Mouse cytochrome P₃-450: complete cDNA and amino acid sequence

Shioko Kimura, Frank J. Gonzalez and Daniel W. Nebert

Laboratory of Developmental Pharmacology, National Institute of Child Health and Human Development, National Institutes of Health, Bethesda, MD 20205, USA

Received 3 January 1984; Revised and Accepted 22 February 1984

ABSTRACT

A full-length cDNA clone (1,894 nucleotides) of mouse cytochrome P₃-450 was isolated with the Okayama-Berg vector and sequenced. An open reading frame spanned positions 61 to 1602. The first 25, and three of the last five, amino acids of P₃-450 are identical to those found in the amino- and carboxy-terminus, respectively, of the rat P-450d protein. Mouse P₃-450 protein has 513 residues, and a molecular weight of 58,223 with six cysteine residues. P₃-450 nucleotides 305 to 352 exhibit 74% homology, and nucleotides 1068 to 1260, 69% homology, with portions of rat P-450b exons 2 and 7, respectively. P₃-450 shows 62% homology in the so-called "highly conserved region" of 39 nucleotides in the rat P-450b and P-450e and the mouse P-450b. These results indicate that P₃-450, P-450b and P-450e arose from a common ancestral gene. Cysteinyl peptide-coding regions were examined: P₃-450 nucleotides 1405 to 1464 exhibit 61% homology, and nucleotides 502 to 552 exhibit 37% homology, when compared with their corresponding regions in the rat P-450b gene. These data support the likelihood that cysteine 456 is the thiolate ligand to the heme iron in the P₃-450 enzyme active-site.

INTRODUCTION

A large portion of drug-metabolizing enzyme activity arises from the membrane-bound multicomponent cytochrome P-450 system (1, 2). Endogenous substrates of P-450 include steroids, fatty acids, biogenic amines, prostaglandins and leukotrienes. Foreign substrates include almost all drugs, chemical carcinogens, and other environmental pollutants numbering in the tens of thousands; hence, the recommended nomenclature for P-450-mediated catalytic activity is "multisubstrate monooxygenase" (3). ["P-450" designates any or all forms of the enzyme. Mouse "P₁-450" and "P₃-450" proteins (4) are defined as those forms of 3-methylcholanthrene-induced P-450 in C57BL/6N liver having the highest turnover number for induced aryl hydrocarbon (benzo[a]pyrene) hydroxylase and acetanilide 4-hydroxylase activity, respectively. Basal and induced P₃-450 protein concentrations are about five times greater than basal and induced P₁-450

protein concentrations, respectively (5). Mouse "P₂-450" is defined as that form of isosafrole-induced P-450 in DBA/2N liver having the highest turn-over number for isosafrole metabolism (6). The size of all three of these hemoproteins is 55,000 daltons, as estimated by sodium dodecyl sulfate polyacrylamide gel electrophoresis (4-6). Mouse P₁-450 and P₃-450 are believed to correspond to rat P-450c and P-450d, respectively, and to rabbit form 6 and form 4, respectively (7). P₂-450 in DBA/2N mice (5) may represent a protein polymorphism with P₃-450 in C57BL/6N mice.]

Purification of mammalian P-450 proteins ($M_r \approx 48,000$ to 60,000) has been difficult, and only very recently has the first one--rabbit liver form 2 P-450--been completely sequenced (8, 9). The recent cloning of near-full length cDNA's for rat P-450b (10) and P-450e (11) has allowed the first comparison for both nucleotide and amino acid sequence of two closely related phenobarbital-inducible forms. The predicted rat P-450b and P-450e protein sequences have been compared with the published rabbit form 2 P-450 (8, 9) and *Pseudomonas putida* P-450_{cam} (12) protein sequences.

The "3-methylcholanthrene-inducible P-450 gene family"--mouse P₁-450, P₂-450 and P₃-450 (3-6)--appears to be quite distinct from the "phenobarbital-inducible P-450 gene family." Phenobarbital-induced mouse liver mRNA does not hybridize on Northern blots (13) or by R₀t analysis (14) to mouse P₁-450 cDNA, nor does 3-methylcholanthrene-induced rat liver mRNA hybridize measurably to rat P-450b (15, 16), rat P-450e (17), or mouse P-450b (18) cDNA clones. The 3-methylcholanthrene-inducible P-450 gene family is regulated by the cytosolic Ah receptor (19). The receptor is known to bind avidly ($K_d \approx 1$ nM) to such polycyclic aromatic inducers as 3-methylcholanthrene, 2,3,7,8-tetrachlorodibenzo-p-dioxin and benzo[a]pyrene, and the inducer-receptor complex undergoes a temperature-dependent nuclear translocation (20) to stimulate P₁-450 mRNA synthesis (21) that includes an intranuclear large-molecular-weight pre-mRNA (13). Transcription of the P₁-450 and P₃-450 genes is activated, and these mRNA levels are induced 15- to 20-fold within 12 h after a single intraperitoneal dose of 3-methylcholanthrene (22). The end result includes enhanced translation by polysomes and insertion of these newly synthesized proteins into the microsomal membrane, as measured by immunoprecipitable radioactivity (4-6) and marked increases in the appropriate catalytic activity (4, 5).

The number of known P-450 inducers includes more than five dozen (1, 2, 23) and is most likely much larger. The number of genetically distinct P-450 proteins is unknown, but estimates range from at least 20 (2) to

hundreds or thousands (24). The mechanisms of diversity by which the P-450 gene superfamily responds to thousands of adverse environmental chemicals are not yet known (25). By means of cloning and sequencing studies, however, it soon should be possible to understand the response of this gene superfamily to multiple chemical stimuli and to follow the evolution of these genes from Pseudomonadeae to mammals. In this report we show the complete cDNA sequence for mouse P₃-450 and compare the nucleotide and predicted amino acid sequences with published rat and rabbit data.

EXPERIMENTAL PROCEDURES

Materials

Reverse transcriptase was purchased from Life Sciences, Inc. (St. Petersburg, FL); restriction endonucleases, DNA polymerase, and DNA polymerase-large fragment from New England Biolabs (Beverly, MA); RNase H, ultrapure urea and phenol from Bethesda Research Laboratories (Rockville, MD); DNase I from Worthington Biochemicals (Freehold, NJ); and T4 DNA ligase from New England Nuclear Corporation (Boston, MA). DNA sequencing reagents and M13 mp10 were obtained from PL Biochemicals, Inc. (Milwaukee, WI). Alpha-[³²P]dATP (400 Ci/mol) was bought from Amersham (Chicago, IL).

Cloning of Full-Length cDNA Complementary to P₃-450 mRNA

We isolated total hepatic polysomes as previously described (15, 26) from 100 male weanling C57BL/6N mice 16 h after they had received a single intraperitoneal injection of 3-methylcholanthrene (200 mg/kg) dissolved in corn oil (25 ml/kg). Polysomes synthesizing P₃-450 were immunoabsorbed by affinity-purified goat anti-(P₃-450) immunoglobulin G as outlined before (15, 26). Poly(A⁺)-enriched RNA was then purified by oligo(dT)-cellulose chromatography (27). Approximately 15 µg of poly(A⁺)-enriched P₃-450 RNA was isolated from 100 g of liver. The mRNA was used to construct a cDNA clone bank by means of the procedure of Okayama and Berg (28, 29) and the parent plasmids pCDV1 and pL1 (29). Colonies containing inserts complementary to P₃-450 mRNA were identified by differential hybridization with [³²P]cDNA synthesized from immunoenriched versus nonenriched poly(A⁺) RNA (15, 26). The colonies containing the largest inserts were identified by plasmid mini-lysis preparations (30, 31) and restriction enzyme mapping. Forty P₃-450 clones were isolated from about 1000 colonies, and all of them contained approximately 1900-bp inserts. One clone, designated pP₃450FL, was processed for sequencing.

Sequencing of P₃-450 cDNA

DNA sequencing was carried out by means of the M13 cloning protocols of Messing *et al.* (32) and the dideoxy technique of Sanger *et al.* (33). pP₃450FL was digested by BamHI, which cleaves on both sides of the insert (29) but does not cleave internally. We purified the fragment by electrophoresis and electroelution, using a 0.7% agarose gel containing ethidium bromide (0.5 µg/ml). Following electroelution, the DNA was extracted successively with phenol and chloroform and finally precipitated with ethanol. A random DNase I shotgun library was prepared by a slight modification of the described protocol (34). DNA was partially digested with DNase I and size-fractionated on a 4% polyacrylamide gel (31, 34). Fragments of 600 to 800 bp were electroeluted, extracted and ethanol-precipitated, as described above, before treatment with the large fragment of DNA polymerase (34). Blunt-end fragments were ligated into either the SmaI or HincII site of M13 mpl0. Transformation of E.coli JM103 was carried out as described (31, 35), and single-stranded DNA was isolated by standard protocols (32). Dideoxy sequencing reactions were performed as detailed in the PL Biochemicals, Inc. protocols accompanying the reaction kits. The nucleotide products were displayed electrophoretically on 6% acrylamide-urea gels (36).

Computer Analysis of DNA and Protein Sequences

Single-strand sequences from both strands were aligned with the aid of the Staden program (37). Nucleotide sequences were analyzed with the SEQ program (38). Protein data were examined by means of the Dayhoff program (39).

RESULTS AND DISCUSSION

Restriction Map of P₃-450 cDNA

The total length of the cDNA insert in the Okayama-Berg vector was estimated to be 1900 bp. P₃-450 mRNA has been determined to be approximately 2100 nucleotides, including the poly(A⁺) tail (40). The sites for HpaII, PstI, HindIII, TaqI, AvaI, PvuII and XbaI are shown in Fig. 1. Restriction enzymes that did not cleave include EcoRI, BamHI, ClaI, SacI, KpnI and SalI.

cDNA and Amino Acid Sequence

Each nucleotide position (Fig. 2) was determined between two and eight times, with an average of 4.4 times. At those positions determined only twice, both strands were sequenced. Whereas reading in the first frame

Fig. 1. Partial restriction map for mouse P₃-450 cDNA.

from position 61 to 1602 gave a continuous translation of 513 amino acids, readings in the second and third frames exhibited multiple termination codons. Although the codon usage frequency phased from position 1 was consistent with the average codon usage determined for 84 other mouse cDNA's, the codon usage frequency phased from either position 2 or position 3 showed several striking exceptions to that found in normal mouse cDNA. Moreover, the first 25 amino acids correspond exactly to the NH₂-terminal sequence of rat P-450d protein (41). Three of the five predicted COOH-terminal amino acids for mouse P₃-450 (Fig. 2) correspond to those of the rat P-450d protein, whereas no such correspondence exists with the COOH-terminal amino acids for rat P-450c protein (41).

Comparison with Other P-450 Nucleotide Sequences

A nucleotide search and comparison with sequences in the published literature produced several interesting findings. Mouse P₃-450 nucleotides 305 to 352 exhibit 74% homology with a portion of rat P-450b exon 2 (16). Mouse P₃-450 nucleotides 1068 to 1260 show 69% homology with most of exon 7 in rat P-450b (10, 16). A 39-nucleotide sequence associated with the "highly conserved tridecapeptide" (9-11, 16, 18, 42) has been described in every P-450 protein or cDNA sequence studied to date and is known to be located in the center of exon 7 in the P-450b (10) and P-450e (11) genes. P₃-450 positions 1147 to 1185 represent this conserved region (Fig. 3). Whereas mouse P-450b (18) is 97% homologous with rat P-450b in the conserved 39-nucleotide region, mouse P₃-450 is 62% homologous with rat P-450b in this region. The P₃-450 protein is 54% homologous with the P-450b protein for these 13 residues (Fig. 3). Based on rat and rabbit amino acid sequence data, the unit evolutionary period (UEP; ref. 43) of P-450b-like genes has been estimated to be 2.1 (44). The data in Fig. 3 therefore suggest that the mouse P₃-450 conserved region has diverged from the mouse P-450b conserved region about 100 million years ago. Since the

30 60 90
GGT CCTGGA CTGA CTCCCA CAA CTCTG CCACT CTCCAG CCCCTG CCCTT CAGTGGTA CAG ATG GGG TTC TCC CAG TAC ATC TCC TTA GCC
MET Ala Phe Ser Gln Tyr Ile Ser Leu Ala

120 150
CCA GAG CTG CTA CTG GCC ACT GCC ATC TTC TGT TTA GTG TTC TGG ATG GTC CAG AGC CTC AAG GAC CCA GGT TCC
Pro Glu Leu Leu Leu Ala Thr Ala Ile Phe Cys Leu Val Phe Trp Met Val Gln Ser Leu Lys Asp Pro Gly Ser

180 210 240
CAA AGG CCT GAA GAA TCC ACC GGG ACC CTG GGG CTT CCC TTC ATT GGG CAC ATG CTG ACT GTG GGG AAG AAC CCA
Gln Arg Pro Glu Glu Ser Thr Arg Thr Leu Gly Leu Pro Phe Ile Gly His Met Leu Thr Val Gly Lys Asn Pro

270 300
CAC CTG TCA CTG ACA GGG CTG AGT CAG CAG TAT GGG GAC GTG CTG CAG ATC GGC ATC GGC TCC ACT CCT GTG GTG
His Leu Ser Leu Thr Arg Leu Ser Gln Gln Tyr Gly Asp Val Leu Gln Ile Arg Ile Gly Ser Thr Pro Val Val

330 360 390
GTG CTG AGC GGC CTG AAC ACC ATC AAG CAG GCC CTG GTG AGG CAG GGA GAT GAC TTC AAG GGC CGA CCA GAC CTC
Val Leu Ser Gly Leu Asn Thr Ile Lys Gln Ala Leu Val Arg Gln Gly Asp Asp Phe Lys Gly Arg Pro Asp Leu

420 450
TAC AGC TTC ACA CTT ATC ACT AAC GGC AAG AGC ATG ACT TTC AAC CCA GAC TCT GGA CCC GTG TGG GCT GCC GGC
Tyr Ser Phe Thr Leu Ile Thr Asn Gly Lys Ser Met Thr Phe Asn Pro Asp Ser Gly Pro Val Trp Ala Ala Arg

480 510 540
GGG GGC CTG GCC CAG GAT GCC CTG AAG AGC TTC TCC ATA GCC TGG GAC CGG ACG TCA GCA TCC TCT TGC TAT TTG
Arg Arg Leu Ala Gln Asp Ala Leu Lys Ser Phe Ser Ile Ala Ser Asp Pro Thr Ser Ala Ser Ser Cys Tyr Leu

570 600
GAG GAG CAC GTG AGC AAG GAG GCT AAC CAT CTC GTC AGC AAG CTT CAG AAG GCG ATG GCA GAG GTT GGC CAC TTC
Glu Glu His Val Ser Lys Glu Ala Asn His Leu Val Ser Lys Leu Gln Lys Ala Met Ala Glu Val Gly His Phe

630 660 690
GAA CCA GTC AGC CAG GTG GTG GAA TCG GTG GCT AAC GTC ATT GGT GCC ATG TGC TTT GGG AAG AAC TTC CCC GGG
Glu Pro Val Ser Gln Val Val Glu Ser Val Ala Asn Val Ile Gly Ala Met Cys Phe Gly Lys Asn Phe Pro Arg

720 750
AAG AGC GAG GAG ATG CTG AAC ATC GTG AAT AAC AGC AAG GAC TTT GTG GAG AAT GTC ACC TCA GGG AAT GCA GTG
Lys Ser Glu Glu Met Leu Asn Ile Val Asn Asn Ser Lys Asp Phe Val Glu Asn Val Thr Ser Gly Asn Ala Val

780 810 840
GAC TTC TTC CGG GTC CTG GGC TAC CTG CCC AAC CGG GCC CTC AAG AGG TTT AAG ACC TTC AAT GAT AAC TTC GTG
Asp Phe Phe Pro Val Leu Arg Tyr Leu Pro Asn Pro Ala Leu Lys Arg Phe Lys Thr Phe Asn Asp Asn Phe Val

870 900
CTG TTT CTG CAG AAA ACT GTC CAG GAG CAC TAC CAA GAC TTC AAC AAG AAC AGT ATC CAA GAC ATC ACA AGT GCC
Leu Phe Leu Gln Lys Thr Val Gln Glu His Tyr Gln Asp Phe Asn Lys Asn Ser Ile Gln Asp Ile Thr Ser Ala

930 960 990
CTG TTC AAG CAC AGC GAG AAC TAC AAA GAC AAT GGC GGT CTC ATC CCC GAG GAG AAG ATT GTC AAC ATT GTC AAT
Leu Phe Lys His Ser Glu Asn Tyr Lys Asp Asn Gly Gly Leu Ile Pro Glu Glu Lys Ile Val Asn Ile Val Asn

1020 1050
GAC ATC TTT GGA GCT GGC TTT GAC ACA GTC ACC ACA GCC ATC ACC TGG AGC ATT TTG CTA CTT GTG ACA TGG CCT
Asp Ile Phe Gly Ala Gly Phe Asp Thr Val Thr Thr Ala Ile Thr Trp Ser Ile Leu Leu Leu Val Thr Trp Pro

1080 1110 1140
AAC GTG CAG AGG AAG ATC CAT GAG GAG CTG GAC ACG GTG GTT GGC AGG GAT GCG CAA CCA GGG CTT TCT GAC GGT
Asn Val Gln Arg Lys Ile His Glu Glu Leu Asp Thr Val Val Gly Arg Asp Arg Gln Pro Arg Leu Ser Asp Arg

1170 1200
CCC CAG CTG CCA TAT CTA GAG GCC TTC ATC CTG GAG ATC TAC CGA TAC ACA TCC TTT GTC CCC TTC ACC ATC CCC
Pro Gln Leu Pro Tyr Leu Glu Ala Phe Ile Leu Glu Ile Tyr Arg Tyr Thr Ser Phe Val Pro Phe Thr Ile Pro

1230 1260 1290
CAC AGC ACA ACG AGG GAC ACC TCA CTG AAT GGC TTC CAC ATT CCC AAG GAG GGC TGT ATC TAC ATA AAC CAG TGG
His Ser Thr Thr Arg Asp Thr Ser Leu Asn Gly Phe His Ile Pro Lys Glu Arg Cys Ile Tyr Ile Asn Gln Trp

1320 1350
CAG GTC AAC CAT GAT GAG AAG CAG TGG AAA GAC CCC TTT GTG TTC GGC CCA GAG GGG TTT GTT ACC AAT AAC AAC
Gln Val Asn His Asp Glu Lys Gln Trp Lys Asp Pro Phe Val Phe Arg Pro Glu Arg Phe Leu Thr Asn Asn Asn

```

1380                                1410                                1440
TGG GCC ATC GAC AAG ACC CAG AGC GAG AAG GTG ATG CTC TTC GGC TTG GGA AAG CGC CGG TGC ATT GGG GAG ATC
Ser Ala Ile Asp Lys Thr Gln Ser Glu Lys Val Met Leu Phe Gly Leu Gly Lys Arg Arg Gln Ile Gly Glu Ile

                                1470                                1500
COG GCC AAG TGG CAA GTC TTC CTC TTC TTA GCC ATC CTG CTG CAG CAT CTG GAG TTT AGT GTG CCA CGG GGT GTG
Pro Ala Lys Trp Glu Val Phe Leu Phe Leu Ala Ile Leu Leu Gln His Leu Glu Phe Ser Val Pro Pro Gly Val

1530                                1560                                1590
AAG GTG GAC CTG ACA CCC AAC TAT GGG TTG ACC ATG AAG CCC GGG ACC TGT GAA CAC GTC CAG GCA TGG CCA CGC
Lys Val Asp Leu Thr Pro Asn Tyr Gly Leu Thr Met Lys Pro Gly Thr Gln Glu His Val Gln Ala Trp Pro Arg

                                1620                                1650                                1680
TTT TCC AAG TGA AGATTGTGAGGCATGGTGGGGCGGTCA CCCTTGTTTCTTTTCTTTTAAAAAAGAGCTTTTCTTTTCTTTT
Phe Ser Lys

1710                                1740                                1770
GAGAGATA CAATTCTTTCCCGATTAAATTCATCTCCAAGCAATTTTCAATAGTGTCTATCATGTTTCA CCCGATAA CCGATA CTCAATAGGA CTTATGA

1800                                1830                                1860
TTTAAGATTCTCTCTA CCGCTGCTTGCTCTG CCGCA CCTCATGCTAATCTAGTTTTTGA CTCAATAGATTGCTCTGCTGCTCTCATATAAATCGA

1890
ATGAATTATGA(A)>100

```

Fig. 2. Nucleotide sequence and predicted amino acid sequence of mouse P₃-450. The underlined region denotes the so-called "highly conserved region" of 39 nucleotides reported for other P-450 cDNA clones. The overlined region represents the putative poly(A⁺) addition signal. The entire length of the poly(A⁺) tract is believed to be between 100 and 200 bp. The (A)₁₆ and (T)₁₂ tracts between nucleotides 1650 and 1690 is an interesting region of potential self-complementarity and may be important in conferring secondary structure to the mRNA.

entire P-450b-like proteins and P₃-450 protein display less than 25% homology, it is likely that these two classes of genes separated more than 150 million years ago.

Comparison with Other P-450 Protein Sequences

The protein predicted from the cDNA sequence has 513 residues and a mass of 58,223 daltons (Table 1). The six cysteines of mouse P₃-450 contrast with five estimated for rat P-450d protein (41). Most of the P₃-450 amino acids are one or more greater than those estimated for rat P-450d protein (41); the largest differences include valine [40 for P₃-450 and 32 for P-450d] and tryptophan [eight for P₃-450 and four for P-450d].

There are two highly conserved cysteinyl peptides in rabbit form 2 (8, 9), rat P-450b (10, 11), and bacterial P-450_{cam} (12) that are under consideration for the source of the thiolate ligand to the heme iron atom near the enzyme active-site. One cysteinyl peptide, in exon 3 of P-450b (10), centers around the cysteine 150 in rabbit form 2 (8). This region (Fig. 4) represents P₃-450 nucleotides 502 to 552 and amino acids 148 to 164. In that region, only 37% nucleotide homology and 18% amino acid homology exists between mouse P₃-450 and rat P-450b.

	1149															
Mouse P ₃ -450	CTG	CCA	TAT	CTA	GAG	GCC	TTC	ATC	CTG	GAG	ATC	TAC	CGA			
	xx	xxx	xx		xx	xx	x	xxx	x	xxx	xx	x	x			
Mouse P-450b	ATG	CCA	TAC	ACT	GAT	GCA	GTT	ATC	CAC	GAG	ATT	CAG	AGG			
Rat P-450b	ATG	CCA	TAC	ACT	GAT	GCA	GTC	ATC	CAC	GAG	ATT	CAG	AGG			
	363															
P ₃ -450 protein	Leu	Pro	Tyr	Leu	Glu	Ala	Phe	Ile	Leu	Glu	Ile	Tyr	Arg			
		x	x			x		x		x	x		x			
P-450b protein	Met	Pro	Tyr	Thr	Asp	Ala	Val	Ile	His	Glu	Ile	Gln	Arg			

Fig. 3. Comparison of nucleotide and amino acid sequences in the "highly conserved region" among mouse P₃-450, mouse P-450b (18), and rat P-450b (16).

The other cysteinyl peptide, located at the beginning of exon 9 of P-450b (10), centers around cysteine 434 in rabbit form 2 (8). This region (Fig. 4) corresponds to P₃-450 nucleotides 1405 to 1464 and amino acids 449 to 471. In this region, there is 61% nucleotide homology and 52% amino acid homology between mouse P₃-450 and rat P-450b. Similar amounts of homology in these two regions are found between mouse P₃-450 protein and rabbit P-450 form 2 protein.

Table 1. Comparison of Amino Acid Composition

Mouse P ₃ -450		Rat P-450d	
Residue	Number of residues	Residue	Number of residues
Cysteine	6	Cysteine	5
Alanine	26	Alanine	26
Arginine	25	Arginine	21
Aspartamine	29	Aspartamine plus	
Aspartic acid	24	aspartic acid	44
Glutamine	24	Glutamine plus	
Glutamic acid	28	glutamic acid	49
Glycine	27	Glycine	37
Histidine	13	Histidine	13
Isoleucine	29	Isoleucine	25
Leucine	50	Leucine	49
Lysine	30	Lysine	32
Methionine	9	Methionine	6
Phenylalanine	34	Phenylalanine	31
Proline	31	Proline	29
Serine	37	Serine	31
Threonine	31	Threonine	24
Tryptophan	8	Tryptophan	4
Tyrosine	12	Tyrosine	10
Valine	40	Valine	32
Total	513	Total	468

M.W. of unmodified chain = 58,223

Estimated M.W. = 52,720

Mouse P ₃ -450	504	ATA	GCC	TGG	GAC	CGG	AGG	TCA	GCA	TCC	TCT	TGC	TAT	TTG	GAG	GAG	CAC	GTG
		x	x				x	x	x	xx	xx	x	xx	xxx	xx	x		x
Rat P-450b	534	CTG	GAA	GAA	CGG	ATT	CAG	GAG	GAA	GCC	CAA	TGT	TTG	GTG	GAG	GAA	CTG	GGG
P ₃ -450 protein	148	Ile	Ala	Ser	Asp	Pro	Thr	Ser	Ala	Ser	Ser	Cys	Tyr	Leu	Glu	Glu	His	Val
												x			x			
P-450b protein	158	Val	Glu	Glu	Arg	Ile	Gln	Glu	Glu	Ala	Gln	Cys	Leu	Val	Glu	Glu	Leu	Arg
Mouse P ₃ -450	1407	TTC	GGC	TTG	GGA	AAG	GGC	GGG	ATT	GGG	GAG	ATC	COG	GCC	AAG	TGG	GAA	GTC
		xxx	x		xxx	xxx	xxx	xx	xx	xx	xx	x	xxx	x	xxx	xxx	xxx	xx
Rat P-450b	1428	TTC	TCC	ACA	GGA	AAG	GGC	ATT	TGT	CTT	GGC	GAA	GGC	ATT	GCC	GGA	AAT	GAA
P ₃ -450 protein	449	Phe	Gly	Leu	Gly	Lys	Arg	Arg	Cys	Ile	Gly	Glu	Ile	Pro	Ala	Lys	Trp	Glu
		x			x	x	x	x	x		x	x			x	x	x	x
P-450b protein	456	Phe	Ser	Thr	Gly	Lys	Arg	Ile	Cys	Leu	Gly	Gly	Ile	Ala	Arg	Asn	Glu	Leu

Fig. 4. Comparison of nucleotide and amino acid sequences of two cysteinyl peptide regions in mouse P₃-450 and rat P-450b.

Fig. 5. The hydropathy index for mouse P₃-450 protein. ○, locations of the six cysteine residues.

The hydropathy index of the P₃-450 protein, as estimated by a sliding window of six amino acids (45), is illustrated in Fig. 5. Enhanced hydrophobicity of the first 20 to 25 amino acids, *i.e.* the "leader sequence," can be seen. Locations of the six cysteine residues are shown. Cysteine 456 is in a more hydrophobic environment than cysteine 158. Because P₃-450 metabolizes very hydrophobic drugs and polycyclic hydrocarbon carcinogens (4, 19), it is reasonable to presume that its enzyme active-site is in a hydrophobic domain. These data, especially when coupled with the homology data in Fig. 4, support the likelihood of cysteine 456 being the thiolate ligand to the heme iron atom near the enzyme active-site.

ACKNOWLEDGMENTS

The computer facilities and technical advice of Jake V. Maizel are greatly appreciated. We thank Hugh A. Privette and Mary Lynn Sienkiewicz for expert technical assistance and Ingrid E. Jordan for valuable secretarial help.

REFERENCES

1. Conney, A.H. (1967) *Pharmacol. Rev.* **19**, 317-366.
2. Lu, A.Y.H. and West, S.B. (1980) *Pharmacol. Rev.* **31**, 277-295.
3. Nebert, D.W., Tukey, R.H., Eisen, H.J. and Negishi, M. (1983) in *Gene Expression: UCLA Symposia on Molecular and Cellular Biology*, New Series Volume VIII, Hamer, D. and Rosenberg, M., Eds., Part 3, pp. 187-206, Alan R. Liss, Inc., New York.
4. Negishi, M. and Nebert, D.W. (1979) *J. Biol. Chem.* **254**, 11015-11023.
5. Negishi, M., Jensen, N.M., Garcia, G.S. and Nebert, D.W. (1981) *Eur. J. Biochem.* **115**, 585-594.

6. Ohyama, T., Nebert, D.W. and Negishi, M. (1984) *J. Biol. Chem.*, in press.
7. Nebert, D.W. and Negishi, M. (1982) *Biochem. Pharmacol.* 31, 2311-2317.
8. Heinemann, F.S. and Ozols, J. (1983) *J. Biol. Chem.* 258, 4195-4201.
9. Tarr, G.E., Black, S.D., Fujita, V.S. and Coon, M.J. (1983) *Proc. Natl. Acad. Sci. U.S.A.* 80, 6552-6556.
10. Mizukami, Y., Sogawa, K., Suwa, Y., Muramatsu, M. and Fujii-Kuriyama, Y. (1983) *Proc. Natl. Acad. Sci. U.S.A.* 80, 3958-3962.
11. Kumar, A., Raphael, C. and Adesnik, M. (1983) *J. Biol. Chem.* 258, 11280-11284.
12. Haniu, M., Armes, L.G., Yasunobu, K.T., Shastry, B.A. and Gunsalus, I.C. (1982) *J. Biol. Chem.* 257, 12664-12671.
13. Tukey, R.H., Nebert, D.W. and Negishi, M. (1981) *J. Biol. Chem.* 256, 6969-6974.
14. Tukey, R.H., Negishi, M. and Nebert, D.W. (1982) *Mol. Pharmacol.* 22, 779-786.
15. Gonzalez, F.J. and Kasper, C.B. (1982) *J. Biol. Chem.* 257, 5962-5968.
16. Fujii-Kuriyama, Y., Mizukami, Y., Kawajiri, K., Sogawa, K. and Muramatsu, M. (1982) *Proc. Natl. Acad. Sci. U.S.A.* 79, 2793-2797.
17. Atchison, M. and Adesnik, M. (1983) *J. Biol. Chem.* 258, 11285-11295.
18. Stupans, I., Kessler, D.J., Ikeda, T. and Nebert, D.W. (1984) *DNA*, in press.
19. Eisen, H.J., Hannah, R.R., Legraverend, C., Okey, A.B. and Nebert, D.W. (1983) in *Biochemical Actions of Hormones*, Litwack, G., Ed., Vol. X, pp. 227-258, Academic Press, New York.
20. Okey, A.B., Bondy, G.P., Mason, M.E., Nebert, D.W., Forster-Gibson, C., Muncan, J. and Dufresne, M.J. (1980) *J. Biol. Chem.* 255, 11415-11422.
21. Tukey, R.H., Hannah, R.R., Negishi, M., Nebert, D.W. and Eisen, H.J. (1982) *Cell* 31, 275-284.
22. Gonzalez, F.J., Tukey, R.H. and Nebert, D.W. (1984) manuscript submitted for publication.
23. Nebert, D.W., Eisen, H.J., Negishi, M., Lang, M.A., Hjelmeland, L.M. and Okey, A.B. (1981) *Annu. Rev. Pharmacol. Toxicol.* 21, 431-462.
24. Nebert, D.W. (1979) *Mol. Cell. Biochem.* 27, 27-46.
25. Nebert, D.W., Negishi, M., Lang, M.A., Hjelmeland, L.M. and Eisen, H.J. (1982) *Advanc. Genet.* 21, 1-52.
26. Gonzalez, F.J. and Kasper, C.B. (1981) *J. Biol. Chem.* 256, 4697-4700.
27. Aviv, H. and Leder, P. (1972) *Proc. Natl. Acad. Sci. U.S.A.* 69, 1408-1412.
28. Okayama, H. and Berg, P. (1982) *Mol. Cell. Biol.* 2, 161-170.
29. Okayama, H. and Berg, P. (1983) *Mol. Cell. Biol.* 3, 280-289.
30. Birnboim, H.C. and Doly, J. (1979) *Nucl. Acids Res.* 7, 1513-1523.
31. Maniatis, T., Fritsch, E.F. and Sambrook, J. (1982) in *Molecular Cloning: A Laboratory Manual*, Cold Spring Harbor Laboratory, New York.
32. Messing, J., Crea, R. and Seeburg, P.H. (1981) *Nucl. Acids Res.* 9, 309-321.
33. Sanger, F., Nicklen, S. and Coulson, A.R. (1977) *Proc. Natl. Acad. Sci. U.S.A.* 74, 5463-5467.
34. Anderson, S. (1981) *Nucl. Acids Res.* 9, 3015-3027.
35. Kushner, S.R. (1978) in *Genetic Engineering*, Boyer, H.B. and Nicosia, S., Eds., p. 17, Elsevier/North-Holland Biomedical Press, Amsterdam.
36. Sanger, F. and Coulson, A.R. (1978) *FEBS Lett.* 87, 107-110.
37. Staden, R. (1980) *Nucl. Acids Res.* 8, 3673-3694.

38. Brutlag, D.J., Clayton, J., Friedland, P. and Kedes, L.H. (1982) Nucl. Acids Res. 10, 279-292.
39. Orcott, B.C., George, D.G., Fredrickson, J.A. and Dayhoff, M.O. (1982) Nucl. Acids Res. 10, 157-174.
40. Ikeda, T., Altieri, M., Chen, Y.-T., Nakamura, M., Tukey, R.H., Nebert, D.W. and Negishi, M. (1983) Eur. J. Biochem. 134, 13-18.
41. Botelho, L.H., Ryan, D.E., Yuan, P.-M., Kutny, R., Shively, J.E. and Levin, W. (1982) Biochemistry 21, 1152-1155.
42. Ozols, J., Heinemann, F.S. and Johnson, E.F. (1981) J. Biol. Chem. 256, 11405-11408.
43. Wilson, A.C., Carlson, S.S. and White, T.J. (1977) Annu. Rev. Biochem. 46, 573-639.
44. Leighton, J.K., DeBrunner-Vossbrinck, B.A. and Kemper, B. (1984) Biochemistry 23, 204-210.
45. Kyte, J. and Doolittle, R.F. (1982) J. Mol. Biol. 157, 105-132.