Am. J. Hum. Genet. 45:889–904, 1989

The Human Debrisoquine 4-Hydroxylase (*CYP2D*) Locus: Sequence and Identification of the Polymorphic *CYP2D6* Gene, a Related Gene, and a Pseudogene

Shioko Kimura,* Morio Umeno,* Radek C. Skoda,† Urs A. Meyer,† and Frank J. Gonzalez*

*Laboratory of Molecular Carcinogenesis, National Cancer Institute, National Institutes of Health, Bethesda, MD; and †Department of Pharmacology, Biocenter, University of Basel, Basel

Summary

The debrisoquine-4-hydroxylase polymorphism is a genetic variation in oxidative drug metabolism characterized by two phenotypes, the extensive metabolizer (EM) and poor metabolizer (PM). Of the Caucasian populations of Europe and North America, 5%–10% are of the PM phenotype and are unable to metabolize debrisoquine and numerous other drugs. The defect is caused by several mutant alleles of the *CYP2D6* gene, two of which are detected in about 70% of PMs. We have constructed a genomic library from lymphocyte DNA of an EM positively identified by pedigree analysis to be homozygous for the normal *CYP2D6* allele. The normal *CYP2D6* gene was isolated; was completely sequenced, including 1,531 and 3,522 bp of 5′ and 3′ flanking DNA, respectively; and was found to contain nine exons within 4,378 bp. Two other genes, designated *CYP2D7* and *CYP2D8P*, were also cloned and sequenced. *CYP2D8P* contains several gene-disrupting insertions, deletions, and termination codons within its exons, indicating that this is a pseudogene. *CYP2D7*, which is just downstream of *CYP2D8P*, is apparently normal, except for the presence, in the first exon, of an insertion that disrupts the reading frame. A hypothesis is presented that the presence of a pseudogene within the *CYP2D* subfamily transfers detrimental mutations via gene conversions into the *CYP2D6* gene, thus accounting for the high frequency of mutations observed in the *CYP2D6* gene in humans.

Introduction

The cytochrome P450 gene superfamily consists of nine gene families in mammals (Nebert et al. 1989). Five of the families code for enzymes expressed in specialized steroidogenic tissues that catalyze steps in steroid biosynthesis. The remaining four families of P450 are expressed in liver and to some extent in extrahepatic tissues, and these enzymes catalyze the oxidation of fatty acids, steroids, and numerous chemicals, including drugs and carcinogens. The hepatic P450 system is responsible for the metabolism and elimination of most foreign chemicals that are ingested. Typically, the P450s will oxidize a hydrophobic chemical to introduce functional groups that can be substrates for various conjugating enzymes. These latter enzymes introduce chemical moieties that render the substrate more hydrophilic so that it can be excreted via urine or bile. Many P450s can also metabolize compounds of diverse structures, and this unique feature allows the organism to process and eliminate scores of chemicals. It is generally thought that many P450s evolve to metabolize plant toxins (Nebert and Gonzalez 1985; Nelson and Strobel 1987; Gonzalez 1988). Thus, P450s may have been required to detoxify poisonous chemicals in plants. As plants or habitats changed, a particular P450 may not have been required for survival and its presence was no longer selected for. This may be the basis for drug oxidation polymorphisms detected in rodents and man (Gonzalez 1988).

The debrisoquine 4-hydroxylase genetic deficiency

Received May 30, 1989; revision received August 1, 1989.

Address for correspondence and reprints: Frank J. Gonzalez, Ph.D., Building 37, Room 3E-24, Laboratory of Molecular Carcinogenesis, National Cancer Institute, National Institutes of Health, Bethesda, MD 20892.

is the most widely studied human drug oxidation defect. Debrisoquine, an adrenergic blocking agent previously used to control hypertension, was found to be inefficiently metabolized by a significant number of individuals. This lack of metabolism resulted in exaggerated response during clinical administration of the drug. Individuals who cannot metabolize debrisoquine are termed *poor metabolizers* (PMs), while those capable of metabolism are called *extensive metabolizers* (EMs). PMs and EMs can be identified by administering subclinical doses of debrisoquine or other drugs and monitoring accumulation of both the parent compound and the hydroxylated metabolite in the urine. By use of urine metabolite analysis, the deficiency in debrisoquine metabolism was found to be inherited as an autosomal recessive trait and to affect 5%–10% of the Caucasian population of Europe and North America (Idle and Smith 1979; Eichelbaum 1986). Numerous other drugs and chemicals have been shown to be subjected to this genetic defect.

P450 IID1 responsible for debrisoquine oxidation has been purified from rats (Larrey et al. 1984; Gonzalez et al. 1987) and man (Distlerath et al. 1985; Gut et al. 1986). The rat (Gonzalez et al. 1987) and human (Gonzalez et al. 1988*b*) IID1 cDNAs have also been cloned and sequenced. The human IID1 cDNA was used to determine that mutant IID1 (*CYP2D6*) genes account for the drug oxidation defect (Gonzalez et al. 1988*a*). Two mutant *CYP2D6* alleles were found that produce incorrectly spliced transcripts. RFLP haplotype analysis was used to identify two mutant alleles in lymphocyte DNA in individuals who cannot metabolize debrisoquine (Skoda et al. 1988). In about 70% of PM individuals at least one mutant allele can be detected by RFLP analysis.

In the present report we have cloned and sequenced the *CYP2D6* gene and two related genes—designated *CYP2D7* and *CYP2D8P*—at the *CYP2D* locus. The three *CYP2D* genes display 92%–97% nucleotide similarities with each other across their introns and exons. The *CYP2D8P* gene was found to be a pseudogene, while the *CYP2D7* gene was found to have in its first exon an insertion that disrupts the reading frame of the protein. The role of the *CYP2D7* and *CYP2D8P* genes in the generation of mutant *CYP2D6* alleles is discussed.

Material and Methods

Material

λEMBL3 DNA was obtained from Promega Biotec. λDASH and Gigapak Gold™ were purchased from Stratagene. Sequenase was from United States Biochemicals and α^{35}S-dATP (500 Ci/mmol) was purchased from New England Nuclear. *Sma*I-digested and phosphatase-treated M13 mp10 was purchased from Amersham.

Isolation and Sequencing of the CYP2D Genes

Two human gene libraries were constructed. One library was constructed from a liver with extensive in vitro metabolism of debrisoquine. DNA from this liver was homozygous for the polymorphic *Xba*I 29-kbp fragment (*Xba*I pattern A of Skoda et al. [1988]), as is the case with most EMs. The DNA was partially digested with *Mbo*I, size-fractionated on an NaCl gradient (5%–25%), and ligated into λEMBL3 arms according to a method described by Kaiser and Murray (1985). The DNA was packaged in vitro using Gigapak Gold™, and packaged phage were plated using *Escherichia coli* strain K802. A second library was constructed using λDASH and DNA from lymphocytes of an EM of debrisoquine who was phenotyped in vivo. This individual was determined to carry two normal *CYP2D* alleles by the segregation of *Bam*HI RFLPs in his family (fig. 1). Complete *Hin*dIII digestion of lymphocyte DNA was carried out, and fragments greater than 15 kbp were isolated and ligated into λDASH that had been digested with *Xba*I and *Hin*dIII. The DNA was processed as described above. Phage were screened at a density of 30,000 pfu/150-mm dish by plaque hybridization using nick-translated IID1 cDNA probe. The specific clones were plaque purified, restriction maps of their insert DNAs were obtained, and the individual *Bam*HI fragments from the inserts were subcloned into pUC9. Insert DNA was isolated by gel purification, self-ligated, and shotgun cloned into M13 mp10 according to a method described elsewhere (Deininger 1983). DNA was sequenced using the dideoxy chain-termination method (Sanger et al. 1977), except that Sequenase was substituted for DNA polymerase Klenow fragment. Sequence data were analyzed using the Beckman Microgenie® program.

Other Procedures

The transcription start site of *CYP2D* was determined by primer extension analysis and S1 nuclease protection assays. In brief, poly(A) RNA was isolated from a liver containing high levels of *CYP2D6* mRNA (Gonzalez et al. 1988*a*). A primer for primer extension (20-mer, +17 to +36; fig. 4) and a probe for S1 mapping (60-mer, −24 to +36, fig. 4) were synthesized using an Applied Biosystems 380B DNA synthesizer. The oli-

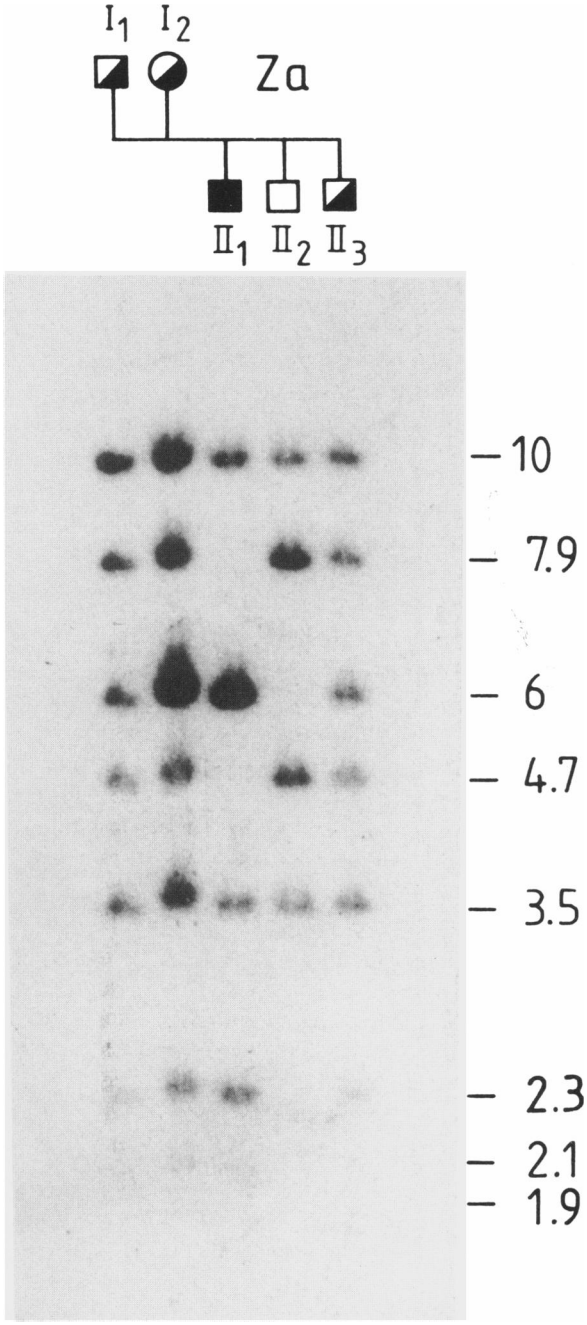


Figure 1 Identification of a homozygous EM by analysis of the segregation of *Bam*HI RFLPs. Each individual within the Za pedigree is placed above the corresponding lane. □ = EM phenotype; ◪ or ◔ = EM phenotype obligate heterozygote by pedigree analysis; ■ = PM phenotype. The 6-kbp fragment together with the 2.3- and the faint 2.1- and 1.9-kbp fragments represent the mutant allele in *CYP2D6* in this family because II_1 is homozygous for these fragments. These fragments behave allelic to the 7.9- and 4.7-kbp fragments which behave as markers for the normal allele. Individual II_2 is homozygous for the normal 7.9/4.7-kbp allele.

gonucleotides were purified by agarose gel electrophoresis, 5′-end labeled with T4 polynucleotide kinase and [γ^{32}P]-ATP, and hybridized to 10 μg poly(A) RNA. The primer was extended using AMV reverse transcriptase. S1 nuclease digestion was performed after the probe was allowed to anneal to Poly(A) RNA. Yeast tRNA was used as control. The reaction products were electrophoresed on an 8% polyacrylamide–50% urea gel. An M13 sequencing ladder was electrophoresed concurrently to determine the size of the extended products. Southern blotting was performed according to a method described by McBride et al. (1986) by using BioTrace RP® nylon membrane (Gelman Sciences, Inc.) and the human IID1 cDNA as a probe (Gonzalez et al. 1988*b*).

Results and Discussion

Isolation of the CYP2D Genes

In the first attempt to isolate the *CYP2D6* gene, a library was constructed in λEMBL3 by using DNA isolated from a human liver. This liver had bufuralol 1′-hydroxylase activity representative of an extensive metabolizer and was homozygous for the *Xba*I 29-kbp fragment (*Xba*I pattern A of Skoda et al. [1988]). Two overlapping phage clones (λ2D-A and λ2D-B) were isolated as shown in figure 2 and were completely sequenced. These contained two tandemly arranged genes that each had nine exons. After careful comparison of their exonic sequences with the *CYP2D6* cDNA (Gonzalez et al. 1988*b*), they were found to be distinct from the *CYP2D6* gene and hence were designated *CYP2D7* and *CYP2D8P*. The exonic sequence of *CYP2D7* and *CYP2D8P* bore 97% and 92% similarities to the *CYP2D6* cDNA, respectively. The restriction maps of these phage DNAs were constructed and compared with the Southern blotting data of Skoda et al. (1988), and unique fragments were identified that were not represented by our restriction map of the cloned *CYP2D7* and *CYP2D8P* genes in figure 2; for example, only two of the three fragments found with *Eco*RI can be explained by the single *Eco*RI site located between *CYP2D8P* and *CYP2D7*. On the basis of these data, a third gene was suspected to be present in EM individuals, and a second library was prepared from lymphocyte DNA taken from an EM who was positively identified by pedigree analysis to carry two normal *CYP2D6* alleles (fig. 1). This was important because heterozygotes cannot be determined by the urine analysis phenotyping procedure and because heterozygotes

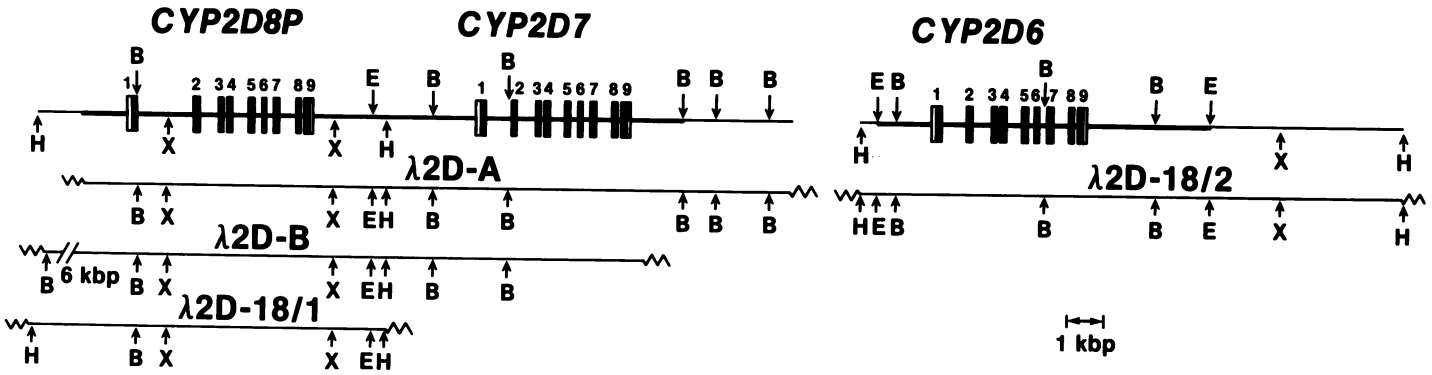


Figure 2 Restriction map of the *CYP2D* locus. A partial restriction map and the position of the exons in the *CYP2D6, CYP2D7,* and *CYP2D8P* genes derived from the sequence data in figs. 4 and 5 are displayed. The individual λ clones used to determine the sequence of the genes are shown below the consensus map. A thick line indicates regions of the clones that were completely sequenced. The exons and the *Eco*RI, *Bam*HI, *Hin*dIII, and *Xba*I restriction-enzyme sites are denoted by black rectangles E, B, H, and X, respectively.

are expected to account for 35%–43% of the Caucasian population. With this library the *CYP2D6* gene was isolated as a 15-kbp *Hin*dIII fragment in λDASH (λ2D-18/2; fig. 2). A second clone (λ2D-18/1) was also isolated that contained the *CYP2D8P* gene. This clone was restriction mapped and not further characterized.

Determination of the Transcription Start Site of CYP2D6

The transcription start site of the *CYP2D6* gene was determined by primer extension and S1 mapping. It must be noted that the three *CYP2D* genes contain virtually identical sequence in their first exons. However, since the *CYP2D8P* gene is a pseudogene (see below) and since we have not detected significant levels of *CYP2D7* in several human liver RNAs, we believe that we are only detecting transcripts derived from the *CYP2D6* gene. One distinct extended fragment was observed after reverse transcriptase treatment of the *CYP2D6* primer that had been annealed with liver RNAs from two individuals, and the position of this extended fragment matched that of the major S1 protected fragment (fig. 3, fragment of 36 bp). The size of the fragments corresponds to a *CYP2D6* gene G residue which was assigned position +1 in figure 4.

Sequence of the CYP2D6 Gene

The complete sequence of the *CYP2D6* gene is presented in figure 4, including 1,531 and 3,522 bp of 5′ and 3′ flanking DNA. The putative transcription start site, determined in figure 3, is designated +1 and is preceded by a TATA box at −24 to −28. No CCAAT box was noted within 200 bp upstream of the start site. The 5′ untranslated portion of the mRNA is 88 bases. The *CYP2D6* gene contains nine exons and spans 4,378 bp from the polymerase start site to the polyadenylation site. Nine exons, including coding first and last exons are typical of other *CYP2* family genes (Gonzalez 1988).

Sequences of the CYP2D7 and CYP2D8P Genes

The sequences of the *CYP2D7* and *CYP2D8P* genes are displayed in figure 5. The two genes are aligned tandemly head to tail and are separated by 4,665 bp of intragenic DNA. The *CYP2D7* gene is located downstream of the *CYP2D8P* gene. *CYP2D7* is apparently a normal gene, except for the presence of a single T insertion at position +226 in the first exon. This insertion disrupts the protein reading frame, indicating that the *CYP2D7* gene would produce an mRNA incapable of translating a functional P450. Screening of eight human liver RNAs with an oligonucleotide specific to the

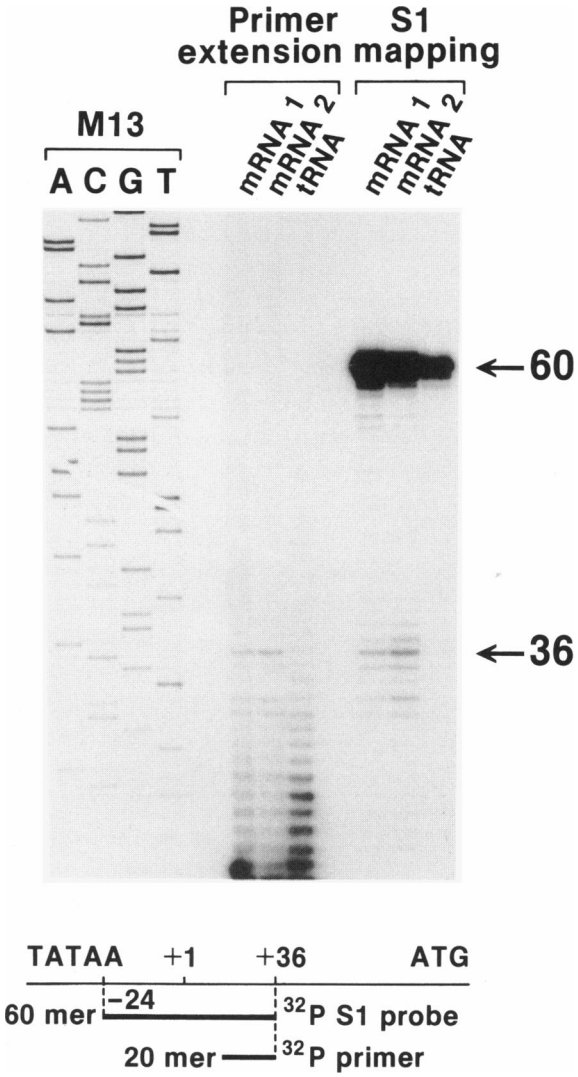
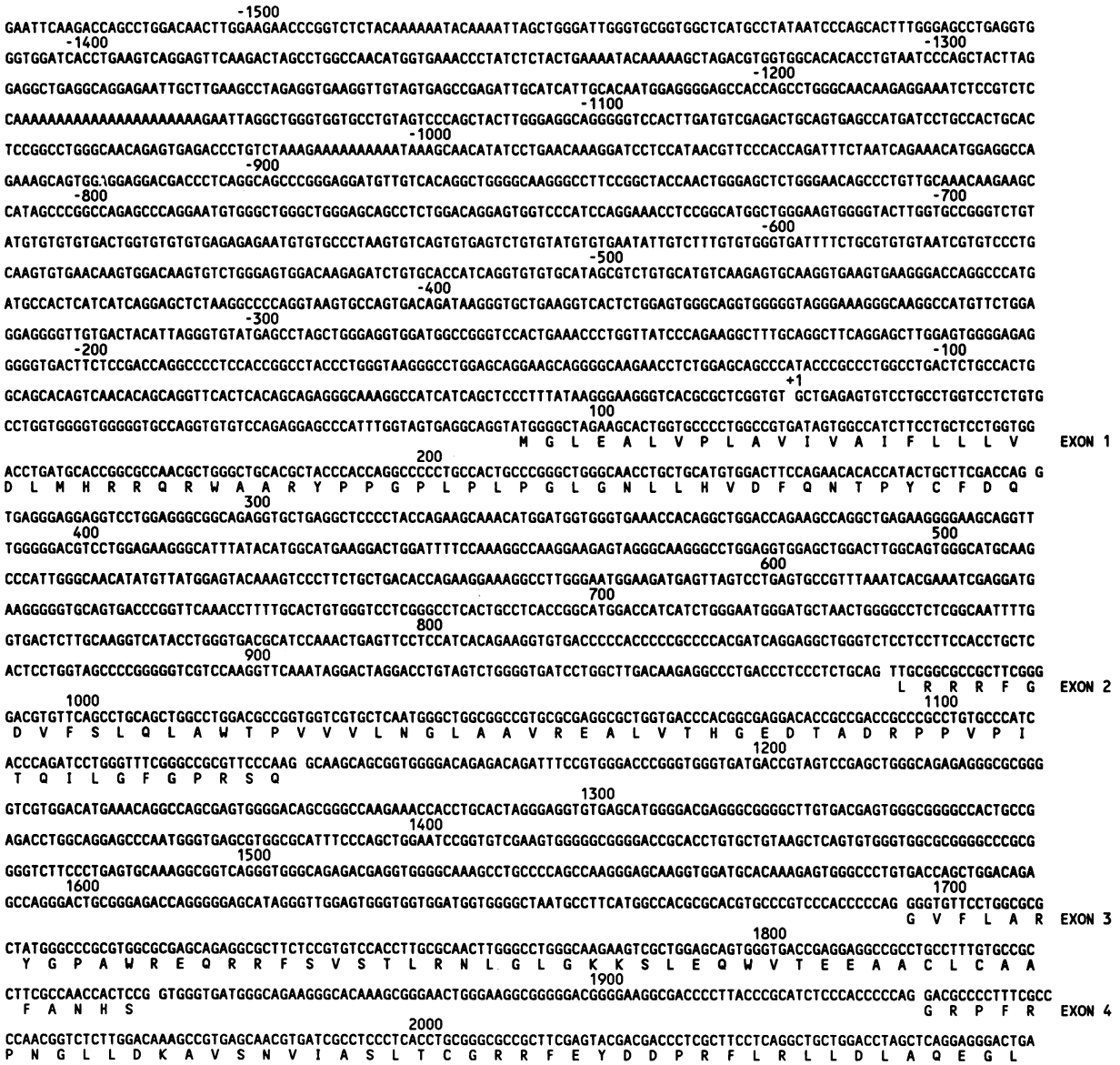


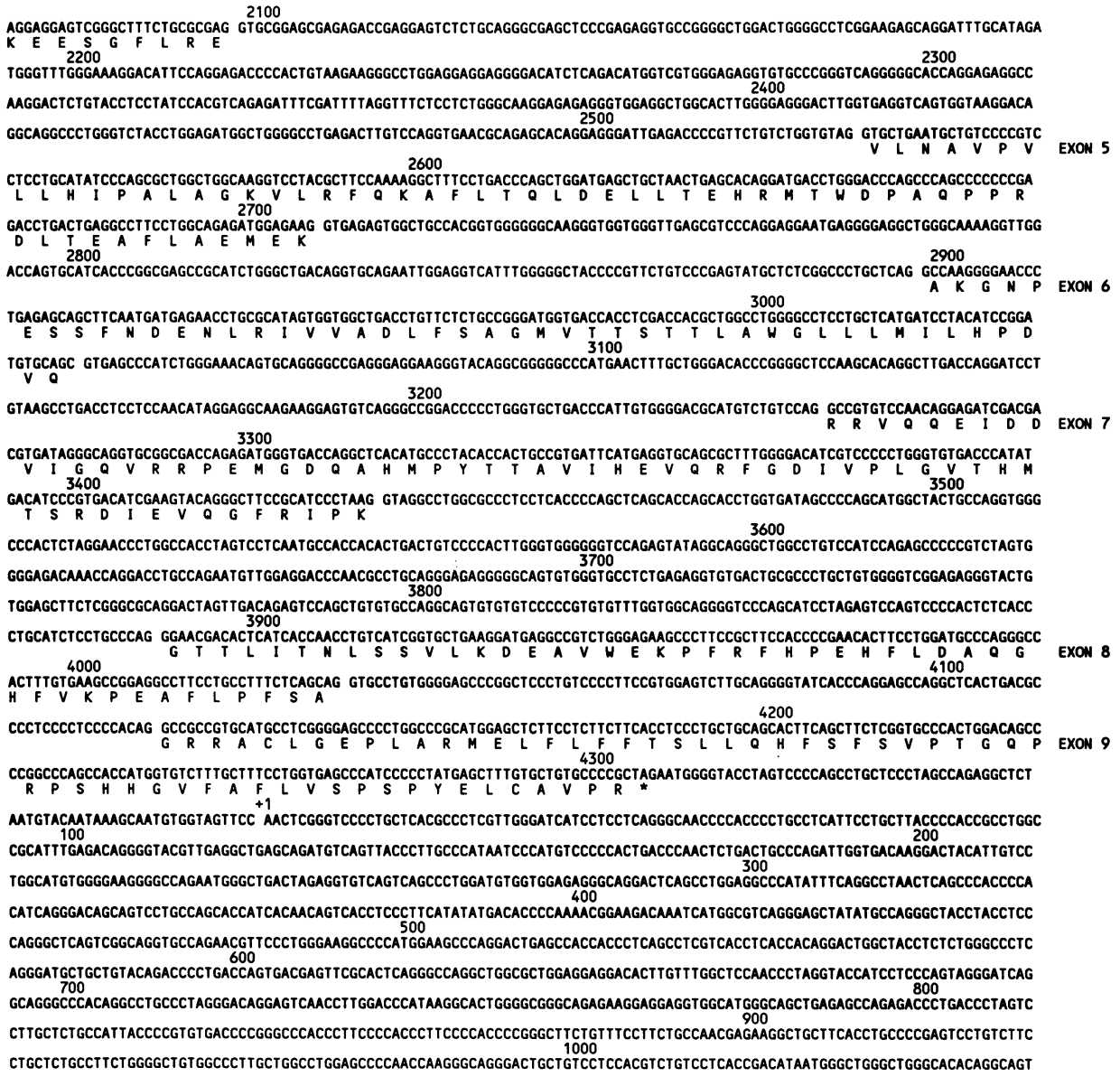
Figure 3 Primer extension and S1 mapping analyses of the *CYP2D6* gene. The primer (20-mer, +17 to +36; fig. 3) for primer extension and S1 probe (60-mer, −24 to +36; fig. 3) for S1 mapping were labeled with [^{32}P], annealed with 10 μg human liver poly(A) RNAs from two individuals, and either extended with reverse transcriptase or treated with S1 nuclease, respectively. Yeast tRNA was used as control. The reaction products were electrophoresed concurrently with a ladder of DNA sequence derived from M13 mp18. mRNA1 and mRNA2 were isolated from different human liver samples. The numbers at the right represent the fragment sizes in base pairs.

putative *CYP2D7* mRNA failed to detect an RNA on a Northern blot (authors' unpublished data). If it is assumed that the *CYP2D7* gene could have the same start site as *CYP2D6* (as shown by (+1) in fig. 5), a TATA box is found at −24 to −28. Among the first 774 bp immediately upstream from the putative cap site, nucleo-



tide sequences of *CYP2D6* and *CYP2D7* genes are 97% similar. Therefore, in spite of the finding of no RNA in a limited survey, it is entirely possible that *CYP2D7* may be expressed in some livers. The presence of a premature termination codon at bases 2587 to 2589 in exon 5 of *CYP2D7* that was introduced by the T insertion in exon 1 would probably render the mRNA unstable (Daar and Maquat 1988). We do not know the frequency of this mutant *CYP2D7* allele in the human population at this time.

The *CYP2D8P* gene resides 4,666 bp upstream of the *CYP2D7* gene, spans 5,267 bp of DNA, and contains nine exons (fig. 5). Upstream DNA of 1,303 bp was also sequenced. When compared with the *CYP2D6* gene, this gene contains multiple deletions and insertions in its exonic sequence, resulting in a fully disrupted



protein open reading frame. The mutations in *CYP2D8P* are shown in table 1. Because of the large number of insertions and deletions, these data would suggest that the *CYP2D8P* gene has been inactive for a longer period of time, suggesting that it is a bona fide pseudogene.

Comparison of the CYP2D Genes

To compare the three *CYP2D* genes, dot-matrix analysis (fig. 6) and direct nucleotide alignment comparisons (table 2) were carried out. The *CYP2D6* gene shares high nucleotide similarity with both *CYP2D7* and *CYP2D8P*. The nucleotide similarities between *CYP2D6* and *CYP2D7* are, in general, greater across both introns and exons than are those between *CYP2D6* and *CYP2D8P* (table 2). Further, the nucleotide similarities between *CYP2D6* and *CYP2D7* extend several

Figure 4 Sequence of the *CYP2D6* gene. The nucleotide sequence of an *Eco*RI fragment encompassing the *CYP2D6* gene was determined. The transcription start site was designated +1 on the basis of the data in fig. 3. The position where the *CYP2D6* gene is polyadenylated was determined from the cDNA clone. Upstream and downstream DNA are numbered beginning with −1 and +1 before the transcription start site and after the polyadenylation site, respectively. The amino acids are designated by the single-letter code.

hundred base pairs both upstream and downstream, whereas the similarities of nucleotide sequence between *CYP2D6* and *CYP2D8P* extended just upstream of the first exon and almost precisely at the end of the termination codon in their ninth exons. These data suggest that the gene duplication events that gave rise to *CYP2D6* and *CYP2D7* involved upstream and downstream intragenic DNA, whereas that giving rise to *CYP2D8P* involved only the structural gene and a small portion of upstream DNA, including the TATA box. Since the transcription regulatory elements in the *CYP2D6* gene have not been defined, it is unknown whether *CYP2D8P* would have the appropriate upstream DNA to be accurately transcribed and regulated.

Another interesting finding is the presence of a large insertion of DNA in the first intron of *CYP2D8P* (figs. 5 and 6). This insertion represents three tandem R.dre.1 (*Alu*) sequences and is flanked by the direct repeats 5′-GAAATCA-3′. These R.dre.1 repeats of about 300 bp display 68%–75% nucleotide similarities with each other. The 900-bp insertion must have occurred subsequent to the formation of the three *CYP2D* genes.

A comparison of the amino acid sequences of the three *CYP2D* genes is shown in figure 7. The alignments were made on the basis of the deduced amino acid sequence of *CYP2D6*, and the percent similarities

Table 1

Mutations in the *CYP2D8P* Gene

Exon 1	3-base deletion	Between 250 and 251
Exon 4	C→T} termination	3000
	C→A}	3002
	2-base insertion	3006–3007
Exon 5	C→T termination	3584
Exon 7	3-base deletion	Between 4177 and 4178
	C→T termination	4205
Exon 9	1-base insertion	5056

NOTE.—The sequence of the *CYP2D8P* gene was compared with that of the *CYP2D6* gene, and detrimental mutations were identified.

Table 2

Comparisons of the Exon and Intron Lengths and Percent Nucleotide Similarities between *CYP2D6*, *CYP2D7*, and *CYP2D8P* Genes

	Length[a] (bp)			% Similarity		
	CYP2D6	*CYP2D7*	*CYP2D8P*	*CYP 6/7*	*CYP 7/8P*	*CYP 6/8P*
Upstream	774	777		97		
		186	183		92	
	189		186			89
Exon 1	268	269	265	97	94	93
Intron 1	703	701	1,620[b]	98	90	89
Exon 2	172	172	172	95	94	91
Intron 2	550	528	546	74	78	77
Exon 3	153	553	153	98	93	92
Intron 3	88	88	88	98	91	93
Exon 4	161	161	161	98	89	91
Intron 4	433	425	449	94	85	86
Exon 5	177	177	177	99	93	92
Intron 5	190	192	186	97	84	83
Exon 6	142	142	142	94	92	96
Intron 6	207	194	204	82	87	90
Exon 7	188	188	185	98	94	95
Intron 7	454	454	449	98	91	91
Exon 8	142	142	142	99	96	96
Intron 8	98	98	96	100	97	97
Exon 9[c]	252	252		94		
		180	181		95	
	180		181			92
3′ Flanking	538	528		97		

Note.—Alignments were performed on the Beckman Microgenie® program. Exons or introns that are the same size among genes are underlined.

[a] Lengths of upstream and downstream DNAs compared were those of maximal nucleotide similarities. DNA further upstream and downstream did not display significant similarities.

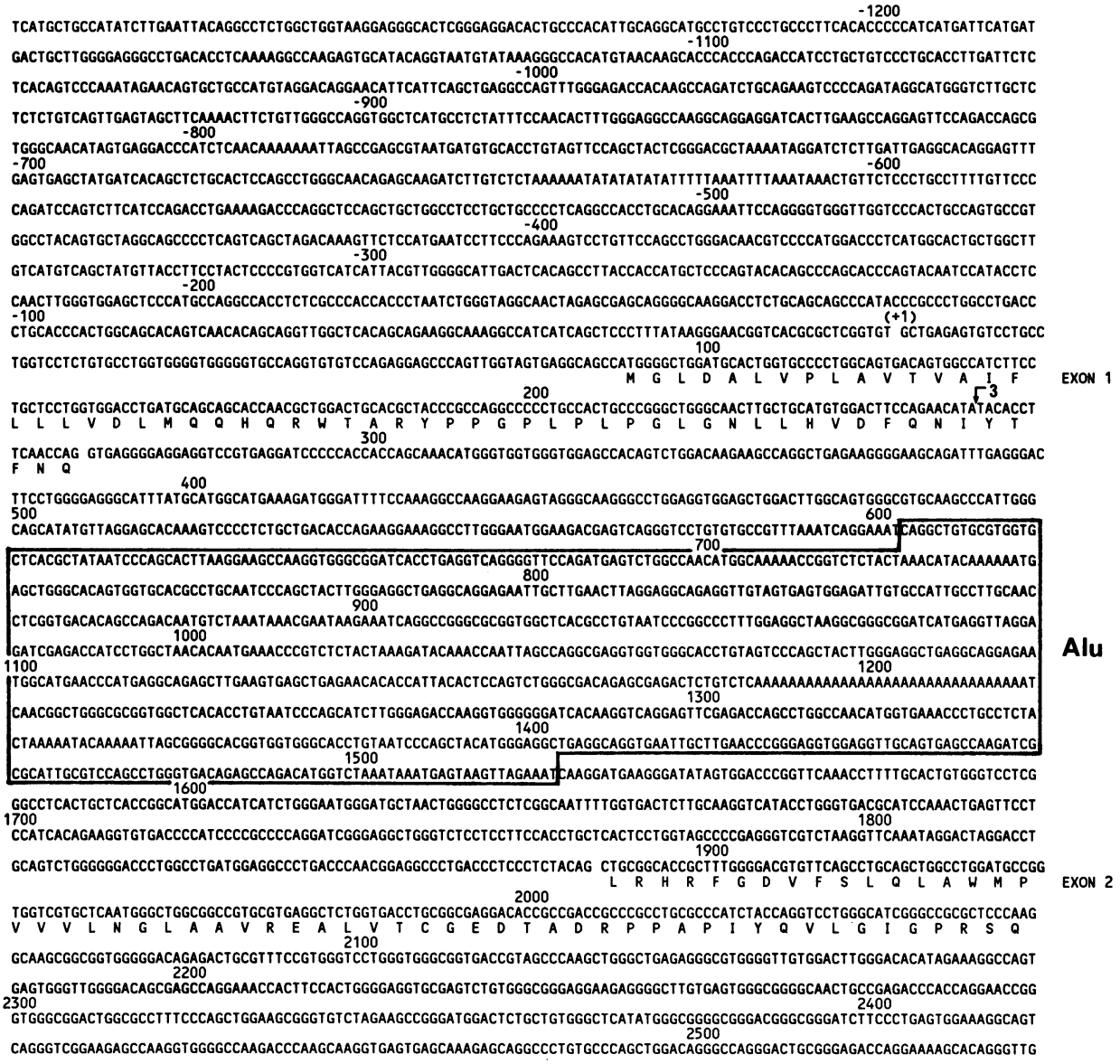
[b] Contains a 920-bp insertion.

[c] When *CYP2D8P* exon 9 sequence was compared with *CYP2D6* and *CYP2D7*, alignment finished 1 base after the termination codon.

reflect that of the *CYP2D* genes' exonic regions. *CYP2D6* displayed 97% and 92% gene-deduced amino acid similarities with *CYP2D7* and *CYP2D8P*, respectively. It is interesting that certain segments of sequence between *CYP2D7* and *CYP2D8P* are more similar to each other than to *CYP2D6;* for example, residues 166–168 are DQA in both *CYP2D7* and *CYP2D8P* and are NHS in *CYP2D6* (fig. 7). Residues 478–481 are SRVV in the former two genes but are HGVF in the latter. These data suggest that gene conversions have occurred between *CYP2D7* and *CYP2D8P* subsequent to the recent formation of *CYP2D6* and *CYP2D7* by gene duplication event. Gene conversions have been seen in several other P450 subfamilies (Gonzalez 1988).

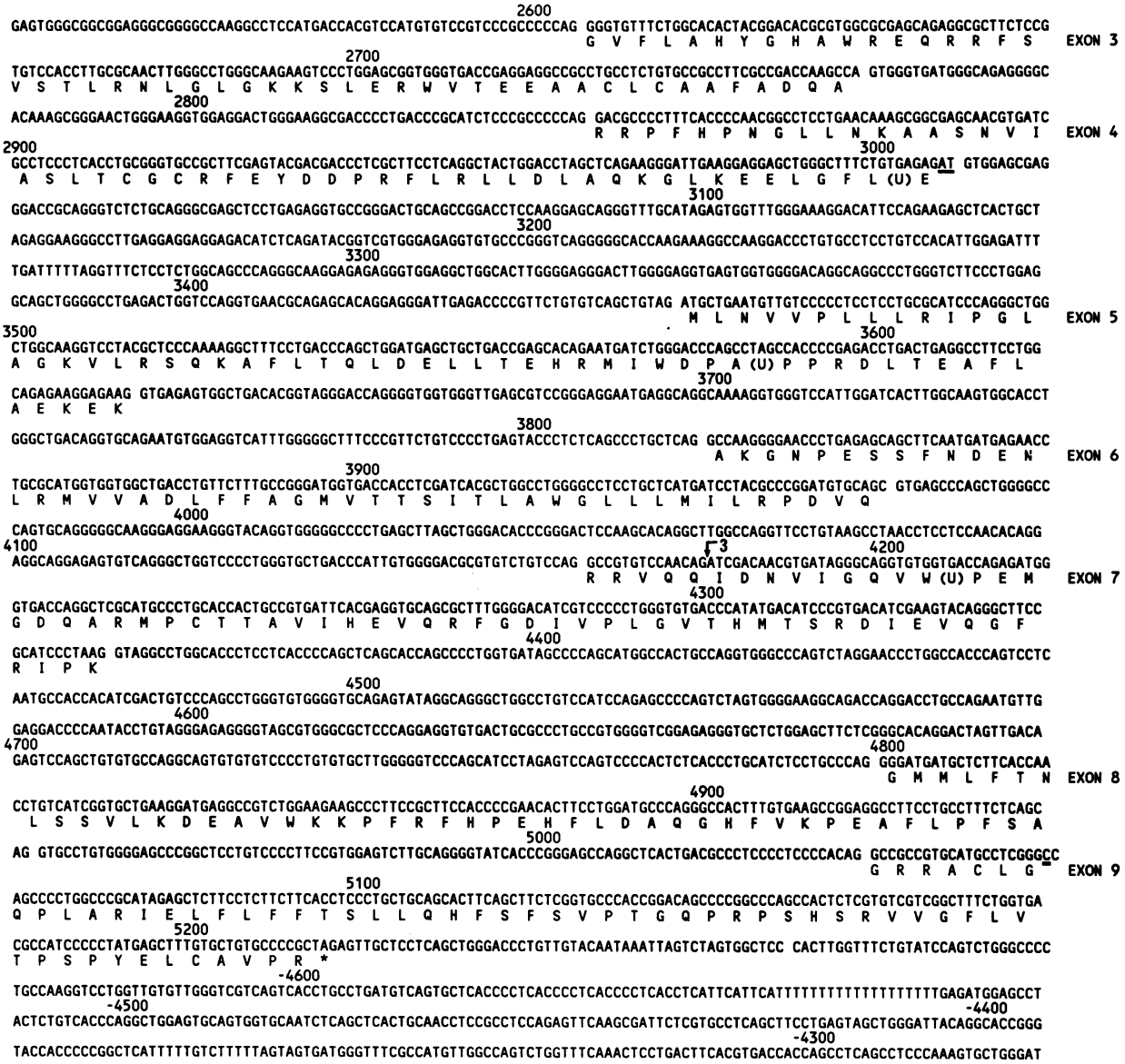
Conclusions

In the present report we have identified three genes that compose the *CYP2D* gene cluster in man. These genes are located distal to IGL on the long arm of chromosome 22 (q11.2-qter) (Gonzalez et al. 1988*b*). The



CYP2D8P gene was found to be a pseudogene, while the *CYP2D7* gene contained a single reading frame–disrupting insertion in its first exon. The *CYP2D6* gene is that associated with polymorphic debrisoquine metabolism. We do not know whether the *CYP2D7* gene product is capable of metabolizing debrisoquine.

The Southern blot data presented elsewhere (Skoda et al. 1988), together with the sequence data presented here, are consistent with *CYP2D6* being located 3′ downstream of *CYP2D8P* and *CYP2D7*; for example, the *BclI* site found −693 bp upstream of the *CYP2D8P* gene (fig. 4) can account for the single 40-kbp frag-

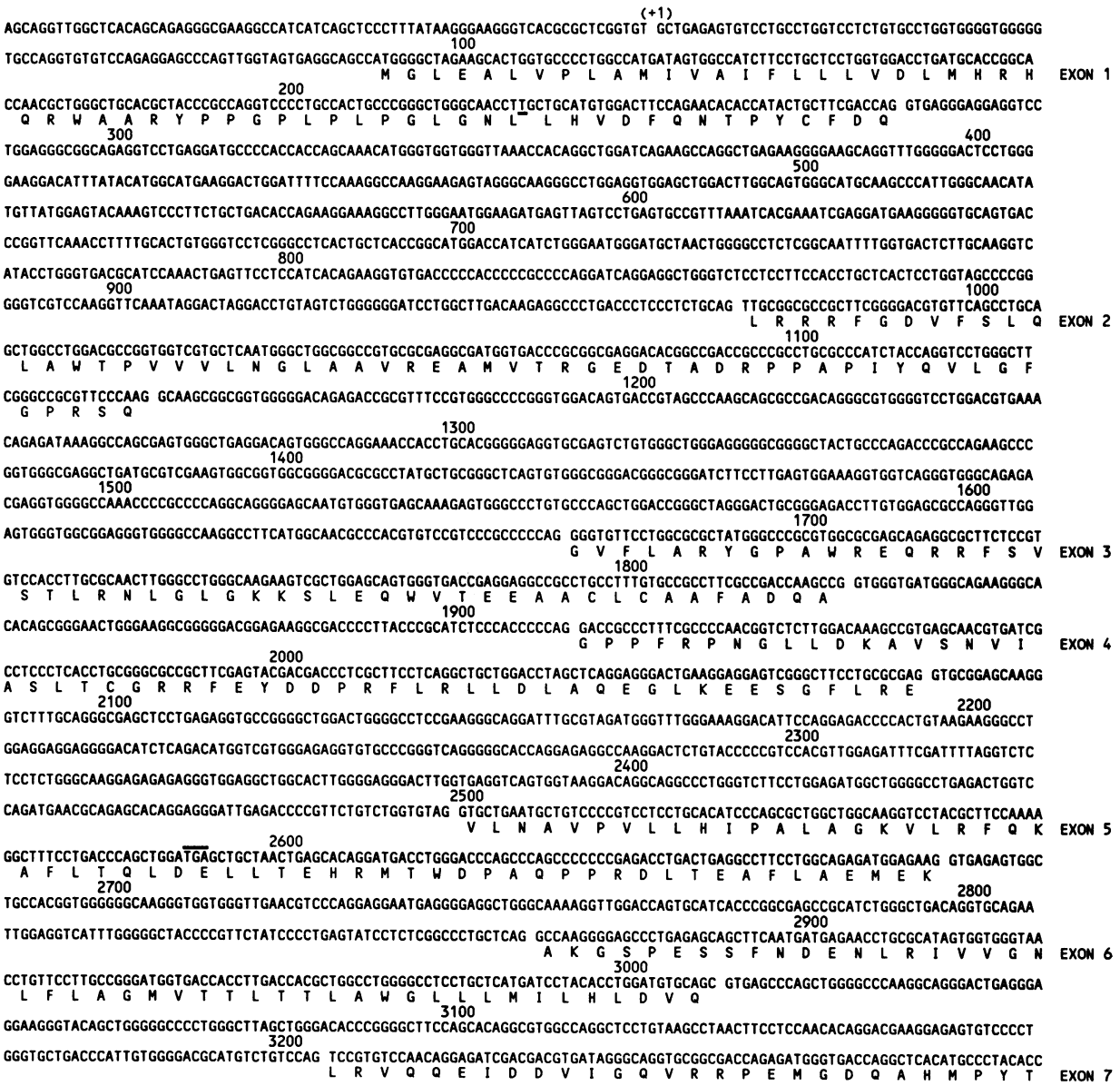


ment detected in Southern blots if *CYP2D6* is downstream of *CYP2D7*. On the other hand, if *CYP2D6* had been located 5′ upstream of *CYP2D8P*, two *BclI* fragments would have been expected. By similar reasoning with data from other restriction enzymes, we believe the placement of the three *CYP2D* genes is as shown in figure 2.

On the basis of both the high frequency of mutant alleles in the population (35%–43%; Gonzalez et al. 1988*a*) and the apparent lack of selection for the presence of the *CYP2D6* gene product (see below), it appears that the *CYP2D6* gene is beginning to vanish. The occurrences of gene conversions between the pseudogene and the *CYP2D6* gene might hasten the in-

troduction of mutations into *CYP2D6*. Indeed, gene conversions between the *CYP21A2* gene and the neighboring *CYP21A1* pseudogene are known to contribute to the mutations at the steroid 21-hydroxylase locus (Miller 1988).

The unexpected finding of two additional genes other than *CYP2D6* allowed us to reexamine some of the variant cDNA transcripts we previously cloned from livers of PM individuals (Gonzalez et al. 1988*a*). Indeed, it appears that our previously described "variant b" could



have been transcribed from the *CYP2D7* gene instead of from *CYP2D6*, since it displayed more nucleotide similarity with the former. Alternatively, gene conversion could have occurred between these two genes, giving rise to a *CYP2D6* allele that is more similar to *CYP2D7*. Either of these possibilities is likely; however, our previously described "variant a" transcript clearly corresponds to the *CYP2D6* gene described herein.

Finally, the question arises as to what is the driving

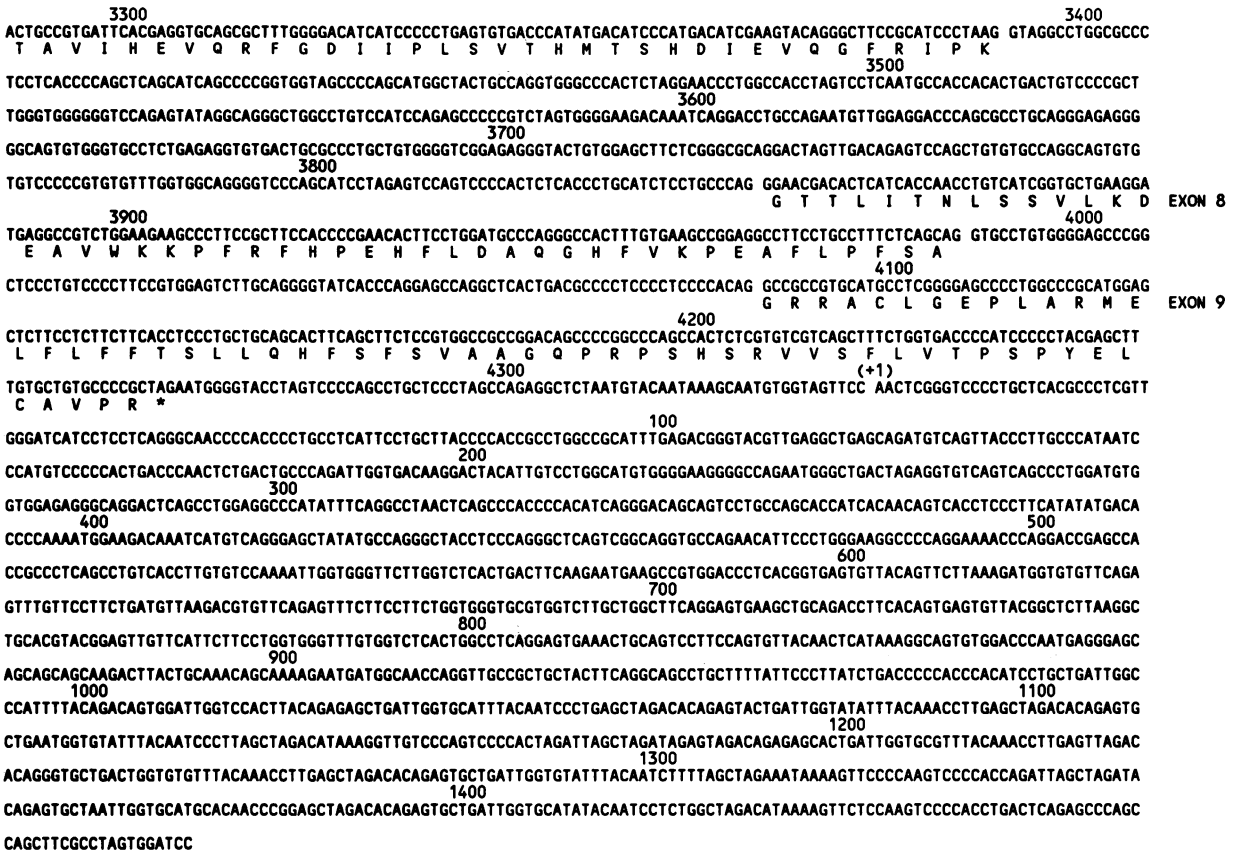


Figure 5 Sequences of the *CYP2D7* and *CYP2D8P* genes. The sequences of several *Bam*HI fragments derived from the λ clones shown in fig. 1 were determined. The putative transcription start sites designated by (+1) and the putative polyadenylation sites were estimated by comparing the sequence similarities between these genes and *CYP2D6*. Although we have not detected mRNA corresponding to these genes, we have made these designations for purpose of comparisons with the *CYP2D6* sequence. The amino acids were displayed on the basis of the *CYP2D6* reading frame in fig. 4, even though these genes have mutations that destroy the normal protein reading frames. The intragenic DNA between *CYP2D8P* and *CYP2D7* genes was numbered beginning at −1 upstream of *CYP2D7*. Base deletion and insertion are shown by an arrow following the number of bases deleted and by an underline, respectively. (U) represents a stop codon due to base changes. The first termination codon introduced in exon 5 of *CYP2D7* because of a single base insertion in exon 1 is overlined. Insertion of Alu repeats seen in first introns of *CYP2D8P* gene is boxed.

force behind the evolution of these drug-metabolizing enzymes. The suggestions that these enzymes evolved to metabolize plant toxins seem quite plausible (Nebert and Gonzalez 1985; Nelson and Strobel 1987; Gonzalez 1988). The current human diets rely almost totally on cultivation, while early man was a hunter-gatherer. Since man no longer relies on the evolution of wild plant fauna and can selectively avoid toxic plants, the detoxifying enzymes may no longer be evolving, and without this selective pressure of additional dietary toxins, many of the P450 genes, previously required for survival, might be lost, as evidenced by the occurrence of the debrisoquine polymorphism and other drug oxidation defects in man (Gonzalez 1988).

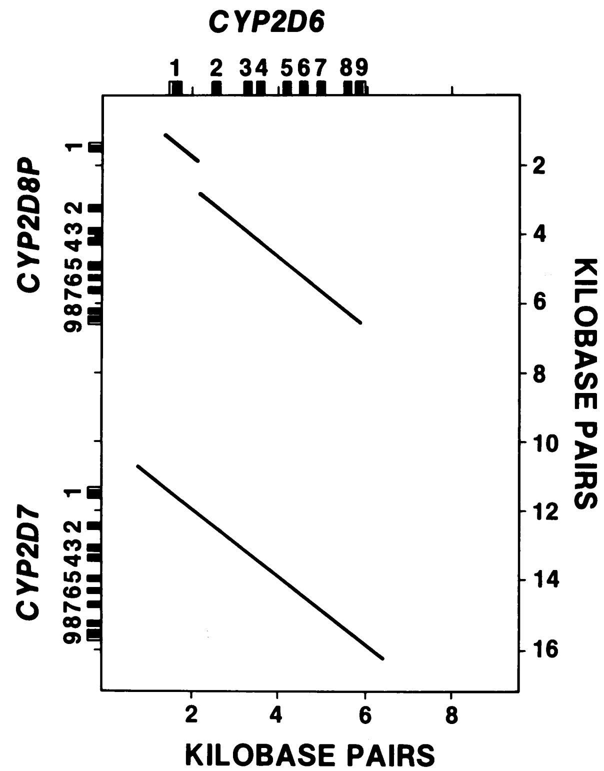


Figure 6 Dot-matrix comparison of the *CYP2D6* gene sequence with the sequence of *CYP2D7* and *CYP2D8P*. Comparisons were made using the Beckman Microgenie program. A dot is generated whenever a stretch of 15 nucleotides displays greater than 85% similarity. The exons are denoted by rectangles. The open rectangles represent the untranslated regions of the mRNAs.

```
CYP2D7            M              H                    ↓1
CYP2D6  MGLEALVPLAVIVAIFLLLVDLMHRRQRWAARYPPGPLPLPGLGNLLHVDFQNTPYCFDQLRRRFGDVFSLQLAWTPVVVLNGLAAVREA   90
CYP2D8     D       T           QQH   T                       I- T N   H            M
                                                              ^
                                                              3

        M  R         A  Y V                                                        DQA P
        LVTHGEDTADRPPVPITQILGFGPRSQGVFLARYGPAWREQRRFSVSTLRNLGLGKKSLEQWVTEEAACLCAAFANHSGRPFRPNGLLDK  180
           C         A  Y V  I          H  H                        R              DQAR   H     N


        AVSNVIASLTCGRRFEYDDPRFLRLLDLAQEGLKEESGFLREVLNAVPVLLHIPALAGKVLRFQKAFLTQLDELLTEHRMTWDPAQPPRD  270
         A          C                 K     L   U M  V  L  R  G       S                 I    U
                                                 ↑
                                                 └2

                      S              GN  L      L             L   L
        LTEAFLAEMEKAKGNPESSFNDENLRIVVADLFSAGMVTTSTTLAWGLLLMILHPDVQRRVQQEIDDVIGQVRRPEMGDQAHMPYTTAVI  360
                K                 M      F       I           R         -  N     WU       R  C
                                                                       ^
                                                                       3

                 I   SV    H                             K
        HEVQRFGDIVPLGMTHMTSRDIEVQGFRIPKGTTLITNLSSVLKDEAVWEKPFRFHPEHFLDAQGHFVKPEAFLPFSAGRRACLGEPLAR  450
                     V                  MM F             K                                   Q
                                                                                             ↑
                                                                                             └1

                          AA        SR VS  T
        MELFLFFTSLLQHFSFSVPTGQPRPSHHGVFAFLVSPSPYELCAVPR                                              497
        I                           SR VG  T
```

Figure 7 Comparison of the deduced amino acid sequences of the human *CYP2D* genes. The amino acid sequences are taken from figs. 4 and 5. Base insertion is shown by an arrow with the number of bases inserted. Base deletion is shown by Λ with the number of bases deleted. A stop codon (U) created by base changes is underlined. Note that the *CYP2D7* and *CYP2D8P* gene reading frames are taken from that of *CYP2D6* for comparison purposes. These genes contain premature termination codons.

References

Daar IO, Maquat LE (1988) Premature translation termination mediates triosphosphate isomerase mRNA degradation. Mol Cell Biol 8:802–813

Deininger PL (1983) Random subcloning of sonicated DNA: application to shotgun DNA sequence analysis. Anal Biochem 129:216–223

Distlerath LM, Reilly PEB, Martin MV, Davis GG, Wilkinson GR, Guengerich FP (1985) Purification and characterization of the human liver cytochromes P-450 involved in debrisoquine hydroxylation and phenactin O-deethylation, two prototypes for genetic polymorphisms in oxidative drug metabolism. J Biol Chem 260:9057–9067

Eichelbaum M (1986) Polymorphic oxidation of debrisoquine and sparteine. Prog Clin Biol Res 214:157–167

Gonzalez FJ (1988) The molecular biology of cytochrome P450s. Pharmacol Rev 40:243–288

Gonzalez FJ, Matsunaga T, Nagata K, Meyer UA, Nebert DW, Pastewka J, Kozak CA, et al (1987) Debrisoquine 4-hydroxylase: characterization of a new P450 gene subfamily, regulation, chromosomal mapping, and molecular analysis of the DA rat polymorphism. DNA 6:149–161

Gonzalez FJ, Skoda RC, Kimura S, Umeno M, Zanger UM, Nebert DW, Gelboin HV, et al (1988*a*) Characterization of the common genetic defect in humans deficient in debrisoquine metabolisms. Nature 331:442–446

Gonzalez FJ, Vilbois F, Hardwick JP, McBride OW, Nebert DW, Gelboin HV, Meyer UA (1988*b*) Human debrisoquine 4-hydroxylase (P450 IID1): cDNA and deduced amino acid sequence and assignment of the *CYP2D* locus to chromosome 22. Genomics 2:174–179

Gut J, Catin T, Dayer P, Kronbach T, Zanger U, Meyer UA (1986) Debrisoquine/sparteine-type polymorphism of drug oxidation: purification and characterization of two functionally different human liver cytochrome P-450 isozymes involved in impaired hydroxylation of the prototype substrate bufuralol. J. Biol. Chem. 261:11734–11743

Idle JR, Smith RL (1979) Polymorphisms of oxidation at carbon centers of drugs and their clinical significance. Drug Metab Rev 9:301–317

Kaiser K, Murray NE (1985) The use of phage lambda replacement vectors in the construction of representative genomic libraries. Glover DM (ed) DNA cloning. Vol 1. IRL, Washington, DC, pp 1–47

Larrey D, Distlerath LM, Dannan GA, Wilkinson GR, Guengerich FP (1984) Purification and characterization of the rat liver microsomal cytochrome P-450 involved in the 4-hydroxylation of debrisoquine, a prototype for genetic variation in oxidative drug metabolism. Biochemistry 23: 2787–2795

McBride OW, Merry D, Givol D (1986) The gene for human p53 cellular tumor antigen is located on chromosome 17 short arm (17p13). Proc Natl Acad Sci USA 83:130–134

Miller WL (1988) Gene conversions, deletions and polymorphisms in congenital adrenal hyperplasia. Am J Hum Genet 42:4–7

Nebert DW, Gonzalez FJ (1985) Cytochrome P450 gene expression and regulation. Trends Pharmacol Sci 6:160–164

Nebert DW, Nelson DR, Adesnik M, Coon MJ, Estabrook RW, Gonzalez FJ, Guengerich FP, et al (1989) The P450 superfamily: update of listing of all genes and recommended nomenclature for the chromosomal loci. DNA 8:1–13

Nelson DR, Strobel HW (1987) Evolution of cytochrome P-450 proteins. Mol Biol Evol 4:572–593

Sanger F, Nicklen S, Coulson AR (1977) DNA sequencing with chain terminating inhibitors. Proc Natl Acad Sci USA 74:5463–5467

Skoda RC, Gonzalez FJ, Demierre A, Meyer UA (1988) Two mutant alleles of the human cytochrome P450dbl gene (P450C2D1) associated with genetically deficient metabolism of debrisoquine and other drugs. Proc Natl Acad Sci USA 85:5240–5243