



(86) Date de dépôt PCT/PCT Filing Date: 2000/06/09
(87) Date publication PCT/PCT Publication Date: 2001/01/04
(85) Entrée phase nationale/National Entry: 2001/12/14
(86) N° demande PCT/PCT Application No.: EP 2000/005582
(87) N° publication PCT/PCT Publication No.: 2001/000864
(30) Priorité/Priority: 1999/06/25 (99202068.5) EP

(51) Cl.Int.⁷/Int.Cl.⁷ C12Q 1/00
(71) Demandeur/Applicant:
VLAAMS INTERUNIVERSITAIR INSTITUUT VOOR
BIOTECHNOLOGIE VZW, BE
(72) Inventeurs/Inventors:
HUYLEBROECK, DANNY, BE;
VERSCHUEREN, KRISTIN, BE;
REMACLE, JACQUES, BE
(74) Agent: FETHERSTONHAUGH & CO.

(54) Titre : LIAISON D'ACIDE NUCLEIQUE A DES FACTEURS DE TRANSCRIPTION DE DOIGT MULTI-ZINC
(54) Title: NUCLEIC ACID BINDING OF MULTI-ZINC FINGER TRANSCRIPTION FACTORS

(57) **Abrégé/Abstract:**

The invention concerns a method of identifying transcription factors comprising providing cells with a nucleic acid sequence at least comprising a sequence CACCT as bait for the screening of a library encoding potential transcription factors and performing a specificity test to isolate said factors. Preferably the bait comprises twice the CACCT sequence, more particularly the bait comprises one of the sequences CACCT-N-CACCT, CACCT-N-AGGTG, AGGTG-N-CACCT or AGGTG-N-AGGTG wherein N is a spacer sequence. The identified transcription factor(s) using the method according to the invention comprises separated clusters of zinc fingers such as for example a two-handed zinc finger transcription factor. The present invention further discloses that at least one such zinc finger transcription factor, denominated as SIP1, induces tumor metastasis by downregulation of the expression of E-cadherin. Compounds interfering with SIP1 activity can thus be used to prevent tumor invasion and metastasis.



(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau



(43) International Publication Date
4 January 2001 (04.01.2001)

PCT

(10) International Publication Number
WO 01/00864 A3

- (51) International Patent Classification⁷: **C12Q 1/00**
- (21) International Application Number: PCT/EP00/05582
- (22) International Filing Date: 9 June 2000 (09.06.2000)
- (25) Filing Language: English
- (26) Publication Language: English
- (30) Priority Data:
99202068.5 25 June 1999 (25.06.1999) EP
- (71) Applicant (*for all designated States except US*):
VLAAMS INTERUNIVERSITAIR INSTITUUT VOOR BIOTECHNOLOGIE VZW [BE/BE]; Rijvisschestraat 120, B-9052 Zwijnaarde (BE).
- (72) Inventors; and
- (75) Inventors/Applicants (*for US only*): **HUYLEBROECK, Danny** [BE/BE]; Lijsterlaan 15, B-1770 Liedekerke (BE). **VERSCHUEREN, Kristin** [BE/BE]; Twee Leeuwenstraat 22, B-3078 Everberg (BE). **REMACLE, Jacques** [BE/BE]; Avenue des Lilas 7, B-4280 Hannut (BE).
- (74) Common Representative: **VLAAMS INTERUNIVERSITAIR INSTITUUT VOOR BIOTECHNOLOGIE VZW**; Rijvisschestraat 120, B-9052 Zwijnaarde (BE).
- (81) Designated States (*national*): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CR, CU, CZ, DE, DK, DM, DZ, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW.
- (84) Designated States (*regional*): ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).
- Published:
— with international search report
- (88) Date of publication of the international search report:
29 November 2001
- For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.*

(54) Title: NUCLEIC ACID BINDING OF MULTI-ZINC FINGER TRANSCRIPTION FACTORS

(57) Abstract: The invention concerns a method of identifying transcription factors comprising providing cells with a nucleic acid sequence at least comprising a sequence CACCT as bait for the screening of a library encoding potential transcription factors and performing a specificity test to isolate said factors. Preferably the bait comprises twice the CACCT sequence, more particularly the bait comprises one of the sequences CACCT-N-CACCT, CACCT-N-AGGTG, AGGTG-N-CACCT or AGGTG-N-AGGTG wherein N is a spacer sequence. The identified transcription factor(s) using the method according to the invention comprises separated clusters of zinc fingers such as for example a two-handed zinc finger transcription factor. The present invention further discloses that at least one such zinc finger transcription factor, denominated as SIP1, induces tumor metastasis by downregulation of the expression of E-cadherin. Compounds interfering with SIP1 activity can thus be used to prevent tumor invasion and metastasis.

WO 01/00864 A3

Nucleic acid binding of multi-zinc finger transcription factors

Field of the invention

The invention concerns a method of identifying transcription factors comprising
5 providing cells with a nucleic acid sequence at least comprising a sequence CACCT
as bait for the screening of a library encoding potential transcription factors and
performing a specificity test to isolate said factors. Preferably the bait comprises twice
the CACCT sequence, more particularly the bait comprises one of the sequences
CACCT-N-CACCT, CACCT-N-AGGTG, AGGTG-N-CACCT or AGGTG-N-AGGTG
10 wherein N is a spacer sequence.

The identified transcription factor(s) using the method according to the invention
comprises separated clusters of zinc fingers such as for example a two-handed zinc
finger transcription factor. The present invention further discloses that at least one
such zinc finger transcription factor, denominated as SIP1, induces tumor metastasis
15 by downregulation of the expression of E-cadherin. Compounds interfering with SIP1
activity can thus be used to prevent tumor invasion and metastasis.

Background of the invention

Zinc fingers are among the most common DNA binding motifs found in eukaryotes. It
20 is estimated that there are 500 zinc finger proteins encoded by the yeast genome and
that perhaps 1 % of all mammalian genes encode zinc finger containing proteins.
These are classified according to the number and position of the cysteine and
histidine residues available for zinc coordination. The CCHH class, which is typified by
the *Xenopus* transcription factor IIIA (19), is the largest. These proteins contain two or
25 more fingers in tandem repeats. In contrast, the steroid receptors contain only
cysteine residues that form two types of zinc-coordinated structures with four (C₄) and
five (C₅) cysteines (28). The third class of zinc fingers contains the CCHC fingers. The
CCHC fingers which are found in *Drosophila*, and in mammalian and retroviral
proteins, display the consensus sequence C-X₂-C-X₄-H-X₄-C (7, 21, 24). Recently, a
30 novel configuration of CCHC finger, of the C-X₅-C-X₁₂-H-X₄-C type, was found in the
neural zinc finger factor/myelin transcription factor family (11, 12, 36). Finally, several
yeast transcription factors such as GAL4 and CHA4 contain an atypical C₆ zinc finger

structure that coordinates 2 zinc ions (9, 32).

Zinc fingers are usually found in multiple copies (up to 37) per protein. These copies can be organized in tandem array, forming a single cluster or multiple clusters, or they can be dispersed throughout the protein. Several families of transcription factors

5 share the same overall structure by having two (or three) widely separated clusters of zinc fingers in their protein sequence. The first, the MBPs/PRDII-BF1 transcription factor family, includes *Drosophila Schnurri* and *Spalt* genes (1, 3, 6, 14, 33). Both MBP-1 (also known as PRDII-BF1) and MBP-2 contain two widely separated clusters of two CCHH zinc fingers. The overall similarity between MBP-1 and MBP-2 is 51%,

10 but the conservation is much higher (over 90%) for both the N-terminal and the C-terminal zinc finger clusters (33). This indicates an important role of both clusters in the function of these proteins. In addition, the N-terminal and C-terminal zinc finger clusters of MBP-1 are very homologous to each other (3). The neural specific zinc finger factor 1 and factor 3 (NZF-1 and NZF-3), as well as the myelin transcription

15 factor 1 (MyT1, also known as NZF-2), belong to another family of proteins containing two widely separated clusters of CCHC zinc fingers (11, 12, 36). Like the MBP proteins, different NZF factors exhibit a high degree of sequence identity (over 80%) between the respective zinc finger clusters, whereas the sequences outside of the zinc finger region are largely divergent (36). In addition, each of these clusters can

20 independently bind to DNA, and recognizes similar core consensus sequences (11). NZF-3 binds to a DNA element containing a single copy of this consensus sequence but was shown to exhibit a marked enhancement in relative affinity to a bipartite element containing two copies of this sequence (36). This suggests that the NZF factors may also bind to reiterated sequences. However, the mechanism underlying

25 the cooperative binding of NZF-3 to the bipartite element is currently unknown. The *Drosophila Zfh-1* and the vertebrate δ EF1 proteins (also known as ZEB or AREB6) belong to a third family of transcription factors. This family is characterized by the presence of two separated clusters of CCHH zinc fingers and a homeodomain-like structure (see Fig. 1A)(4, 5, 35). In δ EF1, the N-terminal and C-terminal clusters are

30 also very homologous and were shown to bind independently to very similar core consensus sequences (10). Recently, it was shown that mutant forms of δ EF1 lacking either the N-terminal or the C-terminal cluster have lost their DNA binding capacity

indicating that both cluster are required for the binding of δ EF1 to DNA (31). The Evi-1 transcription factor was shown to contain 10 CCHH zinc fingers; seven zinc fingers are present in the N-terminal region, and three zinc fingers are in the C-terminal region (22). With this factor the situation is different from the transcription factors
5 described above, because the two clusters bind to two different target sequences, which are bound simultaneously by full-length Evi-1 (20). Binding of full-length Evi-1 is mainly observed when the two target sequences are positioned in a certain relative orientation, but there was no strict requirement for an optimal spacing between these two targets.

10

Cell-cell adhesion is a predominant necessity during cell differentiation, tissue development, and tissue homeostasis. The effect of disrupted cell-cell adhesion is displayed in many cancers, where metastasis and poor prognosis are correlated with loss of cell-cell adhesion. E-cadherin, a homophilic Ca^{2+} -dependent transmembrane
15 adhesion molecule, and the associated catenins are among the major constituents of the epithelial cell-junction system. E-cadherin exerts a potent invasion-suppressing role in tumour cell line systems (46, 47) and in *in vivo* tumourmodel systems (48). Loss of E-cadherin expression during tumour progression has been described for more than 15 different carcinoma types (49). Extensive analyses made clear that
20 aberrant E-cadherin expression as a result of somatic inactivating mutations of both E-cadherin alleles is rare and so far largely confined to diffuse gastric carcinomas and infiltrative lobular breast carcinomas (50, 51). Northern analysis and *in situ* hybridization studies revealed that reduced E-cadherin immunoreactivity in human carcinomas correlates with decreased mRNA levels (52-54). Analysis of mouse and
25 human E-cadherin promoter sequences revealed a conserved modular structure with positive regulatory elements including a CCAAT-box and a GC -box, as well as two E-boxes (CANNTG) with a potential repressor role (55, 56). Mutation analysis of the two E-boxes in the E-cadherin promoter demonstrated a crucial role in the regulation of the epithelial specific expression of E-cadherin. Mutation of these two E-box elements
30 results in the upregulation of the E-cadherin promoter in dedifferentiated cancer cells, where the wild type promoter shows low activity (55, 56).

Brief description of the figures

Figure 1. **Schematic representation of Zfh-1, SIP1 and δ EF1, and alignment of the SIP1 and δ EF1 zinc fingers.** (A) Schematic representation of mouse δ EF1 (1117 amino acids) and SIP1 (1214 amino acids). The filled boxes represent CCHH zinc fingers, the open boxes are CCHC zinc fingers. The homeodomain-like domain (HD) is depicted as an oval. The percentage represents the homology between different domains. SIP1 polypeptides used in this study are depicted with their coordinates. SBD: Smad-binding domain (Verschuere *et al.*, 1999). (B) Alignments of the amino acid sequences from zinc fingers of SIP1 and δ EF1. Vertical bars indicate sequence identity. The conserved cysteine and histidine residues forming the zinc fingers are printed in bold, and indicated by an asterisk. The residues in zinc fingers that can contact DNA are indicated with an arrow. (C) Alignment of the protein sequence of SIP1_{NZF3+NZF4} and SIP1_{CZF2+CZF3}, and of δ EF1_{NZF3+NZF4} and δ EF1_{CZF2+CZF3}, respectively, demonstrating intramolecular conservation of zinc fingers.

Figure 2. **Possible DNA-binding mechanisms for SIP1.** Model 1: SIP1 binds DNA as a monomer. Model 2: SIP1 binds DNA as a dimer.

Summary of the invention

The mechanism of DNA binding remains poorly understood for most of the above-mentioned complex factors. It is our invention to characterize the DNA binding properties of vertebrate transcription factors belonging to the emerging family of two-handed zinc finger transcription factors like δ EF1 and SIP1. SIP1 is a member of this transcription factor family, which was recently isolated and characterized as a Smad-interacting protein (34). Said SIP1 and δ EF1, a transcriptional repressor involved in skeletal development and muscle cell differentiation, belong to the same family of transcription factors. They contain two separated clusters of CCHH zinc fingers, which share high sequence identity (>90%). The DNA-binding properties of these transcription factors have been investigated. The N-terminal and C-terminal clusters of

SIP1 show high sequence homology as well, and according to the invention each binds to a 5'-CACCT sequence. Furthermore, high affinity binding sites for full length SIP1 and δ EF1 in the promoter regions of candidate target genes like Brachyury, α 4-integrin and E-cadherin, are bipartite elements composed of one CACCT sequence and one CACCTG sequence. No strict requirement for the relative orientation of both sequences was observed, and the spacing between them (also denominated as N) may vary from 0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10,..., to at least 44 bp. For binding to these bipartite elements, the integrity of both SIP1 zinc finger clusters is necessary, indicating that they are both involved in binding to DNA. Furthermore, SIP1 binds as a monomer to a CACCT- X_N -CACCTG site, by having one zinc finger cluster contacting the CACCT, and the other zinc finger cluster binding to the CACCTG sequence. This novel mode of binding may be generalised to other transcription factors that contain separated clusters of zinc fingers and may be applied to other Smad-binding proteins. Moreover, the Smad-interacting protein SIP1 shows high expression in E-cadherin-negative human carcinoma cell lines, resulting in downregulation of E-cadherin transcription. Conditional expression of SIP1 in E-cadherin-positive MDCK cells also abrogates E-cadherin-mediated intercellular adhesion and simultaneously induced invasion. Hence, SIP1 can be considered as a potent invasion promoter molecule and compounds, such as anti-SIP1 antibodies, small molecules specifically binding to SIP, anti-sense nucleic acids and ribozymes, which interfere with SIP1 production or activity can prevent tumor invasion and metastasis.

The invention thus concerns a method of identifying transcription factors such as activators and/or repressors comprising providing cells with a nucleic acid sequence at least comprising a sequence CACCT, preferably twice the CACCT sequence as bait for the screening of a library encoding potential transcription factors and performing a specificity test to isolate said factors. In another embodiment the bait comprises one of the sequences CACCT-N-CACCT, CACCT-N-AGGTG, AGGTG-N-CACCT or AGGTG-N-AGGTG wherein N is a spacer sequence. The latter spacer sequence can vary in length and can contain any number of base pairs (bp) from N=0 bp to N= at least 44 bp. Thus, for example N can be 0,1,2,3,4,5,6,7,8,9,10,15,20,25,30,35,40,45,50,60,70,80,90,100,200,300 or 400 bp in length.

The identified transcription factor(s) using the method according to the invention comprises separated clusters of zinc fingers such as for example two-handed zinc finger transcription factors.

5 The sequence above mentioned may originate from any promoter region but preferably from the group (also referred to as target genes, see further) selected from Brachyury, α 4-integrin, follistatin or E-cadherin.

The transcription factors obtainable by above referenced method are part of the present invention as well.

10 In another embodiment the present invention relates to a method of identifying compounds with an interference capability towards transcription factors, obtained as above mentioned, by

a) adding a sample comprising a potential compound to be identified to a test system comprising (i) a nucleotide sequence comprising one of the sequences CACCT-N-CACCT, CACCT-N-AGGTG, AGGTG-N-CACCT or AGGTG-N-AGGTG as bait
15 wherein N is a spacer sequence, (ii) a protein capable to bind said nucleotide sequence,

b) incubating said sample in said system for a period sufficient to permit interaction of the compound or its derivative or counterpart thereof with said protein and

c) comparing the amount and/or activity of the protein bound to the nucleotide
20 sequence before and after said adding.

Comparison of the amount of protein bound to the nucleotide sequence before and after adding the test sample can be accomplished, for example, using a gel band-shift assay or a filter-binding assay. As a next step the compound thus identified can be isolated and optionally purified and further analyzed according to methods known to
25 persons skilled in the art. The protein in step a) (ii) can be any protein capable to bind said nucleotide sequence, but is preferably a Smad-interacting protein such as SIP1.

Compounds identified by the latter method are also part of the present invention. With the terms 'compounds with an interference capability towards transcription factors' are meant compounds which are able to modulate (= i.e. to inhibit, to weaken, to
30 strengthen) the bioactivity of transcription factors. More specifically the latter compounds are able to completely or partially inhibit the production and/or bioactivity of SIP1. Examples of such compounds are small molecules or anti-SIP1 antibodies or

functional fragments derived thereof specifically binding to SIP1 protein or anti-sense nucleic acids or ribozymes binding to mRNA encoding SIP1 or small molecules binding the promoter region bound by SIP1. In this regard, the present invention relates to compounds which modulate regulation of E-cadherin expression by SIP1.

5 More specifically the present invention relates to compounds which, via inhibiting SIP1 production and/or activity, prevent the down-regulation of the expression of the target gene E-cadherin. In other words, the present invention relates to compounds which can be used as a medicament to prevent or treat tumor invasion and/or metastasis which is due to the down-regulation of E-cadherin expression by SIP-1. Methods to
10 produce and use the latter compounds are exemplified further.

To the scope of the present invention also belongs a test kit to perform said method comprising at least (i) an nucleotide sequence comprising one of the sequences CACCT-N-CACCT, CACCT-N-AGGTG, AGGTG-N-CACCT or AGGTG-N-AGGTG wherein N is a spacer sequence and (ii) a protein capable to bind said nucleotide
15 sequence.

In another embodiment the current invention concerns an alternative to the so-called two hybrid screening assay as disclosed in the prior art. Several means and methods have been developed to identify binding partners of proteins. This has resulted in the identification of a number of respective binding proteins. Many of these proteins have
20 been found using so-called two hybrid systems. Two-hybrid cloning systems have been developed in several labs (Chien et al., 1991; Durfee et al., 1993; Gyuris et al., 1993). All have three basic components: Yeast vectors for expression of a known protein fused to a DNA-binding domain, yeast vectors that direct expression of cDNA-encoded proteins fused to a transcription activation domain, and yeast reporter genes
25 that contain binding sites for the DNA-binding domain. These components differ in detail from one system to the other. All systems utilise the DNA binding domain from either Gal4 or LexA. The Gal4 domain is efficiently localised to the yeast nucleus where it binds with high affinity to well-defined binding sites which can be placed upstream of reporter genes (Silver et al., 1986). LexA does not have a nuclear
30 localisation signal, but enters the yeast nucleus and, when expressed at a sufficient level, efficiently occupies LexA binding sites (operators) placed upstream of a reporter gene (Brent et al., 1985). No endogenous yeast proteins bind to the LexA operators.

Different systems also utilise different reporters. Most systems use a reporter that has a yeast promoter, either from the GAL1 gene or the CYC1 gene, fused to lacZ (Yocum et al., 1984). These lacZ fusions either reside on multicopy yeast plasmids or are integrated into a yeast chromosome. To make the lacZ fusions into appropriate
5 reporters, the GAL1 or CYC1 transcription regulatory regions have been removed and replaced with binding sites that are recognised by the DNA-binding domain being used. A screen for activation of the lacZ reporters is performed by plating yeast on indicator plates that contain X-Gal (5-bromo-4-chloro-3-indolyl- β -D-galactoside); on this medium yeast in which the reporters are transcribed produce beta-galactosidase
10 and turn blue. Some systems use a second reporter gene and a yeast strain that requires expression of this reporter to grow on a particular medium. These "selectable marker" genes usually encode enzymes required for the biosynthesis of an amino acid. Such reporters have the marked advantage of providing a selection for cDNAs that encode interacting proteins, rather than a visual screen for blue yeast. To make
15 appropriate reporters from the marker genes their upstream transcription regulatory elements have been replaced by binding sites for a DNA-binding domain. The HIS3 and LEU2 genes have both been used as reporters in conjunction with appropriate yeast strains that require their expression to grow on media lacking either histidine or leucine, respectively. Finally, different systems use different means to express
20 activation-tagged cDNA proteins. In all current schemes the cDNA-encoded proteins are expressed with an activation domain at the amino terminus. The activation domains used include the strong activation domain from Gal4, the very strong activation domain from the Herpes simplex virus protein VP16, or a weaker activation domain derived from bacteria, called B42. The activation-tagged cDNA-encoded
25 proteins are expressed either from a constitutive promoter, or from a conditional promoter such as that of the GAL1 gene. Use of a conditional promoter makes it possible to quickly demonstrate that activation of the reporter gene is dependent on expression of the activation-tagged cDNA proteins.

It is clear from the discussion above that two-hybrid systems for finding binding
30 proteins have been used in the past. However, although the conventional two hybrid system has proven to be a valuable tool in finding proteinaceous molecules that can bind to other proteins it is a (very) artificial system. A characteristic of any two hybrid

system is that a fusion protein is made consisting of a part of which binding partners are sought and a reporter part that enables detection of binding. For finding relevant binding partners several criteria must be met of which one is of course the correct choice of the region in said protein where binding to other proteins occurs. Another
5 criterion which is much more difficult if not impossible to predict accurately on forehand is obtaining correct folding of said region (i.e. a folding of said region sufficiently similar to the folding of said region in the natural protein). Correct folding depends among others on the actual amino-acid sequence chosen for generating said fusion protein. Another factor determining the identification of relevant binding
10 partners is the sensitivity with which binding can be detected.

An alternative to the above mentioned conventional two hybrid system is herewith also provided in the current invention. Thus an alternative object of the invention is to provide an in vivo method and a kit for detecting interactions between proteins and the influence of other compounds on said interaction as such, using reconstitution of
15 the activity of a transcriptional activator. This reconstitution makes use of two, so-called hybrid, chimeric or fused proteins. These two fused proteins each show, independently from one another, a weak affinity towards a nucleic acid sequence comprising one of the sequences CACCT-N-CACCT, CACCT-N-AGGTG, AGGTG-N-CACCT or AGGTG-N-AGGTG wherein N is a spacer sequence. However when both
20 fused proteins are independently being bound to said sequence and the test proteins each available in each of two fused proteins are as a result thereof brought into close proximity, the binding affinity towards said nucleic acid sequence comprising one of the sequences CACCT-N-CACCT, CACCT-N-AGGTG, AGGTG-N-CACCT or AGGTG-N-AGGTG wherein N is a spacer sequence becomes much stronger. If the
25 two test proteins indeed are able to interact, they bring as a consequence thereof into close proximity the two domains of the transcriptional activator. This proximity is sufficient to cause transcription, which can be detected by the activity of a marker gene located adjacent to the nucleic acid sequence comprising one of the sequences CACCT-N-CACCT, CACCT-N-AGGTG, AGGTG-N-CACCT or AGGTG-N-AGGTG
30 wherein N is a spacer sequence. In accordance herewith a method is provided for detecting an interaction between a first interacting protein and a second interacting protein comprising

- a) providing a suitable host cell with a first fusion protein comprising a first interacting protein fused to a DNA binding domain capable to bind a nucleic acid sequence comprising one of the sequences CACCT-N-CACCT, CACCT-N-AGGTG, AGGTG-N-CACCT or AGGTG-N-AGGTG wherein N is a spacer sequence,
- 5 b) providing said suitable host cell with a second fusion protein comprising a second interacting protein fused to a DNA binding domain capable to bind a nucleic acid sequence comprising one of the sequences CACCT-N-CACCT, CACCT-N-AGGTG, AGGTG-N-CACCT or AGGTG-N-AGGTG wherein N is a spacer sequence,
- 10 c) subjecting said host cell to conditions under which the first interacting protein and the second interacting protein are brought into close proximity and
- d) determining whether a detectable gene present in the host cell and located adjacent to said nucleic acid sequence has been expressed to a degree greater than expressed in the absence of the interaction between the first and the second
- 15 interacting protein.

As an example, it should be clear that, in case a binding partner (prey) for a specific protein (bait) has been identified, the first fusion protein containing the bait will for example bind to the sequence CACCT (or AGGTG) of the sequence CACCT-N-AGGTG and that the second fusion protein containing the prey will bind to the

20 sequence AGGTG (or CACCT, respectively) of the sequence CACCT-N-AGGTG so that transcription of a marker gene will occur.

The present invention finally relates to the new sequences CACCT-N-CACCT, CACCT-N-AGGTG, AGGTG-N-CACCT or AGGTG-N-AGGTG wherein N is a spacer sequence as defined above, and to the use of said sequences, in addition to any

25 other sequence at least comprising a sequence CACCT, for the identification, via any method known by a person skilled in the art, of new target genes different from the already described target genes Brachyury, α 4-integrin, follistatin or E-cadherin.

The following definitions are set forth to illustrate and define the meaning and scope of the various terms used to describe the invention herein and their meaning is further

30 elaborated hereunder for sake of clarity.

"Nucleic acid" or "nucleic acid sequence" or "nucleotide sequence" means genomic DNA, cDNA, double stranded or single stranded DNA, messenger RNA or any form of

nucleic acid sequence known to a skilled person.

The terms "protein" and "polypeptide" used in this application are interchangeable.

"Polypeptide" refers to a polymer of amino acids (amino acid sequence) and does not refer to a specific length of the molecule. Thus peptides and oligopeptides are

5 included within the definition of polypeptide. This term does also refer to or include post-translational modifications of the polypeptide, for example, glycosylations, acetylations, phosphorylations and the like. Included within the definition are, for example, polypeptides containing one or more analogs of an amino acid (including, for example, unnatural amino acids, etc.), polypeptides with substituted linkages, as

10 well as other modifications known in the art, both naturally occurring and non-naturally occurring. The proteins and polypeptides described above are not necessarily translated from a designated nucleic acid sequence; the polypeptides may be generated in any manner, including for example, chemical synthesis, or expression of a recombinant expression system, or isolation from a suitable viral system. The

15 polypeptides may include one or more analogs of amino acids, phosphorylated amino acids or unnatural amino acids. Methods of inserting analogs of amino acids into a sequence are known in the art. The polypeptides may also include one or more labels, which are known to those skilled in the art. In this context, it is also understood that the proteins may be further modified by conventional methods known in the art.

20 By providing the proteins it is also possible to determine fragments which retain biological activity, namely the mature, processed form. This allows the construction of chimeric proteins and peptides comprising an amino sequence derived from the mature protein which is crucial for its binding activity. The other functional amino acid sequences may be either physically linked by, e.g., chemical means to the proteins or

25 may be fused by recombinant DNA techniques well known in the art.

The term "derivative", "functional fragment of a sequence" or "functional part of a sequence" means a truncated sequence of the original sequence referred to. The truncated sequence (nucleic acid or protein sequence) can vary widely in length; the minimum size being a sequence of sufficient size to provide a sequence with at least

30 a comparable function and/or activity of the original sequence referred to, while the maximum size is not critical. In some applications, the maximum size usually is not substantially greater than that required to provide the desired activity and/or

function(s) of the original sequence. Typically, the truncated amino acid sequence will range from about 5 to about 60 amino acids in length. More typically, however, the sequence will be a maximum of about 50 amino acids in length, preferably a maximum of about 30 amino acids. It is usually desirable to select sequences of at least about 10, 12 or 15 amino acids, up to a maximum of about 20 or 25 amino acids. The terms "gene(s)", "polynucleotide", "nucleic acid sequence", "nucleotide sequence", "DNA sequence" or "nucleic acid molecule(s)" as used herein refers to a polymeric form of nucleotides of any length, either ribonucleotides or deoxyribonucleotides. This term refers only to the primary structure of the molecule. Thus, this term includes double- and single-stranded DNA, and RNA. It also includes known types of modifications, for example, methylation, "caps" substitution of one or more of the naturally occurring nucleotides with an analog.

A "coding sequence" is a nucleotide sequence which is transcribed into mRNA and/or translated into a polypeptide when placed under the control of appropriate regulatory sequences. The boundaries of the coding sequence are determined by a translation start codon at the 5'-terminus and a translation stop codon at the 3'-terminus. A coding sequence can include, but is not limited to mRNA, cDNA, recombinant nucleotide sequences or genomic DNA, while introns may be present as well under certain circumstances.

With "transcription factor" is meant a class of proteins that bind to a promoter or to a nearby sequence of DNA to facilitate or prevent transcription initiation.

With "promoter" is meant an oriented DNA sequence recognized by the RNA polymerase holoenzyme to initiate transcription. With "RNA polymerase" is meant a multisubunit enzyme that synthesizes RNA complementary to the DNA template. With "holoenzyme" is meant an active form of enzyme that consists of multiple subunits.

The term 'antibody' or 'antibodies' relates to an antibody characterized as being specifically directed against a transcription factor such as SIP-1 or any functional derivative thereof, with said antibodies being preferably monoclonal antibodies; or an antigen-binding fragment thereof, of the F(ab')₂, F(ab) or single chain Fv type, or any type of recombinant antibody derived thereof. The monoclonal antibodies of the invention can for instance be produced by any hybridoma liable to be formed according to classical methods from splenic cells of an animal, particularly of a mouse

or rat immunized against SIP1 or any functional derivative thereof, and of cells of a myeloma cell line, and to be selected by the ability of the hybridoma to produce the monoclonal antibodies recognizing SIP1 or any functional derivative thereof which have been initially used for the immunization of the animals. The monoclonal antibodies according to this embodiment of the invention may be humanized versions of the mouse monoclonal antibodies made by means of recombinant DNA technology, departing from the mouse and/or human genomic DNA sequences coding for H and L chains or from cDNA clones coding for H and L chains. Alternatively the monoclonal antibodies may be human monoclonal antibodies. Such human monoclonal antibodies are prepared, for instance, by means of human peripheral blood lymphocytes (PBL) repopulation of severe combined immune deficiency (SCID) mice as described in PCT/EP 99/03605 or by using transgenic non-human animals capable of producing human antibodies as described in US patent 5,545,806. Also fragments derived from these monoclonal antibodies such as Fab, F(ab)₂ and ssFv ("single chain variable fragment"), providing they have retained the original binding properties, form part of the present invention. Such fragments are commonly generated by, for instance, enzymatic digestion of the antibodies with papain, pepsin, or other proteases. It is well known to the person skilled in the art that monoclonal antibodies, or fragments thereof, can be modified for various uses. The antibodies can also be labeled by an appropriate label of the enzymatic, fluorescent, or radioactive type.

The terms 'small molecules' relate to, for example, small organic molecules, and other drug candidates which can be obtained, for example, from combinatorial and natural product libraries via methods well-known in the art. Random peptide libraries consisting of all possible combinations of amino acids attached to a solid phase support may be used to identify peptides that are able to bind to SIP1 or to the promotor region bound by SIP1. The screening of peptide libraries may have therapeutic value in the discovery of pharmaceutical agents that act to inhibit the biological activity of SIP1.

The terms 'anti-sense nucleic acids' and 'ribozymes' refer to molecules that function to inhibit the translation of SIP1 mRNA. Anti-sense nucleic acids or anti-sense RNA and DNA molecules act to directly block the translation of mRNA by binding to targeted mRNA and preventing protein translation. Ribozymes are enzymatic RNA

molecules capable of catalyzing the specific cleavage of RNA. The mechanism of ribozyme action involves sequence specific hybridization of the ribozyme molecule to complementary target RNA, followed by an endonucleolytic cleavage. Within the scope of the invention are engineered hammerhead motif ribozyme molecules that specifically and efficiently catalyze endonucleolytic cleavage of SIP1 RNA sequences. Specific ribozyme cleavage sites within any potential RNA target are initially identified by scanning the target molecule for ribozyme cleavage sites which include the following sequences, GUA, GUU and GUC. Once identified, short RNA sequences of between 15 and 20 ribonucleotides corresponding to the region of the target gene containing the cleavage site may be evaluated for predicted structural features such as secondary structure that may render the oligonucleotide sequence unsuitable. The suitability of candidate targets may also be evaluated by testing their accessibility to hybridization with complementary oligonucleotides, using ribonuclease protection assays. Both anti-sense RNA and DNA molecules and ribozymes of the invention may be prepared by any method known in the art for the synthesis of RNA molecules. These include techniques for chemically synthesizing oligodeoxyribonucleotides well known in the art such as for example solid phase phosphoramidite chemical synthesis. Alternatively, RNA molecules may be generated by *in vitro* and *in vivo* transcription of DNA sequences encoding the antisense RNA molecule. Such DNA sequences may be incorporated into a wide variety of vectors which incorporate suitable RNA polymerase promoters such as the T7 or SP6 polymerase promoters. Alternatively, antisense cDNA constructs that synthesize anti-sense RNA constitutively or inducibly, depending on the promoter used, can be introduced stably into cell lines.

The above described antibodies, small molecules, anti-sense nucleic acids and ribozymes can be used as 'a medicament' to prevent and/or treat tumor invasion and/or metastasis via inhibiting the down-regulation of E-cadherin expression by SIP-1. Malignancy of tumors implies an inherent tendency of the tumor's cells to metastasize (invade the body widely and become disseminated by subtle means) and eventually to kill the patient unless all the malignant cells can be eradicated. Metastasis is thus the outstanding characteristic of malignancy. Metastasis is the tendency of tumor cells to be carried from their site of origin by way of the circulatory

system and other channels, which may eventually establish these cells in almost every tissue and organ of the body. In contrast, the cells of a benign tumor invariably remain in contact with each other in one solid mass centred on the site of origin. Because of the physical continuity of benign tumor cells, they may be removed
5 completely by surgery if the location is suitable. But the dissemination of malignant cells, each one individually possessing (through cell division) the ability to give rise to new masses of cells (new tumors) in new and distant sites, precludes complete eradication by a single surgical procedure in all but the earliest period of growth. It should be clear that the 'medicament' of the present invention can be used in
10 combination with any other tumor therapy known in the art such as irradiation, chemotherapy or surgery.

With regard to the above-mentioned small molecules, the term 'medicament' relates to a composition comprising small molecules as described above and a pharmaceutically acceptable carrier or excipient (both terms can be used
15 interchangeably) to treat diseases as indicated above. Suitable carriers or excipients known to the skilled man are saline, Ringer's solution, dextrose solution, Hank's solution, fixed oils, ethyl oleate, 5% dextrose in saline, substances that enhance isotonicity and chemical stability, buffers and preservatives. Other suitable carriers include any carrier that does not itself induce the production of antibodies harmful to
20 the individual receiving the composition such as proteins, polysaccharides, polylactic acids, polyglycolic acids, polymeric amino acids and amino acid copolymers. The 'medicament' may be administered by any suitable method within the knowledge of the skilled man. The preferred route of administration is parenterally. In parental administration, the medicament of this invention will be formulated in a unit dosage
25 injectable form such as a solution, suspension or emulsion, in association with the pharmaceutically acceptable excipients as defined above. However, the dosage and mode of administration will depend on the individual. Generally, the medicament is administered so that molecule of the present invention is given at a dose between 1 $\mu\text{g/kg}$ and 10 mg/kg , more preferably between 10 $\mu\text{g/kg}$ and 5 mg/kg , most preferably
30 between 0.1 and 2 mg/kg . Preferably, it is given as a bolus dose. Continuous infusion may also be used and includes continuous subcutaneous delivery via an osmotic minipump. If so, the medicament may be infused at a dose between 5 and

20 µg/kg/minute, more preferably between 7 and 15 µg/kg/minute.

With regard to the antibodies, anti-sense nucleic acids and ribozymes of the present invention, a preferred mode of administration of the 'medicament' for treatment is the use of gene therapy to deliver the above mentioned molecules. Gene therapy means
5 the treatment by the delivery of therapeutic nucleic acids to patient's cells. This is extensively reviewed in Lever and Goodfellow 1995; Br. Med Bull.,51, 1-242; Culver 1995; Ledley, F.D. 1995. Hum. Gene Ther. 6, 1129. To achieve gene therapy there must be a method of delivering genes to the patient's cells and additional methods to ensure the effective production of any therapeutic genes. There are two general
10 approaches to achieve gene delivery; these are non-viral delivery and virus-mediated gene delivery.

The following examples more fully illustrate preferred features of the invention, but are not intended to limit the invention in any way.

15

Examples

- Characterization of nucleic acid sequences at least comprising a CACCT sequence.

20 Introduction and summary

- SIP1 and δEF1 bind to target sites containing one CACCT sequence and one CACCTG sequence

The present invention regards the DNA binding properties of SIP1. As stated above,
25 SIP1, a recently isolated Smad-interacting protein, belongs to the emerging family of two-handed zinc finger transcription factors (34). The organization of SIP1 is very similar to that of δEF1, the prototype member of this family. Both proteins contain two widely separated clusters of zinc fingers, which are involved in binding to DNA. The amino acid sequence homology is very high (more than 90%) within these two zinc
30 finger clusters, whereas it is less evident in the other regions. This finding suggests that both proteins would bind in an analogous fashion to similar DNA targets. Indeed

SIP1 as well as δ EF1 bind with comparable affinities to many different target sites, which always contain two CACCT sequences. For all the target sites tested here, the integrity of both CACCT sequences is absolutely necessary for the binding of either SIP1 or δ EF1.

5 SIP1_{FS} inhibits *Xbra2* expression when overexpressed in the *Xenopus* embryo (34), and SIP1_{FS} binds to the *Xbra2* promoter by contacting two CACCT sequences. Recent studies using *Xenopus* transgenic embryos have shown that 2.1 kb of *Xbra2* promoter sequences suffice to express a reporter protein in the same domain as *Xbra* itself (17). However, a single point mutation within the downstream CACCT site (Xbra-D) in
10 the promoter that disrupts SIP1 binding (as seen in gel retardation assays) has a severe effect. Expression of the marker protein initiates earlier (i.e. at stage 9), and is now found at ectopic sites, e.g. in the majority of ectodermal, mesodermal and endodermal cells (17). This indicates that this nucleotide, which is located within the downstream CACCT site, is required for correct spatial and temporal expression of
15 the *Xbra2* gene. In addition, when a mutation is introduced in the upstream CACCT sequence, we observed the same premature and ectopic expression of *Xbra2* as for the mutation within the downstream CACCT site. Therefore, mutations in either the downstream or upstream CACCT that are known to affect SIP1 or δ EF1 binding in EMSA, give the same phenotype *in vivo*, indicating that a *Xenopus* δ EF1-like protein
20 participates in the regulation of the *Xbra2* gene. In addition, these *in vivo* data support the conclusions from the *in vitro* binding experiments presented here: SIP1/ δ EF1-like transcription factors require two CACCT sites for regulating the expression of the *Xbra2* promoter.

Not all promoter regions containing two CACCT sequences represent SIP1 or δ EF1
25 binding sites. Notably, duplication of the Xbra-F probe, which contains the upstream CACCT sequence present in the Xbra-WT element, is refractory to binding of either SIP1 or δ EF1. Moreover, neither SIP1_{NZF} nor SIP1_{CZF} can bind efficiently to this site (Xbra-F) as monomer or as dimer. Thus other sequences in addition to CACCT may be required for generating a high-affinity binding site. It appears that CACCTG is
30 always a better target site for binding of these zinc finger clusters. Indeed, the high-affinity CACCTG site (Xbra-E) was shown to bind either the SIP1_{NZF} or the SIP1_{CZF} cluster. In addition, modification of the CACCTG site into CACCTA strongly affects the

binding of SIP1_{FS} and δ EF1 to the Xbra promoter, confirming the importance of this 3'-
 guanine residue. By comparing the sequence of all the SIP1 and δ EF1 target sites, a
 minimal consensus sequence was found composed of one CACCT sequence and
 one CACCTG sequence, demonstrating that these two sequences are sufficient to
 5 form a high-affinity binding site for SIP1 or δ EF1.

Although the upstream CACCT sequence is unable to bind SIP1_{CZF} or SIP1_{NZF}, this
 sequence is contacted by full size SIP1 in the context of the Xbra-WT probe. The
 upstream CACCT sequence is a prerequisite for the binding of SIP1_{FS} to the Xbra-WT
 probe. Thus, when the upstream CACCT sequence is combined with another, high-
 10 affinity CACCTG site (Xbra-E), this low affinity site (Xbra-F) becomes committed to the
 binding of SIP1_{FS}. A model in which SIP1_{FS} contacts its target promoter via the binding
 of one of its zinc fingers clusters to a high affinity CACCTG-sequence (e.g. Xbra-E) is
 favoured, which is followed by the contact of the low affinity CACCT site (Xbra-F) by
 the second cluster, and this additional interaction strongly stabilizes SIP1 binding.
 15 Therefore, a CACCT site may still have an important function in the regulation of gene
 expression, while even on its own it neither binds SIP1_{NZF}, SIP1_{CZF} nor SIP1_{FS}.

The DC5 probe from the δ 1-crystallin enhancer was previously shown to bind
 specifically δ EF1 (31). However, this probe contains only one CACCT sequence.
 Therefore, despite having demonstrated here that high affinity binding sites for δ EF1
 20 should contain one CACCT sequence and one CACCTG sequence, it cannot be
 excluded that in particular cases, such as the DC5 probe, one CACCT site would be
 sufficient for the binding of this type of transcription factor.

- Mode of SIP1 DNA binding

When tested independently in EMSA, both the C-terminal as well as the N-terminal
 25 zinc finger clusters of SIP1 or δ EF1 bind to very similar CACCT-containing consensus
 sequences. Both for SIP1 and δ EF1, NZF3 and NZF4 share an extensive amino acid
 sequence homology with CZF2 and CZF3, respectively. This homology may explain
 why these two clusters can bind to similar consensus sequences. In addition, it has
 been shown that SIP1 or δ EF1 require two CACCT sequences for binding to several
 30 potential target sites. Based on these results, it is proposed that SIP1 and δ EF1 would
 bind to their target elements in such a way that one zinc finger cluster contacts one of

the CACCT sites, while the other cluster contacts the second CACCT site (see figure 2, model 1). An alternative model would be that SIP1 or δ EF1 homodimerizes before being able to bind to these target sites with high affinity (model 2). The DNA binding capacity of SIP1_{NZF} is abolished by mutations in either NZF3 or NZF4. Similarly, mutations within CZF2 or CZF3 also affect the binding capacity of SIP1_{CZF}. When these mutations are introduced in the context of the full size SIP1, binding of SIP1_{FS} is not observed any longer. This clearly indicates that the binding activity of both zinc finger clusters is required for the binding of SIP1_{FS} to its target element, containing a doublet of CACCT sites. Similarly, it was previously shown that the integrity of both zinc finger clusters of δ EF1 is also necessary for binding DNA (31). These observations indicate that both zinc fingers clusters are contacting directly the DNA. Therefore, in the dimer model (Fig. 2, model 2), the SIP1_{NZF} of one SIP1 molecule should bind to one CACCT sequence and the SIP1_{CZF} of the second SIP1 molecule should contact the other CACCT sequence. If such a dimer configuration would exist, then it can be assumed that certain combinations of full size SIP1 molecules having different mutations within CZF or NZF, respectively, should allow the formation of functional dimer which is able to bind to its target DNA. None of the possible combinations of the four SIP1_{FS} mutants tested (NZF3mut, NZF4mut, CZF2mut and CZF3mut) gave rise to a DNA/SIP1 complex in EMSAs. This argues against the existence of SIP1 dimers. In addition, using differently tagged SIP1_{FS} molecules, detection of SIP1 dimers in EMSAs was not possible, nor to supershift such dimeric complexes with different antibodies. Therefore support is provided to model 1 in which SIP1 binds as a monomer to a target site, which contains one CACCT sequence and one CACCTG sequence.

It has been shown in this invention 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. This demonstrates that these transcription factors should display a highly flexible secondary structure to accommodate the binding to these different target sites. The long linker region between the two zinc finger clusters within SIP1 and δ EF1 may permit this flexibility in the secondary structure of these proteins. These transcription factors can bind to sites containing CACCT sequences separated by at least 44 bp (Ecad-WT), suggesting that a region of about 50 bp of promoter

sequences might be covered and therefore less accessible to transcriptional activators once SIP1_{FS} or δ EF1 is bound to this promoter. This indicates that SIP1 or δ EF1 could function as transcriptional repressor by competing with transcriptional activators that bind in this region covered by SIP1 or δ EF1.

5 ***- Other families of transcription factors may bind DNA with a similar mechanism as SIP1***

This new mode of DNA binding 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 (1, 3, 6, 29, 33). Like for SIP1 and δ EF1, the
 10 conservation of these zinc finger clusters is very strong between the different members of this family (1). In addition, the C-terminal cluster is very homologous to the N-terminal cluster and, in the case of PRDII-BF1, these clusters bind to the same sequences when tested independently (3). Therefore, this type of transcription factor may bind to two reiterated sequences through the contact of one zinc finger cluster
 15 with one sequence and the other cluster with the second sequence. Similarly, the different members of the NZF family of transcription factors also have two widely separated clusters of zinc fingers (11, 12, 36). MyT1, NZF-1 and NZF-3 all bind to the same consensus element AAAGTTT. Like for SIP1 and δ EF1, which show a significantly higher affinity to elements containing 2 CACCT sequences, an element
 20 containing 2 AAAGTTT sequences demonstrated a markedly higher affinity to NZF-3 (36). This suggests that 2 AAAGTTT sequences are also necessary to create a high-affinity binding site for these transcription factors, and that they may bind DNA with a similar mechanism as SIP1 and δ EF1. Finally, the Evi-1 protein, which contains 7 zinc fingers at the N-terminus and 3 zinc fingers at the C-terminus, binds to two consensus
 25 sequences. It binds to a complex consensus sequence (GACAAGATAAGATAA-N₁₋₂₈-CTCATCTTC) via a mechanism that may involve the binding of the N-terminal zinc finger cluster to the first part and the binding of the C-terminal cluster to the second part (20). In conclusion, the mode of DNA-binding that is described here may not only be applicable to the SIP1/ δ EF1 family of transcription factors, but is more universal.

30 SIP1 was cloned as a Smad1-interacting protein but was also shown to interact with Smad2, 3 and 5 (34). Smad proteins are signal transducers involved in the BMP/TGF-

β signaling cascade (13). 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 migrate to 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 (34). Surprisingly, Evi-1, a transcription factor that may bind DNA with a similar mechanism as SIP1, is a Smad3-interacting protein (15). So far, it was shown that Evi-1 inhibited the binding of Smad3 to DNA but certainly has an effect on target promoters of Evi-1. *Schnurri*, which is the *Drosophila* homologue of the human PRDII-BF1 transcription factor, is a protein that may also bind DNA with a similar mechanism as SIP1 protein. Interestingly, *Schnurri* was proposed to be a nuclear protein target in the dpp-signaling pathway (1, 6). Dpp is a member of the TGF- β family. This makes *Schnurri* a candidate nuclear target for *Drosophila* Mad protein, the *Drosophila* homologue of vertebrate Smads. Therefore, the mode of DNA binding employed by SIP1 can be generalized to other zinc finger containing Smad-interacting proteins, and represents a common feature of several Smad partners in the nucleus.

Based on these results, a novel mode of DNA binding for δ EF1 family of transcription factors is demonstrated. This mode of DNA binding is also relevant to other families of transcription factor that contains separated clusters of zinc fingers.

Materials and methods used in this example

Plasmid constructions.

For expression in mammalian cells, the SIP1 (34) and δ EF1 (5) cDNAs were subcloned into pCS3 (27). In this plasmid, the SIP1 and δ EF1 open reading frames are fused to a (Myc)₆ tag at the N-terminus. SIP1 cDNA was also cloned into pCDNA3 (Invitrogen) as a N-terminal fusion with the FLAG tag. For the expression of SIP1_{NZF} and SIP1_{CZF}, we subcloned into pCS3 the cDNA fragments encoding amino acids 1 to 389 and 977 to 1214, respectively. SIP1_{CZF} (as amino acids 957 to 1156) and SIP1_{NZF} (amino acids 90 to 383) were also produced in *E. coli* as a GST fusion protein (in pGEX-5X-1, Pharmacia) and purified using the GST purification module (Pharmacia). Identical mutations to those made in AREB6 (10) were also introduced in the SIP1

zinc fingers. Mutagenesis of zinc fingers NZF3, NZF4, CZF2 and CZF3 involved substitution of their third histidine to a serine. These mutations were introduced using a PCR based approach with the following primers: SIP1_{NZF3Mut}, 5'-CCACCTGAAAGAATCCCTGAGAATTCACAG; SIP1_{NZF4Mut}, 5'-GGGTCCTACAGTTCATCTATCAGCAGCAAG; SIP1_{CZF2Mut}, 5'-CACCACCTTATCGAGTCCTCGAGGCTGCAC; SIP1_{CZF3Mut}, 5'-TCCTACTCGCAGTCCATGAATCACAGGTAC. The respective mutated clusters were recloned in full size SIP1 in pCS3 in order to produce in mammalian cells the mutated SIP1 proteins named NZF3mut, NZF4mut, CZF2mut and CZF3mut, respectively.

Furthermore, these mutated clusters were subcloned into pGEX5-X2 (Pharmacia), and produced in *E.coli* as a GST fusion protein (GST-NZF3mut, GST-NZF4mut, GST-CZF2mut and GST-CZF3mut). All constructs were confirmed by restriction mapping and sequencing.

Cell culture and DNA transfection.

COS1 cells were grown in DMEM supplemented with 10% fetal bovine serum. Cells were transfected using Fugene according to the manufacturer's protocol (Boehringer Mannheim), and collected 30-48 hrs after transfection.

Gel retardation assay.

The Xbra-WT oligonucleotide covers the region from -344 to -294 of the Xbra2 promoter (16). The region between -412 to -352 of the α 4-integrin promoter is present within the α 4I-WT oligonucleotide (26). The Ecad-WT probe contains the region between -86 to -17 of the human Ecad promoter (2). The sequences of the upper strand of the wild types and mutated double-stranded probes are listed in Table 1.

Double-stranded oligonucleotides were labeled with [³²P]- γ -ATP and T4 polynucleotide kinase (New England Biolabs). Total cell extracts were prepared from COS1 cells (25) transfected with different pCS3 vectors allowing synthesis of full length SIP1, full length δ EF1, and different mutant forms of SIP1 (25), or coproduction of equal amounts of Myc-tagged SIP1 and FLAG-tagged SIP1. GST-SIP1 fusion proteins were purified from *E.coli* extract using the GST purification module (Pharmacia), and tested in gel retardation. The DNA binding assay (20 μ l) was performed 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 (approx. 10⁴ Cerenkov counts) in the

δEF1 binding buffer described previously (30). For supershift experiments, the extracts were incubated with anti-Myc (Santa Cruz) or anti-FLAG (Kodak) antibodies. For competition, an excess of unlabeled double-stranded oligonucleotides was added together with the labeled probe. The binding reaction was loaded onto a 4% polyacrylamide gel (acrylamide/bis-acrylamide, 19:1) prepared in 0.5XTBE buffer. Following electrophoresis, gels were dried and exposed to X-Ray film. All experiments were repeated at least three times.

Methylation interference assay.

The upper and the lower strand of the Xbra-WT probe were labeled separately and annealed with excess of complementary DNA strand. The probes were precipitated and treated with di-methyl-sulfate (8). The methylated probe (10^5 Cerenkov counts) was incubated in a 10 X gel retardation reaction (see above) (200 µl final volume) with 10 µg of total cell extract from COS1 cells expressing either SIP1_{FS} or SIP1_{CZF}. After 20 min. of incubation at 25°C, the products were loaded onto a 4% polyacrylamide gel, and electrophoresis was performed as for the gel retardation assay. Subsequently, the gel was blotted onto DEAE-cellulose membrane; the transfer was performed at 100 V for 30 min. in 0.5XTBE buffer. The membrane was then exposed for one hour, and the bands corresponding to the SIP1_{FS} (or SIP1_{CZF}) and the free probe were eluted at 65°C, using high salt conditions (1M NaCl, 20 mM Tris, pH7.5, 1 mM EDTA). The eluted DNA was precipitated and treated with piperidine (18). After several cycles of solubilization in water and evaporation of the liquid under vacuum, the resulting DNA pellet was dissolved in 10 µl of sequencing buffer (97.5 % deionized formamide, 0.3 % each Bromophenol Blue and Xylene Cyanol, 10 mM EDTA) and denatured for 5 min. at 85 °C. The same amount of counts (1,500 Cerenkov counts) for the free probe and the bound probe was loaded onto a 20% polyacrylamide-8M urea sequencing gel. The gel was run in 0.5XTBE for one hour at 2,000 V. Thereafter, the gel was fixed in 50% methanol/10% acetic acid and dried. The gel was then exposed for autoradiography.

Western blot analysis.

Transfected cells were washed with PBS-O (137 mM NaCl, 2.7 mM KCl, 6.5 mM Na₂HPO₄, 1.5 mM KH₂PO₄), collected in detachment buffer (10 mM Tris pH7.5, 1mM EDTA, 10% glycerol, with protease inhibitors (Protease inhibitor Cocktail tablets,

Boehringer Mannheim)) and pelleted by low spin centrifugation. The cells were then solubilized in 10 mM Tris, pH 7.4, 125 mM NaCl, 1% Triton X-100. For direct electrophoretic analysis, gel sample buffer was added to the cell lysates and the samples were boiled. For other experiments, lysates were first subjected to immunoprecipitation with either anti-Myc or anti-FLAG antibodies. Antibodies were added to aliquots of the cell lysates, which were incubated overnight at 4°C. The antibodies and the bound protein(s) of the cell lysate were coupled as a complex to protein A-Sepharose for 2 hours at 4 °C. The immunoprecipitates were washed 4 times in NET buffer (50 mM Tris pH 8.0, 150 mM NaCl, 0.1% NP40, 1 mM EDTA, 0.25% gelatin), resolved by SDS-polyacrylamide (7.5%) gel electrophoresis, and electrophoretically transferred to nitrocellulose membranes. Membranes were blocked for 2 hours in TBST (10 mM Tris pH 7.5, 150 mM NaCl, 0.1 % Tween-20) containing 3% (w/v) non-fat milk, and incubated with primary antibody (1µg/ml) for 2 hours, followed by secondary antibody (0.5 µg/ml) linked to horseradish peroxidase. Immunoreactive bands were detected with an enhanced chemiluminescence reagent (NEN).

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

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 spermnuclei suspension per egg, at a theoretical concentration of 2 nuclei per 5 nl. *NotI* was used for plasmid linearisation 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 to 80 % completed gastrulation and 20 to 30% developed further into normal swimming tadpoles, if allowed. The transgenic frequency, as analysed by expression, varied between 50 to 90%. Embryos were staged according to Niewkoop and Faber (Niewkoop and Faber, 1967). A minimum of 30 expressing embryos were analysed per construct and shown stage. Whole-mount *in situ* hybridisation 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.

Table 1.

	Oligo	Sequence	Spacing
	Xbra-WT	ATCCAGGCC CACCT AAAAATATAGAATGATAAAGTGACC AGGTGTCAGTTCT	24
5	Xbra-D	-----A-----	
	Xbra-E	TAAAGTGACC AGGTGTCAGTTCT	
	Xbra-F	ATCCAGGCC CACCT AAAAATATAGAATGA	
10	Rdm + Xbra-E	CAATTTAGAGTACTGTGTACTTGGGAGTAAAGTGACC AGGTGTCAGTTCT	
	Xbra-F + AREB6	ATCCAGGCC CACCT AAAAATATAGAATGAGGCTCAGAC AGGTG TAGAATTCGGCG	23
	Rdm + AREB6	CAATTTAGAGTACTGTGTACTTGGGAGGGCTCAGAC AGGTG TAGAATTCGGCG	
	Xbra-WT	ATCCAGGCC CACCT AAAAATATAGAATGATAAAGTGACC AGGTGTCAGTTCT	24
15	Xbra-J	CGA-----	
	Xbra-K	---ACT-----	
	Xbra-L	-----TAA-----	
	Xbra-M	-----CAA-----	
	Xbra-N	-----GCC-----	
20	Xbra-O	-----CCG-----	
	Xbra-P	-----CGC-----	
	Xbra-Q	-----TCC-----	
	Xbra-R	-----GTC-----	
	Xbra-S	-----T-----	
25	Xbra-Z	-----T-----	
	Xbra-WT	ATCCAGGCC CACCT AAAAATATAGAATGATAAAGTGACC AGGTGTCAGTTCT	24
	Xbra-B	ATCCAGGCC CACCTA ____TATAGAATGATAAAGTGACC AGGTGTCAGTTCT	21
	Xbra-C	ATCCAGGCC CACCT AAAAATATAGAATGAT____GTGACC AGGTGTCAGTTCT	21
30	Xbra-U	ATCCAGGCC CACCT AAAAATATA____GTGACC AGGTGTCAGTTCT	14
	Xbra-EE	TAAAGTGACC AGGTGTCAGTTCT TAAAGTGACC AGGTGTCAGTTCT	18
	Xbra-ErE	AGAACTGAC CACCT GGTCACTTTATAAAGTGACC AGGTGTCAGTTCT	20
	Xbra-FrF	ATCCAGGCC CACCT AAAAATATAGAATATTCTATATTT AGGTGGCCTGGAT	24
	Xbra-V	ATCCAGGC AGGTGT AAATATAGAATGATAAAGTGACCC CACCTACAGTTCT	24
35	Xbra-W	ATCCAGGC AGGTGT AAATATAGAATGATAAAGTGACC AGGTGTCAGTTCT	24
	α 4I-WT	GCAGGGC CACCT GGATTGCATTAGAATGAGACTCACTACCCAGTT CAGGTGTGTTGCGT	34
	α 4I-A	-----A-----	
	α 4I-B	-----T-----	
40	Ecad-WT	TGGCCGGC AGGTGA ACCCTCAGCCAATCAGCGGTACGGGGGGCGGTGCTCCGGGGCT CACCTGGCTGCAG	44
	Ecad-A	-----T-----	
	Ecad-B	-----A-----	

45 Table 1. List of all the probes used in this study. The CACCT sequences have been highlighted in bold. The spacing (right column) is the number of nucleotides present between the two CACCT sequences. Underlined gaps correspond to deletions of nucleotides from the wild type probes. For many probes, only the residues that have been changed compared with the wild type probes have been indicated in order to facilitate interpretation of the introduced mutations.

50

The following 8 paragraphs contain some additional –materials and methods- in order to perform the further described experiments:

-Gel retardation assay with different probes from the Xbra2 promoter. The different Xbra ³²P labeled probes (10 pg) were incubated with 1 µg of total protein extract from COS1 cells transfected with pCS3-SIP1_{CZF}, with pCS3-SIP1_{FS} or from mock-transfected cells.

-Two CACCT sites are contacted upon binding of SIP1_{FS} to the Xbra2 promoter.

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 XbraWT abolish SIP1_{FS} binding. Methylation interference assay indicates that SIP1_{FS} contacts both CACCT sequences. XbraWT either labeled in the upper or the lower strand were methylated and incubated with total extract from COS1 cells transfected either with pCS3-SIP1_{FS} or pCS3-SIP1_{CZF}. The DNA retarded in the shifted complex or the unbound DNA (FREE) were purified, cleaved with piperidine and run onto a sequencing gel. Guanine residues are methylated in the free probe. The upstream and the downstream CACCT from the Xbra2 promoter is indicated.

-Two CACCT sequences are necessary for the binding of SIP1_{FS} and δEF1 to the Xbra2, the α4-integrin and the E-cadherin promoters. δEF1 binding to the Xbra2 promoter; SIP1 and δEF1 binding to the α4-integrin promoter.; binding of SIP1 and δEF1 to the α4-integrin promoter, including competition with excess of non-labeled wild type and mutated binding sites; binding of SIP1 and δEF1 to the E-cadherin promoter. In each binding reaction, 10 pg of labeled probes were incubated with 1 µg of a total cell protein extract prepared from COS1 cells transfected with either pCS3-SIP1_{FS} or pCS3-δEF1. In the competition experiments, 5 ng and 50 ng of unlabeled DNA was added at the same time as the labeled probe. Myc-tag directed antibody was added to the binding reaction and the supershifted complex. δEF1 and the SIP1 retarded complexes were demonstrated. For the sequences of all probes, see Table1.

-The spacing and the relative orientation of the CACCT sequences are not critical for the binding of SIP1_{FS} and δEF1 to the Xbra2 promoter. Ten pg of labeled probes were incubated with 1 µg of a total cell protein extract prepared from COS1 cells transfected with either pCS3-SIP1_{FS} or pCS3-δEF1. We used 10 pg of the Xbra-E probe and 10 pg of the Xbra-F probe in the same binding reaction. For

reasons of clear and comparative presentation, we omitted the free probe from the SIP1 binding reactions.

-The integrity of both SIP1 zinc finger clusters is necessary for the binding of SIP1_{FS} to DNA. Mutations within NZF3, NZF4, CZF2, CZF3 abolish the DNA-binding

activity of either the SIP1_{NZF} or SIP1_{CZF} zinc finger clusters. The wild type and mutated

zinc finger clusters were fused to GST and the fusion proteins were produced in

E.coli. After purification, an equal amount of each fusion proteins (0.1 ng) was

incubated with 10pg of labeled Xbra-E probe. Mutations within NZF3, NZF4, CZF2 or

CZF3 affect the binding of SIP1_{FS} to the Xbra-WT probe. Ten pg of labeled Xbra-WT

probe were incubated with 1 µg of a total cell protein extract prepared from COS1

cells transfected with either pCS3-SIP1_{FS}, pCS3-SIP1_{NZF3mut}, pCS3-SIP1_{NZF4mut}, pCS3-

SIP1_{CZF2mut} or pCS3-SIP1_{CZF3mut}. All possible combinations of 2 COS cell extracts (1 µg

of each) expressing different of SIP1 mutants were tested. Myc-tag directed antibody

was added to the binding reaction and the supershifted complex and the SIP1_{FS}

retarded complex are indicated. Mutations within NZF3, NZF4, CZF2 or CZF3 abolish

the binding of SIP1_{FS} to the α4-integrin promoter. Ten pg of labeled α4I -WT probe

were incubated with 1 µg of a total cell protein extract prepared from COS1 cells

transfected with either pCS3-SIP1_{FS}, pCS3-SIP1_{NZF3mut}, pCS3-SIP1_{NZF4mut}, pCS3-

SIP1_{CZF2mut} or pCS3-SIP1_{CZF3mut}. Myc-tag directed antibody were added to the binding

reaction and the supershifted complex and the SIP1_{FS} retarded complex are indicated.

SIP1 mutants are produced in comparable amounts in COS cells. Ten µg of the COS

cell total extract were analyzed by Western blotting using the anti-Myc antibody. SIP1

mutant expression levels are in fact slightly higher that SIP1-WT expression level.

-SIP1_{FS} binds as a monomer to the Xbra-WT probe. 10 pg of labeled Xbra-WT

probe were incubated with 1 µg of total cell protein prepared from COS1 cells

transfected with an equal amount of pCS3-SIP1_{FS} (Myc-tagged) and of pCDNA3-SIP1

(Flag-tagged). Anti-Flag and anti-Myc antibodies were added separately or both anti-

Flag and anti-Myc antibodies were added to the binding assay. The Flag- and the

Myc-supershifted complexes are indicated.

-The integrity of CZF or NZF is necessary for SIP1 repressor activity. SIP1_{FS}

binding to a gel-purified fragment derived from the multiple CACCT-containing artificial

promoter from reporter plasmid p3TP-Lux. Anti-Myc tag antibody were added; the

supershifted complex is indicated. Co-transfection assay of pCS3-SIP1_{FS}, pCS3-CZF3-Mut or pCS3-NZF3-Mut together with the p3TP-Lux reporter vector. The activity is expressed in percentage of full SIP1_{FS} repressor activity, which is 100%.

-Ectopic activity of the mutated *Xbra2* promoter variants (*Xbra2*-Mut) in transgenic frog embryos. SIP1_{FS} binding to the wild-type and mutated (*Xbra*-Mut; see Table I) *Xbra2* promoter elements. Whole-mount *in situ* hybridisation for GFP mRNA of *Xenopus* embryos transgenic for a wild-type or point-mutated 2.1kb *Xbra2* promoter fragment driving a GFP reporter. All embryos were fixed at stage 11 and cleared for better visualisation 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).

Results

15 - SIP1 has a structure similar to δ EF1

SIP1 was recently isolated as a Smad-binding protein and binds Smad1, Smad 5 and Smad2 in a ligand-dependent fashion (in BMP and activin pathways) (34). SIP1 is a new member of the family of two-handed zinc finger/homeodomain transcription factors, which also includes vertebrate δ EF1 and *Drosophila* Zfh-1(4, 5). Like these, SIP1 contains two widely separated zinc finger clusters. One cluster of four zinc fingers (3 CCHH and 1 CCHC fingers) is located at the N-terminal region of the protein and another cluster of three CCHH zinc fingers is present at the C-terminal region (Fig. 1A). Between SIP1 and δ EF1, a high degree of sequence identity is apparent within the N-terminal zinc finger cluster (87 %) and the C-terminal zinc finger cluster (97%)(see Fig.1B), whereas the two proteins are less conserved in the regions outside the zinc finger clusters (34). Therefore, it is assumed that SIP1 and δ EF1 would bind to very similar sequences. In addition, the N-terminal and C-terminal zinc finger clusters of δ EF1 bind to very similar sequences, which contain the core CACCT consensus sequence (10). Within the N-terminal cluster, both δ EF1_{NZF3} and δ EF1_{NZF4} are the main determinants for binding to the CACCT consensus sequence, and δ EF1_{CZF2} and δ EF1_{CZF3} are required for the binding of the C-terminal cluster (10). Moreover, the δ EF1_{NZF3+NZF4} domain shows high homology (67 %) with the

$\delta EF1_{CZF2+CZF3}$ domain and this may explain why these two clusters bind to similar consensus target sites on DNA (Fig.1C). All the residues essential for binding, and which are conserved between $\delta EF1_{NZF3+NZF4}$ and $\delta EF1_{CZF2+CZF3}$, are also conserved between $SIP1_{NZF3+NZF4}$ and $SIP1_{CZF2+CZF3}$. Taken together, these comparisons suggested that the N- and C-terminal zinc finger clusters of SIP1 would also bind to very similar target sequences.

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

CACCT sites are necessary for the binding of SIP1 to the Xbra2 promoter. CACCT sites are necessary for the binding of SIP1 to the Xbra2 promoter. CACCT sites are necessary for the binding of SIP1 to the Xbra2 promoter. CACCT sites are necessary for the binding of SIP1 to the Xbra2 promoter. SIP1 binds to the *Xenopus* Xbra2 promoter and represses expression of Xbra2 mRNA when overexpressed in the *Xenopus* embryo (34). The Xbra2 promoter contains several CACCT sequences, two of which are localized in a region (-381 to -231) necessary for the induction by activin (16). These two sites, an upstream CACCT and a downstream AGGTG (i.e. 5'-CACCT on the other DNA strand) respectively, are separated by 24 bp. To further elucidate the binding requirements of SIP1 to these sites, a corresponding 50 bp-long oligonucleotide (Xbra-WT; for a list of all probes see Table 1) was used as a probe in electrophoretic mobility shift assays (EMSAs). The Xbra-D probe, that contains a mutation of the downstream AGGTG site to AGATG, was included also. A similar mutation was shown previously to abolish the binding of $\delta EF1$ to the $\kappa E2$ enhancer (30). In addition, we also tested the downstream site (probe Xbra-E) and the upstream site (probe Xbra-F) independently as shorter probes. These probes were incubated with total extracts of COS cells expressing the Myc-tagged C-terminal zinc finger cluster of SIP1 ($SIP1_{CZF}$), the Myc-tagged N-terminal zinc finger cluster of SIP1 ($SIP1_{NZF}$), or Myc-tagged full size SIP1 ($SIP1_{FS}$).

When mock-transfected COS cells are used as control with the A probe, two weak complexes and one strong complex are visualized. Using competitor oligonucleotides, the two weak complexes turned out to be non-specific, whereas the strong, fast migrating complex shows specificity for binding to the Xbra probe. The latter

observation suggests that COS cells contain an endogenous protein that can bind to the Xbra-WT probe. When SIP1_{CZF} is present in the extract, we observed a strong and slow migrating complex, in addition to the endogenous binding activity from the COS extract. This complex could be supershifted with an anti-Myc antibody, which confirms
 5 that it results from binding of SIP1_{CZF} to the Xbra-WT probe. Mutation of the downstream site (Xbra-D probe) strongly affected the formation of this SIP1_{CZF} complex. Moreover, SIP1_{CZF} binds to the Xbra-E probe, but not to the Xbra-F probe indicating that the downstream site is essential for binding of SIP1_{CZF}, and SIP1_{CZF} may exclusively bind to this site. The strong complex visualized with the Xbra-F probe
 10 was also present in SIP1_{FS} extracts and in mock extract, and originates from hitherto uncharacterized endogenous COS cells protein binding to the Xbra-F probe. In addition, COS cell extracts containing SIP1_{NZF} displayed similar binding patterns in EMSAs as obtained with SIP1_{CZF}. It is apparent that, like in δ EF1 (10), both zinc finger clusters of SIP1 have similar DNA binding features.

15 A strong complex, corresponding to SIP1_{FS}, is also generated with the Xbra-WT probe. It is important to mention that the SIP1_{CZF} production level in COS cells is approximately 50-fold higher than the SIP1_{FS} level. For each EMSA reaction, we always used the same amount of crude COS cell proteins. The binding of SIP1_{FS} to Xbra-WT probe is as strong as the binding of SIP1_{CZF}. Interestingly, this indicates that
 20 the affinity of SIP1_{FS} for Xbra-WT is at least 50 times higher than this of SIP1_{CZF}.

The SIP1_{FS} complex, similar to SIP1_{CZF} and SIP1_{NZF}, is absent when using the mutated Xbra-D probe. Thus, an intact downstream site is again required for the binding of SIP1_{FS}. In contrast to SIP1_{CZF} and SIP1_{NZF}, which bind with similar affinities to the Xbra-WT and Xbra-E probes, SIP1_{FS} does not bind to the Xbra-E probe. Like
 25 SIP1_{CZF} and SIP1_{NZF}, SIP1_{FS} does not bind to the Xbra-F probe. We conclude that the downstream site (AGGTG) is necessary for SIP1_{FS} to bind to the Xbra2 promoter. However, this site is not sufficient because additional sequences upstream of the Xbra-E probe are necessary for the binding of SIP1_{FS}. One of the reasons for which SIP1_{FS} was unable to bind to the Xbra-E probe may simply be the length of the Xbra-E
 30 probe, because it is shorter than the Xbra-WT probe. To test this, we prepared a probe containing a random sequence (Rdm) upstream of the Xbra-E probe (Table 1) in order to extend it to the same length as Xbra-WT. In contrast to SIP1_{CZF}, which

bound efficiently to Rdm+Xbra-E probe, SIP1_{FS} was unable to bind. This result demonstrates that length of the Xbra-E probe per se is not the cause of the failure of SIP1_{FS} to bind to this probe.

To substantiate that the Xbra-F oligonucleotide also contains sequences necessary
5 for the binding of SIP1_{FS}, we fused this oligonucleotide as well as a random sequence upstream of another CACCT site known to be bound strongly by AREB6 protein (10) (probes Xbra-F + AREB6 and Rdm + AREB6, respectively). SIP1_{CZF} binds, with equal affinity, both the Xbra-F + AREB6 and Rdm + AREB6 probes indicating that the AREB6 sequence is also recognized by SIP1_{CZF}. However, SIP1_{FS} only binds to the
10 Xbra-F + AREB6 probe but not to Rdm + AREB6. This confirms that the Xbra-F oligonucleotide contains sequences necessary for the binding of SIP1_{FS}. In addition, the only common feature between the Xbra-E and the AREB6 probe is the CAGGTGT sequence, suggesting that no other sequences than this CAGGTGT in the Xbra-E probe are necessary for the binding of SIP1_{FS}. One of the reasons for which SIP1_{FS} is
15 unable to bind to the Xbra-E probe might be because the length of the Xbra-E probe is shorter than the length of the Xbra-WT probe. To test this hypothesis, we prepared a probe containing a random sequence upstream of the Xbra-E probe to obtain the same length as the Xbra-WT probe. In contrast to SIP1_{CZF} that binds efficiently to this probe, SIP1_{FS} was unable to bind. This result clearly indicates that the length of the
20 Xbra-E probe was not the reason for which SIP1_{FS} does not bind to this probe. To substantiate that the Xbra-F oligonucleotide also contains sequences necessary for the binding of SIP1_{FS}, we fused that oligonucleotide as well as a random sequence upstream of another CACCT site known to bind strongly AREB6 protein (Xbra-F + AREB6 and Rdm + AREB6, respectively). We observed that SIP_{CZF} binds with equal
25 affinities both the Xbra-F + AREB6 and Rdm + AREB6 probes, indicating that the AREB6 sequence is also recognized by SIP1_{CZF}. However, SIP1_{FS} only binds to the Xbra-F + AREB6 probe and not to the Rdm + AREB6 probe. This confirms that the Xbra-F oligonucleotide contains sequences necessary for the binding of SIP1_{FS}. In addition, the only common denominator between the Xbra-E and the AREB6 probe is
30 the AGGTG sequence, suggesting that no other sequences than this AGGTG in the Xbra-E probe is necessary for the binding of SIP1_{FS}. One of the reasons for which SIP1_{FS} is unable to bind to the Xbra-E probe might be because the length of the Xbra-

E probe is shorter than the length of the Xbra-WT probe. To test this hypothesis, we prepared a probe containing a random sequence upstream of the Xbra-E probe to obtain the same length as the Xbra-WT probe. In contrast to SIP1_{CZF} that binds efficiently to this probe, SIP1_{FS} was unable to bind. This result clearly indicates that the length of the Xbra-E probe was not the reason for which SIP1_{FS} does not bind to this probe. To substantiate that the Xbra-F oligonucleotide also contains sequences necessary for the binding of SIP1_{FS}, we fused that oligonucleotide as well as a random sequence upstream of another CACCT site known to bind strongly AREB6 protein (Xbra-F + AREB6 and Rdm + AREB6, respectively). We observed that SIP_{CZF} binds with equal affinities both the Xbra-F + AREB6 and Rdm + AREB6 probes, indicating that the AREB6 sequence is also recognized by SIP1_{CZF}. However, SIP1_{FS} only binds to the Xbra-F + AREB6 probe and not to the Rdm + AREB6 probe. This confirms that the Xbra-F oligonucleotide contains sequences necessary for the binding of SIP1_{FS}. In addition, the only common denominator between the Xbra-E and the AREB6 probe is the AGGTG sequence, suggesting that no other sequences than this AGGTG in the Xbra-E probe is necessary for the binding of SIP1_{FS}. One of the reasons for which SIP1_{FS} is unable to bind to the Xbra-E probe might be because the length of the Xbra-E probe is shorter than the length of the Xbra-WT probe. To test this hypothesis, we prepared a probe containing a random sequence upstream of the Xbra-E probe to obtain the same length as the Xbra-WT probe. In contrast to SIP1_{CZF} that binds efficiently to this probe (Fig.2, lane 6), SIP1_{FS} was unable to bind (lane3). This result clearly indicates that the length of the Xbra-E probe was not the reason for which SIP1_{FS} does not bind to this probe. To substantiate that the Xbra-F oligonucleotide also contains sequences necessary for the binding of SIP1_{FS}, we fused that oligonucleotide as well as a random sequence upstream of another CACCT site known to bind strongly AREB6 protein (Xbra-F + AREB6 and Rdm + AREB6, respectively). In lanes 4 and 5, we observed that SIP_{CZF} binds with equal affinities both the Xbra-F + AREB6 and Rdm + AREB6 probes, indicating that the AREB6 sequence is also recognized by SIP1_{CZF}. However, SIP1_{FS} only binds to the Xbra-F + AREB6 probe (lane 1) and not to the Rdm + AREB6 probe. This confirms that the Xbra-F oligonucleotide contains sequences necessary for the binding of SIP1_{FS}. In addition, the only common denominator between the Xbra-E and the AREB6 probe is the

AGGTG sequence, suggesting that no other sequences than this AGGTG in the Xbra-E probe is necessary for the binding of SIP1_{FS}. One of the reasons for which SIP1_{FS} is unable to bind to the Xbra-E probe might be because the length of the Xbra-E probe is shorter than the length of the Xbra-WT probe. To test this hypothesis, we prepared a probe containing a random sequence upstream of the Xbra-E probe to obtain the same length as the Xbra-WT probe. In contrast to SIP1_{CZF} that binds efficiently to this probe (Fig.2, lane 6), SIP1_{FS} was unable to bind (lane3). This result clearly indicates that the length of the Xbra-E probe was not the reason for which SIP1_{FS} does not bind to this probe. To substantiate that the Xbra-F oligonucleotide also contains sequences necessary for the binding of SIP1_{FS}, we fused that oligonucleotide as well as a random sequence upstream of another CACCT site known to bind strongly AREB6 protein (Xbra-F + AREB6 and Rdm + AREB6, respectively). In lanes 4 and 5, we observed that SIP_{CZF} binds with equal affinities both the Xbra-F + AREB6 and Rdm + AREB6 probes, indicating that the AREB6 sequence is also recognized by SIP1_{CZF}. However, SIP1_{FS} only binds to the Xbra-F + AREB6 probe (lane 1) and not to the Rdm + AREB6 probe. This confirms that the Xbra-F oligonucleotide contains sequences necessary for the binding of SIP1_{FS}. In addition, the only common denominator between the Xbra-E and the AREB6 probe is the AGGTG sequence, suggesting that no other sequences than this AGGTG in the Xbra-E probe is necessary for the binding of SIP1_{FS}. One of the reasons for which SIP1_{FS} is unable to bind to the Xbra-E probe might be because the length of the Xbra-E probe is shorter than the length of the Xbra-WT probe. To test this hypothesis, we prepared a probe containing a random sequence upstream of the Xbra-E probe to obtain the same length as the Xbra-WT probe. In contrast to SIP1_{CZF} that binds efficiently to this probe (Fig.2, lane 6), SIP1_{FS} was unable to bind (lane3). This result clearly indicates that the length of the Xbra-E probe was not the reason for which SIP1_{FS} does not bind to this probe. To substantiate that the Xbra-F oligonucleotide also contains sequences necessary for the binding of SIP1_{FS}, we fused that oligonucleotide as well as a random sequence upstream of another CACCT site known to bind strongly AREB6 protein (Xbra-F + AREB6 and Rdm + AREB6, respectively). In lanes 4 and 5, we observed that SIP_{CZF} binds with equal affinities both the Xbra-F + AREB6 and Rdm + AREB6 probes, indicating that the AREB6 sequence

is also recognized by SIP1_{CZF}. However, SIP1_{FS} only binds to the Xbra-F + AREB6 probe (lane 1) and not to the Rdm + AREB6 probe. This confirms that the Xbra-F oligonucleotide contains sequences necessary for the binding of SIP1_{FS}. In addition, the only common denominator between the Xbra-E and the AREB6 probe is the
 5 AGGTG sequence, suggesting that no other sequences than this AGGTG in the Xbra-E probe is necessary for the binding of SIP1_{FS}.

To map the sequences within Xbra-F that, in conjunction with the Xbra-E sequence, are required for the binding of SIP1_{FS}, we prepared a series of probes, identical in length to Xbra-WT, containing adjacent triple mutations within the Xbra-F part (see
 10 Table 1). Only three of these mutated probes (i.e. Xbra-L, Xbra-M and Xbra-N) affected the binding of SIP1_{FS}. Indeed, the upstream CACCT sequence, which is intact in the Xbra-F probe, was modified in the L, M and N probes. We also showed that SIP1_{FS} does not bind to the Xbra-S probe, which contains a point mutation, changing the upstream CACCT into CATCT. This mutation is similar to the
 15 downstream AGATG mutation made within the Xbra-D probe.

The results described above are indicative for SIP1_{FS} contacting both CACCT sequences in the Xbra promoter. To further investigate the importance of these sites, a DNA methylation interference assay was carried out. The methylation of three Gs of the downstream AGGTG (SIP_{DO}) and of the two Gs of the upstream CACCT (SIP_{UP})
 20 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 strongly supports that these residues are essential for SIP1_{FS} binding. It has also been observed that the methylation of one of the 2 Gs localized very close to the SIP_{DO} also interfered with the binding of SIP1_{FS}. Consequently it has thus been shown that for SIP1_{FS} two
 25 CACCT sequences and their integrity are required for DNA binding.

- SIP1 and δ EF1 require 2 CACCT sequences for binding to different potential candidate sites.

SIP1 and δ EF1 have a very similar structure with two very highly conserved zinc finger
 30 clusters and it is likely that these two proteins bind DNA in a similar way. We set out whether also δ EF1 binds to the Xbra2 promoter by contacting both CACCT sequences, which has previously not been reported. Myc-tagged δ EF1 was expressed

in COS cells and the corresponding nuclear extracts were tested in EMSA with WT and a panel of mutated Xbra probes. δ EF1 binds strongly to the Xbra-WT probe that contains both CACCT sites. However, like SIP1_{FS}, δ EF1 binds neither the Xbra-E probe comprising only the downstream CACCT site nor the Xbra-F probe containing only the upstream CACCT site. In addition, the point mutation of either the upstream CACCT (Xbra-S) or the downstream CACCT site (Xbra-D) also abolished the binding of δ EF1. Therefore, like SIP1_{FS}, full length δ EF1 requires also the integrity of both CACCT sequences for binding to the Xbra2 promoter. The fact that two CACCT sites are required for the binding of SIP1_{FS} as well as δ EF1 may be unique for the Xbra2 promoter. Therefore, the next question was to analyze whether two CACCT sequences are also necessary for SIP1/ δ EF1 for binding to other target sites. Putative δ EF1 and SIP1 binding elements are present in several promoters. One putative δ EF1 binding element, indeed containing two intact and spaced CACCT sites, was found within the promoter of the human α 4-integrin gene (23). Interestingly, both sites are contained within of E2 boxes. Mutation of these two CACCT sites led to the de-repression of the α 4-integrin gene expression in myoblasts, suggesting that δ EF1 is a repressor of α 4-integrin gene transcription (23). Since these two CACCT sites are closely positioned in the promoter (spacing is 34 bp), we investigated whether both CACCT sequences are required for the binding of δ EF1. For this purpose, a 60 bp-long probe overlapping both CACCT sites of the α 4-integrin promoter was synthesized (α 4I-WT) as well as two mutated versions, i.e. having a point mutation in either the upstream (α 4I-B) or the downstream CACCT site (α 4I-A), respectively (see Table 1). These probes were tested for binding in EMSAs with COS cell extracts of either δ EF1 or SIP1_{FS} transfected cells. Both δ EF1 as well as SIP1_{FS} form strong complexes with the α 4I-WT probe. The δ EF1 complex was entirely supershifted with an anti-Myc antibody, demonstrating its specificity. Both the binding of SIP1 and of δ EF1 is abolished or strongly affected by a mutation of either the upstream or the downstream CACCT site. Moreover, competition experiments revealed that 50 ng of unlabeled α 4I-WT probe was sufficient to abolish the binding of SIP1 or δ EF1 to the α 4I-WT probe, whereas 50 ng of either unlabeled α 4I-A or α 4I-B probes were not. We conclude that SIP1_{FS} as well as δ EF1 require the integrity of two CACCT sites for

binding to the promoter of the $\alpha 4$ -integrin gene.

We also found two closely positioned CACCT sites within the promoter of the human E-cadherin gene. An oligonucleotide comprising both CACCT sites of this E-cadherin promoter was used as a probe (Ecad-WT) together with SIP1_{FS} or δ EF1 extracts in EMSAs. Both SIP1_{FS} as well as δ EF1 form a complex with this probe. However, when either the upstream (Ecad-A probe) or the downstream (Ecad-B probe) CACCT site was mutated (see Table 1, lower part), the binding of SIP1_{FS} and δ EF1 was abolished. This also suggests that the two CACCT sites in this promoter represent a high affinity site for the binding of two-handed zinc finger/homeodomain transcription factors.

From the alignment of the Xbra-WT, $\alpha 4$ I-WT and Ecad-WT probes (see Table 1) we observed no obvious homology, except for one CACCTG site and a second CACCT site. Our results described above and this alignment, indicates that only those sequences participate in the binding of either SIP1_{FS} or δ EF1. We therefore conclude that for binding to target promoters, SIP1_{FS} or δ EF1 require at least one CACCT site and one CACCTG site.

- Spacing variations and orientation of the CACCT sites

Within the Xbra-WT, $\alpha 4$ I-WT and Ecad-WT probes (Table 1), the spacing between the two CACCT sequences was 24 bp, 34 bp and 44 bp, respectively. Since SIP1_{FS} and δ EF1 bind efficiently to these probes, this shows that these proteins can accommodate spacing between the two CACCT sites ranging from 24 bp to at least 44 bp. To further investigate whether the spacing between the two CACCT sites is an important parameter for binding, we generated different Xbra probes with deletions between these sites. Two mutant probes (Xbra-B and Xbra-C) have a deletion of 3 adenines whereas probe Xbra-U has a deletion of 10 nucleotides (Table 1). These probes were tested in EMSA with cell extracts from COS cells expressing either SIP1_{FS} or δ EF1. Both SIP1_{FS} and δ EF1 bind with equal affinity to the Xbra-WT, Xbra-B, Xbra-C and Xbra-U probes. As already suggested by the results shown for different promoters, this indicates that also within the same promoter element, the spacing between the two CACCT sites is not a critical parameter for the binding of these two transcription factors.

By extensive comparison of the Xbra-WT, $\alpha 4$ I-WT and Ecad-WT probes, we observed

that in the case of the Xbra-WT and $\alpha 4I$ -WT probes, the orientation of the two CACCT sites is CACCT-N-AGGTG, whereas in Ecad-WT the orientation is AGGTG-N-CACCT. Because of the non-palindromic feature of the CACCT site, these two sites could be assumed substantially different. However, SIP1_{FS} and δ EF1 bind to these
5 differentially orientated sites with comparable affinities (see above). This suggests that SIP1_{FS} and δ EF1 can bind irrespective of the orientation of the two CACCT sites.

To further investigate the orientation of the two CACCT sites with respect to the DNA binding capacity of SIP1_{FS} and δ EF1, additional probes were designed. Probe Xbra-EE contains a tandem repeat of the Xbra-E probe, whereas probe Xbra-ErE contains
10 an inverted repeat of the same Xbra-E sequence. In addition, we synthesized Xbra-V, 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 the Xbra-W probe, only the downstream site was replaced by the upstream CACCT sequence. All these probes were again tested in EMSAs with extracts prepared from COS cells
15 expressing either SIP1_{FS} or δ EF1. We observed the strongest binding of SIP1_{FS} or δ EF1 to the Xbra-EE probe. Therefore, SIP1_{FS} and δ EF1 cannot bind to Xbra-E, containing a single CACCT site, but bind strongly when this sequence is duplicated, again indicating the requirement for 2 CACCT sites. In addition, it is evident that the two CACCT sites have to be present on the same DNA fragment and not on two
20 separated strands (see below). SIP1 and δ EF1 bind to Xbra-ErE, also suggesting that the respective orientation of the two CACCT sites is not critical for binding. Furthermore, switching both the upstream and the downstream sites (probe Xbra-V) or replacing only the upstream site by a second copy of the downstream site (probe Xbra-W) did not have an effect on SIP1_{FS} and δ EF1 binding. From these experiments,
25 we conclude that neither the spacing between the two CACCT sites nor the respective orientation of these two sites is critical for the binding of two-handed zinc finger/homeodomain transcription factors *in vitro*.

Surprisingly, not all CACCT duplicated sites can bind these factors. In fact, duplication of the Xbra-F sequence, which in combination with the Xbra-E sequence was shown
30 to be necessary for the binding of SIP1_{FS} and δ EF1, is refractory to binding of SIP1_{FS} and δ EF1. This suggests that the CACCT site within the Xbra-F context is a low affinity site and that sequences adjacent to this CACCT site may optimize the affinity.

In addition, the fact that neither the C-terminal cluster nor the N-terminal cluster can bind independently to the Xbra-F probe confirms the assumption that this site displays low affinity. In contrast, the CACCTG site present in the Xbra-E probe can bind SIP1_{CZF} and SIP1_{NZF}, and a duplication of this element creates a high affinity-binding site for both SIP1_{FS} and full length δ EF1. This suggests that the terminal G base in the downstream site may also allow to discriminate between a high and low affinity-binding site. However, the CACCT site in Xbra-F may only bind one of the zinc finger clusters of SIP1_{FS} once the other cluster has occupied the neighboring high affinity CACCTG site (in Xbra-E). To confirm the importance of this terminal G base residue for the binding of SIP1_{FS} and δ EF1, we mutagenized the downstream CACCTG site to CACCTA (probe Xbra-Z). The binding of SIP1_{FS} or δ EF1 to the Xbra-Z probe was strongly decreased (compared with the Xbra-WT probe) suggesting that this G-base residue is important for the generation of a high affinity binding site for both SIP1_{FS} and δ EF1.

Finally, when Xbra-E and Xbra-F probes are mixed prior to addition of SIP1_{FS} or δ EF1, we do not observe any binding, again indicating that both CACCT sites have to be in the *cis* configuration, i.e. on the same DNA.

- The two zinc finger clusters of SIP1 are required and must be intact for binding to DNA

SIP1 and δ EF1 bind to DNA elements containing two CACCT sites and both of these proteins contain two clusters of zinc fingers capable of binding independently to CACCT sites. In subsequent work, we wanted to evaluate the importance of each zinc finger cluster for the binding of SIP1_{FS} to DNA. Mutations destroying either the third or the fourth zinc finger of the N-terminal cluster of δ EF1_{NZF} were shown to abolish the binding of this cluster to the DNA. Similarly, mutagenesis of the second or the third zinc finger in the C-terminal cluster also abolished the binding of δ EF1_{CZF} to CACCT (10). Therefore, we introduced in the SIP1_{NZF} and SIP1_{CZF} clusters mutations similar to those in δ EF1. These mutated and wild type clusters were fused to GST and the fusions proteins were purified from bacteria. We demonstrate that both wild type SIP1_{NZF} and SIP1_{CZF} strongly bind to the Xbra-E probe. However, with the same amount of purified mutant cluster/GST fusion proteins (GST-NZF3, GST-NZF4, GST-

CZF2 and GST-CZF3), no binding to the Xbra-E probe could be detected with any of these fusion proteins. Indeed, these mutations also abolish the capacity of each cluster (SIP1_{NZF} and SIP1_{CZF}) to bind independently to a CACCT site.

Then, we introduced similar mutations in full size SIP1 (NZF3-Mut, NZF4-Mut, CZF2-Mut and CZF3-Mut), and over-expressed these SIP1 mutants in COS cell as Myc-tagged proteins. The expression of the different mutants was established and normalized by Western blot analysis using anti-Myc antibody. By means of EMSAs, we observed that WT SIP1 binds strongly to the Xbra-WT probe, and that the SIP1-complex is super-shifted upon incubation with an anti-Myc antibody. In contrast, none of the mutant forms of full size SIP1 was able to form a SIP1-like complex or a SIP1 super-shifted complex. The same observations were made when the α 4-WT probe was used as a probe. In conclusion, full size SIP1 requires the binding capacities of both intact zinc fingers clusters to bind to its target, which necessarily contains 2 CACCT sites. The effect of these mutations on the repressor activity of SIP1 was tested in a transfection assay together using 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 human collagenase promoter (de Groot and Kruijer, 1990). SIP1 bound to a fragment containing this multimerized element, but neither NZF3-Mut nor CZF3-Mut was able to bind. Over-expression of SIP1 in CHO cells leads to a strong repression of the p3TP-Lux basal transcriptional activity. However, the repression was 6 to 7-fold lower upon overexpression of SIP1 mutants defective in DNA binding (NZF3-Mut or CZF3-Mut). Therefore 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 SIP1 zinc fingers clusters is required for its binding to two CACCT sequences, prompted us to test whether SIP1 binds as a monomer, in which each zinc finger cluster contacts one CACCT site. However, it can be hypothesized also that SIP1 binds to its target sites as a dimer. This may imply that one of the SIP1 proteins of the dimer would bind one CACCT site via its N-terminal zinc finger cluster, while the second SIP1 molecule would contact the DNA via its C-

terminal zinc finger cluster. Consequently, certain combinations of NZF and CZF mutants in a full size SIP1 context (see above) should generate a dimeric configuration that binds DNA. As shown already, in none of the combinations of NZF with CZF mutations tested, binding to the Xbra-WT probe could be detected. Although
5 we cannot rule out that these mutations also would affect dimer formation, it is highly unlikely that the same mutation affects both the DNA binding capacity as well as the monomer-monomer interaction. Moreover, it is highly unlikely that two different mutants, i.e. different mutations within a cluster, would behave identical. Therefore, we considered that SIP1 does not bind to DNA as a dimer. The observation that the
10 integrity of both zinc fingers clusters is required for SIP1 binding to two CACCT sequences, suggests that SIP1 binds as a monomer, in which each zinc finger cluster contacts one CACCT site. However, it can be hypothesized that SIP1 binds its target sites as a dimer. This would imply that one of the SIP1 molecules of the dimer would bind one CACCT site via its N-terminal zinc finger cluster, while the second SIP1
15 molecule would contact the DNA via its C-terminal zinc finger cluster. Since both zinc finger clusters are necessary for binding, the zinc finger cluster not interacting with the DNA would then be involved in dimerization. Consequently, some combinations of NZF and CZF mutants (see above) should generate a dimer configuration that binds DNA. In none of the combinations of NZF and CZF mutations binding to the Xbra-WT
20 probe could be detected. Although we cannot rule out that these mutations also affect potential dimer formation, it is highly unlikely that the same mutation affects both the DNA-binding capacity as well as the protein-protein interaction. Moreover, it is highly unlikely that two different mutants, ie have different mutations within a cluster, would behave the same. These observations indicate that SIP1 does not bind DNA as a
25 dimer. observation that the integrity of both zinc fingers clusters is required for SIP1 binding to two CACCT sequences, suggests that SIP1 binds as a monomer, in which each zinc finger cluster contacts one CACCT site. However, it can be hypothesized that SIP1 binds its target sites as a dimer. This would imply that one of the SIP1 molecules of the dimer would bind one CACCT site via its N-terminal zinc finger
30 cluster, while the second SIP1 molecule would contact the DNA via its C-terminal zinc finger cluster. Since both zinc finger cluster are necessary for binding, the zinc finger cluster not interacting with the DNA would then be involved in dimerization.

Consequently, some combinations of NZF and CZF mutants (see above) should generate a dimer configuration that binds DNA. In none of the combinations of NZF and CZF mutations binding to the Xbra-WT probe could be detected. Although we cannot rule out that these mutations also affect potential dimer formation, it is highly unlikely that the same mutation affects both the DNA-binding capacity as well as the protein-protein interaction. Moreover, it is highly unlikely that two different mutants, ie have different mutations within a cluster, would behave the same. These observations indicate that SIP1 does not bind DNA as a dimer. observation that the integrity of both zinc fingers clusters is required for SIP1 binding to two CACCT sequences, suggests that SIP1 binds as a monomer, in which each zinc finger cluster contacts one CACCT site. However, it can be hypothesized that SIP1 binds its target sites as a dimer. This would imply that one of the SIP1 molecules of the dimer would bind one CACCT site via its N-terminal zinc finger cluster, while the second SIP1 molecule would contact the DNA via its C-terminal zinc finger cluster. Since both zinc finger cluster are necessary for binding, the zinc finger cluster not interacting with the DNA would then be involved in dimerization. Consequently, some combinations of NZF and CZF mutants (see above) should generate a dimer configuration that binds DNA. As shown in Figure 5A, in none of the combinations of NZF and CZF mutations binding to the Xbra-WT probe could be detected. Although we cannot rule out that these mutations also affect potential dimer formation, it is highly unlikely that the same mutation affects both the DNA-binding capacity as well as the protein-protein interaction. Moreover, it is highly unlikely that two different mutants, ie have different mutations within a cluster, would behave the same. These observations indicate that SIP1 does not bind DNA as a dimer. observation that the integrity of both zinc fingers clusters is required for SIP1 binding to two CACCT sequences, suggests that SIP1 binds as a monomer, in which each zinc finger cluster contacts one CACCT site. However, it can be hypothesized that SIP1 binds its target sites as a dimer. This would imply that one of the SIP1 molecules of the dimer would bind one CACCT site via its N-terminal zinc finger cluster, while the second SIP1 molecule would contact the DNA via its C-terminal zinc finger cluster. Since both zinc finger cluster are necessary for binding, the zinc finger cluster not interacting with the DNA would then be involved in dimerization. Consequently, some combinations of NZF and CZF mutants (see above)

should generate a dimer configuration that binds DNA. As shown in Figure 5A, in none of the combinations of NZF and CZF mutations binding to the Xbra-WT probe could be detected. Although we cannot rule out that these mutations also affect potential dimer formation, it is highly unlikely that the same mutation affects both the DNA-binding capacity as well as the protein-protein interaction. Moreover, it is highly unlikely that two different mutants, ie have different mutations within a cluster, would behave the same. These observations indicate that SIP1 does not bind DNA as a dimer. observation that the integrity of both zinc fingers clusters is required for SIP1 binding to two CACCT sequences, suggests that SIP1 binds as a monomer, in which each zinc finger cluster contacts one CACCT site. However, it can be hypothesized that SIP1 binds its target sites as a dimer. This would imply that one of the SIP1 molecules of the dimer would bind one CACCT site via its N-terminal zinc finger cluster, while the second SIP1 molecule would contact the DNA via its C-terminal zinc finger cluster. Since both zinc finger cluster are necessary for binding, the zinc finger cluster not interacting with the DNA would then be involved in dimerization. Consequently, some combinations of NZF and CZF mutants (see above) should generate a dimer configuration that binds DNA. As shown in Figure 5A, in none of the combinations of NZF and CZF mutations binding to the Xbra-WT probe could be detected. Although we cannot rule out that these mutations also affect potential dimer formation, it is highly unlikely that the same mutation affects both the DNA-binding capacity as well as the protein-protein interaction. Moreover, it is highly unlikely that two different mutants, ie have different mutations within a cluster, would behave the same. These observations indicate that SIP1 does not bind DNA as a dimer.

To address this experimentally, we used a combination of differently tagged SIP1 with supershift experiments in EMSAs. First, we produced Myc-tagged and/or FLAG-tagged SIP1_{FS} separately at comparable levels in COS cells, and confirmed that both proteins bind to DNA with similar affinities. The SIP1 complex generated with Myc-tagged SIP1 has a slightly slower migration than the FLAG-tagged complex (the Myc-tag is longer than the FLAG-tag). Extracts prepared from COS cells expressing similar amounts of both Myc-tagged and FLAG-tagged SIP1 were incubated with the Xbra-WT probe and used in EMSAs. We observed the formation of a broad SIP1 complex which is a combination of both the fast migrating FLAG-tagged SIP1 complex with the

slow migrating Myc-tagged SIP1 complex. Using an anti-FLAG antibody, only the lower part of the complex corresponding to FLAG-tagged SIP1 is super-shifted, whereas about 50 % of the radioactivity remains within the Myc-tagged SIP1 complex. This indicates that the latter SIP1 complex is not super-shifted with the anti-FLAG antibody. Conversely, incubating the extract with an anti-Myc antibody super-shifted only the lower part of the complex corresponding to Myc-tagged SIP1 whereas 50% of the radioactivity is retained within the FLAG-tagged SIP1 complex. Again, this indicates that no FLAG-tagged SIP1 is super-shifted with an anti-Myc antibody. Using both antibodies, we observed the same two super-shifted bands, which correspond to the Myc-tagged and the FLAG-tagged super-shifted complex, in the upper part of the gel. If SIP1 dimers would be formed, then at least some heterodimers would be assembled from Myc-tagged SIP1 and FLAG-tagged SIP1. However, no other super-shifted band that would correspond to a potential double super-shift, viz. super-shifted with both anti-Myc- and anti-FLAG-antibodies, is detectable. Hence, this experiment gave no detectable dimer formation between FLAG-tagged SIP1 and Myc-tagged SIP1.

Finally, FLAG-tagged SIP1 in a COS cell extract was immunoprecipitated in the presence of a large excess of DNA binding sites. However, co-immunoprecipitation of Myc-tagged SIP1 was not feasible. The reciprocal experiment, i.e. immunoprecipitating with an anti-Myc antibody and detection with an anti-FLAG antibody, did not show any SIP1 dimer either. Taken together, these observations let us to conclude that SIP1 binds as a monomer to the Xbra-WT probe.

- 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-long Xbra2 promoter fragment was shown 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, for which a regulatory element may be lacking in the reporter cassette tested here.

- 5 A single point mutation within the downstream CACCT site in the promoter, which disrupted SIP1 binding (*Xbra2*-Mut1) 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 ectoderm layer. Mutations within the upstream CACCT sequence (*Xbra2*-Mut4) also affected the SIP1 binding: we observed in all transgenic
- 10 embryos (n>30) the same ectopic expression as for the *Xbra2*-Mut1 mutation. Mutation of the downstream CACCTG to CACCTA (Xbra2-Mut2) also affects SIP1 binding to such probe. This mutation when introduced into the *Xbra2* 2.1kb promoter also led to ectopic expression of GFP mRNA in all transgenic embryos tested (n>30). We also tested a mutation (*Xbra2*-Mut3) that decreased by 3 bp the original 24 bp-
- 15 spacing between the two CACCT sequences. This mutation weakened the interaction of such probe with SIP1. This was also reflected in the corresponding transgene embryos (n=37): while 35% of the embryos showed the same expression pattern as the wild type *Xbra2* 2.1kb promoter fragment, 65% had either patches or weak continuous expression in the inner ectoderm layer.
- 20 A nice correlation between the effect of these mutations on SIP1 binding affinity in 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 an hitherto unknown *Xenopus* SIP1-like repressor regulates
- 25 *Xbra2* gene expression *in vivo*. In addition, it confirms that SIP1-like factors require two intact CACCT sites for regulating target promoters like *Xbra2*.

2. SIP1 induces invasion by downregulation of E-cadherin

Results

- SIP1 binding represses E-cadherin promoter activity through binding on two conserved E-boxes.**

To elucidate whether SIP1 binding affects the transcriptional activity of the human E-cadherin promoter (-308/+41), we transiently co-expressed full-length SIP1 with E-cadherin promoter driven reporter constructs in the E-cadherin positive cell lines NMe (mouse), MDCK (dog) and MCF7/AZ (human). SIP1 expression led to an 80 % decrease of the human E-cadherin promoter activity. To address the binding specificity of SIP1 for the 2 conserved E-boxes, mutagenesis in either the upstream E-box1 (-75) or downstream E-box3 (-25) or simultaneously in both E-boxes was performed. When cotransfection was performed with SIP1 cDNA and the mutant E-cadherin promoter constructs (68), a de-repression of the human E-cadherin promoter activity was consistently shown. In addition, mutated SIP1 constructs, were cotransfected with the human E-cadherin promoter. Mutation of the N-terminal or C-terminal zinc finger clusters resulted only in a slight derepression of the E-cadherin promoter activity. Interestingly, cotransfection of the human E-cadherin promoter and a SIP1 double mutant, affected in both zinc finger clusters, resulted in a considerable loss of SIP1 mediated repression of E-cadherin promoter activity. We can therefore conclude that SIP1 represses the E-cadherin promoter activity by binding to the 2 E-boxes and that the 2 zinc finger clusters are indeed needed for full repression of the E-cadherin promoter activity.

- Inducible expression of SIP1 results in dose-dependent loss of E-cadherin protein and mRNA.**

To elucidate whether SIP1 affects the endogenous E-cadherin expression levels, E-cadherin positive MDCK-Tetoff cells, with high expression of the tTA transactivator was stably transfected with a plasmid expressing a Myc₆-tagged full-length mouse SIP1 cDNA under control of a responsive tTA element. To induce SIP1, cells were grown without tetracycline for 3 days. Analysis of E-cadherin and SIP1 expression by

immunofluorescence of a representative cloned transfectant revealed induced SIP1 in the nucleus, concomitant with total loss of the typical honeycomb E-cadherin expression pattern at cell-cell contacts. Western blot analysis confirmed these results. SIP1 induction occurred at tetracycline concentration equal or lower than 2g/ml. As
5 the tetracycline concentration was gradually decreased, E-cadherin was more strongly repressed and this correlated inversely with SIP1 accumulation. Further, we checked if catenins, linking E-cadherin to the actin cytoskeleton, were influenced by SIP1 expression. Upon a Western blotting neither α E-catenin nor β -catenin appeared to be affected, and this was confirmed by immunofluorescence. Equal amounts of total RNA
10 of both non-induced and induced cells were analysed by Northern blotting. After hybridisation with an E-cadherin-specific probe, the SIP1 expressing cells showed almost no E-cadherin mRNA expression, whereas the non-induced cells (+tet) expressed normal amounts of E-cadherin mRNA. These results validate those of the reporter assays as induction of SIP1 expression affects endogenous E-cadherin
15 expression through mRNA downregulation.

- SIP1 expression in human carcinoma cell lines.

To examine the expression of SIP1 in a panel of E-cadherin-negative and -positive cell lines, Northern blot analyses were performed. To avoid possible cross-
20 hybridizations to other members of the δ EF1 family, appropriate mouse and human SIP1 cDNA fragments were used as probes. A clearcut strong inverse correlation between SIP1 expression and E-cadherin expression was noticed. High expression of SIP1 was found in human fibroblasts and the most prevalent expression of SIP1 was found in E-cadherin-negative carcinoma cells, reported to have a methylated E-
25 cadherin promoter (53). As the expression level of SIP1 in the described cell lines is in common with Snail mRNA expression in E-cadherin negative cell lines (66), we looked for Snail expression levels in our conditional SIP1 expressing cell line MDCK-Tetoff-SIP1. Snail expression could not be detected after SIP1 induction. E-cadherin repression is in our cell system not Snail related.

- **SIP1 enhances the malignant phenotype by promoting loss of cell cell adhesion and invasion.**

As E-cadherin is a well known invasion-suppressor molecule (47), we addressed the question whether SIP1 induction switches the cells to a more invasive phenotype. A cell aggregation assay was performed of non-induced versus induced MDCK-Tetoff-SIP1 cells. The non-induced MDCK-Tetoff-SIP1 cells showed significant aggregation after 30 min, but SIP1 induction abrogated normal cell-cell aggregation to a similar extent as an E-cadherin blocking antibody DECMA-1. Invasion into collagen type-I gels was induced by SIP1 as efficiently as by the DECMA-1 antibody.

- **SIP1-expression results in the reduction of uni-directional cell migration.**

The role of E-cadherin on cell migration was demonstrated by using a blocking E-cadherin with a specific antibody which results in a reduction of uni-directional cell migration (72). The effect of SIP1 expression on different cell migration due to downregulation of E-cadherin was studied in a wound assay in the inducible MDCK-Tetoff SIP1 expressing cell line. We could demonstrate that induction of SIP1 results in a lower uni-directional cell migration. Downregulation of E-cadherin mediated cell-cell contact results in the disturbance of uni-directional migration.

Discussion

Invasion and metastasis are the most crucial steps in tumour progression. Malignancy of carcinoma cells is characterized by loss of both cell-cell adhesion and cellular differentiation and this has been frequently reported to correlate negatively with E-cadherine downregulation. Loss of E-cadherin expression has been attributed to transcriptional dysregulation (52, 73). We show here for the zinc finger protein SIP1 that it represses the E-cadherin expression at the transcriptional level by binding to the conserved E-boxes present in the minimal E-cadherin promoter. The specific binding of SIP1 on the two E-boxes was confirmed by mutagenesis of either the zinc finger clusters of SIP1 or the E-box sequences in the E-cadherin promoter. Indeed, such mutations resulted in the loss of repression of the E-cadherin promoter activity by SIP1. These results are compatible with the finding that comparable mutations of the E-boxes resulted in the upregulation of the E-cadherin promoter activity in E-

cadherin-negative cell lines, where the wild type promoter shows low activity (56, 58). Stable transfection of the transcriptional repressor SIP1 induces downregulation of E-cadherin at both mRNA and protein level. A wound assay demonstrates that SIP1 interferes with the unidirectional migration mediated by a functional E-cadherin cell-cell contact. Weaker cell-cell contact results in more multi-directional migration of the epithelial cells. A striking correlation between downregulated E-cadherin and upregulated SIP1 expression was seen in various human tumour cells. Finally, we demonstrate here that the downregulation of E-cadherin due to SIP1 expression is also associated with a remarkable increase of the invasion capacity. Hence, SIP1 can be considered as an invasion-inducer due to its binding to the E-cadherin promoter. The fact that the transcriptional repressor Snail also specifically binds E-boxes resulting in transcriptional E-cadherin repression (66, 67) raised the question whether the E-cadherin repression in our studies is Snail-mediated. Snail mRNA upregulation could not be detected in the conditional SIP1 expressing MDCK-Tetoff-SIP1 cell line. These data let us to consider SIP1 as the effector of transcriptional E-cadherin repression in our cell system. This idea was supported by the fact that mutations of the E-boxes have a more extensive effect on the decrease of repression of the E-cadherin promoter when cotransfected with SIP1. Derepression of the E-cadherin promoter activity, when cotransfected with SIP1, is already detected with a single E-box mutation. For Snail cotransfection a clear derepression effect was only seen when more E-boxes were mutated in the human E-cadherin promoter (66). The high expression of SIP1 in the breast cancer cell lines MDA-MB435S and MDA-MB231 is remarkable. These tumour cell lines have been described to bear a hypermethylated E-cadherin promoter (53). However, this should not rule out an important role for SIP1 repression of the endogenous E-cadherin promoter. Mutations of the E-boxes reactivate the exogenous E-cadherin promoter activity strongly in these cell lines. Indeed, recent research made clear that many transcription factors function by recruiting multiprotein complexes with chromatin modifying activities to specific sites on DNA (74). It was already shown that another Smad-interacting transcription factor TGIF associates with histone deacetylase (75). DNA methylation and chromatin condensation could therefore act synergistically with histone deacetylation to repress gene transcription (76).

Materials and methods

Cell Culture and reagents

The MDCK-Tetoff cell line was obtained from Clontech (Palo Alto, USA). This cell line is derived from the Madin Darby Canine Kidney (MDCK) Type II epithelial cell line and stably expresses the Tet-off transactivator, tTA (77). MCF7/AZ cell line is a cell line derived from MCF7, a human mammary carcinoma cell line (78). The NMe cell line is an E-cadherin expressing subclone of NMuMG, an epithelial cell line from normal mouse mammary gland (47). MDA-MB231 is a human breast cancer cell line (ATCC, Manassas, VA).

Plasmids

The full-size mouse SIP1 cDNA sequence was cloned into the Myc-tag containing pCS3 eukaryotic expressing vector derived from pCS2 (69). The resulting plasmid was designated pCS3-SIP1FS. Mutagenesis of the zinc finger clusters of the SIP1 is described by Remacle et al. (68). For the construction of the inducible vector pUHD10.3SIP1, an *Cla*I/*Xba*I fragment from pCS3SIP1FS was cloned into the *Eco*RI/*Xba*I-cut pUHD10.3 vector (79). The *Cla*I site of SIP1 fragment and the *Eco*RI site of the vector was blunted using Pfu polymerase (Stratagene; La Jolla, CA). The E-cadherin promoter sequence (-341/+41) was obtained by PCR on genomic DNA from the human MCF7/AZ cell line. PCR-primers used are: 5'-ACAAAAGAACTCAGCCAAGTG-3' and 5'-CCGCAAGCTCACAGGTGC-3'. The GC-melt kit (Clontech; Palo Alto, CA) was used for efficient amplification. The PCR product was blunted, kinased and then cloned into the pGL3basic vector (Promega; Madison, WA) which was opened at the *Srf*I site. By using the *Kpn*I-*Hind*III sites in this luciferase reporter construct the E-cadherin promoter was also transferred to the pGL3enhancer vector. Mutagenesis of the E-boxes in the human E-cadherin promoter was performed by the QuickChange Site-Directed Mutagenesis Kit (Stratagene) by using following primers: forward primer E-box1: 5'-gctgtggccggCAGATGaaccctcag-3'; reverse primer E-box1 : 5'-ctgagggttCATCTGccggccacagc-3'; forward primer E-box3 : 5'-gctccgggctCATCTGgctgcagc-3'; reverse primer E-box3 : 5'-gctgcagcCAGATGagccccggagc-3'

Stable transfection of cells

For stable transfection of the MDCK-Tetoff cell line the LipofectAMINE PLUS™ (Gibco BRL, Rochville, USA) method was used. Two thousands cells were grown on a 75 cm² falcon for 24 h and then transfected with 30 µg of pUHD10.3-SIP1 plasmid plus 3 µg pPHT plasmid. The latter is a derivative of pPNT and confers resistance to hygromycin (80). Stable MDCK-Tetoff transfectants, MDCK-Tetoff-SIP1, were selected by hygromycin-B (150 units/ml) (Duchefa Biochemie, Haarlem, The Netherlands) for a period of 2 weeks. Induction of SIP1 was prevented by adding tetracycline (1 µg/µl) (Sigma Chemicals, USA). Expression of SIP1 was done by washing away tetracycline at the time of subcloning. Stable clones with reliable induction properties were identified by immunofluoresence using anti-Myc tag antibodies.

Promoter reporter assays

MCF7/AZ cells were transiently transfected by using FuGENE 6 (Roche; Basel, Switzerland). NMe and MDA-MB231 were transfected with the LIPOFECTAMINE (Gibco BRL; Rochville, USA) procedure and the parental MCDK cell line was transiently transfected with LIPOFECTAMINEPLUS (Gibco BRL; Rochville, USA). For transient transfection about 200.000 cells were seeded per 10-cm² well. After incubation for 24 h, 600 ng of each plasmid type DNA was transfected. Medium was refreshed 24 h after transfection. Cells were lysed after 3 days in lysis solution of the Galacto-Star™ kit (Tropix, Bedford, MA) Normalisation of transfection was done by measurement of β-galactosidase, encoded by the cotransfected pUT651 plasmid (Eurogentec; Seraing, Belgium). Luciferase substrate is added to each sample. For β-galactosidase detection, a chemiluminescent substrate is supplied (Tropix, Bedford, MA). Luciferase and β-galactosidase activity was assayed in a Topcount microplate scintillation reader (Packard Instrument Co., Meriden, CT).

Northern analysis

Total RNA was isolated with the RNeasy kit (Qiagen; Chatsworth, CA) following the manufacturer's protocol. Total RNA (25 µg) was glyoxylated, size-fractionated on a 1% agarose gel and transferred onto a Hybond-N⁺ membrane (Amersham Pharmacia Biotech, Rainham, UK). Hybridizations were performed as described before (81). The mouse SIP1 probe (459 bp) was generated by a *EcoR*-I digest of the mouse SIP1

cDNA. The human SIP1 probe (707 bp) was created by a *Bst* EII-*Not*I digest on the Kiaa 0569 clone (Kazusa DNA Research Institute). The mouse E-cadherin probe used was a *Sac*I fragment (500 bp) of the mouse E-cadherin cDNA. Two degenerated primers: 5' CTTCCAGCAGCCCTACGAYCARGCNCA 3'; 5' GGGTGTGGGACCGGATRTGCATYTTNAT 3' were used to amplify a fragment of the dog Snail cDNA from a total cDNA population of the MDCK cell line. Cloning and sequencing of the amplified band revealed a 432 bp cDNA fragment. To control the amount of loaded RNA, a GAPDH probe was used on the same blot. The quantification of the radioactive bands was performed by a PhosphorImager 425 (BioRad, Richmond, CA).

Immunofluorescence assays and Antibodies .

Cells of interest were grown on glass coverslips. Fixation was by standard procedures (82). The following antibodies were used: the rat monoclonal antibody DECMA-1 (Sigma; Irvine, UK) recognising both mouse and dog E-cadherin, and the mouse anti-Myc tag antibody (Oncogene, Cambridge, MA). Secondary antibodies used were Alexa 488-coupled anti-rat Ig and Alexa 594-coupled anti-mouse Ig.

Cell Aggregation Assay

Single-cell suspensions were prepared in accordance with an E-cadherin-saving procedure (83). Cells were incubated in an isotonic buffer containing 1.25 mM Ca^{2+} under gyrotory shaking (New Brunswick Scientific, New Brunswick, NJ) at 80 rpm for 30 min. Particle diameters were measured in a Coulter particle size counter LS200 (Coulter, Lake Placid, NY) at the start (N_0) and after 30 min of incubation (N_{30}) and plotted against percentage volume distribution.

Collagen Invasion Assay

Six-well plates were filled with 1.25 ml of neutralized type I collagen (Upstate Biotechnology, Lake Placid, NY) per well. Incubation for at least 1 h at 37°C was needed for gelification. Single-cell suspensions were seeded on top of the collagen gel and cultures were incubated at 37°C for 24 h. Using an inverted microscope controlled by a computer program, the invasive and superficial cells were counted in 12 fields of 0.157 mm². The invasion index expresses the percentage of cells invading the gel over the total numbers of cells (84).

Wound Assay

Wound assay was performed as described before (85). Briefly, wounded mono-layers were cultured for 24 h in serum-deprived medium in the presence or absence of tetracycline. Cell migration was assessed by measuring the distance of the wound. Migration results are expressed as the average of the wound-distance.

5

10

15

20

25

30

REFERENCES

1. **Arora, K., H. Dai, S. G. Kazuko, J. Jamal, O. C. MB, A. Letsou, and R. Warrior.** 1995. The *Drosophila* schnurri gene acts in the Dpp/TGF beta signaling pathway and encodes a transcription factor homologous to the human MBP family. *Cell* **81**:781-90.
2. **Bussemakers, M. J., L. A. Girolodi, A. van Bokhoven, and J. A. Schalken.** 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**:1284-90.
3. **Fan, C. M., and T. Maniatis.** 1990. A DNA-binding protein containing two widely separated zinc finger motifs that recognize the same DNA sequence. *Genes Dev* **4**:29-42.
4. **Fortini, M. E., Z. C. Lai, and G. M. Rubin.** 1991. The *Drosophila* zfh-1 and zfh-2 genes encode novel proteins containing both zinc-finger and homeodomain motifs. *Mech Dev* **34**:113-22.
5. **Funahashi, J., R. Sekido, K. Murai, Y. Kamachi, and H. Kondoh.** 1993. Delta-crystallin enhancer binding protein delta EF1 is a zinc finger- homeodomain protein implicated in postgastrulation embryogenesis. *Development* **119**:433-46.
6. **Grieder, N. C., D. Nellen, R. Burke, K. Basler, and M. Affolter.** 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.
7. **Henderson, L. E., T. D. Copeland, R. C. Sowder, G. W. Smythers, and S. Oroszlan.** 1981. Primary structure of the low molecular weight nucleic acid-binding proteins of murine leukemia viruses. *J Biol Chem* **256**:8400-6.
8. **Hendrickson, W., and R. Schleif.** 1985. A dimer of AraC protein contacts three adjacent major groove regions of the aral DNA site. *Proc Natl Acad Sci U S A* **82**:3129-33.
9. **Holmberg, S., and P. Schjerling.** 1996. Cha4p of *Saccharomyces cerevisiae* activates transcription via serine/threonine response elements. *Genetics* **144**:467-78.
10. **Ikeda, K., and K. Kawakami.** 1995. DNA binding through distinct domains of zinc-finger-homeodomain protein AREB6 has different effects on gene transcription.

Eur J Biochem **233**:73-82.

11. **Jiang, Y., V. C. Yu, F. Buchholz, O. C. S, S. J. Rhodes, C. Candeloro, Y. R. Xia, A. J. Lusic, and M. G. Rosenfeld.** 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**:10723-30.
12. **Kim, J. G., and L. D. Hudson.** 1992. Novel member of the zinc finger superfamily: A C2-HC finger that recognizes a glia-specific gene. Mol Cell Biol **12**:5632-9.
13. **Kretzschmar, M., and J. Massague.** 1998. SMADs: mediators and regulators of TGF-beta signaling. Curr Opin Genet Dev **8**:103-11.
14. **Kuhnlein, R. P., G. Frommer, M. Friedrich, M. Gonzalez-Gaitan, A. Weber, J. F. Wagner-Bernholz, W. J. Gehring, H. Jackle, and R. Schuh.** 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**:168-79.
15. **Kurokawa, M., K. Mitani, K. Irie, T. Matsuyama, T. Takahashi, S. Chiba, Y. Yazaki, K. Matsumoto, and H. Hirai.** 1998. The oncoprotein Evi-1 represses TGF-beta signalling by inhibiting Smad3. Nature **394**:92-6.
16. **Latinkic, B. V., M. Umbhauer, K. A. Neal, W. Lerchner, J. C. Smith, and V. Cunliffe.** 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 [published erratum appears in Genes Dev 1998 Apr 15;12(8):1240]. Genes Dev **11**:3265-76.
17. **Lerchner, W., J. E. Remale, D. Huylebroeck, and J. C. Smith.** Unpublished observations .
18. **Maxam, A. M., and W. Gilbert.** 1980. Sequencing end-labeled DNA with base-specific chemical cleavages. Methods Enzymol **65**:499-560.
19. **Miller, J., A. D. McLachlan, and A. Klug.** 1985. Repetitive zinc-binding domains in the protein transcription factor IIIA from Xenopus oocytes. Embo J **4**:1609-14.
20. **Morishita, K., K. Suzukawa, T. Taki, J. N. Ihle, and J. Yokota.** 1995. EVI-1 zinc finger protein works as a transcriptional activator via binding to a consensus

sequence of GACAAGATAAGATAAN1-28 CTCATCTTC. *Oncogene* **10**:1961-7.

21. **Mount, S. M., and G. M. Rubin.** 1985. Complete nucleotide sequence of the *Drosophila* transposable element copia: homology between copia and retroviral proteins. *Mol Cell Biol* **5**:1630-8.

22. **Nucifora, G.** 1997. The EVI1 gene in myeloid leukemia. *Leukemia* **11**:2022-31.

23. **Postigo, A. A., and D. C. Dean.** 1997. ZEB, a vertebrate homolog of *Drosophila* Zfh-1, is a negative regulator of muscle differentiation. *Embo J* **16**:3935-43.

24. **Rajavashisth, T. B., A. K. Taylor, A. Andalibi, K. L. Svenson, and A. J.**

Lusis. 1989. Identification of a zinc finger protein that binds to the sterol regulatory element. *Science* **245**:640-3.

25. **Ray, D., R. Bosselut, J. Ghysdael, M. G. Mattei, A. Tavitian, and F. Moreau-Gachelin.** 1992. Characterization of Spi-B, a transcription factor related to the putative oncoprotein Spi-1/PU.1. *Mol Cell Biol* **12**:4297-304.

26. **Rosen, G. D., J. L. Barks, M. F. Iademarco, R. J. Fisher, and D. C. Dean.** 1994. An intricate arrangement of binding sites for the Ets family of transcription factors regulates activity of the alpha 4 integrin gene promoter. *J Biol Chem* **269**:15652-60.

27. **Rupp, R. A., L. Snider, and H. Weintraub.** 1994. *Xenopus* embryos regulate the nuclear localization of XMyoD. *Genes Dev* **8**:1311-23.

28. **Schwabe, J. W., and D. Rhodes.** 1991. Beyond zinc fingers: steroid hormone receptors have a novel structural motif for DNA recognition. *Trends Biochem Sci* **16**:291-6.

29. **Seeler, J. S., C. Muchardt, A. Suessle, and R. B. Gaynor.** 1994.

Transcription factor PRDII-BF1 activates human immunodeficiency virus type 1 gene expression. *J Virol* **68**:1002-9.

30. **Sekido, R., K. Murai, J. Funahashi, Y. Kamachi, A. Fujisawa-Sehara, Y. Nabeshima, and H. Kondoh.** 1994. The delta-crystallin enhancer-binding protein delta EF1 is a repressor of E2-box-mediated gene activation. *Mol Cell Biol* **14**:5692-700.

31. **Sekido, R., K. Murai, Y. Kamachi, and H. Kondoh.** 1997. Two mechanisms in the action of repressor deltaEF1: binding site competition with an activator and active

repression. *Genes Cells* **2**:771-83.

32. **Todd, R. B., and A. Andrianopoulos.** 1997. Evolution of a fungal regulatory gene family: the Zn(II)₂Cys₆ binuclear cluster DNA binding motif. *Fungal Genet Biol* **21**:388-405.

5 33. **van 't Veer, L. J., P. M. Lutz, K. J. Isselbacher, and R. Bernards.** 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 U S A* **89**:8971-5.

34. **Verschueren, K., J. E. Remacle, C. Collart, H. Kraft, B. S. Baker, P.**
10 **Tylzanowski, L. Nelles, G. Wuytens, M. T. Su, R. Bodmer, J. Smith, and D. Huylebroeck.** SIP1, a novel zinc finger/homeodomain repressor, interacts with Smad proteins and binds to 5'-CACCT sequences in candidate target genes. *J.Biol.Chem* (1999).

35. **Watanabe, Y., K. Kawakami, Y. Hirayama, and K. Nagano.** 1993.
15 Transcription factors positively and negatively regulating the Na,K-ATPase alpha 1 subunit gene. *J Biochem (Tokyo)* **114**:849-55.

36. **Yee, K. S., and V. C. Yu.** 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**:5366-74.

20 37. **Brent, R. and Ptashne, M.** (1985). A eukaryotic transcriptional activator bearing the DNA specificity of a prokaryotic repressor. *Cell* **43**, 729-736.

38. **Chien, C.T., Bartel, P.L., Sternglanz, R., and Fields, S.** (1991). The two-hybrid system: a method to identify and clone genes for proteins that interact with a protein of interest. *Proc.Natl.Acad.Sci.U.S.A.* **88**, 9578-9582.

25 39. **Durfee, T., Becherer, K., Chen, P.L., Yeh, S.H., Yang, Y., Kilburn, A.E., Lee, W.H., and Elledge, S.J.** (1993). The retinoblastoma protein associates with the protein phosphatase type 1 catalytic subunit. *Genes Dev.* **7**, 555-569.

40. **Gyuris, J., Golemis, E., Chertkov, H., and Brent, R.** (1993). Cdi1, a human G1 and S phase protein phosphatase that associates with Cdk2. *Cell* **75**, 791-803.

30 41. **Silver, P.A., Brent, R., and Ptashne, M.** (1986). DNA binding is not sufficient for nuclear localisation of regulatory proteins in *Saccharomyces cerevisiae*. *Mol.Cell Biol.* **6**, 4763-4766.

42. Yocum, R.R., Hanley, S., West, R.J., and Ptashne, M. (1984). Use of lacZ fusions to delimit regulatory elements of the inducible divergent GAL1-GAL10 promoter in *Saccharomyces cerevisiae*. *Mol. Cell Biol.* 4, 1985-1998.
43. de Groot, R.P. and Kruijer, W. (1990) Transcriptional activation by TGF beta 1 mediated by the dyad symmetry element (DSE) and the TPA responsive element (TRE). *Biochem. Biophys. Res. Commun.*, **168**, 1074-1081.
44. Kroll, K.L. and Amaya, L. (1996) Transgenic *Xenopus* embryos from sperm nuclear transplantations reveal FGF signaling requirements during gastrulation. *Development*, **122**, 3173-3183.
- 10 45. Niewkoop, P.D. and Faber, J. (1967) Normal Table of *Xenopus laevis* (Daudin). Amsterdam, North Holland.
46. Frixen, U. H. *et al.* E-cadherin-mediated cell-cell adhesion prevents invasiveness of human carcinoma cells. *Journal of Cell Biology* **113**, 173-185 (1991).
47. Vleminckx, K., Vakaet Jr, L., Mareel, M., Fiers, W. & van Roy, F. Genetic manipulation of E-cadherin expression by epithelial tumour cells reveals an invasion suppressor role. *Cell* **66**, 107-119 (1991).
- 15 48. Perl, A. K., Wilgenbus, P., Dahl, U., Semb, H. & Christofori, G. A causal role for E-cadherin in the transition from adenoma to carcinoma. *Nature (London)* **392**, 190-193 (1998).
- 20 49. Potter, E., Bergwitz, C. & Brabant, G. The cadherin-catenin system: Implications for growth and differentiation of endocrine tissues. *Endocrine Reviews* **20**, 207-239 (1999).
50. Becker, K. F. *et al.* E-cadherin gene mutations provide clues to diffuse type gastric carcinomas. *Cancer Research* **54**, 3845-3852 (1994).
- 25 51. Berx, G., Nollet, F. & van Roy, F. Dysregulation of the E-cadherin/catenin complex by irreversible mutations in human carcinomas. *Cell Adhesion and Communication* **6**, 171-184 (1998).
52. Brabant, G. *et al.* E-cadherin - a differentiation marker in thyroid malignancies. *Cancer Research* **53**, 4987-4993 (1993).
- 30 53. Graff, J. R. *et al.* E-cadherin expression is silenced by DNA hypermethylation in human breast and prostate carcinomas. *Cancer Research* **55**, 5195-5199 (1995).
54. Yoshiura, K. *et al.* Silencing of the E-cadherin invasion-suppressor gene by CpG

methylation in human carcinomas. *Proceedings of the National Academy of Sciences of the United States of America* **92**, 7416-7419 (1995).

55. Behrens, J., Löwrick, O., Klein-Hitpass, L. & Birchmeier, W. The E-cadherin promoter: functional analysis of a G^oC-rich region and an epithelial cell-specific palindromic regulatory element. *Proceedings of the National Academy of Sciences of the United States of America* **88**, 11495-11499 (1991).

56. Girolidi, L. A. *et al.* Role of E boxes in the repression of E-cadherin expression. *Biochemical and Biophysical Research Communications* **241**, 453-458 (1997).

57. Hennig, G. *et al.* Progression of carcinoma cells is associated with alterations in chromatin structure and factor binding at the E-cadherin promoter *in vivo*. *Oncogene* **11**, 475-484 (1995).

58. Ji, X. D., Woodard, A. S., Rimm, D. L. & Fearon, E. R. Transcriptional defects underlie loss of E-cadherin expression in breast cancer. *Cell Growth & Differentiation* **8**, 773-778 (1997).

59. Hajra, K. M., Ji, X. D. & Fearon, E. R. Extinction of E-cadherin expression in breast cancer via a dominant repression pathway acting on proximal promoter elements. *Oncogene* **18**, 7274-7279 (1999).

60. Miettinen, P. J., Ebner, R., Lopez, A. R. & Derynck, R. TGF-beta induced transdifferentiation of mammary epithelial cells to mesenchymal cells: involvement of type I receptors. *Journal of Cell Biology* **127**, 2021-2036 (1994).

61. Shiozaki, H. *et al.* Effect of epidermal growth factor on cadherin-mediated adhesion in a human oesophageal cancer cell line. *British Journal of Cancer* **71**, 250-258 (1995)

62. Reichmann, E. *et al.* Activation of an inducible c-FosER fusion protein causes loss of epithelial polarity and triggers epithelial-fibroblastoid cell conversion. *Cell* **71**, 1103-1116 (1992).

63. Batsche, E., Muchardt, C., Behrens, J., Hurst, H. C. & Cremisi, C. RB and c-Myc activate expression of the E-cadherin gene in epithelial cells through interaction with transcription factor AP-2. *Molecular and Cellular Biology* **18**, 1-12 (1998).

64. Torban, E. & Goodyer, P. R. Effects of PAX2 expression in a human fetal kidney (HEK293) cell line. *Biochimica et Biophysica Acta - Molecular Cell Research* **1401**, 53-62 (1998).

65. Spath, G. F. & Weiss, M. C. Hepatocyte nuclear factor 4 provokes expression of epithelial marker genes, acting as a morphogen in dedifferentiated hepatoma cells. *Journal of Cell Biology* **140**, 935-946 (1998).
66. Battle, E. *et al.* The transcription factor Snail is a repressor of E-cadherin gene expression in epithelial tumour cells. *Nature Cell Biology* **2**, 84-89 (2000).
67. Cano, A. *et al.* The transcription factor Snail controls epithelial-mesenchymal transitions by repressing E-cadherin expression. *Nature Cell Biology* **2**, 76-83 (2000).
68. Remacle, J. E. *et al.* New mode of DNA binding of multi-zinc finger transcription factors: deltaEF1 family members bind with two hands to two target sites. *EMBO Journal* **18**, 5073-5084 (1999).
69. Verschueren, K. *et al.* SIP1, a novel zinc finger/homeodomain repressor, interacts with Smad proteins and binds to 5'-CACCT sequences in candidate target genes. *Journal of Biological Chemistry* **274**, 20489-20498 (1999).
70. Derynck, R., Zhang, Y. & Feng, X. H. Smads: transcriptional activators of TGFbeta-responses. *Cell* **95**, 737-740 (1998).
71. Massague, J. TGF-beta signal transduction. *Annual Review of Biochemistry* **67**, 753-791 (1998)
72. Andre, F. *et al.* Integrins and E-cadherin cooperate with IGF-I to induce migration of epithelial colonic cells. *International Journal of Cancer* **83**, 497-505 (1999).
73. Hirohashi, S. Inactivation of the E-cadherin-mediated cell adhesion system in human cancers. *American Journal of Pathology* **153**, 333-339 (1998).
74. Bird, A. P. & Wolffe, A. P. Methylation-induced repression - Belts, braces, and chromatin. *Cell* **99**, 451-454 (1999).
75. Wotton, D., Lo, R. S., Lee, S. & Massague, J. A Smad transcriptional corepressor. *Cell* **97**, 29-39 (1999).
76. Cameron, E. E., Bachman, K. E., Myohanen, S., Herman, J. G. & Baylin, S. B. Synergy of demethylation and histone deacetylase inhibition in the re-expression of genes silenced in cancer. *Nature Genetics* **21**, 103-107 (1999).
77. Gossen, M. *et al.* Transcriptional activation by tetracyclines in mammalian cells. *Science (Washington DC)* **268**, 1766-1769 (1995).
78. Bracke, M. E., Van Larebeke, N. A., Vyncke, B. M. & Mareel, M. M. Retinoic acid modulates both invasion and plasma membrane ruffling of MCF-7 human mammary

carcinoma cells in vitro. *British Journal of Cancer* **63**, 867-872 (1991).

79. Gossen, M. & Bujard, H. Tight control of gene expression in mammalian cells by tetracycline-responsive promoters. *Proceedings of the National Academy of Sciences of the United States of America* **89**, 5547-5551 (1992).

5 80. Tybulewicz, V. L. J., Crawford, C. E., Jackson, P. K., Bronson, R. T. & Mulligan, R. C. Neonatal lethality and lymphopenia in mice with a homozygous disruption of the *c-abl* proto-oncogene. *Cell* **65**, 1153-1163 (1991).

81. Bussemakers, M. J. G., Van de Ven, W. J. M., Debruyne, F. M. J. & Schalken, J. A. Identification of High Mobility Group Protein I(Y) as potential progression marker for prostate cancer by differential hybridization analysis. *Cancer Research* **51**, 606-611
10 (1991)

82. .van Hengel, J., Vanhoenacker, P., Staes, K. & van Roy, F. Nuclear localization of the p120^{ctn} Armadillo-like catenin is counteracted by a nuclear export signal and by E-cadherin expression. *Proceedings of the National Academy of Sciences of the United States of America* **96**, 7980-7985 (1999).
15

83. Bracke, M. E. *et al.* Insulin-like growth factor I activates the invasion suppressor function of E-cadherin in MCF-7 human mammary carcinoma cells *in vitro*. *British Journal of Cancer* **68**, 282-289 (1993).

84. Bracke, M. E., Boterberg, T., Bruyneel, E. A. & Mareel, M. M. in *Metastasis Methods and Protocols* (eds. Brooks, S. & Schumacher, U.) In press (Humana Press, Totowa, 1999).
20

85. Andre, F. *et al.* Protein kinase C-gamma and -delta are involved in insulin-like growth factor I-induced migration of colonic epithelial cells. *Gastroenterology* **116**, 64-77 (1999).
25

Claims

1. A method of identifying transcription factors such as activators and/or repressors comprising providing cells with a nucleic acid sequence at least comprising a sequence CACCT, preferably twice a CACCT sequence, as bait(s) for the screening
5 of a library encoding potential transcription factors and performing a specificity test to isolate said factors.
2. A method of identifying transcription factors such as activators and/or repressors comprising providing cells with a nucleic acid sequence comprising one of the
10 sequences CACCT-N-CACCT, CACCT-N-AGGTG, AGGTG-N-CACCT or AGGTG-N-AGGTG as bait wherein N is a spacer sequence.
3. A method according to claims 1 or 2 characterized in that the transcription factor comprises separated clusters of zinc fingers.
4. A method according to any of claims 1 to 3 wherein the sequence originates from a
15 promoter region.
5. A method according to claim 4 wherein the promoter region is selected from Brachyury, α 4-integrin, follistatin or E-cadherin.
6. Transcription factors obtainable by a method according to any of claims 1 to 5.
7. A method of identifying compounds with an interference capability towards
20 transcription factors as defined in claim 6 by
 - a) adding a sample comprising a potential compound to be identified to a test system comprising: (i) a nucleotide sequence comprising one of the sequences CACCT-N-CACCT, CACCT-N-AGGTG, AGGTG-N-CACCT or AGGTG-N-AGGTG as bait wherein N is a spacer, and (ii) a protein capable to bind said nucleotide sequence,
 - 25 b) incubating said sample in said system for a period sufficient to permit interaction of the compound or its derivative or counterpart thereof with said protein,
 - c) comparing the amount and/or activity of the protein bound to the nucleotide sequence before and after said adding and
 - d) identification and optionally isolation and/or purification of the compound.
- 30 8. A method according to claim 7 wherein the protein is a Smad-interacting protein.
9. A method according to claim 8, wherein said Smad-interacting protein is SIP1.
10. A compound obtainable by a method according to any of claims 7 to 9.

11. A compound according to claim 10 that modifies regulation of E-cadherin expression by SIP1.

12. A compound according to any of claims 10 to 11 for use as a medicament.

13. Use of a compound according to any of claims 10 to 11 for the manufacture of a
5 medicament to prevent tumor invasion and/or metastasis.

14. Test kit to perform the method of claim 7 comprising at least (i) a nucleotide sequence comprising one of the sequences CACCT-N-CACCT, CACCT-N-AGGTG, AGGTG-N-CACCT or AGGTG-N-AGGTG wherein N is a spacer sequence and (ii) a protein capable to bind said nucleotide sequence.

10 15. Test kit to perform the method of claim 2 at least comprising a nucleic acid sequence comprising one of the sequences CACCT-N-CACCT, CACCT-N-AGGTG, AGGTG-N-CACCT or AGGTG-N-AGGTG wherein N is a spacer sequence.

16. A method for detecting an interaction between a first interacting protein and a second interacting protein comprising

15 a) providing a suitable host cell with a first fusion protein comprising a first interacting protein fused to a DNA binding domain capable to bind a nucleic acid sequence comprising one of the sequences CACCT-N-CACCT, CACCT-N-AGGTG, AGGTG-N-CACCT or AGGTG-N-AGGTG wherein N is a spacer sequence,

20 b) providing said suitable host cell with a second fusion protein comprising a second interacting protein fused to a DNA binding domain capable to bind a nucleic acid sequence comprising one of the sequences CACCT-N-CACCT, CACCT-N-AGGTG, AGGTG-N-CACCT or AGGTG-N-AGGTG wherein N is a spacer sequence,

25 c) subjecting said host cell to conditions under which the first interacting protein and the second interacting protein are brought into close proximity and determining whether a detectable gene present in the host cell and located adjacent to said nucleic acid sequence has been expressed to a greater degree than if expressed in the absence of the interaction between the first and the second interacting protein.

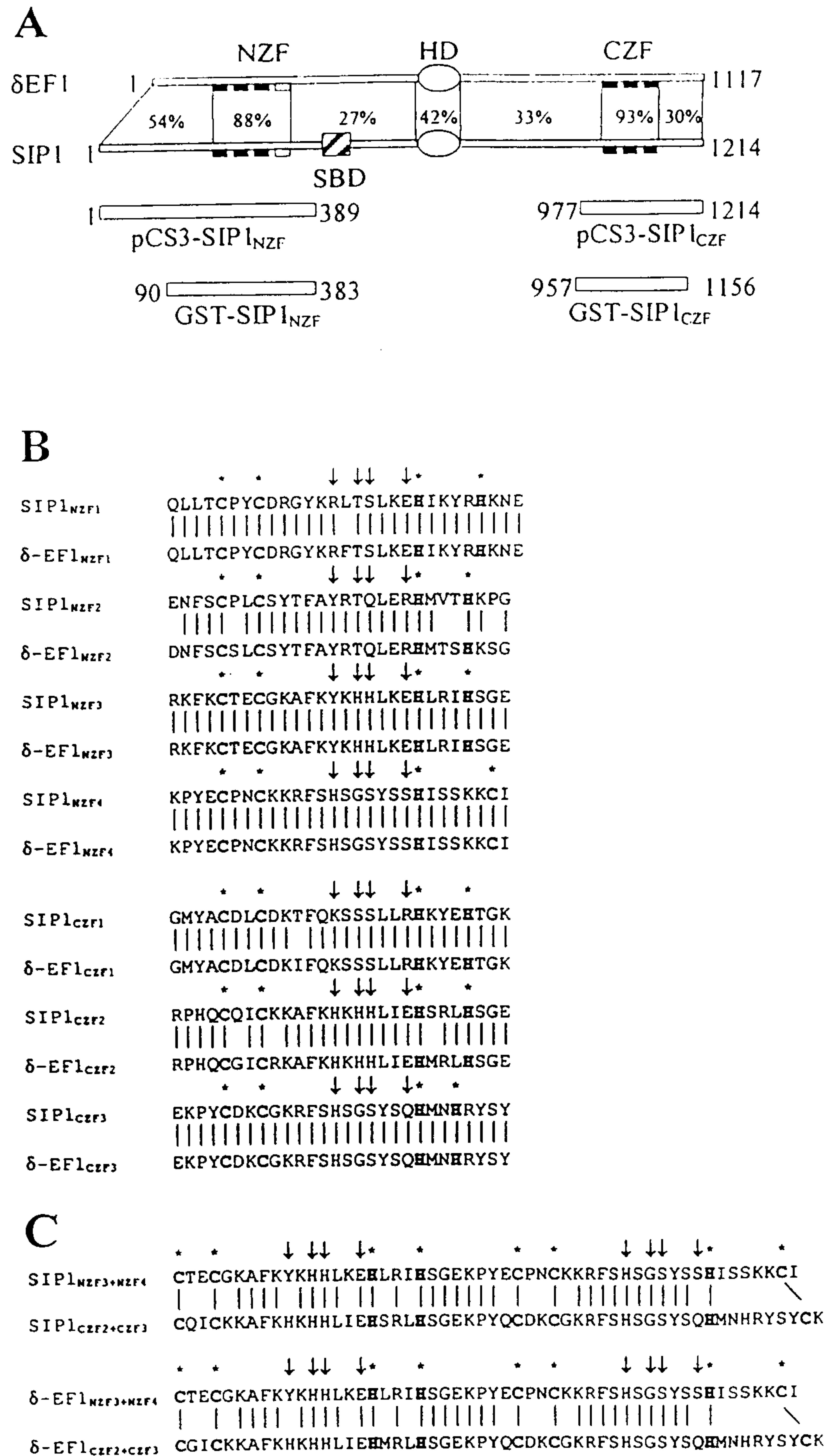
30 17. An isolated nucleic acid sequence comprising one of the sequences CACCT-N-CACCT, CACCT-N-AGGTG, AGGTG-N-CACCT or AGGTG-N-AGGTG wherein N is a spacer.

18. Use of a nucleic acid sequence at least comprising a sequence CACCT, preferably a nucleic acid sequence comprising one of the sequences CACCT-N-

CACCT, CACCT-N-AGGTG, AGGTG-N-CACCT or AGGTG-N-AGGTG wherein N is a spacer, for the identification of new target genes.

1/2

Figure 1



2/2

Figure 2

