Phillips, I. R., Sato, R., & Waterman, M. R. (1989) *DNA 8*, 1-13.
Nelson, D. R., & Strobel, A. W. (1987) *Mol. Biol. Evol. 4*, 572-593.
Omura, T., & Sato, R. (1964) *J. Biol. Chem. 239*, 2379-2385.
Raunio, H., Syngelma, T., Pasanen, M., Juvonen, R., Honkakoski, P., Kairaluoma, M. A., Sotaniemi, E., Lang, M. A., & Pelkonen, O. (1988) *Biochem. Pharmacol. 37*, 3889-3895.
Sanger, R. C., Nicklen, S., & Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. U.S.A. 74*, 5463-5467.
Squires, E. J., & Negishi, M. (1988) *J. Biol. Chem. 263*, 4166-4171.
Towbin, H., Staehlin, T., & Gordin, J. (1979) *Proc. Natl. Acad. Sci. U.S.A. 76*, 4350-4354.
Watson, C. J., & Jackson, J. F. (1985) in *DNA Cloning* (Glover, D. M., Ed.) Vol. 1, pp 79-88, IRL Press, Oxford.
Waxman, D. J. (1988) *Biochem. Pharmacol. 37*, 71-84.
Waxman, D. J., Attisano, C., Guengerich, F. P., & Lapenson, D. P. (1988) *Arch. Biochem. Biophys. 263*, 424-436.
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.

Structure and in Vitro Transcription of the Rat *CYP2A1* and *CYP2A2* Genes and Regional Localization of the *CYP2A* Gene Subfamily on Mouse Chromosome 7

Tamihide Matsunaga,[‡,§] Minoru Nomoto,[‡,∥] Christine A. Kozak,[⊥] and Frank J. Gonzalez*[,‡]

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

Received August 25, 1989; Revised Manuscript Received October 6, 1989

ABSTRACT: The *CYP2A1* and *CYP2A2* genes code for hepatic steroid hydroxylases and differ in their development regulation and expression in male and female rats. In order to explore the mechanism of regulation of these two genes, both genes were isolated and sequenced, their upstream regions compared, and their promoters transcribed in a cell-free system derived from liver. The *CYP2A1* gene was completely sequenced and spanned 12835 bp. The *CYP2A2* gene was sequenced except for 1.5 and 12 kbp in the second and fifth introns, respectively. This gene was about 10 kbp longer than *CYP2A1*. Both genes possess nine exons that displayed overall 93% nucleotide similarity. DNA 4544 and 5529 bp upstream from the *CYP2A1* and *CYP2A2* genes, respectively, was also sequenced, and the transcription start sites were determined. Both genes had typical TATA boxes but did not contain CCAAT boxes within -100 bp of their polymerase start sites. *CYP2A1*, however, contained a reverse CCAAT box between -85 and -90. Search of the Gene Bank revealed a 255 bp region that lies -3 kbp upstream from the transcription start site of *CYP2A1* displaying similarity with retrovirus polymerase. Two regions upstream of *CYP2A2* were also found that displayed 90% sequence similarity with the consensus long interspersed middle repetitive element (LINE). In addition, an unusual 1.6 kbp inserted sequence was detected between -165 and -1779 bp upstream of the *CYP2A2* gene that appears to be a retropseudogene. A nuclear extract derived from adult hepatocytes was used to direct in vitro transcription of the *CYP2A1* and *CYP2A2* gene promoters. Both genes were accurately transcribed in extracts derived from livers of male and female rats. This result is surprising in view of the fact that the *CYP2A1* gene is expressed in adult female rats while the *CYP2A2* gene is expressed exclusively in adult males. The *CYP2A1* promoter was more actively expressed in both extracts than that of *CYP2A2*. By analyzing the segregation pattern of *CYP2A* genes in backcross offspring from an interspecies cross between the laboratory strain NFS/N and the wild mouse *Mus musculus musculus*, the *Cyp2a* subfamily was mapped proximal to the *Gpi-1* locus near the centromere on chromosome 7.

The P450s represent products of a superfamily of genes (Nebert et al., 1989), many of which are under distinct inducer-dependent and developmental control (Gonzalez, 1988). Nine gene families and multiple subfamilies have been described in mammals, and even within closely related members of a single subfamily, P450 genes can be regulated quite differently. For example, the rat *CYP2A* subfamily encodes enzymes having both distinct catalytic activities and regulatory pathways. The most well-studied enzyme in this subfamily is that encoded by the *CYP2A1*[1] gene. This enzyme, designated IIA1, is relatively inactive in the metabolism of various drugs and common P450 substrates but is capable of metabolizing certain steroids such as testosterone (Waxman et al., 1983; Wood et al., 1983) and progesterone (Swinney et al., 1987). IIA1 catalyzes the 7α-hydroxylation of these steroids and also produces low levels of 6α-hydroxysteroid metabolites.

‡ Laboratory of Molecular Carcinogenesis, National Cancer Institute.
§ Present address: Department of Hygienic Chemistry, School of Pharmacy, Hokuriku University, Kanazawa 920-11, Japan.
∥ Present address: Department of Molecular Biology, School of Medicine, University of Occupational and Environmental Health, Fukuoka 807, Japan.
⊥ Laboratory of Molecular Microbiology, National Institute of Allergy and Infectious Diseases.

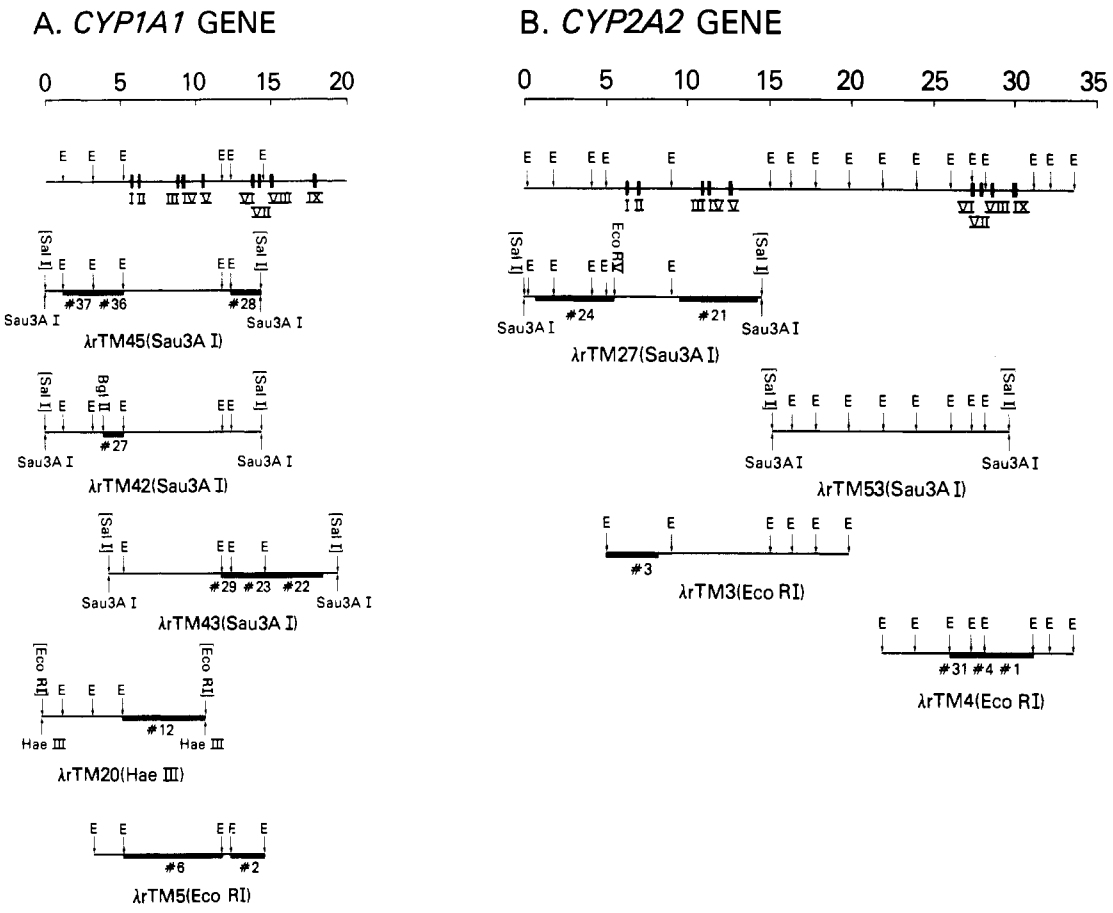


FIGURE 1: Partial restriction maps and structural diagrams of the *CYP2A1* (A) and *CYP2A2* (B) genes. The complete genes are shown above the individual phage clones from which the sequences were derived. The black boxes represent positions of exons. The thick lines delineate sequenced DNA in each phage clone. The numbers below the thick lines designate plasmid subclones. The restriction sites, in parentheses, next to the designated names of each clone, refer to the library from which the clone was isolated (see text). The scale at the top of the figure is in kilobase pairs.

The *CYP2A2* gene product, designated IIA2, was first purified by Jansson et al. (1985) and found to be less specific in its metabolism of testosterone, producing at least 10 distinct metabolites (Jansson et al., 1985; Matsunaga et al., 1988). The *CYP2A1* and *CYP2A2* genes are differentially regulated. The *CYP2A1* gene is activated soon after birth in both males and females and is specifically suppressed in postpubertal males but remains active in females (Nagata et al., 1987). In contrast, the *CYP2A2* gene is activated when males reach puberty but is inactive throughout the life of females (Matsunaga et al., 1988).

To begin to determine the mechanisms of developmental regulation of the *CYP2A* gene subfamily, we isolated and sequenced the *CYP2A1* and *CYP2A2* genes and compared their upstream DNA. Studies were also undertaken to examine if the promoters for these genes could function in a cell-free transcription system. The mouse *CYP2A* subfamily was mapped to a region of chromosome 7 near the centromere using an interspecies backcross.

Materials and Methods

Cloning and Characterization of the CYP2A1 and CYP2A2 Genes. A genomic library was constructed in the vector λEMBL3 using Sprague Dawley DNA partially digested with *Sau*3AI as described previously (Umeno et al., 1988a). The library was screened by plaque hybridization using the IIA1 cDNA probe (Nagata et al., 1987). Two amplified rat genomic libraries, constructed in the vector Charon 4A using *Eco*RI and *Hae*III DNA partial digests, respectively, were also screened. These libraries, constructed in Dr. James Bonner's laboratory at the California Institute of Technology, were provided by Dr. Thomas Sargent of the National Institute of Child Health and Human Development. *Eco*RI fragments, containing the *CYP2A1* and *CYP2A2* genes, were subcloned into pUC9 and then sequenced by using the dideoxy chain termination method (Sanger et al., 1977) in conjunction with shotgun cloning into M13 mp19 (Deininger, 1983). Sequence data were compiled and analyzed with the Beckman Microgenie program.

S1 mapping was performed as described by Sharp et al. (1980). The *CYP2A1* gene was mapped by using an *Acc*I–*Eco*RI fragment (positions 188 to −465, Figure 4). The *CYP2A2* gene was mapped by using an *Acc*I–*Kpn*I fragment (positions 188 to −254, Figure 4). The fragments were labeled with T4 polynucleotide kinase and [γ-^{32}P]ATP (6000 Ci/

[1] The nomenclature used in this report is that described by Nebert et al. (1989). The *CYP2A1* gene product has been designated IIA1 (Matsunaga et al., 1988), P450a (Ryan et al., 1982), P450 3 (Waxman et al., 1983), and P450 UT-F (Guengerich et al., 1982). The *CYP2A2* gene product has been called P450 RLM2 (Jansson et al., 1985) and IIA2 (Matsunaga et al., 1988). According to the P450 nomenclature system for chromosomal loci, the human and mouse subfamilies are designated *CYP2A* and *Cyp2a*, respectively, and genes encoding IIA1 and IIA2 are denoted *CYP2A1* and *CYP2A2*, respectively (Nebert et al., 1989).

mmol; New England Nuclear Corp.). The end-labeled fragments were then annealed with poly(A) RNA and subjected to S1 nuclease digestion. RNA was isolated by the method of Chirgwin et al. (1979), and the poly(A) fraction was purified by using oligo(dT)–cellulose chromatography (Aviv & Leder, 1972). Primer extension was carried out using a 22-mer, 3′-AACCGGAGGGACTCGCAGTACG-5′ (complementary to +62 to +83 in the *CYP2A1* gene, Figure 4). The primer was 5′ labeled with [γ-^{32}P]ATP, annealed with 10 μg of poly(A) RNA, and then extended with reverse transcriptase (Life Sciences, Inc., St. Petersburg, FL). Products of S1 mapping and primer extension reactions were electrophoresed on DNA sequencing gels in parallel with a ladder of sequenced DNA derived from the fragments used for S1 mapping and M13 mp18 clones, respectively.

In Vitro Transcription of the CYP2A1 and CYP2A2 Genes. Nuclear transcription extracts were prepared by using a modification of the procedure of Gorski et al. (1986). Livers from 15-week-old male rats were homogenized in the presence of a mixture of protease inhibitors (10 μg/mL trypsin inhibitor, 5 μg/mL aprotinin, 2 μg/mL antipine, 2 μg/mL chymostatin, 2 μg/mL pepstatin A, and 2 μg/mL leupeptin). Nuclear extracts were quick-frozen in liquid nitrogen and stored in aliquots at −70 °C until use. Transcription reactions (30 μL) contained 90 ng of supercoiled plasmid DNA template (test plasmid), 10 ng of MSV-TK construct (control plasmid), 1 μg of pUC 19 (nonspecific competitor), and 1.6 mg/mL nuclear extract protein in a buffer containing 6 mM $MgCl_2$, 50 mM KCl, 0.6 mM each of ATP, CTP, GTP, and UTP, and 30 units of RNasin (Promega, Madison, WI). The reactions were carried out at 30 °C for 60 min and then terminated by the addition of 170 μL of a stop buffer containing 1% SDS, 10 mM EDTA, 0.2 mg/mL proteinase K, 50 μg/mL yeast tRNA, and 0.3 M sodium acetate, pH 7.5. The reaction mixtures were further incubated at 50 °C for 30 min, extracted with 1:1 phenol/chloroform, and then precipitated with ethanol. The dried pellets were resuspended in 10.5 μL of distilled H_2O. Kinased labeled primers (0.15 ng of 22-mer primer for test plasmid and 0.05 ng of 25-mer TK primer at specific activities of 2×10^9 dpm/μg, respectively) and a hybridization buffer were added to final 0.25 M KCl, 20 mM Tris-HCl, pH 8.0, and 1 mM EDTA (final volume 25 μL). The annealing reactions were carried out at 65 °C for 90 min, and then the mixtures were allowed to cool to room temperature. For primer extension reactions, 50 μL of a buffer containing 15 mM DTT, 150 μg/mL actinomycin D, 0.5 mM of all four dNTPs, 20 mM $MgCl_2$, 40 mM Tris-HCl, pH 8.3, and 5 units of reverse transcriptase was added to the transcription reaction mixture (final volume 75 μL). The reactions were carried out at 37 °C for 45 min and terminated by addition of 75 μL of a stop buffer containing 0.6 M sodium acetate, pH 5.2, and 20 mM EDTA. The mixtures were extracted with 1:1 phenol/chloroform and then precipitated with ethanol. Dried pellets were dissolved in 80% deionized formamide containing 0.01% xylene cyanol and 0.01% bromophenol blue and loaded on a 8% polyacrylamide/50% urea-containing 1.0-mm sequencing gel.

Mapping of the CYP2A Locus in Mouse. Female NFS/N mice were obtained from the Small Animals Section, National Institutes of Health. *Mus musculus musculus* mice, from Skive, Denmark, were kindly provided by Dr. Michael Potter of the National Cancer Institute from his colony at Hazelton laboratories. Female NFS/N were mated with *M. m. musculus*, and F1 progeny were mated with male *M. m. musculus* to obtain backcross mice. DNA isolated from livers of NSF/N and *M. m. musculus* parental and backcross was digested with several restriction enzymes and subjected to Southern blotting and hybridization with the IIA1 cDNA probe. *Eco*RI was found to generate specific fragments that could be distinguished by size between the two mouse strains. A *Bam*HI RFLP was used to type the distribution of *Fes* using v-Fes as a hybridization probe. This probe was provided by Dr. C. Sher of St. Judes Hospital, Memphis, TN. Kidney extracts from the same mice were typed for *Gpi-1* by starch gel electrophoresis.

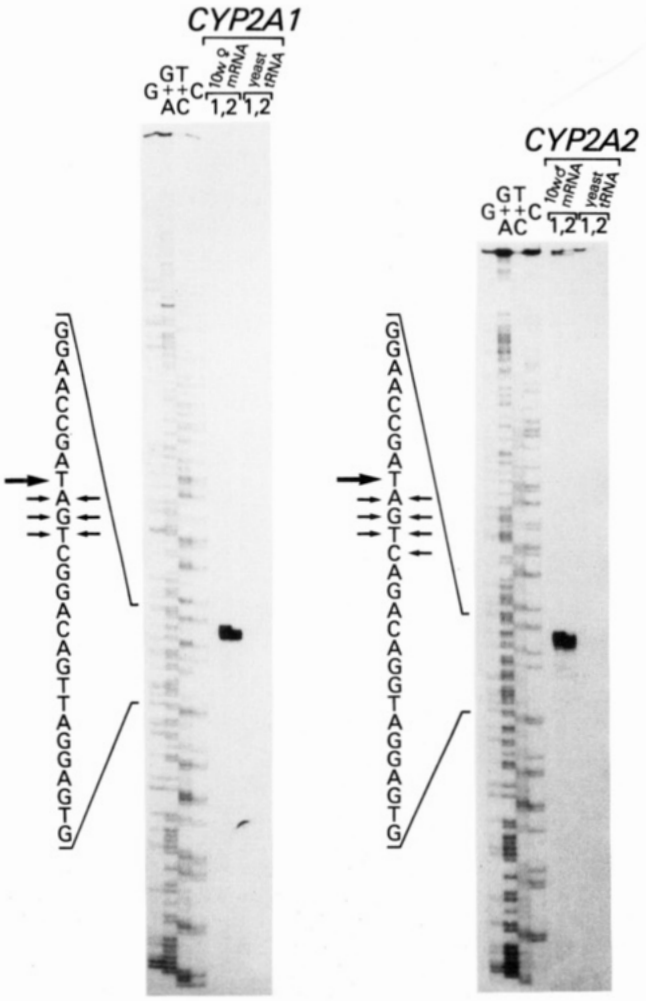


FIGURE 2: S1 nuclease analysis of the *CYP2A1* and *CYP2A2* genes. Kinased single-end-labeled fragments, produced as described under Materials and Methods, were hybridized with adult male or adult female rat liver poly(A)-selected RNA (10 μg) and digested with 20 units (lane 1) and 200 units (lane 2) of S1 nuclease. Yeast tRNA was used as a control. The end-labeled fragments were also sequenced by the Maxam and Gilbert (1980) procedure, and the sequenced fragments were electrophoresed concurrently with the S1-protected fragments. The sequence listed vertically on the edges of each gel is read directly from the Maxam and Gilbert ladder and is the complement of the mRNA strand. The arrows denote the position in the sequence corresponding to each S1 fragment.

Results and Discussion

*Isolation of the Rat CYP2A1 and CYP2A2 Genes. Sau*3AI, *Eco*RI, and *Hae*III genomic libraries were screened to isolate the complete *CYP2A1* and *CYP2A2* genes. The *Eco*RI restriction maps and the intron and exon structures of the *CYP2A1* and *CYP2A2* genes are shown in Figure 1A and Figure 1B, respectively. These data were generated from restriction mapping and sequencing the individual phage clones shown below the composite maps in Figure 1. The complete genes were cloned by screening three separate libraries that

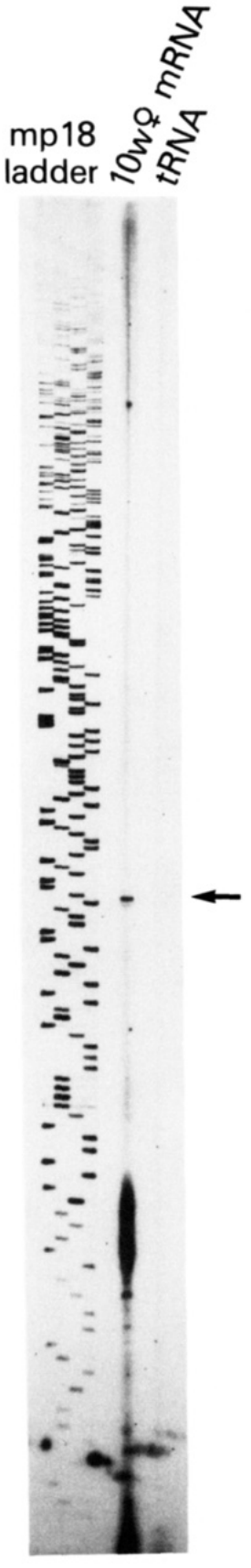


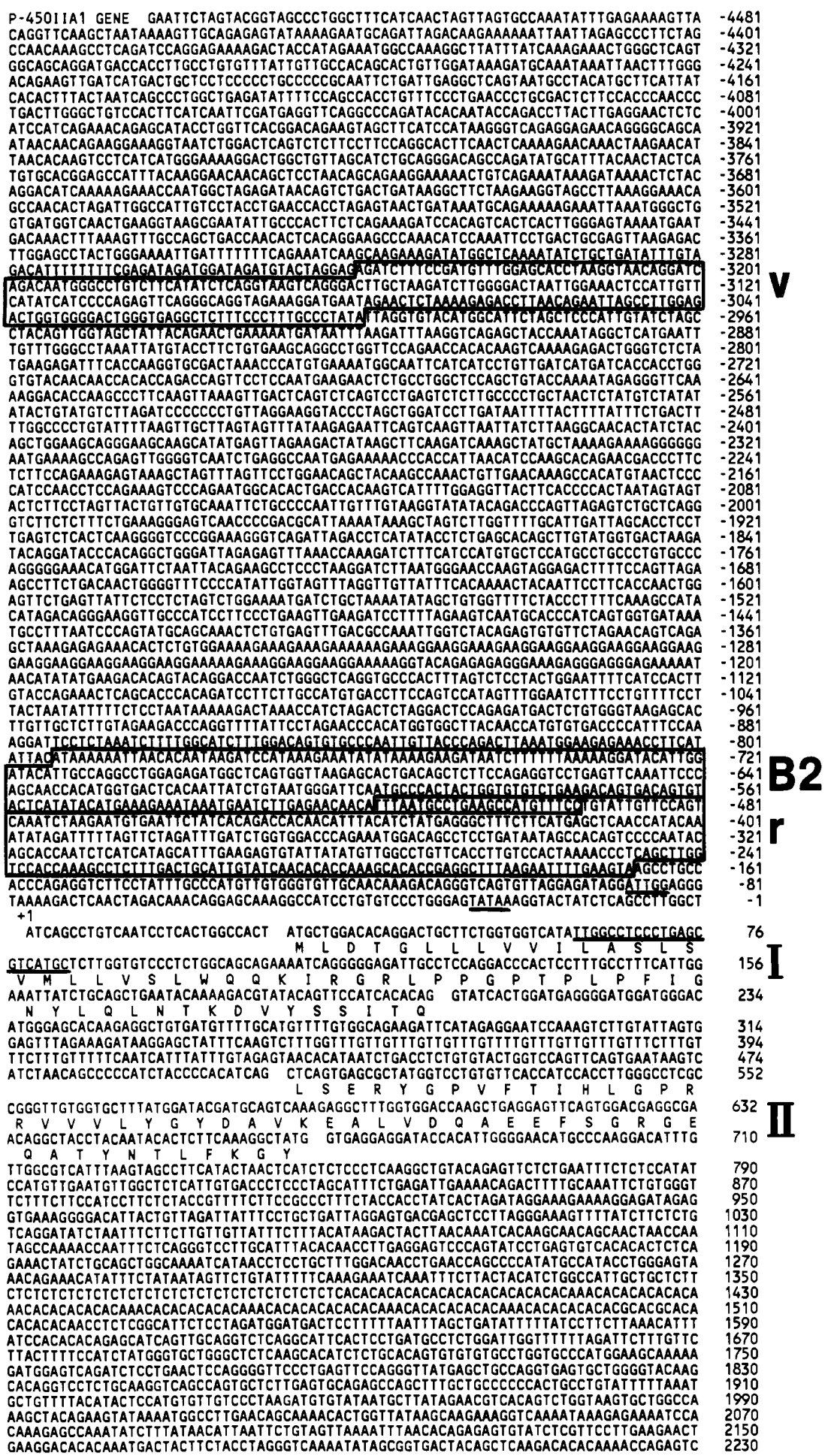
FIGURE 3: Primer extension analysis of the *CYP2A1* gene. The position of the primer used for primer extension is shown in Figure 4. Poly(A)-selected RNA (10 μg) from adult female rats was used. Yeast tRNA was used as a control. A sequencing ladder of M13 mp18 was used as a size standard and electrophoresed concurrently with the extended products. The arrow shows the extended fragment of 82 nucleotides that corresponds to a T residue in the *CYP2A1* gene (Figure 4).

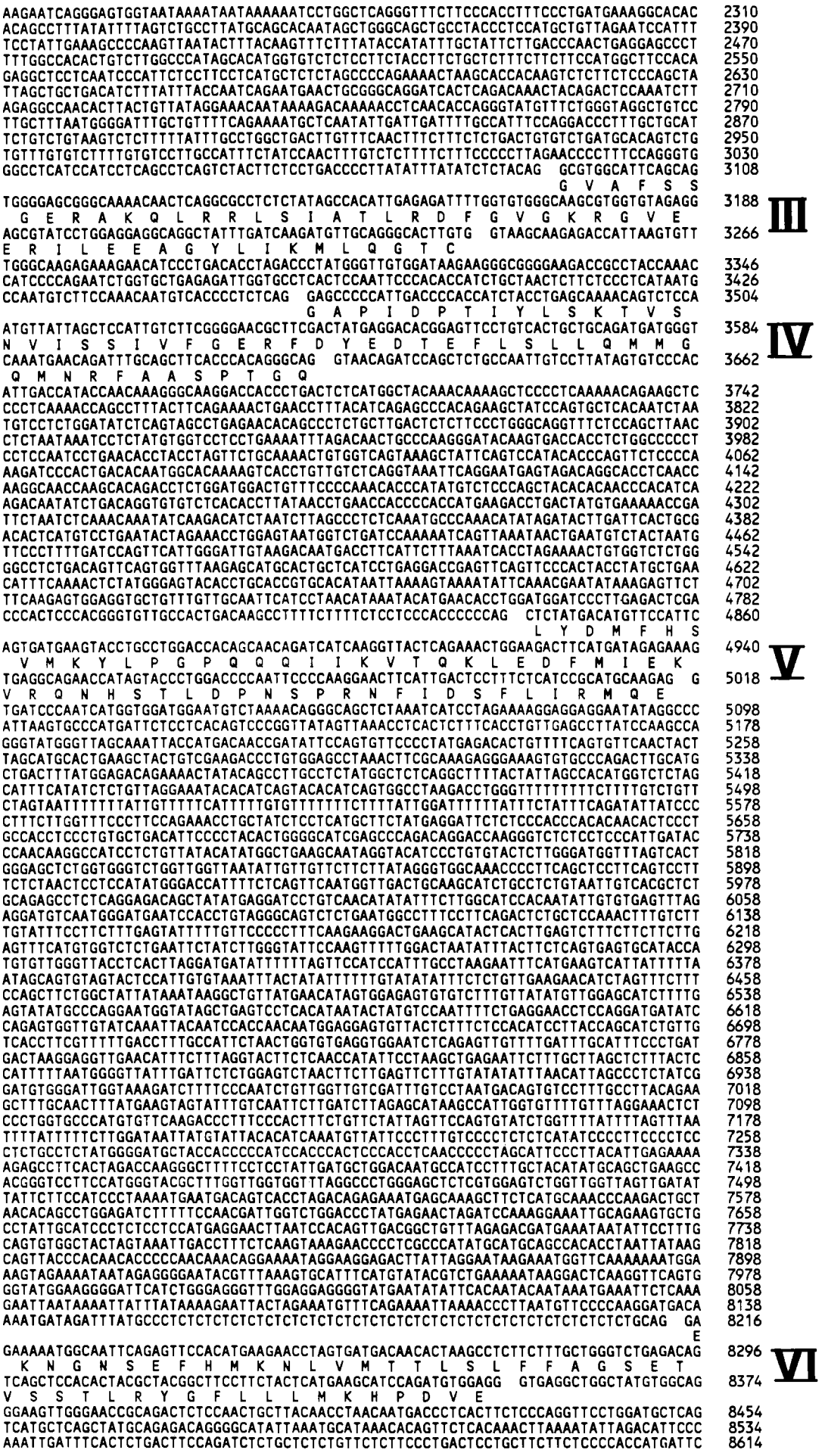
were constructed by using different restriction enzymes. Both genes possess nine exons, and the *CYP2A2* gene is almost twice as long as *CYP2A1*. This is primarily due to a 14 kbp fifth intron in *CYP2A2*. Subclones of each gene were produced in pUC9 and sequenced by using the shotgun cloning and dideoxy methods. The regions of each phage clone sequenced are shown by thick lines in Figure 1.

Determination of the Transcription Start Sites of the CYP2A1 and CYP2A2 Genes. S1 mapping was carried out to determine the transcription start sites of the *CYP2A* genes. The *CYP2A1* gene is expressed in adult females while the *CYP2A2* gene is active only in adult males (Matsunaga et al., 1988). Therefore, poly(A) RNAs, isolated from livers of adult female and male rats, respectively, were used as templates to map the start sites of the *CYP2A1* and *CYP2A2* genes. The adult female RNA protected the *Acc*I–*Eco*RI fragment that was derived from the 5′ end of the *CYP2A1* gene (Figure 2, left panel). Four fragments were produced of 185, 186, 187, and 188 nucleotides after digestion with 20 units of S1 nuclease. Digestion with 200 units of enzyme resulted in elimination of the longest 188-nucleotide fragment. Similar results were obtained when a *Acc*I–*Kpn*I fragment derived from the *CYP2A2* gene was used to protect adult male RNA (Figure 2, right panel). Primer extension analysis of the *CYP2A1* gene was also carried out. An extended fragment of about 82 nucleotides was detected (Figure 3) that appears to correspond in size to the the largest fragment of the S1 mapping products after treatment with 200 units of S1 nuclease (Figure 2). The primer extension result suggests that the multiple fragments detected by S1 mapping are due to S1 "nibbling". It should be noted, however, that the size of the primer-extended fragment cannot be precisely determined since it was measured using a sequenced M13 template and mobilities of DNAs, even on denaturing gels, are known to be sequence dependent. Therefore, this fragment could conceivably be 81 or 83 nucleotides in length. In any case, the start site of transcription was assigned to the longest S1-protected fragments of each gene corresponding to an adenine residue (thick arrow, Figure 2). The start site of each gene is preceded by typical TATA boxes at −24 bp (Figures 4 and 5).

Sequence of the CYP2A1 Gene. The sequence of the *CYP2A1* gene spans 12835 bp from the start of transcription to the first polyadenylation site (Figure 4). Upstream and downstream DNAs of 4544 and 1440 bp, respectively, were also sequenced. As noted in an earlier study, we found that two mRNAs of 2.0 and 3.0 kb reacted with an oligonucleotide probe to IIA1 (Matsunaga et al., 1988). Analysis of the *CYP2A1* gene sequence data revealed another sequence, AATAAG, that is similar to the consensus polyadenylation signal "AATAAA", at 775–780 bp downstream from the polyadenylation site of the putative 2.0-kb transcript previously determined from the IIA1 cDNA sequence (Matsunaga et al., 1988). This may function as a second polyadenylation site giving rise to the 3.0-kb mRNA. The polyadenylation signal producing the 3.0-kb transcript appears to be stronger than the signal giving rise to the 2.0-kb mRNA, since the former transcript is much more abundant than the latter in both untreated and 3-methylcholanthrene-treated rat livers (Matsunaga et al., 1988). The role of other neighboring sequences in alternative polyadenylation of the *CYP2A1* gene transcript needs to be further analyzed.

Upstream sequences were also examined. The *CYP2A1* gene contained a typical TATA box at −27 to −24 bp and a reverse CCAAT box at −85 to −90 bp upstream of the start site. By searching the Gene Bank, we found a sequence of 277 bp displaying 76% similarity to a mouse B2 repeat be-


P-450IIA1 GENE
V
B2
r
I
II


III
IV
V
VI

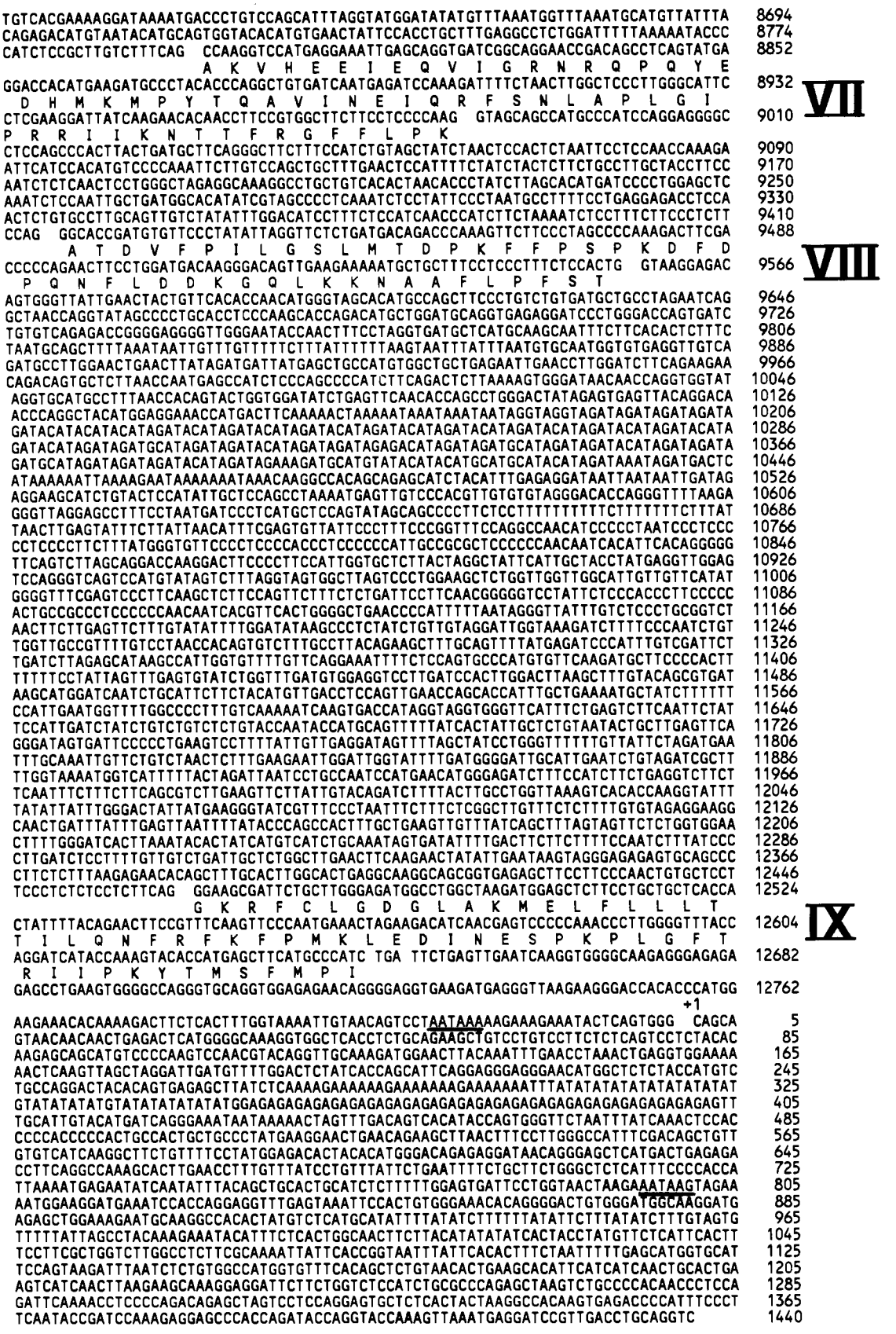
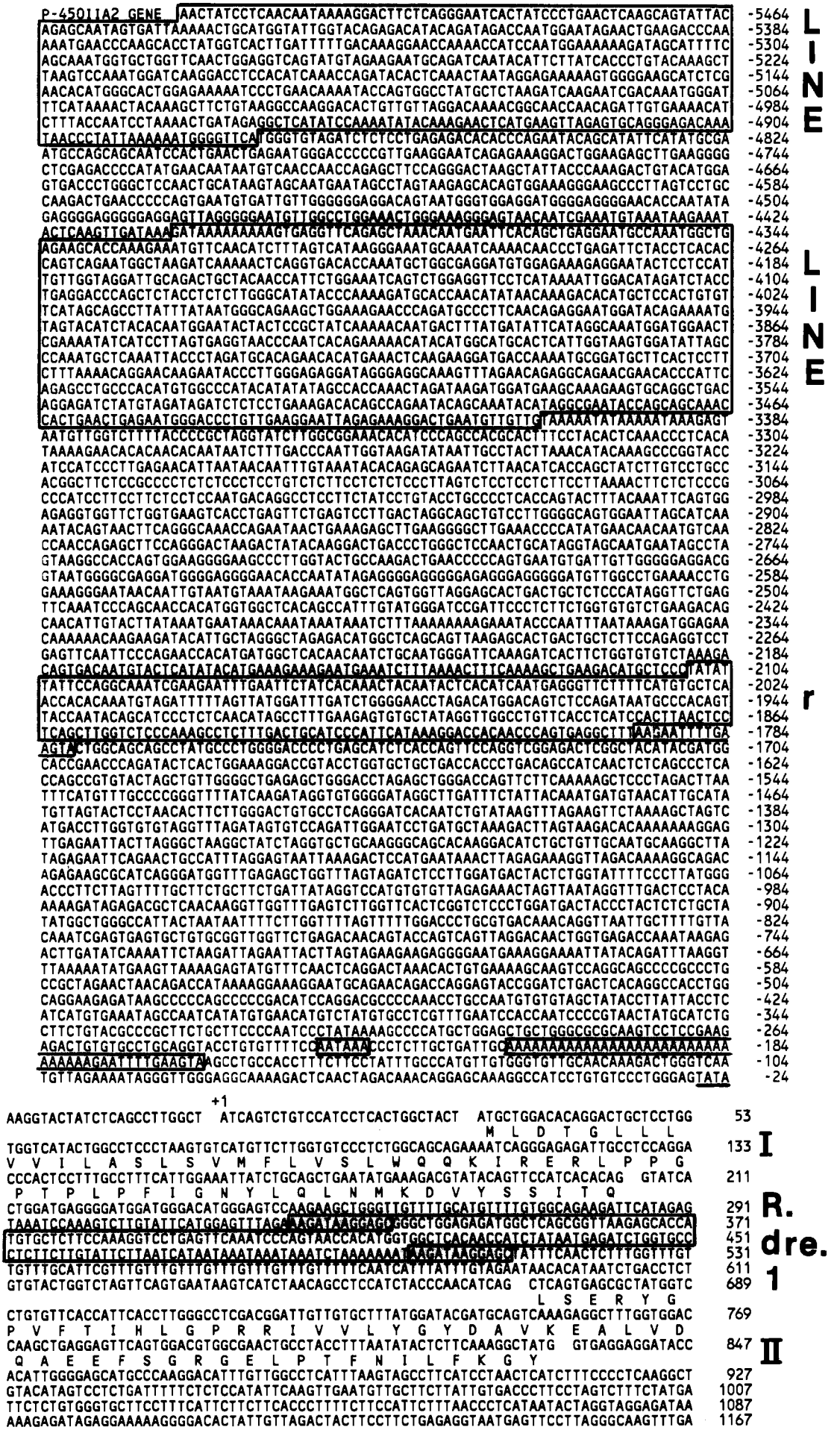
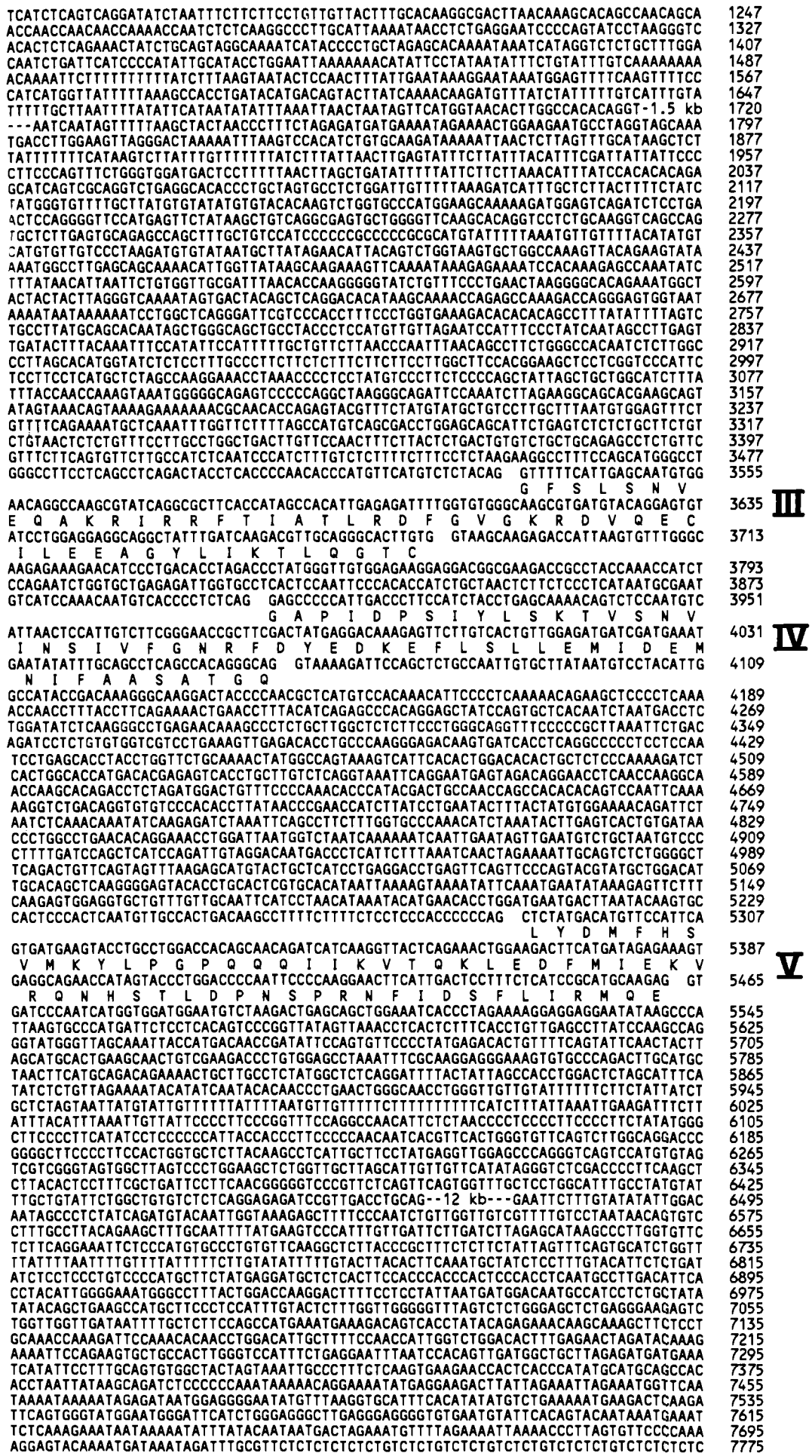


FIGURE 4: Sequence of the *CYP2A1* gene. The complete sequence of the *CYP2A1* gene is shown. The start site at +1 was determined by S1 mapping as shown in Figure 2. The exons are labeled by Roman numerals at the right of the figure. Regions of interest discussed in the text are boxed. These include a viral sequence (v), a B2 repetitive element (B2), and a sequence found upstream of the *CYP2A2* gene (r). The TATA box, the reverse CCAAT box, the sequence from which the primer for primer extension was derived, and the polyadenylation signals (AATAAA and AATAAG) are underlined. The 3′ flanking sequence of the gene is numbered beginning with +1.

ginning at residue −519. A second sequence of 255 bp was found at −3000 that displayed about 73% similarity to the 3′ end of the *pol* gene of the murine leukemia retroviral genomes (Herr, 1984; Etzerodt et al., 1984). This sequence appears


P-450IIA2 GENE
LINE
LINE
r
+1
I
R.
dre.
1
II


III
IV
V

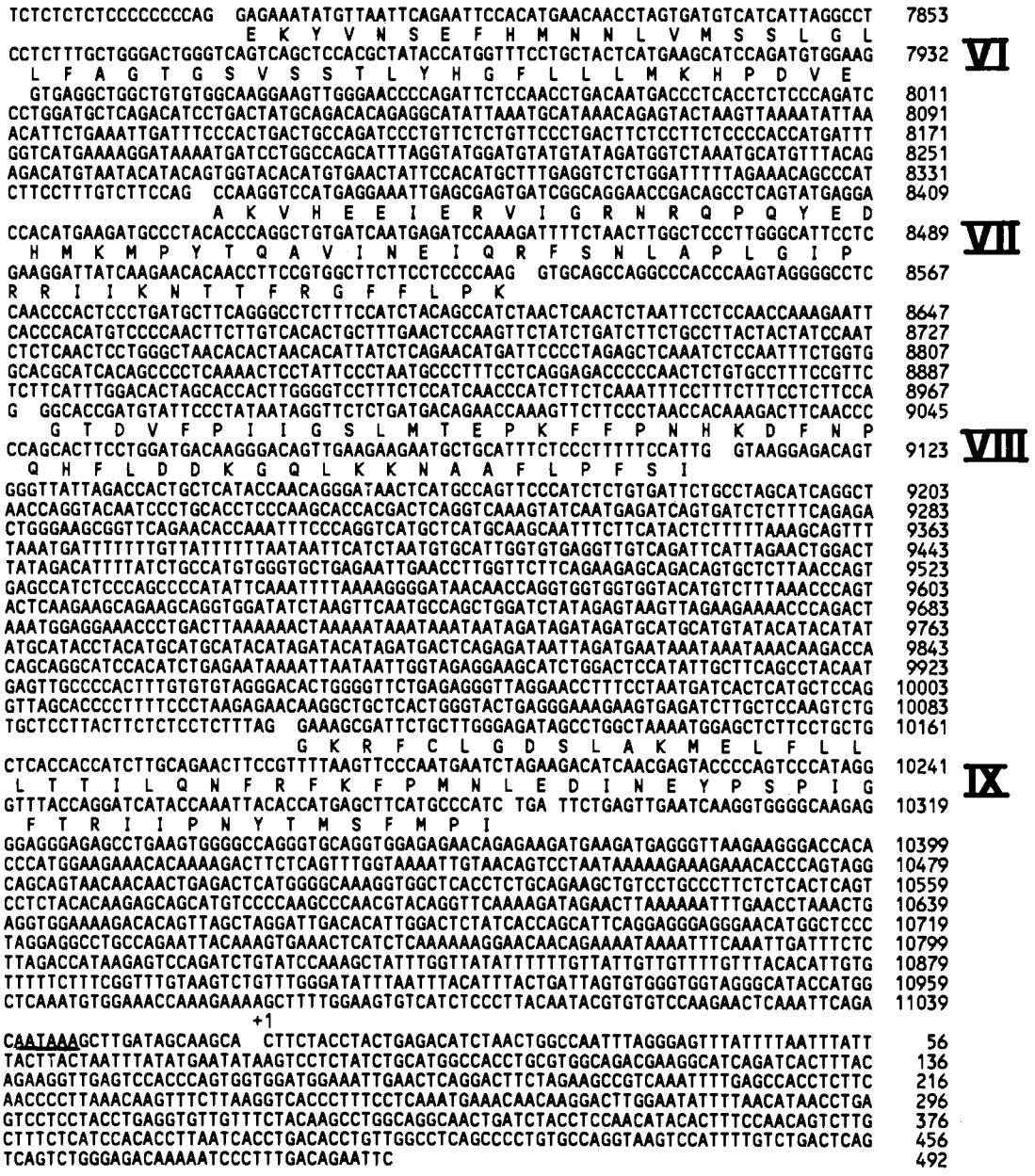


FIGURE 5: Sequence of the *CYP2A2* gene. The sequence of the *CYP2A2* gene is shown, and the exons are labeled by Roman numerals at the right of the figure. The start site at +1 was determined by S1 mapping as described in Figure 2. Regions of interest, including the LINE sequence, a sequence found upstream of the *CYP2A1* gene (r), an R.dre.1 repeat, an upstream poly(A) tract and polyadenylation signal, AATAAA, and direct repeats flanking a putative retroposon are enclosed by lines. The TATA box and the downstream *CYP2A2* gene polyadenylation signal are underlined. The 3′ flanking region of the gene is numbered beginning with +1.

to be a fragment from a rat retrovirus that may have inserted via a transpositional mechanism. The functional significances of the B2 repeat and retroviral fragment, if any, are currently unknown.

Sequence of the CYP2A2 Gene. The sequence of most of the *CYP2A2* gene was determined except for approximately 1.5 and 12 kbp in the second and fifth introns, respectively (Figure 5). Upstream and downstream DNAs of 5529 and 494 bp, respectively, were also sequenced. The *CYP2A2* gene also had a typical TATA box preceding its transcription start site but lacked a CCAAT box. Comparison of the *CYP2A1* and *CYP2A2* upstream DNAs revealed the presence of a 1.6 kbp insertion of DNA between −165 and −1779 upstream of the *CYP2A2* gene (Figure 5, also see Figure 6). This insertion can be visualized by the movement of the r fox found in *CYP2A1* (Figure 4) to further upstream in *CYP2A2* (Figure 5). The 1.6 kbp sequence is flanked by perfect 15 bp direct repeats 5′-AAGAATTTTGAAGTA-3′. It also contains a polyadenylation signal, AATAAA, followed by a segment of 32 adenylate residues immediately before the direct repeat. This sequence did not display similarity to any sequence in the Gene Bank, and no large open reading frame was uncovered. The 1.6 kbp sequence appears to be a retroposon or processed pseudogene that was inserted just upstream of the *CYP2A2* gene after the duplication of the *CYP2A2* locus.

Another notable finding was the presence of segments of the long interspersed middle repetitive element (LINE) upstream of the *CYP2A2* gene (Figure 5). Two segments, spanning 649–1004 bp, were found that were noncollinear with and from different regions of the consensus LINE sequence of Soares et al. (1985). Both are in the same orientation as the consensus sequence. They may have inserted independ-

Table I: Comparison of the Introns and Exons of the *CYP2A1* and *CYP2A2* Genes[a]

	length (bp)		overlap	%
	CYP2A1	*CYP2A2*	(bp)	similarities
exon I	177	177		95
intron 1	299	465	181	90
exon II	162	162		94
intron 2	2425	2711	1571	80
exon III	150	150		86
intron 3	217	217	217	97
exon IV	162	162		89
intron 4	1220	1224	1220	87
exon V	177	177		100
intron 5[b]	3197	1010	1010	72
		1321	1321	84
exon VI	138	138		85
intron 6	441	415	285	85
exon VII	189	189		99
intron 7	432	432	432	71
exon VIII	141	141		91
intron 8	2908	998	421	76
exon IX	180	180		93

[a] Sequences were compared by using the Microgenie program. [b] Two separate sequences of the *CYP2A2* gene fifth intron were individually aligned with the fifth intron the *CYP2A1* gene.

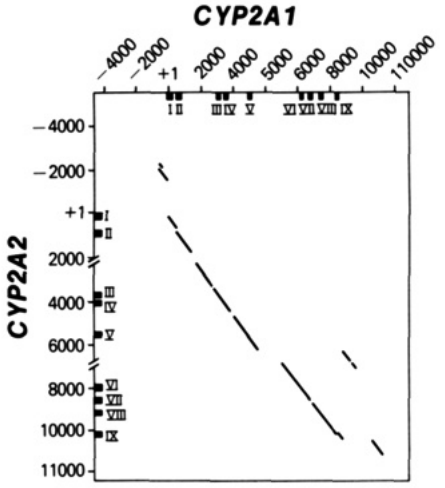


FIGURE 6: Dot matrix analysis of the *CYP2A1* and *CYP2A2* genes. Dot matrix comparisons were generated from the data in Figures 4 and 5 using the Microgenie program. A dot is produced whenever a segment of 15 residues exists between the two genes that displays ≥78% sequence similarity.

ently, upstream of the *CYP2A2* gene, or may have inserted as a single unit and then undergone rearrangement. The functional significance of these repetitive elements in the eukaryote genome is still controversial, although some LINE sequences are transcribed.

Sequence alignments of the first intron in the two genes revealed an extra 174 bp in the *CYP2A2* gene (Figure 5). This represents an insertion of an R.dre.1 repetitive element (Rogers, 1985) that is flanked by a 12 bp direct repeat 5′-AAGATAAGGAGC-3′.

Gene Conversions. In an earlier report, we uncovered regions of high nucleotide similarity interspersed with regions of low similarity between the *CYP1A1* and *CYP1A2* cDNAs (Matsunaga et al., 1988). Dot matrix analysis of the *CYP2A1* and *CYP2A2* gene sequences further confirmed that the similarities between these two genes are quite high and extend in the introns (Figure 6). These data indicate that a rearrangement has occurred between DNA in intron 5 of *CYP2A2* and the 3′ flanking region of the *CYP2A1* gene.

Alignments of sequences within individual exons revealed

LANE	1	2	3	4	5	6	7	8	9	10
TEMPLATE	A1 Tk	A1 —	A1 Tk	A1 Tk	— Tk	A2 Tk	A2 —	A2 Tk	A2 Tk	— Tk
PRIMER	A1 Tk	A1 —	A1 Tk	A2 Tk	— Tk	A2 Tk	A2 —	A2 Tk	A1 Tk	— Tk
EXTRACT	—	+	+	+	+	—	+	+	+	+

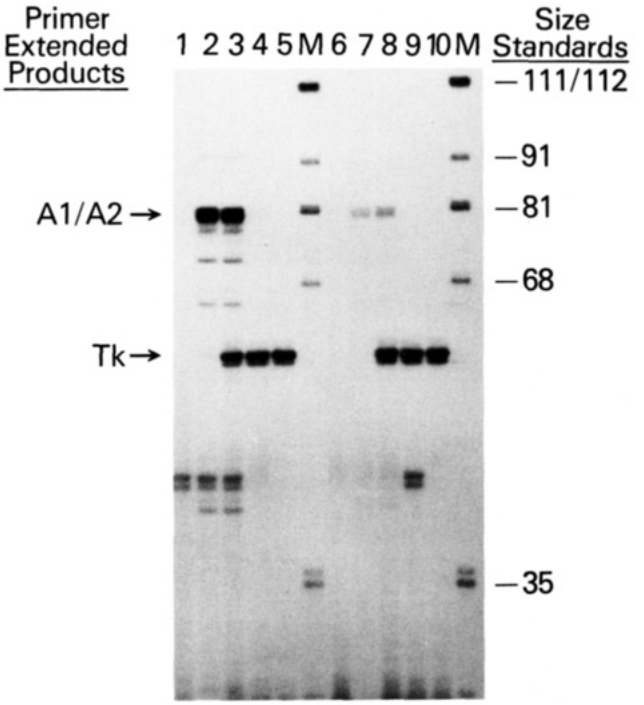


FIGURE 7: Transcription of the *CYP2A1* and *CYP2A2* genes. The *CYP2A1* and *CYP2A2* templates were subcloned from λrTM45 and λrTM27, respectively (Figure 1). The A1 and A2 primer is the same as that used for primer extension in Figure 3. The TK primer was the same as that described by Graves et al. (1985). The molecular weight size standards (M) were derived from a *HpaII* digest of pUC19 and labeled by Klenow polymerase. Two additional fragments of 80 and 90 bp were prepared from restriction fragments of the *CYP2A1* gene. The contents of each transcription reaction and the primers used to analyze the DNAs are shown at the top of the figure. Primer-extended fragments were electrophoresed on an 8% polyacrylamide gel containing 50% urea.

that exons V and VII displayed 100% and 99% similarity, respectively, while all other exons demonstrated from 85% to 95% similarity (Table I). These data indicated that gene conversions may have occurred between portions of the *CYP2A1* and *CYP2A2* that encompased exons V and VII. The introns bracketing these two exons, however, do not display as high a similarity as the exonic DNA. These data indicate either that the putative gene conversions did not involve introns or that these conversions occurred sufficiently long ago so that the introns have begun to diverge. Without clear evidence of intron involvement, the role of gene conversions in generating the high similarity between the two genes within exons V and VII remains speculative. Therefore, it cannot be ruled out that the high nucleotide similarities of exonic DNA are due to functional conservation of mRNA structure.

The role of gene conversions in P450 evolution is still unclear. It has been suggested that conversions are occurring continuously along with other processes of DNA turnover and that converted segments within genes are fixed in populations by genetic drift, natural selection, or some other mechanism (Dover, 1987). The possibility exists that converted segments within P450 genes could enhance the catalytic capacities of

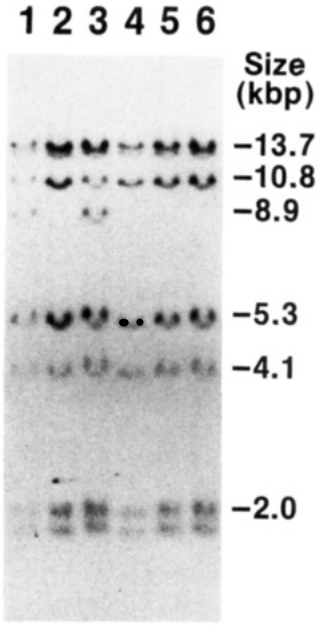


FIGURE 8: Southern blot analysis of DNAs from livers of several backcross mice. The DNA was digested with *Eco*RI, and the blots were hybridized with nick-translated *CYP2A1* cDNA insert. The 9 kbp fragment (arrow) is present in NFS/N parental DNA and not in *M. m. musculus* DNA. Thirty-nine out of 72 backcross mice examined possessed this fragment.

these enzymes toward a new environmental or dietary toxin, hence rendering the organism a selective advantage.

Another noteworthy finding was the low 71% similarity of intron 7 between the two genes (Table I). This is surprising since exons VII and VIII, that bracket this intron, display 99% and 91% similarities, respectively, between *CYP2A1* and *CYP2A2*. Perhaps intron 7 underwent a gene conversion with the seventh intron of yet a third gene in the *CYP2A* subfamily. Alternatively, this intron may have diverged more rapidly than the others.

In Vitro Transcription of the CYP2A1 and CYP2A2 Genes. To determine if the *CYP2A1* and *CYP2A2* promoters were competent in a cell-free nuclear transcription extract, in vitro transcription was carried out using templates derived from the genomic clones. A plasmid construct, derived from *CYP2A1*, containing about −6 kbp upstream and 1.5 kbp downstream, including the first and second exons and part of the second intron, was used to determine the transcriptional activity of the *CYP2A1* promoter. To determine the activity of the *CYP2A2* promoter, a plasmid containing −6 kbp of upstream DNA and 1.5 kbp of downstream DNA was analyzed as a transcription template. Transcription products were measured by using a primer extension assay, which results in precise quantitation of correctly initiated transcripts. As an internal control, the human Herpes simplex virus thymidine kinase (TK) gene under control of the MSV LTR promoter (designated MSV-TK) was analyzed (Graves et al., 1985).

Both genes were accurately transcribed in the cell-free extract derived from adult male liver (Figure 7). In the absence of extract, no RNA was produced from either the *CYP2A1* or the *CYP2A2* promoter (Figure 7, lanes 1 and 6, respectively). When extract was included, primer-extended fragments of 81 and 82 bp, corresponding to the cluster of start sites detected in Figures 2 and 3, were found (Figure 7, lanes 2, 3, 7, and 8). The presence of these extended fragments was absolutely dependent on the appropriate primer as shown in Figure 7, lanes 4 and 9. The *CYP2A1* promoter was about 5-fold more active than the *CYP2A2* promoter. This is most apparent in Figure 7, lanes 3 and 8, in which the MSV-TK promoter is transcribed simultaneously with the *CYP2A1* and *CYP2A2* promoters, respectively. A similar ratio of relative transcription activities between *CYP2A1* and *CYP2A2* genes (or promoters) was also observed when much smaller constructs were used as templates, for example, plasmid constructs containing −74 to +169 from the transcription start site.[2] Since the −74 to +169 construct of the *CYP2A1* gene does not contain the reverse CCAAT sequence described previously, it is questionable whether this sequence is important in the basal transcription activities of the *CYP2A1* gene, at least in the liver in vitro transcription system.

Table II: Linkage Analysis of the *Cyp2a* Gene Subfamily in an Interspecies Backcross[a]

	inheritance of NFS/N allele			no. of mice
	Cypa2a	*Gpi-1*	*Fes*	
parental	+	+	+	21
	−	−	−	14
single recombinant	+	+	+	3
	−	−	+	7
	+	−	−	0
	−	+	+	2

[a] Forty-seven mice of the backcross (NFS/N × *M. m. musculus*) F1 × *M. m. musculus* were typed for RFLPs at *Cyp2a* and *Fes* by Southern blot analysis (Figure 8) and for isozyme variants at *Gpi-1* by starch gel electrophoresis of kidney extracts. The recombination frequencies were the following: *Cyp2a*, *Gpi-1*, $r = 2/47 = 4.3 \pm 2.9$ cM; *Cyp2a*, *Fes*, $r = 12/47 = 25.5 \pm 6.3$ cM; *Gpi-1*, *Fes*, $r = 10/47 = 21.2 \pm 5.9$ cM.

It was previously established that the *CYP2A1* gene was expressed in livers of male and female rats soon after birth and is then suppressed in male liver at the onset of puberty, while the *CYP2A2* gene is expressed only in adult males and not in females (Matsunaga et al., 1988). However, in the in vitro transcription assay, both genes are expressed in adult male liver nuclear extracts. The same result was found in extracts prepared from adult female rats, suggesting that the in vitro system does not reflect the in vivo transcriptional activities of these genes. The possibility exists that the extract system of Gorski et al. (1986) does not contain all the factors necessary for regulation of the *CYP2A* genes. Alternatively, perhaps a higher ordered chromatin structure or remote sequence elements are needed in vivo that cannot be achieved in vitro. In any case, we believe that high basal activity that is removed from regulatory constraints is being expressed in the nuclear transcription extract system of Gorski et al. (1986).

Linkage Analysis Mapping of the CYP2A Locus in Mice. The *CYP2A* locus was regionally localized on mouse chromosome 7 using the linkage analysis strategy of an interspecies backcross of NFS/N and *M. m. musculus* mice as described by Guenet (1986). *Eco*RI digestion of DNAs from parental mice produced multiple fragments reactive with the *CYP2A1* cDNA probe with a single 9 kbp fragment detected only in the NFS/N parental DNA (Figure 8). This fragment was found in 39 of 72 backcross mice, and its inheritance was compared with that of other markers on chromosome 7. These data revealed that the *Cyp2a* locus in mice is closely linked to *Gpi-1* with the gene order centromere *Cyp2a-Gpi-1-Fes* (Table II). This result supports and extends an earlier study using the *CYP2A3* probe in which the *Cyp2a* locus was mapped to chromosome 7 using the somatic cell hybrid mapping strategy (Kimura et al., 1989).

It is likely that the mouse *CYP2A* family is tightly linked to the *CYP2B* subfamily since both gene clusters are found

[2] Nomoto and Gonzalez, unpublished results.

on a single 350 kbp fragment of human chromosome 19 (Miles et al., 1989). The *CYP2E* subfamily, on the other hand, is found on chromosome 7 in mouse but, in contrast to the *CYP2A* and *CYP2B* subfamilies, is located on human chromosome 10 (Umeno et al., 1988b). On the basis of conservation of linkage groups between human and mouse (Nadeau, 1989), these data indicate that the *CYP2E* subfamily is not near to the *CYP2A* and *CYP2B* subfamilies. Two other *CYP2* subfamilies, *CYP2C* and *CYP2D*, are located on different mouse and human chromosomes. The *CYP2C* subfamily is located on human chromosome 10 (Umeno et al., 1988b) and mouse chromosome 19 (Meehan et al., 1988). The *CYP2D* subfamily is found on human chromosome 22 (Gonzalez et al., 1988) and mouse chromosome 15 (Gonzalez et al., 1987). These types of comparisons should provide more insight into the evolution and divergence of P450 gene clusters within the *CYP2* gene family.

Acknowledgments

We thank Dr. Peter Johnson for supplying the MSV-TK construct and for helpful suggestions on the in vitro transcription experiments.

Registry No. DNA (rat liver gene CYP2A1), 124511-64-4; cytochrome P450 (rat liver isoform IIA1 protein moiety reduced), 110069-47-1; DNA (rat liver gene CYP2A2 gene coding region), 120432-64-6; cytochrome P450 (rat liver isoform IIA2 protein moiety reduced), 120432-25-9; cytochrome P450, 9035-51-2; steroid hydroxylase, 9044-53-5; RNA (rat liver clone λrTM45/λrTM5/λrTM43 cytochrome P450 isoform IIA1 specifying messenger), 124511-65-5; RNA (rat liver clone λrTM27/λrTM3/λrTM4 cytochrome P450 isoform IIA2 specifying messenger, 124511-66-6.

References

Aviv, H., & Leder, P. (1972) *Proc. Natl. Acad. Sci. U.S.A. 69*, 1408–1412.

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

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

Dover, G. (1987) *J. Mol. Evol. 26*, 47–58.

Etzerodt, M., Mikkelsen, T., Pedersen, F. S., Kjeldgaard, N. O., & Jorgensen, P. (1984) *Virology 134*, 196–207.

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., 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.

Gorski, K., Carneiro, M., & Schibler, U. (1986) *Cell 47*, 767–776.

Graves, B. J., Eisenman, R. N., & McKnight, S. L. (1985) *Mol. Cell. Biol. 8*, 1948–1958.

Guenet, J. L. (1986) *Curr. Top. Microbiol. Immunol. 127*, 109–113.

Guengerich, F. P., Dannan, G. A., Wright, S. T., Martin, M. V., & Kaminsky, L. S. (1982) *Biochemistry 21*, 6019–6030.

Herr, W. (1984) *J. Virol. 49*, 471–478.

Jansson, I., Mole, J., & Schenkman, J. B. (1985) *J. Biol. Chem. 260*, 7084–7093.

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

Matsunaga, T., Nagata, I., Holsztynska, E. J., Lapenson, D. P., Smith, A., Kato, R., Gelboin, H. V., Waxman, D. J., & Gonzalez, F. J. (1988) *J. Biol. Chem. 263*, 17995–18002.

Maxam, A. M., & Gilbert, W. (1980) *Methods Enzymol. 65*, 499–560.

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

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.

Nagata, K., Matsunaga, T., Gillette, J., Gelboin, H. V., & Gonzalez, F. J. (1987) *J. Biol. Chem. 262*, 2787–2793.

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.

Rogers, J. (1985) *Nature 317*, 765–766.

Ryan, D. E., Thomas, P. E., Reik, L. M., & Levin, W. (1982) *Xenobiotica, 12*, 727–744.

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

Sharp, P. A., Bork, A. J., & Berget, S. M. (1980) *Methods Enzymol. 65*, 750–768.

Soares, M. B., Schon, E., & Efstratiadis, A. (1985) *J. Mol. Evol. 22*, 117–133.

Swinney, D. W., Ryan, D. E., Thomas, P. E., & Levin, W. (1987) *Biochemistry 26*, 7073–7083.

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

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

Waxman, D. J., Ko, A., & Walsh, C. (1983) *J. Biol. Chem. 258*, 11937–11947.

Wood, A. W., Ryan, D. E., Thomas, P. E., & Levin, W. (1983) *J. Biol. Chem. 258*, 8839–8847.