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†

John K. Leighton, Bettina A. DeBrunner-Vossbrinck, and Byron Kemper*

ABSTRACT: We have isolated from rabbit liver three cDNA clones of 1400–1800 base pairs that hybridize selectively to RNA from animals treated with phenobarbital. The nucleotide sequences of the cDNAs have been determined. In the protein coding region the nucleotide sequences of two of the cDNAs are 88% homologous, and the third cDNA is about 72–74% homologous to the other two. All three are 55–60% homologous to rat liver cytochrome P-450b cDNA. The amino acid sequences derived from the cDNA sequences are about 50% homologous to those of rat liver cytochrome P-450b and rabbit liver cytochrome P-450 (form 2). The degree of homology differs substantially in different regions of the protein. The hydrophobicity profiles of these five mammalian cytochromes P-450 are very similar and contain up to eight regions of hydrophobicity that are long enough to span a membrane. These results indicate that these three cDNAs code for rabbit liver cytochromes P-450 which are different from any rabbit liver cytochrome P-450 for which amino acid sequence information is published. These cDNAs are part of a family of genes that are related to rabbit liver cytochrome P-450 (form 2) and rat liver cytochrome P-450b which are the major phenobarbital-inducible forms. The divergence of amino acid sequence between the rat and rabbit forms and the divergence of nucleotide sequences of silent sites in the two most closely related rabbit forms suggest that cytochromes P-450 have a relatively high rate of amino acid divergence compared to many other vertebrate proteins.

Liver cytochromes P-450 are a family of microsomal hemoproteins that are terminal oxidases in the metabolism of many xenobiotic and endogenous compounds (Sato & Kato, 1982). The broad substrate specificity of these enzymes is due in part to the multiplicity of cytochrome P-450 enzymes. Up to 11 forms and 8 forms have been partially characterized in rabbit liver (Aoyama et al., 1981) and rat liver (Guengerich et al., 1982), respectively. Four different rabbit liver cytochromes P-450 have been partially sequenced and clearly are distinct proteins (Koop et al., 1982). The activity of these enzymes can be induced by a number of different agents. In general more than one form of cytochrome P-450 is induced by a given inducer, and different combinations of cytochrome P-450 are induced by different inducers. The mechanism of the induction of cytochrome P-450 is not well understood. Induction by polycyclic hydrocarbons involves a cytosolic receptor (Tukey et al., 1982), but a similar receptor has not been detected for phenobarbital induction. Increased levels of mRNA are present after induction, indicating the increased activity of cytochrome P-450 is due to increased synthesis of the enzyme (Dubois et al., 1979; Kumar & Padmanaban, 1980; Colbert et al., 1979; Adesnik et al., 1981; Gonzales & Kasper, 1982; Dilella et al., 1981). The isolation and characterization of cytochrome P-450 mRNAs and genes can provide definitive evidence concerning the number of cytochrome P-450, comparative structural data to provide insight into the evolution of this family of proteins, and the foundation to study the mechanisms of the induction of these proteins. We have undertaken studies to isolate by molecular cloning the cDNA and genes that code for rabbit liver phenobarbital-inducible cytochrome P-450.

Multiple forms of cytochrome P-450 are induced by phenobarbital in both the rat and rabbit. The major forms are cytochrome P-450b (P-450b)[1] in the rat and cytochrome P-450 (form 2) in the rabbit. In the rat, translation of mRNA results in four forms of cytochrome P-450 that can be distinguished immunochemically (Walz et al., 1982). The determination of the sequence of cloned cDNAs demonstrated that two closely related forms exist (Fujii-Kuriyama et al., 1982), and this was confirmed by isolation and sequence analysis of a minor form of cytochrome P-450 (P-450e) that accounts for the second cDNA (Yuan et al., 1983). The genes of these proteins are part of a gene family of more than six members since the six distinct related genomic forms which have been cloned do not account for all the bands observed by Southern blot analysis of rat genomic DNA (Mizukami et al., 1983). In the rabbit the major phenobarbital-inducible form is present in significant amounts only after induction (Haugen & Coon, 1976). The complete sequence of this protein has been determined (Heinemann & Ozols, 1983). At least one other phenobarbital-inducible isozyme has been detected (Aoyama et al., 1981). Translation of mRNA has resulted in only one immunoprecipitable product with antisera to cytochrome P-450 (form 2) (Dilella et al., 1981).

In the present study we report the isolation and nucleotide sequence of three closely related cDNAs that code for rabbit liver cytochrome P-450 and hybridize selectively to RNA isolated from phenobarbital-treated rabbits. These three cytochrome P-450s are about 50% homologous to rabbit P-450 (form 2) and rat P-450b and are different from any of the four

† From the Department of Physiology and Biophysics and College of Medicine at Urbana-Champaign, University of Illinois, Urbana, Illinois 61801. *Received June 23, 1983.* This work was supported by NIH Grant GM27718.

[1] Abbreviations: P-450b, rat liver cytochrome P-450b; P-450e, rat liver cytochrome P-450e; P-450 (form 2), rabbit liver cytochrome P-450 (form 2) or cytochrome P-450 (LM2); P-450 (form 3b), rabbit liver cytochrome P-450 (form 3b); P-450PBcl, P-450PBc2, and P-450PBc3, rabbit liver cytochromes P-450 predicted from the cDNA clone sequences; P-450 (CAM), camphor-inducible cytochrome P-450 from *Pseudomonas putida*; Tris, tris(hydroxymethyl)aminomethane; EDTA, ethylenediaminetetraacetic acid; Hepes, *N*-(2-hydroxyethyl)piperazine-*N'*-2-ethanesulfonic acid; EGTA, ethylene glycol bis(β-aminoethyl ether)-*N*,*N*,*N'*,*N'*-tetraacetic acid.

0006-2960/84/0423-0204$01.50/0

rabbit liver cytochromes P-450 for which amino acid sequence has been published.

Materials and Methods

Cloning of cDNA. Adult male New Zealand rabbits (2 kg) were administered phenobarbital for 48 h as a 0.1% solution in drinking water. Treated and control animals were starved overnight prior to sacrifice. Livers were removed, cut into about 1-cm chunks, and frozen in liquid N_2. Total RNA was isolated by proteinase K digestion and phenol–chloroform extraction (Lizardi & Engelberg, 1979). DNA was removed by sodium acetate precipitation of RNA as described previously (Stolarsky & Kemper, 1978). Total RNA was isolated by the guanidinium thiocyanate method (Chirgwin et al., 1979) for the RNA dot blot experiments. The RNA was fractionated on a isokinetic 5–30% sucrose gradient in 24 mM Tris-HCl, pH 7.2, 2 mM sodium EDTA, and 1% sodium dodecyl sulfate (McCarty et al., 1974) for 16 h at 22 000 rpm in a Beckman SW27 rotor. Fractions corresponding to the 18S ribosomal RNA peak were collected. Poly(A)-containing RNA was isolated by oligo(dT)–cellulose chromatography as described previously (Stolarsky & Kemper, 1978).

Double-stranded cDNA was synthesized by sequential incubation of poly(A) RNA with reverse transcriptase and *E. coli* DNA polymerase I as described previously (Gordon & Kemper, 1980). The double-stranded cDNA was treated with S_1 nuclease and inserted into the *Pst*I site of pBR322 by homopolymer extension techniques as described previously (Weaver et al., 1982). The annealed recombinant plasmid was used to transform *E. coli* RRI by the calcium shock technique (Mandel & Higa, 1970). About 800 colonies that were resistant to tetracycline and sensitive to ampicillin were transferred to wells of a 96-well microtiter dish containing L-broth. After incubation overnight at 37 °C, an equal volume of 50% glycerol was added, and the cultures were stored at −70 °C.

Colony Filter Hybridization. For colony filter hybridization a 96-prong stainless steel comb was used to transfer colonies from the microtiter plates to nitrocellulose membranes on L-broth agar plates. After growth overnight, the colonies were screened as described by Grunstein & Hogness (1975) by using single-stranded cDNA synthesized from poly(A) RNA isolated from control rabbits and rabbits treated with phenobarbital. Duplicate filters were used for each probe.

Hybrid Selection of mRNA. Colonies that hybridized selectively to cDNA from phenobarbital-treated animals were grown in 100-mL cultures, and plasmid DNA was isolated and used to select mRNA by hybridization as described by Parnes et al. (1981). Selected mRNA was translated in a reticulocyte cell-free translation system (Dorner & Kemper, 1978).

Isolation of Plasmid DNA. For initial restriction enzyme analysis plasmids were isolated from 1-mL cultures by the boiling technique as described by Holmes & Quigley (1981). Larger quantities of plasmid DNA were isolated by lysozyme–Triton X-100 lysis of cells, polyethylene glycol precipitation, and two CsCl gradient centrifugations as described previously (Weaver et al., 1982).

RNA Blot Hybridizations. RNA blot hybridization was carried out on "Gene Screen" (New England Nuclear, Boston MA) as described in the New England Nuclear Gene Screen Manual (hybridization method II). Liver RNA isolated from control and phenobarbital-treated rabbits and tRNA were boiled for 1 min in 5 mM K^+-Hepes, pH 7.4, 1 mM EDTA, and 0.1 mM EGTA and spotted on Gene Screen. Plasmid DNA nick translated to a specific activity of $(1–4) \times 10^8$ cpm/μg was hybridized to the RNA.

Nucleotide Sequence Analysis. The nucleotide sequences

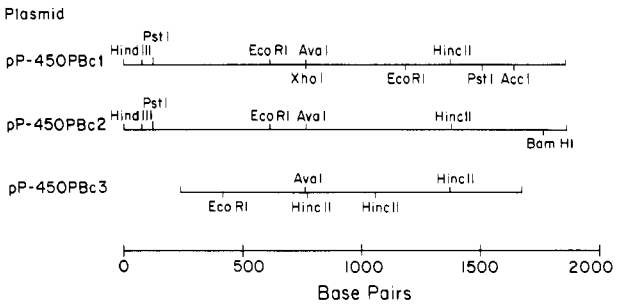

FIGURE 1: Restriction maps of three classes of cloned rabbit liver cDNAs. The sites for restriction enzymes that are common to two or three of the classes are shown above the lines, while restriction sites that are unique for a given class are shown below the line.

of the cDNA inserts were determined by the chemical cleavage method (Maxam & Gilbert, 1980) except that in the C and C + T reaction, 20 μL of acetylacetone was added after the first ethanol precipitation to inactivate residual hydrazine (Jay et al., 1982).

Estimation of Hydrophobicity. The hydrophobicity profile was generated by an Apple II computer program developed by Larson and Messing (University of Minnesota). This program used a seven amino acid sliding window to calculate the hydrophobicity. The index used was the following: −0.5; Arg, 3.0; Asn, 0.2; Asp, 3.0; Cys, −1.0; Gln, 0.2; Glu, 3.0; Gly, O; His, −0.5; Ile, −1.8; Leu, −1.8; Lys, 3.0; Met, −1.3; Phe, −2.5; Pro, O; Ser, 0.3; Thr, −0.4; Trp, −3.4; Tyr, −2.3; Val, −1.5 [adapted from Levitt (1976)].

Results

Our initial aim was to clone cDNA that hybridized selectively to RNA isolated from phenobarbital-treated animals. Poly(A) RNA was isolated from the livers of rabbits given drinking water containing 0.1% phenobarbital, and the RNA sedimenting from 14S to 20S was isolated in an attempt to enrich for mRNA of the correct size to code for cytochrome P-450. Double-stranded cDNA synthesized from this template was incorporated into the *Pst*1 site of pBR322 and cloned. A cDNA bank of 800 colonies that were resistant to tetracycline and sensitive to ampicillin was obtained. These colonies were screened by hybridization to radioactive cDNA synthesized from poly(A) RNA from either control or phenobarbital-treated animals. Twenty-three colonies were selected that possibly hybridized selectively to the cDNA probe from the treated animals. The messenger RNA that hybridized to the plasmids of these colonies was isolated by filter hybridization and translated in a reticulocyte cell free system. Three of the plasmids selected mRNA that coded for proteins of about the size expected for cytochrome P-450. These three plasmids were isolated and hybridized to filter immobolized poly(A) RNA from control and treated animals. One of the three hybridized 6–8-fold more to the RNA from treated animals. This plasmid contained a cDNA insert of only about 500 base pairs and thus was used as a probe to rescreen the original cDNA bank containing 800 colonies. Ten additional colonies were obtained that hybridized to the plasmid.

The sizes of the cDNA inserts of the 10 colonies ranged from about 500 to 1800 base pairs. The five largest plasmids, containing inserts of 1400–1800 base pairs, were analyzed by restriction digestion, and restriction maps were determined as shown in Figure 1. The plasmids fell into three groups with three of the plasmids in group c1 and one each in group c2 and group c3. pP-450PBcl and pP-450PBc2 which appeared most closely related shared five restriction sites but also had unique sites. pP-450PBc3 contained only two sites in common

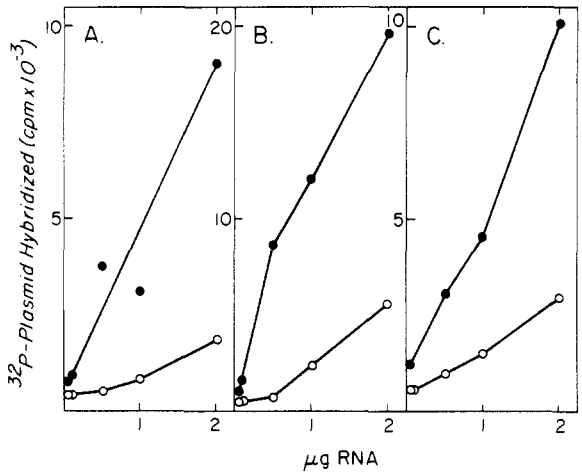

FIGURE 2: Hybridization of plasmid DNA to RNA from phenobarbital-treated and control rabbit liver. Poly(A)-containing RNA was immobilized on Gene Screen, and plasmid DNA, labeled by nick translation, was hybridized to the RNA as described under Materials and Methods. RNA was isolated from either (O) control rabbits or (●) rabbits given 0.1% phenobarbital in drinking water for 48 h. Plasmid DNA was from (A) pP-450PBc2, (B) pP-450PBc1, and (C) pP-450PBc3.

with the other two plasmids, had several unique sites, and appeared to be least closely related to the other cDNA inserts.

The hybridizations of each of the three plasmids to poly(A) RNA from control animals and animals treated with phenobarbital are shown in Figure 2. In each case 6–8 times as much control RNA was required on the filter as RNA from treated animals to produce the same amount of hybridization. Similar results were obtained with total RNA (not shown), eliminating the possibility that fractionation of the RNA may have resulted in selective loss of the hybridizable mRNA from the control preparation. These results using relatively stringent hybridization conditions demonstrated that each of these cDNA inserts, at the least, share extensive homology with phenobarbital-inducible mRNA. Because of the homology between the three cDNA inserts and potentially other cDNAs, additional experiments will be required to establish with certainty that the mRNA corresponding to each to these cDNAs is actually inducible by phenobarbital.

The sequences of each of the cDNA inserts for the three plasmids were determined by the chemical cleavage method of Maxam–Gilbert. The strategy of sequencing is shown in Figure 3. The nucleotide sequences of the three cDNA inserts are shown in Figure 4 and compared to the sequence for the cDNA of the rat phenobarbital-inducible cytochrome P-450 determined by Fujii-Kurijama et al. (1982). Each cDNA contains a single long open reading frame and a TGA termination codon in the same location as the rat sequence. The amino acid sequences derived from the cDNA sequences are shown in Figure 5. Two of the three cloned cDNAs contained poly(A) sequences and, thus, the entire 3′-untranslated region of the mRNA. None of the cloned cDNAs contained the entire 5′ region. pP-450PBcl and pP-450PBc2 were missing the 5′-untranslated region and sequence that codes for the first 11 amino acids based on the rat sequence, and pP-450PBc3 was missing the sequence coding for the first 87 amino acids.

As shown in Table I the nucleotide sequence of each of the three rabbit cDNAs was about 55–60% homologous to that of the rat cytochrome P-450 cDNA in the coding region of the cDNA. There is much less homology in the 3′-untranslated regions. In the coding region only one gap of three nucleotides in pP-450PBc3 was required to maximize homology. The most closely related cDNAs were pP-450PBcl and pP-450PBc2

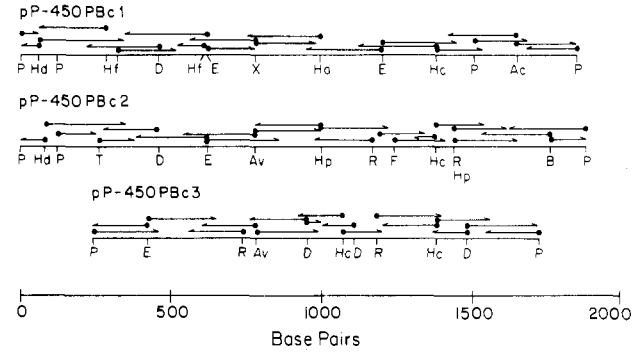

FIGURE 3: Strategy for determining the nucleotide sequence of the cDNA inserts of pP-450PBcl, pP-450PBc2, and pP-450PBc3. The cDNA inserts were sequenced as described by Maxam & Gilbert (1980). Sites of restriction enzymes used in the sequencing are indicated. The closed circles indicates the 5′ ends of the fragments that were labeled with [^{32}P]phosphate, and the arrow indicates the direction and extent of the sequence determined for each fragment. Abbreviations for the restriction enzymes are the following: Ac, *Acc*I; Av, *Ava*I; B, *Bam*HI; D, *Dde*I; E, *Eco*RI; F, *Fnu*4HI; Ha, *Hae*III; Hc, *Hinc*II; Hd, *Hind*III; Hf, *Hin*fI; Hp, *Hpa*II; P, *Pst*I; R, *Rsa*I; T, *Taq*I; X, *Xho*I.

Table I: Homology of Amino Acid and Nucleotide Sequences of Cytochromes P-450PBcl, P-450PBc2, and P-450PBc3[a]

	PBcl	PBc2	PBc3	P-450b
pP-450PBcl		88 (77)	74 (33)	60 (28)
pP-450PBc2	85		72 (31)	56 (31)
pP-450PBc3	66	65		56 (27)
rat P-450b	51	50	49	
rabbit LM_2	48	49	46	76

[a] The percent homologies for the nucleotide sequence in the coding region of the cDNAs are shown in the upper right part of the table. The numbers in parentheses are the homologies in the 3′-noncoding region. In the 3′-noncoding region gaps were introduced to maximize homology between pP-450cl and pP-450c2, but no attempt was made to maximize homology for the other cDNAs. The percent homologies between amino acid sequences are shown in the lower left part of the table.

which were 88% homologous in the coding region and 77% homologous in the 3′-noncoding region although gaps have to be introduced in the 3′-noncoding region to maximize homology. pP-450PBc3 was 65% homologous to pP-450PBcl and pP-450PBc2 and shows much less homology in the 3′-noncoding region. Homology of the derived protein sequence was similar to but slightly lower than that of the DNA sequences.

While the extent of homology extended throughout the coding sequence, there are regions that were more or less homologous. As can be seen in Figures 4 and 5 and in the summary at the bottom of Figure 6, regions of particular variability included the first 100 amino acids at the amino terminus, a region from amino acids 180 to 240 (nucleotides 550–730) and the carboxyl-terminal 40 amino acids (nucleotides 1320–1445). Particularly conserved regions included amino acids 110–150, 310–340, and 410–440. Three cysteines were conserved in all five sequences at 141, 169, and 425. The amino acids around these cysteines are also strongly conserved, and they may play an important role in the structure or function of the protein. The "analogous peptide" region (Ozols et al., 1981) (amino acids 335–347) which was common to rat P-450b and rabbit P-450 (forms 2 and 3b) was well conserved in all the cytochromes P-450. These striking homologies strongly indicate that the cDNA inserts in pP-450PB (c1,c2,c3) code for cytochrome P-450 molecules.

The 3′-untranslated region of the cDNAs was heterogeneous in size. The polyadenylation signal in pP-450PBcl, AATAGA,

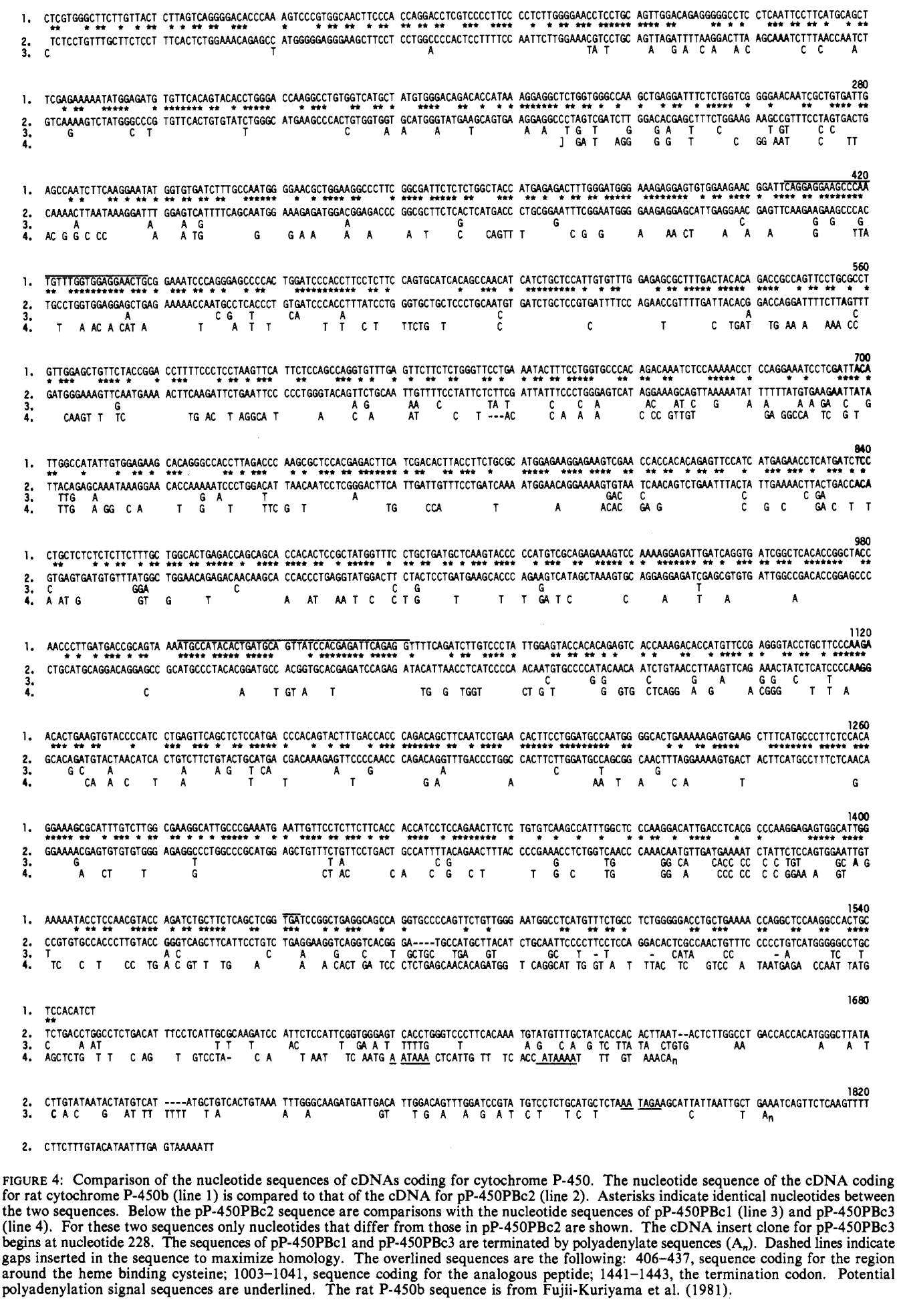

FIGURE 4: Comparison of the nucleotide sequences of cDNAs coding for cytochrome P-450. The nucleotide sequence of the cDNA coding for rat cytochrome P-450b (line 1) is compared to that of the cDNA for pP-450PBc2 (line 2). Asterisks indicate identical nucleotides between the two sequences. Below the pP-450PBc2 sequence are comparisons with the nucleotide sequences of pP-450PBc1 (line 3) and pP-450PBc3 (line 4). For these two sequences only nucleotides that differ from those in pP-450PBc2 are shown. The cDNA insert clone for pP-450PBc3 begins at nucleotide 228. The sequences of pP-450PBc1 and pP-450PBc3 are terminated by polyadenylate sequences (A_n). Dashed lines indicate gaps inserted in the sequence to maximize homology. The overlined sequences are the following: 406–437, sequence coding for the region around the heme binding cysteine; 1003–1041, sequence coding for the analogous peptide; 1441–1443, the termination codon. Potential polyadenylation signal sequences are underlined. The rat P-450b sequence is from Fujii-Kuriyama et al. (1981).

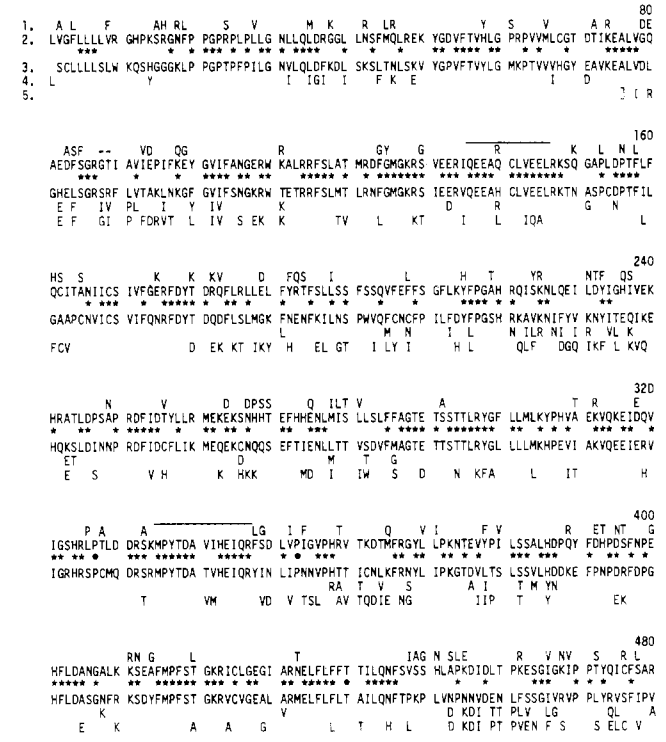

FIGURE 5: Comparison of the amino acid sequences of cytochrome P-450. The sequence of rat cytochrome P-450b (line 2) is compared to the sequence of cytochrome P-450PBc2 (line 3). Asterisks indicate identical amino acids between the two sequences. Above the P-450b sequence is a comparison with the sequence of rabbit liver cytochrome P-450 (form 2) (line 1) in which only amino acids different from those of P-450b are shown. Below the P-450PBc2 sequence is a comparison with sequences of cytochrome P-450PBc1 (line 4) and P-450PBc3 (line 5) in which only the amino acids different from P-450PBc2 are shown. The beginning of the sequence of P-450PBc3 at amino acid 77 is indicated by the reverse bracket. Dashes indicate gaps inserted to maximize homology. The region around the cysteine that is involved in binding the heme moiety and the analogous peptide region are overlined. The first 11 amino acids at the amino terminus of P-450 and P-450 (form 2), are not shown. The rat P-450b sequence is from Fujii-kuriyama et al. (1981), and the rabbit P-450 (form 2) sequence is from Heinemann & Ozols (1983).

differed from the usual AATAAA sequence (Proudfoot & Brownlee, 1976). Fifteen nucleotides around this sequence were conserved in pP-450PBc2, but polyadenylation did not occur at the same site. Presumably a second polyadenylation signal occurs in the mRNA corresponding to pP-450PBc2 but was not present in the cloned cDNA. pP-450PBc3 has a much shorter 3′-untranslated region with a normal polyadenylation signal. Interestingly another consensus polyadenylation site was present 20 nucleotides from the one that is used in pP-450PBc3. Neither of these signals was conserved in the other cDNAs. Clearly the potential exists for multiple polyadenylation sites in each of the cDNAs.

The codon usage for these mRNAs is shown in Table II and was uneven as has been shown with other mRNAs (Grantham et al., 1980). Codons ending with CG were rare, but there was no discrimination against CG in the first two positions in arginine codons. Codons ending with TG were preferred over those ending with TA. In general as might be expected for highly homologous mRNAs similar patterns of codon usage were observed for the three P-450 mRNAs. Exceptions were histidine for which pP-450PBc3 had a preference for CAT in contrast to the other two cDNAs preference for CAC. There were also substantial differences in usage of the leucine and phenylalanine codons.

The hydrophobicity profiles of the cytochrome P-450 molecules are shown in Figure 6 and are compared with that

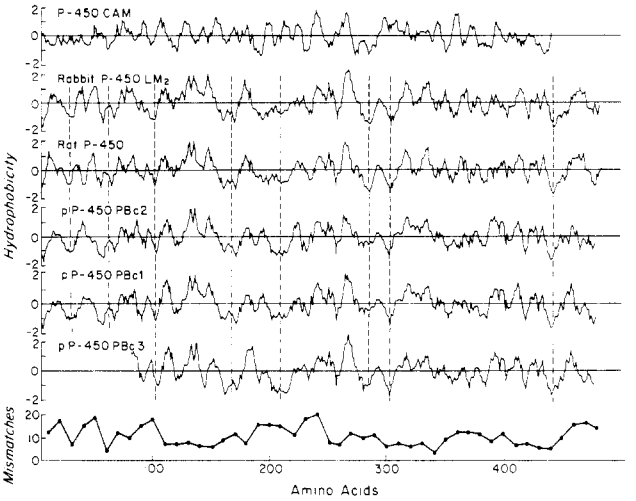

FIGURE 6: Hydrophobicity profiles of cytochrome P-450 and conservation of sequence. The hydrophobicity profile was determined for a sliding window of seven amino acids as described under Materials and Methods. Positive numbers indicate hydrophilic regions, and negative numbers represent hydrophobic regions. In the bottom graph conservation of sequence among the mammalian proteins was estimated for groups of 10 amino acids by summing the number of amino acids used at each position and then subtracting 10. A score of 0 indicates identical sequences for all five proteins and a score of 40 mismatches indicates a different amino acid was used for each position for the 10 amino acid sequence.

of cytochrome P-450 (CAM). The hydrophobicity profiles of the proteins derived from the rabbit cDNAs were remarkably similar to the profiles for P-450 (form 2) and P-450b in spite of the fact that there was only 50% homology in amino acid sequence. Eight hydrophobic regions that were greater than 10–15 residues in length and were conserved in the five proteins are indicated by the dotted lines. A ninth region would probably be at the amino terminus which was missing from our cloned cDNAs. These regions were large enough that they could potentially cross the membrane. P-450 (CAM) which is not an integral membrane protein did not in general have corresponding hydrophobic regions. A striking similarity in the mammalian proteins was present in a double hydrophobic region centered at amino acid 200 that is bounded by strongly hydrophilic regions. A similar profile was observed in this region in P-450 (CAM). A relatively hydrophilic region from amino acid 100 to 160 was also well conserved. The significance of the conservation of these hydrophobic regions is not clear, but they presumably are important in either the association of the protein with the membrane or the three-dimensional structure of the protein. The conservation of hydrophobic regions did not correlate well with the regions of the conservation of amino acid sequence as shown at the bottom of Figure 6.

Discussion

We have isolated three rabbit liver cDNAs by molecular cloning that hybridize selectively to mRNA from phenobarbital-treated animals. The amino acid sequence derived from the cDNA sequence does not correspond to any of the rabbit liver cytochromes P-450 for which amino acid sequence data are published. The striking homology between these sequences and other cytochromes P-450 strongly indicate that these cDNAs code for cytochromes P-450. Each sequence has a terminator codon at the same location as the rat P-450b and rabbit P-450 (form 2). Three cysteines and surrounding regions are well conserved in all five of these mammalian P-450s. Two of these regions around Cys-155 and Cys-452 of P-450 (form 2) are similar to two corresponding regions in the

Table II: Codon Usage for Cytochromes P-450PBc1 (c1), P-450PBc2 (c2), and P-450PBc3 (c3)

amino acid	codon	usage c2	c1	c3	amino acid	codon	usage c2	c1	c3
Ala	GCT	4	6	3	Ile	ATT	16	20	16
	GCC	8	6	2		ATC	8	8	6
	GCA	3	3	5		ATA	2	7	5
	GCG	0	0	1					
					Leu	TTA	6	1	6
Arg	CGT	4	4	2		TTG	4	9	11
	CGC	3	4	3		CTT	9	8	3
	CGA	3	4	1		CTC	7	10	6
	CGG	5	3	3		CTA	5	1	1
	AGA	5	7	6		CTG	20	21	14
	AGG	6	5	3					
					Lys	AAA	15	14	14
Asn	AAT	16	12	9		AAG	14	15	14
	AAC	14	16	7					
					Met	ATG	11	13	8
Asp	GAT	11	10	14					
	GAC	11	13	12	Phe	TTT	14	11	19
						TTC	21	18	12
Cys	TGT	7	4	5					
	TGC	5	6	3	Pro	CCT	10	10	7
						CCC	14	13	11
Gln	CAA	4	4	4		CCA	5	7	8
	CAG	11	12	10		CCG	1	2	0
Glu	GAA	12	12	16	Ser	TCT	5	4	7
	GAG	15	15	14		TCC	5	5	3
						TCA	6	4	1
Gly	GGT	1	3	0		TCG	0	1	1
	GGC	6	7	6		AGT	5	2	6
	GGA	14	16	12		AGC	10	11	3
	GGG	6	5	3					
					Thr	ACT	6	6	9
His	CAT	5	2	10		ACC	9	15	9
	CAC	7	5	3		ACA	10	8	8
						ACG	4	3	0
Trp	TGG	3	3	3					
					Val	GTT	3	7	8
Tyr	TAT	8	9	7		GTC	10	4	7
	TAC	5	7	3		GTA	3	2	2
						GTG	20	17	8

bacterial enzyme, P-450 (CAM) and each of these cysteines has been proposed as the ligand for the heme moiety (Black et al., 1982; Haniu et al., 1982; Gotoh et al., 1983). The analogous peptide region (Ozols et al., 1981), which is conserved in P-450 (form 2), P-450 (form 3b), and P-450b, contains 11 of 13 common amino acids in P-450PBcl and P-450PBc2 and 12 of 13 in P-450PBc3. This sequence in P-450PBc3 is identical with the sequence in P-450 (form 3b). Finally the hydrophobic profiles of these proteins are very similar to those of P-450 (form 2) and P-450b. Because of these homologies, we feel justified in designating these cDNAs as cytochrome P-450 cDNAs. Additional studies will be required to determine whether each corresponds to one of the 11 cytochromes P-450 that have been described in rabbit liver (Aoyama et al., 1981).

The three cDNA clones are part of a family of genes that includes cytochrome P-450 (form 2). There may be additional members of the family which were not present in the cDNA bank we have analyzed. We have not detected cDNA clones for P-450 (form 2) by screening this bank at low stringency with a restriction fragment from pP-450PBc2 or with an oligonucleotide probe made on the basis of the amino acid sequence of the analogous peptide in P-450 (form 3b). Since P-450 (form 2) is generally considered the major phenobarbital-inducible cytochrome P-450 and our cDNA bank contained several hundred clones (even allowing for cell division before plating the transformed bacteria), it is unclear why we did not detect P-450 (form 2) cDNA. A trivial explanation may be that we discarded the P-450 (form 2) mRNA when we fractionated the mRNA by sucrose gradient centrifugation before synthesizing the cDNA for cloning. One of the mRNAs for cytochrome P-450 in mice is considerably larger than expected (Tukey et al., 1981); however, mRNA coding for P-450 (form 2) has been reported to be present in a fraction of RNA containing the 18S rRNA (Treadwell et al., 1980) which is the fraction we selected. The mRNA was isolated 48 h after phenobarbital was added to the drinking water, and chronic stimulation for this length of time may not result in significant accumulation of P-450 (form 2) mRNA. Since we did not detect cDNA clones for P-450 (form 2), which should be a major species, obviously there may also be other members of this family that were not detected.

If the gene for rabbit P-450 (form 2) corresponds to that of rat P-450b and rabbits and rodents diverged 60 million years ago (Fitch & Langley, 1976), then the 24% (29% corrected) difference between these two proteins indicates that the time required for a 1% difference in amino acid sequence is 2.1 million years. This is among the highest rates observed for vertebrate proteins (Wilson et al., 1977). If rabbit P-450 (form 2) and rat P-450b are not corresponding members of the gene family then the estimated rate of divergence could be substantially reduced. For relatively recent times of divergence, silent nucleotide changes in the coding region may be used as an independent estimate of the time of divergence (Perler et al., 1980). For the two most closely related rabbit P-450s, P-450PBcl and P-450PBc2, the corrected divergence of the silent sites in the nucleotide sequence is 32%, calculated as described by Perler et al. (1980). The divergence of silent sites in proinsulin and globin is about 1% per million years (Perler et al., 1980). The change of 32% in P-450PBcl and P-450PBc2 would require 32 million years which is about the same as 36 million years estimated from the amino acid sequence change and a rate of 1% change per 2.1 million years. Additional sequences of cytochrome P-450 from different species will be required to establish the rate of divergence of cytochrome P-450 with certainty. However, a high rate of divergence is consistent with the evolution of a family of cytochromes P-450 with the broad substrate specificity necessary to metabolize and detoxify a wide variety of xenobiotics.

The homology observed in this family of genes can be explained by three gene duplications. The first duplication of an ancestral gene about 140 million years ago separated the genes for rabbit P-450 (form 2) and rat P-450b from those for pP-450(c1,c2,c3). A second duplication about 85 million years ago separated the gene for pP-450c3 from those for pP-450(c1,c2) and a third duplication about 30 million years ago separated the two genes for pP-450c1 and pP-450c2.

The comparative analysis of the sequence of several cytochromes P-450 may be useful in correlating the structure of the protein with its function. Microsomal cytochrome P-450 must interact with the heme moiety, the endoplasmic membrane, and cytochrome P-450 reductase so that these regions might be expected to be conserved. As noted above the region around two cysteines, one of which binds to heme, is conserved in these proteins. Depending on the secondary structure of a peptide 15–20 amino acids are required to pass through a membrane. The hydrophobicity of at least eight regions of this length is conserved in the five mammalian sequences but is not conserved in the bacterial P-450 (CAM) which is not an integral membrane protein. Only two of these regions centered at amino acids 180 and 210, contain sequences of greater than 15 amino acids without a charged amino acid

region. The interspersion of hydrophobic and hydrophilic regions in the N-terminal region as noted by Heinemann & Ozols (1982) is reasonably well conserved. Interestingly the conservation of hydrophobicity does not correlate closely to conservation of amino acid sequence so that conservative substitutions are frequent in some of these regions. Several of these hydrophobic regions are bounded by hydrophilic regions and thus may represent sequences that transverse the membrane. The first 100 N-terminal amino acids and the amino acids from 190 to 250 and from 350 to 400 contain the most variability and probably are not involved in these common functions. If these proteins have different substrate specificities, one of these variable regions may be important in altering the specificity of the enzyme.

Analysis of the structure of rat P-450 genes, which are inducible by phenobarbital, has indicated that these genes are members of a family of more than six genes (Mizukami et al., 1983). In the rabbit, at least two forms of cytochrome P-450 that are inducible by phenobarbital have been described (Aoyama et al., 1981). Sequence analysis of the major form, P-450 (form 2), revealed no sequence heterogeneity (Heinemann & Ozols, 1983). Both these observations suggest the phenobarbital-inducible gene family may be smaller than that of the rat. The demonstration in this paper that at least three additional members of this gene family exist suggests that the gene families in the rat and the rabbit may be similar in size.

Acknowledgments

We thank Olke Uhlenbeck for the gift of T4 polynucleotide kinase, David Shapiro for advice and bacterial strains used in the cDNA cloning, and Linda Paul for technical assistance during the initial stages of this work.

Registry No. Phenobarbital, 50-06-6; cytochrome P-450, 9035-51-2.

References

Adesnik, M., Bar-Nun, S., Maschio, F., Zunich, M., Lippman, A., & Bard, E. (1981) *J. Biol. Chem. 256*, 10340–10345.

Aoyama, T., Imai, Y., & Sato, R. (1982) *Microsomes, Drug Oxidation, and Drug Toxicity* (Sato, R., & Kato, R., Eds.) pp 83–86, Japan Scientific Societies Press, Tokyo, and Wiley-Interscience, New York.

Black, S. D., Tarr, G. E., & Coon, M. J. (1982) *J. Biol. Chem. 257*, 14616–14619.

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

Colbert, R. A., Bresnick, E., Levin, D. E., & Thomas, P. E. (1979) *Biochem. Biophys. Res. Commun. 91*, 886–891.

Dilella, A. G., Chiang, J. Y. L., Hutterer, F., & Steggles, A. W. (1981) *Biochim. Biophys. Acta 652*, 223–227.

Dorner, A. J., & Kemper, B. (1978) *Biochemistry 17*, 5550–5555.

Dubois, R. N., & Waterman, M. R. (1979) *Biochem. Biophys. Res. Commun. 90*, 150–157.

Fitch, W. M., & Langley, C. H. (1976) *Fed. Proc., Fed. Am. Soc. Exp. Biol. 35*, 2092–2097.

Fujii-Kuriyama, Y., Mizukami, Y., Kawajiri, K., Sogawa, K., & Muramatsu, M. (1982) *Proc. Natl. Acad. Sci. U.S.A. 79*, 2793–2797.

Gonzales, F. J., & Kasper, C. B. (1982) *J. Biol. Chem. 257*, 5962–5968.

Gordon, D. F., & Kemper, B. (1980) *Nucleic Acids Res. 8*, 5669–5683.

Gotoh, O., Tagashira, Y., Iizuska, T., & Fujii-Kuriyama, Y. (1983) *J. Biochem. (Tokyo) 93*, 807–817.

Grantham, T., Gautier, C., & Gouy, M. (1980) *Nucleic Acids Res. 8*, 1893–1911.

Grunstein, M., & Hogness, D. S. (1975) *Proc. Natl. Acad. Sci. U.S.A. 72*, 3961–3965.

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

Haniu, M., Armes, L. G., Tanaka, M., Yasunobu, K. T., Shastry, B. S., Wagner, G., & Gunsalus, I. C. (1982) *Biochem. Biophys. Res. Commun. 105*, 889–894.

Haugen, D. A., & Coon, M. J. (1976) *J. Biol. Chem. 251*, 7929–7939.

Heinemann, F. S., & Ozols, J. (1982) *J. Biol. Chem. 257*, 14988–14999.

Heinemann, F. S., & Ozols, J. (1983) *J. Biol. Chem. 258*, 4195–4201.

Holmes, D. S., & Quigley, M. (1981) *Anal. Biochem. 114*, 193–197.

Jay, E., Seth, A. K., Rommens, J., Sood, A., & Jay, G. (1982) *Nucleic Acids Res. 10*, 6319–6329.

Koop, D. R., Morgan, E. T., Tarr, G. E., & Coon, M. J. (1982) *J. Biol. Chem. 257*, 8472–8480.

Kumar, A., & Padmanaban, G. (1980) *J. Biol. Chem. 255*, 522–525.

Levitt, M. (1976) *J. Mol. Biol. 104*, 59–107.

Lizardi, P., & Engelberg, A. (1979) *Anal. Biochem. 98*, 116–122.

Mandel, M., & Higa, A. (1970) *J. Mol. Biol. 53*, 159–162.

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

McCarty, K. S., Jr., Vollmer, R. T., & McCarty, K. S. (1974) *Anal. Biochem. 61*, 165–183.

Mizukami, Y., Fuji-Kuriyama, Y., & Muramatsu, M. (1983) *Biochemistry 22*, 1223–1229.

Ozols, J., Heinemann, F. S., & Johnson, E. F. (1981) *J. Biol. Chem. 256*, 11405–11408.

Parnes, J. R., Velan, B., Felsenfeld, A., Ramanathan, L., Ferini, U., Appella, E., & Seidman, J. G. (1981) *Proc. Natl. Acad. Sci. U.S.A. 78*, 2253–2257.

Perler, F., Efstratiadis, A., Lomedico, P., Gilbert, W., Kolodner, R., & Dodgson, J. (1980) *Cell (Cambridge, Mass.) 20*, 555–566.

Proudfoot, N. J., & Brownlee, G. G. (1976) *Nature (London) 263*, 211–214.

Sato, R., & Kato, R., Eds. (1982) *Microsomes, Drug Oxidations, and Drug Toxicity*, Japan Scientific Societies Press, Tokyo, and Wiley-Interscience, New York.

Stolarsky, L., & Kemper, B. (1978) *J. Biol. Chem. 253*, 7194–7201.

Treadwell, B. V., Mankin, D. P., Persson, A. V., Coon, M. J., & Mankin, H. J. (1980) *Biochemistry, Biophysics and Regulation of Cytochrome P-450* (Gustafsson, J.-A, Carlstedt-Duke, J., Mode, A., & Raften, J., Eds.) pp 389–396, Elsevier/North-Holland Biomedical Press, Amsterdam.

Tukey, R. H., Nebert, D. W., & Negishi, M. (1981) *J. Biol. Chem. 256*, 6969–6974.

Tukey, R. H., Hannah, R. R., Negishi, M., Nebert, D. W., & Eisen, H. J. (1982) *Cell (Cambridge Mass.) 31*, 275–284.

Walz, F. G., Jr., Vlasuk, G. P., Omiecinski, C. J., Bresnick, E., Thomas, P. E., Ryan, D. E., & Levin, W. (1982) *J. Biol. Chem. 257*, 4023–4026.

Wilson, A. C., Carlson, S. S., & White, T. J. (1977) *Annu. Rev. Biochem. 46*, 573–639.

Weaver, C. A., Gordon, D. F., & Kemper, B. (1982) *Mol. Cell. Endocrinol. 28*, 411–424.

Yuan, P.-M., Ryan, D. E., Levin, W., & Shively, J. E. (1983) *Proc. Natl. Acad. Sci. U.S.A. 80*, 1169–1173.