The Human *CYP2F* Gene Subfamily: Identification of a cDNA Encoding a New Cytochrome P450, cDNA-Directed Expression, and Chromosome Mapping†

Patson T. Nhamburo,‡ Shioko Kimura,‡ O. Wesley McBride,§ Christine A. Kozak,∥ Harry V. Gelboin,‡ and Frank J. Gonzalez*,‡

Laboratory of Molecular Carcinogenesis and Laboratory of Biochemistry, National Cancer Institute, and Laboratory of Molecular Microbiology, National Institute of Allergy and Infectious Diseases, National Institutes of Health, Bethesda, Maryland 20892

Received January 3, 1990; Revised Manuscript Received February 22, 1990

ABSTRACT: A cDNA coding for a P450, designated IIF1, was isolated from a human lung λgt11 library by screening with a human IIC9 cDNA probe. The cDNA-encoded IIF1 protein had 491 amino acids and a calculated molecular weight of 55 507. IIF1 cDNA, expressed by using a vaccinia virus vector, produced a cytochrome with a λ_{max} of 454 nm when reduced and complexed with carbon monoxide. This enzyme was able to dealkylate ethoxycoumarin, propoxycoumarin, and pentoxyresorufin but possessed no activity toward ethoxyresorufin and only trace dearylation activity toward benzyloxyresorufin. A variant cDNA, designated IIF1v, was isolated that was identical with IIF1 except for the loss of two segments of 161 and 388 bp within the cDNA coding region. Two mRNAs, consistent with the predicted size of the IIF1 and IIF1v transcripts, were found at very low abundance in lung specimens by Northern blot analysis. A 2-kb transcript, hybridizing with the human IIF1, was also detected as an abundant mRNA in rat lung. The *CYP2F* gene subfamily was localized to human chromosome 19 and mouse chromosome 7. On the basis of Southern blotting analysis with multiple restriction enzymes, we conclude that the *CYP2F1* gene is flanked by a second highly similar gene.

Cytochrome P450s[1] are the terminal components of the mixed-function monooxygenase system. Nine mammalian families of P450s have been described (Nebert et al., 1989). Five of these are involved in steroid biosynthetic pathways while the remaining four families carry out the oxidation of numerous compounds including endogenous steroids and fatty acids and foreign chemicals such as drugs and chemical carcinogens. Either the latter compounds can be activated by a P450 to a highly reactive unstable intermediate capable of binding to DNA and causing cell transformation or conversely P450s can inactivate a potentially dangerous substance. Levels of certain forms of P450 can be influenced by the presence of inducers while other P450s are subjected to developmental and sex-specific regulation (Gonzalez, 1988).

It is well established that genetic factors influence the expression of P450s in both humans and rodents. Genetic differences in P450 expression, referred to as polymorphisms, have been discovered in humans via clinical pharmacological approaches such as metabolic profiles of drugs in population studies. For example, the debrisoquine polymorphism was uncovered due to pronounced hypotensive responses of certain individuals receiving this sympatholytic antihypertensive drug (Mahgoub et al., 1977). Another drug oxidation defect, termed the mephenytoin polymorphism, was discovered through analysis of serum and urine metabolic profiles in a group of individuals who received experimental subtherapeutic doses of the drug (Kupfer & Preisig, 1984). The debrisoquine polymorphism is due to the presence of mutant *CYP2D6* genes in the population (Gonzalez et al., 1988; Skoda et al., 1988) while the molecular basis of the mephenytoin defect is unknown.

On the basis of the known prevalence of P450 gene defects in humans, we began to search for other human P450s that are also polymorphically expressed. The strategy involves screening of human cDNA libraries with various rat and human P450 cDNAs. In particular, we focused on lung, a tissue that comes in direct contact with environmental pollutants and cigarette smoke. In the present report, we describe a cDNA encoding a new P450, designated IIF1, that is expressed in lung tissue. Vaccinia virus was used as a vector to produce IIF1 protein in Hep G2 cells which displayed alkylcoumarin *O*-dealkylase activity. The *CYP2F* gene was localized to human and mouse chromosomes 19 and 7, respectively.

Materials and Methods

Isolation and Characterization of IIF1 and IIF1v. A λgt11 library was previously constructed as described by Watson and Jackson (1985) from poly(A) RNA isolated from a human lung specimen obtained from a kidney donor at the University of Miami Department of Surgery, Division of Transplantation. The library was screened by plaque hybridization using the human IIC9 cDNA probe (Kimura et al., 1987). Two classes of cDNAs were isolated, and the longest of each was subcloned into the *Eco*RI site of pUC9. Each cDNA was subjected to DNA sequencing by using the shotgun cloning method (Deininger, 1983). Briefly, multiple fragments ranging from 400 to 1000 bp were generated from concatemerized cDNA insert by sonication. These were made blunt-ended by using S1 nuclease and DNA polymerase Klenow fragment and then inserted into the *Sma*I site of M13 mp10. Random recom-

† The nucleic acid sequence in this paper has been submitted to GenBank under Accession Number J02906.

‡ Laboratory of Molecular Carcinogenesis, National Cancer Institute.

§ Laboratory of Biochemistry, National Cancer Institute.

∥ Laboratory of Molecular Microbiology, National Institute of Allergy and Infectious Diseases.

[1] The nomenclature used in this report is that described by Nebert et al. (1989). The gene is denoted *CYP2F1*, and the protein, mRNA, and cDNA are designated IIF1. IIF1v represents the variant cDNA. vvWT and vvIIF1 are wild-type and recombinant vaccinia viruses, respectively.

binant M13 clones were subjected to dideoxy nucleotide chain termination sequencing (Sanger et al., 1977). Every base was sequenced at least once in both directions with an average of at least six independent determinations. Sequence data were compiled and analyzed by using the Beckman Microgenie software, and sequence alignments were carried out by use of NUCALN (Wilbur & Lipman, 1983) and FASTA programs (Pearson & Lipman, 1988).

Characterization of the IIF1 and IIF1v mRNAs. Total RNA was isolated as described by Chirgwin et al. (1979) except cesium trifluoroacetic acid was used in place of CsCl. Northern blotting was carried out using 2.2 M formaldehyde–1% agarose gels as described by Lehrach et al. (1977). RNA was transferred to Nytran membranes (Schleicher & Schuell, Inc., Keene, NH), and the membranes were incubated with nick-translated IIF1 cDNA insert according to the procedure of Church and Gilbert (1984). The filters were washed at 65 °C by using 0.3 M NaCl, 30 mM sodium citrate (pH 7.0), and 0.5% SDS, prior to autoradiography.

Expression of IIF1 cDNA Using Vaccinia Virus. The IIF1 cDNA insert was made blunt-ended with Klenow fragment and inserted into the vector pSC11 (Chakrabarti et al., 1985). The pSC11-containing IIF1 cDNA was introduced into vaccinia virus as described by Mackett et al. (1984). The resultant recombinant virus, designated vvIIF1, was used to infect human TK^- 143 or Hep G2 cells (ATCC HB 8065). As a control, TK^- 143 or Hep G2 cells were infected with wild-type vaccinia virus strain WR, designated vvWT. Twenty-four hours after infection, cells were harvested and washed with PBS (Biofluids, Rockville, MD). Spectral analysis was carried out as described by Omura and Sato (1964). Briefly, cells were disrupted by sonication in 0.1 M sodium phosphate buffer, pH 7.2, containing 15% w/v glycerol and 0.3% w/v Emulgen 913. Insoluble material was removed by a brief centrifugation. The sample was divided into two cuvettes, and the sample cuvette was gently bubbled with CO gas. A few crystals of sodium dithionite were added to the two cuvettes, and a difference spectrum was recorded by using an Aminco DW-2000 spectrophotometer. P450 content was calculated by using an extinction coefficient of 91 cm^{-1} mM^{-1} (Omura & Sato, 1964).

Ethoxy- and propoxycoumarin *O*-dealkylase assays were performed at 37 °C for 20 min by using the fluorometric method of Greenlee and Poland (1978). Ethoxy- and pentoxyresorufin *O*-dealkylase and benzyloxyresorufin *O*-dearylase assays were conducted as detailed by Burke and Meyer (1974). These latter assays were performed at 25 °C.

Chromosome Mapping of the CYP2F Subfamily in Humans and Mice. Somatic cell hybrids were used to determine the chromosome location of the *CYP2F* subfamily in mouse and human. To map the human *CYP2F1*[2] gene, we analyzed 27 primary and 14 subclones of human–hamster hybrids and 14 primary and 40 subclones of human–mouse hybrids. Construction and analysis of these hybrids for markers on the various human chromosomes were described in earlier studies (McBride et al., 1982a–c). The mouse–hamster hybrid cell lines used to map the mouse *CYP2F1* gene were previously characterized (Kozak et al., 1975; Kozak & Rowe, 1979, 1980). In order to map the *CYP2F* locus, parental cell DNAs were examined with various restriction enzymes to determine which enzyme produced IIF1 probe-hybridizing fragments that could be electrophoretically resolved: the human–mouse and human–Chinese hamster hybrid DNAs were examined by using *Eco*RI while the mouse–Chinese hamster hybrid DNAs were digested with *Pst*I. Southern blotting conditions for the human gene mapping were as described by McBride et al. (1986). Under these hybridization and washing conditions, only sequences with <10% divergence could be detected, thereby eliminating the possibility that genes in a P450 subfamily other than *CYP2F* would be detected. Lower stringency washings were carried out for mapping the mouse gene.

[2] The *CYP2F1* chromosomal locus in man is designated *CYP2F1* while that of mouse is denoted *Cyp2f1* according to Nebert et al. (1989).

Table I: Amino Acid Sequence Similarities between IIF1 and Other Human P450s[a]

protein	IIA3	IIB7	IIC8	IIC9	IID6	IIE1
% similarity	53	48	50	49	39	48

[a] Amino acid similarities were calculated by using the program FASTA (Pearson & Lipman, 1988). The sequence data were taken from the following references: IIA3, Yamano et al. (1989b); IIB7, Yamano et al. (1989c); IIC8 and IIC9, Kimura et al. (1987); IID6, Gonzalez et al. (1988); IIE1, Song et al. (1986).

Detection of DNA Restriction Fragment Length Polymorphism. DNA was isolated from the peripheral leukocytes of 10 unrelated normal individuals, digested with restriction endonucleases (*Eco*RI, *Hin*dIII, *Bam*HI, *Xba*I, *Taq*I, *Sac*I, *Pvu*II, *Pst*I, *Msp*I, *Bgl*II, *Eco*RV, and *Kpn*I), size-fractionated by gel electrophoresis, transferred to nylon membranes, and hybridized with the IIF1 full-length and subfragment cDNA probes. When RFLPs were detected, DNAs from an additional 29 unrelated individuals were similarly examined. The genotypes at each polymorphic restriction site were readily determined from the Southern blots. The frequency of each allele was then determined by standard methods, i.e., $p = [n_1(11) + n_2/2(12)]/N$. In this assessment, p is the frequency of allele 1, n_1 is the number of homozygotes (i.e., 11) for this allele, n_2 is the number of heterozygotes (12), and N represents the total number of individuals examined. In a two-allele polymorphism, the frequency (q) of allele 2 is simply determined (i.e., $q = 1 - p$).

RESULTS

Isolation and Sequence of IIF1 and IIF1v. A cDNA, designated IIF1, was isolated from a λgt11 library prepared from human lung mRNA. This cDNA spanned 1825 bp and encoded a protein of 491 amino acids (Figure 1). The beginning of the open reading frame of IIF1 was assigned to the 5′-most ATG codon at base 56. The cDNA-deduced protein had a calculated molecular weight of 55 507 and displayed from 39 to 53% amino acid sequence similarities with human P450s within other *CYP2* subfamilies (Table I). The IIF1 protein demonstrated less than 35% sequence similarity with proteins in the human *CYP1* and *CYP3* families. We were unable to find any published P450 sequences displaying more than 50% amino acid similarity with that of IIF1. We are also not aware of any purified P450 preparations in the literature that possessed an amino-terminal sequence with similarity to that of IIF1. These data indicate that IIF1 has not been previously found in any organisms at the protein, cDNA, or gene level.

During screening of the same lung λgt11 library, a second cDNA was identified that differed from IIF1. This clone, designated IIF1v (IIF1 variant), contained 1301 bp and was identical with the IIF1 cDNA except for an additional 25 bp at its 5′ end and two deletions of 161 and 388 bp (Figure 1). The open reading frame of IIF1v contained only 239 amino acid residues; the sequence of the carboxy-terminal 78 amino acid residues differed from IIF1 due to a shift in the translation reading frame generated via the 161 bp deletion. This trun-

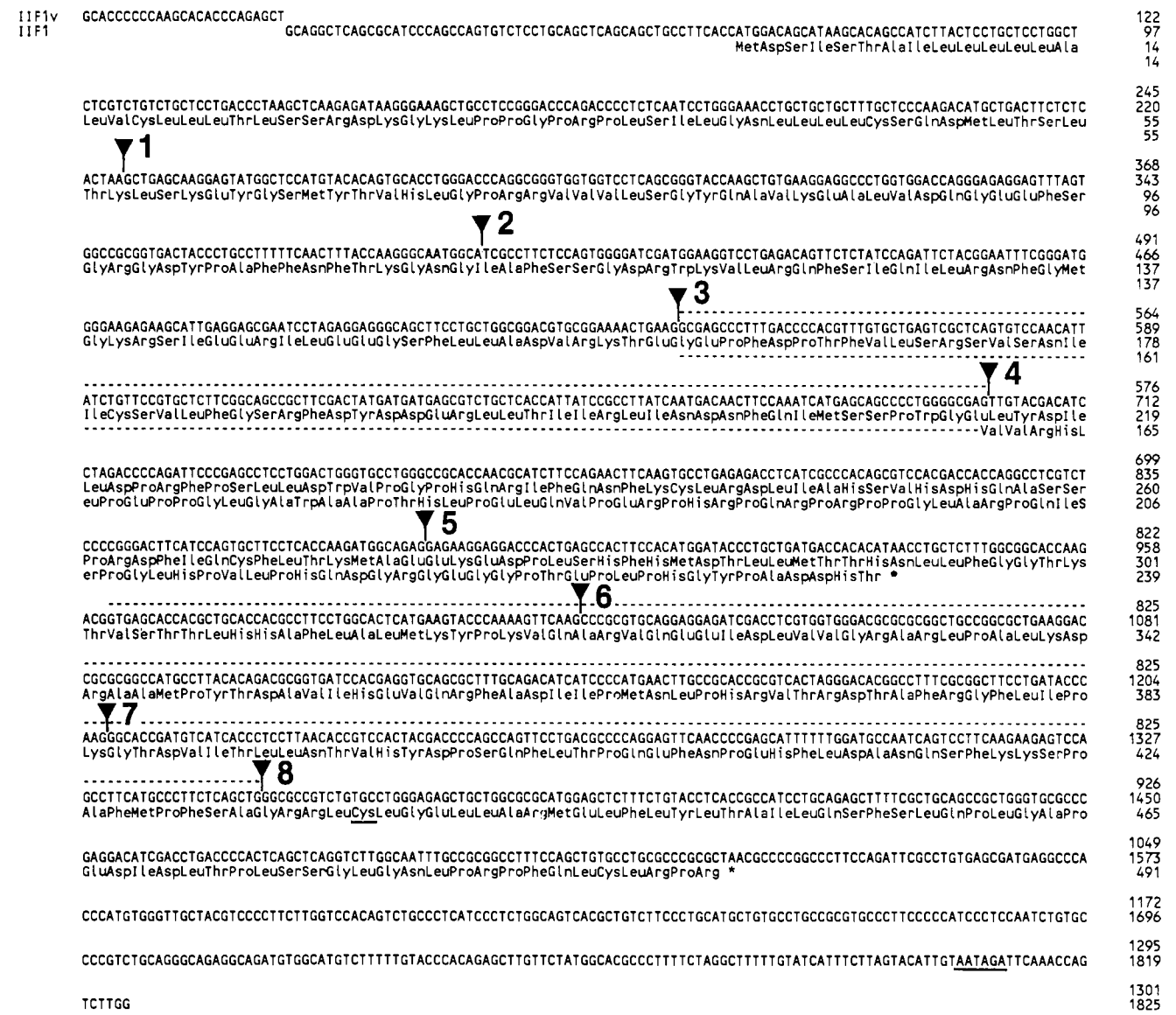

FIGURE 1: Sequence of IIF1 and IIF1v cDNAs. The complete cDNA and deduced amino acid sequences of IIF1 are presented. Only DNA and amino acid sequences in IIF1v that differ from IIF1 are displayed above and below the IIF1 DNA and amino acid sequences, respectively. The dashes in IIF1v represent deleted sequence. The solid triangles and numbers denote the putative exon boundaries of the *CYP2F1* gene. The nucleotides and amino acids are numbered at the right, and the heme binding Cys and polyadenylation signal are underlined.

cated protein is likely to be nonfunctional due to the lack of the cysteine found at residue 436 in IIF1 that serves as the fifth thiolate ligand to the heme iron at the P450 active site.

The deleted regions in IIF1v were compared with the putative splice sites of the *CYP2F1* primary transcript (Figure 1, arrows). These were derived from sequences of other *CYP2* genes, all of which have nine exons (Gonzalez, 1988). As seen in Figure 1, the 161 bp deletion probably results from a skipping of exon 4 since this deletion corresponds exactly to the putative splice junctions. The 3′ boundary of the 388 bp deletion in IIF1v is also conserved with the 3′ and 5′ junctions of putative exons 8 and 9. However, the 5′ base of the 388 bp deletion is inside putative exon 6 of IIF1, suggesting that the 5′ GT in exon 6 is used as an artificial splice donor.

Analysis of IIF1 and IIF1v mRNA Expression in Human Lung and Liver. As predicted from the cDNA sequences, putative IIF1 and IIF1v mRNAs should differ in size by 549 nt. With the addition of a poly(A) tail of 200–300 bp, IIF1 mRNA should be about 2.0 kb, and IIF1v should be around 1.5 kb. Indeed, Northern blotting analysis using the IIF1 cDNA revealed the presence of two mRNAs of 1.5 and 2 kb in total RNA from three separate lung specimens (Figure 2). These RNAs were present at very low levels and were only detectable after long periods of exposure of the filters to autoradiographic film. The levels of these mRNAs varied between tissue specimens taken from different individuals. For example, lung sample 3 had much higher levels of both mRNAs than samples 1 and 2 (Figure 2). Only a single liver, taken from the same individual as lung sample 3, contained a substantial amount of mRNA corresponding to IIF1. Eleven other liver specimens had much lower or undetectable signals. We cannot rule out whether the expression of IIF1 mRNA in this single liver is due to induction via drug administration or smoking.

Expression of CYP2F Genes in Rat. The expression of IIF mRNA was also analyzed in rat tissues. An RNA of about 2.0 kb was found in lung of untreated rats, and its level appeared not to be affected by phenobarbital pretreatment of the animals (Figure 3). A second lower abundant mRNA of approximately 1.5 kb was also detected. No RNA, complementary to the IIF probe, was seen in intestine, and only trace levels were found in liver and kidney. Interestingly, the level of RNA in liver was markedly decreased by phenobarbital treatment, in contrast to that in lung and kidney. However,

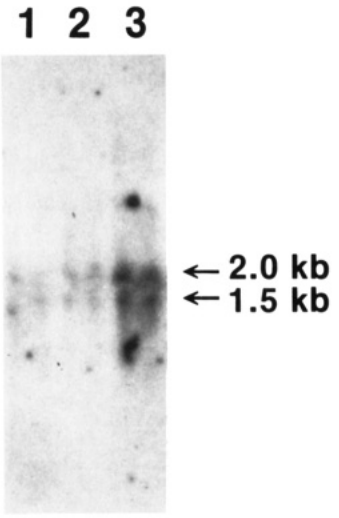

FIGURE 2: Northern blotting analysis of RNA in human lung. Total RNA (40 μg per lane) was electrophoresed on 2.2 M formaldehyde–1% agarose gels and transferred to Nytran membranes. The membranes were hybridized with nick-translated IIF1 cDNA insert, washed, and exposed to autoradiographic film for 5 days at −70 °C with the aid of a Dupont Lightning Plus intensifying screen.

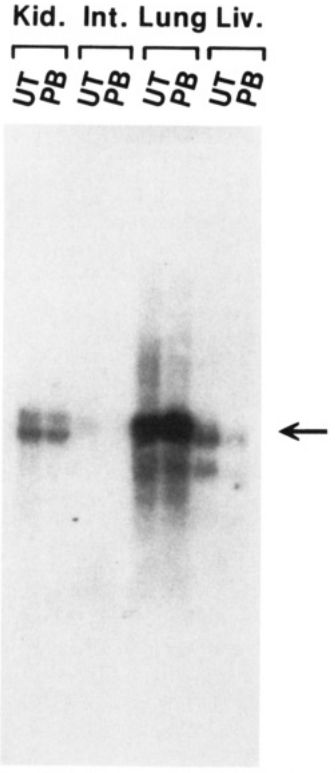

FIGURE 3: Northern blotting analysis of rat RNAs. Ten micrograms of total RNA isolated from kidney (Kid), small intestine (Int), lung, and liver (Liv) of untreated (UT) and phenobarbital-treated (PB) rats was electrophoresed on 2.2 M formaldehyde–1% agarose gels and transferred to Nytran membranes. The filter was hybridized to nick-translated IIF1 cDNA, washed, and exposed to autoradiographic film for 24 h at −70 °C with the aid of a Dupont Lightning Plus intensifying screen. The arrow indicates the transcript detected in lung.

the RNA species in liver and kidney tissues were of slightly different size than that found in lungs, suggesting that they may be transcribed from genes in other *CYP2* gene subfamilies. We have not tested the effect of other P450 inducers on IIF mRNA levels.

Expression of the IIF1 cDNA Using Vaccinia Virus. To determine if the IIF1 cDNA codes for a functional P450,

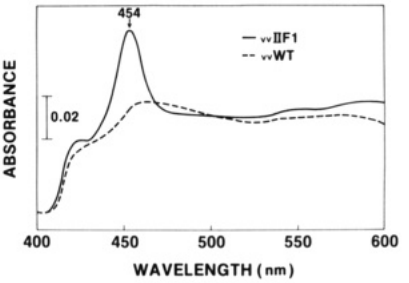

FIGURE 4: Spectral analysis of vaccinia-infected cell lysate. Twenty milligrams of lysate protein from TK⁻ cells infected with vaccinia virus containing the IIF1 cDNA (vvIIF1) or wild-type vaccinia (vvWT) was treated with sodium dithionite, and difference spectra were measured in the presence of carbon monoxide using an Aminco DW-2000 spectrophotometer.

Table II: NADPH-Dependent Enzymatic Activity of cDNA-Expressed Human P450 IIF1[a]

	enzyme activities [pmol min^{-1} (mg of protein)$^{-1}$]			
virus	ethoxy-coumarin *O*-deethylase	propoxy-coumarin *O*-depropylase	pentoxy-resorufin *O*-dependentylase	benzyloxy-resorufin *O*-dearylase
vvIIF1	51.20 ± 1.40	24.30 ± 1.30	22.00 ± 1.20	2.29 ± 0.10
vvWT	0.95 ± 0.04	0.66 ± 0.02	ND[b]	ND[b]

[a] Hep G2 cells were infected for 24 h with recombinant vaccinia virus containing P450 IIF1 or wild-type virus. Cells were harvested, gently sonicated, and centrifuged to sediment the nuclear fraction. The supernatant was then centrifuged at 100000*g* for 60 min to sediment the total membrane fraction. Membranes were resuspended in 0.1 M potassium phosphate buffer (pH 7.4). Enzyme activities were assayed in the presence and absence of 0.4 mM NADPH. Ethoxy- and propoxycoumarin activities were determined at 37 °C after 20-min incubation in the presence of 200 μM substrate, while activities toward resorufin analogues were monitored continuously at 25 °C in the presence of 10 μM substrate. Values represent mean ± SE of three separate experiments. In each experiment, assays were performed in duplicate. [b] ND, not detected.

cDNA expression was carried out with vaccinia virus. Human TK⁻ 143 cells, infected with a recombinant vaccinia virus containing the IIF1 cDNA (vvIIF1), were analyzed for P450 expression. When solubilized total cell protein was treated with sodium dithionite in the presence of CO and compared against a second sample of the same preparation that had only been reduced with dithionite, a Soret absorption band was found with a λ_{max} of 454 nm (Figure 4). No peak was found at 454 nm when cells that had been infected with wild-type vaccinia virus (vvWT) were analyzed. These data demonstrate that IIF1 is capable of producing a spectrally normal cytochrome. The spectral content of P450, calculated from the molar extinction coefficient of 91 cm^{-2} mM^{-1} (Omura & Sato, 1964), was 20 pmol/mg of total cell protein.

To determine if cDNA-expressed IIF1 is capable of oxidizing various common P450 substrates, human Hep G2 cells were infected with vvIIF1. These cells contain both endogenous NADPH–P450 oxidoreductase and cytochrome b_5 needed for P450 electron transfer (Aoyama et al., 1989). Cells infected with vvIIF1 possessed measurable levels of ethoxy- and propoxycoumarin *O*-dealkylase and pentoxyresorufin *O*-dependentylase activities (Table II). Detectable levels of benzyloxyresorufin *O*-dearylase activity were also found. Cells infected with vvWT had only trace levels of activity toward ethoxy- and propoxycoumarin, the same as that found in uninfected Hep G2 cells. Cells infected with vvIIF1 possessed no ethoxyresorufin *O*-deethylase activity. The level of expressed P450 in Hep G2 cell membrane fractions is 75 pmol/mg of protein. Using this value, we calculated a turnover number of ethoxycoumarin of 0.7 min^{-1}. These data indicate

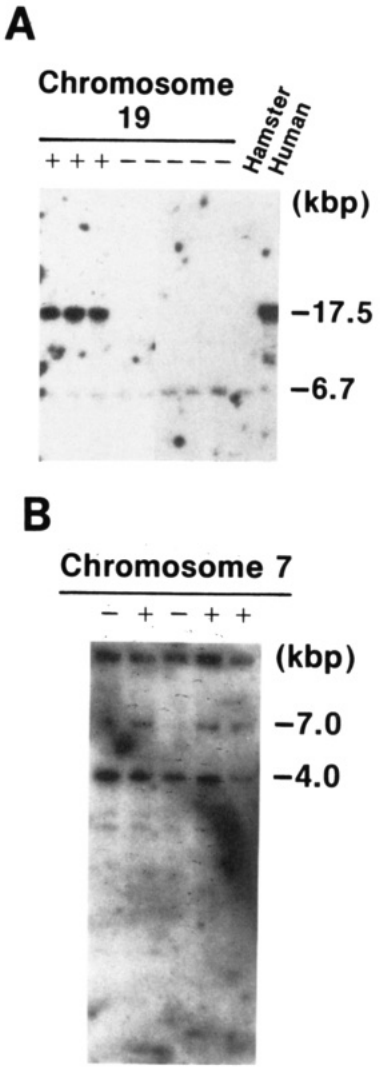

FIGURE 5: Southern blotting analysis of human–hamster (panel A) and mouse–hamster (panel B) somatic cell hybrid DNAs using the IIF1 cDNA probe. The human–hamster and mouse–hamster DNAs were electrophoresed on 0.7% and 1% agarose gels, transferred to nylon membranes, and hybridized with nick-translated probe. The filters were washed under high (<10% divergence) stringency (panel A) and low (<20% divergence) stringency (panel B) and exposed to autoradiographic film for 5 days at −70 °C with the aid of an intensifying screen.

Table III: Correlation between Human Chromosome 19 and the *CYP2F1* Gene among Human–Rodent Somatic Cell Hybrids[a]

human chromosome	gene/chromosome[b] +/+	+/−	−/+	−/−	% discordancy
1	21	3	13	58	17
2	16	8	10	61	19
3	18	6	17	54	24
4	21	3	37	34	42
5	17	7	7	64	15
6	24	0	24	47	25
7	13	11	26	45	39
8	17	7	19	52	27
9	18	6	13	58	20
10	10	14	9	62	24
11	14	10	16	55	27
12	11	13	17	54	32
13	14	10	21	50	33
14	15	9	27	44	38
15	17	7	29	42	38
16	13	11	23	48	36
17	20	4	37	34	43
18	17	7	33	38	42
19	24	0	4	67	4[c]
20	21	3	19	52	23
21	22	2	40	31	44
22	14	10	14	57	25
x	14	10	34	37	46

[a] The human *CYP2F1* gene was detected as a 17.5 kbp fragment in *Eco*RI-digested human–rodent somatic cell hybrid DNAs after Southern hybridization with a full-length IIF1 cDNA probe. This band was well-resolved from 2.9 or 3.6 and 6.7 kbp cross-hybridizing bands in mouse or Chinese hamster (Figure 5A) DNAs, respectively. Detection of the gene is correlated with the presence or absence of each human chromosome in the group of somatic cell hybrids. Discordancy represents the presence of the gene in the absence of the chromosome (+/−) or the absence of the gene despite the presence of the chromosome (−/+), and the sum of these numbers divided by total hybrids examined (×100) represents percent discordancy. The human–hamster hybrids consisted of 27 primary hybrids and 14 subclones (14 positive of 41 total), and human–mouse hybrids represented 14 primary clones and 40 subclones (10 positive of 54 total). [b] Number of hybrid clones in each combination is presented. [c] Three human–mouse clones and subclones and one human–hamster hybrid contained chromosome 19 but did not retain a detectable human *CYP2F1* gene sequence. These four hybrids also did not exhibit detectable hybridization with a human *CYP2B7* probe (Yamano et al., 1989c), but all hybridized very weakly with a probe for the opal suppressor tRNA gene, previously assigned to this chromosome (McBride et al., 1987). These discordancies probably represent differing sensitivities for detection of the different genes although a chromosome 19 break with loss of a portion of the chromosome cannot be excluded.

that none of the tested substrates is rapidly metabolized by IIF1, since favored substrates for other P450s have turnover numbers ranging from 20 to 50 min^{-1} (Guengerich, 1987).

Chromosome Mapping of the CYP2F Gene Subfamily in Humans and Mice. Somatic cell hybrids were used to determine the chromosomal localization of the *CYP2F1* genes in humans and mice. Panels A and B of Figure 5 display representative Southern blots of human–hamster and mouse–hamster hybrid DNAs, respectively. The human gene was contained on a 17.5 kbp *Eco*RI fragment (Figure 5A) while the mouse gene was localized by following the presence of a 7 kbp *Pst*I fragment (Figure 5B). The location of the *CYP2F* locus was determined by correlating the presence or absence of these fragments with the presence or absence of specific human and mouse chromosomes previously characterized for each hybrid cell line as described in earlier reports (see Materials and Methods).

Human chromosome 19 was found to segregate concordantly with the 17.5 kbp *Eco*RI fragment (Table III). Only 4% discordancies were found with chromosome 19 and the *CYP2F* gene. These apparently resulted from failure to detect the *CYP2F1* or several *CYP2B* genes in hybrids retaining human chromosome 19 in only a small fraction of the total cell population (Yamano et al., 1989c; see Table III legend). These cells did, however, possess other markers on 19 such as the human opal suppressor tRNA gene (McBride et al., 1987).

The *CYP2B* (Santisteban et al., 1988; Miles et al., 1988; Yamano et al., 1989c) and *CYP2A* (Phillips et al., 1985; Miles et al., 1989) were previously mapped to human chromosome 19 and, in fact, reside within 350 kbp between 19q12 and 19q13.2 (Miles et al., 1989). Both subfamilies have also been localized to mouse chromosome 7 (Simmons & Kasper, 1983; Kimura, S., et al., 1989). On the basis of these data, we assumed that the *CYP2F1* gene was going to be located on mouse chromosome 7 and therefore used a subpanel of hybrid cell lines, four of which contained this chromosome and the remaining six containing subsets of the other mouse chromosomes, to ascertain the location of the mouse *CYP2F* locus. Indeed, the 7 kbp *Pst*I fragment, derived from the mouse *CYP2F1* gene, segregated concordantly with chromosome 7. No discordancies were found with this chromosome whereas all other chromosomes had discordancies with the 7 kbp

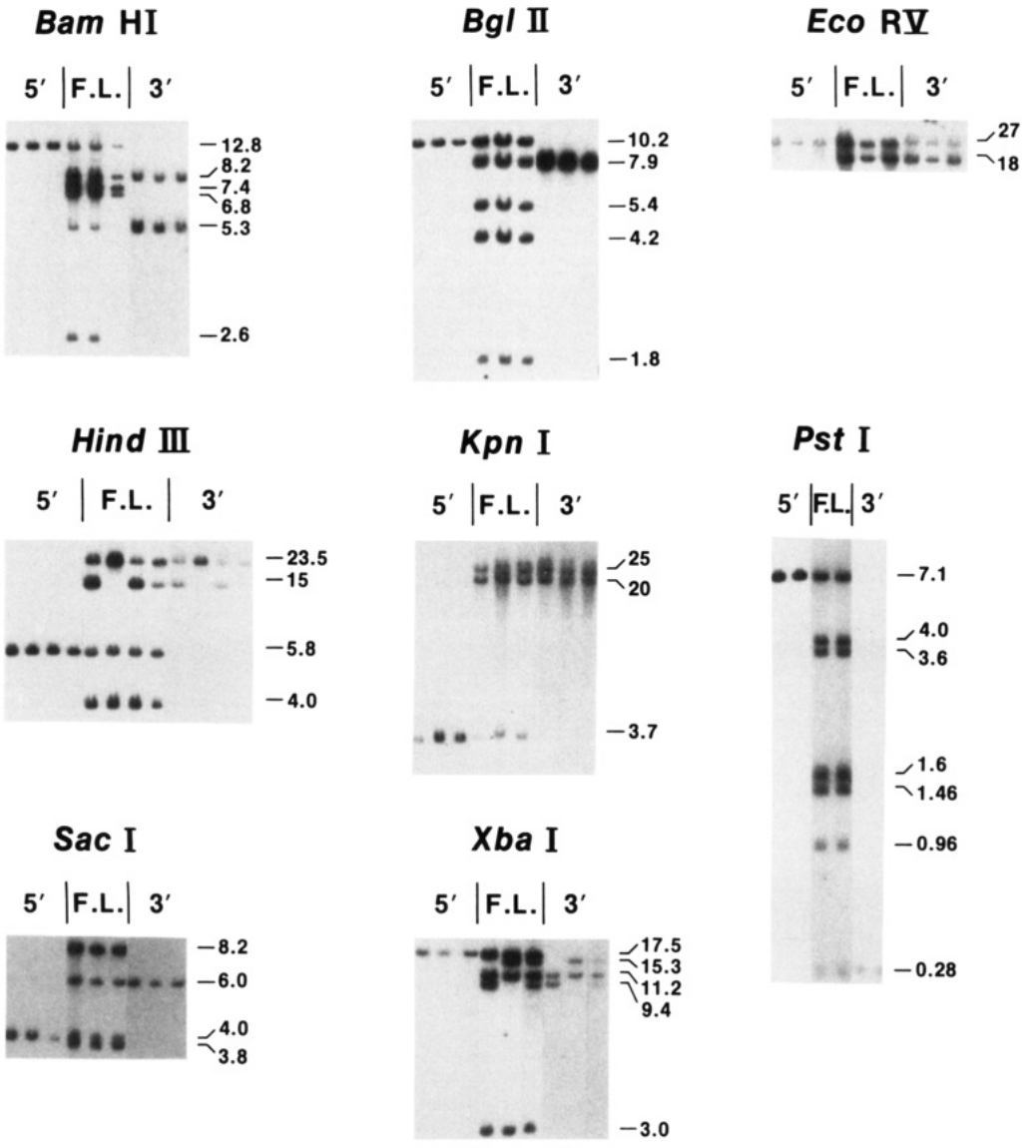

FIGURE 6: Southern blotting analysis of human lymphocyte genomic DNAs digested with eight restriction enzymes. The full-length IIF1 cDNA (F.L.) and probes derived from the 5′ and 3′ ends of the cDNA are described in the text. The fragment sizes are in kilobase pairs. The hybridization patterns with all three probes have been aligned in each photo. Allelic bands of 9.4 and 15.3 kbp were found in *Xba*I digests. Alleles of 15 and 23.5 kbp are shown in *Hin*dIII digests. The molecular sizes of each fragment were calculated from kinase-labeled *Hin*dIII-digested λ and *Hae*III-digested φX174 RF DNA markers electrophoresed and transferred in parallel with genomic DNAs.

fragment ranging from 14 to 80% (Table IV). These results establish that the *CYP2F1* gene is located on human chromosome 19 and mouse chromosome 7.

Evaluation of the Number of Genes within the CYP2F Subfamily. As indicated above, a single 17.5 kbp fragment hybridizing with the IIF1 cDNA was observed in *Eco*RI digests of human–rodent hybrid cell DNAs, and this fragment was localized to chromosome 19. This suggests either a single *CYP2F* gene or multiple gene copies on this chromosome with *Eco*RI restriction fragments which are indistinguishable in size. To assess the latter possibility, DNAs isolated from the peripheral leukocytes of unrelated individuals were digested with eight different restriction endonucleases, and the resultant DNAs were transferred to nylon membranes after fractionation by 0.7% agarose gel electrophoresis. Simple restriction patterns were observed following hybridization with the IIF1 full-length cDNA probe, and the number of hybridizing fragments varied from 1 only in the *Eco*RI digests to 7 fragments in *Pst*I digests (Figure 6) and *Pvu*II digests (not shown). After removal of the hybridized probe by alkaline treatment, these same filters were sequentially hybridized with a 300 bp 5′ cDNA fragment (extending to the *Kpn*I site at nucleotide 293 in exon 2) and a 450 bp 3′ cDNA fragment, extending from the *Sac*I site (nucleotide 1388) in exon 9 to the 3′ terminus. The 3′ cDNA probe was especially useful in this analysis since it apparently does not contain any introns (Figure 1). With this probe, a single hybridizing band should be observed with any restriction digest provided the restriction site is missing from the probe sequence. In contrast, multiple hybridizing fragments might be expected if additional gene copies are present due to lack of conservation of restriction sites in the 3′ region of each gene copy. The results of hybridization with the 3′ cDNA probe strongly suggest the presence of two copies of the *CYP2F* gene at this locus as indicated by the results shown in Figure 6. Two large fragments of 20 and 25 kbp were found in *Kpn*I digests, and a 23.5 kbp constant band as well as 15 and 23.5 kbp allelic fragments were observed in *Hin*dIII digests. Two nonallelic restriction fragments were also detected in *Bam*HI, *Xba*I, *Eco*RV, and *Bgl*II (doublet) digests with the 3′ cDNA probe. Since this probe presumably encompasses no introns (Figure 1) and there are no restriction sites for any of these enzymes in the 3′ cDNA

Table IV: Correlation between Mouse Chromosome 7 and the *CYP2F1* Gene among Mouse–Hamster Somatic Cell Hybrids[a]

mouse chromosome	gene/chromosome +/+	+/−	−/+	−/−	% discordancy
1	4	1	1	1	28
2	3	1	2	1	43
3	0	1	3	1	80
4	3	3	3	0	29
5	0	2	5	0	71
6	2	1	2	1	50
7	4	2	0	0	0
8	1	1	2	1	60
9	2	2	3	0	42
10	0	2	5	0	71
11	0	2	2	0	50
12	3	0	0	2	40
13	4	2	1	0	14
14	0	2	4	0	66
15	3	0	0	2	40
16	0	2	4	0	68
17	2	0	2	2	68
18	2	1	2	1	50
19	2	2	3	0	43
20	3	1	2	1	43

[a] The mouse *CYP2F1* gene was detected as a 7 kbp fragment in *Pst*I-digested hybrid cell DNAs using Southern blots and the human *CYP2F1* probe. A total of 10 hybrid cell lines were tested. This panel was selected from a large number of other cell lines to distinguish chromosome 7, since based on the human mapping results and the known linkage conservation between groups of genes in mouse and human (Nadeau, 1989) and the fact that the *CYP2A* and *CYP2B* subfamilies in mouse and man map to chromosomes 7 and 19, respectively (Gonzalez, 1988). The symbols are described in the legend to Table III.

probe, the two fragments in each case must therefore represent two copies of the gene with different restriction sites in their introns or 3′ flanking sequences. Interestingly, *Pst*I sites are found in the 3′ probe, 27 bp from the 5′ end of this sequence and 287 bp further downstream. The only *Pst*I fragment detected with the 3′ probe is an 0.28 kbp fragment, indicating the absence of introns within this region, as predicted. A single hybridizing fragment was also detected in *Sac*I (6 kbp) digests with the 3′ probe, suggesting conservation of this site in the 3′ regions of these genes. Only single fragments were detected with the 5′ cDNA probe in *Bam*HI, *Hin*dIII, *Xba*I, *Sac*I, *Bgl*II, *Kpn*I, and *Pst*I digests, indicating conservation of these restriction sites within both introns (or 5′ untranslated region for *Pst*I at nucleotide 56) and the 5′ flanking regions of both genes. Taken together, these Southern blotting results suggest the presence of two genes that share multiple conserved restriction sites. It is likely that these genes display high nucleotide similarities due to a recent gene duplication event. Closely related P450s, having diverged less than a few million years ago, are not uncommon in the P450 gene superfamily (Gonzalez, 1988).

Detection of Restriction Fragment Length Polymorphism. RFLPs were found in *Xba*I and *Hin*dIII digests with the 3′ cDNA probe due to the presence or absence of a restriction site (Figure 6). In *Xba*I digests, the allele sizes were 32 kbp[3] (A1), 15.3 kbp (A2), and 9.4 kbp (A3), and the allele frequencies (38 individuals) were A1:A2:A3 = 0.01:0.84:0.15. Two polymorphic alleles of 23.5 kbp (B1) and 15 kbp (B2) were found in *Hin*dIII digests with allele frequencies (39 individuals) of B1:B2 = 0.46:0.54, and an invariant band of 23.5 kbp was also present. This interpretation was based upon band intensity; a homozygote for the 23.5 kbp allele is shown in lane 2, homozygotes for the 15 kbp allele are present in lanes 1 and 3, and a heterozygote is shown in lane 4 of the *Hin*dIII digests detected with the full-length cDNA probe (Figure 6). This interpretation is also strongly supported by the fact that the *Hin*dIII alleles are in Hardy–Weinberg equilibrium whereas this would not be true ($P \ll 0.005$ by χ^2 test) if it were assumed that no invariant band is present (i.e., 32 heterozygotes and 7 homozygotes for the 23.5 kbp allele and no homozygotes for the 15 kbp allele). The details of this analysis are presented in Table V. Mendelian segregation of the *Hin*dIII and *Xba*I alleles in families was also observed (data not shown).

Table V: Observed and Expected Frequencies of Genotypes for the *Hin*dIII Restriction Fragment Length Polymorphism[a]

	number of individuals			
	(A) assuming the presence of 23.5 kbp invariant fragment		(B) assuming no 23.5 kbp invariant fragment	
genotype[b]	observed	expected[d]	observed[c]	expected[d]
11	7	8.3	7	13.6
12	22	19.4	32	18.9
22	10	11.4	0	6.6

[a] Observed genotypes of 39 unrelated individuals are tabulated based upon the assumption that an invariant 23.5 kbp band is present (A) or absent (B) in addition to 23.5 and 15 kbp allelic bands. [b] Genotypes 11, 22, and 12 refer to the presence of only the 23.5 kbp allelic band, only the 15 kbp allelic band, or both alleles, respectively. [c] Observed genotypes (B): 11 indicates the presence of only the 23.5 kbp hybridizing band, 12 indicates the presence of both 23.5 and 15 kbp bands with relative intensities of 23.5 kbp to 15 kbp of about 2:1, and 22 indicates the presence of both hybridizing bands with relative intensities of 23.5 kbp to 15 kbp of about 1:2. [d] Allele frequencies were first determined from the distribution of observed genotypes (see Materials and Methods). The ratio of the frequencies of the 23.5 kbp to 15 kbp alleles is 0.46:0.54 (A) and 0.59:0.41 (B). The expected numbers of individuals with each genotype were then determined by assuming Hardy–Weinberg equilibrium (i.e., the frequencies of genotypes 11, 12, and 22 in the population at equilibrium are p^2, $2pq$, and q^2, respectively). For example, the expected number of homozygotes (genotype 22) for the 15 kbp allele would be $0.54^2 \times 39 = 11.4$ (A) or $0.41^2 \times 39 = 6.6$ (B). The likelihood that the observed distribution of genotypes differs from the expected distribution by change alone can be determined from the χ^2 distribution [i.e., $\chi^2 = \sum(\text{observed} - \text{expected})^2/\text{expected}$] with 2 degrees of freedom. These results indicate that the distribution of genotypes in column A doesd not deviate from Hardy–Weinberg equilibrium ($P \geq 0.5$) whereas the distribution of genotypes in column B differs significantly from Hardy–Weinberg equilibrium ($\chi^2 = 18.88$, $P \ll 0.005$).

Discussion

In the present report, we have identified a new P450 gene subfamily in man, designated *CYP2F*, located on human chromosome 19. Using subfragments of the cDNA as probes for Southern analysis, it appears that the *CYP2F* gene has undergone a relatively recent reduplication in the human genome. Although most restriction sites within the two genes and their 5′ flanking regions have been evolutionarily conserved, some of the sites in the 3′ flanking regions have undergone mutations. The *CYP2F* genes may be linked to the *CYP2A* and *CYP2B* subfamilies within the 19q12 and 19q13.2 region of this chromosome (Miles et al., 1989). Indeed, all three subfamilies are also localized to mouse chromosome 7 near the centromere.[4] Interestingly, other *CYP2* subfamilies are located on different mouse and human chromosomes; the *CYP2C* subfamily is located on human chromosome 10 (Meehan et al., 1988a) and mouse chromosome 19 (Meehan et al., 1988b); the *CYP2D* subfamily is found on human chromosome 22 (Gonzalez et al., 1988) and mouse chromo-

[3] This allele was detected in only a single individual and is not displayed in Figure 6.

[4] Kozak and Gonzalez, unpublished results.

some 15 (Gonzalez et al., 1987); the *CYP2E* subfamily is present on human chromosome 10 (Umeno et al., 1988a) and mouse chromosome 7 (Umeno et al., 1988b). Even though the *CYP2E* gene subfamily resides on mouse chromosome 7, the presence of this gene on human chromosome 10 indicates that it is not closely linked to the mouse *CYP2A*, *CYP2B*, and *CYP2F* subfamilies since they are located on human chromosome 19. Efforts are underway using linkage studies in large human pedigrees and pulse field electrophoresis to determine the distance between the *CYP2F* and the *CYP2A* and *CYP2B* subfamilies.

Interestingly, we isolated two cDNAs representative of a normal mRNA, IIF1 and a variant cDNA, IIF1v. The possible presence of a predicted prematurely terminated protein from this variant suggests that it does not produce a functional P450 protein. The IIF1v transcript that was cloned and sequenced had two deletions. The 5'-most 161 bp deletion appears to be due to exclusion of exon 4. The larger 388 bp deletion is due to a cryptic 5' splice donor site in exon 6 and skipping of exons 7 and 8, using the 3' acceptor site of intron 8. Four possibilities could account for the production of IIF1v: (i) transcription from the second gene in the *CYP2F* subfamily and faulty splicing of the primary transcript; (ii) transcription and then faulty transcript splicing from a mutant allele of the *CYP2F1* gene (that produces IIF1 mRNA); (iii) transcription from a pseudogene; (iv) alternative splicing of the *CYP2F1* gene primary transcript. It seems unlikely that IIF1 and IIF1v mRNAs are products of two separate genes. Although there appears to have been a recent reduplication of the *CYP2F* gene and both copies could be functional, it would be surprising not to find even a single nucleotide difference between them in both the coding and noncoding regions of the mRNAs, other than the two deletions predicted from the IIF1 and IIF1v cDNA sequences. One might also expect at least a few base changes among different *CYP2F1* alleles. In any case, the origin of the variant transcript remains a mystery until further experimentation is carried out.

Two transcripts of similar size to the normal and variant human transcripts were also found in rat lung mRNA. The smaller transcript was considerably less abundant than the larger 2 kb mRNA. If these rat transcripts are the products of two genes, then an independent gene reduplication must have also occurred in this species as well as a mutation(s) resulting in a similarly deleted mutant transcript. This appears quite unlikely. It is tempting to speculate that alternative splicing also occurs in the rat *CYP2F1* gene. However, we cannot at this time rule out the possibility that the shorter mRNA is derived from a second rat *CYP2F* gene.

Unusual P450 mRNA transcripts have been detected in mice (Wong et al., 1987; Noshiro et al., 1988), rats (Kimura, H., et al., 1988, 1989), and humans (Okino et al., 1987; Miles et al., 1988). In all cases, the transcripts were believed to be produced by alternative splicing and to yield RNAs that produce truncated proteins lacking the heme binding region near the P450 carboxy terminus. The truncated protein produced by a variant rat P450 PB-1 (*CYP2C6*) mRNA has, however, been detected by using antibody against a synthetic peptide (Kimura, H., et al., 1989). It remains to be established whether the variant transcripts produced from P450 genes in several subfamilies and many different species of animals are functionally significant.

The IIF1 cDNA was found to code for a P450 that was capable of metabolizing several common monooxygenase substrates. The highest activity was toward ethoxycoumarin. However, we found no correlation between this activity and IIF1 mRNA levels in the four lungs analyzed, indicating IIF1 may not be the major form that metabolizes ethoxycoumarin in human lung. This substrate can be deethylated by several different forms of P450 in rat, indicating a high likelihood that several human P450s also metabolize this substrate. Indeed, human IA2 had a substrate turnover number for ethoxycoumarin about half that of IIF1 (Yamano et al., 1989a). The possibility remains that other compounds exist that are more specific substrates for IIF1. Currently, we are using vaccinia-expressed IIF1 to screen for metabolism of other endogenous and exogenous chemicals.

Northern blot analysis of human lung RNA revealed only very low levels of IIF1 mRNA. This RNA was also weakly expressed in liver. Since lung tissue is composed of a variety of different cell types, the *CYP2F1* gene might be expressed at high levels in a specialized cell or a group of cells. This may account for its low level in total lung RNA. The precise cellular localization and role of IIF1 in lung remain to be established.

References

Aoyama, T., Korzekwa, K., Nagata, K., Gillette, J., Gelboin, H. V., & Gonzalez F. J. (1989) *Eur. J. Biochem. 181*, 331–336.

Burke, M. D., & Meyer, R. T. (1974) *Drug Metab. Dispos. 2*, 583–588.

Chakrabarti, S., Brechling, K., & Moss, B. (1985) *Mol. Cell. Biol. 5*, 3403–3409.

Chirgwin, J. M., Przybyla, A. E., MacDonald, R. J., & Rutter, W. J. (1979) *Biochemistry 18*, 5294–5299.

Church, G. M., & Gilbert, W. (1984) *Proc. Natl. Acad. Sci. U.S.A. 81*, 1991–1995.

Deininger, P. L. (1983) *Anal. Biochem. 129*, 216–223.

Gonzalez, F. J. (1988) *Pharmacol. Rev. 40*, 243–288.

Gonzalez, F. J., Matsunaga, T., Nagata, K., Meyer, U. A., Nebert, D. W., Pastewka, J., Kozak, C. A., Gillette, J., Gelboin, H. V., & Hardwick, J. P. (1987) *DNA 6*, 149–161.

Gonzalez, F. J., Vilbois, F., Hardwick, J. P., McBride, O. W., Nebert, D. W., Gelboin, H. V., & Meyer, U. A. (1988) *Genomics 2*, 174–179.

Greenlee, W. F., & Poland, A. (1978) *J. Pharmacol. Exp. Ther. 205*, 596–608.

Guengerich, F. P. (1987) in *Mammalian Cytochromes P450* (Guengerich, F. P., Ed.) Vol. 1, pp 1–54, CRC Press, Boca Raton, FL.

Kimura, H., Yoshioka, H., Sogawa, K., Sakai, Y., & Fujii-Kuriyama, Y. (1988) *J. Biol. Chem. 263*, 701–707.

Kimura, H., Sogawa, K., Sakai, Y., & Fujii-Kuriyama, Y. (1989) *J. Biol. Chem. 264*, 2338–2342.

Kimura, S., Pastewka, J., Gelboin, H. V., & Gonzalez, F. J. (1987) *Nucleic Acids Res. 15*, 10053–10054.

Kimura, S., Kozak, C. A., & Gonzalez, F. J. (1989) *Biochemistry 28*, 3798–3803.

Kozak, C. A., & Rowe, W. P. (1979) *Science 204*, 69–71.

Kozak, C. A., & Rowe, W. P. (1980) *J. Exp. Med. 152*, 1419–1423.

Kozak, C. A., Nichols, E., & Ruddle, F. H. (1975) *Somat. Cell Genet. 1*, 371–382.

Kupfer, A., & Preisig, R. (1984) *Eur. J. Clin. Pharmacol. 26*, 753–759.

Lehrach, H., Diamond, D., Wozney, J. M., & Boedtker, H. (1977) *Biochemistry 16*, 4743–4751.

Mackett, M., Smith, G. L., & Moss, B. (1984) *J. Virol. 49*, 857–864.

Mahgoub, A., Idle, J. R., Dring, L. G., Lancaster, R., & Smith, R. L. (1977) *Lancet 2*, 584–586.

McBride, O. W., Battey, J., Hollis, G. F., Swan, D. C., Siebenlist, U., & Leder, P. (1982a) *Nucleic Acids Res. 10*, 8155–8170.

McBride, O. W., Hieter, P. A., Hollis, G. F., Swan, D. C., Otey, M. C., & Leder, P. (1982b) *J. Exp. Med. 155*, 1480–1490.

McBride, O. W., Swan, D. C., Santos, E., Barbacid, M., Tronick, S. R., & Aaronson, S. A. (1982c) *Nature (London) 300*, 773–774.

McBride, O. W., Merry, D., & Givol, D. (1986) *Proc. Natl. Acad. Sci. U.S.A. 83*, 130–134.

McBride, O. W., Rajagopalan, M., & Hatfield, D. (1987) *J. Biol. Chem. 262*, 11163–11166.

Meehan, R. R., Gosden, J. R., Rout, D., Hastie, N. D., Friedberg, T., Adesnik, M., Buckland, R., Van Heyningen, V., Fletcher, J., Spurr, N. K., Sweeney, J., & Wolf, C. R. (1988a) *Am. J. Hum. Genet. 42*, 26–37.

Meehan, R. R., Speed, R. M., Gosden, J. R., Rout, D., Hutton, J. J., Taylor, B. A., Hilkens, J., Kroezen, V., Hilgers, J., Adesnik, M., Friedberg, T., Hastie, N. D., & Wolf, C. R. (1988b) *Proc. Natl. Acad. Sci. U.S.A. 85*, 2662–2666.

Miles, J. S., Spurr, N. K., Gough, A. C., Jowett, T. J., McLaren, A., Brook, J. D., & Wolf, C. R. (1988) *Nucleic Acids. Res. 16*, 5783–5795.

Miles, J. S., Bickmore, W., Brook, J. D., McLaren, A. W., Meehan, R., & Wolf, C. R. (1989) *Nucleic Acids Res. 17*, 2907–2917.

Nadeau, J. H. (1989) *Trends Genet. 5*, 82–86.

Nebert, D. W., Nelson, D. R., Adesnik, M., Coon, M. J., Estabrook, R. W., Gonzalez, F. J., Guengerich, F. P., Gunsalus, I. C., Johnson, E. F., Kemper, B., Levin, W., Phillips, I. R., Sato, R., & Waterman, M. R. (1989) *DNA 8*, 1–13.

Noshiro, M., Lasko, M., Kawajiri, K., & Negishi, M. (1988) *Biochemistry 27*, 6434–6443.

Okino, S. T., Quattrochi, L. C., Pendurthi, U. R., McBride, O. W., & Tukey, R. H. (1987) *J. Biol. Chem. 262*, 16072–16079.

Omura, T., & Sato, R. (1964) *J. Biol. Chem. 239*, 2379–2385.

Pearson, W. R., & Lipman, D. J. (1988) *Proc. Natl. Acad. Sci. U.S.A. 85*, 2444–2448.

Phillips, I. R., Shephard, E. A., Povey, S., Davis, M. B., Kelsey, G., Monteiro, M., West, L. F., & Cowell, J. (1985) *Ann. Hum. Genet. 49*, 267–274.

Sanger, F., Nicklen, S., & Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. U.S.A. 74*, 5463–5467.

Santisteban, I., Povey, S., Shephard, E. A., & Phillips, I. R. (1988) *Ann. Hum. Genet. 52*, 129–135.

Simmons, D. L., & Kasper, C. B. (1983) *J. Biol. Chem. 258*, 9585–9588.

Skoda, R. C., Gonzalez, F. J., Demierre, A., & Meyer, U. A. (1988) *Proc. Natl. Acad. Sci. U.S.A. 85*, 5240–5243.

Song, F. J., Gelboin, H. V., Park, S. S., Yang, C. S., & Gonzalez, F. J. (1986) *J. Biol. Chem. 261*, 16689–16697.

Umeno, M., McBride, O. W., Yang, C. S., Gelboin, H. V., & Gonzalez, F. J. (1988a) *Biochemistry 27*, 9006–9013.

Umeno, M., Song, B. J., Kozak, C. A., Gelboin, H. V., & Gonzalez, F. J. (1988b) *J. Biol. Chem. 263*, 4956–4962.

Watson, C. J., & Jackson, J. F. (1985) in *DNA cloning* (Glover, D. M., Ed.) Vol. 1, pp 79–88, IRL Press, Oxford.

Wilber, W. J., & Lipman, D. J. (1983) *Proc. Natl. Acad. Sci. U.S.A. 80*, 726–730.

Wong, G., Kawajiri, K., & Negishi, M. (1987) *Biochemistry 26*, 8683–8690.

Yamano, S., Aoyama, T., McBride, O. W., Hardwick, J. P., Gelboin, H. V., & Gonzalez, F. J. (1989a) *Mol. Pharmacol. 35*, 83–88.

Yamano, S., Nagata, K., Yamazoe, Y., Kato, R., Gelboin, H. V., & Gonzalez, F. J. (1989b) *Nucleic Acids Res. 17*, 4888.

Yamano, S., Nhamburo, P. T., Aoyama, T., Meyer, U. A., Inaba, T., Kalow, W., Gelboin, H. V., McBride, O. W., & Gonzalez, F. J. (1989c) *Biochemistry 28*, 7340–7348.