

## The Zinc Finger-Associated SCAN Box Is a Conserved Oligomerization Domain

AMY J. WILLIAMS, STEPHEN C. BLACKLOW, AND TUCKER COLLINS\*

*Department of Pathology, Brigham and Women's Hospital, Boston, Massachusetts 02115*

Received 14 July 1999/Returned for modification 8 August 1999/Accepted 27 August 1999

**A number of Cys<sub>2</sub>His<sub>2</sub> zinc finger proteins contain a highly conserved amino-terminal motif termed the SCAN domain. This element is an 80-residue, leucine-rich region that contains three segments strongly predicted to be  $\alpha$ -helices. In this report, we show that the SCAN motif functions as an oligomerization domain mediating self-association or association with other proteins bearing SCAN domains. These findings suggest that the SCAN domain plays an important role in the assembly and function of this newly defined subclass of transcriptional regulators.**

Transcription factors frequently consist of modular elements that include a DNA-binding domain and one or more separable effector domains that may activate or repress transcriptional initiation (for a review, see reference 37). Although the majority of the conserved sequence motifs identified in transcription factors are associated with DNA binding, many transcription factors also contain extended motifs that mediate oligomerization to create an active complex. For example, in transcription factors that bind DNA as a dimer, the leucine zipper and helix-loop-helix motifs serve as dimerization domains and increase the potential for functional variation (for a review, see reference 27). In other transcription factors, such as heat shock factor, trimerization is required for specific DNA binding and is controlled by a coiled-coil oligomerization domain (33). Structural modules within transcription factors can regulate subcellular localization, DNA binding, and gene expression by mediating selective association of the transcription factors with each other or with other cellular components.

A variety of modular sequence motifs accompany zinc finger elements in the zinc finger family of transcription factors (20). These motifs include the Kruppel-associated box (KRAB); the finger-associated box (FAX), found in a large number of *Xenopus* zinc finger proteins; the poxvirus and zinc finger (POZ) domain, also known as the BTB domain (Broad-Complex, Tramtrack, and Bric-a-brac); and the SCAN box or leucine-rich region (LeR). These conserved domains have functions that are important in the regulation of the transcription factors.

KRAB is a conserved amino acid sequence motif at the amino-terminal end of proteins that contain multiple Cys<sub>2</sub>-His<sub>2</sub> (C<sub>2</sub>H<sub>2</sub>) zinc fingers at their carboxy termini (4). The KRAB domain is found in almost one-third of the 300 to 700 genes encoding C<sub>2</sub>H<sub>2</sub> zinc fingers. The KRAB domain itself spans approximately 75 amino acids (aa), is divided into A and B boxes, and is predicted to contain two charged amphipathic helices (4). The KRAB domain functions to repress transcription (26, 32, 39, 41) by recruiting the transcriptional corepressor KRAB-associated protein-1 (12) or the KRAB-A interacting protein (19). KRAB-containing zinc finger proteins are likely to play a regulatory role during development.

A second zinc finger-associated protein-protein interaction motif is the BTB or POZ domain. This is an evolutionarily

conserved protein-protein interaction domain that is found at the N terminus of C<sub>2</sub>H<sub>2</sub>-type zinc finger transcription factors and in some proteins having a kelch motif (3). With almost 50 distinct BTB entries in publicly available sequence data bases, it is estimated that 5 to 10% of the zinc finger proteins in man contain these domains. The organization of the human promyelocytic leukemia zinc finger (PLZF) protein is typical of BTB domain proteins, with a single 120-aa BTB domain found at the N terminus of the protein, followed by a central region of several hundred amino acids, and ending with a series of C<sub>2</sub>H<sub>2</sub> Kruppel-type zinc fingers. The crystal structure of the BTB domain of PLZF reveals a tightly intertwined dimer with an extensive hydrophobic interface, in which the central scaffolding of the domain is made up of a cluster of  $\alpha$ -helices flanked by short  $\beta$ -sheets (1, 23). Many BTB proteins are transcriptional regulators that mediate expression through the control of chromatin conformation. The PLZF protein, for example, is a transcriptional repressor in which the BTB domain interacts with several of the components of the histone deacetylase complex (14, 16, 24). PLZF also occurs as a fusion protein with the retinoic acid receptor  $\alpha$  (RAR $\alpha$ ) in a rare t(11;17) form of acute promyelocytic leukemia (6, 10). In this disorder, the PLZF-RAR $\alpha$  fusion protein may repress transcription at retinoic acid-sensitive sites through BTB-mediated recruitment of the histone deacetylase complex (14, 15, 16, 24).

A third type of extended sequence motif found in some zinc finger transcription factors is the SCAN domain, originally identified in ZNF174 (40). The name for this domain, also called LeR because it is a leucine-rich region (32), was derived from the first letters of the names of four proteins initially found to contain this domain (SRE-ZBP, CTfin51, AW-1 [ZNF174], and Number 18 cDNA or ZNF20) (2, 13, 31, 40). The SCAN domain consists of about 80 aa, and the primary amino acid sequence of the domain is not similar to any of the other zinc finger-associated domains.

In this report we define a function for the SCAN domain. The element is capable of mediating association between specific members of the SCAN domain family of zinc finger transcription factors. These findings suggest that the SCAN domain plays an important role in controlling the assembly of complexes that contain this newly defined subfamily of zinc finger proteins.

### MATERIALS AND METHODS

**Plasmid construction.** GAL4 and VP16 fusion gene plasmids were constructed by PCR amplification of the SCAN domains of ZNF174, ZNF165, ZNF191,

\* Corresponding author. Mailing address: Brigham and Women's Hospital, Vascular Research Division, 221 Longwood Ave., Boston, MA 02115. Phone: (617) 732-5990. Fax: (617) 278-6990. E-mail: tcollins@bustoff.bwh.harvard.edu.

ZNF192, ZnF 20, ZnFPFH, CTfin51, FPM315, and SRE-ZBP from human genomic DNA with forward primers that contained a *Bam*HI site and reverse primers that contained an *Xba*I site. PCR products were cloned into TA cloning vector pCR 2.1 (Invitrogen, Carlsbad, Calif.) and then digested with *Bam*HI and *Xba*I and cloned into *Bam*HI-*Xba*I-digested expression vectors pM and pVP16 (Clontech, Palo Alto, Calif.). The SCAN domain regions placed into the expression vectors correspond to the following nucleotides from deposited GenBank cDNA sequences: ZNF174, nucleotides (nt) 707 to 969; ZNF165, nt 363 to 622; ZNF191, nt 301 to 560; ZNF192, nt 318 to 384; ZnF 20, nt 120 to 371; ZnFPFH, nt 32153 to 32415 (corresponds to reported genomic sequence, cDNA sequence not available); CTfin51, nt 314 to 712; FPM315, nt 396 to 655; and SRE-ZBP, nt 2 to 180. ZNF174 SCAN domain mutants were generated by site-directed mutagenesis (Clontech). Constructs containing the FOS and JUN leucine zippers fused to GAL4 were fused to GAL4 and VP16 as described above. Bacterial expression plasmids for ZNF174 wild-type SCAN domain (aa3 to 128) and mutant SCAN domain (aa 44 plus 45 L→P) were made by inserting coding sequences in frame with the *Sma*I site of pGEX 2T and the *Bam*HI-*Eco*RI site of pGEX 6P (Pharmacia) respectively.

**CD spectroscopy.** Circular dichroism (CD) spectra were recorded at 25°C on an Aviv 62DS spectrometer equipped with a thermoelectric temperature controller. Protein concentrations were estimated by the method of Bradford with the Bio-Rad kit. Samples of recombinant ZNF174 wild-type SCAN domain (aa 3 to 128) and mutant SCAN domain (aa 44 plus 45 L→P) at a concentration of 5 μM were prepared in 1 phosphate-buffered saline (PBS) (67 mM phosphate, 150 mM NaCl; pH 7.0) containing 0.2 mM dithiothreitol (DTT). Spectra representing the average of five scans from 260 to 205 nm were measured in a 10-mm path length cuvette by using a step size of 1 nm and a 5-s signal averaging time. All spectra were corrected for the baseline obtained with the buffer alone (5, 25).

**Two-hybrid assay.** COS-7 cells were obtained from the American Type Culture Collection and grown on 10-cm dishes in Dulbecco modified Eagle medium supplemented with 10% fetal bovine serum and 2 mM glutamine. Transfections were performed by the calcium phosphate procedure as described in the manufacturer's insert for the Mammalian Matchmaker two-hybrid assay kit (Clontech). All transfections contained 2 μg of the GAL4×5 CAT reporter construct and 10 μg of each expression construct to be tested. Whole-cell extracts were prepared 48 h after transfection, and 10 μl of extract was assayed for chloramphenicol acetyltransferase (CAT) activity by the two-phase fluor diffusion technique (36).

**Immunoprecipitation assays.** In vitro transcription and translations were performed with a TNT-coupled reticulocyte system (Promega) with [<sup>35</sup>S]methionine (Amersham Corp.) according to the manufacturer's instructions. For each immunoprecipitation sample, 30 μl of protein A/G-agarose (Santa Cruz) was pre-cleared by mixing with 1 μg of mouse immunoglobulin G (IgG) at 4°C for 3 h. The pre-cleared agarose was then mixed with 30 μl of bovine serum albumin (1 mg/ml), 20 μl of AU1 antibody (BabCO), and 5 μl of radiolabeled in vitro translation product in 400 μl of radioimmunoprecipitation assay (RIPA) buffer (50 mM Tris-HCl, 150 mM NaCl, 1% NP-40, 0.5% deoxycholate, 1 mM EDTA). The mixture was incubated at 4°C overnight with gentle rotation. The next day, the protein A/G-agarose was washed three times with cold RIPA buffer. Bound proteins were eluted by boiling for 2 min in 1× sodium dodecyl sulfate (SDS) loading buffer and then separated on a SDS-12% polyacrylamide gel. After electrophoresis, the gels were soaked for 15 min in gel drying solution (10% acetic acid, 30% methanol, 3% glycerol), dried for 1 h at 80°C, and then autoradiographed overnight.

**Bacterial expression and purification of GST fusion proteins.** *Escherichia coli* BL21 bacteria transformed with pGEX-ZNF174 fusion constructs were grown to an optical density at 600 nm of 0.8 to 1.2 at 30°C and then induced with 0.1 mM IPTG (isopropyl-β-D-thiogalactopyranoside) for 1 to 2 h. Expressed proteins were purified by affinity chromatography by using glutathione-Sepharose columns (Pharmacia) according to the manufacturer's instructions. Fusion proteins were cleaved from glutathione *S*-transferase (GST) by using either thrombin or PreScission protease (Pharmacia) in 1× PBS plus 1 mM DTT. The purity of the proteins was assessed by SDS-polyacrylamide electrophoresis (PAGE).

**Size exclusion chromatography.** Recombinant proteins were concentrated by centrifugation by using a VivaSpin 5,000 concentrator and then run over a Bio-Prep SE-100/17 size exclusion chromatography column (8 by 300 mm; Bio-Rad) with the aid of the Biologic HR chromatography system (Bio-Rad). Columns were run in 1× PBS plus 1 mM DTT at a flow rate of 0.2 ml/min. Columns were calibrated under the same conditions with protein standards from Pharmacia (RNase A,  $M_r = 13,700$ ; chymotrypsinogen A,  $M_r = 25,000$ ; ovalbumin,  $M_r = 43,000$ ; BSA,  $M_r = 67,000$ ; aldolase,  $M_r = 158,000$ ).

## RESULTS

**The SCAN domain is a highly conserved zinc finger-associated motif.** Since our initial description of the SCAN box in six zinc finger transcription factors was reported (40), the number of family members has increased substantially. By using the SCAN domain amino acid sequence from ZNF174 to screen the GenBank database, a total of 19 genes were identified with SCAN motifs. Alignment of the amino acid sequences of the

SCAN domains encoded by each of these genes is presented in Fig. 1A. Many additional EST and cosmid sequences were also found to contain SCAN boxes (data not shown). Remarkably, all of the genes with SCAN domains contain C<sub>2</sub>H<sub>2</sub> zinc fingers (except for TRFA, a partial cDNA clone for which the entire open reading frame has not yet been reported). A recent report identified a member of the SCAN domain family as an adipogenic cofactor bound by peroxisome proliferator-activated receptor γ (PPARγ) (5a). This protein, termed PPARγ coactivator 2 (PGC-2), has a partial SCAN domain with homology to the N-terminal 60 residues, and it does not have zinc fingers.

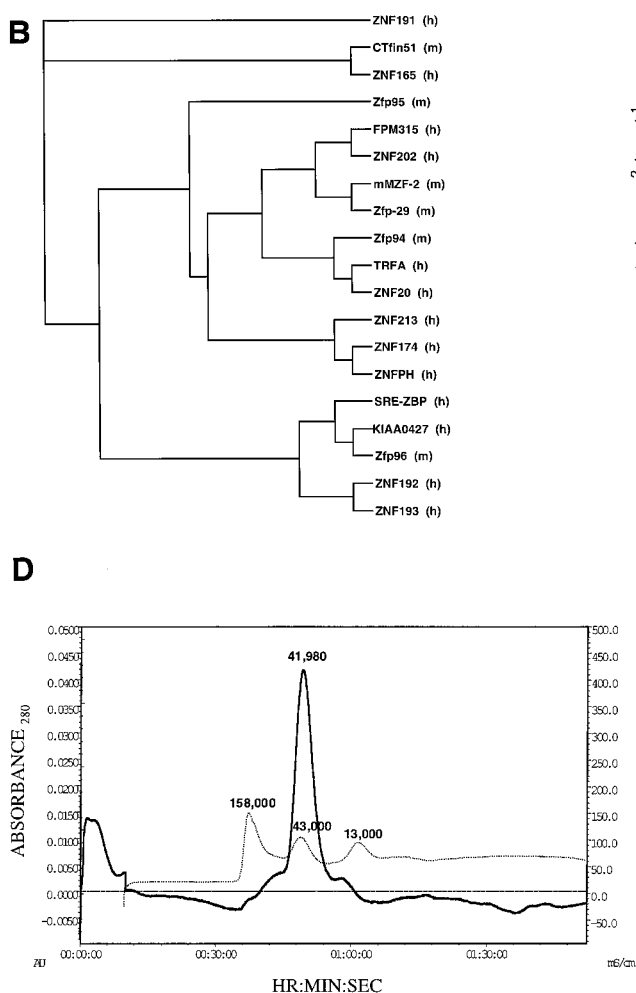
The SCAN domain is always located at the amino terminus of the zinc finger transcription factor. The degree of amino acid identity varies among SCAN domain sequences and ranges from as low as 39% (comparing Zfp-29 and ZnFPFH) to as high as 85% between KIAA0427, SRE-ZBP, and Zfp96. Of the 80 aa that comprise the SCAN domain, 11 are identical among all family members and form the basis for the consensus sequence shown in Fig. 1A. These 11 invariant residues include 3 conserved prolines at positions 16, 33, and 55 (Fig. 1A). Contained in the ZNF174 SCAN domain are two protein kinase C phosphorylation sites, (S/T)-X-(R/K) (SFR and SSK, located at residues 4 and 75, respectively), and one conserved casein kinase II phosphorylation site, (S/T)-X-X-(D/E) (SSKE located at residue 69) (Fig. 1A).

The alignment of human and mouse SCAN domains shown in Fig. 1A was used to generate an unrooted phylogenetic tree (Fig. 1B). Phylogenetic tree analysis reconstructs the history of successive divergences which took place during evolution by comparing the relatedness of different molecular sequences. The results of phylogenetic analysis are depicted as a hierarchical branching diagram (phylogenetic tree), with each branch representing a group of genes derived from a putative single ancestral lineage. The alignment indicates that both orthologs (related genes derived during speciation) and paralogs (related genes found in one genome) are present in the SCAN domain family.

Secondary structure predictions for the SCAN domain (35) suggest that it may form several α-helices (Fig. 1A), with the three conserved prolines in the consensus sequence serving to divide the SCAN domain into at least three predicted α-helices, the first of which is amphipathic (Fig. 1A and reference 40). To test this prediction, we analyzed a recombinant form of ZNF174 (aa 3 to 128) containing the SCAN domain (aa 45 to 124) by CD spectroscopy (Fig. 1C). The characteristic dips in the far UV CD spectrum at wavelengths of 208 and 222 nm clearly indicate that the SCAN domain has substantial helical character in solution (Fig. 1C). Indeed, consideration of the molar ellipticity at 222 nm suggests that the ZNF174 polypeptide spanning residues 3 to 128 contains ca. 25 to 30% helix. Because the recombinant polypeptide spans not only the 80-residue SCAN domain (61% predicted helical content) but also 45 additional flanking residues not predicted to have helical character, the observed helicity, as measured by CD, is consistent with that estimated based on secondary structure prediction algorithms (33a). We next determined whether disruption of the predicted central helix would abolish the α-helical character of the SCAN domain. When the central conserved region of the SCAN domain is altered by mutating two conserved leucines to prolines at positions 44 and 45, the dip in the far UV CD spectrum at 222 nm is lost (Fig. 1C), indicating that the helical character of the domain is substantially disrupted.

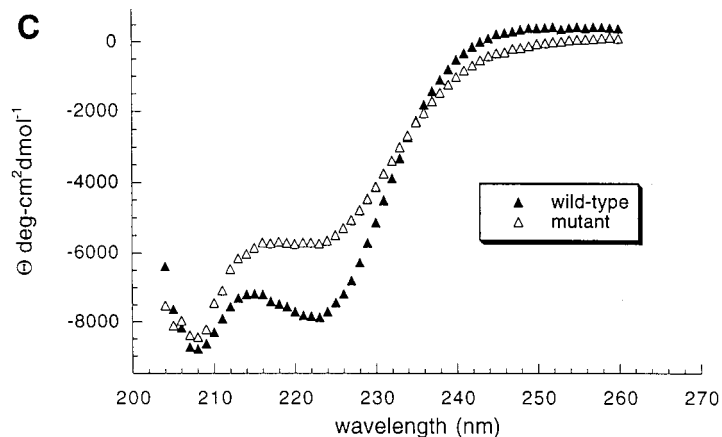
Several analytical approaches were used to investigate the oligomerization properties and stability of the isolated SCAN





domain. Size exclusion chromatography of the purified recombinant SCAN domain (Fig. 1D) demonstrates that the SCAN domain exists as a stable molecular species. A plot of the logarithm of the molecular mass of protein standards against the elution volume predicts the molecular mass of the SCAN domain is ca. 42 kDa. Since the predicted monomer size is ca. 14 kDa, the isolated SCAN domain could be in either a dimeric or trimeric state. We next performed thermally induced unfolding, monitored by CD spectroscopy, to investigate the stability of the SCAN domain. The intact SCAN domain undergoes a single irreversible unfolding transition, whereas the mutant form of the SCAN domain with the proline substitutions does not exhibit the same sharp transition upon heating (data not shown). This type of irreversibility is observed in other multimeric proteins, in which aggregation of unfolded molecules interferes with refolding (18). Collectively, these studies demonstrate that the SCAN domain behaves as a stable oligomeric species under near-physiologic conditions.

**The SCAN domain is a protein-protein interaction motif.** A mammalian two-hybrid assay system was used to test the hypothesis that the SCAN domain mediates protein-protein interactions (Fig. 2A). One construct (the "bait") contains a SCAN domain fused in frame behind a GAL4 DNA-binding domain, and the second construct (the "target") contains a SCAN domain linked to the transcriptional activator VP16.



These constructs were cotransfected into COS-7 cells along with a promoter CAT-reporter construct that contains five GAL4 binding sites (Fig. 2A). Expression of the CAT reporter gene indicates there has been an interaction between the two fusion constructs.

Initially, we examined the ability of each SCAN domain GAL4 fusion protein to activate transcription of the CAT reporter plasmid when transfected alone. With the exception of ZNF191, none of the SCAN-GAL4 DBD fusion proteins tested activated transcription on their own (Fig. 2B, and data not shown). Similarly, none of the SCAN-VP16 fusions were capable of activating transcription when transfected singly. In control studies, levels of each of the SCAN domain containing GAL4 fusions were shown to be comparable, and the ability of each of the fusions to bind a GAL4 site was similar (data not shown). Taken as a whole, the current findings are consistent with previous studies demonstrating that selected SCAN domains do not activate (or repress) transcription (32, 40).

To determine whether the SCAN domain self-associates, the ability of the ZNF174 SCAN-GAL4 fusion to activate the CAT reporter gene in the presence of the ZNF174 SCAN-VP16 fusion was examined. As shown in Fig. 2B, coexpression of both fusion constructs markedly activates transcription compared with that of the empty GAL4 and VP16 vectors or with each of the SCAN domain fusion constructs alone (compare lane 4 with either lane 2 or lane 3). This activation requires DNA binding to the GAL4 sites, since no transcriptional activity is observed when the 5 GAL4 binding sites are removed from the promoter of the reporter gene (Fig. 2B, lane 5).

Next we tested the ability of the SCAN domain from ZNF174 to interact with the leucine zipper motifs of c-Fos and c-Jun to determine if the SCAN domain interacted nonspecifically with other transcription factors that contain amphipathic  $\alpha$ -helices mediating oligomerization. While the leucine zippers from c-Fos and c-Jun clearly associate to activate transcription (Fig. 2B, lanes 10 and 11), no interaction is seen when the SCAN domain of ZNF174 is cotransfected with either of the leucine zippers (Fig. 2B, lanes 6 to 9). This observation indicates that there is specificity in the interface of the ZNF174 SCAN domain that results in self-association.

To determine which SCAN domains have the ability to bind to one another, pairwise combinations of nine SCAN motifs were tested in the mammalian two-hybrid system. The results are presented as relative levels of CAT activation in Fig. 2C. Several general features of SCAN domain interactions can be

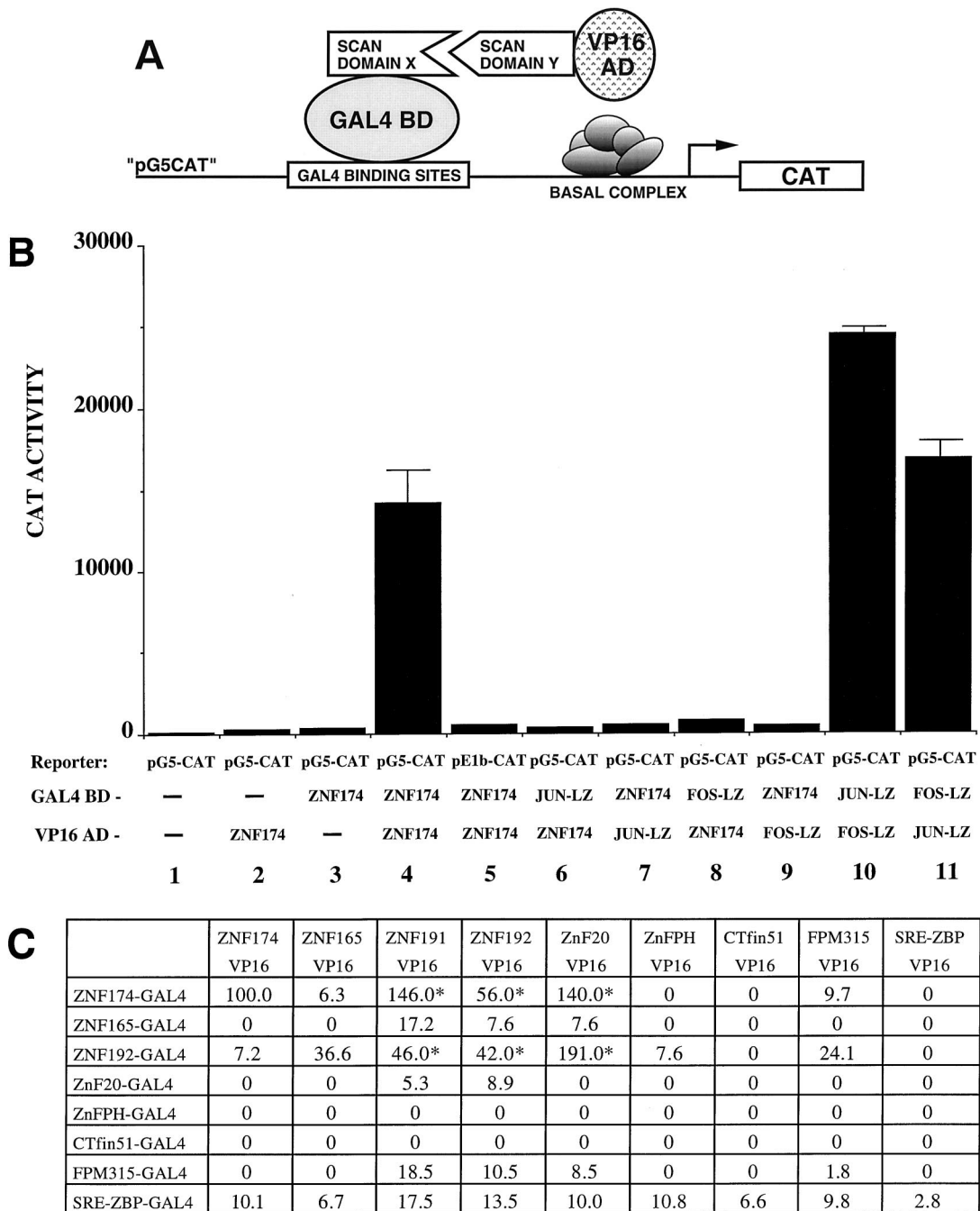


FIG. 2. The SCAN domain is a protein-protein interaction domain. (A) Schematic diagram of the mammalian two-hybrid assay. Specific interactions between SCAN domains from various zinc finger genes were tested by using a mammalian two-hybrid assay system. Plasmids encoding a SCAN domain fused with the yeast GAL4 DNA-binding domain (GAL4 BD) and a SCAN domain fused with the herpes simplex virus VP16 transactivating region (VP16 AD) were transfected into COS-7 cells, along with a reporter plasmid (pG5CAT) containing five consensus GAL4 binding sites and an E1b minimal promoter upstream of the CAT gene. Measured levels of CAT activity can be correlated with the relative affinity of an interaction between hybrid proteins. (B) The SCAN domain of ZNF174 can interact with itself in a mammalian two-hybrid assay system. COS cells were transfected with 2  $\mu$ g of reporter construct and 10  $\mu$ g each of the indicated GAL4 BD and VP16 AD expression plasmids. Reporter construct pE1b-CAT (lane 5) is identical to pG5-CAT except that it lacks GAL4 binding sites. ZNF174 in either the GAL4 BD or VP16 AD vector refers to the SCAN domain region (aa 41 to 128) of this gene, and JUN-LZ and FOS-LZ refer to two-hybrid constructs which contain the leucine zipper regions of these genes. CAT activity is shown in counts per minute. (C) Specificity of SCAN-SCAN protein interactions. The table indicates the relative amounts of CAT activity seen when specific pairwise combinations of SCAN domains are tested for interaction in the mammalian two-hybrid assay, normalized to self-association of the ZNF174 SCAN domain. The results observed with different pairs of SCAN domains are presented relative to this amount. Results from ZNF191-GAL4 are not included here due to the endogenous transcriptional activating activity of ZNF191 when fused to the GAL4 DNA-binding domain. Data are representative of at least three independent experiments. Asterisks indicate that the pooled data from these particular experiments were analyzed by the nonparametric Wilcoxon signed rank test and found to be significantly different ( $P < 0.05$ ) from the level of activity seen with ZNF174-GAL4 and ZNF174-VP16.

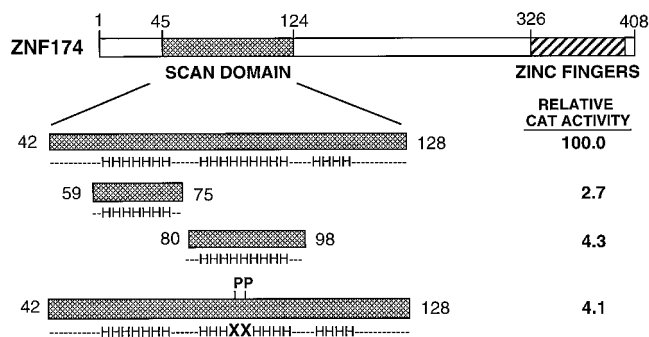


FIG. 3. Structural features of the SCAN domain required for self-association. A schematic representation of ZNF174 is shown at the top. The regions of the SCAN domain that were cloned into GAL4 BD and VP16 AD expression plasmids are indicated below. The first construct contains the full-length wild-type SCAN domain. The next two constructs represent smaller regions of the SCAN domain, with each one expressing individual predicted  $\alpha$ -helices. The last construct contains the full-length SCAN domain in which two conserved leucines have been mutated to prolines.  $\alpha$ -helical regions within the SCAN domain are indicated (H) below each construct. Relative levels of CAT activity between pairs of the expression constructs when tested by two-hybrid analysis are shown on the right.

inferred. First, not all SCAN domains are able to self-associate; in fact, the only two SCAN boxes that exhibit any self-association are those from ZNF174 and ZNF192. Second, interactions between different SCAN boxes are selective. For example, ZNF174 can interact with some, but not all, SCAN domains from other genes. Third, given that the magnitude of the transcriptional response in a two-hybrid assay system can correlate with the affinity of the two protein components for each other (11, 42), the variation between the relative affinities of the SCAN domains is significant. Reactions between some pairs of SCAN domain components were found to be orientation dependent (for example, ZNF174-GAL4 and ZnF20-VP16 interacted strongly, but no interaction was observed with the opposite orientation, ZNF174-VP16 and ZnF20-GAL4). This phenomenon has been observed for numerous protein pairs in two-hybrid systems, although the mechanism responsible for this directionality is unclear (11).

**Self-association of the SCAN domain requires the entire domain.** Predictions of the secondary structure of the SCAN domain suggest the presence of at least three  $\alpha$ -helices that are separated from one another by short looped regions bounded by proline residues. Although we have no evidence that these helices exist or are stably folded, we tested whether two of them might support self-association in isolation. However, neither a sequence spanning residues 59 to 75 nor one from residues 80 to 98 was capable of mediating self-association in the two-hybrid assay (Fig. 3). These findings suggest that a complete intact SCAN domain is required to mediate self-association.

We next determined whether disruption of the predicted central helix would abolish the ability of the ZNF174 SCAN domain to self-associate in the two-hybrid assay. When the central conserved region of the SCAN domain is altered by mutating two conserved leucines to prolines at positions 44 and 45, the helical structure of the domain is disrupted, as judged by CD (Fig. 1C). When this mutant form of the SCAN domain is screened in the two-hybrid assay, partners that bind to the native domain no longer associate with this mutant (Fig. 3). Taken together, these findings strongly suggest that the minimum length functional unit is the entire SCAN domain and

that structural integrity of this domain is required for self-association or association with other SCAN domain partners.

**ZNF174 oligomerizes via the SCAN domain.** Immunoprecipitation studies were performed to demonstrate that the SCAN domain is capable of mediating protein-protein interactions in the context of a full-length form of ZNF174. The immunoprecipitation assay involved cotranslating tagged and nontagged proteins and then looking for an association between the two protein forms by immunoprecipitation with an antibody that recognizes the tag. Coprecipitation of the native protein with the tagged form indicates an association of the two forms of the protein.

When an AU1-tagged form of full-length ZNF174 is cotranslated with two shorter, nontagged forms of ZNF174[1-172] and [-136-408], only the form containing the SCAN domain (ZNF174[1-172]) is coprecipitated in the presence of full-length tagged ZNF174 (Fig. 4A, lane 6). No coprecipitation band is observed with ZNF174[136-408], suggesting that the SCAN domain is required for self-association (Fig. 4A, lane 12). When tagged and nontagged forms of different sizes are mixed together after translation, coprecipitation is inefficient, indicating that limited exchange takes place and that the association is kinetically stable (data not shown).

In a SCAN domain containing two leucine-to-proline substitutions, the  $\alpha$ -helical structure of the domain is disrupted (Fig. 1C), and the ability of the domain to recruit SCAN domain partners in the mammalian two-hybrid assay is lost (Fig. 3). When the ability of this mutated SCAN domain (ZNF174[1-250mutL-P]) to bind to full-length tagged ZNF174 is compared with the native ZNF174[1-250], no coprecipitation band is observed with the mutant SCAN domain (Fig. 4B, lane 12), whereas the native control coprecipitates with the full-length tagged form (Fig. 4B, lane 6).

We next tested by immunoprecipitation assay whether full-length ZNF174 could associate with other members of the SCAN domain family. We chose to look at two SCAN proteins that had previously shown a strongly positive interaction with ZNF174 when tested in the mammalian two-hybrid assay (ZnF20 and ZNF191), as well as a third protein (FPM315) that showed only a weak interaction with ZNF174. When the AU1-tagged form of full-length ZNF174 is cotranslated with shorter, nontagged forms of either ZNF20[1-284], or ZNF191[1-140], coprecipitation bands are seen (Fig. 4C, lanes 6 and 12). In contrast, no coprecipitation band is observed when the tagged full-length ZNF174 was cotranslated with FPM315[1-129] (Fig. 4C, lane 18), a family member that does not interact with ZNF174 in the two-hybrid assay (Fig. 2C). These results demonstrate that ZNF174 can selectively bind other members of the SCAN family and confirm the two-hybrid findings.

## DISCUSSION

In this study we characterized a modular structural element, termed the SCAN domain, that occurs in members of the zinc finger family of transcription factors. This highly conserved 80-aa element is found in a rapidly expanding family of zinc finger proteins. The SCAN domain appears to control association of SCAN domain proteins into noncovalent, multisubunit complexes and may be the primary mechanism underlying partner choice in the oligomerization of these zinc finger transcription factors.

Since this is a newly identified subgroup of zinc finger proteins, we know remarkably little about the SCAN domain-containing proteins and their functions. A summary of what is known about the SCAN family members is presented in Table 1. Many of the SCAN domain family members appear to be

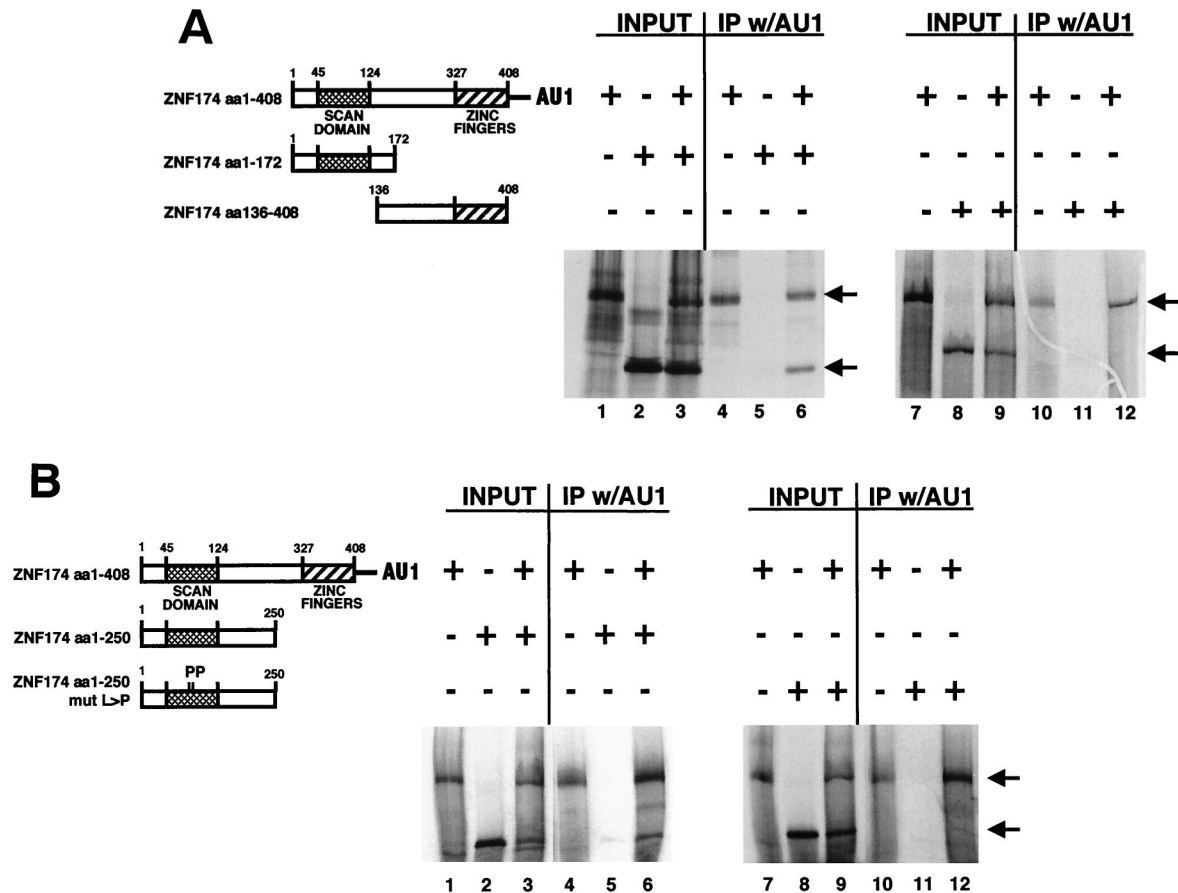


FIG. 4. Selective oligomerization via the SCAN domain. Schematic diagrams of the proteins generated in vitro translation are shown on the left side of the figure. Volumes shown in the input lane are half the amounts used for the immunoprecipitation assays. (A) The SCAN domain mediates self-association. Analysis of [<sup>35</sup>S]methionine-labeled AU1 epitope-tagged full-length ZNF174[1-408] cotranslated with [<sup>35</sup>S]methionine-labeled nontagged shorter forms of ZNF174. Lanes 1 to 6, ZNF174[1-172]; lanes 7 to 12, ZNF174[136-408]. (B)  $\alpha$ -Helical character of the SCAN domain is required for oligomerization. [<sup>35</sup>S]methionine-labeled nontagged wild-type SCAN domain and a mutant form of the SCAN domain in which leucines at positions 44 and 45 were changed to prolines (mutant L>P). Lanes 1 to 6, ZNF174[1-250, wild type]; lanes 7 to 12, ZNF174[1-250, mutant L>P]. (C) Selective interaction of SCAN domains. [<sup>35</sup>S]methionine-labeled nontagged heterologous SCAN domains. Lanes 1 to 6, Znf20[1-284]; lanes 7 to 12, ZNF191[1-140]; lanes 13 to 18, FPM315[1-129].

clustered together in specific chromosomal regions (Table 1). The genes located on chromosomes 3p21, 6p21.3, 11p15.5, and 16p13.3 are of particular interest because these locations are frequently disrupted in a variety of cytogenetic abnormalities. Several SCAN domain genes were cloned as a result of attempts to identify candidate disease genes that lie within these chromosomal regions. The clustered organization of genes for other zinc finger proteins has been reported from the analysis of human and mouse chromosomes (17), but a correlation between genomic organization and expression characteristics has not been established.

Several of the SCAN domain-containing genes are expressed at high levels in the testis and/or ovary (Table 1). Three genes, ZNF165, ZNF202, and Zfp-29, are expressed exclusively in the testis (9, 22, 28, 38). Studies with Zfp-29 and another SCAN family member, CTfin51, suggest that these genes may play a role in the regulation of spermatogenesis (7, 9, 31). CTfin51 (Zipro1, Ru49, or Zfp38) may also be important in lineage determination. In the cerebellar cortex, this SCAN box family member is a marker for the cerebellar granule neuronal lineage and may play a role in the proliferation of granule cell precursors in the developing cerebellum (42a, 43). Additionally, the gene is expressed in skin and increased dosage results in a hair loss phenotype associated with increased

epithelial cell proliferation and abnormal hair follicle development (42a). Collectively, these studies suggest that this family of transcription factors may perform a wide range of functions important in human cell differentiation or development.

The SCAN domain-containing zinc finger proteins can either activate or repress transcription, although recombinant SCAN boxes in isolation generally do not affect transcription. CTfin51 and myeloid zinc finger protein-2 (MZF-2) are both transcriptional activators that contain functional transactivation domains (Table 1). The structure of the activation domains and the potential role of coactivators in increasing transcription are not well understood. At least four of the SCAN domain family members also contain KRAB domains, suggesting that they may function as transcriptional repressors (Table 1). ZNF174 does not contain a consensus KRAB element but has been previously shown to be capable of repressing expression of a promoter-reporter CAT plasmid (40). This raises the possibility that ZNF174 contains a novel repression domain capable of interacting with corepressors to decrease gene expression. Little is known about the authentic DNA binding sites or the target genes that are controlled by the SCAN family members.

The phylogenetic tree analysis presented in Fig. 1B was constructed by using all of the SCAN domains from the genes

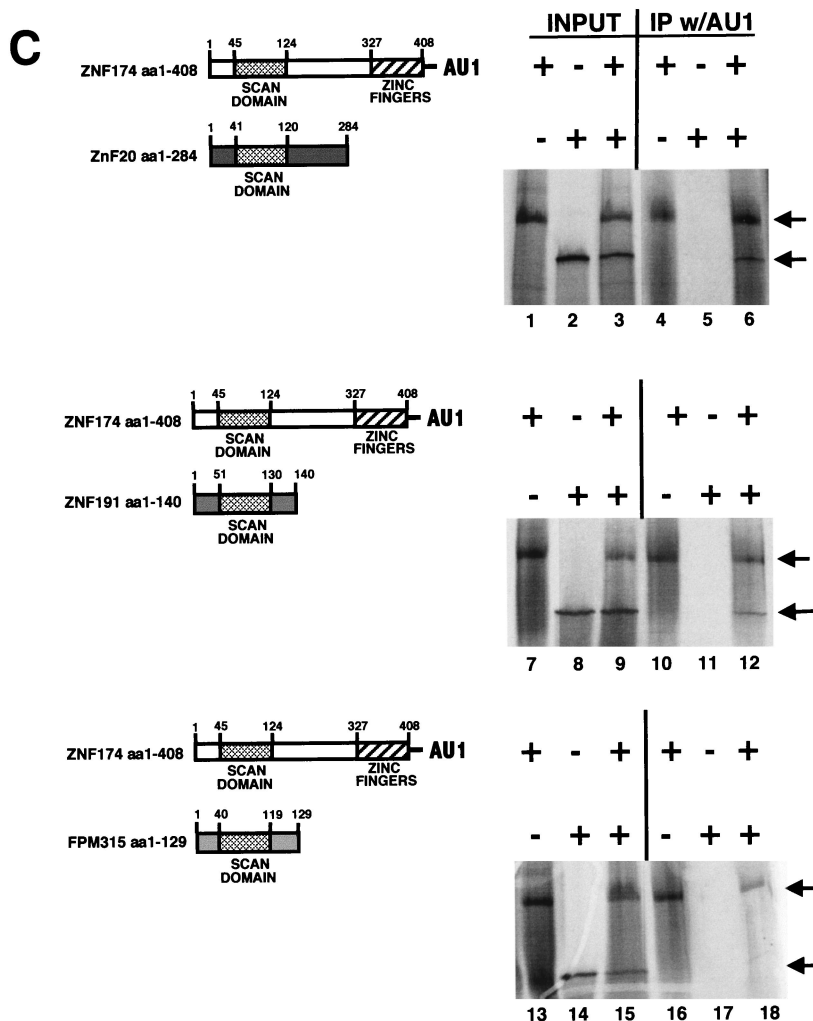


FIG. 4—Continued.

in Table 1, so it includes both human and mouse genes. It is possible that some of the pairings in the tree represent human and mouse homologs or orthologs (CTfin51 with ZNF165 and KIAA0427 with Zfp96, for example). Although the SCAN domains of these genes are quite similar, there is not a lot of sequence homology between these genes outside of the SCAN domain. There are several possible reasons for the sequence differences outside the SCAN domain. First, the genes are not orthologs. Second, these regions could have diverged after a speciation event. Finally, new sequences could have entered into the family by gene rearrangement. There is no strong correlation between the ability of two SCAN domains to interact with each other and their position in the tree. There are, however, several examples of SCAN domains from genes found on the same chromosome that have been paired together on the tree, suggesting that these genes may have arisen by gene duplication. Examples include TRFA and ZNF20 on chromosome 3, ZNF174 and ZNF213 on chromosome 16p13.3, and ZNF192 and ZNF193 on chromosome 6 (Fig. 1B and Table 1). Interestingly, the SCAN domain from ZNF191 is not clustered with any of the other SCAN domains on the phylogenetic tree. It is also the only SCAN domain that has the

ability to activate transcription, so it has clearly diverged from the other SCAN sequences to acquire these characteristics.

Multiple mechanisms may regulate the function of the SCAN domain. First, the domain could be removed by differential mRNA splicing or by specific proteolysis. The human MZF-2 gene has been found to generate two different mRNA transcripts through the alternative use of two transcription initiation sites (29, 30). The longer form contains the SCAN domain, while the shorter form lacks sequence encoding the SCAN domain. Smaller transcript forms of ZNF174, ZNF202, FPM315, ZnF20, and ZnFPH have been detected by Northern blot (references 13, 28, 40, and 44 and unpublished observations), suggesting that these genes may also utilize alternative start sites and/or alternative exon splicing to produce molecules that contain the zinc finger DNA binding regions but lack SCAN domains. Second, the function of the domain could be regulated by covalent modifications such as phosphorylation. All but three of the SCAN domain sequences presented in Fig. 1A contain a potential casein kinase II phosphorylation site [(S/T)XX(D/E)] at residues 69 to 72. Phosphorylation (or dephosphorylation) events may regulate the protein-protein interactions or DNA binding. Third, SCAN domain activity



TABLE 1. Characteristics of SCAN family members<sup>a</sup>

Gene	Chromosome site	Zinc fingers	Function				Expression	
			Activation		Repression		Gonads (testis/ovary)	Other tissues
			YES/NO	Type	YES/NO	KRAB		
ZnF20	3p21	20	?	?	?	Yes	?	Yes
TRFA	3p21	?	?	?	?	Yes	?	Yes
ZNF165	6p21.3	5	?	?	?	No	Yes	No
ZNF192	6p21.3	9	?	?	?	No	Yes	Yes
ZNF193	6p21.3	5	?	?	?	No	Yes	Yes
ZnFPH	11p15.5	4	?	?	?	Yes	Yes	Yes
ZNF202	11q23.3	8	?	?	?	No	Yes	No
ZNF174	16p13.3	3	No	?	Yes	No	Yes	Yes
ZNF213	16p13.3	3	?	?	?	Yes	Yes	Yes
FPM315	16p13.3	9	?	?	?	Yes	Yes	Yes
ZNF191	18q12.2	4	Yes	SCAN	?	No	?	Yes
KIAA0427	?	11	?	?	?	No	?	Yes
SRE-ZBP	?	7	?	?	?	No	?	?
CTfin51	Mouse	7	Yes	Acidic	?	No	Yes	Yes
mMZF-2	Mouse	13	Yes	Acidic	?	No	?	Yes
Zfp94	Mouse	7	?	?	?	No	?	?
Zfp95	Mouse	12	?	?	?	Yes	?	?
Zfp96	Mouse	7	?	?	?	No	?	?
Zfp-29	Mouse	14	?	?	?	No	Yes	No

<sup>a</sup> Chromosomal location, number of zinc fingers, transcription function, and expression data are presented for all the known SCAN domain containing genes. ?, Not determined or unknown.

could be controlled by interaction with proteins that contain the domain. The amino-terminal location of the SCAN domain may facilitate oligomerization. The ability of the domain to participate in heteromeric interactions with other SCAN family members suggests that it should be possible to map residues that determine the specificity of SCAN domain interactions. A hierarchy of affinities between these domains would allow for selective pairing of members of the family. Different combinations of transcription factor partners could create new molecules with altered binding recognition sites and regulatory functions (21). The regulated association of distinct members of a transcription factor family is well illustrated by the differential dimerization of Myc, Max, and Mad (8). Increased expression of Myc is associated with increased cellular proliferation, while in the absence of Myc, Max-Max homodimers and Max-Mad heterodimers repress transcription from growth-regulatory genes. Another example of this type of regulation involves MyoD, a basic DNA-binding domain, helix-loop-helix transcriptional activator of numerous muscle specific genes (34). MyoD function is controlled by Id which, like MyoD, contains a helix-loop-helix motif, but which lacks the basic domain required for DNA binding. During muscle cell differentiation, levels of Id fall, releasing another helix-loop-helix protein (E12 or E47) which dimerizes with MyoD and activates gene expression. Fourth, SCAN domain activity could also be controlled by interaction with proteins that lack the motif. These examples illustrate the diverse mechanisms that might regulate the transcriptional activity of members of the SCAN family of transcription factors. In summary, the current studies provide the first description of the function of the SCAN domain and may provide insights into the interactions among the members of this new group of transcription factors.

#### ACKNOWLEDGMENTS

We thank Michelle Spotnitz for assistance with the calculation of the phylogenetic tree.

These studies were supported by NIH grants R37 HL35716 and HL61001. S.C.B. is a Pew Scholar in the Biomedical Sciences.

#### REFERENCES

- Ahmad, K. F., C. K. Engel, and G. G. Prive. 1998. Crystal structure of the BTB domain from PLZF. *Proc. Natl. Acad. Sci. USA* **95**:12123–12128.
- Attar, R. M., and M. Z. Gilman. 1992. Expression cloning of a novel zinc finger protein that binds to the *c-fos* serum response element. *Mol. Cell. Biol.* **12**:2432–2443.
- Bardwell, V. J., and R. Treisman. 1994. The POZ domain: a conserved protein-protein interaction motif. *Genes Dev.* **8**:1664–1677.
- Bellefroid, E. J., D. A. Poncellet, P. J. Lecocq, O. Revelant, and J. A. Martial. 1991. The evolutionarily conserved Kruppel-associated box domain defines a subfamily of eukaryotic multifingered proteins. *Proc. Natl. Acad. Sci. USA* **88**:3608–3612.
- Blacklow, S. C., M. Lu, and P. S. Kim. 1995. A trimeric subdomain of the simian immunodeficiency virus envelope glycoprotein. *Biochemistry* **34**:14955–14962.
- Castillo, G., R. P. Brun, J. K. Rosenfield, S. Hauser, C. W. Park, A. E. Troy, M. E. Wright, and B. M. Spiegelman. 1999. An adipogenic cofactor bound by the differentiation domain of PPAR $\gamma$ . *EMBO J.* **18**:3676–3687.
- Chen, Z., N. J. Brand, A. Chen, S. J. Chen, J. H. Tong, Z. Y. Wang, S. Waxman, and A. Zelent. 1993. Fusion between a novel Kruppel-like zinc finger gene and the retinoic acid receptor alpha locus due to a variant t(11;17) translocation associated with acute promyelocytic leukemia. *EMBO J.* **12**:1161–1167.
- Chowhurdy, K., M. Goulding, C. Walther, K. Imai, and H. Fickenscher. 1992. The ubiquitous transactivator Zfp-38 is upregulated during spermatogenesis with differential transcription. *Mech. Dev.* **39**:129–142.
- Dang, C. V. 1999. *c-Myc* target genes involved in cell growth, apoptosis, and metabolism. *Mol. Cell. Biol.* **19**:1–11.
- Denny, P., and A. Ashworth. 1991. A zinc finger protein-encoding gene expressed in the post-meiotic phase of spermatogenesis. *Gene* **106**:221–227.
- Dong, S., J. Zhu, A. Reid, P. Strutt, F. Guidez, H. J. Zhong, Z. Y. Wang, J. Licht, S. Waxman, C. Chomienne, Z. Chen, A. Zelent, and S. J. Chen. 1996. Amino-terminal protein-protein interaction motif (POZ-domain) is responsible for activities of the promyelocytic leukemia zinc finger-retinoic acid receptor- $\alpha$  fusion protein. *Proc. Natl. Acad. Sci. USA* **93**:3624–3629.
- Estojak, J., R. Brent, and E. A. Golemis. 1995. Correlation of two-hybrid affinity data with *in vitro* measurements. *Mol. Cell. Biol.* **15**:5820–5829.
- Friedman, J. R., W. J. Fredericks, D. E. Jensen, D. W. Speicher, X. P. Huang, E. G. Neilson, and F. J. Rauscher III. 1996. KAP-1, a novel corepressor for the highly conserved KRAB repression domain. *Genes Dev.* **10**:2067–2078.
- Gonsky, R., J. A. Knauf, R. Elisei, J. W. Wang, S. Su, and J. A. Fagin. 1997. Identification of rapid turnover transcripts overexpressed in thyroid tumors and thyroid cancer cell lines: use of a targeted differential RNA display method to select for mRNA subsets. *Nucleic Acids Res.* **25**:3823–3831.
- Guidez, F., S. Ivins, J. Zhu, M. Soderstrom, S. Waxman, and A. Zelent. 1998. Reduced retinoic acid-sensitivities of nuclear receptor corepressor binding to PML- and PLZF-RAR $\alpha$  underlie molecular pathogenesis and treat-

- ment of acute promyelocytic leukemia. *Blood* **91**:2634–2642.
15. **He, L. Z., F. Guidez, C. Tribioli, D. Peruzzi, M. Ruthardt, A. Zelent, and P. P. Pandolfi.** 1998. Distinct interactions of PML-RARalpha and PLZF-RARalpha with co-repressors determine differential responses to RA in APL. *Nat. Genet.* **18**:126–135.
  16. **Hong, S. H., G. David, C. W. Wong, A. Dejean, and M. L. Privalsky.** 1997. SMRT corepressor interacts with PLZF and with the PML-retinoic acid receptor alpha (RARalpha) and PLZF-RARalpha oncoproteins associated with acute promyelocytic leukemia. *Proc. Natl. Acad. Sci. USA* **94**:9028–9033.
  17. **Hoovers, J. M., M. Mannens, R. John, J. Blik, V. van Heyningen, D. J. Porteous, N. J. Leschot, A. Westerveld, and P. F. Little.** 1992. High-resolution localization of 69 potential human zinc finger protein genes: a number are clustered. *Genomics* **2**:254–263.
  18. **Jaenicke, R., and R. Rudolph.** 1986. Refolding and association of oligomeric proteins. *Methods Enzymol.* **131**:218–250.
  19. **Kim, S. S., Y. M. Chen, E. O'Leary, R. Witzgall, M. Vidal, and J. V. Bonventre.** 1996. A novel member of the RING finger family, KRIP-1, associates with the KRAB-A transcriptional repressor domain of zinc finger proteins. *Proc. Natl. Acad. Sci. USA* **93**:15299–15304.
  20. **Klug, A., and J. W. R. Schwabe.** 1995. Zinc fingers. *FASEB J.* **9**:597–604.
  21. **Lamb, P., and S. L. McKnight.** 1991. Diversity and specificity in transcriptional regulation: the benefits of heterotypic dimerization. *Trends Biochem. Sci.* **16**:417–422.
  22. **Lee, P. L., T. Gelbart, C. West, M. Adams, R. Blackstone, and E. Beutler.** 1997. Three genes encoding zinc finger proteins on human chromosome 6p21.3: Members of a new subclass of the kruppel gene family containing the conserved SCAN box domain. *Genomics* **43**:191–201.
  23. **Li, X., J. M. Lopez-Guisa, N. Ninan, E. J. Weiner, F. J. Rauscher III, and R. Marmorstein.** 1997. Overexpression, purification, characterization, and crystallization of the BTB/POZ domain from the PLZF oncoprotein. *J. Biol. Chem.* **272**:27324–27329.
  24. **Lin, R. J., L. Nagy, S. Inoue, W. Shao, W. H. Miller, Jr., and R. M. Evans.** 1998. Role of the histone deacetylase complex in acute promyelocytic leukaemia. *Nature* **391**:811–814.
  25. **Lu, M., S. C. Blacklow, and P. S. Kim.** 1995. A trimeric structural domain of the human immunodeficiency virus type-1 transmembrane glycoprotein. *Nat. Struct. Biol.* **2**:1075–1082.
  26. **Margolin, J. F., J. R. Friedman, W. K. Meyer, H. Vissing, H. J. Theisen, and F. J. Rauscher III.** 1994. Kruppel-associated boxes are potent transcriptional repression domains. *Proc. Natl. Acad. Sci. USA* **91**:4509–4513.
  27. **McKnight, S. L.** 1996. Transcription revisited: a commentary on the 1995 Cold Spring Harbor Laboratory meeting. *Mechanisms of eukaryotic transcription.* *Genes Dev.* **10**:367–381.
  28. **Monaco, C., M. H. Citterich, E. Caprini, I. Vorechovsky, G. Russo, C. M. Croce, G. Barbanti-Brodano, and M. Negrini.** 1998. Molecular cloning and characterization of ZNF202: A new gene at 11q23.3 encoding testis-specific zinc finger proteins. *Genomics* **52**:358–362.
  29. **Murai, K., H. Murakami, and S. Nagata.** 1997. A novel form of the myeloid-specific zinc finger protein (MZF-2). *Genes Cells* **2**:581–591.
  30. **Murai, K., H. Murakami, and S. Nagata.** 1998. Myeloid specific transcriptional activation by murine myeloid zinc finger protein 2. *Proc. Natl. Acad. Sci. USA* **95**:3461–3466.
  31. **Noce, T., Y. Fujiwara, M. Sezaki, H. Fujimoto, and T. Higashinakagawa.** 1992. Expression of a mouse zinc finger protein in both spermatocytes and oocytes during meiosis. *Dev. Biol.* **153**:356–367.
  32. **Pengue, G., V. Calabro, P. Cannada-Bartoli, A. Pagliuca, and L. Lania.** 1994. Repression of transcriptional activity at a distance by the evolutionarily conserved KRAB domain present in a subfamily of zinc finger proteins. *Nucleic Acids Res.* **22**:2908–2914.
  33. **Peteranderl, R., M. Rabenstein, Y.-K. Shin, C. W. Liu, D. E. Wemmer, D. S. King, and H. C. M. Nelson.** 1999. Biochemical and biophysical characterization of the trimerization domain from the heat shock transcription factor. *Biochemistry* **38**:3558–3569.
  - 33a. **Predict Protein program.** 1999, copyright date. [Online.] <http://dodo.cpmc.columbia.edu/predictprotein>. [6 May 1999, last date accessed.]
  34. **Rawls, A., and E. N. Olson.** 1997. MyoD meets its maker. *Cell* **89**:5–8.
  35. **Rost, B., and C. Sander.** 1993. Improved prediction of protein secondary structure by use of sequence profiles and neural networks. *Proc. Natl. Acad. Sci. USA* **90**:7558–7562.
  36. **Sambrook, J., E. F. Fritsch, and T. Maniatis.** 1989. *Molecular cloning: a laboratory manual*, 2nd ed. Cold Spring Harbor Laboratory Press, Plainview, N.Y.
  37. **Tjian, R., and T. Maniatis.** 1994. Transcriptional activation: a complex puzzle with few easy pieces. *Cell* **77**:5–8.
  38. **Tirosvoutis, K. N., A. Divane, M. Jones, and N. A. Affara.** 1995. Characterization of a novel zinc finger gene (ZNF165) mapping to 6p21 that is expressed specifically in testis. *Genomics* **28**:485–490.
  39. **Vissing, H., W. K. Meyer, L. Aagard, N. Tommerup, and H. J. Thiesen.** 1995. Repression of transcriptional activity by heterologous KRAB domains present in zinc finger proteins. *FEBS Lett.* **369**:153–157.
  40. **Williams, A. J., L. M. Khachigian, and T. Collins.** 1995. Isolation and characterization of a novel zinc-finger protein with transcriptional repressor activity. *J. Biol. Chem.* **270**:22143–22152.
  41. **Witzgall, R., E. O'Leary, A. Leaf, D. Onaldi, and J. V. Bonventre.** 1994. The Kruppel-associated box-A (KRAB-A) domain of zinc finger proteins mediates transcriptional repression. *Proc. Natl. Acad. Sci. USA* **91**:4514–4518.
  42. **Yang, M., Z. Wu, and S. Fields.** 1995. Protein-peptide interactions analyzed with the yeast two-hybrid system. *Nucleic Acids Res.* **23**:1152–1156.
  - 42a. **Yang, X. W., C. Wynder, M. L. Doughty, and N. Heintz.** 1999. BAC-mediated gene-dosage analysis reveals a role for Zipr1 (Ru49/Zfp38) in progenitor cell proliferation in cerebellum and skin. *Nat. Genet.* **22**:327–335.
  43. **Yang, X. W., R. Zhong, and N. Heintz.** 1996. Granule cell specification in the developing mouse brain as defined by expression of the zinc finger transcription factor RU49. *Development* **122**:555–566.
  44. **Yokoyama, M., M. Nakamura, K. Okubo, K. Matsubara, Y. Nishi, T. Matsumoto, and A. Fukushima.** 1997. Isolation of a cDNA encoding a widely expressed novel zinc finger protein with the LeR and KRAB-A domains. *Biochim. Biophys. Acta* **1353**:13–17.