J. Biochem. 122, 1122-1128 (1997)

Structural and Evolutionary Studies on Sterol 14-Demethylase P450 (CYP51), the Most Conserved P450 Monooxygenase: II. Evolutionary Analysis of Protein and Gene Structures[1]

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

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

Received for publication, June 19, 1997

Phylogenetic analyses based on protein sequence data indicated that sterol 14-demethylase P450 (CYP51) and bacterial CYP51-like protein were joined into a distinctive evolutionary cluster, CYP51 cluster, within the CYP protein superfamily. The most probable branch topology of the CYP51 phylogenetic tree was (bacteria, (plants, (fungi, mammals))), which is comparable to the phylogeny of major kingdoms of living matter, suggesting that CYP51 has been conserved from the era of prokaryotic evolution. This may be strong evidence supporting the prokaryotic origin of P450. Structure of flanking regions and the number and insertion sites of introns are quite different between mammalian and fungal *CYP51*s. This fact indicates that different mechanisms are operative in evolution of protein sequences and gene structures. *CYP51* is the first example violating the well-documented rule that the basic structure of a gene, including intron insertion sites, is well conserved in each P450 family. One *CYP51* processed a pseudogene was found in rat genome. Nonsynonymous nucleotide divergence observed between the pseudogene and CYP51 cDNA was less than one-fifth of the synonymous divergence. This unusually low rate of nonsynonymous nucleotide changes in the pseudogene suggests that it may be derived from another *CYP51*, which might have been active for a significant duration in the past.

Key words: molecular evolution, P450, phylogenetic tree, processed pseudogene, sterol 14-demethylase.

P450 thiolate-heme proteins form a gene superfamily consisting of several hundred members distributed in most eukaryotes and many prokaryotes (*1*). Although the structure essential for activating an oxygen atom by a thiolate-ligated heme active center has been conserved, extensive alterations occurring in substrate recognition sites of P450 have produced various enzymes participating in the oxidative metabolism of a wide variety of small molecules (*2, 3*). Such potential flexibility or biodiversity is the most outstanding characteristics of the P450 superfamily, for which the novel term "diversozyme" has been proposed (*4*). Biodiversity of P450 must provide a great advantage to organisms for overcoming environmental stresses and perturbations through the formation of biosignal and biodefense substances and the degradation of adverse chemicals. Therefore, evolutionary consideration of the structural and functional diversification of P450 superfamily may provide information for understanding the strategy of life for adapting to fluid environments. Members of the P450 gene superfamily are classified into numbers of families, and P450s showing more than 40% amino acid sequence identity have been generally joined into one family (*1*). Although this definition was initially operational, an empirical rule that the basic structure of genes belonging to one family is conserved has been accepted (*5, 6*). Therefore, P450s belonging to the same family are considered to have a close evolutionary relationship to one another. However, the functional divergence of P450 is so extensive as to make it difficult to assume orthologous relationship even within the same family (*5*).

Sterol 14-demethylase P450 (CYP51) is an exceptional P450 that is widely distributed in mammals, higher plants and fungi and has the common metabolic role of 14-demethylation of sterol precursors, which is an essential step in sterol biosynthesis by eukaryotes (*3, 7*). As described in the previous papers (*8, 9*), mammalian and fungal CYP51s constitute a unique family (*CYP51* family) that was separated from the closest *CYP7* family long before the

[1] According to the recommendation by the P450 nomenclature committee [Ref. *1* and URL: http://drnelson.utmem.edu/nelsonhomepage.html], each member of P450 monooxygenase superfamily is defined by "CYP" code: Sterol 14-demethylase as CYP51, cholesterol 7α-hydroxylase as CYP7A1, dehydroepiandrosterone 7α-hydroxylase as CYP7B1, prostacyclin synthase as CYP8A1, sterol 12α-hydroxylase as CYP8B1, and benzoate 4-hydroxylase as CYP53A and B.

[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: NonD, nonsynonymous nucleotide divergence; SynD, synonymous nucleotide divergence; ML, maximum likelihood; NJ, neighbor-joining.

divergence of animal and fungi, and CYP51 has been inferred to be one of the most ancient and conserved eukaryotic P450s (*9*). Recently, cDNAs of obtusifoliol 14-demethylase of higher plants, *Sorghum bicolor* (L.) Moench (*10*) and wheat (*Triticum aestivum*, L. cv. Darius) (*11*), were cloned and their orthologous nature to mammalian and fungal CYP51s was suggested. Furthermore, search of nucleotide sequence databases revealed the existence of a gene encoding CYP51-like protein (Z80226) in a bacterium, *Mycobacterium tuberculosis*. The latter fact suggests that distribution of *CYP51* family may be extended to prokaryotes, although the function of the bacterial CYP51-like protein has not yet been studied.

Thus, CYP51 is considered to be the only known P450 widely distributed both in eukaryotes and prokaryotes. This unique characteristic of CYP51 prompted us to make comparative study of the gene structures of *CYP51*s occurring in distant kingdoms to accumulate information for evolutionary consideration of P450 monooxygenases.

This paper presents an evolutionary consideration of *CYP51* family based on the comparison of the protein and gene structures of eukaryotic *CYP51*s and the CYP51-like protein of *M. tuberculosis*. The results suggest that the *CYP51* family might have arisen in the era of prokaryotic evolution and been distributed in both prokaryotic and eukaryotic kingdoms throughout evolution. Comparison of the structures of mammalian and fungal *CYP51* genes revealed that different mechanisms are operative in evolution for conserving protein sequences and altering gene structures.

MATERIALS AND METHODS

Source of the Amino Acid and Nucleotide Sequence Data—The nucleotide sequence of rat *CYP51* gene and its processed pseudogene were determined as described in the preceding paper (*12*). The nucleotide sequence of human *CYP51* gene and its processed pseudogene (*13-15*) and the amino acid sequence of *Sorghum bicolor* CYP51 (U74319) (*10*) were kindly provided by Drs. D. Rozman and M.R. Waterman and by Dr. S. Bak, respectively, prior to their publication. The amino acid and nucleotide sequence data of other *CYP51*s were obtained from nucleotide and protein databases: rat cDNA and protein, D55681; human cDNA and protein, D55653; *Saccharomyces cerevisiae* genomic DNA and protein, M21483; *Candida albicans* genomic DNA and protein, X13296; *C. tropicalis* genomic DNA and protein, M23673; *Schizosaccharomyces pombe* genomic DNA and protein, Z54096; *Penicillium italicum* cDNA, genomic DNA and protein, Z49750; *Ustilago maydis* genomic DNA and protein, Z48164; *Triticum aestivum* cDNA, Y09291.

Methods of Sequence Data Analysis—CYP7, 8, and 51 protein sequences were multiply aligned by the doubly nested randomized iterative refinement (DNR) method (*16*). For the purpose of phylogenetic analysis, the columns containing 10% or more blank characters (deletions) were removed. Thus, the N-terminal *ca.* 60 sites containing the membrane anchor sequence specific to eukaryotic microsomal P450s were omitted from the present analysis. The resulting alignment consisting of 21 sequences and 429 columns was subjected to either neighbor-joining (NJ) (*17*) or maximum-likelihood (ML) (*18*) tree reconstruction analysis, using the programs NEIGHBOR and PROTML distributed in PHYLIP-3.5c (*19*) and MOLPHY 2.2 (*20*) packages, respectively. The bootstrap test with the neighbor-joining method was performed with 100 repetitions of resampling and otherwise the default settings. To limit the calculation time within a realistic range, some plausible constraints were preset to the topologies examined by PROTML; two mammalian (human and rat), three yeast (*S. cerevisiae, C. albicans*, and *C. tropicalis*), and two plant (*Sorghum bicolor* and wheat) CYP51s were presumed to make individual clusters. Similar assumptions were also made for CYP7 and 8 sequences. The bootstrap probabilities reported by PROTML were obtained by a resampling estimated log-likelihood (RELL) method (*18*).

The cDNA and pseudogene sequences were aligned by the DNR method (*16*), and then all insertions in a pseudogene relative to the cDNAs were removed. The numbers of synonymous and nonsynonymous sites and changes were calculated by the method of either Miyata and Yasunaga (*21*) or Nei and Gojobori (*22*). Any codons containing a deletion(s) were omitted from the calculations. The synonymous or nonsynonymous distances in Table I were then obtained using the Jukes-Cantor correction (*23*).

RESULTS AND DISCUSSION

Evolution of CYP51 Family—To search for sequences useful for the evolutionary analysis of *CYP51* family, we conducted a homology search with the TBLASTN program (*24*) against the non-redundant nucleotide sequence database (1,123,353 sequences) by using human CYP51 protein sequence (*9*) as the query. All but one of the top 43 sequences were genes, mRNAs, and ESTs of eukaryotic *CYP51*s, and a large number of sequences of other P450 families followed them. Interestingly, the exceptional one, ranked at 35th, was part of the *M. tuberculosis* genome (Z80226) having a long open reading frame. This finding strongly suggests that the bacterial sequence encodes a protein homologous to CYP51. Then, amino acid sequences of eukaryotic CYP51s, the bacterial protein and a few proteins of *CYP7/8* families, the closest neighbors to *CYP51* family (*9*), were aligned by the DNR method (*16*). The mean amino acid sequence identity calculated from the alignment for the bacterial protein and eukaryotic CYP51s was $31.5 \pm 2.6\%$, whereas that for the bacterial protein and CYP8/7 was $20.3 \pm 0.9\%$. The latter value is comparable to the mean amino acid sequence identity calculated for eukaryotic CYP51 and CYP7/8 ($20.1 \pm 1.0\%$), indicating that the bacterial CYP51-like protein is a member of *CYP51* family.

To examine the phylogenetic relationships among the members of *CYP51* family including the bacterial protein and *CYP7/8* family in more detail, we used two representative tree-reconstruction methods, NJ (*17*) and ML (*18*) methods. A bootstrap test with 100 repetitions of resampling applied to the NJ method, with which no topological constraints were presumed, showed 100% separation of the *CYP7/8* cluster and the *CYP51* cluster. This result indicates that the *CYP7/8* cluster may be used as the outgroup to examine the intra-family relationships of *CYP51* members. The branch topology within the *CYP51* cluster of the most probable tree was (bacterium, (plants, (fungi, mammals))), where the bootstrap probability calculated

for the separation of the (plants, (fungi, mammals)) branch from the bacterial one was 73% (Fig. 1). The same topology was also obtained by the ML method as the most probable one and the bootstrap probability was calculated to be 77.6% (tree, not shown), disregarding the fine topological variations within the fungal members. The branch topology

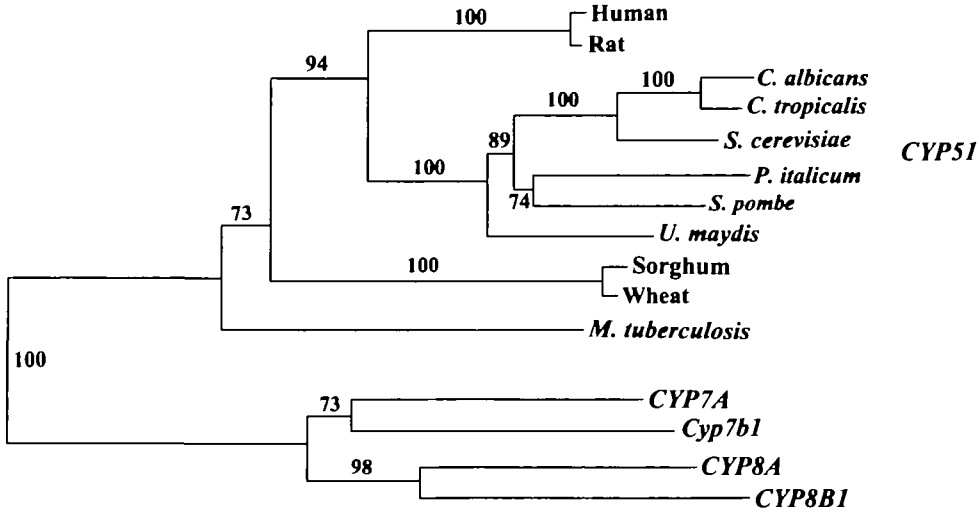

Fig. 1. **Neighbor-joining tree of *CYP51* and a few related *CYP* families.** The bootstrap probability (%) of observing the branch topology shown in the most probable tree is indicated by each branch. Human, rabbit, hamster, rat, and mouse CYP7A1 amino acid sequences are combined to form the single leaf (OTU) CYP7A, and likewise, human and bovine CYP8A1 and rat CYP8A sequences are combined to form the leaf CYP8A. The accession numbers of these sequences are M93133 for human CYP7A1, L10754 for rabbit CYP7A1, L04690 for hamster CYP7A1, J05430 for rat CYP7A1, L23754 for mouse Cyp7a1, D38145 for human CYP8A1, D30718 for bovine CYP8A1, and U53855 for rat CYP8A. Other leaves correspond to individual sequences. The accession numbers for CYP51 sequences are described in "MATERIALS AND METHODS" and that of mouse Cyp7b1 is U36993. The sequence of CYP8B1 has not yet appeared in any database and was cited from the literature (*38*). This species was called CYP12A in the literature because of its catalytic function, sterol 12α-hydroxylase (*38*), but P450 nomenclature committee suggested that this species should be designated as CYP8B1 (see URL: http://drnelson.utmem.edu/nelsonhomepage.html).

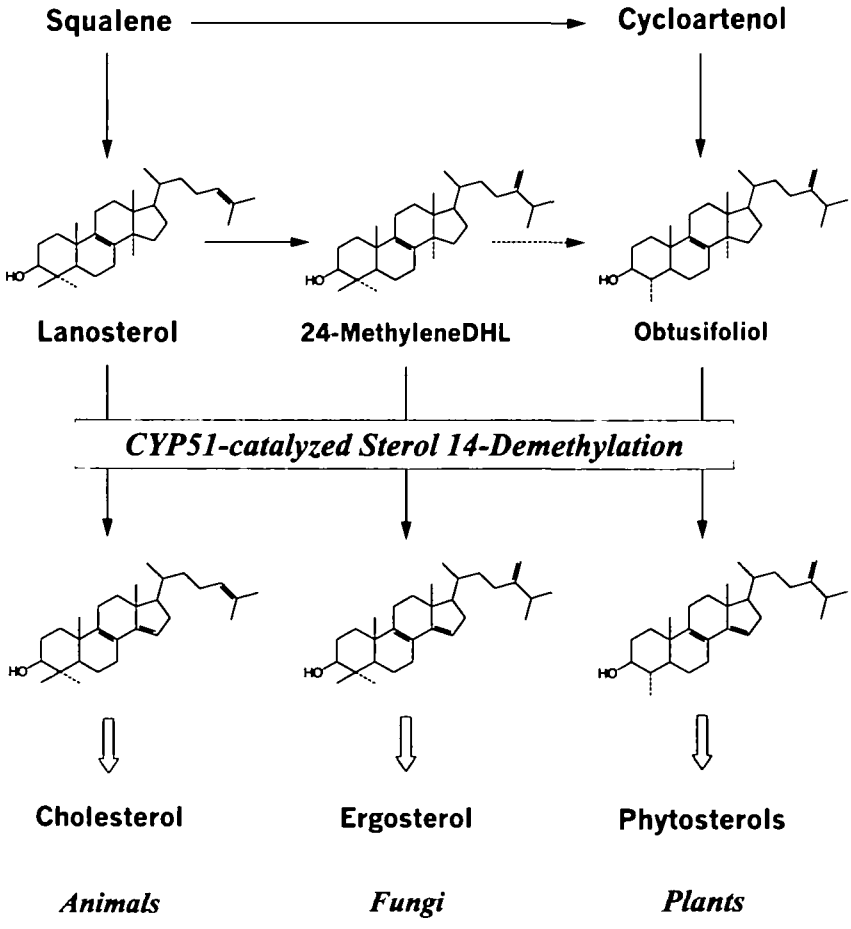

Fig. 2. **Difference in the substrates undergoing 14-demethylation observed among the sterol biosynthetic pathways of mammals, plants, and fungi.**

within the *CYP51* cluster of another NJ tree (not shown) was ((bacterium, plants), (fungi, mammals)), but the bootstrap probability calculated for the occurrence of the (bacterium, plants) branch was 27%, and the bootstrap probability value of the ML tree showing the same branch configuration was 10.5%. The ML method gave other minor tree topologies such as (((bacterium, plants), mammals), fungi) or (bacterium, (fungi, (plants, mammals))) which were not obtained by the NJ method, although the bootstrap probability of each of these minor trees was less than 6%. Consequently, CYP51 protein might have evolved as shown in Fig. 1. The branch topology of this tree is comparable to the most probable phylogeny of major kingdoms (*25, 26*), indicating that CYP51 has evolved in parallel with the biological system as a whole. This is the first evidence showing the occurrence of a distinct P450 species conserved throughout the phylogeny from the prokaryotic era to now, confirming the prokaryotic origin of P450.

Evolution of CYP51 and Alteration of Sterol Biosynthetic Pathway—Although no information is available on the function of *M. tuberculosis* CYP51, the metabolic role of eukaryotic CYP51, 14-demethylation of sterol precursors, has been conserved (*2, 7-9, 27*). The amino-acid sequence identity observed among eukaryotic CYP51s is high, and characteristically conserved regions overlapping the putative substrate recognition sites (SRS) were identified in all eukaryotic CYP51s (*8, 9*). This high conservation must be due to the indispensability of 14-demethylation for producing functional sterols necessary for most eukaryotes (*2, 7, 27*). However, the substrates undergoing 14-demethylation in sterol biosynthetic pathways are different among mammals, plants, and fungi as illustrated in Fig. 2 (*2, 7, 27*). This difference is dependent on different substrate specificities of CYP51 (*6, 28*). Since structural differences among these substrates are small and local (Fig. 2), these differences in substrate specificity can be caused by a few amino acid substitutions occurring in the substrate recognition sites. Actually, some amino acid substitutions exist between the putative substrate-recognition sites (SRS) of mammalian and fungal forms of the enzymes (Fig. 3). The intrinsic substrate of mammalian CYP51, lanosterol, is the most primitive sterol produced by the oxidative cyclization of squalene, whereas the substrate of the fungal enzyme, 24-methylene-24,25-dihydrolanosterol, is a metabolite of lanosterol (Fig. 2). In the phylogenetic tree shown in Fig. 1, the branch lengths of mammalian CYP51s from the mammalian/fungal branching point are shorter than the corresponding length of the fungal orthologues. This observation accords well with the general tendency that fungal proteins evolve more rapidly than their animal and plant orthologues (*26*). Consequently, mammalian CYP51s may retain more of the primordial structure of CYP51 protein than the fungal ones. It can be assumed that the existing fungal form of CYP51 had been formed by the modification of a primordial CYP51 metabolizing lanosterol as the natural substrate. Appearance of this form of CYP51 might contribute to altering the sterol biosynthetic pathway. This may be an example showing that evolutionary alteration of P450 has modified a metabolic pathway (*3*). To get further information, we are performing computer-aided molecular modeling and site-directed point mutations of mammalian and fungal CYP51s.

It is of interest that the products of the closest neighbors of *CYP51*, *CYP7/8*, contain sterol 7α- and 12α-hydoxylases (CYP7A1 and CYP8B1, respectively), because the carbons attacked by oxygen in these reactions are close to 14α-methyl carbon (C_{32}) attacked by oxygen in the CYP51-catalyzed reaction. This may suggest expansion of sterol metabolisms induced by the evolutionary alteration of P450 monooxygenases.

Variation of Exon/Intron Relationship during the Evolution of CYP51 Gene—As described in the preceding paper (*12*), rat *CYP51* gene is divided into ten exons by nine introns. Human *CYP51* gene also consists of the same number of exons (Rozman and Waterman, personal communication and Ref. *15*), and the intron insertion sites of these two mammalian *CYP51* genes were identical. However, no intron was found in the DNA fragment covering the entire coding region of *S. cerevisiae CYP51* gene cloned from a genomic DNA library of the yeast (*29*), and *P. italicum CYP51* consists of four exons interrupted by three short introns (60-72 bp) (*30*). Alignment analysis of amino acid sequences of rat and *P. italicum* CYP51 proteins indicated that the insertion site of the first intron of *P. italicum CYP51* gene corresponded to that of the second intron of the rat and human genes, but other intron insertion sites were different from those of mammalian genes (Fig. 4). These facts suggest that insertion/deletion of introns in the mammalian and fungal *CYP51* genes occurred for the most part independently and randomly after their separation. As discussed previously (*8, 9*), structures that may be necessary for maintaining the catalytic property, such as heme-binding domain and substrate recognition sites, have been conserved. However, the positions of these regions are not correlated with the exon-intron structure of mammalian *CYP51* genes.

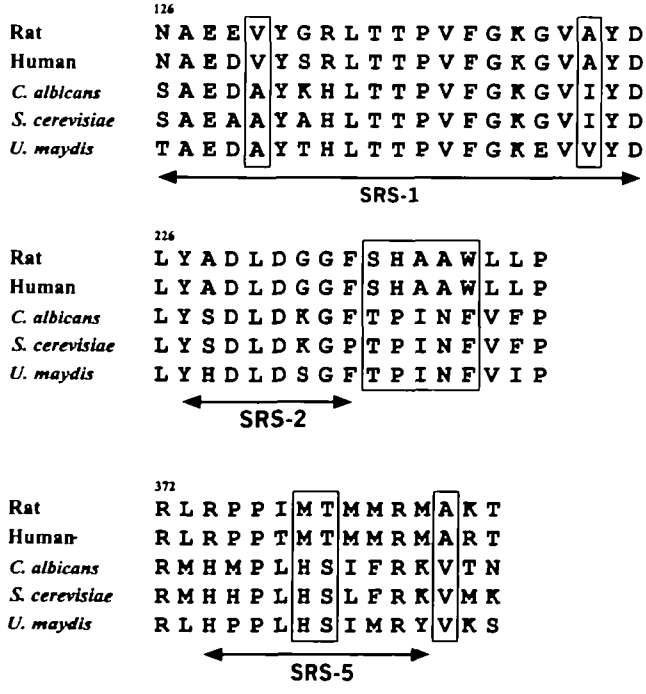

Fig. 3. **Amino acid substitutions observed between SRSs and vicinity of mammalian and fungal CYP51 proteins.** Characteristic residues conserved only within each kingdom are boxed.

Fig. 4. **Comparison of intron insertion sites in *CYP7A1*, *8A1*, and *51* genes.** The protein sequence alignment was obtained as described previously (*9, 16*). Introns inserted between adjacent codons, between the first and the second bases in the following codon, and between the second and the third bases in the previous codon are indicated by the symbols (▼), (▶), and (◀), respectively.

The divergence of the precursor *CYP51* gene into mammalian and fungal forms occurred long after the separation of *CYP51* from an ancestral gene of the *CYP7/8* families (Fig. 1). In accordance with this phylogenetic relationship, the human or rat *CYP51* gene has no intron equivalent to those of the rat *CYP7A* or *CYP8A* gene, whereas three of the five and nine intron insertion sites in *CYP7A* and *CYP8A* genes are equivalent to one another (Fig. 4). Consequently, it is most likely that the insertion of introns into P450 genes occurred after divergence of P450 families. This conclusion strongly supports the previous hypothesis (*6, 31*) that the insertion of introns after divergence of P450 families may be a mechanism for generating variations in gene structures of various P450. On the other hand, the well-documented empirical rule that the basic structure of genes including intron insertion sites in each P450 family is well conserved (*5, 6*) is not applicable to *CYP51* genes. *CYP51* has proven to be the first P450 family violating this rule. Recently, the exon/intron structures of two orthologous fungal P450 genes, *Aspergillus niger CYP53A* and *Rhodotorula minuta CYP53B*, have been found to be completely different (*32*). This is an additional example of a P450 family violating the rule. Furthermore, the human *CYP1B* gene has only one intron in its coding region, and its insertion site is completely different from those of *CYP1A* subfamily genes, although

the first intron in the 5′ untranslated region is commonly found in all the known *CYP1* genes (*33*). As discussed in this and the previous papers (*8, 9*), all members of the *CYP51* family are orthologous, and *CYP53A* and *CYP53B* are also regarded as orthologous to each other (*32*), while *CYP1A* and *1B* are paralogous genes. These facts suggest the need for discussion regarding the definition and meaning of "family" in the case of P450.

Variation Observed in Structures of the 5′-Flanking Region Relating to Transcriptional Initiation—As discussed in the preceding paper (*12*), mammalian *CYP51* is a TATA-box less and GC-element rich gene in accordance with its predicted housekeeping nature and ubiquitous expression in human organs. However, the *CYP51* gene of *S. cerevisiae* was reported to have a TATA box in its transcriptional promoter region (*34*), and little identity was found in the 5′-flanking regions of rat and yeast *CYP51* genes (data not shown). Expression of yeast *CYP51* gene is known to be altered by oxygen concentration of growth medium, and regulatory sites responsible for the oxygen-dependent regulation of yeast *CYP51* gene were suggested (*34*). These facts indicate that the structure of 5′-flanking regions and the regulatory mechanism of gene expression are quite different between mammalian and fungal *CYP51* genes, although their coding regions retain the core structure that might be derived from the common precursor gene.

Evolutionary Consideration on CYP51 Processed Pseudogene—One processed pseudogene, which might be formed by the reverse transcription of CYP51 mRNA in germ line cells, was identified in the rat genome as described in the previous papers (*9, 12*). Similar *CYP51* processed pseudogenes were also reported in the human genome (*13*). Table I shows the synonymous (causing no amino acid change) and nonsynonymous (causing amino acid change) nucleotide divergence observed between given aligned pairs of rat and human CYP51 cDNAs and their processed pseudogenes. These values were obtained by the method of Nei and Gojobori (*22*), and essentially the same results were obtained by the method of Miyata and Yasunaga (*21*). The synonymous divergence (*SynD*) of rat/human cDNA was 0.67 (Table I). The phylogenetic diverging point of rat and human is believed to be some 105 million years ago (*35*). Assuming that *SynD* has increased simply in proportion to time, the *SynD* of rat cDNA/pseudogene (0.098) formally predicts that the pseudogene was formed 15 million years ago. However, if the pseudogene was formed from the second active *CYP51* gene, as discussed below, it might have been formed more recently, as in the cases of human *CYP51* pseudogenes (*13*).

It is generally accepted that the frequency of nucleotide substitution occurring at nonsynonymous sites of a functional gene is suppressed by the selective forces for conserving the structure and function of its coding protein, and nonsynonymous nucleotide divergence (*NonD*) is usually far smaller than *SynD*. Actually, the *NonD/SynD* value calculated for rat/human cDNA pair was 0.051. On the other hand, *NonD/SynD* value for a pseudogene/parent pair may be greater than 0.5, because such selective forces would not affect nucleotide substitutions at nonsynonymous sites along the pseudogene lineage. The *NonD/SynD* values calculated for two human cDNA/pseudogene pairs were 0.75 and 0.91, supporting this expectation. However, the *NonD/SynD* value calculated for the rat cDNA/pseudogene pair was only 0.16, which is unexpectedly small and rather comparable to those obtained for orthologous functional genes. The occurrence of active processed (intronless) genes in mammals is also known (*36, 37*). The pseudogene might be derived from such an active processed gene or another active gene duplicated in the genome. Alternatively, the pseudogene might be derived from an allelic variant, since Asai *et al.* (Asai, K. Tsuchimori, N., Okonogi, K., Gotoh, O., and Yoshida, Y., unpublished) found that an allelic variation of yeast *CYP51* promoted the formation of an altered CYP51 resistant to azole antifungal agents. No evidence for the existence of the second *CYP51* gene was obtained by Southern blot analysis of the present-day rat genomic DNA, as described in the preceding paper (*12*). Therefore, the exact reason for the unexpectedly small NonD/SynD value of the rat CYP51 cDNA/pseudogene pair remains an enigma. However, this low NonD/SynD value strongly suggests that at least two active *CYP51* genes might have been present in the rat genome in the past.

TABLE I. **Synonymous and nonsynonymous nucleotide divergence calculated for given pairs of the coding region sequences of rat and human CYP51 cDNAs and *CYP51* pseudogenes.** The nucleotide sequences corresponding to the coding region of *CYP51* were picked up from the indicated sequence pairs. The nucleotide divergence causing amino acid change (nonsynonymous divergence, NonD) and that causing no amino acid change (synonymous divergence, SynD) observed between each pair of sequences were calculated by the method of Nei and Gojobori (*22*).

Sequence pair	NonD	SynD	NonD/SynD
Human cDNA/Rat cDNA	0.034	0.67	0.051
Rat cDNA/Rat pseudogene	0.016	0.098	0.16
Human cDNA/Human P1	0.025	0.027	0.91
Human cDNA/Human P2	0.030	0.039	0.75

Human P1: human processed pseudogene, *CYP51*P1, Human P2: human processed pseudogene, *CYP51*P2.

REFERENCES

1. Nelson, D.R., Koymans, L., Kamataki, T., Stegeman, J.J., Feyereisen, R., Waxman, D.J., Waterman, M.R., Gotoh, O., Coon, M.J., Estabrook, R.W., Gunsalus, I.C., and Nebert, D.W. (1996) P450 Superfamily: Update on new sequences, gene mapping, accession numbers, and nomenclature. *Pharmacogenetics* **6**, 1-42
2. Coon, M.J., Ding, X.X., Pernecky, S.J., and Vaz, A.D.N. (1992) Cytochrome P450: Progress and predictions. *FASEB J.* **6**, 669-673
3. 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
4. Coon, M.J., Vaz, A.D.N., and Bestervelt, L.L. (1996) Peroxidative reactions of diversozyme. *FASEB J.* **10**, 428-434
5. Nelson, D.R., Kamataki, T., Waxman, D.J., Guengerich, P., Estabrook, R.W., Feyereisen, R., Gonzalez, F.J., Coon, M.J., Gunsalus, I.C., Gotoh, O., Okuda, K., and Nebert, D.W. (1993) The P450 superfamily: Update on new sequences, gene mapping, accession numbers, early trivial names of enzymes, and nomenclature. *DNA Cell Biol.* **12**, 1-51
6. Gotoh, O. (1993) Evolution and differentiation of P-450 genes in *Cytochrome P-450*, 2nd ed. (Omura, T., Ishimura, Y., and Fujii-Kuriyama, Y., eds.) pp. 207-223, Kodansha, Tokyo and VCH, Weinheim
7. Yoshida, Y. (1993) Sterol biosynthesis in *Cytochrome P-450*, 2nd

ed. (Omura, T., Ishimura, Y., and Fujii-Kuriyama, Y., eds.) pp. 93-101, Kodansha, Tokyo and VCH, Weinheim
8. Aoyama, Y., Funae, Y., Noshiro, M., Horiuchi, T., and Yoshida, Y. (1994) Occurrence of a P450 showing high homology to yeast lanosterol 14-demethylase ($P450_{14DM}$) in the rat liver. *Biochem. Biophys. Res. Commun.* **201**, 1320-1326
9. 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
10. Bak, S., Kahn, R.A., Olsen, C.-E., and Halkier, B.A. (1997) Cloning and expression in *Escherichia coli* of the obtusifoliol 14α-demethylase of *Sorghum bicolor* (L.) Monech, a cytochrome P450 orthologous to the sterol 14α-demethylase (CYP51) from fungi and mammals. *Plant J.* **11**, 191-201
11. Cabello-Hurtado, F., Zimmerlin, A., Rahier, A., Taton, M., DeRose, R., Nedelkina, S., Batard, Y., Durst, F., Pallet, K.E., and Werck-Reichhart, D. (1997) Cloning and functional expression in yeast of a cDNA coding for an obtusifoliol 14α-demethylase (CYP51) in wheat. *Biochem. Biophys. Res. Commun.* **230**, 381-385
12. Noshiro, M., Aoyama, Y., Kawamoto, T., Gotoh, O., Horiuchi, T., and Yoshida, Y. (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. *J. Biochem.* **122**, 1114-1121
13. 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
14. 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
15. 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
16. Gotoh, O. (1996) Significant improvement in accuracy of multiple protein sequence alignments by iterative refinement as assessed by reference to structural alignments. *J. Mol. Biol.* **264**, 823-838
17. Saitou, N. and Nei, M. (1987) The neighbor-joining method: A new method for reconstructing phylogenetic trees. *Mol. Biol. Evol.* **4**, 406-425
18. Kishino, H., Miyata, T., and Hasegawa, M. (1990) Maximum likelihood inference of protein phylogeny and the origin of chloroplasts. *J. Mol. Evol.* **31**, 151-160
19. Felsenstein, J. (1992) PHYLIP (Phylogeny Inference Package) Version 3.5c, University of Washington, Seattle
20. Adachi, J. and Hasegawa, M. (1996) MOLPHY Version 2.3. Programs for molecular phylogenetics based on maximum likelihood in *Computer Science Monographs* Vol. 28, The Institute of Statistical Mathematics, Tokyo
21. Miyata, T. and Yasunaga, T. (1980) Molecular evolution of mRNA: A method for estimating evolutionary rates of synonymous and amino acid substitutions from homologous nucleotide sequences and its application. *J. Mol. Evol.* **16**, 23-36
22. Nei, M. and Gojobori, T. (1986) Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. *Mol. Biol. Evol.* **3**, 418-426
23. Jukes, T.H. and Cantor, C.R. (1969) Evolution of protein molecules in *Mammalian Protein Metabolism* (Munro, H.N., ed.) Vol. 3, pp. 21-132, Academic Press, New York
24. Altschul, S.F., Gish, W., Miller, W., Myers, E.W., and Lipman, D.J. (1990) Basic local alignment search tool. *J. Mol. Biol.* **215**, 403-410
25. Baldauf, S.L. and Palmer, J.D. (1993) Animals and fungi are each other's closest relatives: congruent evidence from multiple proteins. *Proc. Natl. Acad. Sci. USA* **90**, 11558-11562
26. Doolittle, R.F., Feng, D.-F., Tsang, S., Cho, G., and Little, E. (1996) Determining divergence times of the major kingdoms of living organisms with a protein clock. *Science* **271**, 470-477
27. Aoyama, Y. and Yoshida, Y. (1991) Different substrate specificities of lanosterol 14α-demethylase ($P\text{-}450_{14DM}$) of *Saccharomyces cerevisiae* and rat liver for 24-methylene-24,25-dihydrolanosterol and 24,25-dihydrolanosterol. *Biochem. Biophys. Res. Commun.* **178**, 1064-1071
28. Aoyama, Y. and Yoshida, Y. (1992) The 4β-methyl group of substrate does not affect the activity of lanosterol 14α-demethylase ($P\text{-}450_{14DM}$) of yeast: Difference between the substrate recognition by yeast and plant sterol 14α-demethylase. *Biochem. Biophys. Res. Commun.* **183**, 1266-1272
29. 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 cDNAs. *Biochem. Biophys. Res. Commun.* **155**, 317-323
30. 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
31. Nishimoto, M., Gotoh, O., Okuda, K., and Noshiro, M. (1991) Structural analysis of the gene encoding rat cholesterol 7α-hydroxylase, the key enzyme for bile acid biosynthesis. *J. Biol. Chem.* **266**, 6467-6471
32. Fujii, T., Nakamura, K., Shibuya, K., Tanae, S., Gotoh, O., Ogawa, T., and Fukuda, H. (1997) Characterization of the structures of the gene and cDNA of the cytochrome P450rm from *Rhodotorula minuta* which catalyzes formation of isobutene and 4-hydroxylation of benzoate. *Mol. Gen. Genet.* **256**, 115-120
33. Tang, Y.M., Wo, Y.-Y.P., Stewart, J., Hawkins, A.L., Griffin, C.A., Sutter, T.R., and Greenlee, W.F. (1996) Isolation and characterization of the human cytochrome P450 *CYP1B1* gene. *J. Biol. Chem.* **271**, 28324-28330
34. 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 (*ERG-11*). *J. Biol. Chem.* **267**, 2046-2056
35. Hedges, S.B., Parker, P.H., Sibley, C.G., and Kumar, S. (1996) Continental breakup and the ordinal diversification of birds and mammals. *Nature* **381**, 226-229
36. Hsu, L.C. and Chang, W.C. (1991) Cloning and characterization of a new functional human aldehyde dehydrogenase gene. *J. Biol. Chem.* **266**, 12257-12265
37. Persson, K., Holm, I., and Heby, O. (1995) Cloning and sequencing of an intronless S-adenosylmethionine decarboxylase gene coding for a functional enzyme strongly expressed in the liver. *J. Biol. Chem.* **270**, 5642-5648
38. Eggertsen, G., Olin, M., Andersson, U., Ishida, H., Kubota, S., Hellman, U., Okuda, K.-I., and Bjorkhem, I. (1996) Molecular cloning and expression of rabbit sterol 12α-hydroxylase. *J. Biol. Chem.* **271**, 32269-32275