Characterization of the Human Lung CYP2F1 Gene and Identification of a Novel Lung-specific Binding Motif*

Brian A. Carr†, Jie Wan‡, Ronald N. Hines§, and Garold S. Yost†‡¶

From the †Department of Pharmacology and Toxicology, University of Utah, Salt Lake City, Utah 84112 and §Department of Pediatrics, Medical College of Wisconsin, Milwaukee, Wisconsin 53226-4801

The CYP2F1 gene encodes a cytochrome P450 enzyme capable of bioactivating a number of pulmonary-selective toxicants. The expression of CYP2F1 is highly tissue-selective; the highest expression is observed in the lung with little or no hepatic expression. The objective of these studies was to elucidate the mechanisms that govern the unique tissue-specific regulation of CYP2F1. Cosmid and bacterial artificial chromosome clones were screened and sequenced to identify a gene that spanned 14 kbp containing 10 exons, including an untranslated exon 1. Primer extension analysis and 5′-rapid amplification of cDNA ends were used to identify the transcription start site. Several sequences homologous to known cis-elements were identified in the 5′-upstream region of the CYP2F1 promoter. Transient transfection studies with luciferase reporter constructs demonstrated significant functional lung-cell-specific CYP2F1 promoter region (from position −129 to +115). DNase footprinting analysis of 1.6 kbp of the upstream sequence with nuclear extracts from human lung tissues revealed one strong DNA-protein complex at −152 to −182. This nuclear protein (called lung-specific factor, LSF) was present only in lung but not liver or heart tissues. Competitive electrophoretic mobility shift assays characterized a DNA consensus site, within the LSF-binding domain, that was highly similar to two E box motifs, but no known "E box" trans-factors were identified. These studies identified a novel LSF and its consensus sequence that may control tissue-specific expression of CYP2F1.

Cytochrome P450 proteins (P450s) are a superfamily of heme-containing enzymes that generally catalyze the metabolism of endogenous and foreign compounds to metabolites that can easily be eliminated from the body. However, for many foreign compounds cytochrome P450 metabolism produces "bio-activated" metabolites, which are highly reactive with endogenous proteins and DNA, causing cell death and gene mutations. Cytochrome P450 expression, which can be important for organ-specific functions, can also lead to tissue-selective bioactivation and toxicity of drugs and other xenobiotic compounds. Cytochrome P450-mediated bioactivation of toxicants is a particularly relevant process to lung diseases because the lungs are exposed directly to environmental pollutants, such as cigarette smoke. Characterizing the mechanisms that regulate tissue-selective P450 expression is vital to understand organ-specific toxicity and individual differences in susceptibility to environmental pollutants and drugs.

Due to the propensity of human lung to bioactivate procarcinogens and other xenobiotics, several screening processes have been performed to identify new P450s that are potentially involved in population susceptibility to cancers caused by cigarette smoke and other environmental pollutants. A cDNA library screen from human lung tissue identified a P450 gene, designated CYP2F1, that was sequenced and mapped to chromosome 19 (3). Expression of recombinant CYP2F1 showed that this enzyme bioactivates two prototype pneumotoxicants, naphthalene and 3-methylindole. CYP2F1 metabolizes naphthalene to its highly pneumotoxic intermediate, naphthalene-1,2-oxide, and 3-methylindole to its dehydrogenated pneumotoxic product, 3-methyleindolene (4–7). CYP2F1 can also bioactivate styrene to its carcinogenic epoxide (8).

The expression of P450 enzymes is controlled by diverse regulatory mechanisms such as inducible transcriptional activation by ligand-activated receptors and constitutive expression by tissue-enriched transcription factors, each of which generally bind to specific regulatory elements in the 5′-upstream regions of genes. It is not known which regulatory factors are responsible for expression of P450 genes in pulmonary tissues or whether these factors might be involved in population susceptibility to lung cancers. However, it has been shown that many P450 genes, CYP1A1, CYP1B1, CYP2B6, CYP2E1, CYP2F1, CYP2S1, CYP3A5, and CYP4A1, are transcribed in lung tissues (9–13), and lung tissues activate carcinogens to produce organ-selective damage (14–17). Despite the extensive knowledge of chemically induced changes in P450 expression, little is known about the transcription factors responsible for constitutive or tissue-selective induction of P450 enzymes. Hepatocyte-enriched transcription factors have been the most extensively studied mechanisms of tissue-selective P450 gene expression, whereas mechanisms of pulmonary-selective gene expression have only received minimal attention (2). One elegant example used transgenic mice with the rat CYP2B1 promoter to drive reporter gene expression in a pulmonary-specific manner (18). Additional studies (19) showed that C/EBP proteins in pulmonary epithelium controlled

* This work was supported by NHLBI Grant HL60143 from the National Institutes of Health. The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked “advertisement”.

† To whom correspondence should be addressed: Dept. of Pharmacology and Toxicology, 30 S. 2000 E., Rm. 201, University of Utah, Salt Lake City, UT 84112. Tel.: 801-581-7956; Fax: 801-685-3945; E-mail: gyost@pharm.utah.edu.

‡ The abbreviations used are: P450s, cytochrome P450 proteins; BAC, bacterial artificial chromosome; LSF, lung-specific factor; C/EBP, CCAAT-enhancer-binding protein; TSS, transcription start site; RACE, rapid amplification of cDNA ends; EMSA, electrophoretic mobility shift assay; AP4, activator protein 4; NFAT, nuclear factor of activated T-cells; eEF1, e-crystallin enhancer-binding protein 1; UTR, untranslated region; BTE, basic transcription element; SIP1, Smad-interacting protein 1; MLC, myosin light chain; MCK, muscle creatine kinase; RT-PCR, reverse transcriptase-coupled PCR.
**CYP2F1** gene expression. Other superb studies (20, 21) have demonstrated that NF1-like factors control nasal-selective expression of CYP1A2 and CYP2A3.

Although there are no examples of endogenous substrates of CYP2F1, its expression appears to be under tight transcriptional control that confines expression predominantly to lung tissues. Therefore, CYP2F1 is an ideal model to elucidate the mechanism of tissue-selective transcription of cytochrome P450 enzymes. Insight into the mechanisms of transcriptional regulation and drug-metabolizing enzymes in lung tissues should provide relevant information regarding tissue-selective toxicity, individual susceptibility to lung cancers, and basic knowledge of constitutive transcriptional mechanisms in lung tissues, where little is known.

**EXPERIMENTAL PROCEDURES**

**Materials**—Human lung tissue was obtained from a 35-year-old male Caucasian donor (Intermountain Donor Services, Salt Lake City, UT). Qiagen plasmid/RNA isolation kits were purchased from Qiagen (Valencia, CA). Cosmid clones were kindly provided by Dr. Harvey Morhensweizer, Lawrence Livermore National Laboratories (Livermore, CA). Human CYP2F1 cDNA in pUC9 (pUC9-2F1) was generously provided by Dr. Frank Gonzalez, NCI, National Institutes of Health (Bethesda, MD). Dr. Michael Lehmann, Institut fur Genetik der Freien Universitat Berlin (Berlin, Germany), provided anti-AP4 antibody. The AP4 expression plasmid was obtained from Dr. Laura Bridgewater, Brigham Young University (Provo, UT) with permission of Dr. Robert Tjian (University of California, Berkeley, CA). The expression plasmid for Six1 and the anti-Six1 antibody were provided by Dr. Hisato Kondoh, Osaka University (Osaka, Japan). Expression plasmids for E47 and E12 were provided by Dr. Cornelis Murre, University of California, San Diego (La Jolla, CA). Anti-E47 and anti-E12 antibodies were purchased from Santa Cruz Biotechnology (Santa Cruz, CA). The p65 luc to amplify exon regions of the dual-luciferase reporter assay system were purchased from Promega (Madison, WI). Marathon-Ready cDNA and Advantage cDNA PCR kits were purchased from Clontech (Palo Alto, CA). Superscript II reverse transcriptase, the TOPO cloning kits, the double-stranded DNA cycling system, Taq polymerase, cell culture media, restriction enzymes, and all other molecular biology reagents were purchased from Invitrogen.

**Cell Culture**—BEAS-2B, A549, and HepG2 cells were obtained from American Type Culture Collection (ATCC, Manassas, VA). BEAS-2B cells were cultured in the serum-free medium, LHC-9 (Biofluids, Rockville, MD). For sub-culturing, cells were trypsin-dissociated and plated onto fibronectin/collagen-coated culture plates (22). A549 cells were maintained with Dulbecco’s modified Eagle’s medium/nutrient mixture F12 containing 10% fetal bovine serum. HepG2 cells were cultured on Dulbecco’s modified Eagle’s medium supplemented with 10% fetal bovine serum and 10 mM sodium pyruvate.

**Cloning and Sequencing of the CYP2F1 Gene**—Cosmid clones F15749, R20003, F16767, R28845, and E7850 (average insert size of 40 kb) that hybridized to CYP2F1 (23) were cleaved with HindIII for Southern blot verification of CYP2F1 using a 298-bp EcoRI/KpnI pUC9-2F1 cDNA fragment as a probe. Positive clones were subsequently sequenced (see below). In addition, BAC clones were identified by screening the CIBT human BAC DNA library (Research Genetics, Huntsville, AL) using “whole cell” PCR (24) with primers designed from the 3‘-flanking region of CYP2F1. The CIBT BAC library represents a 13–17% coverage of the human genome, and BAC clones contain an average insert size of 100–150 kb. Positive BAC clones identified during the library screening and others purchased based on sequence comparisons with the human genome had addresses that corresponded to the following clones: CIBT-HSP-D 2356916 (GenBank accession number AC008892), CIBT-HSP-C 49021 (GenBank accession number AC008537), and CIBT-HSP-D 2451510. All positive clones were sequenced using primers designed based on putative CYP2F1 exonic sequences. All sequencing was performed at the University of Utah core sequencing facility using fluorescent DNA sequencing methods and automated ABI 377 (Applied Biosystems) sequencers.

**Sequencing of CYP2F1 cDNA**—The vector pUC9-2F1 (3) was sequenced to confirm the nucleotide sequence of the CYP2F1 cDNA. First strand cDNA synthesis for reverse transcription-coupled PCR (RT-PCR) was performed using Superscript II reverse transcriptase (Invitrogen) and 5 μg of total cellular RNA isolated from confluent BEAS-2B cells using RNeasy kit (Qiagen) and a Qiashredder microspin homogenizer according to the manufacturer’s recommendations. PCR amplification utilized primers designed to amplify 1728 bp, which spans the entire coding region. The primers were 5′-GCC TCC CAG CAG GTC TTC CTC C-3′ (sense) and 5′-GAA AAG GCC GGC GTC CCA TAG AAC AAG-3′ (antisense), designed for selective binding to the CYP2F1 cDNA sequence (GenBank accession number J02906) using the Oligo 5.0 program (Molecular Biology Insights, Cascade, CO). The PCR products were performed using 2 μl of cDNA reaction, 2.5 units of Pfu DNA polymerase (Invitrogen), 5 μl of 10× PCR buffer, 1.5 μl of 50 mM MglCl2, 1 μl of 10 μM dNTP mix, 0.2 μM each primer, and water to a final volume of 50 μl. PCR conditions were to denature at 94 °C for 3 min, followed by 30 cycles of melting at 94 °C for 1 min, annealing at 55 °C for 1 min, and a 10-min final extension. The PCR product was cloned using Zero Blunt TOPO PCR Cloning Kit (Invitrogen) for sequencing. Clones were sequenced on both strands to verify the sequence of the cDNAs.

**Primer Extension Analysis and 5′-RACE**—To identify the transcription start site (TSS) of CYP2F1, primer extension analysis and rapid amplification of cDNA ends (5′-RACE) were performed. Donor human lung tissue was used to isolate total RNA using an RNeasy miniprep kit (Qiagen). Total RNA (5 μg) was hybridized to a 5′-P-end-labeled primer (5′-GGC TGT GCT TAT GCT GTC CAT-3′) and extended using Superscript II reverse transcriptase (as above). The cDNA product was denatured and analyzed by electrophoretic fractionation on an 8% polyacrylamide gel along with a sequence ladder generated using the same primer and a 10-μb DNA standard ladder. The TSS was determined by comparing the size and sequence of the product with the sequence of the CYP2F1 gene. Human lung Marathon-Ready cDNA (Clontech) was used for 5′-RACE analysis. Briefly, PCR amplification was performed using the gene-specific primer (GSP1) 5′-CAG GAO ACA GTG GCT GGG ATOG-3′, the nested gene-specific primer (GSP2) 5′-GGC TGT GCT TAT GCT GTC CAT-3′, and an adapter primer (AP1) according to the manufacturer’s recommendations. The 5′-RACE products were cloned into the pCR2.1 TOPO-TA vector (Invitrogen) for sequencing using the m13 forward and m13 reverse primers. The resulting sequence was compared with the sequence of the CYP2F1 gene obtained from the CITB clone (CITB-HSP-C 490E21), which contains the CYP2F1 gene. In addition, analysis of the 5′-upstream region was performed using MatInspector version 2.2 and the TRANSFAC 4.0 matrices to identify putative transcription regulatory-binding motifs (25).

**Transient Transfection Studies**—Luciferase reporter assays were performed to identify functional promoter regions. CYP2F1 reporter constructs were produced using PCR amplification with multiple primers that introduced a 5′-SacI at positions 1681, 1468, 1299, 1168, 992, 893, 748, 493, and 129 paired with a single 3′-antisense primer that generated a 3′-BglII restriction site at position +115. After digestion with SacII/BglII, the fragment was cloned into the pGL3 Basic vector. Reporter constructs were transfected into 96-well plates. Cells were lysed 36 h post-transfection, and luciferase activities were assayed using the dual luciferase assay (Promega). Firefly luciferase activity was normalized for transfection efficiency using Renilla luciferase activity (pGL-SV40) and expressed as fold luminescence over the activity observed with the promoter-less pGL3 basic vector. The data are presented as mean fold luminescence ± S.E. for the independent experiments performed in triplicate.

**Preparation of Nuclear Extracts**—Human lung nuclear extracts were prepared using a combination of the protocols described by Ueno and Gonzalez (26) and Dignam (27). All solutions were at 4 °C throughout the procedure and contained a 1:1000 dilution of a protease inhibitor mixture solution (1 mg/ml 4-(2-aminoethyl)benzenesulfonyl fluoride, 0.5 mg/ml aprotonin, 1 mg/ml leupeptin, 1.2 mg/ml bestatin, 1 mg/ml e-aminocaproic acid, 0.5 mg/ml EDTA, 2 μg/ml RNasin, and 0.1 mg/ml aprotinin). The nuclear pool was suspended in 100 ml of homogenization buffer (10 mM HEPES, pH 7.6, 25 mM KCl, 0.15 mM spermine, 0.5 mM spermidine, 1 mM EDTA, 2 μM succrose, 10% glycerol) and homogenized using a motor-driven Teflon glass homogenizer. Tissue homogenate was layered over four 10-ml
cushions of the same buffer and centrifuged for 30 min at 24,000 rpm and 4 °C in an SW28 rotor (Beckman Instruments, Fullerton, CA). After discarding the supernatant fraction, the nuclear pellets were combined, washed with a low salt buffer (20 mM HEPES, pH 7.9, 25% glycerol, 1.5 mM MgCl₂, 0.02 M KCl, 0.2 mM EDTA, 0.2 mM phenylmethylsulfonyl fluoride, and 0.5 mM dithiothreitol), and collected by centrifugation at 3,300 × g for 10 min. After determining the volume of the nuclear pellet, it was resuspended in a 1:1 (v/v) ratio with the same low salt buffer. The nuclear proteins were extracted by addition of an equal volume of the same buffer, adjusted to 1.2 M KCl. The resulting suspension was shaken gently on ice for 30 min and centrifuged in a Ti 50 rotor (Beckman Instruments) at 18,000 rpm for 60 min at 4 °C. The supernatant fraction was dialyzed against 500 volumes of 20 mM HEPES, pH 7.9, 20% glycerol, 0.2 mM EDTA, 100 mM KCl, 0.2 mM phenylmethylsulfonyl fluoride, and 0.5 mM dithiothreitol for 4 h on ice to reduce the KCl concentration to 100 mM. After dialysis, the extract was centrifuged for 20 min at 18,000 rpm to remove precipitated material. The supernatant fraction was aliquoted into working volumes, flash-frozen, and stored at −80 °C. Additional human lung, liver, and heart nuclear extracts were purchased from Geneka Biotechnology (Montreal, Canada). Nuclear extracts from cultured cells were prepared as described by Dignam et al. (27) with addition of the protease mixture solution. Protein concentrations of all nuclear extracts were calculated using the Bio-Rad Protein Assay Kit I.

**DNase Footprinting Assay**—Interactions of 5’ CYP2F1 upstream sequence (position −1,681 to +115) with human lung nuclear proteins were identified using the Core Footprinting System (Promega) with slight modification (20). Binding reaction mixtures (50 µl), which were preincubated on ice for 10 min, contained 20–30 µg of human lung nuclear extract in 10 mM Tris-HCl buffer, pH 7.9, 50 mM NaCl, 1 mM dithiothreitol, 1 mM EDTA, 8% glycerol, 20 mM KCl, 3.5 mM MgCl₂, 7 mM ZnSO₄, and 32P-end-labeled DNA fragments (2 ng, 40,000–50,000 cpm). An equal volume of DNase reaction buffer (10 mM MgCl₂ and 5 mM CaCl₂) was added and incubated at room temperature for 1 min. DNase I was added (0.3 units to probe without protein and 3 units to probe with protein), and the DNase digestion was allowed to proceed for 90–120 s at room temperature. The reactions were terminated by the addition of 100 µl of DNase stop solution (200 mM NaCl, 30 mM EDTA, 1% SDS, 100 µg/ml yeast RNA) provided in the Core Footprinting System. DNA was recovered from the reaction mixture by phenol/chloroform extraction and ethanol precipitation. Precipitated DNA was collected by centrifugation, washed with cold 70% ethanol, air-dried, and resuspended in 10 µl of gel dye loading buffer (1:2 NaOH/formamide (v/v), 0.1% xylene cyanol, 0.1% bromphenol blue). The DNA samples were denatured at 90 °C for 4 min, quickly placed on ice, and analyzed by electrophoretic fractionation on a 6% polyacrylamide, 7 M urea DNA gel.
sequencing gel. DNA sequence ladders of the same DNA were prepared by the method of Maxam and Gilbert (28) and fractionated simultaneously for identification of the DNA-protein-binding site.

Electrophoretic Mobility Shift Assays (EMSA)—EMSA was performed using the gel shift assay system from Promega, essentially as described by the manufacturer. Binding reaction mixtures were preincubated at room temperature for 10 min. The mixtures contained 4 μl of nuclear extract (4 μg for lung tissue and 6 μg for cells) and 2 μl of 5× binding buffer (50 mm Tris-HCl, pH 7.5, 250 mm NaCl, 5 mm MgCl₂, 2.5 mm EDTA, 2.5 mm dithiothreitol, 20% glycerol, 0.25 mg/ml poly(dI-dC)poly(dI-dC)) in a total volume of 10 μl. For competition experiments, a 100-fold molar excess of unlabeled double-stranded oligonucleotide was incubated for 15 min with nuclear extract prior to the addition of 1 μl of 32P-labeled double-stranded probe (0.005–0.01 pmol). The mixtures were incubated for another 20 min at room temperature. Before electrophoresis, gel loading dye (25 mm Tris-HCl, pH 7.5, 0.02% bromphenol blue, 4% glycerol) was added to all binding reaction mixtures. The DNA-protein complexes and unbound probes were separated by electrophoresis using 4% polyacrylamide gels and detected by autoradiography. OCT1, AP1, Sp1, and NFκB double-stranded oligonucleotides were included in the gel shift assay system (Promega). All other oligonucleotide probes were synthesized by Integrated DNA Technologies (Coralville, IA). Sequences of the synthesized DNA probes are listed in Table I. EMSA supershift assays were performed with antibodies to AP1, and the E box factors E47, E12, and αEF1 as described previously (29).

RESULTS

Cloning and Nucleotide Sequence of the Human CYP2F1 Gene—Library screening and sequencing was used to identify one full-length CYP2F1 gene and one pseudogene. Sequencing was performed by walking upstream and downstream from within each exon, using primers based on the published cDNA sequence (GenBank accession number J02906). Three cosmid clones, F16767, F15749, R20003, and one BAC clone, C1TB-HSP-D 2356P16, were used to sequence the full-length CYP2F1 gene. Two other cosmid clones (ED7850 and R28543) and the two BAC clones (C1TB-HSP-C 490E21 and C1TB-HSP-D 2415110) all appeared to contain a pseudo CYP2F1 gene, missing exons 1–4, with mutated exons 5–10, and were not used for further sequencing. Sequencing of CYP2F1 revealed a disparity from what was originally identified as the genomic localization of CYP2F1 (23). Two functional and one pseudo CYP2F1 gene loci were proposed, but our sequencing identified two genes with identity to the CYP2F1 cDNA, one with an incomplete sequence at the centromeric end of the CYP2 family gene cluster (middle of intron 4 through exon 10, CYP2F1P1) and the full-length sequence at the telomeric end of the cluster (CYP2F1) in opposite orientations. This information has subsequently been confirmed by the completion of the human genome and reconstruction of the CYP2 family gene cluster (30).

The full-length CYP2F1 gene was deduced and found to span ~14 kbp and contain 10 exons (Fig. 1). The CYP2F1 gene structure, which contains 10 exons and 9 introns, is unique from all genes of the CYP2 family, which have been shown to contain 9 exons and 8 introns (31). The structure described here recognizes an additional 5′-untranslated (UTR), exon 1, separated from exon 2 by 1685 bp, which was missing from the CYP2F1 gene structures reported earlier (3, 32). Earlier reports concluded that CYP2F1 contained nine exons, but the results presented here demonstrate that the previously reported exon 1 is actually composed of exons 1 and 2, separated by a 1685-bp intron. The cDNA sequence contained 56 bp of 5′-UTR sequence spanning exon 1 and the first 11 bp of exon 2. Similar gene structures, containing a 5′-UTR exon 1, were also observed for CYP2F genes in mouse and rat (NCBI, National Center for Biotechnology Information, genome resources).

Sequence of the Human CYP2F1 cDNA—Sequencing of the CYP2F1 gene revealed several differences from the published cDNA sequence. Therefore, we sequenced clone pUC9-2F1 (used for submission of the sequence, GenBank accession number J02906) along with an RT-PCR product that was obtained from BEAS-2B lung epithelial cells. Comparison of pUC9-2F1 and the sequence obtained from the RT-PCR product confirmed the sequence obtained from the genomic clones, indicating that there were several errors in the original published cDNA sequence. These differences were confirmed by the “curated RefSeq project” at NCBI with submission of CYP2F1, accession number NM_0007744 (33). The correct cDNA sequence is illustrated in Fig. 2A, with the differences in the predicted amino acid sequence illustrated in Fig. 2B. The variant CYP2F1 sequence identified by Nhamburo et al. (3) was not observed in RNA isolated from either BEAS-2B cells or human lung tissue. It was originally proposed that the variant was a product of a pseudogene in the CYP2 cluster. However, the “variant” would appear to be a product of alternate CYP2F1 splicing, because the incomplete gene at the centromeric region of the gene cluster does not contain exons 1–4, and other CYP2F subfamily genes have not been identified.

Identification of the Transcriptional Start Site of CYP2F1—Primer extension analysis and 5′-RACE were used to identify TSSs with mRNA isolated from human lung tissue (Fig. 3) and the BEAS-2B cell line (not shown). The positions of the primer extension-determined TSSs were revealed by comparison of the reverse transcriptase reaction product (primer extension) to a
to regulate differentiation in several tissues and are implicated in regulating pulmonary selective gene expression (35, 36). In addition, several other putative cis-acting elements that have been implicated in regulating surfactant proteins in lung epithelial cells and other genes during lung development (34). Other sites in this region included binding sites for the C/EBP proteins, which are known to regulate pulmonary selective gene expression (35, 36). In addition, several other putative cis-acting elements that have been implicated in regulating constitutive gene expression were also identified (Sp1 and AP2).

Transcriptional Activity of the 5′-Region of CYP2F1—To demonstrate the functionality of the promoter, without potential upstream elements, the −129 to +115 CYP2F1 fragment was cloned into the pGLE vector that contains a strong SV40 enhancer element but no functional promoter (pGLE-129). The transcriptional activity of this construct was assayed in BEAS-2B cells, A549, and HepG2 cells (data not shown). The CYP2F1 reporter constructs demonstrated reasonable functional transcriptional activity in both lung epithelial cell lines but essentially no activity in the human liver HepG2 cell line, despite relatively high HepG2 transfection efficiency (>50%) compared with the BEAS-2B cell line (Fig. 5A). Surprisingly, the CYP2F1-directed reporter activity in the BEAS-2B cells that contain CYP2F1 was slightly lower and more variable than the activity observed in the A549 cells. The cellular variability may have been due to the low transfection efficiency (<15%) observed with BEAS-2B cells compared with the relatively higher efficiency (>40%) observed with the A549 cells. Maximal reporter activity in A549 cells was observed with the construct containing CYP2F1 position −1681 to +115 sequence directing luciferase expression. However, maximal reporter activity in BEAS-2B cells was observed with the construct containing −1168 to +115. Removing the sequence +1 to +115 in several constructs had little to no effect on directing luciferase activity in BEAS-2B cells (data not shown).

To demonstrate the functionality of the promoter, without potential upstream elements, the −129 to +115 CYP2F1 fragment was cloned into the pGLE vector that contains a strong SV40 enhancer element but no functional promoter (pGLE-129). The transcriptional activity of this construct was assayed in BEAS-2B, A549, and HepG2 cells (Fig. 5B). Significant activity was observed in all three cell lines, suggesting the necessary promoter elements for basal activity were present within the first 129 bp upstream of the TSS. Interestingly, the HepG2 cells showed a 2-fold increase in activity, which sug-

sequencing reaction performed using the same primer and a radiolabeled 10-bp DNA ladder (L). This analysis identified two putative TSSs at positions −1781 and −1741 relative to the A of the ATG start codon. Sequence analysis of five clones containing 5′-RACE fragments revealed a single TSS corresponding to the −1741 site. Based on both of these results, we chose to assign the CYP2F1 +1 position to the −1741 site that was identified by both approaches and also agrees with the reported cDNA sequence (GenBank™ accession number NM_0007744).

Nucleotide Sequence Analysis of the 5′-Region of CYP2F1—To identify possible transcription regulators, the sequence upstream of the major TSS was analyzed using the MatInspector version 2.2 and the TRANSFAC matrices 4.0. The major TSSs are marked with right-angled arrows. Exon 1 of the CYP2F1 gene is italicized and underlined. All putative transcription factor-binding sites are highlighted in gray and labeled above the sequence. The 31-bp LSF-binding site in the same 3′-most nucleotide sequence position of the DNA fragments used for construction of the luciferase reporter constructs. All constructs had the same 3′-most nucleotide sequence (→).

Fig. 4. Possible known transcription factor-binding elements of CYP2F1. The sequence from −1775 to +125 relative to the major TSS was analyzed by MatInspector version 2.2 using TRANSFAC matrices 4.0. The major TSSs are marked with right-angled arrows. Exon 1 of the CYP2F1 gene is italicized and underlined. All putative transcription factor-binding sites are highlighted in gray and labeled above the sequence. The 31-bp LSF-binding site is in boldface and underlined. The → indicates the 5′-most nucleotide sequence positions of the DNA fragments used for construction of the luciferase reporter constructs. All constructs had the same 3′-most nucleotide sequence (→).
gested that, in the presence of a strong transcriptional signal, the elements present in this minor regulatory region are sufficient to recruit basal transcriptional machinery, even in liver cells.

**Identification of Lung-specific Binding Factor(s) by DNase I Footprinting**—To investigate potential DNA-protein-binding sites involved in CYP2F1 promoter activity, DNase I footprint analysis was performed using nuclear extracts from human lung tissues. Sequential overlapping DNA fragments of 200–400 bp from position −1681 to +115 were amplified or digested from the largest reporter construct, and 32P-end-labeled on...
either the sense or antisense strands and utilized for DNase I footprinting. Interestingly, only one strong DNA-protein complex (position −182 to −152) was observed (Fig. 6A) over the entire region. This result was confirmed when the opposite strand was similarly analyzed for DNase I protection (Fig. 6B).

Only limited amounts (30 μg/reaction) of nuclear extracts were used, due to the low yields of nuclear protein from human lung tissue preparations. This may explain why only one strongly protected site was observed. The protected site, which is named LSF for lung-specific factor-binding site, is 31 bp long and located 152 bp from the TSS. Analysis of the region using the TRANSFAC data base revealed that the binding site partially included a potential Sp1 site (89.8% similar) and contained two E box-binding elements separated by 2 bp (Fig. 6C). The E box sites scored 94.3% similarity to the human ATPase 1α1 regulatory element-binding protein 6 DNA-binding site, 93.6% to δEF1, 97.3% to myogenic differentiation factor, and 95.3% similarity to AP4. Another site identified, unrelated to E box domains, was NFAT (93.1%).

Investigation of the Binding Specificity of Lung Nuclear Extracts for LSF-binding Motif Using EMSA—The binding specificity of protein(s) to a radiolabeled LSF probe (Table I) was investigated by EMSA analysis using nuclear extracts isolated from lung tissues. Two DNA-protein complexes were observed using 4 μg of lung nuclear extract (Fig. 7A) and were abolished by co-incubation with a 100-fold molar excess of unlabeled LSF probe (Fig. 7B). Further confirmation of the specificity of these DNA-protein complexes was demonstrated by the inability of a 100-fold molar excess of unlabeled OCT1 probe to abolish binding. Tissue-specific binding of the 31-bp LSF radiolabeled probe was demonstrated using 10–12 μg of commercially available nuclear extracts prepared from human lung, liver, and heart.

![Diagram](image_url)
The potential of known trans-acting factors to bind to the LSF motif was investigated with competitive oligonucleotides representing known AP4, NFAT, and E box-binding elements (Table I). The mouse MCK E box consensus sequence (37), which has two E box sites separated by 11 bp versus 2 bp for LSF, strongly competed for protein(s) that bound radiolabeled LSF. Likewise the immunoglobulin κ enhancer element (κE2) containing an E box consensus sequence demonstrated strong competition with LSF. Interestingly, competitive oligonucleotides designed from the human MLC1/3 enhancer (MLC) (38) and the brachyury gene, which both contain E box sites, did not compete with LSF. The oligonucleotides for the MCK, κE2, and brachury E box promoter elements were designed analogously to competition studies that were used to characterize δ-crystallin enhancer-binding protein (δE1F1) (39). In addition, combination with the oligonucleotides containing the NFAT consensus sequence from human aldose reductase promoter (40) or AP4 consensus sequence from human proenkephalin promoter (41) did not compete, suggesting that these factors may not be important. EMSA supershift assays were performed using antibodies to AP4 and the E box specific factors E47/E12, SIP1, a member of δE1F1 family that binds to CACCT motifs (42), and δE1F1 (data not shown). None of the antibodies were able to supershift the LSF complexes with lung nuclear proteins. However, when nuclear extracts were enriched by overexpression (cotransfection) of AP4, E47, E12, or δE1F1 proteins, only the AP4-enriched extracts produced a new protein complex, which was shifted by anti-AP4 (data not shown).

A luciferase reporter construct containing the pGLP-LSF (−182 to −152) was generated to investigate the function of the LSF-binding site in BEAS-2B and HepG2 cells (data not shown). The reporter activity of pGLP, in both cell lines, was not affected by introduction of the LSF-binding motif. This finding was surprising because BEAS-2B cell extracts appeared to contain the LSF. In addition, cotransfection of AP4, E47, E12, and δE1F1, with pGLP-LSF, had no effect on reporter activity compared with cotransfection with pGLP alone.

**DISCUSSION**

CYP2F1 is expressed primarily in lung tissues with little or no expression in hepatic or other extrapulmonary tissues (3, 13, 43). The expression of CYP2F1 has been implicated in the tissue-specific toxicity associated with many pulmonary toxicants including styrene, 3-methylindole, naphthalene, and benzene (6, 8, 44). All of these compounds are environmental and occupational toxicants found in sources such as cigarette smoke, gasoline, and industrial by-products. Understanding the transcriptional regulation of cytochromes P450, which are expressed in tissues that are directly exposed to environmental toxicants, such as the lung, may help predict the susceptibility of an individual to acute toxicities or chemically induced cancer. Despite the importance of pulmonary transcriptional regulation of cytochromes P450, few mechanisms of tissue-specific regulation have been identified (2). Understanding the transcriptional regulation of CYP2F1, which is uniquely expressed in pulmonary tissues, should provide vital information about organ-selective transcription regulatory mechanisms and pulmonary specific bioactivation. Therefore, the primary objectives of the studies described herein were to characterize the CYP2F1 gene and identify specific regulatory motifs and trans-activating factors that may be involved in its expression.
To identify the factors that regulate CYP2F1 transcription in lung cells, it was necessary to clone the gene and sequence its regulatory region. Sequencing of an amplified cDNA product from CYP2F1 mRNA revealed several sequence variations from the published cDNA. It was concluded from sequence analysis that the original published cDNA had several sequencing errors, which were confirmed by sequencing the pUC9-2F1 clone obtained from the original authors (3). The corrected sequence has been updated in GenBank™ as accession number NM_000774. This information may be important for protein modeling studies that depend on correct protein sequences.

Genomic library screening identified a single BAC clone that contained the entire CYP2F1 gene, which spans 14 kbp and contains 10 exons, exon 1 encoding a 5'-UTR. Primer extension analysis was used to identify two major TSSs at positions −1781 and −1741 relative to the ATG start codon in exon 2 and several minor sites upstream of the first exon. However, 5'-RACE only identified the site at position −1741, which is in agreement with the full-length sequence reported in NM_000774. As such, this site was designated as +1. The promoter region of CYP2F1 does not contain a TATA box, but the sequence proximal to the TSS (+1) contains a putative Sp1 site and pyrimidine-rich (transcription initiation element) element, possibly facilitating initiation (45). The promoter also contains potential binding sites for several other well-characterized transcription factors, such as C/EBP. There is substantial evidence that the CYP2B enzyme in pulmonary cells is regulated by C/EBPa and C/EBPβ (46). A 1.3-kbp sequence from the CYP2B promoter was shown to drive lung-specific reporter gene expression in a transgenic mice, and it was shown that C/EBPa and C/EBPβ regulate CYP2B pulmonary expression during differentiation (19). In addition, the promoter contains a BET site that has been demonstrated in numerous studies (35, 36, 47) to regulate CYP1A1 transcription. The BET site of CYP1A1 has been recognized to bind a number of transcription factors from the Sp1/KLF family of factors, which are involved in both constitutive and tissue-specific regulation (48). The functional significance of these binding sites was consistent with the reporter assays that demonstrated promoter activity with the CYP2F1 fragment, position −129 to +115.

CYP2F1 tissue-specific expression was demonstrated using luciferase reporter constructs containing up to −1.6 kbp of 5'-flanking region in two lung epithelial cell lines but not a liver cell line. When the sequence up to −129 bp was inserted into a vector containing a strong SV40 enhancer element, transcription occurred in both lung and liver cells indicating that the region indeed contained the minimal promoter, with the elements necessary for transcriptional initiation. The minimal promoter also showed tissue specificity for lung cells. Although the −129 to +115 fragment appears to contain the minimal promoter, additional experiments are required to identify the cis-elements responsible for the promoter activity. Thus the minimal promoter may be shorter than this fragment. The transient transfection studies demonstrated tissue-specific expression of CYP2F1 in BEAS-2B cells; however, they were not useful in demonstrating functionality of the LSF-binding motif, despite that BEAS-2B cell extracts contained the LSF. An explanation for this result is that LSF in BEAS-2B cells requires another transcription factor-binding site or an additional cofactor protein for activity. We are currently investigating additional basal and tissue-specific binding motifs within this region.

To identify protein-binding sites present within the first 1.6 kbp of 5'-flanking region of the CYP2F1 gene, we performed DNase I footprinting assays using nuclear extracts from human lung tissues. By scanning the 5'-flanking region, we discovered only one strong DNA-binding site at positions −182 to −152, which we termed LSF. Interestingly, the location of this fragment was upstream of the minimal promoter that drove promoter activity in both lung cell lines. Additional DNase I protection regions undoubtedly exist, which could have been identified with higher amounts of nuclear proteins than we could reasonably obtain. DNase I-hypersensitive sites were also observed, which are strong indicators of protein binding-induced conformation changes in DNA structure.

Two DNA-protein binding complexes were observed with the LSF motif using lung nuclear extracts and EMSA analysis in the presence or absence of specific or nonspecific oligonucleotides. Tissue specificity was demonstrated by complex formation with lung nuclear extracts but not liver or heart nuclear extracts. This specificity is consistent with CYP2F1 lung-specific expression. Nuclear extracts from BEAS-2B cells also produced two bands. Surprisingly, no complexes were observed with nuclear extracts from A549 cells despite the observation...
that reporter activity was driven by the CYP2F1 promoter. This difference may explain the lack of CYP2F1 mRNA detected in A549 cells in contrast to the low but detectable expression of CYP2F1 mRNA in BEAS-2B cells. No complexes were observed with HepG2 cell nuclear extract, which is consistent with the lung-specific binding observed with nuclear extracts from human tissues. We defined the core consensus sequence of the LSF-binding site as 5'-CCACGGGACCTTTC-CAGCT-3' (underlined) using mutant oligonucleotide competitors and EMSA with lung nuclear extracts. Analysis of the protected region revealed two putative E box sites, separated by 2 bp.

The consensus sequence of the E box sites showed a high degree of similarity to the binding sites for ATPase 1x1 regulatory element-binding protein 6 (49), SIP1 (Smad-interacting protein 1) (42), myogenic differentiation factor 37, and ESE1 (39). However, there was also considerable similarity to NFAT (40) and AP4 (41). Additional competitive EMSA analyses using oligonucleotides that bind NFAT, AP4, and E box factors were investigated. Only the oligonucleotides that contained E box-binding sites from MCK and E2 (39) competed with LSF binding, suggesting that LSF may be an E box binding transcription factor. Although the lack of competitive binding with the other E box-binding sites may suggest that additional nucleotide interactions, other than the E box core CA(C/G)CT/T G), or additional cofactors are required. It has been shown that two adjacent E box sites are required for the binding of E box factors, SIP1 and ESE1 (50), with weaker binding observed with 3 versus 24 bp spacing between the E boxes. Both LSF and MCK have adjacent E box sites, but the spacing is quite different, 2 bp in LSF and 11 bp in MCK, yet the MCK element was a strong competitor. However, the E2 oligonucleotide, which contains only one E box site, also efficiently abolished LSF binding. Thus, these studies were not informative with respect to the potential need for adjacent E box sites for LSF binding or the possible spacing between these sites if they are required. By using competition studies, we could not conclude whether the established E box factors bind to the LSF sequence. Therefore, EMSA supershift analysis with antibodies directed against two major E box binding factors (E47 and E12) were performed. Neither anti-E47 nor anti-E12 bound to the LSF complex using nuclear lung extracts or extracts from cells that were enriched with E47 or E12 proteins. Similarly, supershift assays that were designed to identify SIP1 and ESE1 showed no reactivity, even with ESE1-enriched extracts.

Due to the high similarity of the consensus AP4 factor-binding site within the 5'-upstream region, and despite the lack of competition of AP4 oligonucleotides with the LSF complex, supershift assays with anti-AP4 were performed. The anti-AP4 antibody was unable to supershift the LSF-binding complexes with lung nuclear extracts. However, nuclear extracts enriched by overexpression of AP4 in the cells produced DNA/protein bands with different mobilities than those observed from normal cell extracts, and the new AP4-LSF complexes were supershifted by anti-AP4. Therefore, AP4 is capable of binding to the LSF-binding sequence when enriched in nuclear extracts, but this result was expected, given the high degree of similarity of the AP4 consensus site within LSF. In fact, recent evidence has shown that the AP4 interacts with the immunoglobulin κ promoter E box of the E47/E12 type with higher affinity than E47, suggesting that AP4 is the major ligand for Ig-κ promoter E boxes (29). Although AP4 is thought to be a ubiquitous transcription factor, we are conducting additional studies to compare the specificity of AP4 and other E box-binding factors for LSF, and to relate multiple factor complexes in the tissue-specific expression of CYP2F1.

These studies provide specific structural characterization of the unique CYP2F1 gene and identification of a tissue-specific promoter. A novel nuclear factor-binding site, which may regulate the selective expression of CYP2F1 in human lung tissue, was also identified. The nuclear protein(s) (LSF) that binds to this domain appears to be a protein uncharacterized previously but may belong to the E box family of transcription factors. The functional consequence of LSF binding to the CYP2F1 promoter was investigated but not elucidated in these studies. Ongoing studies to characterize the LSF binding activity include the following: finding an improved cell culture model, in vitro transcription studies using human lung extracts, and isolation of LSF for identification using mass spectrometry. Characterization of additional regulatory elements of the CYP2F1 gene should yield insight into the biochemical mechanisms of P450 gene expression and the tissue-selective expression of CYP2F1 in human lung.

REFERENCES

1. Guengerich, F. P. (1993) Drug Metab. Dispos. 21, 1–6
2. Ding, X., and Kaminsky, L. S. (2003) Annu. Rev. Pharmacol. Toxicol. 43, 169–173
3. Nhambruto, P. T., Kimura, S., McBride, O. W., Koak, C. A., Gelboin, H. V., and Gonzalez, F. J. (1990) Biochemistry 29, 5491–5499
4. Thornton-Manning, J., Appleton, M. L., Gonzalez, F. J., and Yost, G. S. (1996) J. Biol. Chem. Exp. Ther. 276, 21–29
5. Lanza, D. L., and Yost, G. S. (2001) Drug Metab. Dispos. 29, 950–953
6. Lanza, D. L., Code, E., Crespi, C. L., Gelboin, H. V., and Yost, G. S. (1999) Drug Metab. Dispos. 27, 798–803
7. Yost, G. S. (2001) Adv. Exp. Med. Biol. 500, 53–62
8. Nakajima, T., Elovaa, E., Gonzalez, F. J., Gelboin, H. V., Raunio, H., Pelkonen, O., Vainio, H., and Aoyama, T. (1994) Chem. Res. Toxicol. 7, 891–896
9. Willey, J. C., Coy, E. L., Frampton, M. W., Torres, A., Apostolakos, M. J., Hoehn, G., Schuermann, W. H., Thilly, W. G., Olson, D. E., Hammerle, J. R., Crespi, C. L., and Uettl, M. J. (1997) Am. J. Respir. Cell Mol. Biol. 17, 114–124
10. Rivera, S. P., Saarikoski, S. T., and Hankinson, O. (2002) Mol. Pharmacol. 61, 255–259
11. Anttila, S., Hukkanen, J., Hakkoila, J., Stjernvall, T., Beaune, P., Edwards, R. J., Bobis, A. R., Pelkonen, O., and Raunio, H. (1997) Am. J. Respir. Cell Mol. Biol. 16, 242–249
12. Ceswinski, M., McMenemy, T. L., Gelboin, H. V., and Gonzalez, F. J. (1994) Cancer Res. 54, 1085–1091
13. Willey, J. C., Coy, E., Brelly, C., Uettl, M. J., Frampton, M. W., Hammerle, J., Thilly, W. G., Olson, D., and Cairns, K. (1996) Am. J. Respir. Cell Mol. Biol. 14, 262–271
14. Busters, J. T., Sakai, S., Richter, T., Pineau, T., Alexander, D. L., Savas, U., Doehner, J., Ward, J. M., Jefcoat, C. R., and Gonzalez, F. J. (1999) Proc. Natl. Acad. Sci. U. S. A. 96, 1977–1982
15. Fipari, R., Savela, K., Nurminen, T., Hukkanen, J., Raunio, H., Hakkoila, J., Mantyla, T., Beaune, P., Edwards, R. J., Bobis, A. R., and Anttila, S. (2000) Int. J. Cancer 86, 610–616
16. Smith, T. J., Guo, Z., Gonzalez, F. J., Guengerich, F. P., Stoner, G. D., and Yang, C. S. (1992) Cancer Res. 52, 1757–1763
17. Yost, G. S. (1996) Adv. Exp. Med. Biol. 387, 221–229
18. Sikarin, T., Becher, R., Bucht, A., Duvefelt, K., Bohm, S., Ranneberg-Nilsen, T., Lilleaas, E. M., Schwarze, P. E., and Toftgard, R. (1999) Am. J. Respir. Cell Mol. Biol. 21, 177–184
19. Cassel, T. N., Gustafsson, J. A., and Nord, M. (2000) Mol. Cell Biol. Rev. Commun. 3, 42–47
20. Zhang, J., and Ding, X. (1998) J. Biol. Chem. 273, 23454–23462
21. Zhang, J., Zhang, G. Y., Guo, J., Zhou, Y., and Ding, X. (2000) J. Biol. Chem. 275, 8895–8892
22. Lechner, J. F., and LaVeck, M. A. (1985) J. Tissue Cult. Methods 9, 43–48
23. Hoffman, S. M., Fernandez-Salguero, P., Gonzalez, F. J., and Mohrenweiser, H. W. (1995) J. Mol. Evol. 41, 894–900
24. Kim, U. J., Birren, B. W., Slepak, T., Mancino, V., Boysen, C., Fernandez-Salguero, P., Mancino, V., Boysen, C., and Utell, M. J. (1997) Adv. Exp. Med. Biol. 429, 413–418
25. Hoffman, S. M., Nelson, D. R., and Keeney, D. S. (2001) Proc. Natl. Acad. Sci. U. S. A. 98, 1475–1480
26. Lanza, D. L., and Yost, G. S. (2001) Drug Metab. Dispos. 29, 281–283
27. Ueno, T., and Gonzalez, F. J. (1990) Mol. Cell. Biol. 10, 4485–4495
28. Hoffman, S. M., Nelson, D. R., and Keeney, D. S. (2001) Pharmacogenetics 11, 697–698
29. Nebert, D. W., and Gonzalez, F. J. (1987) Annu. Rev. Biochem. 56, 945–993
32. Chen, N., Whitehead, S. E., Caillat, A. W., Gavit, K., Jephورد, D. R., Kovacevic, D., McCreary, M. B., and Hoffman, S. M. (2002) *Mutat. Res.* **499**, 155–161.

33. Pruitt, K. D., and Maglott, D. R. (2001) *Nucleic Acids Res.* **29**, 137–140.

34. Costa, R. H., Kalinichenko, V. V., and Lim, L. (2001) *Am. J. Physiol.* **280**, L823–L838.

35. Kaczynski, J. A., Conley, A. A., Fernandez Zapico, M., Delgado, S. M., Zhang, J. S., and Urrutia, R. (2002) *Biochem.* **366**, 873–882.

36. Zhang, W., Shields, J. M., Sogawa, K., Fujii-Kuriyama, Y., and Yang, V. W. (1998) *J. Biol. Chem.* **273**, 17917–17925.

37. Wenzlauth, H., Davis, R., Lockshon, D., and Lassar, A. (1990) *Proc. Natl. Acad. Sci. U. S. A.* **87**, 5623–5627.

38. Braun, T., and Arnold, H. H. (1991) *Nucleic Acids Res.* **19**, 5645–5651.

39. Sokido, R., Murai, K., Funahashi, J., Kamachi, Y., Fujisawa-Schara, A., Nabeshima, Y., and Kondo, H. (1994) *Mol. Cell. Biol.* **14**, 5692–5700.

40. Macian, F., Lopez-Rodriguez, C., and Rao, A. (2001) *Oncogene* **20**, 2476–2489.

41. Hu, Y. F., Luscher, B., Admon, A., Mermod, N., and Tjian, R. (1990) *Genes Dev.* **4**, 1741–1752.

42. Verschueren, K., Remacle, J. E., Collart, C., Kraft, H., Baker, B. S., Tytazonowski, P., Nelles, L., Wuytens, G., Su, M. T., Bodmer, R., Smith, J. C., and Huybrechts, D. (1999) *J. Biol. Chem.* **274**, 20489–20498.

43. Philipsen, S., and Suske, G. (1999) *Nucleic Acids Res.* **27**, 2991–3000.

44. Ikeda, K., Halle, J. P., Steiner, G., Meisterernst, M., and Kawakami, K. (1998) *Mol. Cell. Biol.* **18**, 10–18.

45. Remacle, J. E., Kraft, H., Lerchner, W., Wuytens, G., Collart, C., Verschueren, K., Smith, J. C., and Huybrechts, D. (1999) *EMBO J.* **18**, 5073–5084.
Characterization of the Human Lung CYP2F1 Gene and Identification of a Novel Lung-specific Binding Motif
Brian A. Carr, Jie Wan, Ronald N. Hines and Garold S. Yost

J. Biol. Chem. 2003, 278:15473-15483.
doi: 10.1074/jbc.M300319200 originally published online February 21, 2003

Access the most updated version of this article at doi: 10.1074/jbc.M300319200

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

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

This article cites 50 references, 17 of which can be accessed free at http://www.jbc.org/content/278/18/15473.full.html#ref-list-1