Evolution of Cytochrome P-450 Proteins[1]

David R. Nelson and Henry W. Strobel

Department of Biochemistry and Molecular Biology, The University of Texas Medical School at Houston

Thirty-four cytochrome P-450 sequences from one bacterial and six vertebrate species have been aligned with the aid of a computer alignment algorithm. Phylogenetic trees were constructed using the unweighted-pair-group and neighbor-joining methods. The two trees differed at only a single branch point near the base of the tree. The cytochrome P-450 superfamily of proteins clustered into eight families and contained 16 gene-duplication events. The first gene duplication occurred ~1,360 Myr before the present (Mybp) and gave rise to cytochrome P-450s found in two different cellular organelles, the mitochondria and the endoplasmic reticulum. Both groups utilize cholesterol or its metabolites as substrates, implying that cholesterol existed >1,360 Mybp. The fourth gene duplication (~900 Mybp) gave rise to the drug-metabolizing P-450s. These proteins aid in the detoxification of foreign chemicals, as opposed to the metabolism of endogenous compounds. The importance of the capacity to metabolize drugs is reflected in 11 further gene duplications occurring in this lineage. The first occurred ~800 Mybp and gave rise to the two major P-450 families, the phenobarbital and 3-methylcholathrene families. An apparent increase in the rate of cytochrome P-450 evolution is noted between the bird-mammal divergence (300 Mybp) and the mammalian radiation (75 Mybp).

Introduction

Cytochrome P-450 proteins are heme-containing enzymes with an unusual thiolate (S^-) ligand to the heme iron. These proteins utilize electrons from reduced nicotinamide adenine dinucleotide phosphate (NADPH) (or sometimes from reduced nicotinamide adenine dinucleotide) [NADH]) to activate molecular oxygen. The reducing equivalents are relayed to the cytochrome P-450 via one or two additional proteins that form a short electron-transport chain. Cytochrome P-450s are found in diverse organisms, from bacteria to plants and animals, implying the existence of a common ancestor before the eukaryote-prokaryote divergence.

The ability to activate molecular oxygen and subsequently to insert one oxygen atom into a substrate has been exploited by organisms for many purposes. Consequently, there are dozens of different cytochrome P-450 molecules even in a single organism. Each P-450 enzyme has a different substrate specificity, but, unlike the situation with most enzymes, the substrate specificity of a single cytochrome P-450 may be quite broad. The substrates are lipophilic molecules frequently containing

1. Key words: cytochrome P-450, evolution, phylogenetic tree, sequence alignment, evolutionary rate increase. Abbreviations: MC = 3-methylcholanthrene; PB = phenobarbital; PCN = pregnenolone-16-α-carbonitrile; UPGMA = unweighted-pair-group method of analysis; PAMs = accepted point mutations; NADPH = nicotinamide adenine dinucleotide phosphate, reduced.

Address for correspondence and reprints: Dr. Henry W. Strobel, Department of Biochemistry and Molecular Biology, The University of Texas Medical School, P.O. Box 20708, Houston, Texas 77225.

Mol. Biol. Evol. 4(6):572–593. 1987.

0737-4038/87/0406-0002$02.00

multiple ring structures. Some major functions of cytochrome P-450 proteins include the metabolism of drugs and other foreign compounds, steroidogenesis and bile acid production from cholesterol, and ω-oxidation of fatty acids and prostaglandins (for reviews on cytochrome P-450, see White and Coon [1980], Adesnik and Atchison [1985], Nebert and Gonzalez [1985], and Black and Coon [1986]).

Because cytochrome P-450 proteins are so diverse, it will be helpful to categorize them. The first major dichotomy is between eukaryotic and prokaryotic P-450s. The bacterial P-450s form a large class of soluble proteins, in contrast to the eukaryotic enzymes, which are all integral membrane proteins. According to Sligar and Murray (1986), the bacterial cytochrome P-450s "are just as diverse, if not more so, than their eukaryotic counterparts." At present, only one bacterial sequence is known.

Among the eukaryotic P-450s, one bird and 32 mammalian sequences are known. These are found in two cellular locations, the mitochondria (two sequences) and the endoplasmic reticulum (31 sequences). The mitochondrial enzymes catalyze the earliest steps in the conversion of cholesterol into steroid hormones. These proteins have a leader sequence that is removed upon import into the mitochondrion. The endoplasmic-reticulum enzymes are referred to as microsomal P-450s, and they do not have a processed leader sequence, although their N-terminal is very hydrophobic.

The microsomal P-450s can be divided into two groups on the basis of their substrates. The majority of sequenced microsomal P-450s act on compounds foreign to the host cell, such as drugs. They constitute part of a detoxification system and will be referred to as the drug-metabolizing P-450s. The remainder of the microsomal P-450s act on endogenous substrates. They are involved in steroid-hormone biosynthesis and metabolism and in the metabolism of fatty acids and prostaglandins and will be referred to as endogenous P-450s.

Two-thirds of the known P-450 sequences are drug-metabolizing cytochrome P-450s. These proteins can be induced up to 70-fold by the administration of drugs. The proteins fall into two families that are induced by two different classes of drugs, the phenobarbital (PB) family and the 3-methylcholanthrene (MC) family. Table 1 lists the cytochrome P-450 sequences grouped into protein families (nomenclature of Dayhoff [1979, p. 375]). All of these sequences have been described since 1982. There are also several known pseudogene sequences, but they will not be discussed here.

Cytochrome P-450 genes are arranged in multigene families that together constitute a gene superfamily. The actual number of P-450 genes in mammals is not known, but a lower limit may be estimated on the basis of molecular-hybridization studies. Two non-cross-reacting probes have indicated a minimum of 15 genes in the PB family (Atchison and Adesnik 1983; Mizukami et al. 1983*a*). Five rat and six rabbit P-450 sequences are known, so approximately one-third of the genes in these species have been sequenced. It is not known how many of these genes are functional. The MC family has two or three genes. Two sequences each are known in rats, mice, and humans. Three are present in rabbits. Therefore, the majority of MC-family genes in these species may already be sequenced. The glucocorticoid-inducible P-450$_{PCN}$ (also called P-450$_p$) shows six to eight hybridization bands (Gonzalez et al. 1985). P-450 C21 is present as two genes in each of cattle, mice, and humans, with only one gene each functional in mice and man (Chung et al. 1986*a*). The LA$_\omega$ family contains two or three genes (Hardwick et al. 1987). There are at least six additional P-450 proteins in mammals: P-450 17α, P-450$_{scc}$, P-450 11β, P-450 aromatase, 25-hydroxy vitamin D$_3$ 1α-hydroxylase, and cholesterol 26 hydroxylase. Therefore, the minimum number of P-450 genes is 33, with the possibility that there are others. The chromosomal

Table 1
Cytochrome P-450 Sequences

Enzyme (Source[a])	No. of Amino Acids	Gene, mRNA, or Amino Acid Sequence	Reference(s)
PB family:			
P-450 PBc1 (rabbit liver)	480(I)	mRNA	Leighton et al. 1984
P-450 PBc2 (rabbit liver)	479(I)	mRNA	Leighton et al. 1984
P-450 l (rabbit liver)	487	mRNA	Tukey et al. 1985
P-450 f (rat liver)	490	mRNA	Friedberg et al. 1986
		mRNA[b]	Gonzalez et al. 1986
P-450 PB1 (rat liver)	490	mRNA	Friedberg et al. 1986
		mRNA[b]	Gonzalez et al. 1986
P-450 3b (rabbit liver)	490	mRNA	Leighton et al. 1984
		Amino acid[b]	Ozols et al. 1985
P-450 3a (rabbit liver)	492	mRNA	Khani et al. 1987
P-450 j (human liver)	493	mRNA	Song et al. 1986
P-450 j (rat liver)	493	mRNA	Song et al. 1986
P-450 (chicken liver)	491	mRNA	Hobbs et al. 1986
P-450 e (rat liver)	491	mRNA	Fujii-Kuriyama et al. 1982
		mRNA	Affolter and Anderson 1984
		Gene[b]	Mizukami et al. 1983*b*
		Amino acid	Yuan et al. 1983
P-450 b (rat liver)	491	mRNA	Fujii-Kuriyama et al. 1982
		mRNA[b]	Suwa et al. 1985
		Gene	Gotoh et al. 1983
		Amino acid	Yuan et al. 1983
P-450 LM2 (rabbit liver)	491	Amino acid	Heinemann and Ozols 1983
		Amino acid[b]	Tarr et al. 1983
P-450 (human liver)	331(I)	mRNA	Phillips et al. 1985*a*
MC family:			
P-450 c (rat liver)	524	gene[b]	Sogawa et al. 1984
		gene	Hines et al. 1985
		mRNA	Yabusaki et al. 1984
P_1-450 (mouse liver)	524	mRNA	Kimura et al. 1984*a*
P_1-450 (human liver)	512	mRNA[b]	Jaiswal et al. 1985*a*
		Gene	Jaiswal et al. 1985*b*
		Gene	Kawajiri et al. 1986
P-450 6 (rabbit liver)	464(I)	mRNA	Okino et al. 1985
P-450 4 (human liver)	516	mRNA, gene	Quattrochi et al. 1986
P-450 4 (rabbit liver)	424(I)	mRNA	Okino et al. 1985
P-450 LM4 (rabbit liver)	514	Amino acid	Fujita et al. 1984
		Amino acid[b]	Ozols 1986
P-450 d (rat liver)	513	mRNA[b]	Kawajiri et al. 1984
		Gene	Sogawa et al. 1985
		Amino acid	Haniu et al. 1986
P_3-450 (mouse liver)	513	mRNA[b]	Kimura et al. 1984*b*
		mRNA	Kimura and Nebert 1986
17α Hydroxylase family:			
P-450 17α (bovine adrenal cortex microsomes)	509	mRNA	Zuber et al. 1986
P-450 17α (human adrenal cortex microsomes)	508	mRNA	Chung et al. 1987

Table 1 (*Continued*)

Enzyme (Source[a])	Amino Acid	Gene, mRNA, or Amino Acid Sequence	Reference(s)
C21 Hydroxylase family:			
P-450 C21 (human adrenal cortex microsomes)	494	Gene	Higashi et al. 1986
		Gene[b]	White et al. 1986
P-450 C21 (bovine adrenal cortex microsomes)	496	Gene[b]	Chung et al. 1986*a*
		mRNA	John et al. 1986
		mRNA	Yoshioka et al. 1986
P-450 C21 (mouse adrenal cortex microsomes)	487	Gene	Chaplin et al. 1986
PCN family:			
P-450 p (rat liver)	504	mRNA	Gonzalez et al. 1985
P-450 HLp (human liver)	504	mRNA	Molowa et al. 1986
		mRNA[b]	Beaune et al. 1986
scc Family:			
$P\text{-}450_{scc}$ (bovine adrenal cortex mitochondria)	520(P)	mRNA[b]	Morohashi et al. 1984
	481(M)	Amino acid	Chashchin et al. 1986
$P\text{-}450_{scc}$ (human adrenal cortex mitochondria)	521(P)	mRNA	Chung et al. 1986*b*
	482(M)		
Lauric acid ω-hydroxylase family:			
P-450 LA_{ω} (rat liver)	509	mRNA	Hardwick et al. 1987
Bacterial P-450 family:			
$P\text{-}450_{cam}$ (*Pseudomonas putida*)	414	Amino acid	Haniu et al. 1982
		Gene[b]	Unger et al. 1986

NOTE.—I = Incomplete; P = precursor; and M = mature.

[a] Site of high concentration of cytochrome P-450 proteins, a site that may or may not be the source of the cDNA or genomic library used to obtain a sequence.

[b] Sequence used in this alignment when multiple sequences were available (see text for exceptions).

location of these genes is summarized in table 2 (for a more complete discussion of the genetics of cytochrome P-450, see Adesnik and Atchison 1985).

The present study presents an alignment of 34 sequences and a phylogenetic tree based on both the unweighted-pair-group method of analysis (UPGMA; Sneath and Sokal 1973, pp. 230–234) and the neighbor-joining method of Saitou and Nei (1987). The cytochrome P-450 superfamily of proteins has undergone many gene-duplication events, especially in the PB family of P-450s. Some of the gene duplications appear to have occurred much earlier than previously suspected.

Material and Methods

Table 1 lists the cytochrome P-450 sequences used in the present study. Sixteen identical or very similar sequences were determined by more than one laboratory or from more than one of the following sources: protein, mRNA, or genomic DNA. Discrepancies were found at 90 of the 7,947 amino acids in these proteins. These differences are listed in the Appendix. Twelve differences are due to incorrect translation of codons into amino acids (P450b alignment positions 82, 187, 230, 270, 295, and 322 [Suwa et al. 1985]; P-450e positions 210, 365, and 460 [Fujii-Kuriyama et al.

Table 2
Chromosomal Location of Cytochrome P-450 and Related Genes

Gene or Gene Family	Chromosome		Reference(s)
	Mouse	Human	
P-450e and b-like genes	7	19	Simmons and Kasper 1983, Phillips et al. 1985*b*
MC family	9	15	Tukey et al. 1984, Hildebrand et al. 1985
Ah receptor[a]	17, 12	...	Legraverend et al. 1984, Cobb et al. 1987
17α Hydroxylase	...	10	Matteson et al. 1986
C21 Hydroxylase	17	6	White et al. 1984, Carroll et al. 1985
PCN family	6	...	Simmons et al. 1985
$P\text{-}450_{scc}$	...	15	Chung et al. 1986*b*
NADPH cytochrome P-450 reductase[b]	6	...	Simmons et al. 1985

[a] Required for the induction of the MC family of P-450s. Two loci appear to be involved.
[b] The electron-transfer protein common to all microsomal P-450s.

1982]; P-450c positions 318 and 337 Hines et al. [1985]; and P1h 444 [Jaiswal et al. 1985*b*]). Four differences (LM2 116, scc bovine 58, cam 403, and cam 580) were probably caused by deamidation of GLN or ASN during peptide isolation. Two highly conserved amino acids (LM2 120 and LM2 121) and four moderately conserved amino acids (LM2 175; LM2 176, LM4 338, and LM4 339) appear to have been inverted. Four amino acids (LM2 124, LM2 125, cam 91, and cam 92) were not detected during amino acid sequencing, but they were present in another sequence. In the paper of Sogawa et al. (1985) a printing error made GLY look like GLU (position 105). Position 526 in $P\text{-}450_{cam}$ has been confirmed as HIS by means of X-ray crystallography (Poulos et al. 1985). The remaining 62 differences were not as easily explained. They may have been caused by sequencing distinct genes or polymorphic alleles of the same gene, or they could be due to sequencing error. Thirteen differences occurred at positions of 100% conservation within the alignment of the PB or MC families (PB1 326, b 474, LM4 28, LM4 59, LM4 194, LM4 270, LM4 281, d 44, d 158, d 305, d 469, P3 449, and P1h 444). The last was a codon-translation error mentioned above. Three additional differences (LM4 509, 23/25 sequences; LM4 581, 8/9 sequences; and C21h 520, 32/33 vertebrate sequences) showed very high conservation of a particular amino acid. These 16 differences may represent sequencing errors, although a genetic difference at these positions is possible. In these cases, the conserved amino acid was used in the alignment. The remaining 46 differences belong to two different types of sequences, those in families known to have multiple genes and those in families with one or a very low number of functional genes; P-450s f, PB1, 3b, e, LM2, and HLp fall into the first group, and P-450s c, P1h, LM4, and C21 bovine are members of the second group. Differences (not already discussed) between sequences in the first group probably represent differences between distinct genes, although some differences may be due to polymorphic alleles or sequencing errors; differences (not already discussed) between sequences in the second group are unlikely to represent differences between distinct genes but are more likely due either to differences between polymorphic alleles or to sequencing errors. The sequences chosen for use in this alignment

when more than one sequence were available are indicated by a superscript b in table 1, with the following exceptions: The complete P-450 3b amino acid sequence of Ozols et al. (1985) was used except at positions 115, 281, 418, and 517, where the incomplete cDNA sequence of Leighton et al. (1984) more closely resembled sequences 1–5. The complete LM4 sequence of Ozols (1986) was used except at positions 28, 194, 338, and 339, where the sequence of Fujita et al. (1984) was favored for reasons already mentioned. The P-450d sequence of Kawajiri et al. (1984) was used except at position 305, where PHE is favored over SER, as per Haniu et al. (1986) and Sogawa et al. (1985). The human p-like sequence (HLp or nifedipine oxidase) of Beaune et al. (1986) was used except at position 343, where LYS was favored over a gap in the sequence. The chosen sequences more closely resembled other sequences in their respective families.

Alignment of Amino Acid Sequences

Alignment of P-450 sequences within families was done manually because these sequences were >40% identical in the PB family and >60% identical in the MC family. Sequence alignment between these two families was achieved by comparing cytochrome P-450c with cytochrome P-450e by means of the alignment algorithm of Needleman and Wunsch (1970) as modified by Gotoh (1982). The gap-weighting factor in this program was $2k + 3$, where k is the gap width. The computer-generated alignment of P-450c and P-450e was used to align the two large sets of sequences. Some judgment was used in this process. The more distantly related sequences—P-450s 17α, C21b, p, scc, and cam—were aligned with cytochrome P-450e by computer. The gap-weighting factor was increased in some instances to $4k + 3$ or to $5k + 3$ to make gaps less likely. Once the computer-generated alignments were available, the sequences were individually incorporated into the larger body of sequences by using the computer alignment as a guide. Because of the large number (561) of pairwise comparisons involved, computer alignment of all the pairs was neither practical nor deemed necessary. After the initial alignment was completed, the overall length of the alignment was shortened by 10%, from 660 positions to 594. The longest sequence was 524 amino acids. In this shortening procedure, the number of gaps was reduced by 50%. The alignment of all 34 sequences is shown in figure 1.

The Phylogenetic Tree of Cytochrome P-450 Proteins

The 34 cytochrome P-450 sequences were analyzed for the number of identities in 561 pairwise comparisons. Only those portions of the molecules that overlapped were compared. Twenty nine of the 34 sequences are complete. Two others (PBc1 and PBc2) are >97% complete. P-450 6 and P-450 4 are ~90% and ~82% complete, respectively. The human P-450 sequence of Phillips et al. (1985*a*) is two-thirds complete. The alignment contains a total of 16,659 amino acids. Gaps were not counted as differences and did not contribute to the overall length of the sequences compared. Once a percentage difference had been obtained for each pair, the evolutionary distance in accepted point mutations (PAMs) was estimated by using table 36 of Dayhoff (1979, p. 375). To simplify conversion of percentage difference values to PAM values, the following three equations, which closely approximate the PAM values over given ranges of percentage difference, were used:

$$-0.808 \ln(2 - 2.5x) + 0.56$$
$$0.01 \leq x \leq 0.60; \tag{1}$$

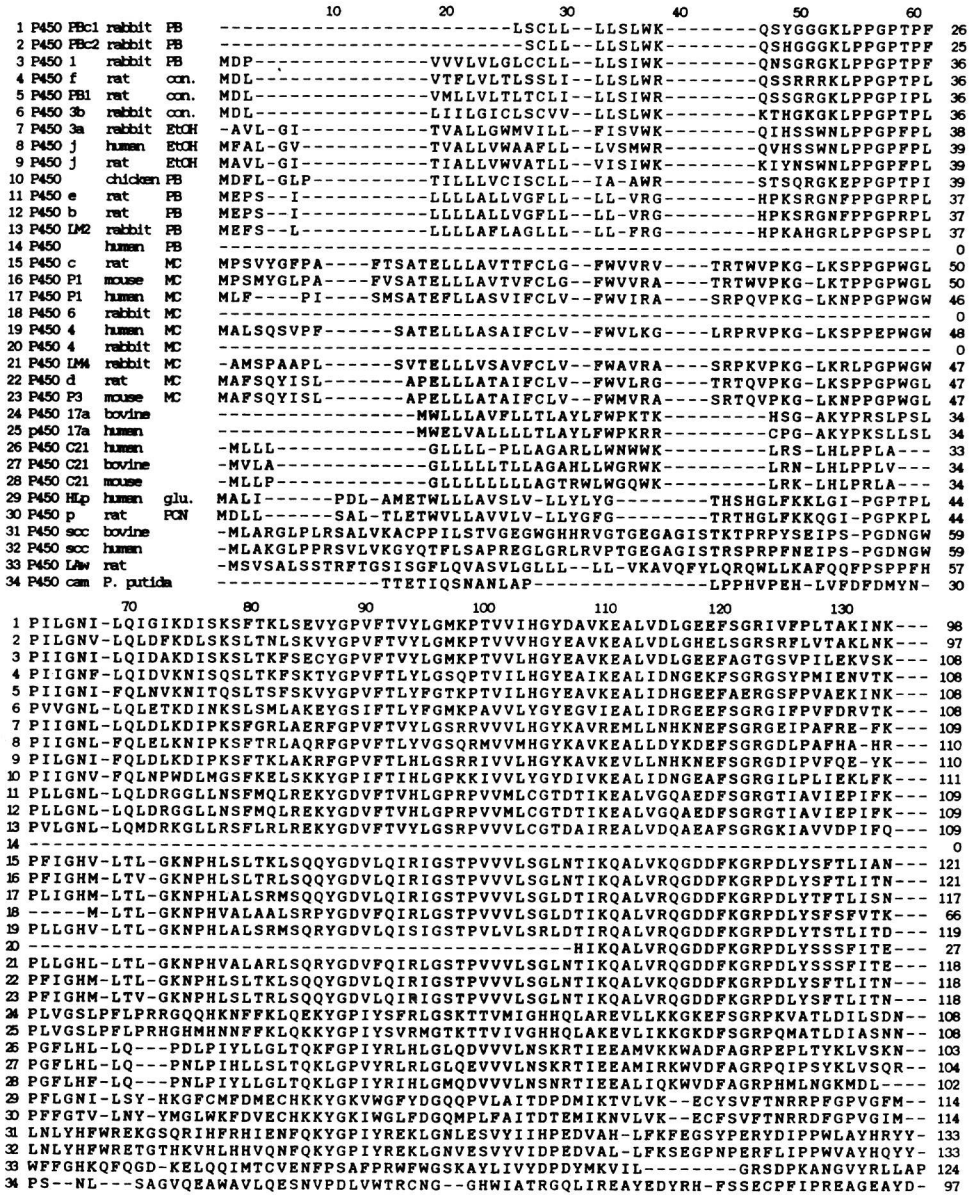

FIG. 1.—Alignment of cytochrome P450 sequences. The column following the organism indicates the inducer class. Con. = constitutive; glu. = glucocorticoids; PCN = pregnenolone-16α-carbonitrile. Numbers on the far right are the number of amino acids up to that position.

$$0.1372e^{3.5x} \qquad 0.50 \le x \le 0.70; \tag{2}$$

$$0.1833 \exp \frac{1.3}{1.3-x} \qquad 0.70 < x \le 0.89\,. \tag{3}$$

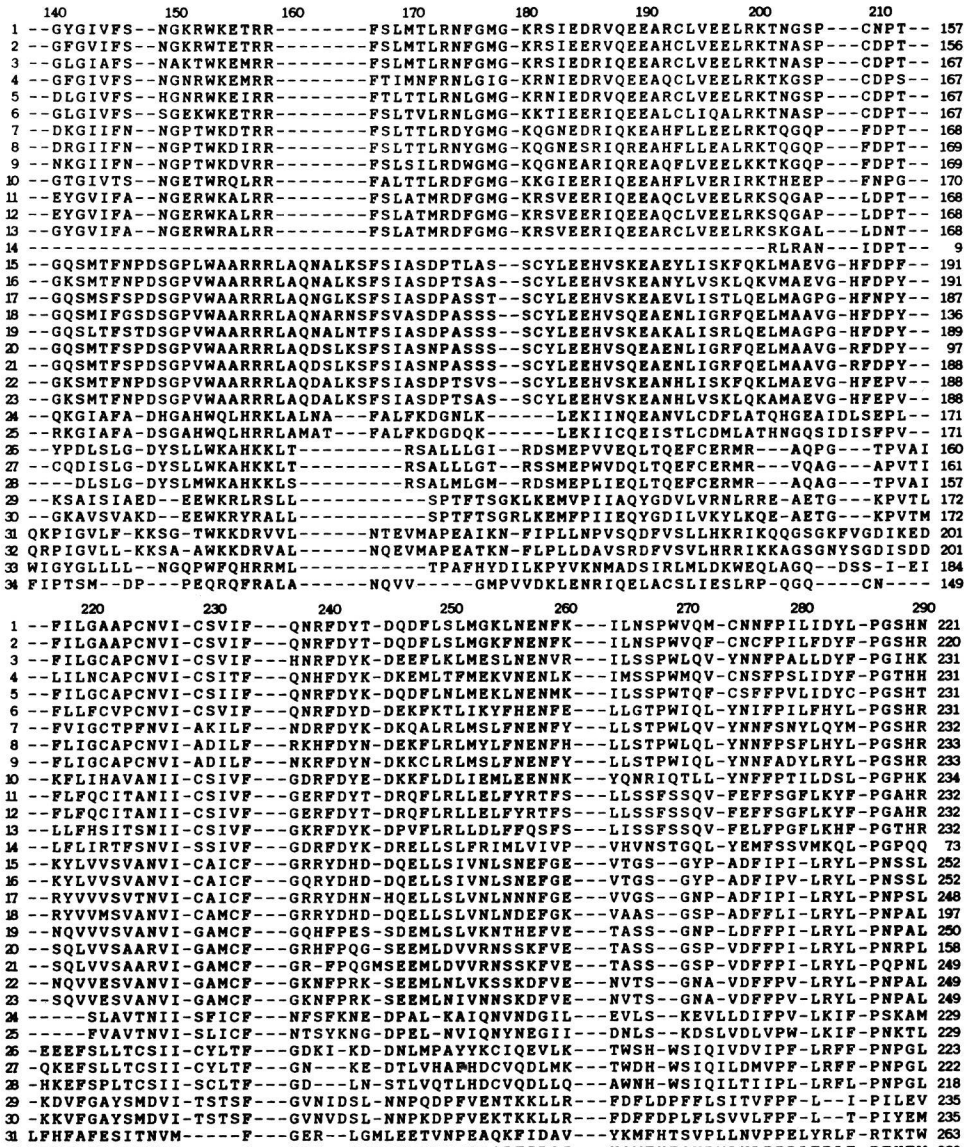

FIG. 1 (*Continued*)

These equations yield PAMs/100, where x is the percentage difference/100. Table 3 contains the percentage differences between all the pairs of cytochrome P-450 sequences (above the diagonal) and the number of identities between pairs (below the diagonal). A UPGMA phylogenetic tree was constructed by using the evolutionary-distance (d) values obtained by applying equations (1–3) to the percentage difference data of table 3. A second phylogenetic tree was constructed by using the neighbor-joining method of Saitou and Nei (1987). This method is a minimum-evolution method; that is, it does not assume a constant rate of evolutionary change in all branches of the tree. The program was obtained in FORTRAN and converted to BASIC

```
          300        310        320        330        340        350        360
 1 -KILRN-NIYIR-NYVLEKIKEHQETLDI--NNPRD---FIDCFLI-KMEQEK--DNQQSE----FTIENLMTTVT 282
 2 -KAVKN-IFYVK-NYITEQIKEHQKSLDI--NNPRD---FIDCFLI-KMEQEK--CNQQSE----FTIENLLTTVS 281
 3 -TLLKN-ADYIK-NFIMEKVKEHQKLLDV--NNPRD---FIDCFLI-KMEQEN--N---LE----FTLESLVIAVS 289
 4 -KIAKN-INYMK-SYLLKKIEEHQESLDV--TNPRD---FVDYYLI-KQKQAN--NIEQSE----YSHENLTCSIM 292
 5 -TLAKN-VYHIR-NYLLKKIKEHQESLDV--TNPRD---FIDYYLI-KWKQEN--HNPHSE----FTLENLSITVT 292
 6 -QLFKN-IDGQI-KFILEKVQEHQESLDS--NNPRD---FVDHFLI-KMEKEK--HKKQSE----FTMDNLITTIW 292
 7 -KVIKN-VSEIK-EYTLARVKEHHKSLDP--SCPRD---FIDSLLI-EMEKDK--HSTEPL----YTLENIAVTVA 293
 8 -KVIKN-VAEVK-EYVSERVKEHHQSLDP--NCPRD---LTDCLLV-EMEKEK--HSAERL----YTMDGITVTVA 294
 9 -KIMKN-VSEIK-QYTLEKAKEHLQSLDI--NCARD---VTDCLLI-EMEKEK--HSQEPM----YTMENVSVTLA 294
10 -TLIKN-TETVD-DFIKEIVIAHQESFDA--SCPRD---FIDAFIN-KMEQEK--EN--SY----FTVESLTRTTL 293
11 -QISKN-LQEIL-DYIGHIVEKHRATLDP--SAPRD---FIDTYLL-RMEKEK--SNHHTE----FHHENLMISLL 293
12 -QISKN-LQEIL-DYIGHIVEKHRATLDP--SAPRD---FIDTYLL-RMEKEK--SNHHTE----FHHENLMISLL 293
13 -QIYRN-LQEIN-TFIGQSVEKHRATLDP--SNPRD---FIDVYLL-RMEKDK--SDPSSE----FHHQNLILTVL 293
14 -QAFQL-LQGLE-DFIAKKVE-HNTPLDP--NSPRD---FIDSFLI-RMQEEE--KNPNTE----FYLEKLVMTSL 133
15 -DAFKD-LNKKFYSFMKKLIKEHYRTFEK--GHIRD---ITDS-LIEHCQDRRLDENANVQ----LSDDKVITIVF 316
16 -DAFKD-LNDKFYSFMKKLIKEHYRTFEK--GHIRD---ITDS-LIEHCQDRKLDENANVQ----LSDDKVITIVL 316
17 -NAFKD-LNEKFYSFMQKMVKEHYKTFEK--GHIRD---ITDS-LIEHCQEKQLDENANVQ----LSDEKIINIVL 312
18 -DTFKD-LNERFYSFTQERVKEHCRSFEK--GHIRD---ITDS-LIKHYRVDRLDENANVQ----VSDEKTVGIVL 261
19 -QRFKA-FNQRFLWFLQKTVQEHYQDFDK--NSVRD---ITGA-LFKHSKKGP-RASGNL-----IPQEKIVNLVN 312
20 -QRFKD-FNQRFLRFLQKTVREHYEDFDR--NSIQD---ITGA-LFKHSEKNS-KANGGL-----IPQEKIVNLVN 220
21 -RRFKD-FNQRFLRFLQKTVREHYEDFDR--NSIQD---ITGA-LFKHSEKN--KANGGL-----IPQEKIVNLVN 310
22 -KRFKN-FNDNFVLFLQKTVQEHYQDFNK--NSIQD---ITGA-LFKHSENY--KDNGGL-----IPQEKIVNIVN 310
23 -KRFKT-FNDNFVLFLQKTVQEHYQDFNK--NSIQD---ITSA-LFKHSENY--KDNGGL-----IPEEKIVNIVN 310
24 -EKMKG-CVQTRNELLNEILEKCQENFSS--DSITN---LLHI-LIQAKVNADNNNAGPDQDSKLLSNRHMLATIG 297
25 -EKLKS-HVKIRNDLLNKILENYKEKFRS--DSITN---MLDT-LMQAKMNSDNGNAGPDQDSELLSDNHILTTIG 297
26 -RRLKQ-AIEKRDHIVEMQLRQHKESLVA--GQWRD---MMDY-MLQGVAQPSMEE-GSGQ----LLEGHVHMAAV 286
27 -WRLKQ-AIENRDHMVEKQLTRHKESMVA--GQWRD---MTDY-MLQGVGRQRVEE-GPGQ----LLEGHVHMSVV 285
28 -QKLKQ-IQESRDHIVKQQLKQHKESLVA--GQWKD---MIDY-MLQGVEKQRDGK-DEEQ----LHEGHVHMSVV 281
29 -LNICV-FPREVTNFLRKSVKRMKESRLE--DTQKH---RVD-FLQLMIDSQKNSKETESHKA--LSDLELVAQSI 301
30 -LNICM-FPKDSIEFFKKFVYRMKETRLD---SVQKH---RVD-FLQLMMNAHNDSKDKESHTA--LSDMEITAQSI 301
31 -RDHVA-AWDTIFNKAEKYTEIFYQDLRRK-TEFRNYPGILYCLL--KSEK--------------MLLEDVKANIT 320
32 -KDHVA-AWDVIFSKADIYTQNFYWELRQKGSVHHDYRGMLYRLL--GDSK--------------MSFEDIKANVT 321
33 AHDHTDGVIKLRKDQLQNAGE--LEKVKK--KRRLD---FLDILLLARME------NG---DS--LSDKDLRAEVD 315
34 PDGSMT-FAEAK-EALYDYLIPIIEQRRQ--KPGTD---AISIVANGQV-------NGRPI----TSDEAKRM-CG 243
     370        380        390        400        410        420        430        440
 1 DVFGAGTETTSTTLRYGLLLLMKHPEVIA-KVQEEIERVIG----RHRSPCMQDRSRM----PYTDATVHEIQRYI 349
 2 DVFMAGTETTSTTLRYGLLLLMKHPEVIA-KVQEEIERVIG----RHRSPCMQDRSRM----PYTDATVHEIQRYI 348
 3 DLFGAGTETTSTTLRYSLLLLLKHPEVAA-RVQEEIERVIG----RHRSPCMQDRSRM----PYTDAVIHEIQRFI 356
 4 DLIGAGTETMSTTLRYALLLMLKYPHVTA-KVQEEIDRVIG----RHRSPCMQDRKHM----PYTDAMIHEVQRFI 359
 5 DLFGAGTETTSTTLRYALLLLLKCPEVTA-KVQEEIDRVVG----KHRSPCMQDRSRM----PYTDAHDHEVQRFI 359
 6 DVFSAGTDTTSNTLKFALLLLKHPEITA-KVQEEIEHVIG----RHRSPCMQDRSRM----PYTDAVMHEIQRYV 359
 7 DMFFAGTETTSTTLRYGLLILLKHPEIEE-KLHEEIDRVIG----PSRMPSVRDRVQM----PYMDAVVHEIQRFI 360
 8 DLFFAGTETTSTTLRYGLLILMKYPEIEE-KLHEEIDRVIG----PSRIPAIKDRQEM----PYMDAVVHEIQRFI 361
 9 DLFFAGTETTSTTLRYGLLILMKYPEIEE-KLHEEIDRVIG----PSRVPAVRDRLDM----PYMDAVVHEIQRFI 361
10 DLFLAGTGTTSTTLRYGLLILLKHPEIEE-KMHKEIDRVVG----RDRSPCMADRSQL----PYTDAVIHEIQRFI 360
11 SLFFAGTETGSTTLRYGFLLMLKYPHVTV-KVQKEIDQVIG----SHRPPSLDDRTKM----PYTDAVIHEIQRFA 360
12 SLFFAGTETSSTTLRYGFLLMLKYPHVAE-KVQKEIDQVIG----SHRLPTLDDRSKM----PYTDAVIHEIQRFS 360
13 SLFFAGTETTSTTLRYGFLLMLKYPHVTE-RVQKEIEQVIG----SHRPPALDDRAKM----PYTDAVIHEIQRLG 360
14 NLFIGGTETVSTTLRYGFLLLIKHPGVEA-KVHEEIDRVIG----KNRQPKFEDRAKM----PYMEAMIHEIQRFG 200
15 DLFGAGFDTITTAISWSLMYLVTNPRIQR-KIQEELDTVIG----RDRQPRLSDRPQL----PYLEAFILETFRHS 383
16 DLFGAGFDTVTTAISWSLMYLVTNPRVQR-KIQEELDTVIG----RDRQPRLSDRPQL----PYLEAFILETFRHS 383
17 DLFGAGFDTVTTAISWSLMYLVMNPRVQR-KIQEELDTVIG----RSRRPRLSDRSHL----PYMEAFILETFRHS 379
18 DLFGAGFDTVTTAISWSLMYLVTKRQHTREKIQEELDAVVG----RARRPRFSDRPQL----PYLEAVIMETFRHT 329
19 DIFGAGFDTVTTAISWSLMYLVTKPEIQR-KIQKELDTVIG----RERRPRLSDRPQL----PYLEAFILETFRHS 379
20 DIFGAGFDTITTALSWSLMYLVTNPRRQR-KIQEELDAVVG----RARQPLLSDRPQL----PYLEAFILELFRHT 287
21 DIFGAGFDTITTALSWSLMYLVTNPRRQR-KIQEELDAVVG----RARQPRLSDRPQL----PYLEAFILELFRHT 377
22 DIFGAGFETVTTAIFWSILLLVTEPKVQR-KIHEELDTVIG----RDRQPRLSDRPQL----PYLEAFILEIYRYT 377
23 DIFGAGFDTVTTAITWSILLLVTWPNVQR-KIHEELDTVVG----RDRQPRLSDRPQL----PYLEAFILEIYRYT 377
24 DIFGAGVETTTSVIKWIVAYLLHHPSLKK-RIQDDIDQIIG----FNRTPTISDRNRL----VLLEATIREVLRIR 364
25 DIFGAGVETTTSVVKWTLAFLLHNPQVKK-KLYEEIDQNVG----FSRTPTISDRNRL----LLLEATIREVLRLR 364
26 DLLIGGTETTANTLSWAVVFLLHHPEIQQ-RLQEELDHELGPGASSSRVP-YKDRARL----PLLNATIAEVLRLR 356
27 DLFIGGTETTASTLSWAVAFLLHHPEIQR-RLQEELDRELGPGASCSRVT-YKDRARL----PLLNATIAEVLRLR 355
28 DLFIGGTETTATTLSWAVAFLLHHPEIQK-RLQEELDLKLGPG----SQLLYRNRMQL----PLLMATIAEVLRLR 348
29 IFIFAGYETTSSVLSFIMYELATHPDVQQ-KLQEEIDAVLP----NKAPPTYDTVLQM----EYLDMVVNETLRLF 368
30 IFIFAGYEPTSSTLSFVLHSLATHPDTQK-KLQEEIDRALP----NKAPPTYDTVMEM----EYLDMVLNETLRLY 368
31 EMLAGGVNTTSMTLQWHLYEMARSLNVQE-MLREE---VLN-----ARRQAEGDISKMLQMVPLLKASIKETLRLH 387
32 EMLAGGVDTTSMTLQWHLYEMARNLKVQD-MLRAE---VLA-----ARHQAQGDMATMLQLVPLLKASIKETLRLH 388
33 TFMFEGHDTTASGVSWIFYALATHPKHQQ-RCREEVQSVLG-----DGSSITWDHLD--QI-PYTTMCIKEALRLY 382
34 LLLVGGLDTVVNFLSFSMEFLAKSPEHRQ----ELIER-------PERIPAACE----------------ELLRRF 292
```

FIG. 1 (*Continued*)

for use on an IBM XT with a BASIC compiler. To produce a rooted tree for comparison with the UPGMA tree, a set of dummy data was included as an outgroup.

Results

The Phylogenetic Tree

Figure 2 depicts the UPGMA phylogenetic tree of cytochrome P-450 proteins. The branching pattern is the same as that obtained before reducing the number of gaps by half, and it is the same whether gaps are not counted in the computation of percentage difference or are counted as single replacements. If a gap of length n is

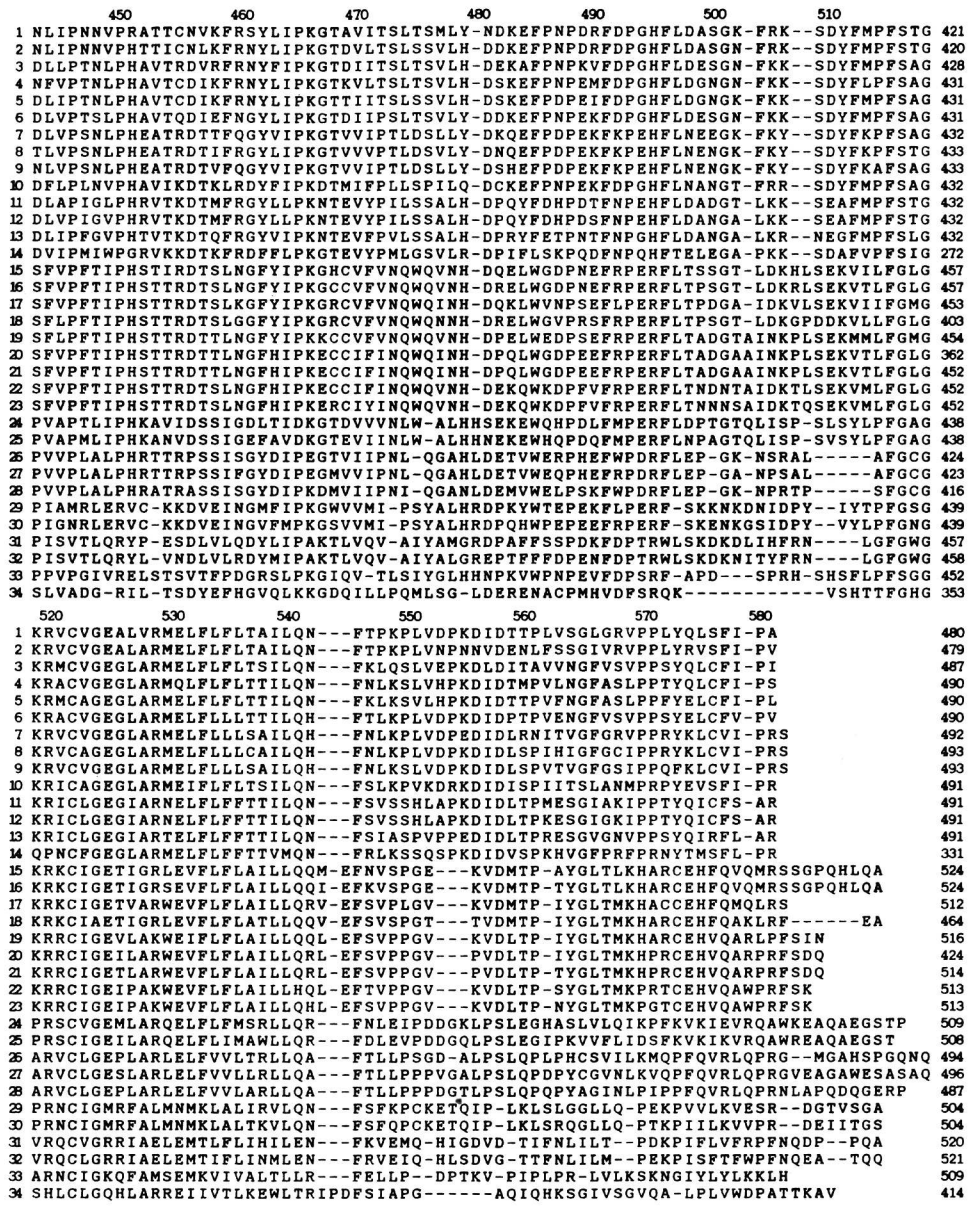

FIG. 1 (*Continued*)

counted as n replacements, an extreme case, the pattern has only one change. The lauric acid ω-hydroxylase (LA_{ω}) sequence clusters with the two side-chain cleavage sequences ($P\text{-}450_{scc}$) rather than forming a separate branch.

The phylogenetic tree produced by the neighbor-joining method, with an outgroup included, is remarkably similar to the UPGMA tree. When gaps are not included in the calculation of percentage difference, the neighbor-joining method produces a tree with only one difference in the branching pattern. Again, the LA_{ω} sequence clusters with the two $P\text{-}450_{scc}$ sequences. This pattern also remains the same if gaps of length n are counted as single replacements or as n replacements. These two independent

Table 3
Percentage Difference (above diagonal) and Number of Identities (below diagonal) between Pairs of Cytochrome P-450 Sequences

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
1....		15.7	26.0	34.4	28.7	34.4	44.1	44.3	43.4	44.9	49.5	48.9	49.1	57.1	68.2	67.6
2....	404		27.9	36.1	32.2	34.2	45.2	45.2	46.0	45.2	49.6	49.2	49.6	57.1	67.1	67.1
3....	353	343		30.8	26.1	32.9	43.1	44.9	44.7	44.0	48.6	48.4	49.2	53.4	68.7	67.9
4....	315	306	337		24.1	39.0	49.2	47.9	47.2	49.1	50.3	50.5	52.4	56.5	67.8	67.4
5....	342	325	360	372		35.1	45.9	44.8	45.0	43.7	48.9	49.1	51.9	54.1	67.8	67.0
6....	315	315	327	299	318		45.1	43.6	44.4	45.6	52.4	52.4	51.7	56.8	67.4	67.6
7....	268	262	276	248	264	268		20.9	18.9	49.6	49.7	49.5	51.3	56.5	69.0	68.4
8....	267	262	268	255	270	276	389		21.3	48.3	52.0	51.4	53.7	57.7	67.4	67.2
9....	271	258	269	258	269	272	399	388		49.9	51.4	50.8	54.1	58.6	68.9	68.3
10....	263	261	272	248	274	265	246	253	245		51.7	51.3	53.0	54.7	67.7	67.7
11....	242	241	250	243	250	233	246	235	238	236		2.9	23.0	50.2	67.3	66.0
12....	245	243	251	242	249	233	247	238	241	238	477		22.6	49.5	67.1	65.8
13....	244	241	247	233	235	236	238	227	225	230	378	380		52.3	68.5	66.9
14....	142	142	153	144	152	143	144	140	137	149	165	167	158		70.3	69.3
15....	149	154	149	154	154	156	149	157	150	155	157	158	151	96		6.9
16....	152	154	153	156	158	155	152	158	153	155	163	164	159	99	488	
17....	148	149	154	150	150	150	147	150	144	153	157	160	155	102	406	410
18....	134	138	138	136	133	138	136	136	134	144	146	145	138	97	352	353
19....	139	143	147	144	142	153	142	148	138	151	156	155	151	92	347	350
20....	117	117	124	122	120	127	118	127	117	128	126	127	119	93	280	280
21....	141	143	150	146	146	151	147	149	142	152	159	160	151	93	342	345
22....	146	148	158	150	149	154	146	153	147	153	158	157	152	95	357	362
23....	142	143	155	148	149	150	146	153	146	152	156	158	152	92	348	365
24....	131	125	139	136	133	133	130	127	131	130	133	132	130	73	146	146
25....	125	118	126	129	136	125	126	124	127	126	130	129	120	76	153	154
26....	123	122	133	131	128	134	133	132	134	127	125	125	125	76	140	137
27....	118	116	127	132	130	122	128	131	132	121	120	121	126	76	148	145
28....	114	112	121	119	120	127	132	132	131	131	123	123	127	77	136	134
29....	112	110	115	115	111	117	114	112	111	118	122	121	122	71	115	119
30....	104	105	115	113	106	117	114	114	107	120	118	118	118	72	119	118
31....	96	94	98	99	100	94	93	97	100	92	96	97	95	61	103	105
32....	84	87	87	92	89	85	84	86	86	86	89	88	95	57	97	98
33....	103	97	94	94	92	92	92	92	89	99	94	95	98	65	101	102
34....	79	80	79	75	80	84	84	91	81	77	86	87	86	46	77	80

NOTE.—Sequence numbers are as in fig. 1.

methods give essentially the same tree, with only one minor difference in tree topology. The integrity of this branching pattern under these extremely varied assumptions argues for its correctness.

A Time Scale for the Phylogenetic Tree

Since most of the sequences in figure 1 are mammalian, calibration of d with time will be most accurate for times more recent than the mammalian radiation. To fix the time scale for the more ancient part of the tree, the divergence between bacterial and eukaryotic P-450s (d = 2.5) has been set at 1,400 Mybp, the date of the earliest eukaryotic microfossils (Vidal 1984). If this and zero were the only points used in the calibration, the divergences between rabbits, humans, and rodents would fall at ~150 Mybp (d for eight divergences 0.27 ± 0.04). Clearly this is unsatisfactory, since the mammalian radiation is commonly considered as having occurred ~75 Mybp (McLaughlin and Dayhoff 1972). To avoid this problem we have chosen the bird-mammal divergence time of 300 Mybp (Dickerson 1971; Romero-Herrera et al. 1978)

17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34
68.4	69.5	70.3	70.7	69.7	68.7	69.6	71.8	73.1	72.7	73.7	74.4	75.2	76.9	78.5	81.2	76.6	79.1
68.2	68.6	69.4	70.7	69.2	68.2	69.3	73.0	74.5	72.8	74.1	74.8	75.6	76.7	78.9	80.5	77.9	78.8
67.6	68.3	69.1	68.7	68.2	66.7	67.3	70.3	73.1	70.8	72.1	73.2	74.9	74.9	78.5	80.9	79.0	79.3
68.7	69.0	69.9	69.4	69.3	68.6	69.0	71.1	72.6	71.4	71.2	73.8	75.1	75.5	78.3	79.8	79.1	80.5
68.7	69.7	70.3	69.9	69.3	68.8	68.8	71.8	71.1	72.1	71.6	73.6	75.9	77.0	78.1	80.5	79.5	79.2
68.7	68.6	68.0	68.2	68.2	67.7	68.6	71.8	73.5	70.7	73.4	72.0	74.6	74.6	79.4	81.4	79.5	78.2
69.2	69.0	70.4	70.4	69.2	69.5	69.5	72.4	73.2	71.0	72.1	71.0	75.2	75.2	79.7	81.7	79.6	78.2
68.7	69.0	69.2	68.2	68.8	68.1	68.1	73.0	73.7	71.2	71.5	71.0	75.7	75.3	78.9	81.3	79.6	76.4
69.9	69.5	71.3	70.7	70.3	69.4	69.6	72.2	73.0	70.8	71.2	71.2	75.9	76.8	78.2	81.3	80.3	79.0
67.9	67.0	68.5	67.8	68.1	68.0	68.2	72.2	73.1	72.1	73.5	71.0	74.3	73.9	80.0	81.3	78.1	79.9
67.2	66.7	67.4	68.4	66.6	66.9	67.4	71.7	72.3	72.7	73.8	72.9	73.6	74.5	79.0	80.5	79.2	77.7
66.5	67.0	67.6	68.2	66.4	67.2	66.9	71.9	72.6	72.7	73.6	72.9	73.8	74.5	78.8	80.7	78.9	77.4
67.6	68.6	68.5	70.2	68.3	68.2	68.2	72.3	74.5	72.7	72.5	72.0	73.6	74.5	79.2	79.2	78.3	77.7
68.4	70.0	71.4	71.1	70.9	70.4	71.3	77.3	76.4	75.8	75.7	75.3	77.7	77.4	79.8	81.1	78.8	82.0
20.7	24.0	32.6	33.8	33.2	30.3	32.0	70.2	68.8	70.5	68.8	71.1	76.0	75.2	78.4	79.7	78.1	80.8
19.9	23.8	32.0	33.8	32.6	29.3	28.7	70.2	68.6	71.1	69.5	71.5	75.2	75.4	78.0	79.5	77.9	80.0
	24.7	27.2	33.1	31.1	32.2	32.0	70.7	68.9	70.6	69.0	71.8	77.1	77.8	79.4	79.7	77.9	78.6
347		31.8	32.6	30.2	36.5	36.7	72.3	71.1	71.2	69.0	71.5	77.5	79.2	79.8	81.0	79.2	80.2
370	313		20.8	22.2	25.1	27.7	71.1	71.9	71.3	70.9	72.2	76.3	77.6	78.0	79.3	79.6	79.9
281	283	336		1.9	24.4	25.1	72.5	68.9	70.6	70.1	71.0	75.8	75.5	77.0	79.4	77.1	77.9
348	319	399	414		25.4	25.4	72.4	70.3	70.5	69.6	71.3	75.3	76.2	77.4	79.5	78.7	78.4
343	291	384	319	381		6.8	70.7	69.7	72.0	71.0	71.6	75.3	76.0	77.9	79.6	78.9	81.3
344	290	371	316	381	478		71.2	69.9	72.7	71.2	71.9	75.1	76.0	78.8	80.5	78.6	80.1
141	124	140	112	133	141	139		28.9	69.6	67.4	69.8	75.5	76.8	81.0	82.9	79.5	82.2
150	129	136	127	143	146	145	361		68.9	66.9	69.5	74.8	74.8	80.5	80.7	75.7	81.7
137	125	134	116	137	130	127	143	146		20.5	27.4	76.9	77.2	80.8	81.1	78.1	80.1
144	134	136	118	142	135	134	154	156	391		27.6	77.2	77.4	81.4	79.9	77.9	81.3
130	122	129	113	133	131	130	141	142	350	352		76.3	76.5	81.9	81.7	78.3	80.8
108	97	112	95	116	116	117	115	118	107	106	109		27.0	78.8	80.3	77.1	81.9
105	90	106	96	112	113	113	109	118	106	105	108	368		79.1	78.8	77.5	82.6
96	85	104	88	107	104	100	88	90	87	85	82	99	98		27.5	79.6	85.4
95	80	98	79	97	96	92	79	89	86	92	83	92	99	377		78.8	85.2
101	85	94	85	98	97	98	92	109	96	97	94	105	103	96	100		82.7
85	72	80	73	86	74	79	69	71	74	70	71	70	67	56	57	66	

to calibrate the d scale from the chicken-mammal divergence (d = 0.71) to the prokaryote-eukaryote divergence (d = 2.51). To calibrate the low end of the scale, the rat-mouse divergence (d = 0.072) was assumed to have occurred ~20 Mybp (Miyata et al. 1982; Shoshani et al. 1985). By this choice, the mammalian radiation (estimated at d = 0.27) occurred ~75 Mybp. The break in the time scale may reflect an increase in evolutionary rate that occurred since the bird-mammal divergence. The estimates of time in the oldest part of the tree are subject to the largest error. It is to be hoped that sequences of P-450s from more nonmammalian sources will soon be available to improve the estimation of a time scale in this part of the tree.

Features of the Tree

There are two major divisions among sequences in this phylogenetic tree. The first involves the separation between eukaryotic and prokaryotic sequences. This branch point represents the base of the tree. The second division is between endogenous P-450s (sequences 24–33) and drug-metabolizing P-450s (sequences 1–23). The gene duplication giving rise to the drug-metabolizing P-450s occurred ~900 Mybp. This

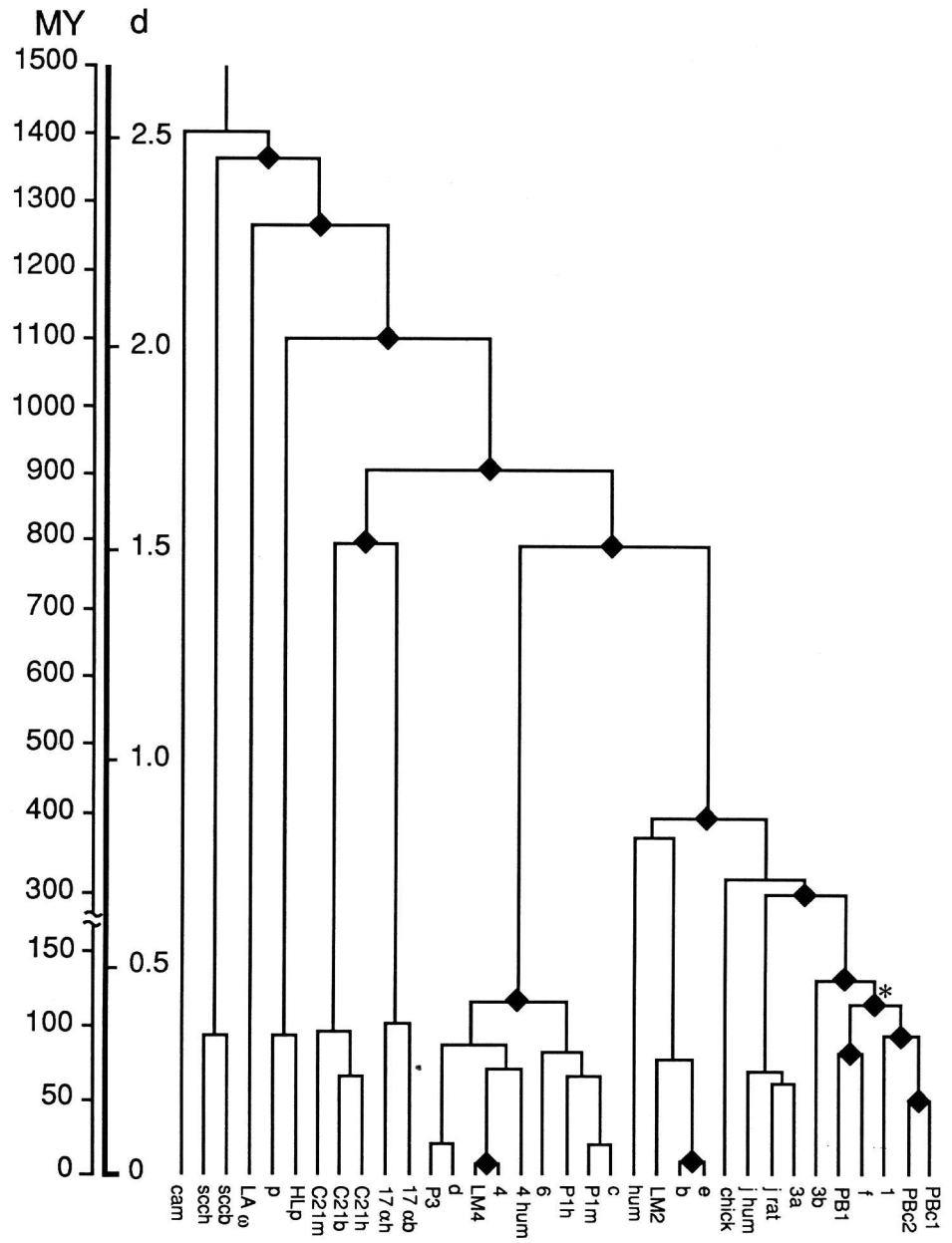

FIG. 2.—UPGMA phylogenetic tree of 34 cytochrome P-450 proteins. Black diamonds indicate gene-duplication events because the same species occurs on both sides of the branch point. An asterisk marks a presumed gene duplication, because the time of the event is too early to represent a species divergence (i.e., rats did not diverge from rabbits 120 Mybp). Unmarked branch points represent species divergences.

group subsequently separated into two distinct families ~800 Mybp. No further branchings occur until ~400 Mybp, when the PB family began to undergo a phenomenal ramification. This period in time corresponds to the Devonian period of the Paleozoic era, the period when vertebrate ancestors began to colonize the land.

Rates of Cytochrome P-450 Evolution

The rate of cytochrome P-450 evolution was calculated from the d values for species-divergence events in figure 2 and from geologic dates for the various divergence events. The rates in amino acid changes per site per year and the unit evolutionary period are listed in table 4. The rates are well below the maximum rate of $\sim 9 \times 10^{-9}$ reported for fibrinopeptides (McLaughlin and Dayhoff 1972). There is, however, an approximately twofold difference between the maximum and minimum rates. All the rates for mammals diverging from mammals are greater than the rates both for mammals diverging from birds and for prokaryotes diverging from eukaryotes.

Discussion

The early history of cytochrome P-450 proteins appears to be related to the evolution of cholesterol and its metabolites. The earliest eukaryotic P-450 sequences yet known are the mitochondrial $P\text{-}450_{scc}$ sequences. These enzymes catalyze cholesterol side-chain cleavage to produce pregnenolone, the precursor of all steroid hormones. The evolution of $P\text{-}450_{scc}$ must have preceded the divergence of plants and animals,

Table 4
Rate of Evolution of Cytochrome P-450 Lineages

Divergence Time (Mybp),[a] Lineages Diverging	d	Changes/ Site/Year[b]	Unit Evolutionary Period[c] (Mybp)
1,400, Prokaryote-eukaryote	2.51	0.90×10^{-9}	5.58
300, Bird-mammal	0.71	1.18×10^{-9}	4.23
85, Bovine-human and bovine-rodent:			
C21b-C21h	0.24	1.41×10^{-9}	3.54
sccb-scch	0.34	2.00×10^{-9}	2.50
17αb-17αh	0.36	2.12×10^{-9}	2.36
C21m-(C21b and C21h)	0.34	2.00×10^{-9}	2.54
Average ± SD	0.32 ± 0.05	$1.88 \pm 0.28 \times 10^{-9}$	2.74 ± 0.47
75, Human-rabbit-rodent:			
LM2-(e and b)	0.27	1.80×10^{-9}	2.78
6-(Plm, c, and Plh)	0.29	1.93×10^{-9}	2.59
(4hum, LM4, and 4)-(d and P3) ~~LM4, and 4-P3~~	0.31	2.07×10^{-9}	2.42
3a-j rat	0.22	1.47×10^{-9}	3.41
jhum-(jrat and 3a)	0.25	1.67×10^{-9}	3.00
Plhum-(c and Plm)	0.24	1.60×10^{-9}	3.13
4hum-(LM4 and 4)	0.25	1.67×10^{-9}	3.00
HLp-p	0.33	2.20×10^{-9}	2.27
Average ± SD	0.27 ± 0.04	$1.80 \pm 0.23 \times 10^{-9}$	2.83 ± 0.36
20, Rat-mouse:			
d-P3	0.072	1.80×10^{-9}	2.78
c-Plm	0.073	1.83×10^{-9}	2.74

[a] Sources: prokaryote-eukaryote, Vidal 1984; bird-mammal, Romero-Herrera et al. 1978 and Dickerson 1971; bovine-human and bovine-rodent, Shoshani et al. 1985; human-rabbit-rodent, McLaughlin and Dayhoff 1972; and rat-mouse, Shoshani et al. 1985.

[b] Calculated from the equation $d = 2\lambda$, where λ is the rate of change (Tajima and Nei 1984).

[c] Time required for a 1% difference in sequence to occur, calculated by dividing the divergence time by the evolutionary distance × 100.

because it is found in plants that make cardiac glycosides. Therefore, plants diverged from animals before the first gene duplication shown in figure 2 (~1,360 Mybp). An even older connection exists between P-450 and cholesterol, because a P-450 enzyme is required to synthesize cholesterol, e.g., to remove the 14 methyl group from lanosterol. Apparently, P-450 proteins were adapted for use in the biosynthesis of cholesterol even before the divergence of prokaryotes from eukaryotes, because there are three types of bacteria known to contain—and presumably synthesize—cholesterol or partially demethylated lanosterol derivatives (Bloch 1983).

The only eukaryotic sequence of the 10 endogenous P-450s in figure 2 that does not act on cholesterol or its derivatives is P-450 LA_{ω}. This sequence shows some ambiguity in its branching on the tree, sometimes clustering with the $P\text{-}450_{scc}$ sequences. Some bacterial P-450s are ω-hydroxylases of alkanes (Sligar and Murray 1986). Furthermore, P-450 cholesterol 26 hydroxylase, a mitochondrial enzyme, also hydroxylates the end of a long lipid. These two proteins, one mitochondrial and one microsomal, may be related in their function. It is not possible to tell whether the LA_{ω} sequence is a bacterial relic or a more recent branch from the cholesterol-related P-450s.

As mentioned earlier, eukaryotic P-450s are found in two cellular locations, the mitochondria and the endoplasmic reticulum. These two types of P-450 proteins receive electrons from two very different sources. The mitochondrial P-450s are reduced by an iron-sulfur protein called adrenodoxin, which is first reduced by adrenodoxin reductase, a flavoprotein. The microsomal P-450s are reduced by a single protein called NADPH cytochrome P-450 reductase, which contains two different flavins. The prokaryotic $P\text{-}450_{cam}$, like the mitochondrial P-450s, receives electrons from an iron-sulfur protein that is first reduced by a flavoprotein. The similarity between the bacterial and mitochondrial systems suggests that the mitochondrial P-450s derived from a bacterial ancestor. At the same time, the difference between andrenodoxin and NADPH cytochrome P-450 reductase implies a different origin for the microsomal P-450s. One possible explanation is that the two different types of P-450s came from the two unique lineages that united to form the first eukaryote. More sequences of bacterial and mitochondrial P-450s and their reductase proteins will be needed to decide whether this speculation is correct.

Perhaps the most significant event in P-450 evolution was the gene duplication ~900 Mybp. One of the resultant lineages continued to function as an endogenous type of P-450. The two different genes in this branch code for enzymes in the biosynthetic pathway of cortisol. The other branch began a new function for cytochrome P-450, the detoxification of lipid-soluble foreign compounds. This later lineage then soon (~800 Mybp) divided into the two major P-450 families, the PB and MC families. Approximately 400 Mybp the PB family began diversifying. Since then, there have been eight gene duplications in this branch of the tree. The impetus for this striking diversification may have been the abrupt exposure of vertebrates to toxic chemicals in land plants. Plants preceded vertebrates on land by several million years, giving the plants ample time to develop chemicals never before experienced by the previously water-bound vertebrates. The evolution of detoxification enzymes such as the drug-metabolizing P-450 proteins would have been highly advantageous.

The date estimated here for the divergence of the PB and MC families is twice as far in the past as the date estimated by Sogowa et al. (1985). Similarly, estimates of the time of divergence between the PCN family (P-450s HLp and p) and the lineage leading to the drug-metabolizing P-450s are much lower than depicted in figure 2. Gonzalez et al. (1985) estimate that the PCN family divergence occurred at much

more than 200 Mybp. Hobbs et al. (1986) estimate that it occurred 600 Mybp. In figure 2, this divergence occurs ~1,100 Mybp. The reason for these differences lies in the fact that figure 2 has a break in the time scale. This break exists because it was not possible, when using a single scale, to reconcile the d values with estimated dates for either the divergence of reptiles and mammals (equivalent to the chicken-mammal divergence) or divergences between mammals. For example, if 17 Mybp is assumed for the rat-mouse divergence (as in Gonzalez et al. 1985), then the PB/MC split falls at ~370 Mybp but the chicken-mammal divergence falls at ~175 Mybp. Similarly, use of a rodent-human divergence time of 60–80 Mybp (as in Kawajiri et al. 1984; Gonzalez et al. 1985) will result in the PB/MC split occurring at ~340–450 Mybp but the bird-mammal divergence at ~160–210 Mybp. Both Dickerson (1971) and Romero-Herrera et al. (1978) give the bird-mammal divergence as ~300 Mybp. There are two ways to resolve this problem. Either the d value for the bird-mammal divergence is underestimated by ~80% or there must be a break in the time scale. An error of this magnitude is hardly believable for a d value of 0.7. The variation in 12 mammal-mammal divergences in table 4 is only 15%. This variation reflects the random nature of evolutionary change. Certainly, the error at the chicken-mammal divergence is not much greater than this. Consequently, the assumption that there is a break in the time scale of figure 2 seems justified.

The most likely reason for a break in the time scale is an acceleration in the evolutionary rate after the bird-mammal divergence. Table 4 indicates the rate of evolutionary change for 14 mammal-versus-mammal divergence events as being $1.83 \pm 0.23 \times 10^{-9}$ changes · site^{-1} · year^{-1}. The bird-mammal and the prokaryote-eukaryote divergences yield rates of 1.18×10^{-9} and 0.90×10^{-9} changes · site^{-1} · year^{-1}, respectively. Between the bird-mammal divergence and the mammalian radiation, there appears to have been an increase of 55%–100% in the cytochrome P-450 evolutionary rate. Even if the rates for the earlier divergences are in error, the magnitude of the error seems unlikely to be 55%–100%. The observation of an increase in the rate of evolution is not unique to cytochrome P-450. Such an increase in evolutionary rate in the mammalian lineage has been proposed during the same time period for cytochrome c, α and β hemoglobins, and myogloblin (Baba et al. 1981; Goodman 1981; Shoshani et al. 1985). The increase in rate has been linked to the rapid divergence of species during major adaptive radiations. Recently, Catzeflis et al. (1987) have given evidence that the evolutionary rate in rodents is 10 times the rate in hominoids and birds. The reason for the change in rate is unknown, but these authors propose that differences in DNA repair mechanisms may be involved. Britten (1986) finds a similar difference in rate (i.e., approximately fivefold) between rodents, hominoids, and birds. If the rate of evolution can vary 5–10-fold, then the need for multiple time calibrations is clearly indicated. Furthermore, Kunisawa et al. (1987) have found that a molecular clock based on vertebrate amino acid sequences of cytochrome c greatly underestimates divergence times among bacteria. These authors suggest that a multiclock model, one that posits a slower rate of evolution for earlier times periods, is needed to make reliable phylogenetic trees from amino acid data. A break in the time scale, identical to that seen in figure 2, is seen in figure 3 of Shoshani (1986).

On the basis of these considerations, the time scale in figure 2 seems justified. The earlier dates for divergence of the PB and MC families and other gene duplications are probably more accurate than dates derived by extrapolation from divergences among mammals. Certainly there is need for further revision of the time scale as more sequences of earlier-existing vertebrate and nonvertebrate P-450s become available.

Acknowledgments

Many thanks are due to Dr. J. C. Stephens for numerous discussions, a critic reading of the manuscript, and a copy of his UPGMA-tree computer program. D Naruya Saitou wrote the FORTRAN version of the modified Needleman-Wunsch a gorithm according to Gotoh (1982) and the neighbor-joining program. This study w supported by grant AU1067 from the Robert A. Welch Foundation and by gra CA37148 from the National Cancer Institute, Department of Health and Huma Welfare.

APPENDIX

Differences between Very Similar Sequences

f: 113LV 171NT 387ML 388LM 466-8 KGT (RRA) 492DV 494-5 GH(WP) 499G 560PA
PB1: 173-4 RN(TD) 326DN
3b: 115DY 281FG 418MS 422ST 517AT
e: 210DE 365LP 460RQ 562MK
b: 82QG 187IV 230ST 270SE 295KL 322SN 474LR
LM2: 116QE 120-1 FS(SF) 124-5 GK(--) 175-6 FG(GY) 182SG 246PK 278PS 376T 550-4 SPVPP (GNLSL)
c: 64IM 318EG 337EG
P1h: 444FL 527VI
LM4: 28CS 59WS 61WV 194LF 196GS 221SV 270GS 281LV 338-9 HS(SH) 509S 581RS 585QK
d: 44RK 105GE 158RH 305FS 469CR
P3: 449 IM
C21h: 520 RP
C21b: 28AS 489HY 526SC
HLp: 111TL 127RE 203AR 231TS 240IV 251FL 309SA 342-3 QK(H-)
sccb: 58DN
cam: 91-2 WT(--) 403EQ 526HS 580DN

The numbers refer to position in the present alignment. The letters are the amin acids found in that position. The first letter (or group of letters) was used in th alignment.

LITERATURE CITED

ADESNIK, M., and M. ATCHISON. 1985. Genes for cytochrome P-450 and their regulation. CR Crit. Rev. Biochem. **19**:247–305.

AFFOLTER, M., and A. ANDERSON. 1984. Segmental homologies in the coding and 3′ nor coding sequences of rat liver cytochrome P-450e and P-450b cDNAs and cytochrome I 450e-like genes. Biochem. Biophys. Res. Commun. **118**:655–662.

ATCHISON, M., and M. ADESNIK. 1983. A cytochrome P-450 multigene family: characterizatio of a gene activated by phenobarbital administration. J. Biol. Chem. **258**:11285–11295.

BABA, M. L., L. L. DARGA, M. GOODMAN, and J. CZELUSNIAK. 1981. Evolution of cytochrom c investigated by the maximum parsimony method. J. Mol. Evol. **17**:197–213.

BEAUNE, P. H., D. R. UMBENHAUER, R. W. BORK, R. S. LLOYD, and F. P. GUENGERICH. 198(Isolation and sequence determination of a cDNA clone related to human cytochrome P-45 nifedipine oxidase. Proc. Natl. Acad. Sci. USA **83**:8064–8068.

BLACK, S. D., and M. J. COON. 1986. Comparative structures of P-450 cytochromes. Pp. 161 216 *in* P. R. ORTIZ DE MONTELLANO, ed. Cytochrome P-450: structure, mechanism an biochemistry. Plenum, New York.

BLOCH, K. E. 1983. Sterol structure and membrane function. CRC Crit. Rev. Biochem. **14**:47–92.

BRITTEN, R. J. 1986. Rates of DNA sequence evolution differ between taxonomic groups. Science **231**:1393–1398.

CARROLL, M. C., R. D. CAMPBELL, and R. R. PORTER. 1985. Mapping of steroid 21-hydroxylase genes adjacent to complement component C4 genes in HLA, the major histocompatibility complex in man. Proc. Natl. Acad. Sci. USA **82**:521–525.

CATZEFLIS, F. M., F. H. SHELDON, J. E. AHLQUIST, and C. G. SIBLEY. 1987. DNA-DNA hybridization evidence of the rapid rate of muroid rodent DNA evolution. Mol. Biol. Evol. **4**:242–253.

CHAPLIN, D. D., L. J. GALBRAITH, J. G. SEIDMAN, P. C. WHITE, and K. L. PARKER. 1986. Nucleotide sequence analysis of murine 21-hydroxylase genes: mutations affecting gene expression. Proc. Natl. Acad. Sci. USA **83**:9601–9605.

CHASHCHIN, V. L., V. N. LAPKO, T. B. ADAMOVICH, A. G. LAPKO, N. S. KUPRINA, and A. A. AKHREM. 1986. Primary structure of the cholesterol side-chain cleavage P-450 from bovine adrenocortical mitochondria and some aspects of its functioning on a structural level. Biochim. Biophys. Acta **871**:217–223.

CHUNG, B.-C., K. J. MATTESON, and W. L. MILLER. 1986*a*. Structure of a bovine gene for P-450c21 (steroid 21-hydroxylase) defines a novel cytochrome P-450 gene family. Proc. Natl. Acad. Sci. USA **83**:4243–4247.

CHUNG, B.-C., K. J. MATTESON, R. VOUTILAINEN, T. K. MOHANDAS, and W. L. MILLER. 1986*b*. Human cholesterol side-chain cleavage enzyme P-450$_{scc}$: cDNA cloning, assignment of the gene to chromosome 15, and expression in placenta. Proc. Natl. Acad. Sci. USA **83**: 8962–8966.

CHUNG, B.-C., J. PICADO-LEONARD, M. HANIU, M. BIENKOWSKI, P. F. HALL, and W. L. MILLER. 1987. Cytochrome P-450c17 (steroid 17α-hydroxylase/17,20 lyase): cloning of human adrenal and testis cDNAs indicates the same gene is expressed in both tissues. Proc. Natl. Acad. Sci. USA **84**:407–411.

COBB, R. R., T. A. STOMING, and J. B. WHITNEY III. 1987. The aryl hydrocarbon hydroxylase (Ah) locus and a novel restriction fragment length polymorphism (RFLP) are located on mouse chromosome 12. Biochem. Genet. **25**:401–413.

DAYHOFF, M. O. 1979. Atlas of protein sequence and structure. Vol. **5**, suppl. 3. National Biomedical Research Foundation, Silver Spring, Md.

DICKERSON, R. E. 1971. The structure of cytochrome c and the rates of molecular evolution. J. Mol. Evol. **1**:26–45.

FRIEDBERG, T., D. J. WAXMAN, M. ATCHISON, A. KUMAR, T. HAAPARANTA, C. RAPHAEL, and M. ADESNIK. 1986. Isolation and characterization of cDNA clones for cytochromes P-450 immunochemically related to rat hepatic P-450 form PB-1. Biochemistry **25**:7975–7983.

FUJII-KURIYAMA, Y., Y. MIZUKAMI, K. KAWAJIRI, K. SOGAWA, and M. MURAMATSU. 1982. Primary structure of a cytochrome P-450: coding nucleotide sequence of phenobarbital-inducible cytochrome P-450 cDNA from rat liver. Proc. Natl. Acad. Sci. USA **79**:2793–2797.

FUJITA, V. S., S. D. BLACK, G. E. TARR, D. R. KOOP, and M. J. COON. 1984. On the amino acid sequence of cytochrome P-450 isozyme 4 from rabbit liver microsomes. Proc. Natl. Acad. Sci. USA **81**:4260–4264.

GONZALEZ, F. J., S. KIMURA, B.-J. SONG, J. PASTEWKA, H. V. GELBOIN, and J. P. HARDWICK. 1986. Sequence of two related P-450 mRNAs transcriptionally increased during rat development. J. Biol. Chem. **261**:10667–10672.

GONZALEZ, F. J., D. W. NEBERT, J. P. HARDWICK, and C. B. KASPER. 1985. Complete cDNA and protein sequence of a pregnenolone 16α-carbonitrile-induced cytochrome P-450. J. Biol. Chem. **260**:7435–7441.

GOODMAN, M. 1981. Globin evolution was apparently very rapid in early vertebrates: a reasonable case against the rate-constancy hypothesis. J. Mol. Evol. **17**:114–120.

GOTOH, O. 1982. An improved algorithm for matching biological sequences. J. Mol. Biol. **162**:705–708.

GOTOH, O., Y. TAGASHIRA, T. IIZUKA, and Y. FUJII-KURIYAMA. 1983. Structural characteristics of cytochrome P-450: possible location of the heme-binding cysteine in determined amino-acid sequences. J. Biochem. **93**:807–817.

HANIU, M., L. G. ARMES, K. T. YASUNOBU, B. A. SHASTRY, and I. C. GUNSALUS. 1982. Amino acid sequence of the *Pseudomonas putida* cytochrome P-450. J. Biol. Chem. **257**:12664–12671.

HANIU, M., D. E. RYAN, W. LEVIN, and J. E. SHIVELY. 1986. The primary structure of cytochrome P-450d purified from rat liver microsomes: prediction of helical regions and domain analysis. Arch. Biochem. Biophys. **244**:323–337.

HARDWICK, J. P., B.-J. SONG, E. HUBERMAN, and F. J. GONZALEZ. 1987. Isolation, complementary DNA sequence, and regulation of rat hepatic lauric acid ω-hydroxylase (cytochrome $P\text{-}450_{LA\omega}$): identification of a new cytochrome P-450 gene family. J. Biol. Chem. **262**:801–810.

HEINEMANN, F. S., and J. OZOLS. 1983. The complete amino acid sequence of rabbit phenobarbital-induced liver microsomal cytochrome P-450. J. Biol. Chem. **258**:4195–4201.

HIGASHI, Y., H. YOSHIOKA, M. YAMANE, O. GOTOH, and Y. FUJII-KURIYAMA. 1986. Complete nucleotide sequence of two steroid 21-hydroxylase genes tandemly arranged in human chromosome: a pseudogene and a genuine gene. Proc. Natl. Acad. Sci. USA **83**:2841–2845.

HILDEBRAND, C. E., F. J. GONZALEZ, O. W. MCBRIDE, and D. W. NEBERT. 1985. Assignment of the human 2,3,7,8-tetrachlorodibenzo-p-dioxin-inducible cytochrome P_1-450 gene to chromosome 15. Nucleic Acids Res. **13**:2009–2016.

HINES, R. N., J. B. LEVY, R. D. CONRAD, P. L. IVERSEN, M.-L. SHEN, A. M. RENLI, and E. BRESNICK. 1985. Gene structure and nucleotide sequence for rat cytochrome P-450c. Arch. Biochem. Biophys. **237**:465–476.

HOBBS, A. A., L. A. MATTSCHOSS, B. K. MAY, K. E. WILLIAMS, and W. H. ELLIOTT. 1986. The cDNA and protein sequence of phenobarbital-induced chicken cytochrome P-450. J. Biol. Chem. **261**:9444–9449.

JAISWAL, A. K., F. J. GONZALEZ, and D. W. NEBERT. 1985*a*. Human dioxin-inducible cytochrome P-450: complementary DNA and amino acid sequences. Science **228**:80–83.

———. 1985*b*. Human P_1-450 gene sequence and correlation of mRNA with genetic differences in benzo[α]pyrene metabolism. Nucleic Acids Res. **13**:4503–4520.

JOHN, M. E., T. OKAMURA, A. DEE, B. ADLER, M. C. JOHN, P. C. WHITE, E. R. SIMPSON, and M. R. WATERMAN. 1986. Bovine steroid 21-hydroxylase: regulation of biosynthesis. Biochemistry **25**:2846–2853.

KAWAJIRI, K., O. GOTOH, K. SOGAWA, Y. TAGASHIRA, M. MURAMATSU, and Y. FUJII-KURIYAMA. 1984. Coding nucleotide sequence of 3-methylcholanthrene-inducible cytochrome P-450d cDNA from rat liver. Proc. Natl. Acad. Sci. USA **81**:1649–1653.

KAWAJIRI, K., J. WATANABE, O. GOTOH, Y. TAGASHIRA, K. SOGAWA, and Y. FUJII-KURIYAMA. 1986. Structure and drug inducibility of the human cytochrome P-450c gene. Eur. J. Biochem. **159**:219–225.

KHANI, S. C., P. G. ZAPHIROPOULOS, V. S. FUJITA, T. D. PORTER, D. R. KOOP, and M. J. COON. 1987. cDNA and derived amino acid sequence of ethanol-inducible rabbit liver cytochrome P-450 isoenzyme 3a ($P\text{-}450_{ALC}$). Proc. Natl. Acad. Sci. USA **84**:638–642.

KIMURA, S., F. J. GONZALEZ, and D. W. NEBERT. 1984*a*. Mouse cytochrome P_3-450: complete cDNA and amino acid sequence. Nucleic Acids Res. **12**:2917–2928.

———. 1984*b*. The murine Ah locus, comparison of the complete cytochrome P_1-450 and P_3-450 cDNA nucleotide and amino acid sequences. J. Biol. Chem. **259**:10705–10713.

KIMURA, S., and D. W. NEBERT. 1986. cDNA and complete amino acid sequence of mouse $P_2$450: allelic variant of mouse $P_3$450 gene. Nucleic Acids Res. **14**:6765–6766.

KUNISAWA, T., K. HORIMOTO, and J. OTSUKA. 1987. Accumulation pattern of amino acid substitutions in protein evolution. J. Mol. Evol. **24**:357–365.

LEGRAVEREND, C., S. O. KÄRENLAMPI, S. W. BIGELOW, P. A. LALLEY, C. A. KOZAK, J. E. WOMACK, and D. W. NEBERT. 1984. Aryl hydrocarbon hydroxylase induction by benzo[a]anthracene: regulatory gene localized to the distal portion of mouse chromosome 17. Genetics **107**:447–461.

LEIGHTON, J. K., B. A. DEBRUNNER-VOSSBRINK, and B. KEMPER. 1984. Isolation and sequence analysis of three cloned cDNAs for rabbit liver proteins that are related to rabbit cytochrome P-450 (Form 2), the major phenobarbital-inducible form. Biochemistry **23**:204–210.

MCLAUGHLIN, P. J., and M. O. DAYHOFF. 1972. Evolution of species and proteins: a time scale. Pp. 47–52 *in* M. O. DAYHOFF, ed. Atlas of protein sequence and structure. Vol. **5**. National Biomedical Research Foundation, Silver Spring, Md.

MATTESON, K. J., J. PICADO-LEONARD, B.-C. CHUNG, T. K. MOHANDAS, and W. L. MILLER. 1986. Assignment of the gene for P-450c17 (steroid 17α-hydroxylase/17,20 lyase) to human chromosome 10. J. Clin Endocrinol. Metab. **63**:789–791.

MIYATA, T., H. HAYASHIDA, R. KIKUNO, M. HASEGAWA, M. KOBAYASHI, and K. KOIKE. 1982. Molecular clock of silent substitution: at least six-fold preponderance of silent changes in mitochondrial genes over those in nuclear genes. J. Mol. Evol. **19**:28–35.

MIZUKAMI, Y., Y. FUJII-KURIYAMA, and M. MURAMATSU. 1983*a*. Multiplicity of deoxyribonucleic acid sequences with homology to a cloned complementary deoxyribonucleic acid coding for rat phenobarbital-inducible cytochrome P-450. Biochemistry **22**:1223–1229.

MIZUKAMI, Y., K. SOGAWA, Y. SUWA, M. MURAMATSU, and Y. FUJII-KURIYAMA. 1983*b*. Gene structure of a phenobarbital-inducible cytochrome P-450 in rat liver. Proc. Natl. Acad. Sci. USA **80**:3958–3962.

MOLOWA, D. T., E. G. SCHUETZ, S. A. WRIGHTON, P. B. WATKINS, P. KREMERS, G. MENDEZ-PICON, G. A. PARKER, and P. S. GUZELIAN. 1986. Complete cDNA sequence of a cytochrome P-450 inducible by glucocorticoids in human liver. Proc. Natl. Acad. Sci. USA **83**: 5311–5315.

MOROHASHI, K., Y. FUJII-KURIYAMA, Y. OKADA, K. SOGAWA, T. HIROSE, S. INAYAMA, and T. OMURA. 1984. Molecular cloning and nucleotide sequence of cDNA for mRNA of mitochondrial cytochrome P-450$_{scc}$ of bovine adrenal cortex. Proc. Natl. Acad. Sci. USA **81**:4647–4651.

NEBERT, D. W., and F. J. GONZALEZ. 1985. Cytochrome P-450 gene expression and regulation. Trends Pharmacol. Sci. **6**:160–164.

NEEDLEMAN, S. B., and C. D. WUNSCH. 1970. A general method applicable to the search for similarities in the amino acid sequence of two proteins. J. Mol. Biol. **48**:443–453.

OKINO, S. T., L. C. QUATTROCHI, H. J. BARNES, S. OSANTO, K. J. GRIFFIN, E. F. JOHNSON, and R. H. TUKEY. 1985. Cloning and characterization of cDNAs encoding 2,3,7,8-tetrachlorodibenzo-p-dioxin-inducible rabbit mRNAs for cytochrome P-450 isozymes 4 and 6. Proc. Natl. Acad. Sci. USA **82**:5310–5314.

OZOLS, J. 1986. Complete amino acid sequence of a cytochrome P-450 isolated from β-naphthoflavone-induced rabbit liver microsomes. J. Biol. Chem. **261**:3965–3979.

OZOLS, J., F. S. HEINEMANN, and E. F. JOHNSON. 1985. The complete amino acid sequence of a constitutive form of liver microsomal cytochrome P-450. J. Biol. Chem. **260**:5427–5434.

PHILLIPS, I. R., E. A. SHEPHARD, A. ASHWORTH, and B. R. RABIN. 1985*a*. Isolation and sequence of a human cytochrome P-450 cDNA clone. Proc. Natl. Acad. Sci. USA **82**:983–987.

PHILLIPS, I. R., E. A. SHEPHARD, S. POVEY, M. B. DAVIS, G. KELSEY, M. MONTEIRO, L. F. WEST, and J. COWELL. 1985*b*. A cytochrome P-450 gene family mapped to human chromosome 19. Ann. Hum. Genet. **49**:267–274.

POULOS, T. L., B. C. FINZEL, I. C. GUNSALUS, G. C. WAGNER, and J. KRAUT. 1985. The 2.6Å crystal structure of *Pseudomonas putida* cytochrome P-450. J. Biol. Chem. **260**:16122–16130.

QUATTROCHI, L. C., U. R. PENDURTHI, S. T. OKINO, C. POTENZA, and R. H. TUKEY. 1986.

Human cytochrome P-450 4 mRNA and gene: part of a multigene family that contains Alu sequences in its mRNA. Proc. Natl. Acad. Sci. USA **83**:6731–6735.

ROMERO-HERRERA, A. E., H. LEHMANN, K. A. JOSEY, and A. E. FRIDAY. 1978. On the evolution of myoglobin. Philos. Trans. R. Soc. Lond. [Biol.] **283**:61–163.

SAITOU, N., and M. NEI. 1987. The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol. Biol. Evol. **4**:406–425.

SHOSHANI, J. 1986. Mammalian phylogeny: comparison of morphological and molecular results. Mol. Biol. Evol. **3**:222–242.

SHOSHANI, J., M. GOODMAN, J. CZELUSNIAK, and G. BRAUNITZER. 1985. A phylogeny of rodentia and other eutherian orders: parsimony analysis utilizing amino acid sequences of alpha and beta hemoglobin chains. Pp. 191–210 *in* W. P. LUCKETT and J.-L. HARTENBERGER, eds. Evolutionary relationships among rodents. NATO ASI series A: Life sciences. Vol. **92.** Plenum, New York.

SIMMONS, D. L., and C. B. KASPER. 1983. Genetic polymorphisms for a phenobarbital-inducible cytochrome P-450 map to the Coh locus in mice. J. Biol. Chem. **258**:9585–9588.

SIMMONS, D. L., P. A. LALLEY, and C. B. KASPER. 1985. Chromosomal assignments of genes coding for components of the mixed-function oxidase system in mice. J. Biol. Chem. **258**:515–521.

SLIGAR, S. G., and R. I. MURRAY. 1986. Cytochrome P-450$_{cam}$ and other bacterial P-450 enzymes. Pp. 429–503 *in* P. R. ORTIZ DE MONTELLANO, ed. Cytochrome P-450 structure, mechanism and biochemistry. Plenum, New York.

SNEATH, P. H. A., and R. R. SOKAL. 1973. Numerical taxonomy. W. H. Freeman, San Francisco.

SOGAWA, K., O. GOTOH, K. KAWAJIRI, and Y. FUJII-KURIYAMA. 1984. Distinct organization of methylcholanthrene- and phenobarbital-inducible cytochrome P-450 genes in the rat. Proc. Natl. Acad. Sci USA **81**:5066–5070.

SOGAWA, K., O. GOTOH, K. KAWAJIRI, T. HARADA, and Y. FUJII-KURIYAMA. 1985. Complete nucleotide sequence of a methylcholanthrene-inducible cytochrome P-450 (P-450d) gene in the rat. J. Biol. Chem. **260**:5026–5032.

SONG, B.-J., H. V. GELBOIN, S.-S. PARK, S. C. YANG, and F. J. GONZALEZ. 1986. Complementary DNA and protein sequences of ethanol-inducible rat and human cytochrome P-450s. J. Biol. Chem. **261**:16689–16697.

SUWA, Y., Y. MIZUKAMI, K. SOGAWA, and Y. FUJII-KURIYAMA. 1985. Gene structure of a major form of phenobarbital-inducible cytochrome P-450 in rat liver. J. Biol. Chem. **260**:7980–7984.

TAJIMA, F., and M. NEI. 1984. Estimation of evolutionary distance between nucleotide sequences. Mol. Biol. Evol. **1**:269–285.

TARR, G. E., S. D. BLACK, V. S. FUJITA, and M. J. COON. 1983. Complete amino acid sequence and predicted membrane topology of phenobarbital-induced cytochrome P-450 (isozyme 2) from rabbit liver microsomes. Proc. Natl. Acad. Sci. USA **80**:6552–6556.

TUKEY, R. H., P. A. LALLEY, and D. W. NEBERT. 1984. Localization of cytochrome P_1-450 and P_3-450 genes to mouse chromosome 9. Proc. Natl. Acad. Sci. USA **81**:3163–3166.

TUKEY, R. H., S. OKINO, H. BARNES, K. J. GRIFFIN, and E. F. JOHNSON. 1985. Multiple gene-like sequences related to the rabbit hepatic progesterone 21-hydroxylase cytochrome P-450 1. J. Biol. Chem. **260**:13347–13354.

UNGER, B. P., I. C. GUNSALUS, and S. G. SLIGAR. 1986. Nucleotide sequence of the *Pseudomonas putida* cytochrome P-450$_{cam}$ gene and its expression in *Escherichia coli*. J. Biol. Chem. **261**:1158–1163.

VIDAL, G. 1984. The oldest eukaryotic cells. Sci. Am. **250**:48–57.

WHITE, P. C., X. CHAPLIN, J. H. WEIS, B. DUPONT, M. I. NEW, and J. G. SEIDRIAN. 1984. Two steroid hydroxylase genes are located in the murine S-region. Nature **312**:465–467.

WHITE, R. E., and COON, M. J. 1980. Oxygen activation by cytochrome P-450. Annu. Rev. Biochem. **49**:315–356.

WHITE, P. C., M. I. NEW, and B. DUPONT. 1986. Structure of human steroid 21-hydroxylase genes. Proc. Natl. Acad. Sci. USA **83**:5111–5115.

YABUSAKI, Y., M. SHIMIZU, H. MURAKAMI, K. NAKAMURA, K. OEDA, and H. OHKAWA. 1984. Nucleotide sequence of a full-length cDNA coding for 3-methylcholanthrene-induced rat liver cytochrome P-450MC. Nucleic Acids Res. **12**:2929–2938.

YOSHIOKA, H., K.-I. MOROHASHI, K. SOGAWA, M. YAMANE, S. KOMINAMI, S. TAKEMORI, Y. OKADA, T. OMURA, and Y. FUJII-KURIYAMA. 1986. Structural analysis of cloned cDNA for mRNA of microsomal cytochrome P-450(C21) which catalyzes steroid 21-hydroxylation in bovine adrenal cortex. J. Biol. Chem. **261**:4106–4109.

YUAN, P.-M., D. E. RYAN, W. LEVIN, and J. E. SHIVELY. 1983. Identification and localization of amino acid substitutions between two phenobarbital-inducible rat hepatic microsomal cytochromes P-450 by micro sequence analyses. Proc. Natl. Acad. Sci. USA **80**:1169–1173.

ZUBER, M. X., M. E. JOHN, T. OKAMURA, E. R. SIMPSON, and M. R. WATERMAN. 1986. Bovine adrenocortical cytochrome P-450 17α regulation of gene expression by ACTH and elucidation of primary sequence. J. Biol. Chem. **261**: 2475–2482.

MASATOSHI NEI, reviewing editor

Received December 29, 1986; revision received July 16, 1987.