MOLECULAR AND CELLULAR BIOLOGY, July 1996, p. 3338–3349 Vol. 16, No. 7
0270-7306/96/$04.00+0

Determinants of DNA-Binding Specificity of ETS-Domain Transcription Factors

PAUL SHORE,[1]† ALAN J. WHITMARSH,[2] R. BHASKARAN,[3] ROGER J. DAVIS,[2]
JONATHON P. WALTHO,[3] AND ANDREW D. SHARROCKS[1]*

Department of Biochemistry and Genetics, The Medical School, University of Newcastle upon Tyne, Newcastle upon Tyne NE2 4HH,[1] and The Krebs Institute, Department of Molecular Biology and Biotechnology, University of Sheffield, Sheffield S10 2UH,[3] United Kingdom, and Program in Molecular Medicine, Department of Biochemistry and Molecular Biology, Howard Hughes Medical Institute, University of Massachusetts Medical School, Worcester, Massachusetts 01605[2]

Received 17 January 1996/Returned for modification 27 February 1996/Accepted 27 March 1996

Several mechanisms are employed by members of transcription factor families to achieve sequence-specific DNA recognition. In this study, we have investigated how members of the ETS-domain transcription factor family achieve such specificity. We have used the ternary complex factor (TCF) subfamily as an example. ERK2 mitogen-activated protein kinase stimulates serum response factor-dependent and autonomous DNA binding by the TCFs Elk-1 and SAP-1a. Phosphorylated Elk-1 and SAP-1a exhibit specificities of DNA binding similar to those of their isolated ETS domains. The ETS domains of Elk-1 and SAP-1a and SAP-2 exhibit related but distinct DNA-binding specificities. A single residue, D-69 (Elk-1) or V-68 (SAP-1), has been identified as the critical determinant for the differential binding specificities of Elk-1 and SAP-1a, and an additional residue, D-38 (Elk-1) or Q-37 (SAP-1), further modulates their DNA binding. Creation of mutations D38Q and D69V is sufficient to confer SAP-1a DNA-binding specificity upon Elk-1 and thereby allow it to bind to a greater spectrum of sites. Molecular modelling indicates that these two residues (D-38 and D-69) are located away from the DNA-binding interface of Elk-1. Our data suggest a mechanism in which these residues modulate DNA binding by influencing the interaction of other residues with DNA.

DNA-binding proteins are grouped into families based on the structure of their DNA-binding domains (reviewed in reference 34). Within each family, the primary sequence of the DNA-binding domain is highly conserved. This sequence conservation is often reflected by the similarity of the sites to which individual proteins bind. Differences between the sequences of the DNA-binding domains dictate the differential binding specificities of individual family members. Many transcription factors, including members of the bZIP, MADS-box, and helix-turn-helix families, utilize α-helices as the principal DNA-binding determinants. In some instances, single residues within α-helices dictate differential binding specificities (e.g., the homeodomain protein Bcd [13, 48; reviewed in reference 9]). Alternatively, residues located outside these helices can also dramatically alter DNA-binding specificities (e.g., serum response factor [SRF]/MEF2A [31, 43]). However, it is still unclear how different transcription factors achieve differential sequence recognition using a series of highly related DNA-binding motifs. Indeed, the ETS-domain transcription factors represent one such family.

The ETS-domain family of eukaryotic transcription factors is currently composed of more than 30 members from organisms as diverse as sponges and humans (5, 25). 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 (21). The conserved region, termed the ETS domain, corresponds to the DNA-binding domain of these proteins (reviewed in references 20 and 52). Minimal polypeptides (≈86 amino acids) that encompass the ETS domain bind DNA with high degrees of both affinity and specificity (reviewed in references 20 and 52). The ETS domains of three representative members of this family consist of a mixture of α-helical and β-strand structural elements (6, 27, 44). Moreover, the three-dimensional structures of the ETS domains from Fli-1 and Ets-1 have recently been determined and shown to be a variation of the winged helix-turn-helix motif (7, 26, 53). Protein-DNA contacts in the major groove appear to be mediated mainly by a recognition α-helix within this motif.

Footprinting analyses indicate that Ets-1 makes DNA contacts over a 20-bp region centered on a GGA trinucleotide motif. Major groove contacts between Ets-1 and the two guanine residues in this motif are inferred (32). Indeed, 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 references 20 and 52). Indeed, a series of studies using pools of random double-stranded oligonucleotides to select high-affinity binding sites has identified consensus sequences for Ets-1 (8, 32, 51, 55); ER81, ER71, and GABPα (2); Fli-1 (29); and Elk-1/SAP-1a (46), all of which contain a GGA motif. However, ETS-domain proteins select different DNA sequences surrounding this central trinucleotide motif. Individual ETS-domain proteins therefore possess distinct DNA-binding specificities. Residues within the recognition helix (1) and C terminal to this α-helix (47) determine, at least in part, the recognition specificity of Ets-1.

The ternary complex factors (TCFs) are a subgroup of the ETS-domain family and are defined by the presence of a conserved N-terminal ETS domain and the ability to form complexes with SRF and derivatives of the c-*fos* serum response element (SRE) (12; reviewed in reference 49). Three human TCFs have been identified, Elk-1 (37) SAP-1 (4), and SAP-2

* Corresponding author. Phone: 44-191 222 8800. Fax: 44-191 222 7424. Electronic mail address: a.d.sharrocks@ncl.ac.uk.

† Present address: The Walter and Eliza Hall of Medical Research, The Royal Melbourne Hospital, Melbourne, Victoria 3050, Australia.

(36). TCF homologs also exist in mice (12, 28). The TCFs share three regions of primary sequence similarity: the N-terminal ETS DNA-binding domain (19, 38); the B-box, which mediates direct protein-protein interactions with SRF (45); and a C-terminal domain which is the target for mitogen-activated protein (MAP) kinase signal transduction pathways (22, 30, 36, 54, 57). In the case of Elk-1, MAP kinase phosphorylation of the C-terminal domain is required for transcriptional activation by TCFs (14, 17, 18, 22, 36, 54) and stimulation of their DNA-binding activity (10, 11, 40, 54, 56). SAP-1a also appears to be a target for the extracellular signal-regulated kinase (ERK) pathway (17, 36, 54), whereas, to date, it is unclear which pathways target SAP-2. Members of both the ERK and Jun N-terminal kinase (JNK) MAP kinase subfamilies target Elk-1 (54, 56). Deletion of the C-terminal phosphorylation domain allows TCFs to bind constitutively to SRF-SRE complexes (4, 12, 19, 28, 36, 45). Further deletion of the B-box region blocks the ability of TCFs to form ternary complexes (4, 19). TCFs can also bind to DNA in an SRF-independent manner. In the case of SAP-1, deletion of the B-box augments its binding to certain DNA sequences, such as the c-*fos* SRE (4). In contrast, full-length and truncated Elk-1 derivatives cannot bind efficiently to the c-*fos* SRE (19). However, truncated derivatives of SAP-1a and Elk-1 can bind to other high-affinity ets sites, such as the *Drosophila* E74 site (17, 19, 45). Moreover, DNA-binding site selection demonstrates that the Elk-1 ETS domain exhibits a more stringent DNA-binding site specificity than the corresponding domain of SAP-1 (46). Elk-1 selects sites conforming to the consensus sequence ACCGGAAGTR, whereas SAP-1 selects the more degenerate consensus A(C/t)CG GA(A/t)(G/a)(T/c)N. The ETS domain of SAP-1 can therefore recognize and efficiently bind a series of sites that Elk-1 cannot, thereby indicating that the two proteins have different DNA-binding specificities.

In this study, we have investigated the molecular mechanisms which determine the differential binding specificities of the TCF ETS-domain proteins Elk-1 and SAP-1a. In common with Elk-1, SAP-1a is stimulated by ERK2 kinase phosphorylation to bind DNA in both an SRF-dependent and an SRF-independent manner. Moreover, active full-length, phosphorylated Elk-1 and SAP-1a exhibit DNA-binding specificities that are similar to those exhibited by their isolated ETS domains. A single amino acid (D-69 or V-68) modulates the DNA-binding specificity of Elk-1 and SAP-1a. An additional residue (D-38 or Q-37) further alters DNA binding by these proteins. Together, these two residues are sufficient to confer SAP-1a DNA-binding specificity on the Elk-1 ETS domain and thereby allow it to bind to a greater spectrum of sites. We propose a model in which these two residues indirectly modulate specific DNA contacts made by other amino acids. Similar mechanisms may be used by other ETS-domain proteins to modulate their DNA-binding specificities.

MATERIALS AND METHODS

Plasmid constructs. pQE6/16Elk and pAS275 encode His-tagged full-length Elk-1 and SAP-1a under the control of the T5 and T7 promoters, respectively (19, 40, 54). pAS96 encodes ElkAB (amino acids 1 to 168 of Elk-1), pAS168 encodes SAPAB (amino acids 1 to 156 from SAP-1a), and pAS72 encodes the ETS domain of Elk-1 (Elk-93; amino acids 1 to 93) (45). pAS136 encodes the ETS domain of SAP-1a (SAP-92RI; amino acids 1 to 92 with a Q-to-E mutation at position 32 [mutation Q32E]) (46). This plasmid contains a unique *Eco*RI cleavage site introduced at position 94 in the DNA sequence (numbering from the translational start).

(i) pAS185. The Elk-1 sequence from pAS96 (amino acids 93 to 168) was amplified with primers ADS115 and ADS170, and the resulting product was cleaved with *Bam*HI and *Eco*RI and ligated into the same sites in pAS168. The resulting plasmid, pAS185, encodes the chimeric protein SEAB.

(ii) pAS186. The Elk-1 sequence from pAS96 (amino acids 1 to 93) was amplified with primers ADS106 and ADS171, and the resulting product was cleaved with *Bam*HI and *Nco*I and ligated into the same sites in pAS168. The resulting plasmid, pAS186, encodes the chimeric protein ESAB.

(iii) pAS280. The *Eco*RI-*Nco*I fragment from pAS183 (encoding SAP-1 amino acids 1 to 92) (46) was ligated into pAS37 cleaved with the same enzymes.

(iv) pAS170. The SAP-2 sequence from pT7SAP2 (36) was amplified with primers ADS172 and ADS173, cleaved with *Nco*I and *Bam*HI, and ligated into pAS37 cleaved with the same enzymes. The resulting plasmid encodes the ETS domain of SAP-2 (SAP2-92; amino acids 1 to 92).

(v) pAS171. The SAP-2 sequence from pT7SAP2 (36) was amplified with primers T7 and ADS174, cleaved with *Nco*I and *Bgl*II, and ligated into pAS37 cleaved with *Nco*I and *Bam*HI. The resulting plasmid encodes the ETS domain and B box of SAP-2 (SAP2AB; amino acids 1 to 152).

(vi) pAS137. pAS72 was cleaved with *Eco*RI, and the resultant fragment was ligated into pAS136 cleaved with the same enzyme. The resulting plasmid, pAS137, encodes the chimeric ETS domain SEE. The first 32 amino acids are derived from SAP-1a, and the remaining 61 are from Elk-1.

(vii) pAS138. pAS136 was cleaved with *Eco*RI, and the resulting fragment was ligated into pAS72 cleaved with the same enzymes. The resultant plasmid, pAS138, encodes the chimeric ETS domain ESS. The first 32 amino acids are derived from Elk-1, and the remaining 60 are from SAP-1a.

(viii) pAS180. Plasmid pAS180 was constructed by using PCR overlap extension. The sequence that encodes the C-terminal 25 amino acids of the SAP-1 ETS domain was amplified from pAS136 by PCR with the oligonucleotide ADS136 and FOR primer. The amplified product was used in a second PCR with REVL primer and plasmid pAS137 as a template. The amplified product was cleaved with *Nco*I and *Bam*HI and ligated into pAS37 cleaved with the same enzymes. The resultant plasmid encodes the hybrid ETS domain SES. The 32 N-terminal amino acids and the 25 C-terminal amino acids are derived from SAP-1a. The intervening 36 residues originate from Elk-1.

All point mutations were introduced by a two-step PCR protocol using a mutagenic primer and the flanking oligonucleotides FOR and REVL as described previously (42). Details of the mutagenic primers can be supplied on request.

(ix) pAS155. pAS157 was cleaved with *Eco*RI, and the released fragment was ligated into pAS72 cut with *Eco*RI. The resultant plasmid encodes the Elk-1 ETS domain (MElk) with two point mutations, D38Q and D69V.

The following plasmids encode derivatives of the chimeric ETS domain SEE, containing single or double amino acid substitutions in the Elk-1-derived sequence.

(x) pAS171a to pAS174. Plasmids pAS171a to pAS174 were constructed by using the following mutagenic primers on the template pAS137, resulting in the indicated mutations: pAS171a, ADS161, D69V; pAS172, ADS160, S77N; pAS173, ADS159, T53P; and pAS174, ADS162, D38Q.

(xi) pAS175 and pAS176. pAS175 and pAS176 were constructed by using the mutagenic primers ADS159 and ADS160 on pAS171a as a template. The resultant proteins contain the two point mutations D69V-T53P and D69V-S77N.

(xii) pAS177. pAS177 was constructed by using the mutagenic primer ADS160 and pAS173 as a template. The resultant protein contains the mutations T53P and S77N.

(xiii) pAS157, pAS178, and pAS179. Plasmids pAS157, pAS178, and pAS179 were constructed by using pAS174 as a template and mutagenic primers ADS161, ADS164, and ADS160, respectively. The resultant plasmids encode proteins containing the following amino acid substitutions in the Elk-1-derived sequence: pAS157, D38Q and D69V; pAS178, D38Q and L48I; and pAS179, D38Q and S77N.

Plasmids derived from PCR mutagenesis were sequenced by the dideoxy termination method. Other plasmids were verified by restriction enzyme analysis.

Protein purification. The mutant Elk-1 ETS-domain polypeptide MElk-93 was expressed in *Escherichia coli* as a fusion protein with glutathione *S*-transferase and purified essentially as described previously for Elk-93 (44), 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 plasmids pQE6/16Elk and pAS275 by nickel affinity chromatography as described previously (40, 54). Bacterially expressed core^{SRF} was purified as described previously (45). Phosphorylation reactions were carried out in the presence of recombinant ERK2 as described previously (40). All in vitro-translated proteins were synthesized by sequential in vitro transcription and translation from plasmids encoding polypeptides derived from the pBsSK$^+$ derivative pAS37 (40) under the control of the T3 promoter unless stated otherwise. The following proteins were derived from the indicated plasmids which had been linearized with *Bam*HI: Elk-93, pAS72; SAP-92, pT7SAP-1a; SAP-92RI, pAS36; SAP2-92, pAS170; ElkAB, pAS96; MElk, pAS155; SEE, pAS137; ESS, pAS138; SES, pAS180; SEE(D69V), pAS171; SEE(S77N), pAS172; SEE(T53P), pAS173; SEE(D38Q), pAS174; SEE(D69V/T53P), pAS175; SEE(D69V/S77N), pAS176; SEE(T53P/S77N), pAS177; SEE(D32Q/D69V), pAS157; SEE(D32Q/L48I), pAS178; SEE(D32Q/S77N), pAS179; and Elk-93(D38Q/D69V), pAS155. The following proteins were derived from plasmids linearized with *Xba*I: SAPAB, pAS168; SEAB, pAS185; ESAB, pAS186; and SAP2AB, pAS171. SAP-92 was derived from pAS280 linearized with *Eco*RI.

DNA-binding assays. Gel retardation assays were performed as described previously (41). DNA-binding sites derived from synthetic oligonucleotides were

radiolabelled by incorporation of [α-^{32}P]dCTP using the Klenow fragment of DNA polymerase I according to standard protocols (39). Oligonucleotides used to construct the E74 and SRE binding sites have been described elsewhere (45). Mutant E74 binding sites were made by annealing the following complementary oligonucleotides: E74-WT, ADS153 5′-GATAACCGGAAGTAA-3′ and ADS154 5′-GTTACTTCCGGTTAT-3′; E74-M1, ADS177 5′-GATAACCGGATGTAA-3′ and ADS178 5′-GTTACATCCGGTTAT-3′; E74-M2, ADS175 5′-GATAACAG GAAGTAA-3′ and ADS176 5′-GTTACTTCCTGTTAT-3′; E74-M3, ADS185 5′-GATAAACGGAAGTAA-3′ and ADS184 5′-GTTACTTCCGTTTAT-3′; E74-M4, ADS179 5′-GATAACAGGATGTAA-3′ and ADS180 5′-GTTACATCCTGT TAT-3′; E74-M5, ADS191 5′-GATAAACGGATGTAA-3′ and ADS190 5′-GTTA CATCCGTTTAT-3′; and E74-M6, ADS193 5′-GATAAACGGATGCAA-3′ and ADS192 5′-GTTGCATCCGTTTAT-3′.

Individual DNA-binding sites derived from the site selection procedure were synthesized and radiolabelled by PCR as described previously (46). All DNA-binding sites were routinely purified on 10% nondenaturing polyacrylamide gels.

In standard reactions, proteins 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 to equilibrium for 20 min at room temperature before the mixtures were loaded on a 5% nondenaturing polyacrylamide gel. Competition assays were carried out as described previously (45). Gels were fixed, dried, and visualized by autoradiography or phosphorimaging (425S Phosphorimager; Molecular Dynamics). DNA-protein complexes were quantified by analysis of phosphorimage data (ImageQuant software; Molecular Dynamics). Absolute DNA-binding affinities were not calculated. However, all reactions are carried out to achieve binding of <50% of the DNA in DNA-protein complexes. Under these conditions, relative binding affinities within an experiment can be calculated by direct quantification of DNA-protein complexes. The scoring of these relative binding affinities is indicated in the figure legends.

DNA-binding site selection. DNA-binding sites were selected from a pool of random-sequence double-stranded oligonucleotides by using bacterially expressed and purified MElk as previously described (46). Individual binding sites were ligated into the T-vector pT7blue (Novagen) as described previously (46).

Molecular modelling. Two models for the Elk-93–DNA complex were built, the first using the sequence similarity between Elk-1 and Fli-1 (62% identical) based on the solution structure of Fli-1 (26) and the second using Ets-1 (56% identical) based on the solution structure of the Ets-1–DNA complex (53). Where differences occur between the proteins, new loop conformations were derived from the Brookhaven Protein Databank by using the program INSIGHT (Biosym/MSI). In the model based on Fli-1, restrained rigid-body molecular dynamics calculations were used to determine the relative positions of Elk-93 and the DNA. These calculations were carried out by using the program XPLOR (3) and structural constraints derived from the nuclear magnetic resonance study of Fli-1 (26). Protein residues corresponding to those for which intermolecular nuclear Overhauser effects were observed between Fli-1 and DNA (i.e., residues 7-283, 52-327, 55-330, 63-338, 66-341, 80-355, and 81-356; the former numbers are for Elk-93 residues and the latter are for Fli-1) were restrained to within 5 Å (0.5 nm) of the DNA molecule; the constraints were assumed to be between the hydrogen atoms of the protein and those of the central 8 bp of the DNA. The sum distance averaging option of XPLOR was used in the calculation since the exact atoms involved in the constraints were unknown. The molecular dynamics simulation was performed for 10 ps. The two residues central to this study, D-38 and D-69 of Elk-93, do not contact DNA in either model.

RESULTS

ERK2 stimulates sequence-specific DNA binding by Elk-1 and SAP-1. Elk-1 binds DNA in both an SRF-dependent and an SRF-independent manner (19, 54, 57). Ternary complex formation with SRF and the c-*fos* SRE is stimulated by phosphorylation of Elk-1 by both the ERK1/2 (10, 40, 54) and JNK (54) MAP kinases. In addition, phosphorylation of Elk-1 by these MAP kinases also stimulates autonomous binding of Elk-1 (40, 54). The ability of ERK2 to stimulate ternary complex formation and DNA binding by SAP-1a was therefore tested. Full-length SAP-1a was incubated with active ERK2 in the presence and absence of ATP, and the resulting proteins were tested for binding to either SRF-SRE complexes or the *Drosophila* E74 site (Fig. 1B). Both Elk-1 and SAP-1a are stimulated by ERK2 phosphorylation (approximately threefold) to form ternary complexes with SRF and the c-*fos* SRE (Fig. 1B, lanes 1 to 4). Moreover, ERK2 stimulates the formation of a lower-mobility ternary complex containing SAP-1a on a 134-bp c-*fos* SRE-containing promoter fragment (data not shown). In addition, ERK2 stimulates both Elk-1 and SAP-1a to bind autonomously to the E74 site (≈5- and 13-fold, respectively) (Fig. 1B, lanes 5 to 8). In contrast, JNK1 stimulates neither autonomous DNA binding nor ternary complex formation by SAP-1a (data not shown). This is consistent with the observation that SAP-1a is a good ERK2 substrate but a poor JNK1 substrate in vitro (54).

It has previously been demonstrated that the ETS domains of Elk-1 and SAP-1a exhibit different DNA-binding specificities (46). Full-length phosphorylated Elk-1 and SAP-1a were therefore tested for binding to a range of sites which had been selected from random double-stranded oligonucleotide pools by the SAP-1a ETS domain (46). Phosphorylated SAP-1a binds efficiently to all the sites tested (Fig. 1D, lanes 10, 12, 14, and 16). S2 was previously shown to be one of the weakest selected sites and to bind the ETS domain of SAP-1a with an approximately threefold-lower affinity than the S27 site (46). Full-length phosphorylated SAP-1a also binds with an approximately threefold-lower affinity (Fig. 1C). In contrast, phosphorylated Elk-1 binds to each of the sites with various efficiencies (Fig. 1D, lanes 2, 4, 6, and 8). Indeed, 2 orders of magnitude separate the relative affinities of binding of Elk-1 to the weak S2 site and the strong S27 site (Fig. 1C). The range of relative affinities of binding to each site closely mirrors the range observed with the isolated ETS domain of Elk-1 (46). A second, higher-mobility DNA-protein complex which represents a degradation product of the TCFs that contains the ETS domain but lacks the C-terminal phosphorylation domain is also observed (40, 54). Again, the range of binding affinities of the full-length TCFs is similar to those exhibited by these truncated proteins.

ERK2 clearly stimulates DNA binding of SAP-1a to the sites tested (5- to 20-fold). However, the stimulation of Elk-1 binding is more modest (1.2- to 3-fold). This is consistent with previous results showing that ternary complex formation between Elk-1, SRF, and a 134-bp SRE-containing probe is stimulated to a lesser extent than formation with a shorter SRE (54).

These results therefore demonstrate that ERK2 stimulates both autonomous and SRF-dependent DNA binding by Elk-1 and SAP-1a. Furthermore, they demonstrate that active phosphorylated full-length Elk-1 and SAP-1a show virtually identical spectra of relative affinities of binding of their respective isolated ETS domains to a series of sites. In addition, full-length SAP-1a binds efficiently to a more degenerate series of sites than does Elk-1.

The effect of the ETS domain on the strength of ternary complexes. The ETS DNA-binding domains of Elk-1 and SAP-1 exhibit differential DNA-binding specificities (4, 19, 46). However, the role of such differential binding within ternary TCF-SRF-SRE complexes is unclear. Ternary complex formation by Elk-1 and SAP-1a derivatives was therefore tested. Truncated versions of Elk-1 (ElkAB) and SAP-1 (SAP1AB) which contain the ETS (DNA-binding) and B-box (SRF interaction) domains were used (Fig. 2A). These recombinant proteins bind constitutively to binary SRF-SRE complexes in a MAP kinase-independent manner (4, 19, 46) and act as dominant-negative regulators of c-*fos* SRE transcriptional activity in vivo (17, 54; our unpublished data).

Ternary complexes between ElkAB or SAP1AB and binary SRF-SRE complexes were allowed to form. Subsequently, the strength of complex formation was assessed by competition with excess unlabelled E74 DNA-binding site. The E74 site acts as an efficient target for autonomous binding of ElkAB and SAP1AB (45; our unpublished data). ElkAB and SAP1AB formed ternary complexes with similar efficiencies (Fig. 2B, lanes 1 and 16). However, after addition of the competitor E74, ternary complexes containing ElkAB decay much faster

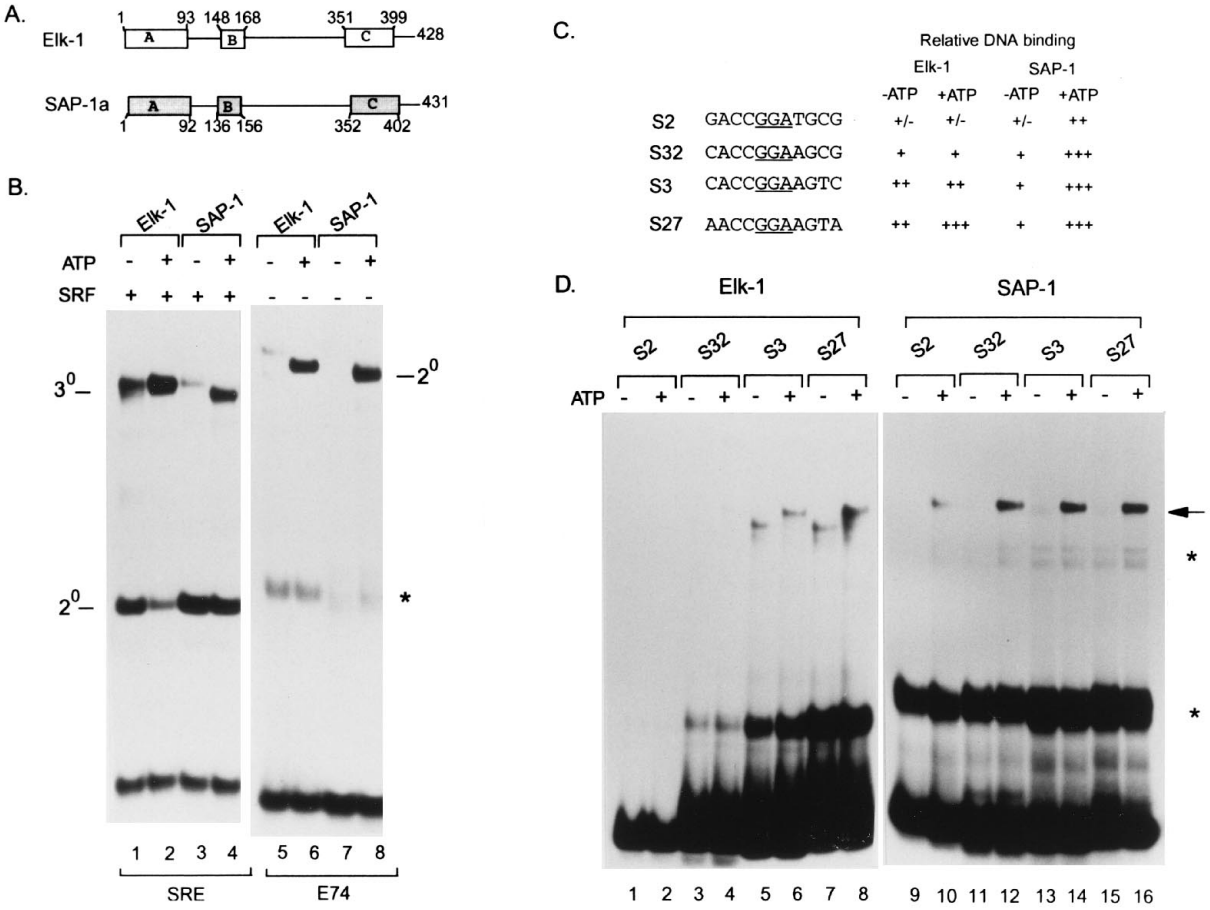

FIG. 1. ERK2 stimulates sequence-specific DNA binding by Elk-1 and SAP-1a. (A) Schematic representation of full-length Elk-1 and SAP-1a. Domains A (ETS DNA binding), B (SRF interaction), and C (MAP kinase phosphorylation) and the amino acids that define the endpoints of each domain are indicated. (B) Gel retardation analysis of Elk-1 and SAP-1a binding to SRF-SRE complexes (lanes 1 to 4) and the E74 site (lanes 5 to 8). Elk-1 and SAP-1a were incubated with active ERK2 in the absence (−) or presence (+) of ATP prior to incubation with DNA in the presence or absence of SRF, as indicated above the lanes. The locations of binary (2°) SRF-SRE and ternary (3°) TCF-SRF-SRE complexes are indicated on the left. The locations of binary (2°) TCF-E74 complexes are indicated on the right. The DNA-protein complex indicated by an asterisk is probably a degradation product of the TCFs. (C) Sequences of the 11 nucleotides centered on the conserved GGA motif (underlined) in SAP-1-selected sites. DNA-binding affinities relative to maximal SAP-1a binding (on the S27 site) or Elk-1 binding (on the S27 site) are indicated as follows: +++, 65 to 100% binding; ++, 25 to 65% binding; +, 5 to 25% binding; +/−, 1 to 5% binding; and −, <1% binding. (D) Gel retardation analysis of full-length Elk-1 (lanes 1 to 8) or full-length SAP-1a (lanes 9 to 16) binding to several SAP-1-selected DNA-binding sites (indicated above the lanes). Elk-1 and SAP-1a were incubated with active ERK2 in the absence (−) or presence (+) of ATP prior to incubation with DNA. The locations of the full-length TCF-DNA complexes (arrow) and DNA-protein complexes which probably contain a degradation product of the TCFs (asterisks) are indicated. The binding of full-length TCFs is quantified in panel C.

(half-life, <1 min) than those containing SAP1AB (half-life, >3 min) (Fig. 2B, compare lanes 1 to 5 with lanes 16 to 20). This indicates that ternary complexes containing SAP1AB are more stable than those containing ElkAB and is consistent with the observation that the SAP-1a ETS DNA-binding domain binds the SRE with higher affinity than the corresponding domain from Elk-1. It is therefore likely that the ETS domains of Elk-1 and SAP-1a dictate this differential binding stability. Alternatively, the interactions of the B-box regions of SAP-1a and Elk-1 with SRF may confer this differential ternary complex stability. To test these hypotheses, two chimeric proteins in which the ETS domains were exchanged between Elk-1 and SAP-1a were constructed (Fig. 2A).

ESAB (containing the Elk-1 ETS domain) and SEAB (containing the SAP-1 ETS domain) form ternary complexes with similar efficiencies (Fig. 2B, lanes 6 and 11). However, competition analysis demonstrates that ESAB decays rapidly at a rate similar to that of ElkAB (half-life, <1 min) (Fig. 2B, lanes 6 to 10), whereas SEAB decays more slowly, at a rate similar to that of SAP1AB (half-life, >3 min) (Fig. 2B, lanes 11 to 15). Therefore, as observed with the wild-type polypeptide, the chimera containing the SAP-1a ETS domain forms a stronger ternary complex. These results therefore demonstrate that the ETS domain of the TCFs rather than the B-box contributes to the strength of DNA binding within ternary complexes containing SRF and the c-*fos* SRE. Within this complex, the ETS domain of SAP-1a apparently mediates stronger ternary complex formation than the corresponding domain of Elk-1.

Identification of residues which mediate the differential DNA-binding specificities of Elk-1 and SAP-1a. Both full-length and truncated Elk-1 and SAP-1a derivatives show differential DNA-binding specificities. In addition, on the c-*fos* SRE, ternary complexes containing SAP-1 are stronger than those containing Elk-1. Taken together, these results indicate that the ETS domains play major roles in dictating DNA binding by TCFs.

The primary sequences of the Elk-1 and SAP-1a ETS domains differ by 20 amino acids. In order to identify the deter-

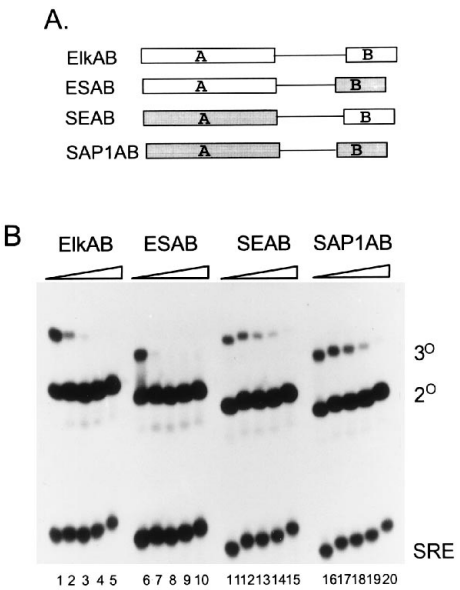

FIG. 2. The SAP-1 ETS domain mediates strong DNA binding in the ternary TCF-SRF-SRE complex. (A) Schematic representation of the Elk-1, SAP-1, and chimeric TCF derivatives which include the A (ETS DNA-binding) and B (SRF-binding) domains. Elk-1-derived domains (open boxes) and SAP-1-derived domains (shaded boxes) are indicated. (B) Competition gel retardation analysis of Elk-1, SAP-1, and chimeric TCF derivatives within ternary TCF-SRF-SRE complexes. Ternary complexes were allowed to form with ElkAB, ESAB, SEAB, or SAP1AB as indicated above the lanes. Equal molar quantities of each TCF derivative were used in DNA-binding reactions. A 150-fold molar excess of unlabelled E74 competitor DNA was subsequently added, and the resulting complexes were analyzed after 0 (lanes 1, 6, 11, and 16), 1 (lanes 2, 7, 12, and 17), 3 (lanes 3, 8, 13, and 18), 7 (lanes 4, 9, 14, and 19) or 15 (lanes 5, 10, 15, and 20) min. The positions of the binary (2°) (SRF-SRE) and ternary (3°) (TCF-SRF-SRE) complexes are indicated. The increase in time of competition is indicated schematically above each set of lanes.

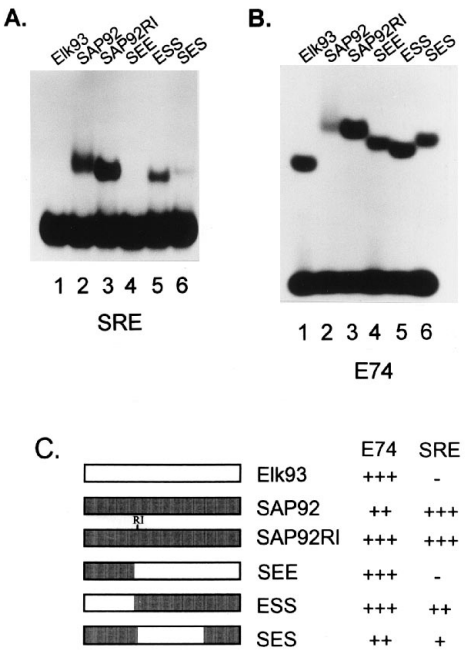

FIG. 3. Mapping the DNA-binding specificity determinants in the TCF ETS domain. (A and B) Gel retardation analysis of Elk-1, SAP-1, and chimeric ETS domains bound to the c-*fos* SRE and the E74 site, respectively. (C) Schematic diagram of the sequences contained within each chimeric molecule. Elk-1-derived sequences (open boxes) and SAP-1-derived sequences (shaded boxes) are indicated. SEE contains the N-terminal 32 amino acids of SAP-1 and the 60 C-terminal amino acids of the Elk-1 ETS domain. ESS contains the N-terminal 33 amino acids of Elk-1 and 60 C-terminal amino acids of the SAP-1 ETS domain. SES contains the N-terminal 32 amino acids and C-terminal 25 amino acids from the SAP-1 ETS domain and residues 34 to 68 from Elk-1. SAP-92 was synthesized from plasmid pT7SAP1 by runoff translation. Subsequent experiments used pAS280 to produce SAP-92, which contains translational stop codons. Relative DNA-binding affinities in comparison with SAP-92 binding to each site are shown as follows: +++, 65 to 100% binding; ++, 25 to 65% binding; +, 15 to 25% binding; and −, <1% binding.

minants of the differential DNA-binding specificities of Elk-1 and SAP-1a, we first constructed a series of chimeric ETS domains by exchanging homologous regions of these two polypeptides (Fig. 3C) and compared their binding with that of the wild-type proteins to the c-*fos* SRE and the *Drosophila* E74 site.

Both SAP-92 (containing SAP-1a amino acids 1 to 92) and SAP-92RI (containing the Q32E mutation) bind efficiently to the c-*fos* SRE (Fig. 3A, lanes 2 and 3), whereas Elk-93 (containing Elk-1 amino acids 1 to 93) binding to the SRE is not detectable (Fig. 3A, lane 1). The chimeric protein ESS also binds efficiently to the SRE, whereas SEE does not (Fig. 3A, lanes 4 and 5). These data therefore localize the specificity-determining residues to the C-terminal 60 amino acids of the ETS domain. To further define the SRE-binding specificity determinants, the nonbinding chimera SEE was further mutated by replacing the C-terminal 25 amino acids with the equivalent region from the SAP-1 ETS domain. The resulting chimera, SES, binds the SRE, albeit with reduced binding affinity compared with that of ESS (Fig. 3A, lane 6), indicating that the C-terminal 25 amino acids of the SAP-1 ETS-domain contribute to DNA binding. All these wild-type and chimeric polypeptides bind efficiently to the E74 site, indicating that all the proteins are active DNA-binding proteins (Fig. 3B). Taken together, these results indicate that a residue(s) within the C-terminal 25 amino acids of SAP-92 dictates binding to the c-*fos* SRE, but a residue(s) within the central region of the ETS domain is also required for maximal binding.

In order to identify residues which dictate the differential DNA-binding specificities of Elk-1 and SAP-1, a series of point mutations were made in the Elk-1 ETS domain. Each mutation converted an Elk-1 residue into the corresponding residue in SAP-1a. Mutations were created in the context of the SEE or SES chimeras, since a role for the N-terminal end of the ETS domain in stabilizing DNA binding could not be excluded. The binding of each of the mutant proteins was tested (Fig. 4B and C). The results are summarized in Fig. 4A. All single and double point mutants of SEE and SES efficiently bound the E74 site (Fig. 4C, lanes 1 to 13). In contrast, SRE binding could be detected only for polypeptides containing the D69V mutation. Single point mutations at other positions did not confer SRE binding upon the SEE polypeptide. Maximal binding to the SRE is obtained with the SEE chimeric protein containing the double mutation D38Q/D69V (Fig. 4B, lane 6). Of these two mutations, only D69V gave detectable DNA binding to the SRE (Fig. 4B, lane 3), suggesting that the two residues work in a combinatorial manner. SEE(T53P/D69V) and SES(L48I/T53P) both show significant SRE binding, albeit at a lower level than that exhibited by SEE(D38Q/D69V) (Fig. 4B, lanes 8 and 12). As both of these mutant proteins contain the T53P/D69V double mutation, this result indicates that the T53P mutation can also combine with the D69V mutation to provide significant SRE binding. A similar scenario can be envisaged for the mutant protein SEE(D69V/S77N), which also weakly binds the SRE (Fig. 4B, lane 10), in which S77N has a combi-

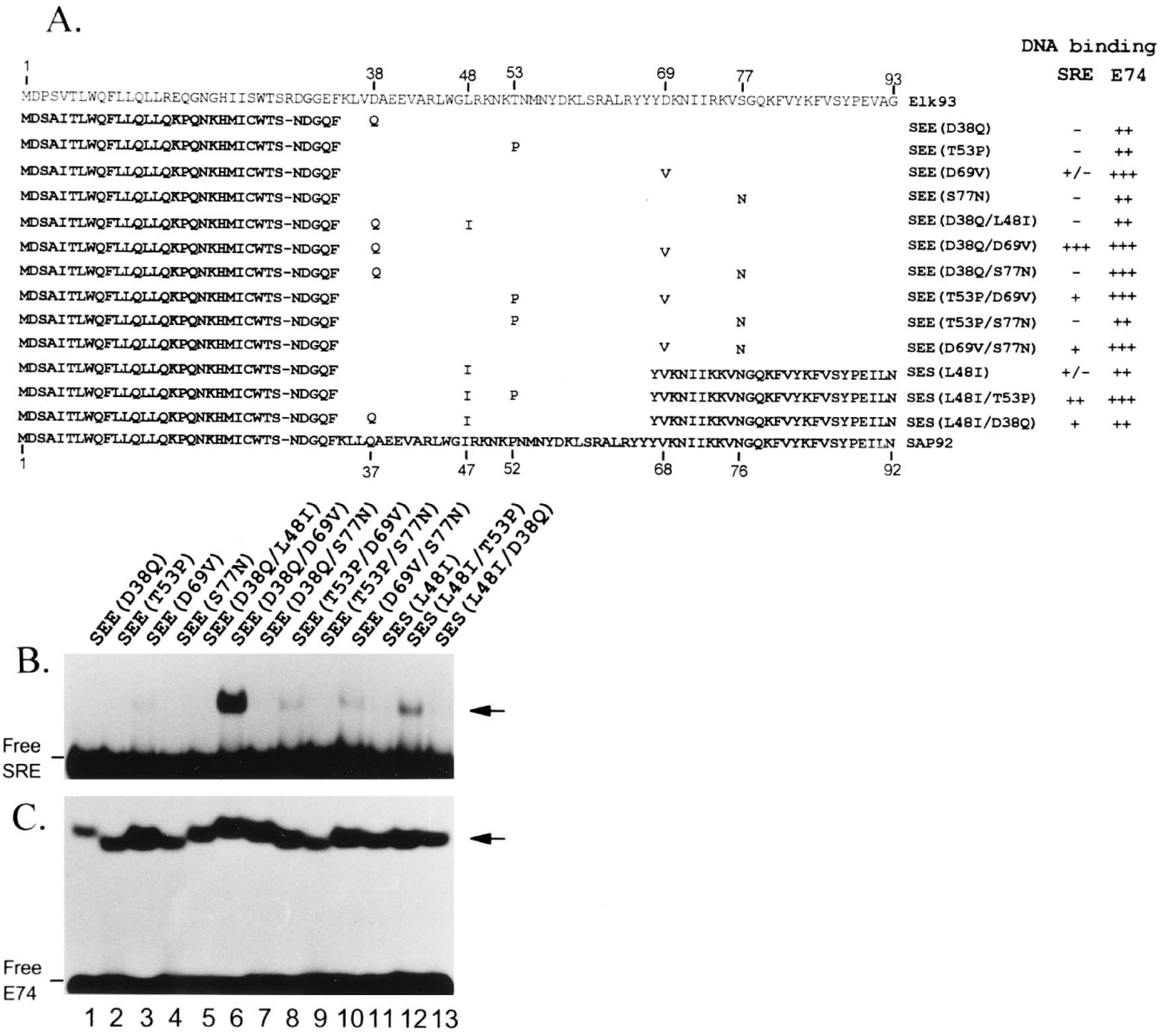

FIG. 4. Mapping the DNA-binding specificity-determining residues of the SAP-1 ETS domain. (A) Sequences of the mutant and wild-type ETS domains. SAP-1-derived amino acids are indicated in boldface. The positions of the mutated residues with respect to the native proteins are indicated above (Elk-1) and below (SAP-1) the wild-type sequences. Relative DNA-binding affinities in comparison with SEE(D38Q/D69V) binding to each site are shown as follows: +++, 65 to 100% binding; ++, 25 to 65% binding; +, 5 to 25% binding; and −, <1% binding. (B and C) Gel retardation analysis of mutant, chimeric ETS domains on the c-*fos* SRE and the E74 site, respectively. Equal molar quantities of each protein were used in binding the individual DNA-binding sites. Ten times more protein was used in binding to the SRE than to the E74 site. The bands corresponding to protein-DNA complexes are indicated (arrows).

national effect in the context of the D69V mutation. All the mutant proteins bound efficiently to the E74 site (Fig. 4C), indicating that they are fully functional DNA-binding proteins. These data therefore demonstrate that the mutation D69V alters the DNA-binding specificity of Elk-1 to that of SAP-1. However, this mutation works efficiently only in the presence of other single point mutations (e.g., D38Q).

The double mutation D38Q/D69V was introduced into the context of the wild-type ETS domain of Elk-1. The resulting protein, MElk (Fig. 5A), was tested for SRE binding. MElk efficiently bound to both the SRE (Fig. 5B, lane 4) and the E74 site (Fig. 5B, lane 8), whereas Elk-93 bound efficiently only to the E74 site (Fig. 5B, lanes 1 and 5). Therefore, the introduction of the two mutations D38Q and D69V into the context of Elk-1 is sufficient to confer SRE binding upon this ETS domain.

MElk displays a binding specificity similar to that of SAP-92. The mutant Elk-93 derivative MElk efficiently binds to the c-*fos* SRE and therefore appears to have a binding specificity similar to that of SAP-92. In order to further investigate the DNA-binding specificity of MElk, its binding to a series of mutated E74 sites was determined. The mutated E74 sites were created by replacing individual nucleotides with the corresponding nucleotides of the c-*fos* SRE (Fig. 5D). MElk DNA binding closely mirrored that observed with SAP-92, indicating that its DNA-binding specificity is similar to that of SAP-1 (Fig. 5C). In contrast, Elk-93 binds only to the wild-type E74 site with an efficiency similar to that of SAP-92 (Fig. 5C, lane

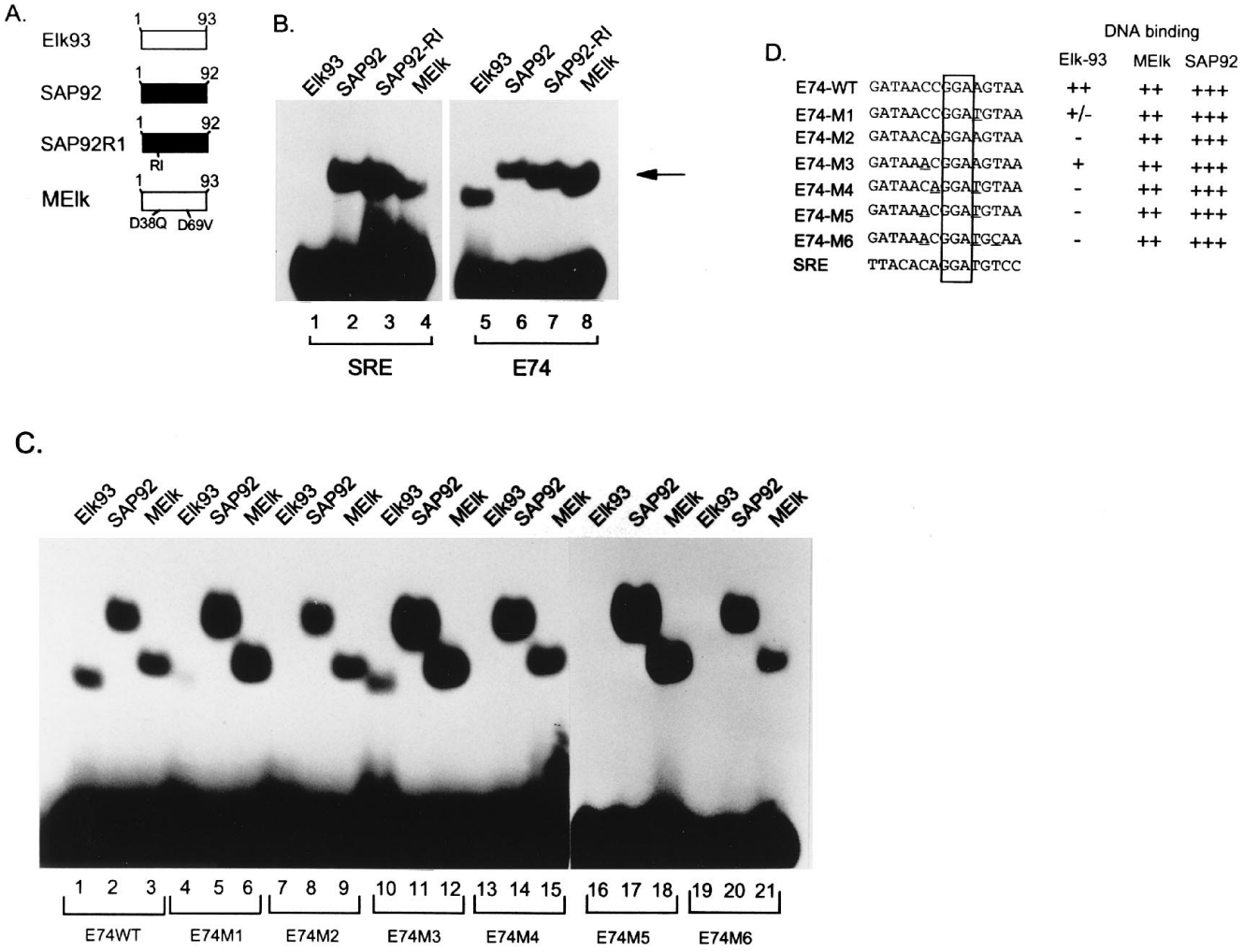

FIG. 5. MElk shows binding site specificity similar to that of SAP-92. (A) Schematic representation of the Elk-93, SAP-92, SAP-92RI, and MElk ETS domains. Elk-1-derived sequences (open boxes) and SAP-1-derived sequences (shaded boxes) and mutated residues in MElk are indicated. (B) Gel retardation analysis of wild-type and mutant ETS domains bound to the c-*fos* SRE (lanes 1 to 4) and the E74 site (lanes 5 to 8). SAP-92 was synthesized from plasmid pAS280. The positions of protein-DNA complexes are indicated (arrow). (C) Gel retardation analysis of Elk-93, SAP-92, and MElk bound to wild-type and mutant E74-binding sites. The identities of the proteins and of the DNA-binding sites are indicated above and below the lanes, respectively. Equal molar quantities of each protein were used in each binding reaction. The DNA-binding sites are not labelled to equal specific activities. (D) Sequences of the wild-type and mutant E74 sites. Residues which are mutated in each site (underlined) and the central conserved GGA motif (boxed) are indicated. Relative DNA-binding affinities are shown. Binding is calculated relative to SAP-92 binding to each site and is indicated as follows: +++, 65 to 100% binding; ++, 25 to 65% binding; +, 5 to 25% binding; +/−, 1 to 5% binding; and −, <1% binding.

1). Elk-93 binds with a much lower affinity to the E74-M1 and E74-M3 sites (Fig. 5C, lanes 4 and 10), and binding to the other sites was not detectable (Fig. 5C, lanes 7, 13, 16, and 19). These results indicate that the D38Q and D69V mutations allow Elk-93 to bind to a spectrum of sites similar to that of SAP-92. Moreover, these results further substantiate previous observations that Elk-1 binds efficiently only to sites containing the optimal Elk-1 binding sequence (e.g., E74), whereas SAP-1 can efficiently bind to a more degenerate series of sites (46).

To further characterize the specificity of MElk DNA binding, sites were selected from a pool of random-sequence double-stranded oligonucleotides. The site selection procedure was monitored by gel retardation analysis after each round of selection (Fig. 6A). After three rounds of selection, efficient complex formation was observed (Fig. 6A, lane 5). The sequences of the selected sites are shown in Fig. 6B. All the selected sequences contain the invariant GGA motif. We have numbered residues on the basis of this motif, with the central guanine residue designated as position 0. After three rounds of selection, MElk selects sites based on the consensus sequence A(C/t)CGGA(A/t)(G/a)YR, of which four residues are invariant (C−2, G−1, G0, and A+1). This consensus sequence represents a relaxed version of the consensus sequence selected by Elk-93 and closely resembles that selected by SAP-92 (see Fig. 9 and Discussion). In particular, the degeneracy at the +2, +3, +4, and +5 positions is similar to the sites selected by SAP-92. In order to further analyze the differential binding of MElk, SAP-92, and Elk-93, the binding of MElk to sites selected by SAP-92 (Fig. 7A, lanes 1 to 10) and Elk-93 (Fig. 7A, lanes 11 to 15) was investigated. All these sites were efficiently bound by MElk (Fig. 7A) and SAP-92 (45) but showed variable binding by Elk-93 (46). Moreover, a comparison of Elk-93, SAP-92, and MElk binding to sites selected by MElk reveals that with the exception of M10, all the sites are bound efficiently by MElk (Fig. 7B, lanes 21 to 30) and SAP-92 (Fig. 7B, lanes 11 to 20) but are bound with variable efficiency by Elk-93 (Fig. 7B,

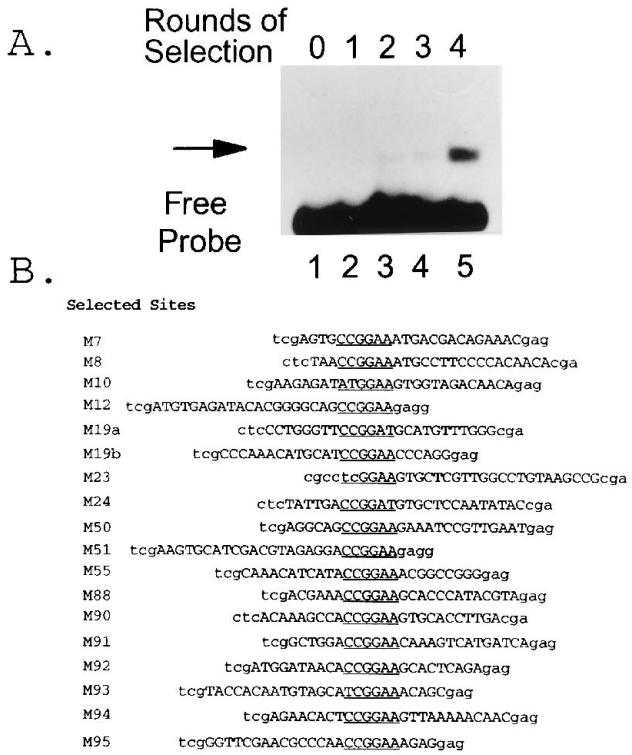

FIG. 6. Selection of a consensus DNA-binding site for MElk. (A) Gel retardation analysis of the selected pools of binding sites by bacterially expressed MElk (lanes 2 to 5). The starting oligonucleotide pool is shown in lane 1. The DNA probe represents the oligonucleotide pool after the indicated number of rounds of selection. The position of the protein-DNA complex is shown (arrow). The DNA from the complexes in lane 4 was amplified and cloned for sequence analysis. (B) Sequences of DNA-binding sites selected by MElk after three rounds of selection. Nucleotides derived from the random sequence (uppercase) and from the constant flanking primers (lowercase) are indicated. The central 6 bp of the ets-binding motif are underlined in each sequence. Sites are aligned and oriented according to this motif. A summary of the selected sites is given below the sequence. Nucleotide positions within sites are numbered with respect to the central guanine within the GGA trinucleotide motif (designated 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, lowercase letter indicates that this nucleotide is present in >80% of the remaining sites. R, selected purine; Y, selected pyrimidine. The M10 site is omitted from the summary of sites as this was subsequently shown not to be bound efficiently by MElk.

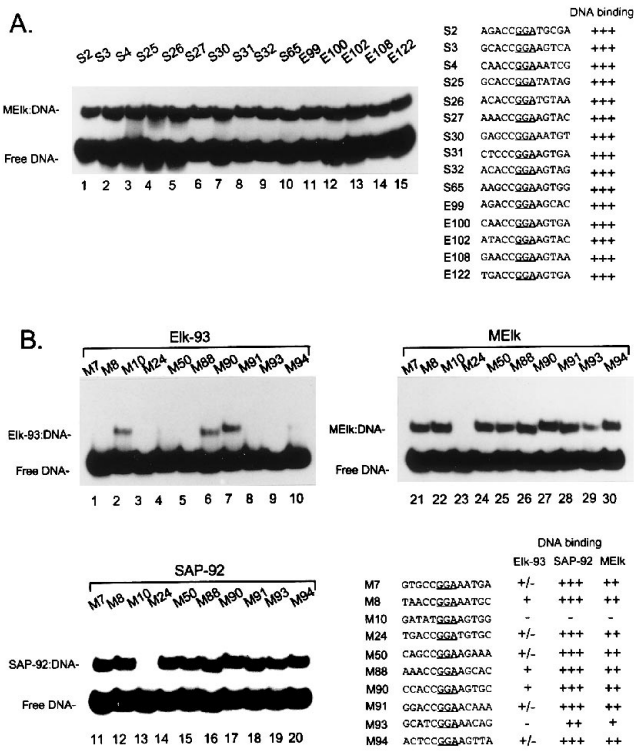

FIG. 7. Binding of TCF ETS domains to selected sites. (A) Gel retardation analysis of the binding of in vitro-translated MElk to sites selected by SAP-92 (lanes 1 to 10) and Elk-93 (lanes 11 to 15). The identities of the DNA-binding sites are indicated above the lanes. The locations of the free DNA and MElk-DNA complexes are indicated on the left. The sequences of the 13 nucleotides centered on the conserved GGA motif (underlined) are shown on the right. DNA-binding affinities relative to the optimal site tested (MElk on S26) are indicated (+++, 65 to 100% binding). (B) Gel retardation analysis of the binding of in vitro-translated Elk-93, SAP-92, and MElk to sites selected by MElk. The identities of the DNA-binding sites are indicated above the lanes. The locations of the free DNA and protein-DNA complexes are indicated. Equal molar amounts of each ETS domain were used in all the DNA-binding reactions. The sequences of the 13 nucleotides centered on the conserved GGA motif (underlined) are shown. DNA-binding affinities relative to the optimal site tested (SAP-92 on M90) are indicated as follows: +++, 65 to 100% binding; ++, 25 to 65% binding; +, 5 to 25% binding; +/−, 1 to 5% binding; and −, <1% binding.

lanes 1 to 10). These results confirm that the MElk-selected sites reflect the binding specificity of this protein. Together, these data demonstrate that the D38Q and D69V mutations in Elk-93 are sufficient to confer a DNA-binding specificity that is essentially identical to that of SAP-92.

The DNA-binding specificity of the SAP-2 ETS domain. In addition to Elk-1 and SAP-1, a third human TCF, SAP-2, has been identified (36). Truncated SAP-2 proteins bind to sites containing a high-affinity ets motif but are incapable of binding autonomously to the c-*fos* SRE (36). The DNA-binding characteristics of the isolated ETS domain of SAP-2 (SAP2-92, containing amino acids 1 to 92 from SAP-2) have not previously been investigated. The binding of SAP2-92 to the c-*fos* SRE was therefore compared with that of SAP-92 and Elk-93. In comparison with SAP-92, SAP2-92 bound very weakly to the SRE (<5% of SAP-92 binding) (Fig. 8B, lane 3) but with higher affinity than Elk-93. This result is consistent with the notion that SAP-2 binds DNA in a manner similar to that of Elk-1 but may have a slightly different mode of binding. A truncated SAP-2 derivative (SAP2AB) which contains the A (ETS DNA-binding) and B-box (SRF interaction) domains forms ternary complexes with an efficiency similar to that of SAP1AB (Fig. 8C, lanes 1 and 6). However, upon competition with excess unlabelled E74 binding site, the complex containing SAP2AB decayed more rapidly (half-life, <1 min) than the complex containing SAP1AB (half-life, >3 min) (Fig. 8C, compare lanes 1 to 5 with lanes 6 to 10). The stability of the ternary complex containing SAP2AB was similar to that of the analogous Elk-1 derivative (ElkAB) (Fig. 2). These data indicate that in the context of the ternary complex with SRF and the

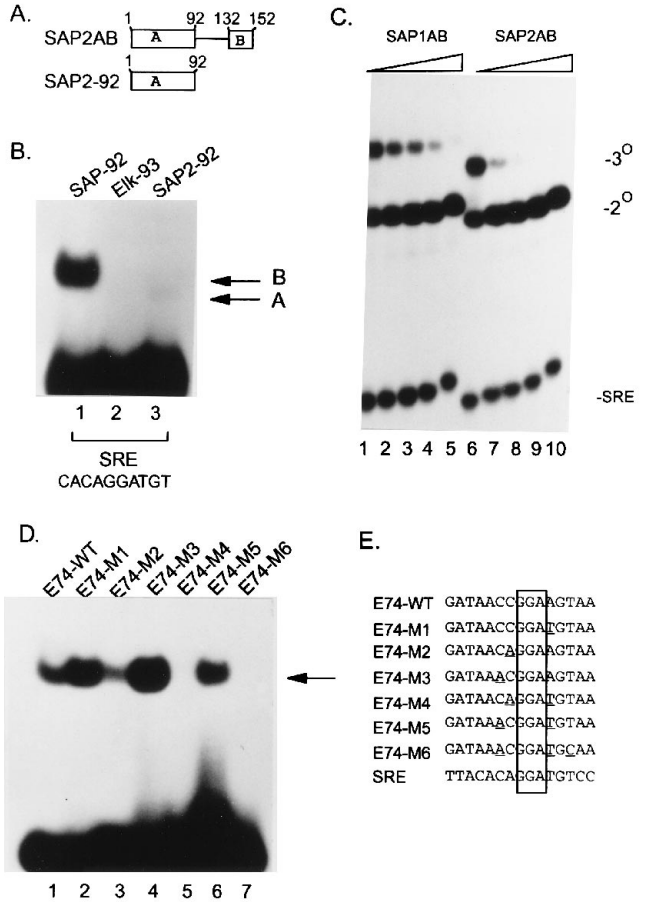

FIG. 8. DNA binding by the SAP-2 ETS domain. (A) Schematic representation of the SAP-2 derivatives which include either domain A (ETS DNA binding) (SAP2-92) or the A (ETS DNA-binding) and B (SRF-binding) domains (SAP2AB). The numbers of the amino acids which delineate the borders of these domains are indicated. (B) Gel retardation analysis of SAP-92, Elk-93, and SAP2-92 binding to the c-*fos* SRE. Equal molar quantities were used in each binding reaction. The locations of protein-DNA complexes are indicated (arrows). The sequence surrounding the ets motif in the c-*fos* SRE is shown. (C) Competition gel retardation analysis of SAP-1 and SAP-2 derivatives within ternary TCF-SRF-SRE complexes. Ternary complexes were allowed to form between SRF, the c-*fos* SRE, and either SAP1AB or SAP2AB as indicated above the lanes. A 150-fold molar excess of unlabelled E74 competitor DNA was subsequently added, and the resulting complexes were analyzed after 0 (lanes 1 and 6), 1 (lanes 2 and 7), 3 (lanes 3 and 8), 7 (lanes 4 and 9), and 15 (lanes 5 and 10) min. The positions of the binary (2°) (SRF-SRE) and ternary (3°) (TCF-SRF-SRE) complexes are indicated. The increase in time of competition is indicated schematically above each set of lanes. (D) Gel retardation analysis of SAP2-92 bound to wild-type and mutant E74 binding sites. The identities of the DNA-binding sites are indicated above the lanes. The locations of SAP2-92–DNA complexes are indicated (arrow). The DNA-binding sites are the same as used for Fig. 5 and are not labelled to equal specific activities. (E) Sequences of the wild-type and mutant E74 sites. Residues which are mutated in each site (underlined) and the central conserved GGA motif (boxed) are indicated.

c-*fos* SRE, both SAP-2 and Elk-1 form weaker complexes than SAP-1. This provides further evidence to indicate that SAP-2 has DNA-binding characteristics similar to those of Elk-1. To further investigate the DNA-binding specificity of SAP-2, the binding of SAP2-92 to a series of wild-type and mutant E74 sites was examined (Fig. 8D and E). Efficient DNA binding by SAP2-92 to all the sites tested except E74-M4 (Fig. 8D, lane 5) and E74-M6 (Fig. 8D, lane 7), which showed negligible binding, was seen. In common with Elk-93, SAP2-92 efficiently binds to the E74 site and binds with negligible affinity to the E74-M4 and E74-M5 sites (Fig. 8D, lanes 1, 5, and 7). In contrast, SAP2-92 shows increased binding to the E74-M1, -M2, -M3, and -M5 sites (Fig. 8D, lanes 2, 3, 4, and 6). Therefore, although SAP-2 shows binding characteristics similar to those of Elk-1 (e.g., the c-*fos* SRE and a subset of mutant E74 sites), SAP-2 binds efficiently to additional sites.

Elk-1	A	A	C	C	G	G	A	A	G	T	Ga	
SAP-1	N	A	C	C	G	G	A	At	Ga	Tc	N	
MElk	N	A	Ct	C	G	G	A	At	Ga	Y	R	
	-5	-4	-3	-2	-1	0	+1	+2	+3	+4	+5	

FIG. 9. Comparison of consensus sites selected by Elk-1, SAP-92 (46), and MElk (this study). The central conserved GGA motif is boxed. Uppercase letters represent nucleotides that are present in >65% of sites. A second, lowercase letter indicates that this nucleotide is present in >80% of the remaining sites. R, selected purine; Y, selected pyrimidine; N, no preferred nucleotide. Nucleotide positions within sites are numbered with respect to the central guanine within the GGA trinucleotide motif (designated 0).

DISCUSSION

The ETS-domain transcription factors all contain a common DNA-binding motif which is conserved both in the primary sequence (reviewed in references 20 and 52) and, at least in the case of Fli-1 and Ets-1, in the secondary (6, 27) and tertiary (7, 26, 53) structure elements. This amino acid sequence conservation is reflected by the similarity of the sites bound by ETS-domain transcription factors. However, it is unclear how ETS-domain transcription factors achieve sequence-specific DNA recognition. The ETS domains of the TCFs Elk-1 and SAP-1a exhibit differential DNA binding specificities (46). Both Elk-1 and SAP-1a are induced by ERK MAP kinase to efficiently form ternary complexes with SRF and the c-*fos* SRE and bind autonomously to sites containing high-affinity ets motifs. These full-length, phosphorylated proteins exhibit similar specificities of binding to their respective isolated ETS DNA-binding domains. Thus, the ETS domain contains all the information required to direct sequence-specific DNA binding. Indeed, the replacement of two residues in the Elk-1 ETS domain by the corresponding residues from SAP-1 is sufficient to confer the DNA-binding specificity of SAP-1 upon the Elk-1 ETS domain. The consensus DNA-binding site selected by the mutant ETS domain containing the D38Q and D69V mutations, MElk, is very similar to that selected by SAP-1 (Fig. 9). In contrast, Elk-1 selects a very constrained sequence in which the central nine residues are invariant (46) (Fig. 9). The two mutated residues, D38Q and D69V, are therefore sufficient to confer the relaxed binding specificity on Elk-1. Close inspection of the selected sites reveals that MElk can tolerate base changes within the +2, +3, +4, and +5 positions and at the −5 position, although a central conserved CCGGA core is recognized by both proteins. The D38Q and D69V mutations therefore alter the ability of the ETS domain to recognize residues flanking the central CCGGA core, implying that they induce some flexibility in the recognition determinants within the ETS domain.

The locations of these two critical amino acids, D-38 and D-69, within the model of the ETS domain of Elk-1 based on the Ets-1–DNA structure, are shown in Fig. 10. D-69 is located after the C terminus of the recognition helix (H3), whereas D-38 is located between β-strand 2 and α-helix H2. The locations of these residues suggest that neither is in a position to

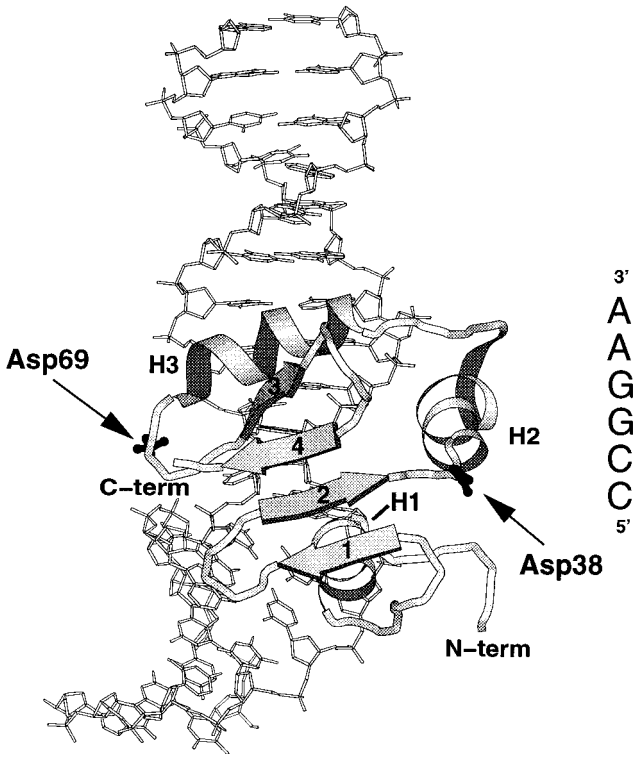

FIG. 10. Model of the Elk-1 ETS domain (Elk-93) bound to DNA, based on the solution structure of the Ets-1–DNA complex (53). The Elk-1 structure is shown as a ribbon by using the program MOLSCRIPT (23). The locations of residues D-38 and D-69 and the secondary structure elements (α-helices H1 to H3 and β-strands 1 to 4) are indicated. The nucleotide sequence of the central core of the binding site is shown on the right, opposite to the locations of the bases in the model.

make direct hydrophobic or H-bond contacts with the DNA. This is in agreement with the structure of the Ets-1 ETS domain complexed to DNA, in which the equivalent residues (D-57 and D-88) are not implicated in direct DNA contacts (53). Moreover, the equivalent residues in the Fli-1 ETS domain (D-313 and D-344) are also not implicated in direct DNA contacts but do experience nuclear magnetic resonance chemical shift changes on DNA binding (26). We therefore favor the hypothesis that the mutations D38Q and D39V play roles in affecting the orientations of other amino acids which are involved in DNA contacts. Indeed, D-38 is positioned to form H bonds with the N terminus of α-helix 2 (H2) to form an N-terminal cap. The substitution D38Q is likely to have a very significant influence on such helix capping, and this could significantly alter the protein conformation in this region. The side chain of D-69 appears to be substantially exposed to solvent, but it is spatially close to Y-66 in the model based on Fli-1 and Y-67 in the model based on Ets-1. The residues to homologous Y-66 in the Fli-1 (Y-341) (26) and Ets-1 (Y-85) (53) DNA-binding domains have been implicated in direct DNA contacts to the methyl group of the thymidine which is complementary to the adenine at the +2 position in the GGAA core of the DNA-binding site. Indeed, in the Ets-1–DNA structure (53), the side chain of this residue is very close to the methyl group of both thymidines complementing the GGAA sequence. The D69V mutation may therefore alter the orientation of Y-66 either directly or indirectly by the bulky valine residue, disfavoring the tight turn connecting the C terminus of the H3 α-helix with β-strand 3. The side chain of Y-67 appears to be integral to this turn. In both cases, the net result of the D69V mutation would be an alteration in the contacts made at the +1 and +2 positions, which is corroborated by the observation that MElk no longer has a strict requirement for an adenine at the +2 position. The mutation of K-78 in Ets-1 also leads to a change in binding specificity at the +2 position in the GGAA core binding motif (1). However, K-78 is located at the N-terminal end of the recognition helix and instead appears to be close enough to form H-bond contacts between its amino group and the +2-position adenine ring (53). On the basis of these structural predictions, the net result of introducing both the D38Q and the D69V mutations would be to allow greater flexibility in the SAP-1a ETS domain, which would potentially allow recognition of a larger number of sites. This is consistent with DNA-binding site selection data showing that both MElk (this study) and SAP-1 (46) exhibit less stringent binding specificities than Elk-1 (Fig. 9). However, although unlikely, we cannot rule out that an additional role for Asp-38 and Asp-69 in DNA binding may be brought about by water-mediated hydrogen bonds. Such bonds are important in the DNA-bound Trp repressor complex (33). Again, in this alternative scenario, the substitution of the carboxylate residues would significantly influence the ability of these residues to mediate DNA contacts.

Several DNA-binding site selections on ETS-domain proteins have been carried out (reviewed in reference 46). However, no clear trend linking particular amino acids at the positions equivalent to D-38 and D-69 with the generation of a particular consensus sequence has emerged. The derivation of consensus sequences may mask more-striking differences in binding specificities of ETS-domain proteins for individual binding sites. It is not surprising, however, that a simple relationship does not emerge because of the intimate interaction we propose for D-38 and D-69 with other spatially close amino acids. Indeed, the residues surrounding D-38 and C terminal from D-69 are conserved among individual family members. Moreover, the introduction of the mutation T53P in conjunction with D69V also strengthens SRE binding by Elk-1, albeit at a lower level than the combination of D38Q with D69V. This residue (T-53) is located between α-helices H2 and H3 and may also play a role in modulating the precise alignment of the DNA-protein interface. It therefore appears to be a general theme that residues interspersed between the structural elements play a role in modulating DNA binding by the ETS domain. Indeed, it is likely that different intramolecular interactions may play key roles in other ETS proteins in modulating their DNA-binding specificity. One striking example is Ets-1, in which amino acid I-445 plays a key role in modulating DNA binding (47). The equivalent residue in Elk-1 (I-72) is close to D-69 in the loop connecting the recognition helix H3 with β-strand 3, and its mutation is proposed to cause a local conformational change in the ETS domain. Furthermore, the R74D mutation in Elk-1 causes a temperature-sensitive reduction in DNA binding by Elk-1 (19). Mutation of the corresponding residue in Ets-1 causes a similar temperature-sensitive effect on its DNA-binding affinity (24). However, it is unknown whether the binding specificities of the Ets-1 I445 and H447D mutants and the Elk-1 R74D mutant are altered in comparison with those of the wild-type proteins.

The different binding specificities of Elk-1 and SAP-1a suggest that they may target different spectra of sites in vivo. Indeed, it might be predicted that because of its relaxed specificity, SAP-1a would target a greater number of sites, only some of which would represent Elk-1 targets. As Elk-1 apparently does not bind any sites better than SAP-1a, it seems unlikely that Elk-1 has any unique in vivo targets. It appears

that the DNA-binding specificity of SAP-2 is different from those of both Elk-1 and SAP-1a. However, DNA-binding site selections are required to determine the full spectrum of sites that are efficiently bound by SAP-2. The two critical residues identified in this study which modulate DNA binding are D-68 (conserved in Elk-1) and K-37 in SAP-2. The latter residue is a glutamine in SAP-1a and an aspartate in Elk-1 and may therefore play a role in directing the differential binding specificity of SAP-2. Moreover, one prediction from our results is that the V68D mutation (the mutation reciprocal to that of MElk) would disrupt the binding of SAP-1a to the SRE. This is in fact the case in SAP-2, which contains an aspartate at the equivalent position and binds the SRE very weakly. Both Elk-1 and SAP-1a form SRF-dependent complexes on the c-*fos* SRE (4, 16), whereas full-length SAP-2 is incapable of forming such complexes (36). It is currently not known whether TCFs can function in vivo in an SRF-independent manner, although at such sites the DNA-binding specificities of individual TCFs would play a major role in promoter targeting. In addition, a role for differential TCF binding at SRF-dependent sites other than the c-*fos* SRE, at which TCF binding would be the primary event followed by SRF recruitment, can be envisaged. It is also tempting to speculate that the interaction of TCFs with other proteins may alter their DNA-binding specificities, potentially by inducing allosteric changes within the ETS domain. Such changes could turn Elk-1 into a more promiscuous DNA-binding protein. Indeed, within ternary complexes with SRF, Elk-1 can bind a greater spectrum of sites (50), although in this case, this is probably because of a stabilization of the complex by additional protein-protein interactions rather than an effect on the Elk-1 DNA-binding domain.

Phosphorylation of SAP-1a by ERK2 MAP kinase stimulates its recruitment into ternary complexes with SRF and its autonomous binding to high-affinity ets sites. Similarly, phosphorylation of Elk-1 by both ERK2 and JNK1 also induces both modes of DNA binding (40, 54). In contrast, compared with Elk-1, JNK1 MAP kinase neither targets SAP-1a efficiently (54) nor stimulates its DNA-binding capacity in vitro (our unpublished data). These data suggest that Elk-1 and SAP-1a may be differentially targeted by different MAP kinase pathways. Indeed, it has been previously suggested that SAP-1a can respond to signal transduction pathways different from those of Elk-1 (15). However, further in vivo experiments are required to substantiate this theory. Phosphorylated full-length TCFs appear to have specificities of binding similar to those of their respective ETS DNA-binding domains. Therefore, phosphorylation appears to increase the efficiency of DNA binding rather than cause a change in specificity. A model has been proposed on the basis of the original grappling-hook model (50) in which phosphorylation of Elk-1 causes a change in its conformation from a closed complex to an open complex in which the DNA and protein binding interfaces are exposed (40). On the basis of our results, a similar model can be envisaged for SAP-1a.

The mechanisms by which individual members of transcription factor families achieve sequence-specific DNA binding appear to be increasingly diverse. A large number of amino acids are conserved among family members in order to maintain the overall structural fold of the DNA-binding domain. Moreover, many protein-DNA contacts are also conserved. Nonconserved amino acids then direct unique protein-DNA contacts which confer a particular DNA-binding specificity. This can be achieved by creating new direct H bonds between the residues on recognition helices and DNA bases. A notable example is a single amino acid on the recognition helix of the Bcd homeodomain which determines the DNA-binding specificity of this protein (13, 48; reviewed in reference 9). Alternatively, residues which contact DNA but are located away from the face of the recognition helix can have profound effects on DNA-binding specificity. Notable examples in this context are within the basic region of the MADS domains of SRF and MEF2 (31, 35, 43). In this study, we have identified residues within the ETS domain which modulate the DNA-binding specificity of Elk-1. In this case, the important residues appear not to contact DNA but rather to modulate the way other residues interact with the DNA. This type of mechanism may be commonly used to subtly modify the DNA-binding specificity of transcription factors from other families.

ACKNOWLEDGMENTS

We thank Margaret Bell for excellent technical assistance, Catherine Pyle for excellent secretarial assistance, and Bob Liddell for DNA sequencing and oligonucleotide synthesis. We are grateful to Peter Shaw and Richard Treisman for reagents and discussion of data prior to publication. We thank Steve Fesik for the Fli-1 coordinates and also Lawrence McIntosh and Marius Clore for the Ets-1 coordinates. We also thank members of our laboratories for stimulating discussions and critical comments on the manuscript. We also thank Peter Shaw for help, advice, and use of his laboratory for a short-term EMBO Fellowship (Paul Shore).

Paul Shore was supported by a short-term EMBO Fellowship and the NECRC. Roger Davis is an investigator of the Howard Hughes Medical Institute. This work was also supported by grants from the Wellcome Trust, Royal Society, and National Cancer Institute (CA65861 and CA58396).

ADDENDUM

While the manuscript was under review, a paper on the structure of the ETS-domain protein PU.1 bound to DNA was published (21a). The orientations of the ETS domain and of the DNA contacts differ from those in the Ets-1 structure. On the basis of the PU.1 model, the specificity-determining residues in the Elk-1 ETS domain would be located away from the protein-DNA interface. The study by Kodandapani et al. therefore supports our proposed mechanism of function for these two residues, although the DNA-contacting residues which are affected by the specificity-determining residues in Elk-1 would differ depending on which structure our data are modelled on.

REFERENCES

1. **Bosselut, R., J. Levin, E. Adjadj, and J. Ghysdael.** 1993. A single amino-acid substitution in the Ets domain alters core DNA binding specificity of Ets1 to that of the related transcription factors Elf1 and E74. Nucleic Acids Res. **21:**5184–5191.
2. **Brown, T. A., and S. L. McKnight.** 1992. Specificities of protein: protein and protein-DNA interactions of GABPα and two newly defined Ets-related proteins. Genes Dev. **6:**2502–2512.
3. **Brunger, A. T.** 1992. XPLOR, a system for X-ray crystallography and NMR. Yale University Press, New Haven, Conn.
4. **Dalton, S., and R. Treisman.** 1992. Characterisation of SAP-1, a protein recruited by serum response factor to the c-*fos* serum response element. Cell **68:**597–612.
5. **Degnan, B. M., S. M. Degnan, T. Naganuma, and D. E. Morese.** 1993. The *ets* multigene family is conserved throughout the Metazoa. Nucleic Acids Res. **21:**3479–3484.
6. **Donaldson, L. W., J. M. Petersen, B. J. Graves, and L. P. McIntosh.** 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.
7. **Donaldson, L. W., J. M. Petersen, B. J. Graves, and L. P. McIntosh.** 1996. Solution structure of the ETS-domain from murine Ets-1: a winged helix-turn-helix motif. EMBO J. **15:**125–134.
8. **Fisher, R. J., S. Koizumi, A. Kondoh, J. M. Mariano, E. Mavrothalassitis, N. K. Bhat, and T. S. Papas.** 1992. Human Ets-1 oncoprotein. J. Biol. Chem. **267:**17957–17965.
9. **Gehring, W. J., Y. Q. Qian, M. Billeter, K. Furukubo-Tokunaga, A. F. Schier,**

D. Resendez-Perez, M. Affolter, E. Otting, and K. Wuthrich. 1994. Homeodomain-DNA recognition. Cell **78:**211–223.
10. **Gille, H., M. Kortenjann, O. Thomae, C. Moomaw, C. Slaughter, M. H. Cobb, and P. E. Shaw.** 1995. EMBO J. **14:**951–962.
11. **Gille, H., A. D. Sharrocks, and P. E. Shaw.** 1992. Phosphorylation of transcription factor $p62^{TCF}$ by MAP kinase stimulates ternary complex formation at the c-*fos* promoter. Nature (London) **358:**415–417.
12. **Giovane, A., A. Pintzas, S.-M. Maira, P. Sobieszcuk, and B. Wasylyk.** 1994. Net, a new *ets* transcription factor that is activated by Ras. Genes Dev. **8:**1502–1513.
13. **Hanes, S. D., and R. Brent.** 1989. DNA specificity of the bicoid activator protein is determined by homeodomain recognition helix residue 9. Cell **57:**1275–1283.
14. **Hill, C. S., R. Marais, S. John, J. Wynne, S. Dalton, and R. Treisman.** 1993. Functional analysis of a growth factor-responsive transcription factor complex. Cell **73:**395–406.
15. **Hipskind, R. A., M. Baccarini, and A. Nordheim.** 1994. Transient activation of RAF-1, MEK, and ERK2 coincides kinetically with ternary complex factor phosphorylation and immediate-early gene promoter activity in vivo. Mol. Cell. Biol. **14:**6219–6231.
16. **Hipskind, R. A., V. N. Rao, C. G. F. Mueller, E. S. P. Reddy, and A. Nordheim.** 1991. Ets-related protein Elk-1 is homologous to the c-*fos* regulatory factor $p62^{TCF}$. Nature (London) **354:**351–354.
17. **Janknecht, R., W. H. Ernst, and A. Nordheim.** 1995. SAP1a is a nuclear target of signaling cascades involving ERKs. Oncogene **10:**1209–1216.
18. **Janknecht, R., W. H. Ernst, V. Pingoud, and A. Nordheim.** 1993. Activation of ternary complex factor Elk-1 by MAP kinases. EMBO J. **12:**5097–5104.
19. **Janknecht, R., and A. Nordheim.** 1992. Elk-1 protein domains required for direct and SRF-assisted DNA-binding. Nucleic Acids Res. **20:**3317–3324.
20. **Janknecht, R., and A. Nordheim.** 1993. Gene regulation by Ets proteins. Biochim. Biophys. Acta **1155:**346–356.
21. **Karim, F. D., L. D. Urness, C. S. Thummel, M. J. Klemsz, S. R. McKercher, A. Celada, C. Van Beueren, R. A. Maki, C. V. Gunther, J. A. Nye, and B. J. Graves.** 1990. The ETS-domain: a new DNA-binding motif that recognises a purine-rich core DNA sequence. Genes Dev. **4:**1451–1453.
21a. **Kodandapani, R., F. Pio, C.-Z. Ni, G. Piccialli, M. Klemsz, S. McKercher, R. A. Maki, and K. R. Ely.** 1996. A new pattern for helix-turn-helix recognition revealed by the PU.1 ETS-domain-DNA complex. Nature (London) **380:**456–460.
22. **Kortenjann, M., O. Thomae, and P. E. Shaw.** 1994. Inhibition of v-*raf*-dependent c-*fos* expression and transformation by a kinase-defective mutant of the mitogen-activated protein kinase Erk2. Mol. Cell. Biol. **14:**4815–4824.
23. **Kraulis, P. J.** 1991. MOLSCRIPT: a program to produce both detailed and schematic plots of protein structures. J. Appl. Crystallogr. **24:**946–950.
24. **Kraut, N., J. Frampton, K. M. McNagny, and T. Graf.** 1994. A functional Ets DNA-binding domain is required to maintain multipotency of hematopoietic progenitors transformed by Myb-Ets. Genes Dev. **8:**33–44.
25. **Laudet, V., C. Niel, M. Duterque-Coquillaud, D. Leprince, and D. Stehelin.** 1993. Evolution of the ets gene family. Biochem. Biophys. Res. Commun. **190:**8–14.
26. **Liang, H., X. Mao, E. T. Olejniczak, D. G. Nettersheim, L. Yu, R. P. Meadows, C. B. Thompson, and S. W. Fesik.** 1994. Solution structure of Fli-1 when bound to DNA. Nature Struct. Biol. **1:**871–876.
27. **Liang, H., E. T. Olejniczak, X. Mao, D. G. Nettesheim, L. Yu, C. B. Thompson, and S. W. Fesik.** 1994. The secondary structure of the ets domain of human Fli-1 resembles that of the helix-turn-helix DNA-binding motif of the *Escherichia coli* catabolite gene activator protein. Proc. Natl. Acad. Sci. USA **91:**11655–11659.
28. **Lopez, M., P. Oettgen, Y. Akbarali, U. Fendorter, and T. A. Liberman.** 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.
29. **Mao, X., S. Miesfeldt, H. Yang, J. M. Leiden, and C. B. Thompson.** 1994. J. Biol. Chem. **269:**18216–18222.
30. **Marais, R., J. Wynne, and R. Treisman.** 1993. The SRF accessory protein Elk-1 contains a growth factor-regulated transcriptional activation domain. Cell **73:**381–393.
31. **Nurrish, S. J., and R. Treisman.** 1995. DNA binding specificity determinants in MADS-box transcription factors. Mol. Cell. Biol. **15:**4076–4085.
32. **Nye, J. A., J. M. Peterson, C. V. Gunther, M. D. Jonsen, and B. J. Graves.** 1992. Interaction of murine Ets-1 with GGA-binding sites establishes the ETS domain as a new DNA-binding motif. Genes Dev. **6:**975–990.
33. **Otwinowski, Z., R. W. Schevitz, R. G. Zhang, C. L. Lawson, A. Joachimiak, R. Q. Marmorstein, B. F. Luisi, and P. B. Sigler.** 1988. Crystal structure of trp repressor/operator complex at atomic resolution. Nature (London) **335:** 321–329.
34. **Pabo, C. O., and R. T. Sauer.** 1992. Transcription factors: structural families and principles of DNA recognition. Annu. Rev. Biochem. **61:**1035–1095.
35. **Pellegrini, L., S. Tan, and T. J. Richmond.** 1995. Structure of serum response factor bound to DNA. Nature (London) **376:**490–498.
36. **Price, M. A., A. E. Rogers, and R. Treisman.** 1995. Comparative analysis of the ternary complex factors Elk-1, SAP-1a and SAP-2 (ERP/NET). EMBO J. **14:**2589–2601.
37. **Rao, V. N., K. Huebner, M. Isobe, A. Ab-Rushidi, C. M. Croce, and E. S. P. Reddy.** 1989. Elk-1, tissue-specific Ets-related genes on chromosomes X and 14 near translocation breakpoints. Science **244:**66–70.
38. **Rao, V. N., and E. S. P. Reddy.** 1992. A divergent *ets*-related protein, Elk-1, recognizes similar c-*ets*-1 proto-oncogene target sequences and acts as a transcriptional activator. Oncogene **7:**65–70.
39. **Sambrook, J., E. F. Fritsch, and T. Maniatis.** 1989. Molecular cloning: a laboratory manual, 2nd ed. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.
40. **Sharrocks, A. D.** 1995. ERK2/p42 MAP kinase stimulates both autonomous and SRF-dependent DNA binding by Elk-1. FEBS Lett. **368:**77–80.
41. **Sharrocks, A. D., H. Gille, and P. E. Shaw.** 1993. Identification of the amino acids essential for DNA binding and dimerization in $p67^{SRF}$: implications for a novel DNA-binding motif. Mol. Cell. Biol. **13:**123–132.
42. **Sharrocks, A. D., and P. E. Shaw.** 1992. Improved primer design for PCR-based, site-directed mutagenesis. Nucleic Acids Res. **20:**1147.
43. **Sharrocks, A. D., F. Von Hesler, and P. E. Shaw.** 1993. The identification of elements determining the different DNA binding specificities of the MADS-box proteins $p67^{SRF}$ and RSRFC4. Nucleic Acids Res. **21:**215–221.
44. **Shore, P., L. Bisset, J. Lakey, J. P. Waltho, R. Virden, and A. D. Sharrocks.** 1995. Characterization of the Elk-1 ETS DNA-binding domain. J. Biol. Chem. **270:**5805–5811.
45. **Shore, P., and A. D. Sharrocks.** 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.
46. **Shore, P., and A. D. Sharrocks.** 1995. The ETS-domain transcription factors Elk-1 and SAP-1 exhibit differential DNA binding specificities. Nucleic Acids Res. **23:**4698–4706.
47. **Soudant, N., O. Albagli, P. Dhordain, A. Flourens, D. Stehelin, and D. Leprince.** 1994. A residue of the ETS domain mutated in the *v-ets* oncogene is essential for the DNA-binding and transactivating properties of the ETS-1 and ETS-2 proteins. Nucleic Acids Res. **22:**3891–3899.
48. **Treisman, J., P. Gonczy, M. Vashishtha, E. Harris, and C. Desplan.** 1989. A single amino acid can determine the DNA binding specificity of homeodomain proteins. Cell **59:**552–562.
49. **Treisman, R.** 1994. Ternary complex factors: growth regulated transcriptional activators. Curr. Opin. Genet. Dev. **4:**96–101.
50. **Treisman, R., R. Marais, and J. Wynne.** 1992. Spatial flexibility in complexes between SRF and its accessory proteins. EMBO J. **11:**4631–4640.
51. **Urness, L. D., and C. S. Thummel.** 1990. Molecular interactions within ecdysone regulatory hierarchy: DNA binding properties of the Drosophila ecdysone-inducible *E74A* protein. Cell **63:**47–61.
52. **Wasylyk, B., S. L. Hahn, and A. Giovane.** 1993. The Ets family of transcription factors. Eur. J. Biochem. **211:**7–18.
53. **Werner, M. H., G. M. Clore, C. L. Fisher, R. J. Fisher, L. Trinh, J. Shiloach, and A. M. Gronenborn.** 1995. The solution structure of the human ETS1-DNA complex reveals a novel mode of binding and true side chain intercalation. Cell **83:**761–771.
54. **Whitmarsh, A. J., P. Shore, A. D. Sharrocks, and R. J. Davis.** 1995. Integration of MAP kinase signal transduction pathways at the serum response element. Science **269:**403–407.
55. **Woods, D. B., J. Ghysdael, and M. J. Owen.** 1992. Identification of nucleotide preferences in DNA sequences recognised by c-Ets-1 protein. Nucleic Acids Res. **20:**699–704.
56. **Zinck, R., M. A. Cahill, M. Kracht, C. Sachsenmaier, R. A. Hipskind, and A. Nordheim.** 1995. Protein synthesis inhibitors reveal differential regulation of mitogen-activated protein kinase and stress-activated protein kinase pathways that converge on Elk-1. Mol. Cell. Biol. **15:**4930–4938.
57. **Zinck, R., R. A. Hipskind, V. Pingoud, and A. Nordheim.** 1993. c-*fos* transcriptional activation and repression correlate temporally with the phosphorylation status of TCF. EMBO J. **12:**2377–2387.