Proc. Natl. Acad. Sci. USA
Vol. 81, pp. 5066–5070, August 1984
Biochemistry

Distinct organization of methylcholanthrene- and phenobarbital-inducible cytochrome P-450 genes in the rat

(intron/exon/gene cloning/monooxygenase/molecular evolution)

KAZUHIRO SOGAWA*, OSAMU GOTOH†, KANAME KAWAJIRI†, AND YOSHIAKI FUJII-KURIYAMA*

*Department of Biochemistry, Cancer Institute, Japanese Foundation for Cancer Research, Toshima-ku 170, Japan; and †Department of Biochemistry, Saitama Cancer Center Research Institute, Ina-machi, Saitama 362, Japan

Communicated by Minor J. Coon, April 30, 1984

ABSTRACT **The complete nucleotide sequence of the methylcholanthrene-inducible cytochrome P-450c gene was determined by sequence analysis of cloned genomic DNA and the sequence, consisting of 524 amino acids, of the protein was deduced therefrom. The gene for the cytochrome was approximately 6.0 kilobases long and was split into seven exons. Comparison of the gene with that of the phenobarbital-inducible cytochrome P-450e showed that the gene structures for the two types of cytochrome P-450 differ greatly; the location, number, and size of intervening sequences are very dissimilar. However, the sequence homology between the two types of cytochrome suggests that the two genes have evolved from a common ancestor.**

Recent studies involving protein chemistry and molecular cloning technology (1–8) have shown that multiple forms of cytochrome P-450 are present in rat liver microsomes and that their syntheses are induced in different ways by the administration of various drugs. It is of great interest to understand the molecular mechanism underlying the multiplicity and selective drug induction of the cytochrome P-450 family.

We have already cloned several phenobarbital (PB)-inducible cytochrome P-450 genes and clarified the structure of one of them (the P-450e gene) (4, 9). It has been reported that there are at least two forms of methylcholanthrene (MC)-inducible cytochrome P-450 in rat liver, P-450c and P-450d (8). These two forms of the cytochrome show partial immunological cross-reactivity with each other, but they have different substrate specificities and different NH_2-terminal sequences.

Recently, we cloned and analyzed a cDNA for one of the MC-inducible cytochrome P-450 species (P-450d) and deduced the complete amino acid sequence (10, 11).

In this study, we have isolated genomic clones for another form of MC-inducible cytochrome P-450 (P-450c, a major species in MC-treated rat liver) from a rat gene library by cross-hybridization with a P-450d cDNA and determined the complete nucleotide sequence of the gene.

MATERIALS AND METHODS

Restriction endonucleases were obtained from Takara Shuzo (Kyoto, Japan), Bethesda Research Laboratories, and New England Biolabs. Polynucleotide kinase, reverse transcriptase and nuclease S1 were from Takara Shuzo, Life Sciences (St. Petersburg, FL), and P-L Biochemicals, respectively. [γ-^{32}P]ATP (3000 Ci/mmol; 1 Ci = 3.7 × 10^{10} Bq), [α-^{32}P]ddATP (5000 Ci/mmol), and [α-^{32}P]dCTP (2000–3000 Ci/mmol) were from the Radiochemical Centre.

DNA Preparation and Blot-Hybridization Experiments. Recombinant plasmid and phage DNAs were purified as described (3, 4). Blot-hybridization analyses were carried out according to the published procedure (12).

DNA Sequence Analysis. DNA sequences were analyzed by the procedure of Maxam and Gilbert (13).

Isolation of Genomic Clones for Cytochrome P-450. A rat *Eco*RI gene library of λ Charon 4A was provided by T. D. Sergent, R. B. Wallace, and J. Bonner. Plaque hybridization was carried out as described (4).

Primer Extension and Nuclease S1 Mapping. Primer extension (3) and nuclease S1 protection mapping (9) were carried out as described.

RESULTS AND DISCUSSION

Cloning and Identification of the MC-Inducible Cytochrome-P-450c Gene. Approximately 1 × 10^6 plaques from the rat *Eco*RI gene library were screened using the cloned P-450d cDNA (pcP-450mc-3; ref. 11) as a probe.

Two different clones were finally obtained, and they contained a common 5.3-kilobase (kb) *Eco*RI fragment that hybridized to the cDNA probe. Their restriction maps are shown in Fig. 1*A*. One clone (λP-450c-1) was used for further characterization. Clone λP-450c-1 was subjected to Southern blot analysis after cleavage by *Eco*RI, *Bam*HI, or *Hin*dIII. As shown in Fig. 1*B*, one band from the insert hybridized with the cDNA probe in each lane, and the blots of total DNA digested with the three restriction enzymes showed hybridized fragments of the same size as observed with the insert DNA, indicating that the organization of the cloned cytochrome P-450 gene reflects its native structure in the chromosome. Some other bands were observed in the autoradiogram from the total DNA. These bands have been shown, by cloning and subsequent characterization of the fragments, to originate from a gene for another form of MC-inducible P-450 (P-450d) (unpublished data).

The 5.3-kb *Eco*RI fragment and the upstream 1.6-kb (from near the *Kpn* I site to the *Eco*RI site) fragment were sequenced by the strategy in Fig. 2. The nucleotide sequence is shown in Fig. 3. The nucleotide sequence coding for amino acids was deduced by using the sequence homology to cytochrome P-450d (11), short amino acid sequences containing cysteine residues of P-450c (6), and the consensus sequence of the exon–intron boundary (14). Because the NH_2-terminal sequence deduced from the nucleotide sequence agreed completely with that of the purified protein of cytochrome P-450 MC-1 reported by Kuwahara *et al.* (15) and with the recently corrected sequence of 18 amino acids in cytochrome P-450c except for the second amino acid (histidine for proline) (6), the cloned genomic DNA was identified as the P-450c gene. Cytochrome P-450c is composed of 524 amino acids and its calculated M_r is 59,380. The overall homologies in amino acid and coding nucleotide sequences between cytochromes P-450c and P-450d are 68% and 75%, respectively.

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

Abbreviations: PB, phenobarbital; MC, methylcholanthrene; kb, kilobase(s); bp, base pair(s).

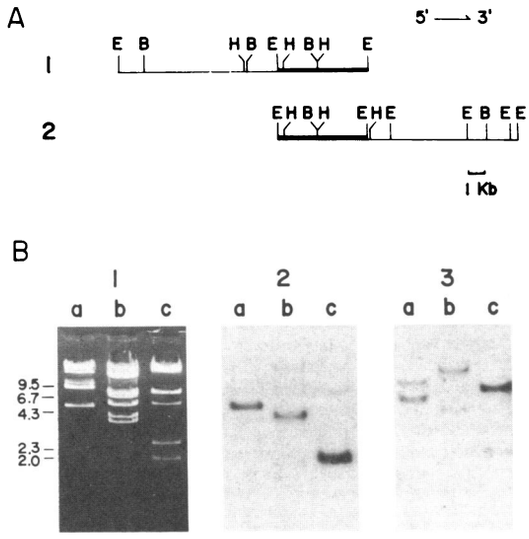

FIG. 1. Restriction maps of overlapping genomic clones of the rat MC-inducible cytochrome P-450 gene (*A*) and blot-hybridization analysis of the cloned genomic DNA and total DNA (*B*). (*A*) Two different clones (1, λP-450c-1, and 2, λP-450c-2) were screened from about 1×10^6 recombinant phages with the cDNA insert from plasmid pcP-450mc-3 as a probe. (*B*) Phage DNA (λP-450c-1) (1 μg) or total liver DNA (10 μg) was digested with *Eco*RI (lanes a), *Bam*HI (lanes b), or *Hin*dIII (lanes c) for electrophoresis in agarose and then transferred to nitrocellulose filters. The filters were hybridized to the ^{32}P-labeled cDNA probe and washed as described (5). Group 1, the ethidium bromide-stained agarose gel; group 2, autoradiogram of the digested cloned DNA; group 3, autoradiogram of digested total DNA. Lengths (in kb) of size markers are shown on the left. Faint bands in group 2 lanes are not reproducible and are probably due to partially digested DNA fragments.

Characteristics of the Nucleotide Sequence of the Cytochrome P-450c Gene. As shown in Fig. 4*A*, the leader sequence of the mRNA was estimated by the primer extension method to be approximately 100 bases long. Comparison of the sequence (Fig. 4*B*) of the primer-extended cDNA fragment with that of the corresponding part of the gene shows that the gene structure encoding the leader sequence of the mRNA is split 14 base pairs (bp) upstream from the initiation codon by an intervening sequence(s). This result was confirmed by the nuclease S1 mapping experiment (data not shown).

As shown in Fig. 2, sequencing downstream from the *Bst*EII site revealed the presence of the 5′ part of the leader sequence, which was determined by the primer-extension method and was followed by the typical splicing signal of the 5′ end of the intervening sequence, G-T-G-A-G (14). The size of the first exon estimated from the primer-extension experiment located the cap site of the mRNA at about 85 bp upstream from the 5′ end of the intron. Subsequently, nuclease S1 protection mapping confirmed this estimation and localized the transcription-initiation site at an adenosine 87 bp upstream from the splicing site (Fig. 5*A*). A modified "TATA" box, C-A-T-A-T-A, was located 27 bp upstream from the cap site. In the promoter region, a sequence, G-T-G-G-A-A-A-G, a core sequence of the enhancer elements (16), and a Z-DNA forming repetitive sequences (17), $(G\text{-}T)_{21}$, lie approximately 380 and 220 bp upstream from the cap site, respectively. There exist two other kinds of repetitive sequences of dinucleotide, $(C\text{-}T)_{24}$ and $(A\text{-}G)_{19}$, in the first intervening sequence and also one kind of the repetitive sequence $(C\text{-}T)_{19}$ in the upstream part of the promoter region. These two kinds of repetitive sequences, $(C\text{-}T)_{24 \text{ or } 19}$ and $(A\text{-}G)_{19}$, are complementary with each other, suggesting a possibility of forming two types of stem-and-loop structures.

When the two types of cytochrome P-450, the MC-inducible P-450c and the PB-inducible P-450e (data on the P-450e promoter sequence will be published elsewhere) are compared in the nucleotide sequences of the promoter region, a short conserved sequence, A-G-G-A-G-G-C-G-T-G, is noticeable, and it is located 55 and 78 bp upstream from the transcription initiation sites of P-450c and P-450e, respectively. In other parts of the promoter regions, however, no marked sequence homology is observed between them.

In the DNA sequence coding for the trailer portion of the mRNA, there occur two typical poly(A) addition signals, A-A-T-A-A-A. They are 470 and 950 bp downstream from the termination codon, respectively. A nuclease S1 mapping experiment was carried out to determine which signal functions. As shown in Fig. 5*B*, a single major protected band was detected 700 bp downstream from the *Sst* I site together with several faint bands. This result indicates that the second signal is the sole functional poly(A) addition signal of the gene. Practically no band corresponding to the first signal was observed. The reason why the first poly(A) signal is silent in the P-450c gene is not known. Nuclease S1 protection mapping together with sequence analysis of the cloned cDNA showed that the poly(A) sequence was attached to either the guanosine or one of the adenosines 12–14 bp downstream from the second signal.

On the whole, the cytochrome P-450c gene is approximately 6.0 kb long from the transcription-initiation site to the poly(A) attachment site and contains seven split exonic sequences.

From the gene framework as described above and in Fig. 2, the complete architecture of the P-450c mRNA can be inferred. The total length of the mRNA is estimated to be approximately 2650 bases. This estimate was supported by an RNA blot experiment using poly(A)$^+$ RNA from MC-induced rat liver (data not shown).

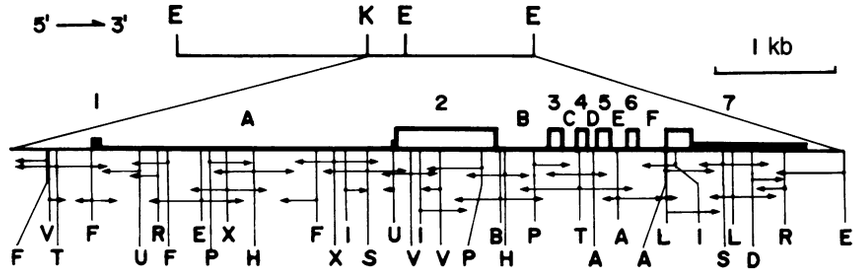

FIG. 2. Organization of the rat MC-inducible P-450 gene and strategy for sequencing. A cloned genomic DNA (14.9 kb) containing the P-450c gene is represented by a bar, oriented in the 5′ to 3′ direction. Only the restriction cleavage sites used for DNA sequencing are shown. A, *Acc* I; B, *Bam*HI; D, *Dde* I; E, *Eco*RI; F, *Fok* I; H, *Hin*dIII; I, *Pst* I; K, *Kpn* I; L, *Hpa* II; P, *Pvu* II; R, *Dra* I; S, *Sst* I; T, *Bst*EII; U, *Sau*3AI; V, *Ava* II; X, *Xba* I. Coding sequences are indicated by open boxes and DNA segments coding for leader and trailer sequences of the mRNA are indicated by closed boxes. Exons are numbered from 1 to 7 and introns are indicated alphabetically (A, B, . . . , F above the bar representing the gene). Arrows indicate direction and length of DNA sequences analyzed. The scale is for the gene structure as shown below.

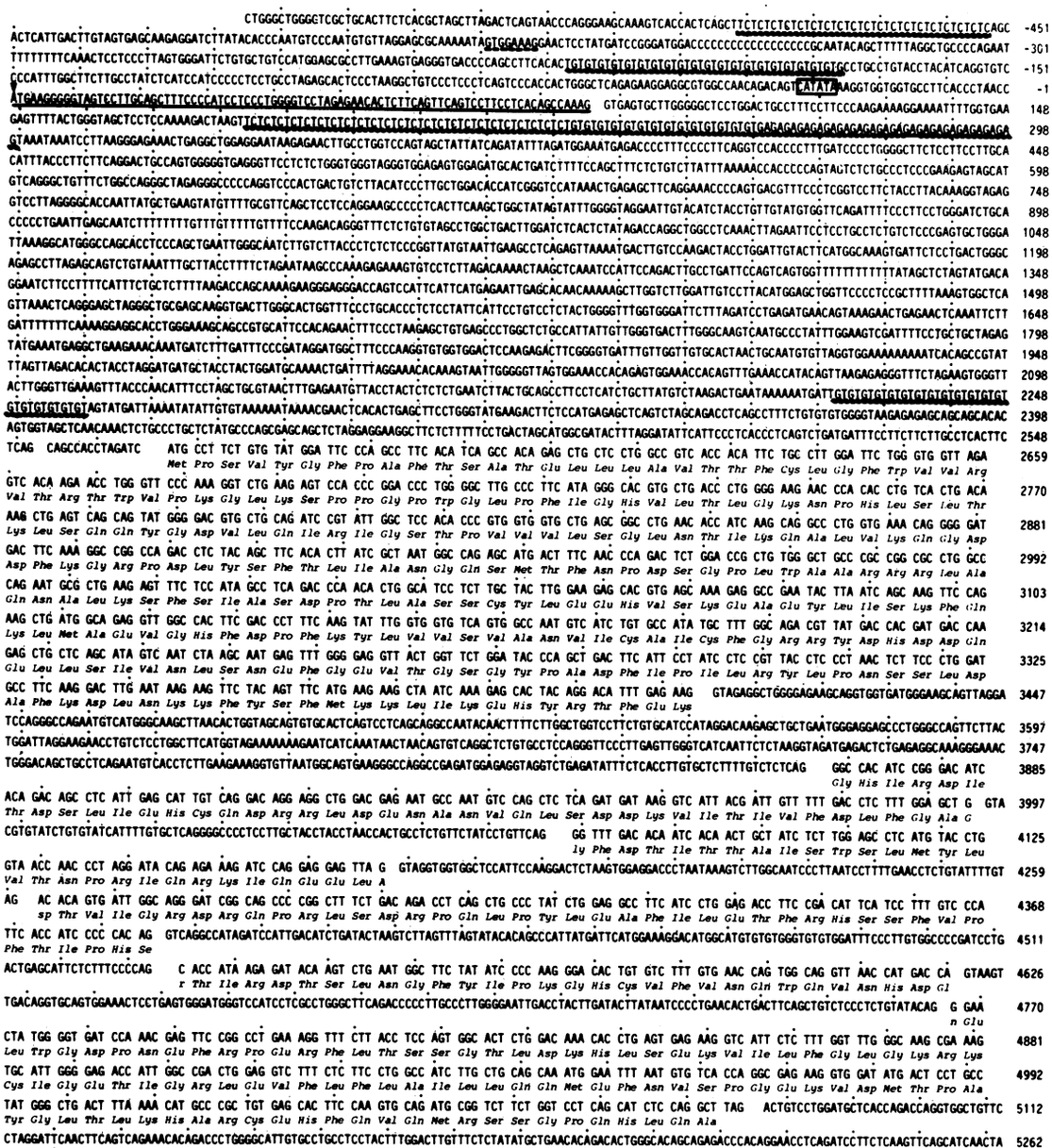

FIG. 3. Nucleotide sequence of the rat MC-inducible cytochrome P-450 gene. The strategy for sequencing is shown in Fig. 2. Nucleotides are numbered from the adenosine of the cap site. The predicted amino acid sequence is shown below the nucleotide sequence. The first exon is underlined. The sequences C-A-T-A-T-A and A-A-T-A-A-A in the 5′ and 3′ noncoding region, respectively, are indicated by enclosures. The repetitive sequences of dinucleotides $(G\text{-}T)_n$, $(C\text{-}T)_n$, and $(G\text{-}A)_n$, are underlined with wavy lines. The sequence homologous to the core sequence of enhancer elements is underlined with a broken line. Arrowheads indicate termini of the probes generated by nuclease S1 mapping procedures (see Fig. 5).

Homology in the Amino Acid Sequences. The overall homology in amino acid sequence of cytochrome P-450c with other species of MC-inducible cytochrome P-450d and with the PB-inducible counterparts (P-450b and -e) are 68% and 29%, respectively. The complete primary structures of cytochromes P-450c, P-450d, and P-450e are compared in Fig. 6. The putative heme-binding cysteine (located at position 461 from the NH_2-terminal methionine) in one of the two conserved segments (11, 19) and its surrounding sequence are also highly conserved in P-450c. On the basis of the amino acid replacement rate in the evolution of P-450 molecules as calculated previously (11), we can estimate that the ancestors of the PB-inducible P-450 (P-450b and e) and the MC-inducible P-450 (P-450c and d) diverged some 400 million years ago in the Devonion period of the Paleozoic Era and then the divergence leading to P-450c and P-450d occurred 120 million years ago. The accuracy of this estimate might be improved by examining the structure of P-450 genes in other species of vertebrates, especially in lower vertebrates.

Structural Comparison of MC-Inducible and PB-Inducible P-450 Genes. Gilbert (20) has proposed that introns exist as intervening sequences to separate portions of the coding re-

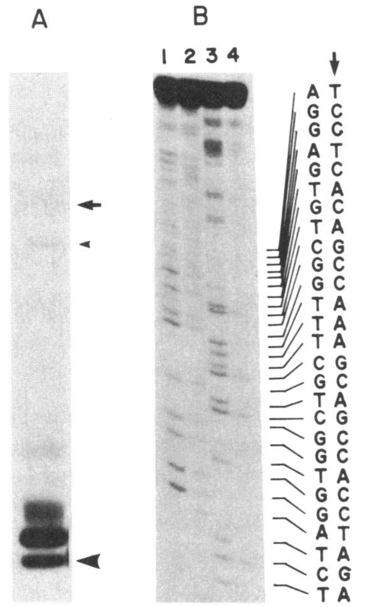

FIG. 4. Estimation of the length of the P-450c mRNA leader sequence (*A*) and analysis of this sequence (*B*). (*A*) The terminally labeled anticoding strand was prepared from the *Sau*3AI/*Ava* II fragment (140 bp) by strand separation and used as the primer. Extension of the primer was carried out using 10 μg of the poly(A)$^+$ RNA as the template and the products were analyzed as described (2). The size was estimated using appropriate sequence ladders as size markers. The arrow indicates the position of the extended DNA fragment. The small arrowhead indicates a contaminant DNA fragment, and the large arrowhead indicates a fragment not extended. (*B*) A synthetic oligonucleotide (5′ A-A-T-C-C-A-T-A-C-A-C-A-G 3′) was used as a primer. The extended DNA fragment was eluted from the gel and used for sequencing. Lanes: 1, 2, 3, and 4, G, G+A, T+C, and C degradation products, respectively.

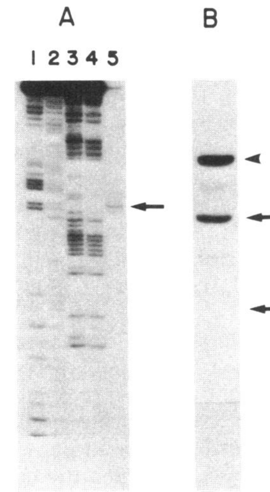

FIG. 5. Nuclease S1 mapping of the 5′ (*A*) and 3′ (*B*) termini of the MC-inducible P-450 mRNA. (*A*) A 280-bp *Bst*EII/*Ava* II fragment was labeled at the 5′ termini with ^{32}P and then strand separated after cleavage with *Dde* I. The single-stranded fragment was hybridized with the poly(A)$^+$ RNA (10 μg) from MC-treated rat liver, and then the mixture was treated with nuclease S1 as described (5). For chain-length markers, the relevant DNA fragment was degraded by the Maxam–Gilbert procedure, and fragments were electrophoresed in parallel. Lanes: 1, 2, 3, and 4, G, G+A, T+C, and C degradation products, respectively; lane 5, fragments protected by hybridization with poly(A)$^+$ RNA. (*B*) An *Sst* I restriction fragment was labeled at the 3′ termini with ^{32}P and then cleaved with *Bam*HI (the *Bam*HI site is located in the vector). The 1.4-kb *Sst* I/*Bam*HI fragment was hybridized with the poly(A)$^+$ RNA (10 μg) and then treated as in *A*. The size was determined in relation to the mobilities of appropriate ^{32}P-labeled fragments in the same gel. The upper arrow indicates the location of the protected fragment and the lower arrow indicates the location of the putative protected fragment corresponding to the first poly(A) addition signal. The arrowhead indicates the undigested fragment.

gions of genes according to the structural–functional domains for their corresponding proteins. The gene structures for β-globin (21), chicken ovomucoid (22), and a heavy chain of immunoglobulin (23) support this proposition. Accordingly, the numbers and positions of multigene families such as the globins (21) and vitelogenins (24) are rigidly conserved in the course of divergent evolution. The situations, however, have become somewhat complicated, because observations that appear to be contradictory with this proposition have been accumulating. These include actin genes of various organisms (25–29), myosin genes in the nematode (30), and the supergene family of ovalbumin and α_1-antitrypsin (31). Now,

```
2.5Kb
MPSVYGFPAF TS-ATELLLA VTTFCLGFWV VRVTRTWVPK GLKSPPGPWG LPFIGHVLTL GKNPHL-SLT KLSQQYGDVL QIRIGSTPVV VLSGLNTIKQ ALVKQGDDFK   108
    MAFSQY ISLAPELLLA TAIFCLVFWV LRGTRTQVPK GLKSPPGPWG LPFIGHMLTL GKNPHL-SLT KLSQQYGDVL QIRIGSTPVV VLSGLNTIKQ ALVKQGDDFK   105
           MEPTILLLLA LL---VGFLL LLVRGHPKS- RGNFPPGPRP LPLLGNLLQL DRGGLLNSFM QLREKYGDVF TVHLGPRPVV MLCGTDTIKE ALVGQAEDFS    96
                                                                           3.2Kb

GRPDLYSFTL IANGQSMTFN PDSGPLWAAR RRLAQNALKS FSIASDPTLA SSCYLEEHVS KEAEYLISKF QKLMAEVGHF DP-FKYLVVS VANVICAICF GRRYDHDDQE   217
GRPDLYSFTL ITNGKSMTFN PDSGPVWAAR RRLAQDALKS FSIASDPTSV SSCYLEEHVS KEANHLISKF QKLMAEVGHF EP-VNQVVES VANVIGAMCF GKNFPRKSEE   214
GRGTIAVIEP IFKEYGVIF- -ANGERWKAL RRFSLATMRD FGMGKRS--- ----VEERIQ EEAQCLVEEL RK--SQGAPL DPTFLFQCIT -ANIICSIVF GERFDYTDRQ   194
              0.3Kb                                                      0.5Kb         2.3Kb                                84bp

LLSIVML-SN EF-GEVTGSG YPADFIP-IL RYLPNSSLDA FKDLNKKFYS FMKKLIKEHY RTFEKGHIRD ITDS-LIEHC QDRRLDENAN VQLSDDKVIT IVFDLFGAGF   323
MLNLVKS-SK DF-VENVTSG NAVDFFP-VL RYLPNPALKR FKNFNDNFVL SLQKTVQEHY QDFNKNSIQD ITGA-LFKHS EN--YKDNGG L-IPQEKIVN IVNDIFGAGF   317
FLRLLELFYR TFSLLSSFSS QVFEFFSGFL KYFPGAHRQI SKNL-QEILD YIGHIVEKHR ATLDPSAPRD FIDTYLLRME KE--KSNHHT E-FHHENLMI SLLSLFFAGT   300
                        0.8Kb       93bp                                    0.15Kb       0.5Kb           0.15Kb

DTITTAISWS LMYLVTNPRI QRKIQEELDT VIGRDRQPRL SDRPQLPYLE AFILETFRHS SFVPFTIPHS TIRDTSLNGF YIPKGHCVFV NQWQVNHDQE LWGDPNEFRP   433
ETVTTAIFWS ILLLVTEPKV QRKIHEELDT AIGRDRQPRL SDRPQLPYLE AFILEIYRYT SFVPFTIPHS TTRDTSLNGF HIPKECCIFT NQWQVNHDEK QWKDPFVFRP   427
ETGSTTLRYG FLLMLKYPHV TVKVQKEIDQ VIGSHRPPSL DDRTKMPYTD AVIHEIQRFA DLAPIGLPHR VTKDTMFRGY LLPKNTEVYP ILSSALHDPQ YFDHPDTFNP   410
                        1.6Kb                                                          0.3Kb

ERFLTSSGT- LDKHLSEKVI LFGLGKRKCI GETIGRLEVF LFLAILLQQM EFNVSPGEK- VDMTPAYGLT LKHARCEHFQ VQMRSSGPQH LQA   (P-450c)   524
ERFLTNDNTA IDKTLSEKVM LFGLGKRRCI GEIPAKWEVF LFLAILLHQL EFTVPPGVK- VDLTPSYGLT MKPRTCEHVK AWPRFSK       (P-450d)   513
EHFL-DADGT LKK--SEAFM PFSTGKRICL GEGIARNELF LFFTTILQNF SVSSHLAPKD IDLTPKESGI GKIPPTYQIC FSAR          (P-450e)   491
                        3.2Kb
```

FIG. 6. Primary structures of cytochromes P-450c, P-450d, and P-450e and locations of introns in relation to the structures. Amino acids are represented by the single-letter code (18). Gaps are introduced to obtain maximum homology. Sites of introns are indicated at the appropriate amino acids by downward and upward arrows for P-450c and P-450e, respectively, and their sizes are shown in kb or bp. The exact locations of introns are as follows: just before the first nucleotide of the codon, the second intron for P-450c and the first, fourth, fifth, and seventh for P-450e; between the first and the second nucleotide of the codon, the third and fourth for P-450c and the second, third, sixth, and eighth for P-450e; between the second and the third nucleotide of the codon; the fifth and sixth for P-450c. Introns are numbered from left to right (NH_2 to COOH terminal). The first intron of the P-450c gene is localized 14 bp upstream from the initiation codon.

the genes for two species of cytochrome P-450 [P-450c and P-450e (ref. 9)] have been sequenced and the exact numbers and locations of introns and exons have been established in their respective genes. Therefore, we are able to compare these two gene structures precisely. The locations of introns in P-450c and P-450e are marked in the amino acid sequences in Fig. 6. The first intron of P-450c is located in the leader sequence of the mRNA and therefore marked in front of the initiator methionine. As shown in the figure, none of the six intervening sequences of the P-450c gene is located at a site equivalent to one of the eight intervening sequences in the P-450e gene.

At present, it is impossible to determine conclusively whether cytochromes P-450c and P-450e arose by convergent or divergent evolution. The lack of similarity in gene organization between the two proteins would not be unexpected if they were derived from different ancestors by convergent evolution. However, when the homology in coding nucleotide and amino acid sequences between P-450c and P-450e is taken into consideration together with their interaction with the common reductase, divergent evolution appears to be more likely than convergent. If divergent evolution is the case with the P-450 gene family, then how can we explain the totally different numbers and locations of intervening sequences in the two P-450 genes?

Roughly speaking, two possibilities can be considered. First, the ancestor gene to cytochromes P-450c and P-450e may have been duplicated before the insertion or deletion of intervening sequences. The insertion or deletion of the DNA sequence may occur independently in the duplicated genes. It has been suggested that some introns are vestiges of transposon-like sequences that have been inserted into genes, become fixed, and then diverged in nucleotide sequence (28). In the case of the P-450 genes, however, we have not yet found DNA sequences reminiscent of the insertion of transposon-like elements in the intervening sequences. The DNA sequences in the introns may have diverged so rapidly after insertion into the preexisting gene that their characteristic features may not have been recognizable.

Second, the ancestral gene to these two proteins may have contained 14 or more introns and then, during the course of gene evolution, given rise to two or more genes by random deletion of introns. Precise deletion of an entire intron has been reported for the rat insulin gene (32) and the mouse α-globin gene (33). In the case of the cytochrome P-450 gene, whose primordial gene is presumed to contain as many as or more than 14 introns, a number of introns would be expected to share common sites in the two genes, if random deletion of introns were a real process. The observation that there are no common sites of intervening sequences between two P-450 genes makes the latter alternative less likely. Elucidation of the gene structures for other types of cytochrome P-450, such as those involved in the biosynthesis of steroid hormones and those induced by other kinds of inducers, may help us to understand the origin of introns during the course of gene evolution.

After completion of this manuscript, the sequence of rat P-450c cDNA was published (34) and it concurs with the coding nucleotide sequence in this paper.

We thank Drs. T. D. Sergent, R. B. Wallace, and J. Bonner for the gift of a rat gene library. We also thank Dr. M. Nobuhara (Mochida Pharmaceutical, Tokyo) for supplying the synthetic primer. This work was partly supported by grants-in-aid for scientific research from the Ministry of Education, Science, and Culture of Japan; by funds obtained under the Life Science Project from the Institute of Physical and Chemical Research, Japan; and by funds from the Society for Promotion of Cancer Research, Japan, and from the Princess Takamatsu Cancer Research Fund.

1. Sato, R. & Omura, T., eds. (1978) *Cytochrome P-450* (Kodansha, Tokyo).
2. Lu, A. Y. H. & West, S. B. (1980) *Pharmacol. Rev.* **31,** 277–295.
3. Fujii-Kuriyama, Y., Mizukami, Y., Kawajiri, K., Sogawa, K. & Muramatsu, M. (1982) *Proc. Natl. Acad. Sci. USA* **79,** 2793–2797.
4. Mizukami, Y., Fujii-Kuriyama, Y. & Muramatsu, M. (1983) *Biochemistry* **22,** 1223–1229.
5. Hardwick, J. P., Gonzalez, F. J. & Kasper, C. B. (1983) *J. Biol. Chem.* **258,** 10182–10186.
6. Haniu, M., Yuan, P.-M., Ryan, D. E., Levin, W. & Shively, J. E. (1984) *Biochemistry* **23,** 2478–2482.
7. Atchison, M. & Adesnik, M. (1983) *J. Biol. Chem.* **258,** 11286–11295.
8. Botelho, L. H., Ryan, D. E., Yuan, P.-M., Kutny, R., Shively, J. E. & Levin, W. (1982) *Biochemistry* **21,** 1152–1155.
9. Mizukami, Y., Sogawa, K., Suwa, Y., Muramatsu, M. & Fujii-Kuriyama, Y. (1983) *Proc. Natl. Acad. Sci. USA* **80,** 3958–3962.
10. Kawajiri, K., Sogawa, K., Gotoh, O., Tagashira, Y., Muramatsu, M. & Fujii-Kuriyama, Y. (1983) *J. Biochem.* **94,** 1465–1473.
11. Kawajiri, K., Gotoh, O., Sogawa, K., Tagashira, Y., Muramatsu, M. & Fujii-Kuriyama, Y. (1984) *Proc. Natl. Acad. Sci. USA* **81,** 1649–1653.
12. Southern, E. (1975) *J. Mol. Biol.* **98,** 503–517.
13. Maxam, A. M. & Gilbert, W. (1977) *Proc. Natl. Acad. Sci. USA* **74,** 560–564.
14. Sharp, P. A. (1981) *Cell* **23,** 643–646.
15. Kuwahara, S., Harada, N., Yoshioka, H., Miyata, T. & Omura, T. (1984) *J. Biochem.* **95,** 703–714.
16. Khoury, G. & Gruss, P. (1983) *Cell* **33,** 313–314.
17. Wang, A. H.-J., Quigley, G. J., Kalpal, F. J., Crawford, J. L., Van Boom, J. M., Vander Marel, G. & Rich, A. (1979) *Nature (London)* **282,** 680–689.
18. IUPAC-IUB Commission on Biochemical Nomenclature (1968) *Eur. J. Biochem.* **5,** 151–153.
19. Gotoh, O., Tagashira, Y., Iizuka, T. & Fujii-Kuriyama, Y. (1983) *J. Biochem.* **93,** 807–817.
20. Gilbert, W. (1978) *Nature (London)* **271,** 501.
21. Maniatis, T., Fritsch, E. F., Lauer, J. & Lawn, R. M. A. (1980) *Rev. Genet.* **14,** 145–178.
22. Stein, J. P., Catterall, J. F., Kristo, P., Means, A. R. & O'Malley, B. W. (1980) *Cell* **21,** 681–687.
23. Sakano, H., Rogers, J. H., Huppi, K., Brack, C., Traunecker, A., Maki, R., Wall, R. & Tonegawa, S. (1979) *Nature (London)* **277,** 627–633.
24. Wahli, W., Dawid, I. B., Wyler, T., Wever, R. & Ryffel, G. U. (1980) *Cell* **20,** 107–117.
25. Gallwitz, D. & Sures, I. (1980) *Proc. Natl. Acad. Sci. USA* **77,** 2546–2550.
26. Ng, R. & Abelson, J. (1980) *Proc. Natl. Acad. Sci. USA* **77,** 3912–3916.
27. Firtel, R., Rimm, R., Kimmel, A. R. & McKeown, M. (1979) *Proc. Natl. Acad. Sci. USA* **76,** 6202–6210.
28. Fyrberg, E. A., Bond, B. J., Hershey, N. D., Mixter, K. S. & Davidson, N. (1981) *Cell* **24,** 107–116.
29. Durica, D. S., Schloss, J. A. & Crain, W. R., Jr. (1980) *Proc. Natl. Acad. Sci. USA* **77,** 5683–5687.
30. Karn, J., Brenner, S. & Barnett, L. (1983) *Proc. Natl. Acad. Sci. USA* **80,** 4253–4257.
31. Leicht, M., Long, G. G., Chandra, T., Kurachi, K., Kidd, V. J., Mace, M., Davie, E. W. & Woo, S. L. C. (1982) *Nature (London)* **297,** 655–659.
32. Lomedico, P., Rosenthal, N., Efstratiadis, A., Gilbert, W., Kolodner, R. & Tizard, R. (1979) *Cell* **18,** 545–558.
33. Nishioka, Y., Leder, A. & Leder, P. (1980) *Proc. Natl. Acad. Sci. USA* **77,** 2806–2809.
34. Yabusaki, Y., Shimizu, M., Murakami, H., Nakamura, K., Oeda, K. & Ohkawa, H. (1984) *Nucleic Acids Res.* **12,** 2929–2938.