796–802 *Nucleic Acids Research, 1998, Vol. 26, No. 3*

Identification of the DNA sequence that interacts with the gut-enriched Krüppel-like factor

Janiel M. Shields[1,+] and Vincent W. Yang[1,2,*]

[1]Department of Medicine and [2]Department of Biological Chemistry, The Johns Hopkins University School of Medicine, Baltimore, MD 21205, USA

Received October 1, 1997; Revised and Accepted November 26, 1997

ABSTRACT

The gut-enriched Krüppel-like factor (GKLF) is a recently identified eukaryotic transcription factor that contains three C_2H_2 zinc fingers. The amino acid sequence of the zinc finger portion of GKLF is closely related to several Krüppel proteins, including the lung Krüppel-like factor (LKLF), the erythroid Krüppel-like factor (EKLF) and the basic transcription element binding protein 2 (BTEB2). The DNA sequence to which GKLF binds has not been definitively established. In the present study we determined the DNA binding sequence of GKLF using highly purified recombinant GKLF in a target detection assay of an oligonucleotide library consisting of random sequences. Upon repeated rounds of selection and subsequent characterization of the selected sequences by base-specific mutagenesis a DNA with the sequence 5′-$G/_AG/_AGGC/_TGC/_T$-3′ was found to contain the minimal essential binding site for GKLF. This sequence is present in the promoters of two previously characterized genes: the CACCC element of the β-globin gene, which interacts with EKLF, and the basic transcription element (BTE) of the *CYP1A1* gene, which interacts with Sp1 and several Sp1-like transcription factors. Moreover, the selected GKLF binding sequence was capable of mediating transactivation of a linked reporter gene by GKLF in co-transfection experiments. Our results establish GKLF as a sequence-specific transcription factor likely involved in regulation of expression of endogenous genes.

INTRODUCTION

The zinc finger is a common structural motif used by transcription factors to bind DNA (1,2). The Cys_2His_2 (C_2H_2) type zinc fingers, initially identified in the *Xenopus laevis* transcription factor TFIIIA (3), represent the most abundant DNA binding motif of zinc finger-containing proteins (4). A characteristic three-dimensional structure of a C_2H_2 zinc finger consists of two antiparallel β-strands followed by an α-helix (5). This structure in turn contacts 3 bp of DNA (2). Most zinc finger transcription factors contain multiple tandem repeats of the fingers, which confer the sequence specificity of the DNA that the individual factor recognizes.

A subset of C_2H_2 zinc finger proteins contains additional homology to the *Drosophila* segmentation gene product Krüppel (6). This homology is present in the region between two adjacent fingers and contains the highly conserved sequence TGEKP$^Y/_F$X. Examples of Krüppel-related proteins include Sp1 (7), zif268/Egr-1 (8), WT-1 (9) and EKLF (10). These proteins are collectively involved in diverse aspects of eukaryotic gene regulation during growth, development and differentiation.

We recently identified a novel C_2H_2 zinc finger protein with Krüppel homology, which we named the gut-enriched Krüppel-like factor (GKLF) (11). Expression of GKLF is enriched in epithelial cells of the gastrointestinal tract (11,12) and in the epidermal layer of the skin (12). In cultured cells expression of GKLF is induced under conditions that promote growth arrest, such as serum deprivation and contact inhibition (11). In addition, enforced expression of GKLF in transfected cells results in inhibition of DNA synthesis (11). Taken together, these findings suggest that GKLF may have an important function in regulating proliferation of epithelial tissues.

Although the amino acid sequence of GKLF outside the zinc finger region is unique, that of its three zinc fingers is closely related to several Krüppel proteins, including LKLF (13), EKLF (10) and BTEB2 (14). Recently, by comparing the amino acid sequences necessary for nuclear localization, we observed that GKLF is more closely related to LKLF and EKLF than to BTEB2 (15). The DNA sequences with which these transcription factors interact also appear to share some degree of similarity; GKLF was shown to interact with the CACCC sequence that binds EKLF and the basic transcription element (BTE) that binds BTEB2 (12). Nevertheless, no methodical analysis of the DNA recognition sequence for GKLF has been performed to date. We therefore undertook the task of empirically determining the GKLF binding sequence, since we deemed this information essential for eventual deciphering of the biological functions of GKLF.

*To whom correspondence should be addressed at: Ross 918, The Johns Hopkins University School of Medicine, 720 Rutland Avenue, Baltimore, MD 21205, USA. Tel: +1 410 955 9691; Fax: +1 410 955 9677; Email: vyang@welchlink.welch.jhu.edu

+Present address: Department of Pharmacology, University of North Carolina, Mary Ellen Jones Building, Chapel Hill, NC 27599, USA

MATERIALS AND METHODS

Production of recombinant GKLF

cDNA encoding the C-terminus of GKLF between amino acids 350 and 483, which contains the three zinc fingers, was subcloned into the *Xho*I–*Bam*HI sites of the bacterial expression vector PET-16b (Novagen; Madison, WI) to produce a GKLF fusion protein tagged with 10 histidine residues at the N-terminus. This protein was named His–GKLFZn. *Esherichia coli* strain BL21(DE3)pLysS (Novagen) was transformed with the recombinant plasmid and induced with 2 mM isopropyl-β-D-thiogalactopyranoside (IPTG) to produce recombinant His–GKLFZn protein. Five hundred milliliters of logarithmically growing bacteria were pelleted by centrifugation and resuspended in 20 ml buffer containing 20 mM Tris–HCl, pH 7.9, 0.5 M NaCl, 6 M urea and 5 mM imidazole. The suspension was sonicated with a Fisher Scientific 550 Sonic Dismembrator at a setting of 50% for 20 s at a time for a total of 20 times, chilling to 4°C between sonications. The sample was then flowed through a Ni–NTA–agarose column (Qiagen, Santa Clarita, CA) equilibrated with the above buffer. After washing the column with a buffer containing 20 mM Tris–HCl, pH 7.9, 0.5 M NaCl, 6 M urea and 60 mM imidazole the bound protein was eluted with the same buffer with the exception that the concentration of imidazole was raised to 1 M. The eluted fractions containing His–GKLFZn were dialyzed exhaustively against a solution of 10 mM Tris–HCl, pH 7.4, 100 mM NaCl, 10 µM $ZnCl_2$ and 10% glycerol. The protein was stored at –70°C at an approximate concentration of 2 mg/ml.

Target detection assay

The target detection assay (TDA) was performed based on a protocol by Thiesen and Bach (16), with some modifications. A library of single-stranded oligonucleotides containing the sequences 5′-CAAGCTTACTGCAGATGC$(N)_{14}$CGTAGGATCCATCTAGAGT-3′ (N is any nucleotide) was generated. The invariable 5′- and 3′-flanking sequences contained unique restriction sites. Fifteen micrograms of the library were made double-stranded with 5 µg reverse primer of sequence 5′-ACTCTAGATGGATCCTACG-3′ in a reaction that contained 200 µM each of the four deoxynucleoside triphosphates, 25 mM Tris–HCl, pH 8.3, 50 mM KCl, 2 mM $MgCl_2$, 1 mM dithiothreitol (DTT) and 20 U Taq DNA polymerase (Boehringer Mannheim, Indiannoplis, IN) for one cycle of 98°C for 3 min, 94°C for 1 min, 47°C for 2 min, 72°C for 1 min and 72°C for 30 min in a Perkin Elmer thermocycler. Fifty nanograms of the double-stranded library were end-labeled with T4 polynucleotide kinase and 100 µCi [γ-^{32}P]ATP in a reaction containing 70 mM Tris–HCl, pH 7.6, 10 mM $MgCl_2$ and 5 mM DTT at 37°C for 30 min. Four hundred nanograms of purified His–GKLFZn were added to the labeled library in a binding buffer (BB) of 10 mM Tris–HCl, pH 7.5, 150 mM NaCl, 20 mM KCl, 10 mM $MgCl_2$ and 5 µM $ZnCl_2$, at 4°C for 20 min, following which the mixture was electrophoresed in a 7% non-denaturing polyacrylamide gel in 0.5× TBE (1× TBE is Tris–HCl, pH 8.0, 89 mM boric acid and 2 mM EDTA). After autoradiography the region of the gel above the free probe was excised and the DNA eluted electrophorectically and concentrated by ethanol precipitation. The eluted DNA was then amplified by PCR in a reaction containing 10 ng end-labeled reverse primer, 10 ng forward primer with sequence 5′-ACAAGCTTACTGCAGATGC-3′, 25 mM Tris–HCl, pH 8.3, 50 mM KCl, 2 mM $MgCl_2$, 1 mM DTT and 10 U Taq DNA polymerase for 30 PCR cycles of 96°C for 5 min, 96°C for 1 min, 50°C for 1 min and 72°C for 1 min. Two hundred nanograms of purified His–GKLFZn were then used in a gel shift reaction with the newly created probe as described above. The same procedure was repeated for four additional rounds, at which time shifted oligonucleotides were digested with *Pst*I and *Bam*HI and subcloned into pBluescript digested with the same enzymes. After transformation of competent host cells DNA from individual clones was isolated and the insert within each clone sequenced.

Electrophoretic mobility shift assay (EMSA)

EMSA of synthetic oligonucleotides was performed in BB using 1 pmol end-labeled double-stranded oligonucleotide and 100 ng purified His–GKLFZn per reaction. Control reactions contained 100 ng bovine serum albumin. In reactions that included unlabeled oligonucleotides as competitors a 1- to 20-fold molar excess of unlabeled DNA was added to the protein in BB at 4°C for 10 min before addition of the probe. In reactions in which antiserum or preimmune serum was used the serum was incubated with the protein in BB at 4°C for 10 min before addition of the probe.

Transfection

Transient transfection of COS-1 cells with the eukaryotic expression vector containing the full-length GKLF, PMT3-GKLF, was performed by the lipofection method as previously described (11). Twenty four to 48 h following transfection whole cell extracts were prepared by resuspending centrifuged cell pellets in BB containing 0.05% NP40 and sonincating the cell suspensions at 20% intensity for 20 s in a 550 Sonic Dismembrator. After a brief centrifugation glycerol was added to 10% to the supernatant and the extracts were stored at –70°C. Twenty micrograms of crude extracts were used for each EMSA reaction.

Reporter assay

A lucifierase reporter plasmid driven by a minimal TATA box, pGL2-TATA-Luc, was constructed by ligating the adenovirus E_{1b} TATA box derived from plasmid E_{1b}-CAT (17) into the *Xho*I and *Bgl*II sites of the pGL2-Basic vector (Promega, Madison, WI). Two tandem copies of either the wild-type GKLF binding sequence as determined by TDA or a mutated sequence which failed to bind GKLF (M6; Fig. 3) were subcloned into pGL2-TATA-Luc, giving rise to TDA(WT)×2-pGL2-TATA-Luc or TDA(M6-Mut)×2-pGL2-TATA-Luc. Co-transfection experiments were performed with 5 µg/10 cm dish each of the luciferase reporter and PMT3 expression constructs containing GKLF, together with 1 µg/dish internal standard pCMV-SPORT-β-galactosidase (Life Technologies, Gaithersburg, MD). In addition to full-length GKLF, two truncated GKLF constructs in expression vector PMT3 were used in the experiments. They included PMT3-GKLF(1–401), which had all three zinc fingers deleted, and PMT3-GKLF(350–483), which contained the zinc finger region only (15). Lucifierase activity was determined as recommended by Promega (Madison, WI). Transfected cells were lysed with Cell Culture Lysis Reagent (Promgega) and the lysates cleared by microcentrifugation. Assays were performed on cell lysates using Luciferase Assay Substrate and Luciferase Assay Buffer (Promega) in a Monolight 2010 luminometer (Analytical Luminescence Laboratory, San Diego, CA). β-Galactosidase

activity was determined by chemiluminescent assay (18) using Lumi-Gal 530 (Lumigen Inc., Southfield, MI). All luciferase activities were standardized against β-galactosidase activities in transfected cells.

RESULTS

TDA of GKLF binding sequence

To empirically determine the DNA sequence to which GKLF binds we performed TDA on a library of oligonucleotides with randomized sequences at 14 nt positions. Multiple rounds of EMSA using highly purified recombinant protein containing the zinc finger portion of GKLF and amplification by PCR of the oligonucleotides shifted by the recombinant protein were performed to enrich for GKLF recognition sequences. After a total of five rounds the shifted oligonucleotides were subcloned into the multiple cloning sites of the pBluescript plasmid and the sequence of the 14 nt insert in individual clones was determined. Inserted sequences from 80 clones were analyzed and the ability of each insert to bind GKLF was verified by EMSA. Figure 1 shows the inserted sequences from 49 clones, each of which was able to bind GKLF with high affinity. With the exception of two invariable guanine residues at the 3′-most location of 46 of the 49 oligonucleotide inserts, the sequences selected by GKLF were fairly relaxed, although adenine and guanine residues were evidently preferred over thymidine and cytosine residues. It would appear that the core binding sequence for GKLF may include at least an additional 2 nt, CG, situated at the beginning of the 3′-flanking sequence (the shaded area on the right hand side of Fig. 1). This assertion was substantiated by the observation that the same invariable sequence 5′-GGCG-3′ was also present independent of the 5′- and 3′-flanking sequences in three individual clones (#30, #32 and #54; Fig. 1). Shown at the bottom of Figure 1 is the relative frequency of appearance of each nucleotide in the 14 positions of the 46 clones and a compiled sequence of the 14 nt.

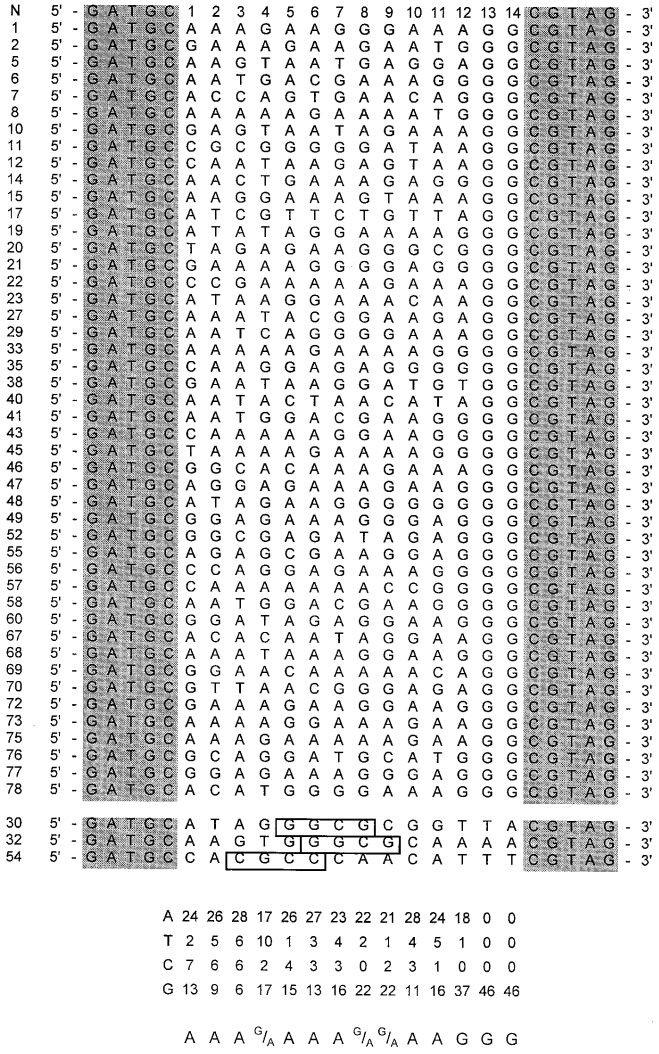

Figure 1. Compilation of GKLF binding sequences. TDA of GKLF binding sequences was performed as described in Materials and Methods. After five rounds of selection shifted oligonucleotides were subcloned into pBluescript and their sequences determined. Shown are 49 clones that gave rise to high affinity binding to GKLF (high affinity binding was defined as ≥50% conversion of radioactivity from the free probe to the shifted complex). The shaded boxes contain the 5′- and 3′-flanking sequences present in the oligonucleotide library. The 'core' binding sequence is present toward the 3′-end of the inserts in a majority of clones, although in three clones it is present in the center of the insert (identified by a box in clones 30, 32 and 54). The table below the figure summarizes the frequency with which each nucleotide is represented in the 14 positions in the 46 clones. The bottom sequence represents the compiled most favored sequence in the 14 nucleotide positions.

Specificity of binding of GKLF to the selected sequences

We next examined the specificity of the selected sequence in binding to GKLF by performing competition experiments. As a probe we selected the insert sequence from clone 47 (5′-AGGA-GAAAGAAGGG-3′), which represented a high affinity binding site for GKLF. A double-stranded oligonucleotide containing this sequence, dubbed the TDA sequence, was synthesized, labeled to high specific activity and analyzed by EMSA. One picomole of probe and 100 ng recombinant GKLF were used in each reaction, which also contained variable amounts of unlabeled specific or non-specific competing DNA. As shown in Figure 2, incubation of GKLF with the probe alone without any competitors resulted in formation of a single DNA–protein complex (C, lanes 1 and 5). Addition of increasing amounts of unlabeled probe (WT, Fig. 1) resulted in a gradual diminution of complex formation (lanes 2–4). In contrast, the addition of an unlabeled non-specific oligonucleotide, poly(dI·dC), failed to compete for formation of the DNA–protein complex (lanes 6–8). The specificity of binding was evident from the fact that at a 5-fold molar excess the wild-type oligonucleotide was able to compete for the majority of binding (lane 4), yet at a molar ratio as high as 200-fold (lane 8) poly(dI·dC) failed to compete for binding entirely. These results indicate that the sequence selected by TDA binds to GKLF in a highly specific manner.

Identifying the minimal essential binding sequence for GKLF

Although each insert of the 49 clones shown in Figure 1 was capable of binding to GKLF with high affinity, it was unclear whether each of the 14 nt selected was necessary for binding. To determine the minimal essential nucleotide sequence required for binding a series of mutant oligonucleotides, each with base substitution at two positions, were synthesized and analyzed by

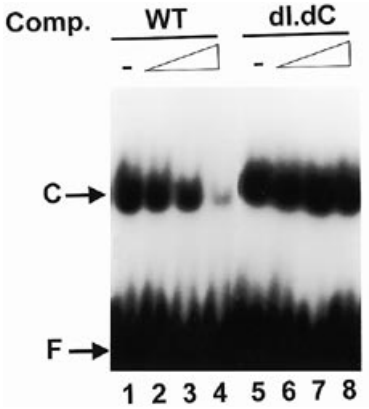

Figure 2. EMSA of GKLF binding sequence. An oligonucleotide, called the TDA oligonucleotide, that contained the insert and the flanking sequence of clone 47 (Fig. 1; see also Fig. 3 for the exact sequence of the oligonucleotide) was used as probe. One picomole of labeled probe and 100 ng recombinant GKLF were used in each reaction. Lanes 1 and 5 contained no competitors; lanes 2–4 contained a 1-, 2- and 5-fold molar excess of the unlabeled wild-type (WT) oligonucleotide respectively; lanes 6–8 contained 10, 100 and 1000 ng poly(dI·dC) (Sigma, St Louis, MO) respectively. With an average molecular weight of 5000 these quantities of poly(dI·dC) represent an ~2-, 20- and 200-fold molar excess of the labeled probe. C indicates complex and F indicates free probe.

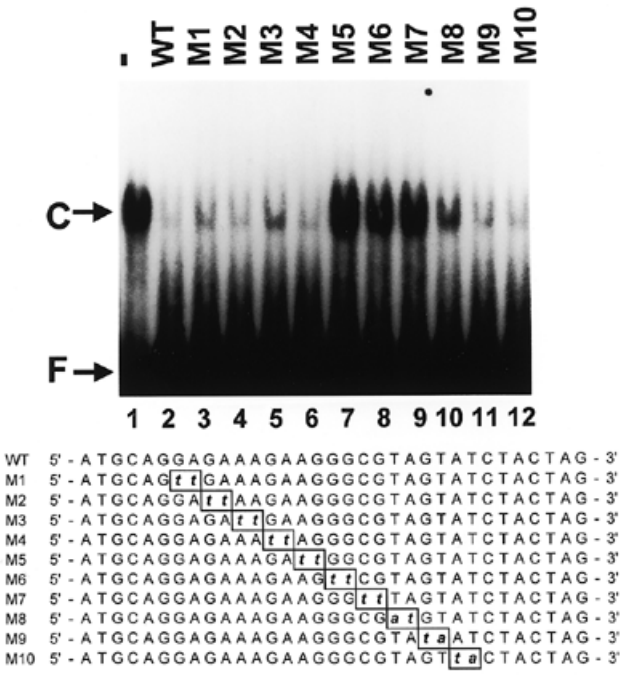

Figure 3. Mutational analysis of GKLF binding sequence. EMSA was performed with 1 pmol labeled GKLF TDA oligonucleotide (WT) and 100 ng recombinant GKLF. Unlabeled competitor oligonucleotides were added in 10-fold molar excess over the probe. The boxed lower case italic letters indicate the mutated sequences in each mutant oligonucleotide. C is DNA–protein complex and F is free probe.

competition experiments. In all, 10 mutant oligonucleotides were obtained, which extended from the beginning of the 14 nt insert to eight bases into the 3′-flanking sequence. Figure 3 shows the result of one such competition experiment. When present the amount of competitor DNA was in 10-fold molar excess over the probe. As seen, the unlabeled wild-type oligonucleotide (lane 2) competed efficiently for formation of the DNA–protein complex, as were mutants M1–M4, M9 and M10, suggesting that the mutations within these oligonucleotides were not essential for binding. In contrast, mutants M5–M7 failed to compete (lanes 7–9). Mutant M8 (lane 10) competed to some extent, although not as efficiently as the wild-type sequence. This partial competition by the M8 mutant oligonucleotide has been repeatedly observed. The results from Figure 3 indicate that the sequence altered in M5–M7 and in part in M8, i.e. 5′-AGGGCGTA-3′, contains the minimal essential binding site for GKLF.

As four of the eight minimal essential nucleotides for GKLF were derived from the 3′-flanking sequence, we sought to further clarify whether each of these four bases was absolutely required for binding. A series of mutant oligonucleotides with single base substitutions in the four 3′-flanking positions were synthesized and analyzed by competition experiments. The M7 series of mutants involved changes in the 5′-CG dinucleotide and the M8 series of mutants involved changes in the 3′-TA dinucleotide. Figure 4 shows the result of one such competition experiment.

As in Figure 3, the unlabeled wild-type oligonucleotide competed effectively for formation of the complex (WT, Fig. 4, lane 2). Again, the M7 mutant oligonucleotide failed to compete (M7, Fig. 4, lane 3). A mutation changing the C residue in the CG dinucleotide to a T residue was able to partially compete (M7A1, Fig. 4, lane 4), but a C→G change abolished binding (M7A2, Fig. 4, lane 5). Similarly, a mutation changing the G residue in the CG dinucleotide to a T residue also adversely affected binding (M7B, Fig. 4, lane 6). These results indicate that the CG dinucleotide is essential for binding, although a T residue is permitted in place of the C residue.

Figure 4, lanes 7–12, shows the potential participation of the TA flanking residues in binding to GKLF. Although a double mutation at this position to AT affected binding to an appreciable extent (M8, Fig. 4, lane 9), changing TA to CA (M8A2, lane 11) or TT (M8B, lane 12) had no effect on binding. In contrast, changing TA to AA seemed to decrease the binding activity somewhat (M8A1, lane 10), but not to the extent seen with the M8 mutant. These results suggest that a T or C residue is preferred over an A residue at the first position of the TA dinucleotide and that the last A residue in the dinucleotide is probably not critical for binding.

Binding of GKLF to previously established DNA sequences that interact with known transcription factors

The minimal essential binding sequence for GKLF established by the mutagenesis experiments is similar to a number of previously established DNA sequences which interact with known transcription factors. One example is BTE (basal transcription element; 19), which is present in the promoters of and is essential for basal transcription of a number of genes encoding the superfamily of cytochrome P450 (CYP) enzymes, including *CYP1A1* (19,20). A second example is the CACCC sequence within the promoter of the β-globin gene (21) with which EKLF interacts (10). To determine whether GKLF binds to these two sites competition experiments were performed using synthetic oligonucleotides containing published sequences. As shown in

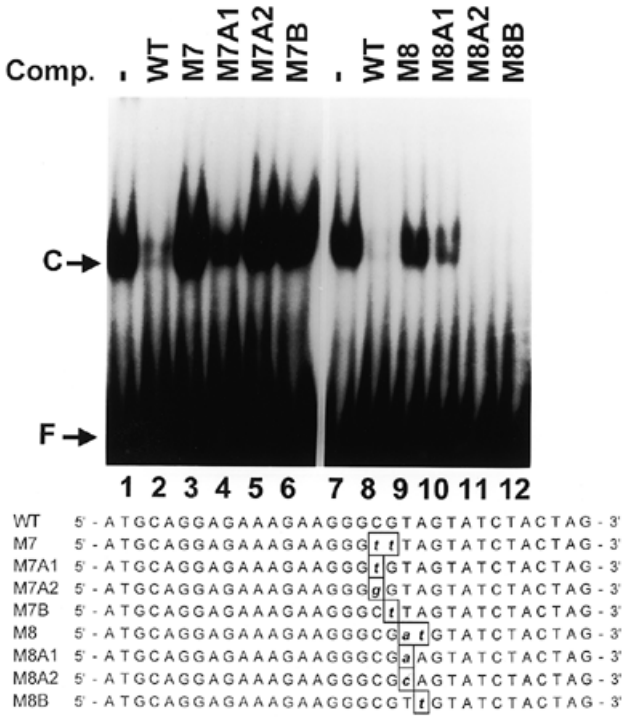

Figure 4. Additional mutational analysis of GKLF binding sequence. EMSA was performed as in Figure 3. Unlabeled oligonuceotides were added in 10-fold molar excess over the probe. The boxed lower case italic letters indicate the mutated sequences in each mutant oligonucleotide. C is complex and F is free probe.

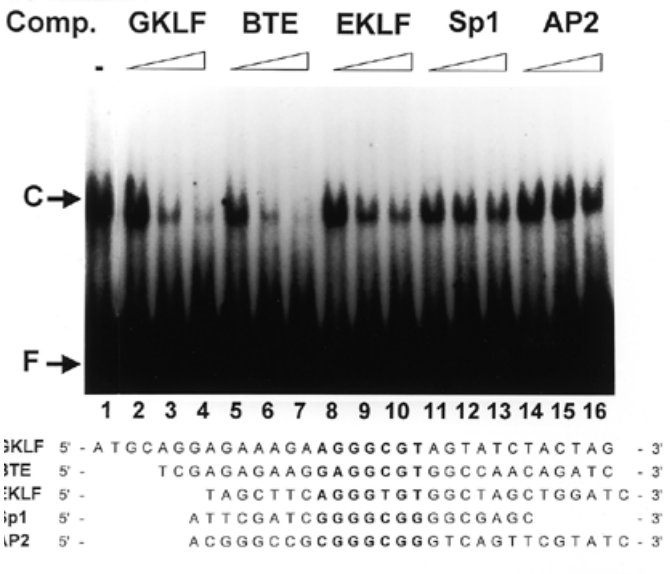

Figure 5. Competition experiments using established transcription factor binding sites. EMSA was performed as in Figures 3 and 4 using 1 pmol labeled GKLF TDA oligonucleotide and 100 ng recombinant GKLF. With the exception of lane 1, all lanes contained unlabeled competitor oligonucleotides. For each established binding sequence increasing amounts of competitor DNA, in the order 1-, 5- and 10-fold molar excess over the probe, were added to the reaction. The sequence shown for EKLF contains an inverted CACCC box sequence. The oligonucleotides containing the Sp1 and AP2 binding sites were purchased from Promega. The minimal essential binding sequence for GKLF and the corresponding sequences in the other transcription factor binding sites are in bold.

Figure 5, BTE (lanes 5–7) competed as efficiently for binding as the wild-type GKLF TDA sequence (lanes 2–4). Although to a somewhat lesser degree, the CACCC sequence (EKLF, lanes 8–10) also competed. The slightly lower efficiency of binding of the CACCC sequence to GKLF is consistent with the result in Figure 4, when the M7A1 mutant oligonucleotide, which also contains an inverted CACCC sequence, was used as competitor (Fig. 4, lane 4). In contrast, a synthetic oligonucleotide containing the binding site for Sp1 (lanes 11–13) or AP2 (lanes 14–16) competed poorly for binding. Of note is that a major difference in the binding sites for GKLF/BTE/EKLF and for Sp1/AP2 is the presence of a T residue in the 3′-most position in the minimal essential binding sequence (bold, Fig. 5) in the former group. These results therefore establish that this T is an important residue for binding to GKLF. Lastly, the first two nucleotides in the minimal essential binding sequence of GKLF (AG) are interchangeable with GA, as demonstrated by the equal binding affinity of GKLF to the TDA sequence and to BTE (Fig. 5).

Full-length GKLF also binds to the sequence selected by TDA

The TDA used to select the GKLF binding sequence involved a truncated form of GKLF, a portion that contained only the zinc fingers. To determine whether full-length GKLF also binds to the same sequence extracts were prepared from COS-1 cells transfected with an expression vector containing full-length GKLF (PMT3-GKLF) and analyzed by EMSA using the GKLF TDA oligonucleotide as probe. As a control extracts were prepared from similarly transfected cells with vector (PMT3) alone. Figure 6 shows that a DNA–protein complex was formed between the probe and extracts from PMT3-GKLF-transfected cells (lane 1). This complex was completely disrupted when the reaction was performed in the presence of an anti-GKLF serum (lane 3), but not in the presence of a preimmune serum (lane 2). In contrast, no complexes were apparent when extracts from PMT3-transfected cells were analyzed (lanes 4–6). These results indicate that full-length GKLF is capable of binding to the selected GKLF binding sequence and that anti-GKLF serum is able to interact with the formed complex and disrupt its formation.

GKLF transactivates a reporter gene driven by the TDA-selected sequence

To demonstrate that the sequence selected by TDA is important for GKLF-mediated transcriptional activity co-transfection experiments were performed in CHO cells using an expression vector containing full-length or truncated GKLF and a reporter gene driven by two tandem copies of the wild-type or a mutated site that no longer binds GKLF (M6, Fig. 3). As shown in Figure 7, full-length GKLF was able to activate the reporter gene driven by the wild-type GKLF binding site (shaded bar, lane B) but not by the mutated GKLF binding site (open bar, lane B). In contrast, neither empty vector (lane A) nor a truncated GKLF construct that lacked either the zinc fingers (lane C) or the region outside the zinc fingers which contains the putative transactivation domain (lane D) of GKLF was able to activate the reporter gene driven by either the wild-type (shaded bars) or mutated (open bars) sequence. These results indicate that full-length GKLF alone is capable of activating transcription mediated by the selected binding sequence.

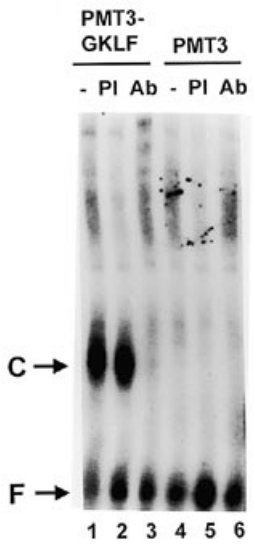

Figure 6. EMSA of GKLF TDA sequence with full-length GKLF. Extracts were prepared from transiently transfected COS-1 cells with PMT3-GKLF or PMT3 as described in Materials and Methods. Twenty micrograms were used for each reaction, which also contained 1 pmol labeled probe. In reactions containing serum 1 µl IgG-enriched polyclonal anti-GKLF serum (11) or preimmune serum was added 10 min prior to addition of probe. Lanes 1 and 4 contained no added serum.

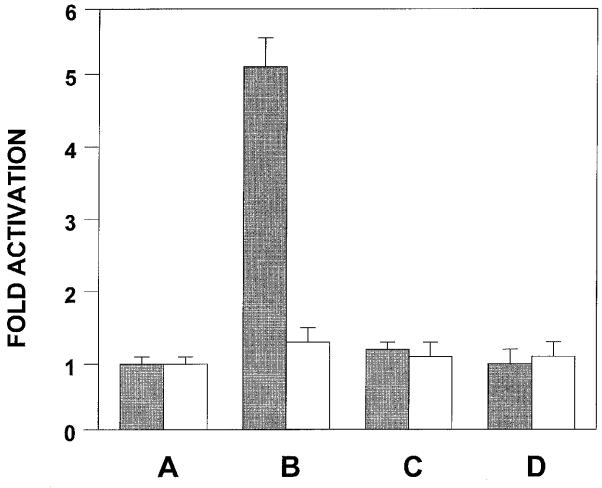

Figure 7. GKLF transactivates a reporter gene driven by the wild-type GKLF binding sequence. CHO cells were co-transfected with a PMT3 construct that contained no insert (PMT3; lane A), full-length GKLF (PMT3-GKLF; lane B), a truncated GKLF that lacked the zinc fingers [PMT3-GKLF(1–401); lane C] or a truncated GKLF that lacked the putative transactivation domain [PMT3-GKLF(350–483); lane D] and a luciferase reporter gene driven by two copies in tandem of either the wild-type (shaded bars) or a mutated (M6; open bars) GKLF binding site in conjunction with a TATA box. Shown are the means of fold activation of the reporter activity from four independent experiments in the presence and absence of each expression construct after normalizing for activity of a co-transfected internal control, pCMV-SPROT-β-galactosidase. Lines on top of the bars represent standard errors.

DISCUSSION

Within the last decade molecular techniques such as TDA (16) and other closely related methods, including SAAB (selected and amplified binding sites; 22) and CASTing (cyclic amplification and selection of targets; 23), have proven to be potent tools in identifying the binding sequences for transcription factors. In particular, a considerable number of publications have utilized this principal of target site selection and established the putative binding sites for a diverse group of zinc finger-containing transcription factors (16,21,24–27). Similarly, in the present study we were able to establish a binding site for GKLF using highly purified recombinant GKLF that contained the three zinc fingers. The identified binding sequence should facilitate characterization of the biological functions of GKLF, including the target genes that it regulates.

The TDA and subsequent mutational analyses (Figs 3 and 4) helped established a minimal essential binding sequence (5′-$^{G}/_{A}{}^{G}/_{A}$GG$^{C}/_{T}$G$^{C}/_{T}$-3′) to which GKLF exhibited high affinity binding. This sequence not only bound to recombinant GKLF that contained the zinc finger portion only of GKLF but to full-length GKLF (Fig. 6). Moreover, only the full-length and not truncated forms of GKLF was able to activate a reporter gene driven by the selected GKLF binding sequence (Fig. 7). These results indicate that the selected sequence is able to mediate binding and transactivation by the full-length protein.

The minimal essential sequence selected by TDA is very similar to the previously predicted binding sequence for GKLF based on conservation of amino acid sequences in the zinc finger region between GKLF and several known Krüppel-like transcription factors, including BTEB2, EKLF and Sp1 (11). The ability of GKLF to bind to the previously identified binding sites for BTEB2 and EKLF was substantiated by the competition experiment, as illustrated in Figure 5. However, GKLF appeared to bind to an established Sp1 binding site much more poorly, despite the highly similar nature of the binding sites for these two transcription factors (Fig. 5). We attributed this difference to the single T residue at the 3′-most end of the GKLF binding site, which was substituted with a G residue in the Sp1 binding site. A conversion of this T residue to an A residue also reduced binding affinity, although a change to a C residue was much better tolerated (Fig. 4). These results indicate that GKLF binds to only a subset of previously established ‘GC box’-containing sequences (28,29), due to its ability to discriminate single base changes.

The ability of GKLF to bind to the BTEB2 and EKLF binding sites but not to the Sp1 binding site is of particular interest. It has previously been shown that Sp1 is capable of binding to both BTEB2 (20) and EKLF (21) binding sequences. The failure of GKLF to bind to a Sp1 binding site then suggests that GKLF is more restrictive in terms of its binding requirement. This result is reminiscent of a previous observation that EKLF recognizes its cognate CACCC binding sequence much better than a Sp1 binding site (21). These findings further substantiate the close relationship between GKLF and EKLF, as previously established by sequence homology in the zinc finger region (11) and in the nuclear localization signal (15).

The ability of GKLF to interact with BTE (Fig. 5) raises another interesting facet of the present study. BTE was initially identified as a *cis* sequence element that is required for basal promoter activity of the rat cytochrome P-450c gene, also known as *CYP1A1* (19). Since its identification, BTE has been shown to interact with a multitude of transcription factors, including Sp1

(20), BTEB (20) and BTEB2 (14). It now appears that GKLF is another transcription factor that recognizes this sequence with high affinity. Recent studies from our laboratory suggest that the base contacts between GKLF and BTE (Zhang and Yang, unpublished observations) are similar to those between BTEB and BTE as well as between Sp1 and BTE (30). Of note is that BTE is a highly conserved sequence element and is present in several other cytochrome P-450 genes, including P-450b and P-450e (31). Of further interest is that these cytochrome P-450 genes are highly expressed in the epithelial cells of the intestinal tract (32–34) with a very similar distribution to GKLF (11). It is therefore a formal possibility that GKLF is a major regulator of this family of genes in the intestinal epithelium. Studies to understand the contribution of GKLF to expression of the cytochrome P-450 family of genes are currently in progress in our laboratory.

ACKNOWLEDGEMENTS

This work was in part supported by grants from the National Institutes of Health (to V.W.Y.). J.M.S. was supported by a National Research Service Award from the National Institutes of Health.

REFERENCES

1 Klug,A. and Schwabe,J.W.R. (1995) *FASEB J.*, **9**, 597–604.
2 Berg,J.M. and Shi,Y. (1996) *Science*, **271**, 1081–1085.
3 Pelham,H.R.B. and Brown,D.D. (1980) *Proc. Natl. Acad. Sci. USA*, **77**, 4170–4174.
4 Hoovers,J.M.N., Mannens,M., John,R., Bliek,J., van Heyningen,V., Porteous,D.J., Leschot,N.J., Westerveld,A. and Little,P.F.R. (1992) *Genomics*, **12**, 254–263.
5 Berg,J.M. (1988) *Proc. Natl. Acad. Sci. USA*, **85**, 99–102.
6 Schuh,R., Aicher,W., Gual,U., Cote,S., Preiss,A., Maier,D., Seifert,E., Nauber,U., Shroder,C., Kemler,R. and Jackle,H. (1986) *Cell*, **47**, 1025–1032.
7 Kadonaga,J.T., Carner,K.R., Masiarz,F.R. and Tjian,R. (1987) *Cell*, **51**, 1079–1090.
8 Christy,B.A., Lau,L.F. and Nathans,D. (1988) *Proc. Natl. Acad. Sci. USA*, **85**, 7857–7861.
9 Call,K.M., Glaser,T., Ito,C.Y., Buckler,A.J., Pelletier,J., Haber,D.A., Rose,E.A., Kral,A., Yeger,H., Lewis,W.H., Jones,C. and Housman,D.E. (1990) *Cell*, **60**, 509–520.
10 Miller,I.J. and Bieker,J.J. (1993) *Mol. Cell. Biol.*, **13**, 2776–2786.
11 Shields,J.M., Christy,R.J. and Yang,V.W. (1996) *J. Biol. Chem.*, **271**, 20009–20017.
12 Garrett-Sinha,L.A., Eberspaecher,H., Seldin,M.F. and de Crombrugghe,B. (1996) *J. Biol. Chem.*, **271**, 31384–31390.
13 Anderson,K.P., Kern,C.B., Crable,S.C. and Lingrel,J.B. (1995) *Mol. Cell. Biol.*, **15**, 5957–5965.
14 Sogawa,K., Imataka,H., Yamasaki,Y., Kusume,H., Abe,H. and Fujii-Kuriyama,Y. (1993) *Nucleic Acids Res.*, **21**, 1527–1532.
15 Shields,J.M. and Yang,V.W. (1997) *J. Biol. Chem.*, **272**, 18504–18507.
16 Thiesen,H.-J. and Bach,C. (1990) *Nucleic Acids Res.*, **18**, 3203–3209.
17 Potter,J.J., Cheneval,D., Dang,C.V., Resar,L.M.S., Mezey,E. and Yang,V.W. (1991) *J. Biol. Chem.*, **266**, 15457–15463.
18 Beale,E.G., Deeb,E.A., Handley,R.S., Akhavan-Tafti,H. and Schaap,A.P. (1992) *BioTechniques*, **12**, 320–323.
19 Yanagida,A., Sogawa,K., Yasumoto,K.-I. and Fujii-Kuriyama,Y. (1990) *Mol. Cell. Biol.*, **10**, 1470–1475.
20 Imataka,H., Sogawa,K., Yasumoto,K.-I., Kikuchi,Y., Sasano,K., Lobayashi,A., Hayami,M. and Fujii-Kuriyama,Y. (1992) *EMBO J.*, **11**, 3663–3671.
21 Hartzog,G.A. and Myers,R.M. (1993) *Mol. Cell. Biol.*, **13**, 44–56.
22 Blackwell,T.K. and Weintraub,H. (1990) *Science*, **250**, 1104–1110.
23 Wright,W.E., Binder,M. and Funk,W. (1991) *Mol. Cell. Biol.*, **11**, 4104–4110.
24 Gogos,J.A., Hsu,T., Bolton,J. and Kafatos,F.C. (1992) *Science*, **257**, 1951–1954.
25 Morris,J.F., Hromas,R. and Rauscher,F.J. (1994) *Mol. Cell. Biol.*, **14**, 1786–1795.
26 Swirnoff,A.H. and Milbrandt,J. (1995) *Mol. Cell. Biol.*, **15**, 2275–2287.
27 Jamieson,A.C., Wang,H. and Kim,S.-H. (1996) *Proc. Natl. Acad. Sci. USA*, **93**, 12834–12839.
28 Dynan,W.S. and Tjian,R. (1985) *Nature*, **316**, 774–778.
29 Mitchell,P.J. and Tjian,R. (1989) *Science*, **245**, 371–378.
30 Sogawa,K., Kikuchi,Y., Imataka,H. and Fujii-Kuriyama,Y. (1993) *J. Biochem.*, **114**, 605–609.
31 Fujii-Kuriyama,Y., Imataka,H., Sogawa,K., Yasumoto,K.-I. and Kikuchi,Y. (1992) *FASEB J.*, **6**, 706–710.
32 Traber,P.G., Chianale,J., Florence,R., Kim,K., Wojcik,E. and Gumucio,J.J. (1988) *J. Biol. Chem.*, **263**, 9449–9455.
33 Traber,P.G., Wang,W. and Yu,L. (1992) *Am. J. Physiol.*, **263**, G215–G223.
34 Traber,P.G., McDonnell,W.M., Wang,W. and Florence,R. (1992) *Biochim. Biophys. Acta*, **1171**, 167–175.