SIP1, a Novel Zinc Finger/Homeodomain Repressor, Interacts with Smad Proteins and Binds to 5′-CACCT Sequences in Candidate Target Genes*

(Received for publication, March 5, 1999)

Kristin Verschueren,a Jacques E. Remacle,a Clara Collart,a,b,c Harry Kraft,a,b,d,e Betty S. Baker,f Przemko Tylzanowski,a,b,g Luc Nelles,b Gunther Wuytens,a,b,c Ming-Tsan Su,b,h Rolf Bodmer,b,h James C. Smith,f,k and Danny Huylenbroeck,a,b,t

From the aDepartment of Cell Growth, Differentiation and Development (VIB-07), Flanders Interuniversity Institute for Biotechnology (VIB), Herestraat 49, B-3000 Leuven, Belgium, the bLaboratory of Molecular Biology (CELGEN), University of Leuven, Herestraat 49, B-3000 Leuven, Belgium, the cDivision of Developmental Biology, National Institute for Medical Research, The Ridgeway, London NW7 1AA, United Kingdom, and the dDepartment of Biology, University of Michigan, Ann Arbor, Michigan 48109-1048

Activation of transforming growth factor β receptors causes the phosphorylation and nuclear translocation of Smad proteins, which then participate in the regulation of expression of target genes. We describe a novel Smad-interacting protein, SIP1, which was identified using the yeast two-hybrid system. Although SIP1 interacts with the MH2 domain of receptor-regulated Smads in yeast and in vitro, its interaction with full-length Smads in mammalian cells requires receptor-mediated Smad activation. SIP1 is a new member of the 5′EF1/Zfh-1 family of two-handed zinc finger/homeodomain proteins. Like 5′EF1, SIP1 binds to 5′-CACCT sequences in different promoters, including the Xenopus brachyury promoter. Overexpression of either full-length SIP1 or its C-terminal zinc finger cluster, which bind to the Xbra2 promoter in vitro, prevented expression of the endogenous Xbra gene in early Xenopus embryos. Therefore, SIP1, like 5′EF1, is likely to be a transcriptional repressor, which may be involved in the regulation of at least one immediate response gene for activin-dependent signal transduction pathways. The identification of this Smad-interacting protein opens new routes to investigate the mechanisms by which transforming growth factor β members exert their effects on expression of target genes in responsive cells and in the vertebrate embryo.

Ligands of the TGF-β1 family exert their biological effects by activating serine/threonine kinase receptor complexes, which in turn activate intracellular mediators, the Smad proteins. Smads were initially identified by means of genetic studies in Drosophila and Caenorhabditis elegans as Mad and Smg gene products, respectively. Nine different vertebrate Smads have been isolated (reviewed in Refs. 1–3; Ref. 4). These proteins are characterized by a three-domain structure containing conserved N-terminal and C-terminal domains, the MH1 and MH2 domains, which flank a more variable, proline-rich linker region. The Smads can be classified into three subgroups based on their distinct functions. The receptor-regulated Smads (Smad1, 2, 3, 5, and 8) contain a conserved SSXS motif at their extreme C-terminal end. Upon ligand stimulation, two serines in this motif are directly phosphorylated by specific type I receptors. Once activated, these Smads associate with Smad4, a common mediator Smad, and the heteromeric complexes translocate to the nucleus where they mediate responses to specific ligands. Smads 1, 5, and 8 act in bone morphogenetic protein (BMP) pathways, whereas Smads 2 and 3 act in activin and TGF-β pathways. A third group of Smads, the inhibitory Smads (Smad6 and Smad7), prevent the activation of receptor-regulated Smads or their heteromerization with Smad4. Functional homologues of inhibitory Smads and the common mediator Smad in Drosophila have been identified as Dad and Medea, respectively (1–3).

In the absence of signaling, Smads are kept in a latent conformation through an intramolecular interaction between the MH1 and MH2 domains. Activation of receptor-regulated Smads has been proposed to disrupt this autoinhibition, allowing the MH1 and MH2 domains to exert distinct functions in the nucleus (1–3). Smad4 and the MH1 domain of activated

* This work was supported by the Flanders Interuniversity Institute for Biotechnology, the Fund of Scientific Research-Flanders (G.0296.98), and EU-Training and Mobility of Researchers (CT98-0216). The VIB-07 group kindly acknowledges the support from Inno-
genetics S. A. (Gent, Belgium) within the framework of collaboration agreements with the University of Leuven and the Flanders Interuni-
iversity Institute for Biotechnology. The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked “advertisement” in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

a Recipient of a Pre-doctoral Fellowship from the Flemish Institute for the Promotion of Industrial Research and Technology Transfer (IWT).

b Present address: Innogenetics S. A., Industriepark 7 box 4, B-9052 Zwijnaarde, Belgium.

c Recipient of a Post-doctoral Fellowship from the National Fund of Scientific Research (NFWO) and of the Research Council of the University of Leuven.

d Present address: Laboratory for Skeletal Development and Joint Disorders, University of Leuven, Herestraat 49, B-3000 Leuven, Belgium.

e Supported by a Fellowship from the Organogenesis Center at the University of Michigan.

f Supported by grants from the National Institutes of Health and the American Heart Association.

h Supported by the UK Medical Research Council and was an Inter-
national Fellow of the Howard Hughes Medical Institute.

i Supported by the Howard Hughes Medical Institute.

j Supported by VIB. To whom correspondence should be addressed. Tel: +32 16 34 59 16; Fax: +32 16 34 59 33; E-mail: dhu@sgi.celgen.
ku leuven.ac.be.

1 The abbreviations used are: TGF-β, transforming growth factor β; bHLH, basic helix-loop-helix; BMP, bone morphogenetic protein; bra, brachyury; C2F, C-terminal zinc finger cluster; DBD, DNA-binding domain; GST, glutathione S-transferase; LacZ, β-galactosidase product of the E. coli LacZ gene; NZF, N-terminal zinc finger cluster; SBD, Smad-binding domain; SIP, Smad-interacting protein; PCR, polymerase chain reaction; X, Xenopus; dpc, days post coitum.
Smad3 can directly bind DNA. Smad-binding elements in the promotors of different immediate response genes such as JunB and PAI-I contain 5′-CAGA boxes, and multimerization of such elements creates a TGF- β -inducible enhancer (5–8). The crystal structure of the Smad3 MH1 domain bound to a Smad-binding element revealed that 5′-GTCT represents the minimal DNA-binding sequence (9). However, promoter studies on other direct target genes, such as vestigial and tinman in Drosophila and goosecoid in the mouse, have implicated GC-rich sequences as direct DNA targets for Mad and/or Medea and for Smad3 and/or Smad4 (10–12). Together, these data suggest that Smads display a low DNA-binding affinity and specificity but are able to achieve highly specific regulation of target promoters through physical or functional interaction with nearby bound transcription factors (12–20). This has been exemplified through detailed studies of activin/TGF- β response elements (ARE) in the promotors of Xenopus Mix.2 and mouse goosecoid which bind the forkhead transcription factors FAST1 and FAST2, respectively (12–16). It has been proposed that upon ligand stimulation, FAST1 or FAST2 recruit heteromeric Smad2/4 complexes to the Mix.2 or goosecoid promotors through their interaction with the MH2 domain of activated Smad2. This promoto binding of Smad4 to an adjacent site, resulting in enhanced transcriptional activation (12, 16).

The MH2 domain of Smads appear to mediate the association with transcription factors and although the majority of documented interactions involve the induction of gene expression, some block transcriptional responses to ligand stimulation. For example, the transcription factor and oncoprotein Evi-1 specifically interacts with activated Smad3, thereby preventing Smad3 from binding DNA and blocking TGF- β -induced growth arrest in certain cell types (20). Recruitment of Smad5/Smad4 heteromeric complexes to the mouse goosecoid promoter blocks, rather than induces, transcription of the gene (12). Overall, these data indicate that, once activated and targeted to the nucleus, Smads are able to undergo multiple interactions with DNA and/or with different transcription factors to cause both activation and repression of gene expression.

Previously, we have shown that overexpression of the Xenopus Smad1 MH2 domain induces ventral cell types in Xenopus embryos. Because this domain does not have DNA-binding capacity, we anticipated that it would interact with transcription factors in the nucleus to elicit its biological effect (21). Therefore, a search for Smad-interacting proteins (SIPs) was initiated using two-hybrid screening in yeast. As bait, the XSmad1 MH2 domain was fused to the DNA-binding domain of the yeast transcription factor GAL4 (GAL4\textsuperscript{BD}). As source of preys, we used a 12.5-dpc mouse embryo cDNA library fused to the GAL4 transactivation domain (GAL4\textsuperscript{AD}). This screen identified several SIPs, one of which, SIP1, is characterized here.

**Experimental Procedures**

cDNA Cloning and Mammalian Expression Plasmids—Mouse Smad1 and Smad2 cDNAs were identified by low stringency screening of an oligo-dT-primed λEcoX library made from 12-dpc mouse embryo (Novagen), using Smad5 (MLP1.2 clone; Ref. 21) as a probe. This library was also used to screen for SIP1 cDNAs other than th1 cDNA, yielding λExTW6. The 3.6-kilobase TW6 cDNA overlapped with th1 and contained additional 3′-coding sequences including an in-frame stop codon. The complete SIP1 open reading frame was reconstituted by fusing TW6 cDNA with a SIP1 sequence including the ATG translation initiation codon, obtained in an independent screen for mouse homologues of Zfh-1. For expression in mammalian cells and Xenopus, the SIP1 cDNA was subcloned into pCS2 and pCS3 (22). In the latter, the SIP1 open reading frame was fused to a Myc tag at the N terminus. For expression of SIP1\textsubscript{G23}, we subcloned a cDNA fragment encoding amino acids 977–1214 into pCS3.

Yeast Two-hybrid Cloning and Assays—XSmad1 full-size and MH2 domain bait plasmids were constructed using the previously described EcoRI-XhoI inserts (21) and cloned between the EcoRI and SalI sites of the bait vector pGBT9-9 (Matchmaker I, CLONTECH), such that in-frame fusions with GAL4\textsuperscript{BD} were obtained. Similar bait plasmids with mouse Smad1, Smad2, and Smad5 were generated by PCR starting with the respective cDNA fragments encoding the MH2 domain. The G418-resistant colonies were identified by X-Gal staining after growth of the bacteria harboring the respective plasmids. Restriction of this plasmid with EcoRI generated a 2059G TAG represents the mini- DNA-binding Direct Partner of Receptor-regulated Smads.
DNA-binding Direct Partner of Receptor-regulated Smads

1 h at 4 °C. Beads were collected by centrifugation and washed four times with lysis buffer at 4 °C, and bound proteins were visualized as described above in the pull-down experiments.

Electrophoretic Mobility Shift Assays—The sequence of the upper strand of the double-stranded oligonucleotide probes used in this work are shown in Figs. 6 and 7. The wild type and mutant XEF1 sequences and the X. brachyury-binding site and MyoD-binding site were taken from Sekido et al. (25). The ARE6-binding site (26), the NII-2a-binding site (27), and the GATA2-binding site (28) were identical to those described previously. Double-stranded oligonucleotides were end-labeled with T4 polynucleotide kinase and [γ-32P]ATP and purified by polyacrylamide gel electrophoresis. Gel retardation assays were carried out with either or with cell extracts from COS1 cells transiently transfected with expression constructs encoding Myc-tagged SIP1 proteins. Extracts were made from those cells as described in the GST pull-down experiments using solubilization buffer. Electrophoretic mobility shift assay was carried out according to Sekido et al. (25). The GST-FLAG1 fusion protein, used as a negative control, was a gift from M. Voz (Flanders Interuniversity Institute for Biotechnology, Dept. VIB-04, Leuven, Belgium).

Experiments in Xenopus—RNA encoding SIP1C2F, SIP1TH1, and full-length SIP1 was prepared by linearizing the appropriate pCS2 plasmids with AspT18 and carrying out transcription reactions according to (29). Xenopus embryos were obtained by in vitro fertilization (30). They were maintained in 10% Normal Amphibian Medium (31) and staged according to Nieuwkoop and Faber (32). Embryos at the 2- to 4-cell stage were injected with 1 ng of RNA dissolved in 14 nl of water as described (33). They were cultured to early gastrula stage 10.5 and processed for whole mount in situ hybridization according to the method of Harland (33), using a probe specific for Xbra (34).

RESULTS

Two-hybrid Cloning of Smad-interacting Proteins—To carry out the two-hybrid screening, the coding sequence of the MH2 domain of XSmad1 was fused to the GAL4DBD in the plasmid pGBT-9. This GAL4DBD-Smad1 bait protein, when tested on its own, did not give detectable levels of GAL4-dependent synthesis of HIS3 and LacZ in the yeast strain used. As a source of prey cDNAs, a random primed library was constructed in a modified pACT2 vector using polyadenylated RNA isolated from 12.5-dpc mouse embryos. Screening of about 4 million yeasts using this bait and the prey plasmids yielded approximately 500 colonies expressing both the HIS3 marker and LacZ reporter genes. Rescreening of these colonies identified 81 in which expression of the two genes required the presence of prey as well as bait cDNAs. One of the prey cDNAs, th72, encoded a protein in which the GAL4 transactivation domain was fused in-frame to Smad4, which started from amino acid 252 in the proline-rich domain (data not shown). Smad4 is known to interact with other receptor-activated Smad proteins (1–3), and the isolation of this Smad4 cDNA confirmed the feasibility of our two-hybrid approach toward identifying Smad-interacting proteins.

The cDNA insert of another positive prey plasmid, th1, encoded a polypeptide of 626 amino acids, named SIP1TH1. Whereas th72 (Smad4) was isolated only once from the initial collection of 81 positive colonies, two additional SIP1 clones, identical to SIP1TH1, were obtained. Sequence analysis revealed that SIP1TH1 has similarities to the vertebrate δ-cystallin enhancer binding protein (δEF1) and Drosophila Zfh-1 (25, 35). These proteins, like SIP1TH1, contain a homedomain sequence. The Zfh-1 homedomain is a canonical domain containing highly conserved residues in helix 3/4 critical for DNA binding, such as a conserved asparagine and arginine at positions 10 and 12 within the helix (36). These critical amino acids are, however, not conserved in the corresponding regions of δEF1 or SIP1, suggesting that their homedomain cannot bind directly to DNA. We therefore prefer to call this domain a homedomain-like sequence. Because Zfh-1 is involved in patterning of mesoderm-derived tissues (35), including muscle and a subset of cells in the heart, and δEF1 is required for normal development of T cells and certain skeletal elements in the mouse (38, 39), it is possible that SIP1, which is expressed during mouse embryogenesis (data not shown), also plays a role in embryonic development. Therefore, SIP1TH1 was subjected to further analysis.

Analysis of SIP1TH1/Smad Interactions in Yeast—Interaction between SIP1TH1 and different Smad proteins were first examined using the yeast two-hybrid system. Interaction of SIP1TH1 with the MH2 domain of XSmad1 was maintained upon removal of the homedomain-like segment of SIP1TH1 (data not shown), and similar approaches enabled us to position the Smad-binding domain (SBD) of SIP1TH1 to a region within the first 192 amino acids. Strikingly, we did not observe an interaction between SIP1TH1 and full-length XSmad1 in yeast (Fig. 1). This was not because of inefficient expression of full-length Smad1 in yeast because other Smad-interacting polypeptides, that are not related to SIP1, interacted efficiently with this bait (data not shown). Additional experiments showed that SIP1TH1 did not interact with the MH1 domain of XSmad1 nor with the MH2 domain from which the last 43 amino acids were deleted (Δ424–466) (Fig. 1). A truncated Mad similar to the Δ424–466 mutant has been shown to cause loss-of-function phenotypes in Drosophila, whereas a similar truncation of Smad4 (dp4) in a loss-of-heterozygosity background is associated with pancreatic carcinomas (40, 41). In contrast, SIP1TH1 did interact with a modified XSmad1 MH2 domain having a single amino acid substitution (G418S, Fig. 1). This mutation affects a conserved glycine residue and has been reported to render the Smad homologue of Drosophila inactive and to abolish BMP-dependent phosphorylation of Smad1 in mammalian cells (40, 42).

Despite their very high degree of sequence similarity, the MH2 domains of Smad1 and Smad2 display striking differences in biological effects when overexpressed in Xenopus embryos; the former induces ventral mesoderm whereas the latter induces dorsal tissues (1–3). Recently, Smad5 has also been shown to induce ventral fates in the Xenopus embryo and to be a target for phosphorylation by activated BMP type I receptors; it thus shares certain activities with Smad1 (43–44). To investigate whether SIP1TH1 interacts specifically with MH2 domains of different Smads, we tested the ability of SIP1TH1 to interact with the MH2 domains of mouse Smad1, 2, and 5 in a yeast two-hybrid assay. SIP1TH1 was found to interact with the MH2 domain of all three Smad proteins tested (Fig. 1). Therefore, SIP1 may be a common binding protein for these receptor-regulated Smads.

SIP1 Is a Novel δEF1-related DNA-binding Protein—The complete SIP1 cDNA sequence was obtained from the sequence of the SIP1TH1 insert and by screening additional cDNA libraries. Mouse SIP1 cDNA was also isolated in an independent

\[\text{SIP1TH1} \sim \text{Smad} \sim \text{δEF1} \sim \text{DNA-binding Protein}\]

2 M.-T. Su, M. Fujioka, and R. Bodmer, unpublished results.
screen for mammalian homologues of Drosophila Zfh-1. A strong sequence conservation between certain segments of SIP1, δEF1, and Zfh-1 can be observed, and the three proteins display a similar organization of putative functional domains (Fig. 2A). As in δEF1 and Zfh-1, the homeodomain-like (HD) segment of SIP1 is flanked by two zinc finger (ZF) clusters, one (NZF) located in the N-terminal part and one (CZF) in the C-terminal part. Despite the very high sequence identity of these zinc finger clusters in the mammalian proteins, mouse SIP1 is distinct from mouse δEF1. Gray boxes reflect identical or biochemically similar amino acids. The position of the C2H2 type zinc fingers in SIP1 is indicated by a bold overline, other zinc fingers in SIP1 are indicated by thin double overlines.

A Smad-binding Domain of 51 Amino Acids Is Essential for Interaction of SIP1 with the XSmad1 MH2 Domain—The SBD of SIP1 was mapped to a segment spanning amino acids 315–507 (see above). This region of 192 amino acids was sufficient to interact with the XSmad1 MH2 domain in yeast (data not shown). To further delineate the SBD, progressive deletions were made within this segment, and the resulting truncated polypeptides were tested in yeast for interaction with the XSmad1 MH2 domain. Mutant SIP1 SBD constructs containing amino acids 437–507, as well as 315–487, still sustained interaction with the bait, whereas mutants encoding amino acids 457–507 and 315–467 did not (Fig. 3). Therefore, the isolated MH2 domain of Smad proteins, lacks both the NZF and the CZF. The SBD, as defined above, maps to amino acids 315–507 in the full-length protein.

FIG. 2. A, schematic representation of the domain structure of SIP1 protein and its similarities with mouse δEF1. The putative zinc fingers are shown (dark gray boxes for C2H2-type zinc fingers and light gray boxes for C3H type) together with the homeodomain-like sequence (HD). The oval indicates the domain essential for interaction of SIP1 with the Smad MH2 domain (SBD). SIP1 cDNAs used in this study are indicated (SIP1 full-length shown in red, SIP1TH1 and SIP1CZF shown in yellow). The sequence of SIP1 is available from GenBank (AF033116). B, amino acid sequence comparison between mouse SIP1 and mouse δEF1. Gray boxes reflect identical or biochemically similar amino acids. The position of the C2H2 type zinc fingers in SIP1 is indicated by a bold overline, other zinc fingers in SIP1 are indicated by thin double overlines.

M.-T. Su, M. Liu, and R. Bodmer, unpublished results.
minimal domain required for interaction with the MH2 domain of XSmad1 was defined as a 51-amino acids-long region encompassing the segment 437–487 of full-length SIP1. To confirm that this domain is involved in mediating Smad binding in the context of full-length SIP1, a SIP1 mutant was constructed in which this minimal SBD was deleted (SIP1_{SBDD51}). As expected, this mutant SIP1 protein gave no detectable interaction with the XSmad1 MH2 bait in yeast (Fig. 3).

SIP1 Interacts with the MH2 Domain of Receptor-regulated Smads in Vitro—Like SIP1_{TH1}, full-length SIP1 protein interacted with the MH2 domain of XSmad1, but not with full-length XSmad1 in yeast (data not shown). We went on to analyze this interaction further in vitro using glutathione S-transferase (GST) pull-down assays. GST-Smad MH2 domain fusion proteins were produced in E. coli and coupled to glutathione-Sepharose beads. An unrelated GST-fusion protein (GST fused to the intracytoplasmic domain of CD40, the receptor for the tumor necrosis factor-related CD40 ligand; Ref. 46) and GST itself were used as negative controls. Full-length SIP1 protein, epitope-tagged (with six Myc epitopes) at its N-terminal end, was produced in COS1 cells. Using GST-Smad beads, we pulled down this SIP1 protein from cell lysates, as shown by Western blotting using anti-Myc antibody (Fig. 4, lane 1). This interaction was specific because a polypeptide consisting of six consecutive Myc tags alone was not pulled down from extracts of transfected COS1 cells (data not shown) nor was SIP1 pulled down by GST alone or by the GST-CD40 fusion protein (Fig. 4, lanes 2 and 3). In addition, full-length SIP1 protein interacted with the MH2 domain of Smad2, Smad5, and the XSmad1G4185 mutant, confirming the results obtained in the yeast two-hybrid assay (Fig. 4, lanes 4–6). Moreover, SIP1_{TH1}, which lacks both zinc finger clusters, behaved like full-length SIP1 in these experiments (data not shown).

SIP1 Interacts with Activated, but Not Latent Full-length Smads in Mammalian Cells—To further examine the functional relevance of the detected interaction between SIP1 and the MH2 domain of XSmad1, we analyzed whether SIP1 could interact with this domain and with latent or activated, full-length Smad1 in mammalian cells. In the first series of experiments, SIP1 with the MH2 domain of XSmad1 was studied using mammalian GST pull-down assays. cDNA encoding the GST-XSmad1_{SBDD1} fusion protein used in the in vitro pull-down assay was cloned in a mammalian expression vector, and this construct was used to transiently transfect COS1 cells, together with an expression construct for Myc-tagged SIP1. The GST-fusion protein was subsequently purified from cell extracts with glutathione-Sepharose beads.

As assayed by immunoblotting, SIP1 could be co-purified with the XSmad1 MH2 domain in these experiments (Fig. 5A, middle panel, lane 1). The specificity of the interaction was further confirmed using the SIP1 mutant which lacks the 51-amino acid-long SBD (SIP1_{SBDD1}), as defined in yeast two-hybrid assays. As expected from previous results, we failed to detect any interaction of this SBD deletion mutant with the MH2 domain of XSmad1 in mammalian cells (Fig. 5A, middle panel, lane 2).

Second, binding of SIP1 with full-length Smad1 was analyzed by co-immunoprecipitation experiments. Expression constructs for Myc-tagged SIP1 and Flag-tagged full-length Smad1 were co-transfected in HEK293T cells, with or without constitutively active ALK-6 (or BMPR-IB), a well characterized BMP type I receptor, which has been shown previously to phosphorylate and activate Smad1 (42, 47). Flag-tagged Smad1 was first immunoprecipitated using anti-Flag monoclonal antibodies, and the resulting precipitate was then probed for the presence of SIP1 by Western analysis using anti-Myc monoclonal antibodies. As demonstrated in Fig. 5B (lanes 1–8), SIP1 could...
be specifically co-immunoprecipitated with full-length Smad1 but only after co-transfection of the cells with constitutively active ALK-6. This shows that SIP1 did not detectably interact with latent full-length Smad1 in mammalian cells and that activation of Smad1 does allow interaction.

To confirm that the detected interaction was a direct consequence of Smad1 activation, SIP1/Smad1 complex formation was analyzed after co-transfection of cells with expression constructs for wild type ALK-6 or constitutively active ALK-4 (ActR-IB). The former cannot signal in the absence of an appropriate type II receptor and without ligand stimulation, whereas the latter specifically phosphorylates Smad2 and not Smad1 (48, 49). Very weak (Fig. 5B, lane 10) or no interaction (lane 11) of SIP1 with full-length Smad1 was detected under these conditions, strongly indicating that complex formation occurred as a result of Smad1 activation by constitutively active ALK-6. In addition, as was shown for the XSmad1 MH2 domain, full-length activated Smad1 did not interact with SIP1 lacking the 51-amino acids-long SBD (Fig. 5B, lane 9).

Our results, obtained both in yeast and in vitro, show that SIP1 interacts with the MH2 domain of several different Smads. To extend these data, we analyzed the association of SIP1 with Smads 2, 3, and 4 by co-immunoprecipitation. SIP1 weakly bound full-length Smad2 and Smad3 in the absence of co-transfected constitutively active ALK-4 (Fig. 5B, lanes 12 and 14). This weak interaction may have resulted from autocrine signaling in HEK293T cells through pathways that activate both Smad2 and Smad3, but not Smad1. Significantly, however, co-transfection of constitutively active ALK-4 greatly enhanced complex formation between SIP1 and both Smad2

**Fig. 5.** A, interaction of SIP1 with the MH2 domain of XSmad1 in mammalian cells. An expression construct encoding a fusion between GST and the MH2 domain of XSmad1 was transfected in COS1 cells together with expression constructs for Myc-tagged SIP1. As shown by immunoblotting of pulled-down material from cell extracts, SIP1 specifically interacted with the GST-XSmad1 fusion protein (middle panel, lane 1), whereas deletion of the SBD disrupted the interaction (middle panel, lane 2). Comparable affinity purification of the GST-fusion protein and equal expression of SIP1 were confirmed by immunoblotting of the pulled down material using polyclonal anti-GST antibody (upper panel) and of total cell extracts using monoclonal anti-myc antibody (lower panel), respectively. B, ligand-dependent interaction of SIP1 with full-length Smads in mammalian cells. Lanes 1–5, lanes 6–11, and lanes 11–17 contain data from three independent experiments. HEK293T cells were transiently transfected with various combinations of expression constructs encoding Myc-tagged SIP1, Flag-tagged Smads, and type I receptors, as indicated. Cell lysates were immunoprecipitated with anti-Flag antibodies, and the precipitated proteins were visualized by SDS-polyacrylamide gel electrophoresis and immunoblotting using anti-Myc (upper panel) or anti-Flag (middle panel) antibodies. The middle panel shows the comparable immunoprecipitations of Flag-tagged Smads in each experiment, whereas the lower panel shows immunoblotting of total cell extracts using anti-Myc antibody, to confirm comparable expression of SIP1. *, indicates the heavy chain of the anti-Flag antibody used in the immunoprecipitations; ca, constitutively active; wt, wild type; ΔSBD indicates SIP1 in which the 51-amino acids-long SBD was deleted.
Fig. 6. Interaction of the DNA-binding domain of SIP1 (SIP1_CZF) with target sites of different promoters by gel retardation analysis. A, the fusion protein GST-SIP1_CZF (10 ng) was incubated with the indicated 32P-labeled double-stranded oligonucleotides. No binding was observed with negative control GST-PLAG1DBD (Control; the Bra-binding site was used here). B, interaction of SIP1_CZF with E2 box sequences. The experiment was carried out as described in panel A, but MyoD and MyoD-Mut labeled probes were used. Competition experiments were carried out with the oligonucleotides listed above the lanes. In the negative control lanes (Control), both MyoD and MyoD-Mut probes were incubated with GST-PLAG1DBD and no competitor was added. No, no competitor added.

and Smad3 (Fig. 5B, lanes 13 and 15). Interaction with Smad3 was stronger than with Smad2. In contrast, we never detected any direct interaction between SIP1 and Smad4 in this experiment (Fig. 5B, lanes 16 and 17). In conclusion, these experiments demonstrate that SIP1 is a common binding protein for different receptor-regulated Smads in mammalian cells and that interaction of full-length Smads with SIP1 is driven by activation of these Smads by specific type I receptors.

SIP1 Has the Same DNA-binding Specificity as εEF1—δEF1 is a repressor of transcription which binds specifically to the sequence 5′-CACCT, as determined by target site selection (50, 51, 25). This (underlined) sequence is also part of the E2 box (5′-CACCTG), which is the binding site for a subgroup of bHLH DNA-binding proteins, such as E2A, E47, and MyoD. It has been proposed that εEF1 may regulate cell type-specific gene expression by competing with these activators for binding. For example, εEF1-mediated repression has been proposed as the primary mechanism for silencing the IgH enhancer in non-B cells. δEF1 is also present in B-cells, but its activity is counteracted by E2A, a B-cell specific bHLH factor (51). Similarly, δEF1 represses the 1gs enhancer where it competes with bHLH factor E47 for binding (25).

The C-terminal zinc finger cluster (CZF) of δEF1 is required for binding to E2 box sequences and for competition with activators (25). Bearing in mind the high similarity of the SIP1 and δEF1 CZF domains, we decided to test whether they have similar DNA-binding specificities. The DNA-binding properties of the CZF fragment of SIP1 (named SIP1_CZF) were analyzed by gel retardation assays using a bacterially expressed and purified GST-SIP1_CZF fusion protein. Larger GST-SIP1 fusion proteins could not be produced because they were subject to proteolytic degradation in E. coli (data not shown).

Purified GST-SIP1_CZF bound to the E2 box of the 1gs enhancer (κE2 probe) (Fig. 6A). A mutation of this site (Mut1), which was previously shown to affect binding of the bHLH factor E47 but not of δEF1 (25), did not affect binding of SIP1_CZF. Two other mutations in this κE2 site (Mut2 and Mut4) which abolished binding of δEF1_CZF (25), also abolished binding of SIP1_CZF. SIP1_CZF also bound to the Nil-2a-binding site of the interleukin-2 (IL-2) promoter, as well as the AREB6-binding site, both of which have previously been shown to bind δEF1 (26, 27). Moreover, as previously proposed for δEF1_CZF, SIP1_CZF bound to a palindromic 5′-CACCT sequence, which also constitutes a binding site for the transcriptional activator X. brachyury (Fig. 6A) (52). Fig. 6B extends these analyses to a site recognized by the bHLH factor MyoD. SIP1_CZF was able to bind to a probe which encompasses the muscle creatine kinase (MCK) enhancer E2 box, and this complex was competed by the E2 box oligonucleotide or by other SIP1-binding sites. Additionally, a point mutation within this E2 box, similar to the previously used κE2-Mut4 site, also abolished binding of SIP1_CZF (Fig. 6B). These experiments show that the GST-SIP1_CZF fusion protein displays the same DNA-binding specificity as the GST-fusion protein made with the CZF region of δEF1 (25) and binds to 5′-CACCT sequences.

SIP1 Binds to the Promoter of the Xbra Gene and Down-regulates Transcription of This Gene upon Overexpression in the Xenopus Embryo—Having characterized SIP1 as a DNA-binding protein, we verified whether promoters of immediate early target genes for signaling by TGF-β members contain 5′-CACCT sequences. Examination of the Xenopus Xbra2 promoter sequence revealed the presence of two potential SIP1-binding sites localized in a 153-bp-long region that confers responsiveness to FGF and activin (53). These sites are arranged in a palindrome and are separated by 24 nucleotides. This prompted us to investigate whether SIP1_CZF or full-length SIP1 can bind to this Xbra2 promoter element.

Myc-tagged SIP1_CZF or full-length SIP1 (SIP1FL) were ex-
pressed in COS1 cells and cell extracts used in gel retardation assays, together with a 50-bp-long Xbra2 promoter sequence encompassing the two 5′-CACCT sequences (Xbra2-WT probe). As shown in Fig. 7, lane 5, cell extracts of mock-transfected cells contain endogenous proteins that bind to this probe. They are visualized as two weak and slowly migrating complexes and one strong, faster migrating complex. When extracts from cells expressing either full-length SIP1 or SIP1CZF were used, an additional complex could be seen (Fig. 7, lanes 1 and 3). In both cases, these complexes could be supershifted with anti-Myc antibody, and the signal could be competed by incubation with an excess of unlabeled Xbra2-WT oligonucleotide (data not shown). This indicated that the complex represents binding of SIP1CZF or full-length SIP1, respectively, to the Xbra2-WT probe and that binding is specific. Integrity of at least the downstream 5′-AGGTG sequence appeared crucial for binding because a single nucleotide substitution in that sequence abolished binding of SIP1CZF or SIP1FL to the Xbra2-D probe (Fig. 7, lanes 2 and 4). A similar mutation disrupted binding of SIP1CZF to the e2f2-binding site (Fig. 6A, lane 5).

These results demonstrate that SIP1 binds specifically to an element in the region of the Xbra2 promoter which mediates fibroblast growth factor and activin induction. They raise the possibility that SIP1 could affect Xbra expression in the Xenopus embryo. To address this question, RNA encoding SIP1CZF or full-length SIP1 was microinjected into one blastomere of Xenopus embryos at the 2- or 4-cell stage. The embryos were allowed to develop to early gastrula stage (stage 10.5), and expression of Xbra was analyzed by whole-mount in situ hybridization. In uninjected embryos, Xbra was expressed throughout the mesoderm, and its expression pattern formed an uninterrupted circle when viewed from the vegetal hemisphere of the embryo (Fig. 8A). As a negative control, embryos were injected with RNA encoding a mutant version of XOt2(K→E) (54). As was shown previously (53), overexpression of this mutant protein had no effect on Xbra expression (Fig. 8C). In embryos injected with RNA encoding either SIP1CZF or full-length SIP1, however, there were gaps in the Xbra expression domains, suggesting that these SIP1 polypeptides abolished transcription of the endogenous Xbra gene (Fig. 8, B and D).

**DISCUSSION**

We describe here the isolation of SIP1, a Smad-interacting protein that is a new member of the family of two-handed zinc finger/homeodomain transcription factors. SIP1 was isolated as a mouse cDNA encoding a protein that interacted with the MH2 domain of Smads1 in yeast. It was subsequently shown to bind in mammalian cells, albeit not as strongly, full-length receptor-regulated Smads, and not to bind to the common mediator Smad4. These findings identify SIP1 as a potential new component of signal transduction pathways triggered by members of the TGF-β superfamily and add SIP1 to the list of transcription factors able to physically interact with Smads, including FAST1 and FAST2, Evi-1, c-Jun, and c-Fos (14, 15, 16, 17, 20). These factors have been identified as mediators/modulators of signaling because they either bound to ligand-responsive elements in promoters of immediate target genes (12, 13, 17) or, as is the case for Evi-1, could counteract the growth inhibitory effect of TGF-β (20).

SIP1 is the first novel Smad-interacting protein identified by a two-hybrid screening in yeast. Our observation that the isolated MH2 domain of Smads could bind to SIP1 in the absence of signaling, but that full-length Smads needed to be activated to interact, supports the view that the Smad MH1 domain exerts an inhibitory effect on the MH2 domain. These autoinhibitory interactions are known to be disrupted upon phos-
 phosphorylation (1–3), thus allowing the MH2 domain to associate with transcription factors. Interestingly, we were unable to detect binding of SIP1 to full-length XSmad1 in yeast. Our data indicate that, in yeast, full-length Smads are also in a latent conformation, preventing the MH2 domain from interacting with certain proteins. Whether unfolding of receptor-regulated Smads and their nuclear translocation is sufficient to drive association with SIP1 in mammalian cells, or whether high affinity interaction of SIP1 with these Smads also depends on the phosphorylation status of the C-terminal serines remains to be determined. In this context, it is important to mention that these serines map in a region shown to be critical for binding of the XSmad1 MH2 domain to SIP1. This region encompasses the last 43 amino acids of XSmad1, and thus their deletion abolishes the interaction. In contrast, substitution of the conserved Gly-418 localized in a loop required for association of Smads with activated type I receptors (55) does not significantly affect interaction of the XSmad1 MH2 domain with SIP1. Thus, interactions of the XSmad1 MH2 domain with type I receptors and SIP1 occur through distinct sequences.

We have also mapped a 51-amino acids-long Smad-binding domain in SIP1 that is essential for its association with Smads. The SIP1 zinc finger clusters are dispensable for binding to Smads, in contrast to the C-terminal zinc finger cluster of Evi-1, which is necessary for interaction with Smad3 (20). Furthermore, there is no sequence similarity between the SBD and the Smad-binding domain of FAST1 (SID). Ref. 14. Smads therefore appear to be able to bind to a wide range of amino acid sequences.

We have demonstrated that SIP1 can interact with different, receptor-activated Smads in mammalian cells. However, in our experiments, SIP1 bound more strongly to Smad3 than to Smad2, both of which were activated by constitutively active ALK-4 (or Act-R-IB), an activin type I receptor. This suggests that the affinity of SIP1 for these Smads differ in vivo. In addition, this aspect may contribute to functional differences between Smad2 and Smad3, as observed before in HaCat keratinocyte cells (56). Alternatively, activation and nuclear translocation may occur more efficiently for Smad3-containing Smad complexes in HEK293T cells. It has been demonstrated previously that Smad6 can inhibit receptor-regulated phosphorylation of Smad1 and Smad2, but not Smad3 (57). Thus, differential activation of Smad2 and Smad3 could be regulated by the inhibitory Smad6 in these cells. Nuclear accumulation of Smads is also modulated by cross-talk between different signaling pathways. For example, phosphorylation of the proline-rich linker region in Smad1 by the Erk family of mitogen-activated protein kinases prevents nuclear accumulation of Smad1/4 heteromeric complexes (58). In general, which Smads are targeted to the nucleus, and consequently interact with binding proteins such as SIP1, may indeed largely depend on cellular context.

SIP1 displays sequence similarities with vertebrate δEF1 and Drosophila Zfh-1, which are both involved in the control of cell type specification during embryonic development (37–39). δEF1 was originally identified as a chicken δ1-crystallin enhancer binding protein. Likely homologues of chicken δEF1 have also been cloned from other species. These include mouse δEF1 or MEB1, human AREB6, ZEB, or Nil-2a, hamster BZP and rat Zfhep (Ref. 46, and reviewed in Ref. 59). δEF1 has been characterized as a repressor of E2 box-mediated gene activation. It binds to 5'-CACCT sequences, and in doing so it can repress transcription by two mechanisms. Either it blocks activity of transcription factors such as c-Myb and Ets bound at nearby sites in target promoters (60) or it competes for common binding sites (E2 boxes) with a subset of bHLH activators, such as E2A and MyoD. This competition has been proposed to provide a genetic switch in which the activity of Ig and muscle-specific enhancers is dictated by the relative levels of δEF1 and these bHLH activators in the nucleus (45, 51).

Analysis of the DNA-binding activities of SIP1, as presented here and currently being analyzed in a separate study involving dissection of DNA target sites and structure-function analysis of the zinc finger clusters, has shown that SIP1, like δEF1, binds to 5’-CACCT sequences. Interestingly, it has been proposed, based on the phenotypes of δEF1 knock-out mice, that another transcription factor with a similar DNA-binding specificity as that of δEF1 exists (39). Indeed, despite the characterization of δEF1 as a negative regulator of muscle differentiation in vitro and its abundant expression in somites, these mice do not display any detectable phenotype in developing muscle (39). Perhaps SIP1, with its overlapping DNA-binding specificities and partially overlapping expression pattern (data not shown), compensates for the loss of δEF1 in certain tissues of the knock-out mice.

We have identified 5’-CACCT sequences in the promoter of X. brachyury (Xbra2), an immediate response gene for mesoderm inducing factors such as activin. These sequences can bind SIP1CZF or full-length SIP1 in vitro. Moreover, overexpression of the SIP1 CZF domain or of full-length SIP1 protein in Xenopus embryos abolished expression of Xbra in presumptive mesoderm. These data indicate that SIP1, when overexpressed, can act as a transcriptional repressor of Xbra in vivo. Interestingly, the SIP1-binding sites in the Xbra2 promoter, the regulation of which is very complex in the amphibian embryo, map to a region that is required for transcriptional induction of the gene by activin (53). Results from a recent study using transgenic Xenopus embryos have shown that these SIP1-binding sites are required for the correct spatial and temporal expression patterns of Xbra2 reporter constructs. Thus, if endogenous SIP1-like proteins are present in the Xenopus embryo, they may participate together with co-repressors (as shown for δEF1, Ref. 61) in the regulation of Xbra2 expression during early development. Overall, our results suggest that transcriptional induction of Xbra in response to mesoderm inducing factors such as activin could (at least in part) occur through interaction of activated Smads with a SIP1-like protein, thereby preventing the latter from binding to the Xbra2 promoter and inhibiting its expression. Activation might therefore be viewed more as relief of repression. Both the Smad-SIP1 and SIP1-DNA interaction are, however, very difficult to analyze biochemically in the amphibian embryo, and cell-based systems in which the Xbra2 promoter can be functionally tested are not available. Further experiments, including the isolation of a Xenopus SIP1 homologue, analysis of its expression pattern, and identification of direct target promoters which can be analyzed in cell lines, will help to unravel further the relevance of SIP1/Smad interactions in signal transduction pathways triggered by members of the TGF-β superfamily.

Acknowledgments—We thank Stefan Pyke and Marianne Voz for providing the GST-CD40 and GST-PLAG1-fusion proteins, Jeff Wrana and Susumu Itoh for various expression constructs encoding Flag-tagged Smads and type I receptors, and Peter Vanderbelle for HEK293T cells.

REFERENCES
1. Heldin, C. H., Miyazono, K., and ten Dijke, P. (1997) Nature 389, 465–471
2. Kretzschmar, M., and Massague´, J. (1998) Curr. Opin. Genet. Dev. 8, 103–111
3. Whitman, M. (1996) Genes Dev. 10, 2245–2246
4. LeSteur, J. A., and Graff, J. M. (1999) Development 126, 137–146

4 W. Lerchner, B. V. Latinkic, L. Faurclough, J. E. Remacle, D. Huybreteek, and J. C. Smith, unpublished results.
