New mode of DNA binding of multi-zinc finger transcription factors: δ EF1 family members bind with two hands to two target sites

Jacques E.Remacle, Harry Kraft,
Walter Lerchner¹, Gunther Wuytens,
Clara Collart, Kristin Verschuere,
James C.Smith¹ and Danny Huylebroeck²

Department of Cell Growth, Differentiation and Development (VIB-07), Flanders Interuniversity Institute for Biotechnology (VIB) and Laboratory of Molecular Biology (CELGEN), University of Leuven, Herestraat 49, B-3000 Leuven, Belgium and ¹Division of Developmental Biology, National Institute for Medical Research, The Ridgeway, London NW7 1AA, UK

²Corresponding author
e-mail: dhu@sgi.kuleuven.ac.be

SIP1, a Smad-interacting protein, and δ EF1, a transcriptional repressor involved in skeletal and T-cell development, belong to the same family of DNA binding proteins. SIP1 and δ EF1 contain two separated clusters of zinc fingers, one N-terminal and one C-terminal. These clusters show high sequence homology and are highly conserved between SIP1 and δ EF1. Each zinc finger cluster binds independently to a 5'-CACCT sequence. However, high-affinity binding sites for full-length SIP1 and δ EF1 in the promoter regions of candidate target genes like *Xenopus Xbra2*, and human α 4-integrin and E-cadherin, are bipartite elements composed of one CACCT and one CACCTG sequence, the orientation and spacing of which can vary. Using transgenic *Xenopus* embryos, we demonstrate that the integrity of these two sequences is necessary for correct spatial expression of a *Xbra2* promoter-driven reporter gene. Both zinc finger clusters must be intact for the high-affinity binding of SIP1 to DNA and for its optimal repressor activity. Our results show that SIP1 binds as monomer and contacts one target sequence with the first zinc finger cluster, and the other with the second cluster. Our work redefines the optimal binding site and, consequently, candidate target genes for vertebrate members of the δ EF1 family.

Keywords: δ EF1/Smad-interacting protein/Smad proteins/transgenic *Xenopus*/two-handed zinc finger proteins

Introduction

Zinc fingers are among the most common DNA binding motifs found in eukaryotes. It is estimated that there are 500 zinc finger proteins encoded by the yeast genome and perhaps 1% of all mammalian genes encode such proteins. Zinc fingers can be found in many copies (up to 37) in proteins, and are often organized in tandem array, forming a single or multiple clusters. The prototype members of several families of transcription factors have the same overall structure by having two or three widely separated

clusters of zinc fingers. One, the MBP/PRDII-BF1 family, includes *Drosophila Schnurri* and *Spalt* genes (Fan and Maniatis, 1990; van't Veer *et al.*, 1992; Kuhnlein *et al.*, 1994; Arora *et al.*, 1995; Grieder *et al.*, 1995). Both MBP-1 (PRDII-BF1) and MBP-2 contain two separated clusters of two zinc fingers of the CCHH type. The neural-specific zinc finger factors 1 and 3 (NZF-1 and NZF-3), as well as the myelin transcription factor 1 (MyT1, NZF-2), belong to another family whose members contain two separated clusters of CCHC zinc fingers (Kim and Hudson, 1992; Jiang *et al.*, 1996; Yee and Yu, 1998). The *Drosophila* Zfh-1 and the vertebrate δ EF1 proteins (also known as ZEB or AREB6, in human) belong to a third family, which is characterized by the presence of two separated clusters of CCHH zinc fingers and a homeodomain-like segment (Fortini *et al.*, 1991; Funahashi *et al.*, 1993; Watanabe *et al.*, 1993; Genetta *et al.*, 1994).

Common features can be seen for all the members of these three families of multi-zinc finger transcription factors. In each family, the different members have high sequence conservation within their respective zinc finger clusters. Furthermore, for all these proteins, the C- and N-terminal zinc finger clusters are highly conserved, and these clusters have been shown to bind to very similar target sequences. This suggests that these factors would bind to reiterated sequences, but in fact this specific feature remains to be demonstrated.

In general, the precise mode of DNA binding remains poorly understood for these multi-zinc finger proteins. Here we describe the DNA binding properties of δ EF1 and SIP1. SIP1 is a novel member of the δ EF1 family, which we recently isolated and characterized as a DNA binding transcriptional repressor that binds in a ligand-dependent fashion to receptor-activated Smads involved in bone morphogenetic protein (BMP) and activin/transforming growth factor (TGF)- β pathways (Verschuere *et al.*, 1999). Like δ EF1, SIP1 contains two widely separated zinc finger clusters. One cluster (referred to as SIP1_{NZF}) of four zinc fingers (three CCHH fingers and one CCHC finger) is located in the N-terminal part of the protein, and another cluster (named SIP1_{CZF}) of three CCHH zinc fingers is present in the C-terminal part (Verschuere *et al.*, 1999). A high degree of sequence identity exists within the NZF (88%) and CZF (93%) clusters between SIP1 and δ EF1, whereas the two proteins are significantly less conserved in the regions outside the zinc fingers (Verschuere *et al.*, 1999). This indicates that SIP1 and δ EF1 may bind to similar DNA target sites and that each cluster plays an important role in the binding of the full-length proteins. Indeed, SIP1_{CZF} was shown to bind to several CACCT-containing sequences known to bind δ EF1 (Verschuere *et al.*, 1999). In addition, the δ EF1_{NZF3+NZF4} subdomain shows high homology (67%) with the δ EF1_{CZF2+CZF3} subdomain (Funahashi *et al.*,

Table I. List of all the DNA probes used in this study

Oligo	Sequence	Spacing
Xbra-WT	ATCCAGGCC CACCT AAAAATATAGAATGATAAAGTGACC AGGTG TCAGTTCT	24
Xbra-D	-----A-----	
Xbra-E	TAAAGTGACC AGGTG TCAGTTCT	
Xbra-F	ATCCAGGCC CACCT AAAAATATAGAATGA	
Rdm + Xbra-E	CAATTAGAGTACTGTGTACTTGGGAGTAAAGTGACC AGGTG TCAGTTCT	
Xbra-F + AREB6	ATCCAGGCC CACCT AAAAATATAGAATGAGGCTCAGAC AGGTG TAGAATTCGGCG	23
Rdm + AREB6	CAATTAGAGTACTGTGTACTTGGGAGGCTCAGAC AGGTG TAGAATTCGGCG	
Xbra-WT	ATCCAGGCC CACCT AAAAATATAGAATGATAAAGTGACC AGGTG TCAGTTCT	24
Xbra-J	CGA-----	
Xbra-K	---ACT-----	
Xbra-L	-----TAA-----	
Xbra-M	-----CAA-----	
Xbra-N	-----GCC-----	
Xbra-O	-----CCG-----	
Xbra-P	-----CGC-----	
Xbra-Q	-----TCC-----	
Xbra-R	-----GTC-----	
Xbra-S	-----T-----	
Xbra-WT	ATCCAGGCC CACCT AAAAATATAGAATGATAAAGTGACC AGGTG TCAGTTCT	24
Xbra-B	ATCCAGGCC CACCTA TATAGAATGATAAAGTGACC AGGTG TCAGTTCT	21
Xbra-C	ATCCAGGCC CACCT AAAAATATAGAATGAT GTGACC AGGTG TCAGTTCT	21
Xbra-U	ATCCAGGCC CACCT AAAAATATA GTGACC AGGTG TCAGTTCT	14
Xbra-EE	TAAAGTGACC AGGTG TCAGTTCTTAAAGTGACC AGGTG TCAGTTCT	18
Xbra-ErE	AGAACTGAC CACCT GGTCACTTTATAAAGTGACC AGGTG TCAGTTCT	20
Xbra-FrF	ATCCAGGCC CACCT AAAAATATAGAATATTCTATATTT AGGTG GCCTGGAT	24
Xbra-V	ATCCAGGCC AGGTG TAAATATAGAATGATAAAGTGACC CACCT ACAGTTCT	24
Xbra-W	ATCCAGGCC AGGTG TAAATATAGAATGATAAAGTGACC AGGTG TCAGTTCT	24
Xbra-Mut1	-----A-----	24
Xbra-Mut2	-----T-----	24
Xbra-Mut3	ATCCAGGCC CACCTA TATAGAATGATAAAGTGACC AGGTG TCAGTTCT	21
Xbra-Mut4	-----TGG-----	24
α4I-WT	GCAGGGCA CACCT GGATTGCATTAGAATGAGACTCACTACCCAGTTC AGGTG TGTTGCGT	34
α4I-A	-----A-----	
α4I-B	-----T-----	
Ecad-WT	TGGCCGGC AGGTG AACCTCAGCCAATCAGCGGTACGGGGGGCGGTGCTCCGGGGCT CACCT GGCTGCAG	44
Ecad-A	-----T-----	
Ecad-B	-----A-----	

Only the sequence of one strand of the probes is given. The CACCT and AGGTG sequences have been highlighted in bold. The spacing (right column) is the number of nucleotides, which separate the two CACCT sequences. Underlined gaps in some Xbra probes correspond to deletions of nucleotides from the wild-type (XbraWT) probe. For many probes, only the residues that have been changed compared with the wild-type probes are given in order to facilitate interpretation of the introduced mutations.

1993) and like $\delta\text{EF1}_{\text{CZF}}$, $\delta\text{EF1}_{\text{NZF}}$ binds to very similar CACCT consensus sequences. The same homology can also be seen for SIP1, suggesting that SIP1 and δEF1 may bind to reiterated CACCT elements.

We report here that high-affinity binding sites for full-length SIP1 and δEF1 are composed of two widely spaced CACCT sequences. In addition, the integrity of both zinc finger clusters is necessary for SIP1 to bind to its target, and SIP1 binds as a monomer. A new model for DNA binding for this family of transcription factors is proposed that may be extended to other factors containing multiple zinc finger clusters.

Results

Two CACCT sites are necessary for the binding of SIP1 to the Xbra2 promoter

Overexpression of either SIP1_{CZF} or full-length SIP1 (SIP1_{FS}) in *Xenopus* embryos repressed endogenous

Xbra2 expression (Verschuere *et al.*, 1999). Both SIP1_{CZF} and full-length SIP1 were shown to bind to a 50 bp probe from the *Xbra2* promoter (XbraWT; see Table I) containing two CACCT sites (Verschuere *et al.*, 1999). In addition, mutation of the downstream CACCT was shown to abolish the binding of both full-length SIP1 and SIP1_{CZF} (Verschuere *et al.*, 1999). Therefore, this downstream site is necessary for the binding of full-length SIP1. To elucidate further the binding requirements of SIP1 polypeptides to this *Xbra2* promoter, we compared their binding to XbraWT with binding to probes containing either the upstream (XbraF) or downstream (XbraE) CACCT sequence (Figure 1). We observed that SIP1_{CZF} bound to XbraE (Figure 1A, lane 2) and XbraWT (lane 1) with equal affinity, but does not bind to XbraF (lane 3). Therefore, the downstream AGGTG site is essential for binding of SIP1_{CZF} , and SIP1_{CZF} binds exclusively to this site. Extracts containing SIP1_{NZF} displayed binding patterns similar to SIP1_{CZF} (data not shown). As in δEF1 (Ikeda and

Fig. 1. Gel retardation assay with different probes from the *Xbra2* promoter. (A) Binding of SIP1_{CZF} and of proteins present in an extract prepared from mock-transfected COS1 cells. (B) Binding of SIP1_{FS}. (C) Expression levels of SIP1_{CZF} and SIP1_{FS}. Ten micrograms of the COS1 cell extract were analyzed by Western blotting using anti-Myc antibody. (D) Competition of the SIP1 binding to XbraWT. (E) Binding of SIP1_{CZF} and SIP1_{FS} to different hybrid probes. The different Xbra³²P-labeled probes (10 pg) were incubated with 1 µg of total protein extract from COS1 cells transfected with pCS3-SIP1_{CZF}, pCS3-SIP1_{FS} or from mock-transfected cells. The SIP1_{CZF}-specific complexes are indicated with gray arrows and the SIP1_{FS}-specific complexes are indicated with a black arrow. In competition experiments, 5 and 50 ng of unlabeled DNA were added together with the labeled probe. All other (non-specific) complexes originated from DNA binding activities also present in mock-transfected COS1 cells.

Kawakami, 1995; Sekido *et al.*, 1997), both zinc finger clusters of SIP1 thus have similar DNA binding features when tested individually.

SIP1_{FS} was shown previously to bind to XbraWT (Verschuere *et al.*, 1999; included in Figure 1B, lane 1, as positive control). The amount of SIP1_{CZF} produced in COS cells was ~50-fold higher than that of SIP1_{FS} (Figure 1C). However, the binding of SIP1_{FS} and SIP1_{CZF}

to this probe was equally strong. This indicates that the affinity of SIP1_{FS} for XbraWT is at least 50 times higher than SIP1_{CZF}. Remarkably, and in contrast to SIP1_{CZF} and SIP1_{NZF}, which bound with similar affinities to XbraWT and XbraE, SIP1_{FS} did not bind to XbraE at all (Figure 1B, lane 2). Like SIP1_{CZF} and SIP1_{NZF}, SIP1_{FS} did not bind to XbraF (lane 3). Moreover, competition experiments (Figure 1D) revealed that the presence of 50 ng of unlabeled XbraWT strongly reduced the binding of SIP1, whereas 50 ng of either unlabeled XbraE, XbraF or an unrelated GATA binding site did not affect it. We conclude that the downstream site (AGGTG) in the *Xbra2* promoter, which was previously shown to be needed for binding of SIP1_{FS} (Verschuere *et al.*, 1999), is not sufficient because additional sequences upstream of XbraE are necessary.

One reason why SIP1_{FS} is unable to bind to XbraE may simply be the length of this probe, because it is shorter than XbraWT. To test this, we prepared another probe, equal in length to XbraWT, but containing a random sequence (Rdm) upstream of XbraE (Rdm+XbraE, Table I). In contrast to SIP1_{CZF}, which bound efficiently to Rdm+XbraE (Figure 1E, lane 6), SIP1_{FS} was unable to bind (lane 3). This result demonstrates that the length of the XbraE probe *per se* is not the cause of the failure of SIP1_{FS} to bind.

To substantiate whether XbraF also contains sequences necessary for the binding of SIP1_{FS}, we fused its sequence as well as a random sequence upstream of another CACCT site known to bind AREB6 protein strongly (Ikeda and Kawakami, 1995) (probes XbraF+AREB6 and Rdm+AREB6, respectively). SIP1_{CZF} bound to both probes (Figure 1E, lanes 4 and 5). However, SIP1_{FS} bound only to XbraF+AREB6 (Figure 1E, lane 1) but not to Rdm+AREB6 (lane 2). The only feature common to the XbraE and the AREB6 probes is the CAGGTGT sequence. We therefore conclude that no sequences other than this CAGGTGT in XbraE are necessary for the binding of SIP1_{FS}.

To map the sequences within XbraF that, in conjunction with XbraE, are required for the binding of SIP1_{FS}, we prepared a series of probes identical in length to XbraWT and containing adjacent triple mutations within the XbraF part. Only three of these triple mutations (i.e. L, M and N) affected the binding of SIP1_{FS} (Figure 2A). These mutations all destroyed the upstream CACCT site present in XbraF. We also showed that SIP1_{FS} does not bind to XbraS, which contains a point mutation, changing CACCT into CATCT (Figure 2A, lane 12). This mutation is thus similar to the downstream AGATG mutation within XbraD, which was previously shown not to bind SIP1_{FS} (Verschuere *et al.*, 1999; included in Figure 2A, lane 11, as control). The results indicate that SIP1_{FS} contacts both CACCT sites in the *Xbra2* promoter.

To investigate the importance of these sites further, a DNA methylation interference assay was carried out (Figure 2B). The methylation of the three Gs of the downstream AGGTG (SIP_{DO}) and the three Gs of the upstream CACCT (SIP_{UP}) was significantly lower in the SIP1_{FS} bound versus unbound probe, suggesting that the methylation of these Gs interfered with the binding of SIP1_{FS}. This is strong evidence that these residues are essential for SIP1_{FS} binding. We also observed

that the methylation of one of the two Gs localized very close to the SIP_{DO} also interfered with the binding of SIP1_{FS} (middle lane, right panel). Consequently, we showed for SIP1_{FS} that two CACCT sequences, and their integrity, are required for DNA binding.

Using protein extracts prepared from COS cells expressing δ EF1, we demonstrated that δ EF1 also binds to Xbra^{WT} and, as for SIP1, the integrity of both CACCT sequences is required for binding (Figure 2C).

Mutations in either the upstream or downstream CACCT lead to ectopic activity of the Xbra2 promoter in transgenic frog embryos

SIP1 binds to the Xbra2 promoter and represses expression of endogenous Xbra2 mRNA when overexpressed in *Xenopus* embryos (Verschuere *et al.*, 1999). To analyze the importance of CACCT sequences in the regulation of the Xbra2 promoter *in vivo*, we tested whether mutations of these would affect Xbra2 promoter activity in transgenic embryos. Xbra2 promoter sequences were fused upstream of the green fluorescent protein (GFP) gene and this reporter cassette was used for transgenesis. A 2.1 kb Xbra2 promoter fragment was shown to be sufficient to yield the reporter protein synthesis in the same domain of the embryo (85% of the embryos, stage 11, $n = 57$) as compared with endogenous Xbra mRNA (which is in the marginal zone) except in the organizer region (W.Lerchner, personal communication), for which a regulatory element may be lacking in the reporter cassette tested here.

A single point mutation within the downstream CACCT site in the promoter, which disrupted SIP1 binding (Xbra2-Mut1; Figure 3A, lane 2) and is identical to XbraD, had a severe effect on spatial production of the reporter protein. All embryos ($n > 30$) showed ectopic expression in the inner ectodermal layer (Figure 3B). Mutations within the upstream CACCT sequence (Xbra2-Mut4) also affected SIP1 binding (Figure 3A, lane 3); we observed in all transgenic embryos ($n > 30$) the same ectopic expression as for the Xbra2-Mut1 mutation (Figure 3B). Mutation of the downstream CACCTG to CACCTA (Xbra2-Mut2) also affects SIP1 binding to the probe (Figure 3A, lane 4). This mutation when introduced into the Xbra2 2.1 kb promoter also led to ectopic expression of GFP mRNA in all transgenic embryos tested ($n > 30$; Figure 3B). We also tested a mutation (Xbra2-Mut3) that decreased by 3 bp the original 24 bp spacing between the two CACCT sequences. This mutation weakened the interaction of the probe with SIP1 (Figure 3A, lane 5). This was also reflected in the corresponding transgenic embryos ($n = 37$): while 35% of the embryos showed the same expression pattern as the wild-type Xbra2 2.1 kb promoter fragment, 65% had either patches or weak continuous expression in the inner ectodermal cells (Figure 3B).

A good correlation between the effect of these mutations on SIP1 binding affinity in electrophoretic mobility shift assay (EMSA) and the phenotype (ectopic expression of the reporter gene), and its penetrance *in vivo*, was thus obtained, indicating the importance of the SIP1 target sites in the normal regulation of Xbra2 expression in *Xenopus* development (stage 11). It also suggests that a hitherto unknown *Xenopus* SIP1-like repressor regulates Xbra2

Fig. 2. Two CACCT sites are contacted upon binding of SIP1_{FS} to the Xbra2 promoter. (A) Only mutations within the upstream CACCT sequence (as revealed by scanning mutagenesis, see Table I) or the downstream CACCT sequence (see elsewhere in Table I) of Xbra^{WT} abolish SIP1_{FS} binding. (B) Methylation interference assay indicates that SIP1_{FS} contacts both CACCT sequences. Xbra^{WT} either labeled in the upper (left panel) or the lower (right panel) strand were methylated and incubated with total extract from COS1 cells transfected with pCS3-SIP1_{FS}. The DNA retarded in the shifted complex or the unbound DNA (FREE) were purified, cleaved with piperidine and run onto a sequencing gel. The arrows indicate the G residues that are methylated in the free probe. SIP_{UP} and SIP_{DO} indicate the upstream and downstream CACCT from the Xbra2 promoter, respectively. (C) δ EF1 binds to the Xbra2 promoter.

gene expression *in vivo*. In addition, it confirms that SIP1-like factors require two intact CACCT sites for regulating target promoters like *Xbra2*.

Binding of full-length SIP1 and δ EF1 to other candidate target promoters

The fact that two CACCT sites are required for the binding of full-length SIP1 and δ EF1 may be unique for the *Xbra2* promoter. Two CACCT sequences that are also part of E2 boxes are present in the promoter of the human α 4-integrin gene (Postigo and Dean, 1997). Combined mutation of these two CACCT sites led to the derepression of α 4-integrin gene expression in myoblasts, and overexpression of δ EF1 downregulated α 4-integrin mRNA levels (Postigo and Dean, 1997). Since these two sites are closely positioned in the promoter (the spacing is 34 bp), we investigated whether they are required for the binding of δ EF1 and SIP1. A 60 bp probe encompassing both sites

Fig. 3. Ectopic activity of the mutated *Xbra2* promoter variants (*Xbra2*-Mut) in transgenic frog embryos. (A) SIP1 binding to the wild-type and mutated (*Xbra2*-Mut; see Table I) *Xbra2* promoter elements. (B) Whole-mount *in situ* hybridization for GFP mRNA of *Xenopus* embryos transgenic for a wild-type or point-mutated 2.1 kb *Xbra2* promoter fragment driving a GFP reporter. All shown embryos were fixed at stage 11 and cleared for better visualization of the signal. Percentages are indicative of intermediary phenotype (i.e. 35% of transgenic embryos displayed the normal *Xbra2* expression pattern and 65% showed ectopic expression). Marginal zone (MZ), inner ectodermal layer (IEL) and outer ectodermal layer (OEL) are indicated.

of the α 4-integrin promoter was synthesized (α 4IWT) together with two mutated versions, of the same length but with a point mutation in either the upstream (α 4IB) or the downstream (α 4IA) CACCT site, respectively (see Table I). SIP1_{FS} forms a strong complex with the α 4IWT probe, which is entirely supershifted in the presence of anti-Myc antibody (Figure 4A). Although SIP1_{CZF} is at least 50-fold more concentrated than SIP1_{FS} (see above), we see that SIP1_{CZF} shifted approximately the same amount of probe, indicating that (like for *Xbra2*WT) SIP1_{CZF} has

Fig. 4. Two CACCT sequences are necessary for the binding of SIP1_{FS} and δ EF1 to the *Xbra2*, α 4-integrin and E-cadherin promoters. (A) SIP1_{FS} and SIP1_{CZF} bind to the α 4-integrin (α 4I) promoter. (B) Mutations in either the upstream (α 4IB) or downstream (α 4IA) CACCT affect binding of SIP1 and δ EF1. (C) Binding of SIP1 and δ EF1 to the α 4-integrin promoter, including competition with excess of non-labeled wild-type and mutated binding sites. (D) Binding of SIP1 and δ EF1 to the E-cadherin promoter (Ecad). In competition experiments, 5 and 50 ng of unlabeled DNA were added together with the labeled probe. In lane 4 (A) and lane 7 (C), anti-Myc tag antibody was added to the binding reaction; the supershifted complex is indicated by an asterisk (*). The black arrows, grey arrows and filled diamonds indicate the retarded complex, caused by δ EF1, SIP1_{CZF} and SIP1_{FS} binding, respectively.

a 50-fold lower affinity than SIP1_{FS} for the α 4IWT site. Both δ EF1 (Figure 4B, lane 4) and SIP1 (lane 1) formed a strong complex with the α 4IWT probe. The δ EF1 complex was supershifted with an anti-Myc antibody (Figure 4B, lane 7). The binding of both SIP1 and δ EF1 was abolished or strongly affected, respectively, by the specific mutation of either the upstream or the downstream site (lanes 2–3 and 5–6). Moreover, competition experiments (Figure 4C) revealed that 50 ng of unlabeled α 4IWT were sufficient to abolish the binding of SIP1 or δ EF1, whereas 50 ng of either unlabeled α 4IA or α 4IB mutant probes were not. We conclude that SIP1 and δ EF1 require the integrity of two CACCT sites for binding to the corresponding segment of the α 4-integrin promoter.

Two CACCT sites (spaced by 44 bp) are present in the human E-cadherin promoter. A probe comprising these (EcadWT) was used in EMSA (Figure 4D). Both SIP1 (lane 1) and δ EF1 (lane 4) formed a complex with this probe. When either the upstream (in EcadA) or the downstream (in EcadB) CACCT site was mutated, the binding of SIP1 and δ EF1 was abolished or very strongly reduced. This confirms, with different promoters, that the two spaced CACCT sites represent a high-affinity site for the binding of SIP1 and δ EF1.

From the alignment of the Xbra, α 4I and Ecad wild-type probes (Table I), we observed no obvious homology, except for one CACCTG site (also an E2 box) and a second CACCT site. Our results described above and this alignment therefore strongly indicate that only those promoter sequences participate in the binding of either SIP1 or δ EF1, which requires at least one CACCT site and one CACCTG site.

Spacing variations and orientation of the CACCT sites

Within the Xbra2, α 4-integrin and E-cadherin promoters, the spacing between the two CACCT sequences is 24, 34 and 44 bp, respectively. SIP1 and δ EF1 bind to the respective probes, suggesting that these proteins can accommodate variable inter-CACCT spacing. To investigate further whether this spacing is important, we generated another panel of Xbra2 probes. Two (XbraB and -C) have a deletion of 3 bp in the spacing whereas XbraU has one of 10 bp. The three probes were tested in EMSA with SIP1 or δ EF1 (Figure 5). Both proteins bound XbraWT, -B, -C and -U probes *in vitro* (lanes 1–4). This indicates that within the same promoter element, the spacing between the two CACCT sites is not critical for the binding of these two factors.

By sequence comparison of the Xbra, α 4I and Ecad wild-type probes, we observed that in the case of Xbra and α 4I, the relative orientation of the two CACCT sites is CACCT-N_n-AGGTG, whereas in Ecad it is AGGTG-N_n-CACCT. Because of the non-palindromic nature of the CACCT site, these two sites could be assumed to be substantially different. However, SIP1 and δ EF1 bound to both promoter segments with comparable affinity (see above). This suggests that full-length SIP1 and δ EF1 can bind, irrespective of the relative orientation of the two CACCT sites. To investigate further the orientation of the two CACCT sites with respect to the DNA binding capacity of SIP1 and δ EF1, additional probes were

Fig. 5. Spacing and relative orientation of the CACCT sequences are not critical for the binding of SIP1_{FS} and δ EF1 to the Xbra2 promoter. For details on the design of the different probes, see Results. In lane 10, we mixed 10 pg of XbraE with 10 pg of XbraF. For reasons of clear and comparative presentation, we omitted the free probe from the SIP1-binding reactions (upper part).

designed. XbraEE contains a tandem repeat of XbraE, whereas probe XbraErE contains an inverted repeat of XbraE. We also made XbraV, in which the upstream CACCT site (plus one extra base pair on each side) was replaced by the downstream AGGTG sequence, and vice versa. Finally, in XbraW, only the upstream site was replaced by the downstream CACCT sequence. We observed the strongest binding of both proteins to XbraEE (Figure 5, lane 5). Therefore, full-length SIP1 and δ EF1 cannot bind to XbraE containing a single CACCT site (Figures 1B and 2C, respectively), but bind strongly when this sequence is duplicated. Both proteins bound to XbraErE (Figure 5, lane 6), demonstrating that the relative orientation of both CACCT sequences is not critical for binding. Switching of the sites (XbraV and -W) did not affect the binding of both proteins either. Thus, neither the spacing between the two CACCT sites nor their respective orientation is critical for DNA binding of SIP1 and δ EF1 *in vitro*.

Not all duplicated and spaced CACCT sites are able to bind these proteins. XbraF, which in combination on the same DNA with XbraE (i.e. forming the XbraXW probe) was shown to be necessary for binding, was refractory to binding when duplicated [Figure 5, lane 9 for the inverted repeat (XbraFrF) and (data not shown) for the direct repeat]. This suggests that the CACCT site within the XbraF context is a low-affinity site and that sequences adjacent to it optimize the binding. In addition, the fact that neither SIP1_{CZF} (Figure 1A) nor SIP1_{NZF} (data not shown) could bind independently to XbraF confirms the assumption that this upstream site displays low affinity. In contrast, the downstream CACCTG site in XbraE can bind SIP1_{CZF} and SIP1_{NZF}. In fact duplication of this element creates a high-affinity binding site for both SIP1 and δ EF1 (see above). This suggests that the extra G in

Fig. 6. The integrity of zinc finger clusters is necessary for the binding of SIP1_{FS} to DNA. (A) Mutations within NZF3, NZF4, CZF2 and CZF3 abolish DNA binding of short SIP1_{NZF} and SIP1_{CZF} polypeptides. The wild-type and mutated zinc finger clusters (as NZF or CZF polypeptide) were fused to GST, and the fusion proteins were produced in *Escherichia coli*. After purification, an equal amount of each fusion protein (0.1 ng) was incubated with 10 pg of labeled Xbra-E probe. (B) Mutations of either NZF3 (lanes 2 and 7), NZF4 (lane 3), CZF2 (lane 4) or CZF3 (lanes 5 and 8) in full-length SIP1 affect the binding of SIP1 to XbraWT. In lanes 9–14, all possible combinations of two extracts (1 μg of each) expressing different SIP1 mutants were tested. In lanes 6–8, we added anti-Myc tag antibody to the binding reaction. The supershifted complex is indicated by an asterisk. The arrow indicates the SIP1_{FS}-containing complex (lanes 1 and 6). (C) Mutations within NZF3, NZF4, CZF2 or CZF3 abolish the binding of SIP1_{FS} to α4IWT. Lane order is identical to Figure 6B. (D) SIP1 mutants are produced in similar amounts in COS1 cells. Ten micrograms of the cell extract were analyzed by Western blotting using anti-Myc antibody.

the downstream site may discriminate a high- from a low-affinity site. Mutation in the Xbra2 probe of the downstream site to CACCTA (Xbra2-Mut2) strongly affected the binding of SIP1_{FS} (Figure 3A). This supports the importance of this G for high-affinity binding. Finally, when the short XbraE and XbraF probes were mixed prior to addition of SIP1 or δEF1, we did not observe any binding, again indicating that both CACCT sites have to be in the *cis* configuration (Figure 6, lane 10).

Mutations from the protein side: the two zinc finger clusters of SIP1 are required, and must be intact for binding to DNA and repressor activity

In this part of our work, we wanted to evaluate the importance of each zinc finger cluster for the binding of SIP1 to DNA. Mutations destroying either the third or the fourth zinc finger of δEF1_{NZF} were shown to abolish the binding of this isolated cluster to DNA. Similarly, mutagenesis of the second or the third zinc finger in

Fig. 7. The integrity of CZF or NZF is necessary for SIP1 repressor activity. (A) SIP1_{FS} binding to a gel-purified fragment derived from the multiple CACCT-containing artificial promoter from reporter plasmid p3TP-Lux. In lane 2, we added anti-Myc tag antibody; the supershifted complex is indicated by an asterisk. (B) Effect of transfection of pCS3-SIP1_{FS}, pCS3-SIP1_{FS}-NZF3-Mut, pCS3-SIP1_{FS}-CZF3-Mut or an expression vector encoding the double mutant with the p3TP-Lux reporter vector. The activities are expressed in relative light units, and were normalized for transfection efficiency.

δEF1_{CZF} also abolished the binding to CACCT (Ikeda and Kawakami, 1995). Therefore, we first introduced in SIP1_{NZF} and SIP1_{CZF} similar mutations. The mutated and wild-type clusters were fused to glutathione *S*-transferase (GST) and the fusion proteins were purified from bacteria. Figure 6A shows that both wild-type SIP1_{NZF} (lane 1) and SIP1_{CZF} (lane 4) bound strongly to XbraE. However, binding could not be detected with the same amount of purified mutant cluster–GST fusion proteins (GST–NZF3, GST–NZF4, GST–CZF2 and GST–CZF3; lanes 2–3 and 5–6). Thus, these mutations abolish the capacity of each cluster (SIP1_{NZF} and SIP1_{CZF}) to bind independently to a CACCT site.

We subsequently introduced these mutations in full-length, Myc-tagged SIP1 (NZF3-Mut, NZF4-Mut, CZF2-Mut and CZF3-Mut). The production of the mutant proteins in COS cells was confirmed, and their amounts normalized by Western blotting (Figure 6D). We observed (Figure 6B) that wild-type SIP1 bound strongly to XbraWT (lane 1); the SIP1 complex was supershifted with an anti-Myc antibody (lane 6). In contrast, none of the mutant forms of SIP1 was able to form a complex either tested individually (lanes 2–5, 7 and 8) or in combination (lanes 9–14, see below). The same results were obtained with α4IWT probe (Figure 6C). In conclusion, full-length SIP1 requires the binding capacities of both (intact) zinc finger clusters to bind to its target.

The effect of these mutations on the repressor activity of SIP1 was tested in a transfection assay together with p3TP-Lux reporter plasmid. This plasmid contains three copies, each of which has one CACCT, of a sequence covering the –73 to –42 region of the human collagenase promoter (de Groot and Kruijer, 1990). SIP1 bound to a fragment containing this multimerized element (Figure 7A), but neither NZF3-Mut nor CZF3-Mut was

able to bind (data not shown). Overexpression of SIP1 in CHO cells leads to a strong repression of the p3TP-Lux basal transcriptional activity (Figure 7B). SIP1 mutants defective in DNA binding (NZF3-Mut or CZF3-Mut) are strongly affected in their repressor activity (Figure 7B), but the SIP1 repressing activity is totally abolished only when using a double NZF3/CZF3 SIP1 mutant. We conclude that the integrity of both zinc finger clusters is necessary for both the DNA binding and optimal, i.e. wild-type repressor activity of SIP1.

SIP1 binds to DNA as a monomer

The observation that the integrity of both zinc finger clusters of SIP1 is required for its binding to two CACCT sequences, prompted us to test whether SIP1 binds as a monomer, implying a model in which each zinc finger cluster contacts one of the two CACCT sites. However, it can be hypothesized that SIP1 also binds as a dimer. In the latter model, one of the SIP1 proteins would bind one CACCT site via its NZF, while the second would contact DNA via its CZF. According to this model, appropriate combinations of protein extracts prepared from cells producing either NZF3-Mut or NZF4-Mut with CZF2-Mut or CZF3-Mut full-length SIP1 (see above) should generate a SIP1 dimer still competent to bind DNA. As shown for the individual mutants (see Figure 6B), none of these combinations resulted in binding to XbraWT, indicating that SIP1 binds as a monomer.

To address this further, we used a combination of differentially tagged SIP1 proteins in supershift experiments in EMSA. First, we produced Myc-tagged and/or FLAG-tagged SIP1 separately to comparable levels in COS cells, and confirmed that both SIP1 proteins bound to DNA with similar affinity (data not shown). The SIP1 complex generated with Myc-tagged SIP1 had a slightly slower migration than the FLAG-tagged complex (the Myc tag is indeed longer than the FLAG tag). Similar amounts of both Myc-tagged and FLAG-tagged SIP1 were incubated with XbraWT. We observed the formation of a broad SIP1 complex (Figure 8A, lane 1), which was a combination of both the somewhat faster-migrating FLAG-tagged SIP1 complex and the slower Myc-tagged SIP1 complex. Using an anti-FLAG antibody, only the lower complex containing FLAG-tagged SIP1 was supershifted, whereas ~50% of the signal persisted within the Myc-tagged SIP1 complex (Figure 8, lane 2). This indicates that the latter SIP1 complex is not supershifted with the anti-FLAG antibody. Conversely, incubating the extract with an anti-Myc antibody supershifted only the upper part of the complex corresponding to Myc-tagged SIP1, whereas 50% of the signal was retained within the FLAG-tagged SIP1 complex (lane 3). These results show that no FLAG-tagged SIP1 was co-supershifted with SIP1 shifted by the anti-Myc antibody. Using both antibodies at the same time (lane 4), we observed the same two slower-migrating supershifted bands, which correspond to the Myc-tagged and FLAG-tagged supershifted complexes. If SIP1 dimers could be formed, then at least some heterodimers would be assembled from Myc-tagged SIP1 and FLAG-tagged SIP1. However, no other supershifted band that would correspond to a potential double supershift, i.e. supershifted with both anti-Myc- and anti-FLAG antibodies, was detected. Hence, this experiment

Fig. 8. SIP1 binds as a monomer to the XbraWT probe. (A) XbraWT probe was incubated with an extract prepared from COS1 cells expressing equal amounts of Myc-SIP1_{FS} and Flag-SIP1_{FS}. In lanes 2 and 3, we added anti-FLAG and anti-Myc tag antibodies, respectively, and both antibodies were added to the binding assay in lane 4. The FLAG- and the Myc-epitope containing supershifted complexes are indicated with an asterisk and a bullet, respectively. (B) The same extract was incubated with large excess of unlabeled XbraWT (50 ng) in gel retardation buffer. Next, Myc-SIP1_{FS} or Flag-SIP1_{FS} were immunoprecipitated (IP) using anti-Myc (left panel) or anti-Flag (right panel) antibody. Subsequently, these immunoprecipitates were subjected to Western blot (WB) analysis using either anti-Flag (left panel) or anti-Myc antibody (right panel).

gave no detectable formation of FLAG-tagged and Myc-tagged SIP1 heterodimers.

Finally, we incubated a large excess of XbraWT with equal amounts of FLAG-tagged and Myc-tagged SIP1. Subsequently, this reaction was immunoprecipitated with anti-FLAG antibody and the precipitate was analyzed by Western blotting using anti-Myc antibody. The co-immunoprecipitation of a Myc-tagged SIP1 within the FLAG-tagged SIP1 will support the formation of a dimer between Myc-tagged and FLAG-tagged SIP1 on the DNA. The reciprocal experiment, i.e. immunoprecipitation with an anti-Myc antibody and detection with an anti-FLAG antibody was also carried out. In both experiments, co-immunoprecipitation of Myc-tagged and FLAG-tagged SIP1 was not observed (Figure 8B). This argues against the formation of SIP1 dimer in the presence of DNA. Together, these results support a model in which SIP1 binds as a monomer to the XbraWT probe.

Discussion

SIP1 and δ EF1 bind target DNA sites containing one CACCT and one CACCTG sequence

We have studied the DNA binding properties of SIP1, a recently isolated Smad-interacting protein and a new vertebrate member of the emerging family of two-handed zinc finger transcription factors (Verschuere *et al.*, 1999). The organization of δ EF1, the prototype member of this family, and SIP1 is identical. Both proteins contain two widely separated clusters of zinc fingers, which are involved in binding to DNA (Ikeda and Kawakami, 1995; Sekido *et al.*, 1997; this study). The amino acid sequence

homology is highest (>90%) within these two clusters, suggesting that both proteins would bind in an identical fashion to similar DNA targets. We demonstrate here that full-length SIP1 and δ EF1 bind with high and comparable affinities to many different target sites, which always contain two CACCT sequences. For all the targets tested here, the integrity of both CACCT sequences is absolutely necessary for the binding of either SIP1 or δ EF1. When mutations (either in the upstream or in the downstream CACCT) affecting the binding of SIP1 were introduced in the 2.1 kb *Xbra2* promoter, we observed ectopic expression of the GFP marker in transgenic *Xenopus* embryos. The marker protein is then found in the majority of ectodermal and mesodermal cells, whereas *Bra* protein and GFP mRNA (in the *Xbra*WT transgene) are found exclusively in mesoderm. We believe that expression in these ectopic sites is the result of the absence of binding of a SIP1-like protein and its transcriptional repression. In addition, these *in vivo* data confirm our *in vitro* evidence, i.e. SIP1/ δ EF1-like transcription factors require two intact CACCT sites for binding to and regulation of the *Xbra2* promoter.

Not all promoter regions containing two CACCT sequences represent SIP1 or δ EF1 binding sites. Notably, a duplicated *XbraF* probe, which contains the upstream CACCT of *Xbra*WT, is refractory to binding of SIP1 and δ EF1. Although this upstream CACCT sequence is unable to bind SIP1_{CZF} or SIP1_{NZF}, it is contacted by full-length SIP1 in the context of the *Xbra*WT probe. Indeed, this upstream sequence is a prerequisite for the binding of SIP1_{FS} to *Xbra*WT. Thus, when the upstream CACCT is combined *in cis* with another high-affinity CACCTG site (*XbraE*), this low-affinity site (*XbraF*) can bind SIP1_{FS}. We therefore propose a model in which SIP1_{FS} contacts its target promoter via the binding of one of its zinc finger clusters to a high-affinity CACCTG sequence (e.g. *XbraE*), which is followed by the contact of the low-affinity CACCT site (*XbraF*) by the second cluster, and this additional contact strongly stabilizes SIP1 binding. Therefore, a CACCT site may still have an important function in the regulation of gene expression, while even on its own it does not bind SIP1_{NZF}, or SIP1_{CZF} and SIP1_{FS}.

The DC5 probe from the δ 1-crystallin enhancer was previously shown to bind δ EF1 specifically (Sekido *et al.*, 1997). This probe contains only one CACCT sequence. Therefore, despite having demonstrated here that high-affinity binding sites for δ EF1 should contain one CACCT sequence and one CACCTG sequence, which is new for the field, we cannot exclude that in particular cases (such as the DC5 probe) one CACCT site would be sufficient for the binding of this type of transcription factor. However, for testing the repressor activity of δ EF1, exclusively a reporter made with eight copies of the DC5 element was used (Sekido *et al.*, 1997). Therefore, the multimerization of this element generated reiterated CACCT sequences that may fortuitously have created a potential high-affinity site for δ EF1. SIP1_{FS} does not bind to *XbraE* whereas SIP1_{CZF} and SIP1_{NZF} bind to this probe. None of the full-length SIP1 proteins with zinc finger mutations (i.e. NZF3-Mut, NZF4-Mut, CZF2-Mut and CZF3-Mut) binds to *Xbra*WT containing two CACCT sites, despite the mutant proteins having one intact zinc finger cluster. It can be speculated that SIP1_{CZF} and SIP1_{NZF} are hindered in the

Fig. 9. Possible DNA binding mechanisms for SIP1. SIP1 binds DNA as a monomer to bipartite CACCT elements.

context of full-length wild-type SIP1 via intramolecular association. This association would be impaired only when the two CACCT sites are present, allowing both zinc finger clusters to bind DNA.

In addition, we demonstrated convincingly that the integrity of both zinc finger clusters is necessary for the binding of SIP1 to several bipartite elements. Similarly, it was shown previously that the integrity of both zinc finger clusters of δ EF1 is also necessary for its binding to DNA and repressor activity (Sekido *et al.*, 1997). However, mutant mice carrying CZF-deleted δ EF1 show a defect in T-cell development (Higashi *et al.*, 1997) while those completely lacking δ EF1 display multiple skeletal defects as well (Takagi *et al.*, 1998). Therefore, it was concluded that a form of δ EF1 lacking CZF can still support skeletal development normally. Consequently, we cannot exclude the possibility that some functionality of these transcription factors may result from the binding of a single zinc finger cluster to one CACCT target. In this respect, we saw that a mutant version of SIP1 containing only one intact zinc finger cluster still displayed a weak repressor activity in the 3TP-lux assay upon overexpression. This may result from the low-affinity binding of this overexpressed SIP1 mutant to one CACCT site.

Mode of SIP1 DNA binding

The independent CZF and NZF clusters of SIP1 or δ EF1 bind to CACCT-containing consensus sequences because within SIP1 and δ EF1, NZF3 and NZF4 share an extensive amino acid sequence homology with CZF2 and CZF3, respectively. We have shown that full-length SIP1 and δ EF1 require two CACCT sequences for binding to several potential target sites. In addition, we demonstrated that the integrity of both zinc finger clusters is necessary for the binding of SIP1 to these bipartite elements and that SIP1 binds as a monomer. We propose that SIP1 and δ EF1 would bind to their target elements by having one zinc finger cluster contacting one of the CACCT sites, while the other cluster contacts the second CACCT site (Figure 9).

We showed also that neither the relative orientation of the two CACCT sequences nor the spacing between these sequences is critical for the binding of SIP1_{FS} or δ EF1 *in vitro*. This suggests that the structure of these transcription factors is highly flexible. The long linker region between the two zinc finger clusters within SIP1 and δ EF1 may confer this flexibility. These transcription factors can bind to sites containing CACCT sequences separated by at least 44 bp (in *Ecad*), suggesting that a region of ~50 bp of promoter sequences might be less accessible to other transcription activators. This might be one mechanism by which SIP1 or δ EF1 could function as transcriptional repressors.

Other families of transcription factors may bind DNA with a mechanism similar to SIP1/ δ EF1

The new mode of DNA binding presented here for SIP1 and δ EF1 may also be generalized to other transcription factor families which, like SIP1 and δ EF1, contain separated clusters of zinc fingers, like those of the MBP/PRDII-BF1 family (Fan and Maniatis, 1990; van't Veer *et al.*, 1992; Seeler *et al.*, 1994; Arora *et al.*, 1995; Grieder *et al.*, 1995). As in SIP1 and δ EF1, the conservation of their zinc finger clusters is very high between the different members of this family (Arora *et al.*, 1995). In addition, their C-terminal cluster is very homologous to their N-terminal cluster and, in the case of PRDII-BF1, these clusters bind to the same sequences when tested independently (Fan and Maniatis, 1990). Therefore, PRDII-BF1 may indeed bind to two reiterated sequences via two zinc finger clusters. Similarly, well-studied members of the NZF family also have two widely separated clusters of zinc fingers (Kim and Hudson, 1992; Jiang *et al.*, 1996; Yee and Yu, 1998), and MyT1, NZF-1 and NZF-3 all bind to the same consensus element AAAGTTT. Like SIP1 and δ EF1, which show a significantly higher affinity to promoter segments containing two spaced CACCT sequences, an element containing two AAAGTTT sequences demonstrated a markedly higher affinity for NZF-3 (Yee and Yu, 1998). This suggests that two AAAGTTT sequences are also necessary to create a high-affinity binding site for these factors, and that they may bind DNA similarly to SIP1 and δ EF1. The Evi-1 protein, another type of zinc finger protein, with seven zinc fingers at the N-terminus and three zinc fingers at the C-terminus, binds to two targets. It binds to a complex consensus sequence (GACAAGATAAGATAA-N₁₋₂₈-CTCATCTTC) via a mechanism that may involve the binding of its NZF to the first part and the binding of its CZF to the second part (Morishita *et al.*, 1995). In conclusion, the novel mode of DNA binding described here for SIP1 and δ EF1 may be more universal.

SIP1 was cloned as a Smad1-interacting protein but also interacts with Smad2, Smad3 and Smad5 (Verschuere *et al.*, 1999). Smad proteins are signal transducers involved in the TGF- β family signaling cascade (Kretschmar and Massagué, 1998). Upon binding of TGF- β ligands to the serine/threonine kinase receptor complex, the receptor-regulated Smad proteins are phosphorylated by type I receptors and accumulate in the nucleus where they modulate transcription of target genes. The interaction between SIP1 and Smads is only observed upon ligand stimulation, indicating that Smads need to be activated before they are capable of interacting with SIP1 (Verschuere *et al.*, 1999). Evi-1, a transcription factor (Nucifora, 1997) that may bind DNA by a mechanism similar to that of SIP1, is a Smad3-interacting protein (Kurokawa *et al.*, 1998). So far, it has been shown that Evi-1 inhibits the binding of Smad3 to DNA but it remains to be investigated what the effect of nuclear Smad3 on binding of Evi-1 is on its targets in TGF- β -treated cells.

Schnurri, which is the *Drosophila* homolog of the human PRDII-BF1 transcription factor, is a protein that may also bind DNA by a mechanism similar to that of the SIP1 protein. Interestingly, *Schnurri* was proposed to be involved in the dpp signaling pathway (Arora *et al.*, 1995; Grieder *et al.*, 1995). Dpp is a member of the

TGF- β family. More recent genetic studies in *Drosophila* indicate strongly that *Schnurri* is likely to be a nuclear target for the Mad protein, the homolog of the vertebrate Smads (Henderson *et al.*, 1999). Therefore, the novel mode of DNA binding presented here may be shared by other multi-zinc finger transcription factors interacting with Smads.

Materials and methods

Plasmid constructions

For expression in mammalian cells, (Myc)₆-tagged SIP1 (Verschuere *et al.*, 1999) and δ EF1 (Funahashi *et al.*, 1993) cDNAs were subcloned into pCS3 (Rupp *et al.*, 1994). SIP1 cDNA was also cloned into pCDNA3 (Invitrogen) as an N-terminal fusion with the FLAG tag. For the production of SIP1_{NZF} and SIP1_{CZF}, we subcloned into pCS3 the cDNA fragments encoding amino acids 1–389 and 977–1214, respectively (Verschuere *et al.*, 1999). SIP1_{CZF} (as amino acids 957–1156) and SIP1_{NZF} (amino acids 90–383) were also produced in *E. coli* as a GST fusion protein and purified using the GST purification module (Amersham/Pharmacia Biotech).

Mutagenesis of NZF3, NZF4, CZF2 and CZF3 modified CCHH zinc fingers to CCSC (NZF3, CZF2 and CZF3) and the CCHC finger (NZF4) to CCSC (Ikeda and Kawakami, 1995). Mutations (underlined) were introduced using PCR with the following primers: SIP1_{NZF3Mut}, 5'-CCA-CCTGAAAGAATCCCTGAGAATTCACAG; SIP1_{NZF4Mut}, 5'-GGG-TCTACAGTTCATCTATCAGCAGCAAG; SIP1_{CZF2Mut}, 5'-CACCAC-CTTATCGAGTCTCGAGGCTGCAC; SIP1_{CZF3Mut}, 5'-TCCTACTC-GCAGTCCATGAATCACAGGTAC.

Mutated clusters were recloned in full-length SIP1 in order to produce the SIP1 proteins named NZF3mut, NZF4mut, CZF2mut and CZF3mut, respectively. They were also subcloned into pGEX5-X2 (Amersham/Pharmacia Biotech), and produced in *E. coli* as GST fusion proteins.

The p3TP-Lux and the Xbra2-GFP vectors have been described previously (Wrana *et al.*, 1992; Latinkic *et al.*, 1997). Mutants of the Xbra2 2.1 kb promoter fragment were produced using the following oligonucleotides: Xbra2-Mut1, 5'-GTAAGAACTGACATCTGGTCACTTTTATCAT; Xbra2-Mut2, 5'-GTAAGAACTGACACCTAGTCACTTTTATCAT; Xbra2-Mut3, 5'-CACTT-TATCATTCTATATAGGTGCGCTGGATAG; Xbra2-Mut4, 5'-CACTTTATCATTCTATATTTTCCATGGCTGGATAGATTG.

Cell culture, DNA transfection and luciferase assay

COS1 cells were grown in Dulbecco's modified Eagle's medium (DMEM) supplemented with 10% fetal bovine serum (FBS). CHO cells were grown in HamF12 medium with 10% FBS. COS1 and CHO cells were transfected using Fugene according to the manufacturer's protocol (Boehringer Mannheim), and collected 30–48 h after transfection. Luciferase activity produced from 3TP-Lux was measured using the luciferase assay system (Promega). The same limiting amount of pCMV- β Gal expression vector was also co-transfected in each sample, for normalization of the luciferase activity with the transfection efficiency.

Electrophoretic mobility shift and methylation interference assays

The XbraWT probe covers the region from –344 to –294 of the Xbra2 promoter (Latinkic *et al.*, 1997). The region between –412 and –352 of the human α 4-integrin promoter is present within α 4IWT (Rosen *et al.*, 1994). Probe EcadWT contains the region between –86 and –17 of the human E-cadherin promoter (Bussemakers *et al.*, 1994). Double-stranded oligonucleotides were labeled with [γ -³²P]ATP and T4 polynucleotide kinase (New England Biolabs). The 150 bp probe from p3TP-Lux (Wrana *et al.*, 1992) was obtained after digestion of the plasmid with *Asp*718 and *Bam*HI, dephosphorylation of the fragment with shrimp alkaline phosphatase (Boehringer Mannheim), and labeling with [γ -³²P]ATP and T4 polynucleotide kinase.

Total cell extracts were prepared from COS1 cells (Ray *et al.*, 1992) transfected with different pCS3 vectors allowing synthesis of full-length SIP1 and δ EF1, and different mutant forms of SIP1, or co-production of equal amounts of Myc-tagged SIP1 and FLAG-tagged SIP1. GST–SIP1 fusion proteins were purified from *E. coli* extracts (see above), and tested in gel retardation. The DNA binding assay (20 μ l final volume) was carried out at 25°C, with 1 μ g of COS1 total cell protein, 1 μ g of poly (dI–dC), 10 pg of ³²P-labeled double-stranded oligonucleotide (10⁴

Cerenkov counts) in δ EF1 binding buffer (Sekido *et al.*, 1994). For supershift experiments, the extracts were incubated with anti-Myc (Santa Cruz) or anti-FLAG (Kodak) antibody. The mixtures were loaded onto a 4% polyacrylamide gel prepared in $0.5\times$ TBE buffer. Following electrophoresis, gels were dried and exposed to X-ray film. All experiments were repeated at least three times.

For the methylation interference assay, the upper and lower strands of XbraWT were labeled separately and annealed with excess of complementary DNA strand. Subsequent steps were performed according to standard procedures (Ausubel *et al.*, 1998).

***Xenopus laevis* transgenesis and whole-mount *in situ* hybridization**

Xenopus embryos transgenic for *Xbra2*-GFP were generated as described previously (Kroll and Amaya, 1996), with the following modifications. A Drummond Nanoinject was used for injecting a fixed volume of 5 nl of sperm nuclei suspension per egg, at a theoretical concentration of two nuclei per 5 nl. *NotI* was used for plasmid linearization and nicking of sperm nuclei. Approximately 800 eggs were injected per egg extract incubation. The procedure resulted in a successful cleavage of the embryo, which rates between 10 and 30%. Of these, 50–80% completed gastrulation and 20–30% developed further into normal swimming tadpoles, if allowed. The transgenic frequency, as analysed by expression, varied between 50 and 90%. Embryos were staged according to Nieuwkoop and Faber (1967). A minimum of 30 expressing embryos were analysed per construct and shown stage. Whole-mount *in situ* hybridization for the GFP reporter gene was as described previously (Latinkic *et al.*, 1997). After colour detection, embryos were dehydrated and cleared in a 2:1 mixture of benzyl alcohol/benzyl benzoate.

Acknowledgements

We thank H.Kondoh for providing δ EF1 expression plasmid, and L.Nelles, L.van Grunsven and M.Jones are greatly acknowledged for their suggestions and criticisms. D.H., K.V. and J.E.R. are supported by VIB, H.K. was holder of a post-doctoral fellowship from the National Fund of Scientific Research and the Research Council of the University of Leuven (F/97/37). C.C. and G.W. are IWT predoctoral fellows, and W.L. was supported by the Boehringer Ingelheim Fund. This work was supported by VIB (VIB-07), FWO-Flanders (G.0296.98) and EU-Training and Mobility of Researchers (CT98-0216). The VIB-07 group kindly acknowledges the support from Innogenetics S.A. (Gent, Belgium) within the framework of a collaboration agreement with the University of Leuven (1991–1997) and VIB (1998–2000).

References

- Arora,K., Dai,H., Kazuko,S.G., Jamal,J., O'Connor,M.B., Letsou,A. and Warrior,R. (1995) The *Drosophila* *Schnurri* gene acts in the Dpp/TGF β signaling pathway and encodes a transcription factor homologous to the human MBP family. *Cell*, **81**, 781–790.
- Ausubel,F.M., Brent,R., Kingston,R.E., Moore,D.D., Seidman,J.G., Smith,J.A. and Struhl,K. (1998) *Current Protocols in Molecular Biology*, Volume 2. John Wiley & Sons, Inc., New York, NY.
- Bussemakers,M.J., Giroldi,L.A., van Bokhoven,A. and Schalken,J.A. (1994) Transcriptional regulation of the human E-cadherin gene in human prostate cancer cell lines: characterization of the human E-cadherin gene promoter. *Biochem. Biophys. Res. Commun.*, **203**, 1284–1290.
- de Groot,R.P. and Kruijer,W. (1990) Transcriptional activation by TGF- β 1 mediated by the dyad symmetry element (DSE) and the TPA responsive element (TRE). *Biochem. Biophys. Res. Commun.*, **168**, 1074–1081.
- Fan,C.M. and Maniatis,T. (1990) A DNA binding protein containing two widely separated zinc finger motifs that recognize the same DNA sequence. *Genes Dev.*, **4**, 29–42.
- Fortini,M.E., Lai,Z.C. and Rubin,G.M. (1991) The *Drosophila* *zfh-1* and *zfh-2* genes encode novel proteins containing both zinc-finger and homeodomain motifs. *Mech. Dev.*, **34**, 113–122.
- Funahashi,J., Sekido,R., Murai,K., Kamachi,Y. and Kondoh,H. (1993) δ -crystallin enhancer binding protein δ EF1 is a zinc finger-homeodomain protein implicated in postgastrulation embryogenesis. *Development*, **119**, 433–446.
- Genetta,T., Ruezinsky,D. and Kadesch,T. (1994) Displacement of an E-box binding repressor by basic helix–loop–helix proteins: implications for B-cell specificity of the immunoglobulin heavy-chain enhancer. *Mol. Cell. Biol.*, **14**, 6153–6163.
- Grieder,N.C., Nellen,D., Burke,R., Basler,K. and Affolter,M. (1995) *Schnurri* is required for *Drosophila* Dpp signaling and encodes a zinc finger protein similar to the mammalian transcription factor PRDII-BF1. *Cell*, **81**, 791–800.
- Henderson,K.D., Isaac,D.D. and Andrew,D.J. (1999) Cell fate specification in the *Drosophila* salivary gland: the integration of homeotic gene function with the DPP signaling cascade. *Dev. Biol.*, **205**, 10–21.
- Higashi,Y., Moribe,H., Takagi,T., Sekido,R., Kawakami,K., Kitutani,H. and Kondoh,H. (1997) Impairment of T cell development in δ EF1 mutant mice. *J. Exp. Med.*, **185**, 1467–1479.
- Ikeda,K. and Kawakami,K. (1995) DNA binding through distinct domains of zinc-finger-homeodomain protein AREB6 has different effects on gene transcription. *Eur. J. Biochem.*, **233**, 73–82.
- Jiang,Y., Yu,V.C., Buchholz,F., O'Connell,S., Rhodes,S.J., Candeloro,C., Xia,Y.R., Lusi,A.J. and Rosenfeld,M.G. (1996) A novel family of Cys-Cys, His-Cys zinc finger transcription factors expressed in developing nervous system and pituitary gland. *J. Biol. Chem.*, **271**, 10723–10730.
- Kim,J.G. and Hudson,L.D. (1992) Novel member of the zinc finger superfamily: A C2-HC finger that recognizes a glia-specific gene. *Mol. Cell. Biol.*, **12**, 5632–5639.
- Kretzschmar,M. and Massagué,J. (1998) SMADs: mediators and regulators of TGF- β signaling. *Curr. Opin. Genet. Dev.*, **8**, 103–111.
- Kroll,K.L. and Amaya,L. (1996) Transgenic *Xenopus* embryos from sperm nuclear transplantations reveal FGF signaling requirements during gastrulation. *Development*, **122**, 3173–3183.
- Kuhnlein,R.P., Frommer,G., Friedrich,M., Gonzalez-Gaitan,M., Weber,A., Wagner-Bernholz,J.F., Gehring,W.J., Jackle,H. and Schuh,R. (1994) Spalt encodes an evolutionarily conserved zinc finger protein of novel structure which provides homeotic gene function in the head and tail region of the *Drosophila* embryo. *EMBO J.*, **13**, 168–179.
- Kurokawa,M., Mitani,K., Irie,K., Matsuyama,T., Takahashi,T., Chiba,S., Yazaki,Y., Matsumoto,K. and Hirai,H. (1998) The oncoprotein Evi-1 represses TGF- β signalling by inhibiting Smad3. *Nature*, **394**, 92–96.
- Latinkic,B.V., Umbhauer,M., Neal,K.A., Lerchner,W., Smith,J.C. and Cunliffe,V. (1997) The *Xenopus* Brachyury promoter is activated by FGF and low concentrations of activin and suppressed by high concentrations of activin and by paired-type homeodomain proteins. *Genes Dev.*, **11**, 3265–3276 [erratum in *Genes Dev.*, **12**, 1998, p. 1240].
- Morishita,K., Suzukawa,K., Taki,T., Ihle,J.N. and Yokota,J. (1995) EVI-1 zinc finger protein works as a transcriptional activator via binding to a consensus sequence of GACAAGATAAGATAAN_{1–28}CTCATCTTC. *Oncogene*, **10**, 1961–1967.
- Nieuwkoop,P.D. and Faber,J. (1967) Normal Table of *Xenopus laevis* (Daudin). Amsterdam, North Holland.
- Nucifora,G. (1997) The EVI1 gene in myeloid leukemia. *Leukemia*, **11**, 2022–2031.
- Postigo,A.A. and Dean,D.C. (1997) ZEB, a vertebrate homolog of *Drosophila* Zfh-1, is a negative regulator of muscle differentiation. *EMBO J.*, **16**, 3935–3943.
- Ray,D., Bosselut,R., Ghysdael,J., Mattei,M.G., Tavittian,A. and Moreau-Gachelin,F. (1992) Characterization of Spi-B, a transcription factor related to the putative oncoprotein Spi-1/PU.1. *Mol. Cell. Biol.*, **12**, 4297–4304.
- Rosen,G.D., Barks,J.L., Iademarco,M.F., Fisher,R.J. and Dean,D.C. (1994) An intricate arrangement of binding sites for the Ets family of transcription factors regulates activity of the α 4 integrin gene promoter. *J. Biol. Chem.*, **269**, 15652–15660.
- Rupp,R.A., Snider,L. and Weintraub,H. (1994) *Xenopus* embryos regulate the nuclear localization of XMyoD. *Genes Dev.*, **8**, 1311–1323.
- Seeler,J.S., Muchardt,C., Suessle,A. and Gaynor,R.B. (1994) Transcription factor PRDII-BF1 activates human immunodeficiency virus type 1 gene expression. *J. Virol.*, **68**, 1002–1009.
- Sekido,R., Murai,K., Funahashi,J., Kamachi,Y., Fujisawa-Sehara,A., Nabeshima,Y. and Kondoh,H. (1994) The δ -crystallin enhancer binding protein δ EF1 is a repressor of E2-box-mediated gene activation. *Mol. Cell. Biol.*, **14**, 5692–5700.
- Sekido,R., Murai,K., Kamachi,Y. and Kondoh,H. (1997) Two mechanisms in the action of repressor δ EF1: binding site competition with an activator and active repression. *Genes Cells*, **2**, 771–783.
- Takagi,T., Moribe,H., Kondoh,H. and Higashi,Y. (1998) δ EF1, a zinc finger and homeodomain-containing factor is required for normal skeleton patterning in multiple lineages. *Development*, **125**, 21–32.

- van't Veer, L.J., Lutz, P.M., Isselbacher, K.J. and Bernards, R. (1992) Structure and expression of major histocompatibility complex binding protein 2, a 275-kDa zinc finger protein that binds to an enhancer of major histocompatibility complex class I genes. *Proc. Natl Acad. Sci. USA*, **89**, 8971–8975.
- Verschuere, K. *et al.* (1999) SIP1, a novel zinc finger/homeodomain repressor, interacts with Smad proteins and binds to 5'-CACCT sequences in candidate target genes. *J. Biol. Chem.*, **274**, 20489–20498.
- Watanabe, Y., Kawakami, K., Hirayama, Y. and Nagano, K. (1993) Transcription factors positively and negatively regulating the Na,K-ATPase $\alpha 1$ subunit gene. *J. Biochem. (Tokyo)*, **114**, 849–855.
- Wrana, J.L., Attisano, L., Carcamo, J., Zentella, A., Doody, J., Laiho, M., Wang, X.F. and Massagué, J. (1992) TGF β signals through a heteromeric protein kinase receptor complex. *Cell*, **71**, 1003–1014.
- Yee, K.S. and Yu, V.C. (1998) Isolation and characterization of a novel member of the neural zinc finger factor/myelin transcription factor family with transcriptional repression activity. *J. Biol. Chem.*, **273**, 5366–5374.

*Received May 19, 1999; revised June 21, 1999;
accepted July 27, 1999*