Structural Analysis of the Gene Encoding Rat Cholesterol α-Hydroxylase, the Key Enzyme for Bile Acid Biosynthesis*

(Received for publication, September 11, 1990)

Masazumi Nishimoto, Osamu Gotoh, Kyuichiro Okuda, and Mitsuhide Noshiro

From the Department of Biochemistry, Hiroshima University School of Dentistry, Kasumi 1-2-3, Hiroshima 734 and the

Department of Biochemistry, Saitama Cancer Center Research Institute, Ina-machi, Saitama 362, Japan

The gene encoding cholesterol 7α-hydroxylase (P450VIIA) was isolated from rat genomic DNA. The gene spanned about 11 kilobases and contained six exons. Blotting analysis of genomic DNA and complete matching of restriction maps of several isolated genomic clones indicated that there appeared to be only one gene in the rat genome. The putative transcription initiation site was present 61 base pairs upstream from the ATG codon. The typical TATA sequence and CCAAT promoter element were found at 24 and 47 base pairs upstream from the transcription initiation site, respectively.

Alignment of several P450 proteins showed that the cholesterol 7α-hydroxylase gene shared location of introns with none of the other P450 genes except for intron 5, which was in the same position as intron 10 of the gene encoding P450IVA1. The alignment also indicated that the distal helix of cholesterol 7α-hydroxylase contained an asparagine in place of the well conserved threonine that is postulated to be involved in the O2 binding site. Unusual residues, Asn-126 and Thr-442, were also found at the sites where all other P450s have positively charged amino acids, which are considered to be involved in interaction with heme propionate. These replacements may be related to the unique function and unusual lability of the hydroxylase. Analysis of evolutionary distance between the cholesterol 7α-hydroxylase gene and other known P450 genes indicated that yeast P450OLIA is most closely related to P450VIIA. This finding suggests that the cholesterol 7α-hydroxylase gene is an evolutionary old P450 gene.

Cholesterol 7α-hydroxylase catalyzes the initial hydroxylation of cholesterol at the 7α-position, which constitutes a rate-limiting step for cholesterol catabolism (Björkhem, 1985). The enzyme is a microsomal monoxygenase system comprising cytochrome P450VIIA1 and NADPH-cytochrome P450 reductase (Wada et al., 1968; Björkhem et al., 1974). The enzyme activity in the rat exhibits a circadian rhythm, with maximum activity at night and minimum activity during the day (Myant and Mitropoulos, 1977), and is regulated by the level of bile acids returning to liver via enterohepatic circulation. The enzyme activity is also reported to be affected by administration of hormones such as adrenal cortical hormone and thyroxine (Balasubramaniam et al., 1975), cytosolic factors (Kwok et al., 1981), and phosphorylation (Scallen and Sanghvi, 1983).

Recently P450VIIA has been purified to homogeneity and its enzymatic natures characterized to show strict substrate specificity and stereo- and regioselectivity (Ogishima et al., 1987). Using specific antibodies against the enzyme as a probe, a cDNA clone encoding P450VIIA has been isolated and its complete amino acid sequence predicted from the nucleotide sequence of cDNA (Noshiro et al., 1989). Subsequent studies using the specific antibodies and the cDNA have demonstrated that the circadian rhythm of cholesterol 7α-hydroxylase activity is mainly pretranslationally regulated and that both the enzyme protein and mRNA of cholesterol 7α-hydroxylase show rapid turnover (Noshiro et al., 1990). For further understanding of such enzyme regulation, characterization of the protein structure as well as the regulatory elements of the gene may be needed. In this study, we determined the structure of the rat cholesterol 7α-hydroxylase gene and performed an alignment analysis of the protein sequence.

EXPERIMENTAL PROCEDURES

RESULTS AND DISCUSSION

Number of Cholesterol 7α-Hydroxylase Genes in Rat Genome—Eight positive clones were isolated from 2 × 106 plaques by screening EMBL genomic libraries using the insert of p7α-11 cDNA as a probe and then subjected to restriction mapping analysis. We constructed the cholesterol 7α-hydroxylase gene by overlapping two clones (AG-3 and AG-4) as shown in Fig. 1. Other isolated clones showed the restriction sites of BamHI, EcoRI, HindIII, PstI, and XbaI overlapping with parts of the constructed gene (data not shown). To examine the size and complexity of the cholesterol 7α-hydroxylase gene, we performed Southern blot analysis of total genomic DNA using a total insert of p7α-11 cDNA or fragment I or II (shown in Fig. 1) as a probe. The hybridized signals with the total insert of cDNA showed a simple hybridization pattern on digestion of genomic DNA by BamHI, EcoRI, or HindIII (Fig. 2A). When fragment I, which contains
the first exon and the flanking sequence (700 base pairs) prepared from λg-4, was used as a probe, HindIII digests of genomic DNA showed only one hybridized band (2.2 kbp) corresponding to the expected DNA fragment of the cholesterol 7α-hydroxylase gene (Fig. 1). When fragment II, a 3′-noncoding region of p7a-11 cDNA (1.1 kbp), was used to hybridize EcoRI digests of genomic DNA, only one hybridized band (4.0 kbp), which was also expected from the restriction map, was observed (Fig. 2C). Sizes of the hybridized bands with fragments I and II were the same as the corresponding DNA fragments obtained from HindIII-digested λg-4 and EcoRI-digested XG-3, respectively (data not shown). Taken together, the data show that there appears to be only one gene encoding P450VIIA in rat genome. Accordingly, the two clones, λg-3 and λg-4, may have been derived from the same gene covering an entire transcriptional region.

Structure of Cholesterol 7α-Hydroxylase Gene—As shown in Fig. 1, the rat cholesterol 7α-hydroxylase gene spanned approximately 11 kbp in size and contained six exons. Exons 1–6 contained 141, 241, 587, 131, 176, and 2295 base pairs, respectively (Fig. 3). Nucleotide sequences of the exons were identical with that of p7α-11 cDNA (Noshiro et al., 1990) except for a few nucleotide differences in the noncoding region as indicated in Fig. 3. It was confirmed that the inconsistency was due to errors in the sequencing of cDNA. The number of exons in the cholesterol 7α-hydroxylase gene was less compared with other mammalian P450 species, which were reported to contain 7–13 exons. As listed in Fig. 5, microsomal P450IA1 (Sogawa et al., 1984), P450IB1 (Suwa et al., 1985), P450XVIIIA (Picado-Leonard and Miller, 1987), P450XXIB (Higashi et al., 1986; White et al., 1986), P450VIIA (Kimura et al., 1989), and P450IX (Means et al., 1989) genes contain 7, 9, 8, 10, 13, and 19 exons, respectively, and mitochondrial P450XIIA1 (Morohashi et al., 1987) and P450XIIIB1 (Hashimoto et al., 1989) genes contain 9 exons. Like most other P450s, the heme binding domain and the entire 3′-untranslated region of cholesterol 7α-hydroxylase gene are contained in the last exon.

Introns 1–5 were 1.7, 0.7, 0.9, 1.9, and 0.5 kbp long, respectively, and were not so large and not too different from one another in size. Nucleotide sequences of all exon/intron boundaries followed the canonical G-T/A-G rule. Flanking sequences of 5′- and 3′-splicing sites in the introns were well in accordance with the splice junction consensus sequences proposed by Mount (1982) except for the sequence (5′-tatttctg-3′) of the 3′-splicing site of the fourth intron, which was rich in nucleotide A. This sequence may be usable as a putative cap site, sequence TATAAA, the expected TATA box, occurred between positions −29 and −24. The sequence ATTTG complementary to the CCAAT consensus sequence is present between positions −51 and −47. However, we have not yet discovered a regulatory element, which may be sensitive to bile acid(s), a putative negative feedback effector for cholesterol catabolism. The region further upstream is now under investigation.

Analysis of Amino Acid Sequence—Li et al. (1990) and Jelinek et al. (1990) have also recently reported the nucleotide sequence of cDNA for rat cholesterol 7α-hydroxylase. Their coding sequences are identical to our sequence except for a few nucleotide differences, which do not alter the amino acids encoded. These are likely due to the different strains of rats used. Accordingly, the amino acid sequences of P450VIIA reported by all three laboratories are completely identical. Fig. 5 shows the alignment of P450 amino acid sequences with the exon/intron boundaries superimposed. No intron in the cholesterol 7α-hydroxylase gene is equivalent to any intron in other P450 genes with respect to their locations on the aligned amino acid sequences except for intron 5 of the cholesterol 7α-hydroxylase gene, which is the same position upon alignment as intron 10 of the gene encoding P450IA1. This poor correlation favors the hypothesis that the addition of introns after divergence of P450 families may be a mechanism for generation of these variations in gene structure (Sogawa et al., 1984).

Several amino acid residues are highly conserved in all 450 species to maintain the function of hemoprotein as monoxygenase (Gotoh and Fujii-Kuriyama, 1989). Most of the conserved amino acids were also observed in P450VIIA, as indicated at the bottom of each alignment in Fig. 5. However, the alignment analysis indicated that the distal helix of P450VIIA contained an asparagine in place of the corresponding threonine residue, which is common to almost all P450s so far sequenced, except for P450IIC2 with asparagine by site-directed mutagenesis (Okuda, 1990). Furthermore, the amino acid sequence of P450VIIA, which was predicted from the cDNA (Noshiro and Okuda, 1990), has shown only one major mRNA species, among which are likely to interact with heme propionate (Poulos et al., 1985), i.e. positions 157 and 508 of rat P450VIIA are replaced by asparagine and threonine 442, respectively.

To define the start site for transcription, S1 nuclease mapping was performed. As shown in Fig. 4, the signal of S1 nuclease-resistant fragment corresponding to nucleotide G, which was considered to be the nucleotide capped in the mRNA, was observed 61 base pairs upstream from the ATG translation initiation codon. In the region preceding the putative cap site, sequence TATAAA, the expected TATA box, occurred between positions −29 and −24. The sequence ATTTG complementary to the CCAAT consensus sequence is present between positions −51 and −47. However, we have not yet discovered a regulatory element, which may be sensitive to bile acid(s), a putative negative feedback effector for cholesterol catabolism. The region further upstream is now under investigation.
cholesterol 7α-hydroxylase. These substitutions may explain the unique function and unusual lability of cholesterol 7α-hydroxylase described by Ogishima et al. (1987).

Fig. 6 shows a phylogenetic tree constructed based on the analysis of the P450 protein sequences. Among the known P450 families, P450VIIA is most closely related to yeast P450LIA. These two P450s are estimated to have diverged 1.0–1.2 × 10⁸ years ago. It has been proposed (Gotoh and Fujii-Kuriyama, 1989) that the substrate specificity of ancestral enzyme might have been stringent, but duplications and mutations produced P450 species with relaxed specificities. Accordingly, early divergence of P450VIIA may be a reason for its stringent substrate specificity (Ogishima et al., 1987).

It has been reported that the circadian rhythm of cholesterol 7α-hydroxylase activity (Noshiro et al., 1990) and the modulation of the activity by bile acid(s) (Jelinek et al., 1990) are mainly regulated at the pretranslational level. The enzyme activity is also modulated by various other factors such as thyroxine, glucocorticoids, and insulin (Myant and Mitropoulos, 1977). Characterization of the regulatory sequences in the gene responsive for these factors should be helpful for understanding the multifactorial regulation and tissue-specific expression of liver cholesterol 7α-hydroxylase, which may lead to the ultimate clarification of the etiology of a number of important human diseases related to disorders of cholesterol metabolism.

REFERENCES

Aviv, A., and Leder, P. (1972) Proc. Natl. Acad. Sci. U. S. A. 69, 1408–1412

Balasubramaniam, S., Mitropoulos, K. A., and Myant, N. B. (1975) in Advances in Bile Acids Research (Matern, S., Hackenschmidt, J., Back, P., and Gerock, W., eds) Vol. 3, pp. 61–67, F. K. Schattauer Akademie-Verlag, Berlin, FRG

Bjorkhem, I., Danielsson, H., and Wikvall, K. (1974) FEBS Lett. 448, 570–589

Bjorkhem, I., Black, D. L., LeMaster, D. A., and Steitz, J. A. (1985) J. Biol. Chem. 260, 7435–7441

Bjorkhem, I. (1985) in Sterols and Bile Acids (Danielsson, H., and Sjövall, J., eds) pp. 231–278, Elsevier Science Publishers, Amsterdam

Bjorkhem, I., Danielsson, H., and Wikvall, K. (1974) Biochem. Biophys. Res. Commun. 61, 934–941

Bozak, K. R., Yu, H., Sirevag, R., and Christoffersen, R. F. (1990) Proc. Natl. Acad. Sci. U. S. A. 87, 3904–3908

Chabot, B., Black, D. L., LeMaster, D. A., and Steitz, J. A. (1985) Science 230, 1344–1349

Cox, R. A. (1966) Methods Enzymol. 12B, 120–129

Dayhoff, M. O., Schwartz, R. M., and Orcott, B. C. (1978) in Atlas of Protein Sequence and Structure (Dayhoff, N. O., ed) Vol. 5, Suppl. 3, pp. 345–352, National Biochemical Foundation, Silver Spring, MD

Frischauf, A. M., Lehbruck, H., Poutska, A., and Murry, N. (1983) J. Mol. Biol. 170, 827–842

Gonzalez, F. J., Nebert, D. W., Hardwick, J. P., and Kasper, C. B. (1985) J. Biol. Chem. 260, 7435–7441

Gotto, O. (1990) Bull. Math. Biol. 52, 359–373

Gotoh, O., and Fujii-Kuriyama, Y. (1989) in Frontiers in Biotransformation (Ruckpaul, K., and Rein, H., eds) Vol. 1, pp. 195–245, Akademie-Verlag, Berlin, FRG

Hashimoto, T., Morohashi, K., and Omura, T. (1989) J. Biochem. (Tokyo) 105, 676–679

Henikoff, S. (1984) Gene (Amst.) 28, 351–359

Higashi, Y., Yoshioka, H., Yamane, M., Gotto, O., and Fujii-Kuriyama, Y. (1989) Proc. Natl. Acad. Sci. U. S. A. 83, 2841–2845

Imai, Y., and Nakamura, M. (1989) Biochem. Biophys. Res. Commun. 158, 717–722

Jelinek, D. F., Anderson, S., Slaughter, C. A., and Russell, D. W. (1990) J. Biol. Chem. 265, 8190–8197

Kimura, S., Hanio, N., Matsuura, E., and Gonzalez, F. J. (1989) DNA (N. Y.) 8, 503–516

Kwok, C. T., Burnett, W., and Hardie, J. R. (1981) J. Lipid Res. 22, 570–579

Labbé, D., Jean, A., and Anderson, A. (1988) DNA (N. Y.) 7, 253–260

Li, Y. C., Wang, D. P., and Chiang, J. Y. L. (1990) J. Biol. Chem. 265, 12012–12019

Maniatis, T., Hardison, R. C., Lacy, E., Lauer, J. O., Connell, C., Quon, D. S., and Elstratiades, A. (1978) Cell 15, 687–701

Matsunaga, T., Nagata, K., Holsztynska, E. J., Lapenson, D. P., Smith, A., Kato, R., Gelboin, H. V., Waxman, D. J., and Gonzalez, F. J. (1988) J. Biol. Chem. 263, 17995–18002

Means, G. D., Mahendroo, M. S., Corbin, C. J., Mathis, J. M., Powell, F. E., Mendelson, C. R., and Simpson, E. R. (1989) J. Biol. Chem. 264, 19385–19391

Messing, J., Gronenborn, B., Muller-Hill, B., and Hofschneider, P. H. (1977) Proc. Natl. Acad. Sci. U. S. A. 74, 3642–3646

Morohashi, K., Sogawa, K., Omura, T., and Fujii-Kuriyama, Y. (1987) J. Biochem. (Tokyo) 101, 879–887

Mount, S. M. (1982) Nucleic Acids Res. 10, 459–472

Myant, N. B., and Mitropoulos, K. A. (1977) J. Lipid Res. 18, 135–153

Nebert, D. W., Nelson, D. R., Adesnik, M., Coon, M. J., Estabrook, R. W., Gonzalez, F. J., Guengerich, F. P., Gunsalus, I. C., Johnson, E. F., Kemper B., Levine, W., Phillips, I. R., Sato, R., and Waterman, M. R. (1989) DNA (N. Y.) 8, 1–13

Noshiro, M., and Okuda, K. (1990) FEBS Lett. 268, 137–140

Noshiro, M., Nishimoto, M., Morohashi, K., and Okuda, K. (1989) FEBS Lett. 257, 97–100

Noshiro, M., Nishimoto, M., and Okuda, K. (1990) J. Biol. Chem. 265, 10036–10041

Ogishima, T., Deguchi, S., and Okuda, K. (1987) J. Biol. Chem. 262, 7646–7650

Picado-Leonard, J., and Miller, W. L. (1987) DNA (N. Y.) 6, 439–448

Poulos, T. L., Finzel, B. C., Gunsalus, I. C., Wagner, G. C., and Kraut, J. (1985) J. Biol. Chem. 260, 16122–16130

Poulos, T. L., Finzel, B. C., and Howard, A. J. (1987) J. Mol. Biol. 192, 887–900

Sanger, F., Coulson, A. R., Barrell, B. G., Smith, A. J. H., and Roe, B. A. (1980) J. Mol. Biol. 143, 161–178

Scaleni, T. J., and Sanghvi, A. (1983) Proc. Natl. Acad. Sci. U. S. A. 80, 2477–2480

Sneath, P. H. A., and Sokal, R. R. (1973) Numerical Taxonomy, W. H. Freeman & Co., New York

Sogawa, K., Gotto, O., Kawajiri, K., and Fujii-Kuriyama, Y. (1984) Proc. Natl. Acad. Sci. U. S. A. 81, 5066–5070

Southern, E. M. (1975) J. Mol. Biol. 98, 503–517

Suwa, Y., Mizukami, Y., Sogawa, K., and Fujii-Kuriyama, Y. (1985) J. Biol. Chem. 260, 7980–7984

Wada, F., Hirata, K., Nakao, K., and Sakamoto, Y. (1988) J. Biochem. (Tokyo) 64, 415–417

White, P. C., New, M. I., and Dupont, B. (1986) Proc. Natl. Acad. Sci. U. S. A. 83, 5111–5115

Wong, G., Kawajiri, K., and Negishi, M. (1987) Biochemistry 26, 8683–8690

Continued on next page.
**Cholesterol 7α-Hydroxylase Gene**

**Supplemental Material to**

**Structural Analysis of the Gene Encoding Rat Cholesterol 7α-Hydroxylase, the Key Enzyme for bile Acid Biosynthesis**

Masanori Shimizu, Tsumo Gotoh, Koichiro Okada, and Mitsunori Shimizu

---

**Materials and Methods**

Materials: 1,2,3-Diacylglycerol (DAG), 2,3-diacetoxy-1-propanol (DAP), and 2,3-dihydroxy-1-propene (DHP) were obtained from Sigma Chemical Co. 

**Experimental Procedures**

Materials: 1,2,3-Diacylglycerol (DAG), 2,3-diacetoxy-1-propanol (DAP), and 2,3-dihydroxy-1-propene (DHP) were obtained from Sigma Chemical Co.

**Preparation of poly(A)+</b> RNA and Genomic DNA**

Total RNA was prepared from livers of normal male rats using RNAzol B method (Innogenetics, 1994). Only poly(A) RNA was enriched by chromatography using oligo-dT-cellulose column (Sambrook 1989). Genomic DNA was isolated from rat liver by the procedure described by Wang (1997).

**Construction of genomic DNA library**

Genomic DNA from rat was partially digested with HpaI and ligated into BamHI site of pUC119 using vector packaged, and transferred to blue colonies (Eliot, 1993) by the method described by Pichon et al. (1993). The ligated DNA was transformed into competent E. coli cells and plated on LB agar plate. A genomic library was constructed from rat liver.

**DNA sequence analysis**

The subcloned genomic DNA were digested with the proper restriction enzymes in order to ligate to the YII vector (Moseng et al. 1993) as described by the manufacturer. The XbaI and SalI digestion were performed using restriction enzymes (New England BioLabs, Beverly, MA) according to the manufacturer's instructions. DNA sequencing was performed using the dideoxynucleotide chain termination method.

**Southern hybridization analysis**

DNA blotting analysis was performed under the conditions required to prevent hybridization. The probes were hybridized with 32P-labeled BamHI and HindIII and subjected to autoradiography in 3% gel electrophoresis. The separated DNA fragments were transferred to nitrocellulose and hybridized with the probes. The membranes were hybridized with the probes for 24 h at 65°C and washed in 0.1X SSC containing 0.1% SDS and exposed to x-ray films.

**Determination of transcriptional initiation site**

The position of the transcriptional start sites were determined by primer extension analysis using the primer 5'-ATGCTGGGATTATGTTTAC-3'. Autoradiographic signal was amplified by the Quantum Yellow Kit (Sangon). The RNA samples were treated with pancreatic RNase by the manufacturer's instructions. RNase treatment was performed using the appropriate dideoxynucleotide chain termination method.

---

**Figure 1. Restriction enzyme map of genomic clones of rat FUG3**

Two different clones 10G-3 and 10G-4 were isolated by restriction enzyme digestion with XhoI, PstI, EcoRI, HindIII, and BglII.

**Figure 2. Southern hybridization analysis of rat genomic DNA**

Southern hybridization analysis was performed under the conditions required to prevent hybridization. The probes were hybridized with 32P-labeled BamHI and HindIII and subjected to autoradiography in 3% gel electrophoresis. The separated DNA fragments were transferred to nitrocellulose and hybridized with the probes. The membranes were hybridized with the probes for 24 h at 65°C and washed in 0.1X SSC containing 0.1% SDS and exposed to x-ray films.

**Figure 3. Determination of transcriptional initiation site**

A picture of an electrophoretic gel showing the 10G-3 clone fragment on a 8% polyacrylamide gel is shown. The RNA samples were treated with pancreatic RNase by the manufacturer's instructions. RNase treatment was performed using the appropriate dideoxynucleotide chain termination method.

---

**Figure 4. Determination of transcriptional initiation site**

A picture of an electrophoretic gel showing the 10G-3 clone fragment on a 8% polyacrylamide gel is shown. The RNA samples were treated with pancreatic RNase by the manufacturer's instructions. RNase treatment was performed using the appropriate dideoxynucleotide chain termination method.

---

**Figure 5. Northern hybridization of rat liver cholesterol 7α-hydroxylase gene**

The nucleotide sequence corresponding to the six exons and their flanking regions were amplified as described in Fig. 1. Lanes 1-3 indicate nucleotide positions from the transcription start site. The positions of RNA size markers are indicated on the right of the gel. The consensus sequence that is conserved among all the mammal species is indicated. The consensus sequence for all the mammal species is indicated.
