Cloning the cDNA for a New Human Zinc Finger Protein Defines a Group of Closely Related Krüppel-like Transcription Factors*

(Received for publication, June 26, 1998, and in revised form, August 4, 1998)

Nobukyuki Matsumoto‡, Friedrich Laubs§, Rafael Aldabe§, Wen Zhang§, Francesco Ramirez§, Teruhiko Yoshida‡, and Masaaki Terada‡

From the Genetics Division, National Cancer Center Research Institute, 5-1-1 Tsukiji, Chuo-Ku, Tokyo 104-0045, Japan and the Brookdale Center for Developmental and Molecular Biology, Mount Sinai School of Medicine, New York, New York 10029

We have identified a novel zinc finger protein that has been named ubiquitous Krüppel-like factor (UKLF) based on structural considerations and the pattern of gene expression. UKLF was isolated by the polymerase chain reaction approach using degenerate oligonucleotides corresponding to the DNA-binding domain of erythroid Krüppel-like factor (EKLF) and cDNA prepared from human vascular endothelial cells. The carboxyl-terminal portion of UKLF contains three zinc fingers of the Cys_{2}-His_{2} type and binds in vitro to the CAACCC motif of the β-globin promoter and to the Sp1 recognition sequence. The amino-terminal portion of UKLF consists of a hydrophobic region rich in serines and a negatively charged segment with several glutamic acid residues. The first 47 amino acids of the acidic region are nearly identical to the amino-terminal portion of another Krüppel-like factor, the so-called core promoter-binding protein (CPBP) or Zf9. Like CPBP/Zf9, UKLF can function as a transcription activator in co-transfection assays. However, this activity is lost when the highly conserved amino-terminal segment is deleted. These findings indicate that UKLF and CPBP/Zf9 represent a distinct subgroup of closely related Krüppel-like activators of transcription. Mapping of the UKLF gene to chromosome 2 suggested that UKLF and CPBP/Zf9 translocated to different chromosomes following duplication from an ancestral gene.

Proteins that bind to specific DNA sequences regulate a variety of developmental programs and cellular activities by activating or repressing transcription. There are several categories of transcription regulators, which are divided according to the structural motif that contacts the DNA (1). One such motif is the zinc finger motif, in which a zinc atom tetrahedrally coordinates by four residues participates in the formation of a 12-amino acid loop (2). The sequence, number, and organization of the zinc fingers, as well as the overall composition and functional contribution of the remainder of the molecule, segregate the zinc finger proteins into distinct classes and groups. One of these classes is characterized by the TFIIIA-like zinc finger Cys_{2}-His_{2} and additional sequence homology to the Drosophila segmentation gene product Krüppel (3, 4). Functional and structural features have identified a distinct group within the Krüppel-like class of mammalian nuclear factors. Members of this group exhibit a highly conserved carboxyl-terminal region with three zinc fingers and a more divergent amino-terminal domain rich in proline residues; they also transactivate gene expression after binding to similar G/C-rich recognition sequences (5–12). This group of Krüppel-like factors includes nuclear proteins specific for or enriched in erythroid cells (EKLF), lung (LKLF), gut (GKLF/EZF), and placenta (BTEB2), and a basic factor with wide tissue distribution (BKLF) (5–10).

EKLF, the founding member of this group of zinc finger proteins, was originally isolated as an erythroid-specific factor by subtractive hybridization (5). Subsequent gene targeting experiments in the mouse corroborated this notion, in that they documented the critical role of EKLF in regulating γ to β globin switching and liver erythropoiesis (13, 14). The DNA-binding domain of EKLF was subsequently employed to isolate the LKLF gene by cross-hybridization (7). Despite being mostly expressed in the lung and spleen, LKLF appears to be essential for the survival of naïve T cells and for the formation and stabilization of blood vessels during morphogenesis (15, 16). BKLF and GKLF/EZF were isolated by cross-hybridization with the zinc finger region of EKLF and by polymerase chain reaction (PCR) amplification using degenerate sequences of the same motif (8–10). Both genes were shown to be more broadly expressed than EKLF and LKLF (8–10). The cross-hybridization strategy was employed to isolate BTEB2 from a human placental library, and more recently, a sixth member of the EKLF group of proteins has been independently isolated from murine and human sources (6, 11, 12).

One group of investigators had previously identified an upstream sequence that acts as an initiator-like element in the

* This work was supported by a grant-in-aid for the 2nd Term Comprehensive 10-Year Strategy for Cancer Control and Health Sciences Research Grants from the Ministry of Health and Welfare of Japan, by the Mitsui Life Social Welfare Foundation, and by Grants AR-38648 and AR-42841 from the National Institutes of Health. This is article 268 from the Brookdale Center for Developmental and Molecular Biology. The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked “advertisement” in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

The nucleotide sequence(s) reported in this paper has been submitted to the GenBank™/EMBL Data Bank with accession number(s) AB015132.

‡ To whom correspondence should be addressed: Brookdale Center for Developmental and Molecular Biology, Mount Sinai School of Medicine, One Gustave L. Levy Pl., Box 1126, New York, NY 10029. Tel.: 212-241-7984; E-mail: ramirf01@doc.mssm.edu.

§ The abbreviations used are: BKLF, basic Krüppel-like factor; BTEB2, basic transcription element binding protein 2; CPBP, core promoter binding protein; CAT, chloramphenicol acetyltransferase; EMSA, electrophoretic mobility shift assay; GKF, gut Krüppel-like factor; EKLF, erythroid Krüppel-like factor; FISH, fluorescence in situ hybridization; GFP, green fluorescent protein; LKLF, lung Krüppel-like factor; LUC, luciferase; ORF, open reading frame; NLS, nuclear localization signal; RT, reverse transcriptase; PCR, polymerase chain reaction; RACE, rapid amplification of cDNA ends; TK, thymidine kinase; UKLF, ubiquitous Krüppel-like factor; HUVEC, human umbilical vein endothelial cell; PSG, pregnancy-specific glycoprotein; DAPI, 4′,6-diamidino-2-phenylindole.
TATA-less promoter of the pregnancy-specific glycoprotein 5 (PSG5) gene (17). This sequence was subsequently used to isolate a cDNA from a human placental cDNA expression library, the product of which was named core promoter binding protein (CPBP) (11). The other group of investigators identified the same transcript (and named it Zf9) in a subtraction hybridization screening for genes induced in a rat model of liver fibrosis (12, 18). CPBP/Zf9 contains a unique amino-terminal domain that is rich in serine/proline clusters and leucines (11, 12). Recombinant CPBP/Zf9 binds in vitro to the core promoter element of the TATA-less PSG5 gene and to the recognition sequence for Sp1 (11, 12). In co-transfection assays, CPBP/Zf9 increases severalfold the transcription from a heterologous promoter that harbors the PSG5 element or G/C boxes (11, 12). CPBP/Zf9 is expressed in all tissues to some degree and is maximally expressed in placental cells (11). Finally, CPBP/Zf9 biosynthesis increases markedly in response to injury in activated liver stellate cells (12).
Toward this end, we have initiated a systematic search for potential regulators that are expressed in chondrocytes and endothelial cells. In this search, we have employed the PCR approach and degenerate primers corresponding to the DNA-binding domains of each of the major categories of transcription factors. In our initial study, we have looked for new members of the EKLF group of proteins among the transcripts of vascular endothelial and rat chondrosarcoma cells. While the results of the cartilage-specific factor will be reported elsewhere, here we describe the isolation from endothelial cells of a new member of the EKLF group of proteins that is structurally related to CPBP/Zf9. The kinship between the two proteins is most evident in the overall composition of the amino-terminal domain and in the near identity of the first 47 residues. This highly conserved region maps up most of the transactivating segment of the protein that we have identified. Pending the results of additional analyses, we have preliminarily named it ubiquitin-related protein that we have identified.

EXPERIMENTAL PROCEDURES

Materials—DNA modifying enzymes, Taq polymerase, and Seque-dNTPs were purchased from Takara Bio Inc. (Beverly, MA); Agilent Technologies; Promega Corp. (Madison, WI); Boehringer Mannheim, New England Biolabs Inc. (Beverly, MA); United States Biochemical Corp. (Cleveland, OH); and Perkin-Elmer; radioisotopes were purchased from NEN Life Science Products; media and serum were purchased from Life Technologies, Inc. (Rockville, MD); and Hyclone Laboratories (Logan, UT). Human cell lines were obtained from the following sources: Cascade Biologics, Inc. (Portland, OR) and Human Epithelial Cells, Clonetics Corp. (Walkersville, MD); kidney mesangial cells and umbilical vein endothelial cells (HUVECs); Cell Systems (Kirkland, WA), keratinocytes, hepatocytes, and vascular smooth muscle cells; and the ATCC repository (Camden, NJ), WI-38 lung fibroblasts. The His-tagged prokaryotic expression vector pET30a (+) was purchased from Novagen (Madison, WI), whereas the eukaryotic expression vector pEGFP-C3 containing the jellyfish green fluorescent protein (GFP) gene was purchased from CLONTECH Laboratories (Palo Alto, CA). The following plasmids were used for the transcription analysis: pDN3A3, which contains the cytomegalovirus promoter (Inovitrogen, Carlsbad, CA); pcEG1, which contains the SV40 promoter and the sequence coding for the GALA DNA-binding domain (amino acids 1-147); pBGX1-GAL4, which contains the SV40 promoter and the whole GAL4 gene; pG5BTKCAT, which contains five GAL4 binding sites placed 5' of the basal TATA box promoter and reporter gene; and pG5BTKCAT, which contains five GAL4 binding sites placed 5' of the minimal thymidine kinase (TK) promoter and reporter gene. The GAL4-containing plasmids were generously supplied by Dr. M. Ptashne (Memorial Sloan-Kettering Cancer Institute, New York, NY). DNA fragments and PCR products were subcloned into pBluescript II SK (Stratagene, La Jolla, CA), PCR-Blunt (Inovitrogen, Carlsbad, CA), and pGEM-T Easy (Promega).

Cloning UKLF—cDNA was derived from HUVEC DNA using the reverse transcriptase (RT) reaction and employed for PCR amplification with degenerate primers (forward: 5'-ACHCAAYCHGNAGARAARCC-3'; reverse: 5'-TTCAATRGTGANGNCAGRTG-3'; see Fig. 1A). Denaturation was at 94 °C for 1 min, annealing was at 60 °C for 1 min, and extension was at 72 °C for 1 min. Amplified products were visualized after agarose gel electrophoresis by ethidium bromide staining; the expected 134-bp band was isolated and subcloned as cDNA fragment U64Z17 (19). To generate 3' overlapping clones, the 3' RACE (rapid amplification of cDNA ends) was performed using HUVEC RNA as a template and nested primers that correspond to the 5' sequences of U64Z17 (5'-TTGGAGGAGATGCGGC-3' and 5'-GGAGGAGATGCGGC-3') in combination with 3' DTT primer; i.e., an oligo(dT) stretch followed by the 7mer (20). To generate 5' overlapping clones, the circular first-strand cDNA-mediated RACE (cRACE) was employed on the HUVEC RNA template using the 5' full RACE core set (Takara Corp.) and the following primers: 5'-GATGCTGATGCTC-3' for the gene-specific reverse transcription; 5'-TTGGCATGCAGGAGATGTC-3' and 5'-GCAGGAGATGCGGC-3' for the first round of PCR; and 5'-TTGGCATGCAGGAGATGTC-3' and 5'-GATGCTGATGCTC-3' for the nested PCR (21). To generate the single clone (pUL8) that spans from the 5' and 3' untranslated regions and includes the entire open reading frame (ORF), the reverse transcriptase PCR amplification protocol was employed.

Laub et al., manuscript in preparation.
of HUVEC RNA was performed with the sequence information derived from appropriate 3' RACE and cRACE products. Clone pUL8 was chosen after verification of the sequences of 9 clones from the 5' cRACE, 11 clones from the 3' RACE, and 14 clones from the full-sized RT-PCR amplification. A BLAST search of expressed sequence tags localized the mouse sequence to expressed sequence tag AA146238.

Northern Analysis and Gene Mapping—Total RNA was isolated from various cell lines as described (22). Twenty micrograms of total RNA was electrophoresed through formaldehyde gels, transferred to Hybond N+ nylon filters (Amersham Pharmacia Biotech), and hybridized to the 465-bp 3'-untranslated region probe of pUL8 (20). The same DNA fragment was used to probe commercial filters (Human Multiple Tissue Northern, CLONTECH Laboratories) containing poly(A)^+ RNA from adult and fetal tissues. The procedure for fluorescence in situ hybridization (FISH) mapping was performed by a commercial firm (SeeDNA, North York, Ontario, Canada) and according to the published protocol (23). Briefly, chromosomal slides of synchronized human lymphocytes were baked at 55 °C for 1 h. After RNase treatment, they were denatured in 70% formamide in 2× SSC for 2 min at 70 °C and then dehydrated with ethanol. A 2.2-kb 3’ RACE fragment coding for the last 44 amino acids of the zinc finger protein was biotinylated with dATP using the BioNick labeling kit (Life Technologies, Inc.). After denaturation at 75 °C for 5 min in 50% formamide and 10% dextrane sulfate, the probe was loaded on the denatured chromosomal slides. Following overnight hybridization, FISH signals and DAPI banding patterns were recorded separately by photography; assignment of the FISH mapping to a specific chromosomal segment was achieved by superimposing the two sets of images.

Recombinant UKLF Protein Expression and in Vitro DNA Binding—The sequence for the zinc finger domain of UKLF (amino acids 212–302) was fused in-frame to the sequence for the amino-terminal histidine of the pET30a (+) vector in the pET-D2 plasmid. The His-tagged fusion protein was expressed in the BL21 (DE3) strain of E. coli and purified with Ni-NTA resin (Qiagen, Santa Clarita, CA). The recombinant protein was purified according to the manufacturer’s recommendations, first by eluting it from the resin in a 50 mM sodium phosphate buffer, pH 6.0 (300 mM NaCl, 10% glycerol, 0.5 mM imidazole, 1 mM phenylmethylsulfonyl fluoride, 20 μg/ml aprotinin, and 100 μg/ml leupeptin), and subsequently by overnight dialysis at 4 °C against 10 mM Tris-Cl, pH 7.4, 100 mM NaCl, 10 mM ZnCl2, 10% glycerol, 1 mM phenylmethylsulfonyl fluoride, 20 μg/ml aprotinin, and 100 μg/ml leupeptin. The recombinant protein was divided into 20-μl aliquots at an estimated concentration of 50 ng/μl and stored at –80 °C. For the electrophoretic mobility shift assay (EMSA), double-stranded oligonucleotides were

![Fig. 3. Expression of UKLF mRNA in human tissues.](image-url)

The relative intensity of the bands was quantified by the public domain NIH Image program (http://rsb.info.nih.gov/nih-image/). Relative UKLF expression is the sum of the intensity of the β-actin band. Because the intensity varies depending on many factors, the histograms of Figs. 3 and 4 cannot be compared with each other.

![Fig. 4. Expression of UKLF RNA in human primary cell lines.](image-url)

The signals were quantified as described in the legend to Fig. 3.
labeled with Klenow enzyme and [α-32P]dCTP or with polynucleotide kinase and [γ-32P]dATP (19). Double-stranded oligonucleotides included the β-globin promoter sequence containing the core CACCC site (5'-AGCTAGCCACCCCTGAAGCT-3'), the Sp1 binding site (5'-ATTCGATCGGGGCGGGGCGAGC-3'), and the GATA-1 recognition sequence (5'-CACTTGATAACAGAAAGTGATAACTCT-3'). In a typical assay, 10,000–20,000 cpm of probe were incubated for 30 min on ice with 250 ng of recombinant pET-D2 product in the presence of 3 μg of poly(dI-dC)*poly(dI-dC) as described (19). The reaction products were then fractionated through a 6% polyacrylamide gel electrophoresis in 0.5× Tris borate-EDTA buffer (24).

Cellular Localization of Recombinant UKLF Proteins—The coding sequence for the UKLF protein (BglII/XhoI fragment of pBX-FL; see below) was fused in-frame to the GFP gene in the pEGFP-C3 vector in the construct pGFP-FL. Likewise, the inserts of plasmids pET-D2 (see above) and pBX-D3 (see below) were fused in-frame to GFP in the deletion mutants pGFP-D2 and pGFP-D3, respectively; the UKLF sequence begins at amino acid 212 in the former construct and at residue 221 in the latter. Self-ligation of XbaI-digested pGFP-FL DNA yielded the pGFP-D4 plasmid, in which the UKLF sequence ends at residue 120. About 10 μg of each construct was transiently transfected into 70% confluent NIH-3T3 cells using the LipofectAMINE procedure according to the manufacturer's instructions (Life Technologies, Inc.). The cellular localization of the various UKLF-GFP fusion proteins was monitored 36 h later by autofluorescence microscopy.

Transcriptional Activity—Various segments of the UKLF gene were fused in-frame to the sequence for the DNA-binding domain of GAL4 (DBDG4, residues 1–143) in the expression vector pBXG1. The fragments included the full-length protein (AflIII/NcoI digest of pUL8 subcloned into the blunted BamHI site of pBXG1; pBX-FL), the 5' deletion from amino acids 1–72 (HindIII/NcoI digest of pUL8 subcloned into the Smal site of pBXG1; pBX-D1), the 5' deletion from amino acids 1–117 (XbaI digest of pBX-FL subcloned into the Smal site of pBXG1; pBX-D2); and the 5' deletion from amino acids 1–225 (BamHI/BamHI digest of pBX-D1 subcloned into the blunted BamHI site of pBXG1; pBX-D3). The whole UKLF sequence was also subcloned in the expression vector pcDNA3. Either 0.5 or 2.5 μg of each expression vector was transfected together with 2 μg of a control plasmid and 5 μg of the reporter plasmid containing five GAL4 binding sites. Chloroamphenicol acetyltransferase (CAT) activities were measured using a liquid scintillation assay, and the resulting values were normalized against the luciferase (LUC) activity of the co-transfected control plasmid pSVLUC or pTKLUC (19). Transfections were performed multiple times and in duplicate by lipofection (human embryonic 293T kidney cells) or calcium phosphate precipitation (mouse NIH-3T3 fibroblasts); results were evaluated by the Student's t test (24).

RESULTS

Cloning UKLF—A comparative RT-PCR strategy on transcripts from a variety of cells was initially used in order to identify one that is specifically expressed in HUVECs and of the size and composition predicted for a peptide of the EKLF type. Aside from HUVECs, RNA samples included those pre-

FIG. 5. UKLF gene mapping. A, an example of FISH signals (arrow) on chromosome 2 (top); the bottom panel shows the same mitotic figure stained with DAPI. B, diagram summarizing the FISH mapping results; each dot represents a double FISH signal on chromosome 2.
pared from fibroblasts and skeletal muscles, from erythroleukemia and hepatocarcinoma lines, and from pancreatic and gastric tumors. The expected 134-bp-long product was observed in the samples amplified from HUVECs and the erythroleukemia and hepatocarcinoma lines. Sequencing of 25 randomly chosen subclones identified transcripts from the BKLF (four clones) and CPBP/Zf9 (two clones) genes, as well as from four non-Krüppel zinc finger genes; the remaining clones encoded proteins unrelated to zinc finger factors. The same experiment was repeated after increasing the annealing temperature from 51 to 60 °C and only with HUVEC RNA. Eight of the 134-bp subclones were sequenced; three were found to correspond to GKLF, one to BKLF, one to CPBP, and one to a protein without zinc fingers. The last two subclones yielded identical sequences (U64Z17), potentially coding for a new peptide of the EKLF group (Fig. 1A).

Consistent with this observation, in vitro transcription/translation of pUL8 yielded a single product with an estimated molecular mass virtually identical to the predicted 33.36 kDa of the conceptual cDNA translation product (data not shown). The primary structure of the 34-kDa product consists of three distinct domains (Fig. 1D). At the carboxyl terminus of the molecule (residues 221–302), there is a potential DNA-binding domain that is made of three zinc fingers of the Cys2-His2 type separated by spacers adhering to the consensus sequence TGEKP(Y/F)X (Fig. 1B). Pairwise comparisons of the putative DNA-binding domain scored the closest homology to CPBP/Zf9, followed by BTEB2, EKLF, and GKLF (Fig. 2A). The central portion of the 34-kDa protein (residues 76–220) is characterized by a high concentration of hydrophobic residues (particularly alanines, leucines, and valines) and by numerous serines. Together, these four amino acids make up more than 50% of the hydrophobic domain. The homology search also identified a potential lucine zipper sequence between amino acids 98 and 119 and the nuclear localization signal (NLS) KKRVHR (residues 215–220) immediately before the first zinc finger (Fig. 1B). The third domain (residues 1–75) is negatively charged and enriched in serines, alanines, leucines, and glutamic acid. The amino terminus of this domain (residues 1–47) is nearly identical to the amino terminus of CPBP/Zf9 (Fig. 1C) (11, 12). This same sequence has been reported in one of the
two ORFs of the Bcl oncogene, the other ORF of which encodes a single Cys$_2$HIs$_2$ zinc finger with homology to BTEB2 (25).

The pattern of expression of the pUL8 transcript was preliminarily assessed by Northern analysis of RNA prepared from adult and embryonic human tissues, as well as from a representative group of human cell lines (Figs. 3 and 4). The 3′-untranslated region probe of pUL8 identified three major transcripts (8.3, 4.6, and 2.6 kilobases in size) in virtually all tissues and cell lines. The multiple mRNAs may represent alternative splicing, alternative polyadenylation sites, or a combination of both. Although their significance is currently unclear, it is interesting to note that the relative level of expression of the mRNAs varies in different tissues (Fig. 4). Major tissues that express the pUL8 transcript appear to be kidney and brain in the embryo and brain and spinal cord in the adult organism (Fig. 3). Keratinocytes and lung fibroblasts are the cell lines that express pUL8 the most, followed by kidney mesangial cells and HUVECs (Fig. 4). The same probe was used to establish the chromosomal location of the UKLF gene using the FISH approach (23). Of 100 mitotic figures examined, 74 displayed a positive signal on one autosomal pair (Fig. 5A). DAPI banding correlated the signal to the long arm of chromosome 2; to be precise, to band q32 (Fig. 5B). Pending completion of ongoing developmental analyses, the product of the pUL8 cDNA has been preliminarily named UKLF for ubiquitous Krüppel-like factor.

**UKLF Is a Nuclear Protein That Binds DNA**—In the next set of experiments, we sought experimental proof for UKLF binding to DNA. To this end, we first expressed the full-sized and 5′ deletions of UKLF fused in-frame to GFP and monitored the cellular localization of the fusion products. Autofluorescence analysis of the transfected cells demonstrated that full-sized UKLF-GFP (pGFP-FL) translocates and accumulates in the nucleus (Fig. 6). The same result was obtained with the pGFP-D2 plasmid, which lacks the first 211 residues of UKLF (Fig. 6). By contrast, the fusion product that contains only the zinc finger domain of UKLF (pGFP-D3) accumulated in the cytosol, and in the nucleus to a lesser extent than the other two fusion products (Fig. 6). This last result is consistent with the presence of the NLS between residues 215 and 220 (Fig. 1B). Expression of the truncated protein without the NLS and the whole zinc finger domain (pGFP-D4) completely eliminated accumulation of the GFP fusion product in the NIH-3T3 nuclei (Fig. 6). We interpret the finding as suggesting that one or more NLSs are located within the zinc finger domain of UKLF, carboxyl-terminal of the major NLS. These transfection assays therefore demonstrated that UKLF is a nuclear protein, in addition to identifying the region responsible for nuclear translocation.

The near identity of the zinc finger domains equates with the binding of the EKLF group of proteins to closely related G/C-rich sequences (Fig. 2B) (5–7, 9–12, 26, 27). To test the ability of UKLF to bind to G/C-rich DNA, we generated a recombinant fusion protein (pET-D2) that contains the amino-terminal His-tagged peptide fused in-frame to the Cys$_2$HIs$_2$ zinc finger. Increasing amounts of the fusion protein were incubated in vitro with oligonucleotides containing CACCC, GATA-1, and Sp1 binding sites (see Fig. 2B) and analyzed by EMSA. 250 ng of recombinant UKLF were competed with increasing amounts (50-, 250-, 500-, and 1000-fold excess) of unlabeled sequences for CACCC, Sp1 binding site, and an unrelated sequence (Unrel). The specificity of binding was demonstrated by competing each of the two complexes with excess of the respective unlabeled sequence (Fig. 7B). The data of the EMSAs therefore indicated that UKLF has the propensity to bind to G/C-rich sequences of DNA.

**UKLF Is a Transcriptional Activator**—As a final step in our initial characterization of UKLF, we determined whether this DNA-binding protein has any effect on gene transcription. Accordingly, various segments of UKLF fused in-frame to the DNA-binding domain of GAL4 were co-transfected in 293T cells together with a reporter gene that contains five GAL4 binding sites in front of the basal TATA box promoter. In addition to the expression plasmid for the full-length UKLF (pBX-FL), we tested those encoding UKLF products without the conserved region of the acidic domain (deletion 1–72, pBX-D1), without the acidic domain and the amino-terminal portion of the hydrophobic domain (deletion 1–117, pBX-D2), and with only the zinc finger domain (deletion 1–225, pBX-D3). The full-length UKLF fusion was the sole product that activated transcription of the reporter gene (Fig. 8A). Transcriptional activation by UKLF was about 30% the level of GAL4 (data not shown). The same experiment was repeated with NIH-3T3 fibroblasts and using the reporter plasmid that contains the minimal TK promoter preceded by five GAL4 binding sites. Aside from assessing UKLF activity within a different cellular context, the test was designed to determine whether deletion of the conserved region in the acidic domain may unmasks a repressing activity in UKLF. The results demonstrated once again that the trans-activating property of the full-sized protein resides in its amino-terminal region, in addition to excluding repression by the amino-terminally deleted product (Fig. 8B). Transcriptional activation by UKLF was about 2-fold more than by GAL4 (data not shown). In conclusion, the transfection assays strongly suggested that the highly conserved acidic domain is likely to
account for most if not all of the transactivating property of UKLF.

**DISCUSSION**

In the present study, we have described the identification and initial characterization of UKLF, a novel member of the EKLF group of transcription factors. UKLF is expressed broadly and has transactivating property. Like other members of the EKLF group, a bacterially expressed fusion protein that contains the zinc finger domain of UKLF binds in vitro to a G/C-rich sequence. Work in progress is aimed at identifying genes that are potential targets EKLF by screening a CpG island library for natural UKLF binding sites (28). Similarly, we are trying to determine the functional significance, if any, of the multiple mRNA species observed in the Northern analysis. One exciting possibility is that they may represent alternatively spliced products coding for isoforms of UKLF with distinct, if not opposite, regulatory properties.

The amino-terminal region of UKLF is as acidic as those of many transactivators, including yeast GAL4 (29). We have provided experimental support for this structural correlation. First, UKLF fused to a heterologous DNA-binding domain exhibits strong activating properties; and second, deletion of the acidic domain abrogates UKLF ability to function as a transcription activator. Additional mutations may document the importance of other sequences, such as the hydrophobic residues or the alleged leucine-zipper, in conferring full regulatory activity to UKLF (30, 31). For example, functional analyses of selected GKLF/EZF and EKLF deletions have recently demonstrated that both factors harbor activation and repression subdomains (27, 32).

The expression pattern of UKLF is apparently inconsistent with our original goal to identify endothelial cell-specific genes involved in determining the formation and/or function of vascular tissue. It has been argued that the EKLF group consists of structurally and functionally related regulators of cell differentiation and tissue morphogenesis (7, 9, 27). Indeed, ablation of EKLF and LKLF gene function has documented the critical role that these factors play in erythropoiesis, lymphopoiesis, and vasculogenesis (13–16). However, the consequence of LKLF loss on mouse development is rather restricted compared with its broad tissue distribution (7, 15, 16). Thus, the pattern of gene expression may not always predict the role of these factors in selected cellular and morphogenetic programs. Accordingly, the ubiquitous connotation that we have given to our 34-kDa protein should not be construed as indicative of function. Further clues are expected from ongoing in situ hybridizations, which may identify the cellular source of UKLF in the organs in which transcript accumulation is most substantial, such as brain and kidney. This information will complement the identification of UKLF genomic targets by correlating the spatiotemporal pattern of expression of the effector and target genes.

Structural and functional considerations place UKLF and CPBP/Zf9 close together within the EKLF group. They include the high homology of the DNA-binding domains at the carboxyl terminus and the near identity of the activation domains at the amino terminus. Conservation of major and minor NLSs have been similarly used to cluster together GKLF, LKLF, and EKLF (33). CPBP/Zf9 is a transcriptional activator with G/C-rich sequence binding preference and with low expression in most tissues except placenta (11, 12). One study has suggested that CPBP/Zf9 may participate in the regulation and constitu-
tive expression of the placental PSG5 gene (11). The authors have argued that the acidic character of the amino terminus of CPBP/Zf9 may favor interaction with the basic domains of general transcription factors, such as TFIIB (11). They have further hypothesized that CPBP/Zf9 may therefore be a general regulator of TATA-less genes (11).

Another study has documented the positive biosynthetic response of CPBP/Zf9 during the tissue repair process that follows liver injury in vivo (12). This process requires activation of resident mesenchymal cells, which in the liver are the stellate cells located in the subendothelial space of the sinusoid (34). Activation of stellate cells leads to induction of genes coding for structural proteins, such as the collagens, and for cytokines and their receptors, such as TGF-β and cognate receptors (34). The authors therefore proposed that CPBP/Zf9 may mediate critical intracellular signals by the stellate cells in response to liver injury (12). Interestingly, the analogous role of stellate cells in the repair process of kidney and lung is performed by mesangial cells and fibroblasts, both of which are high UKL family proteins.

The human CPBP/Zf9 gene has been recently mapped to chromosome 10 p15 and re-named core promoter element binding protein (37). The localization of UKL to chromosome 2q32 implies that the two genes are evolutionarily ancient, in that they translocated to different chromosomes after duplicating from a common progenitor gene. We are currently searching for a gene in the DNA of invertebrate species. The mapping of the UKL gene has also raised the possibility that it may be causally involved in the etiopathogenesis of one of the disorders that have been linked to chromosome 2q32. They include wrinkle skin syndrome, familial persistent pulmonary embolism, and a familial form of heart failure (OMIM; http://www.ncbi.nlm.nih.gov/Omim/). Along these lines, we are currently involved in comparing the expression patterns of the mouse UKL and CPBP/Zf9 genes during embryogenesis in order to gain insight into their possible function(s). This kind of information will be also invaluable to interpret the phenotypes of transgenic mice that express UKL and CPBP/Zf9 ectopically and of mice that lack either or both of these genes. These genetic studies and the full biochemical characterization of the proteins will eventually elucidate the functional role(s) of this new subgroup of transcription factors.

Acknowledgments—We thank Drs. J. Bieker, S. Friedman, H. Ide, S. Ohnishi, and H. Sasaki for many helpful discussions and suggestions, Dr. M. Ptashne for providing invaluable reagents, C. Else for excellent technical assistance, and K. Johnson for typing the manuscript.

REFERENCES

Cloning the cDNA for a New Human Zinc Finger Protein Defines a Group of Closely Related Krüppel-like Transcription Factors

Nobukyuki Matsumoto, Friedrich Laub, Rafael Aldabe, Wen Zhang, Francesco Ramirez, Teruhiko Yoshida and Masaaki Terada

doi: 10.1074/jbc.273.43.28229

Access the most updated version of this article at http://www.jbc.org/content/273/43/28229

Alerts:
- When this article is cited
- When a correction for this article is posted

Click here to choose from all of JBC's e-mail alerts

This article cites 37 references, 17 of which can be accessed free at http://www.jbc.org/content/273/43/28229.full.html#ref-list-1