Nebert, D. W., Adesnik, M., Coon, M. J., Estabrook, R. W., Gonzalez, F. J., Guengerich, F. P., Gunsalus, I. C., Johnson, E. F., Kemper, B., Levin, W., Phillips, I. R., Sato, R., & Waterman, M. R. (1987) *DNA 6*, 1-11.
Okayama, H., & Berg, P. (1982) *Mol. Cell. Biol. 2*, 161-170.
Palmiter, R. D. (1974) *Biochemistry 13*, 3606-3615.
Poulos, T. L., Finzel, B. C., Gunsalus, I. C., Wagner, G. C., & Kraut, J. (1985) *J. Biol. Chem. 260*, 16122-16130.
Ryan, D. E., Thomas, P. E., Koreniwski, D., & Levin, W. (1979) *J. Biol. Chem. 254*, 1365-1374.
Ryan, D. E., Thomas, P. E., & Levin, W. (1982) *Arch. Biochem. Biophys. 216*, 272-288.
Sanger, F., Coulson, A. R., Barrell, B. G., Smith, A. J., & Roe, B. A. (1980) *J. Mol. Biol. 143*, 161-178.
Sato, R., Aoyama, T., & Imai, Y. (1982) in *Oxygenases and Oxygen Metabolism* (Ishimura, Y., Yamamoto, S., & Nozaki, M., Eds.) pp 321-332, Academic, New York.
Southern, E. (1975) *J. Mol. Biol. 98*, 503-518.
Suwa, Y., Mizukami, Y., Sogawa, K., & Fujii-Kuriyama, Y. (1985) *J. Biol. Chem. 260*, 7980-7984.
Taniguchi, H., Imai, Y., Iyanagi, T., & Sato, R. (1979) *Biochim. Biophys. Acta 550*, 341-355.
Taniguchi, T., Ohno, S., Fujii-Kuriyama, Y., & Muramatsu, M. (1980) *Gene 10*, 11-15.
Tarr, G. E., Black, S. D., Fujita, V. S., & Coon, M. J. (1983) *Proc. Natl. Acad. Sci. U.S.A. 80*, 6552-6556.
Thomas, P. S. (1980) *Proc. Natl. Acad. Sci. U.S.A. 77*, 5201-5205.
Tsunasawa, S., Kondo, J., & Sakiyama, F. (1985) *J. Biochem. (Tokyo) 97*, 701-704.
Van der Hoeven, T. A., Haugen, D. A., & Coon, M. J. (1974) *Biochem. Biophys. Res. Commun. 60*, 569-675.
Vlasuk, W., & Walz, F. G., Jr. (1982) *Biochemistry 21*, 789-793.

Comparison of Primary Structures Deduced from cDNA Nucleotide Sequences for Various Forms of Liver Microsomal Cytochrome P-450 from Phenobarbital-Treated Rabbits†

Yoshio Imai,* Masayuki Komori,‡ and Ryo Sato
Institute for Protein Research, Osaka University, Suita, Osaka 565, Japan
Received April 13, 1987; Revised Manuscript Received July 30, 1987

ABSTRACT: cDNA clones, termed pHP2, b32-3, b43, and b43-1, encoding cytochromes P-450 that are expressed in the liver of phenobarbital- (PB-) treated rabbits were isolated, and their nucleotide sequences were determined. pHP2 cDNA contains an open reading frame for a 490-residue protein and is a full-length counterpart of pP-450PBc2 [Leighton, J. K., Debrunner-Vossbrinck, B. A., & Kemper, B. (1984) *Biochemistry 23*, 204-210]. The b32-3 insert has a sequence for a protein whose primary structure is 91% similar to that of progesterone 21-hydroxylase P-450 1, though this cDNA lacks the sequence encoding the amino-terminal 110 residues. The overlapping clones b43 and b43-1 together encode an ethanol-inducible form of cytochrome P-450, though the amino-terminal five or more residues are missing in the composite b43/b43-1 sequence. Northern blot analysis showed that the b43/b43-1 protein is more strongly inducible by polycyclic aromatic hydrocarbons and isosafrole than by PB, in contrast to the case of the HP2 and b32-3 proteins. A comparison of the primary structures of eight forms of cytochrome P-450, including the HP2, b32-3, and b43/b43-1 proteins, that are expressed in the liver of PB-treated rabbits showed that 149 out of 487-492 amino acid residues are conserved in these cytochromes P-450. The eight forms can be assigned to three rabbit cytochrome P-450 gene subfamilies, P450IIB, P450IIC, and P450IIE. It was also shown that the members of the rabbit P450IIC subfamily can be further classified into three subgroups on the basis of their sequence similarity.

In attempts at elucidating the biological significance of the multiplicity of liver microsomal cytochrome P-450 at the DNA level, a number of cDNA and genomic clones for different forms of cytochrome P-450 have been isolated and their structures analyzed [e.g., Fujii-Kuriyama et al. (1982), Mizukami et al. (1983), Kawajiri et al. (1984), Sogawa et al. (1984, 1985), Gonzalez et al. (1984, 1985, 1986a,b), Leighton et al. (1984), Suwa et al. (1985), Tukey et al. (1985), Govind et al. (1986), Song et al. (1986), Hardwick et al. (1987), Imai (1987), and Kagawa et al. (1987)]. These studies have provided evidence that there are several distinct cytochrome P-450 gene families, such as polycyclic aromatic hydrocarbon inducible, phenobarbital- (PB-)[1] inducible, steroid-inducible, and peroxisome proliferator inducible ones. Nebert et al. (1987) have recently proposed a classification of the cytochrome P-450 gene superfamily, comprising not only liver cytochromes P-450 but also those from all other sources, into families and subfamilies and recommended a nomenclature for them.

In the preceding paper (Komori et al., 1988), we reported that microheterogeneity exists in major PB-inducible forms of rabbit liver cytochrome P-450, which belong to the rabbit P450IIB gene subfamily. Leighton et al. (1984) and Govind et al. (1986) have isolated cDNA and genomic clones, respectively, encoding rabbit liver cytochromes P-450 that are distinct from major PB-inducible ones but expressed in the

† This work was supported in part by a grant-in-aid (5806002) from the Ministry of Education, Science and Culture of Japan.
* Author to whom correspondence should be addressed.
‡ Present address: Faculty of Pharmaceutical Sciences, Hokkaido University, Sapporo 060, Japan.

[1] Abbreviations: PB, phenobarbital; SDS, sodium dodecyl sulfate.

0006-2960/88/0427-0080$01.50/0

liver of PB-treated rabbits and analyzed their structures. Imai (1987) has also obtained a rabbit liver cDNA clone, pHP3, coding for a similar cytochrome P-450, expressed it in yeast cells, and purified its protein product. This paper reports the isolation of one full-length cDNA clone, termed pHP2, and three clones of shorter insert lengths, called b32-3, b43, and b43-1, all encoding cytochromes P-450 other than major PB-inducible ones but that occur in the liver of PB-treated rabbits. The primary structures deduced from their nucleotide sequences were compared with those of five forms of cytochrome P-450 expressed in the liver of PB-treated rabbits. It was thus found that they can be assigned to three rabbit cytochrome P-450 gene subfamilies, P450IIB, P450IIC, and P450IIE, and that members of the P450IIC subfamily can further be classified into three subgroups.

Materials and Methods

Materials. Livers of rabbits (2.5–3.5 kg) that had received i.p. injections of one of the following drugs (dissolved in olive oil) were kindly supplied by Dr. N. Kagawa of this laboratory: 3-methylcholanthrene (daily dose, 20 mg/kg of body weight) for 2 consecutive days; β-naphthoflavone (70 mg/kg) twice at a 3-day intervals; isosafrole (100 mg/kg) for 7 consecutive days. All the other materials, including a liver from a PB-treated rabbit and the enriched mRNA used for cDNA synthesis, were described in the preceding paper (Komori et al., 1988).

Construction and Screening of cDNA Libraries. The construction of two cDNA libraries by methods I and II with enriched mRNA from liver microsomes of a PB-treated rabbit as template was described in the preceding paper (Komori et al., 1988). In this study a third cDNA library was prepared from the same enriched mRNA by method III. In this method, the "small primer", obtained from pSV7186 as a by-product in preparation of the vector primer of Okayama and Berg (1982), was used as a primer to synthesize the first strand of cDNA by reverse transcriptase. The RNA–DNA hybrid thus obtained was oligo(dC)-tailed and annealed with *Pst*I-cut, oligo(dG)-tailed pBR322. The RNA strand was then replaced by DNA with the aid of the repair system of Okayama and Berg (1982). The recombinant plasmids were used to transform *Escherichia coli* DH1 to obtain the third library. Screening of this library was performed as described in the preceding paper (Komori et al., 1988) with a 390-bp *Bgl*II fragment of pcP-450pb4, a cDNA clone for rat P-450b (Fujii-Kuiryama et al., 1982), as a probe, and washing was made twice with 2× SSC and 0.2% sodium dodecyl sarcosinate at 55 °C for 30 min. The cDNA library constructed by method II was also screened under more stringent conditions (0.1× SSC and 0.1% sodium dodecyl sarcosinate at 65 °C for 30 min) with a 190-bp *Eco*RI/*Hin*cII fragment of b32 or a 220-bp *Fok*I/*Pst*I fragment of b43 as a probe. b32 and b43 are cytochrome P-450 cDNA clones isolated in the preceding paper (Komori et al., 1988) from the library constructed by method II.

Other Methods. Total RNA used for RNA blot analysis was prepared from livers of untreated or variously drug-treated rabbits by the guanidine thiosulfate/CsCl method (Raymond & Shore, 1979). All the other procedures, including Southern and RNA blot analyses and nucleotide sequencing, were carried out as described in the preceding paper (Komori et al., 1988).

Results and Discussion

cDNA Clones Isolated and Examined. A cDNA library, constructed by method III (see Materials and Methods) with enriched mRNA from liver microsomes of a PB-treated rabbit as template, was screened for clones carrying cytochrome P-450 cDNAs by in situ colony hybridization under relatively mild washing conditions. A 390-bp *Bgl*II fragment of pcP-450pb4 (a cDNA clone for rat P-450b) containing the highly conserved HR2 region was used as a probe. Among the 12 positive clones thus selected, pHP2 hybridized with the probe only weakly and was found to contain an insert of about 2 kbp long. In the preceding paper (Komori et al., 1988), we reported two cDNA clones, b32 and b43, that hybridized with the pcP-450pb4 probe weakly. In this study the cDNA library prepared by method II was rescreened under more stringent conditions with a 190-bp *Eco*RI/*Hin*cII fragment (containing the HR2 region) of b32 as a probe, resulting in the isolation of pHP2-1 and b32-3. The insert of pHP2-1 exhibited the same restriction map as pHP2 (Figure 1), but the former was longer by about 500 bp toward the 3′-direction. The restriction map of b32-3 cDNA differed clearly from those of pHP2 (Figure 1) and pHP3 (Imai, 1987), but a 400-bp *Pst*I/*Eco*RI fragment of b32-3 hybridized with an HR1-containing fragment of pHP3 (Imai, 1987) on Southern blot analysis. The insert of b43, which gave only a faint hybridization signal with the pcP-450pb4 probe, showed a restriction map that is clearly different from those of pHP2 and bp32-3 (Figure 1). Rescreening of the cDNA library constructed by method II under highly stringent conditions with a 220-bp *Fok*I/*Pst*I fragment (containing the HR1 region) of b43 as a probe led to the isolation of several clones. One of them, called b43-1, exhibited the same restriction map as b43 in their overlapping region and was extended furthest toward the 5′-direction (Figure 1). The inserts of b43 and b43-1 together could cover a length of about 1.9 kbp. In this study, structural analysis was made on five cDNAs contained in pHP2, pHP2-1, b32-3, b43, and b43-1, the restriction maps of which (except for pHP2-1) are shown in Figure 1.

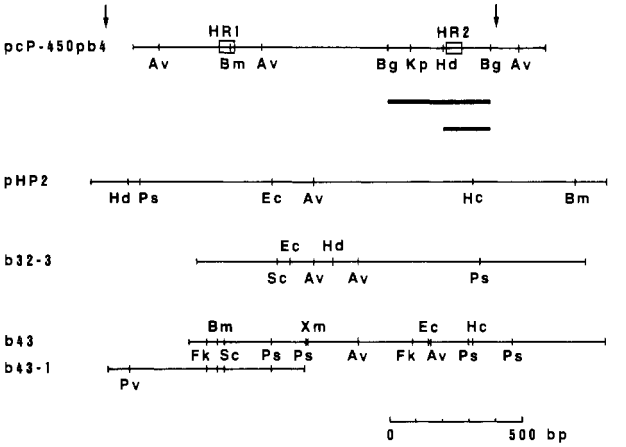

FIGURE 1: Restriction enzyme maps of cDNAs cloned and analyzed in this study (pHP2, b32-3, b43, and b43-1). The map of pcP-450pb4, a cDNA clone for rat P-450b, is also shown at the top. The *Bgl*II fragment and *Hin*dIII/*Bgl*II fragment of pcP-450pb4, used as probes for colony hybridization, are indicated by thick bars. The highly conserved regions, HR1 (left) and HR2 (right), are shown by open boxes. The positions of both initiation and termination codons are indicated by vertical arrows. Av, *Ava*I; Bg, *Bgl*II; Bm, *Bam*HI; Ec, *Eco*RI; Fk, *Fok*I; Hc, *Hin*cII; Hd, *Hin*dIII; Kp, *Kpn*I; Sc, *Sac*I; Ps, *Pst*I; Pv, *Pvu*II; Xm, *Xma*I.

Sequence Analysis of pHP2 and pHP2-1. Figure 2 shows the strategy adopted for nucleotide sequencing of the pHP2 insert. The complete nucleotide sequence thus determined and the amino acid sequence deduced therefrom are shown in Figure 3. pHP2 cDNA (1947 bp) consists of a 58-bp 5′-flanking sequence and a 416-bp 3′-noncoding region in addition

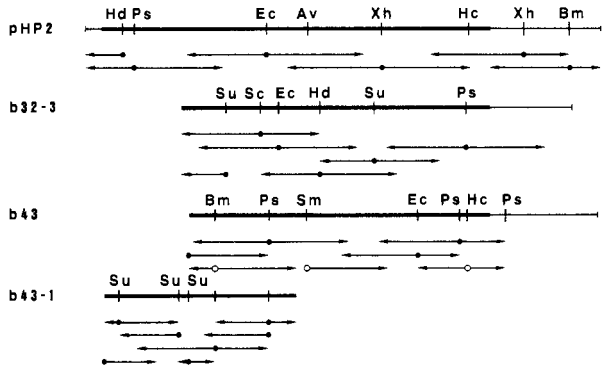

FIGURE 2: Sequencing strategies for cytochrome P-450 cDNAs (pHP2, b32-3, b43, and b43-1). Arrows with a closed circle indicate fragments that were subcloned into either M13mp10 or M13mp11 and seqeunced by the dideoxy chain termination method in the direction and extents shown by arrows. Fragments sequenced by the chemical cleavage method are indicated by arrows with an open circle. Su, *Sau*3A; Sm, *Sma*I; Xh, *Xho*I. The abbreviations for the other restriction enzymes are shown in Figure 1.

to an open reading frame encoding a polypeptide containing 490 amino acid residues. A molecular weight of 55 804 was calculated for the encoded polypeptide. The nucleotide sequence determined for the pHP2-1 insert was in perfect match with that of pHP2 in their overlapping region (data not shown). The pHP2-1 insert has a 17-bp 5′-noncoding region and a 3′-noncoding sequence that is about 500 bp longer than that of pHP2. A poly(A) sequence is seen at the 3′ end of pHP2 cDNA, which also contains two possible poly(A) addition signals, AATAGA and AATTAA (boxed in Figure 3). The occurrence of two or more poly(A) addition signals in cytochrome P-450 cDNAs has been reported (Leighton et al., 1984; Yabusaki et al., 1984). The nucleotide sequence determined for pHP2 cDNA is virtually identical with that reported for pP-450PBc2 (Leighton et al., 1984), except for two nucleotide changes at nucleotide 1470 (T to C) leading to the conversion of Leu to Pro and at nucleotide 1780 (C to T) in the 3′-noncoding region. The sequence upstream from nucleotide 90, which is missing in pP-450PBc2 cDNA, is again identical with the corresponding region in the P-450PBc2 gene cloned and sequenced by Govind et al. (1986). Although these workers have suggested that the most likely start site of P-450PBc2 mRNA is 23 bp upstream from the initiator ATG codon, our cDNA possesses a 58-bp 5′-noncoding region, indicating that the mRNA can start at least 35 bp upstream from the suggested start site. The finding that the 3′-noncoding region of pHP2-1 cDNA is longer than that of pHP2 cDNA, even though the latter contains a poly(A) sequence, suggests that the mRNA can be polyadenylated at two or more different sites. The existence of two or more mRNAs of different lengths encoding the same form of cytochrome P-450 has been reported (Tukey et al., 1986).

The deduced primary structure for the HP2 protein (encoded by pHP2 cDNA) is 81% similar to that of HP3 (Imai, 1987); 399 out of 490 residues are identical (underlined in Figure 3). Amino acid differences between these two proteins are unevenly distributed, and several highly variable regions are noticed, especially the regions spanning residues 93–111, 214–249, and 464–486, which are located in exons 2, 4, and 9, respectively, of the P-450PBc2 gene (Govind et al., 1986), in contrast to the finding that the differences between rat P-450b and P-450e occur most frequently in exon 7 (Suwa et al., 1985). The primary structures of HP2, which is virtually identical with P-450PBc2, and HP3 are about 85% similar to that of the protein (P-450PBc1) encoded by pP-450PBc1, another cDNA clone for a rabbit cytochrome P-450 (Leighton et al., 1984). A notable fact is that in these three proteins the amino-terminal 30-residue segment is highly conserved, even though this segment is usually not well conserved in the other microsomal cytochromes P-450, as will be discussed later in this paper. It is evident that the three proteins, HP2, HP3, and P-450PBc1, are closely related to one another.

Sequence Analysis of b32-3. The nucleotide sequence of the b32-3 insert, determined by adopting the strategy depicted in Figure 2, and the primary structure deduced therefrom are shown in Figure 4. This cDNA does not cover the entire sequence of a cytochrome P-450 and lacks the sequence encoding the amino-terminal portion, though it contains both HR1 (nucleotides 95-154) and HR2 (nucleotides 947–1006) sequences. The deduced primary structure (378 residues) is 91% similar to that of the segment from residue 111 to the carboxyl-terminus of rabbit liver progesterone 21-hydroxylase cytochrome P-450 1, a cDNA for which has been studied by Tukey et al. (1985). It is evident that the b32-3 insert encodes a cytochrome P-450 closely related to the 21-hydroxylase, although the cDNA obtained in this study lacks the stretch coding for the amino-terminal 110-residue segment. The similarity between the b32-3 protein and the 21-hydroxylase on one hand and PH2, PH3, and P-450PBc1 on the other is 72–76%, and the sequences of the former proteins have a three-residue deletion at residues 276–278 of the latter sequences.

Sequence Analysis of b43 and b43-1. As mentioned earlier, the inserts of b43 and b43-1 show the same restriction map in their overlapping region and together cover a cDNA of about 1.9 kbp long. When their nucleotide sequences were determined according to the strategy shown in Figure 2, it was actually found that they have exactly the same sequence in their overlapping region. The two sequences were, therefore, combined, and the composite nucleotide sequence (b43/b43-1 sequence) is shown in Figure 5, together with the primary structure deduced therefrom. From a sequence comparison with other microsomal cytochromes P-450, it is reasonable to assume that the b43/b43-1 sequence lacks a stretch encoding the amino-terminal several residues. In fact, residues 1–20 of the b43/b43-1 sequence are identical with residues 6–25 of rabbit P-450 LM3a (form 3a), an ethanol-inducible form, the amino-terminal 25 residues of which have been protein chemically sequenced (Koop et al., 1982). Moreover, the b43/b43-1 sequence is 83% similar to that of rat P-450j (Song et al., 1986), which is thought to be the rat counterpart of P-450 LM3a. After we reported the b43/b43-1 sequence preliminary [for an abstract, see Komori et al. (1986)], Khani et al. (1987) reported a sequence analysis of P-450 LM3a cDNA. The complete primary structure they constructed from the nucleotide sequences of two overlapping cDNAs and the protein chemically determined amino-terminal 25 residues (Koop et al., 1982) is in perfect match with that of the b43/b43-1 protein. The nucleotide sequences are also identical with each other except for one silent substitution (C to T at nucleotide 1338 in the b43/b43-1 sequence). It can be concluded that the b43/b43-1 protein is identical with P-450 LM3a. When compared with the primary structures of HP2, HP3, and P-450PBc1, there is one deletion at residue 105 and one more residue (Ser) is added at the carboxyl-terminus of the b43/b43-1 sequence. The sequence similarity between these two sets of proteins is about 55%. The primary structure of the b43/b43-1 protein, on the other hand, is about 50% similar to those of LM2 (Tarr et al., 1983) and HP1 (Komori et al., 1988), but its similarity to polycyclic aromatic hydro-

```
                                                                                                                    AGGAGTGTTATAAAAGCCTGAGCTTGCAGCTTCCAGTTGTCAGGAAGAGAAGGCTCCC

                                                                10                                                                  20
ATG GAT CTG GTG GTA GTG CTG GGG CTC TGT CTC TCC TGT TTG CTT CTC CTT TCA CTC TGG AAA CAG AGC CAT GGG
Met Asp Leu Val Val Val Leu Gly Leu Cys Leu Ser Cys Leu Leu Leu Leu Ser Leu Trp Lys Gln Ser His Gly

                    30                                                                  40                                                                  50
GGA GGG AAG CTT CCT CCT GGC CCC ACT CCT TTT CCA ATT CTT GGA AAC GTC CTG CAG TTA GAT TTT AAG GAC TTA
Gly Gly Lys Leu Pro Pro Gly Pro Thr Pro Phe Pro Ile Leu Gly Asn Val Leu Gln Leu Asp Phe Lys Asp Leu

                                                                60                                                                  70
AGC AAA TCT TTA ACC AAT CTG TCA AAA GTC TAT GGG CCC GTG TTC ACT GTG TAT CTG GGC ATG AAG CCC ACT GTG
Ser Lys Ser Leu Thr Asn Leu Ser Lys Val Tyr Gly Pro Val Phe Thr Val Tyr Leu Gly Met Lys Pro Thr Val

                    80                                                                  90                                                                  100
GTG GTG CAT GGG TAT GAA GCA GTG AAG GAG GCC CTA GTC GAT CTT GGA CAC GAG CTT TCT GGA AGA AGC CGT TTC
Val Val His Gly Tyr Glu Ala Val Lys Glu Ala Leu Val Asp Leu Gly His Glu Leu Ser Gly Arg Ser Arg Phe

                                                                110                                                                 120
CTA GTG ACT GCA AAA CTT AAT AAA GGA TTT GGA GTC ATT TTC AGC AAT GGA AAG AGA TGG ACG GAG ACC CGG CGC
Leu Val Thr Ala Lys Leu Asn Lys Gly Phe Gly Val Ile Phe Ser Asn Gly Lys Arg Trp Thr Glu Thr Arg Arg

                    130                                                                 140                                                                 150
TTC TCA CTC ATG ACC CTG CGG AAT TTC GGA ATG GGG AAG AGG AGC ATT GAG GAA CGA GTT CAA GAA GAA GCC CAC
Phe Ser Leu Met Thr Leu Arg Asn Phe Gly Met Gly Lys Arg Ser Ile Glu Glu Arg Val Gln Glu Glu Ala His

                                                                160                                                                 170
TGC CTG GTG GAG GAG CTG AGA AAA ACC AAT GCC TCA CCC TGT GAT CCC ACC TTT ATC CTG GGT GCT GCT CCC TGC
Cys Leu Val Glu Glu Leu Arg Lys Thr Asn Ala Ser Pro Cys Asp Pro Thr Phe Ile Leu Gly Ala Ala Pro Cys

                    180                                                                 190                                                                 200
AAT GTG ATC TGC TCC GTG ATT TTC CAG AAC CGT TTT GAT TAC ACG GAC CAG GAT TTT CTT AGT TTG ATG GGA AAG
Asn Val Ile Cys Ser Val Ile Phe Gln Asn Arg Phe Asp Tyr Thr Asp Gln Asp Phe Leu Ser Leu Met Gly Lys


                                                                210                                                                 220
TTC AAT GAA AAC TTC AAG ATT CTG AAT TCC CCC TGG GTA CAG TTC TGC AAT TGT TTT CCT ATT CTC TTC GAT TAT
Phe Asn Glu Asn Phe Lys Ile Leu Asn Ser Pro Trp Val Gln Phe Cys Asn Cys Phe Pro Ile Leu Phe Asp Tyr

                    230                                                                 240                                                                 250
TTC CCT GGG AGT CAT AGG AAA GCA GTT AAA AAT ATT TTT TAT GTG AAG AAT TAT ATT ACA GAG CAA ATA AAG GAA
Phe Pro Gly Ser His Arg Lys Ala Val Lys Asn Ile Phe Tyr Val Lys Asn Tyr Ile Thr Glu Gln Ile Lys Glu

                                                                260                                                                 270
CAC CAA AAA TCC CTG GAC ATT AAC AAT CCT CGG GAC TTC ATT GAT TGT TTC CTG ATC AAA ATG GAA CAG GAA AAG
His Gln Lys Ser Leu Asp Ile Asn Asn Pro Arg Asp Phe Ile Asp Cys Phe Leu Ile Lys Met Glu Gln Glu Lys

                    280                                                                 290                                                                 300
TGT AAT CAA CAG TCT GAA TTT ACT ATT GAA AAC TTA CTG ACC ACA GTG AGT GAT GTG TTT ATG GCT GGA ACA GAG
Cys Asn Gln Gln Ser Glu Phe Thr Ile Glu Asn Leu Leu Thr Thr Val Ser Asp Val Phe Met Ala Gly Thr Glu

                                                                310                                                                 320
ACA ACA AGC ACC ACC CTG AGG TAT GGA CTT CTA CTC CTG ATG AAG CAC CCA GAA GTC ATA GCT AAA GTG CAG GAG
Thr Thr Ser Thr Thr Leu Arg Tyr Gly Leu Leu Leu Leu Met Lys His Pro Glu Val Ile Ala Lys Val Gln Glu

                    330                                                                 340                                                                 350
GAG ATC GAG CGT GTG ATT GGC CGA CAC CGG AGC CCC TGC ATG CAG GAC AGG AGC CGC ATG CCC TAC ACG GAT GCC
Glu Ile Glu Arg Val Ile Gly Arg His Arg Ser Pro Cys Met Gln Asp Arg Ser Arg Met Pro Tyr Thr Asp Ala

                                                                360                                                                 370
ACG GTG CAC GAG ATC CAG AGA TAC ATT AAC CTC ATC CCC AAC AAT GTG CCC CAT ACA ACA ATC TGT AAC CTT AAG
Thr Val His Glu Ile Gln Arg Tyr Ile Asn Leu Ile Pro Asn Asn Val Pro His Thr Thr Ile Cys Asn Leu Lys

                    380                                                                 390                                                                 400
TTC AGA AAC TAT CTC ATC CCC AAG GGC ACA GAT GTA CTA ACA TCA CTG TCT TCT GTA CTG CAT GAC GAC AAA GAG
Phe Arg Asn Tyr Leu Ile Pro Lys Gly Thr Asp Val Leu Thr Ser Leu Ser Ser Val Leu His Asp Asp Lys Glu

                                                                410                                                                 420
TTC CCC AAC CCA GAC AGG TTT GAC CCT GGC CAC TTC TTG GAT GCC AGC GGC AAC TTT AGG AAA AGT GAC TAC TTC
Phe Pro Asn Pro Asp Arg Phe Asp Pro Gly His Phe Leu Asp Ala Ser Gly Asn Phe Arg Lys Ser Asp Tyr Phe

                    430                                                                 440                                                                 450
ATG CCT TTC TCA ACA GGA AAA CGA GTG TGT GTG GGA GAG GCC CTG GCC CGC ATG GAG CTG TTT CTG TTC CTG ACT
Met Pro Phe Ser Thr Gly Lys Arg Val Cys Val Gly Glu Ala Leu Ala Arg Met Glu Leu Phe Leu Phe Leu Thr

                                                                460                                                                 470 *
GCC ATT TTA CAG AAC TTT ACC CCG AAA CCT CTG GTC AAC CCA AAC AAT GTT GAT GAA AAT CCA TTC TCC AGT GGA
Ala Ile Leu Gln Asn Phe Thr Pro Lys Pro Leu Val Asn Pro Asn Asn Val Asp Glu Asn Pro Phe Ser Ser Gly

                    480                                                                 490
ATT GTC CGT GTG CCA CCC TTG TAC CGG GTC AGC TTC ATT CCT GTC TGA GGAAGGTCAGGTCACGGGATGCCATGCTTACATCT
Ile Val Arg Val Pro Pro Leu Tyr Arg Val Ser Phe Ile Pro Val

GCAATTCCCCTTCCTCCAGGACACTCGCCAACTGTTTCCCCCTGTCATGGGGGCCTGCTCTGACCTGGCCTCTGACATTTCCTCATTGCGCAAGATCCA

TTCTCCATTCGGTGGGAGTCACCTGGGTCCCTTCACAAATGTATGTTTGCTATCACCACACTTAATACTCTTGGCCTGACCACCACATGGGCTTATACT
                *
TGTATAATACTATGTTATATGCTGTCACTGTAAATTTGGGCAAGATGATTGACATTGGACAGTTTGGATCCGTATGTCCTCTGCATGCTCTAAATAGAA

GCATTATTAATTGCTGAAATCAGTTCTCAAGTTTTCTTCTTTGTACATAATTTGAGTAAAAATTAAAGGAAACAGATTCCCAAG(A)n
```

FIGURE 3: Nucleotide sequence of pHP2 cDNA and the amino acid sequence deduced therefrom. The amino acid sequence is numbered from the initiator methionine. The boxed cysteinyl residue is thought to serve as the fifth ligand to the heme iron. The boxed sequences in the 3′-noncoding region are potential poly(A) addition signals. Asterisks show nucleotide differences in pP-450PBc2 (Leighton et al., 1984). Identical amino acid residues between the HP2 and HP3 proteins are underlined.

```
                                                                                            (G)n A
                                    10                                      20
CTC GGA ATC CTT TTC AGC AAT GCA AAC ACA TGG AAG GAG ATG CGG CGC TTC TCG CTC ATG ACC CTG CGG AAT TTC
Leu Gly Ile Leu Phe Ser Asn Ala Asn Thr Trp Lys Glu Met Arg Arg Phe Ser Leu Met Thr Leu Arg Asn Phe

                30                                      40                                      50
GGG ATG GGG AAG AGA AGC ATT GAG GAC CGA GTT CAA GAG GAG GCC CGC TGC CTG GTG GAG GAG CTG AGA AAA ACC
Gly Met Gly Lys Arg Ser Ile Glu Asp Arg Val Gln Glu Glu Ala Arg Cys Leu Val Glu Glu Leu Arg Lys Thr

                                    60                                      70
AAT GCC TCA CCC TGT GAT CCC ACC TTT ATC CTG GGC TGT GCT CCC TGC AAT GTG ATC TGC TCC ATT ATT TTC CAT
Asn Ala Ser Pro Cys Asp Pro Thr Phe Ile Leu Gly Cys Ala Pro Cys Asn Val Ile Cys Ser Ile Ile Phe His

                80                                      90                                      100
AAT CGC TTT GAT TAT AAA GAT GAG CAT TTT CTT AAA CTG ATG GAA AAA TTC AAT GAA AAT GTT AGG ATT CTG AGC
Asn Arg Phe Asp Tyr Lys Asp Glu His Phe Leu Lys Leu Met Glu Lys Phe Asn Glu Asn Val Arg Ile Leu Ser

                                    110                                     120
TCT CCA TGG TTG CAG ATT TGC AAT AAT TTC CCT GTT CTC ACT GAC TAT TTA CCG GGA ATT CAT AAC ACC TTA GTA
Ser Pro Trp Leu Gln Ile Cys Asn Asn Phe Pro Val Leu Thr Asp Tyr Leu Pro Gly Ile His Asn Thr Leu Val

                130                                     140                                     150
AAA AAT ATT GAA TAT ACA AAA AAT TTT ATT ATG GAG AAA GTG AAA GAA CAC CAA AAG TCT CTG GAT GTT AAC AAC
Lys Asn Ile Glu Tyr Thr Lys Asn Phe Ile Met Glu Lys Val Lys Glu His Gln Lys Ser Leu Asp Val Asn Asn

                                    160                                     170
CCT CGG GAC TTT ATT GAT TGC TTC TTG ATC AAA ATG GAT CAA GAA AAC CAT TTG GAG TTC ACT CTT GAA AGC TTG
Pro Arg Asp Phe Ile Asp Cys Phe Leu Ile Lys Met Asp Gln Glu Asn His Leu Glu Phe Thr Leu Glu Ser Leu

                180                                     190                                     200
GTA ACC ACT GTG TCT GAT TTG TTT GGG GCT GGG ACT GAG ACA ACG AGC ACA ACG CTG AGC ATA TCC CTC CTG CTC
Val Thr Thr Val Ser Asp Leu Phe Gly Ala Gly Thr Glu Thr Thr Ser Thr Thr Leu Ser Ile Ser Leu Leu Leu

                                    210                                     220
CTG CTG AAG CAC CCC GAG GTC GCA GCT AAA GTG CAG GAG GAG ATT GAG CGT GTG ATT GGC AGG CAC CGG AGC CCC
Leu Leu Lys His Pro Glu Val Ala Ala Lys Val Gln Glu Glu Ile Glu Arg Val Ile Gly Arg His Arg Ser Pro

                230                                     240                                     250
TGC ATG CAG GAC AGG AGC CGC ATG CCC TAC ACG GAT GCT GTG ATA CAT GAG ATC CAG AGA TAC ATT GAC CTC ATT
Cys Met Gln Asp Arg Ser Arg Met Pro Tyr Thr Asp Ala Val Ile His Glu Ile Gln Arg Tyr Ile Asp Leu Ile

                                    260                                     270
CCC ATT AAC CTG CCC CAT GCA GTG ACC AGA GAC ATT AAA TTC AGA AAC TAC TTT ATC CCC AAG GGA ATG AAC ATC
Pro Ile Asn Leu Pro His Ala Val Thr Arg Asp Ile Lys Phe Arg Asn Tyr Phe Ile Pro Lys Gly Met Asn Ile

                280                                     290                                     300
ATA ACA TCA CTG ACA TCT GTG TTA CAT GAT GAA AAG GAA TTT CCT AAC CCA AAG GTA TTT GAC CCT GGA CAC TTT
Ile Thr Ser Leu Thr Ser Val Leu His Asp Glu Lys Glu Phe Pro Asn Pro Lys Val Phe Asp Pro Gly His Phe

                                    310                                     320
CTG GAT GAG AGT GGC AAC TTC AAG AAG AGT GAC TAC TTC ATG CCT TTC TCA GCA GGA AAA AGG ATG TGT GTG GGA
Leu Asp Glu Ser Gly Asn Phe Lys Lys Ser Asp Tyr Phe Met Pro Phe Ser Ala Gly Lys Arg Met Cys Val Gly

                330                                     340                                     350
GAG GGC CTG GCC CGC ATG GAG CTG TTT TTG TTC CTG ACC TCC ATT TTG CAG AAC TTT AAA CTG CAA TCT CTG GTT
Glu Gly Leu Ala Arg Met Glu Leu Phe Leu Phe Leu Thr Ser Ile Leu Gln Asn Phe Lys Leu Gln Ser Leu Val

                                    360                                     370
GAA CCA AAG GAC CTG GAC ATC ACT GCA GTT GTC AAT GGA TTC GCT TCT GTG CCA CCT GCC TAC CAG CTC TGT TTC
Glu Pro Lys Asp Leu Asp Ile Thr Ala Val Val Asn Gly Phe Ala Ser Val Pro Pro Ala Tyr Gln Leu Cys Phe

        378
AGT CCT GTT TGA AAGGGGACAAACTGGCTGCCGCTGTGCTGTTGTCTGCCATCA----------
Ser Pro Val
```

FIGURE 4: Nucleotide sequence of b32-3 cDNA and the amino acid sequence deduced therefrom. This sequence lacks the amino-terminal 110 residues. The heme-binding cysteinyl residue is boxed. Identical amino acid residues between the b32-3 protein and progesterone 21-hydroxylase P-450 1 are underlined.

Table I: Similarity of Amino Acid Sequences among Eight Forms of Microsomal Cytochrome P-450 That Are Expressed in the Liver of PB-Treated Rabbits[a]

form	P-450 1	b32-3	PBc1	HP2	HP3	form 3b	b43/b43-1	HP1
HP1	51	52	51	50	50	49	49	
b43/b43-1	57	58	56	54	55	55		
form 3b	67	69	65	66	67			
HP3	76	75	84	81				
HP2	72	74	85					
PBc1	75	74						
b32-3	91							
P-450 1								

[a] Numbers indicate percentages of matched residues. For calculation the sequence data reported for HP1 (Komori et al., 1987), form 3b (Ozols et al., 1985), HP3 (Imai, 1987), P-450PBc1 (Leighton et al., 1984), and P-450 1 (Tukey et al., 1985) were used in addition to those reported in this paper for HP2, b32-3, and b43/b43-1.

carbon inducible forms of rabbit liver cytochrome P-450 (Ozols, 1986; Kagawa et al., 1987) is only 20–25%.

Effects of Drug Treatments on Cytochrome P-450 mRNA Levels in Liver. Total RNA was isolated from the liver of an untreated or variously drug-treated rabbit, and the levels of mRNAs coding for the HP1, HP2, and b43/b43-1 proteins were examined by RNA blot analysis. The probes used were a *Xma*I fragment of pHP1 cDNA [see Figure 1 in Komori

```
                                                                              (G)n
                              10                                      20
ACC GTC GCC CTG CTG GGG TGG ATG GTC ATC CTC CTG TTC ATA TCC GTC TGG AAG CAG ATC CAC AGC AGC TGG AAC
Thr Val Ala Leu Leu Gly Trp Met Val Ile Leu Leu Phe Ile Ser Val Trp Lys Gln Ile His Ser Ser Trp Asn

            30                                      40                                      50
CTG CCC CCA GGA CCT TTC CCA CTG CCC ATC ATC GGG AAT CTT CTC CAG TTG GAT TTG AAG GAT ATT CCC AAG TCC
Leu Pro Pro Gly Pro Phe Pro Leu Pro Ile Ile Gly Asn Leu Leu Gln Leu Asp Leu Lys Asp Ile Pro Lys Ser

                                          60                                      70
TTT GGC AGG CTG GCA GAG CGC TTT GGG CCG GTG TTC ACT GTG TAC CTG GGC TCC AGG CGT GTT GTG GTT CTG CAC
Phe Gly Arg Leu Ala Glu Arg Phe Gly Pro Val Phe Thr Val Tyr Leu Gly Ser Arg Arg Val Val Val Leu His

            80                                      90                                      100
GGC TAC AAG GCG GTG AGG GAG ATG CTG TTG AAC CAC AAG AAC GAG TTC TCT GGG CGT GGC GAG ATC CCT GCT TTC
Gly Tyr Lys Ala Val Arg Glu Met Leu Leu Asn His Lys Asn Glu Phe Ser Gly Arg Gly Glu Ile Pro Ala Phe

                                          110                                     120
CGG GAG TTT AAG GAC AAG GGG ATC ATT TTC AAC AAT GGA CCC ACC TGG AAG GAC ACT CGG CGG TTC TCC CTG ACC
Arg Glu Phe Lys Asp Lys Gly Ile Ile Phe Asn Asn Gly Pro Thr Trp Lys Asp Thr Arg Arg Phe Ser Leu Thr

            130                                     140                                     150
ACC CTC CGG GAC TAT GGG ATG GGG AAA CAG GGC AAC GAG GAC CGG ATC CAG AAG GAG GCC CAC TTC CTG CTG GAG
Thr Leu Arg Asp Tyr Gly Met Gly Lys Gln Gly Asn Glu Asp Arg Ile Gln Lys Glu Ala His Phe Leu Leu Glu

                                          160                                     170
GAG CTC AGG AAG ACC CAG GGC CAG CCC TTC GAC CCC ACC TTT GTC ATC GGC TGC ACA CCC TTC AAC GTC ATC GCC
Glu Leu Arg Lys Thr Gln Gly Gln Pro Phe Asp Pro Thr Phe Val Ile Gly Cys Thr Pro Phe Asn Val Ile Ala

            180                                     190                                     200
AAA ATC CTC TTC AAT GAC CGC TTT GAC TAT AAG GAC AAG CAG GCT CTG AGG CTG ATG AGT TTG TTC AAC GAG AAC
Lys Ile Leu Phe Asn Asp Arg Phe Asp Tyr Lys Asp Lys Gln Ala Leu Arg Leu Met Ser Leu Phe Asn Glu Asn

                                          210                                     220
TTC TAC CTG CTC AGT ACT CCT TGG CTG CAG GTT TAC AAT AAT TTT TCA AAC TAT CTA CAG TAC ATG CCT GGA AGT
Phe Tyr Leu Leu Ser Thr Pro Trp Leu Gln Val Tyr Asn Asn Phe Ser Asn Tyr Leu Gln Tyr Met Pro Gly Ser

            230                                     240                                     250
CAC AGG AAA GTA ATA AAA AAT GTG TCT GAA ATA AAA GAG TAC ACA CTC GCA AGA GTG AAG GAG CAC CAC AAG TCG
His Arg Lys Val Ile Lys Asn Val Ser Glu Ile Lys Glu Tyr Thr Leu Ala Arg Val Lys Glu His His Lys Ser

                                          260                                     270
CTG GAC CCC AGC TGC CCC CGG GAC TTC ATT GAC AGC CTG CTC ATA GAA ATG GAG AAG GAC AAA CAC AGC ACG GAG
Leu Asp Pro Ser Cys Pro Arg Asp Phe Ile Asp Ser Leu Leu Ile Glu Met Glu Lys Asp Lys His Ser Thr Glu

            280                                     290                                     300
CCC CTG TAC ACG CTG GAA AAC ATT GCT GTG ACT GTG GCG GAC ATG TTC TTT GCG GGC ACG GAG ACC ACC AGC ACC
Pro Leu Tyr Thr Leu Glu Asn Ile Ala Val Thr Val Ala Asp Met Phe Phe Ala Gly Thr Glu Thr Thr Ser Thr

                                          310                                     320
ACG CTG CGA TAT GGG CTC CTG ATC CTG CTG AAG CAC CCC GAG ATC GAA GAG AAA CTT CAT GAA GAA ATC GAC AGG
Thr Leu Arg Tyr Gly Leu Leu Ile Leu Leu Lys His Pro Glu Ile Glu Glu Lys Leu His Glu Glu Ile Asp Arg

            330                                     340                                     350
GTG ATT GGG CCG AGC CGA ATG CCT TCT GTC AGG GAC AGG GTG CAG ATG CCC TAC ATG GAC GCT GTG GTA CAT GAG
Val Ile Gly Pro Ser Arg Met Pro Ser Val Arg Asp Arg Val Gln Met Pro Tyr Met Asp Ala Val Val His Glu

                                          360                                     370
ATT CAG CGA TTC ATC GAT CTC GTG CCC TCC AAT CTG CCG CAC GAA GCC ACA CGG GAC ACC ACC TTC CAA GGA TAC
Ile Gln Arg Phe Ile Asp Leu Val Pro Ser Asn Leu Pro His Glu Ala Thr Arg Asp Thr Thr Phe Gln Gly Tyr

            380                                     390                                     400
GTC ATC CCC AAG GGC ACT GTT GTA ATC CCG ACT CTG GAC TCC CTT TTG TAT GAC AAG CAA GAA TTC CCT GAT CCC
Val Ile Pro Lys Gly Thr Val Val Ile Pro Thr Leu Asp Ser Leu Leu Tyr Asp Lys Gln Glu Phe Pro Asp Pro

                                          410                                     420
GAG AAG TTC AAA CCA GAG CAC TTT CTG AAT GAG GAG GGG AAG TTC AAG TAT AGC GAC TAC TTC AAG CCG TTT TCC
Glu Lys Phe Lys Pro Glu His Phe Leu Asn Glu Glu Gly Lys Phe Lys Tyr Ser Asp Tyr Phe Lys Pro Phe Ser

            430                                     440                                     450
GCA GGA AAA CGC GTG [TGT] GTT GGA GAA GGC CTG GCT CGC ATG GAG TTG TTT CTG CTC CTG TCC GCC ATT CTG CAG
Ala Gly Lys Arg Val [Cys] Val Gly Glu Gly Leu Ala Arg Met Glu Leu Phe Leu Leu Leu Ser Ala Ile Leu Gln

                                          460                                     470
CAT TTT AAC CTC AAG CCT CTC GTT GAC CCA GAG GAC ATT GAC CTT CGC AAT ATT ACG GTG GGC TTT GGC CGT GTC
His Phe Asn Leu Lys Pro Leu Val Asp Pro Glu Asp Ile Asp Leu Arg Asn Ile Thr Val Gly Phe Gly Arg Val

            480                         487
CCA CCA CGC TAC AAA CTC TGT GTC ATT CCC CGC TCG TAA ACCCAAGGGCAGCACCCAGAGGCCACTCTTCTCCTCGAGTGCCCCTG
Pro Pro Arg Tyr Lys Leu Cys Val Ile Pro Arg Ser

GGGAGGCTCCTGCTGCAG----------
```

FIGURE 5: Composite nucleotide sequence of two overlapping cDNAs, b43 and b43-1, and the amino acid sequence deduced therefrom. This primary structure lacks the amino-terminal five or more residues. The heme-binding cysteinyl residue is boxed. Two highly conserved regions, HR1 (upper) and HR2 (lower), are underlined.

et al. (1988)], a *Hin*dIII/*Pst*I fragment of pHP2 cDNA (see Figure 1), and a *Pst*I/*Eco*RI fragment of b43 cDNA (see Figure 1) for HP1, HP2, and b43/b43-1 mRNAs, respectively. These probes were ^{32}P-labeled by nick translation, and hybridization was visualized by autoradiography. As shown in Figure 6, the pHP1 and pHP2 probes hybridized most strongly with the RNA from the PB-treated liver, followed by that from the isosafrole-treated one. On of the other hand, the RNAs from untreated and 3-methylcholanthrene- and β-naphthoflavone-treated livers gave only very weak hybridization signals.

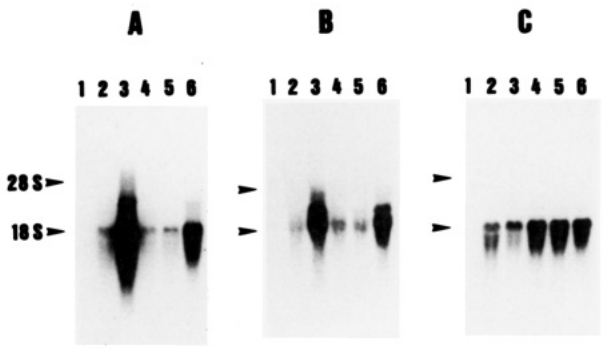

FIGURE 6: RNA blot analysis of mRNAs from livers of untreated and variously drug-treated rabbits. Total RNAs prepared from five types of liver were subjected to agarose gel electrophoresis under denaturing conditions (formaldehyde-containing gels) and then transferred to nitrocellulose filters. The blotted RNAs were hybridized with a ^{32}P-labeled probe from pHP1 (panel A), pHP2 (panel B), or b43 (panel C), and hybridized mRNAs were visualized by autoradiography. In each panel: (lane 1) calf liver ribosomal RNA as size markers (indicated by arrows); (lanes 2–6) RNAs from untreated and PB-, 3-methylcholanthrene-, β-naphthoflavone-, and isosafrole-treated rabbit livers, respectively.

On the contrary, the b43 probe hybridized more strongly with the RNAs from 3-methylcholanthrene-, β-naphthoflavone- and isosafrole-treated rabbits than with those from untreated and PB-treated animals. These results indicate that the b43/b43-1 gene is quite different from the HP1 and HP2 genes in its drug-mediated expression, even though the sequence similarity among their protein products falls in the range of 49–54% (Table I).

Comparison of Primary Structures of Cytochromes P-450. In this and two preceding papers (Imai, 1987; Komori et al., 1988), we predicted the primary strutures of five forms of cytochrome P-450 that are expressed in the liver of PB-treated rabbits, i.e., HP1, HP2, HP3, b32-3, and b43/b43-1, although the sequences of the latter two are still incomplete. In addition, P-450PBc1 (Leighton et al., 1984), P-450 1 (progesterone 21-hydroxylase) (Tukey et al., 1985), and P-450 form 3b, which has been protein chemically sequenced (Ozols et al., 1985) and is identical with P-450PBc3 (Leighton et al., 1984), also seem to be expressed in the liver of PB-treated rabbits. These eight forms of cytochrome P-450 are composed of 487–492 amino acid residues[2] and thus differ from polycyclic aromatic hydrocarbon inducible forms of rabbit liver cytochrome P-450 consisting of 516–518 residues (Kagawa et al., 1987). The primary structures of these eight forms of cytochrome P-450 are aligned in Figure 7, in which numbers denote the residue numbers of the HP1 sequence. All the sequences, except for that of PH1, have a deletion of one amino acid, Gly, at position 22. As mentioned earlier, a three-residue deletion is seen in the sequences of b32-3 and P-450 1 (residues 276–278), whereas a deletion of one amino acid at position 105 and an addition of serine at the carboxyl-terminus are noticed in the b43/b43-1 sequence. In the eight forms, 149 out of 491 residues (30%) are conserved. Of these 149 conserved residues, 69 can also be found at the corresponding positions in the sequences of two polycyclic aromatic hydrocarbon inducible forms of rabbit liver cytochrome P-450 (asterisked in Figure 7) (Kagawa et al., 1987). A notable fact is that in the eight forms examined highly conserved and variable regions are unevenly distributed along the sequence. For instance, in the amino-terminal 30-residue segment only one residue, Leu-6, is conserved. It is, however, to be noted that in all the eight sequences residues 2 is acidic, residue 21 is basic, and the stretch between these two residues consists of only uncharged residues, many of which are hydrophobic. A study in our laboratory using a cDNA for a polycyclic hydrocarbon inducible rabbit liver cytochrome P-450 has shown that the amino-terminal short segment (<29 residues) functions not only as a signal required for cotranslational insertion of the cytochrome into the microsomal membrane but also as a stop-transfer sequence that halts the translocation of the rest of the polypeptide across the membrane (Sakaguchi et al., 1987). The occurrence of the highly conserved HR1 and HR2 regions (Gotoh et al., 1983) as well as the so-called "analogous peptides" (Ozols et al., 1981) in the eight sequences can be recognized. In addition to these regions, two highly conserved segments are also noticed, i.e., a proline cluster (Leu-Pro-Pro-Gly-Pro-X-Pro-X-Pro) at residues 30–38 and a threonine-serine cluster (Phe-X-Ala-Gly-Thr-X-Thr-Thr-Ser-X-Thr-Leu) at residues 296–307. Moreover, the latter cluster is followed by a hydrophobic pentapeptide (Phe-Leu-Leu-Leu-Leu in HP1 and Leu-Leu-Leu-Leu-Met in HP2, HP3, P-450PBc1, etc.) at residues 311–315. Similar proline and threonine–serine clusters can also be found at the corresponding positions in the sequences of two polycyclic aromatic hydrocarbon inducible rabbit liver cytochromes P-450 (Ozols, 1986; Kagawa et al., 1987) as well as in those of rat liver microsomal cytochromes P-450 so far sequenced (Fujii-Kuriyama et al., 1982; Kawajiri et al., 1984; Yabusaki et al., 1984; Gonazalez et al., 1986a; Song et al., 1986). It is reasonable to assume that these highly conserved regions are involved in functions that are common for all the cytochromes P-450, such as the interaction with NADPH–cytochrome P-450 reductase. In fact, the HR2 region has been shown to act as the heme-binding site (Gotoh et al., 1983; Poulos et al., 1985), and the threonine–serine cluster has been located in the region containing the distal helix just above the heme in *Pseudomonas putida* P-450cam (Poulos et al., 1985). In the eight sequences the regions spanning residues 89–120, 160–183, 195–251, and 270–295 as well as the carboxyl-terminal segment starting from residue 458 are especially variable. In these regions practically no sequence similarities can be detected between the polycyclic aromatic hydrocarbon inducible rabbit liver cytochromes P-450 on the one hand and the eight forms shown in Figure 7 on the other. In some of these regions, especially in the segment spanning residues 215–250, amino acid differences are frequently seen among the three closely related cytochromes P-450, HP2, HP3, and P-450PBc1. It is likely that some of these highly variable regions are responsible for functions that are specific for individual cytochromes P-450. It has been suggested that the segment spanning residues 165–191 is involved in substrate binding in liver microsomal cytochromes P-450 (Gotoh et al., 1983).

Grouping of Cytochromes P-450 on the Basis of Sequence Similarity. Table I summarizes the similarity in the primary structures of the eight forms of rabbit liver microsomal cytochrome P-450 discussed above. The fact that they are more than 49% similar to one another indicates that they all belong to the same cytochrome P-450 gene family, because it has recently been defined that in a given species any cytochrome P-450 protein in one gene family should be less than 36% similar to a protein in any of the other families (Nebert et al., 1987). From the results reported in the preceding paper (Komori et al., 1988), it is evident that the HP1 protein, which

[2] The primary structure reported by Khani et al. (1987) for P-450 LM3a, which is identical with the b43/b43-1 protein, consists of 492 amino acid residues, but its amino-terminal residue is not methionine but alanine. This indicates that its full-length cDNA codes for a polypeptide composed of 493 or more residues.

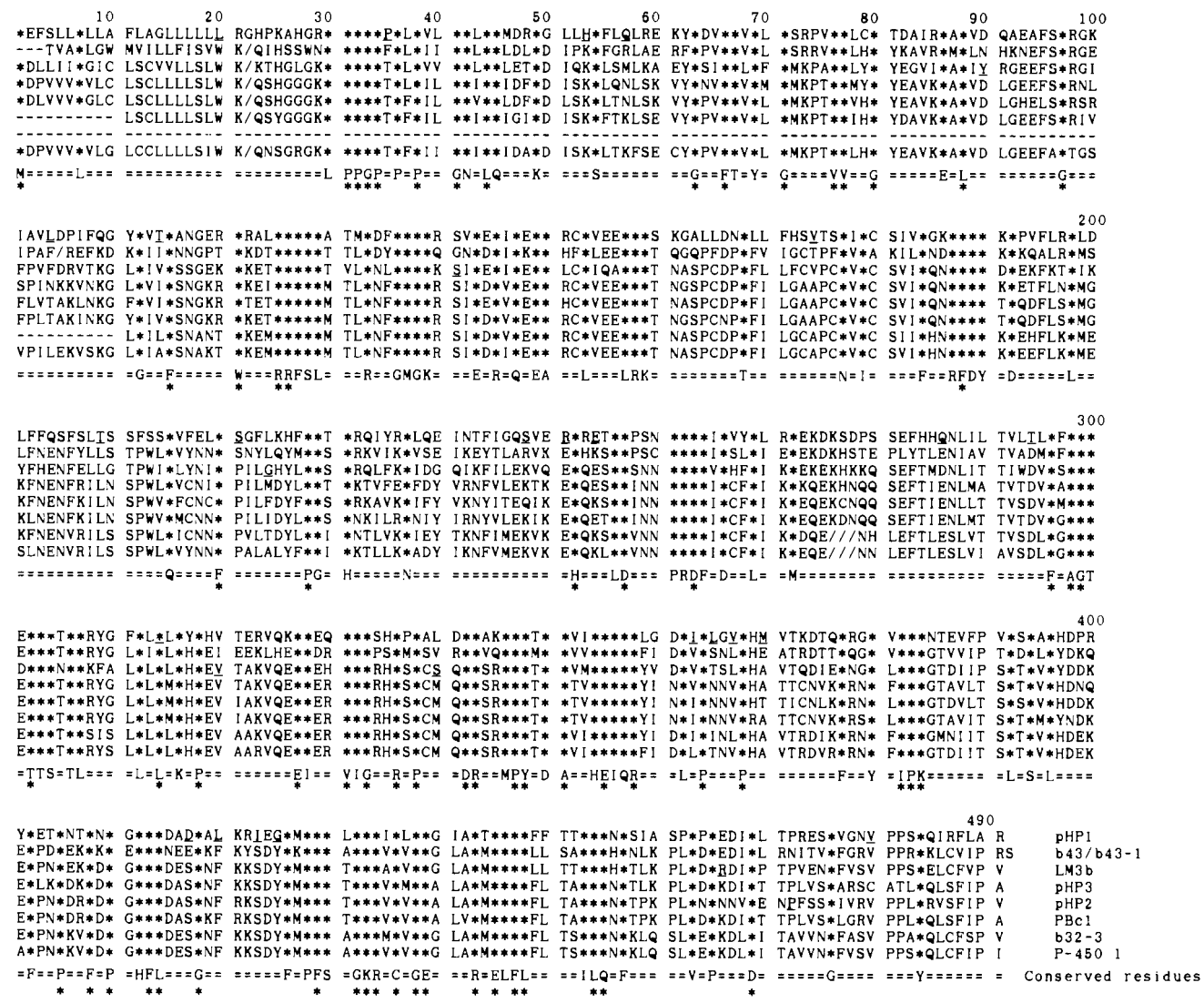

FIGURE 7: Comparison of amino acid sequences of eight forms of cytochrome P-450 that are expressed in the liver of PB-treated rabbits. Amino acids are shown in single-letter abbreviations, but those conserved throughout the eight forms are indicated by asterisks. These conserved residues are shown in single-letter abbreviations in the bottom line (conserved residues). Amino acids that are also conserved in the two polycyclic aromatic hydrocarbon inducible forms of rabbit liver cytochrome P-450 (Kagawa et al., 1987) are marked with asterisks below the bottom line. Slashes indicate the gaps inserted to maximize the sequence similarity. Dashes indicate the amino-terminal residues that could not be deduced because of the incomplete lengths of the cloned cDNAs. The sequences of HP1 (pHP1), form 3b (LM3b), HP3 (pHP3), P-450PBc1 (PBc1), and progesterone 21-hydroxylase P-450 1 (P-450 1) are from Komori et al. (1987), Ozols et al. (1985), Imai (1987), Leighton et al. (1984) and Tukey et al. (1985), respectively.

is a major PB-inducible form, is a member of the rabbit P450IIB gene subfamily according to the recommended nomenclature (Nebert et al., 1987). Since HP1 is only 49–52% similar to the other seven forms, the latter should be assigned to subfamilies other than P450IIB. The similarity of b43/b43–1, which is identical with P-450 LM3a (form 3a), to the other forms (except for HP1) is 54–58%, warranting that b43/b43-1 forms a separate group, which has been called the rabbit P450IIE subfamily (Nebert et al., 1987). The other six forms, i.e., form 3b, HP2, HP3, PBc1, b32-3, and P-450 1, are 65–91% similar to one another and thus seem to belong to the same subfamily. Actually, Nebert et al. (1987) have assigned PBc1, PBc2 (virtually identical with HP2), and form 3b (identical with PBc3) to the P450IIC subfamily. However, these six forms can be further classified into three subgroups. Form 3b is only 65–69% similar to the other five forms and thus can be separated from the others. As mentioned above, HP2, HP3, and PBc1 are closely related to one another and thus form another subgroup, in which the sequence similarity is 81–85%. A third subgroup comprises b32-3 and P-450 1, which are 91% similar to each other. It is, therefore, evident that the rabbit P450IIC subfamily consists of at least three clearly distinguishable subgroups, the sequence similarity among which is 65–76%. These findings, together with the microheterogeneity observed in the P450IIB subfamily (Komori et al., 1988), suggest that evolution of cytochrome P-450 genes, especially those belonging to the P450II family, is more complicated than previously expected.

ACKNOWLEDGMENTS

We thank Dr. Y. Fujii-Kuriyama of the Cancer Institute, Japanese Foundation for Cancer Research, Tokyo, for kindly supplying us with pcP-450pb4.

Registry No. pHP2 (DNA), 111769-76-7; pHP2 (peptide), 111769-74-5; cytochrome P-450, 9035-51-2; isosafrole, 120-58-1; phenobarbital, 50-06-6.

REFERENCES

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

Gonzalez, F. J., Mackenzie, P. I., Kimura, S., & Nebert, D. W. (1984) *Gene 29*, 281–292.

Gonzalez, F. J., Nebert, D. W., Hardwick, J. P., & Kasper, C. B. (1985) *J. Biol. Chem. 260*, 7435-7441.
Gonzalez, F. J., Song, B.-J., & Hardwick, J. P. (1986a) *Mol. Cell. Biol. 6*, 2969-2976.
Gonzalez, F. J., Kimura, S., Song, B.-J., Pastewka, J., Gelboin, H. V., & Hardwick, J. P. (1986b) *J. Biol. Chem. 261*, 10667-10672.
Gotoh, O., Tagashira, Y., Iizuka, T., & Fujii-Kuriyama, Y. (1983) *J. Biochem. (Tokyo) 93*, 807-817.
Govind, S., Bell, P. A., & Kemper, B. (1986) *DNA 5*, 371-382.
Hardwick, J. P., Song, B.-J., Huberman, E., & Gonzalez, F. J. (1987) *J. Biol. Chem. 262*, 801-810.
Imai, Y. (1987) *J. Biochem. (Tokyo) 101*, 1129-1139.
Kagawa, N., Mihara, K., & Sato, R. (1987) *J. Biochem. (Tokyo) 101*, 1471-1479.
Kawajiri, K., Gotoh, O., Sogawa, K., Tagashira, Y., Muramatsu, M., & Fujii-Kuriyama, Y. (1984) *Proc. Natl. Acad. Sci. U.S.A. 81*, 1649-1653.
Khani, S. C., Zaphiropoulos, P. G., Fujita, V. S., Portier, T. D., Koop, D. R., & Coon, M. J. (1987) *Proc. Natl. Acad. Sci. U.S.A. 84*, 638-642.
Komori, M., Imai, Y., & Sato, R. (1986) *Seikagaku 58*, 660 (in Japanese).
Komori, M., Imai, Y., Tsunasawa, S., & Sato, R. (1988) *Biochemistry* (preceding paper in this issue).
Koop, D. R., Morgan, E. T., Tarr, G. E., & Coon, M. J. (1982) *J. Biol. Chem. 257*, 8472-8480.
Leighton, J. K., Debrunner-Vossbrinck, B. A., & Kemper, B. (1984) *Biochemistry 23*, 204-210.
Mizukami, Y., Sogawa, K., Suwa, Y., Muramatsu, M., & Fujii-Kuriyama, Y. (1983) *Proc. Natl. Acad. Sci. U.S.A. 80*, 3958-3962.
Nebert, D. W., Adesnik, M., Coon, M. J., Estabrook, R. W., Gonzalez, F. J., Guengerich, F. P., Gunsalus, I. C., Johnson, E. F., Kemper, B., Levin, W., Phillips, I. R., Sato, R., & Waterman, M. R. (1987) *DNA 6*, 1-11.
Okayama, H., & Berg, P. (1982) *Mol. Cell. Biol. 2*, 161-170.
Ozols, J. (1986) *J. Biol. Chem. 261*, 3965-3979.
Ozols, J., Heinemann, F. S., & Johnson, E. F. (1981) *J. Biol. Chem. 256*, 11405-11408.
Ozols, J., Heinemann, F. S., & Johnson, E. F. (1985) *J. Biol. Chem. 260*, 5427-5434.
Poulos, T. L., Finzel, B. C., Gunsalus, I. C., Wagner, G. C., & Kraut, J. (1985) *J. Biol. Chem. 260*, 16122-16130.
Raymond, Y., & Shore, G. C. (1979) *J. Biol. Chem. 254*, 9335-9338.
Sakaguchi, M., Mihara, K., & Sato, R. (1987) *EMBO J. 6*, 2425-2431.
Sogawa, K., Gotoh, O., Kawajiri, K., & Fujii-Kuriyama, Y. (1984) *Proc. Natl. Acad. Sci. U.S.A. 81*, 5066-5070.
Sogawa, K., Gotoh, O., Kawajiri, K., Harada, T., & Fujii-Kuriyama, Y. (1985) *J. Biol. Chem. 260*, 5026-5032.
Song, B.-J., Gelboin, H. V., Park, S.-S., Yang, C. S., & Gonzalez, F. J. (1986) *J. Biol. Chem. 261*, 16689-16697.
Suwa, Y., Mizukami, Y., Sogawa, K., & Fujii-Kuriyama, Y. (1985) *J. Biol. Chem. 260*, 7980-7984.
Tarr, G. E., Black, S. D., Fujita, V. S., & Coon, M. J. (1983) *Proc. Natl. Acad. Sci. U.S.A. 80*, 6552-6556.
Tukey, R. H., Okino, S., Barnes, H., Griffin, K. J., & Johnson, E. F. (1985) *J. Biol. Chem. 260*, 13347-13354.
Tukey, R. H., Quattrochi, I. C., Pendurthi, U. R., & Okino, S. K. (1986) *Fed. Proc., Fed. Am. Soc. Exp. Biol. 45*, 1663.
Yabusaki, Y., Shimizu, M., Murakami, H., Nakamura, K., Oeda, K., & Ohkawa, H. (1984) *Nucleic Acids Res. 12*, 2929-2938.

Characterization of Individual Tryptophan Side Chains in Proteins Using Raman Spectroscopy and Hydrogen-Deuterium Exchange Kinetics†

Takashi Miura, Hideo Takeuchi, and Issei Harada*
Pharmaceutical Institute, Tohoku University, Aobayama, Sendai 980, Japan
Received May 14, 1987; Revised Manuscript Received September 8, 1987

ABSTRACT: Two Raman bands at 880 and 1360 cm^{-1} of tryptophan (Trp) side chains have been found useful in structural studies of the side chains in proteins. The frequency of the 880-cm^{-1} band reflects the strength of H bonding at the N_1H site of the indole ring: the lower the frequency is, the stronger the H bonding is. The intensity of the 1360-cm^{-1} band, on the other hand, is a marker of the hydrophobicity of the environment of the indole ring: particularly strong in hydrophobic environments. It is also demonstrated that a combination of stepwise deuteriation of the tryptophan side chains and difference spectrum techniques is useful to observe these marker bands due to each side chain separately. The states of six tryptophans in lysozyme revealed by this Raman spectroscopic method in solution are compared with those by X-ray diffraction in crystal. The Raman data on the outer four Trp's are consistent with the X-ray structure, whereas significant differences between solution and crystal are suggested for the strength of H bonding of the most and second most buried Trp's. Characterization of four Trp's in α-lactalbumin shows that the two outer Trp's are moderately H bonded to solvent water and closely surrounded by aliphatic side chains while the inner two are not H bonded nor closely surrounded by aliphatic side chains.

In order to understand the role of individual amino acid side chains in the function and structure of a protein, it is important to know the states of the side chains in addition to their locations in the protein. Raman spectroscopy has been a useful tool for such purposes. A vibrational mode localized at a side chain of a protein is often influenced by a specific environmental or structural factor (e.g., H bonding or conformation), and the corresponding Raman band reflects the state of the

† Work supported in part by a Grant-in-Aid for General Scientific Research (62430004) from the Ministry of Education, Science and Culture of Japan.

0006-2960/88/0427-0088$01.50/0