Characterization of the DNA-Binding Properties of the Myeloid Zinc Finger Protein MZF1: Two Independent DNA-Binding Domains Recognize Two DNA Consensus Sequences with a Common G-Rich Core

JENNIFER F. MORRIS, ROBERT HROMAS, AND FRANK J. RAUSCHER III*

Wistar Institute of Anatomy and Biology, Philadelphia, Pennsylvania 19104, and Departments of Medicine and Biochemistry, Walther Oncology Center, Indiana University Medical Center, Indianapolis, Indiana 46202

Received 4 August 1993/Returned for modification 14 October 1993/Accepted 17 December 1993

The myeloid zinc finger gene 1, MZF1, encodes a transcription factor which is expressed in hematopoietic progenitor cells that are committed to myeloid lineage differentiation. MZF1 contains 13 C2H2 zinc fingers arranged in two domains which are separated by a short glycine- and proline-rich sequence. The first domain consists of zinc fingers 1 to 4, and the second domain is formed by zinc fingers 5 to 13. We have determined that both sets of zinc finger domains bind DNA. Purified, recombinant MZF1 proteins containing either the first set of zinc fingers or the second set were prepared and used to affinity select DNA sequences from a library of degenerate oligonucleotides by using successive rounds of gel shift followed by PCR amplification. Surprisingly, both DNA-binding domains of MZF1 selected similar DNA-binding consensus sequences containing a core of four or five guanine residues, reminiscent of an NF-κB half-site: 1-4, 5'-AGTGGGA-3'; 5-13, 5'-CGGguAGGGGAA-3'. The full-length MZF1 protein containing both sets of zinc finger DNA-binding domains recognizes synthetic oligonucleotides containing either the 1-4 or 5-13 consensus binding sites in gel shift assays. Thus, we have identified the core DNA consensus binding sites for each of the two DNA-binding domains of a myeloid-specific zinc finger transcription factor. Identification of these DNA-binding sites will allow us to identify target genes regulated by MZF1 and to assess the role of MZF1 as a transcriptional regulator of hematopoiesis.

Hematopoiesis is a complex process of cell differentiation and renewal in which the entire population of blood cells are derived from a few pluripotent stem cells. Regulation of this process is largely via cytokines which interact with specific cell surface receptors on the stem cells. Following cytokine binding, these pluripotent stem cells must be able to transduce signals from the cell surface to the nucleus and initiate the process of differentiation along specific cell lineages by regulating gene expression. Thus, transcription factors that act at specific stages of hematopoietic differentiation can be seen as critical regulators of the differentiation process. Intuitively, loss of the regulation of gene expression at key stages during hematopoietic differentiation can ultimately lead to leukemogenesis.

The family of C2H2 zinc finger genes represents a class of DNA-binding proteins, many of which have been demonstrated to have roles in regulating transcription during developmental processes. Zinc finger genes encode proteins that contain highly conserved 28- to 30-amino-acid (aa) domains that tetrahedrally bind a molecule of zinc through two cysteines and two histidines (reviewed in reference 5). The zinc finger forms a DNA-binding domain which interacts with specific nucleotides along the major groove of DNA (24, 25). Krüppel, a Drosophila gap gene, represents a subclass of the C2H2 zinc finger genes that is characterized by multiple zinc fingers containing the conserved amino acid motif CX2CX3FXLX3H3H3 separated by highly conserved 7-aa linkers (H-C links) of the form TG(EPYX (9, 29).

This article describes a novel transcription factor, encoded by myeloid zinc finger gene 1 (MZF1), that may play a central role in regulating hematopoiesis. MZF1 was isolated by screening a cDNA library prepared from the peripheral blood leukocytes of a patient with chronic myelogenous leukemia by using an oligonucleotide probe which recognizes the conserved H-C link (15). MZF1 encodes a 485-aa protein (Fig. 1) that contains 13 zinc fingers arranged in two distinct sets: the first set contains zinc fingers 1 to 4, and the second set, in the carboxy terminus, contains zinc fingers 5 to 13. This protein configuration suggests that MZF1 contains two separate DNA-binding domains that may recognize two independent DNA target sequences. Several lines of evidence suggest that MZF1 plays a key role in regulating transcription during differentiation. First, a survey of MZF1 mRNA abundance in a variety of cell lines detected expression almost exclusively in early myeloid progenitor cells, including HL60, KG1, HEL, and K562 (15). MZF1 was also detected in early myeloid lineage progenitor cells by in situ hybridization of bone marrow cells from normal donors (4). In contrast, MZF1 RNA was not detected in cell lines representing other hematopoietic cell lineages, including T-cell leukemia (C10), B-cell leukemia (Burkitt and Daudi), megakaryocytic leukemia (M07E), hairy cell leukemia (ESKOL), or differentiated monocytic leukemia (U937) cells (15). Thus, MZF1 mRNA expression is limited to progenitor cells committed to differentiating along the myeloid lineage.

Evidence that MZF1 regulates hematopoiesis is provided by the pattern of MZF1 mRNA expression in the myeloid-committed progenitor cell line HL60 following retinoic acid induction of granulocytic differentiation. The HL60 cell line expresses detectable amounts of MZF1 mRNA before induc-
The zinc fingers of MZF1 are aligned and numbered to illustrate the arrangement of the 13 zinc fingers into two DNA-binding domains; the first domain contains zinc fingers 1 to 4, and the second contains zinc fingers 5 to 13. The Cys and His amino acids involved in the chelation of Zn$^{2+}$ molecules are underlined. Putative DNA contact amino acids (on the basis of the Egr1 crystal structure [24]) within each zinc finger are outlined. The acidic region (aa 60 to 72) and the glycine-proline-rich region separating zinc fingers 4 and 5 (aa 214 to 237) are italicized and underlined.

![Image](https://via.placeholder.com/150)

**FIG. 1.** Amino acid sequence of MZF1. The zinc fingers of MZF1 are aligned and numbered to illustrate the arrangement of the 13 zinc fingers into two DNA-binding domains; the first domain contains zinc fingers 1 to 4, and the second contains zinc fingers 5 to 13. The Cys and His amino acids involved in the chelation of Zn$^{2+}$ molecules are underlined. Putative DNA contact amino acids (on the basis of the Egr1 crystal structure [24]) within each zinc finger are outlined. The acidic region (aa 60 to 72) and the glycine-proline-rich region separating zinc fingers 4 and 5 (aa 214 to 237) are italicized and underlined.

**TABLE 1.** PCR primers used to produce recombinant MZF1 proteins

<table>
<thead>
<tr>
<th>Protein (aa)</th>
<th>Sense</th>
<th>Antisense</th>
</tr>
</thead>
<tbody>
<tr>
<td>rZN 1-4 (108-240)</td>
<td>5'TATAGGATCCCGAGCTGATGATTG-3'</td>
<td>5'CCAAAAAGCTTCACGGTCAAGAAA-3'</td>
</tr>
<tr>
<td>rZN 5-8 (218-364)</td>
<td>5'TATAGGATCCCGAGCTGATGATTG-3'</td>
<td>5'CCAAAAAGCTTCACGGTCAAGAAA-3'</td>
</tr>
<tr>
<td>rZN 8-10 (305-416)</td>
<td>5'TATAGGATCCCGAGCTGATGATTG-3'</td>
<td>5'CCAAAAAGCTTCACGGTCAAGAAA-3'</td>
</tr>
<tr>
<td>rZN 11-13 (389-485)</td>
<td>5'TATAGGATCCCGAGCTGATGATTG-3'</td>
<td>5'CCAAAAAGCTTCACGGTCAAGAAA-3'</td>
</tr>
</tbody>
</table>
growing culture with 1 mM IPTG (isopropyl-β-D-thiogalactopyranoside) for 2 h. The bacteria from a 2-ml culture were lysed in 100 μl of Laemmli loading buffer (20), boiled for 10 min, and clarified by centrifuging for 5 min in a microcentrifuge. The supernatant (10 μl) was analyzed on a sodium dodecyl sulfate (SDS)–10% polyacrylamide gel and stained with Coomassie blue to determine the presence of recombinant protein. DNA was prepared from selected colonies and sequenced with 32p-dATP (New England Nuclear) and Sequenase (U.S. Biochemical) to confirm the correct MZF1 sequence. A convenient Smal restriction site in the MZF1 sequence 5′ to zinc finger domain 5 allowed for the subcloning of fingers 5 to 13 without PCR amplification. Briefly, ZNF1.8 was sequentilly digested with Smal and EcoRI, and the 1.0-kb DNA fragment was cloned into a polylinker variant of the pDS56 vector.

To engineer an MZF1 expression vector for efficient in vitro transcription and translation, the 5′ untranslated region of MZF1 was deleted and a Kozak consensus sequence was introduced (19). The 5′ sense primer 5′-ATATATCGATTCC-3′ cloned sequence fingers 5 with ClaI, filling in fragment DNA pDS56 primer site DNA. The purified DNA sequence cloned into the Escherichia coli expression vector MZF1-7ZF+. The Escherichia coli expression vector for full-length MZF1 protein was produced by digesting MZF1-7ZF+ with ClaI, filling in the 5′ overhang with the large fragment of DNA polymerase (Klenow), and digesting with HindIII. The MZF1 fragment was subsequently subcloned into the polylinker variant of the pDS56 expression vector (14) and verified by DNA sequence analysis.

**Purification of recombinant proteins.** One hundred milliliters of a saturated bacterial culture was used to inoculate 1 liter of Luria broth containing antibiotics. After 45 min of growth, expression of protein was induced for 2 h with 1 mM IPTG. The bacteria were harvested and lysed in 20 ml of 6 M guanidine-HCl (pH 8.0)–50 mM sodium phosphate for 2 h at room temperature or overnight at 4°C. The lysates were clarified by centrifugation at 40,000 × g for 30 min, and the histidine fusion proteins were purified by nickel chelate affinity chromatography (1, 14). The purified proteins were dialyzed extensively in binding buffer (25 mM HEPES [N-2-hydroxyethylpiperazine-N′-2-ethanesulfonic acid]–KOH [pH 7.5], 50 mM KCl, 10 μM ZnSO4, 10% glycerol, 0.1% Nonidet P-40, and 1 mM dithiothreitol) containing 1 mM phenylmethylsulfonyl fluoride or phosphate-buffered saline containing 10 μM ZnSO4 and analyzed for purity on SDS–10% or 15% polyacrylamide gels.

**Design of degenerate oligonucleotide libraries.** An oligonucleotide library containing diverse binding sites (6) was created by mixing all four nucleotides (A, C, G, and T) at a defined number of positions during synthesis. To optimize selection of DNA-binding sites for each domain and to simplify analysis of sequences for a DNA consensus, two separate libraries were prepared. The first library, N14, was synthesized with 14 consecutive degenerate nucleotide positions and contained 414 possible nucleotide sequences. The second library, N20, was synthesized with 20 degenerate oligonucleotide positions and contained 420 possible nucleotide sequences. To facilitate PCR amplification and subsequent cloning of the oligonucleotide library, the degenerate nucleotide positions were flanked by defined sequences containing either a BamHI or an EcoRI restriction endonuclease site (see Fig. 3A). A 5′ sense primer and a 3′ antisense primer complementary to the oligonucleotide flanking sequences were synthesized and used for PCR amplification.

**Selection of DNA-binding sites.** The degenerate oligonucleotide library was end labeled with [γ-32p]ATP (New England Nuclear) and T4 polynucleotide kinase (New England Biolabs) and used as a DNA probe in gel shift assays (GSA). Five to 500 ng of purified, recombinant protein was incubated for 10 min in binding buffer containing 0.1 μg of poly(dI-dC) (Pharmacia) per μl. One hundred thousand counts per minute of labeled oligonucleotide probe was added to each reaction mixture, and the mixture was incubated for an additional 10 min at room temperature. The DNA-protein complex was electrophoresed on native 5% polyacrylamide gels in running buffer with 0.5 × Tris-borate-EDTA at 4°C. The gels were dried on Whatman paper and exposed to film, and after alignment of the film with the gel the DNA-protein complex was excised from the dried gels. The gel slice was rehydrated in 10 mM Tris-CI (pH 7)–100 mM NaCl and crushed with an Eppendorf pestle, and the DNA was eluted at 37°C during shaking for 2 h. The DNA-containing supernatant was filtered through a 0.45-μm-pore-size syringe filter membrane, ethanol precipitated, resuspended in 50 μl of distilled water, and further purified on a G-25 spin column (Boehringer Mannheim). One- to 5-μl aliquots of the extracted DNA were used as templates for PCR amplification using 200 μM each dNTP, 1 U of Pfu polymerase in buffer 1, 2 mM MgCl2, and 50 pmol each of primers A and B (see Fig. 3A). The PCR conditions optimized amplification of short oligonucleotides (30). Briefly, each cycle consisted of 30-s DNA denaturation at 93°C, 2-min primer annealing at 45°C, a 2-min increase of temperature from 45 to 67°C, and a 2-min elongation at 67°C for 25 cycles. The PCR product was gel purified, ethanol precipitated, and applied to a G-25 spin column. DNA recovery was quantitated by UV A260. Two hundred nanograms of DNA was end labeled with [γ-32p]ATP and used as a DNA probe in subsequent rounds of GSA and PCR amplification.

**Screening bacterial colonies.** After three to five rounds of gel shift selection and PCR amplification, the affinity-selected oligonucleotides were extensively digested with BamHI and EcoRI and cloned into pGEM7Zf+ (Promega). Individual bacterial colonies were screened for cloned oligonucleotide insert by PCR amplification of miniprep DNA. Primers A and B (see Fig. 3A) were used to amplify a 45-bp fragment that was present in the clones which contained an inserted oligonucleotide. The PCR product from each clone was prepared for use as a DNA probe in GSA by one of two methods. Either the product was gel purified and 32p end labeled with T4 polynucleotide kinase or the PCR was performed with 32p-end-labeled PCR primers.

**DNA sequencing and analysis of selected oligonucleotides.** Both strands of the cloned oligonucleotides that bound to the recombinant proteins in GSA were sequenced with 32p-dATP and Sequenase (U.S. Biochemical) by using commercially available T7 and SP6 primers (Promega). The degenerate positions of the oligonucleotide were aligned by using the Genetics Computer Group (University of Wisconsin) (GCCG) sequence analysis Pileup program (13). After the degenerate oligonucleotide sequences were aligned, the flanking sequences were included in the derivation of the consensus sequence.

**GSA and binding site competitions.** Five to 500 ng of purified rZNF1.4 or rZNF5-13 protein in binding buffer (25 mM HEPES–KOH [pH 7.5], 50 mM KCl, 10 μM ZnSO4, 10% glycerol, 0.1% Nonidet P-40, and 1 mM dithiotheriol) was incubated with 0.1 μg of poly(dI-dC) per μl for 10 min at room temperature. A 32p-end-labeled synthetic oligonucleotide containing the binding site for either the rZNF1-4 protein (5′-GATCTAAAAGTGGGAGAAA-3′) or the rZNF5-13 protein (5′-GATCCGGCTGGTGGGAGAAA-3′) was
added and incubated for an additional 10 min. In competition assays, 50 ng of rZN 1-4 or rZN 5-13 protein was incubated with poly(dI-dC). A mixture of 10−10 M 32P-end-labeled oligonucleotide and a 100- to 1,000-fold excess of unlabeled oligonucleotide either homologous (1-4 or 5-13 synthetic oligonucleotide) or heterologous (NIL oligonucleotide 5′-GCC AATTATCTGTCTAAACCG-3′) was added to the protein and incubated for 10 min at room temperature. DNA-protein complexes were analyzed on a native 5% polyacrylamide gel, and the free probe and bound probe were quantitated on a PhosphorImager (Molecular Dynamics). Calculations of the apparent (app) equilibrium constants were performed by the method of Carey (8): K_{app} = [protein] × [DNA]/[protein-DNA complex]. Under conditions of limiting [DNA], [protein]_{app} = [protein]_{eq} and the protein concentration required for half-maximal binding approximates the K_{app}.

Southwestern (DNA-protein) blots. Ten micrograms of affinity-purified rZN 1-4 (in binding buffer) or rZN 5-13 (in 8 M urea−50 mM sodium phosphate) protein was separated on an SDS−12.5% polyacrylamide gel and transferred to nitrocellulose (BA45; Schleicher & Schuell) by semidy electroblotting (Sartorius) at 70 mA for 90 min. Filters were blocked overnight at 4°C in binding buffer (see “GSA and binding site selections”) containing 5% dry milk. Filters were washed twice with binding buffer for 10 min and probed for 4 h at 4°C with 106 cpm of a 32P-end-labeled oligonucleotide per ml in binding buffer containing 10 μg of poly(dI-dC) per ml. The probe was decanted, and filters were washed twice with binding buffer for 10 min and air dried.

RESULTS

Expression and purification of MZF1 histidine fusion proteins. The amino acid sequence of MZF1 is shown in Fig. 1 (15). The 13 zinc fingers occur in two separate domains: the first domain contains zinc fingers 1 to 4, and the second domain contains zinc fingers 5 to 13 in the COOH terminus. Recombinant proteins containing either set of fingers were produced by cloning zinc fingers 1 to 4 (aa 108 to 240) or zinc fingers 5 to 13 (aa 218 to 485) into a histidine fusion protein expression vector (Fig. 2A) (14). The recombinant histidine fusion proteins were expressed in E. coli, purified by nickel chelate chromatography (1), and examined by SDS-polyacrylamide gel electrophoresis (SDS-PAGE) (Fig. 2B). In order to identify the individual zinc fingers critical for DNA binding, subdomains of the rZN 5-13 DNA-binding region were produced and purified (Fig. 2). These proteins contained zinc fingers 5 to 8, 8 to 13, 8 to 10, or 11 to 13. Each protein was stably expressed in E. coli and, with the exception of rZN 5-13, was purified to greater than 95% homogeneity (Fig. 2B).

Strategy for DNA-binding site selection. The strategy we used for the isolation of a DNA consensus binding site for the MZF1 DNA binding domains is outlined in Fig. 3. Briefly, synthetic oligonucleotides contained either 14 or 20 degenerate bases, flanked by defined nucleotide sequences containing restriction sites and complementary to PCR primers (Fig. 3A). The oligonucleotides were amplified by PCR, and the resulting double-stranded oligonucleotide libraries were end labeled with 32P and incubated with the purified recombinant zinc finger proteins (Fig. 3B). The DNA-protein complexes were

FIG. 2. Production of recombinant MZF1 histidine fusion proteins. (A) Domains present in the recombinant MZF1 proteins. Recombinant MZF1 fusion proteins that contained the amino acid sequence NH2-MSRGHHHHHHH followed by the indicated amino acids of MZF1 were produced. (B) SDS-PAGE of purified recombinant proteins. Recombinant MZF1 histidine fusion proteins were expressed in E. coli and purified as described in Materials and Methods. The purified proteins were analyzed on an SDS−10% polyacrylamide gel and stained with Coomassie blue. MW, molecular weight markers.
Identification of a binding site for MZF1 fingers 1 to 4. The rZN 1-4 protein and the N14 oligonucleotide library were used to enrich for DNA sequences recognized by fingers 1 to 4 in GSA. Each round of affinity selection with the rZN 1-4 protein produced a significant enrichment of sequences in the heterogeneous oligonucleotide library that bound to the protein (Fig. 4A). The enrichment was detected by two criteria. First, the formation of DNA-protein complexes after each round of selection required lower concentrations of protein. Second, an increased percentage of the heterogeneous oligonucleotide library was capable of forming a complex with the rZN 1-4 protein. Thus, the selection procedure enriched for DNA-binding sites from the N14 oligonucleotide library for the rZN 1-4 protein.

After three rounds of selection, the affinity-selected oligonucleotides were cloned into pGEM7Zf+, and individual clones were analyzed by GSA (Fig. 4B). Approximately 60% of the clones contained oligonucleotides which bound to the rZN 1-4 protein. The plasmids that contained DNA-binding sites for the rZN 1-4 protein were sequenced, and the region containing the original 14-base degeneracy was aligned with the GCG Pileup program (13). The 20 oligonucleotide sequences used in the rZN 1-4 binding site analysis are listed in Fig. 4C. More than 90% of the individual clones contained four contiguous G residues followed by an A and usually preceded by 5'-AGT-3'. The derived DNA consensus binding site for the first MZF1 DNA-binding domain of zinc fingers 1 to 4 is 5'-AGTGGGGA-3'.

Identification of a binding site for MZF1 fingers 5 to 13. The procedure for selecting a DNA-binding site from the oligonucleotide library for the rZN 5-13 protein was done, with two exceptions, in a fashion similar to that used for the rZN 1-4 protein. First, in order to accommodate a longer contiguous set of nine fingers, an oligonucleotide library with a degeneracy of 20 nucleotides was used with the rZN 5-13 protein. Second, five successive rounds of GSA and PCR amplification were performed before the oligonucleotide population was cloned. Again, each successive round of GSA selection and PCR amplification enriched the N20 oligonucleotide library for DNA sequences that bind to the rZN 5-13 protein (Fig. 5A). As with the selection of a binding site for the rZN 1-4 protein, enrichment of DNA sequences recognized by the rZN 5-13 protein is detected by the formation of DNA-protein complexes at lower protein concentrations and by an increased percentage of the oligonucleotide library forming a DNA-protein complex in GSA.

After five consecutive rounds of GSA selection, the oligonucleotides were cloned into pGEM7Zf+. Sixty percent of the individual bacterial colonies contained oligonucleotide sequences that bind to the rZN 5-13 protein in GSA (Fig. 5B). Alignment of 16 oligonucleotide clones was done with the GCG Pileup program (13) on the basis of the 20 degenerate nucleotides, and the sequences are shown in Fig. 5C. The nine-zinc finger DNA-binding domain recognizes a longer DNA-binding site. A 9-nucleotide core sequence, 5'-GAGGGGAA-3', that contains specific nucleotide bases that occur more than 60% of the time is present. This core DNA sequence is preceded by three nucleotides (5'-CGGn-3') which are present in 40 to 50% of the oligonucleotides analyzed, suggesting that these positions are also contributing to DNA binding. Thus, the derived DNA consensus binding site for the second MZF1 DNA-binding domain of zinc fingers 5 to 13 is 5'-CGGNGAGGGGAA-3'.

Enrichment for both rZN 1-4 and rZN 5-13 DNA-binding sites by affinity selection with either rZN 1-4 or rZN 5-13 proteins. On the basis of the analysis described above, the rZN
1-4 and rZN 5-13 proteins appear to select similar DNA consensus binding sites from two separate libraries of degenerate oligonucleotides. To determine whether the enrichment of DNA-binding sites with one protein (rZN 1-4) was also enriching for a population of oligonucleotides recognized by the other protein (rZN 13), we used the uncloned populations of DNA from each round of GSA selection by the rZN 1-4 protein of the N14 library also enriched for DNA-binding sites for the rZN 5-13 protein. Figure 6A clearly shows that enrichment of DNA-binding sites for the rZN 1-4 protein from the N14 oligonucleotide library also enriched for DNA-binding sites for the rZN 5-13 protein. The reciprocal is also true (Fig. 6B), since affinity selection of DNA-binding sites for the rZN 5-13 protein from the N20 oligonucleotide library concurrently enriched for DNA binding sites for the rZN 1-4 protein. Thus, both the rZN 1-4 and the rZN 5-13 proteins select similar G-rich DNA-binding sites from a heterogeneous population of oligonucleotides.

rZN 1-4 and rZN 5-13 proteins bind to similar DNA sequences. Oligonucleotides which contain the consensus binding sites for either the rZN 1-4 or the rZN 5-13 proteins were synthesized (Fig. 7) and used as probes in DNA binding assays. As expected, each of the rZN 1-4 and rZN 5-13 proteins bound to the oligonucleotide containing its derived DNA consensus binding site in both GSA (Fig. 7) and Southwestern blots (Fig. 8). The apparent equilibrium binding constants of rZN 1-4 and rZN 5-13 proteins for their respective binding-site oligonucleotides are 10 and 50 μM (data not shown). Binding to the oligonucleotide is specific and can be inhibited in GSA with an excess of unlabeled homologous oligonucleotide (containing the binding site) but not by an excess of a heterologous oligonucleotide (Fig. 9).

The similar DNA consensus sequences isolated by the two MZF1 DNA-binding domains and the coenrichment of DNA-binding sites for the rZN 1-4 and rZN 5-13 proteins from two separate oligonucleotide libraries suggest that the two domains recognize the same core DNA sequence. Indeed, both the rZN 1-4 and the rZN 5-13 proteins can also bind to the oligonucleotide that contains the DNA consensus binding site derived for the other DNA-binding domain (Fig. 7). Thus, MZF1 contains 13 zinc fingers arranged in two separate DNA-binding domains which recognize similar core DNA sequences.

The Southwestern blot demonstrated that the intact form of the rZN 5-13 protein is the active DNA-binding protein. To determine whether only a subset of the nine zinc fingers in the second MZF1 DNA-binding domain was sufficient for binding, additional recombinant MZF1 proteins that expressed subsets of fingers 5 to 13 (Fig. 2) were produced. The recombinant proteins containing zinc finger domain 5-8, 8-13, 8-10, or 11-13 were produced (Fig. 2A) and purified to greater than 95% homogeneity (Fig. 2B). Each recombinant zinc finger protein was tested for DNA binding with each synthetic oligonucleotide containing the derived binding site for either fingers 1 to 4 or fingers 5 to 13. None of the recombinant proteins expressing subsets of the zinc fingers from the second DNA-binding domain (fingers 5 to 13) recognized either of the two synthetic oligonucleotides in GSA (Fig. 7). The truncated proteins also did not bind to the individual DNA clones (Fig. 4B and 5B) that were isolated during the selection of the DNA-binding sites for the rZN 1-4 and rZN 5-13 proteins (data not shown). In addition, the rZN 8-13 protein was used in combination with the N20 oligonucleotide library to select a DNA-binding site. While the rZN 8-13 protein was able to
produce a detectable DNA-protein complex in the GSA, no enrichment of DNA sequences was observed, even after five rounds of GSA selection and PCR amplification (data not shown). Therefore, while this is not a comprehensive analysis, these preliminary results suggest that a full complement of the nine zinc fingers in the second MZF1 DNA-binding domain may be necessary to bind the derived DNA consensus binding site.

Full-length MZF1 binds both of the derived DNA consensus binding sites. The DNA-binding sites were isolated by using recombinant proteins which expressed only one of the two potential MZF1 DNA-binding domains. It was important to confirm that the full-length MZF1 protein also binds to the derived DNA consensus binding sites. Since attempts to purify recombinant full-length MZF1 protein were unsuccessful, E. coli lysates which express the full-length MZF1 protein were used. The full-length MZF1 protein bound to each synthetic oligonucleotide containing the derived DNA consensus binding site for zinc fingers 1 to 4 or 5 to 13 in GSA (Fig. 10). Thus, full-length MZF1 can bind the derived DNA-binding sites for each of the two DNA-binding domains.

**DISCUSSION**

We have produced recombinant proteins that contain either the first or the second DNA binding domain of MZF1. These recombinant proteins were used to isolate specific DNA sequences from libraries of degenerate oligonucleotides by using successive rounds of GSA and PCR amplification. We have shown that both sets of zinc fingers are able to bind DNA and confirm that the full-length MZF1 protein also binds to the derived DNA consensus binding sites. Since attempts to purify recombinant full-length MZF1 protein were unsuccessful, E. coli lysates which express the full-length MZF1 protein were used. The full-length MZF1 protein bound to each synthetic oligonucleotide containing the derived DNA consensus binding site for zinc fingers 1 to 4 or 5 to 13 in GSA (Fig. 10). Thus, full-length MZF1 can bind the derived DNA-binding sites for each of the two DNA-binding domains.
have identified a DNA consensus binding site for each of the DNA-binding domains of MZF1: rZN 1-4 binds 5'-AGTG GGGA-3', and rZN 5-13 binds 5'-CCGnGAGGGGGAA-3'. These two core DNA consensus sequences are very similar, despite the fact that the two DNA-binding domains of MZF1 are markedly different in amino acid sequence. Thus, MZF1 is a bipartite DNA-binding protein that recognizes two distinct DNA consensus sequences with a common G-rich core.

On the basis of our current understanding of zinc finger recognition of DNA, it was unexpected that fingers 5 to 13 would recognize a G-rich DNA sequence. Unlike fingers 1 to 4, the amino acids in the recognition helices of fingers 5 to 13 that are likely to contact DNA are mostly glutamine, an amino acid predicted to recognize an A-rich DNA sequence (11, 18). There are several possible interpretations to these findings. One is that zinc finger domains expressed in the context of multiple zinc fingers (i.e., more than 3 or 4) may no longer interact with the DNA as totally independent domains but may act cooperatively with the other zinc fingers to influence DNA recognition. Indeed, the inability of subdomains of zinc fingers 5 to 13 to bind to synthetic oligonucleotides containing the DNA consensus binding site for either zinc fingers 1 to 4 or zinc fingers 5 to 13 (Fig. 7) is preliminary evidence that suggests that all nine zinc fingers of the second MZF1 DNA-binding domain may be necessary for recognition of the core DNA consensus sequence. A second interpretation for the isolation of a G-rich binding site for the second MZF1

FIG. 8. Southwestern blot with synthetic oligonucleotides and recombinant MZF1 proteins. Approximately 10 μg each of rZN1-4 and rZN5-13 proteins was separated on an SDS–12.5% PAGE and transferred to nitrocellulose by semidry electrophotography. Filters were probed with 10^6 cpm of a [32P]-end-labeled oligonucleotide containing the binding site for either rZN1-4 or rZN5-13 per ml. (A) Coomassie blue-stained acrylamide gel of the proteins used in the Southwestern blots. (B) The intact rZN1-4 and rZN5-13 proteins bind to the oligonucleotides containing the consensus binding site for fingers 1 to 4 or that for fingers 5 to 13.

FIG. 9. The rZN 1-4 and rZN 5-13 proteins specifically bind to the consensus binding site oligonucleotides. GSA were done with 50 ng of protein preincubated in binding buffer containing 0.1 μg of poly(dI-dC) per μl. A mixture containing a [32P]-end-labeled binding site oligonucleotide and a 100- to 1,000-fold excess of unlabeled oligonucleotide, either homologous or heterologous (NIL) (see Materials and Methods), was added to the protein mixture. (A) rZN 1-4 specifically binds to the oligonucleotide containing the 1-4 binding site. (B) rZN 5-13 specifically binds to the oligonucleotide containing the 5-13 binding site. –, no competitor. Lane 1, no protein.
DNA-binding domain is that the previously identified amino acids that contact DNA within the Egr1 zinc finger domains may not be in the identical positions in proteins containing multiple contiguous zinc fingers. Identification of a DNA-binding site for the second MZF1 DNA-binding domain provides a valuable reagent to help elucidate the rules governing DNA recognition by multiple zinc finger domains.

The consensus sequences for the two MZF1 DNA-binding domains were isolated from a relatively short oligonucleotide (N14 or N20). This length would bias the selection towards a short sequence and preclude the isolation of a discontinuous binding site spanning more than 20 nucleotides. Thus, these consensus binding sites may represent only a core binding site. A preliminary search of the promoters of myeloid-specific genes has identified several potential candidate genes which contain multiple copies of both of these core binding elements. The identification of MZF1 DNA consensus binding sites in the context of putative target gene promoters provides the means to quantitatively analyze the recognition of DNA target sequences by the bipartite, full-length MZF1 protein.

The DNA binding sites recognized by the two MZF1 DNA-binding domains are reminiscent of the NF-kB DNA-binding site found in the promoters of major histocompatibility class I and beta interferon (reviewed in reference 2). It is of interest that NF-kB binding sites (5'-GGGACCTTCC-3') are recognized by an increasing number of proteins which can be divided into two general families: the Rel family (p50, p65, c-Rel, and v-Rel) and a zinc finger family (PRDII-BF1 [12], also known as MBP1 [3], and MBPII [28]). PRDII-BF1 (MBP1) contains two separate DNA-binding domains, each containing two Cys-His zinc fingers. As with MZF1, each of the two DNA-binding domains can independently bind to the same DNA sequence. However, unlike MZF1, the amino acid positions in both sets of the zinc fingers of PRDII-BF1 that may contact DNA are very similar (12). We are currently investigating the potential recognition and regulation of the NF-kB binding site by MZF1.

Another Cys-His zinc finger gene that has two separate DNA-binding domains and has been implicated in the regulation of hematopoiesis is Evi-1. Evi-1 contains 10 zinc fingers arranged in two separate domains. The amino-terminal domain contains one zinc finger separated by 25 amino acids from six contiguous zinc fingers, each separated by the conserved Krüppel-like H-C link (23). Three hundred eighty amino acids separate the second DNA-binding domain of three zinc fingers. Like MZF1, the two DNA-binding domains of Evi-1 recognize similar DNA sequences (15a). Thus, two Cys-His zinc finger genes with separate DNA-binding domains which have a potential role in transcriptional regulation of hematopoiesis have been identified.

Recently, two other genes which contain structural homologies to MZF1 have been reported. Kid-1 (kidney, ischemia, and developmentally regulated gene) was isolated from a rat cDNA library by using a degenerate oligonucleotide probe for the H-C link (32). Like MZF1, Kid-1 contains 13 zinc fingers arranged in two domains: fingers 1 to 4 and fingers 5 to 13. Fingers 4 and 5 are separated by 32 amino acids in Kid-1 and 24 amino acids in MZF1. There is 45% homology within the zinc finger domains (primarily the conserved amino acid positions in the Cys-His zinc finger motif) and no significant homology outside the zinc finger domains. Expression of mRNA for both Kid-1 and MZF1 is developmentally regulated and restricted to specific tissues. Thus, the 4-9 bipartite DNA-binding domain may represent a new subclass of Cys-His zinc finger genes that regulate tissue-specific gene expression during differentiation.

The second gene that has reported homology with MZF1 is promyelocytic leukemia zinc finger (PLZF) containing a single DNA-binding domain of nine contiguous zinc fingers (33). PLZF was isolated as a translocation fusion protein infrequently found in patients with acute promyelocytic leukemia and is produced by a rare t(11;17)(q23;q21.1) translocation with retinoic acid receptor-a gene. There is greater than 50% homology of the nine zinc fingers of PLZF with both of the DNA-binding domains of MZF1 and no significant sequence homology outside the zinc finger domains. Of great interest is the fact that both PLZF and MZF1 have mRNA expression limited to cells of the myeloid lineage. In contrast to MZF1, expression of PLZF mRNA is down-regulated by retinoic acid-induced differentiation of the human myeloid progenitor cell line HL60 (33). Thus, the homology of PLZF and MZF1 and their myeloid lineage-restricted expression suggest that these two genes may be members of a new family of zinc finger genes that are coordinately regulated to control hematopoietic differentiation.

MZF1 contains two DNA-binding domains which bind two distinct, albeit similar, DNA consensus target sequences. The ability of a transcription factor to bind more than one DNA sequence at the same time can potentially provide multiple levels for differentially regulating target gene expression. One level is at the organization of the target gene promoter; promoters that contain either a single binding site or multiple binding sites can be differentially regulated at the level of transcription. A second level could be the developmentally regulated production of alternative mRNA transcripts of MZF1 that contain only one or both of the DNA-binding domains. Thus, expression of specific subsets of target genes could be differentially regulated by producing a monor bipartite DNA-binding protein. A final level in regulating gene expression could be that the two DNA-binding domains of MZF1 may allow for coordinate regulation of two independent target gene promoters. This would allow MZF1 to orchestrate the simultaneous regulation of a family of genes that may be necessary to initiate the altered program of gene expression which commits the progenitor cell to differentiate along a specific myeloid cell lineage. We are currently investigating the roles of both the MZF1 DNA-binding domains in DNA binding and the role of MZF1 in the transcriptional regulation of target genes during myeloid differentiation.
TRANSCRIPTIONAL REGULATION OF HEMATOPOIESIS

ACKNOWLEDGMENTS

We gratefully acknowledge the expert preparation of the manuscript by M. Marinelli. We thank S. Berger, N. Gaill, and D. Cook for critically reviewing the manuscript and J. Ihle for helpful discussions and suggestions.

These studies were supported in part by grants HL48914 (R.H.), CA52009 (F.J.R.), and CA47983 (F.J.R.) and core grant CA10815 (F.J.R.) from the National Institutes of Health and grants from the W. W. Smith Charitable Trust (F.J.R.), the Irving A. Hansen Memorial Foundation (F.J.R.), and the Mary A. H. Rumsey Foundation (F.J.R.). F.J.R. is a Pew Scholar in the Biomedical Sciences. J.F.M. is a Cancer Research Institute Fellow.

REFERENCES

16. Ihle, J. Personal communication.

Downloaded from http://mcb.asm.org/ on January 4, 2018 by guest