The porcine taurochenodeoxycholic acid 6 α -hydroxylase (*CYP4A21*) gene: evolution by gene duplication and gene conversion

Kerstin LUNDELL¹

Division of Biochemistry, Department of Pharmaceutical Biosciences, University of Uppsala, Box 578, S-751 23 Uppsala, Sweden

Porcine taurochenodeoxycholic acid 6 α -hydroxylase, cytochrome P450 4A21 (*CYP4A21*), differs from other members of the *CYP4A* subfamily in terms of structural features and catalytic activity. *CYP4A21* participates in the formation of hyocholic acid, a species-specific primary bile acid in the pig. The *CYP4A21* gene was investigated and found to be approx. 13 kb in size and split into 12 exons. The intron–exon organization of the *CYP4A21* gene corresponds to that of *CYP4A* fatty acid hydroxylase genes in other species. Comparison with a genomic segment of a pig *CYP4A* fatty acid hydroxylase gene (*CYP4A24*) revealed a sequence identity with *CYP4A21* that extends beyond the exons, indicating a

common origin by gene duplication. A pronounced sequence identity was found also within the proximal 5'-flanking regions, whereas the patterns of mRNA expression of *CYP4A21* and *CYP4A* fatty acid hydroxylases in pig liver differ. Sequence comparison aiming to elucidate the origin of the unique features of *CYP4A21* revealed a region of decreased sequence identity from exon 6 to exon 8, strongly suggesting that gene conversion could have contributed to the evolution of *CYP4A21*.

Key words: *CYP4A* subfamily, *CYP4A21*, *CYP4A24*, gene conversion, gene duplication, hyocholic acid.

INTRODUCTION

The cytochrome P450 4A gene subfamily (*CYP4A*) encodes a number of enzymes, which are characterized by ω - and (ω -1)-hydroxylase activities towards various fatty acids, including arachidonic acid, and prostaglandins [1,2]. *CYP4A21* in the domestic pig (*Sus scrofa*) is distinct from other *CYP4A* enzymes both in its substrate specificity and in structural features [3]. *CYP4A21* is unique among *CYP4A* enzymes in that it participates in bile acid biosynthesis by performing a 6 α -hydroxylation of chenodeoxycholic acid (or its conjugate). The hyocholic acid (3 α ,6 α ,7 α -trihydroxy-5 β -cholanoic acid) thus formed is a major primary bile acid in the pig. Whereas the bile of adult pig is devoid of cholic acid, this species has a biosynthetic balance between hyocholic acid and chenodeoxycholic acid which is similar to that between cholic acid and chenodeoxycholic acid in most mammals [4].

The *CYP4A21* gene shares high sequence identity with the pig *CYP4A* fatty acid hydroxylases, *CYP4A24* and *CYP4A25* (94 % overall nucleotide identity) [5]. To elucidate the relationship of *CYP4A21* to *CYP4A* fatty acid hydroxylases, the gene encoding *CYP4A21* was cloned. The intron–exon organization of the gene was compared with that of *CYP4A* genes in rat [6], rabbit [7,8] and human [9,10]. Also, a segment of a pig *CYP4A* fatty acid hydroxylase gene (*CYP4A24*) was obtained and used for comparison. The high similarity previously observed between the porcine *CYP4A* cDNAs was found also between introns as well as between the proximal 5'-flanking regions. A distinct regulation of these genes in pig liver was, however, indicated by differences in the mRNA expression levels. Sequence comparison aiming to elucidate the origin of the unique features of *CYP4A21* indicates that gene conversion has occurred during evolution of *CYP4A21*.

EXPERIMENTAL

Materials

Livers from 2-week-old, 6-week-old and 3-month-old pigs were obtained from the Funbo-Lövsta Research Centre, Department of Animal Breeding and Genetics, Swedish University of Animal Science, Uppsala, Sweden. Livers and kidney from 6-month-old male pigs were obtained from the local slaughterhouse. RNeasy Midi Kit and Qiagen plasmid kit were from Qiagen. Porcine genomic library, NucleoSpin Tissue kit, Universal Genome Walker kit and Advantage genomic PCR kit were obtained from Clontech Laboratories. Oligonucleotides were from Thermo Hybaid (Interactiva Division, Ulm, Germany). *Taq*Gold polymerase and oligo(dNTP) mix were from PE Applied Biosystems. A 1 kb DNA ladder was from Life Technologies. Megaprime DNA Labelling System, [γ -³²P]dATP and [α -³²P]dCTP were from Amersham Biosciences. The pGEM-T Vector was from Promega and the pBluescript KS II vector was purchased from Stratagene. All chemicals used were of the highest grade.

Isolation and analysis of *CYP4A21* genomic clones

A porcine genomic library constructed in the EMBL3 SP6/T7 vector was screened according to the manufacturer's instructions (Lambda Library User Manual, Clontech Laboratories). About 5×10^5 clones were screened using a full-length ³²P-labelled *CYP4A21* cDNA (Megaprime DNA Labelling System). A ³²P-labelled probe containing intronic sequences, spanning the *CYP4A21* gene from exon 6 to 9, was subsequently used for screening of 4×10^5 additional clones. Selected hybridization-positive plaques were subjected to successive rounds of plating

Abbreviations used: Sp1, specificity protein 1; YY1, Yin-Yang-1.

¹ E-mail Kerstin.Lundell@farmbio.uu.se

The compiled genomic sequences of *CYP4A21* and *CYP4A24*, from exon 2 to 12 and from exon 2 to intron 11, have been submitted to the EMBL database under accession numbers AJ586618 and AJ586619 respectively. The sequences of the 5'-flanking regions and adjacent nucleotides in exon 1 of the *CYP4A21* and *CYP4A24* genes have been deposited in the EMBL database under accession numbers AJ586859 and AJ586860, respectively.

Table 1 Oligonucleotides used in PCR amplification of genomic sequences

The primers are denoted according to the following principles: the first letter refers to a forward (F) or reverse (R) primer. EX is an abbreviation for exon-derived primers, IN for intron-derived primers and P for a primer from the 5'-flanking region. The number following EX and IN denotes the exon or intron number, respectively. In cases where more than one primer is assigned the same name, based on the principles above, a final digit (e.g. 1 or 2) distinguishes them. Position refers to the cDNA sequence of CYP4A21 (accession number AJ278474) numbered relative to the start-code adenine (+1). The position of a primer in the 5'-flanking region is numbered upstream from the start-code adenine. The positions of primers within introns are written in italics.

Name	Position	Sequence
FP	-184	5'-CTTTGTCCCTCTCGCGGACAAAGCTT-3'
REX1-1	154	5'-AGTGGGAAGGAGGGGACGGGAAGTCT-3'
REX1-2	129	5'-CTGGAGGGCTTTGAGCAGCCACTGTCT-3'
FEX2	247	5'-TGGCTGTGGGGAACAAGAGC-3'
FEX4	415	5'-ATGCTACCCAGCCTTCCACT-3'
REX4	484	5'-CTTGGACAGACTTGGCCATG-3'
FEX5	512	5'-TCGCCAGGACCCAGCTCT-3'
RIN5	52	5'-TCCTTCCCAGCTGTGTGAGAACAAGCAG-3'
FEX6	689	5'-TTCTTCAGAAATGACATCATCTA-3'
FEX8	1006	5'-GAGCATCAGCAGAGATGTCG-3'
FEX9	1095	5'-CACCACCATGTGCATCAAGGAG-3'
REX9	1206	5'-ACCTGCAGGTAAAGAGCGCTCCA-3'
FIN9	806	5'-CACGGGAACCATCAGTCTTC-3'
REX10	1238	5'-TGAAGGCCGTAAATGGAGAGCG-3'
FEX11	1300	5'-TCTGCTCGCCACAGCCATGCTTT-3'
REX11	1302	5'-AGAACCCTGGTGCAAAACCGG-3'
RIN11	122	5'-CCTCCTTCTCAACACACAA-3'
FEX12	1353	5'-CATTGGGAAGCAGTTTGCCA-3'
REX12-1	1538	5'-GCTTGCTTGTGCCCCACA-3'
REX12-2	2362	5'-ACACAGCGAATTTATCAGGG-3'

to ensure single plaque purity. Plaque extracts were analysed by PCR using a *CYP4A21*-specific reverse primer RNON-CONS [3] combined with an upstream forward primer FEX6 (Table 1). PCR cycles were as follows: 94 °C for 30 s, 50 °C for 30 s and 72 °C for 2 min for 30 cycles, followed by 72 °C for an additional 10 min. Two *CYP4A21*-positive recombinant phages, denoted 4111 and 1732, were identified and isolated from lambda lysate by polyethylene glycol and CsCl precipitation, according to the manual (Lambda Library User Manual, Clontech Laboratories). Insert DNAs were excised using *Xho*I, and the nucleotide sequence of the gene was obtained from a restriction fragment (4111-*Sac*I) subcloned into the pBluescript vector and from PCR-amplified fragments (4111-A to 4111-C and 1732-A to 1732-D) as depicted in Figure 1(A). The primers used for PCR (Table 1) were designed from putative exon sequences, based on the intron-exon structure of the human *CYP4A* genes, as well as from subsequently obtained intron sequences and from a sequence in the 5'-flanking region. PCR cycles were as described previously with annealing temperatures of 55 or 60 °C for the different primer pairs. Sequencing was performed using Dye Terminator cycle sequence kit for the sequence reaction and a Megabase 1000 sequenator (Amersham Biosciences) for analysis.

Isolation and analysis of a *CYP4A24* genomic clone

Hybridization-positive phages from the genomic library screening described above were also analysed by PCR using a reverse primer directed against the consensus sequence found in all *CYP4As* except for *CYP4A21* [3], RCONS-1 (5'-CTGTGGTGTTCATGACCCTCGA-3'), combined with the forward primer FEX6 (Table 1). The PCR cycles were as described above with an annealing temperature of 50 °C. Three RCONS-1-positive

A. *CYP4A21* gene

B. *CYP4A24* gene

Figure 1 Organization of the cloned porcine *CYP4A* genes

The *CYP4A21* gene was compiled from the two recombinant phages 4111 and 1732 (A), and the segment of the *CYP4A24* gene was obtained from the recombinant phage 124 (B) as described in the Experimental section. Exons are indicated with solid boxes and identified by numbers. (A) *CYP4A21* gene: the locations of fragments used for sequencing of the recombinant phages 4111 and 1732 are indicated. A restriction fragment generated by *Sac*I cleavage of the purified phage 4111 insert is denoted 4111-*Sac*I. PCR amplification of the purified inserts of phages 4111 and 1732 was performed using the following primer pairs (see Table 1 for sequences): 4111-A, FEX12 and REX12-2; 4111-B, FEX11 and REX12-1; 4111-C, FEX8 and REX11; 1732-A, FEX4 and RIN5; 1732-B, FEX2 and REX4; 1732-C, FP and REX1-2. (B) *CYP4A24* gene: the following primers (see Table 1) were used for PCR amplification to obtain fragments for sequencing of the recombinant clone 124: 124-A, FIN9 and RIN11; 124-B, FEX9 and REX10; 124-C, FEX5 and REX9; 124-D, FEX4 and RIN5; 124-E, FEX2 and REX4.

recombinant phages were selected for further PCR amplification, generating a genomic sequence spanning exons 2–4, using the forward primer FEX2 and the reverse primer REX4 (Table 1). One recombinant phage was assigned to *CYP4A24* by *Eco*RI digestion of the gel-purified sequence. A genomic sequence of this clone, denoted 124, spanning exons 2–11, was obtained from overlapping PCR fragments (124-A to 124-E), amplified using exon- and intron-derived primers (Table 1), as illustrated in Figure 1(B). PCR cycles and sequencing of the purified PCR fragments were as described previously.

Amplification of pig *CYP4A* promoter sequences by genome walking

Genomic DNA from pig kidney was purified using NucleoSpin kit (Clontech Laboratories). GenomeWalker DNA libraries were constructed and screened, using reverse nested primers derived from the first exon of pig *CYP4As*, REX1-1 and REX1-2 (Table 1), according to the manual (Universal Genome Walker Kit, Clontech Laboratories). A PCR product of about 500 bp was amplified, gel-purified and subsequently cloned into the pGEM-T vector and propagated in *Escherichia coli* DH5α cells. Multiple plasmid DNAs were purified using Qiagen plasmid kit and subjected to restriction-enzyme analysis using *Bsg*I. One *Bsg*I-digestion-positive and one *Bsg*I-digestion-negative plasmid were selected for sequencing as described above.

Northern blot analysis using oligo-probes

Total RNA was isolated from pig livers using RNeasy Midi Kit (Qiagen). Two matching Northern blots were prepared as described previously [3] using total RNA (30 µg) from livers of pigs of different ages. The *CYP4A21*-specific oligonucleotide (RNON-CONS) [3] and an oligonucleotide specific for the conserved sequence common to *CYP4A* fatty acid hydroxylases, RCONS (5'-TGGTGTTCATGACCCTCGAA-3'), were ³²P-labelled using

Figure 2 Alignment of the 5'-flanking regions and adjacent nucleotides in exon 1 of the *CYP4A21* and *CYP4A24* genes

The two genomic sequences were assigned to *CYP4A21* and *CYP4A24* based on the two differing nucleotides within exon 1 that are indicated by bold arrows. A silent nucleotide substitution within the *CYP4A21* sequence is marked *. Nucleotides common to both sequences are shaded. The start codons are boxed and the sequences are numbered relative to the start-codon adenine (+1). The location of the transcription start (TS) site in *CYP4A21* [3] is indicated, as well as the binding sites for Sp1 and YY1 that were predicted by the Transcription Element Search Software (TESS). The positions of the forward primer FP and the reverse primer REX1-2 used in PCR amplification of fragment 1732-C (Figure 1) are shown by arrows above the aligned sequences.

the DNA 5'-End Labelling System (Promega) and used as probes. The membranes were hybridized with the 5'-labelled oligo-probes overnight at 47 °C as described in [11]. Samples of reversed-transcribed cDNAs of *CYP4A21* and *CYP4A24* were used to assess the ability of the oligo-probes to recognize each distinct RNA, as described previously [11].

RESULTS

Cloning and characterization of the *CYP4A21* gene

One recombinant phage that contained the *CYP4A21*-specific sequence was obtained by screening the porcine genomic library using a full-length *CYP4A21* cDNA probe. This clone, designated 4111, was found to span the *CYP4A21* gene from the fourth intron to the 3'-untranslated region. A second clone, designated 1732, containing the upstream sequence of the *CYP4A21* gene, was subsequently obtained by screening the genomic library with a genomic *CYP4A21* probe. A genomic sequence of *CYP4A21*, approx. 13 kb, with the exception of nucleotides within the first intron and adjacent exon sequences, was obtained. Attempts to amplify over the first intron using primers derived from exons 1 and 2 were unsuccessful, indicating that this intron resists PCR amplification, due to size or compression. The presence of a first intron in pig *CYP4A* genes, with a conserved position in the coding sequence, was, however, confirmed using GenomeWalker DNA libraries (result not shown).

A physical map of the gene encoding *CYP4A21* and the locations of fragments acquired for sequencing is shown in Figure 1(A). The intron-exon boundaries of the genomic sequence were determined by alignment of the compiled sequence with that of *CYP4A21* cDNA (accession number AJ278474). The genomic exon sequences all agreed with that of the *CYP4A21* cDNA with two exceptions: a silent nucleotide substitution within exon 1, indicated in Figure 2, and one divergent nucleotide in the fourth exon. The latter caused a replacement of Gln by Pro at position 135 in the deduced amino acid sequence of *CYP4A21*.

The *CYP4A21* gene is split into 12 exons and has consensus GT and AG [12] at all the 5' and 3' splice sites, respectively. The position of each intron within the coding sequence was highly conserved compared with those of the human, rabbit and rat *CYP4A* genes. The full-length gene sequences of rat *CYP4A1* (accession numbers M57718 and M33937), rat *CYP4A2* (accession numbers M57719 and M33938), human *CYP4A11* (accession number AF525488) and human *CYP4A22* (accession number AF208532), currently annotated as *CYP4A11* [9,13], are available, whereas only parts of the intronic sequences of rabbit *CYP4A6* and *CYP4A4* are published [7,8]. Similar to *CYP4A21*, these genes possess 12 exons, with the exception of rat *CYP4A1* and rabbit *CYP4A6*, where the 3'-untranslated region is split by a twelfth intron, yielding a total of 13 exons [6,7]. Among the full-length *CYP4A* genomic sequences, the human genes are most similar to *CYP4A21* in terms of intron sizes and sequence similarity.

Cloning and characterization of a segment of the *CYP4A24* gene and comparison with the *CYP4A21* gene

Among the hybridization-positive phages obtained from the genomic library screening several clones with sequences related to the *CYP4A* subfamily were identified. The pig fatty acid hydroxylase *CYP4A24* sequence is characterized by an *EcoRI* restriction site [5] that is located in the third exon. One recombinant phage, denoted 124, was selected based on *EcoRI* digestion of a PCR fragment spanning exons 2-4 (Fragment 124-E, Figure 1). The genomic sequence in clone 124, from exons 2 to 11, was subsequently obtained from overlapping PCR fragments as depicted in Figure 1(B). Attempts to amplify a sequence within exon 1 failed, indicating that the genomic sequence in clone 124 ends within the first intron.

The exon sequences identified within this segment conformed to the cDNA of *CYP4A24* (accession number AJ318096) with some differences. Seven amino acid replacements were encoded by the

Table 2 Genomic sequence identity

Sequence identity between the genomic sequences assigned to *CYP4A21* and *CYP4A24*, from intron 2 to exon 11. The second and fifth columns show the number of bp in each exon or intron of *CYP4A21* and *CYP4A24*; where the two genes differ, two numbers show the bp in *CYP4A21* and *CYP4A24* respectively. Under Nucleotide identity, values in parentheses indicate the percentage identity.

Exon no.	Length of <i>CYP4A21</i> / <i>CYP4A24</i> (bp)	Nucleotide identity	Intron no.	Length of <i>CYP4A21</i> / <i>CYP4A24</i> (bp)	Nucleotide identity
3	45	40 (89 %)	2	669	657 (98 %)
4	128	125 (98 %)	3	477	475 (100 %)
5	125	124 (99 %)	4	1000	992 (99 %)
6	155	137 (83 %)	5	370	366 (99 %)
7	107	93 (87 %)	6	431/432	337 (78 %)
8	191	179 (94 %)	7	83/84	77 (92 %)
9	134	134 (100 %)	8	113	110 (97 %)
10	65	65 (100 %)	9	847/844	841 (99 %)
11	77	77 (100 %)	10	145	142 (98 %)

genomic DNA in clone 124 (Ala-202-Val, Ile-224-Thr, Asn-246-Ser, Arg-293-Gln, Glu-382-Ser, Ser-384-Glu, Thr-404-Ile). These amino acids account for less than 2 % of the 366 residues obtained. The genomic sequence in clone 124 could thus represent an allelic variant of the *CYP4A24* gene. The intron–exon distribution within the cloned genomic segment, assigned to *CYP4A24*, was found to agree fully with that of the *CYP4A21* gene. As shown in Table 2, a high degree of overall nucleotide identity is seen between the *CYP4A21* and *CYP4A24* gene sequences. A region of higher divergence, which is clearly visible in an alignment of the *CYP4A21* and *CYP4A24* genes, is found between exons 6 and 8. The decreased nucleotide identity over this region could imply that gene conversion has occurred during the evolution of *CYP4A21*, since both exons and introns are affected.

Characterization of the 5'-flanking regions of *CYP4A21* and *CYP4A24*

GenomeWalker DNA libraries were used to obtain the 5'-flanking regions of the *CYP4A21* and *CYP4A24* genes. Using reverse nested primers based on nucleotides in the first exon combined with likewise-nested GenomeWalker adaptor primers, a PCR fragment of about 500 bp was amplified. The reverse primers hybridized to exon 1 sequences in both *CYP4A21* and *CYP4A24*. A restriction site for *BsgI* was used to distinguish between sequences derived from *CYP4A21* (*BsgI*-negative) and *CYP4A24* (*BsgI*-positive), respectively. Two sequences were obtained which could be assigned to *CYP4A21* and *CYP4A24* based on minor nucleotide differences in the first exon, as indicated in Figure 2. These nucleotides also exclude the exon 1 sequence of the fatty acid hydroxylase *CYP4A25* [5]. The nucleotide sequences shown in Figure 2 have been deposited in the EMBL database under accession numbers AJ586859 and AJ586860, respectively. The sequence of the amplified fragment obtained from clone 1732, denoted 1732-C in Figure 1, completely agreed with the sequence assigned to *CYP4A21* (Figure 2).

The 5'-flanking regions of the *CYP4A21* and *CYP4A24* genes are identical except for two nucleotides. Similar to what has been reported for *CYP4A* genes in other species, there is no consensus TATA box within the 40 bp sequence upstream from the transcription start site, which has been localized to a position 37 bp upstream of the start codon in *CYP4A21* [3]. Using the Transcription Element Search Software (TESS; <http://www.cbil.upenn.edu/>) multiple binding sites for the ubiquitous mammalian transcription factor Sp1 (specificity protein 1) were identified

within the proximal 100 bp sequence upstream of the transcription start site. Promoters of TATA-less genes that exhibit this feature are most commonly housekeeping genes [14], i.e. genes whose products are required in all cells at low levels. Sp1 sites in the promoter of the housekeeping gene dihydrofolate reductase (*Dhfr*) have been proposed to be a platform for assembling of the basal transcription machinery [15]. Two binding sites for the transcription factor YY1 (Yin-Yang-1) were likewise identified as indicated in Figure 2. The YY1 binds a subset of control elements called initiator (Inr) [16] and may act as a co-activator for Sp1 [17]. The high sequence identity within this short 5'-flanking region, however, hampered the identification of *cis*-acting elements involved in gene-specific regulation of *CYP4A21* and *CYP4A24*.

Northern blot analysis using oligo-probes

Samples of total RNA from livers of pigs of different ages were subjected to Northern blot analysis using a *CYP4A21*-specific oligonucleotide and an oligonucleotide based on a sequence common to all *CYP4A* fatty acid hydroxylases, respectively. As shown in Figure 3, a distinct pattern of mRNA expression was obtained with these two oligo-probes. The stronger hybridization signals obtained with the *CYP4A21*-specific oligo-probe indicate a more active transcription of the *CYP4A21* gene in liver compared with the vague signals obtained with the *CYP4A* fatty acid hydroxylase-directed oligo-probe. This difference is seen in all samples and appears not to be dependent on circumstances related to age and diet, i.e. suckling versus weaned pigs.

The ability of the probes to discriminate between *CYP4A21* and *CYP4A* fatty acid hydroxylases was examined using reverse-transcribed RNA samples of *CYP4A21* and *CYP4A24* cDNAs, respectively. As shown in the control panel in Figure 3, the probes clearly recognize each distinct RNA sample. The oligo-probes differ in seven out of nineteen nucleotides, as shown in the alignment in Figure 4, and hybridizes to position 958 (on the bottom strand) in the coding regions of *CYP4A21* and *CYP4A24*, respectively.

DISCUSSION

The *CYP4* family is one of the oldest families in the cytochrome P450 superfamily of enzymes (1.25 billion years), and is suggested to have been involved in maintaining membrane integrity in early eukaryotes [1]. Although members of the *CYP4A* subfamily are structurally conserved enzymes, specialized in fatty acid hydroxylation, a comparison between species and *CYP4A* isoenzymes indicates a species-specific evolution of this gene subfamily. The number of *CYP4A* enzymes and the substrate specificity towards various fatty acids and prostaglandins differ among isoenzymes and species, as do the regulation of these genes [1,2].

The *CYP4A21* is clearly an atypical member of the *CYP4A* subfamily with a unique role in bile acid biosynthesis in the pig. The results presented here show that the overall organization of the *CYP4A21* gene is highly conserved when compared with available *CYP4A* genes from other species. The cDNAs of *CYP4A21* and the porcine *CYP4A* fatty acid hydroxylase *CYP4A24* share 94 % nucleotide identity. The apparent overall sequence similarity and the conserved intron–exon organization of the genes shown here indicate that the *CYP4A21* and the pig *CYP4A* fatty acid hydroxylase *CYP4A24* have evolved from an ancestral gene by gene duplication. The classical model for the origin of new gene functions by gene duplication proposes that one copy of a particular gene maintains the original function, whereas the other copy can accumulate mutations for further evolution of

Figure 3 Northern blot analysis and control hybridization using oligo-probes

Northern blot analysis of total RNA isolated from livers of eight different pigs of varying ages is shown. The Northern blots were probed with a ^{32}P -labelled CYP4A21-specific oligonucleotide (**A**) and a ^{32}P -labelled CYP4A-derived oligo based on nucleotides in a conserved sequence common to all CYP4As except for CYP4A21 (**B**), respectively, as described in the Experimental section. The positions of 28 S (4.8 kb) and 18 S (1.9 kb) ribosomal RNAs used as built-in molecular-mass markers and the estimated size of hybridization-positive bands (2.5 kb) are indicated. Four samples from unweaned pigs (2 and 6 weeks old) and four from weaned pigs (3 and 6 months old) were loaded as indicated in the Figure. Ethidium bromide staining of 18 S rRNA on the blotted filters is shown below each blot. Control hybridization with each ^{32}P -labelled oligo-probe using reversed transcribed cDNAs of CYP4A21 and CYP4A24, applied to Hybond-N membranes, is shown along with each blot. Circles mark the positions of the applied RNAs.

RNON-CONS probe	TGGAGTCGTGACCTGCAGC
RCONS probe	TGCTGTGATGACCTGCGAA

Figure 4 Sequences of the oligo-probes used for Northern blotting

An alignment between the two oligo-probes (RNON-CONS and RCONS) that were used for Northern blot is shown. Nucleotides common to both sequences are boxed.

new functions [18]. CYP4A21 and the ability to form hyocholic acid in pigs have probably provided a selective advantage, as this gene is fixed in the pig genome and not silenced as a null allele, which is a more common fate of duplicated genes [19].

The structural features governing the ability of CYP4A21 to 6 α -hydroxylate chenodeoxycholic acid is not fully known. An identified anomaly in the deduced amino acid sequence of CYP4A21 is found around residue 315, within a conserved signature sequence common to CYP4A fatty acid hydroxylases [20]. Among the diverging amino acids is a glutamic acid residue that has been shown to participate in covalent binding of haem in CYP4 proteins [21–23]. This residue is replaced by alanine in CYP4A21. Introduction of the conserved signature sequence, including the glutamic acid residue, has been shown to abolish the 6 α -hydroxylase activity of the recombinant expressed enzyme, demonstrating that residues within this sequence have significant impact on the catalytic activity [3]. The CYP4A signature motif or its counterpart in CYP4A21 is located within exon 8. A region with a high degree of divergence between the cDNAs and deduced amino acid sequences of CYP4A21 and CYP4A24 coincides with the nucleotides found in exons 6–8. The intronic sequences within this region are likewise less conserved between the *CYP4A21* and *CYP4A24* genes compared with the upstream and downstream sequences (Table 2). Gene duplication as well as gene conversion events are considered important in the evolution of cytochrome P450 enzymes [24–27]. Gene conversion allows genes to procure DNA segments from other more or less distantly related genes

and even non-functional pseudogenes, thus producing genes encoding proteins with novel functions. Whereas the overall sequence identity and conserved intron–exon distribution between *CYP4A21* and *CYP4A24* indicate a common origin by gene duplication, gene conversion could account for the region of decreased sequence identity observed from exon 6 to exon 8. The catalytic properties of CYP4A21 could thus originate from a recombination event affecting this region in a duplicated copy of a *CYP4A* gene. The notion that gene conversion has occurred in the evolution of *CYP4A21* is strongly supported by the fact that the region of decreased sequence identity is restricted to a segment of the gene and includes both exons and introns. The pig genome project is ongoing [28] and will, when completed, hopefully provide the additional sequence data needed to fully elucidate the origin of the *CYP4A21* gene.

Dietary habits have been proposed as factors governing the evolution of P450 enzymes [29]. An adaptation to circumstances related to lipid components in the diet could have contributed to the apparent species-specific diversification of the CYP4A subfamily, at least to the evolution CYP4A21 in pig. According to Haslewood [4] there is an association between mammalian bile salts and the type of diet. Carnivores have preferably trihydroxy bile acids, whereas herbivores generally have dihydroxy or keto-hydroxy acids, and omnivores a good proportion of these various types. The domestic pig of today is an omnivore and has a similar balance between tri- and di-hydroxy bile acids to humans. Hyocholic acid generated by CYP4A21 can theoretically fulfil a basic requirement for trihydroxy bile acids in the pig in the absence of cholic acid. Both cholic acid and hyocholic acid are trihydroxy bile acids, but the position of the third hydroxyl group (12 α and 6 α , respectively) has a profound effect on the hydrophobic/hydrophilic properties of these bile acids [30,31]. Hence, the presence of hyocholic acid in pig bile, instead of cholic acid, will probably also have an impact on cholesterol and lipid metabolism [32–34].

Despite the overall similarity between the *CYP4A21* and *CYP4A24* genes in the proximal 5'-flanking region, distinct

expression patterns of these genes were shown. CYP4A21 appears to be highly expressed in pig liver, with a concomitant low basal expression of CYP4A fatty acid hydroxylases. The lack of a consensus TATA box is common to all CYP4A genes hitherto cloned. Combined with the presence of multiple binding sites for the transcription factor Sp1 upstream of the transcription start site, the promoters of the pig CYP4A genes exhibit features resembling those of housekeeping genes. Multiple putative Sp1-binding sites are present also in the promoters of CYP4A genes from other species. The CYP4A enzymes present today could thus be descendants from an ancestral housekeeping gene with a maintained basal gene expression level, but with the addition of specific *cis*-acting elements conferring tissue-, and isoenzyme-specific regulation. Binding sites for YY1 were found in the proximal promoter region of pig CYP4As. YY1 and Sp1 have been shown to interact physically and to act synergistically in the transcriptional activation of promoters containing multiple Sp1-binding sites [17].

In summary, the present study shows that the pig CYP4A21 is a distinct gene which exhibits a high sequence identity and equivalent intron–exon distribution compared with the pig CYP4A fatty acid hydroxylase gene (CYP4A24), indicating a common origin by gene duplication. Sequence comparison indicates that the unique catalytic properties of CYP4A21 could originate from gene-conversion events affecting the genomic sequence from exon 6 to exon 8. Despite a high sequence identity between CYP4A21 and CYP4A24 within the proximal 5'-flanking region, the patterns of expression of the CYP4A21 and CYP4A fatty acid hydroxylase genes in pig liver differ. The pig CYP4A genes with distinct pattern of expression could serve as future tools to elucidate mechanisms for expression of CYP4A genes and also contribute to new aspects on their evolution.

This work was supported by the Swedish Research Council – Medicine (project 03X-218).

REFERENCES

- Simpson, A. E. C. M. (1997) The cytochrome P450 4 (CYP4) family. *Gen. Pharmacol.* **28**, 351–359
- Okita, R. T. and Okita, J. R. (2001) Cytochrome P450 4A fatty acid omega hydroxylases. *Curr. Drug Metab.* **2**, 265–281
- Lundell, K., Hansson, R. and Wikvall, K. (2001) Cloning and expression of a pig liver taurochenodeoxycholic acid 6 α -hydroxylase (CYP4A21). *J. Biol. Chem.* **276**, 9606–9612
- Haslewood, G. A. D. (1967) Bile salt evolution. *J. Lipid Res.* **8**, 535–550
- Lundell, K. (2002) Cloning and expression of two novel pig liver and kidney fatty acid hydroxylases [cytochrome P450 (CYP)4A24 and CYP4A25]. *Biochem. J.* **363**, 297–303
- Kimura, S., Hanioka, N., Matsunaga, E. and Gonzalez, F. J. (1989) The rat clofibrate-inducible CYP4A gene subfamily I. Complete intron and exon sequence of CYP4A1 and CYP4A2 genes, unique exon organization, and identification of a conserved 19-bp upstream element. *DNA* **8**, 503–516
- Muerhoff, A. S., Griffin, K. J. and Johnson, E. F. (1992) Characterization of a rabbit gene encoding a clofibrate-inducible fatty acid ω -hydroxylase: CYP4A6. *Arch. Biochem. Biophys.* **296**, 66–72
- Palmer, C. N. A., Griffin, K. J. and Johnson, E. F. (1993) Rabbit prostaglandin ω -hydroxylase (CYP4A4): gene structure and expression. *Arch. Biochem. Biophys.* **300**, 670–676
- Bellamine, A., Wang, Y., Waterman, M. R., Gainer, J. V., Dawson, E. P., Brown, N. J. and Capdevila, J. H. (2003) Characterization of the CYP4A11 gene, a second CYP4A gene in humans. *Arch. Biochem. Biophys.* **409**, 221–227
- Kawashima, H., Naganuma, T., Kusunose, E., Kono, T., Yasumoto, R., Sugimura, K. and Kishimoto, T. (2000) Human fatty acid ω -hydroxylase, CYP4A11: determination of complete genomic sequence and characterization of purified recombinant protein. *Arch. Biochem. Biophys.* **378**, 333–339
- Lundell, K. and Wikvall, K. (2003) Gene structure of pig 12 α -hydroxylase (CYP8B1) and expression in fetal liver: comparison with expression of taurochenodeoxycholic acid 6 α -hydroxylase (CYP4A21). *Biochim. Biophys. Acta* **1634**, 86–96
- Proudfoot, N. J., Furger, A. and Dye, M. J. (2002) Integrating mRNA processing with transcription. *Cell* **108**, 501–512
- Savas, Ü., Hsu, M. H. and Johnson, E. F. (2003) Differential regulation of human CYP4A genes by peroxisome proliferators and dexamethasone. *Arch. Biochem. Biophys.* **409**, 212–202
- Carey, M. and Smale, S. T. (2000) *Transcriptional Regulation in Eucaryotes*, pp. 1–50, Cold Spring Harbour Press, Cold Spring Harbour, NY
- Slansky, J. E. and Farnham, P. J. (1996) Transcriptional regulation of the dihydrofolate reductase gene. *BioEssays* **18**, 55–62
- Smale, S. T. (1997) Transcription initiation from TATA-less promoters within eucaryotic protein-coding genes. *Biochim. Biophys. Acta* **1351**, 73–88
- Lee, J. S., Galvin, K. M. and Shi, Y. (1993) Evidence for physical interaction between the zinc-finger transcription factors YY1 and Sp1. *Proc. Natl. Acad. Sci. U.S.A.* **90**, 6145–6149
- Long, M. (2001) Evolution of novel genes. *Curr. Opin. Genet. Dev.* **11**, 673–680
- Walsh, B. J. (1995) How often do duplicated genes evolve new functions? *Genetics* **139**, 421–428
- Gasser, R. and Philpot, R. M. (1989) Primary structures of cytochrome P-450 isozyme 5 from rabbit and rat and regulation of species-dependent expression and induction in lung and liver: identification of cytochrome P-450 gene subfamily IVB. *Mol. Pharmacol.* **35**, 617–625
- Henne, K. R., Kunze, K. L., Zheng, Y. M., Christmas, P., Soberman, R. J. and Rettie, A. E. (2001) Covalent linkage of prosthetic heme to CYP4A family P450 enzymes. *Biochemistry* **40**, 12925–12931
- Hoch, U. and Ortiz de Montellano, P. R. (2001) Covalent linked heme in cytochrome P4504A fatty acid hydroxylases. *J. Biol. Chem.* **276**, 11339–11346
- LeBrun, L. A., Hoch, U. and Ortiz de Montellano, P. R. (2002) Autocatalytic mechanism and consequences of covalent heme attachment in the cytochrome P4504A family. *J. Biol. Chem.* **277**, 12755–12761
- Gonzalez, F. J. and Nebert, D. W. (1990) Evolution of the P450 gene superfamily: animal-plant 'warfare', molecular drive and human genetic differences in drug oxidation. *Trends Genet.* **6**, 182–186
- Atchison, M. and Adesnik, M. (1986) Gene conversion in a cytochrome P-450 gene family. *Proc. Natl. Acad. Sci. U.S.A.* **83**, 2300–2304
- Lewis, D. F. V., Watson, E. and Lake, B. G. (1998) Evolution of the cytochrome P450 superfamily: sequence alignments and pharmacogenetics. *Mutat. Res.* **410**, 245–270
- Degtyarenko, K. N. and Archakov, A. I. (1993) Molecular evolution of P450 superfamily and P450-containing monooxygenase systems. *FEBS Lett.* **332**, 1–8
- Li, H. (2000) Gene sequencing. China, Denmark team up to tackle the pig. *Science* **290**, 913–914
- Nebert, D. W. (1997) Polymorphisms in drug-metabolizing enzymes: what is their clinical relevance and why do they exist? *Am. J. Hum. Genet.* **60**, 265–271
- Armstrong, M. J. and Carey, M. C. (1982) The hydrophobic-hydrophilic balance of bile salts. Inverse correlation between reverse-phase high performance liquid chromatographic mobilities and micellar cholesterol-solubilizing capacities. *J. Lipid Res.* **23**, 70–80
- Heuman, D. M. (1989) Quantitative estimation of the hydrophilic-hydrophobic balance of mixed bile salt solutions. *J. Lipid Res.* **30**, 719–730
- Hofmann, A. F. and Roda, A. (1984) Physicochemical properties of bile acids and their relationship to biological properties: an overview of the problem. *J. Lipid Res.* **25**, 1477–1489
- Heuman, D. M., Hylemon, P. B. and Vlahcevic, Z. R. (1989) Regulation of bile acid synthesis. III. Correlation between biliary bile salt hydrophobicity index and the activities of enzymes regulating cholesterol and bile acid synthesis in the rat. *J. Lipid Res.* **30**, 1161–1171
- Cohen-Solal, C., Parquet, M., Férézou, J., Sérougne, C. and Lutton, C. (1995) Effects of hyodeoxycholic acid and α -hyocholic acid, two 6 α -hydroxylated bile acids, on cholesterol and bile acid metabolism in the hamster. *Biochim. Biophys. Acta* **1257**, 189–197

Received 30 October 2003; accepted 25 November 2003

Published as BJ Immediate Publication 25 November 2003, DOI 10.1042/BJ20031657