

Isolation and characterization of a gene encoding human Kruppel-like factor 5 (IKLF): binding to the CAAT/GT box of the mouse lactoferrin gene promoter

Huiping Shi, Zhiping Zhang, Xin Wang, Shiguang Liu and Christina T. Teng*

Gene Regulation Group, Laboratory of Reproductive and Development Toxicology, National Institute of Environmental Health Sciences, National Institutes of Health, Research Triangle Park, NC 27709, USA

Received July 23, 1999; Revised September 24, 1999; Accepted October 25, 1999

DBJ/EMBL/GenBank accession nos[†]

ABSTRACT

The mouse lactoferrin gene promoter includes a CAAT/GT box, GGGCAATAGGGTGGGGCCAGCCC, which functions as the epidermal growth factor response element (EGFRE) in human endometrial carcinoma RL95-2 cells (RL95). A positive clone, EGFREB, of 2575 bp length, was isolated from an expression library of RL95 cells with a multimer of the EGFRE sequence. In this work, we have identified that EGFREB encodes the C-terminus of Kruppel-like factor 5 (KLF5). This mRNA is most abundant in human colon and small intestine. A full-length cDNA clone was isolated from a human colon library using EGFREB as the hybridization probe. The full-length cDNA consists of 3336 bp with a 302 bp 5'-UTR, a 1663 bp 3'-UTR, and a 1371 bp sequence coding for a 457 amino acid polypeptide. Based on its tissue distribution and sequence homology to the mouse IKLF, we renamed this protein IKLF. DNase I footprinting and electrophoresis mobility shift assay confirmed the binding of IKLF to the EGFRE. The human IKLF gene spans >20 kb in length and is organized into four exons, whose intron/exon junctions follow the GT/AG rule. The three zinc fingers are encoded by three exons. Nuclear localization of IKLF was demonstrated by green fluorescence protein (GFP)-tagged IKLF in transfection experiments and western analysis. Overexpression of IKLF in RL95 cells represses the activity of reporter constructs containing the CAAT/GT box of the mouse lactoferrin gene. These findings imply that IKLF is a nuclear transcription factor that binds to the CAAT/GT box, and functions as a modulator of the mouse lactoferrin gene promoter activity.

INTRODUCTION

Transcriptional activity of a eukaryotic gene is governed by a complex interaction between transcription factors and DNA

response elements (1,2). A major DNA-binding motif of transcription factors is the zinc finger domain (3,4). There are several types of zinc finger domains, and transcription factors are classified into families based on their motif-type. The Kruppel-like factor (KLF) represents a subfamily of C2H2 zinc finger genes that contain multiple zinc finger motifs separated by seven conserved amino acids (5–7). Members in this subfamily (7–12) share high sequence homology within the three zinc finger motifs, but very little homology within the other parts of the protein. These transcription factors display tissue-selective expression patterns, for example EKLF (8) is highly restricted to erythroid cells whereas LKLF (7) is predominantly expressed in lung. Likewise, GKLF (9,10) and IKLF (12) are primarily in the gut, while BKLF (11) is widely but not ubiquitously expressed in cell lines and tissues. Proteins in this family bind GC/GT-rich sequences (13) and the DNA co-crystal structure of zif268, a protein closely related to the Kruppel-like factors, predicted CCNCNCCCN as the consensus DNA binding site (14). Thus, regulation of genes containing such binding sequences may be dictated by the site of expression rather than by the binding specificity of the Kruppel-like factor in different tissues.

It is well documented that the Kruppel-like factors act as both positive and negative regulators, and are involved in diverse aspects of eukaryotic gene expression (13,15,16). EKLF was shown to activate the β -globin locus control region both *in vitro* (8,17) and *in vivo* (18). The activating domain was mapped to the proline-rich N-terminal half of the protein (19,20). BKLF and IKLF are also able to activate a reporter construct containing their binding element in transfected cells (11,12). In contrast, GKLF is highly expressed in growth-arrested cells and not in exponentially growing cells. Furthermore, overexpression of GKLF in transfected cells inhibits DNA synthesis. Thus, GKLF is capable of repressing the genes to which it binds in the gut (9). A functional domain which is capable of mediating repression has been identified in the *Drosophila* Kruppel protein (21) and recently in the human counterpart of mouse GKLF (10,20).

We have shown previously that a CAAT/GT box in the mouse lactoferrin gene termed the epidermal growth factor response element (EGFRE) mediates epidermal growth factor (EGF)-induced transcriptional activation of the promoter in

*To whom correspondence should be addressed. Tel: +1 919 541 0344; Fax: +1 919 541 0696; Email: teng@niehs.nih.gov

[†]BankIt256860, AF132818

transiently transfected human endometrial carcinoma RL95-2 (RL95) cells (22,23). We then used a multimerized form of EGFRE and isolated a partial cDNA clone called EGFREB from the RL95 expression library (24). This study describes the identification of a protein that binds to the EGFRE. A full-length cDNA was isolated from a colon cDNA library with EGFREB as probe, and then the full-length cDNA was used to isolate the gene from a human genomic library. Sequence analysis suggests that the cDNA encodes a member of the Kruppel-like family of transcription factors. This newly cloned factor shares 93% sequence identity at the amino acid level with mouse IKLF (12), suggesting that the full-length EGFREB is the human counterpart of mouse IKLF. The human IKLF has an open reading frame of 457 amino acids which includes the BTEB2 sequence (25) at its C-terminus. This work demonstrates that IKLF binds to the CAAT/GT box of the mouse lactoferrin gene. Overexpression of IKLF in RL95 cells inhibits lactoferrin promoter activity in a dose-dependent manner. Therefore, IKLF is likely to be a modulator of its target gene(s) activity.

MATERIALS AND METHODS

Materials

The human umbilical vein endothelial cell genomic DNA library and FLASH non-radioactive gene mapping kit were obtained from Stratagene (La Jolla, CA). The human colon cDNA library was kindly provided by Dr R. Dubois (Vanderbilt). The ABI Prism™ dye terminator cycle sequencing ready reaction kit for automatic sequencing was purchased from Perkin Elmer (Branchburg, NJ); reagents for manual sequencing from US Biochemical (Cleveland, OH); radio-labeled compounds from either Dupont NEN (Boston, MA) or Amersham (Arlington Heights, IL); the radipack GST purification module from Pharmacia (Piscataway, NJ); human endometrial carcinoma cell line RL95 from ATCC (Rockville, MD); tissue culture components from Life Technologies Inc. (Gaithersburg, MD).

Cloning and sequencing of cDNAs

The RL95 cDNA expression library (26) was screened with labeled EGFRE concatenated oligonucleotides as described (24,27). Four specific clones were identified after screening 1.5×10^6 plaques plated at a density of 5×10^4 p.f.u./filter. One of the clones, EGFREB, was completely sequenced by the dideoxynucleotide chain termination method of Sanger (28) and by automatic sequencing with the ABI Prism™ dye terminator cycle sequencing kit on both strands of the DNA. Three additional clones from a human colon cDNA library, constructed in the *EcoRI* and *XhoI* sites of plasmid pBK-CMV (Stratagene, La Jolla, CA), were isolated and the longest, after sequencing and characterization, was designated IKLF.

Northern blot analysis

Human adult multiple-tissue northern (MTN™) Blot I and II (Clontech, Palo Alto, CA) containing 2 µg poly(A)⁺ RNA from the indicated tissues were hybridized with the [³²P]dCTP-labeled EGFREB cDNA as previously described (23). After the radioactivity had decayed (checked with X-ray film as negative), the filters were reprobed with the digoxigenin-

labeled 231 bp 5'-untranslated region (5'-UTR) of IKLF. The fragment was excised from the pBK-IKLF expression plasmid by *EcoRI* and *SacI* digestion, and labeled with digoxigenin with a non-radioactive labeling and detection kit for nucleic acids (Boehringer Mannheim, Indianapolis, IN). The autoradiograms of the northern blots were scanned with a Chemi-Imager™ 4000 (Alpha Inotech Corp), and the intensity of the bands quantitated with AlphaEase™ (v.3.3) software.

Isolation of genomic clones and Southern blot analysis

Genomic clones containing the IKLF gene were isolated from a human umbilical vein endothelial cell genomic DNA library (λ DASH II; Stratagene, La Jolla, CA). We screened 2×10^6 phage with IKLF cDNA and obtained eight positive clones. These clones were further probed with oligonucleotides from different regions of the cDNA. Based on the hybridization results, three clones (H3, H5 and H6) were analyzed further. The smallest insert (clone H6) was isolated by *NotI* digestion and subcloned into Bluescript SK⁺ forming a plasmid named IKLF-Gn. The 470 bp 5'-region of IKLF-Gn was obtained by *SstI/NotI* digestion, subcloned and sequenced. The 5'-genomic sequence was obtained from this plasmid. Restriction maps of the three λ genomic clones were obtained using the FLASH non-radioactive gene mapping kit from Stratagene. The intron/exon junctions were determined by automatic sequencing with oligonucleotide primers to the cDNA. Intron lengths were determined by restriction mapping and PCR analysis of the genomic λ clones and verified with the human placental genomic DNA. Southern blot analyses were carried out with restriction enzyme digestions of DNAs from human placenta (Promega, Madison, WI) and genomic clones followed by agarose gel separation, DNA transfer and hybridization to the IKLF cDNA and 5'-UTR.

Expression and purification of glutathione S-transferase fusion protein and antibody production

The longest IKLF cDNA clone (encoding the entire IKLF from amino acid 1 to 457) and a 3'-truncated cDNA (IKLFA, encoding IKLF from amino acid 1 to 128) were added into a glutathione S-transferase (GST) fusion protein expression system. The constructs were transformed into bacterial strain HB101 and the fusion proteins induced by IPTG after overnight incubation. GST fusion proteins were purified with a Redipack GST purification module according to the manufacturer's specifications (Pharmacia, Piscataway, NJ). Purified GST fusion proteins were analyzed by 10% SDS-PAGE and their purity determined by Coomassie Blue staining. The proteins were aliquoted and stored at -70°C until use. Three female New Zealand White rabbits were immunized with 500 µg gel-purified GST-IKLF or GST-IKLFA per rabbit followed by booster injections with 200 µg gel-purified fusion protein at regular intervals.

Preparation of nuclear protein extract, electrophoresis mobility shift assay (EMSA) and western blotting

Nuclear protein extracts prepared from the tissues of human prostate, monkey endometrium, mouse uterus and mammary gland, and the cell lines Comma-D (mouse mammary gland tumor), HB100 (human breast cancer), HEC-1B (human endometrial carcinoma), HeLa (human cervix epitheloid carcinoma) and RL95-2 (human endometrial carcinoma) were

described previously (23,26). The double-stranded EGFRE or MRU (containing the EGFRE and CRE; the top strand of the nucleotide sequences are presented in Table 2) of the mouse lactoferrin gene was used as the probe in EMSA. The oligonucleotides were labeled with [³²P]dGTP and Klenow fragment at the 3'-ends by fill-in reaction. GST-IKLF fusion protein (1.5–2.0 µg) was incubated with 0.3 ng labeled probe (1.5 × 10⁴ c.p.m.) in a binding reaction at room temperature for 15 min before electrophoresis. Antibody supershift of the protein–DNA complex was conducted by incubating the antibody with the fusion protein at room temperature for 15 min prior to the addition of probe and reaction buffer. For competition analyses, the labeled probe and 100-fold molar excess of unlabeled oligonucleotide were added to the incubation reaction at the same time. Following binding, the reaction mixtures were electrophoresed on a 5% non-denaturing polyacrylamide gel in 0.5× TBE buffer. The gel was dried and exposed to X-ray film at –70°C overnight. For western blotting, the nuclear proteins were separated on 4–12% NuPAGE (Novex, San Diego, CA), blotted onto a nitrocellulose membrane, and the presence of IKLF in the nuclear protein was tested with IKLF antiserum and ECL detection (Amersham, Arlington Heights, IL).

DNase I footprinting analysis

The 104 bp upstream fragment of the mouse lactoferrin gene (+1 to –103; see Table 2 for nucleotide sequence) was subcloned into the *Eco*RI site of Bluescript SK⁺ by blunt-end ligation. The desired DNA fragment was isolated with either *Hind*III/*Sma*I or *Hinc*II/*Bam*HI digestion and end-labeled with [α -³²P]dATP at the 3'-ends of both top and bottom strands according to a previously described protocol (26). Specific activity of the labeled DNA was 3–6 × 10⁷ c.p.m./µg. Protein binding and DNase I protection assays were performed according to Roesler *et al.* (29). The samples were analyzed on a 6% denaturing polyacrylamide sequencing gel. Chemical DNA sequencing reactions for G and G+A were performed on the same DNA fragment and were included as markers (28).

Confocal microscopy

The IKLF cDNA fragment (nucleotide positions 163–2667), which includes the entire peptide coding region, was excised from plasmid pBK-IKLF by *Taq*I digestion. The green fluorescence protein (GFP)–IKLF fusion protein was generated by introducing the IKLF fragment into the pEGFP-C1 (Clontech, Palo Alto, CA) expression vector using an *Acc*I site. The construct was verified by restriction enzyme digestion and automatic sequencing. The IKLF–GFP fusion plasmid (10 µg) was transiently transfected into HeLa cells by electroporation at a capacitance of 960 µF at 250 V using a Gene Pulser II system (Bio-Rad, Hercules, CA). The cellular localization of the IKLF–GFP fusion protein was viewed 24 h later with a Zeiss LSM-310 confocal microscope equipped with an argon/krypton laser.

Cell culture, transfection and CAT assay

Human endometrial carcinoma cells (RL95-2, ATCC no. CRL1671) were maintained in Dulbecco's modified Eagle's medium/Ham's F12 (1:1) supplemented with 10% fetal calf serum (FCS) (Life Technologies Inc.) as previously described (22). The cells were co-transfected with mLFO.1–CAT reporter

plasmids (2.5 µg/3.5 cm well), β-galactosidase, plasmid pCH110 (0.25 µg) and pBK-IKLF expression plasmid or pBK vector (1.0 µg) by the calcium phosphate method (30). Cell extract preparations and CAT enzyme assays were as described previously (22,23). Relative CAT activity, after normalization for β-galactosidase activity, is presented as means ± SD. All experiments were repeated at least three times with duplicated samples in each experiment.

GenBank accession number

The nucleotide and the deduced amino acid sequences of IKLF have been submitted to GenBank. The accession numbers are BankIt256860 and AF132818.

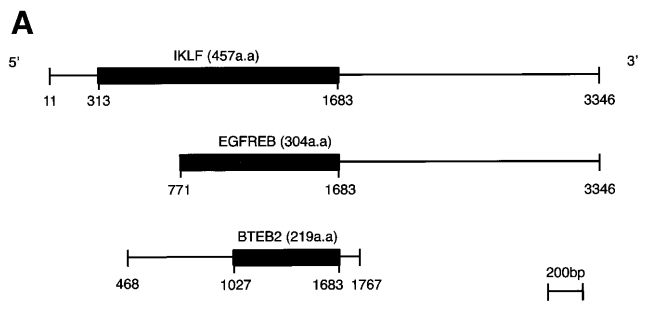
RESULTS

Cloning and characterization of the human IKLF cDNA

To identify transcription factors involved in the EGF responsiveness of the lactoferrin gene, a cDNA expression library from RL95-2 cells was screened with a concatenated EGFRE sequence (24). Among the 1.5 × 10⁶ clones screened, four positive clones were identified. The longest one, EGFREB, was 2575 bp (Fig. 1A). Sequence analysis of EGFREB showed a perfect match to the BTEB2 sequence at the overlapping region (25) with two exceptions: there is a 1 nt addition at the 5'-end of EGFREB (nucleotide position 849 in Fig. 1B) and the last three nucleotides of BTEB2 (CCC) are changed to AAA in EGFREB (nucleotide positions 1754–1756 in Fig. 1B). The nucleotide sequence of EGFREB is 303 bp shorter at its 5'-end and 1580 bp longer at its 3'-end than the BTEB2 sequence (Fig. 1A). Although the polypeptide encoded by both BTEB2 and EGFREB ends with the same C-terminal amino acid sequence (position 1683 in Fig. 1B), EGFREB has an open reading frame of 304 amino acids whereas BTEB2 consists of 219 amino acids (Fig. 1A).

Since EGFREB encodes a continuous open reading frame from the beginning of the cDNA, we believe that this cDNA might be incomplete. To obtain a full-length cDNA, EGFREB was used to screen the human colon cDNA library under high stringency hybridization and washing conditions which resulted in the identification of three positive clones, the longest of which was 3336 bp. This clone, like EGFREB, has no in-frame stop codon at its 5'-end. Using a PCR–RACE procedure, we were unable to generate any longer product and thus it was assumed that this clone is full length. The presumed full-length cDNA clone has 93% sequence homology to mouse IKLF (12), therefore, it was named human IKLF (Fig. 1A). A subsequent RNase protection assay confirmed that its 5'-end is 10 bp away from the transcription initiation site (arrow in Fig. 1B). The first methionine codon appears 302 bp from the beginning of the cDNA clone (position 313 in Fig. 1B) and the open reading frame continues for 457 amino acids before encountering a stop codon. The nucleotide sequence surrounding the methionine codon conforms to Kozak's rule for translation initiation (31). The IKLF cDNA has 1663 bp of 3'-untranslated sequence. A polyadenylation signal (32) appears 25 bp before the beginning of the poly(A) tract.

The deduced IKLF amino acid sequence is rich in proline residues (~13%) which is randomly distributed throughout the protein outside the zinc finger region. However, two short



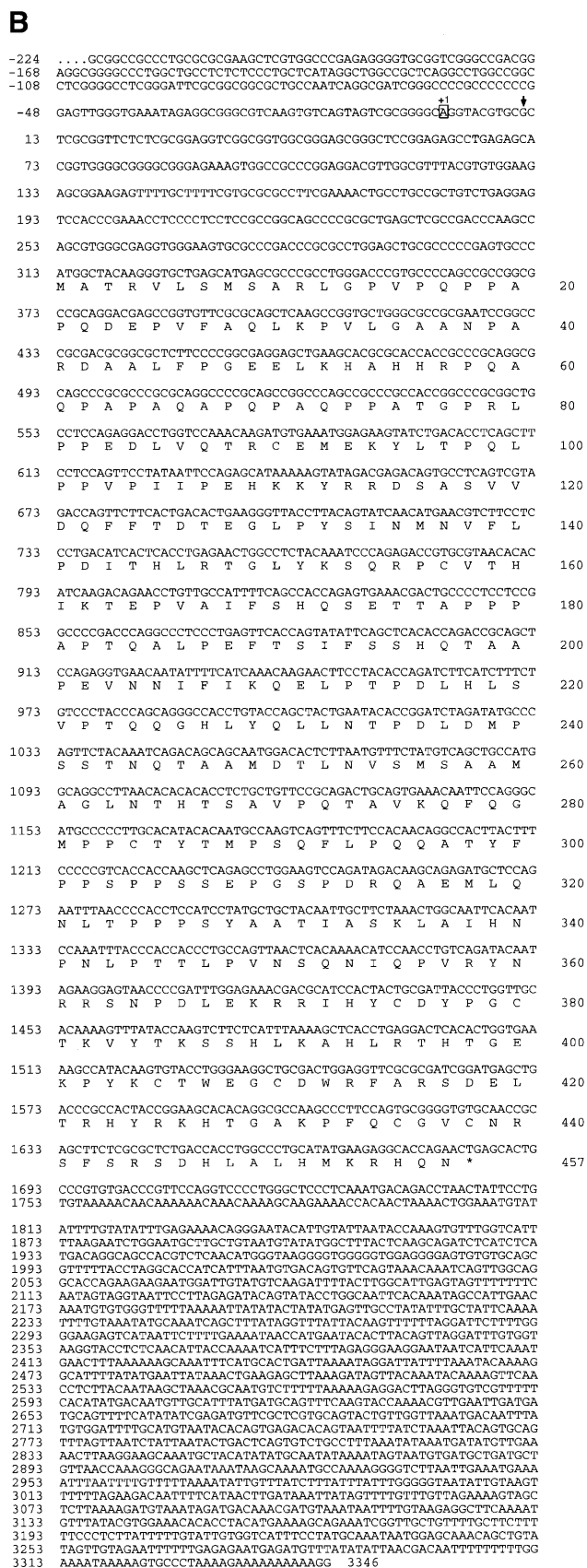
stretches of Pro/Ala/Gln (17 amino acids) and Pro/Ser (10 amino acids) are present. In addition, there are nine potential phosphorylation sites and three glycosylation sites in IKLF. Interestingly, six of the phosphorylation sites are distributed within the N-terminal region. IKLF has three zinc finger motifs near its C-terminal end. The amino acids in these zinc fingers conform to the consensus zinc finger sequence in the Kruppel-like gene family (7). The seven amino acids in between the fingers are conserved and conform to the H-C link consensus sequence in the family. A diagrammatic presentation of the IKLF peptide structure is shown in Figure 1C.

A search of the computer databases for nucleotide and amino acid sequences similar to IKLF was performed. The only identical match found was BTEB2, although many sequences were similar to IKLF within its C2H2 zinc finger domains. The Kruppel-like family members (7,8,10–12,33) share 80–85% homology over the three zinc finger domains with IKLF (Fig. 1D). Outside the zinc finger domains, IKLF shares very little sequence homology with the other Kruppel-like factors, although the N-terminal regions of these proteins are proline rich.

Expression of the IKLF mRNA and protein

Tissue-specific expression of IKLF was examined using northern blot analyses of poly(A)⁺ RNA from human tissues (human adult multiple-tissue Blot I and II; Clontech) with EGFREB cDNA as the probe (Fig. 2A). A single transcript 3.4 kb in length was present in all tissues except testis, where the major mRNA was 1.5 kb in length. This smaller mRNA transcript may be formed by alternative splicing and/or different usage of poly(A) adenylation signals. It is not clear whether the IKLF protein produced from the testis-specific mRNA is smaller than the protein in other tissues. The IKLF mRNA is selectively expressed in human tissues and is most prominently expressed in the colon and small intestine; in brain, liver, spleen and white blood cells the IKLF transcript is undetectable. Northern blot data were quantitated by phosphorimager analysis under the assumption that IKLF poly(A)⁺ RNA is recovered equally from various human tissues. By normalizing all values to the lowest signal (heart tissue), the following relative IKLF mRNA levels were calculated: heart, 1; kidney, thymus and ovary, 1.2–1.4; pancreas, 2.8; lung, 2.9; placenta, skeletal muscle, prostate and testis, 5.0–5.7; small intestine, 9.5; colon, 11.5.

To confirm IKLF mRNA expression, northern blots were rehybridized with an IKLF-specific probe. The 5'-UTR of the IKLF cDNA (nucleotides 11–241 in Fig. 1B), which is unique to IKLF and absent from BTEB2, was prepared as a digoxigenin-labeled probe. Similar patterns of expression were



detected with this probe and the EGFREB cDNA probe in all tissues (data not shown).

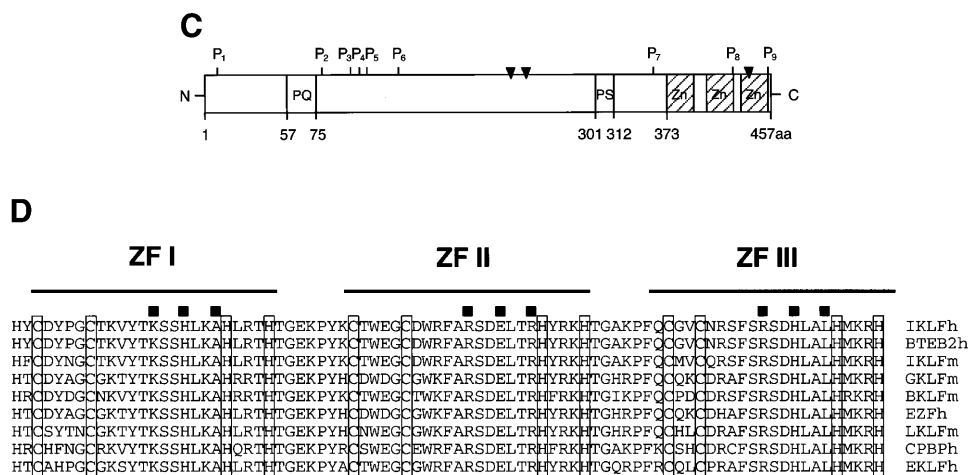


Figure 1. (Opposite and above) The cDNA and 5' genomic sequence of IKLF. (A) Relationship between the cDNA clones of IKLF, EGFREB and BTEB2. The lines indicate the untranslated regions of each cDNA; the numbers under the lines indicate the nucleotide position assigned according to the transcription initiation site (+1). The filled boxes indicate the open reading frame of each cDNA. (B) Complete nucleotide sequence and deduced amino acid sequence of IKLF. Numbers on the right refer to the peptide and on the left to the nucleotide sequence. The transcription initiation site is boxed and marked as +1 and the beginning of the longest IKLF cDNA clone is marked by an arrow. The poly(A) adenylation site is underlined. (C) A schematic presentation of the IKLF polypeptide. Nine potential phosphorylation sites are labeled P1–P9, and the three glycosylation sites are marked by solid arrowheads. The amino acid (aa) positions of the three zinc fingers and Pro/Ala/Gln (PQ) and Pro/Ser (PS) clusters are indicated. (D) Amino acid sequence alignment of the zinc finger DNA-binding domain of IKLF and related proteins. The three zinc finger regions are aligned according to the position of two C and two H residues in human IKLF [IKLFh (B)], human BTEB2 (BTEB2h; 25), mouse IKLF (IKLFm; 12), mouse GKLF (GKLFm; 9), mouse BKLF (BKLFm; 11), human EZF (EZFh; 10), mouse LKLF (LKLFm; 7), human CPBP (CPBP; 47) and human EKLf (EKLfh; 8). The C2H2 zinc finger amino acids are boxed, and the DNA contact amino acids (35) are indicated by solid square boxes above the sequence.

IKLF protein was detected in nuclear extracts of Comma-D, RL95, mouse uterus and mammary gland tissues by western blot analyses (Fig. 2B). Antiserum generated against GST–IKLF (amino acids 1–457) identified a major band of 57 kDa (lanes 5–8). Antiserum against GST–IKLFA (amino acids 1–128), which includes only the N-terminal 128 amino acids of IKLF, also detected the 57 kDa protein (lanes 1–4). In Comma-D nuclear extract, a smaller band at 39 kDa was detected by both antisera. Other nuclear extracts that showed the 57 kDa protein were HeLa, HEC-1B and HB100 cells, human prostate and monkey endometrium (data not shown). Based on the deduced amino acid sequence of IKLF, the predicted size is only 50.7 kDa; therefore, the slower migrating protein species may carry a post-translational modification, such as phosphorylation or glycosylation. To confirm that IKLF is mainly present in the nucleus, we transfected a fusion plasmid of IKLF and GFP into HeLa cells. Twenty four hours after transfection, the fusion protein was localized to the nucleus, as visualized by confocal microscopy (Fig. 2C). This confirms that IKLF is a nuclear factor.

Isolation of the IKLF gene

Three clones for the IKLF gene were isolated from a human umbilical vein endothelial cell genomic DNA library with the IKLF cDNA. The genomic clones were characterized by Southern blotting, PCR and restriction enzyme mapping. Two of the clones, H5 and H6, include the first three exons of IKLF, and the third clone, H3, includes a partial exon 2 and complete exons 3 and 4. Although the IKLF gene spans >20 kb in length, it has only four exons. The organization of the gene is presented in Figure 3. Primer extension experiments failed to identify the transcription start site due to its high GC content. However, the transcription start site was identified by RNase

protection. Using mRNA from three different cell lines, HeLa, HEC-1B and RL95-2, the transcription initiation site was mapped to a position 10 bp upstream from the full-length cDNA clone (data not shown); this nucleotide is designated +1 in Figure 1B. A partial genomic sequence (224 bp) including the beginning of exon 1 showed a high GC content and several potential Sp1 binding sites were found in this region.

The IKLF gene is organized into four exons and its intron/exon junctions conform to the conserved GT/AG dinucleotide pattern (Table 1). The first exon is 573 bp and encodes the first 87 amino acids. The 874 bp second exon encodes the 291 amino acids that make up the major N-terminal portion of the protein and the first four amino acids of the first zinc finger. The third exon is the smallest (60 bp) and encodes 20 amino acids including the rest of the first zinc finger. The second and third zinc fingers are encoded by exon 4 (59 amino acids). A stop codon follows immediately after the third zinc finger. It is interesting to find that the splicing patterns in the zinc finger region are identical (junctions of the second and third introns in Table 1) and that they may reflect a duplication event involving a small group of amino acids (2). The intron between the third and fourth exons is huge, spanning >14 kb. There is a long, untranslated sequence after the coding region and before the polyadenylation site, making the fourth exon 1839 bp in length. Southern analysis of the human genomic DNA with cDNA and 5'-UTR probes showed that IKLF is a single copy gene (data not shown).

IKLF binds the EGFRE and MRU of the mouse lactoferrin gene

To characterize the ability of IKLF to bind to the EGFRE, we expressed IKLF as a GST fusion protein in bacteria and carried

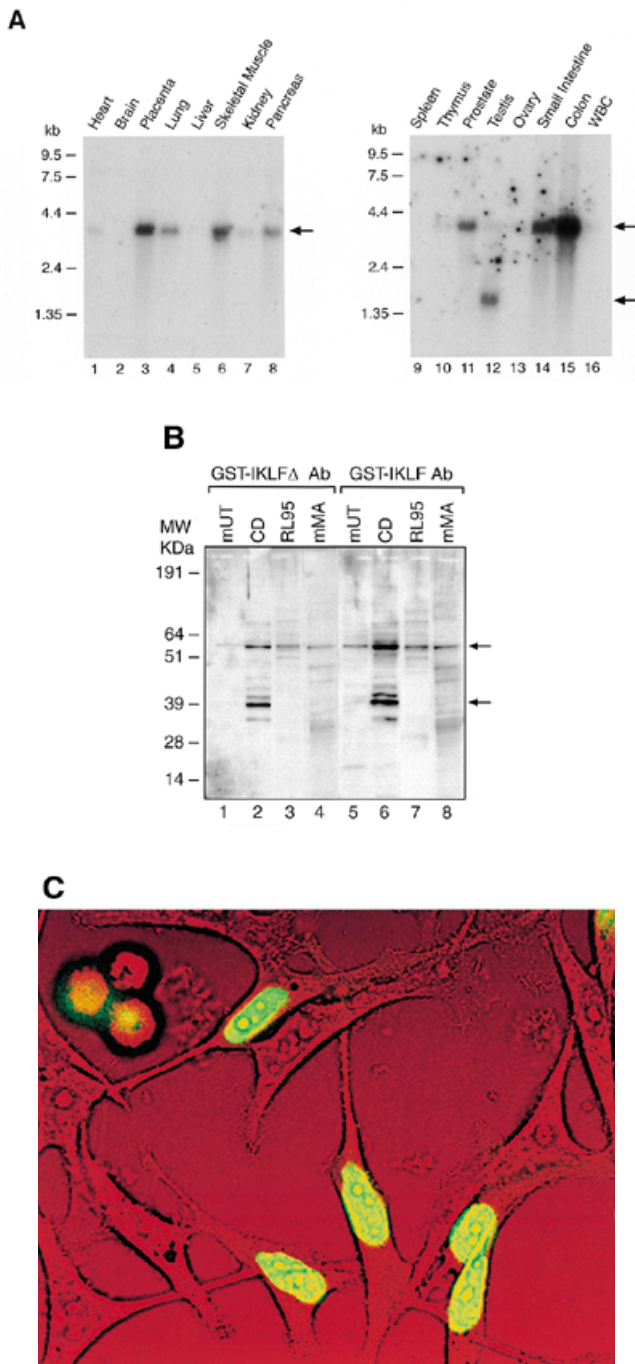


Figure 2. The IKLf mRNA and protein. (A) Tissue distribution of human IKLf mRNA. Two micrograms of poly(A)⁺ mRNA from each of the various human tissues on the northern blots (Clontech) were hybridized to the ³²P-labeled EGFRE cDNA. The IKLf mRNAs are indicated by arrows. (B) Western analysis of IKLf protein in human cell lines and mouse tissues. Nuclear protein extract from Comma-D (CD) and RL95-2 cells and the mouse uterine and mammary gland tissues were separated on 4–12% NuPAGE (Novex) and blotted onto a nitrocellulose membrane. IKLf was detected using a 10 000-fold dilution of the polyclonal rabbit antiserum against GST-IKLf (rabbit 4723) or GST-IKLf Δ (rabbit 4721) and an ECL detection kit. A molecular weight marker (Novex, see blue marker) is indicated on the left and IKLf is marked by an arrow. (C) IKLf is a nuclear protein. IKLf-GFP fusion constructs were transfected into HeLa cells by electroporation. The fluorescence was detected with a Zeiss LSM-310 confocal microscope. The image was acquired on the microscope at 512 × 512 resolution over a 32 s exposure period.

out EMSA experiments. Purified fusion protein was used in the EMSA with the double-stranded oligonucleotides shown in Table 2. Figure 4 shows that the fusion protein bound strongly to the EGFRE (Fig. 4A, lane 1, and Fig. 4B, lane 1), and the protein-DNA complex was supershifted by antiserum raised against GST-IKLf (Fig. 4B lane 2, arrow). The binding was eliminated by a 100-fold molar excess of unlabeled competitor EGFRE (Fig. 4A, lane 2) and by a BTEB2 binding element (lane 6) but not by EGFRE sequences carrying mutations (lanes 3–5), suggesting that the DNA binding was sequence specific.

Table 1. DNA sequence of the exon/intron boundaries of IKLf

Exon	Exon Size (bp)	Exon/Intron Junction	Intron Size (kbp)
1	573	GAC.CTG.GTC.CAgttaggaa.....tctttgttagACA.AGA.TGT.GAA D L V Q T R C E	>2.1
2	874	TGC.GAT.TAC.CCT.Ggtatgtg.....ccttatttagGT.TGC.ACA.AAA C D Y P G C T K	>1.2
3	60	AGG.ACT.CAC.ACT.Gstatgta.....gttgcacacGT.GAA.AAG.CCA R T H T G E K P	>14
4	1839		

Table 2. Sequences of the oligonucleotides

Top Strand of mLF0.1 (-103 to +1, ref.22)

-103 GGCAATGGGGCTGGAAGGCAGGCCTATTGGGCAATAGGGTGGGGCCAGCCC
GGTGAGGTACCCAGCACAGATAAAGGGCCCCGGGGAGAGGGCAGAAGCCAGG +1

Top Strand of mLF Oligonucleotides Used in EMSA (GATC is included in both strands, ref. 22,23)

MRU -75 GGGCAATAGGGTGGGGCCAGCCCCTGGAGGTCACCC -40

EGFRE GATCGGGCAATAGGGTGGGGCCAGCCC

CRE GCCCCTGAGGTCACCCAGCA

m6 -----cta-----

m10 -----a-c-----

m11 -----at-----

m12 -----ttt-----

m15 -----cgtgact-----

GC Rich Response Elements Used in EMSA

BTEB2 5'-GATCGAGAAGGAGGCGTGGCCAAC-3'
Ref.25 3'-CTCTTCCCTCCGACCCGGTTGCTAG

GERE 5'-GATCTGGGGCGGGTGGGGGAT-3'
Ref.37 3'-ACCCGCCCCACCCCTACTAG-5'

Myc-specific 5'-CCTTCCCCACCCTCCACCTC-3'
Ref.38 3'-GGTGGGAGGGTGGGAGGGTA-5'

Sp1 5'-ATTCTGATCGGGCGGGCCAGC-3'
Ref.48 3'-CTAGCCCCGCCCTCGCTTA-5'

The extent of the DNA region bound by GST-IKLf was examined using DNase I footprinting on a DNA fragment from the mouse lactoferrin gene promoter (Fig. 5). The region from -103 to +1 was used in the footprinting analyses (Table 2). The results indicate that both strands of the entire EGFRE are protected by GST-IKLf (lanes 5 and 10). The nuclear protein extract (NPE) from RL95 cells protects the CRE and GATAAA regions in addition to the EGFRE (lanes 4 and 9). The nucleotides protected by GST-IKLf are shown under the gel (Fig. 5).

IKLf binding to the EGFRE in the context of MRU (containing both the EGFRE and CRE; sequence in Table 2) was also investigated by EMSA using MRU as probe with GST-IKLf fusion protein. As expected, IKLf bound to MRU

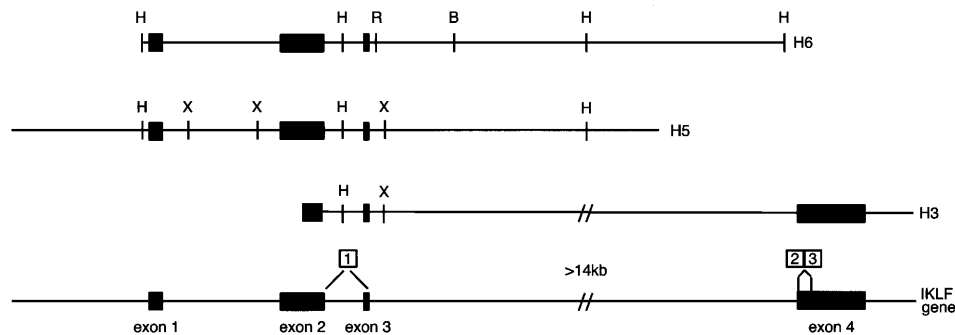


Figure 3. Organization of the human IKLF gene. Partial restriction endonuclease maps of the genomic clones H3, H5 and H6 containing the IKLF exon sequences are shown. The lines represent introns, and the filled boxes represent exons. The boxes with numbers indicate the zinc finger coding exons; the drawings are not to scale. H, *HindIII*; X, *XbaI*; R, *EcoRI*; B, *BamHI*.

(Fig. 6, lane 2) and a 100-fold molar excess of unlabeled EGFRE or MRU oligonucleotide competed equally well with the labeled probe (lanes 3 and 4). It was interesting to find that IKLF formed a doublet when bound to MRU; it is possible that IKLF binds MRU as a homodimer. In the competition studies, MRU oligonucleotides were made with mutations in either the EGFRE (m10 and m15) or CRE (m6) and the mutant oligonucleotides were used as competitors in EMSA (Fig. 6, lanes 5–7). Oligonucleotides containing mutations in the EGFRE region failed to compete (m10, lane 5; m15, lane 6). In contrast, a competitor with a mutation in the CRE region (m6) competed as well as the wild-type sequence competitor (lane 7). These results imply that IKLF binds to the EGFRE region. Other GC-rich binding elements such as the BTEB2 binding element (lane 8) and Sp1 binding site (lane 11) competed very efficiently with the wild-type EGFRE, whereas the gastrin EGF response element (GERE) competed with reduced efficiency (lane 9). The Myc binding element is a weak competitor (lane 10).

The mouse lactoferrin gene promoter is repressed by IKLF in RL95-2 cells

To study the effect of IKLF on the mouse lactoferrin gene promoter *in vivo*, a plasmid expressing IKLF was co-transfected with a reporter construct containing the mouse lactoferrin promoter (mLF 0.1–CAT) into RL95 cells. Overexpression of IKLF repressed the basal transcriptional activity of the mouse lactoferrin gene promoter in a dose-dependent manner (Fig. 7).

DISCUSSION

This work has identified a new member of the GLI-Kruppel family of genes (16,34 and references therein) which is the human homolog of mouse IKLF. The mouse lactoferrin EGFRE is a target sequence of IKLF, and it was isolated based on its ability to bind the EGFRE sequence. In RL95 cells, IKLF acts as a repressor of mouse lactoferrin promoter activity.

Kruppel-like family genes are usually expressed in a tissue-selective manner. IKLF and GKLf have similar patterns of expression; they are predominantly expressed in the small intestine and colon, and are hardly detectable in the liver,

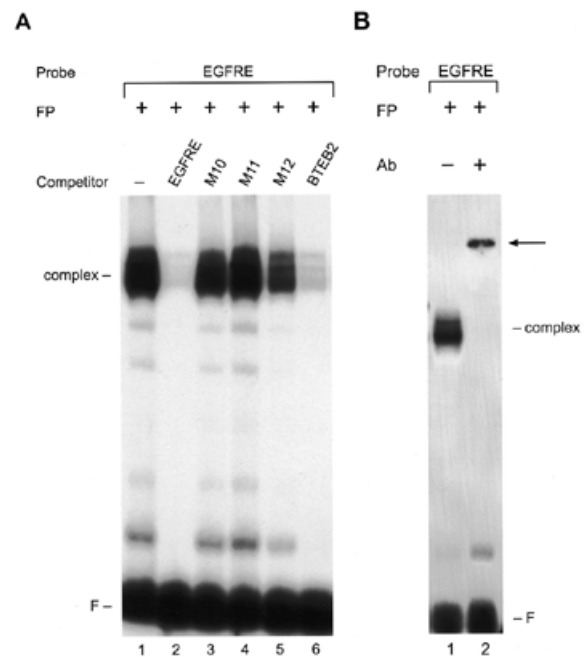


Figure 4. Specific interaction between GST-IKLF and the EGFRE. (A) Two micrograms of GST-IKLF fusion protein (FP) were incubated with the labeled EGFRE oligonucleotide. A 100-fold molar excess of unlabeled EGFRE, mutant EGFRE (m10, m11 or m12) or BTEB2 binding element (BTEB2) was used as the competitor DNA. The nucleotide sequences of the competitors are indicated in Table 2. (B) One microliter of rabbit antiserum to GST-IKLF supershifted the IKLF-EGFRE complex (arrow). F indicates the position of free probe.

kidney, brain and spleen. However, IKLF is distinct from GKLf. Recently, the human homolog of mouse GKLf was cloned and reported to share 91% sequence identity at the amino acid level with mouse GKLf (10). A similar degree of sequence identity also exists between human and mouse IKLF (12), suggesting that the protein is evolutionarily conserved. The tissue distribution of IKLF in human and mouse is similar, but not identical. The level of IKLF mRNA in skeletal muscle is substantially higher in human. Rabbit BTEB2 (35), which shares 98% sequence homology with human BTEB2 (25), also showed a comparable pattern of expression to human IKLF.

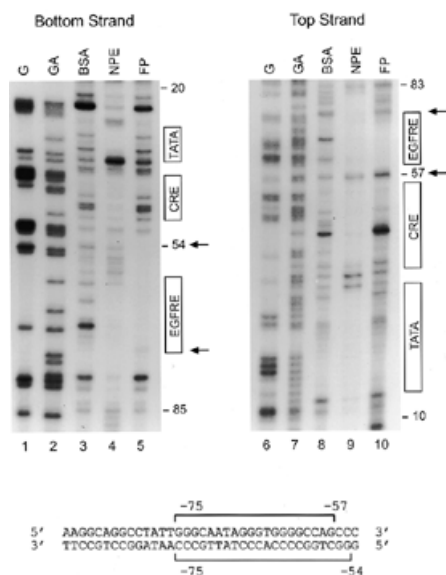


Figure 5. DNase I footprinting of the mouse lactoferrin EGFRE region by GST-IKLF. The RL95-2 nuclear protein extract (5 μ g) protected regions are indicated by the boxes and GST-IKLF (2 μ g) protected regions are indicated by arrows. Five micrograms of BSA was included in the control sample. Chemical sequencing reactions (G and G+A) performed on the same DNA fragment are shown as markers. The nucleotides protected by IKLF are indicated beneath the gel.

The diverse sequences of the Kruppel-like family members outside the zinc finger regions suggests that they may have different physiological functions in various cell types.

According to Pavletich and Pabo's model, the Kruppel-like gene family binds to a consensus site with the sequence CCNCNCCN (14) and each zinc finger can form specific contacts with a triplet of nucleotides in the DNA binding element. The interaction involves a one-to-one correspondence between the 3 bp of the binding element and the three amino acids in a zinc finger in an antiparallel orientation (36). The opposite orientation of the EGFRE sequence, AGGGTGGGG, fits the consensus binding element of Kruppel-like factors. Thus, the binding experiments shown in Figures 4 and 6 are consistent with the hypothesis that IKLF is a functional Kruppel-like gene family member. This analysis is also supported by the fact that sequence variants with mutations made within this region (m11, m12 and m15 in Table 2) lose their ability to compete with the wild-type sequence probe. Moreover, the CAAT of the EGFRE is also necessary for binding of IKLF to the EGFRE, since mutant 10 (Table 2) could not compete with the wild-type EGFRE alone or in the context of MRU. The present findings are in agreement with earlier studies indicating that the CAAT at the 5'-end of the EGFRE plays an important role in the EGF-responsiveness of RL95 cells. Deletion or mutation within this region significantly reduces the EGF-stimulated activity (22,23). Therefore, IKLF binding requires more than the consensus core. Furthermore, the GERE (37) and myc-specific GC element (Myc; 38), which contain the base triplet recognition sequence, did not compete as well as MRU and EGFRE in EMSA whereas BTE (BTEB2; 25), which differs from the core element, competes efficiently. Taken together, the results demonstrate that IKLF binding requires both the 5'-CAAT element and the base triplet

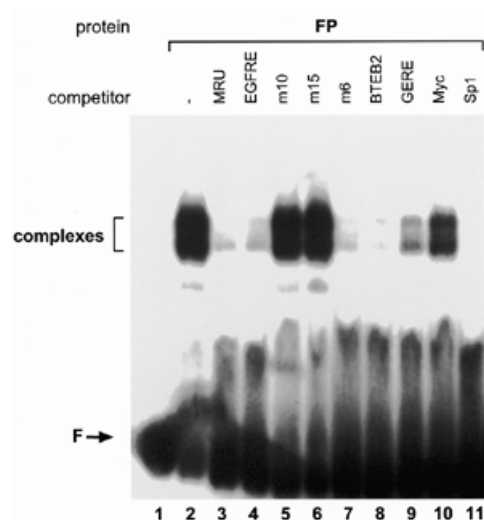


Figure 6. Specific interaction between GST-IKLF and MRU. Two micrograms of GST-IKLF fusion protein (FP) were incubated with the labeled MRU oligonucleotide. A 100-fold molar excess of unlabeled oligonucleotide was added as competitor (lanes 2–11). The nucleotide sequences of the competitors are indicated in Table 2. Lane 1, probe only. F indicates free probe.

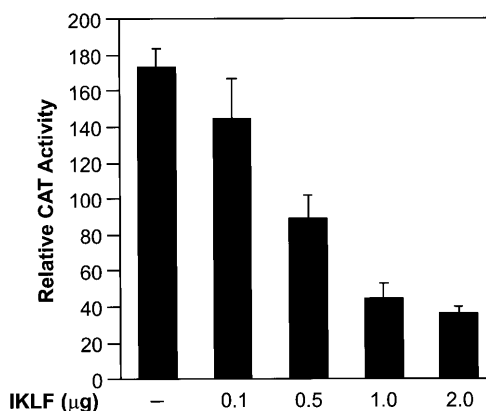


Figure 7. IKLF function as a repressor. The RL95-2 cells were co-transfected with an IKLF expression vector plasmid or vector plasmids in conjunction with the mLF0.1-CAT reporter constructs. Increasing amounts of IKLF expression vector in RL95 cells inhibit CAT reporter activity. The data shown are means \pm standard deviation of three experiments with duplicated samples.

recognition core element. This region includes the entire EGFRE that has been characterized (23), and this characterization might distinguish IKLF from other Kruppel-like family members. A similar requirement for an AT-rich 5'-flanking sequence for GC box binding has previously been observed for another zinc finger protein, MIG1 of the yeast *Saccharomyces cerevisiae* (39). Binding of MIG1 to a natural MIG1 binding site in the SUC2 promoter is determined by both base triplet recognition in the GC box and by the presence of an adjacent AT box. Few target genes have been established for the members of the family of Kruppel-like factors, although they have been shown to bind CACCC elements (7,10–12,33,35). It is important to examine whether subtle differences determine the binding specificity of each Kruppel-like factor interacting with an endogenous gene.

Most of the zinc finger transcription factors activate transcription of their target genes upon binding to their cognate *cis*-acting elements (reviewed in 1,2), and few of them act as repressors. The Kruppel family of proteins, however, have been shown to both activate and repress genes (9,13,18,20,40–44). Recently, the repression and activation domains of human EKLF (20) and hEZF (10), a human counterpart of mouse GKLf, were mapped in different regions of the proteins. Thus, the Kruppel family of proteins may modulate the target gene activity both positively and negatively. Therefore, that IKLF acts as a repressor of the mouse lactoferrin promoter in RL95 cells is not surprising. In fact, we have observed that IKLF can activate mouse lactoferrin promoter activity in other cell lines (unpublished observations).

The mechanism by which IKLF exerts its repression activity is not clear. Nonetheless, possession of repression domains, protein modification and competition for binding could play important roles in the observed transfection results. Like GKLf, EKLF and EZF, IKLF may contain a potential activation domain within its proline-rich N-terminal region and a repression domain within the C-terminal region rich in basic amino acids located 5' to the first zinc finger. The non-repetitive proline-rich region of IKLF could act as a 'sticky arm' and interact with other transcription factors (39). Competition for binding to a DNA element is a common way that transcription factors modulate each others functions (45). Data presented here (Fig. 6) suggest that IKLF can compete with Sp1, a known activator, for binding to the EGFRE response element. Therefore, IKLF may function as a repressor by competing with Sp1 at the level of DNA binding, thereby inhibiting Sp1-mediated activation.

Modification of IKLF by either phosphorylation or glycosylation could modulate its biological functions. In view of the number of potential phosphorylation sites in IKLF (Fig. 1), it is likely that IKLF is phosphorylated *in vivo*. Recently, Zhang and Bieker (46) reported that EKLF is acetylated by CBP/p300. Acetylated EKLF could serve as the bridging factor in chromatin structure modification, and thus induce β -globin expression. Further elucidation of the protein factors that are interacting with IKLF and post-translational modification of IKLF in response to EGF will shed light on the molecular mechanisms of EGF action in regulating the lactoferrin gene.

ACKNOWLEDGEMENTS

We gratefully acknowledge Raymond DuBois for generously providing us with the human colon cDNA library. We thank Wesley Gladwell for sequencing the 5' genomic region of IKLF and Yuji Mishnia and Minoru Koi for critically reading the manuscript. The paper has been edited by Dr Miriam Sander of Page One Editorial Service and Clara Beard.

REFERENCES

- Johnson,P.F. and McKnight,S.L. (1989) *Annu. Rev. Biochem.*, **58**, 799–839.
- Mitchell,P.J. and Tjian,R. (1989) *Science*, **245**, 371–378.
- Brown,R.S., Sander,C. and Argos,P. (1985) *FEBS Lett.*, **186**, 271–274.
- Miller,J., McLachlan,A.D. and Klug,A. (1985) *EMBO J.*, **4**, 1609–1614.
- Schuh,R., Alcher,W., Gual,U., Cote,S., Preiss,A., Maier,D., Seifert,E., Nauber,U., Shroder,C., Kemier,R. and Jackle,H. (1986) *Cell*, **47**, 1025–1032.
- Chowdhury,K., Deutsch,U. and Gruss,P. (1987) *Cell*, **48**, 771–778.

- Anderson,K.P., Kern,C.B., Crable,S.C. and Lingrel,J.B. (1995) *Mol. Cell. Biol.*, **15**, 5957–5965.
- Miller,J. and Bieker,J.J. (1993) *Mol. Cell. Biol.*, **13**, 2776–2786.
- Shields,J.M., Christy,R.J. and Yang,V.W. (1996) *J. Biol. Chem.*, **271**, 20009–20017.
- Yet,S.F., McA'Nulty,M.M., Folta,S.C., Yen,H.W., Yoshizumi,M., Hsieh,C.M., Layne,M.D., Chin,M.T., Wang,H., Perrella,M.A., Jain,M.K. and Lee,M.E. (1998) *J. Biol. Chem.*, **273**, 1026–1031.
- Crossley,M., Whitelaw,M., Perkins,A., Perkins,G., Fujiwara,Y. and Orkin,S.H. (1996) *Mol. Cell. Biol.*, **16**, 1695–1705.
- Conkright,M.D., Wani,M.A., Anderson,K.P. and Lingrel,J.B. (1999) *Nucleic Acids Res.*, **27**, 1263–1270.
- Liu,C., Calogero,A., Ragona,G., Adamson,E. and Mercola,D. (1996) *Crit. Rev. Oncogen.*, **7**, 101–125.
- Pavletich,N.P. and Pabo,C.O. (1991) *Science*, **252**, 809–817.
- Frasch,M. and Levine,M. (1987) *Genes Dev.*, **1**, 981–995.
- Ruppert,J.M., Kinzler,K.W., Wong,A.J., Bigner,S.H., Kao,F.T., Law,M.L., Seuznez,H.N., O'Brian,S.J. and Vogelstein,B. (1988) *Mol. Cell. Biol.*, **8**, 3104–3113.
- Donze,D., Townes,T.M. and Bieker,J.J. (1995) *J. Biol. Chem.*, **270**, 1955–1959.
- Gillemans,N., Tewari,R., Lindeboom,F., Rottier,R., de Wit,T., Wijgerde,M., Grosveld,F. and Philipsen,S. (1998) *Genes Dev.*, **12**, 2863–2873.
- Bieker,J.J. and Southwood,C.M. (1995) *Mol. Cell. Biol.*, **15**, 852–860.
- Chen,X. and Bieker,J.J. (1996) *EMBO J.*, **15**, 5888–5896.
- Licht,J.D., Grosse,M.J., Figge,J. and Hansen,U.M. (1990) *Nature*, **346**, 76–79.
- Shi,H. and Teng,C.T. (1994) *J. Biol. Chem.*, **269**, 12973–12980.
- Shi,H. and Teng,C. (1996) *Mol. Endocrinol.*, **10**, 732–741.
- Teng,C., Shi,H., Yang,N. and Shigeta,H. (1998) *Adv. Exp. Med. Biol.*, **433**, 65–78.
- Sogawa,K., Imataka,J., Yamasaki,Y., Kusume,H., Abe,H. and Fujii-Kuriyama,Y. (1993) *Nucleic Acids Res.*, **21**, 1527–1532.
- Yang,N., Shigeta,H., Shi,H. and Teng,C.T. (1996) *J. Biol. Chem.*, **271**, 5795–5804.
- Vinson,C.R., LaMarco,K.L., Johnson,P.F., Landschulz,W.H. and McKnight,S.L. (1988) *Genes Dev.*, **2**, 801–806.
- Ausubel,F.M., Brent,R.E., Kingston,R.E., Moore,D.D., Seidman,J.K., Smith,J.A. and Struhl,K. (1990) *Current Protocols in Molecular Biology*, Wiley, New York, NY, Ch. 7, pp. 1.1–1.8.
- Roesler,W.J., Vandenbark,G.R. and Hanson,R.W. (1989) *J. Biol. Chem.*, **264**, 9657–9664.
- Sambrook,J., Fritsch,E.F. and Maniatis,T. (1989) *Molecular Cloning: A Laboratory Manual*, 2nd Edn. Cold Spring Harbor Laboratory, Cold Spring Harbor, NY, Ch. 16, pp. 16.33–16.36.
- Kozak,M. (1987) *Nucleic Acids Res.*, **15**, 8125–8146.
- Proudfoot,N.J. and Brownlee,G.G. (1976) *Nature*, **263**, 211–214.
- Shields,J.M. and Yang,V.W. (1998) *Nucleic Acids Res.*, **26**, 796–802.
- Turner,J. and Crossley,M. (1999) *Trends Biochem. Sci.*, **24**, 236–241.
- Watanabe,N., Kurabayashi,M., Shimomura,Y., Kawai-Kowase,K., Hoshino,Y., Manabe,I., Watanabe,M., Aikawa,M., Kuro-o,M., Suzuki,T., Yazaki,Y. and Nagai,R. (1999) *Circ. Res.*, **85**, 182–191.
- Klevit,R.E. (1991) *Science*, **253**, 1367–1393.
- Merchant,J.L., Demedick,B. and Brand,S.J. (1991) *Mol. Cell. Biol.*, **11**, 2686–2696.
- Postel,E.H., Mango,S.E. and Flint,S.J. (1989) *Mol. Cell. Biol.*, **9**, 5123–5133.
- Lundin,M., Nehlin,J.O. and Ronne,H. (1994) *Mol. Cell. Biol.*, **14**, 1979–1985.
- Williamson,M.P. (1994) *Biochem. J.*, **297**, 249–260.
- Shi,Y., Seto,E., Chang,L.S. and Shenk,T. (1991) *Cell*, **67**, 377–388.
- Merchant,J.L., Iyer,G.R., Taylor,B.R., Kitchen,J.R., Mortensen,E.R., Wang,Z., Flintoft,R.J., Michel,J.B. and Bassel-Duby,R. (1996) *Mol. Cell. Biol.*, **16**, 6644–6653.
- Zhang,W., Shields,J.M., Sogawa,K., Fujii-Kuriyama,Y. and Yang,V.W. (1998) *J. Biol. Chem.*, **273**, 17917–17925.
- Kojima,S., Kobayashi,A., Gotoh,O., Ohkuma,Y., Fujii-Kuriyama,Y. and Sogawa,K. (1997) *J. Biochem.*, **121**, 389–396.
- Liu,Y.H., Yang,N. and Teng,C.T. (1993) *Mol. Cell. Biol.*, **13**, 1836–1846.
- Zhang,W. and Bieker,J.J. (1998) *Proc. Natl Acad. Sci. USA*, **95**, 9855–9860.
- Koritschoner,N.P., Bocco,J.L., Panzetta-Dutari,G.M., Dumur,C.I., Flury,A. and Patrino,L.C. (1997) *J. Biol. Chem.*, **272**, 9573–9580.
- Gidoni,D., Dyman,W.S. and Tjian,R. (1984) *Nature*, **312**, 409–413.