MOLECULAR AND CELLULAR BIOLOGY, Nov. 1995, p. 5957–5965 Vol. 15, No. 11
0270-7306/95/$04.00+0

Isolation of a Gene Encoding a Functional Zinc Finger Protein Homologous to Erythroid Krüppel-Like Factor: Identification of a New Multigene Family

KATHLEEN P. ANDERSON,* CHRISTINE B. KERN, SCOTT C. CRABLE, AND JERRY B. LINGREL

Department of Molecular Genetics, Biochemistry, and Microbiology, University of Cincinnati College of Medicine, Cincinnati, Ohio 45267-0524

Received 28 April 1995/Returned for modification 26 May 1995/Accepted 7 August 1995

We have identified and characterized the gene for a novel zinc finger transcription factor which we have termed lung Krüppel-like factor (LKLF). LKLF was isolated through the use of the zinc finger domain of erythroid Krüppel-like factor (ELKF) as a hybridization probe and is closely related to this erythroid cell-specific gene. LKLF is expressed in a limited number of tissues, with the predominant expression seen in the lungs and spleen. The gene is developmentally controlled, with expression noted in the 7-day embryo followed by a down-regulation at 11 days and subsequent reactivation. A high degree of similarity is noted in the zinc finger regions of LKLF and EKLF. Beyond this domain, the sequences diverge significantly, although the putative transactivation domains for both LKLF and EKLF are proline-rich regions. In the DNA-binding domain, the three zinc finger motifs are so closely conserved that the predicted DNA contact sites are identical, suggesting that both proteins may bind to the same core sequence. This was further suggested by transactivation assays in which mouse fibroblasts were transiently transfected with a human β-globin reporter gene in the absence and presence of an LKLF cDNA construct. Expression of the LKLF gene activates this human β-globin promoter containing the CACCC sequence previously shown to be a binding site for EKLF. Mutation of this potential binding site results in a significant reduction in the reporter gene expression. LKLF and EKLF can thus be grouped as members of a unique family of transcription factors which have discrete patterns of expression in different tissues and which appear to recognize the same DNA-binding site.

Zinc finger domains have emerged as a major class of DNA-binding motifs. Since their initial discovery with the *Xenopus* transcription factor TFIIIA (7, 37), hundreds of proteins possessing these domains have been described (5, 25). At least five subgroups based on the type and spacing of the amino acids involved in the zinc chelation have been defined within the broad description of zinc finger proteins. TFIIIA and zif268, a mouse immediate-early protein (10), for example, have two cysteine and two histidine residues that contact the zinc ion, in the form of Cys-X_{2-4}-Cys-X_{12}-His-X_{3-4}-His (37, 42). Two other subclasses both utilize four cysteines to chelate the zinc, but the spacing is quite different, leading to some variation in the three-dimensional structure and the function of these finger regions (16, 17, 19, 48, 49, 53). Proteins that fall into these two categories include steroid receptors and the GATA transcription factor family. As examples of the spacing differences seen between these subgroups, in the estrogen receptor the sequence for its two zinc fingers follows the format Cys-X_{2-5}-Cys-X_{9-13}-Cys-X_2-Cys while the two finger regions in GATA-1 have the form Cys-X_2-Cys-X_{17}-Cys-X_2-Cys. Although these spacing differences may not appear to be dramatic, 6 of the 17 amino acids in the loop region of the GATA fingers are invariant, adding to the specificity of this structure as compared with the steroid receptor group. A fourth category has the structure Cys-X_5-Cys-X_{12}-His-X_4-Cys and is exemplified by myelin transcription factor I (MyTI) (28). Finally, Gal4, a positive regulator of galactose metabolism in *Saccharomyces cerevisiae*, has a cluster of six cysteines chelating the zinc ions (27).

While these proteins can be included in a superfamily of zinc finger proteins, the other domains of the individual proteins are totally unrelated. This is despite the fact that these proteins all function to bind DNA and in some manner influence transcriptional activity. The activation domains may be localized regions rich in a particular amino acid, such as glutamine, proline, or acidic residues. These modular activation domains are seen in several members of the Cys_2-His_2 class of zinc finger proteins (38). Alternatively, in the Cys_4 GATA family of transcription factors, the activation domain does not fall into one of these prototypic categories. Rather, there are three regions scattered throughout the molecule, on both sides of the DNA-binding domain, that are important in the transactivation function of these proteins (52).

Erythroid Krüppel-like factor (EKLF) is an erythroid cell-specific transcription factor first described by Miller and Bieker (36), which contains a zinc finger motif matching the Cys_2-His_2 class of DNA-binding domains. Another member of this class termed zif268 has been crystalized in the presence of its consensus nucleotide-binding site (42), and this structure has provided a foundation for developing predictions concerning the consensus binding site for EKLF and other Cys_2-His_2 zinc fingers (18, 30, 39). On the basis of this model, EKLF was expected to bind the site CCN CNC CCN. This sequence encompasses the CACCC site which is present in the 5′ region of the human and mouse β-globin promoter, as well as other erythroid cell-specific genes. Subsequent studies have shown that EKLF does indeed bind this site and transactivate the globin gene promoter (6, 18, 36). The activation domain in this

* Corresponding author. Mailing address: Department of Molecular Genetics, Biochemistry, and Microbiology, University of Cincinnati College of Medicine, 231 Bethesda Ave., Cincinnati, OH 45267-0524.

transcription factor contains a large number of proline residues and retains its function when produced as a fusion protein with a different DNA-binding domain, indicating the modular nature of these regions of the protein (6). EKLF appears to function as a regulator of β-globin gene expression by binding to the CACCC element, which has been previously shown to be critical for transcription from this gene (2, 12, 22, 23, 45). In addition, recent studies both in vitro (15) and in vivo (40, 44) have demonstrated a potential role for EKLF in the developmental switch between the fetal γ-globin gene and the adult β-globin gene. Furthermore, the proximal promoter region for EKLF has been cloned and analyzed. A binding site for the erythroid cell-specific transcription factor GATA-1 at −60 from the EKLF transcription start site was determined to be critical in the regulation of the EKLF gene (14).

The expression of EKLF thus appears to be dependent on, or coincident with, the expression of GATA-1, another erythroid cell-specific transcription factor. GATA-1 was the first member of a multigene family named for the consensus nucleotide-binding site. As mentioned above, these genes all encode proteins containing two zinc fingers with a Cys_4 configuration for the chelation of the zinc ion. Several genes have been isolated thus far as members of this family; all of them have homology in the zinc finger region but diverge significantly in the remainder of the molecule (17, 21, 33, 34, 48, 49, 51, 53). Although the core DNA-binding site is similar for all these proteins, the tissue distribution pattern differs for each GATA gene, dictating their site of action in association with other factors. GATA-1 and -2, for example, are expressed in hematopoietic cells, in mast cells, and in megakaryocytes (17, 49, 51), while GATA-3 is expressed predominantly in T lymphocytes and brain (21, 33) and GATA4/5/6 are transcribed in embryonic heart and gut (34).

We describe here the isolation and characterization of a novel zinc finger gene related to EKLF. Similar to the GATA family of transcription factors, the identification of this gene denotes the emergence of a multigene family with homology to EKLF. We have named this factor lung Krüppel-like factor (LKLF) to specify the site of predominant expression and underscore the relationship to EKLF.

MATERIALS AND METHODS

Cloning of the LKLF gene and cDNA. A mouse genomic library from strain 129 was obtained (3). The library was constructed in the Lambda DASH vector (Stratagene, La Jolla, Calif.), and the manufacturer's protocol was followed for plating and lifting the filters. After being baked, the filters were washed in 50 mM Tris-HCl (pH 8)–1 M NaCl–1 mM EDTA–0.1% sodium dodecyl sulfate (SDS) for 1 h at 65°C with shaking. A single solution composed of 6× SSC, 5× Denhardt's solution, 0.1% SDS, and 0.1 mg of denatured DNA per ml was used for the prehybridization and hybridization buffer (1× SSC is 0.15 M NaCl plus 0.015 M sodium citrate; 1× Denhardt's solution is 0.02% polyvinylpyrrolidone, 0.02% Ficoll 400, and 0.02% bovine serum albumin). The filters were prehybridized for >2 h at 65°C. A denatured, radiolabelled probe was then added directly to this solution, and the hybridization was continued at 65°C overnight. The probe for this genomic screening was a reverse transcriptase-PCR product spanning EKLF cDNA nucleotides 852 to 1234 (36), which includes the three zinc fingers. The template for this PCR was mouse erythroleukemia cell RNA. After hybridization with this probe, the filters were washed to a stringency of 0.5× SSC–0.1% SDS at 65°C. Sixty-seven positive phage clones were obtained from this screening. Of these, 29 corresponded to the mouse EKLF gene, while the majority of the remaining 38 clones fell into a second class representing LKLF.

A PCR fragment was generated from one of these LKLF genomic clones for the screening of a mouse lung cDNA library. The probe was 249 nucleotides in length, containing 76 bp of zinc finger exonic sequence corresponding to nucleotides 858 to 934 in the cDNA sequence shown in Fig. 3. The mouse lung cDNA library was purchased from Stratagene, La Jolla, Calif., and was prepared in the Lambda ZAP II vector. The library was screened as described above. Thirty positive clones were isolated and characterized after excision of the pBSSK plasmid containing the cDNA insert from the phage vector.

Northern blot analysis of LKLF. Poly(A) RNA Northern (RNA) blots of mouse tissues were purchased from Clontech Laboratories, Inc., Palo Alto, Calif. The mouse multiple-tissue blot contained approximately 2 μg of mRNA per lane from mouse heart, brain, spleen, lung, liver, skeletal muscle, kidney, and testis. The developmental blot had 2 μg of poly(A) RNA from mouse embryos at days 7, 11, 15, and 17 of gestation. The samples labeled 7 day and 11 day also included placenta and yolk sac. The manufacturer's protocol was followed concerning hybridization solutions and stripping and reuse of the blots. A rat tissue poly(A) Northern blot was also prepared. RNA was extracted from the stomach, small intestine, large intestine, and uterus, in addition to the series of tissue samples described for the mouse multiple-tissue blot. All blots were probed with fragments of the LKLF gene. The mouse multiple-tissue blot was hybridized with a 3.5-kb *Eco*RI fragment which included LKLF coding sequence from the 5′ end through the first zinc finger. The blots containing RNA from the mouse embryos and the rat tissues were probed with the PCR fragment used to screen the cDNA library as described above. In addition, all three Northern blots were probed with the EKLF reverse transcriptase-PCR product used to screen the mouse genomic library and with an actin control probe supplied with the blots. No cross-hybridization between the EKLF and LKLF probes was noted under the conditions used in these analyses.

Nucleotide sequence analysis. Mouse genomic subclones and lung cDNA clones were subjected to double-stranded nucleotide sequence analysis with a model 373 DNA sequenator and the *Taq* DyeDeoxy sequencing protocol (Applied Biosystems, Inc., Foster City, Calif.). The FASTA (43) and BLAST (1) programs were used to search the GenBank and EMBL nucleotide databases and the Swiss-Prot and PIR/NBRF protein sequence databases.

Determination of the 5′ end of the LKLF transcript. Rapid amplification of cDNA ends (RACE) (4, 20) was performed to isolate the 5′ end of the LKLF transcript and clone this product for analysis. A 5′-AmpliFINDER RACE kit was used (Clontech Laboratories, Inc.) with some modifications. Total mouse lung RNA was isolated by the method of Chomczynski and Sacchi (9). Poly(A) RNA was then prepared with an Oligotex-dT mRNA mini-kit (Qiagen, Chatsworth, Calif.). cDNA was synthesized according to the AmpliFINDER protocol, with 5′-GGCGCGGGAGGCCCGT-3′ (nucleotides 671 to 686) as the LKLF-specific primer. After ligation of the anchor sequence, a PCR amplification was performed with the corresponding anchor primer and a nested LKLF primer, 5′-GAGGCGCAAGGAGGAAGC-3′ (nucleotides 374 to 392). A 1-μl aliquot of this reaction mixture was subsequently subjected to an additional round of PCR amplification with the same anchor primer and a third LKLF-specific primer, 5′-GTTGTTTAGGTCCTCATCCG-3′ (nucleotides 152 to 171). A fragment of approximately 170 bp was recovered and subjected to nucleotide sequence analysis. This sequence was found to include a cDNA sequence matching our genomic and cDNA clones, and it extended 39 nucleotides 5′ of our longest LKLF cDNA clone.

Plasmid constructs. A full-length LKLF cDNA clone was prepared by using PCR to amplify a 76-bp 5′ region to extend the longest cDNA clone obtained from the library screen to the translation start site. This fragment, extending from the *Nco*I site at the start of translation to the *Eco*47III site, was cut and ligated to replace the smaller 5′ end of the existing clone. This full-length LKLF cDNA fragment was subcloned into the expression vector pBK-CMV (Stratagene, La Jolla, Calif.), which utilizes the cytomegalovirus immediate-early promoter. The reporter construct used in the initial transactivation assays is HS2-βCAT, containing a 1.6-kb *Kpn*I-*Pvu*II HS2 fragment. The human β-globin promoter region from −265 to the translation start site was obtained by PCR amplification. Restriction sites were added to the primers such that this fragment could be cloned with *Sal*I-*Hin*dIII. By using these sites, the fragment was subcloned into a Bluescript vector (Stratagene) containing the chloramphenicol acetyltransferase (CAT) gene in the *Hin*dIII-*Bam*HI sites. The HS2 fragment was subsequently added. All PCR amplified fragments were verified by nucleotide sequence analysis.

Since a CACCC site is present in the HS2 fragment, a second reporter construct, βCAT, which lacks this region was also prepared. Mutations were then introduced to destroy the two consensus CACCC elements in the β-globin promoter. A 15-nucleotide portion surrounding the more proximal 3′ element was deleted by digestion with *Dra*III and *Avr*II. Religation of the plasmid produced mut3′βCAT. The 5′-distal CACCC was mutated by PCR amplification. A mutant primer, 5′-CCGAT**ATCTCTAGA**TCTAAGTGATGACAGC-3′, and the T3 universal primer in the vector were used to amplify from −265 to −99 at the *Dra*III site. The original βCAT plasmid was then digested with *Sal*I and either *Dra*III or *Avr*II. The resulting constructs contained either the 5′ CACCC mutation alone, mut5′βCAT, or a double mutation of both CACCC elements, mut5′/3′βCAT. Nucleotide sequence analysis was performed to verify all the plasmid constructs.

Transfections and CAT assays. Mouse NIH 3T3 fibroblasts were plated at a density of 1.5×10^6 cells per 150-mm-diameter dish. The following day, DNA was introduced by $CaPO_4$ precipitation (29) with 20 μg of HS2-βCAT plasmid or 15 μg of the plasmids lacking the HS2 fragment and 10 μg of pSV2luc control plasmid used in each transfection. Different amounts of the LKLF cDNA were included as indicated in the figure legends. The total amount of DNA in each transfection was brought to 50 μg with salmon sperm or plasmid DNA. The medium was changed 24 h after the $CaPO_4$ treatment, and the cells were harvested 48 h after transfection. Cell extracts were prepared by freeze-thaw lysis in 0.1 M Tris (pH 7.8). An aliquot was removed for analysis of the luciferase activity from the transfection control plasmid. The remaining extract was heat inacti-

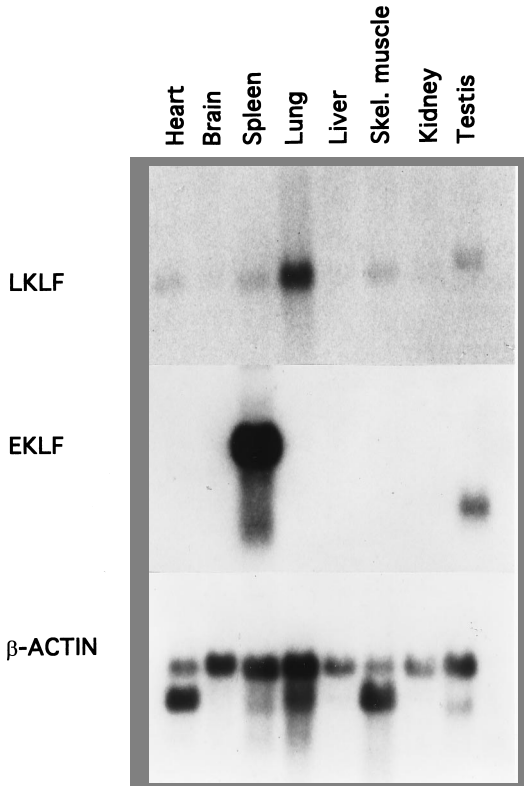

FIG. 1. Northern blot analysis of LKLF expression in mouse tissues. A Northern blot containing 2 μg of poly(A) RNA per lane was obtained from a commercial supplier. The blot was probed with a genomic fragment of LKLF containing the 5′ end through the first zinc finger region. After stripping, the blot was reprobed with a PCR fragment corresponding to the three zinc finger regions of EKLF. Lastly, a β-actin probe was used as a control for the integrity of the poly(A) RNA.

vated at 65°C for 10 min. Protein concentrations of the extracts were determined by the bicinchoninic acid assay (Pierce, Rockford, Ill.). Extract amounts were normalized for transfection efficiencies, and CAT assays were performed with a 1-h incubation. The thin-layer chromatography plates were exposed to a PhosphorImager plate (Molecular Dynamics, Sunnyvale, Calif.) for quantitation. Normalized values for the CAT activity are based on percent conversion of [^{14}C]chloramphenicol substrate to the acetylated forms, corrected for transfection efficiency.

Nucleotide sequence accession number. The LKLF cDNA sequence shown in Fig. 3 has been deposited in GenBank and given accession number U25096.

RESULTS

LKLF is expressed predominantly in the lung and is developmentally regulated. The LKLF gene was isolated from a mouse genomic library using the EKLF zinc finger region as a hybridization probe (unpublished data). Nucleotide sequence analysis revealed a closely related zinc finger motif in LKLF. A fragment from this genomic clone, including the first zinc finger region, was then used as a hybridization probe on a tissue distribution Northern blot. As shown in Fig. 1, the predominant site of expression of LKLF is the adult mouse lung. Some expression is also seen in the heart, spleen, skeletal muscle, and testis. The transcript size is approximately 1.5 kb in all tissues except the testis, where the mRNA is slightly larger at 1.65 kb. These signals were quantitated by PhosphorImager analysis. By normalizing to the weaker hybridization signal in the heart, the following numbers were calculated for the relative mRNA expression levels in the tissues shown: heart, 1; skeletal muscle, 1.9; testis, 2.3; spleen, 2.5; and lung, 10.9. Since the blot was prepared with approximately 2 μg of poly(A) RNA, these numbers represent a reasonable comparison of the level of LKLF expression in each RNA population.

The same Northern blot was stripped and rehybridized with the EKLF zinc finger probe, originally used in the genomic library screen. These results are shown in Fig. 1. As expected, EKLF is highly expressed in the mouse spleen. Interestingly, EKLF is also apparently transcribed in the testis. As seen with LKLF, the testis transcript differs in size from the predominant erythroid form of EKLF. In this case, EKLF mRNA is 1.5 kb while the testis transcript is only 0.85 kb. Finally, the blot was stripped and reprobed with the β-actin control, as shown in Fig. 1. This probe recognizes two forms of β-actin with mRNA sizes of 1.8 and 2 kb. Because different tissues may express actin isoforms in differing amounts, the actin control serves primarily to ensure the presence and integrity of the poly(A) RNA.

An extended tissue distribution analysis was conducted with rat tissue. We switched to rats for this experiment to obtain reasonable amounts of poly(A) RNA. A rat tissue Northern blot probed with the LKLF zinc finger region is shown in Fig. 2. As seen in the mouse, the lung is a principal site of LKLF expression, with lower levels observed in the heart and skeletal muscle. Although LKLF expression was detected in mouse spleen, the level of this expression in the rat is nearly equivalent to that seen in the lung tissue. Quantitation of the hybridization signal from a PhosphorImager analysis gave the following normalized values: uterus, 1; skeletal muscle, 1.4; heart, 1.7; spleen, 4.4; and lung, 5.2. This blot was stripped and reprobed with the β-actin hybridization probe, as shown in Fig. 2. This result demonstrates that intact mRNA is present in all the lanes and, with the possible exception of the liver sample, in roughly equivalent amounts. Although exact quantitation and comparison of mRNA levels between tissue types are difficult, both blots establish a similar pattern of expression in the 12 tissues analyzed, with the lung expressing the highest level of LKLF in each case.

A similar Northern blot analysis was also performed to assess the developmental expression pattern of LKLF. Poly(A)

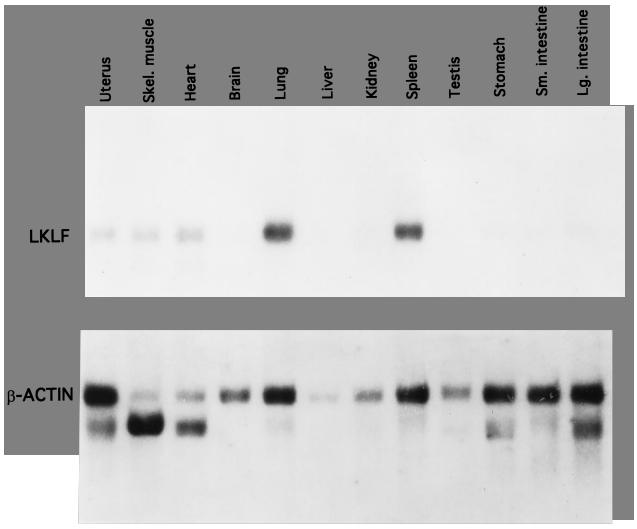

FIG. 2. LKLF expression in rat tissues. Poly(A) RNA was extracted from 12 different rat tissues. Approximately 2 μg of each sample was loaded per lane in preparing a Northern blot. The blot was probed with a genomic fragment of mouse LKLF. The blot was subsequently stripped and rehybridized with a β-actin probe.

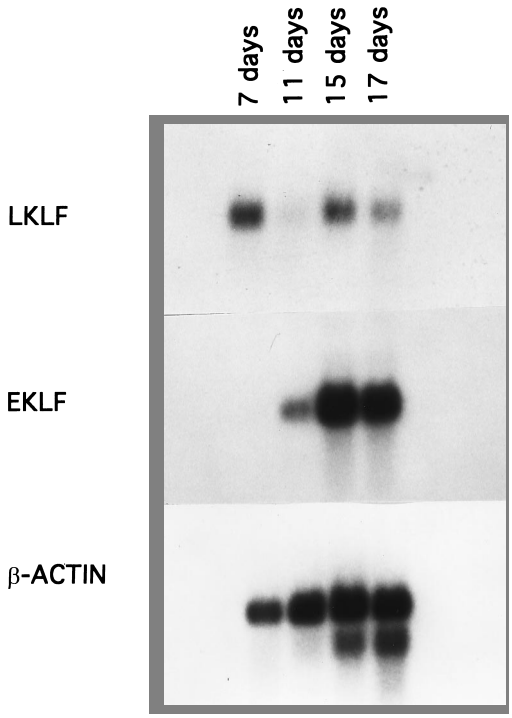

FIG. 3. Northern blot analysis of LKLF developmental expression. Poly(A) RNA was prepared from mouse embryos at 7, 11, 15, and 17 days of gestation. A Northern blot was produced with 2 μg of poly(A) RNA (Clontech) per lane. The blot was probed with a genomic fragment of LKLF followed by a reverse transcriptase-PCR fragment of EKLF and thereafter with a β-actin control probe as described in Materials and Methods.

RNA prepared from mouse embryos at 7, 11, 15, and 17 days of gestation was blotted and hybridized with probes specific for LKLF, EKLF, and β-actin. The results are shown in Fig. 3. LKLF is expressed at the early primitive streak, or gastrulation stage, at day 7 and then shuts off in the day 11 embryo. The gene is then reactivated in the day 15 embryo and continues in the later-stage day 17 embryo. This is in contrast to the pattern of expression of EKLF, also shown in Fig. 3. The transcript for EKLF is present in the day 11 embryo when hematopoiesis is occurring in the yolk sac and then increases in abundance in the day 15 and 17 embryos. On long exposures, a very low level of EKLF expression can also be seen in the day 7 embryonic sample.

Isolation of the LKLF cDNA. Having established the lung as the primary site of expression of LKLF, we obtained and screened a mouse lung cDNA library. Thirty overlapping clones were isolated and characterized. On the basis of data obtained from these clones, the nucleotide and deduced amino acid sequence of the LKLF cDNA are presented in Fig. 4. Figure 4 also includes approximately 30 nucleotides of genomic sequence at the 5′ end which was derived from a genomic LKLF clone and is included to indicate the position of a consensus TATAA site. Although several of the cDNA clones included the 3′ poly(A) tract, none extended to the full length of the transcript. However, on the basis of comparisons with EKLF and the known transcript size for LKLF from our Northern blot analyses (Fig. 1 to 3), one clone was predicted to be within 100 nucleotides of the 5′ end. Nested primers were prepared to this clone, and the RACE procedure was used to determine the 5′ end of the LKLF transcript in lung mRNA. Nucleotide sequence analysis of the PCR product obtained from this procedure exactly matched our genomic clone from the 5′ region of the LKLF gene. The start site derived from this RACE product is marked in Fig. 4. In addition, a primer extension analysis was performed with lung mRNA and indicated the same transcription start site (data not shown). At 31 nucleotides upstream of this initiation site is a consensus TATAA in the genomic DNA sequence for binding of the RNA polymerase to this promoter. The first ATG codon matches the Kozak consensus sequence (32), and 7 of the first 10 deduced amino acids are analogous to those in EKLF. This transcript would thus be predicted to have only 14 nucleotides of 5′ untranslated sequence, similar to the EKLF cDNA with only 55 nucleotides of 5′ untranslated region. The 3′ untranslated region consists of 429 nucleotides. The entire LKLF cDNA encodes a protein with 354 amino acids with a calculated molecular mass of 37,698 Da and an estimated pI of 9.1.

Comparison of LKLF with other zinc finger proteins. A search of the computer databases for nucleotide and amino acid sequence comparisons was conducted. The only matches found were to the three Cys_2-His_2 zinc finger domains of LKLF. These regions are underlined in Fig. 4. The highest matches were to EKLF (36), a transcription factor cloned from human placenta and termed BTEB2 (47), the Wilms' tumor locus protein WT (8), and, to a lesser degree, Sp1 (26). An amino acid alignment of the LKLF zinc finger region with EKLF, BTEB2, Sp1, and BTEB is shown in Fig. 5. BTEB is a protein that has been shown to bind to the basal transcription element of the rat cytochrome P-4501A1 (CYP1A1) gene (24). The zinc finger-encoding domain of the BTEB cDNA was used as the hybridization probe in the isolation of the BTEB2 cDNA (47), hence the name assignment for this placental factor. As seen in Fig. 5, the entire zinc finger region including the number of amino acids in the spacer segments is highly conserved between LKLF, EKLF, and BTEB2. A lower but still significant homology is observed between LKLF and BTEB and Sp1. The amino acids which have been shown to contact the DNA strand in other zinc finger proteins (30) are marked by boxes. These residues are identical among LKLF, EKLF, and BTEB2, while they differ by only one position from BTEB and Sp1. This suggests that all these proteins would bind to a similar and perhaps identical DNA sequence. It has already been shown, for example, that in in vitro binding assays, EKLF and Sp1 will bind the same CACCC site in the β-globin gene promoter (23, 36). Indeed, the differences we have observed between these zinc finger alignments are along the amino-terminal half of each finger, away from the DNA contact sites. This similarity between all five amino acid sequences is quantitated in Fig. 6, which presents the percent amino acid identity and the percent homology if conservative amino acid changes are also included in the calculation. From this analysis, EKLF and LKLF are clearly related family members, and it appears that BTEB2 is more closely related to these zinc finger proteins than it is to BTEB or Sp1. The *Drosophila* Krüppel factor was also selected in the database survey, but, as shown by Miller and Bieker (36), the match, while significant, is only 45% similarity in the zinc finger regions.

The amino-terminal regions of EKLF, LKLF, and BTEB2 are also related. Similar to the GATA family of transcription factors, the nucleotide and amino acid sequence comparison is highest in the DNA-binding domain but diverges significantly in the rest of the molecule. Nevertheless, all three cDNAs encode a proline-rich region in the 5′ half of the molecule which, by analogy with other classes of transcription factors, presumably functions as the transactivation domain. LKLF has the highest percentage with 18% proline residues overall and 23% excluding the zinc finger region. These proline residues

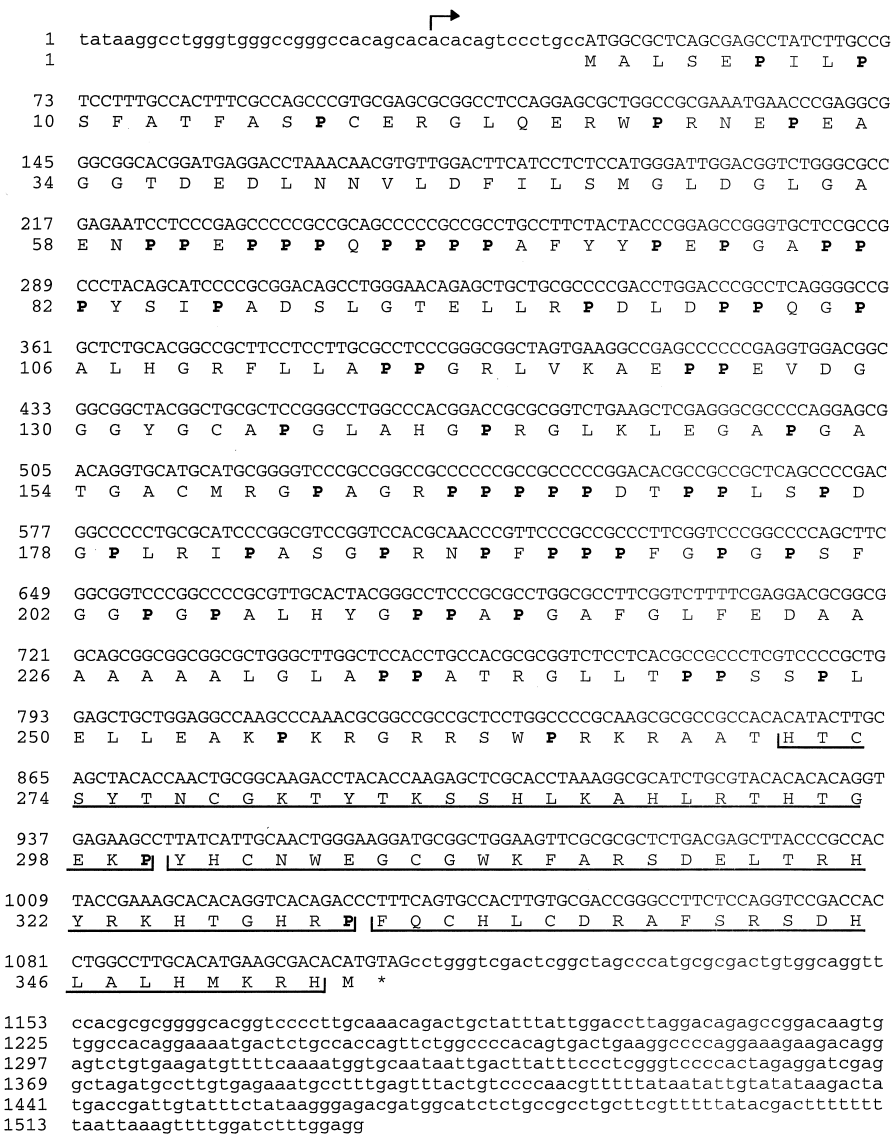

FIG. 4. Nucleotide and deduced amino acid sequence of the LKLF cDNA. The 5′ and 3′ untranslated nucleotide regions are in lowercase letters; the coding region is denoted by capital letters. The transcription start site as determined by the RACE protocol and primer extension analysis is indicated by an arrow. Amino acids are identified by their one-letter code. Proline residues are presented in boldface type. The three zinc finger regions are underlined. Both nucleotides and amino acids are numbered at the beginning of each line.

are often clustered in LKLF, with several repeats of three residues and runs of four and five prolines in a row. This arrangement is similar to that seen in the activation domain of WT-1 (8). Of additional note is a stretch of seven consecutive alanine residues centered in a 20-amino-acid region from positions 215 to 235. Protein structure algorithms predict a strong α-helical structure for this region (46). Interestingly, this sequence occurs between the proline-rich putative activation domain and the zinc finger DNA-binding domain in LKLF.

LKLF can transactivate a human β-globin reporter gene. To test the transactivation potential of this new member of the EKLF family, we reconstructed a full-length LKLF cDNA in an expression vector. Mouse NIH 3T3 fibroblasts were transfected with a human β-globin construct in which the human β-globin gene promoter containing a CACCC site is used to drive the expression of CAT in the absence and presence of increasing amounts of the LKLF expression plasmid. This reporter construct and the LKLF expression construct are shown in Fig. 7A. After 48 h, the cells were harvested and the CAT activity in the cell extracts was measured. The data from these experiments are presented in Fig. 7B, with the normalized values graphed in Fig. 7C. Cotransfection with the LKLF cDNA results in a 118-fold increase in the expression of CAT from the human β-globin promoter. This transactivation occurs in a dose-dependent manner, with increasing CAT activity evident as more of the LKLF cDNA is included in the cotransfection.

On the basis of the conserved amino acid residues in the zinc finger region of EKLF and LKLF, one would predict that these proteins would recognize a common nucleotide-binding site. To test this prediction, we prepared a series of constructs in which the potential CACCC elements in the β-globin promoter were mutated. Two sites fit the consensus CCNCNC CCN derived for EKLF binding, and point mutations in each

```
LKLF    HTCSYTNCGKTYTKSSHLKAHLRTHTGEKP
EKLF    ---GHEG---S-S----------------
BTEB2   -Y-D-PG-T-V----------------
BTEB    -K-P-SG---V-G----I---Y-V----R-
SP1     -I-HIQG---V-G-T---R----W----R-

LKLF    YHCNWEGCGWKFARSDELTRHYRKHTGHRP
EKLF    -A-S-D--D-R----------------
BTEB2   -K-T----D-R----------------AK-
BTEB    FP-T-PD-IK--S----------T---EKQ
Sp1     FM---SY--KR-T-----Q--K-T---EKK

LKLF    FQC..HLCDRAFSRSDHLALHMKRH
EKLF    -C-..G--P----------
BTEB2   ---..GV-N-S----------
BTEB    -R-..PI-EKR-M-----TK-AR--
Sp1     -A-..PE-PKR-M-----SK-I-T-
```

FIG. 5. Alignment of the zinc finger regions of LKLF with other related proteins. The three zinc finger regions are aligned according to the positions of the cysteines and histidines. These residues are boxed. LKLF is compared with EKLF (36), BTEB2 (47), BTEB (24), and Sp1 (26). Dashes represent identity to LKLF. Black squares highlight the DNA contact amino acids as determined by the model of Klevit (30).

of these sites have been associated with thalassemias. The elements were therefore mutated individually and in combination as described in Materials and Methods and as outlined in Fig. 8A. Transactivation assays were carried out with NIH 3T3 cells in the absence and presence of LKLF as described above. The normalized CAT activities are presented in Fig. 8B. LKLF is able to induce the activity of the β-globin promoter in the wild-type state or with either one of the CACCC elements destroyed. The disruption of both CACCC sites, however, severely impairs the transactivation by LKLF.

DISCUSSION

LKLF represents the second member of a new gene family. By using a probe from the DNA-binding domain of the erythroid cell-specific zinc finger gene EKLF, we have isolated a closely related gene which we have termed LKLF. LKLF encodes a protein with 88% conserved amino acid homology in the zinc finger motif but deviates from EKLF in terms of nucleotide and amino acid sequence alignment in the remainder of the molecule. LKLF is expressed primarily in the lungs, with reduced levels evident in the spleen, skeletal muscle, testis, heart, and uterus. This conservation in the DNA-binding domain with divergence in the putative transactivation domain and differences in the sites of expression of these related proteins suggest that EKLF and LKLF are part of a multigene family similar to that seen with the GATA family of transcription factors.

	LKLF	EKLF	BTEB2	BTEB
EKLF	**83(88)**	100	---	---
BTEB2	**81(86)**	**81(87)**	100	---
BTEB	60 (73)	59 (71)	63 (75)	100
SP1	58 (67)	59 (70)	59 (69)	**69(78)**

FIG. 6. Homology comparison of LKLF and other zinc finger proteins. Only the zinc finger regions shown in Fig. 4 were considered in this comparison. The first number in each pair is the percent identity with LKLF. The number in parentheses is the percent homology, which includes conservative amino acid changes (46). The values in boldface type highlight the similarity of LKLF, EKLF, and BTEB2 and the homology between Sp1 and BTEB.

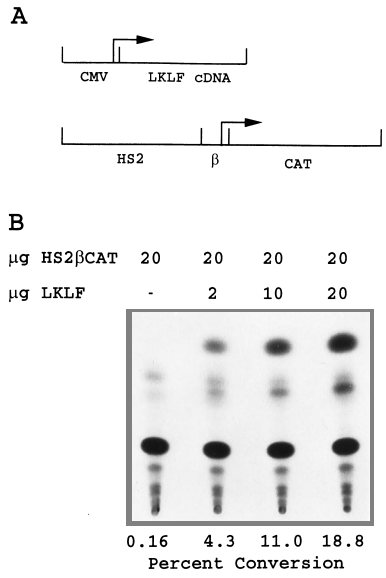

FIG. 7. Transactivation of the human β-globin gene promoter by LKLF. (A) Schematic diagram of LKLF cDNA expression construct and the HS2βCAT reporter construct. These plasmids were cotransfected into mouse NIH 3T3 cells with 20 μg of the CAT reporter plasmid and different amounts of the LKLF cDNA. (B) Autoradiograph of the thin-layer chromatography plate from a CAT assay with the amounts of added LKLF cDNA and HS2βCAT indicated above each sample. The percent conversion of chloramphenicol to the acetylated product was quantitated with a PhosphorImager and is included beneath each sample. (C) The values from the CAT assay were normalized and plotted to indicate the fold induction observed when increasing amounts of LKLF cDNA were included in the cotransfection.

A case can also be considered for including BTEB2 in this family, perhaps as placental Krüppel-like factor (PKLF). As seen in Fig. 5 and 6, the zinc finger domain of BTEB2 exhibits the same high degree of similarity to EKLF and LKLF and is much more closely related to these molecules than it is to BTEB. Additionally, the transactivation domain of BTEB falls into a different class from that of BTEB2. Different categories of activation regions have been noted as more transcription factors have been identified and characterized in recent years (38, 41). These groupings are based on the prevalence of a particular amino acid in the transcriptional activation domain of the molecule. Thus, Sp1 contains a glutamine-rich region, BTEB includes a highly acidic region, and EKLF, LKLF, and BTEB2 all possess a proline-rich domain. Hence, although

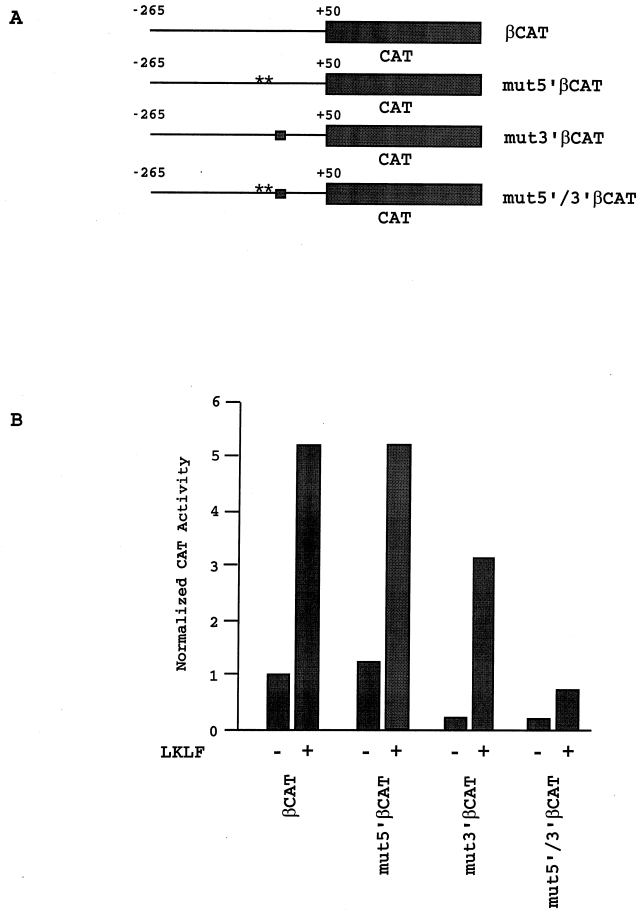

FIG. 8. Disruption of CACCC elements significantly impairs transactivation by LKLF. (A) A diagram of the β-globin–CAT reporter constructs. Point mutations in the distal CACCC element are indicated by the asterisks. Deletion of the proximal CACCC site is shown by a box. The construct with both sites mutated contains both of these symbols. These constructs were cotransfected into mouse NIH 3T3 cells in the absence and presence of 5 μg of the LKLF expression plasmid. (B) Bar graph of the values from a CAT assay, normalized for transfection efficiency. The activity obtained with the βCAT construct was arbitrarily set at 1 for comparison of the mutation constructs.

BTEB2 was cloned on the basis of homology to the BTEB DNA-binding domain, its zinc finger region specifically and the overall structure of this transcription factor more generally bears more resemblance to members of the EKLF family.

The proline-rich transactivator domains of these proteins can also be subcategorized on the basis of the distribution of the prolines in this region. For example, the prolines in EKLF, BTEB2, and NGF1-C, an immediate-early transcription factor (13), are relatively evenly distributed throughout the domain, while the proteins encoded by LKLF, WT-1 (8), and CTF (35) contain clusters of proline residues. LKLF, for example, has runs of four and five prolines. It has been proposed that rather than being a random coil composed of α-helix breakers, this sequence would adopt a structure known as a polyproline II helix (50). This helix is more tightly packed than an α-helix, with 3 residues per turn in the polyproline structure compared with 3.6 in the α-helix. The structure itself is envisioned as a constrained extension from the protein that functions to bind other proteins. RNA polymerase II has such a domain, and data are accumulating to suggest that this region interacts with the TFIID transcription initiation complex (31). A model that arises from this structural and binding data suggests that multiple proteins may interact via these proline-rich elements to form an optimized preinitiation complex for transcription. This systems affords great flexibility, since the polyproline II helix can be formed from many different linear sequences and the interaction of this structure with other proteins appears to depend more on the contact with a similar structure than with a specific amino acid-binding site. These properties are in keeping with the observation that many activation sites are modular and can be switched between transcription factors with different DNA-binding domains yet can still carry out their function. Additionally, this predicts that site-directed mutagenesis of a proline-rich activation domain would not be a productive means of identifying important residues. In fact, for RNA polymerase II, deletion of a substantial portion of the proline-rich region still results in a functional protein, as long as a minimal segment is left that can form the polyproline II helix (11).

LKLF and EKLF may share a common binding site. The zinc finger regions of LKLF and EKLF are identical at the putative contact sites with the DNA helix as modeled by Klevit (30). On the basis of this similarity, one would anticipate that both proteins could bind the same core *cis* element. Although one could envision additional influences such as binding-site context, binding affinity, or interaction with another protein, each playing a role to enhance the specificity for slightly different sites by closely related transcription factors, this has been seen in some but not all systems. In the GATA family of proteins, for example, each GATA protein apparently binds the same site with equivalent affinity in vitro. Their specificity for particular genes is governed more by the differing location of expression of the GATA proteins. Thus, in the case of GATA-1 and GATA-2 where there is overlap in their expression in erythroid cells, there may be some functional interchange as well.

A similar scenario may also be in effect with EKLF and LKLF expression in the spleen. Although the major site of expression of LKLF is the lung, transcripts were also detected in mRNA from the spleen, particularly in our analysis of rat tissues. Transactivation assays in which the β-globin gene promoter was used in the presence and absence of LKLF expression demonstrate that LKLF can transactivate this promoter. Additionally, mutation of the consensus CACCC elements in the β-globin promoter, i.e., the EKLF-binding site, significantly reduces the activity of LKLF in this transactivation assay. Although it is difficult to make direct comparisons, the 118-fold level of transactivation of the β-globin promoter obtained with LKLF appears to be significant when compared with the studies with EKLF. Donze et al. (15) demonstrated a 1,000-fold activation of a similar β-globin promoter construct in transient-expression assays of K562 cells with the inclusion of an EKLF expression construct. In another study, EKLF up-regulated a thymidine kinase promoter with a CACCC-binding site by 13-fold in CV-1 cells (36). It is possible, therefore, that LKLF represents a level of redundancy in the regulation of β-globin. Alternatively, this related transcription factor could be playing a specific role in β-globin gene expression. It has been suggested that EKLF participates in the γ- to β-globin gene switch during development (15), and gene-targeting experiments that ablate the EKLF gene in mice have resulted in a fatal anemia when the switch to fetal liver erythropoiesis should occur (40, 44). Perhaps LKLF is also required to fully realize the β-globin gene levels in the adult animal. Although these are intriguing possibilities, investigations concerning the exact site of expression of LKLF in the spleen will first have to be conducted to ascertain which lineages of cells

are expressing the factor. Questions concerning which other genes LKLF may influence may best be addressed with targeted recombination to disrupt the LKLF gene in animals and analysis of the phenotype of the resulting mice. Such studies will also help determine the physiological consequence of the transactivation of the β-globin promoter by LKLF.

In summary, we have identified, through cloning and a literature search, three potential members of a multigene family. The original protein, EKLF (36), is erythroid cell specific. We have isolated a cDNA encoding a closely related protein, LKLF, expressed principally in the lungs. A search of the computer databases revealed a third protein, BTEB2 (47), which is also highly homologous to EKLF and LKLF and is expressed in placental tissue. All three appear to be transcription factors with nearly identical zinc finger DNA-binding domains and proline-rich activation domains. The adult β-globin gene has been shown to be a downstream target for EKLF (6, 18, 36). Studies presented here indicate that LKLF is also expressed in the spleen and can transactivate the β-globin gene promoter as well.

ACKNOWLEDGMENTS

We thank Gary Shull for generously providing the rat tissue Northern blot. We thank Michael Reder and Michael Barton for their excellent technical assistance in screening the genomic and cDNA libraries.

This work was supported by National Institutes of Health (NIH) grant DK-39585-05. C.B.K. was the recipient of NIH postdoctoral fellowship DK-08902.

REFERENCES

1. **Altschul, S. F., W. Gish, W. Miller, E. W. Myers, and D. J. Lipman.** 1990. Basic local alignment search tool. J. Mol. Biol. **215:**403–410.
2. **Antoniou, M., and F. Grosveld.** 1990. β-Globin dominant control region interacts differently with distal and proximal promoter elements. Genes Dev. **4:**1007–1013.
3. **Askew, G. R., T. Doetschman, and J. B. Lingrel.** 1993. Site-directed point mutations in embryonic stem cells: a gene targeting tag-and-exchange strategy. Mol. Cell. Biol. **13:**4115–4124.
4. **Belyavsky, A., T. Vinogradova, and K. Rajewsky.** 1989. PCR-based cDNA library construction: general cDNA libraries at the level of a few cells. Nucleic Acids Res. **17:**2919–2932.
5. **Berg, J. M.** 1990. Zinc fingers and other metal-binding domains. J. Biol. Chem. **265:**6513–6516.
6. **Bieker, J. J., and C. M. Southwood.** 1995. The erythroid Kruppel-like factor transactivation domain is a critical component for cell-specific inducibility of a β-globin promoter. Mol. Cell. Biol. **15:**852–860.
7. **Brown, R. S., C. Sander, and P. Argos.** 1985. The primary structure of transcription factor TFIIIA has 12 consecutive repeats. FEBS Lett. **186:**271–274.
8. **Call, K. M., T. Glaser, C. Y. Ito, A. J. Buckler, J. Pelletier, D. A. Haber, E. A. Rose, A. Kral, H. Yeger, W. H. Lewis, C. Jones, and D. E. Housman.** 1990. Isolation and characterization of a zinc finger polypeptide gene at the human chromosome 11 Wilms' tumor locus. Cell **60:**509–520.
9. **Chomczynski, P., and N. Sacchi.** 1987. A single-step method of RNA isolation by acid guanidium thiocyanate-phenol-chloroform extraction. Anal. Biochem. **162:**156–159.
10. **Christy, B. A., L. F. Lau, and D. Nathan.** 1988. A gene activated in mouse 3T3 cells by serum growth factor encodes a protein with "zinc finger" sequences. Proc. Natl. Acad. Sci. USA **85:**7857–7861.
11. **Corden, J. L.** 1990. Tails of RNA polymerase II. Trends Biochem. Sci. **15:** 383–387.
12. **Cowie, A., and R. M. Myers.** 1988. DNA sequences involved in transcriptional regulation of the mouse β-globin promoter in mouse erythroleukemia cells. Mol. Cell. Biol. **8:**3122–3128.
13. **Crosby, S. D., J. J. Puetz, K. S. Simburger, T. J. Fahrner, and J. Milbrandt.** 1991. The early response gene NGFI-C encodes a zinc finger transcriptional activator and is a member of the GCGGGGGCG (GSG) element-binding protein family. Mol. Cell. Biol. **11:**3835–3841.
14. **Crossley, M., A. P. Tsang, J. J. Bieker, and S. H. Orkin.** 1994. Regulation of the erythroid Kruppel-like factor (EKLF) gene promoter by the erythroid transcription factor GATA-1. J. Biol. Chem. **269:**15440–15444.
15. **Donze, D., T. M. Townes, and J. J. Bieker.** 1995. Role of erythroid Kruppel-like factor in human γ- to β-globin gene switching. J. Biol. Chem. **270:**1955–1959.
16. **Evans, R. M.** 1988. The steroid and thyroid hormone receptor superfamily. Science **240:**889–895.
17. **Evans, T., and G. Felsenfeld.** 1989. The erythroid-specific transcription factor Eryf1: a new finger protein. Cell **58:**877–885.
18. **Feng, W. C., C. M. Southwood, and J. J. Bieker.** 1994. Analyses of β-thalassemia mutant DNA interactions with erythroid Kruppel-like factor (EKLF), an erythroid cell-specific transcription factor. J. Biol. Chem. **269:**1493–1500.
19. **Freedman, L. P., B. F. Luisi, Z. R. Korszun, R. Basavappa, P. B. Sigler, and K. R. Yamamoto.** 1988. The function and structure of the metal coordination sites within the glucocorticoid receptor DNA binding domain. Nature (London) **334:**543–546.
20. **Frohman, M. A., M. K. Dush, and G. R. Martin.** 1988. Rapid production of full-length cDNAs from rare transcripts: amplification using a single gene-specific oligonucleotide. Proc. Natl. Acad. Sci. USA **85:**8998–9002.
21. **George, K. M., M. W. Leonard, M. E. Roth, K. H. Lieuw, D. Kioussis, F. Grosveld, and J. D. Engel.** 1994. Embryonic expression and cloning of the murine GATA-3 gene. Development **120:**2673–2686.
22. **Gumucio, D. L., H. Heilstedt-Williamson, T. A. Gray, S. A. Tarle, D. A. Shelton, D. A. Tagle, J. L. Slightom, M. Goodman, and F. S. Collins.** 1992. Phylogenetic footprinting reveals a nuclear protein which binds to silencer sequences in the human γ and ε globin genes. Mol. Cell. Biol. **12:**4919–4929.
23. **Hartzog, G. A., and R. M. Myers.** 1993. Discrimination among potential activators of the β-globin CACCC element by correlation of binding and transcriptional properties. Mol. Cell. Biol. **13:**44–56.
24. **Imataka, H., K. Sogawa, K. Yasumoto, Y. Kikuchi, K. Sasano, A. Kobayashi, M. Hayami, and Y. Fuji-Kuriyama.** 1992. Two regulatory proteins that bind to the basic transcription element (BTE), a GC box sequence in the promoter region of the rat P-4501A1 gene. EMBO J. **11:**3663–3671.
25. **Jacobs, G., and G. Michaels.** 1990. Zinc finger gene database. New Biol. **2:**583. (Letter.)
26. **Kadonaga, J. T., K. R. Carner, F. Masiarz, and R. Tjian.** 1987. Isolation of cDNA encoding transcription factor Sp1 and functional analysis of the DNA binding domain. Cell **51:**1079–1090.
27. **Keegan, L., G. Gill, and M. Ptashne.** 1986. Separation of DNA binding from the transcription-activating function of a eukaryotic regulatory protein. Science **231:**699–704.
28. **Kim, J. G., and L. D. Hudson.** 1992. Novel member of the zinc finger superfamily: a C2-HC finger that recognizes a glia-specific gene. Mol. Cell. Biol. **12:**5632–5639.
29. **Kingston, R. E.** 1987. Calcium phosphate transfection, p. 9.1.1–9.1.3. *In* F. M. Ausubel, R. Brent, R. E. Kingston, D. D. Moore, J. G. Seidman, J. Smith, and K. Struhl (ed.), Current protocols in molecular biology. Greene-Wiley Interscience, New York.
30. **Klevit, R. E.** 1991. Recognition of DNA by Cys2,His2 zinc fingers. Science **253:**1367, 1393.
31. **Koleske, A. J., S. Buratowski, M. Nonet, and R. A. Young.** 1992. A novel transcription factor reveals a functional link between the RNA polymerase II CTD and TFIID. Cell **69:**883–894.
32. **Kozak, M.** 1991. An analysis of vertebrate mRNA sequences: intimations of translational control. J. Cell Biol. **115:**887–903.
33. **Labastie, M.-C., D. Bories, C. Chabret, J.-M. Gregoire, S. Chretien, and P.-H. Romeo.** 1994. Structure and expression of the human GATA3 gene. Genomics **21:**1–6.
34. **Laverriere, A. C., C. MacNeill, C. Mueller, R. E. Poelmann, J. B. E. Burch, and T. Evans.** 1994. GATA-4/5/6, a subfamily of three transcription factors transcribed in developing heart and gut. J. Biol. Chem. **269:**23177–23184.
35. **Mermod, N., E. A. O'Neill, T. J. Kelly, R. Kelly, and R. Tjian.** 1989. The proline-rich transcriptional activator of CTF/NF-1 is distinct from the replication and DNA binding domain. Cell **58:**741–753.
36. **Miller, I. J., and J. J. Bieker.** 1993. A novel, erythroid cell-specific murine transcription factor that binds to the CACCC element and is related to the Kruppel family of nuclear proteins. Mol. Cell. Biol. **13:**2776–2786.
37. **Miller, J., A. D. McLachlan, and A. Klug.** 1985. Repetitive zinc-binding domains in the protein transcription factor IIIA from Xenopus oocytes. EMBO J. **4:**1609–1614.
38. **Mitchell, P. J., and R. Tjian.** 1989. Transcriptional regulation in mammalian cells by sequence-specific DNA binding proteins. Science **245:**371–378.
39. **Nardelli, J., T. J. Gibson, C. Vesque, and P. Charnay.** 1991. Base sequence discrimination by zinc-finger DNA-binding domains. Nature (London) **349:** 175–178.
40. **Nuez, B., D. Michalovich, A. Bygrave, R. Ploemacher, and F. Grosveld.** 1995. Defective haematopoiesis in fetal liver resulting from inactivation of the EKLF gene. Nature (London) **375:**316–318.
41. **Pabo, C. O., and R. T. Sauer.** 1992. Transcription factors: structural families and principles of DNA recognition. Annu. Rev. Biochem. **61:**1053–1095.
42. **Pavletich, N. P., and C. O. Pabo.** 1991. Zinc finger-DNA recognition: crystal structure of a zif268-DNA complex at 2.1A. Science **252:**809–817.
43. **Pearson, W. R., and D. J. Lipman.** 1988. Improved tools for biological sequence comparison. Proc. Natl. Acad. Sci. USA **85:**2444–2448.
44. **Perkins, A. C., A. H. Sharpe, and S. H. Orkin.** 1995. Lethal β-thalassaemia in mice lacking the erythroid CACCC-transcription factor EKLF. Nature

(London) **375:**318–322.
45. **Reddy, P. M. S., and C. K. J. Shen.** 1993. Erythroid differentiation of mouse erthroleukemia cells results in reorganization of protein-DNA complexes in the mouse β^{maj} globin promoter but not its distal enhancer. Mol. Cell. Biol. **13:**1093–1103.
46. **Smith, E. L.** 1970. Evolution of enzymes, p. 267–339. *In* P. D. Boyer (ed.), The enzymes, vol. I. Academic Press, Inc., New York.
47. **Sogawa, K., H. Imataka, Y. Yamasaki, H. Kusume, H. Abe, and Y. Fuji-Kuriyama.** 1993. cDNA cloning and transcriptional properties of a novel GC box-binding protein, BTEB2. Nucleic Acids Res. **21:**1527–1532.
48. **Trainor, C. D., T. Evans, G. Felsenfeld, and M. S. Boguski.** 1990. Structure and evolution of a human erythroid transcription factor. Nature (London) **343:**92–96.
49. **Tsai, S. F., D. L. 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.
50. **Williamson, M. P.** 1994. The structure and function of proline-rich regions in proteins. Biochem. J. **297:**249–260.
51. **Yamamoto, M., L. J. Ko, M. W. Leonard, H. Beng, S. H. Orkin, and J. D. Engel.** 1990. Activity and tissue-specific expression of the transcription factor NF-E1 multigene family. Genes Dev. **4:**1650–1662.
52. **Yang, H.-Y., and T. Evans.** 1992. Distinct roles for the two cGATA-1 finger domains. Mol. Cell. Biol. **12:**4562–4570.
53. **Zon, L. I., S. F. Tsai, S. Burgess, P. Matsudaira, G. A. Bruns, and S. H. Orkin.** 1990. The major human erythroid DNA-binding protein (GF-1): primary sequence and localization of the gene to the X chromosome. Proc. Natl. Acad. Sci. USA **87:**668–672.