Four of the Seven Zinc Fingers of the *Evi-1* Myeloid-Transforming Gene Are Required for Sequence-Specific Binding to GA(C/T)AAGA(T/C)AAGATAA

RUUD DELWEL,^{1,2} TETSUNORI FUNABIKI,^{1†} BRENT L. KREIDER,¹ KAZUHIRO MORISHITA,^{1‡}
AND JAMES N. IHLE^{1,3*}

Department of Biochemistry, St. Jude Children's Research Hospital, Memphis, Tennessee 38105¹; Institute of Hematology, Erasmus University, 3000 DR Rotterdam, The Netherlands²; and Department of Biochemistry, University of Tennessee, Memphis, Tennessee 38163³

Received 2 March 1993/Returned for modification 31 March 1993/Accepted 21 April 1993

Expression of the *Evi-1* gene is activated in murine myeloid leukemias by retroviral insertions and in human acute myelogenous leukemia by translocations and inversions involving chromosome band 3q26 where the gene resides. Aberrant expression of the *Evi-1* gene has been shown to interfere with myeloid differentiation, which is proposed to be the basis for its role in leukemias. The *Evi-1* gene encodes a 145-kDa DNA-binding protein containing two domains of seven and three Cys₂-His₂ zinc fingers. Previous studies identified a portion of the consensus DNA-binding sequence for the first domain of zinc fingers. The experiments presented here extend these studies and demonstrate that the first domain recognizes a consensus of 15 nucleotides consisting of GA(C/T)AAGA(T/C)AAGATAA. The first three fingers of the first domain do not detectably bind DNA but contribute to the binding by conferring a relative specificity for GACAA verses GATAA in the first position. The first three fingers also contribute to optimal binding of the 15-nucleotide consensus sequence.

The *Evi-1* gene was first identified as a gene that is activated following retroviral insertions in the *Evi-1* common site of viral integrations in myeloid leukemias and myeloid leukemia cell lines (1, 2, 17). It has subsequently been shown that aberrant expression of the *EVII* gene occurs in human acute myelogenous leukemias and in myelodysplastic diseases as a consequence of chromosomal translocations or inversions involving chromosome band 3q26 where the human gene resides (16). The gene encodes a 145-kDa DNA-binding protein that contains two domains of zinc fingers (12, 17). The first domain of zinc fingers consists of six contiguous fingers and an amino-terminal finger that is separated from the others by 25 amino acids. The carboxyl-terminal domain contains three contiguous zinc fingers.

The normal function of the *Evi-1* gene is not known. Expression studies have shown that the gene is primarily expressed in the kidney and developing oocytes in adult mice (15). During mouse embryonic development, the *Evi-1* gene shows a spatially and temporally restricted pattern of expression (20), which has suggested that the gene may play a role in development. On the basis of the phenotypic properties of the myeloid leukemias in which *Evi-1* is aberrantly expressed, it has been hypothesized that the *Evi-1* protein interferes with the terminal differentiation of hematopoietic cells. Consistent with this hypothesis, aberrant expression in 32Dcl3 myeloid cells has been shown to block their ability to terminally differentiate to granulocytes in response to granulocyte-specific colony-stimulating factor (G-CSF) (14). More recently, inappropriate expression of the *Evi-1* gene in erythroid progenitor cell lines or in normal

bone marrow cells can block erythropoietin-supported differentiation and/or proliferation (11).

To extend our knowledge of the normal function of the *Evi-1* gene and the mechanisms by which it may cause myelogenous leukemia, it is important to determine the DNA sequence specificity for binding of the two domains of zinc fingers. By using an oligonucleotide construct containing a core of 10 random nucleotides, an 11-nucleotide consensus binding site for the first domain was determined (19). However, since the first domain contains seven zinc fingers, each of which may bind 3 to 5 nucleotides, it was quite possible that the consensus binding site that was determined was only a portion of the recognition site. For these reasons, we have further characterized the DNA-binding properties of the first domain. The results demonstrate that within the first domain, zinc fingers 4 to 7 are involved in the binding of a consensus sequence that contains three repeats of a core recognition sequence of GA(T/C)AA. The first three fingers do not detectably bind DNA but contribute to the ability of the first domain to bind to the 15-bp consensus sequence and to influence the specificity for GACAA relative to GATAA in the first elements in the recognition sequence.

MATERIALS AND METHODS

Construction and production of *Evi-1* GST fusion proteins.

Evi-1 domain 1 was fused, in-frame, with glutathione S-transferase (GST) in pGEX-3 (Pharmacia). The *Evi-1* domain 1 fragments were obtained by the polymerase chain reaction (PCR) with full-length *Evi-1* cDNA as a template. The oligonucleotides used for the various constructs were (i) 5'-AGG ATC CTG ACA TCC ACG AAG ACC G, 5' of zinc finger 1; (ii) 5'-AGG ATC CTC TCC GAG AGA TCC ATG GCA A, 5' of zinc finger 2; (iii) 5'-AGG ATC CTG AGG AGA GGG AAT ACA AGT, 5' of zinc finger 3; (iv) 5'-AGG ATC CAC ATG ACA GTG GAA AGC AC, 5' of zinc finger

* Corresponding author.

† Present address: Department of Pediatrics, Yokohama City University School of Medicine, 3-9 Fukuura, Kanazawaku, Yokohama 236, Japan.

‡ Present address: Biology Division, National Cancer Center Research Institute, Chuoh-ku, Tokyo 104, Japan.

4; (v) 5'-AGG ATC CTC AGC ATG TTG GTG CCC GG, 5' of zinc finger 5; (vi) 5'-AGG ATT CAC ACT CAT AGT GCT TTC, 3' of zinc finger 3; (vii) 5'-AGA ATT CTC GGG GCA AGC ATG AGC CC, 3' of zinc finger 4; (viii) 5'-AGA ATT CAC AGA TAA AGG GCT TCAC, 3' of zinc finger 5; (ix) 5'-AGA ATT CCA CAC TCT TTG CAC TTG A, 3' of zinc finger 6; (x) 5'-AGA ATT CAA TCC ACC TGC CGC AAA ATG, 3' of zinc finger 7; (xi) 5'-GGA TCC CCT GAC ATC CAC GAA CGG, 5' of domain 1; and (xii) 5'-GAA TTC AGC AGG ATG CCT ATT GGT GCC, 3' of domain 1. Amplified fragments were purified, digested with *Bam*HI and *Eco*RI, and ligated into the *Bam*HI-*Eco*RI-digested pGEX-3 vector. The constructs were cloned into *Escherichia coli* BL-21 cells. Following the 3-h incubation of a 10-ml culture with 0.1 mM isopropyl- β -D-thiogalactopyranoside (IPTG), the bacteria were collected by centrifugation, washed twice in cold phosphate-buffered saline (PBS), resuspended in 0.5 ml of PBS, and lysed by sonication. The fusion proteins were isolated from the supernatants with 500 μ l of a slurry of 50% glutathione-Sepharose beads (Pharmacia). Following four washings with ice-cold PBS, the fusion proteins were eluted with 5 mM reduced glutathione in 50 mM Tris-HCl, pH 8.0. Alternatively, the washed beads, containing the fusion proteins, were aliquoted, stored at 4°C, and used for binding and amplification reactions. The protein concentrations were determined with a protein assay (Bio-Rad), the protein sizes were analyzed by sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE), and the DNA sequences were verified by DNA sequencing (United States Biochemical Corp. Sequenase kit).

Selection of Evi-1-GST fusion protein-binding oligonucleotides from random oligonucleotides. The 55- and 75-base, single-stranded, synthetic oligonucleotides with the sequence CTG GAT CCT AAG ATT CCC TG-N₁₅ or N₃₅-AGG CTC AAA GCT GAA TTC CT were used as templates for second-strand synthesis. In volumes of 50 μ l, 200 pmol of the single-stranded oligonucleotides was incubated with 600 pmol of reverse primer (5'-AGG AAT TCA GCT TTG AGC CT)-200 μ M of each deoxynucleoside triphosphate-polymerase reaction buffer with MgCl₂ (Perkin Elmer-Cetus)-5 U of *Taq* polymerase (Perkin Elmer-Cetus). The mixtures were treated as follows: 1-min incubation at 94°C, 2-min incubation at 50°C, and 2-min incubation at 72°C. Of the mixture, 2 μ l was analyzed on a 2% agarose gel, and 10 μ l was used for a binding reaction mixture with 50 μ l of a 50% slurry consisting of Evi-1-GST fusion protein and glutathione-Sepharose beads in binding buffer (100 mM KCl, 10 mM MgCl₂, 0.1 mM ZnSO₄, 25 mM HEPES [N-2-hydroxyethylpiperazine-N'-2-ethanesulfonic acid], 0.1% Nonidet P-40, 1 mM dithiothreitol, 5% glycerol), 100 μ g of poly(dI-dC) (Boehringer-Mannheim) per ml, and 100 μ g of bovine serum albumin (Sigma) per ml in a final volume of 125 μ l. After 30 min of incubation at room temperature, the beads were washed four times with binding buffer. The selected oligonucleotides were eluted from the beads in 50 μ l of H₂O by boiling for 5 min. The supernatant was harvested, and 10 μ l was used for PCR amplification in a 100- μ l mixture containing 200 pmol of forward primer (5'-CTG GAT CCT AAG ATT CCC TG), 200 pmol of reverse primer, and 5 U of *Taq* polymerase. Following the first binding reaction, 20 cycles of the PCR were done, whereas 15 PCR cycles were used after the second binding reaction and 10 PCR cycles were used after the subsequent binding reactions. After six rounds of binding and amplification by PCR, *Bam*HI-*Eco*RI-digested oligonucleotides were ligated into *Bam*HI-*Eco*RI-digested Bluescript (pBS-SK), cloned into *E. coli* DH5 α cells, and

sequenced by using T7 primer (Stratagene) and the United States Biochemical Corp. DNA sequencing kit.

Gel shift assay with oligonucleotides. Double-stranded oligonucleotides containing GACAAGATAA, GATAAGAT AA, GATAAGACAA, or GACAAGACAA sequences were generated by using 15 μ g of single-stranded oligonucleotide 5'-ACC CAT GGA GCT TTG AGC CCTG-[(TTATCTTG TC), (TTATCTTATC), (TTGTCTTATC), or (TTGTCTTG TC)]-AGGGAAGTGTACCATGGA and 5 μ g of the reverse primer (5'-TCCATCCTACACTTCCCT) in 350 μ l of mM NaCl-10 mM Tris-HCl (pH 8.0)-0.1 mM EDTA. Following an incubation for 5 min at 90°C, the mixture was placed at room temperature for 20 min. Then, 4 μ l of 1 M MgCl₂-10 μ l of 10 mM (each) deoxynucleoside triphosphate-40 U of Klenow-DNA polymerase I fragment (Promega) was added to the mixture and incubated at room temperature for 3 h. Double-stranded DNA was isolated on a 8% nondenaturing polyacrylamide gel by backshadowing by using shortwave UV, followed by elution and precipitation.

The oligonucleotide (500 ng) was radiolabeled with 10 U of T₄ polynucleotide kinase (Promega) and 50 μ Ci of [γ -³²P] ATP (Dupont). Radiolabeled oligonucleotides were purified by using Nunc trap push columns and a push column beta shield device (Stratagene). In 20 μ l of a reaction mixture containing 100 mM KCl, 10 mM Tris-HCl (pH 7.6), 5 mM MgCl₂, 1 mM EDTA, 12.5% glycerol, 1 mM dithiothreitol, and 1 μ g of poly(dI-dC), the ³²P-radiolabeled double-stranded oligonucleotide (2 $\times$ 10⁴ cpm; approximately 5 to 10 ng) was mixed with the Evi-1-GST fusion protein (for the concentrations indicated, see Results). Following a 15-min incubation on ice and a 15-min incubation at room temperature, the free probe was separated from the bound complexes on a nondenaturing 5% acrylamide gel at 4°C. The gel was vacuum dried and exposed to film.

RESULTS

Previous studies identified an 11-bp consensus recognition sequence (TGACAAGATAA) for the first domain of the zinc fingers of the *Evi-1* gene by binding and PCR amplification from a 50-bp oligonucleotide construct containing an internal core of 10 random nucleotides (19). Since the first domain contains seven zinc fingers, six of which are contiguous and the most 5' finger is separated from the remainder by 25 amino acids (Fig. 1), it was anticipated that the recognition sequence would be longer than 11 nucleotides. We therefore sought to determine whether a larger consensus sequence existed by using a construct containing a core of 35 random nucleotides. We also noted that in the initial studies the construct used for selection of a consensus sequence contained an AGATA in the constant sequence. Since this is a portion of the consensus sequence, the possibility existed that it contributed to the selection. We therefore used a construct that lacked any of the known consensus sequence in the sequences flanking the random core.

For the binding and amplifications reactions, a bacterial expression construct that contained the first domain of the zinc fingers and 40 carboxyl-terminal amino acids fused to GST was made. Production of the fusion protein was under the control of an inducible promoter in these constructs. Figure 2 illustrates the expression construct used as well as several of the fusion proteins that are used in studies described below. The fusion protein was purified by affinity binding to glutathione-Sepharose, and the Sepharose-bound protein was used for binding and amplification reactions as described in Materials and Methods. The sequences of 27

FIG. 1. Structure (A) and amino acid composition (B) of the zinc fingers of the murine *Evi-1* gene. In panel A, the hatched boxes represent the zinc fingers and are localized into two domains as indicated. An acidic domain is indicated by the solid box. In panel B, the amino acid numbers at the beginning and end of each finger are indicated. The boxed amino acids are the conserved zinc finger amino acids. The complete sequence of the *Evi-1* gene has been published previously (17).

sequence was frequently found flanking the consensus sequence.

The second group of clones contained a consensus sequence of 10 or 11 nucleotides that contained two repeats of the core element GA(T/C)AA. As described above, GACAA was frequently found in the first position (50%) and less frequently in the second position (21%). Importantly, in this set of clones, all the clones contained one or two GATA (TATC) or GACA (TGTC) sequences flanking the consensus sequence, suggesting that during the selection from a construct containing a core of 35 random nucleotides, binding and/or selection may occur with noncontiguous core elements.

To determine which zinc fingers contributed to the binding of the consensus sequence, a series of GST fusion proteins that contained groups of three zinc fingers fused to GST were made (Fig. 2). Initially, the fusion proteins were assayed for binding in gel shift assays. As illustrated in Fig. 4, only fusion proteins containing zinc fingers 5 to 7 or 4 to 6 bound an oligonucleotide with the sequence GACAA GATAAGATAA. Very weak binding, which is not readily seen in Fig. 4, was seen with the fusion protein containing fingers 3 to 5, and no detectable binding was seen with the fusion proteins containing fingers 1 to 3 or 2 to 4. These results demonstrated that the zinc fingers 4 to 7 were responsible for binding the consensus sequence.

The lack of apparent binding by zinc fingers 1 to 3 and 2 to 4 could be due to several reasons. It is possible that the zinc finger constructs for zinc fingers 1 to 3 and 2 to 4 are structurally unstable or do not fold correctly to retain binding activity. Alternatively, zinc fingers 1 to 4 may contribute to the consensus sequence seen in binding and amplification reactions, but the constructs may have a lower affinity so that binding cannot be detected in gel shifts. Lastly, it is possible that zinc fingers 1 to 4 lack either DNA-binding activity or sequence specificity. To distinguish between some of these possibilities, the various fusion

FIG. 2. Production of GST fusion constructs containing various zinc fingers (ZF) from the first zinc finger domain of Evi-1. (A) The pGEX expression construct (Pharmacia) was used for expression of the GST fusion proteins. (B) Expression constructs containing the entire first domain and the flanking 45 amino acids were made (data not shown). The regions of the first domain used in the other constructs are indicated and included constructs with zinc fingers 1 to 3, 2 to 4, 3 to 5, 4 to 6, and 5 to 7. (C) SDS-PAGE analysis of the induced fusion proteins. The various constructs were introduced into *E. coli*, and expression was induced in exponentially growing cultures with IPTG. The cells were collected and lysed, and the extract was incubated with glutathione-Sepharose. The Sepharose beads were washed, the bound fraction was eluted by boiling with SDS-PAGE buffer, and the bound proteins were resolved by SDS-PAGE. The sizes of standards (in kilodaltons) are indicated on the right.

FIG. 3. Sequences of clones obtained by binding and amplification with a GST fusion protein containing the first domain of zinc fingers from the *Evi-1* gene. (A) GST fusion protein containing the first domain of zinc fingers was produced as described in Materials and Methods. The fusion protein, bound to glutathione-Sepharose beads, was used for binding and amplification reactions. The Sepharose-bound protein was incubated with a oligonucleotide construct containing a core of 35 random nucleotides, as described in Materials and Methods, in the presence of poly(dI-dC) (1 mg/ml). The beads were washed, and the bound nucleotides were eluted and amplified by PCR under the conditions described in Materials and Methods. Following six rounds of binding and amplification, the products were cloned into pBluescript and sequenced. The sequences of the 35-nucleotide core region from 27 unique clones are shown. The top section shows the sequences of clones that had a detectable 15-bp consensus sequence, while the bottom section shows the sequences of clones that had a 10-bp consensus sequence. The nucleotide frequency in each set of clones is indicated below the clone sequences.

proteins were used in binding and amplification reactions with a construct containing a core of 15 random nucleotides. As illustrated in Fig. 5, following the first binding and amplification reactions, PCR-amplified products could be detected with all the fusion proteins, as well as with GST alone, indicating that this was not specific binding. After the second binding and amplification reactions however, products were not detected in reaction mixtures containing GST or containing fusion proteins with zinc fingers 1 to 3 or 2 to 4. After the third round of binding and amplification, PCR products were detected with the fusion proteins containing only zinc fingers 5 to 7 or 4 to 6. PCR products were also obtained from the fusion protein containing zinc fingers 3 to 5 but only after additional cycles of PCR. These results suggest that fusion proteins containing zinc fingers 1 to 3 or 2 to 4 do not detectably bind DNA under the conditions used, that the fusion protein containing zinc fingers 3 to 5 binds DNA weakly, and that fusion proteins containing zinc fingers 4 to 6 and 5 to 7 bind DNA with high affinity.

To determine whether the fusion proteins containing isolated triplets of zinc fingers maintained the specificity seen with the first domain construct and to better define the

FIG. 4. Zinc fingers 4 to 7 bind the consensus sequence. GST-fusion proteins containing the first domain of zinc fingers (1-7) or fingers 5-7, 4-6, 3-5, 2-4, 1-3 were used in gel shift assays with an oligonucleotide containing the consensus sequence GACAAGATA AGATAA. Controls included the oligonucleotide alone (C) or GST protein (GST). The binding reactions and conditions of electrophoresis of the complexes are described in Materials and Methods.

FIG. 5. Binding and amplification reactions with constructs containing various zinc fingers from the first domain. Binding and amplification reactions were done with GST fusion proteins containing zinc fingers 5 to 7, 4 to 6, 3 to 5, 2 to 4, or 1 to 3. As a control, reactions were also done with GST protein alone. The conditions of the reactions are described in Materials and Methods. Following the first binding and amplification reaction, which involved 20 cycles of PCR, the products were resolved on agarose gels and stained with ethidium bromide, and the gels were photographed. Photographs of the products following a second round of binding and amplification with 15 cycles of PCR are shown. The products following the third round of binding and amplification with either 10 or 18 cycles of PCR are shown at the bottom. The positions of the 55-bp oligonucleotide construct are indicated by the arrows. The lower band in each gel is the primer.

regions involved in binding to the consensus sequence, the PCR products were cloned and sequenced. As shown in Fig. 6A, the fusion protein containing zinc fingers 5 to 7 gave a consensus sequence that contained the core element seen in the sequences obtained with the entire first domain. The core element GA(T/C)AA was readily evident, was repeated twice in all the clones, and primarily contained the GATAA sequence (93% occurrence in the first position and 100% occurrence in the second position). Similarly, the core consensus elements were readily apparent in the sequences of the PCR products from the binding and amplification reactions done with the fusion protein containing zinc fingers 4 to 6 (Fig. 6B). However, many clones contained three core elements, and both GACAA and GATAA sequences were present. In these cases, GACAA occurred 25% of the time in the first or second position and never in the third position. In addition, clones that contained two contiguous repeats were obtained, and in these cases, GACAA occurred in the first position in 26% of the clones and did not occur in the second position. It should be noted that the consensus sequence in these constructs often included nucleotides from the constant flanking sequences. Clones that were obtained from binding and amplification reactions with the fusion protein containing zinc fingers 3 to 5 contained no apparent consensus sequence (data not shown). These data suggest that the major binding occurs with zinc fingers 4 to 7 and that binding preference for GACAA relative to GATAA is shown more often by zinc fingers in the amino-terminal region.

The possibility existed that some binding specificity or a binding component was lost by using fusion proteins containing only three fingers. Since it was clear that the binding thus far detected resided in the carboxyl-terminal zinc fingers, we examined a series of fusion proteins that incorpo-

rated progressively more zinc fingers starting with fingers 5 to 7. In addition, the initial fusion protein containing domain 1 also contained approximately 40 amino acids beyond the seventh zinc finger, which might have contributed to binding; therefore, a fusion protein lacking this region was also used. The structures of the fusion proteins are shown in Fig. 7, and the results of sequencing of clones from binding and amplification reactions with these fusion proteins are shown in Fig. 8. For these studies, an oligonucleotide construct containing a core of 35 random nucleotides was used. As illustrated, the fusion protein containing fingers 1 to 7 gave results comparable to those obtained with the fusion protein containing the first domain and, in addition, 40 carboxyl-terminal amino acids. Approximately half of the clones contained a 15-bp contiguous consensus sequence consisting of the GA(T/C)CC core sequence elements. Although not shown, most of the clones contained additional GATAA or GACAA elements in the flanking sequences as described above. The fusion protein containing zinc fingers 2 to 7 selected considerably fewer clones with a contiguous 15-bp consensus sequence, but rather there was a shift to a 10-bp contiguous consensus, although still consisting of the GA(T/C)AA core sequence. The predominant core consensus sequence with the constructs containing fingers 3 to 7 or 4 to 7 contained the 10-bp contiguous core although an increase in clones containing only a single GA(T/C)AA element was clearly evident. In these clones, all contained additional GA(T/C)AA elements in the flanking sequences (data not shown). Together, the results suggest that the first three zinc fingers do not contribute to the consensus recognition sequence but may contribute to the confirmation of the first domain in such a way as to allow optimal binding and selection of a contiguous consensus sequence of 15 nucleotides.

The above studies demonstrated that the first domain binds a consensus sequence of 15 bp containing three elements of GA(T/C)AA. The results also indicated that the order of preference for selection in the binding and amplification reactions was GACAAGATAAGATAA > GATAA GACAAGATAA > GATAAGATAAGACAA, consistent with the finding of the previous studies which indicated that the binding preference for the GACAA element precedes that for the GATAA element. To assess this further, a series of gel shift assays were done with several of the above fusion proteins. As shown in Fig. 9, the fusion protein containing the entire first domain and an additional 40 amino acids showed a clear preference. In particular, comparable binding was seen with GACAAGATAA (Fig. 9, lanes C/T) and GACAAGACAA (Fig. 9, lanes C/C), and a much lower level of binding was seen with GATAAGATAA (Fig. 9, lanes T/T) and GATAAGACAA (Fig. 9, lanes T/C). A similar preference for binding was seen with the fusion protein containing only zinc fingers 1 to 7. However, the fusion protein containing zinc fingers 2 to 7 showed less specificity in binding, and the fusion protein containing zinc fingers 3 to 7 had no apparent preference among the four oligonucleotides examined. Similarly, fusion proteins containing zinc fingers 4 to 7 and 5 to 7 bound the four oligonucleotides comparably. Of note with the fusion proteins containing zinc fingers 5 to 7 and 4 to 7 was a concentration-dependent shift in electrophoretic mobility that was not seen with other fusion proteins. The basis for the changes in mobility is not known but may be due to formation of dimers or aggregates. Together, the results suggest that the amino-terminal region of the first domain containing finger 1 and the spacer region contribute to specificity in the binding of the carboxyl-terminal zinc

B

Zinc Fingers 4-6

c t	gACTAGATATGATAT	Gagg
c	tgATTAGATAAGATA	GTagg
c t	gATAAGATATGATAC	Aagg
c tGT	AGATAAGATAAAGATAG	G
c t g	AGATAAGACTAGATAa	gg
c c t	AGATATGATATGATAC	ag
c t g	GGACATGATAAAGATAa	gg
c	t gATATGACAAGATAA	Tagg

$$\begin{array}{ccccccc} A^3 & & & & C^2 \\ T^4 G A T^6 A^8 A^5 G A T^9 A^7 A^5 G A T A & & C^2 \\ G^1 & & C^2 T^2 T^3 & & C^2 T^1 T^3 & & G^2 T^1 \end{array}$$

ctgAA	AGATAAGATAG	TTagg
cctgT	AGATAAGATAA	GACag
c	tgACTAGATAT	ATCATTagg
ct	gACTAGATAT	ATCTGTagg
cctCCAAC	AGATAAGATAC	ag
ctgAA	TGATAAGATAT	GAagg
ctg	AGATAAGATAA	TTAGagg
ctAAGATTA	AGATAAGATag	g
cctAA	AGATAAGATAA	TACag
ctgGAAT	AGACAAGATAG	Tagg
ctgAATG	AGATAAGATAA	agg
	tgATATGATAT	ATCTATagg
ctgGG	AGATAAGATAG	GACagg
cct	AGATAAGATAA	TAAACag
cctGAAT	AGACATGATAG	Cag
ctgACA	AGACATGATAG	Cagg
cc	tGACAAGATAT	ATCTTCag
cctT	AGATAAGATAA	TAACag
cctATTT	AGATAAGATAA	Cag

$$\begin{array}{ccccccccccc} A^{14} & G & A & T^{14} & A^{17} & A^{16} & G & A & T & A & T^5 \\ T^5 & & & C^5 & T^2 & T^3 & & & & & A^7 \\ & & & & & & & & & & G^6 \\ & & & & & & & & & & C^1 \end{array}$$

The results support and extend previous studies (19) that addressed the DNA-binding specificity of the first domain of zinc fingers of the *Evi-1* myeloid-transforming gene. A consensus sequence of TGACAAGATAA had been previously identified by binding and amplification with an oligonucleotide construct that contained a core of 10 random nucleotides. The first T in the deduced consensus sequence was most often contributed by the constant nucleotides in the oligonucleotide construct and thus may not be part of the consensus recognition sequence. By binding and amplification with an oligonucleotide construct containing a core of 35 random nucleotides, it was possible to confirm and extend the consensus recognition sequence. In particular, the results demonstrate that the first domain recognizes a consensus sequence of 15 nucleotides that contains three repeats of a core consensus sequence of GA(T/C)AA. There was also a relative order of preference for GACAA relative to GATAA such that it was found most frequently in the first position, often in the second position, and seldom in the last position, consistent with the results of the previous studies. However, the results did not indicate a strong preference for T in the first position of the previously defined consensus sequence.

FIG. 7. GST fusion proteins containing progressively fewer zinc fingers. (A) Structure of the first domain of zinc fingers. The hatched boxes indicate the positions of the zinc fingers. The position of the amino terminus is indicated. The construct containing the entire first domain included 40 amino acids (40 A.A.) beyond the indicated point. The structures of fusion proteins containing zinc fingers 1 to 7, 2 to 7, 3 to 7, 4 to 7 and 5 to 7 are shown. (B) SDS-PAGE analysis of fusion proteins. The fusion proteins were purified as described in Materials and Methods and were resolved by SDS-PAGE. The positions of molecular size markers (in kilodaltons) are indicated on the right.

The consensus sequence, derived by binding and amplification, is primarily recognized by zinc fingers 4 to 7 in gel shift mobility assays. Assuming that each zinc finger recognizes 3 bp (18), it would be anticipated that the recognition sequence for zinc fingers 4 to 7 would consist of 12 bp. It is possible that the zinc fingers of the first domain of the *Evi-1* gene do not conform to the structures that have been obtained with proteins containing three contiguous zinc fingers. Alternatively, it is possible that within the context of the entire first domain, zinc finger 1, 2, or 3 contributes to the binding and selection of the complete consensus recognition sequence.

Recent studies have provided structural information detailing interactions of zinc fingers with DNA (13, 18). The results demonstrate that the charged residues within the

alpha helical region of a zinc finger make contact with bases in the major groove of DNA. On the basis of the known recognition sequences, a series of rules have been proposed for binding (3, 4, 18). The potential DNA contact residues in the first domain of the zinc fingers of the *Evi-1* gene, reading in the carboxyl-to-amino direction (zinc fingers 7 to 1), are KST RNQ QGT RND RNW KSD DES. It is interesting to note that zinc fingers 3, 4, and 6 share the potential contact residues RN which, in several zinc finger proteins, have been shown to recognize GA (4, 10, 18). Consistent with this, the second domain in the *Evi-1* gene contains three fingers with potential contact residues of RNQ, RNI, and RNR, and these fingers recognize the sequences GAA, GAT, and GAG, respectively (5). Of particular interest is the sixth finger in the first domain, which contains the recogni-

ZF1-7	ZF2-7	ZF3-7	ZF4-7	ZF5-7
AGACAAGATAAGATAG	AGATATGATAAGATAC	AGATAAGACAG	GATAAGACAAGATTA	CGATAAGATAT
GGACAAGATAAGATAT	AGACAAGATATGATAC	GGACAAGATAT	TGATAAGATTA	AGATACGATAA
AGATAAGATAACACAA	GGACAAGAGAAGATAC	AGACAAGATAA	GATAAGATTA	CGATATGATAA
AGATAAGATATGACAG	AGATAAGATAC	AGATATGATAT	AGATAAGATAC	AGATAAGAGAA
GGACAAGACAAGATAC	AGATAAGATAA	AGACAAGACAT	TGACAAGATAA	AGACTAGATGA
CGACATGATAAGACAT	AGACATGACAA	AGATAAGACAG	TGATAAGATAA	AGACTAGATCA
AGATAAGATAAGACAG	AGATAAGATAT	TGATATGATCT	TGACAAGATAG	AGATAG
GGACATGATAAGACAA	AGATAAGACAA	TGATAGATAGT	AGATAT	TGATAG
AGATAAGATAAGATGA	AGATAAGATAA	GGACAAGATAA	TGACAC	AGATAA
TGACGAGATAAGACAA	AGATATGATAG	AGATAAGATAA	TGATAT	AGATAT
AGACAAGACAAGATAA	AGATAAGATAG	GGACTAGATAT	GGATAC	AGATAT
TTACAAGATAAGATAT	GATAAGATAC	TGATGAGACAA	AGACAC	AGATAA
AGATAAGACATGATAA	TGATAAGATAC	AGATATGATAA	AGACAG	AGATAC
AGATAAGATAA	AGATAAGATAC	TGACAAGATAA	GGACAG	GGATAG
AGACAAGATAA	AGATAAGATAT	AGATAA	AGATAA	AGATAG
TTATAAGATAA	AGATAAGATA	CGATAC	TGACAA	TGATAC
TGATAAGATAA	TGATAAGATAG	TGACAG	AGATAA	AGACAA
AGATAAGATAA	TGATAAGATAG	CGATAC	GGACAT	AGATAT
AGATAAGATAG	TGACATGATAG	AGATAG	AGACAA	TGATAG
AGATAAGATAA	AGATAAGACAC	AGATAG	AGATAC	TGACAT
AGATAAGATAG	AGACAAGATAG	CGACAG		
AGATAAGATAG	AGACAAGACAA	TGACAG		
GGATAAGATAA	TGACAAGACAG	TGATAG		
CGATAAGATAA	AGATCAGATAG	TGATAT		
AGATAAGATAA	AGATAG	AGATAA		
AGACAAGATAA	AGATAA	AGATAA		
GGATAAGATAT	AGATAG			
TGATTAGATAG	AGACAC			
AGATAAGATAT	AGATAA			
AGATAAGATGT	CGATAT			
AGATAAGATAT				

FIG. 8. Sequence of oligonucleotides isolated by binding and amplification with various first domain zinc finger (ZF) constructs. The zinc finger fusion proteins described in the legend to Fig. 7 were used in binding and amplification reactions with an oligonucleotide construct that contained a core of 15 random nucleotides. Following six rounds of binding and amplification, the products were cloned and sequenced. The consensus sequences of the clones are indicated.

FIG. 9. Gel mobility shift assays with various zinc finger constructs. Gel shift assays were done as described in Materials and Methods. GST fusion proteins containing zinc fingers (ZF) 5 to 7, 4 to 7, 3 to 7, 2 to 7, or 1 to 7 or the entire first domain were used as indicated on the left. The fusion proteins were assayed with two amounts of protein (5 and 1.7 µg), as indicated at the top. The oligonucleotides used for the assays included GACAAGATAA (C/T), GATAAGATAA (T/T), GATAAGACAA (T/C), and GACAAGACAA (C/C) as indicated at the bottom. 40AA, 40 amino acids.

tion helix sequence QFSNLCR. In studies to define rules relating zinc finger protein sequences and DNA-binding site preference, it was shown that within the similar recognition helix QSSNLQR, the RNQ contact residues conferred a high specificity for the sequence GAT (4). Thus, the existing rules suggest multiple binding sites for GA and a possible binding site for GAT, consistent with the consensus sequence obtained. However, it is not possible to totally reconcile the complete consensus sequence obtained by binding and amplification reactions with the current binding rules.

The results demonstrate that the first three zinc fingers do not contribute to binding to the consensus sequence nor do they bind DNA with sufficient affinity to isolate clones in binding and amplification reactions. The lack of usage of all zinc fingers has also been shown for the eukaryotic transcriptional repressor PRDI-BF1, which utilizes only two of five contiguous zinc fingers for sequence-specific DNA binding of a 14-bp sequence (9). The roles of zinc fingers 1 to 3 in the *Evi-1* gene are not known; however, the results suggest that they contribute to optimal binding and confer some sequence specificity for binding. This was particularly evident in the gel shift assays (Fig. 9) in which removal of the first finger alone eliminated the binding specificity for GACA relative to GATAA. Therefore, it is tempting to speculate that the amino-terminal zinc fingers contribute to

FIG. 10. Competition assays for binding of the consensus sequence by the first domain. Gel shift assays were done as described in Materials and Methods. A GST fusion protein, containing the entire first domain of zinc fingers, was used for all binding and competition assays. The labeled probe is indicated at the left and consisted of GACAAGATAA (CT), GATAAGATAA (TT), or GACAAGATAAGATAA (CTT). Gel shifts were done with reaction mixtures containing no competitor (NC) or with a 1-, 10-, or 100-fold (1X, 10X, or 100X, respectively) excess of unlabeled competitor as indicated.

the overall structure of the first domain and thereby influence DNA binding.

Evi-1 fusion proteins that lack the first three or four zinc fingers gave different results in binding and amplification reactions by using constructs with a core of either 35 or 15 random nucleotides. In particular, with a core of 35 random nucleotides, the products of binding and amplification reactions often contained two or more separated GA(T/C)AA elements. The basis for this difference is not known. One possibility is the relative sizes of the pools. The constructs with 15 random nucleotides have 1.1×10^9 possibilities while the constructs with 35 random nucleotides have 1.2×10^{21} possibilities; 1.1×10^9 molecules is 1.8×10^{-3} pmol, while 1.2×10^{21} molecules is 2.0×10^9 pmol. Since the binding and amplification reaction mixtures contain 200 pmol, all possibilities would be expected to be represented in the case of 15 random nucleotides, while this is not the case for the constructs containing 35 random nucleotides. This cannot completely account for the differences, however, since the expected sequences were obtained from the constructs containing 35 random nucleotides with fusion proteins containing zinc fingers 1 to 7, 2 to 7, and 3 to 7.

It is also possible that the truncated first domain constructs are able to bind and amplify oligonucleotides with noncontiguous elements. This could be due to the ability of either a single protein to bind two noncontiguous elements or two proteins to bind independently to the two sites. We have examined the significance of a second, noncontiguous GATA on binding in competition assays, by using an oligonucleotide containing either a single GATA or GATA sites separated by 4, 8, or 12 nucleotides. The binding of an oligonucleotide containing two GATA sequences separated by 12 nucleotides was equivalent to the binding of constructs containing two sites spaced by 4, 8, or 12 nucleotides and was stronger than the binding of an oligonucleotide with a single site. Conversely, the binding to an oligonucleotide with a single site competed equally well with binding to all oligonucleotides. Therefore, a second noncontiguous site can contribute to the binding when single sites are involved.

However, when binding to GACAAGATAA containing an oligonucleotide was examined, the presence or absence of a second, noncontiguous single site in the competing oligonucleotide had no effect on competition. Therefore, noncontiguous sites can contribute when single sites are involved. However, it is much less apparent that a noncontiguous site contributes to the binding affinity when an oligonucleotide contains contiguous sites. In the latter case, the selection of oligonucleotides with noncontiguous sites could be due to the binding through two proteins under the conditions of the binding and amplification reactions, namely, the high protein concentrations on beads. This might be expected to occur more frequently with the constructs containing 35 random nucleotides since more possibilities for noncontiguous sequences exist.

A potential role for zinc fingers 1 to 3 in contributing to the structure of domain 1 rather than for defining specificity of DNA binding is also suggested by comparing the zinc finger amino acid sequences of the murine *Evi-1* gene with that of the *Drosophila Evi-1* gene (6). In particular, all of the critical contact residues in fingers 4 to 7 are conserved between the two species. However, none of the contact residues are conserved in zinc finger 2 or 3; the amino acid sequence of the first zinc finger of the *Drosophila* gene has not been determined. More strikingly, recent studies have identified the *Caenorhabditis elegans* homolog of the murine *Evi-1* gene (7). As with the *Drosophila* gene, the contact residues in the second domain are completely conserved, while most of the contact residues in zinc fingers 5 to 7 of the first domain are conserved. However, the *C. elegans* homolog lacks zinc fingers 1 to 4.

Our studies have utilized a series of fusion proteins containing the first domain of zinc fingers or parts thereof. It is therefore important to determine whether the full-length protein binds this sequence in vivo and whether this binding has a biological consequence. Attempts to produce sufficient full-length protein in bacteria or insect cells for binding and amplification reactions have not succeeded. However, experiments which address the ability of Evi-1 to recognize the consensus sequence in vivo have been done (11). In particular, we have examined the ability of Evi-1 to activate chloramphenicol acetyltransferase constructs containing the first domain consensus sequence. Using the first domain consensus sequence alone or in combination with the second domain consensus binding site, we have failed to see transcriptional activation in fibroblasts. However, the first domain consensus sequence contains within it the consensus sequence for GATA-1 [(T/A)GATA(A/G)] (8, 22). Consistent with this, chloramphenicol acetyltransferase constructs containing the first domain consensus sequence can be transcriptionally activated by cotransfection with GATA-1 expression constructs. Cotransfection of GATA-1 and full-length *Evi-1* expression constructs results in the suppression of the GATA-1-dependent transcription. This was specific since the *Evi-1* expression construct did not block transcription from a similar CAT construct in which transcription was driven from the adenovirus E4 promoter by cotransfection with an E1A expression construct.

The ability of *Evi-1* to recognize sequences which contain GATA-1 sites in vivo has significance to the possible mechanisms of transformation. In particular, the aberrant expression of the *Evi-1* gene is associated with myeloid leukemias and myelodysplastic syndromes in mice and humans. Consistent with this, the aberrant expression of the *Evi-1* gene in 32Dcl myeloid leukemia cells blocks their ability to terminally differentiate in responses to G-CSF (14). Similarly,

inappropriate expression of the *Evi-1* gene in an erythroid progenitor cell line or in normal bone marrow progenitors blocks their ability to respond to erythropoietin and to differentiate (11). Since GATA-1 has been shown to be essential for erythroid differentiation (21), it can be hypothesized that inappropriate expression of the *Evi-1* gene blocks erythropoiesis by repressing the transcription of a subset of GATA-1 target genes. Whether comparable effects occur within the granulocytic lineage is not known.

ACKNOWLEDGMENTS

This work was supported in part by National Cancer Institute Cancer Center Support (CORE) grant P30 CA21765, by grant CA51020 to J.N.I. from the National Cancer Institute, by funds from the Dutch Cancer Society (KWF) and the Dutch Organization of Scientific Research (NWO) to R.D., and by the American Lebanese Syrian Associated Charities.

REFERENCES

1. Bartholomew, C., and J. N. Ihle. 1991. Retroviral insertions 90 kilobases proximal to the *Evi-1* myeloid transforming gene activate transcription from the normal promoter. *Mol. Cell. Biol.* 11:1820-1828.
2. Bartholomew, C., K. Morishita, D. Askew, A. Buchberg, N. A. Jenkins, N. G. Copeland, and J. N. Ihle. 1989. Retroviral insertions in the CB-1/Fim-3 common site of integration activate expression of the *Evi-1* gene. *Oncogene* 4:529-534.
3. Desjarlais, J. R., and J. M. Berg. 1992. Redesigning the DNA-binding specificity of a zinc finger protein: a data base-guided approach. *Proteins* 12:101-104.
4. Desjarlais, J. R., and J. M. Berg. 1992. Toward rules relating zinc finger protein sequences and DNA binding site preferences. *Proc. Natl. Acad. Sci. USA* 89:7345-7349.
5. Funabiki, T., R. Delwel, K. Morishita, and J. N. Ihle. Unpublished data.
6. Funabiki, T., and J. N. Ihle. Unpublished data.
7. Garriga, G., C. Guenther, and H. R. Horvitz. Unpublished data.
8. Ho, I.-C., P. Vorhees, N. Marin, B. K. Oakley, S.-F. Tsai, S. H. Orkin, and J. M. Leiden. 1991. Human GATA-3: a lineage-restricted transcription factor that regulates the expression of the T cell receptor α gene. *EMBO J.* 10:1187-1192.
9. Keller, A. D., and T. Maniatis. 1992. Only two of the five zinc fingers of the eukaryotic transcriptional repressor PRDI-BF1 are required for sequence-specific DNA binding. *Mol. Cell. Biol.* 12:1940-1949.
10. Kleit, R. 1991. Recognition of DNA by Cys₂, His₂ zinc fingers. *Science* 253:1367-1393.
11. Kreider, B. L., S. H. Orkin, and J. N. Ihle. Loss of erythropoietin responsiveness in erythroid progenitors due to expression of the *Evi-1* myeloid transforming gene. *Proc. Natl. Acad. Sci. USA*, in press.
12. Matsugi, T., K. Morishita, and J. N. Ihle. 1990. Identification, nuclear localization, and DNA-binding activity of the zinc finger protein encoded by the *Evi-1* murine myeloid transforming gene. *Mol. Cell. Biol.* 10:1259-1264.
13. Min, S. L., G. P. Gippert, K. V. Soman, D. A. Case, and P. E. Wright. 1989. Three-dimensional solution structure of a single zinc finger DNA-binding domain. *Science* 245:635-637.
14. Morishita, K., E. Parganas, T. Matsugi, and J. N. Ihle. 1992. Expression of the *Evi-1* zinc finger gene in 32Dcl3 myeloid cells blocks granulocytic differentiation in response to granulocyte colony-stimulating factor. *Mol. Cell. Biol.* 12:183-189.
15. Morishita, K., E. Parganas, D. Parham, T. Matsugi, and J. N. Ihle. 1990. The *Evi-1* zinc finger myeloid transforming gene is normally expressed in the kidney and in developing oocytes. *Oncogene* 5:1419-1423.
16. Morishita, K., E. Parganas, C. L. Willman, M. H. Whittaker, H. Drabkin, J. Oval, R. Taetle, and J. N. Ihle. 1992. Activation of EVI1 gene expression in human acute myelogenous leukemias by translocations spanning 300-400 kilobases on chromosome 3q26. *Proc. Natl. Acad. Sci. USA* 89:3937-3941.

17. Morishita, K., D. S. Parker, M. L. Mucenski, N. A. Jenkins, N. G. Copeland, and J. N. Ihle. 1988. Retroviral activation of a novel gene encoding a zinc finger protein in IL-3-dependent myeloid leukemia cell lines. *Cell* 54:831-840.
18. Pavletich, N. P., and C. O. Pabo. 1991. Zinc finger-DNA recognition: crystal structure of a Zif268-DNA complex at 2.1 Å. *Science* 252:809-817.
19. Perkins, A. S., R. Fishel, N. A. Jenkins, and N. G. Copeland. 1991. *Evi-1*, a murine zinc finger proto-oncogene, encodes a sequence-specific DNA-binding protein. *Mol. Cell. Biol.* 11: 2665-2674.
20. Perkins, A. S., J. A. Mercer, N. A. Jenkins, and N. G. Copeland. 1991. Patterns of *Evi-1* expression in embryonic and adult tissues suggest that *Evi-1* plays an important regulatory role in mouse development. *Development* 111:479-498.
21. Pevny, L., M. C. Simon, E. Robertson, W. H. Klein, S.-F. Tsai, V. D'Agati, S. H. Orkin, and F. Costantini. 1991. Erythroid differentiation in chimaeric mice blocked by a targeted mutation in the gene for transcription factor GATA-1. *Nature (London)* 349:257-260.
22. Tsai, S.-F., D. I. K. Martin, L. I. Zon, A. D. D'Andrea, G. G. Wong, and S. H. Orkin. 1989. Cloning of cDNA for the major DNA-binding protein of the erythroid lineage through expression in mammalian cells. *Nature (London)* 339:446-451.