A Common DNA-binding Site for SZF1 and the BRCA1-associated Zinc Finger Protein, ZBRK1

Hongzhuang Peng, Lei Zheng, Wen-Hwa Lee, John J. Rux, and Frank J. Rauscher III

The Wistar Institute, Philadelphia, Pennsylvania 19104 [H. P., J. J. R., F. J. R.], and Department of Molecular Medicine and Institute of Biotechnology, University of Texas Health Science Center at San Antonio, San Antonio, Texas 78245 [J. J. R., W. H. L.]

ABSTRACT

More than 220 Kruppel-associated box-zinc finger protein (KRAB-ZFP) genes are encoded in the human genome. KRAB-ZFPs function as transcriptional repressors by binding DNA through their tandem zinc finger motifs. Gene silencing is mediated by the highly conserved KRAB domain, which recruits histone deacetylase complexes, histone methylases, and heterochromatin proteins. However, little is known of the biological programs regulated by KRAB-ZFPs, in large part because of the difficulty in identifying DNA-binding sites recognized by long arrays of zinc fingers. In an attempt to identify the natural target genes for a KRAB-ZFP, we chose SZF1, a hematopoietic progenitor-restricted, KRAB-ZFP that contains only four C2H2 zinc finger motifs. Using recombinant SZF1 protein and a PCR-based binding site selection strategy, we identified a 15-bp consensus DNA sequence recognized by SZF1. Remarkably, this sequence is similar to the core DNA-binding site described recently for ZBRK1, a KRAB-ZFP that binds to BRCA1 and is involved in coordinating the cellular DNA damage response. The SZF1 and ZBRK1 proteins bind to both the experimentally derived SZF1 site and the canonical ZBRK1 site. The KRAB domain from SZF1 bound directly to the Kap-1 corepressor and displayed intrinsic silencing activity. Moreover, full-length SZF1 repressed a promoter containing ZBRK1 recognition sequences. Thus, SZF1 and ZBRK1 may regulate a common set of target genes in vivo.

INTRODUCTION

The ~220 KRAB-ZFP genes in the human genome have been proposed to have important regulatory roles during cell differentiation and development. Each KRAB-ZFP is composed of a 75-amino acid KRAB domain at the NH2 terminus and tandem C2H2 class zinc fingers at the COOH terminus (Fig. 1). The tandem zinc finger modules can number from 3 to upwards of 40 in a single protein and presumably participate in DNA recognition. The KRAB domain was originally identified as a conserved motif at the NH2 terminus of ZFPs (1) and was shown to be a potent, DNA binding-dependent transcriptional repression module (2–4). The KRAB domain only occurs in higher vertebrates and can be classified into three subfamilies based on amino acid sequence alignment (5, 6): subfamilies containing a KRAB A box alone, both A and B boxes, or A box with a divergent B box. The KRAB domain consists of ~75 amino acid residues and is predicted to fold into two amphipathic helices that are involved in protein/protein interactions (Fig. 1B; Refs. 1, 7). Silencing mediated by the KRAB domain occurs through recruitment of the corepressor protein, Kap-1 (8–10), which in turn recruits the NURD histone deacetylase complex, histone methylases, and the HP1 family of heterochromatin proteins (11–13). Thus, although much is known of the KRAB silencing mechanism, comparatively little is known about KRAB-ZFP function in the organism.

The spatial and temporal expression of KRAB-containing zinc finger transcription factors suggests that their biological functions could include influences on embryonic development, cell differentiation, and cell transformation (1, 14–18). For example, some KRAB-containing ZFPs are mainly restricted to lymphoid cells and may play a specific role in lymphoid differentiation (14, 19), whereas others are expressed and specifically down-regulated during myeloid differentiation (1). A number of KRAB-ZFPs are candidate genes for human diseases based on their chromosomal locations (20, 21). There are more than 40 KRAB-ZFP-encoding genes that have been identified on chromosome 19p13 and >10 KRAB-ZFP genes clustered on chromosome 19q13 (22, 23), many of which exhibit hematopoietic-specific expression (24). Intriguingly, some of the KRAB-ZFPs in these clusters are coordinately regulated in specific cell lineages (24).

The tandem C2H2 zinc fingers in KRAB-ZFPs are presumed to recognize specific DNA targets. The analyses of two crystal structures of the C2H2 zinc finger proteins, Zif268/Egrl and GLI-1, with their cognate DNA-binding sites have provided some rules about DNA recognition by zinc finger domains (25–28). Each finger has a conserved βα structure, and amino acids on the surface of the α-helix contact bases of the DNA. Each finger recognizes a three-nucleotide sequence along the major groove of the DNA helix. In the case of GLI-1, not all of the five zinc finger domains contact the DNA, suggesting that not all zinc finger domains contribute equally to DNA recognition. The linker region that separates neighboring C2H2 zinc fingers is usually of the form TGEKPYX (X representing any amino acid), is an important structural element that helps control the spacing of the fingers along the DNA site, and is required for high-affinity DNA binding (28–31).

Thus, the crystal structures of 3- and 5-fingered proteins have clearly provided a preliminary set of rules for zinc finger DNA interaction. However, the great challenge lies in determining whether these rules will also govern the binding of long array zinc finger arrays, e.g., 10–20 tandem units. Only a few target DNA consensus sequences have been identified for the long array KRAB-ZFP family, and these already suggest that the rules may be different. For instance, the 8-fingered ZBRK1 protein binds a consensus of GGGxxx-CAGxxxTTT (32). The 8-fingered ZNF202 protein binds a consensus of GGGGT (17), and the 10-fingered KS1 protein binds the 27-bp consensus TCCTACAGTGCAACCCTACAGTAA (33). These results suggest that: (a) not all fingers bind DNA; and (b) not all fingers contribute specificity for sequence recognition.

To simplify the problem of identifying target sequences (and promoters) regulated by KRAB-ZFPs, we have chosen a short array, lineage-restricted KRAB-ZFP, SZF1. SZF1 is a KRAB-zinc finger gene expressed predominantly in CD34+ stem/progenitor cells. The gene was isolated by screening a cDNA library prepared from human
bone marrow CD34<sup>+</sup> cells (16). It encodes a protein containing a canonical KRAB domain at the NH<sub>2</sub> terminus and four contiguous zinc fingers of the C<sub>2</sub>H<sub>2</sub> type at the COOH terminus. Two alternatively spliced transcripts were isolated; SZF1-1 is a truncated form resulting in an apparently incomplete fourth zinc finger (Fig. 1C) and is expressed only in CD34<sup>+</sup> cells; SZF1-2 encodes a consensus fourth zinc finger, a novel COOH terminus, and is ubiquitously expressed. Both forms showed the potential for transcriptional repression of a CD34<sup>+</sup>-specific promoter (16). Our experimental strategy was to define the DNA site recognized by SZF1 zinc fingers as a means to identify potential target genes regulated in CD34<sup>+</sup> hematopoietic cells.

**MATERIALS AND METHODS**

**Construction of Expression Plasmids.** The bacterial expression plasmid for His-SZF1-1-ZF (amino acids 220–348) was created by PCR using the primers SZF1-1-ZFfor (BamHI) 5′-CGC AGG ATC CGC TTT TAA CCA GAA GTC AAA C-3′ and SZF1-1-ZFrev (SalI) 5′-CAC TGG TCG ACT CAC TGG TGC TTA ATG AGC TCT GAC-3′. The resulting PCR products were digested with the restriction endonucleases indicated in brackets and cloned into the corresponding sites of the pQE30 vector (Qiagen) for His-SZF1-1-ZF.

---

*Fig. 1.* A, a diagram illustrating the architecture of the KRAB-ZFPs, including SZF1-1, ZBRK1, and KOX1. The numbers represent amino acid positions. ZF or Zn refers to a zinc finger motif. *Column to the right,* expressed proteins. Database accession numbers are: SZF1-1 (AF114816); ZBRK1 (AF295096); and KOX1 (XP_031850). B, amino acid alignment of the KRAB domains from SZF1, ZBRK1, and KOX1 proteins. The numbers refer to amino acid positions in the corresponding residues. The KRAB domain consensus residues are highlighted in black. The periods represent spaces introduced to obtain maximal alignment. The consensus sequence was previously derived (35). C, amino acid alignment of the zinc finger motifs from SZF1-1 and ZBRK1 proteins. The numbers on the left of the alignment refer to the orders of the zinc finger motifs within each protein. The numbers (−1, +3, and +6) under the alignment refer to the position of the amino acids within the α-helical region of each zinc finger. The consensus Cys<sub>2</sub>-His<sub>2</sub>-Cys<sub>2</sub>-His<sub>2</sub> residues of each zinc finger are highlighted in black. The other consensus residues of zinc fingers and linker regions are highlighted in gray. The identity and similarity of zinc finger and linker regions between SZF1-1 and ZBRK1 are indicated under the alignment. D, homology model of the first zinc finger domain of SZF1-1 (yellow) bound to DNA (red) highlighting the −1, +3, and +6 positions of the zinc finger (green). The model was prepared with the Swiss-PdbViewer protein-modeling environment (40) and is based upon the closely related Zif268-DNA complex structure (26). E, the phylogenetic tree of zinc finger motifs of SZF1-1 and ZBRK1 proteins. The unrooted tree was calculated with the program CLUSTALX (45, 46) using the distance algorithm of Kimura (47) and displayed with Tree View (48). Branch lengths are proportional to the number of substitutions, and the scale bar represents 10 mutations per 100 sequence positions.
the pGEX-4T-1 vector (Amersham Pharmacia Biotech) for GST-SZF1-KRAB, and the pM2 vector for GAL4-SZF1-KRAB. All plasmids generated by PCR were confirmed by DNA sequencing on both strands to verify the appropriate reading frame and the integrity of flanking sequences. The pNeoSZF1-1 (amino acids 1–421) expression plasmid was kindly provided by Donald Small (The Johns Hopkins Oncology Center, The Johns Hopkins University School of Medicine, Baltimore, MD). The pCHP-ΔZBRK1 (amino acids 1–532) and the GST-ΔZBRK1-1ZF (amino acids 144–439) expression plasmids have been described previously (32). The bacterial expression plasmid for His-ZBRK1-1ZF (amino acids 144–439) was created by PCR using the primers ZBRK1-1ZFfor (BamHI) 5'-CCA CGA GGA TCC AAC CAG AGC AAA GGC TAT GAA A-3' and ZBRK1-1ZFrev (HindIII) 5'-CCA CGA AAG CTT TCA GGC AGG ATT TTC CAC CTT-3'. The GST-KOX1-1KRAB (amino acids 1–144) and the GAL4-KOX1-KRAB (amino acids 1–90) protein complexes were described previously (2, 8).

Purification of Recombinant Proteins. Escherichia coli SG13009 cells (Qigene) bearing the desired plasmid were propagated with agitation at 37°C in 1 liter of Luria broth to an A600 of ~0.6. Isopropyl-1-thio-β-D-galactopyranoside was added to 1 mM, and ZnSO₄ was added to 100 μM. The growth at 18°C was continued for 20 h. The cells were harvested by centrifugation. The His-SZF1-1ZF and His-ZBRK1-1ZF proteins expressed from bacteria were purified at 4°C using nondenaturing conditions as recommended by the manufacturer (Qigene). Briefly, the bacterial pellet was resuspended in P300 buffer (pH 7.0), 10% (v/v) glycerol, 10 mM imidazole, and lysed by sonication (34). The cell extract was centrifuged at 12,000 × g for 20 min. The supernatant fraction containing soluble protein was incubated in batch with Ni-NTA resin for 1 h. The resin was washed four times with P300 buffer, 10% glycerol, and for the last wash, the imidazole was increased to 40 mM. The resin was loaded into a column, and the protein was eluted with P300 buffer, 10% glycerol, and 300 mM imidazole. The eluted protein was dialyzed against three changes of P300, 10% glycerol, 20 μM ZnSO₄, and 0.5 mM DTT.

GST Association Assays. The preparation of the GST fusion proteins and the GST association assays were performed essentially as described previously (35, 36). Briefly, 5 μM of freshly prepared GST fusion protein immobilized on glutathione-Sepharose were incubated with 5 μM of Ni-NTA-purified recombinant His-tagged protein in 100 μl of BB500 buffer [20 mM Tris (pH 7.9), 500 mM NaCl, 0.2 mM EDTA, 10% glycerol, 0.2% NP40, 1 mM phenylmethylsulfonyl fluoride, 500 μg of BSA (BSA; fraction V)] for 1 h at room temperature. The protein complexes were washed four times with BB750 [20 mM Tris (pH 7.9), 750 mM NaCl, 0.2 mM EDTA, 10% glycerol, 0.2% NP40, and 1 mM phenylmethylsulfonyl fluoride], and the bound proteins were eluted in 5× Laemmli buffer, resolved by SDS-PAGE, and visualized with Coomassie Blue stain.

Generation of Oligonucleotide Library and Binding Site Selection. A single-stranded oligonucleotide 5′-AGACGAGTCCATTGGA-CGGATACCGA-3′ bearing 15- and 16-nucleotide fixed-end sequences flanking 14 random nucleotides was synthesized. Recognition sequences for the BamHI and EcoRI restriction enzymes (underlined) were incorporated to facilitate the cloning of the binding sites. The oligonucleotides were amplified by PCR using primers complementary to the fixed-end sequences. The resulting double-stranded DNA library was purified by electrophoresis on a 12% native polyacrylamide gel, followed by elution and precipitation. The purified oligonucleotide library was end-labeled with [32P]ATP and incubated with the purified recombinant SZF1-1-ZF protein. The selection and amplification of the binding site was performed as described previously (37). The DNA-protein binding reactions were performed in a buffer containing 20 mM HEPES (pH 7.6), 50 mM NaCl, 50 μM ZnSO₄, 0.5 mM MgCl₂, 0.5 mM DTT, 10% glycerol, 0.5 μg deoxyinosinic-deoxyxycytidic acid in a total volume of 15 μl (38).

DNA Sequencing and Analysis of Selected Oligonucleotides. After three rounds of EMSA selection and PCR amplification, the affinity-selected oligonucleotides were extensively digested with BamHI and EcoRI and ligated into the pGEMTZI+ plasmid (Promega) at the corresponding sites. After bacterial transformation and clone selection, the DNAs were sequenced. The degenerate positions of the oligonucleotide were aligned using the GCG sequence analysis software package (39).

EMSA and Binding Site Competitions. Increasing amounts of purified SZF1-1-ZF and ZBRK1-1ZF proteins in binding buffer were incubated with a 32P-end-labeled synthetic oligonucleotide containing the binding sites for either the wild-type or the mutant probe for 5 h. The wild-type probes included the SZF1-1-ZF (designated as SZF1 probe, 5'-GATCCCAGGGTAGA-CAGCGTTTT-3') and the ZBRK1-ZF (designated as ZBRK1 probe, 5'-GATCCAAGGGAGCCAGTTTTGTTGCGCG-3'). Three sets of mutant SZF1 oligonucleotides were synthesized. The triplet GGG was mutated to CTC in the first set (designated as SZF1-mut1 probe, 5'-GATCCCACTCTAA-CACCGCGTTTG-3'). The triplet CAG was mutated to TTC in the second set (designated as SZF1-mut2 probe, 5'-GATCCCAAGGGATATCCCGTTTG-3'). Both triplets GGG and CAG were mutated to AAA in the third set (designated as SZF1-mut3 probe, 5'-GATCCCAATTTAAAGGGCGTTTG-3'). In competition assays, 1 μg of SZF1-1-ZF or 400 ng of ZBRK1-1ZF protein were incubated with a mixture of 32P-end-labeled oligonucleotide and a 25-100-fold molar excess of unlabeled oligonucleotide for 15 min at 30°C. The DNA-protein complexes were analyzed on a native polyacrylamide gel in 0.5× TBE running buffer. The gels were dried and visualized by autoradiography.

Transient Transfection Luciferase Assays. Transient transfection and luciferase assays were performed with NIH/3T3 cells as described previously (36). Briefly, NIH/3T3 cells were transiently cotransfected with a luciferase reporter plasmid, a pcDNA3-β-galactosidase expression plasmid, and effecter plasmids for 5 h. For reporter assays with GAL4 fusion proteins, the luciferase reporter plasmids consisted of either 0 or 5 copies of the GAL4 UAS in front of a minimal herpes simplex virus TK promoter. Reporter assays with ZBRK1 and SZF1 used the pGL3p-E reporter plasmid, which contains four copies of the ZBRK1 consensus binding site upstream of the SV40 promoter driving expression of a luciferase gene as described previously (32). At 18 h after transfection, the cells were collected and assayed for luciferase activity using the Luciferase Assay System (Promega), and values were normalized for transfection efficiency using β-galactosidase activity.

RESULTS

Domain Analysis for KRAB-ZFPs and Purification of Recombinant Proteins. KRAB-ZFP genes are very abundant in the human genome. However, only a few detailed biochemical and biophysical analyses of KRAB-ZFP function have been undertaken (34, 35). The potential target sequences have been identified for only a few members of the KRAB-ZFP family (17, 32, 33). We began our studies by aligning and analyzing the protein domains for the members of the KRAB-ZFPs, including SZF1, ZBRK1, and KOX1 (Fig. 1). As expected, the KRAB domain shows remarkable homology among the members (Fig. 1B). In KRAB-ZFPs, the zinc fingers conform to a “classic” C₃Hₓ zinc finger motif first described in the Kruppel protein in Drosophila. SZF1-1 encodes four consecutive zinc finger motifs, whereas ZBRK1 encodes eight consecutive zinc finger motifs (Fig. 1C). Both proteins contain the canonical linker region between fingers composed of the sequence TGKEPKYX. Because the zinc finger regions are the determinant for DNA-binding specificity (Fig. 1D; Ref. 40), comparisons of the SZF1-1 and ZBRK1 zinc finger region sequences were made. As expected, the zinc finger motifs are closely related; comparative analyses indicate that SZF1-1 displays 57% identity and 72% similarity with ZBRK1 over the zinc finger motifs (Fig. 1C). A distance-based phylogenetic analysis of zinc finger sequences was performed (Fig. 1E). That the sequences are all closely related is evident from the similarity of the distances between the sequences. However, an examination of the branching pattern reveals that the sequences cluster into three groups of four sequences each. In addition, all four SZF1-1 zinc fingers completely segregate from those of ZBRK1, indicating that they are more similar to one another than they are to those of ZBRK1. Finally, it is also interesting to note that in both SZF1-1 and ZBRK1, the first and third zinc finger domains cluster together, whereas second finger is always more divergent. The degree to which this pattern relates to the observed overlap in DNA-binding specificity of these proteins is not yet known and will require further study.

For the binding site selection experiments, recombinant protein
containing four zinc fingers of SZF1-1 was produced by cloning the DNA encoding the four zinc fingers (amino acids 220–348) into a protein expression vector (designated as SZF1-1-ZF; Fig. 1A). The recombinant SZF1-1-ZF protein was expressed in Escherichia coli as a 6-histidine fusion protein and purified by Ni-NTA chromatography under non-denaturing conditions. SDS-PAGE analysis showed that the protein was purified to near homogeneity (Fig. 2A). Recombinant protein containing eight zinc fingers of ZBRK1 was produced by cloning the DNA encoding amino acids 144–439 into a GST fusion protein expression vector (designated as GST-ZBRK1-ZF; Fig. 1A). The recombinant GST-ZBRK1-ZF protein was expressed in E. coli, purified by GST chromatography, and examined by SDS-PAGE (Fig. 2A). These purified SZF1-1-ZF and GST-ZBRK1-ZF proteins were used in binding site selection experiments.

Identification of a Consensus SZF1-1 DNA Binding Sequence. On the basis of the fact that one zinc finger module has the capability of binding to 3 bp of DNA (26), the four zinc fingers of SZF1-1 would be predicted to bind to a minimum of 12 bp of DNA. Therefore, a library of double-stranded DNA was designed to contain 14 bp of random sequence, flanked by sequences containing restriction enzyme sites and which were complementary to the PCR primers. Purified, recombinant SZF1-1-ZF protein was then used to affinity-select DNA sequences from the library by using successive rounds of EMSA. At each step, the bound oligonucleotides were recovered and subjected to PCR amplification. The first round of affinity selection with SZF1-1-ZF protein produced a significant enrichment of sequences in the heterogeneous oligonucleotide library that bound to the protein (Fig. 2B). However, the second and the third rounds of selection produced little further enrichment of DNA-binding activity. Therefore, after three rounds of selection, the affinity-selected oligonucleotides were subcloned and sequenced. A consensus sequence of 5′-CCAGGG-TAACACCGC-3′ was derived from sequence alignment of individual subclones (shown in Fig. 2C) are 12 sequences of 80 individual subclones. Surprisingly, the selected consensus sequence for SZF1-1 is very similar to the consensus binding site recently described for ZBRK1 (5′-CACGGGAGCGGTTTTG-3′; Fig. 2D). Both consensus sequences share the core binding site, 5′-GGGXX-CAGXX-3′. By comparison of the sequence selected from SZF1-1-ZF protein binding and the reported STK-140 promoter sequence (16), we found that the STK-140 promoter has only a partial SZF1-1 consensus sequence (CAGGGXXXXGCG-3′). This finding is consistent with the fact that SZF1-1 has a weaker repression activity with the STK-140 promoter reporter than SZF1-2. Consideration that SZF1-1 has more activity on the ZBRK-like site suggests that the alternate splicing that generates SZF1-1 may select for a different set of target genes, although this has not been thoroughly examined.

To determine whether SZF1-1 and ZBRK1 proteins specifically bind to both consensus binding site DNA sequences selected from both proteins, oligonucleotides that contain the selected consensus binding sites for SZF1-1-ZF and ZBRK1-ZF proteins were synthesized (Fig. 3A) and used as probes in the EMSA analysis. Because the SZF1 site is a subset of the site required for ZBRK1 binding, we added the additional three nucleotides, TTT to the 3′ end of the core SZF1-1 consensus. The SZF1-1-ZF protein was tested by EMSA for its ability to bind to the SZF1 probe (5′-GATCCCCAGGTAA-CAGCCGTTTG-3′) and to the ZBRK1 probe (5′-GATCCAGGG-GACGAGGTGTTTTGTTCCG-3′). As shown in Fig. 3B, left panel, the SZF1-1-ZF protein bound to both consensus probes to form protein-DNA complexes and yields the mobility shifts indicated by the arrow. The binding ability of SZF1-1-ZF protein to both probes is comparable and occurs in a protein concentration-dependent manner (Fig. 3B, left panel). The ZBRK1-ZF also bound to both consensus probes to form DNA-protein complexes (Fig. 3B, right panel), and the apparent binding affinity of the ZBRK1-ZF protein to both of the probes is comparable. Thus, both SZF1-1-ZF and ZBRK1-ZF proteins select similar core DNA binding sites from a heterogeneous population of oligonucleotides.

To address the specificity for the SZF1-1-ZF and ZBRK1-ZF proteins binding to their consensus sequences, we used three sets of mutant SZF1 oligonucleotides to bind to both His-tagged SZF1-1-ZF and ZBRK1-ZF proteins (Fig. 3, A and C). The triplet GGG was mutated to CTC (SZF1-mut1), the triplet CAG was mutated to TTC (SZF1-mut2), and both GGG and CAG mutated to AAA (SZF1-mut3). We observed that these sets of mutant oligonucleotides significantly reduced but did not completely abolish the binding of SZF1-1-ZF protein in the EMSA assay (Fig. 3C, left panel). These data indicate the specificity of the SZF1-1-ZF protein in the recognition of the consensus sequence, the GGG and CAG triplets. However, the SZF1-1-ZF protein may also interact with the bp between the triplet GGG and CAG or other triplets in the sequence because mutations of both GGG and CAG did not completely abolish the binding of SZF1-1-ZF protein. We also observed that the three sets of mutant oligonucleotides completely abolished the binding of ZBRK1-ZF protein in the EMSA assay (Fig. 3C, right panel). The data are consistent with the previous observation in competition EMSA assay.
with a mutant probe using GST-ZBRK1-ZF protein (32). The data indicate that the ZBRK1-ZF protein more specifically recognizes the triplets GGGxXCAxGxXTTT in the sequence.

The specificity of SZF1-1-ZF and ZBRK1-ZF protein binding to their consensus sequence was further revealed by testing the ability of a molar excess (25–100-fold) of unlabeled probes to compete for binding to the labeled probes. The results of these competition EMSA analyses indicated that the SZF1 probe and the ZBRK1 probe, but not the GAL4 probe, could effectively compete for SZF1-1-ZF protein binding to its consensus sequence (Fig. 3D, left panel). A similar result was obtained for the ZBRK1-ZF protein (Fig. 3D, right panel). Therefore, these data suggest that both SZF1-1 zinc fingers and the ZBRK1 zinc fingers comprise sequence-specific DNA-binding domains that specifically recognize similar core consensus DNA sequences.

Direct Interaction between the SZF1-KRAB Domain and KAP-1-RBCC Domain. Previous studies had shown the potential for SZF1 in transcriptional repression of a CD34⁺-specific promoter (16). The KRAB domain of SZF1 shows a high degree of homology with the KRAB domain of other KRAB-ZFPs (Fig. 1B). Thus, it would be expected that the SZF1-KRAB domain would confer transcription repression activity via KAP-1 binding, as demonstrated for other KRAB proteins (8, 34, 35). To confirm this, we analyzed the KRAB domain ofSZF1 using biochemical approaches. We have shown previously that the E. coli-expressed KOX1-KRAB domain was able to directly bind to the RBCC (RING-B box-Coiled-coil) domain of the KAP-1 protein in GST association and in EMSA analyses (34, 35). We used the GST association assay to test the ability of SZF1-KRAB to bind to the purified KAP-1-RBCC protein. Significant binding of the KAP-1-RBCC protein was observed for the GST-SZF1-KRAB protein but was negative for the control GST protein (Fig. 4). The affinity of the SZF1-KRAB protein for binding to the KAP-1-RBCC protein is comparable with that of the KOX1-KRAB protein, supporting our previous results demonstrating that the interaction between the KRAB domain and the RBCC domain of KAP-1 is direct and specific.

Assessment of SZF1-KRAB Transcriptional Regulation with GAL4 Reporter Assays. Previous studies demonstrate that the KRAB domains from other zinc finger proteins repressed transcription when fused to a heterologous DNA-binding domain (2, 4, 41). To assess the transcriptional regulatory properties of SZF1-KRAB, the KRAB domain of SZF1 was fused to a heterologous GAL4 DNA-binding domain. The GAL4-SZF1-KRAB expression plasmid was cotransfected with a 5xGAL4-TK-Luc reporter plasmid into NIH/3T3 cells. As a positive control for transcriptional repression, the GAL4-KOX1-KRAB protein strongly repressed luciferase activity in a dose-dependent manner (Fig. 5B). The GAL4-SZF1-KRAB protein also strongly repressed luciferase activity (17-fold over vector alone, at the highest input level
of GAL4-SZF1-KRAB tested; Fig. 5B). On the basis of the similar levels of expression in mammalian cells (data not shown), the repression activity by GAL4-SZF1-KRAB is comparable with the repression activity by GAL4-KOX1-KRAB.

**SZF1 Is a Sequence-specific Transcriptional Repressor.** The studies of the GAL4-SZF1-KRAB chimera illustrate the transcriptional repression activity of the SZF1-KRAB domain, and the EMSA analysis indicated that the SZF1-1-ZF and the ZBRK1-ZF proteins specifically bind to similar consensus sequences. To evaluate the ability of SZF1-1 to regulate transcription according to its intrinsic DNA-binding activity, the SZF1-1 expression vector was cotransfected with the pGL3p-E luciferase reporter plasmid into NIH/3T3 cells. This luciferase reporter plasmid contains four copies of the ZBRK1 consensus binding site upstream of the SV40 promoter driving expression of a luciferase reporter gene relative to that obtained with the indicated expression plasmid. All activities were normalized for transfection efficiency based on ß-galactosidase activity. Bars, SD for three independent transfections.

**DISCUSSION**

A mandatory prerequisite for understanding the function of a putative transcription factor (or family of transcription factors) is to identify the DNA sequence recognized by the cognate DNA-binding domain in the protein. This key discovery allows application of a myriad of experimental approaches, including the ability to predict target genes and pathways, to detect binding activity in complex mixtures and to test the potential of the protein as an activator or repressor of gene expression in vivo. This principle is well illustrated by the studies of the MYC-MAX, FOS-JUN, and HOMEODOMAIN protein families wherein identification of core DNA-binding consensus sequences was the key catalyst that allowed full characterization of their functions. A paradigm that emerged from these studies is that often multiple members of families that contain a common type of DNA-binding domain will recognize a common DNA-binding site. For example, the FOS-JUN-CREB family of bZIP proteins has the capability of forming >100 biologically distinct complexes via combinatorial heterodimerization; yet, all of them bind to subtle variants of the AP1 site. A great challenge has been to determine whether this paradigm will extend to other classes of DNA-binding proteins.
We have been studying the C2H2 zinc finger class of DNA-binding proteins. The C2H2 finger is the second most abundant motif found in the human genome (6, 42). This 24–30 amino acid motif, which contains tandem cysteine and histidine residues, functions by chelating a zinc atom resulting in a tightly folded modular unit. In most zinc finger proteins, these modules are found arranged in tandem arrays from 3 to upwards of 40 zinc finger units. Crystal structures of 3- and 5-fingered proteins have suggested that each module recognizes ~3 bp of DNA (26, 27). Thus, in a many-fingered protein, there is a potential for tremendous DNA-binding specificity and target affinity. Unfortunately, these predictions have not been thoroughly tested because only a few DNA-binding sites for multifingered proteins have been identified.

To simplify this problem, we have focused on a subclass of C2H2 zinc finger proteins, the KRAB-ZFPs. In this study, we have presented evidence for sequence-specific DNA binding by the KRAB-ZFP, SZF1. We chose SZF1 for a number of reasons: (a) it encodes only four zinc fingers, which should limit the complexity of the DNA recognition sequence; (b) SZF1 is predominantly expressed in CD34+ hematopoietic progenitor cells, which should serve to limit the number of potential target genes; and (c) elucidation of its function may provide evidence for sequence-specific DNA binding by the KRAB-ZFP, ZBRK1. We also provided that each protein will probably use different modes of molecular recognition to accomplish binding the same sequence. Solving these types of comparative molecular recognition problems (via atomic structure determination) will certainly be a great aid in correcting and extending the code hypothesis for zinc finger DNA recognition.

A key finding of this study is that SZF1 and ZBRK1 show a common binding site. By analogy to the known functions of ZBRK1, this finding immediately suggests both potential target genes for SZF1 and physiological functions. ZBRK1 was cloned in a two-hybrid screen for proteins that interact with the BRCA1 protein (32). BRCA1 is the familial breast/ovarian cancer tumor suppressor, and germ-line mutations in the gene account for ~50% of this inherited cancer predisposition syndrome. The BRCA1 protein is a nuclear protein, the abundance, posttranslational modification, and intranuclear location of which is dynamically regulated by many different cellular stimuli. Roles for BRCA1 in direct DNA repair, DNA damage sensing, transcriptional regulation/chromatin remodeling, and DNA replication have emerged (43). Most likely, BRCA1 serves as a scaffolding protein that coordinates these diverse activities in response to DNA damage stimuli. Although BRCA1 is a nuclear protein with gene-specific regulatory effects, it contains no apparent intrinsic DNA-binding activity or recognition domain. The key issue of how BRCA1 is tethered to DNA was partly resolved by the cloning of ZBRK1 as a BRCA1-binding protein. Moreover, the discovery that ZBRK1 bound the consensus sequence 5'-GGGAGGACAGTTT-3' found in many genes regulated by stress and DNA damage further suggested the relevance of ZBRK1 to BRCA1 function. BRCA1 has been shown to function as a corepressor for ZBRK1, and it was suggested that DNA damage-induced phosphorylation could relieve repression, thereby leading to activation of DNA damage response genes (43, 44). Whether the KAP-1 corepressor is involved in the ZBRK1/BRCA1 complex remains to be clarified.

How does SZF1 fit into this model? It is critical to determine whether SZF1 also binds to BRCA1. Although these experiments are ongoing, we think it is highly unlikely: (a) the COOH terminus of zinc fingers of ZBRK1, which contributes to the BRCA1 binding surface, are the most divergent in sequence when compared with the SZF1 fingers; (b) there is no significant homology between the COOH-terminal, non-finger region of ZBRK1 and the COOH terminus of either SZF1-1 or SZF1-2. In ZBRK1, this COOH-terminal extension in addition to the zinc fingers are required for BRCA1 binding (32). Thus, we think it unlikely that there is direct cross-talk between ZBRK1 and SZF1 for BRCA1 binding. A more likely scenario may simply be competition for binding with proteins containing these recognition sequences. Because the SZF1 consensus is shorter than the full ZBRK1 consensus, theoretically SZF1 has the capability for binding every ZBRK1 target gene. It should also be noted that ZBRK1 will also bind the shorter SZF1 consensus, albeit with lower apparent affinity (data not shown). It will be interesting to determine whether the protein with the longer array of fingers (ZBRK1) is able to functionally displace a SZF1 bound to a consensus sequence in vivo.

The current model for ZBRK1 function suggests that the protein is very stable when bound to target genes in vivo and that various cellular stress/DNA damage signals regulate the association of coactivator or corepressor complexes with the bound ZBRK1 protein, thereby regulating gene activation or repression. That BRCA1 is found in both types of regulatory complexes in vivo is consistent with this. However, ZBRK1 isolated from cells appears to be stably associated with both BRCA1 and KAP-1 (the caveat to this result being that there may be separate pools of ZBRK1 in the cells indistinguishable in the assays used). If indeed a ternary KAP-1:ZBRK1:BRCA1 complex is formed, it will be quite interesting to determine how BRCA1 when functioning as a coactivator in this complex can over-
ride gene silencing that is presumably mediated by the KAP-1 corepressor, which is bound to the spatially separate KRAB domain on ZBRK1. One hypothesis is that SZF1 is functioning at ZBRK1 targets as a pure repressor; it does not toggle between activation and repression (similar to ZBRK1), because it probably does not bind BRCA1 (or presumably other coactivator complexes). In this context, the SZF1-ZBRK1 system may be functioning like other binary regulation systems whereby basal (repressed) gene activity is controlled by a DNA-binding protein with less specificity, whereas derepression is controlled by a complex with a DNA-binding subunit with more specificity (i.e., a large number of zinc fingers) that contains coactivators. To resolve this, it will be critical to determine whether SZF1 and ZBRK1 regulate the same or overlapping sets of target genes, experiments that are currently underway using cDNA microarrays.

In summary, this discovery sheds new light on a potential mechanism for regulating target genes under the control of BRCA1.

ACKNOWLEDGMENTS

We thank D. Small for the SZF1 plasmids and W. J. Fredericks, D. C. Schultz, and M. S. Lechner for helpful discussions. We thank the Wistar Institute DNA Core Facility for DNA sequencing.

REFERENCES


A Common DNA-binding Site for SZF1 and the BRCA1-associated Zinc Finger Protein, ZBRK1


Updated version  Access the most recent version of this article at: http://cancerres.aacrjournals.org/content/62/13/3773

Cited articles  This article cites 48 articles, 30 of which you can access for free at: http://cancerres.aacrjournals.org/content/62/13/3773.full#ref-list-1

Citing articles  This article has been cited by 9 HighWire-hosted articles. Access the articles at: http://cancerres.aacrjournals.org/content/62/13/3773.full#related-urls

E-mail alerts  Sign up to receive free email-alerts related to this article or journal.

Reprints and Subscriptions  To order reprints of this article or to subscribe to the journal, contact the AACR Publications Department at pubs@aacr.org.

Permissions  To request permission to re-use all or part of this article, contact the AACR Publications Department at permissions@aacr.org.