Gene Structure of a Major Form of Phenobarbital-inducible Cytochrome P-450 in Rat Liver*

Yorimasa Suwa†, Yuzuru Mizukami‡, Kazuhiro Sogawa, and Yoshiaki Fujii-Kuriyama‡

From the Department of Biochemistry, Cancer Institute, Japanese Foundation for Cancer Research, Toshima-ku, Tokyo 170, Japan

The gene structure of cytochrome P-450b, a major form of phenobarbital-inducible cytochrome P-450 in rat livers was elucidated by sequence analysis of the cloned genomic DNAs and was compared with the previously determined gene structures of cytochrome P-450e, a minor form of phenobarbital-inducible cytochrome P-450 and two forms of 3-methylcholanthrene-inducible cytochrome P-450 (P-450c and -d). The gene for cytochrome P-450b is 23 kilobase pairs (kb) long and is separated into 9 exons by 8 intervening sequences. This gene structure is very similar to that of cytochrome P-450e except for the first intron, the first intron being much longer in cytochrome P-450b gene (approximately 12 kb) than in cytochrome P-450e gene (3.2 kb), but differs greatly from the gene structures of two 3-methylcholanthrene-inducible cytochromes P-450 as pointed out previously (Sogawa, K., Gotoh, O., Kawajiri, K. & Fujii-Kuriyama, Y. (1984) Proc. Natl. Acad. Sci. U. S. A. 81, 5066–5070).

The nucleotide sequences in all 9 exons and their flanking regions in introns show very close homology between the two phenobarbital-inducible cytochrome P-450 genes. Forty base substitutions are found in approximately 1900 nucleotides of all exonic sequences, and 15 of them result in 14 amino acid replacements. These base substitutions occur in relatively limited regions of the gene sequences. Most of them are found in exons 6, 7, 8, and 9, most frequently in exon 7 as described previously (Mizukami, Y., Sogawa, K., Suwa, Y., Muramatsu, M. & Fujii-Kuriyama, Y. (1983) Proc. Natl. Acad. Sci. U. S. A. 80, 3958–3962). The close sequence homology between the two phenobarbital-inducible cytochrome P-450 genes is also found to extend to the promoter region with one notable exception. The simple repeated sequences of (CA), which is present at −254 position in cytochrome P-450e gene is also observed at the equivalent position in cytochrome P-450b gene, but the repetitiveness is greatly reduced in cytochrome P-450b gene ((CA)₅, for P-450b versus (CA)₁₀ for P-450e), and this may somehow be related to the difference in the level of cytochrome P-450b and P-450e in the inductive phase of phenobarbital administration.

Cytochrome P-450 plays an important role in the oxidative metabolism of a wide variety of structurally unrelated compounds of endogenous as well as exogenous origin (1–3). It has been shown that this versatility of the cytochrome system results from the presence of multiple forms of the hemoprotein, each of which exhibits different but rather broad substrate specificities. The syntheses of some of them are induced in specific manners by the administration of various kinds of drugs (1–3).

In livers of phenobarbital-treated rats, the amounts of at least two forms of cytochrome P-450 are remarkably increased, and they are purified to a homogenous state. These two forms of cytochrome P-450, designated P-450b and P-450e, show some differences in catalytic activities and electrophoretic mobility with each other, although they are immunologically indistinguishable (4).

Recent studies involving molecular cloning technology and protein chemistry have demonstrated that the primary structures of these two hemoproteins are very similar. There are only 14 amino acid replacements in the limited portion of the molecules which consist of 491 amino acids (5–7). Analyses of genetic cross experiments have indicated that the genes for cytochrome P-450b and P-450e are nonallelic and are closely linked to each other on the same autosome (8). Under variety of conditions, the syntheses of cytochrome P-450b and P-450e were always coinduced, and the level of the former was constantly 2 or more times greater than that of the latter (8, 9). It would be of interest to study how these two forms of cytochrome P-450 are coinduced by the administration of phenobarbital and many other drugs such as γ-chlordare, SKF525A (9, 10), and 1,1-di(P-chlorophenyl)-2,2-dichloroethylene (11) and how their levels in the hepatocytes are regulated.

We have previously reported the gene structure of cytochrome P-450e, a minor form of phenobarbital-inducible cytochrome P-450(7). In the present paper, we report the gene structure of a major form of phenobarbital-inducible cytochrome P-450, P-450b, and compare its gene structure with that of cytochrome P-450e.

EXPERIMENTAL PROCEDURES

Materials — Restriction endonucleases were obtained from Takara Shuzo (Kyoto, Japan), Bethesda Research Laboratories, and New England Biolabs (Beverly, MA). Escherichia coli DNA polymerase I and polynucleotide kinase were purchased from Boehringer (Mannheim, Federal Republic of Germany) and Takaka Shusyo, respectively. T4 DNA ligase and bacterial alkaline phosphatase were from Takara Shuzo. [γ-³²P]ATP (5000 Ci/mmol), [α-³²P]dATP (5000 Ci/mmol),

7980
cloned genomic DNA. The result of the sequence analyses of the corresponding EcoRII/HindIII region in the P-450b cDNA contained most frequently base substitutions from the cytochrome P-450e gene. We considered that this cloned genomic DNA (clone 14) was similar but clearly different from those of cytochrome P-450b and P-450e gene and the sequencing strategy. The arrow above each cloned region show the direction and length of the fragments sequenced by the procedure of Maxam and Gilbert (17). Closed boxes show exons sequences. Fragments I and V were subcloned from clone 31 and 22, respectively, and the other fragments (II, III, and IV) were from clone 14. A fragment (VI) of the promoter region of cytochrome P-450e gene was derived from clone 6. A, AccII; Ac, Accl; B, BamHI; Bg, BglII; E, EcoRI; He, HaeIII; H, HindIII; Hf, HinfI; Hp, HpaII; K, KpnI; P, PstI; S, SstI; S96, Sau96I; X, XbaI; Xh, Xhol.

RESULTS AND DISCUSSION

Isolation of Cytochrome P-450b Gene—When we cloned cytochrome P-450b gene (7), we isolated another clone (clone 14), which also hybridized strongly with the DNA probe (the whole insert of pcP-450-pb4, Ref. 5) from a rat HaeIII gene library. As shown in Fig. 1, the cloned genomic DNA contained the 2.5-kb BamHI/EcoRI and the 1.2-kb EcoRI/HindIII DNA fragments which were commonly found with cytochrome P-450e gene, but its detailed restriction cleavage maps in the other region were clearly different from those of cytochrome P-450e gene. We considered that this cloned genomic DNA (clone 14) was similar but clearly different from cytochrome P-450e gene (clone 11).

Since it is known that the 7th and 8th exons in the 1.2-kb EcoRI/HindIII fragment of cytochrome P-450e contained most frequently base substitutions from the cytochrome P-450b cDNA sequence (7), we determined the nucleotide sequence of the corresponding EcoRI/HindIII region in the cloned genomic DNA. The result of the sequence analyses showed that the sequences of this part of the cloned DNA conformed with those of pcP-450pb1 and -4 (cytochrome P-450b cDNA clones). We, therefore, concluded that the genomic DNA isolated here was the gene for cytochrome P-450b.
7982 Genotype for Major Form of Phenobarbital-inducible Cytochrome P-450

Restriction fragments containing the other exon sequences were identified by Southern blot analysis, using the labeled cDNA of cytochrome P-450b which appropriately fragmented by restriction enzymes, as probes. It was found that the cloned genomic DNA lacked the sequences for the first and the last 9 exons. Then, we searched for the genomic clone containing the missing parts of cytochrome P-450b gene in the isolated genomic DNA by screening rat gene library again and genomic DNA lacked the sequences for the first and the last were used as probes in the screenings. Two of them (clones TACCAGATCTGCTTCTCAGCTCGGTGATCCGGCTGAG~AGCCATGTGCCCCAGTTCTGTTGGG~TGGCCTCATGTTTCTGCCTCTGGGGGACCTGCT~C

gagtcttgtgcctttacttcacataatattattctaggttcctgtgttctacaggccacagtcacacacattc~------

GTGATCTGTGCTAATGGACTcTGTATATGGTCTCAGTGCTATGTCTACAGACTTACATAGTATGTATGGTTCAGGT~CAGMTCACAGAGTGTG~gcttcg

TATTTTGACATGTGACAGAGATTTCATGAGTCCACATCTCATGCTGAGTTACTTCCCTCTTCCTCCTMCAGCCCATGTCCCCAGTTA~AGCCCTCCAT~TCT

CAGGCTCCAAGGCCACTGCTCACATCTTCCTATTGCAGT

TTCTTCACCACCATCCTCCAGAACTTCTTCTCTGTGTC~~CAT~GGCTCCCAACGGACATTGACCTCACGCCCMGGAGAGTG~ATTG~TACCTC~CG

CATTGGCCATATTGTGGAGAAAAGCACAGGGCCACCTTA~CCCAACGCGCTCCAC~~CTTCATCGACACTTACCTTCTGCGCATGGAGAAGgtgdgtCCtgCatg

cctggggtgggagagaa-"

cctcttatcctgcctcctcatcctccagGAGCCCCACTGGATCCCACCTTCCTCTTCCAGTGCATCACAGC~CA~TCTGCTC

FIG. 3. Comparison of nucleotide and amino acid sequence of the cytochrome P-46Ob and P-4SOe gene. The deduced amino acid sequence and the nucleotide sequence for cytochrome P-450b are shown in the first and second line, respectively. For cytochrome P-450e gene, only nucleotide substitutions from those of the cytochrome P-450b gene are indicated in the third line, and the resulting amino acid replacements are shown in underlined. Asterisks show nucleotide substitutions from cytochrome P-450b cDNA (PCP-450bp4).
TABLE I
Summary of nucleotide and amino acid substitutions between
cytochrome P-450b and P-450e

| exon | substitutions in nucleotide | amino acid |
|------|----------------------------|------------|
|      |                            | P-450b     | P-450e     |
| 1    | 1 / 201                    | -          | -          |
| 2    | 3 / 163                    | -          | -          |
| 3    | 0 / 150                    | -          | -          |
| 4    | 0 / 161                    | -          | -          |
| 5    | 1 / 177                    | -          | -          |
| 6    | 3 / 122                    | Ser - Gly  | Ala - Thr  |
| 7    | 16 / 208                   | 6          | 6          |
| 8    | 9 / 142                    | 4          | 4          |
| 9    | 9 / 540                    | 2          | 2          |
| total| 40 / 1864                  | 14 / 491   |            |

22 and 41) contained the 9th exon.

Detailed restriction cleavage of these cloned DNAs confirmed that they contained partially overlapping regions with one another and altogether covered a stretch of approximately 40 kb of the chromosomal DNA. These results are represented in Fig. 1. The closed boxes indicate the localization of the exons which were first assigned by DNA blot analysis (data not shown) using the labeled cDNA as a probe and then were finally determined by sequence analysis of the genomic DNA. The general structure of cytochrome P-450b gene is very similar to cytochrome P-450e gene except for the length of the first intron. The first intron of cytochrome P-450b gene is much longer than that of cytochrome P-450e gene. This could result either from occurrence of insertion or deletion of large DNA fragment(s) in the first intron of phenobarbital-inducible cytochrome P-450 genes. Accordingly, the total length of cytochrome P-450b gene which spans from the putative transcription initiation site to the polyadenylation site comes to 12 kb, 9 kb longer than that of cytochrome P-450e gene. The gene is split into 9 exons by 8 intervening sequences.

Nucleotide Sequence of Cytochrome P-450b Gene—Then, we determined the nucleotide sequence of all 9 exons and their flanking sequences in the introns by using the method of Maxam and Gilbert (17). The strategy for sequencing is outlined in Fig. 2. The determined sequence is consistent with that of PCP-450bp1 and -4, cytochrome P-450b cDNA clones, except for the two neutral base substitutions at the third letter of the codons (codon 333 and 354). These substitutions could be more likely due to allelic polymorphism than sequencing errors, because they were confirmed by sequencing from different sites. Therefore, we concluded that the cloned genomic DNA, the sequence of which was determined here, is the gene for cytochrome P-450b. The result is shown in Fig. 3.

The sequence of cytochrome P-450b gene is very similar to that of cytochrome P-450e gene in the partial flanking sequences in introns as well as in the nine exonic sequences. The putative transcription-initiation site was assigned at an A, 30 base pairs upstream from the initiation codon ATG by analogy from cytochrome P-450e gene (7), and the polyadenylation site was determined by comparison with the nucleotide sequence of cytochrome P-450b cDNA (7). A typical poly(A) addition signal AATAAA was not found, but a possible equivalent sequence, GGTAAA, was present at the corresponding position of the gene, as is also the case with cytochrome P-450e gene (7). This unusual poly(A) addition signal is considered to function as such, because the identical sequence was actually present at the same position in the cDNA sequence.

The insertion site of all 8 introns in relation to the coding nucleotide is also rigidly conserved between the two genes, and all the splicing junctions satisfy the canonical GT/AG rule (19). All the results described above confirmed the previous conclusion that the genes for cytochrome P-450b and -e diverged in the relatively recent past, some 8 million years ago (20).

Only 40 base substitutions were found in all exon sequences of the two genes, and 15 of them resulted in 14 amino acids replacements. As summarized in Table I, these substitutions are not scattered evenly among 9 exonic sequences but show biased distribution in 6, 7, 8, and 9th exons, notably in exon 7 and 8. This nonrandom distribution of base substitutions

![Fig. 4. Comparison of the nucleotide sequence of the 5' flanking region of the cytochrome P-450b and P-450e genes. The nucleotide sequence for the cytochrome P-450e gene is shown in the first line. Only the nucleotide replacements in cytochrome P-450e gene are shown under the corresponding nucleotide of cytochrome P-450b gene. Dashes show deletion of nucleotides.](image-url)
may be generated from a gene conversion event as suggested by Affolter and Anderson (21).

From their genetic cross experiments using inbred and outbred rat strains/colonies, Rampersaud and Waltz (8) have reported that there are two closely linked loci encoding cytochrome P-450b and P-450e with at least 6 alleles at the cytochrome P-450b locus and at least 2 alleles at the P-450e locus. Simmons and Kasper (22) have also demonstrated that phenobarbital-inducible cytochrome P-450 genes are located in a cluster near the Coh locus on mouse chromosome 7. The close proximity of phenobarbital-inducible cytochrome P-450 genes to one another may be responsible for the increased levels of nonallelic genes and this may also explain the high degree of apparent polymorphism observed at the cytochrome P-450e and P-450b loci.

When the gene organization of the two phenobarbital-inducible cytochrome P-450s is compared with that of two 3-methylcholanthrene-inducible cytochromes P-450, P-450c, and P-450d, they look very different from each other as pointed out previously (22, 24). None of the insertion sites of 8 introns for phenobarbital-inducible cytochrome P-450 genes occur at the equivalent positions of 6 introns of 3-methylcholanthrene-inducible ones, although on the basis of the similarity of the amino acid sequence, all these forms of cytochrome P-450 are reasonably assumed to be evolved from a common ancestor (20, 23, 24). The implication of this different gene organization in the evolution of cytochrome P-450 gene family has been previously discussed (24, 25).

**Nucleotide Sequence of 5' Flanking Region of Cytochrome P-450b and P-450e Genes—**Many structurally unrelated compounds such as phenobarbital, γ-chlordane, SKF-525A, and trans-stilbene are known to induce remarkably the synthesis of cytochrome P-450e gene as described. Both genes have the same modified TATA sequence, CATAAAA, 20 bases upstream from the transcription-initiation site. The nucleotide sequence CACACA--- was found approximately 260 bases upstream from the transcription-initiation site in the two genes.

The chain length of the simple repeated sequence, however, is much longer in cytochrome P-450b gene than in cytochrome P-450e gene. Since the alternating purine-pyrimidine sequence has a potential to form the Z-DNA structure (28-32) which has been suggested to play a role in the regulation of gene expression (33), the difference in chain length of the repeated sequence between cytochrome P-450b and P-450e genes may be somehow related to the increased levels of the two proteins by the drug administration.

Recently, Guarente and Mason (33, 34) have reported that iso-1-cytochrome c gene was controlled by the 5' flanking region of the TATA box and was subjected to the down-regulation possibly due to the sequence containing an extensive run of alternating purine-pyrimidine residues. Deletion of this sequence increased the basal level of transcription to 2- to 3-fold (33, 34). These phenomena appear to be analogous to the expression of cytochrome P-450b and P-450e genes, in that the cytochrome P-450b gene with the very short alternating purine-pyrimidine stretch in its 5' flanking region was expressed at a constantly higher level than cytochrome P-450e gene with the much longer repeated sequence. But it remains to be seen whether the Z conformation of DNA region of the TATA box and was subjected to the down-regulation in the case of phenobarbital-inducible cytochrome P-450 genes as suggested with iso-1-cytochrome c gene. This problem will be tested experimentally.

Acknowledgments—We thank Drs. L. L. Jagodzinsky and J. Bonner for kindly supplying a rat gene library.

REFERENCES
1. Sato, S. & Omura, T. (eds) (1978) Cytochrome P-450, Kodansha, Tokyo and Academic Press, New York
2. Guengerich, F. P. (1976) Pharmacol. Ther. 6, 99-121
3. Lu, A. Y. H. & West, S. B. (1986) Pharmacol. Rev. 31, 227-295
4. Ryan, D. E., Thomas, P. E. & Levin, W. (1982) Arch. Biochem. Biophys. 216, 272-286
5. Fuji-Kuriyama, Y., Minzakami, Y., Kawajiri, K., Sogawa, K. & Muramatsu, M. (1982) Proc. Natl. Acad. Sci. U.S.A. 79, 2793-2797
6. Yuan, P-M., Ryan, D. E., Levin, W. & Shively, J. E. (1983) Proc. Natl. Acad. Sci. U.S.A. 80, 1169-1173
7. Minzakami, Y., Sogawa, K., Sawa, Y., Muramatsu, M. & Fuji-Kuriyama, Y. (1983) Proc. Natl. Acad. Sci. U.S.A. 80, 3958-3962
8. Rampersaud, A. & Waltz, F. G., Jr. (1983) Proc. Natl. Acad. Sci. U.S.A. 80, 8542-8546
9. Vlasuk, G. P., Ryan, D. E., Thomas, P. E., Levin, W. & Waltz, F. G., Jr. (1982) Biochemistry 21, 6298-6299
10. Thomas, P. E., Reik, L. M., Ryan, D. E. & Levin, W. (1981) J. Biol. Chem. 256, 1044-1052
11. Minzakami, Y., Yoshikawa, H., Sogawa, K., Fuji-Kuriyama, Y. & Omura, T. (1984) J. Biochem. (Tokyo) 95, 949-957
12. Benton, W. D. & David, R. W. (1977) Science 196, 180-182
13. Minzakami, Y., Fuji-Kuriyama, Y. & Muramatsu, M. (1983) Biochemistry 22, 1223-1229
14. Rudie, P. W. J. & Dieckmann, M., Rhodes, C. (1985) Mol. Biol. 113, 237-251
15. Blattner, R. R., Williams, G. E., Bled, A. E., Bubolz, D. A. & Smithies, O. (1977) Science 196, 161-169
16. Southern, E. (1970) J. Mol. Biol. 58, 503-517
17. Maxam, A. M. & Gilbert, W. (1977) Proc. Natl. Acad. Sci. U.S.A. 74, 566-1664
18. Yousaf, S., K., Carroll, A. R. & Clarke, B. E. (1984) Gene 37, 309-313
19. Brechta, H. & Champon, P. (1981) Annu. Rev. Biochem. 50, 340-383
20. Kawajiri, K., Gotoh, O., Sogawa, K., Tagashira, Y., Muramatsu, M. & Fuji-Kuriyama, Y. (1984) Proc. Natl. Acad. Sci. U.S.A. 81, 1649-1653
21. Affolter, M. & Anderson, A. (1984) Biochem. Biophys. Res. Commun. 118, 655-662
22. Simmon, D. L. & Kasper, C. B. (1983) J. Biol. Chem. 258, 9585-9588
23. Sogawa, K., Gotoh, O., Kawajiri, K. & Fuji-Kuriyama, Y. (1984) Proc. Natl. Acad. Sci. U.S.A. 81, 5066-5071
24. Sogawa, K., Gotoh, O., Kawajiri, K. & Harada, T. & Fuji-Fuyumia, Y. (1985) J. Biol. Chem. 260, 5026-5032
25. Adenamk, M., Bar-Nun, S., Maschico, F., Zinich, M., Lippman, A. & Bard, F. (1981) J. Biol. Chem. 256, 10340-10345
26. Gonzalez, F. J. & Kasper, C. B. (1982) J. Biol. Chem. 257, 959-968
27. Reich, J. P., Gonzalez, F. J. & Kasper, C. B. (1983) J. Biol. Chem. 258, 8081-8085
28. Wang, A. H-J., Quigley, G. J., Knapk, F. J., Crawford, J. L., Van Boon, J. H., van der Marel, G. & Rich, A. (1979) Nature 282, 680-686
29. Vorkeckoe, M., Kypr, J., Stokrove, S. & Sponar, J. (1982) Nucleic Acids Res. 10, 1071-1080
30. Zimmer, C., Tymen, S., Macek, C. & Gueschmaner, W. (1982) Nucleic Acids Res. 10, 1081-1091
31. Singleton, C. R., Kilpatrick, M. W. & Wells, R. D. (1984) J. Biol. Chem. 259, 1963-1967
32. Nordheim, A. & Rich, A. (1983) Nature 303, 674-679
33. Guarente, L. & Mason, T. (1983) Cell 32, 1279-1286
34. Guarente, L. (1984) Cell 36, 799-800