J. Biochem. 122, 1114-1121 (1997)

Structural and Evolutionary Studies on Sterol 14-Demethylase P450 (CYP51), the Most Conserved P450 Monooxygenase: I. Structural Analyses of the Gene and Multiple Sizes of mRNA[1]

Mitsuhide Noshiro,* Yuri Aoyama,† Takeshi Kawamoto,* Osamu Gotoh,‡ Tadao Horiuchi,† and Yuzo Yoshida§,2

*Department of Biochemistry, Hiroshima University School of Dentistry, Hiroshima 734; †Department of Bioengineering, Faculty of Engineering, Soka University, Hachioji, Tokyo 192; ‡Department of Biochemistry, Saitama Cancer Center Research Institute, Ina-machi, Saitama 362; and §School of Pharmaceutical Sciences and Interdisciplinary Research Institute for Biosciences, Mukogawa Women's University, Nishinomiya 663

Received for publication, June 19, 1997

The structure of rat *CYP51* gene encoding sterol 14-demethylase was examined. The *CYP51* gene spanned about 18 kb and contained 10 exons. The copy number of *CYP51* in the rat genome was determined to be one. In addition, one *CYP51* processed (intron-less) pseudogene covering the coding and *ca.* 600-bp 3′-noncoding sequences of CYP51 cDNA was found in the rat genome. Multiple transcription initiation sites were predicted by primer extension and 5′-RACE methods using poly(A)$^+$ RNA from liver, ovary, and testis, and the major ones were located at 126 and 123 nucleotides upstream from the initiation ATG codon. The primer extension also showed several minor sites around the major ones. In addition to these sites, other minor initiation sites were also predicted at around 330 and 460 nucleotides upstream from the initiation ATG codon. No TATA box was found in the putative promoter region, but multiple GC boxes were found around the cap sites, supporting the previously inferred housekeeping nature of *CYP51* gene and the existence of the multiple transcription initiation sites. A few consensus transcription regulatory elements such as CRE were found in the 5′-flanking region. Four polyadenylation signals were found in the 3′-noncoding region by the 3′-RACE method. Three of them were used to generate 3.1-, 2.7-, and 2.3-kb mRNAs in liver and ovary. The remaining one was used only in testis to generate 1.9-kb mRNA having an unusually short trailer sequence, suggesting a specific regulatory mechanism for generating CYP51 mRNA in testis different from that in liver and ovary.

Key words: gene structure, P450, processed pseudogene, sterol 14-demethylase, transcription-initiation site.

Sterol 14-demethylase P450 (CYP51) catalyzing the removal of 14-methyl group of 14-methylsterols has been found in mammals, higher plants, and fungi (*1*). Removal of the 14-methyl group is essential for generating functional eukaryotic steroids, because no functional steroid has a 14-methyl group in its structure. Therefore, eukaryotic sterol 14-demethylases participating in *de novo* synthesis of sterols are considered to be functionally equivalent orthologous enzymes, although the known natural substrates for these enzymes vary somewhat from one organism to another (*1, 2*). Amino acid sequence alignment and phylogenetic analysis have demonstrated that sterol 14-demethylases of divergent phyla form one unique family (*CYP51*) clearly distinct from other members of the P450 gene superfamily (*CYP* superfamily), supporting the notion that all CYP51s might be orthologous enzymes derived from a common ancestor with a conserved metabolic role (*3*). Thus, *CYP51* is considered to be an ancient P450 species that arose before the divergence of eukaryotes and has been conserved throughout the eukaryotic evolution (*3*). *CYP51* is the only P450 family known to be distributed in eukaryotic phyla from fungi to mammals.

It is generally accepted that the basic structure of genes in each P450 family is well conserved (*4*). From this, there arises the interesting question of whether this empirical rule is applicable to the widely distributed *CYP51* family members. The comparison of gene structures of *CYP51*s of evolutionarily distant phyla is necessary to solve this problem. Although gene structures of a few fungal *CYP51*s have been reported (*5-8*), little information is available on the mammalian *CYP51*.

As preliminarily described in the previous paper (*3*) and also reported by Rozman *et al.* (*9*), mammalian *CYP51* is considered to be expressed in germ line cells. In addition,

[1] Sequence data reported in this paper have been entered in the DDBJ/EMBL/GenBank database under the following numbers: D78390 and D87997 for rat CYP51 pseudogene clones RG3 and RG41, respectively; AB004087-AB004096 for exons 1-10 of rat CYP51 gene, respectively.

[2] To whom correspondence should be addressed. Tel: +81-798-47-1212 (Ext. 531), Fax: +81-798-41-2792, E-mail: yoshiday@mwu.mukogawa-u.ac.jp

Abbreviations: CRE, cyclic AMP responsive element; RACE, rapid amplification of cDNA ends protocol; bp, base pair(s); kb, kilobases.

ubiquitous expression of CYP51 in every human tissue studied thus far has been described (*10*). These facts suggested the ubiquitous expression of CYP51 as a housekeeping enzyme both in somatic and germ line cells. This is another unique characteristic of CYP51, because most known P450s responsible for various differentiated functions are expressed in limited organs or tissues related to their functions (*2*), and most of these P450s are inducible enzymes. Therefore, analysis of the structure of the *CYP51* gene responsible for the regulation of its expression is another interesting problem to be solved.

This paper gives a full account of the results of structural analysis of the functional gene and a processed pseudogene of rat *CYP51*. It also describes the occurrence of multiple sizes of CYP51 mRNA in liver, ovary, and testis.

MATERIALS AND METHODS

Preparation of Poly(A)$^+$ RNA and Genomic DNA—Total RNAs were prepared from liver, ovary, and testis of normal Wistar rats at 10:00 p.m. by guanidine HCl method (*11*), and poly(A)$^+$ RNA was enriched by chromatography using oligo(dT)-cellulose column (*12*). Genomic DNA was isolated from rat spleen by the procedure described by Wong (*13*).

Construction of Genomic DNA Library—Genomic DNA from rat was partially digested with *Mbo*I and ligated into *Bam*HI site of EMBL3 phage vector (Stratagene, La Jolla, CA), packaged, and transfected to host *Escherichia coli* XL1-Blue MRA(P2) by the method described by Frischauf *et al.* (*14*). The libraries were plated and screened with ^{32}P-labeled total insert of CYP51 cDNA (pRT11) (*3*) by the plaque hybridization method (*15*). Positive clones were isolated, digested with proper restriction enzymes, and subcloned into pBluescript II SK(−) phagemid vector (Stratagene).

DNA Sequence Analysis—The subcloned genomic DNAs were digested with the proper restriction endonucleases in order to ligate them to the pBluescript II SK(−) phagemid vector. Double stranded DNAs of the subcloned pBluescript were used for sequencing templates. Chain termination reaction was performed according to the method described by Sanger *et al.* (*16*) using either [^{35}S]dCTP and a Sequenase 7-deaza-dGTP DNA Sequencing Kit (Amersham Life Science, Buckinghamshire, England) or fluorescent-labeled sequence primers and a BcaBEST dideoxy sequencing kit (Takara Shuzo, Otsu), and analyzed by polyacrylamide gel electrophoresis or an autosequencer (Hitachi SQ-5500), respectively.

Southern and Northern Hybridization Analyses—Rat genomic DNA (20 μg) and isolated genomic clones (10 ng) were digested with *Eco*RI or *Pst*I and subjected to agarose (0.8%) gel electrophoresis by the Southern method (*17*). The separated DNA fragments were transferred to Hybond-N nylon membrane (Amersham Life Science). The membrane was hybridized with the ^{32}P-labeled *Eco*RI-*Hin*dIII fragment (1.7 kb) of pRT11 in hybridization solution containing 6×SSC, 5×Denhardt's, 10 mM EDTA, 1% SDS, and 0.5 mg/ml sonicated salmon sperm DNA at 68°C. The membranes were washed with 0.1×SSC containing 0.5% SDS and exposed to X-ray films. For Northern hybridization, poly(A)$^+$ RNA (5 μg) for each sample was denatured with 2.2 M formaldehyde and 50% formamide, electrophoresed on 1% agarose gel containing 2.2 M formaldehyde as described by Thomas (*18*), then transferred to Hybond-N nylon membrane. Conditions for hybridization and washing were the same as described above.

Determination of Transcriptional Initiation Site—The location of the cap site of CYP51 mRNA was determined by a primer extension method. Primers DM30 (5′-AAGCTGGAGGCCGAGGTCCCTGCCGCTTCT-3′, corresponding to +70 to +99, 0.5 pmol) and DM40 (5′-TCCCGACCGCTGCTTCGATCCGCCCGGTAA-3′, corresponding to +98 to +127, 0.5 pmol) were endlabeled by ^{32}P (1×10^5 cpm) with T4 polynucleotide kinase, annealed to 1 μg of poly(A)$^+$ RNA prepared from rat liver and testis, and extended with 200 units of SuperScript II (Life Technologies, Gaithersburg, MD) according to the method of Trienzenberg (*19*). The size of radioactive extended DNA was analyzed on 6% acrylamide gel, and sequencing ladders were primed from the same synthesized oligonucleotides.

5′- and 3′-Rapid Amplification of cDNA End (5′- and 3′-RACE)—Double stranded cDNAs were synthesized from 1 μg of poly(A)$^+$ RNA preparations by use of a Marathon kit (Clontech, Palo Alto, CA) according to the instruction manual. Both 5′- and 3′-ends of the cDNA were ligated to the 44-mer adapter containing two overlapping primer sites for AP1 (5′-CCATCCTAATACGACTCACTATAGGGC-3′) and AP2 (5′-ACTCACTATAGGGCTCGAGCGGC-3′) supplied by the manufacturer. For 5′-RACE, PCR was performed with the cDNA using Klen Taq polymerase (Clontech), an antisense primer DM24 (5′-TAGACCTCTTCCGCATTCAGGTCTT-3′) and AP1 primer as 5′-end primer (Fig. 5). The PCR products were used for the 2nd PCR using a nested antisense primer DM26 (5′-ACACAGGTCCGTACTTCTCATATGC-3′) and the nested 5′-end primer AP2. For 3′-RACE, the 1st PCR was primed using a sense primer DM1 (5′-TGAATGCGGAAGAGGTCTAC-3′) and AP1 as a 3′-end primer, and the 2nd PCR was primed with a nested sense primer DM25 (5′-AAGCTGCAGCTTGGCAGAGAATGAA-3′) or DM15 (5′-AGAGCCCATCGAGAGATCAA-3′) for liver and testis products, respectively, and AP2 as a nested 3′-end primer. After denaturing the template cDNA at 94°C for 1 min, PCR was carried out by denaturation at 94°C for 30 s, annealing at 60°C for 30 s and elongation at 68°C for 4 min for each cycle. After repeating the cycle for 30 times, prolonged elongation was carried out at 68°C for 7 min. For the nested PCR, the cycle was repeated for 25 times. The PCR products were separated by agarose gel electrophoresis and analyzed by Southern hybridization (*17*). The second PCR products were subcloned into pGEM-T-Vector (Promega, Madison, WI) for sequencing analysis.

Materials—[^{32}P]dCTP (111 TBq/mmol), [^{32}P]ATP (111 TBq/mmol), and [^{35}S]dCTP (50 TBq/mmol) were obtained from Du Pont-New England Nuclear. Oligolabeling kit was a product of Pharmacia LKB Biotechnology. Oligonucleotides were synthesized by Kurabo (Osaka).

RESULTS AND DISCUSSION

Number of CYP51 Gene in Rat Genome—Ten positive clones were isolated from 3×10^6 plaques by screening EMBL3 genomic libraries using the insert of the rat CYP51 cDNA (pRT11) (*3*) as a probe and subjected to restriction mapping analysis. Seven out of 10 isolated clones showed

restriction enzyme sites of *Bam*HI, *Eco*RI, and *Sal*I at least partly overlapping with each other, and the entire structure of the *CYP51* gene was finally constructed by overlapping 2 clones (RG25 and RG40) as shown in Fig. 1. Three other clones (RG3, RG32, and RG41) showed different restriction enzyme sites to the constructed gene. The restriction enzyme map of the largest clone (RG41) is shown in the figure for comparison. As already suggested (*3*) and discussed in the next section, these clones are derived from a processed pseudogene of *CYP51*.

To confirm the size and complexity of *CYP51* gene, we performed Southern blot analysis of total rat genomic DNA and the isolated genomic clones using *Eco*RI-*Hin*dIII fragment (1.7 kb) of the CYP51 cDNA (pRT11) as a probe. The hybridized signals with the cDNA showed a simple hybridization pattern on the *Eco*RI or *Pst*I digests of the genomic DNA (Fig. 2A). The hybridized *Eco*RI-digested fragments were 7.9, 3.7, 2.5, 2.3, 2.1, 1.6, and 0.6 kb in size. The sizes of these hybridized bands, except that of 3.7 kb, are in good agreement with those of the *Eco*RI fragments of either RG25 (8.1, 2.3, 2.0, 1.6, and 0.62 kb) or RG40 (2.5, 2.2, 2.0, and 0.62 kb) (Fig. 2B). The hybridized *Pst*I-digested fragments were 5.7, 2.6, 1.5, 0.63, 0.55, 0.53, and 0.30 kb in size. Similarly, the sizes of these hybridized bands, except that of 1.5 kb, are in good agreement with those of the *Pst*I fragments of either RG25 (5.7, 2.6, 0.63, 0.53, and 0.30 kb) or RG40 (2.6, 0.63, 0.55, 0.53, and 0.30 kb) (Fig. 2C). The 3.7-kb *Eco*RI-digested and 1.5-kb *Pst*I-digested bands corresponded to the 3.7-kb *Eco*RI- and 1.5-kb *Pst*I-fragments of RG41, respectively, which was identified as a processed pseudogene of *CYP51*, as discussed in detail in the next section. The processed (intron-less) pseudogene nature of these fragments was supported by their high density in the Southern hybridization (Fig. 2A), since the intron-less fragments contain longer hybridizable areas with the cDNA probe as shown in Fig. 1. These facts indicate that the rat genome probably includes a single copy of the functional *CYP51* gene and one intron-less processed pseudogene.

Characterization of a CYP51 Processed Pseudogene—In the previous paper (*3*), we described that RG3 contained a

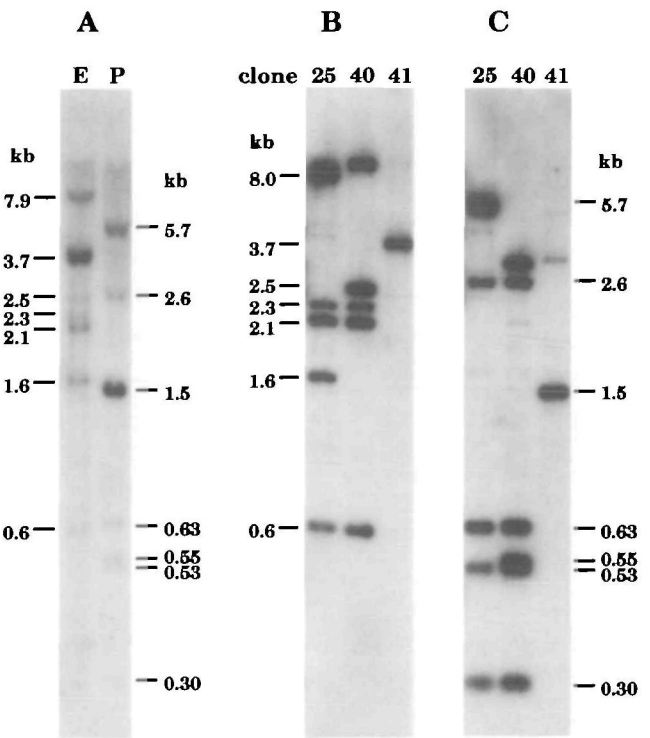

Fig. 2. **Southern hybridization analysis of rat total genomic DNA and the isolated genomic clones.** Experimental conditions are described in "MATERIALS AND METHODS." The membranes were hybridized with ^{32}P-labeled *Eco*RI-*Hin*dIII fragment (1.7 kb) of pRT11 cDNA (5×10^7 cpm of total radioactivity). Sizes of hybridized bands in total genomic DNA and the corresponding fragments of the isolated clones are indicated on both sides. A, genomic DNA digested with *Eco*RI (E) or *Pst*I (P). B, the isolated genomic clones (RG26, RG40, and RG41) digested with *Eco*RI; and C, with *Pst*I. kb, kilobase pairs.

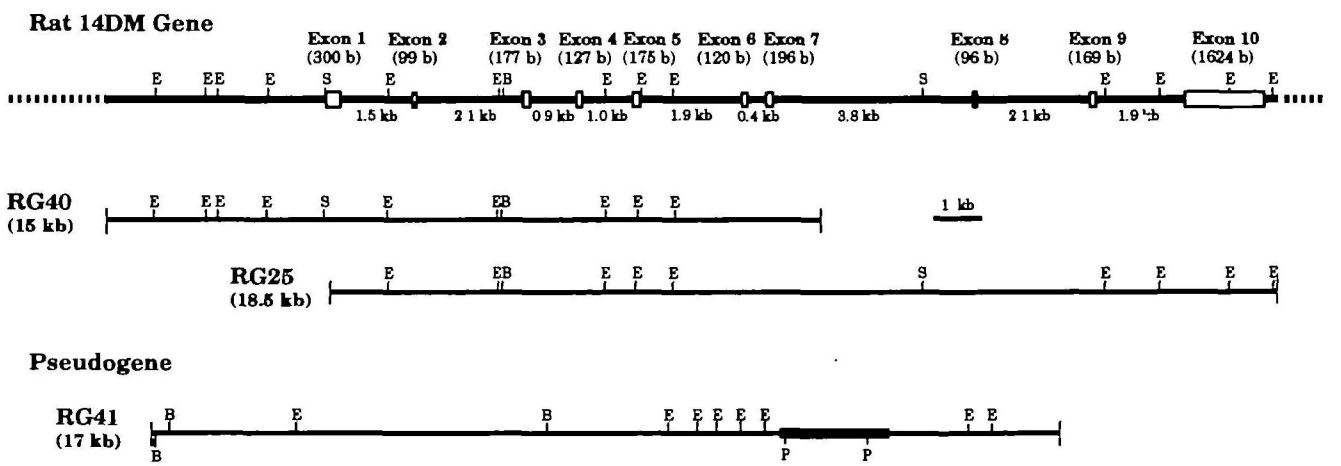

Fig. 1. **Restriction enzyme map of genomic clones of rat lanosterol 14-demethylase.** Two different clones (RG25 and RG40) isolated from rat genomic libraries were used to construct an entire map of lanosterol 14-demethylase gene as described in "MATERIALS AND METHODS." A map of a clone containing the processed pseudogene (RG41) is also indicated for comparison. Restriction enzyme sites shown in the figure are as follows: E, *Eco*RI; B, *Bam*HI; P, *Pst*I; S, *Sal*I. Open boxes in the constructed gene indicate exons, sizes of which are shown in the parenthesis. Thick bar (*ca.* 2.2-kb long) in RG41 indicates highly identical region to the cDNA sequence. kb, kilobase pairs; b, base pairs.

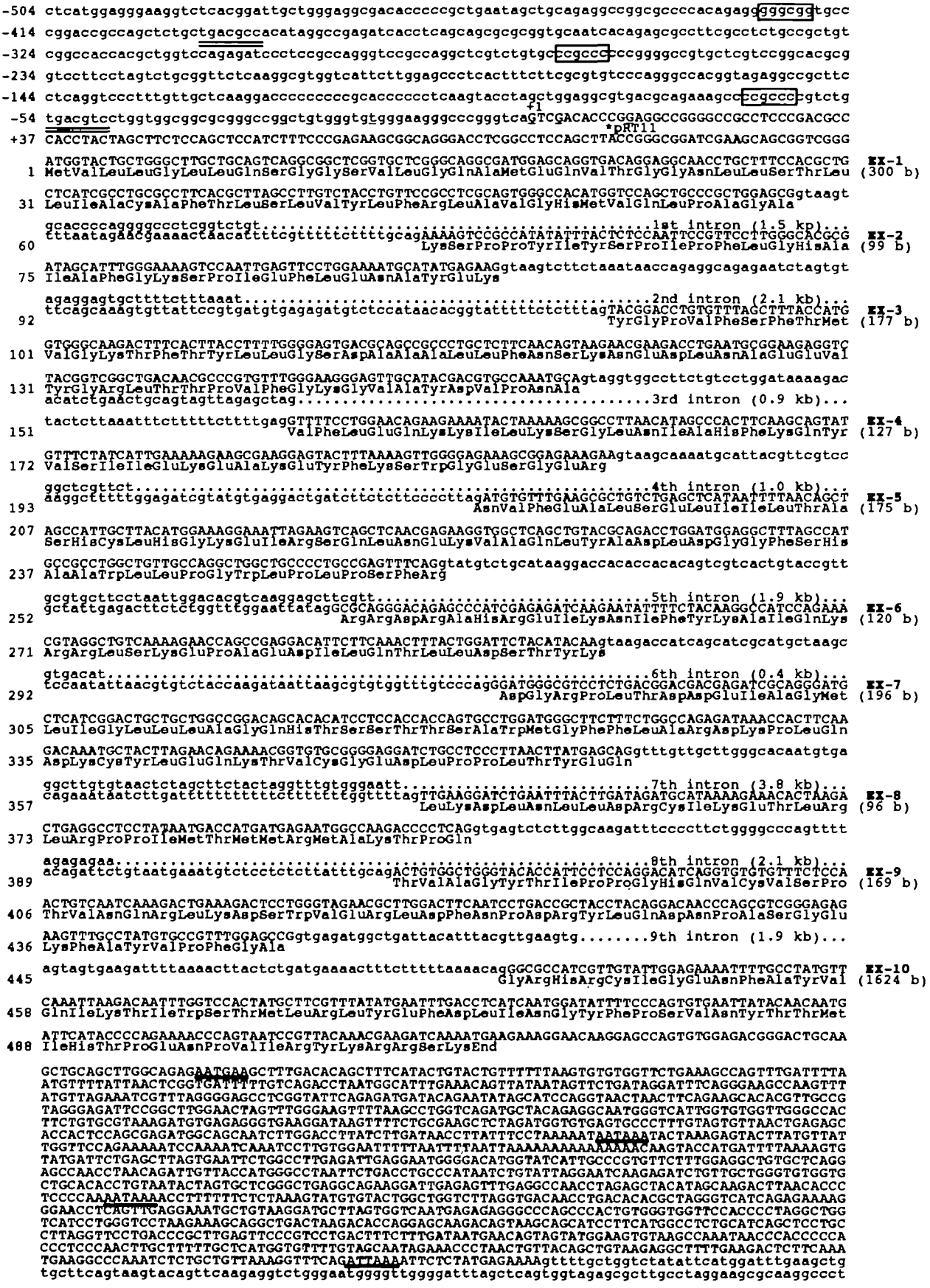

Fig. 3

nucleotide sequence (DDBJ accession number D78390) covering the entire coding region of CYP51 cDNA, and identified it as a processed intron-less pseudogene of *CYP51*. To obtain further information about the *CYP51* processed pseudogene, the nucleotide sequence of the 3.7-kb *Eco*RI fragment of RG41, which corresponds to the distinct band observed in the Southern blot analysis of the *Eco*RI digests of total rat genomic DNA (Fig. 2), was determined (sequence: DDBJ accession number D87997). The nucleotide sequence of this fragment could be aligned with the CYP51 cDNA (pRT11) and the overlapping region covered from the initiating ATG to the 2nd poly-adenylation signal (AATAAA) of the cDNA (Figs. 3 and 5). However, the open reading frame was disturbed by considerable numbers of deletions, insertions, and substitutions; the numbers of deleted, inserted, and substituted nucleotides observed in the coding region of RG41 were 32, 6, and 49, respectively. The nucleotide sequence difference between the aligned regions of RG41 and the cDNA was calculated to be 3.7%. These facts clearly indicated that the 3.7-kb *Eco*RI fragment of RG41 contained a processed intron-less pseudogene of *CYP51* that might be formed by the reverse transcription of a CYP51 mRNA corresponding to the 2.3-kb Northern hybridization band (Fig. 7). The sequences flanking the pseudogene are rich in dispersed repetitive sequences such as SINEs (B1-elements) (*20*) and possible relics of integrated retrovirus genome (data not shown). Recently, Rozman *et al.* reported the occurrence of two *CYP51* processed pseudogenes in the human genome that were localized in different chromosomes (*9*). We isolated three clones (RG3, RG32, and RG41) that might contain a *CYP51* processed pseudogene from the rat genomic DNA library. However, no significant difference was observed among the nucleotide sequences corresponding to the coding region of these clones. Furthermore, only a 3.7-kb band corresponding to the *Eco*RI fragment of RG41 was identified as a pseudogene band in the Southern blotting of the *Eco*RI-digested total genomic DNA (Fig. 2). These facts suggested that clones RG3 and RG32 contain only fragments of the pseudogene existing in RG41. Therefore, at present, we conclude that the rat genome has only one *CYP51* processed pseudogene.

Organization of Rat CYP51 Gene—As shown in Fig. 1, the functional rat *CYP51* gene spanned approximately 18 kb in size and contained 10 exons. Exons 1 through 10 consisted of 300, 99, 177, 127, 175, 120, 196, 96, 169, and 1624 bp, respectively (Figs. 1 and 3). The size of exon 1 varies because of the multiple transcriptional initiation sites as described in the next section. Nucleotide sequences of all exon/intron boundaries followed the typical GT/AG rule (Fig. 3). Flanking sequences of 5′- and 3′-splicing sites in the introns are well in accordance with the splice junction consensus sequences proposed by Mount (*21*). Nucleotide sequences of the exons were completely identical with those of the CYP51 cDNA (pRT11) (*3*) and the RACE products shown in Fig. 5. The number of exons of rat *CYP51* gene was in the range (3-13 exons) of those reported for other mammalian P450 genes (*4*). Like most other P450s, the heme-binding region and the entire 3′-untranslated region of rat CYP51 gene are contained in the last exon. The number of exons and the intron insertion sites are completely identical with those of the human orthologue (Rozman and Waterman, personal communication) (*22*), but are quite different from those of lower eukaryotic *CYP51*s, as discussed in the accompanying paper.

Transcriptional Initiation Sites and Related Structures in the 5′-Flanking Region—To define the start site for transcription, a primer extension was performed using poly(A)$^+$ RNA of rat liver and testis. When template mRNA preparations were primed with DM40 oligonucleotide, two major extended fragments ending at nucleotides G (its position was designated as +1) and G (+4) were observed at 126 and 123 nucleotides upstream from the ATG translation initiation codon, respectively, as shown in Fig. 4. Several minor extended fragments were also observed around the major ones as indicated in the figure. When the mRNA templates were primed with DM30 oligonucleotide, the extended fragments were localized to the same positions determined with DM40 primer (data not shown), confirming the reproducibility of the multiple

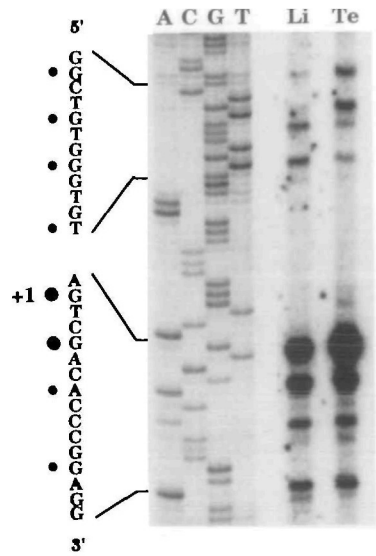

Fig. 4. **Determination of transcriptional initiation site.** The primer extension experiments were made with an antisense primer DM40 and poly(A)$^+$ RNA prepared from rat liver and testis, as described in "MATERIALS AND METHODS." To determine the position, the genomic fragment of RG40 was subjected to dideoxy chain termination reaction by using the same 30mer primer DM40 and electrophoresed on the same acrylamide gel. The nucleotide sequences around the start site sequence are represented by the 5′ to 3′ orientation from top to bottom. The large closed circles indicate the major initiation sites of transcription and the small ones indicate the minor ones. The first major initiation site was designated as +1.

Fig. 3. **Nucleotide sequence of rat lanosterol 14-demethylase gene.** The nucleotide sequence corresponding to the ten exons and their flanking regions were determined. Numbers to nucleotide in the first exon and 5′-flanking region indicate nucleotide positions from the major transcription start site. The putative transcriptional start sites determined in the experiment of Fig. 4 are indicated by underlines at positions of +1 and +4. The exonic sequences are shown by capital letters; the 5′- and 3′-flanking sequences and intron sequences are indicated by small letters. The predicted amino acid sequence is indicated below the corresponding nucleotide sequence, and the numbers from the initiation Met are indicated at the left side. Three putative GC boxes are boxed. Two putative CRE-like sequences are double underlined. Four poly(A)$^+$ additional signals determined in the experiments of Fig. 8 are underlined in the 10th exon. kb, kilobase pairs; b, base pairs.

transcriptional initiation sites. The positions of these potential cap sites were also assured by the 5′-RACE method. A PCR using double-stranded template cDNA prepared by the Marathon method, the antisense primer DM24 and the 5′-end primer AP1 generated a few DNA fragments. The second PCR was performed using the first PCR products and the nested primers DM26 and AP2. The final products were subjected to the Southern blot analysis using a 140-bp 5′-end fragment of pRT11 cDNA and a 287-bp DNA fragment, corresponding to positions +10 to +149 and −331 to −45 of the gene, respectively (Figs. 3 and 5). A major (*ca.* 450 bp) and a minor (*ca.* 650 bp) band were hybridized to the 140-bp probe as shown in Fig. 6A, although the minor 650-bp band was not obvious for ovary and testis. When the same samples were hybridized with the 287-bp probe and the exposure time to the X-ray film was prolonged, the 650- and 780-bp bands were observed in all three samples (Fig. 6B). The length of the major fragment (450 bp) corresponds to the sum of the length between the above-mentioned major cap sites at +1 and the 3′-end of DM26 (+412) and the size of the attached adapter (36 bases), confirming the position of the major cap sites assumed by the primer extension. Based on the same estimation, the 5′-end of minor two bands (650 and 780 bp) corresponded to the positions of around at 330 and 200 bp upstream from the major transcriptional initiation site (+1). We also observed the minor extended bands at positions −219, −167, −156, and −155, corresponding to the 5′-end position of 650-bp 5′-RACE product, when mRNA from the three tissues was primed with DM34 oligonucleotide (−74 to −45, data not shown). These results indicate the existence of minor potential transcriptional initiation sites upstream of the major ones.

To confirm the initiation sites, the major 450-bp product of 5′-RACE and the minor 650- and 780-bp products from liver, ovary, and testis were subcloned into pGEM-T-Vector. Several clones were isolated for each size and source and subjected to nucleotide sequencing. The 5′-end nucleotide sequences of the clones derived from the 450-bp products of liver, ovary, and testis were all started at positions between −18 and +19 corresponding to the region of major transcriptional initiation sites. The 5′-end sequences of the minor 650-bp product from liver and the 780-bp product from testis started at positions −198, −195, and −181 for the 650-bp product and at −330, −322, and −319 for the 780-bp product, respectively. These results confirm the actual positions of the major and minor transcriptional initiation sites and also indicated that these sites start at the same positions in liver, ovary, and testis.

In the region preceding the putative cap site, the sequence TATA(A/T)A, the expected TATA box, was not observed. Instead, the 5′-flanking region is rich in GC nucleotides, and the sequences GGGCGG or CCGCCC, possible candidates for a GC box (SP1 site), are found at positions −425, −261, and −66. The nucleotide-sequence alignment with the 5′-flanking region of human *CYP51* gene

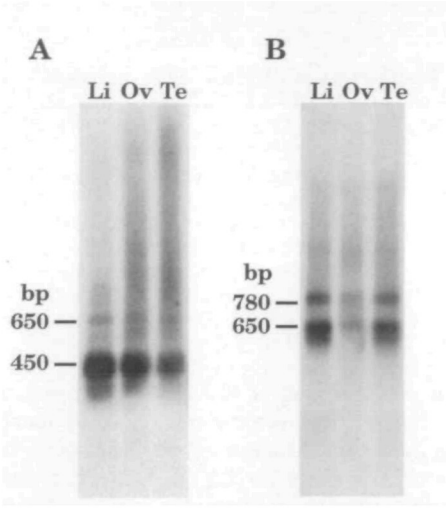

Fig. 6. **5′-RACE analysis for CYP51 mRNA of rat liver.** After denaturing the template cDNA from rat liver at 94°C for 1 min, PCR was carried out with DM24 and AP1 primers as shown in Fig. 5 and described in "MATERIALS AND METHODS." The nested PCR was primed with DM26 and AP2. The final products were subjected to 1.8% agarose gel electrophoresis and Southern blot analysis with ^{32}P-labeled 140-bp 5′-end fragment of pRT11 (corresponding to +10 to +149) (A), or 287-bp fragment (corresponding to −331 to −45) (B). The hybridized membranes were exposed to the X-ray films for 1 and 5 h for experiments A and B, respectively. The sizes of hybridized bands are indicated at left of figures. Li, liver; Ov, ovary; Te, testis; bp, base pairs.

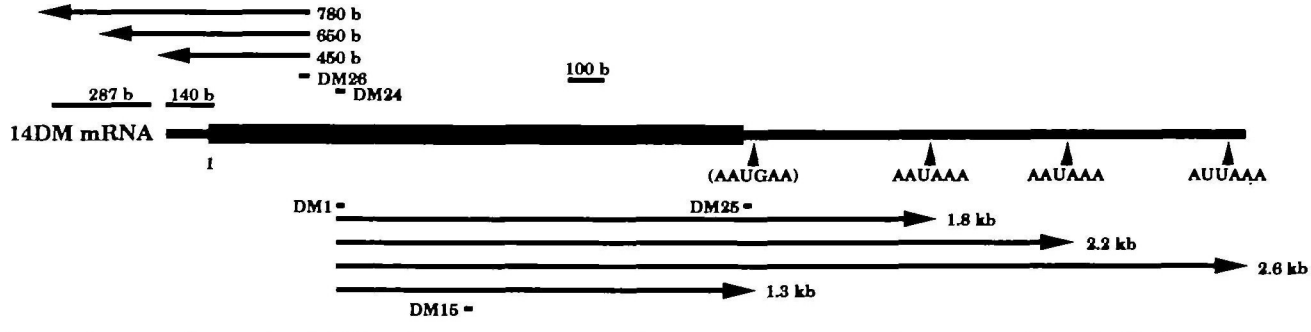

Fig. 5. **Strategy of 5′- and 3′-RACE and resultant positions of poly adenylation sites.** Double stranded cDNAs from liver, ovary, and testis were primed with DM24 and DM1 for 5′- and 3′-RACE, respectively, as described in "MATERIALS AND METHODS." The 1st PCR products were subjected to the 2nd PCR with the nested primers DM26 and DM25 for 5′- and 3′-RACE, respectively. For 3′-RACE of testis, the 2nd PCR was carried out with DM15. AP1 and AP2 primers were used as counter primers for the 1st and 2nd PCR, respectively. The arrows indicate the positions and sizes of the resultant fragments. The full-length cDNA is shown by a bar, and the coding region is indicated by a thick bar. The bars of 287 b and 140 b indicate the probes used to analyze the 5′-RACE products. kb, kilobase pairs; b, base pairs.

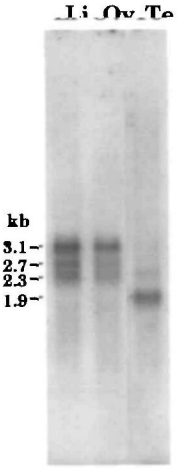

Fig. 7. **Northern hybridization with lanosterol 14-demethylase cDNA of poly(A)$^+$ RNA from liver, ovary, and testis of rats.** Poly(A)$^+$ RNA samples (5 μg per lane) were electrophoresed on 1% agarose gel containing 2.2 M formaldehyde and transferred to Hybond-N nylon membrane. The membrane was hybridized with the oligolabeled ^{32}P-*Eco-Hin*d (1.7 kb) fragment of cDNA pRT11 (5×10^7 cpm of total radioactivity), washed with 0.1×SSC containing 0.5% SDS, then exposed to X-ray film for 14 h at −80°C. The sizes of hybridized bands are indicated in the figure. Li, liver; Ov, ovary; Te, testis; kb, kilobases.

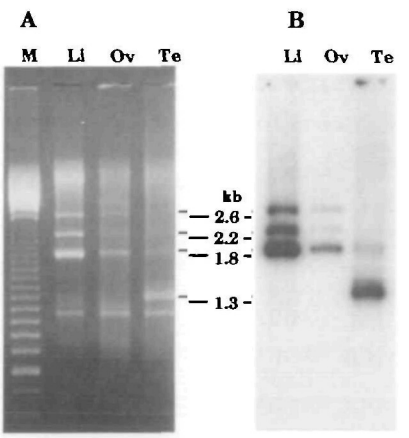

Fig. 8. **3′-RACE analysis to detect polyadenylation sites in liver, ovary, and testis.** Template cDNAs from liver, ovary, or testis were amplified with primers DM1 and AP1 as shown in Fig. 5, and described in "MATERIALS AND METHODS." The products were subjected to 1.2% agarose gel electrophoresis and Southern blot analysis with ^{32}P-labeled *Eco*RI-*Hin*dIII (1.7 kb). The sizes of hybridized bands are indicated in the figure. A, ethidium bromide staining; B, hybridization with the ^{32}P-probe. M, marker; Li, liver; Ov, ovary; Te, testis; kb, kilobase pairs.

determined by Rozman *et al.* (*22*) showed that the first and third GC box sequences were conserved at nearly the same positions in the human gene. Since human *CYP51* gene was also reported to have no TATA box (*22*), these GC box motif(s) may play an important role in the transcription of mammalian *CYP51* genes. Most P450 genes studied thus far, except for *CYP8* (*23*) and *CYP19* (*24*), have typical TATA boxes. *CYP51* is the third example of a rare TATA-less P450 gene. Multiple copies of the GC box in the transcriptional initiation region without a TATA box have been found dominantly in ubiquitously expressed genes, and this structure is considered to be characteristic of housekeeping genes (*25*). Therefore, the structure of the transcriptional initiation site of mammalian *CYP51* gene coincides well with its predicted housekeeping nature (*3*) and ubiquitous expression in human organs (*10*). Homology search against transcriptional factor sequence databases showed the occurrence of a few consensus transcription regulatory elements including CRE in the 5′-flanking region of the rat *CYP51* gene (Fig. 3). Two typical CREs, TGACGCC and TGACGTC found at −398 and −54, respectively, were located close to two GC boxes (−425 and −66) commonly found both in the rat and human genes, and the sequence TGACGC was conserved at the same sites of the human gene. This fact seems to suggest some regulatory role of these putative CREs. Recently, it was suggested that the lanosterol 14-demethylase activity of rat ovaries was regulated by gonadotropins (*26*). Since gonadotropins may act through an intracellular signal transduction system including cAMP, these CREs may participate in this regulation.

Multiple Poly-Adenylation Sites and Their Tissue-Specific Usage—Three distinct bands (3.1, 2.7, and 2.3 kb) for CYP51 mRNA were observed in Northern blotting with rat liver and ovary poly(A)$^+$ RNA preparations, whereas one major 1.9-kb band and a faint 2.3-kb band were observed in the same experiment with poly(A)$^+$ RNA preparations from testis (Fig. 7). Similar results were reported by Stromstedt *et al.* (*10*) with the human preparations. These findings indicate that at least four mRNA species of different sizes are formed for mammalian CYP51, and the mechanisms for determining the mRNA size in testis may be different from those in liver and ovary. Since the difference between the various types of mRNA (*ca.* 400 bp) is too large to be attributed to possible differences in the size of the 5′-noncoding region, as described in the preceding section, these mRNAs must have different lengths of the 3′-noncoding regions due to the multiple poly-adenylation signals in the last exon. To confirm this possibility, 3′-RACE was carried out by using combination of DM1 and AP1 primers. When PCR was primed with DM1 and AP1 (Fig. 5), 3′-RACE products from liver and ovary poly(A)$^+$ RNA hybridizable to the cDNA probe (pRT11) were 1.8, 2.2, and 2.6 kb in size, whereas major 1.3-kb and faint 1.8-kb bands were obtained from the testis mRNA (Fig. 8). As the primer DM1 is located at the position of 491–510 nucleotides of the cDNA (Fig. 5), the sizes of those 3′-RACE products are in good agreement with the sizes of mRNAs observed by the Northern blotting (Fig. 7). The three 3′-RACE products from the liver specimen and the 1.3-kb product from testis were subjected to the 2nd PCR using AP2 paired with DM25 or DM15, respectively. The products were subcloned into pGEM-T-Vector and sequenced. The largest 3′-RACE product from the liver specimen exhibited an extra 831 nucleotides with a 27-base poly A tail that were extended from the 3′-end of the longest CYP51 cDNA clone (pRT11) (*3*). The sequence of the extended 804 nucleotides was completely identical to that of the genomic DNA immediate downstream of the region overlapping the 3′-end of the cDNA. Based on these facts, we have ascribed the position corresponding to the end of the extended 804 nucleotides as the 3′-end of exon

10. Sequences matching the 3′-end parts of other shorter products from the liver specimen were found inside of exon 10, and a 20- to 30-bp long poly A tail was attached to the 3′-end of each product. The two typical (AATAAA) and one modified (ATTAAA) polyadenylation signals existing in exon 10 (Figs. 3 and 5) were found at 17-22 bp upstream from the poly A tail of each of the three 3′-RACE products. It can thus be concluded that the three different sizes of CYP51 mRNAs occurring in liver and ovary (Fig. 7) are attributed to the existence of multiple polyadenylation signals. Sequencing of the 3′-RACE products from testis specimen revealed that another polyadenylation signal (AATGAA) close to the coding region (Fig. 5) was used in testis to generate the 1.9-kb mRNA. Since the sequence of the testis 3′-RACE product was identical to that of the overlapping region of the cDNA, the 1.9-kb mRNA expressed in testis is a transcript from the same gene as in liver and ovary.

The biological significance of the existence of multiple mRNAs with different sizes of the 3′-noncoding region may be related to the stability of mRNA. The unusually short trailer sequence of the mRNA expressed in testis must have some meaning for the exertion of particular functions of lanosterol 14-demethylase in this tissue. Germ line cells occupy a considerable portion of testis, and the *CYP51* gene must be expressed in germ line cells as previously suggested (*3*). Therefore, the short CYP51 mRNA specifically observed in testis may represent the particular regulatory mechanism for the transcription and/or post transcriptional processing of *CYP51* gene in male germ line cells. More experiments are clearly needed to examine this hypothesis.

Dr. K. Nakamasu, Dr. M. Shen of Hiroshima University and Mr. T. Kashiyama of Soka University are acknowledged for their technical assistant in cloning and sequencing of rat CYP51 genes.

REFERENCES

1. Yoshida, Y. (1993) Sterol biosynthesis in *Cytochrome P450* (Schenkman, J.B. and Grein, H., eds.) pp. 627-639, Springer-Verlag, Berlin
2. Yoshida, Y. and Aoyama, Y. (1994) The P450 superfamily: A group of versatile hemoproteins contributing to the oxidation of various small molecules in *Regulation of Heme Protein Synthesis* (Fujita, H., ed.) pp. 75-88, AlphaMed Press, Dayton
3. Aoyama, Y., Noshiro, M., Gotoh, O., Imaoka, S., Funae, Y., Kurosawa, N., Horiuchi, T., and Yoshida, Y. (1996) Sterol 14-demethylase P450 (P45014DM) is one of the most ancient and conserved P450 species. *J. Biochem.* **119**, 926-933
4. Gotoh, O. (1993) Evolution and differentiation of P-450 genes in *Cytochrome P-450* (Omura, T., Ishimura, Y., and Fujii-Kuriyama, Y., eds.) pp. 207-223, Kodansha, Tokyo and VCH, Weinheim
5. Kalb, V.F., Woods, C.W., Turi, T.G., Dey, C.R., Sutter, T.R., and Loper, J.C. (1987) Primary structure of the P450 lanosterol demethylase gene from *Saccharomyces cerevisiae*. *DNA* **6**, 529-537
6. Ishida, N., Aoyama, Y., Hatanaka, R., Oyama, Y., Imajo, S., Ishiguro, M., Oshima, T., Nakazato, H., Noguchi, T., Maitra, U.S., Mohan, V.P., Sprinson, D.B., and Yoshida, Y. (1988) A single amino acid substitution converts cytochrome P450(14DM) to an inactive form, cytochrome P450SG1: complete primary structures deduced from cloned DNAs. *Biochem. Biophys. Res. Commun.* **155**, 317-323
7. Turi, T.G. and Loper, J.C. (1992) Multiple regulatory elements control expression of the gene encoding the *Saccharomyces cerevisiae* cytochrome P450, lanosterol 14α-demethylase (*ERG11*). *J. Biol. Chem.* **267**, 2046-2056
8. van Nistelrooy, J.G., van den Brink, J.M., van Kan, J.A., van Gorcom, R.F., and de Waard, M.A. (1996) Isolation and molecular characterization of the gene encoding eburicol 14α-demethylase (CYP51) from *Penicillium italicum*. *Mol. Gen. Genet.* **250**, 725-733
9. Rozman, D., Stromstedt, M., and Waterman, M.R. (1996) The three human cytochrome P450 lanosterol 14α-demethylase (CYP51) genes reside on chromosomes 3, 7, and 13: structure of the two retrotransposed pseudogenes, association with a Line-1 element, and evolution of the human CYP51 family. *Arch. Biochem. Biophys.* **333**, 466-474
10. Stromstedt, M., Rozman, D., and Waterman, M.R. (1996) The ubiquitously expressed human CYP51 encodes lanosterol 14α-demethylase, a cytochrome P450 whose expression is regulated by oxysterols. *Arch. Biochem. Biophys.* **329**, 73-81
11. Cox, R.A. (1968) The use of guanidium chroride in the isolation of nucleic acids. *Methods Enzymol.* **12B**, 120-129
12. Aviv, H. and Leder, P. (1972) Purification of biologically active globin messenger RNA by chromatography on oligothymidylic acid-cellulose. *Proc. Natl. Acad. Sci. USA* **69**, 1408-1412
13. Wong, G., Kawajiri, K., and Negishi, M. (1987) Gene family of male-specific testosterone 16α-hydroxylase (C-P-450(16α)) in mouse liver: cDNA sequences, neonatal imprinting, and reversible regulation by androgen. *Biochemistry* **26**, 8683-8690
14. Frischauf, A.M., Lehrach, H., Poustka, A., and Murray, N. (1983) Lambda replacement vectors carrying polylinker sequences. *J. Mol. Biol.* **170**, 827-842
15. Sambrook, J., Fritsch, E.F., and Maniatis, T. (1989) *Molecular Cloning*, 2nd ed., Cold Spring Harbor Laboratory, New York
16. Sanger, F., Coulson, A.R., Barrell, B.G., Smith, A.J., and Roe, B.A. (1980) Cloning in single-stranded bacteriophage as an aid to rapid DNA sequencing. *J. Mol. Biol.* **143**, 161-178
17. Southern, E.M. (1975) Detection of specific sequences among DNA fragments separated by gel electrophoresis. *J. Mol. Biol.* **98**, 503-517
18. Thomas, P.S. (1980) Hybridization of denatured RNA and small DNA fragments transferred to nitrocellulose. *Proc. Natl. Acad. Sci. USA* **77**, 5201-5205
19. Triezenberg, S.J. (1987) Primer extension in *Current Protocols in Molecular Biology* (Ausubel, F.M., Brent, R., Kingdston, R.E., Moore, D.D., Seidmen, J.G., Smith, J.A., and Struhl, K., eds.) Vol. 1, pp. 4.8.1-4.8.5, John Wiley & Sons, New York
20. Kalb, V.F., Glasser, S., King, D., and Lingrel, J.B. (1983) A cluster of repetitive elements within a 700 base pair region in the mouse genome. *Nucleic Acids Res.* **11**, 2177-2184
21. Mount, S.M. (1982) A catalogue of splice junction sequences. *Nucleic Acids Res.* **10**, 459-472
22. Rozman, D., Stromstedt, M., Tsui, L.-C., Scherer, S.W., and Waterman, M.R. (1996) Structure and mapping of the human lanosterol 14α-demethylase gene (CYP51) encoding the cytochrome P450 involved in cholesterol biosynthesis; comparison of exon/intron organization with other mammalian and fungal CYP genes. *Genomics* **38**, 371-381
23. Yokoyama, C., Yabuki, T., Inoue, H., Tone, Y., Hara, S., Hatae, T., Nagata, M., Takahashi, E.-I., and Tanabe, T. (1996) Human gene encoding prostacyclin synthase (PTGIS): genomic organization, chromosomal localization, and promoter activity. *Genomics* **36**, 296-304
24. Harada, N., Utsumi, T., and Takagi, Y. (1993) Tissue-specific expression of the human aromatase cytochrome P-450 gene by alternative use of multiple exons 1 and promoters, and switching of tissue-specific exons 1 in carcinogenesis. *Proc. Natl. Acad. Sci. USA* **90**, 11312-11316
25. Kadonaga, J.T., Jones, K.A., and Tjian, R. (1986) Promoter-specific activation of RNA polymerase II transcription by SP1. *Trends Biochem. Sci.* **11**, 20-23
26. Yoshida, Y., Yamashita, C., Noshiro, M., Fukuda, M., and Aoyama, Y. (1996) Sterol 14-demethylase P450 activity expressed in rat gonads: contribution to the formation of mammalian meiosis-activating sterol. *Biochem. Biophys. Res. Commun.* **223**, 534-538