Stepwise evolution of Elk-1 in early deuterostomes

Janice Saxton, Zoltan Ferjentsik§, Charles Ducker and Peter E Shaw*

School of Life Sciences,

University of Nottingham,

Queen's Medical Centre,

Nottingham, NG7 2UH, UK.

Phone: +44 115 82 30 120

Fax: +44 115 82 30 142

Email: peter.shaw@nottingham.ac.uk

*author for correspondence

§present address: Adaptimmune Ltd, Milton Park, Abingdon, OX14 4RY, UK

Running title: Elk-1 antecedents

Word count: 6,337

Abbreviations:

Electrospray tandem mass spectrometry (ES-MS/MS)

Extracellular signal-Regulated Kinase (ERK)

Fibroblast growth factor (FGF)

human embryonic stem cells (hESC)

Myocardin-Related Transcription Factors (MRTFs)

murine embryonic fibroblasts (MEFs)

Primordial germ cells (PGCs),

Saccoglossus kowalevskii Elk (*Sk*-Elk)

Serum Response Factor (SRF)

SRF DNA-binding core (core^{SRF})

Ternary Complex Factors (TCFs)

transcription factor (TF)

trans-activation domain (TAD)

vascular smooth muscle (VSM)

ABSTRACT

Metazoans have multiple ETS paralogues with overlapping or indiscriminate biological functions. Elk-1, one of three mammalian Ternary Complex Factors (TCFs), is a well-conserved, ETS domain-containing transcriptional regulator of mitogen-responsive genes that operates in concert with Serum Response Factor (SRF). Nonetheless, its genetic role remains unresolved because the *elk-1* gene could be deleted from the mouse genome seemingly without adverse effect. Here we have explored the evolution of Elk-1 to gain insight into its conserved biological role.

We identified antecedent Elk-1 proteins in extant early metazoans and used amino acid sequence alignments to chart the appearance of domains characteristic of human Elk-1. We then performed biochemical studies to determine whether putative domains apparent in the Elk-1 protein of a primitive hemichordate were functionally orthologous to those of human Elk-1.

Our findings imply the existence of primordial Elk-1 proteins in primitive deuterostomes that could operate as mitogen-responsive ETS transcription factors but not as TCFs. The role of TCF was acquired later, but presumably prior to the whole genome duplications in the basal vertebrate lineage. Thus its evolutionary origins link Elk-1 to the appearance of mesoderm.

BACKGROUND

It is now recognised that several classes of transcription factor (TF) originated with metazoans. Whereas some metazoan TFs resemble non-metazoan relatives, e.g. homeobox analogues, others appear to be novel. Importantly, in both cases early gene duplications gave rise to conserved TF families and provided the diversity that allowed the evolution of differentiated cells types and body patterning characteristic of higher metazoans [1].

One family of metazoan TFs that appears to lack protozoan antecedents unites the ETS proteins, originally identified on the basis of homology to the E26 avian oncogene [2]. In an early study, Degnan and colleagues successfully used degenerate PCR primers to amplify ETS domain sequences from the genomes of several lower metazoans, but not from fungi, plants or protozoans, concluding that ETS proteins are exclusive to metazoan animals [3].

The advent of next generation sequencing accelerated in-depth analysis of multiple genomes exemplifying extant early metazoans. All phyla depicted in the phylogram shown in figure 1a have multiple *ETS* genes. Sponges (Porifera) and cnidarians such as *Nematostella vectensis* (starlet sea anemone) consist of ecto- and endodermal cell layers whereas all other metazoans have a third, mesodermal, cell layer. In general, such triploblastic animals are either protostomes, in which the blastopore develops into a mouth, or deuterostomes, in which it develops into an anus. Primitive extant deuterostomes include the echinoderm *Strongylocentrotus purpuratus* (sea urchin) and hemichordate *Saccoglossus kowalevskii* (acorn worm). Also related to the vertebrates are the tunicates (urochordates) exemplified by *Ciona intestinalis*, albeit in larval form only, and cephalochordates such as *Branchiostoma lanceolatum* (amphioxus) [4].

Ternary Complex Factors (TCFs) represent an idiosyncratic sub-family of ETS proteins. The first example was characterised as a companion transcription factor for Serum Response Factor (SRF) and subsequently identified as the *Elk-1* gene product [5-7]. Although implicated in mitogen-dependent expression of key genes early in the cell cycle [8], gene deletion experiments in mice failed to ascribe any fundamental role to TCFs during development, possibly reflecting a degree of redundancy among the three TCF genes, *Elk-1*, *Elk-3* (Net) and *Elk-4* (Sap1a) [9-12]. In contrast,

Elk-1 knockdown in *Xenopus* impacted on mesoderm development [13]. Identifying the origin of Elk-like genes may shed light on its ancestral role in development. In this study, we first identified antecedent Elk-1 genes in several extant early metazoans. We then used amino acid sequence alignments to compare their domain structures and biochemical studies to confirm that characteristic domains present were functionally orthologous to those of human Elk-1. Our findings link the evolutionary origin of Elk-1 to the advent of deuterostomes and its co-option as a TCF to the appearance of mesoderm.

RESULTS

Putative Elk orthologues among early metazoan ETS proteins

The degree of early *ETS* gene expansion is illustrated in Table 1. NCBI lists 16 predicted *ETS* genes for both the sea anemone and sea urchin; higher vertebrates have as many as 28, a probable consequence of genome duplications [1]. This extensive early gene expansion notwithstanding, there is a high degree of ETS-domain amino acid sequence conservation. Nonetheless, sequence comparisons suggest the early diversification of ETS protein paralogues and their subsequent evolutionary conservation. On this basis, potential Elk-1 antecedents can be identified in the sea anemone (cnidarian) and primitive deuterostomes. The distance relationships between these putative Elk proteins and several vertebrate Elks are indicated in figure 1b. In the ETS domain alignment shown in figure 1c, secondary structure elements are indicated and identical residues in the loops connecting helices and strands of the $\alpha\beta\beta\alpha\alpha\beta\beta$ fold are highlighted to indicate the extensive conservation.

ETS domain amino acid conservation strongly reflects ETS domain structure and primary function. Regardless of its position within the protein, the winged helix-turn-helix domain retains its characteristic 3D form and recognises a core purine-rich DNA sequence (C/AGGA) [14]. Thus other features of ETS proteins, for example the pointed domain [15], characterise the different sub-families and provide additional indicators of evolutionary conservation. The three mammalian Elks have in common an amino-terminal ETS domain, a SRF interaction motif (B-box), MAPK docking motifs (D-box and DEF) that frame a trans-activation domain (TAD) with multiple MAPK consensus phosphorylation sites (figure 1d). In addition Elk-1 has SUMOylation sites involved in transcriptional repression and a cryptic degron implicated in protein stability [16, 17].

To determine the evolutionary origin of Elk-1 we sought to identify these characteristic features in potential orthologues from extant early metazoans. Although there are 16 sea anemone (cnidarian) ETS proteins, sequence comparisons revealed no domains orthologous to those listed above, including the putative protein used in figure 1 that has the highest homology to vertebrate Elks within its ETS domain. In contrast, the same approach revealed that in primitive representatives of all four deuterostome lineages (figure 1a) putative Elk proteins share additional homology regions with Elk-1. However, as we describe below, the domain composition of these Elk antecedents is incomplete, which may reflect the adapting role of Elk during the evolutionary transition to vertebrates. To test whether our inferences from sequence data correlate with function, we performed functional studies comparing human Elk-1 (*Hs*-Elk-1) with a putative orthologue from the hemichordate acorn worm *Saccoglossus kowaleskii* (*Sk*-Elk).

ETS domain binding to E74 and c-*fos* SRE sites

A full-length cDNA for *Sk*-Elk, a 384aa protein, was synthesised, cloned and used to express recombinant proteins. Analogous *Hs*-Elk-1 proteins were produced in parallel. All proteins were purified by means of histidine tags using immobilised metal affinity chromatography (IMAC). Because the C-terminal domain of Elk-1 can have a negative impact on DNA binding we first generated recombinant Elk proteins truncated C-terminally to the B-box (*Sk*-Elk-AB and *Hs*-Elk-AB; figure 2a and b) and used them to assay for DNA binding to well-characterised Elk-1 binding sites.

Neither *Sk*-Elk-AB nor *Hs*-Elk-AB bound to a control DNA duplex lacking a consensus ETS binding site (figure 2c, lanes 3 and 4) whereas both proteins bound to a DNA duplex encoding the high-affinity E74 ETS binding site (figure 2c, lanes 7 and 8), a site to which *Hs*-Elk-1 is known to bind directly [18]. The lack of any binding activity associated with a control protein expressed in the same bacterial strain and purified by IMAC (figure 2c, lane 6) indicates that the complex formed by *Sk*-Elk-AB is specific. Thus the ETS domains of *Sk*-Elk and *Hs*-Elk-1 both bind to the E74 motif.

Early deuterostome Elks are not TCFs

Mammalian Elks interact with SRF through a short region (B-box) that forms a β-strand flanked by short α10 helices and contacts the DNA-binding core of SRF (core^{SRF}) [19]. Alanine scanning mutagenesis identified five key residues within the B-box, including two tyrosine residues [20]. Comparison of sequences C-terminal to each ETS domain revealed a high degree of B-box conservation among vertebrates but significantly less in tunicate, echinoderm, hemichordate and early vertebrate representatives (figure 2d). As the SRF core is 95.6% conserved between acorn worm and mammals and the region recognised by the Elk-1 B-box is 100% conserved (figure 2e) these sequences suggest the absence of a SRF interaction domain in Elk-like ETS proteins of primitive deuterostomes.

To assay for ternary complex formation the same truncated proteins used to assess E74 binding were incubated with the c-*fos* SRE either alone or in the presence of recombinant core^{SRF}. Neither protein, nor the non-specific control, bound to the SRE directly (lanes 10-12). Furthermore, whereas core^{SRF} recruited *Hs*-Elk-AB efficiently into a ternary complex (lane 16), it did not recruit *Sk*-Elk-AB (lane 15). As echinoderm, tunicate and amphioxus Elk proteins have similarly poor B-box consensus sequences (figure 2d), this result suggests that none of these Elk proteins is recruited to DNA by SRF. Thus primitive Elks appear not to operate as TCFs with SRF.

Early deuterostome Elks resemble MAPK targets

Figure 3a shows an alignment of the *Hs*-Elk-1 TAD with sequences of orthologues from echinoderm, hemichordate, tunicate and vertebrate species. The core TAD of *Hs*-Elk-1, defined functionally by Janknecht and Nordheim [21], contains an IHFW motif (blue), two ERK phosphorylation sites (green) and an FxFP (DEF) docking motif (red) [22]. Additional ERK consensus sites are present flanking the core TAD, which are also phosphorylated in mitogen-stimulated cells [23]. This arrangement is wholly conserved in vertebrates and echinoderms and largely conserved in putative Elks from the tunicate *Ciona* and the hemichordate acorn worm. In the *Ciona* ETS protein most closely related to *Hs*-Elk-1, instead of 15 residues separating the IHFW and FxFP motifs there are 47, with a single intervening ERK consensus site. In the acorn worm, the IHFW motif is imperfect (FPFW) but the spacing to the FxFP motif and the two intervening ERK consensus sites are well conserved. However, there are two additional FxFP motifs, one overlapping the imperfect IHFW motif, such that three direct repeats, each consisting of FxFP motif and two ERK consensus sites, are apparent (figure 3b).

Several observations suggest that the core TAD defines the first ancestral Elk, i.e. in which ERK signalling regulates an ETS protein. Firstly, sequences resembling the most highly conserved core motif (IHFWSTLSP, see figure 3a) are present in fungi, in which MAPK cascades exist, but also in bacteria, in which they do not (Table 2). Moreover, fungal proteins bearing such motifs do not appear to be MAPK substrates. This implies that Elks did not originate by recombination-based convergence between an ETS domain and an existing ERK substrate. Instead an ERK docking motif with adjacent phosphorylation sites (dock-phos cassette) appears to have evolved in a single ETS gene, undergone duplication in the hemichordate lineage, or optimisation as a core TAD as reflected in echinoderms and vertebrates (figure 3b).

Phosphorylation of Elk-1 by ERK and other MAPKs is aided by a second

docking motif, the D-box, consisting of basic and non-polar elements [24]. Although not essential for the phosphorylation of key serines in Elk-1, the D-box contributes to the binding of ERK to Elk-1 and may participate in Elk-1 recruitment of ERK to chromatin [25, 26]. The degree of conservation, especially of several key residues within the basic patch and LxL motif (figure 3c) in all four lineages suggests that it could contribute to ERK phosphorylation of primitive deuterostome Elk proteins.

Early deuterostome Elks are good targets for ERK phosphorylation

To determine if *Sk*-Elk serves as an ERK substrate, we compared *Sk*-Elk with *Hs*-Elk-1 in an *in vitro* kinase assay. Initial attempts to prepare full-length *Sk*-Elk from bacteria were unsuccessful because the protein was rapidly cleaved by endogenous proteases. To circumvent this problem, we used baculoviral vectors to express *Hs*-Elk-1 and *Sk*-Elk proteins in insect cells. Purified *Sk*-Elk migrated at 45kDa in SDS-PAGE, significantly faster than human Elk-1 (figure 4a), but ES-MS/MS indicated that both purified proteins were intact (Table 3).

Recombinant, active ERK2 phosphorylated both proteins efficiently. At initial time points *Sk*-Elk phosphorylation compared well with phosphorylation of *Hs*-Elk-1 (figure 4b, compare upper and lower panels; figure 4c). This implies that *Sk*-Elk possesses functional ERK docking motifs. However, after 20 minutes, phosphate incorporation began to plateau and it was apparent that more phosphate was incorporated into *Hs*-Elk-1 (after 30 min average *Hs*/*Sk* ratio = 1.25). This difference could reflect fewer good consensus ERK phosphorylation sites in *Sk*-Elk or a conformational flexibility less amenable to multiple phosphorylation events. Although we cannot yet distinguish between these possibilities, the result nonetheless indicates that *Sk*-Elk is likely to be a *bona fide* ERK substrate.

As ERK phosphorylation influences Elk-1 DNA binding, we compared DNA binding of unphosphorylated and ERK-phosphorylated proteins. Neither protein bound to control DNA, regardless of phosphorylation status (figure 4d, lanes 2-5). In line with an earlier report [18], ERK phosphorylation stimulated E74 binding by *Hs*-Elk-1 efficiently (compare lanes 9 and 10) but only enhanced *Sk*-Elk binding weakly (compare lanes 7 and 8). Thus C-terminal domains of *Hs*-Elk-1 and *Sk*-Elk repress ETS domain DNA binding and ERK phosphorylation relieves this repression, albeit inefficiently in the case of *Sk*-Elk.

The impact of phosphorylation on SRF-dependent binding of *Sk*-Elk and *Hs*-Elk-1 to the SRE was also tested. *Hs*-Elk-1 bound only with core^{SRF}, whereby binding increased slightly upon phosphorylation and the ternary complex was further retarded. In all cases binding of *Sk*-Elk to SRE was undetectable, as anticipated. In summary, sequence comparisons and biochemical data indicate that Elk proteins of primitive deuterostomes have functional ETS domains, are phosphorylated by MAPKs but cannot be recruited to SREs by SRF.

Early deuterostome Elks have Ras-responsive TADs

To see if *Sk*-Elk harboured a mitogen-responsive TAD, the *Sk*-Elk dock-phos cassette trimer (aas 288-384) was expressed as a GAL4 fusion in mammalian cells and tested for its ability to drive UAS-luciferase expression in response to mitogen signalling. Expression of active Ras caused a >40x increase in reporter activity by GAL4-SKE288 (figure 5a). In comparison, the analogous region of Elk-1 was significantly more active in the absence or presence of active Ras. Relative phosphorylation levels (figure 4b, c) may contribute to this disparity but it more likely reflects the high level of GAL4-Elk336 fusion expression (figure 5a, lower panel).

Mammalian Elk-1 proteins depend on MED23, a subunit of the Mediator co-activator complex, for transcriptional activation of target genes [27] and the acorn worm genome carries a *MED23* orthologue encoding a protein with 68.4% identity to murine MED23. We therefore assessed the potency of the *Sk*-Elk trans-activation domain in MED23+/+ and MED23-/- murine embryonic fibroblasts (MEFs). Whereas the *Hs*-Elk-1

TAD was active only in mitogen-stimulated MED23+/+ MEFs (figure 5b), the *Sk*-Elk TAD functioned equally well in both MED23+/+ and MED23-/- MEFs (figure 5c). We conclude that transcriptional activation by *Sk*-Elk does not depend on MED23 in murine cells, although it remains possible that *Sk*-Elk requires *Sk*-MED23 in acorn worm cells.

DISCUSSION

Although first identified 25 years ago and since examined in considerable molecular detail, Elk-1 has retained an aura of uncertainty about its biological role, not least because removal of the *Elk-1* gene from the mouse genome incurred no strong phenotype [9]. On the other hand, interference with Elk-1 expression in *Xenopus* led to disruption of mesoderm development, particularly posterior mesoderm, formation of which is dependent on fibroblast growth factor (FGF) signalling [13]. Indeed, Elk-1 knockdown mimicked the defects associated with FGF inhibition in *Xenopus* [28]. Here we sought additional clues to the biological role of Elk-1 from its evolutionary history. We discovered orthologues of Elk-1 in invertebrate species at the base of deuterostomes, but not among protostomes. Conservation of *Elk-1* genes from lower deuterostomes to vertebrates suggests it regulates a fundamental biological process. Interestingly, our results from sequence comparisons and functional studies reveal that the canonical Elk-1 activities, i.e. those associated with findings in mammals, emerged stepwise as complexity evolved within the lineage that led to vertebrates.

Elk-1 is the paradigmatic ERK-responsive transcription factor originally shown, in conjunction with SRF, to confer mitogen responsiveness on genes with SREs in their promoters (reviewed in [29, 30]). It is now known that Elk-1 recruits ERK to these gene promoters, where it phosphorylates MED14, a core subunit of the Mediator co-activator complex as part of the transcriptional activation process [26, 31]. In line with these mechanistic revelations, ChIP-seq data identified genes co-occupied by ERK and Elk-1 in human embryonic stem cells (hESCs) as being involved in metabolism and cell cycle progression [32].

Our exploration of databases for invertebrate antecedents of vertebrate Elks identified single genes encoding proteins with Elk characteristics in all four deuterostome lineages. Representatives of each lineage contain an amino-terminal ETS domain with amino acid sequences most closely related to mammalian Elk-3. In the case of *Sk*-Elk, ETS domain DNA binding was consistent with that ascribed to Elk-1, i.e. direct binding to E74 but not SRE sequences. Invertebrate Elk antecedents also harbour a recognisable TAD with an ERK 'dock-phos' cassette. As exemplified again by *Sk*-Elk, such TADs repress ETS domain DNA binding, they are phosphorylated by ERK and activate transcription in response to mitogen stimulation. Thus the primordial Elk possessed features conserved from echinoderms and hemichordates to mammals and were equipped to direct mitogen (FGF) signals to ETS target genes.

Invertebrate deuterostomes also have genes for SRF and MED23. Nonetheless Elk proteins from these species lack a B-box and dependence on MED23. In murine cells these features have been related to the repression of ventral mesoderm (smooth muscle). The B-box contributes to vascular smooth muscle (VSM) repression by enabling Elk-1 to compete with Myocardin-Related Transcription Factors (MRTFs) [33]. This antagonism also appears to depend on MED23, an essential co-factor for Elk-1, because MED23-/- MEFs have qualities resembling VSM cells [27]. The dependence of mammalian Elk-1 on MED23 is not shared by Elk-3 because the Elk-3 TAD functions in MED23-/- MEFs [34]. Together with sequence analysis, these data suggest that Elk-3, not Elk-1, represents the founding member of the vertebrate Elk family.

The absence of a B-box from the Elk-like protein of amphioxus and other non-chordate deuterostomes suggests that this domain evolved within the putative Elk ancestor prior to the whole genome amplifications in the basal vertebrate lineage and expansion of the Elk family (figure 6) [1]. The plasticity of

Elk family members may relate to the role of the ancestral Elk-1 as a transcriptional activator of genes in response to ERK signalling. ERK acquired novel roles as the complexities associated with vertebrate development evolved. The parallel evolution of structural complexity in Elk-1 would have facilitated the integration of nuclear ERK signalling with that of other developmental pathways to coordinate the differentiation and expansion of diverse tissues during vertebrate development.

ERK signalling in response to FGF is required to maintain pluripotency in hESCs and for the induction of mesendoderm [35, 36]. Mirroring the changes seen upon loss of ERK signalling, Elk-1 depletion in hESCs induced loss of pluripotency, up-regulation of differentiation genes and adoption of a fibroblast-like morphology [32]. Our findings indicate that the requisite up-regulation of pluripotency genes downstream of ERK signalling represents the ancestral role of Elk proteins, whereas Elk-1 interactions with SRF and MED23 are vertebrate innovations. The specification of ventral mesoderm, which is dependent on SRF and MRTFs, was incompatible with pluripotency and through the evolution of its B-box Elk-1 could be co-opted as repressor of SRF/MRTFs to maintain the integrity of a pluripotent stem cell population. Consistent with this notion, Elk-1 appears to require mitogens and MED23 in its role as an antagonist of MRTFs [27].

Recent work showed that the interruption of ERK signalling during early development of axolotl embryos causes a highly selective loss of primordial germ cells (PGCs), specification of which establishes the germ line during development [37]. The findings described here are therefore consistent with a hypothesis that the ancestral role of Elk proteins is to inhibit the differentiation of pluripotent cells prior to PGC specification in early development. This idea implicates Elk-1 in a process that is fundamental to metazoan evolution, explaining the conservation of its core structure and biochemical function from simple deuterostomes to mammals. A logical prediction from a model of Elk-1 as germ line guardian and differentiation repressor would be that in vertebrates close to the evolutionary path to mammals, deletion of Elk-1 early in development would result in loss of germ line and mesoderm expansion, not unlike the outcome in *Xenopus* [13].

MATERIALS AND METHODS
Genome analysis and gene identification

The draft *Ambystoma mexicanum* and *Branchiostoma lanceolatum* genomes were searched using the Basic Local Alignment Search Tool (BLAST). Searches were performed using amino acid sequences predicted on the basis of homologous Elk-1 genes from the most closely related taxon for which such a sequence was available on the publicly accessible NCBI protein database. Each identified gene was confirmed by the construction of phylogenetic trees of amino acid sequences using Elk-1 domain sequence alignments.

Gene identifiers for Elk-1 protein sequences from other species used in the study are as follows:
Human [*Homo sapiens*]
GI: 166362732; NP_001107595.1
Mouse [*Mus musculus*]
GI: 116292182; NP_031948.4
Xenopus laevis
GI: 194373047; ACF59754.1
Zebrafish [*Danio rerio*]
GI: 292616696; XP_696294.3
Ciona intestinalis
GI: 171916107; NP_001116439.1
Acorn worm [*Saccoglossus kowalevskii*]
GI: 291235822; XP_002737846.1
Sea urchin [*Strongylocentrotus purpuratus*]
GI: 47551223; NM_214627.2
Sea anemone [*Nematostella vectensis*]
GI: 156374032; XP_001629613.1

Protein alignments and phylograms were produced using ClustalOmega (EMBL-EBI).

Protein Expression and mass spectrometry
The coding sequence for acorn worm (*Saccoglossus kowalevskii*) Elk (*Sk*-Elk, 1-384, MW 42,921) with murine codon usage was synthesised by Gene ART. Amino-terminal regions of *Hs*-Elk-1 (1-189) and *Sk*-Elk (1-204) were

expressed from truncated cDNAs as bacterial proteins with C-terminal histidine tags (*Hs*-Elk-AB, MW 21,386; *Sk*-Elk-AB, MW 24,689). *Sk*-Elk and *Hs*-Elk-1 were expressed with C-terminal histidine tags in insect Sf9 cells from full-length cDNAs inserted into pFastBac1. Proteins were purified by immobilised metal affinity chromatography. Generation of recombinant core^{SRF} has been described previously [23].

Protein bands were excised from coomassie-stained gels, destained, reduced (10mM dithiothreitol), alkylated (20mM iodoacetamide) and digested with trypsin. Extracted peptides were evaporated to dryness and submitted to LC-MS/MS on an Ultimate 3000 Nano LC system (Dionex, UK) coupled to a LTQ-FT Ultra mass spectrometer (Thermo Scientific). Spectra were acquired in positive ion mode for a 400-2000 m/z range at a nominal resolution of 100,000. Identification of peptides was conducted in data-dependent mode. The 5 most intense ions for each scan were isolated and subjected to CID using a nominal energy of 35.0. Raw data files obtained from each LC-MS/MS acquisition were processed using Bioworks software (Thermo Scientific), searching each file in turn against a custom database including Elk sequences (*Homo sapiens* and *Saccoglossus kowalevskii*).

In vitro kinase and DNA binding assays

Recombinant Elk-1 proteins (0.5µg) were incubated with recombinant active Extracellular signal-Regulated Kinase (ERK)2 and ^{32}P-[γ]-ATP (Perkin Elmer, 3000Ci mmol^{-1}) as described [26]. Reactions were resolved by SDS-PAGE. Dried gels were exposed to imaging plates (Fujifilm); images were analysed by densitometry (AIDA).

Recombinant proteins (25-50ng) were pre-incubated with radio-labelled oligo-nucleotide control, E74 or SRE probes and the complexes formed were resolved on native 5% polyacrylamide gels as described earlier [23].

Cell culture and gene reporter assays

HEK293T cells were cultured in DMEM supplemented with 10% foetal calf serum, 100U ml^{-1} penicillin, 100µg ml^{-1} streptomycin and 2mM *L*-glutamine.

MED23+/+ and -/- MEFs were cultured in Dulbecco's MEM (Sigma D5671) supplemented with 15% FCS, 4mM *L*-glutamine, 100U ml^{-1} penicillin and 100µg ml^{-1} streptomycin. HEK293T cells were transfected by DNA-calcium phosphate co-precipitation; MEFs were transfected with K2 Transfection System (Biontex, Martinsried). Transfections included a (G5E4) UAS-luc3 reporter, TK-Rluc renilla control, expression vectors for GAL4 (pSG424) derivative and active (V^{12}) Ras. Sequences corresponding to *Hs*-Elk-1 (336-428) and *Sk*-Elk (288-386) TADs were inserted in frame with the GAL4 DNA-binding domain (1-147) in pSG424). All data shown derive from experiments performed at least twice independently, each with triplicate points and normalised to the Renilla control. Expression of GAL4 fusions was confirmed by immunoblotting with a mouse monoclonal antibody to the GAL4 DNA-binding domain (Santa Cruz Biotech, Dallas TX; sc-510).

AUTHOR CONTRIBUTIONS

JS generated recombinant proteins and performed experimental work; ZF provided the cDNA clone for *Sk*-Elk and performed genome searches; CD performed MS analyses; PES conceived and designed the study, performed sequence analyses and experimental work, prepared figures and wrote the manuscript.

REFERENCES

1. Cañestro, C., Albalat, R., Irimia, M. & Garcia-Fernàndez, J. (2013) Impact of gene gains, losses and duplication modes on the origin and diversification of vertebrates., *Semin Cell Dev Biol.* **24**, 83-94.

2. Watson, D. K., McWilliams-Smith, M. J., Nunn, M. F., Duesberg, P. H., O'Brien, S. & Papas, T. S. (1985) The ets sequence from the transforming gene of avian erythroblastosis virus, E26, has unique domains on human chromosomes 11 and 21: both loci are transcriptionally active., *Proc Natl Acad Sci U S A.* **82**, 7294-8.

3. Degnan, B. M., Degnan, S. M., Naganuma, T. & Morse, D. E. (1993) The ets multigene family is conserved throughout the Metazoa, *Nucleic Acids Res.* **21**, 3479-84.

4. Kimball, J. W. (1993) *Biology*, William C Brown.

5. Shaw, P. E., Schroeter, H. & Nordheim, A. (1989) The Ability of a Ternary Complex to Form Over the Serum Response Element Correlates with Serum Inducibity of the Human c-*fos* Promoter, *Cell.* **56**, 563-572.

6. Hipskind, R. A., Rao, V. N., Mueller, C. G. F., Reddy, E. S. P. & Nordheim, A. (1991) Ets-related protein Elk-1 is homologous to the c-fos regulatory factor p62TCF, *Nature.* **354**, 531-534.

7. Rao, V. N., Huebner, K., Isobe, M., Ar-Rushdi, A., Croce, C. M. & Reddy, E. S. P. (1989) elk, tissue specific ets-related genes on chromosome X and 14 near translocation breakpoints, *Science.* **244**, 66-70.

8. Gille, H., Sharrocks, A. D. & Shaw, P. E. (1992) Phosphorylation of transcription factor p62TCF by MAP kinase stimulates ternary complex formation at c-fos promoter, *Nature.* **358**, 414-417.

9. Cesari, F., Brecht, S., Vintersten, K., Vuong, L. G., Hofmann, M., Klingel, K., Schnorr, J. J., Arsenian, S., Schild, H., Herdegen, T., Wiebel, F. & Nordheim, A. (2004) Mice deficient for the ets transcription factor elk-1 show normal immune responses and mildly impaired neuronal gene activation., *Mol Cell Biol.* **24**, 294-305.

10. Ayadi, A., Zheng, H., Sobieszczuk, P., Buchwalter, G., Moerman, P., Alitalo, K. & Wasylyk, B. (2001) Net-targeted mutant mice develop a vascular phenotype and up-regulate *egr-1*., *EMBO J.* **20**, 5139-5152.

11. Costello, P. S., Nicolas, R. H., Watanabe, Y., Rosewell, I. & Treisman, R. (2004) Ternary complex factor SAP-1 is required for Erk-mediated thymocyte positive selection., *Nat Immunol.* **5**, 289-98.

12. Costello, P. S., Nicolas, R. H., Willoughby, J. E., Wasylyk, B., Nordheim, A. & R., T. (2010) Ternary complex factors SAP-1 and Elk-1, but not net, are functionally equivalent in thymocyte development., *J Immunol.* **185**, 1082-92.

13. Nentwich, O., Dingwell, K. S., Nordheim, A. & Smith, J. C. (2009) Downstream of FGF during mesoderm formation in Xenopus: the roles of Elk-1 and Egr-1., *Dev Biol.* **336**, 313-26.

14. Mo, Y., Vaessen, B., Johnston, K. & Marmorstein, R. (2000) Structure of the Elk-1-DNA complex reveals how DNA-distal residues affect ETS domain recognition of DNA, *Nature Struct Biol.* **7**, 292-297.

15. Slupsky, C. M., Gentile, L. N., Donaldson, L. W., Mackereth, C. D., Seidel, J. J., Graves, B. J. & McIntosh, L. P. (1998) Structure of the Ets-1 pointed domain and mitogen-activated protein kinase phosphorylation site., *Proc Natl Acad Sci U S A.* **95**, 12129-34.

16. Yang, S. H., Jaffray, E., Hay, R. T. & Sharrocks, A. D. (2003) Dynamic interplay of the SUMO and ERK pathways in regulating Elk-1 transcriptional activity., *Mol Cell.* **12**, 63-74.

17. Evans, E. L., Saxton, J., Shelton, S. J., Begitt, A., Holliday, N. D., Hipskind, R. A. & Shaw, P. E. (2011) Dimer formation and conformational flexibility ensure cytoplasmic stability and nuclear accumulation of Elk-1. , *Nucl Acids Res.* **39**, 6390-402.

18. Sharrocks, A. D. (1995) ERK2/p42 MAP kinase stimulates both autonomous and SRF- dependent DNA-binding by Elk-1, *FEBS Lett.* **368**, 77-80.

19. Hassler, M. & Richmond, T. J. (2001) The B-box dominates SAP-1-SRF interactions in the structure of the ternary complex, *EMBO J.* **20**, 3018-3028.

20. Ling, Y., Lakey, J. H., Roberts, C. E. & Sharrocks, A. D. (1997) Molecular characterization of the b-box protein-

protein interaction motif of the ets-domain transcription factor elk-1, *EMBO J.* **16**, 2431-2440.

21. Janknecht, R., Zinck, R., Ernst, W. H. & Nordheim, A. (1994) Functional dissection of the transcription factor Elk-1, *Oncogene.* **9**, 1273-1278.

22. Jacobs, D., Glossip, D., Xing, H., Muslin, A. J. & Kornfeld, K. (1999) Multiple docking sites on substrate proteins form a modular system that mediates recognition by ERK MAP kinase, *Genes Dev.* **13**, 163-175.

23. Gille, H., Kortenjann, M., Thomae, O., Moomaw, C., Slaughter, C., Cobb, M. H. & Shaw, P. E. (1995) ERK phosphorylation potentiates ELK-1-mediated ternary complex formation and transactivation, *EMBO J.* **14**, 951-962.

24. Yang, S. H., Whitmarsh, A. J., Davis, R. J. & Sharrocks, A. D. (1998) Differential targeting of map kinases to the ets-domain transcription factor elk-1, *EMBO J.* **17**, 1740-1749.

25. Burkhard, K. A., Chen, F. & Shapiro, P. (2011) Quantitative analysis of ERK2 interactions with substrate proteins: roles for kinase docking domains and activity in determining binding affinity., *J Biol Chem.* **286**, 2477-85.

26. Zhang, H. M., Li, L., Papadopoulou, N., Hodgson, G., Evans, E., Galbraith, M., Dear, M., Vougier, S., Saxton, J. & Shaw, P. E. (2008) Mitogen-induced recruitment of ERK and MSK to SRE promoter complexes by ternary complex factor Elk-1, *Nucl Acids Res.* **36**, 2594-607.

27. Yin, J. W., Liang, Y., Park, J. Y., Chen, D., Yao, X., Xiao, Q., Liu, Z., Jiang, B., Fu, Y., Bao, M., Huang, Y., Liu, Y., Yan, J., Zhu, M. S., Yang, Z., Gao, P., Tian, B., Li, D. & Wang, G. (2012) Mediator MED23 plays opposing roles in directing smooth muscle cell and adipocyte differentiation., *Genes Dev.* **26**, 2192-205.

28. Fletcher, R. B., Baker, J. C. & Harland, R. M. (2006) FGF8 spliceforms mediate early mesoderm and posterior neural tissue formation in Xenopus., *Development* **133** 1703-14.

29. Shaw, P. E. & Saxton, J. (2003) Ternary Complex Factors: Prime Nuclear Targets For Mitogen-Activated Protein Kinases., *Int J Biochem Cell Biol.* **35**, 1210-1226.

30. Buchwalter, G., Gross, C. & Wasylyk, B. (2004) Ets ternary complex transcription factors., *Gene.* **324**, 1-14.

31. Galbraith, M., Saxton, J., Li, L., Shelton, S. J., Espinosa, J. M., Zhang, H. M. & Shaw, P. E. (2013) ERK phosphorylation of MED14 in promoter complexes during mitogen-induced gene activation by Elk-1., *Nucleic Acids Res* **41**, 10241-53.

32. Göke, J., Chan, Y. S., Yan, J., Vingron, M. & Ng, H. H. (2013) Genome-wide kinase-chromatin interactions reveal the regulatory network of ERK signaling in human embryonic stem cells., *Mol Cell* **50** 844-55.

33. Wang, Z., Wang, D. Z., Hockemeyer, D., McAnally, J., Nordheim, A. & Olson, E. N. (2004) Myocardin and ternary complex factors compete for SRF to control smooth muscle gene expression, *Nature.* **428**, 185-9.

34. Balamotis, M. A., Pennella, M. A., Stevens, J. L., Wasylyk, B., Belmont, A. S. & Berk, A. J. (2009) Complexity in transcription control at the activation domain-mediator interface., *Sci Signal* **2**, ra20.

35. Xu, R. H., Peck, R. M., Li, D. S., Feng, X., Ludwig, T. & Thomson, J. A. (2005) Basic FGF and suppression of BMP signaling sustain undifferentiated proliferation of human ES cells., *Nat Methods.* **2**, 185-90.

36. Yu, P., Pan, G., Yu, J. & Thomson, J. A. (2011) FGF2 sustains NANOG and switches the outcome of BMP4-induced human embryonic stem cell differentiation., *Cell Stem Cell* **8**, 326-34.

37. Chatfield, J., O'Reilly, M. A., Bachvarova, R. F., Ferjentsik, Z., Redwood, C., Walmsley, M., Patient, R., Loose, M. & Johnson, A. D. (2014) Stochastic specification of primordial germ cells from mesoderm precursors in axolotl embryos., *Development* **141**, 2429-40.

FIGURE LEGENDS

Figure 1. <u>Elk-1 proteins found in primitive extant deuterostomes and vertebrates.</u> (**a**) Simplified evolutionary tree showing the four deuterostome lineages that emerged from triploblastic metazoans with extant species used in this study indicated below. (**b**) Phylogram indicating distance relationships for the known and predicted full-length Elk-1 protein sequences. (**c**) ClustalOmega alignment of ETS domain sequences from Elk-1 proteins in (b). Helices and β-strands of the winged-helix-turn-helix structure are over-lined; conservation within the intervening loops is highlighted in green. (**d**) Diagram of *Hs*-Elk-1 domain organisation.

Figure 2. <u>DNA binding and recruitment of SRF by hemichordate and mammalian Elk-1.</u> (**a**) Diagram of C-terminal truncations of *Sk*-Elk and *Hs*-Elk-1. (**b**) Preparations of recombinant *Sk*-Elk-AB and *Hs*-Elk-AB used in this study. Arrowheads indicate the full-length truncated proteins. NB, although ~3kDa larger, *Sk*-Elk-AB migrates faster than *Hs*-Elk-AB. (c) Comparison of *Sk*-Elk-AB and *Hs*-Elk-AB binding to DNA by electrophoretic mobility shift assay. Radio-labelled control (lanes 1-4), E74 (lanes 5-8) and SRE DNA duplexes (lanes 9-16) were incubated alone, with control protein, recombinant *Sk*-Elk-AB or *Hs*-Elk-AB, along with core^{SRF} where indicated. DNA-protein complexes were resolved on native polyacrylamide gels. (**d**). ClustalOmega alignment of SRF-interaction domain (B-box) sequences, including all deuterostome Elk-1 proteins in figure 1b, generated from sequences corresponding to 156 residues C-terminal to each ETS domain. (**e**) ClustalOmega alignment of SRF sequences encoding the Elk-1 interaction surface from deuterostome species in figure 1b. The dimerisation interface and Elk-1 binding surface are over-lined by solid and dashed lines respectively. Non-conserved residues are highlighted in colour.

Figure 3. <u>Conservation of the Elk-1 C-terminal domain in extant deuterostomes.</u> (**a**) ClustalOmega alignment of C-terminal sequences from deuterostome Elk-1 proteins in figure 1b.

Conserved motifs are highlighted in colour: the core IHFW motif (blue); ERK docking (DEF) motif (red) and consensus ERK phosphorylation sites (green). (**b**) Motif organisation in acorn worm (hemichordate) and vertebrate Elk-1 C-terminal domains. The bracket indicates the ERK dock-phos cassette repeated in the hemichordate Elk-1. (**c**) ClustalOmega alignment of basic/hydrophobic D-box docking motif sequences from deuterostome Elk-1 proteins in figure 1b. Conserved basic (magenta) and hydrophobic (red) residues are highlighted; asterisks show functionally important residues mapped by mutations [24].

Figure 4. <u>Hemichordate Elk-1 C-terminal domain is a substrate for ERK phosphorylation.</u>
(**a**) Preparations of recombinant *Sk*-Elk and *Hs*-Elk-1 used in this study. Arrowheads indicate the full-length proteins. (**b**) Recombinant full-length *Hs*-Elk-1 (upper panel) and *Sk*-Elk (lower panel) proteins were incubated with recombinant, active ERK2 and ^{32}P-labelled ATP for the times indicated. Reactions were stopped by heat inactivation and examined by SDS-PAGE and phosphor-imaging of the dried gels. (**c**) Quantification of ^{32}P-labelled phosphate incorporation into *Hs*-Elk-1 (blue) and *Sk*-Elk (red) as shown in (b). (**d**) Comparative assay of unphosphorylated and phosphorylated *Sk*-Elk and *Hs*-Elk-1 binding to DNA by electrophoretic mobility shift assay. Radio-labelled control (lanes 1-5), E74 (lanes 6-10) and SRE DNA duplexes (lanes 11-20) were incubated with control protein, unphosphorylated (-) or phosphorylated (+) *Sk*-Elk, unphosphorylated (-) or phosphorylated (+) or *Hs*-Elk-1, in the presence of core^{SRF} where indicated. DNA-protein complexes were resolved on native polyacrylamide gels.

Figure 5. <u>Hemichordate Elk-1 harbours a Ras-responsive, MED23-independent TAD.</u> (**a**) LH panel: HEK293 cells were co-transfected with UAS-firefly and control renilla luciferase reporters, expression plasmids for the Gal4 DBD alone (1-147)

or the same fused to the C-terminal region of *Sk*-Elk (aas 288-384) and for oncogenic Ras (+) or control empty vector (-). Data is expressed as ratio of firefly/renilla and represents the average of 3 independent experiments. RH panel: the same but including a Gal4 fusion with *Hs*-Elk-1 (aas 336-428). Data is expressed as $\log_{10}$ of the ratio of firefly/renilla (average of 3 independent experiments). Relative levels of Gal4 fusion expression were determined by immunoblotting cell lysates from one of the experiments (lower panel). (**b**) MED23+/+ and -/- MEFs were co-transfected with UAS-firefly and control renilla luciferase reporters, expression plasmids for the Gal4 DBD alone (1-147) or the same fused to the C-terminal region of *Hs*-Elk-1. Cells were serum starved for 24h and mock stimulated or treated with TPA for 6h. (**c**) As in (b) except the Gal4-*Sk*-Elk fusion was used instead of Gal4-Elk336.

Figure 6. Evolutionary time line for acquisition of Elk-1 characteristics. *ETS* genes are a metazoan innovation. The genome of the last common ancestor with cnidarians carried multiple *ETS* genes. Versions of the ERK dock-phos cassette are present in all four deuterostome lineages, consistent with its evolution before the divergence of hemichordates and echinoderms. The presence of multiple Elk proteins in vertebrates all containing a B-box suggests that this characteristic feature evolved prior to the whole genome duplications that preceded the emergence of agnathostomes. These molecular events coincide with the evolution of mesoderm (shaded box).

Table 1. Elk gene complements in cnidarians and deuterostomes

Phylum	Example	ETS genes	Elk genes	B Box	D Box	dock-phos cassette
Cnidarians	sea anemone	16	1	NO	NO	NO
Tunicates	*Ciona intestinalis*	17	1	NO	?	YES
Echinoderms	sea urchin	16	1	NO	?	YES
Hemichordates	acorn worm	11	1	NO	?	YES
Vertebrates	axolotl	?	3	YES	YES	YES
	human	30	3	YES	YES	YES

Table 2. Fungal and bacterial proteins with pseudo-core TAD motifs

Species	Protein	GI no.	Motif sequence
Pseudozyma hubeiensis	hypothetical	501308678	APFWSTLSP
Chaetomium globosum	predicted	116192667	QDFWSTLSP
Roseiflexus sp.	asparagine synthase	148657749	FTFWSTLSP
Deinococcus radiodurans	hypothetical	15806922	SAFWSTLSP

Table 3. ES-MS/MS analysis of recombinant Elk-1 proteins*

Identification	Sequence	% Coverage
Saccoglossus kowalevskii ETS domain-containing protein Elk-3-like	MQVEQLQVDDTRHGLKRTLAEMESKDEMSSAESQLKKRCLMP FVKPKGLESNITLWQFLLELLMDKNNQPLITWTSNDGEFKLV NAEEVARRWGLRKNKTNMNYDKLSRALRYYYDKNIIKKVMGQ KFVYKFVSFPEIVKTETKVPFRVKMESIEPQVSRPHSYPGAD NSTTTTSSNSVLPRTISSLVSPQNRQTSYSNPDLTVKSQAGH VNHSRPLFSPNVSNFTQLTSDVHLVSSTGKPISTVCTTMPSP ISRPQPIQLSIPMPVSKPSMTLPELRTTMASPSYTGFHPSHF TFPGTPVVLSSPIISQGTPTPLFSFPFWSTMSPMALSPQMSS MNHFQFPIISNINGHMAAMQGFMPPVSAFNAVSPVLLSPTSH KPIIAS	23.4
Homo Sapiens Elk-1	MDPSVTLWQFLLQLLREQGNGHIISWTSRDGGEFKLVDAEEV ARLWGLRKNKTNMNYDKLSRALRYYYDKNIIRKVSGQKFVYK FVSYPEVAGCSTEDCPPQPEVSVTSTMPNVAPAAIHAAPGDT VSGKPGTPKGAGMAGPGGLARSSRNEYMRSGLYSTFTIQSLQ PQPPPHPRPAVVLPNAAPAGAAAPPSGSRSTSPSPLEACLEA EEAGLPLQVILTPPEAPNLKSEELNVEPGLGRALPPEVKVEG PKEELEVAGERGFVPETTKAEPEVPPQEGVPARLPAVVMDTA GQAGGHAASSPEISQPQKGRKPRDLELPLSPSLLGGPGPERT PGSGSGSGLQAPGPALTPSLLPTHTLTPVLLTPSSLPPSIHF WSTLSPIAPRSPAKLSFQFPSSGSAQVHIPSISVDGLSTPVV LSPGPQKP	33.9

*Peptide sequences with positive identifications from MS/MS data for tryptic digests of protein bands excised from the polyacrylamide gel shown in figure 4a (arrowheads) are underlined.

FIGURE 1ab.

a

sea
anemone

sea
urchin

acorn
worm

amphioxus

*Ciona
intestinalis*

Xenopus,
axolotl,
zebrafish,
mouse,
human

b

Ciona	0.33796
sea anemone	0.21086
human	0.07209
mouse	0.07077
axolotl	0.21634
zebrafish	0.24797
Xenopus	0.19762
sea urchin	0.27286
acorn worm	0.24848
amphioxus	0.28461

FIGURE 1cd.

c

```
                                              α1                β1        β2
human         -----------MDPSVTLWQFLLQLLREQGNG-HIISWTSRDGGEFKLVDAEEVA
mouse         -----------MDPSVTLWQFLLQLLREQGNG-HIISWTSRDGGEFKLVDAEEVA
zebrafish     -----------MDPSITLWQFLLHLLDDQNQK-HLISWTSGD-GEFKLLDAEEVA
axolotl       -----------MDPSVTLWQFLLQLLEEQSNE-HLISWTSND-GEFKLVDAEEVA
Xenopus       -----------MDPAGTLWQFLLQLLEEQTHQ-HLISWTSND-GEFKLLDAEEVA
Ciona         MQQLGAHLPRVLDMNVTLWQFLLELLMDPSSNSHLISWTSAD-GEFKLHNSEEVA
amphioxus     -----------MEPNLTLWQFLLELLLDPKNA-NLIAWTSND-GEFKLNSAEEVA
acorn worm    RCLMPFVKPKGLESNITLWQFLLELLMDKNNQ-PLITWTSND-GEFKLVNAEEVA
sea urchin    ---MQHIKPKGLETNITLWQFLLELLMDKARQ-HLITWTGNE-GEFKLLNAEEVA
sea anemone   ---------FQRRGSLQLWQFLVALLDEP-ECSSFIAWTGRG-LEFKLIDPEEVA
                        .    **:***.** :     :*:**     *****  . ****

              α2            α3            β3        β4
human         RLWGLRKNKTNMNYDKLSRALRYYYDKNIIRKVSGQKFVYKFVSYPEVAGCSTED
mouse         RLWGLRKNKTNMNYDKLSRALRYYYDKNIIRKVSGQKFVYKFVSYPEVAGCSTED
zebrafish     RLWGLRKNKTNMNYDKLSRALRYYYDKNIIKKVSGQKFVYKFVTYPDPNSADGLK
axolotl       RLWGLRKNKTNMNYDKLSRALRYYYDKNIIKKVSGQKFVYKFVSYPDPVVMETMK
Xenopus       RLWGLRKNKTNMNYDKLSRALRYYYDKNIIKKVSGQKFVYKFVCYPDPGSSEVPK
Ciona         RLWGLRKNKTNMNYDKLSRALRYYYDKNIIKKVNGQKFVYKFVSFPEIIKTETKI
amphioxus     RLWGLRKNKTNMNYDKLSRALRYYYDKNIIKKVMGQKFVYKFVNLPDVLKPDVKP
acorn worm    RRWGLRKNKTNMNYDKLSRALRYYYDKNIIKKVMGQKFVYKFVSFPEIVKTETKV
sea urchin    RLWGLRKNKTNMNYDKLSRALRYYYDKNIIKKVMGQKFVYKFVSFPEIVKTENKI
sea anemone   RRWGLQKNRPAMNYDKLSRSLRYYYEKGIMQKVAGERYVYKFVCSPEALFSMAFP
              * **    *::    ***************:****:** **::*****   *
```

d

FIGURE 2.

FIGURE 2.

e

```
                               _____   ----------------
human        LSTLTGTQVLLLVASETGHVYTFATRKLQPMITSETGKALIQTCLNSP
mouse        LSTLTGTQVLLLVASETGHVYTFATRKLQPMITSETGKALIQTCLNSP
axolotl      LSTLTGTQVLLLVASETGHVYTFATRKLQPMITSETGKALIQTCLNSP
Xenopus      LSTLTGTQVLLLVASETGHVYTFATRKLQPMITSETGKALIQTCLNSP
zebrafish    LSTLTGTQVLLLVASETGHVYTFATRKLQPMITSETGKALIQTCLNSP
amphioxus    LSTLTGTQVMLLVASETGHVYTFATRKLQPMITSEAGKALIQTCLNSP
acorn worm   LSTLTGTQVMLLVASETGHVYTFATRKLQPMITSESGKALIQTCLNSP
sea urchin   LSTLTGTQVMLLVASETGHVYTFATRKLQPMITSESGKALIQTCLNSP
Ciona        LATLTGTQVMLLVASETGHVYTFATRKLQPILSSEAGKQLIQTCLNSP
             *:*******:********************:::**:**  *********
```
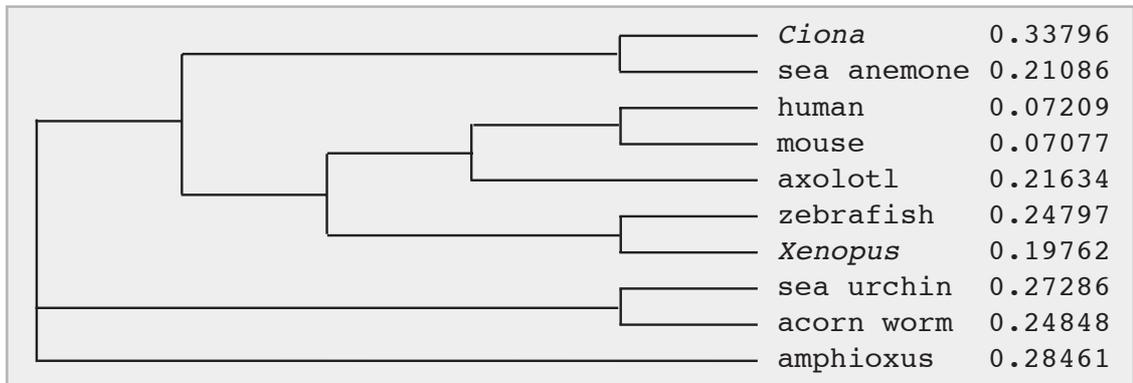
FIGURE 3.

a

```
human       ----PALTPSLLPT-HTLTPVLLT--PSSLPPSIHFWSTLSPIAPRSP---------------
mouse       ----PALTPSLLPT-HTLTPVLLT--PSSLPPSIHFWSTLSPIAPRSP---------------
zebrafish   ----NTLTPTVITS-HSLTPVLLT--PSPLPSTIHFWSTLSPIAPRSP---------------
axolotl     ----NSLTPTVITS-HALTPVLLT--PSSLPPSIHFWSTLSPIAPRSP---------------
Xenopus     ----GSASSSVITS-HSLTPLLLT--PGSLPPSIHFWSTLSPIAPRSP---------------
Ciona       ------KTPSDTNATFFPTSVIMT-PSPMVIPSITFWSTLSPMPPHGLGKTDLLAGVREGCSS
amphioxus   -----AQTPIMVTS--SPSPLIPNGQTGTPVIPLHFWSTLSPAAALSPRF-------------
acorn worm  -HFTFPGTPVV-----LSSPII-SQGTPTPLFSFPFWSTMSPMA-LSPQM-------------
sea urchin  GHSPLPGTPIM-----VASPLL-GQN-GTPLVPIHFWSTLSPLT-LSPSL-------------
                :           :  ::        :  ****:**       .
```

```
human       -----------------AKLSFQFPSSGSAQVHIPSI-----SVDGL--STPVVLSPGPQKP-
mouse       -----------------AKLSFQFPSSGSAQVHIPSI-----SVDGL--STPVVLSPGPQKP-
zebrafish   -----------------AKLSFQFPSNGSNQIQIPAL-----SVDGL--STPVVLSPGPQKP-
axolotl     -----------------AKLSFQFPTNGNNQIHIPTL-----SVDGL--STPVVLSPGPQKP-
Xenopus     -----------------AKLSFQFPTNGNNQIHVPTL-----SVDGL--STPVVLSPGLQKP-
Ciona       KSTNNDSKDGEGEKDQKAGSVFQFPTVVNGQMTFAGV----------PVR-------------
amphioxus   ---------------TPTSNHFQFPAITGSGSMPVAIGVPVNSFTGLP-LSPLILSPGFQKVY
acorn worm  ---------------SSMNHFQFPIISNINGHMAAMQGFMPPVSAFNAVSPVLLSPTSHKP-
sea urchin  ---------------NSNSSFQFPTLVSNGLSS-PL-----TVNGVPVTPTVLLSPASTKP-
                :    ****                    :
```

b

dock-phos cassette

hemichordate

vertebrate
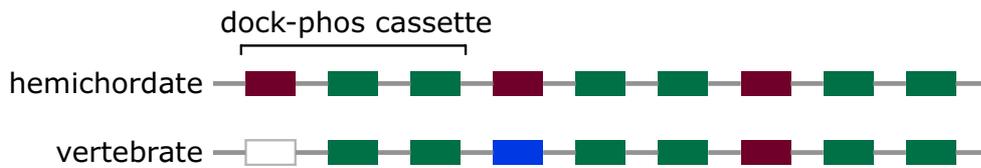
c

D-Box

```
                  **    *  *
human       ----ISQPQKGRKPRDLELPLSPSL-
mouse       ----ITQPQKGRKPRDLELPLSPSL-
zebrafish   ----DVQPPKSKKPRVLELPSSSTL-
axolotl     ----EGQPLKSRKPRDLELPFCPSL-
Xenopus     ----EGQLVKPKDLELPYSAVI-
Ciona       KRRLTGKKSIDRKPSPIDLSKPTRTE
amphioxus   SAGSTIHRPAVSRHRPHFISIPNIIS
acorn worm  ----TTMPSPISRPQPIQLS------
sea urchin  S-PLVTGSSLGNKPKPLQLS------
```
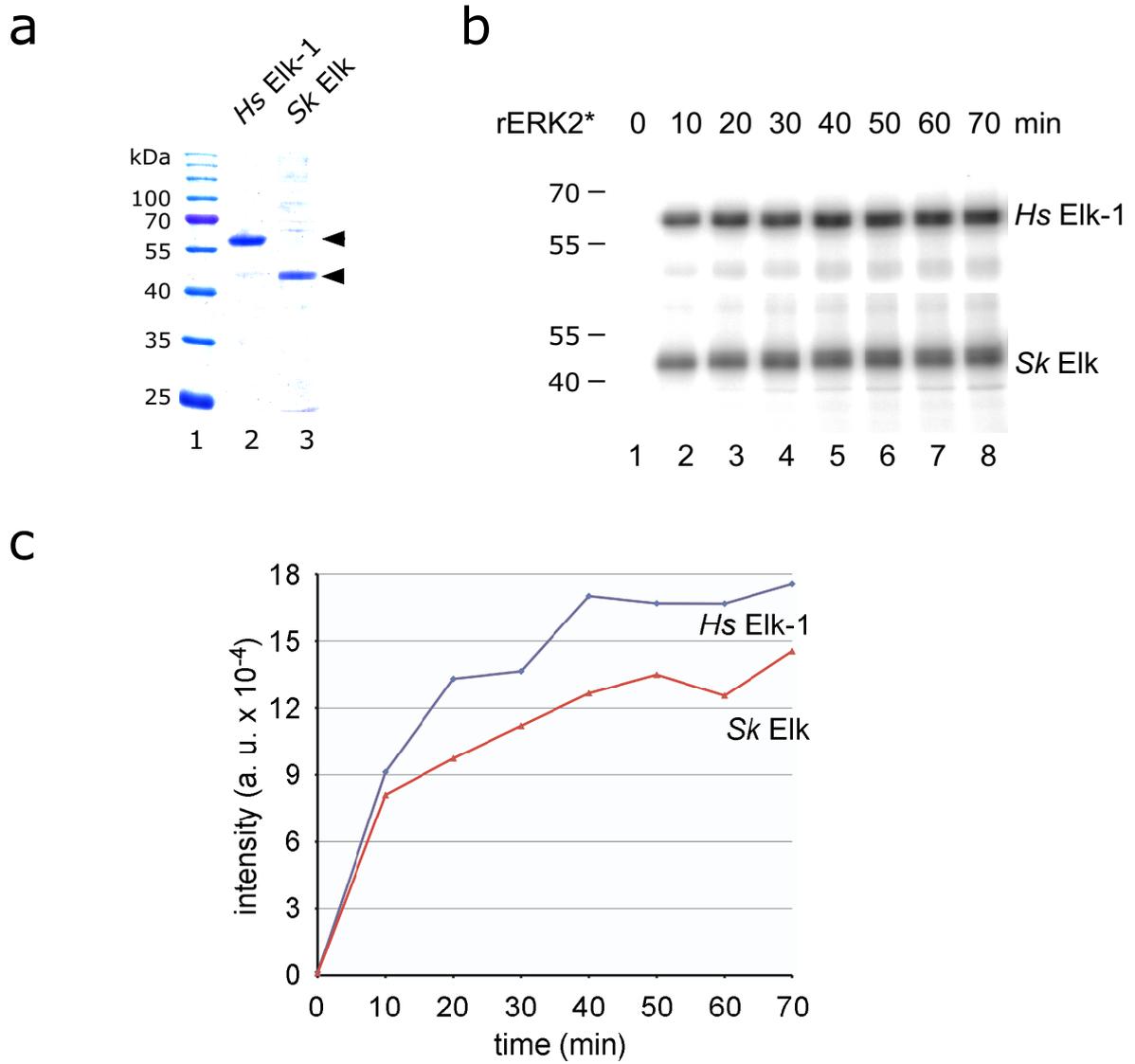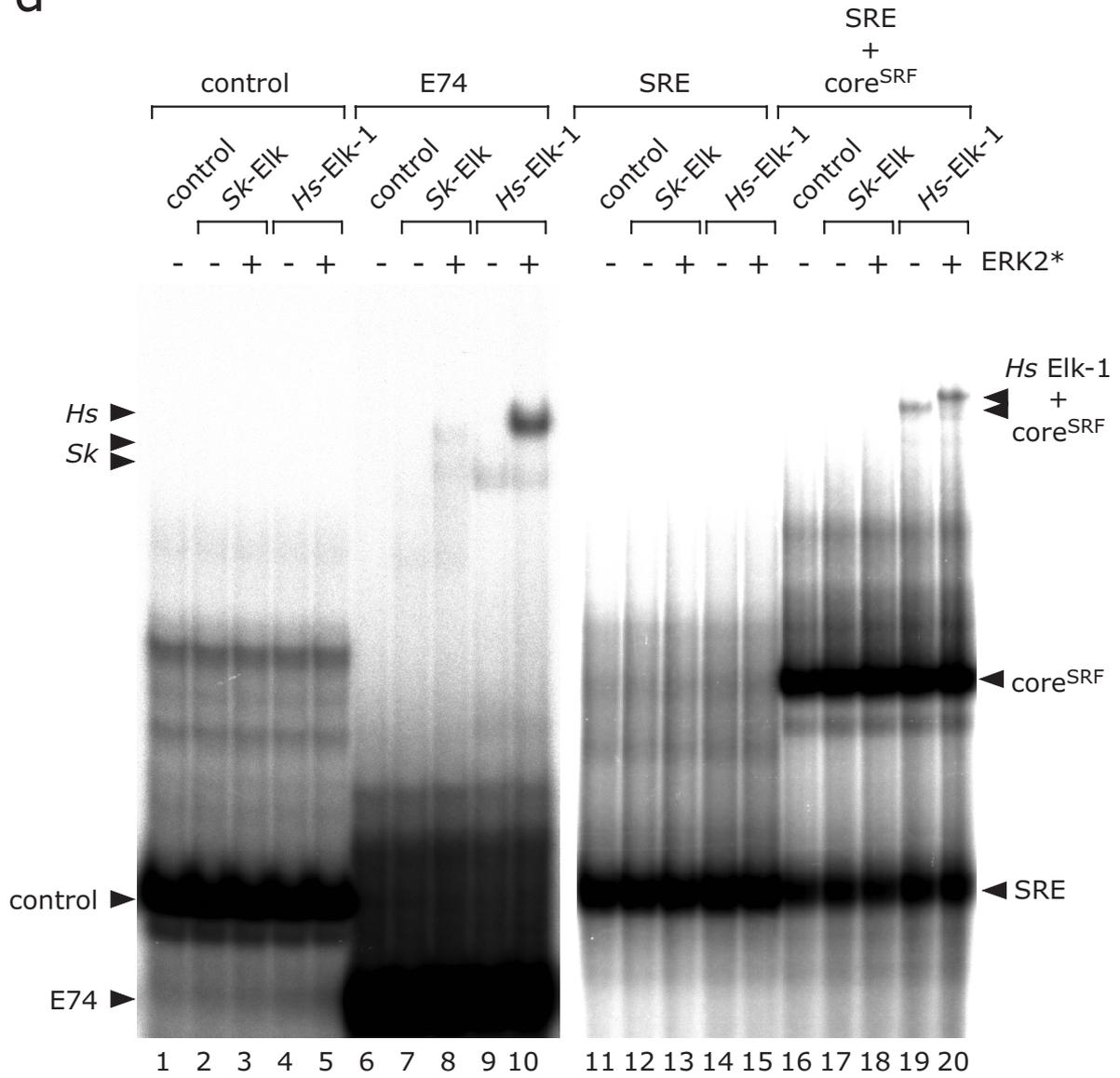
FIGURE 4ac.
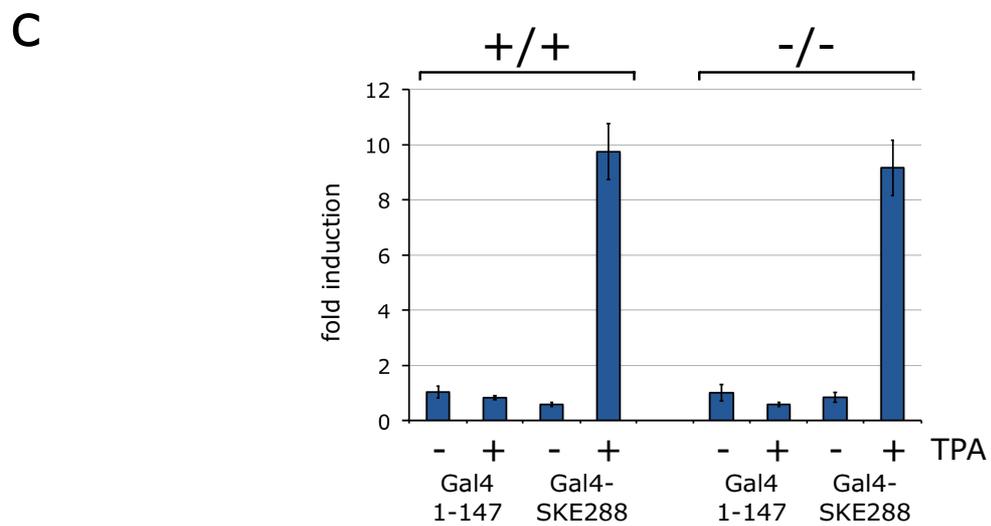
a

b

c

FIGURE 4d.

d

control   E74   SRE   SRE + core^SRF

control Sk-Elk Hs-Elk-1 control Sk-Elk Hs-Elk-1 control Sk-Elk Hs-Elk-1 control Sk-Elk Hs-Elk-1

- - + - +   - - + - +   - - + - +   - - + - +   ERK2*

Hs ▶
Sk ▶

Hs Elk-1 + core^SRF ◀

core^SRF ◀

control ▶

SRE ◀

E74 ▶

1 2 3 4 5   6 7 8 9 10   11 12 13 14 15 16 17 18 19 20

21

FIGURE 5.

FIGURE 6.