection are shown

A

**Selected sites (Round 4)**

```
E13      tcgCGTCCGATAAGAACCGGAAGTAACCAgag
E17 tcgCAGTATCCGCCAGATAAACCGGAAGTAgag
E18           tcgCAAGAACCGGAAGTTGACTCACAGCCgag
E41    tcgCTGCACCACCACGAACCGGAAGTAATgag
E42             tcgGCAACCGGAAGTGAGACTGGAACTAAgag
E43          tcgGGAGCACAACCGGAAGTGCCTGATGCgag
E44        tcgTGCGAATTGAACCGGAAGTGATCGGAgag
E45             tcgCAACCGGAAGTTGACAGCGAGCCCCTgag
E46  tcgTGTGCTAGAAACGCGAACCGGAAGTAgag
E48    tcgAAGCTGGGGGGAAAACCGGAAGTAGGgag
E75 tcgGGAACTATAGCATCAGAACCGGAAGTgag
E76           tcgAAGGTGACCGGAAGTGCGTATAGAAAgag
E78        ctcTAGAGGCTAAACCGGAAGTGCTCAGGcga
E79           tcgAAGGTGACCGGAAGTAAAAAGCTAACgag
E81    tcgCGAGAGTTAGCGTAACCGGAAGTGATgag
```

**Summary of selected sites (Round 4)**

```
G  5  6  2  -  -  - 15 15  -  - 15  -  7  5  4
A  5  3 13 15  -  -  -  - 15 15  -  -  6  7  5
T  3  3  -  -  -  -  -  -  -  -  - 15  2  -  4
C  2  3  -  - 15 15  -  -  -  -  -  -  -  3  2
  -7 -6 -5 -4 -3 -2 -1  0 +1 +2 +3 +4 +5 +6 +7
Consensus:  A  A  C  C  G  G  A  A  G  T  R
```

B

**Selected sites (Round 3)**

```
E99              tcgAGAAGACCGGAAGCACTACGACAAATgag
E100             tcgTCATCAACCGGAAGTGATGATTGGAAgag
E102             tcgACGGATACCGGAAGTACATTAGCCCGgag
E105 tcgAGAGGGTACAAGCGCCAACCGGAAGTAgag
E106   tcgCCAAAATCAATGGATACCGGAAGTGAgag
E108          tcgAAAAGGAACCGGAAGTAACCCGAGCAgag
E122   ctcTTCTGATCTGATCTGACCGGAAGTGAGTGCTcga
E123      tcgACAACTGTAGAAACCGGAAGTAGTAAgag
E125            tcgTGCAAAACCGGAAGTGTTAAGAGCGAgag
E127            tcgAACACGACCGGAAGTTGAAACCCGAGgag
E128                 tcgACCGGAAGTCAGAGCAGGCTCTAGTAgag
```

**Summary of selected sites (Round 3)**

```
G  4  1  4  -  -  - 11 11  -  - 11  -  4  3  3
A  3  5  5 11  -  -  -  - 11 11  -  -  5  5  3
T  2  1  2  -  -  -  -  -  -  -  - 10  1  1  4
C  2  4  -  - 11 11  -  -  -  -  -  1  1  2  1
  -7 -6 -5 -4 -3 -2 -1  0 +1 +2 +3 +4 +5 +6 +7
Consensus:     A  C  C  G  G  A  A  G  T  R
```

Uppercase letters indicate nucleotides derived from the random sequence whereas lowercase letters are from the constant flanking primers. The central 6 base pairs of the ets binding motif are underlined in each sequence. Sites are aligned and orientated according to this motif. 88% of the sites were in the same orientation with respect to the flanking primers. After three rounds of selection, ~50% of sites contained no discernible ets motif (data not shown). Gel retardation analysis using several of these 'none-ets' sites showed no detectable binding of Elk-93 (data not shown). A summary of the selected sites is given below the sequences. Nucleotide positions within the sites are numbered with respect to the central guanine within the GGA trinucleotide motif (designated as '0'). Consensus sequences for sites after four (A) or three (B) rounds of selection are indicated. Nucleotides present in >80% of sites are shown as upper case letters. R represents a conserved purine residue (G or A).

**Table 2.** Sites selected by SAP-92: the sequences of DNA-binding sites selected by SAP-92 after three rounds of selection are shown

**Selected sites**

```
S1          tcgAAACGATCAACCGGAACACGGACACCgag
S2                tcgTAGACCGGATGCGATGCCAGTATCTCgag
S3       tcgTTCACAAGTAGCACCGGAAGTCAgag
S4      tcgGAGATTCCAGACAACCGGAAATCGTAgag
S5  ctcACTATCCCCCAATGGGACCGGAAGCACga
S25       ctcTGCGGATGCGCACCGGATATAGCACTcga
S26    tcgTACGGGTGCTCAACACCGGATGTAAGgag
S27       tcgAAAGAAGGCAAACCGGAAGTACCATCgag
S28 tcgACAACCATCAATTCCAGACCGGAAGCgag
S29         tcgGCAAATGTCCACCGGATGTACGgag
S30               tcgGTGAGCCGGAAATGTACCATACAAGGgag
S31 tcgTTTGATAAGAGATTGCTCCCGGAAGTgag
S32  tcgGGACGGCACACATCACACCGGAAGCGgag
S34                 tcgACACCGGAAGTAGATAAGTGTGTTAGgag
S35             tcgCCTTGACCGGAAGTGGGGCTTGGGTGgag
S63                   tcgACCGGAAGTCATAAAGCAGCATCACTgag
S64                   ctcTCCGGAAGCACAATAGCACCCATGCcga
S65  tcgTACCAAGGTAGACAAAGCCGGAAGTGgag
S67                  tcgTACCGGAAGTCAACATGAGCAGGAGTgag
S72            tcgAGGACAACCGGAAACGCCCATACAGCgag
```

**Summary of selected sites**

```
G  3  6  5  2  -  - 20 20  -  - 15  -  8  7  8
A  5  7  6 16  -  -  -  - 20 16  4  1  7  7  6
T  6  1  2  1  -  -  -  -  -  4  - 13  -  1  3
C  6  6  7  1 20 20  -  -  -  -  1  6  5  5  3
  -7 -6 -5 -4 -3 -2 -1  0 +1 +2 +3 +4 +5 +6 +7
Consensus:     A  C  C  G  G  A  At Ga Tc
```

Uppercase letter indicate nucleotides derived from the random sequence whereas lowercase letters are from the constant flanking primers. The central 6 base pairs of the ets binding motif are underlined in each sequence. Sites are aligned and orientated according to this motif. 85% of the sites were in the same orientation with respect to the flanking primers. A summary of the selected sites is given below the sequence. Nucleotide positions within the sites are numbered with respect to the central guanine within the GGA trinucleotide motif (designated as '0'). A consensus sequence for sites after three rounds of selection is indicated. Nucleotides present in >65% of sites are shown as uppercase letters. A second lower case letter indicates that this nucleotide is present in >80% of the remaining sites.

technique which purifies high affinity binding sites from a pool of random double-stranded oligonucleotides (38,40). The strategy used is shown in Figure 2A. Either Elk-93 or SAP-92 were incubated with a pool of double-stranded oligonucleotides containing 26 random bases (38). Protein–DNA complexes were separated from free DNA by gel retardation analysis and the protein-bound DNA was subsequently recovered from the gel and amplified using PCR. This procedure was repeated for a further 3 or 4 rounds. The resulting protein-bound DNA fragments were isolated and ligated into pT7-blue. Individual clones were then analysed by dideoxy sequencing. The site-selection procedure was monitored after each round by gel retardation analysis using isolated radio-labelled DNA as a probe. Enrichment of binding sites was apparant as an increased efficiency of complex formation as the number of rounds of selection increased for both Elk-93 and SAP-92 (Fig. 2B). After three (Fig. 2B lane 4)

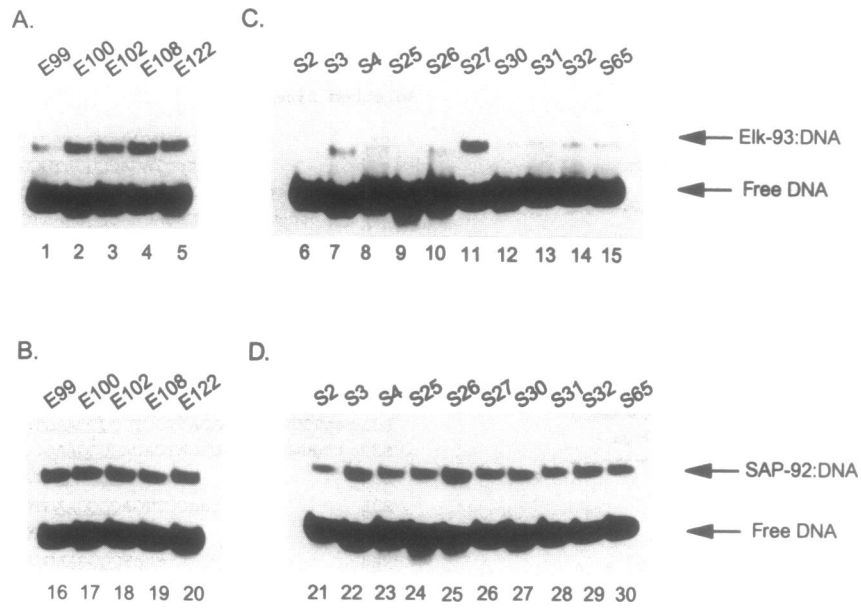


**Figure 3.** Binding of Elk-93 and SAP-92 to selected sites. The binding of *in vitro* translated Elk-93 (A and C ; lanes 1–15) and SAP-92 (B and D; lanes 16–30) to sites selected by Elk-93 (**A** and **B**) and SAP-92 (**C** and **D**) was analysed by gel retardation analysis. Lanes designated as S2-S65 contain SAP-92 selected sites whereas E99-E122 contain Elk-93 selected sites. Equal amounts of Elk-93 and SAP-92 were used in the binding reactions. For sequences of individual sites see Tables 1 and 2. The positions of the free DNA and DNA–protein complexes are indicated by arrows.

(SAP-92) or four rounds of selection (Fig. 2B lane 10) (Elk-93), efficient complex formation is observed.

The sequences of the sites selected by Elk-93 (after three and four rounds) and SAP-92 (after three rounds) are shown in Table 1 and Table 2, respectively. All the selected sequences contain the invariant GGA motif. We have numbered residues based on this motif, with the central guanine residue being designated as position '0'. After four rounds of selection (Table 1A), Elk-93 selects sites based on the consensus sequence AACCGGAAGTR of which the central nine residues are invariant. This consensus sequence is exactly conserved within the 'high affinity' *Drosophila* E74 site. In addition, an adenine located at the –5 position, is conserved in 13 of the 15 selected sequences.

We also analysed the sequences selected from the previous round (Table 1B) in an attempt to identify a more degenerate set of binding sites for Elk-93. The consensus sequence derived, ACCGGAAGTR, was virtually identical to that obtained after a further round of selection. In this case, eight nucleotides were invariant, with a slight increase in heterogeneity at the –5 and +4 positions. Such increases in heterogeneity are expected as the number of rounds of selection is decreased (38). Moreover, a high proportion of sequences with no obvious binding sites were obtained after three rounds of selection, to which no Elk-93 binding could be detected (data not shown). These results therefore indicate that the sites selected by Elk-93 in round four represent a true reflection of the sites which are efficiently bound by Elk-93.

After three rounds of selection, SAP-92 selects the consensus sequence ACCGGAA/tG/aT/c of which five nucleotides are invariant (C-3, C-2, G-1, G0 and A+1). This consensus represents a more relaxed version of the consensus sequence selected by Elk-93. Indeed, the majority of individual selected sites contain several deviations from the overall 10 base pairs of the Elk-93 consensus sequence with only one site conforming exactly to this sequence. Both Elk-93 and SAP-92 also show a degree of selectivity at positions +5 and +6 where thymines are strongly selected against. However, little selectivity is observed at the –6, –7 and +7 positions, indicating that sequence selectivity is limited to a contiguous 12 base pair region.

These results therefore suggest that Elk-93 and SAP-92 have similar DNA-binding specificities but in the case of SAP-92, the site-selectivity is more relaxed, enabling the protein to bind efficiently to a greater spectrum of sites.

## Binding of Elk-93 and SAP-92 to the selected sites

Several different sites selected by either SAP-92 or Elk-93 were analysed for binding by the two ETS-domains. These sites should all act as targets for the proteins which selected them. This is indeed the case, as both Elk-93 (Fig. 3A) and SAP-92 (Fig. 3D) efficiently bind to the sites they have selected. This confirms that the selected sites are a true reflection of the binding specificities of the two ETS-domains. However, only a subset of the sites selected by SAP-92 are bound efficiently by Elk-93 (Fig. 3C). In contrast SAP-92 efficiently binds to all the sites selected by Elk-93 that were tested (Fig. 3B).

The sequences of the sites selected by SAP-92 which are efficiently bound by Elk-93 (Fig. 3C) all conform closely to the consensus site for Elk-93. For example, the highest affinity site S27 conforms exactly to the 10 base pair consensus sequence AACCGGAAGT, whereas, S3 which binds Elk-93 with lower affinity, contains one deviation from this consensus sequence at the –5 position. Sites which contain two deviations such as S32 (at positions –5 and +4) bind with yet lower affinity. Sites which contain a thymine at the +2 position (as in the *c-fos* SRE) (Table 2) are very inefficiently bound by Elk-93 (S2, S25 and S26) (Fig. 3C). This is consistent with the observation that in

αcontrast to SAP-92, Elk-93 selects an invariant adenine at this position. However, SAP-92 can efficiently bind to all the sites tested which were selected by either itself (Fig. 3D) or by Elk-93 (Fig. 3B) including those with a thymine at position +2 (S2, S25 and S26). Differences in relative binding affinities for individual sites show a ~3-fold range for SAP-92 whereas Elk-93 binding affinities vary over several orders of magnitude. Indeed, binding of Elk-93 to the S2 site is not detectable. It is noteworthy that although the S27 site is the highest affinity binding site for Elk-93, S26 is the highest affinity site for SAP-92 (compare Fig. 3C and D). These results therefore further underline the notion that SAP-92 shows a wider DNA binding site selectivity than Elk-93.

In contrast to SAP-92, longer versions of SAP-1, in which sequences C-terminal to the ETS-domain are included, are incapable of binding to the *c-fos* SRE (19). We constructed a truncated SAP-1 derivative, SAP1-156, which contains the ETS-domain and C-terminal sequences up to the end of the B-box homology region (Fig. 4A). SAP1-156 is incapable of efficient autonomous binding to the *c-fos* SRE but efficiently binds to the *Drosophila* E74 site (19; our unpublished data). This is consistent with the hypothesis that the B-box inhibits binding to the *c-fos* SRE (19). However, as SAP1-156 efficiently binds to the 'high affinity' E74 site, this suggests that the B-box may play a role in modulating autonomous binding of the ETS-domains to a subset of sites. Moreover, it is possible that truncation of the TCFs immediately downstream from the ETS-domain may have artificially created domains with an altered DNA-binding specificity. Such effects have been shown to occur in other DNA binding proteins (e.g. 36,41). To test these hypotheses, we investigated the binding of Elk1-168 (24) and SAP1-156 to the sites selected by SAP-92.

SAP1-156 bound efficiently to all the sites that were selected by SAP-92 (Fig. 4C). As observed with SAP-92, the relative binding affinities for individual sites varied over a ~3-fold range. In contrast, Elk1-168 efficiently bound to a subset of the sites selected by SAP-92 (Fig 4B). The relative binding affinities of Elk1-168 to individual sites were virtually identical to those observed with Elk-93 (compare Fig. 4B with Fig. 3C). Binding affinities for Elk1-168 again ranged over several orders of magnitude between the S2 site (not detectable) and the S27 site (equivalent to binding of SAP1-156). Furthermore, S27 was the highest affinity site for Elk1-168 whereas S26 was the highest affinity site for SAP1-156.

It is possible that further C-terminal regions of the TCFs may affect their DNA binding specificity by intramolecular interactions with the ETS-domain. To test this hypothesis, the binding of full-length Elk-1 and SAP-1a was tested on a series of sites that had been selected by SAP-92 (Fig. 5). Sites were chosen which bound Elk-93 with varying efficiency but which were all bound efficiently by SAP-92 (Fig. 3). Full-length SAP-1a bound efficiently to all the sites tested (Fig. 5, lanes 1–5). In contrast, Elk-1 only bound efficiently to the S27 site (Fig. 5, lanes 6–10). These results essentially mirror those obtained with the isolated ETS-domains and SAP1-156/Elk1-168 in which all the SAP-92 selected sites represent efficient SAP-1 targets whereas only S27 is an efficient Elk-1 target (see Figs 3 and 4).

These results indicate that the sites selected by the ETS-domains of SAP-1 and Elk-1 represent true high affinity targets for C-terminally extended proteins. The B-box of SAP-1 does not appear to modulate the binding of SAP1-156 to high affinity sites in contrast to its effect on the lower affinity *c-fos* SRE. This may

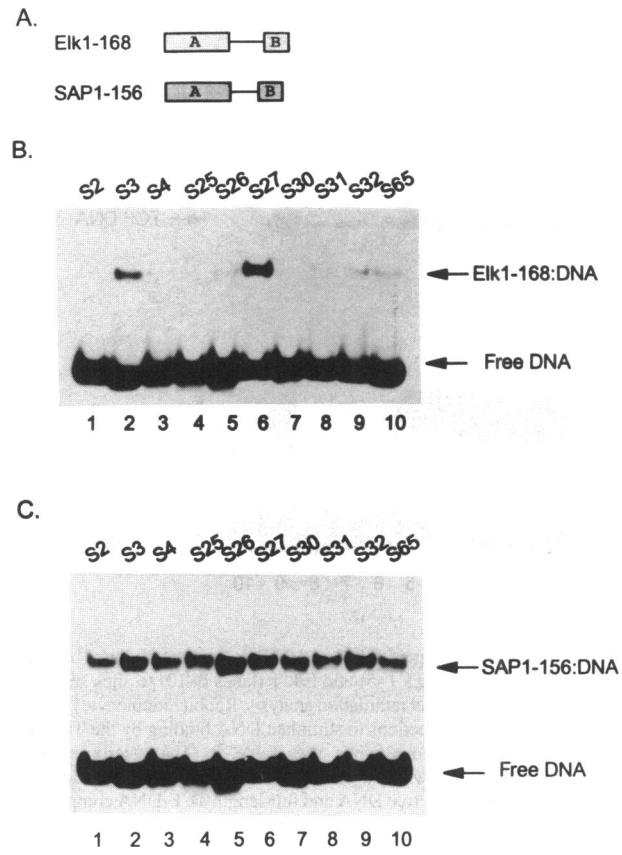


**Figure 4.** Binding of Elk1-168 and SAP1-156 to selected sites. (**A**) Schematic representation of the domain structure of Elk1-168 (open boxes) and SAP1-156 (shaded boxes). The A-box (DNA binding) and B-Box (SRF interaction) are indicated. (**B**) and (**C**) Gel retardation analysis of *in vitro* translated Elk1-168 (B; lanes 1–10) and SAP1-156 (C; lanes 1–10) on the binding sites selected by SAP-92. Equal amounts of Elk-168 and SAP1-156 were used in the binding reactions. The positions of the free DNA and protein–DNA complexes are indicated by arrows. The binding sites are the same as those used in Figure 3C and D.

reflect a role for the B-box in modulating autonomous SAP-1 binding to sites of lower affinity than those investigated in this study. Furthermore, full-length Elk-1 and SAP-1a also exhibit similar DNA binding specificities to those exhibited by their isolated ETS-domains.

## DISCUSSION

The ETS DNA-binding domain shows remarkable sequence conservation amongst species and amongst different family members. This sequence conservation is reflected by the observation that all ETS-domains appear to bind to sequences centred upon the GGA trinucleotide motif (reviewed in refs 4 and 5) with the flanking nucleotides determining the binding specificity of individual proteins. The ETS-domains of Elk-1 (Elk-93) and SAP-1 (SAP-92) are no exception to this rule. The amino acid sequences of Elk-93 and SAP-92 are 77% identical. Both proteins bind to similar sites based on the consensus sequence AAC-CGGAAGTR. However, SAP-92 displays a more relaxed binding specificity and will efficiently bind to a greater spectrum of sites than Elk-93. Indeed Elk-93 selects sites which do not differ by

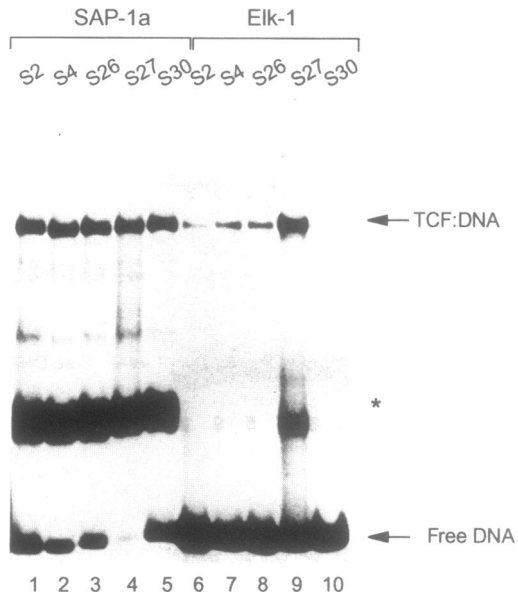


**Figure 5.** Binding of full-length Elk-1 and SAP-1a to selected sites. The binding of full-length SAP-1a (lanes 1–5) and Elk-1 (lanes 6–10) to sites selected by SAP-92 was analysed by gel retardation analysis. Rabbit reticulocyte lysate was included in the binding reactions to stimulate DNA binding by the TCFs (32). For sequences of individual sites see Table 2. The asterix represents DNA–protein complexes which probably contain a degradation product of the TCFs. The positions of the free DNA and full-length TCF-DNA complexes are indicated.

**Table 3.** Comparison of consensus sequences determined by DNA binding site selection using different ETS-domain proteins

| | -5 | -4 | -3 | -2 | -1 | 0 | +1 | +2 | +3 | +4 | +5 | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Elk-1 | A | A | C | C | G | G | A | A | G | T | Ga | (This study) |
| SAP-1a | N | A | C | C | G | G | A | At | Ga | Tc | N | (This study) |
| Ets-1 | N | A | Ca | Ca | G | G | A | AT | GA | T | T | (11) |
| Ets-1 | ? | GA | GC | Ca | G | G | A | A | G | T | TC | (10) |
| Ets-1 | ? | A | C | Ca | G | G | A | AT | Ga | TC | N | (12) |
| Fli-1 | GA | A | C | C | G | G | A | A | Ga | Tc | AG | (15) |
| E74A | A | A | CT | C | G | G | A | A | Ga | T | Ga | (13) |
| GABPα | ? | GA | Cg | Ca | G | G | A | At | Ga | Tc | N | (14) |
| ER81 | ? | Ga | Gc | Ca | G | G | A | AT | Ga | Tc | N | (14) |
| ER71 | ? | GC | CG | Ca | G | G | A | TA | Ga | Tc | C | (14) |

The consensus sequences are aligned using the conserved central GGA motif (boxed). Numbers indicating nucleotide positions are relative to the central guanine in this motif (designated as '0'). Uppercase letters indicates that a nucleotide contributes to the majority of sites selected. Two uppercase letters at a given position indicates that both nucleotides contribute approximately equally to the spectra of selected sites. One uppercase and one lower case letter indicates that these two nucleotides contribute to the majority of sites but the lowercase nucleotide is less favoured. Binding site selections were carried out with the following proteins and randomised oligonucleotide pools: Elk-1/SAP-1 ETS-domains (26 random nucleotides) (this study); full length Ets-1 (26 random nucleotides) (11); full length Ets-1 (fixed central GGAA with five random nucleotides on each side) (10); Ets-1 ETS-domain (fixed central GGA with three random upstream nucleotides and four random downstream nucleotides) (12); Fli-1 C-terminally extended ETS-domain (amino acids 267–452) (26 random nucleotides) (15); full length E74A (25 random nucleotides) (13); GABPα, ER81 and ER71 ETS-domains (fixed central GA with four random nucleotides on each side) (14). The numbers of reference in which these studies were carried out are shown on the right.

more than one nucleotide from the central 9 base pairs of its consensus sequence and binds poorly to sites which differ more significantly from this sequence. SAP-92 in contrast consistently selects and efficiently binds to sites which differ by at least two nucleotides in the same region of the consensus sequence.

Elk-1 also binds to DNA in a secondary protein-dependent manner to form ternary complexes in conjunction with SRF (reviewed in ref. 17). In this mode of binding, Elk-1 utilises both DNA–protein (22) and protein–protein interactions (24) to target specific DNA sequences. Selection of binding sites for Elk-1 in ternary complexes with SRF identified the consensus sequence RC/aC/aGGAA/tRT/c (42). This consensus sequence represents a relaxed version of that selected by autonomously binding Elk-93 and closely resembles the consensus sequence selected by SAP-92. The selection of a thymine at position +2 is particularly noteworthy as this occurs in ternary complexes containing Elk-1 and binary complexes containing SAP-92. Elk-93 binds very weakly to sites containing a thymine at this position. The formation of ternary complexes with SRF therefore allows Elk-1 to efficiently bind to a greater spectrum of sites. Protein–protein interactions with SRF presumably compensate for the loss of affinity associated with the more degenerate Elk-1 binding sites within ternary complexes (42). An alternative but not exclusive hypothesis is that the interaction of Elk-1 with SRF may promote a conformational change in the ETS-domain which allows it to bind with a more relaxed specificity as exhibited by the SAP-1 ETS-domain.

A comparison of the consensus sequence selected by Elk-93 (Table 1) with the sequence of the *c-fos* SRE (Fig. 1B) reveals that there are only three differences (at positions –5, –2 and +2) within a 10 base pair region (–5 to +4). These differences are therefore likely to account for the inability of Elk-93 to bind the *c-fos* SRE. However, SAP-92 can bind efficiently to sites containing both a cytosine at position –5 and a thymine at position +2 as contained in the *c-fos* SRE (e.g. site S26; Table 2). In contrast, sites selected by SAP-92 contain an invariant cytosine at position -2 (Table 2). Thus SAP-92 binding to the *c-fos* SRE might be predicted from our results but with a lower affinity binding due to the lack of a cytosine at position –2. The ets-motif within the *c-fos* SRE therefore appears to be a derivative of a high affinity TCF binding site. Two hypotheses might explain why such a lower affinity motif is favoured. First, the requirement for concomitant binding to SRF may impart promoter-specific recognition. Secondly, a lower affinity site might allow a tighter regulation of DNA binding. This second hypothesis is consistent with the observation that MAP kinase phosphorylation of Elk-1 (28,33,34) and SAP-1 (our unpublished data) regulates their DNA binding activity.

A comparison of the consensus binding sites selected by Elk-93, SAP-92 and other ETS-domain proteins is shown in Table 3. Inspection of the consensus sequences indicates that they all contain a central GGA motif. In addition, similar nucleotide

preferences are observed in the flanking regions. The closest consensus to the Elk-1 selected sequence is that selected by the *Drosophila* E74A protein (13) whereas Ets-1 (11,12) and GABPα (14) selected similar sequences to SAP-1. One of the most striking differences is at the +2 position at which Elk-1, E74A (13) and Fli-1 (15) almost exclusively select adenine whereas SAP-1, Ets-1 (11, 12), GABPα, ER71 and ER81 (14) all select sites which contain either adenine or thymine at this position. In addition, the extent of the selected sequences varies amongst proteins. In general, selection occurs over an 11 base pair region centred on the GGA motif. However, SAP-1 clearly shows sequence selectivity over a narrower region which also appears to be the case for Ets-1 (11,12), GABPα and ER81 (14). This may also reflect a less stringent DNA binding specificity by these proteins in common with SAP-1. However, due to differences in the site-selection protocols, direct comparisons with SAP-1 are difficult as the stringency of the selection procedure used will clearly influence its outcome. It should also be emphasised that although the consensus sites for given ETS-domain proteins appear similar, the binding efficiency of proteins to individual sites may vary considerably. For example Elk-1 efficiently binds the S27 site but only poorly binds to the S2 site (Fig. 3C). However both sites are contained within the SAP-1 consensus sequence and are bound efficiently by SAP-1 (Table 3). It is likely that investigations of the binding of other ETS-domain proteins to a panel of binding sites will reveal similar differences. Moreover, binding sites are likely to exist which are either unique high-affinity targets for individual ETS-domain proteins or binding sites which are bound efficiently by several different proteins. Thus, within the cell it is likely that target sites exist which can potentially bind either individual ETS-domain proteins or a subset of this transcription factor family. However, DNA binding alone is unlikely to represent the sole targeting mechanisms for ETS-domain proteins as protein–protein interactions with other transcription factors clearly play an important role in promoter recognition *in vivo* (reviewed in refs 4,5).

The TCFs contain three domains which have been defined by both primary sequence homology and functional similarities (reviewed in ref. 17). The N-terminal ETS-domain dictates sequence-specific DNA binding whereas the inclusion of additional C-terminal residues which encompass the B-box domain allows binding with a bipartite specificity to DNA-bound SRF. The inclusion of the B-box apparently reduced the binding affinity of Elk-1 to high affinity sites (43) and dramatically reduces the binding of SAP-1 to lower affinity sites (19, our unpublished data). The C-terminal domain acts as a target for phosphorylation by MAP kinases (reviewed in refs 17,28,33) and regulates the binding of TCFs to DNA. Phosphorylation or deletion of this domain stimulates the autonomous and protein dependent DNA-binding of TCFs (20,28,33,34). Such C-terminally deleted constructs are constitutively targeted to the *c-fos* SRE *in vivo* (28,31, our unpublished data). Inclusion of the B-box however, did not significantly alter the binding of Elk-1 and SAP-1 to the selected sites, indicating that C-terminal residues do not affect the binding selectivity of the isolated ETS-domains. Moreover, full-length Elk-1 and SAP-1a also exhibit the same DNA binding site selectivity as their respective isolated ETS-domains. It therefore appears that the ETS-domains contain all the information required to specifically target the proteins to DNA although it remains possible that specific phosphorylation events may play an additional modulatory role. In the case of SAP-1, our results apparently contradict the observation that the SAP-1 B-box severely inhibits its autonomous DNA binding. However, the sites which we have investigated are all 'high-affinity' whereas the *c-fos* SRE is bound by SAP-1 with a significantly lower affinity [~10-fold lower, Fig. 1; compare lanes 2 (high affinity site) and 4 (low affinity site)]. This may reflect a role for the B-box in modulating SAP-1 binding to low affinity sites. As SAP-1 shows more promiscuous binding than Elk-1, such a regulatory role for the B-box could aid specific targeting *in vivo*.

In summary, we have demonstrated that SAP-1 and Elk-1 have similar binding specificities. However, SAP-1 clearly shows a less stringent selectivity. This may reflect that SAP-1 has a greater range of potential *in vivo* targets. It remains to be seen whether TCFs bind promoters in a secondary protein-independent manner. However, our data will aid in identifying potential high affinity target sites within promoters. In addition, promoters may be found in which the arrangement of binding sites is the inverse of that observed at the *c-fos* SRE that is a strong ets motif and weak CArG box. Such a scenario has been suggested for the *nur77* gene (44) in which TCF targeting would be the primary event followed by SRF recruitment. At such promoters, the different binding specificities of individual TCFs would be of paramount importance. Indeed it is likely that TCFs target a range of promoters *in vivo*. Further studies are required to assess the roles of TCFs within promoter contexts other than the *c-fos* SRE.

## ACKNOWLEDGEMENTS

We thank Margaret Bell for excellent technical assistance and Sheila Gibbs-Barton and Catherine Pyle for secretarial assistance. We are grateful to Richard Treisman for the SAP-1 cDNA. We also thank Bob Liddell for sequencing. We thank members of the laboratory for critical discussions on the manuscript. This work was supported by grants from the North of England Cancer Research Campaign.

## REFERENCES

1 Degnan,B.M., Degnan,S.M., Naganuma,T. and Morese,D.E. (1993) *Nucleic Acids Res.*, **21**, 3479–3484.
2 Laudet,V., Niel,C., Duterque-Coquillaud,M., Leprince,D. and Stehelin,D. (1993) *Biochem. Biophys. Res. Commun.*, **190**, 8–14.
3 Karim,F.D., Urness,L.D., Thummel,C.S., Klemsz,M.J., McKercher,S.R., Celada,A., Van Beueren,C., Maki,R.A., Gunther,C.V., Nye,J.A. and Graves,B.J. (1990) *Genes Dev.*, **4**, 1451–1453.
4 Janknecht,R. and Nordheim,A., (1993) *Biochem. Biophys. Acta*, **1155**, 346–356.
5 Wasylyk,B., Hahn,S.L. and Giovane,A., (1993) *Eur. J. Biochem.*, **211**, 7–18.
6 Donaldson,L.W., Peterson,J.M., Graves,B.J. and McIntosh,L.P. (1994) *Biochemistry*, **33**, 13 509–13 516.
7 Liang,H., Olejniczak,E.T., Mao,X., Nettesheim,D.G., Yu,L., Thompson,C.B. and Fesik, S.W. (1994) *Proc. Natl. Acad. Sci. USA*, **91**, 11 655–11 659.
8 Shore,P., Bisset,L., Lakey,J., Waltho,J.P., Virden,R. and Sharrocks,A.D. (1995) *J. Biol, Chem.*, **270**, 5805–5811.
9 Liang,H., Mao,X., Olejniczak,E.T., Nettesheim,D.G., Yu,L., Meadows,R.P. Thompson,C.B. and Fesik, S.W. (1994) *Nature Struct. Biol.*, **91**, 11 655–11 659.
10 Fisher,R.J., Koizumi,S., Kondoh,A., Mariano,J.M., Mavrothalassitis,E., Bhat,N.K. and Papas,T.S (1992) *J. Biol. Chem.*, **267**, 17 957–17 965.
11 Woods,D.B., Ghysdael,J. and Owen,M.J. (1992) *Nucleic Acids Res.*, **20**, 699–704.
12 Nye,J.A., Peterson,J.M., Gunther,C.V., Jonsen,M.D. and Graves,B.J. (1992) *Genes Dev.*, **6**, 975–990.
13 Urness,L.D. and Thummel,C.S. (1990) *Cell*, **63**, 47–61.
14 Brown,T.A. and McKnight,S.L. (1992) *Genes Dev.*, **6**, 2502–2512.

15 Mao,X., Miesfeldt,S., Yang,H., Leiden,J.M. and Thompson,C.B. (1994) *J. Biol. Chem.*, **269**, 18 216–18 222.
16 Giovane,A., Pintzas,A., Maira,S-M., Sobieszcuk,P. and Wasylyk,B. (1994) *Genes Dev.*, **8**, 1502–1513.
17 Treisman,R. (1994) *Curr. Opin. Gene Dev.*, **4**, 96–101.
18 Rao,V.N., Huebner,K., Isobe,M., Ab-Rushidi,A., Croe,C.M. and Reddy,E.S.P. (1989) *Science*, **244**, 66–70.
19 Dalton,S. and Treisman,R. (1992) *Cell*, **68**, 597–612.
20 Price,M.A., Rogers,A.E. and Treisman,R. (1995) *EMBO J.*, **14**, 2589–2601.
21 Lopez,M., Oettgen,P., Akbarali,Y., Dendorfer,U. and Liberman,T.A. (1994) *Mol. Cell. Biol.*, **14**, 3292–3309.
22 Janknecht,R and Nordheim,A (1992) *Nucleic Acids Res.*, **20**, 3317–3324.
23 Rao,V.N. and Reddy,E.S.P. (1992) *Oncogene*, **7**, 65–70.
24 Shore,P and Sharrocks,A.D. (1994) *Mol. Cell. Biol.*, **14**, 3283–3291.
25 Marais,R., Wynne,J. and Treisman,R. (1993) *Cell*, **73**, 381–393.
26 Zinck,R., Hipskind,R.A., Pingoud,V. and Nordheim,A. (1993) *EMBO J.*, **12**, 2377–2387.
27 Kortenjann,M., Thomae,O. and Shaw,P.E. (1994) *Mol. Cell. Biol.*, **14**, 4815–4824.
28 Whitmarsh,A.J., Shore,P., Sharrocks,A.D. and Davis,R.J. (1995) *Science*, **269**, 403–407.
29 Janknecht,R., Ernst,W.H., Pingoud,V. and Nordheim,A. (1993) *EMBO J.*, **12**, 5097–5104.
30 Hill,C.S., Marais,R., John,S., Wynne,J., Dalton,S. and Treisman,R. (1993) *Cell*, **73**, 395–406.
31 Janknecht,R., Ernst,W.H. and Nordheim,A. (1995) *Oncogene*, **10**, 1209–1216.
32 Gille,H., Sharrocks,A.D. and Shaw,P.E. (1992) *Nature*, **358**, 415–417.
33 Gille,H., Kortenjann,M., Thomae,O., Moomaw,C., Slaughter,C, Cobb,M.H. and Shaw,P.E. (1995) *EMBO J.*, **14**, 951–962.
34 Sharrocks,A.D. (1995) *FEBS Lett.*, **368**, 77–80.
35 Sharrocks,A.D. and Shore,P. (1995) *Nucleic Acids Res.*, **23**, 2442–2449.
36 Sharrocks,A.D., Von Hesler,F. and Shaw,P.E. (1993) *Nucleic Acids Res.*, **21**, 215–221.
37 Sharrocks,A.D., Gille,H. and Shaw,P.E. (1993) *Mol. Cell. Biol.*, **13**, 123–132.
38 Pollock,R. and Treisman,R. (1990) *Nucleic Acids Res.*, **18**, 6197–6204.
39 Sambrook,J., Fritsch,E.F. and Maniatis,T. (1990) *Molecular Cloning: A Laboratory Manual*. 2nd edn, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
40 Blackwell,T.K. and Weintraub,H. (1990) *Science*, **250**, 1104–1110.
41 Nurrish,S.J. and Treisman,R. (1995) *Mol. Cell. Biol.*, **15**, 4076–4085.
42 Treisman,R., Marais,R. and Wynne,J. (1992) *EMBO J.*, **11**, 4631–4640.
43 Janknecht,R., Zinck,R., Ernst,W.H. and Nordheim,A. (1994) *Oncogene*, **9**, 1273–1278.
44 Williams,G.T. and Lau,L. (1993) *Mol. Cell. Biol.*, **13**, 6124–6136