0026-895X/07/7106-1630–1639$20.00
MOLECULAR PHARMACOLOGY
Copyright © 2007 The American Society for Pharmacology and Experimental Therapeutics
Mol Pharmacol 71:1630–1639, 2007

Vol. 71, No. 6
31302/3209005
Printed in U.S.A.

Identification and Functional Analysis of a Novel Human *CYP2E1* Far Upstream Enhancer

Jeff D. Shadley, Karthika Divakaran, Kimber Munson, Ronald N. Hines, Kirk Douglas,[1] and D. Gail McCarver

Clinical Pharmacology, Pharmacogenetics & Teratology, Department of Pediatrics, Medical College of Wisconsin, Milwaukee, Wisconsin

Received October 6, 2006; accepted March 9, 2007

ABSTRACT

Both transcriptional and post-transcriptional *CYP2E1* regulatory mechanisms are known, resulting in 20-fold or greater variation in CYP2E1 expression. To evaluate functional regulatory elements controlling transcription, *CYP2E1* promoter constructs were used to make adenovirus vectors containing *CYP2E1* promoter-driven luciferase reporters for analyses in both primary human hepatocytes and HepG2 cells. A 1.2-kilobase pair portion of the *CYP2E1* promoter was associated with 5- to 10-fold greater luciferase activity. This upstream region contained five direct repeats of 59 base pairs (bp) that increased thymidine kinase-driven luciferase reporter activity in HepG2 cells more than 5-fold, regardless of orientation. Electrophoretic mobility shift assays (EMSAs) identified sequence-specific nuclear protein binding to the 59-bp repeats that was dependent on a 17-bp sequence containing a canonical GATA binding site (WGATAR). Competitive and supershift EMSA identified the participation of GATA4, another GATA family member or GATA-like factor, and a third factor unrelated to the GATA family. Involvement of the tricho-rhino-phalangeal syndrome-1 factor, which also binds a GATA sequence, was eliminated. Rather, competitive EMSA using known binding sequences for the orphan nuclear receptors, steroidogenic factor-1 (or NR5A1), and fetoprotein transcription factor (or NR5A2) implicated an NR5A member in binding a sequence overlapping the canonical GATA. Chromatin immunoprecipitation assay demonstrated in vivo binding of NR5A2 to the enhancer sequence in human hepatocytes. The enhancer sequence is conserved within the human population but seems species-specific. The identification of this novel enhancer and its putative mechanism adds to the complexities of human *CYP2E1* regulation.

CYP2E1 is highly expressed in liver, in which it metabolizes a large number of small molecular weight endogenous or exogenous compounds (Lieber, 1999). CYP2E1 converts some chemicals to a more toxic form; for example, it is the major enzyme responsible for the oxidation of acetaminophen to the reactive quinone, NAPQI. The more than 20-fold interindividual variation in human CYP2E1 activity has been attributed to diverse regulatory mechanisms acting at both transcriptional and post-transcriptional levels (Lieber, 1999). *CYP2E1* genetic variants have been found within the structural gene and upstream sequences; however, the majority have not demonstrated an in vivo functional impact (Carrière et al., 1996). An exception is a polymorphism in the *CYP2E1* promoter *CYP2E1*1D*, which has been associated with increased in vivo metabolic activity with ethanol intake and obesity (McCarver et al., 1998). In part because of the emphasis on post-transcriptional regulation, knowledge regarding mechanisms responsible for the control of *CYP2E1* transcription has lagged. We report here the use of adenovirus representing the *CYP2E1*1D* allele and the reference sequence *CYP2E1*1C* to identify a novel enhancer in the far *CYP2E1* upstream region (*CYP2E1*1C* position −3690 to −3386; *CYP2E1*1D* position −3786 to −3482) that has a significant impact on constitutive promoter activity. The dis-

This investigation was supported by Public Health Service grant AA11636 and support from the Children's Research Institute, Children's Hospital of Wisconsin.

Portions of this work were presented at the Federation of American Societies Experimental Biology annual meeting, Washington D.C. (2004), and at the 27th annual meeting of the Research Society on Alcoholism, Vancouver, BC, Canada (2004).

[1] Current affiliation: Department of Internal Medicine, Wayne State University, Detroit, Michigan.

Article, publication date, and citation information can be found at http://molpharm.aspetjournals.org.

doi:10.1124/mol.106.031302.

ABBREVIATIONS: FBS, fetal bovine serum; ChIP, chromatin immunoprecipitation; C_T, threshold cycle; EMSA, electrophoretic mobility shift assay; FTF, fetoprotein transcription factor; HNF, hepatocyte nuclear factor; NR5A, nuclear receptor family 5 group A; PSA, prostate-specific antigen; *Ptk*, promoter of thymidine kinase gene; RT-PCR, real-time polymerase chain reaction; SF1, steroidogenic factor 1; TRPS1, tricho-rhino-phalangeal type I protein; bp, base pair; PCR, polymerase chain reaction; MOI, multiplicity of infection; PBS, phosphate-buffered saline; IP, immunoprecipitate; LRH, liver receptor homolog; ANOVA, analysis of variance.

covery of the distal enhancer adds to the complexities of *CYP2E1* regulation and provides an opportunity to better understand the major factors responsible for the wide variation in *CYP2E1* expression.

Materials and Methods

Cells and Cell Culture. Normal human hepatocytes plated on a collagen matrix in 24-well plates were obtained from the Liver Tissue Procurement and Distribution System (Dr. Steve Strom, University of Pittsburgh, Pittsburgh, PA). Upon receipt, the medium was aspirated and replaced with 0.5 ml/well fresh Williams' E medium (Sigma, St. Louis, MO), pH 7.3, supplemented with 2.2 g/l sodium bicarbonate, 25 mM HEPES, 10% fetal bovine serum (FBS), 1% antibiotic/antimycotic, 1 μM dexamethasone, 1 μM hydrocortisone, 2.5 μg/ml human transferrin, 20 ng/ml epidermal growth factor, 20 μM niacinamide, and 0.5 mM ornithine. HepG2 human hepatoma cells (Aden et al., 1979) were obtained from Dr. Barbara Knowles (The Jackson Laboratory, Bar Harbor, ME) and were cultured in Williams' E medium, pH 7.3, supplemented with 2.2 g/l sodium bicarbonate, 25 mM HEPES, and 10% FBS. Human embryonic kidney 293A cells (Quantum Biogene, Montreal, QC, Canada) were cultured in Dulbecco's modified Eagle's medium, pH 7.3, supplemented with 3.7 g/l sodium bicarbonate, 25 mM HEPES, and 5% FBS. All cultures were incubated at 37°C in a 5% CO_2 humidified atmosphere.

Amplification and Cloning of the *CYP2E1* Promoter. Human genomic DNA from a person having the *CYP2E1*1C/*1D* genotype was used as a template for nested polymerase chain reaction (PCR) amplification of the *CYP2E1* promoter using the Advantage Genomic Polymerase Kit (Clontech, Mountain View, CA). The outer primer pair, 5′-AGA GCC ATA CCT GCA CAC-3′ (*CYP2E1*1C* position −3761 to −3744; *CYP2E1*1D* position −3857 to −3840) and 5′-GCT CCA GGA TGC TAT CAA-3′ (*CYP2E1* position +353 to +336) produced amplicons of 4114 and 4214 bp, respectively. These fragments were used as templates in reactions with the inner primer pair, 5′-CG*A CGC GT*T CCT GGA AGC AGC AAG AGT G-3′ (*CYP2E1*1C* position −3710 to −3691; *CYP2E1*1D* position −3806 to −3787), and 5′-CAA CTG GAA GAG GTT CCC GAT GAT-3′ (*CYP2E1* position +179 to +156). The upstream inner primer incorporated an MluI recognition site (italicized bases in above sequence) at the 5′-end for cloning purposes. The gel purified amplicons were digested with BglI to remove the *CYP2E1* translation start site, and flush ends were created by treatment with T4 DNA polymerase and finally digested with MluI. The resulting fragments were cloned into MluI/SmaI-digested pGL3Basic to produce the constructs pDGM20 and pDGM21 containing *CYP2E1*1C* position −3710 to +25 and *CYP2E1*1D* position −3806 to +25, respectively, directing luciferase expression. To construct expression plasmids with the distal *CYP2E1*1C* promoter deleted, an 1161-bp fragment containing *CYP2E1*1C* position −3710 to −2555 and *CYP2E1*1*1D* position −3806 to −2651 was removed from pDGM20 and pDGM21, respectively, by digestion with MluI/Bsu36I. After creating flush ends, the isolated 7359- and 8616-bp MluI/Bsu36I vector fragments were ligated to produce pDGM34 (*CYP2E1*1C* position −2554 to +25) and pDGM35 (*CYP2E1*1D* position −2650 to +25), respectively, directing luciferase expression.

To prepare vector constructs for recombinant adenovirus using the Quantum Biogene pAdEasy System (see *Construction of Recombinant Adenovirus*), the unique PacI sites in pDGM20 and pDGM21 were eliminated by digestion with PacI followed by treatment with T4 DNA polymerase to create flush ends. The religated plasmids pDGM22 and pDGM23 were digested with NotI/SalI, resulting in 5877- and 5973-bp fragments, respectively, that carried the upstream terminator, *CYP2E1*1C* position −3710 to +25 or *CYP2E1*1D* position −3806 to +25, respectively, the firefly luciferase gene, and the SV40 polyadenylation recognition sequence. These fragments were cloned into the NotI/SalI-digested shuttle transfer vector (Quantum Biogene). The corresponding NotI/SalI fragments from pDGM34 and pDGM35 also were cloned into the shuttle transfer vector to make adenovirus lacking the distal region of the *CYP2E1* promoter.

Construction of Recombinant Adenovirus. Recombinant adenoviruses were generated using the pAdEasy vector system (Quantum Biogene). PmeI-linearized shuttle transfer vector carrying the *CYP2E1*1C* or *CYP2E1*1D* promoters and firefly luciferase gene and pAdEasy-1 were used to cotransform recombination competent *Escherichia coli* BJ5183 to transfer the NotI/SalI fragment from the shuttle transfer vector to the pAdEasy-1 adenovirus plasmid. Kanamycin-resistant clones were screened with PacI, and potential recombinant adenovirus plasmids were used to transform *E. coli* DH5α for expansion of plasmid DNA and verification of recombinants by restriction enzyme analysis. Once verified, recombinant adenovirus plasmid was linearized with PacI and transfected into 293A cells (Quantum Biogene) using Lipofectamine Plus (Invitrogen, Carlsbad, CA). The transfected cells were overlaid with 1.25% SeaPlaque agarose (FMC, Rockland, ME). Resulting plaques were picked and the viral particles expanded in 293A cells. The maintenance of the recombinant plasmid was verified by restriction enzyme analysis. Viral titers were determined using the multiplicity of infection (MOI) assay per manufacturer's instructions (Quantum Biogene) and were approximately 10^6 viral particles per microliter of infected cell lysate. Recombinant adenovirus constructs for promoterless luciferase (negative control) were generated from pGL3Basic (Promega, Madison, WI) using the same procedure.

Recombinant Adenovirus Infection. At the time of infection, culture medium was removed from wells and replaced with 0.2 ml of Williams' E medium with 2% FBS containing a volume of viral particles to give an MOI of 10. After 6 h, the volume of medium in the well was increased to 0.5 ml with Williams' E medium. Approximately 20 h after infection, the cells were processed for luciferase activity using the Luciferase Assay System (Promega). A Dynex MLX 96-well plate luminometer was used to measure chemiluminescence expressed as relative luminescence units. Separate infections with adenovirus carrying the CMV promoter-driven β-galactosidase reporter gene (pInfect+; Quantum Biogene) were essentially 100% efficient for infection of both human hepatocytes and HepG2 hepatoma cells as demonstrated by in situ staining with 5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside.

Plasmid Constructs. Plasmids were constructed to test for the ability of the *CYP2E1*1C* distal promoter to act as an enhancer with the heterologous thymidine kinase promoter (*Ptk*). An MluI/StyI fragment from pDGM22 (described above) containing 330 bp of the distal promoter (*CYP2E1*1C* position −3710 to −3386) was isolated, and flush ends were created by treatment with the Klenow fragment of DNA polymerase I (3′ to 5′ exonuclease) (New England Biolabs, Beverly, MA) and inserted into the SmaI site of pBlueScript II KS- to make pDGM58. A 198-bp HindIII/XhoI fragment from pBLcat2 (Luckow and Schutz, 1987) containing the Ptk was inserted into HindIII/XhoI-digested pDGM58. The resulting plasmid, pDGM60, has the *CYP2E1*1C* distal promoter upstream of Ptk. A 594-bp SacI/XhoI fragment from pDGM60 containing the *CYP2E1*1C* distal promoter, and Ptk was inserted into SacI/XhoI-digested pGL3Basic vector, resulting in the pDGM62 reporter construct. Thus, pDGM62 consists of the *CYP2E1*1C* distal promoter (*CYP2E1*1C* position −3690 to −3386) upstream of Ptk and the luciferase gene. Expression plasmid pDGM65 carries the same elements as pDGM62; however, the 330-bp element is in reverse orientation. pDGM61 was made by inserting a 396-bp SacI/HindIII fragment from pDGM58 into the SacI/HindIII-cut pGL3Basic vector. pDGM63 was created by digesting pDGM62 with XbaI, religating the 1884-bp Ptk:luciferase fragment with the 3107-bp vector fragment, and selecting a clone with the desired orientation. Thus, pDGM63 lacks the *CYP2E1*1C* distal promoter sequence of pDGM62.

Transient Transfections. HepG2 cells were plated at 2×10^5 cells/well in 24-well plates coated with poly(L-lysine) (Sigma). Forty-eight hours after plating, the cells in each well were transfected with 1 to 2 μg of total plasmid DNA (test plasmid, 0.8; pCMVβgal, 0.2) using Lipofectamine 2000 (Invitrogen) at 1 to 4 μg/well. Cells were processed for luciferase activity 48 h after transfection. Plasmid DNA for transfection was obtained using the High-Purity Plasmid MIDI Prep System (Marligen BioSciences, Ijamsville, MD).

Electrophoretic Mobility Shift Assay. A 123-bp Hinf I fragment (*CYP2E1*1C* position −3623 to −3501) was labeled using the Klenow fragment of DNA polymerase I (3′ to 5′ exonuclease) (New England Biolabs) with 25 μCi [α-^{32}P]dCTP (3000 mCi/mmol) (PerkinElmer Life and Analytical Sciences, Boston, MA) in a 25-μl final reaction volume containing 10 mM Tris-HCl, pH 7.5, 5 mM $MgCl_2$, 7.5 mM dithiothreitol, and 250 μM each dATP, dGTP, and dTTP at 37°C for 30 min. A 30-bp double-stranded oligonucleotide representing *CYP2E1*1C* position −3548 to −3519 was labeled using T4 DNA polynucleotide kinase (New England Biolabs) with 25 μCi [γ-^{32}P]ATP (3000 mCi/mmol) (PerkinElmer Life and Analytical Sciences) in a final reaction volume of 25 μl containing 70 mM Tris-HCl, pH 7.6, 10 mM $MgCl_2$, and 5 mM dithiothreitol at 37°C for 30 min. Reactions were stopped by adding Na_2EDTA and salmon sperm DNA to final concentrations of 1.0 mM and 0.1 mg/ml, respectively. Unincorporated nucleotides were separated from product using a 1-ml Sephadex G-50 fine grade (GE Healthcare, Chalfont St. Giles, Buckinghamshire, UK) spin column centrifuged at 500*g* for 3 min at room temperature. A 2-μl aliquot of the eluate was applied to a dry scintillant disk (Ready-Cap; Beckman Coulter, Fullerton, CA), and specific activity was determined using a Wallac 1410 scintillation counter (PerkinElmer Life and Analytical Sciences). Nuclear protein extracts were obtained from HepG2 cells using the NE-PER Nuclear and Cytoplasmic Extraction Reagents (Pierce, Rockford, IL) following the manufacturer's instructions. Protein content was measured using the Micro BCA Protein Assay Reagent Kit (Pierce). Aliquots of the nuclear protein extracts were stored at −80°C. Binding reactions (25 μl final volume) between aliquots of nuclear protein and labeled oligonucleotide probes were performed in 10 mM HEPES, pH 7.9, 0.1 mM Na_2 EDTA, 100 mM KCl, 25 μg/ml bovine serum albumin, 5% glycerol, 4 μg of poly[d(I/C)] with protease inhibitors (aprotinin, pepstatin A, antipain, leupeptin at 1 μg/ml each, 250 μg/ml benzamidine, and NaF, $NaMoO_4$, dithiothreitol, and phenylmethylsulfonyl fluoride at 1 mM each). For binding studies, nuclear extract (5 or 10 μg of protein) was incubated with 3 to 4 fmol (30 bp double-stranded oligonucleotide) or 7 to 8 fmol (123-bp Hinf I fragment) for 30 min. For competition studies, unlabeled DNA sequences (25-, 50-, or 100-fold molar excess over probe) were preincubated with nuclear protein extract for 30 min, followed by the addition of probe. Oligonucleotides used in this study were purchased from MWG Biotech, Inc. (High Point, NC). For supershift experiments, antibody (GATA4 sc-1237X; Santa Cruz Biotechnology, Inc., Santa Cruz, CA) was added after the probe/nuclear protein extract incubation time, and the tubes were left on ice for 60 min. DNA/protein complexes were fractionated by electrophoresis using 4% (123-bp Hinf I probe) or 6% (30-bp double-strand oligonucleotide probe) nondenaturing polyacrylamide gels in running buffer (0.25 M Trizma Base, 1.9 M glycine, and 0.01 M Na_2EDTA) at 12°C. Gels were pre-electrophoresed for 30 min at 100 V; sample separation was performed for 1.5 to 2 h at 300 V, 50 mA. After electrophoresis, the gel was transferred to Whatman filter paper, covered with plastic wrap, and dried at 90°C for 45 min using a Bio-Rad 583 gel dryer (Bio-Rad Laboratories, Hercules, CA). Dried gels were exposed to X-ray film overnight at −80°C.

Chromatin Immunoprecipitation. The chromatin immunoprecipitation (ChIP) method described in Hatzis and Talianidis (2001) was used with slight modifications. Plated primary human hepatocytes were treated with 1% formaldehyde (Sigma, St. Louis, MO) in PBS for 10 min at room temperature, followed by the addition of one-tenth volume of 1.25 M glycine (Roche Diagnostics, Indianapolis, IN) for 10 min. The hepatocytes were washed twice with ice-cold PBS, scraped into ice-cold PBS-containing protease inhibitors (1 mM phenylmethylsulfonyl fluoride, 1 μg/ml aprotinin, pepstatin A, and antipain), and transferred to centrifuge tubes on ice. The tubes were centrifuged 1000 RPM for 5 min at 4°C. Subsequently, the cell pellet was resuspended in lysis buffer (50 mM Tris-HCl, pH 6.7, 10% glycerol, 1.72% SDS, and 0.33% each of Nonidet P-40 and sodium deoxycholate) with protease inhibitors. Aliquots (0.5 ml) were sonicated in siliconized microfuge tubes maintained in ice water to prevent overheating. A Misonix 3000 sonicator (Misonix. Inc., Farmingdale, NY) equipped with a microtip was used to deliver a series of 30-s pulses at power level three. After sonication, the tubes were centrifuged at 13,000 RPM for 30 min at 4°C. A 50-μl aliquot of each sonication supernatant was removed for checking sonication efficiency by gel electrophoresis. The remainder of the supernatant was snap-frozen and stored at −80°C. Sonicates containing predominantly DNA fragments between 100 and 650 base pairs were further evaluated. After thawing on ice, protein G PLUS agarose beads (Santa Cruz Biotechnology) were added for 1.5 h. After centrifugation at 4000 RPM for 2 min at 4°C, 100-μl aliquots of the supernatant were transferred to 1.5-ml siliconized tubes. NR5A2 IP was performed by adding 2 μl of rabbit anti-human LRH serum (Dr. Iannis Talianidis, Crete, Greece). Nonspecific IP was performed by the addition of 1 μg of normal rabbit IgG (sc-2027; Santa Cruz Biotechnology) to a separate aliquot. A third aliquot was set aside as a source of input DNA. The IP reactions were incubated overnight on ice with gentle rocking. The next morning, 50 μl of protein G PLUS agarose beads was added to each IP reaction, and the reactions were incubated for an additional 2 h. The washing of the beads, elution of complexes, decross-linking, and DNA isolation were as described in Hatzis and Talianidis (2001). DNA pellets were dissolved in 50 μl of Tris-HCl, pH 8.0, and EDTA.

Real Time Polymerase Chain Reaction. Reactions were set up using the iQ SYBR Green Supermix (Bio-Rad) following the manufacturers' instructions with the exception that the total volume was reduced to 20 μl. Quadruplicate wells were set up for each DNA template isolated from the LRH and IgG IP reactions and from sonicate without IP used as the input DNA. An upstream primer, 5′-CTT CAG TGC CCT GAC TGT GTC ATC-3′ (*CYP2E1*1C* position −3713 to −3690), and downstream reverse primer, 5′-GAG TCC TGG AAG CAG CAA GAG TG-3′ (*CYP2E1*1C* position −3314 to −3337), were used to amplify a 373-bp product that includes all five upstream repeats. Real-time polymerase chain reaction (RT-PCR) was performed using the iCycler iQ system (Bio-Rad) with the following protocol: initial denaturation at 95°C for 3 min, followed by 40 cycles consisting of denaturation at 95°C for 0.10 min, annealing at 58°C for 0.30 min, and extension at 72°C for 1.3 min. The PCR was immediately followed by a DNA melting protocol: 1.0 min at 95°C, 1.0 min at 55°C, followed by 79 additional 1.0-min cycles increasing 0.5°C in each successive cycle. Software analysis by the iCycler iQ system was used to determine the baseline during cycles 2 through 10, the threshold (defined as 10 times the standard deviation of the baseline), and the threshold cycle (C_T) (the cycle at which the threshold was crossed). The criteria of Aparicio et al. (2005) were used to determine the suitability of inclusion and to calculate binding differences. A net C_T was calculated by subtracting the mean C_T of the non-IP sonicate DNA (input) from the mean C_T of each IP DNA. The difference in net C_T (ΔC_T) was calculated by subtracting the net C_T of the selective IP (rabbit anti-human LRH) from that of the nonselective (rabbit IgG). The fold enrichment was calculated by raising the mean slope of the linear portion of the amplification curves from the input wells (Sm_{INPUT}) by the exponent of the resulting delta C_T as shown in the following formula: Fold enrichment = $(Sm_{INPUT})2^{\Delta CT}$.

DNA Sequence Analysis and Single-Nucleotide Polymorphism Discovery. For single-nucleotide polymorphism discovery, a panel of DNA samples from 24 unrelated individuals and representative of the human population's ethnic diversity (Collins et al., 1998)

was obtained from the Coriell Institute (Camden, NJ). Sequence analysis was carried out essentially as described by Hines et al. (2003) but was performed using the Beckman Coulter CEQ DTCS dye-labeled dideoxynucleotide terminator cycle sequencing protocol with analysis on a Beckman Coulter CEQ 8000 gene analyzer. The same primer pair described in the RT-PCR section above was used for both gene-specific PCR DNA amplification and sequencing. Sequencing results were analyzed with SeqManII software (DNAstar, Madison, WI) to identify potential discrepancies relative to the reference sequence (National Center for Biotechnology Information locus Link ID number 1571 based on contig NT_017795.17, build 35.1).

Statistical Analysis. Significant differences in reporter gene activity between adenovirus or plasmid constructs was determined using one-way ANOVA with Tukey-Kramer multiple comparison post test using InStat (version 3.05; GraphPad Software, Inc., San Diego, CA).

Results

The Distal Region of the *CYP2E1* Promoter Is Associated with Higher Basal Expression and Contains a 59-Bp Direct Repeat Sequence. As part of ongoing studies on the functional components of the human *CYP2E1* promoter, DNA fragments representing sequences from both the *CYP2E1*1C* (position −3710 to +25) and **1D* (position −3806 to +25) alleles were cloned into an adenovirus vector for studies in primary human hepatocytes. In addition to the full-length sequences, adenovirus vectors lacking the distal end of the *CYP2E1*1C* (position −3710 to −2555) and **1D* (position −3806 to −2651) promoters were constructed. When these four adenovirus vectors were used to infect primary human hepatocytes, 10-fold (**1C*) and 6-fold (**1D*) reductions (ANOVA, $p < 0.001$) in luciferase reporter activity were noted with constructs missing the distal end of the promoter compared with the full-length promoter constructs (Fig. 1A). When these adenovirus vectors were used to infect HepG2 cells, 7-fold (**1C*) and 3-fold (**1D*) reductions (ANOVA, $p < 0.001$) in luciferase reporter activity also were observed if the distal end of the promoter was absent (Fig. 1B). Thus, reduction in reporter activity was observed regardless of the *CYP2E1* promoter allele (**1C* or **1D*) or cells infected (primary hepatocytes or HepG2 hepatoma cells). These data suggested the presence of a potent, constitutively acting, positive regulatory domain within the deleted sequence.

Inspection of the 1155 bp of the distal promoter revealed five direct repeats of 59 bp that exhibit high sequence identity (86–96%) between position −3690 and −3386 (**1C*) and position −3746 and −3482 (**1D*) (Fig. 2). A search of the human genome using the consensus sequence of these five repeats (Fig. 2) and the National Center for Biotechnology Information BLAST program (Altschul et al., 1997) revealed significant identity with the *CYP2E1* promoter sequence only (E value = 1×10^{-19}), the next closest similarity being orders of magnitude less (E value = 1.3).

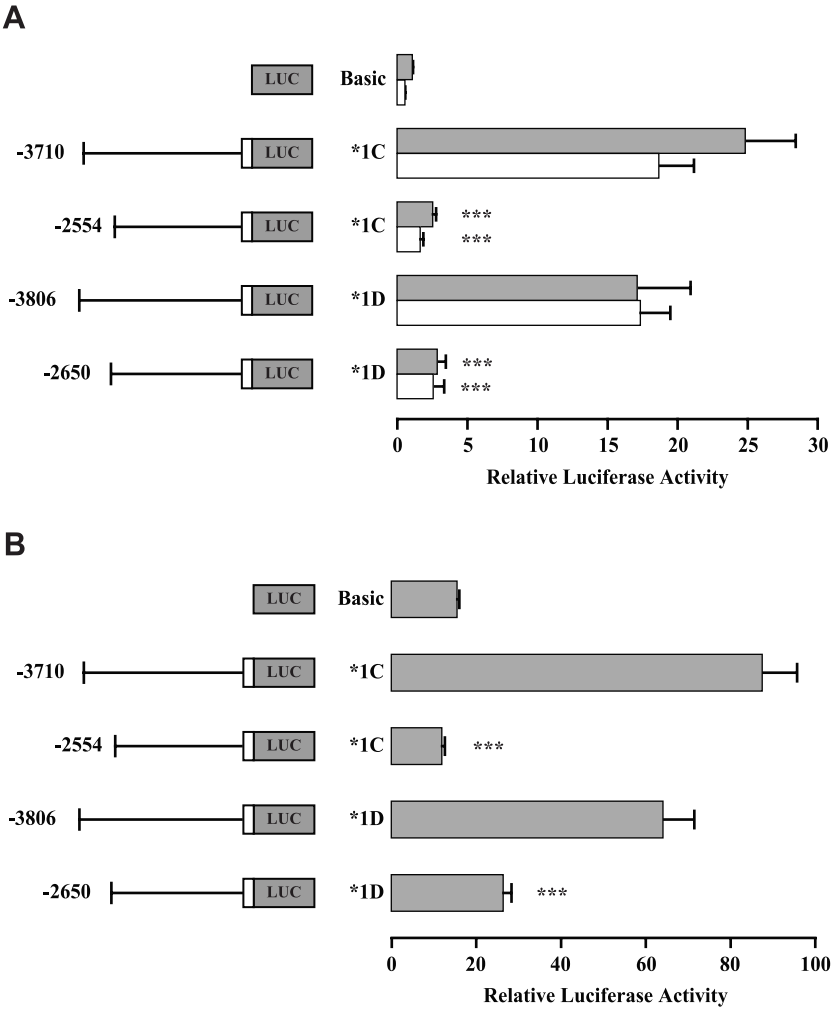

Fig. 1. Deletion of the distal end of the human *CYP2E1*1C* and **1D* promoter results in reduced luciferase activity in primary human hepatocytes and HepG2 cells. Relative luciferase activity with recombinant adenovirus expression vectors containing no promoter (Basic), full-length promoter for *CYP2E1*1C* position −3710 to +25 (−3710 **1C*), or for *CYP2E1*1D* position −3806 to +25 (−3806 **1D*), or their respective corresponding distal deletions, position −2554 to +25 (−2554 **1C*), or position −2650 to +25 (−2650 **1D*) is shown. Salient features of each luciferase reporter construct are depicted to the left of each bar graph. Data shown are mean ± S.D. of results obtained 20 h after infection with an MOI of 10. A, primary human hepatocytes: □, donor HH996, 54-year-old male; ■, donor HH997, 44-year-old male. Single experiment for each donor, triplicate wells for each treatment. B, HepG2 cells, passage 9, data shown are from one of three replicate experiments, quadruplicate wells for each treatment. A significant reduction in reporter activity was seen in each construct lacking distal *CYP2E1* compared with the corresponding full-length construct (***, $p < 0.001$, ANOVA with Tukey-Kramer post test).

The Distal *CYP2E1* Promoter 59-Bp Repeats Behave as an Enhancer. To determine whether the 59-bp repeats were involved in the observed positive regulation, the ability of the repeats to enhance the activity of a minimal *Ptk* directing luciferase expression was tested by transient transfection into HepG2 cells. pDGM63, containing only the Ptk, resulted in luciferase activity approximately 10-fold greater than the promoterless control ($p < 0.001$) (Fig. 3, note log scale). pDGM61, containing the *CYP2E1* 59-bp repeats and no promoter, resulted in luciferase activity not significantly different from pGL3Basic (Fig. 3). Thus, the *CYP2E1* 59-bp repeats per se do not exhibit promoter-like activity. However, pDGM62 containing the *CYP2E1* 59-bp repeats immediately upstream of Ptk resulted in luciferase activity that was 10-fold higher ($p < 0.001$) than that of pDGM63 (Fig. 3). pDGM65, which has the repeats in reverse orientation, resulted in Ptk-driven luciferase activity that was significantly lower ($p < 0.001$) than that for pDGM62 but was still 5-fold higher ($p < 0.001$) than that of pDGM63, which contains Ptk alone. Thus, the *CYP2E1*1C* 59-bp repeat region positively regulates the basal Ptk in an orientation-independent manner.

GATA Factor Binding Sites in the *CYP2E1* 59-Bp Repeats Participate in Sequence-Specific Nuclear Protein Binding. Because the enhancer effect was similar in direction and magnitude in HepG2 and human hepatocytes (Fig. 1, A and B), subsequent electrophoretic mobility shift assays (EMSAs) to identify sequence-specific nuclear protein binding to the *CYP2E1* 59-bp repeats were performed using HepG2 nuclear extract. Initial experiments used a 123-bp Hinf I fragment (*CYP2E1*1C* position −3623 to −3501) from the repeat region as a probe (Fig. 4A). Three sequence-specific binding complexes were formed with HepG2 nuclear extract (Fig. 5, arrowheads A, B, and C, lanes 2 and 3).

To narrow the location of the functionally important DNA sequence within the 123-bp Hinf I fragment, five overlapping double-stranded oligonucleotides were synthesized that spanned the entire probe: positions −3623 to −3594, − 3598 to −3569, − 3573 to −3544, − 3548 to −3519, and −3523 to −3501 (Fig. 4B). The two double-stranded oligonucleotides representing *CYP2E1* positions −3623 to −3594 and −3548 to −3519 competed with the 123-bp Hinf I probe for sequence-specific binding, whereas the other three did not (Fig. 5, arrowheads A, B, and C for lanes 4–8). Both of these double-stranded oligonucleotides have a 17-bp sequence, 5′-ATC CAG CCC **TTG ATA AA**-3′, that includes a GATA transcription factor consensus binding sequence, WGATAR, at −3601 to −3596 and −3539 to −3534, respectively (Fig. 4B).

GATA4 and an Unrelated Factor Participate in Binding to the 59-Bp Repeats. Sequence-specific binding of known GATA transcription factor(s) to the repeat sequence was examined using the −3548 to −3519 double-stranded oligonucleotide as an EMSA probe. Several sequence-specific complexes were noted, two doublets designated D and E, and two apparent singlets designated F and G (Fig. 6, lanes 2–4) that were eliminated by the mutation of the GATA core sequence (Fig. 6, lanes 5 and 6). The GATA factor consensus binding sequence competed for specific binding of complexes D and F (Fig. 6, lanes 7 and 8). Inclusion of antibody to GATA4 in the binding reaction produced a supershift (Fig. 6, lane 9, open arrowheads) with concomitant loss of the same two complexes, D and F. Using the same probe, no supershift was observed with antibody to GATA1 or GATA6, and no supershift was observed with GATA4 antibody when the probe included the TAGCAA mutation (data not shown). Taken together, these results suggest that GATA4 is capable of specific binding to the repeat sequences and is involved in both complexes D and F.

Additional sequence-specific complexes were observed other than those involving GATA4 (Fig. 6, complexes E and G, lanes 3 and 4 compared with 7–9). However, the double-stranded oligonucleotide containing the TAGCAA mutation also failed to compete for these complexes. Furthermore, the migration of complexes E or G was not altered by inclusion of GATA antibodies in the binding reaction (Fig. 6, lane 9). Together, these results are consistent with additional transcription factors(s) not in the GATA factor family but requiring all or part of the GATA sequence to bind to the *CYP2E1* 59-bp repeats.

The Prostate-Specific Antigen Enhancer Sequence Competes with the *CYP2E1* −3548 to −3519 Probe for Specific Binding. Tricho-rhino-phalangeal syndrome type I protein (TRPS1) is an atypical vertebrate GATA protein containing only one GATA-type zinc finger. TRPS1 binds to consensus GATA motifs and acts as a transcriptional repressor (van den Bemd et al., 2003). van den Bemd et al. (2003) demonstrated sequence-specific binding of purified human recombinant TRPS1 to an inverted GATA motif within a 35-bp sequence of the far upstream enhancer of the prostate-specific antigen (*PSA*) gene. This 35-bp *PSA* enhancer sequence was used in competitive EMSA to test for sequence-specific binding of TRPS1 to the *CYP2E1* 59-bp repeats. Although TRPS1 sequence-specific competition with a *CYP2E1* −3548 to −3519 probe was observed that included complex E (Fig. 7, lane 4), this binding was not eliminated when the inverted GATA was mutated to TAAG (Fig. 8A, lanes 6 and 7). When the 35-bp *PSA* enhancer sequence was used as a probe, sequence-specific competition for HepG2 nuclear protein was lost when the inverted GATA was mutated to TAAG (data not shown), similar to the observation of van den Bemd et al. (2003). Thus, these data suggest


```
R1 AGATGAGTCGTCAGGTAAATCAGCCCTCGATAAAATAGCAGGAACCTGTTCACCCAAAA-  94%
R2 AAATGAGTCGTCAGGTAAATCAGCCCTTGATAAAACAGCAGGAACCTGTTCACCCAAAA-  96%
R3 AAATGAGTTGTCAGGTAAATCAGACCTTGATAAAATAGCAG-CGCCTGTTCACCCGAAAC  90%
R4 AAATGAGTCATCAGGTAAATCAGCCCTGGATAAAATAGCAGGAACCTGTTCCGCCGAAA-  93%
R5 AAATGAGTCGTCAGGTAAATCAGCCCTTGATAAAATAACA-GCACTTGTTGACCAGAAGC  86%

Consensus

AAATGAGTCGTCAGGTAAATCAGCCCTTGATAAAATAGCAGGAACCTGTTCACCCGAAA
```

Fig. 2. The five 59-bp sequences of the *CYP2E1* promoter position −3690 to −3386 (**1C*) and position −3786 to −3482 (**1D*). Alignment of repeat sequences one through five (R1–R5) is shown. Percentage of sequence identity and consensus sequence were determined by multiway alignment using Align Plus 5, version 5.02 (Science and Education Software, Cary, NC).

that TRPS1 is not the factor involved in the formation of complex E.

Oligonucleotides Containing Binding Sites for the Nuclear Receptor Family 5 Group A Compete with *CYP2E1* −3548 to −3519 Probe for Sequence-Specific Binding. Two members of the nuclear receptor family 5 group A (NR5A), steroidogenic factor 1 (SF1 or NR5A1), and fetoprotein transcription factor (FTF or NR5A2) are found in human liver and hepatoma cells (Nitta et al., 1999; Gilbert et al., 2000; Sirianni et al., 2002). To test for sequence-specific binding at the potential NR5A site in the *CYP2E1* position −3548 to −3519 probe, competitive EMSA was performed with double-stranded oligonucleotides representing known binding sites for SF1 (Pincas et al., 2001) or FTF (Galarneau et al., 1996). Double-stranded oligonucleotides representing authentic SF1 or FTF binding sites exhibited sequence-specific competition with the *CYP2E1* −3548 to −3519 probe, primarily involving complex E (Fig. 7, lanes 5 and 7). Competition was not observed with double-stranded oligonucleotides in which the SF1 and FTF binding sites were mutated (SF1 mutant and FTF mutant, Fig. 7, lanes 6 and 8). These same mutations were shown previously to eliminate specific SF1 and FTF binding on the rat gonadotropin-releasing hormone receptor and α-fetoprotein genes, respectively (Galarneau et al., 1996; Pincas et al., 2001). Together these results implicate one or more members of NR5A in sequence-specific binding to the *CYP2E1* 59-bp repeats and participating in the formation of complex E.

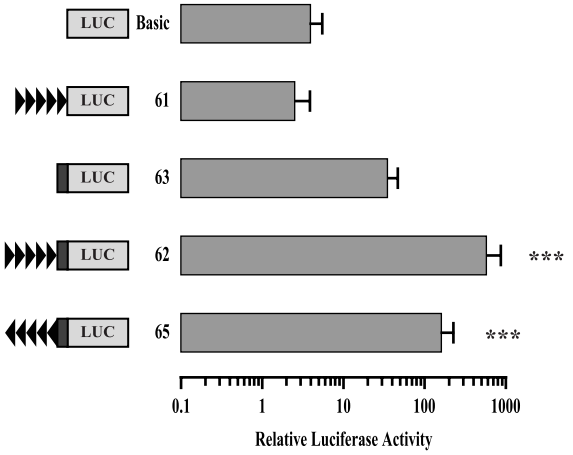

Fig. 3. The *CYP2E1*1C* 59-bp repeats enhance *Ptk*-driven luciferase activity in an orientation-independent manner. Relative luciferase activity in HepG2 cell extracts 48 h after transfection with pGL3Basic, pDGM61, (*CYP2E1*1C* 59-bp repeats, denoted as arrowheads); pDGM63, (*Ptk,* denoted as closed dark gray box), pDGM62 (*CYP2E1*1C* 59-bp repeats + *Ptk*), or pDGM65 (inverted *CYP2E1*1C* 59-bp repeats + *Ptk*). The data shown are mean ± S.D. from two replicate transfections. A significant difference in *Ptk*-driven luciferase activity was seen for both pDGM62 and pDGM65 compared with pDGM63 (***, $p < 0.001$, ANOVA with Tukey-Kramer post test).

The ability of the *PSA* enhancer sequence to compete with the *CYP2E1* −3548 to −3519 probe independent of an intact GATA binding site seemed incongruent with the requirement of a GATA site for competition with the *CYP2E1* −3548 to −3519 double-stranded oligonucleotide. However, this apparent conflict was resolved upon closer inspection of the sequences of the oligonucleotides used in the competition experiments. The 35-bp *PSA* enhancer and the *CYP2E1* sequences with GATA site intact have a potential binding site for members of the NR5A family. The putative NR5A binding site in the *CYP2E1* −3548 to −3519 sequence is immediately upstream of the GATA site, overlapping it by 1 base pair (Fig. 8B). Mutation of the GATA site in the *CYP2E1* −3548 to −3519 double-stranded oligonucleotide (**G**ATA to **A**GCA) also mutated the putative NR5A site (CCTT**G** to CCTT**A**). The putative NR5A binding site in the 35-bp PSA enhancer sequence is 6 bp upstream of the inverted GATA and was not altered by mutation of the latter element. Thus, the difference in the relative physical location of the GATA and NR5A binding sites between the *CYP2E1* −3548 to −3519 and 35-bp *PSA* enhancer sequences could explain the observed difference in competitive abilities when their respective GATA sites were mutated.

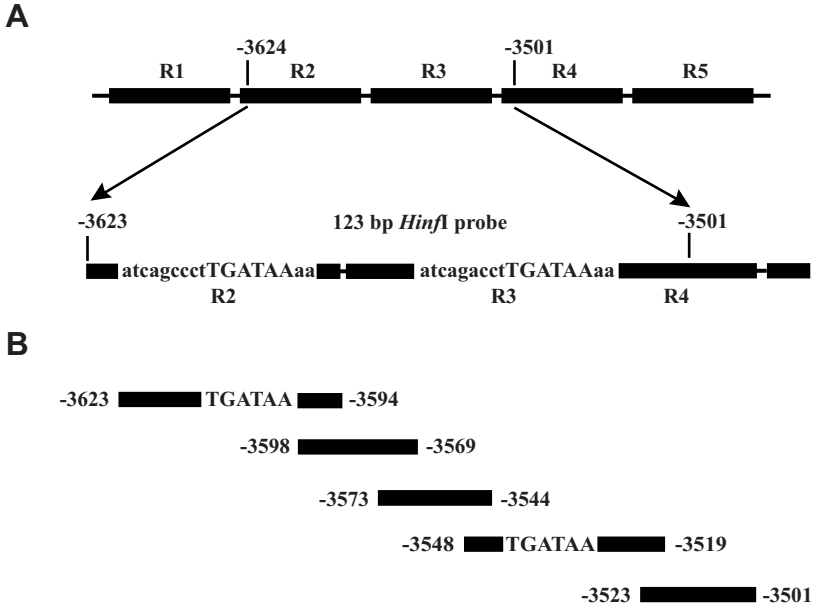

Fig. 4. Locations of the *CYP2E1*1C* 59-bp repeats, the 123-bp Hinf I fragment, and the five double-stranded oligonucleotides used to identify sequences involved in binding nuclear proteins. A, the *CYP2E1* 59-bp repeats (R1–R5) and the locations of the Hinf I restriction sites used to obtain the 123-bp probe sequence are depicted. The putative GATA factor binding sites are shown in uppercase letters within the 17-bp sequence shared between double-stranded oligonucleotides covering positions −3623 to −3594 and −3548 to −3519. B, the overlapping double-stranded oligonucleotides spanning the 123-bp Hinf I probe sequence used in competitive EMSA are shown along with the relative locations of the consensus GATA elements.

ChIP Demonstrates in Vivo Binding of NR5A2 to the *CYP2E1* 59-Bp Repeats in Primary Human Hepatocytes. The EMSA experiments presented above demonstrated in vitro binding of NR5A members to the *CYP2E1* 59-bp repeats in HepG2 cells. However, demonstration of in vivo binding in primary human hepatocytes is critical. Plated human hepatocytes from a 15-month-old male were treated with formaldehyde to form DNA-protein cross-links. Cell lysates were sonicated to fragment the chromatin, and ChIP was performed as described in the *Materials and Methods* section using either nonspecific rabbit IgG or rabbit anti-human LRH (NR5A2) for immunoprecipitation. In two independent ChIP experiments, the RT-PCR amplification of the NR5A2 immunoprecipitated DNA crossed the amplification threshold 2.4 and 2.2 cycles earlier compared with the amplification reactions of the nonspecific immunoprecipitation DNA (Fig. 9). These cycle number differences yield a 4- and 5-fold increase in NR5A2 binding, respectively, in the two experiments.

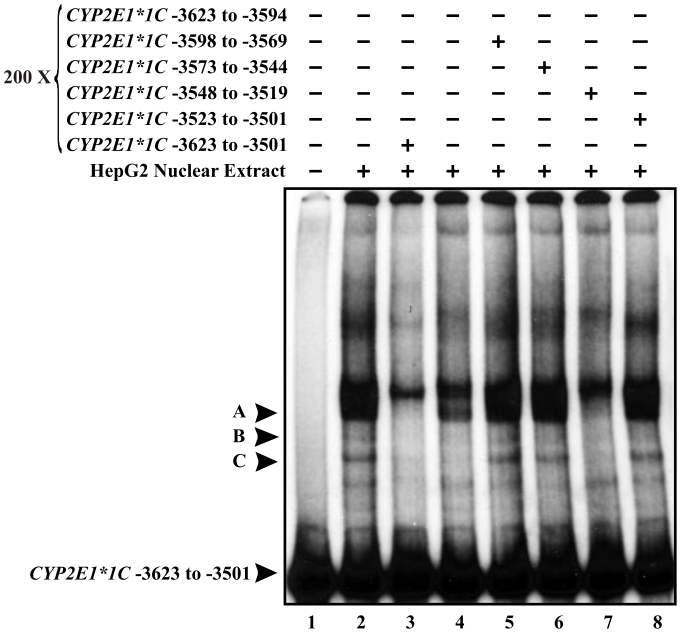

Fig. 5. Double-stranded oligonucleotides representing *CYP2E1*1C* positions −3623 to −3594 and −3548 to −3519 compete with the *CYP2E1*1C* −3623 to −3501 Hinf I probe for sequence-specific protein binding. Competition for HepG2 nuclear protein sequence-specific binding between the *CYP2E1*1C* 123-bp Hinf I probe and a 200-fold molar excess of unlabeled probe sequence (lane 3) or each of the five overlapping double-stranded oligonucleotides (lanes 4–8) is shown. Arrowheads A, B, and C denote sequence-specific nuclear protein binding complexes based on competition with both the unlabeled *CYP2E1*1C* −3623 to −3501 probe and the double-stranded oligonucleotides representing *CYP2E1*1C* positions −3623 to −3594 (lane 4) and −3548 to −3519 (lane 7).

Genetic Variation within the *CYP2E1* 59-Bp Repeats. Given the potent constitutive enhancer activity demonstrated by the *CYP2E1* 59-bp repeat elements, we questioned whether genetic variation in the GATA and/or NR5A elements within the repeats might contribute to intersubject variation in *CYP2E1* constitutive expression. To address this question, these sequences were examined in the 24-subject panel from the Coriell Polymorphism Discovery Resource (Collins et al., 1998). However, no variants were observed within either element in any of the DNA samples examined, suggesting that these regulatory elements are highly conserved.

Discussion

We report the first identification of an upstream enhancer with a significant impact on *CYP2E1* constitutive gene expression. Deletion of the distal 1155 bp of upstream *CYP2E1* sequence resulted in decreased luciferase reporter activity in primary human hepatocytes and HepG2 hepatoma cells. The

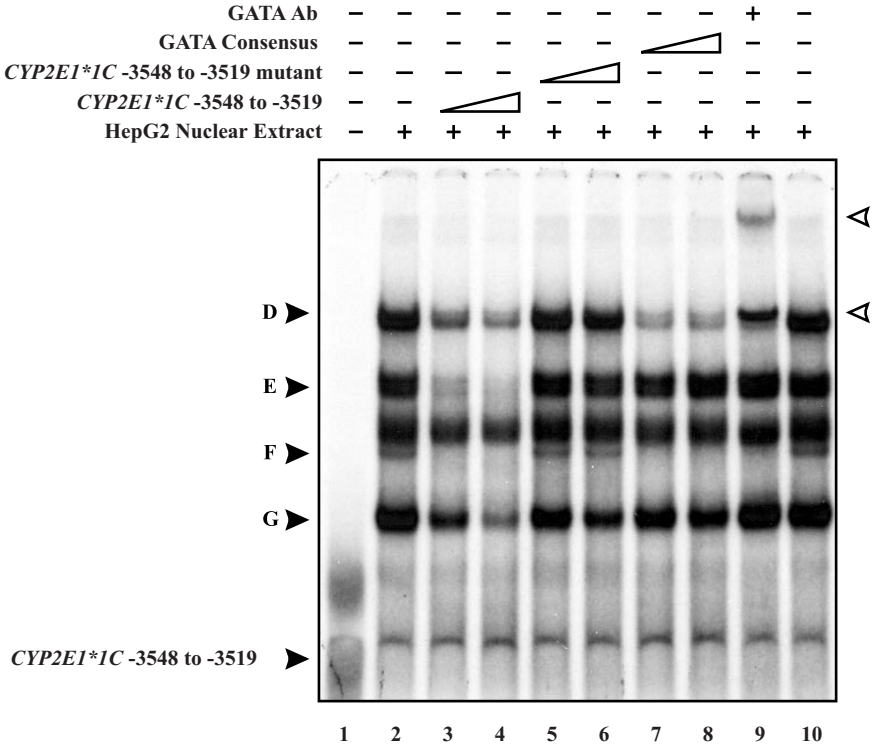

Fig. 6. GATA4 binds the *CYP2E1*1C* position −3548 to −3519 enhancer element. Competition for HepG2 nuclear protein sequence-specific binding was examined between the *CYP2E1*1C* position −3548 to −3519 probe and a 50- or 100-fold molar excess of unlabeled sequence with GATA factor binding site intact (lanes 3 and 4) or mutated to AGCA (lanes 5 and 6), or a 50- or 100-fold molar excess of a GATA factor consensus binding sequence (lanes 7 and 8). Arrowheads D, E, F, and G denote sequence-specific nuclear protein binding complexes. Lane 9 shows the supershift (open arrowheads) observed by inclusion of 4 μg of antibody to GATA4 in the binding reaction.

domain within the 1155-bp fragment contains five direct 59-bp repeat sequences and enhanced thymidine kinase promoter-driven activity in HepG2 cells and in the human embryonic kidney cell line 293A (data not shown) in an orientation-independent manner. Finally, nuclear protein extracts from HepG2 cells formed sequence-specific complexes dependent on a 17-bp region containing a canonical GATA factor binding site. Two sequence-specific complexes were demonstrated to involve GATA4. In addition, a binding site for NR5A orphan nuclear receptors was found to overlap the GATA factor binding site, with sequence-specific binding occurring for both steroidogenic factor 1 (NR5A1) and fetoprotein transcription factor (NR5A2). Furthermore, such binding was demonstrated both in vitro and in vivo.

Cytochrome P450 genes are regulated by liver-enriched transcription factors, including members of the hepatocyte nuclear factor families (HNF-1, -3, -4, and -6) and CCAAT/enhancer binding proteins (Schrem et al., 2002). Combinations of these HNFs and ubiquitous transcription factors are believed to be necessary for liver-specific gene expression. The GATA transcription factors GATA4 and GATA6 also play a role in regulating liver-specific gene expression (Molkentin, 2000). Liver-selective genes regulated by GATA4 include albumin (Bossard and Zaret, 1998), fetoprotein transcription factor (Pare et al., 2001), and erythropoietin (Dame et al., 2004). GATA4 and HNF3β bind to the albumin enhancer element, resulting in alteration of nucleosomal architecture (Bossard and Zaret, 1998). Whether GATA4 binding to the *CYP2E1* enhancer acts in a similar fashion is being addressed in our laboratory, although we could not demonstrate binding of HNF3β to the *CYP2E1* enhancer. However, using competitive and supershift EMSA, we also eliminated GATA6, CCAAT/enhancer binding protein β, HNF1α, HNF1β, and HNF4 as factors binding to the *CYP2E1* enhancer.

Nuclear receptors are known to regulate several cytochrome P450 genes (Waxman, 1999). Our study identified potential orphan nuclear receptor binding sites in the *CYP2E1* enhancer. These are the first bona fide nuclear receptor sites reported in the *CYP2E1* promoter with supporting data beyond sequence identity. SF1 and FTF are expressed in liver with FTF being predominant in human

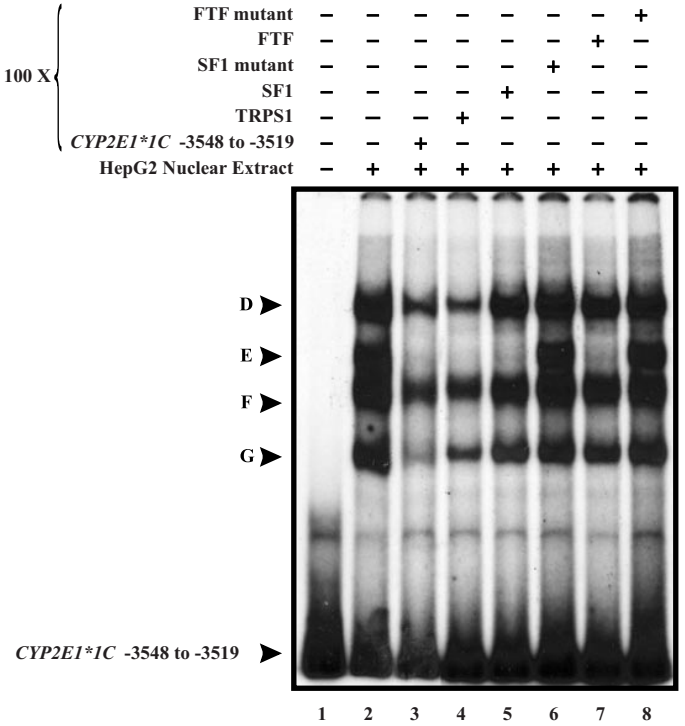

Fig. 7. Oligonucleotides containing binding sites for TRPS1, SF-1, or FTF compete for sequence-specific binding with the *CYP2E1*1C* −3548 to −3519 probe. Competition for HepG2 nuclear protein sequence-specific binding was examined between the *CYP2E1*1C* position −3548 to −3519 probe and a 100-fold molar excess of unlabeled sequence (lane 3), a 35-bp fragment from the *PSA* enhancer containing an inverted GATA site required for TRPS1 binding (lane 4), double-stranded oligonucleotides containing binding sites for the orphan nuclear receptors SF-1 (lane 5) or FTF (lane 7), and their respective double-stranded oligonucleotides with mutated binding sites, SF1 mutant (lane 6) and FTF mutant (lane 8). Arrowheads D–G denote sequence-specific nuclear protein binding complexes.

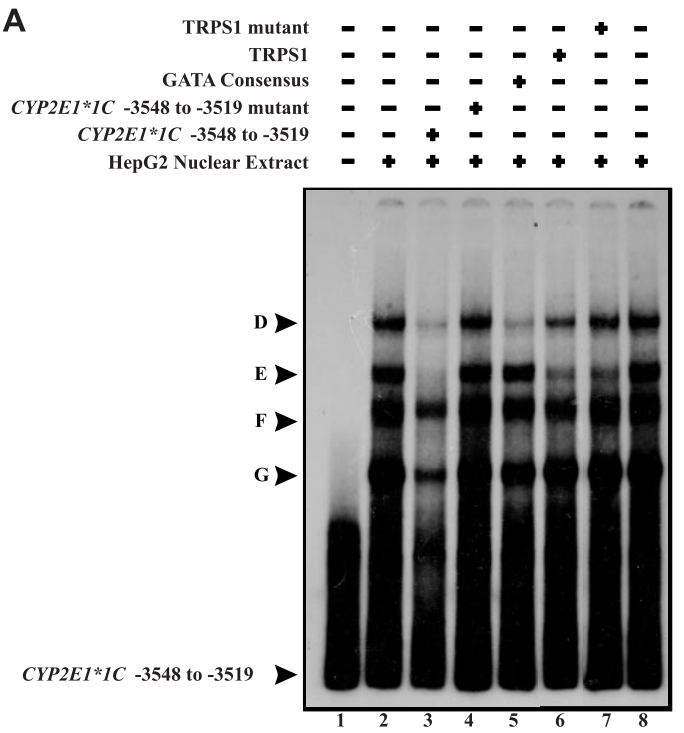

Fig. 8. Oligonucleotides containing binding sites for TRPS1 and TRPS1 mutant compete for sequence-specific binding with the *CYP2E1*1C* −3548 to −3519 probe. A, competition for HepG2 nuclear protein sequence-specific binding was examined between the *CYP2E1*1C* position −3548 to −3519 probe and a 100-fold molar excess of unlabeled sequence with GATA factor binding site intact (lane 3) or mutated to AGCA (lane 4), a GATA factor consensus binding sequence (lane 5), a 35-bp fragment from the *PSA* enhancer containing an inverted GATA site required for TRPS1 binding intact (lane 6), or mutated to TAAG (lane 7). Arrowheads D, E, F, and G denote sequence-specific nuclear protein binding complexes. B, sequences of *CYP2E1*1C* position −3548 to −3519 with the GATA site intact (upper sequence) or mutated (underlined bases in lower sequence) are shown. The NR5A binding sequence is shown in italics. The asterisk denotes the single base overlap between the NR5A and GATA sites in *CYP2E1*1C* position −3548 to −3519.

(Sirianni et al., 2002) and rat (Falender et al., 2003). SF1 and FTF are expressed in HepG2 cells with SF1 being more abundant (Gilbert et al., 2000). Moreover, FTF is altered compared with that observed in normal human liver (Galarneau et al., 1996). Both factors can bind to the same sequence and both act as transcription activators. Thus, the specific NR5A factor acting at the distal *CYP2E1* enhancer may differ between primary human hepatocytes and hepatoma cells but result in the same functional effect.

GATA4 and SF1/FTF can act together to regulate gene expression (Tremblay and Viger, 1999; Flück and Miller, 2004). The SF1/FTF and GATA4 binding sites overlap in the *CYP2E1* enhancer. Proximity and functional interaction between GATA4 and SF1 sites have been found in the human *CYP17* promoter (Flück and Miller, 2004), and the mouse Müllerian Inhibiting Substance gene promoter (Tremblay and Viger, 1999). Together, these findings suggest that the proximity of GATA4 and SF1/FTF binding sites may play a role in *CYP2E1* gene regulation.

The *CYP2E1* upstream enhancer sequence seems to be highly conserved within the human population. Thus, differences in the enhancer sequence per se are unlikely to explain intersubject variation in constitutive *CYP2E1* expression. Known functional genetic variants in the transcription factors binding within this region, such as GATA4 (Garg et al., 2003) and FTF (Nitta et al., 1999), or in cooperating cofactors, such as FOG2 (Pizzuti et al., 2003) and SHP (Nishigori et al., 2001), may contribute to intersubject variation in constitutive *CYP2E1* expression. Furthermore, FTF and GATA4 are known to play pivotal roles in the differentiation and development of endodermally derived tissues, such as liver and intestine (Molkentin, 2000; Fayard et al., 2004). Thus, the observed ontogenic (Johnsrud et al., 2003) and tissue-specific CYP2E1 variation (Lieber, 1999) may be in part a function of the interaction of the genetic variants of GATA4, FTF, and their cofactors acting at the *CYP2E1* upstream enhancer.

Before this report, the major contributor to *CYP2E1* basal expression had been assigned to the proximal promoter, specifically to the HNF1α binding site at position -112 to -95 (Ueno and Gonzalez, 1990; Liu and Gonzalez, 1995). However, studies supporting this conclusion only examined the first 1500 bp of the *CYP2E1* upstream region. Thus, the contribution of the more distally located 59-bp repeats would have been missed. Hu et al. (1999) used 5′-nested deletions of a 3800-bp human *CYP2E1* upstream sequence to identify regions having an effect on luciferase reporter gene expression in human hepatoma cells. They proposed the existence of a negative acting element between positions -3712 and -3205, because loss of that region resulted in a 2-fold increase in luciferase activity. This region includes the distal *CYP2E1* enhancer reported in this article. Although the disparity between their observation and the strong enhancer effect observed in the present study is difficult to reconcile, it may be due in part to the particular human hepatoma cell line used by Hu et al. (1999) and its prolonged maintenance at confluence before transfection. We are not aware of additional confirming evidence for the proposed negative regulatory element reported by Hu et al. (1999). The consistency of our observation in both primary human hepatocytes and HepG2 cells, enhancer activity in conjunction with the heterologous promoter, and the identification of sequence-specific binding by known transcription factors argues strongly for a positive function for the 59-bp repeats.

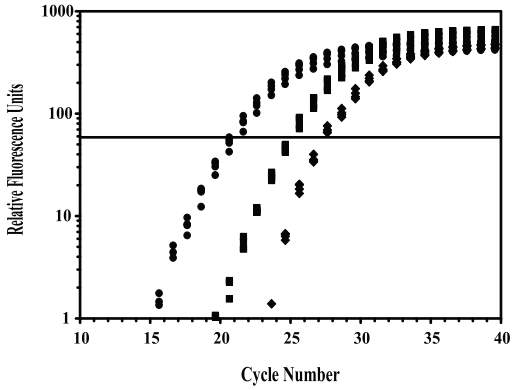

Fig. 9. ChIP RT-PCR demonstrates in vivo binding of NR5A2 to the *CYP2E1* 59-bp repeats human hepatocytes. DNA isolated from non-IP (input), selective IP (NR5A2), and nonselective IP (IgG) from sonicated human hepatocyte chromatin was subjected to RT-PCR to amplify *CYP2E1* position -3714 to -3341 containing the 59-bp repeats. Data shown are from one of two independent immunoprecipitations. Circles, input; squares, NR5A2 IP; diamonds, IgG IP. Multiple curves of the same symbol represent replicate wells. Horizontal line across plot designates the threshold value.

Both human and rodent *CYP2E1* exhibit constitutive and inducible expression, and both transcriptional and post-transcriptional mechanisms contribute in both species. Although protein stabilization may be a dominant mechanism in rodents and has been observed in a HepG2 human hepatoma cell line (Carroccio et al., 1994), multiple studies of human liver tissue and in vivo human CYP2E1 metabolic activity are consistent with differential transcription as a substantial regulatory step (Takahashi et al., 1993; McCarver et al., 1998; Raucy et al., 1999). Sequence comparisons between rodent and human also support species differences in *CYP2E1* regulation. A BLAST search of the 59-bp repeat consensus sequence found significant sequence identity only in the human *CYP2E1* upstream region, consistent with the transcriptional regulation of constitutive *CYP2E1* expression being different across species. Consistent with this, Hu et al. (1999) did not find significant sequence identity between rat and human *CYP2E1* in the region corresponding to the location of the 59-bp repeats. Although, high sequence identity between rat and human *CYP2E1* was found within the first 150 bp of the proximal promoter (Umeno et al., 1988; Hu et al., 1999), the overall *CYP2E1* promoter sequence similarity was only 50%, and homology was largely restricted to two areas (Hu et al., 1999). Furthermore, cross-species transient transfection expression studies resulted in low activity compared with studies restricted to the same species (Hu et al., 1999). Our data demonstrating a positive functional effect of the 59-bp repeats on basal levels of reporter gene activity combined with their uniqueness to the human further calls into question the degree of functional conservation of constitutive *CYP2E1* transcriptional regulation between humans and other species.

In summary, we report the identification of a unique far upstream human *CYP2E1* element, consisting of five nearly identical 59-bp repeats, that enhances constitutive *CYP2E1* expression through multiple mechanisms. Several pieces of evidence reported herein support a GATA4-dependent mechanism, whereas additional evidence implicates an NR5A-

related mechanism. Additional factors acting at this element are as yet unidentified. The magnitude of the enhancer effect in vitro was substantial. The intraspecies differences in this sequence in conjunction with the magnitude of the enhancer's effect add to concern regarding the validity of animal models for studies of *CYP2E1* regulation.

Acknowledgments

We acknowledge Dr. Iannis Talianidis (Institute of Molecular Biology and Biotechnology, Foundation for Research and Technology Hellas, Crete, Greece) for his generous gift of rabbit anti-human LRH serum. The Liver Tissue Procurement and Distribution System (Dr. Steve Strom, University of Pittsburgh, Pittsburgh, PA) is funded by National Institutes of Health contract number N01-DK92310.

References

Aden DP, Fogel A, Plotkin S, Damjanov I, and Knowles BB (1979) Controlled synthesis of HBsAg in a differentiated human liver carcinoma-derived cell line. *Nature (Lond)* **282:**615–616.

Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W, and Lipman DJ (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* **25:**3389–3402.

Aparicio O, Geisberg JV, Sekinger E, Yang A, Moqtaderi Z, and Struhl K (2005) Chromatin immunoprecipitation for determining the association of proteins with specific genomic sequences in vivo, in *Current Protocols in Molecular Biology* (Ausubel FA, Brent R, Kingston RE, Moore DD, Seidman JG, Smith JA, and Struhl K eds) pp 21.3.1–21.3.30, John Wiley & Sons, New York.

Bossard P and Zaret KS (1998) GATA transcription factors as potentiators of gut endoderm differentiation. *Development* **125:**4909–4917.

Carrière V, Berthou F, Baird S, Belloc C, Beaune P, and de Waziers I (1996) Human cytochrome P450 2E1 (CYP2E1): from genotype to phenotype. *Pharmacogenetics* **6:**203–211.

Carroccio A, Wu D, and Cederbaum AI (1994) Ethanol increases content and activity of human cytochrome P4502E1 in a transduced HepG2 cell line. *Biochem Biophys Res Commun* **203:**727–733.

Collins FS, Brooks LD, and Chakravarti A (1998) A DNA polymorphism discovery resource for research on human genetic variation. *Genome Res* **8:**1229–1231.

Dame C, Sola MC, Lim KC, Leach KM, Fandrey J, Ma Y, Knopfle G, Engel JD, and Bungert J (2004) Hepatic erythropoietin gene regulation by GATA-4. *J Biol Chem* **279:**2955–2961.

Falender AE, Lanz R, Malenfant D, Belanger L, and Richards JS (2003) Differential expression of steroidogenic factor-1 and FTF/LRH-1 in the rodent ovary. *Endocrinology* **144:**3598–3610.

Fayard E, Auwerx J, and Schoonjans K (2004) LRH-1: an orphan nuclear receptor involved in development, metabolism and steroidogenesis. *Trends Cell Biol* **14:**250–260.

Flück CE and Miller WL (2004) GATA-4 and GATA-6 modulate tissue-specific transcription of the human gene for P450c17 by direct interaction with Sp1. *Mol Endocrinol* **18:**1144–1157.

Galarneau L, Pare JF, Allard D, Hamel D, Levesque L, Tugwood JD, Green S, and Belanger L (1996) The alpha1-fetoprotein locus is activated by a nuclear receptor of the Drosophila FTZ-F1 family. *Mol Cell Biol* **16:**3853–3865.

Garg V, Kathiriya IS, Barnes R, Schluterman MK, King IN, Butler CA, Rothrock CR, Eapen RS, Hirayama-Yamada K, Joo K, et al. (2003) GATA4 mutations cause human congenital heart defects and reveal an interaction with TBX5. *Nature (Lond)* **424:**443–447.

Gilbert S, Galarneau L, Lamontagne A, Roy S, and Belanger L (2000) The hepatitis B virus core promoter is strongly activated by the liver nuclear receptor fetoprotein transcription factor or by ectopically expressed steroidogenic factor 1. *J Virol* **74:**5032–5039.

Hatzis P and Talianidis I (2001) Regulatory mechanisms controlling human hepatocyte nuclear factor 4α gene expression. *Mol Cell Biol* **21:**7320–7330.

Hines RN, Luo Z, Hopp KA, Cabacungan ET, Koukouritaki SB, and McCarver DG (2003) Genetic variability at the human FMO1 locus: significance of a basal promoter yin yang 1 element polymorphism (FMO1*6). *J Pharmacol Exp Ther* **306:**1210–1218.

Hu Y, Hakkola J, Oscarson M, and Ingelman-Sundberg M (1999) Structural and functional characterization of the 5′-flanking region of the rat and human cytochrome P450 2E1 genes: identification of a polymorphic repeat in the human gene. *Biochem Biophys Res Commun* **263:**286–293.

Johnsrud EK, Koukouritaki SB, Divakaran K, Brunengraber LL, Hines RN, and McCarver DG (2003) Human hepatic CYP2E1 expression during development. *J Pharmacol Exp Ther* **307:**402–407.

Lieber CS (1999) Microsomal ethanol-oxidizing system (MEOS): the first 30 years (1968–1998)—a review. *Alcohol Clin Exp Res* **23:**991–1007.

Liu S-Y and Gonzalez FJ (1995) Role of the liver-enriched transcription factor HNF-1 alpha in expression of the CYP2E1 gene. *DNA Cell Biol* **14:**285–293.

Luckow B and Schutz G (1987) CAT constructions with multiple unique restriction sites for the functional analysis of eukaryotic promoters and regulatory elements. *Nucleic Acids Res* **15:**5490.

McCarver DG, Byun R, Hines RN, Hichme M, and Wegenek W (1998) A genetic polymorphism in the regulatory sequences of human *CYP2E1*: association with increased chlorzoxazone hydroxylation in the presence of obesity and ethanol intake. *Toxicol Appl Pharmacol* **152:**276–281.

Molkentin JD (2000) The zinc finger-containing transcription factors GATA-4, -5, and -6. Ubiquitously expressed regulators of tissue-specific gene expression. *J Biol Chem* **275:**38949–38952.

Nishigori H, Tomura H, Tonooka N, Kanamori M, Yamada S, Sho K, Inoue I, Kikuchi N, Onigata K, Kojima I, et al. (2001) Mutations in the small heterodimer partner gene are associated with mild obesity in Japanese subjects. *Proc Natl Acad Sci USA* **98:**575–580.

Nitta M, Ku S, Brown C, Okamoto AY, and Shan B (1999) CPF: an orphan nuclear receptor that regulates liver-specific expression of the human cholesterol 7α-hydroxylase gene. *Proc Natl Acad Sci USA* **96:**6660–6665.

Pare JF, Roy S, Galarneau L, and Belanger L (2001) The mouse fetoprotein transcription factor (FTF) gene promoter is regulated by three GATA elements with tandem E box and Nkx motifs, and FTF in turn activates the Hnf3β, Hnf4α, and Hnf1α gene promoters. *J Biol Chem* **276:**13136–13144.

Pincas H, Amoyel K, Counis R, and Laverriere JN (2001) Proximal cis-acting elements, including steroidogenic factor 1, mediate the efficiency of a distal enhancer in the promoter of the rat gonadotropin-releasing hormone receptor gene. *Mol Endocrinol* **15:**319–337.

Pizzuti A, Sarkozy A, Newton AL, Conti E, Flex E, Digilio MC, Amati F, Gianni D, Tandoi C, Marino B, et al. (2003) Mutations of ZFPM2/FOG2 gene in sporadic cases of Tetralogy of Fallot. *Hum Mutat* **22:**372–377.

Raucy JL, Schultz ED, Kearins MC, Arora S, Johnston DE, Omdahl JL, Eckmann L, and Carpenter SP (1999) CYP2E1 expression in human lymphocytes from various ethnic populations. *Alcohol Clin Exp Res* **23:**1868–1874.

Schrem H, Klempnauer J, and Borlak J (2002) Liver-enriched transcription factors in liver function and development. Part I: the hepatocyte nuclear factor network and liver-specific gene expression. *Pharmacol Rev* **54:**129–158.

Sirianni R, Seely JB, Attia G, Stocco DM, Carr BR, Pezzi V, and Rainey WE (2002) Liver receptor homologue-1 is expressed in human steroidogenic tissues and activates transcription of genes encoding steroidogenic enzymes. *J Endocrinol* 174: R13–R17.

Takahashi T, Lasker JM, Rosman AS, and Lieber CS (1993) Induction of cytochrome P-4502E1 in the human liver by ethanol is caused by a corresponding increase in encoding messenger RNA. *Hepatology* **17:**236–245.

Tremblay JJ and Viger RS (1999) Transcription factor GATA-4 enhances Mullerian inhibiting substance gene transcription through a direct interaction with the nuclear receptor SF-1. *Mol Endocrinol* **13:**1388–1401.

Ueno T and Gonzalez FJ (1990) Transcriptional control of the rat hepatic *CYP2E1* gene. *Mol Cell Biol* **10:**4495–4505.

Umeno M, McBride OW, Yang CS, Gelboin HV, and Gonzalez FJ (1988) Human ethanol-inducible P450IIE1: complete gene sequence, promoter characterization, chromosome mapping and cDNA-directed expression. *Biochemistry* **27:**9006–9013.

van den Bemd GJ, Jhamai M, Brinkmann AO, and Chang GT (2003) The atypical GATA protein TRPS1 represses androgen-induced prostate-specific antigen expression in LNCaP prostate cancer cells. *Biochem Biophys Res Commun* **312:**578–584.

Waxman DJ (1999) P450 gene induction by structurally diverse xenochemicals: central role of nuclear receptors CAR, PXR, and PPAR. *Arch Biochem Biophys* **369:**11–23.

Address correspondence to: Dr. Jeff D. Shadley, MFRC 5017, Medical College of Wisconsin, 8701 W. Watertown Plank Road, Milwaukee, WI 53226-4801. E-mail: jshadley@mcw.edu