Structure and Expression Analysis of Nuclear Receptor Class 1 Genes in

Sus scrofa

ブタの核内受容体クラス１遺伝子の構造解析と発現解析

PATTANAPONG THADTHA

2008

ABSTRACT

Structure and Expression Analysis of Nuclear Receptor Class 1 Genes in

Sus scrofa

BY

PATTANAPONG THADTHA

Nuclear receptors are ligand-inducible transcription factors that have been known as transcriptional regulators of genes involved in growth and development, reproduction, metabolism, and immunity. Because studies of nuclear receptors in farm animals, especially in swine, are very few, our objective was to identify and sequence cDNAs encoding nuclear receptors *CAR*, *PXR*, *LXRα*, and *RORγ* in pigs. We searched the cDNA clones in the PEDE database (Pig EST Data Explorer) to identify nuclear receptors *CAR*, *PXR*, *LXRα*, and *RORγ* in swine by using human sequences as queries. In present study, we report the primary sequence data of full-length cDNA clones encoding porcine nuclear receptors *CAR*, *PXR*, *LXRα*, and *RORγ*.

The constitutive androstane receptor (*CAR*; NR1I3) and the pregnane X receptor (*PXR*; NR1I2) are members of the nuclear receptor superfamily that act as transcription factors. They regulate the expression of several important genes that encode enzymes responsible for the metabolism of endogenous and exogenous substances such as *CYP3A4*, *CYP2B6*, and *UGT1A1*. The porcine *CAR* gene has a 1407 bp mRNA, which contains a 216-bp 5'-untranslated region (UTR), a 144-bp 3'-UTR, and a 1047-bp coding region that encodes 348 amino acids. The mRNA sequence of porcine *PXR* is composed of 2567 bp that contains a 273-bp 5'-UTR, a

1266-bp putative coding region encoding 421 amino acids, and a 1028-bp 3'-UTR. The porcine CAR and PXR proteins showed a high degree of sequence identity in their DNA-binding domain (DBD) and ligand-binding domain (LBD) with 80%-90% amino acid identity in comparison with those of humans. The mRNA expression of both porcine *CAR* and *PXR* was detected in liver, small intestine, and kidney by using reverse-transcription (RT)-PCR.

The nuclear receptor liver X receptor alpha (*LXRα*; NR1H3) is a ligand-activated transcription factor that belongs to a member of the nuclear receptor superfamily. LXRα functions as a transcriptional regulator of several important genes involved in the metabolism of cholesterol, fatty acids, and glucose. We have identified and sequenced three porcine *LXRα* transcripts. The porcine *LXRα-1* and *-2* are identical in open reading frames and 3'-UTRs, but differ in 5'-UTRs (nucleotide positions 1-145 in *LXRα-1* and 1-178 in *LXRα-2*). The partial porcine *LXRα-3* transcript represents an incomplete transcript that has an in-frame TGA stop codon, resulting in a truncated protein lacking amino acid residues downstream from the DBD. Analysis of amino acid sequences revealed that porcine LXRα DBD and LBD is 97% and 100% identical to human LXRα DBD and LBD, respectively. By using RT-PCR, the expression of the porcine *LXRα-1* transcript was detected in the liver, kidney, small intestine, heart, muscle, thymus, spleen, and brain, whereas the porcine *LXRα-2* transcript was expressed in the thymus and spleen.

The retinoic acid receptor-related orphan receptor gamma (*RORγ*; NR1F3) is a member of nuclear receptor superfamily that has been proposed to involve in the T lymphocyte and lymph node development. We have identified and sequenced four porcine *RORγ* transcripts, designated porcine *RORγ-1*, *RORγ-2*, *RORγ-3*, and *RORγ-4*. All four transcripts are different in the non-coding 5'-UTR, but highly similar in 3'-

UTR. The porcine *RORγ-1* transcript has an additional-26 nucleotide in 5'-UTR and an additional-63 base in the open reading frame. The 63-base insertion encodes additional 21 amino acid residues in N-terminal of the D domain when compared to porcine *RORγ-2*. The transcript variant 3 and 4 encode porcine RORγ isoforms lacking functional DBD and/or LBD. Compared to human RORγ receptors, porcine RORγ-1 is similar to human RORγ-1 isoform, whereas porcine RORγ-2 is similar to human RORγ-2 isoform. The DBD and LBD of porcine RORγ are 97% and 93% identical to human RORγ DBD and LBD, respectively. The mRNA expression of the porcine *RORγ-1* transcript was detected in the liver, kidney, small intestine, heart, muscle, thymus, spleen, and brain, whereas the porcine *RORγ-2* transcript was expressed in the thymus and spleen.

Keywords: nuclear receptor, *PXR*, *CAR*, *LXRα*, *RORγ*, cDNA sequence, expression, transcript, swine

ACKNOWLEDGMENTS

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

1.1. Nuclear Receptors

The nuclear receptor (NR) superfamily comprises of several transcription factors that have been known to play important roles in many important biological processes, including growth and development, metabolism, reproduction, and homeostasis (Mangelsdorf *et al.,* 1995; Giguere, 1999; Robinson-Rechavi *et al.,* 2003). NRs are found in all animals from invertebrates to mammals. The predicted nuclear receptors in the human (*Homo sapiens*), mouse (*Mus musculus*), rat (*Rattus norvegicus*), fruit fly (*Drosophila melanogaster*), and nematode (*Caenorhabditis elegans*) are 48, 49, 47, 21, and 270 nuclear receptors, respectively (Maglich *et al.,* 2001; Sluder and Maina, 2001; Zhang *et al.,* 2004).

NRs are classified and grouped, based on phylogenetic tree constructed from the two well-conserved domains of NRs (the DNA-binding domain and the ligand-binding domain), into six-gene groups (Nuclear Receptors Nomenclature Committee, 1999). NRs are composed of four structural domains, from amino acid terminus to carboxyl terminus, the A/B domain, the DNA-binding domain (DBD), the D domain, and the ligand-binding domain (LBD) (Figure 1.1). The A/B domain represents the most variable domain, both in length and sequence, of nuclear receptors. It is thought that the A/B domains also confer the transcriptional activation function, called ligand-

independent activation function-1 (AF-1). The highly conserved DBD is responsible for the recognition of specific DNA sequences in the response element of target genes. The DBD is the most conserved domain of nuclear receptors and contains two zinc-finger motifs (Figure 1.2). The first zinc motifs ($Cys-X_2-Cys-X_{13}-Cys-X_2-Cys$) determine the specificity of the DNA recognition, and the second zinc motifs ($Cys-X_5-Cys-X_9-Cys-X_2-Cys$) are involved in dimerization (Mangelsdorf *et al.,* 1995; Giguere, 1999; Khorasanizadeh and Rastinejad, 2001; Robinson-Rechavi *et al.,* 2003). The DBDs also contain the highly conserved P-box (CEGCKGFF), the T-box, and the carboxyl-terminal extension (CTE) (Figure 1.2). The conserved DBD has approximately 70 amino acids in length. The D domain is variable in both length and sequence identity percentage. It is thought that the D region is involved in dimerization and nuclear transport. The LBD is conserved but less than the DNA-binding domain. The LBD is a site for ligand binding, a dimerization function, and also responsible for the interaction with both coactivator and corepressor proteins. The ligand-binding domains also contain the ligand-dependent activation-2 (AF-2) (Mangelsdorf *et al.,* 1995; Giguere, 1999; Robinson-Rechavi *et al.,* 2003).
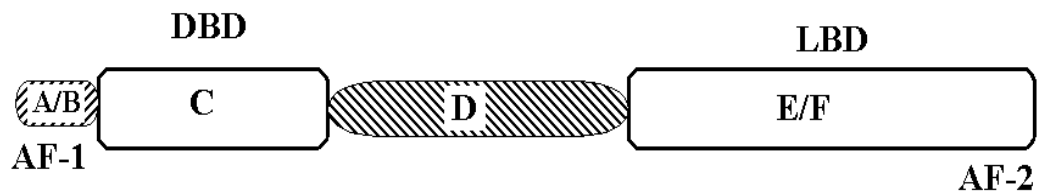
Figure 1.1. Schematic representing structural domains of NRs.

The nuclear receptor comprises of four structural domains, the A/B, the DNA-binding domain (DBD), the hinge D, and the ligand-binding domain (LBD).
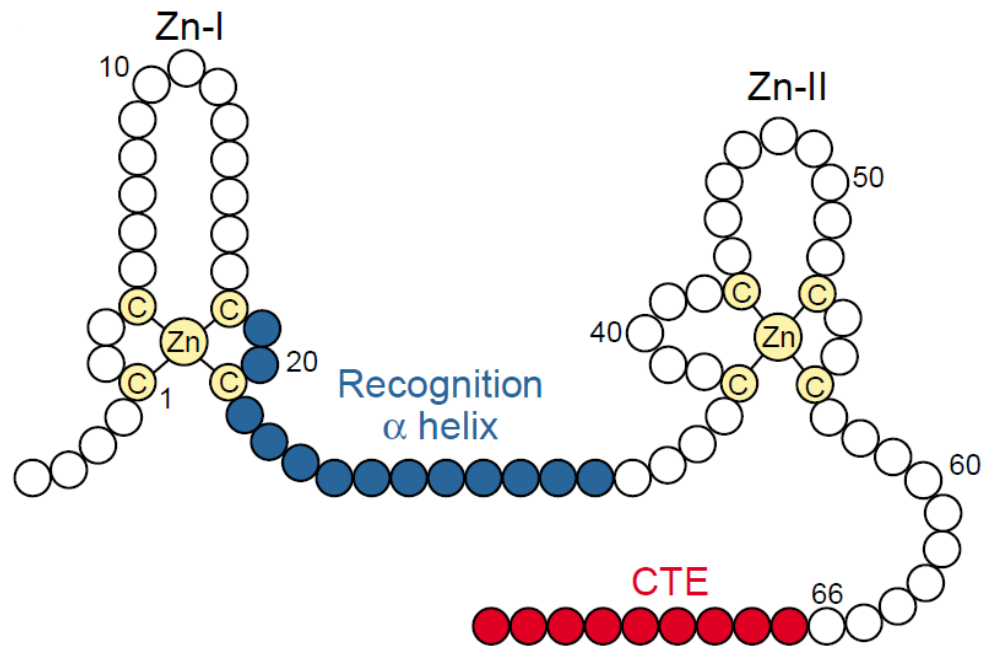
Figure 1.2. Schematic representing the structure of DBD (Khorasanizadeh and Rastinejad, 2001).

The DBDs consist of two zinc-finger motifs (Zn-I and Zn-II), a conserved recognition α helix, and a C-terminal extension (CTE). The first zinc motifs determine the specificity of the DNA recognition and the second zinc motifs are involved in dimerization.

NRs are activated by ligands such as fatty acids, bile acids, steroids, xenobiotics, and other lipid soluble compounds (Figure 1.3). Ligand binding changes the conformation of the LBD and allows the AF-2 domain to interact with co-activators such as steroid receptor coactivator-1 (SCR-1) and nuclear receptor coactivator-2 (NCoA-2), resulting in the transcriptional activation of target genes (Mangelsdorf *et al.,* 1995; Giguere, 1999; Robinson-Rechavi *et al.,* 2003). Several of NRs form heterodimers with the retinoid X receptor (RXR) and bind to specific DNA sequences called nuclear receptor response elements (NRREs) located at the promoter regions of target genes (Mangelsdorf and Evans, 1995).
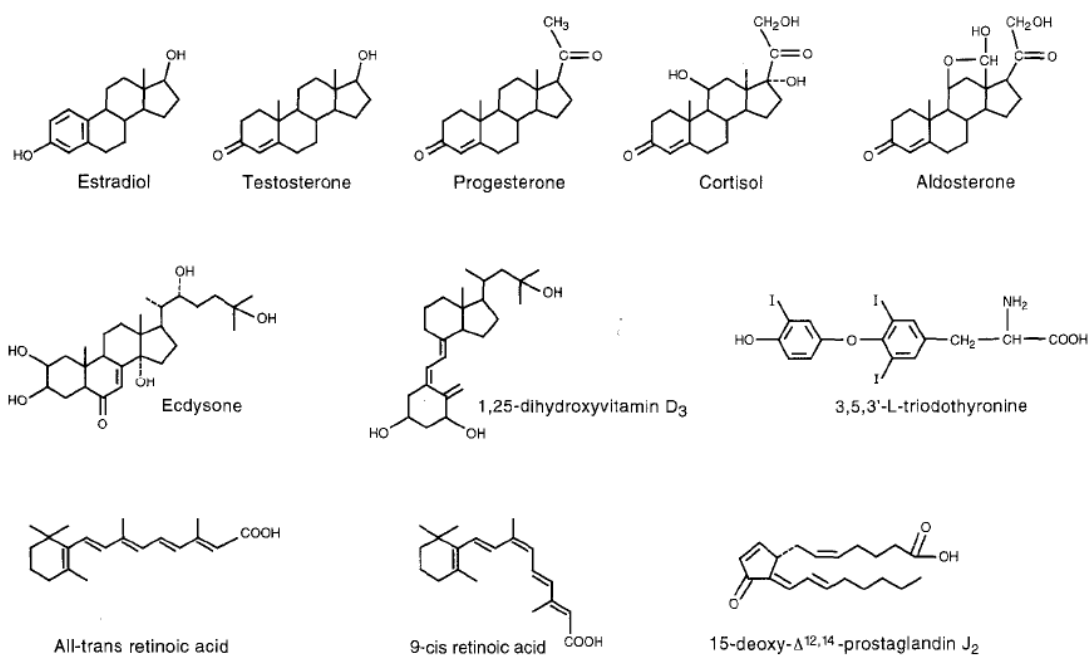
Figure 1.3. Structures of the known ligands for nuclear receptors (Mangelsdorf *et al.,* 1995). NRs are activated by ligands such as fatty acids, bile acids, steroids, xenobiotics, and other lipid soluble compounds.

The nuclear receptors NR1F3 (retinoic acid receptor-related orphan receptor gamma, *RORγ*), NR1H3 (liver X receptor alpha, *LXRα*), NR1I2 (pregnane X receptor, *PXR*), and NR1I3 (constitutive androstane receptor, *CAR*) are members of the nuclear receptor subfamily 1 that have been implicated in the regulation of genes involved in metabolism, reproduction, and immunity (Peet *et al.,* 1998a; Kliewer *et al.,* 2002; Eberl and Littman, 2003) (Table 1.1). These nuclear receptors can be activated by numerous ligands. The PXR receptors can be activated by a large number of ligands such as steroid hormones (e.g., 17β-estradiol, 5β-pregnane-3,20-dione, corticosterone), bile acids (e.g., lithocholic acids, 6-keto lithocholic acid), xenobiotics (e.g., rifampicin, pregnenolone 16α -carbonitrile) (Bertilsson *et al.,* 1998; Lehmann *et al.,* 1998; Moore *et al.,* 2000; Kliewer *et al.,* 2002; Timsit and Negishi, 2007; Xue *et al.,* 2007). In addition, the endocrine disrupting chemicals (EDCs) are also PXR ligands such as pesticides (methoxychlor, endosulfan, dieldrin, DDT), phthalic acid, and the

plasticizer nonylphenol (Masuyama *et al.,* 2000; Mikamo *et al.,* 2003; Wyde *et al.,* 2003; Kretschmer and Baldwin, 2005; Lemaire *et al.,* 2006). Moreover, the plant-derived phytoestrogens (flavonoids and lignans) and the fungi-biosynthesized mycoestrogen zearalenone are also the activators of PXR receptor proteins (Jacobs *et al.,* 2005; Ding *et al.,* 2006; Mnif *et al.,* 2007). The CAR receptors can be activated by several ligands such as 5β-pregnane-3,20-dione, androstanol (5α-androstan-3α-ol), androstenol ([16(5α)-androsten-3α-ol]), phenobarbital (PB), 1,4-bis[2-(3,5-dichloropyridyloxy)]benzene (TCPOBOP), EDCs (e.g., nonylphenol and DDE), and drugs (Forman *et al.,* 1998; Moore *et al.,* 2000; Wyde *et al.,* 2003; Hernandez *et al.,* 2007; Timsit and Negishi, 2007). The receptor LXRα is activated by naturally occurring oxysterols, which are the cholesterol metabolites in the cholesterol biosynthetic pathway such as 22 (R)-hydroxycholesterol, 24(S),25-epoxycholesterol, 24(S)-hydroxycholesterol (Lehmann *et al.,* 1997; Janowski *et al.,* 1999), and a synthetic compound T0314407, which is a selective agonist of LXRs (Schultz *et al.,* 2000). The ligand for RORγ has not yet been identified. After ligand binding, CAR, PXR, and LXRα receptors change the conformation of the LBD and allows the AF-2 domain to interact with co-activators such as steroid receptor coactivator-1 (SCR-1) and nuclear receptor coactivator-2 (NCoA-2), resulting in the transcriptional activation of target genes (Peet *et al.,* 1998a; Orans *et al.,* 2005; Timsit and Negishi, 2007). These receptors form a heterodimer with RXR and binds to response element (AGGTCA) located in the promoter of target genes known to be involved in the metabolism (e.g., *CYP3A4, CYP2B6, UGT1A1, SREBP-1c,* and *CYP7A1*) (Bertilsson *et al.,* 1998; Lehmann *et al.,* 1997; Kliewer *et al.,* 1998; Lehmann *et al.,* 1998; Peet *et al.,* 1998b; Kawamoto *et al.,* 1999; Schultz *et al.,* 2000; Xie *et al.,* 2003). The binding of ROR to response elements in target genes has been shown to bind as a monomer to

a single core motif AGGTCA preceded by an A/T rich region (Ortiz *et al.,* 1995), and

bind to a direct repeat AGGTCA separated by four and five nucleotides (Medvedev *et*

al., 1996), which is the same response element identified as the binding site for the

thyroid hormone receptors (TRs) and the retinoic acid receptors (RARs).

Table 1.1. NR1F3, NR1H3, NR1I2, and NR1I3 members of nuclear receptor subfamily 1.

Gene	Identified ligand	Target gene	Function
NR1F3 (*ROR*)	not yet identified	*TCR-Jα, Bcl-xL*	T lymphocyte and lymph node development
NR1H3 (*LXRα*)	Oxysterols, synthetic chemical (T0901317)	*CYP7A1, SREBP-1c*	Cholesterol metabolism, lipid & glucose metabolism
NR1I2 (*PXR*)	5β-pregnane-3,20-dione, 17β-estradiol, bile acids, drugs, EDCs	*CYP3A4, UGT1A1*	Metabolism of hormones, bile acids, and exogenous compounds
NR1I3 (*CAR*)	Androstanol, androstenol, 5β-pregnane-3,20-dione, TCPOBOP, drugs, EDCs	*CYP2B6, CYP2B10*	Metabolism of hormone & exogenous compounds

Both structure and function of these nuclear receptor genes have been

extensively studied in humans and mice. However, cDNAs encoding of these

important transcriptional regulators have not been cloned in farm animals, especially

in swine. The objective of this work was to isolate and sequence the full-length cDNA

clones encoding *CAR*, *PXR*, *LXRα*, and *RORγ* in porcine genome. We present the

results of this study as follows; chapter 1: introduction, chapter 2: materials and

methods, chapter 3: the nuclear receptor CAR and PXR, chapter 4: the nuclear

receptor LXRα, chapter 5: the nuclear receptor RORγ, and chapter 6: conclusion and

discussion.

CHAPTER 2

MATERIALS AND METHODS

2.1. Identification of PEDE Clones

The cDNA clones for sequencing of nuclear receptor mRNAs were from the PEDE (Pig EST Data Explorer) database (http://pede.dna.affrc.go.jp/) (Uenishi *et al.,* 2004). The clones used in the sequencing are shown in Table 2.1. The plasmids were kindly provided by Dr. Hirohide Uenishi (National Institute of Agribiological Sciences, Japan).

Table 2.1. PEDE clones used in the sequencing of porcine nuclear receptor genes.

Gene	PEDE clone	Animal	
		Tissue	Breed
RORγ 1	THY01_0063_F07	adult thymus	Landrace + Large White + Duroc
RORγ 2	THY01_0119_H06	adult thymus	Landrace + Large White + Duroc
RORγ 3	LVRM1_0186_D10	adult liver	Chinese Meishan
RORγ 4	THY01_0100_B07	adult thymus	Landrace + Large White + Duroc
LXRα 1	LVR01_0077_E03	adult liver	Landrace + Large White + Duroc
LXRα 2	LVR01_0034_E04	adult liver	Landrace + Large White + Duroc
LXRα 3	OVRM1_0199_G01	adult ovary	Chinese Meishan
PXR	LVR01_0073_E03	adult liver	Landrace + Large White + Duroc
CAR	LVRM1_0034_H03	adult liver	Chinese Meishan

Note: The PEDE clones were identified by using human cDNA sequences of *CAR*, *PXR*, *LXRα*, and *RORγ*.

2.2. Transformation

The transformation of *Escherichia coli* (*E. coli*) by electroporation method was used to introduce circular plasmid vector containing inserted DNA fragment into competent cells. Five microliters (μl) of purified plasmids was applied to a sterile 1.5 milliliter (ml) microcentrifuge tube (ice cold) and put on ice until needed. The aliquots of the JM 109 competent *E. coli* host cells (NIPPON GENE, Japan) were thawed according to used samples, and thawed competent cells were maintained on ice until used. Twenty-five microliters of JM 109 *E. coli* cells was added to chilled sample tube, followed by mixing using micropipette. Samples were incubated on ice for 10 min in order to transform circular plasmid vector into *E. coli* competent cell. Increasing transformation efficiency was performed by using the BTX Electroporation System ECM399 (NEPA GENE, Japan). Thirty microliters of transformed reactions was transferred to chilled BTX cuvette and run on the electroporation machine. Two milliliters of Luria-Bertani (LB) liquid medium was added to each sample in the BTX cuvette. The reaction mixture was then applied to test tube. Samples in test tubes were incubated and shaken at 200-250 revolutions per minute (rpm) at 37 ℃ for 1 h in an air incubator (EYELA incubator FMS). After one-hour incubation, fifty microliters of sample solution was spread in the solid media for bacterial growth, LB agar plates, supplemented with 50 μg/ml ampicillin by using a stirring stick. This procedure was done under an aseptic condition. The LB agar plates were sat on the bench for 2-3 min until inoculums of cells were completely absorbed into the LB agar. The plates were incubated by upside down the LB agar plates at 37 ℃ overnight.

2.3. Colony PCR

The colonies on LB agar plates were picked to the master plate using inoculating loop. After inoculation of colonies to the master plate, the loops were sterilized by heat flaming. The master plates were incubated at 37 ℃ overnight in an EYELA air incubator by inverted the cover of the plate so that LB agar faced down. The colony direct PCR was then performed. The bacterial colony from master plate was picked and transferred to a 0.5 ml PCR tube, which contains 20 μl of LB liquid medium containing 50 μg/ml of antibiotic ampicillin. The 0.5 ml tubes of sample solution were then incubated at 37 ℃ for 2 hours in the BIOSHAKER M.BR-022. After two hours of incubation, the samples were spun down and 2 μl of solution was applied to a new 0.5 ml tube containing 1.375 μl of pure water on ice. The mixture was briefly centrifuged and heated using a thermal cycler (Techne; TECHNE, UK). The incubation was at 95 ℃ for 10 min, followed by a final hold at 4 ℃. The solution was used as a template for PCR. The PCR reaction mixture containing 0.125 μl of *ExTaq* Hot Start, 2.5 μl of PCR reaction buffer, 2 μl of dNTPs, primers M13-M4 (5'-GTTTTCCCAGTCACGAC-3') and T7 (5'-TAATACGACTCACTATAGGG-3') each 1 μl, and 15 μl of pure water, was added to template tubes. This step was done on ice. The sample tubes were briefly centrifuged and subjected to PCR amplification (Techne thermal cycler). The PCR cycles were performed as follows; initial denaturation at 95 ℃ for 5 min, 40 cycles of denaturation at 95 ℃ for 30 s, annealing at 53 ℃ for 30 s, and elongation at 72 ℃ for 1 min, followed by 72 ℃ for 10 min and a final hold at 4 ℃. Five microliters from colony direct PCR was subjected to 1.2% agarose gel electrophoresis. After the electrophoresis, the DNA bands were visualized by using UV transilluminator.

2.4. Plasmid DNA Extraction

The colony of plasmid with inserted DNA fragment was transferred to test tube containing 2 ml of LB liquid medium. Samples were incubated and shaken at 200-250 rpm at 37 ℃ in an EYELA air incubator overnight. The purification of plasmid DNA was done using the QIAprep Spin Miniprep Kit (QIAGEN, Japan). The procedures were done according to the manufacturer's instructions. The test tube was thoroughly shaken and transferred solution to a new 1.5 ml microcentrifuge tube. To collect bacterial pellet, the centrifugation of sample tubes was done at 12,000 rpm for 3 min. The supernatant was carefully discarded without disturbing the pellet by using a micropipette. The bacterial pellet was resuspended by adding 250 µl of P1 buffer (resuspension buffer). The RNase A (Ribonuclease), provided by the manufacturer, was added to resuspension buffer P1 before used. The sample tubes were shaken at high speed for 2 min in the BIOSHAKER M.BR-022, followed by brief centrifugation. Two hundred and fifty microliters of lysis buffer P2 was added to sample tubes, followed by gently inverting. Three hundred and fifty microliters of buffer N3 (neutralization buffer) was added, and the tube was immediately mixed by gently inverting. The samples were then centrifuged at 13,000 rpm for 10 min. Following centrifugation, the supernatant was carefully applied to the QIAprep spin column. The QIAprep spin column was centrifuged at 13,000 rpm for 1 min, followed by discarding the flow-through in the collection tube. After adding 500 µl of PB buffer (binding buffer), the column was centrifuged at 13,000 rpm for 1 min, followed by discarding the flow-through. Seven hundred and fifty microliters of PE buffer (wash buffer) was added to sample tube and centrifuged at 13,000 rpm for 1 min, followed by discarding the flow-through. The QIAprep spin column was subjected to centrifugation at 13,000 rpm for 1 min, followed by discarding the collection tube

with the flow-through. The QIAprep spin column was placed in a clean 1.5 ml microcentrifuge tube. The dilution of plasmid DNA was performed by adding 50 µl of buffer EB (elution buffer; 10 mM Tris-Cl, pH 8.5), and the tube was sat at room temperature for 1 min, followed by centrifugation at 13,000 rpm for 1 min. Plasmid DNA solution was transferred to labeled collection tubes. Five microliters of plasmid DNA was analyzed on 1.2% agarose gel staining with EtBr. After the electrophoresis, the DNA bands were visualized by using UV transilluminator. Quantitation of plasmid DNA was performed using UV/Visible spectrophotometer (See 2.7. total RNA extraction).

2.5. Sequencing

Purified plasmids of CAR, PXR, LXRα, and RORγ were subjected to the sequencing by a primer walking sequencing method using the ABI PRISM BigDye Terminator v3.1 Cycle Sequencing Ready Reaction Kit and ABI PRISM 310 (Applied Biosystems, Japan). The purification of PCR products for sequencing reactions was performed using ExoSAP-IT (USB Corporation, USA) to remove unincorporated dNTPs and primers. Twenty microliters of PCR product was mixed with 3 µl ExoSAP-IT mixture, containing exonuclease I and shrimp alkaline phosphatase, in a 0.5 ml PCR tube. The reaction was carried out in a Techne thermal cycler with the following conditions; 37 ℃ for 30 min (incubation), followed by 80 ℃ for 15 min (inactivation of ExoSAP-IT). The purified PCR product was then used in the sequencing reactions. The sequencing reaction (Terminator sequencing kit; ABI Big Dye Ver.1) was composed of 1.8 µl of BigDye, 6.2 µl of 2.5X sequencing buffer, 1.6 µl of primer, 2 µl of purified template DNA, and 8.4 µl of sterilized distilled water.

The reaction conditions were as follows; 25 cycles of 96 ℃ for 30 sec, 50 ℃ for 15 sec, and 60 ℃ for 4 min, followed by a final hold at 4 ℃.

To remove excess dye-labeled dideoxynucleotides, the purification of sequencing reactions was performed using the AutoSeq G-50 (Amersham Biosciences, UK). One hundred microliters of 10.5 mM EDTA (pH 8.0) was added to the AutoSeq G-50 column. To resuspend the resin, the AutoSeq G-50 spin columns were spun for 1 min at 2000 x g. The spin column was then placed in a new 1.5 ml microcentrifuge tube, followed by adding 20 μl of sample to the center of the angled surface of the compacted resin bed. The spin columns were centrifuged at 2000 x g for 1 min. The purified sample was collected in the bottom of the 1.5 ml microcentrifuge tube and was used in the automated sequencing at the laboratory of plant virology, faculty of agriculture, Saga University.

2.6. Sample Collection for Expression Analysis

Tissues from swine were used as starting materials for total ribonucleic acid (RNA) extraction. A six months old pig, which belongs to the first generation (F1) of the crosses between a male Landrace and a female Large white (a mature Landrace x Large White female crossbred pig), was used in this study. Total RNA extraction from pig tissues such as liver, pancreas, kidney, heart, small intestine, and muscle was performed (See 2.7. total RNA extraction). Total RNA samples from the thymus, spleen, and brain (a one month-old Landrace pig) were purchased from the UNITECH (Japan).

2.7. Total RNA Extraction

Total RNA was extracted from pig's tissues (e.g., liver, pancreas, kidney, heart, small intestine and muscle) using the QIAGEN RNeasy Mini and Fibrous Tissue Mini Kits (QIAGEN, Japan).

Total RNA was extracted from liver, pancreas, kidney, and small intestine using RNeasy Mini Kit as described by the manufacturer. Frozen animal tissue was removed from deep freezer storage. Thirty milligrams (mg) of each tissue was used as starting material. After removal of sample, it was immediately placed in liquid nitrogen and ground thoroughly with a mortar and pestle under liquid nitrogen. Properly ground tissue was decanted into an RNase-free 1.5 ml microcentrifuge tube, which contained 600 μl of guanidine isothiocyanate (GITC) containing lysis buffer (buffer RLT) and β-mercaptoethanol (β-ME). Beta-mercaptoethanol was added to RLT buffer in order to ensure the immediate inactivation of RNases in tissue samples. Ten microliters of beta-mercaptoethanol was added to 1 ml buffer RLT. Tissue lysate in 1.5 ml microcentrifuge tube was carefully mixed using stainless spatulas. Centrifugation of tissue lysate was performed at 20-25 ℃ in a refrigerated laboratory microcentrifuge for 3 min at 15,000 rpm. Supernatant was carefully transferred to a new 1.5 ml microcentrifuge by pipetting. Six hundred microliters of 70% ethanol was added to the cleared lysate and immediately mixed by pipetting and inverting. Sample was then applied to an RNeasy mini spin column placed in a 2 ml collection tube, which supplied by manufacturer, and centrifuged for 15 seconds (s) at 10,000 rpm, followed by discarding the flow-through. In order to get high quality total RNA, DNase treatment was performed. Three hundred and fifty microliters of buffer RW 1 (wash buffer) was added to the RNeasy spin column and centrifuged at 10,000 rpm for 15 s, followed by discarding the flow-through. Eighty microliters of DNase I incubation

mix was directly added to RNeasy silica-gel membrane and incubated on the benchtop at room temperature (20-30 ℃) for 15 min. DNase I incubation mix was prepared by adding 10 μl DNase I stock solution to 70 μl buffer RDD. After 15 min of incubation, another 350 μl buffer RW1 was added to the RNeasy mini column and centrifuged at 10,000 rpm for 15 s, followed by discarding both the flow-through and collection tube. The RNeasy spin column was then placed in a new 2 ml collection tube. Five hundred microliters of buffer RPE (wash buffer) was added to the RNeasy spin column and centrifuged for 15 s at 10,000 rpm, followed by discarding the flow-through. Buffer RPE was supplied as 11 ml concentrate stock solution. To obtain 55 ml working solution of buffer RPE, 44 μl of 99.5% ethanol was added. Another 500 μl RPE buffer was added to the RNeasy spin column and centrifuged at 10,000 rpm for 2 min, followed by discarding the collection tube with the flow-through. RNeasy spin column was transferred to a new 2 ml collection tube. To eliminate ethanol that may be carried over in the dilution step, the centrifugation of spin column at 10,000 rpm for 1 min was performed, followed by discarding the collection tube with the flow-through. The RNeasy mini spin column was then placed on a new 1.5 ml collection tube. The elution of RNA was performed by adding 30 μl of RNase-free water to the RNeasy silica-gel membrane, followed by incubation for 1 min and centrifugation at 10,000 rpm for 1 min. Another 20 μl RNase-free water was directly added to the RNeasy silica-gel membrane, incubated 1 min, and centrifuged at 10,000 rpm for 1 min. Finally, 50 μl of total RNA solution was transferred into a new and properly labeled 1.5 ml collection tube. Sample was stored at -85 ℃ deep freezer. Quantitation of total RNA was performed using UV/Visible spectrophotometer (Ultrospec 3000; Biochrom Ltd., UK). The cuvette was washed by using distilled water and then added 400 μl of distilled water, and run on spectrophotometer as a

blank control. Five microliters total RNA solution was mixed with 395 μl distilled water in a new 1.5 ml microcentrifuge tube. The solution was transferred to a cuvette and run on spectrophotometer. The concentration of total RNA was calculated and data were analyzed before using in downstream reactions.

In the case of fiber-rich tissues such as muscle and heart, the RNeasy Fibrous Tissue Mini Kit was used. The procedures were the same as RNeasy Mini Kit, except for the lysis and homogenization of tissue samples.

The RNeasy Fibrous Tissue Mini Kit was used to extract total RNA from heart and muscle of pig. Tissues were disrupted and ground under liquid nitrogen. Ground tissue was decanted to 1.5 ml microcentrifuge tube, which contains 300 μl of lysis buffer RLT. Tissue lysate in 1.5 ml microcentrifuge tube was carefully mixed using stainless spatulas. Five hundred and ninety microliters of RNase-free water was added to the homogenate, and 10 μl of proteinase K solution was added and mixed thoroughly by pipetting. The treatment of sample with proteinase K is for the digestion of contractile proteins, connective tissue, and collagen proteins. Sample was then heated in 55 ℃ water bath for 10 min. Centrifugation of tissue lysate was performed at 20-25 ℃ in a refrigerated microcentrifuge for 3 min at 10,000 x g. Supernatant was carefully transferred to a new 1.5 ml microcentrifuge by pipetting. Four hundred and fifty microliters of 99.5% ethanol was added to the cleared lysate and immediately mixed by pipetting and inverting. Sample was applied to an RNeasy mini spin column placed in a 2 ml collection tube and centrifuged for 15 s at 10,000 rpm, followed by discarding the flow-through. DNase treatment was also performed and procedures were the same as RNeasy mini kit.

2.8. Expression Analysis

The mRNA expression pattern of the porcine *CAR* and *PXR* genes was determined using Ready To Go RT-PCR beads (Amersham Biosciences, Japan) in a 0.2 ml PCR tube containing 2 μl of 10 mM porcine *CAR* or *PXR* gene-specific primers in a 50 μl-reaction mixture. One-tube reverse transcription (RT)-PCR for *β-actin* gene was also carried out as a positive control. The reaction mixture was the same as those for the porcine *CAR* and *PXR* genes. The reaction cycles were as follows: reverse transcription at 42 ℃ for 30 min, followed by PCR amplification, which involves an initiation denaturation step at 95 ℃ for 5 min, 40 cycles of denaturation at 95 ℃ for 30 s, annealing at 60.5 ℃ (porcine *CAR*), 58 ℃ (porcine *PXR*) or 61.5 ℃ (*β-actin*) for 30 s, elongation at 72 ℃ for 1 min, final elongation at 72 ℃ for 10 min, and a final hold at 4 ℃. Primer sets used in this study were as follows: porcine *CAR* forward primer, 5'-GAA CAG TCA ACA AAA GCA CCA G-3' and porcine *CAR* reverse primer, 5'-CCC AGG AGT ATC TGG ACT AAC G-3'; porcine *PXR* forward primer, 5'-CCG CTA CTT CAG GGC TAC TG-3' and porcine *PXR* reverse primer, 5'-CCC GAT CTG TTC TCG TTT CT-3'; and *β-actin* forward primer 5'-GGA CCT GAC CGA CTA CCT CA-3' and *β-actin* reverse primer 5'-GAG GTC CTT CCT GAT GTC CA-3'. These primer sets were designed from cDNA regions which are expected to include the intron using Primer 3 software (http://frodo.wi.mit.edu/cgi-bin/primer3/primer3_www.cgi) (Rozen and Skaletsky, 2000).

The mRNA expression pattern of the porcine *LXRα-1* and *-2* transcripts was determined using the two step RT-PCR method. The first-strand cDNA was transcribed from the total RNA sample of each porcine tissue using the Avian Myeloblastosis Virus (AMV) reverse transcriptase and the oligo dT-3 sites adaptor

primer from the 3'-full RACE core set (TAKARA, Japan). Single-strand cDNA was used as a template for the PCR amplification reaction; the 25 μl PCR reaction mixture contained 0.025U/μl of TAKARA *Ex Taq* Hot Start version DNA polymerase, 1× *Ex Taq* buffer (Mg^{2+} plus), 200 μM of each dNTP (TAKARA, Japan), and 0.8 μM of porcine *LXRα-1* and *-2* transcript-specific primers. The PCR amplification of *β-actin* genes was also carried out as a positive control. The reaction mixture used was the same as that used for the porcine *LXRα* gene. The PCR amplification cycles were as follows: an initial denaturation step at 95 ℃ for 3 min, followed by 40 cycles of denaturation at 95 ℃ for 30 s, annealing at 60.5 ℃ (porcine *LXRα-1* transcript), 63.5 ℃ (porcine *LXRα-2* transcript), or 61.5 ℃ (*β-actin*) for 30 s, elongation at 72 ℃ for 1 min, final elongation at 72 ℃ for 10 min, and a final hold at 4 ℃. The primer sets used in this study were as follows: porcine *LXRα-1* primers (forward) 5'-GGA CAA GGG ACT GCA CCA T-3' and (reverse) 5'-GCT CAG CAC GTT GTA GTG GA-3'; and porcine *LXRα-2* primers (forward) 5'-CAC ATG GCC TGG TCA CCT A-3' and (reverse) 5'-GCT CAG CAC GTT GTA GTG GA-3'. The primers for *β-actin* were the same as porcine *CAR* and *PXR* genes. These primer sets were designed from cDNA regions, which are expected to specific for each transcript and include the intron, using Primer 3 software.

The mRNA expression analysis of the porcine *RORγ-1* and *-2* transcripts was also performed using the same procedures as described in the expression analysis of porcine *LXRα* transcripts. The PCR amplification cycles were as follows: an initial denaturation step at 95 ℃ for 3 min, followed by 40 cycles of denaturation at 95 ℃ for 30 s, annealing at 62 ℃ (both porcine *RORγ* transcripts), or 61.5 ℃ (*β-actin*) for 30 s, elongation at 72 ℃ for 3 min, final elongation at 72 ℃ for 10 min, and a final

hold at 4 ℃. The primer sets used in this study were as follows: porcine *RORγ-1* primers (forward) 5'-GTC CAA GAA GCA AAG GGA CA-3' and (reverse) 5'-GAT CCT CCT GCT GTC CTG AG-3'; and porcine *RORγ-2* primers (forward) 5'-CAT GTC CCG AGA TGC TGA G-3' and (reverse) 5'-AGC TTC TGA GCC CAG AGT TG-3'. The primers for *β-actin* were the same as porcine *CAR* and *PXR* genes. Primers were designed using web interface of Primer 3. The PCR products were then analyzed in a 1.2% agarose gel stained with ethidium bromide and visualized using a UV transilluminator.

2.9. Agarose Gel Electrophoresis

The PCR products were analyzed using agarose gel electrophoresis stained with ethidium bromide and visualized by using a UV transilluminator. Concentrations of agarose gel were used according to expected DNA bands. Mostly, 1.2% agarose gel was used. The 1.2 grams (g) of agarose gel S (NIPPON GENE, Japan) was weighted and dissolved in 200 ml Erlenmeyer flask, which contained 100 ml of 0.5X Tris-Borate-EDTA (TBE) electrophoresis buffer. One hundred of 0.5X TBE electrophoresis buffer was diluted from 5X TBE stock solution by mixing 10 ml of 5X TBE with 80 ml of distilled water and adding distilled water to make up 100 ml final volume of 0.5X TBE electrophoresis buffer. To visualize DNA in agarose gel, it was stained by the fluorescent dye ethidium bromide (EtBr). One microliter of 10 mg/ml EtBr was then added to solution. Agarose gel solution was melted in a microwave oven until bubbles were appeared. The flask was removed from the microwave oven and gently swirled the flask to resuspend any settled powder and gel pieces. The flask was then heated again in the microwave oven until the agarose was completely

dissolved. After gel melting, the flask was sat at room temperature to let the solution cooled to 50-60 ℃. The gel casting platform and gel comb were set in order to make the sample wells. Warm agarose gel solution was then poured into a gel casting apparatus and allowed to harden at room temperature (20-30 min). If the air bubbles were formed, it was removed by using a sterilized micropipette tip. After gel hardening, the gel comb was carefully removed, and the gel was placed in a horizontal electrophoresis tank. Four hundred milliliters of 0.5X TBE buffer was then poured directly on to the top of the harden gel. Five microliters of DNA size marker was loaded to a well. Two microliters of 6X gel-loading buffer was mixed with 5 μl of sample on the paraffin paper, and the mixture was slowly and carefully loaded to a gel well by using a micropipette. The electrophoresis was run for 40-50 min at 100 volts DC power supply. The gel was then carefully transferred to a clean plastic tray and visualized and photographed by using an ultraviolet benchtop transluminator (FAS-III-mini FAS-301; TOYOBO, Japan).

2.10. Sequence Analysis

2.10.1. Analysis of Nucleotide Sequences

The open reading frames (ORF) of nucleotide sequences were deduced using the ORF Finder analysis tool (http://www.ncbi.nlm.nih.gov/gorf/gorf.html), and standard genetic codes were used for this analysis. The secondary structure of 5'-and 3'-untranslated regions (UTRs) was predicted using web interface of the RNAfold program at the Vienna RNA secondary structure server (Hofacker, 2003; http://rna.tbi.univie.ac.at/cgi-bin/RNAfold.cgi). The identification and characterization of functional elements in the non-coding UTR were performed using the UTRdb

(Mignone *et al.,* 2005; http://www2.ba.itb.cnr.it/UTRSite/). The codon usage of porcine nuclear receptor ORFs was analyzed by using the sequence manipulation suite (SMS: http://bioinformatics.org/sms2/) locally runs on personal computer (Stothard, 2000). The Kozak consensus sequences were drawn using WebLogo service at http://weblogo.berkeley.edu/logo.cgi (Crooks *et al.,* 2004). General bioinformatics analysis of nucleotide sequences was performed using BioEdit locally runs on PC (http://www.mbio.ncsu.edu/BioEdit/bioedit.html).

The nucleotide sequences used in the analysis were porcine *CAR* (GenBank accession no. AB214979), chicken *CXR* (NM_204702), Japanese quail *CAR* (AB104462), mouse *CAR* (NM_009803), human *CAR* (NM_005122), bovine *CAR* (NM_001079768), porcine *PXR* (AB214980), mouse *PXR* (NM_010936), human *PXR* (NM_003889), bovine *PXR* (BC133405), porcine *LXRα* (AB254405 and AB254406), chicken *LXRα* (NM_204542), bovine *LXRα* (NM_001014861), human *LXRα* (NM_005693), human *RORγ* (NM_005060 and NM_001001523), bovine *RORγ* (NM_001083451), and mouse *RORγ* (NM_011281 and AF163668).

2.10.2. Analysis of Amino Acid Sequences

The structural domains of the predicted proteins were predicted by using the simple modular architecture research tool (SMART; http://smart.embl-heidelberg.de/) (Schultz *et al.,* 1998; Letunic *et al.,* 2004). Analysis of protein properties and domains was performed using ScanProsite at http://br.expasy.org/prosite/ and ProtParam tool at http://br.expasy.org/tools/protparam.html. Basic amino acid sequence analysis was performed using BioEdit. The protein three-dimensional (3-D) model was predicted using the service of the Swiss Model (http://swissmodel.expasy.org/). The predicted 3-D models of porcine PXR (UniProtKB/TrEMBL accession no. Q2V0W2), CAR

(Q2V0W3), and LXRα (A4UVT2) were retrieved from the Swiss Model Repository (http://swissmodel.expasy.org/repository/). In the case of porcine RORγ, the Swiss Model Automatic Modeling Mode (First Approach mode) was used instead (Schwede *et al.,* 2003). The protein 3-D models were displayed using the UCSF Chimera (Pettersen *et al.,* 2004).

The protein sequences used in the analysis were porcine CAR (GenBank accession no. BAE54304), chicken CXR (NP_990033), Japanese quail CAR (BAF57043), mouse CAR (NP_033933), human CAR (NP_005113), bovine CAR (NP_001073236), porcine PXR (BAE54305), mouse PXR (NP_035066), human PXR (NP_003880), bovine PXR (AAI33406), porcine LXRα (BAF57044 and BAF57045), chicken LXRα (NP_989873), bovine LXRα (NP_001014861), human LXRα (NP_005684), human RORγ (NP_005051 and NP_001001523), bovine RORγ (NP_001076920), and mouse RORγ (NP_035411 and AAD46913).

2.10.3. Prediction of Exon-Intron Structure

The porcine genomic DNA sequences were identified using BLAST pig sequences(http://www.ncbi.nlm.nih.gov/projects/genome/seq/BlastGen/BlastGen.cgi?pid=10718). In the case of pig genomic DNA was not available, bovine DNA was used instead. The bovine DNA was identified using BLAST cow sequences (http://www.ncbi.nlm.nih.gov/projects/genome/seq/BlastGen/BlastGen.cgi?pid=10708). The Sim 4 alignment program (Florea *et al.,* 1998; http://pbil.univ-lyon1.fr/sim4.php) was used in the alignment of porcine mRNA with genomic DNA.

CHAPTER 3

THE NUCLEAR RECEPTORS CAR AND PXR

3.1. Introduction

The constitutive androstane receptor (*CAR*; NR1I3) and the pregnane X receptor (*PXR*; NR1I2) are members of nuclear receptor superfamily that have been known to control the expression of the cytochrome P450 gene family, which encodes enzymes involved in the metabolism of both endogenous and exogenous substances (Honkakoski and Negishi, 2000; Wei *et al.,* 2000; Kliewer *et al.,* 2002). Studies on the CAR and PXR in humans and mice revealed that they are related nuclear receptors that can be activated by the same ligand and can regulate the same gene family (Moore *et al.,* 2000; Xie *et al.,* 2000; Maglich *et al.,* 2002; Moore *et al.,* 2002). Both CAR and PXR proteins have been shown to be activated by endogenous hormones such as androstanol (5α-androstan-3α-ol) and androstenol [$16(5\alpha)$-androsten-3α-ol] (Forman *et al.,* 1998; Moore *et al.,* 2000), the progesterone metabolite 5β-pregnane-3,20-dione (Moore *et al.,* 2000), and 17β-estradiol (Xue *et al.,* 2007), and exogenous substances such as drugs (Kliewer *et al.,* 2002; Moore *et al.,* 2003; Timsit and Negishi, 2007). In addition, certain reports showed that endocrine disrupting chemicals (EDCs) are potential ligands for the CAR and PXR proteins such as pesticides (methoxychlor, endosulfan, dieldrin, DDT), phthalic acid, and the plasticizer nonylphenol (Masuyama *et al.,* 2000; Mikamo *et al.,* 2003; Wyde *et al.,* 2003; Kretschmer and

Baldwin, 2005; Lemaire *et al.,* 2006; Hernandez *et al.,* 2007). Moreover, the plant-derived phytoestrogens (flavonoids and lignans) and the fungi-biosynthesized mycoestrogen zearalenone are also the activators of PXR receptor proteins (Jacobs *et al.,* 2005; Ding *et al.,* 2006; Mnif *et al.,* 2007). Activated CAR and PXR proteins form a heterodimer with the 9-*cis* retinoic acid receptor (RXR; NR2B) and bind to a specific DNA region of target genes referred to as nuclear receptor response element (NRRE) (Mangelsdorf and Evans, 1995). Several CAR and PXR target genes, such as *CYP3A, UGT1A1,* and *CYP2B* genes, have been identified (Bertilsson *et al.,* 1998; Kliewer *et al.,* 1998; Lehmann *et al.,* 1998; Kawamoto *et al.,* 1999; Xie *et al.,* 2003). *PXR* and *CAR* genes are primarily expressed in the liver and intestine (Baes *et al.,* 1994; Choi *et al.,* 1997; Bertilsson *et al.,* 1998; Blumberg *et al.,* 1998; Kliewer *et al.,* 1998; Lehmann *et al.,* 1998).

The structure as well as the function of the *CAR* and *PXR* genes has been extensively studied in humans and mice. In contrast, in agricultural animals, these important genes have been reported only in chickens (Handschin *et al.,* 2000) and the Japanese quail (Yamada *et al.,* 2006). In this chapter, we report the identification and sequencing cDNAs encoding porcine *CAR* and *PXR* genes.

3.2. Nucleotide Sequences of Porcine *CAR* and *PXR*

3.2.1. Nucleotide Sequence of Porcine *CAR*

The full-length cDNA clone encoding porcine *CAR* was identified and sequenced from a PEDE (Pig EST Data Explorer) clone using the primer walking sequencing method. Swine *CAR* cDNA has 1,407 bp mRNA (52.6% G+C content), which contains 216 bp of 5'-untranslated region (5'-UTR) (nucleotide positions 1-

216), 144 bp of 3'-UTR (nucleotide positions 1264-1407), and 1047 bp (nucleotide positions 217-1263) of coding region encoding 348 amino acid residues (Figure 3.1). The full-length cDNA sequence of porcine *CAR* has been submitted to the DDBJ/EMBL/GenBank databases under the accession number AB214979 (Thadtha *et al.,* 2005). Comparison of nucleotide sequences revealed that full-length porcine *CAR* shares 88% identity with bovine *CAR*, 84% identity with human *CAR*, 34% identity with Japanese quail *CAR*, and 30% identity with chicken *CXR* (Table 3.1). The nucleotide composition of porcine *CAR* is also shown in Table 3.2.

```
  1 CAGCTGAAGGCCACAGAGGTAGAAGTTCCTTGTTTTCCAGGTACTAAAGGCTATAATCCC
 61 TAACTTTAAATACTGGAAGCTCCTGAGATCAAAGGGAAAGCAGGGTTACAGTGGGAGTAC
121 GTGGAGAGGAATCCAGCACCAGATTCTGTGGCCTGCAGGTGACGTGCTGCCTAAGAGGAA
181 CAGGGGCCTGTGACCACCATCCCAACACCTGAAGCCATGGCCAGCGGGGAAGATGAGCCA
                                            M  A  S  G  E  D  E  P    8
241 AGGAACTGTGCTGTGTGCGGGGACCGAGCCACAGGCTATCACTTCCATGCCTTGACTTGT
     R  N  C  A  V  C  G  D  R  A  T  G  Y  H  F  H  A  L  T  C    28
301 GAGGGCTGCAAGGGTTTCTTCAGGCGAACAGTCAACAAAAGCACCAGTCTCATCTGCCCC
     E  G  C  K  G  F  F  R  R  T  V  N  K  S  T  S  L  I  C  P    48
361 TTTGCTGGAAGCTGTAAGGTCAATAAGGCCCAGAGGCGCCACTGCCCAGCCTGCAGGTTG
     F  A  G  S  C  K  V  N  K  A  Q  R  R  H  C  P  A  C  R  L    68
421 CAGAAGTGCCTAGATGCTGGCATGAAGAAAGACATGATCCTATCAGCAGAAGTCCTGGCA
     Q  K  C  L  D  A  G  M  K  K  D  M  I  L  S  A  E  V  L  A    88
481 TTGCGGCGAGCAAGACAGGTTCAGCGCCGGGCACAGCAAGCATCACTGCAGCTGAGTAAG
     L  R  R  A  R  Q  V  Q  R  R  A  Q  Q  A  S  L  Q  L  S  K    108
541 GAGCAGAAAGCGTTAGTCCAGATACTCCTGGGGGGCCCATACCCGCCATATGGGCACTATG
     E  Q  K  A  L  V  Q  I  L  L  G  A  H  T  R  H  M  G  T  M    128
601 TTTGATCAGTTTGTGCAGTTCAGGCCTCCAGCTCATCTGTTCATCCATCACCAGCACTTG
     F  D  Q  F  V  Q  F  R  P  P  A  H  L  F  I  H  H  Q  H  L    148
661 CCACCCCTGGTGCCTGAACTGTCTCTGCTCATGCATTTCGCGGACATCAACACTTTCATG
     P  P  L  V  P  E  L  S  L  L  M  H  F  A  D  I  N  T  F  M    168
721 ATACAGCAAATTATCAAGTTCACCAAGGATCTGCCCCTCTTCCGGTCCCTGCCCATGGAG
     I  Q  Q  I  I  K  F  T  K  D  L  P  L  F  R  S  L  P  M  E    188
781 GACCAGATCTCCCTTCTCAAGGGAGCAGCTGTAGAAATCTGTCAGATCGTACTCAATACC
     D  Q  I  S  L  L  K  G  A  A  V  E  I  C  Q  I  V  L  N  T    208
841 ACTTTCTGTCTGCAAACACAAAAATTCCTCTGTGGGCCTCTTCGCTACACAATAGAAGAT
     T  F  C  L  Q  T  Q  K  F  L  C  G  P  L  R  Y  T  I  E  D    228
901 GGAGCGCATGTGGGGTTCCAGGAAGAGTTTTTGGAGTTGCTCTTTGGCTTCCATAAGACA
     G  A  H  V  G  F  Q  E  E  F  L  E  L  L  F  G  F  H  K  T    248
961 CTTCGGCGACTGCAGCTCCAGGAGCCTGAGTATGTGCTCATGGTTGCTGTGGCCCTCTTC
     L  R  R  L  Q  L  Q  E  P  E  Y  V  L  M  V  A  V  A  L  F    268
```

24

```
1021 TCTCCTGACCGGCCTGGGGTAACCCAGAGGAAGGAGATTGATCAGTTGCAAGAGGAGATG
      S   P   D   R   P   G   V   T   Q   R   K   E   I   D   Q   L   Q   E   E   M   288
1081 GCACTGACTCTGCAGAGCTACATCAAAGGGCAGCAGCCAAGTCTCCGGGACAGGTTTCTC
      A   L   T   L   Q   S   Y   I   K   G   Q   Q   P   S   L   R   D   R   F   L   308
1141 TATGCAAAGCTGCTGGGCCTATTGGCTGAGCTCCGAAGCATTAACAAAGAATACTGGTAC
      Y   A   K   L   L   G   L   L   A   E   L   R   S   I   N   K   E   Y   W   Y   328
1201 CAAATCCAGAACATCCAGGGACTGTCCACCATGATGCCGCTGCTCCAGGAGATCTGCAGC
      Q   I   Q   N   I   Q   G   L   S   T   M   M   P   L   L   Q   E   I   C   S   348
1261 TGAGGCCCAGTTTGCCTCCTTCCCCAGCCCACCTGCACATGATGGACTGTGTACGGCATC

1321 CAGGGGAAAGTGCTGGGAGCCGTCAAAAGGGAGCCCACTGGTTGCAATGAAAGATTAAAG
1381 CAATAACTGCCAAAAAAAAAAAAAAAA   1407
```

Figure 3.1. Nucleotide and deduced amino acid sequences of porcine *CAR*.

The numbers on the left and right indicate the nucleotide base and the deduced amino acid in porcine *CAR*, respectively. The ORF of porcine *CAR* was deduced using the ORF Finder. The deduced amino acid sequence is shown in single-letter code below the nucleotide sequence. The ATG start codon (positions 217-219) and the TGA stop codon (positions 1261-1263) are printed in bold. The predicted DNA-binding domain (amino acid residues 8-79) is boxed. The putative ligand-binding domain of porcine CAR is underlined. The polyadenylation signal (ATTAAA) is underlined.

Table 3.1. Comparison of nucleotide sequences of *CAR* cDNAs in farm animals and human.

Gene	mRNA			% identity	Length
	5'-UTR	ORF	3'-UTR		
Porcine *CAR*	216	1047	144		1407
Bovine *CAR*	82	1047	163	88	1292
Quail *CAR*	130	1158	77	34	1365
Chicken *CXR*	47	1176	117	30	1340
Human *CAR*	203	1047	131	84	1381

Note: Nucleotide sequence identities were determined by performing comparisons using the pairwise alignment in BioEdit with default parameters. The percent identity of each *CAR* mRNA sequence was compared to porcine *CAR*.

Table 3.2. Nucleotide composition of porcine *CAR* and *PXR* cDNA sequences.

Sequence	Purine base			Pyrimidine base			Total (bp)
	C	G	C+G	A	T	A+T	
Full-length *CAR* cDNA	371	369	52.6%	366	301	47.4%	1407
5'-UTR	51	60	51.4%	62	43	48.6%	216
CDS	284	274	53.3%	255	234	46.7%	1047
3'-UTR	36	35	49.3%	49	24	50.7%	144
Full-length *PXR* cDNA	683	657	52.2%	683	544	47.8%	2567
5'-UTR	69	68	50.2%	88	48	49.8%	273
CDS	348	344	54.7%	320	254	45.3%	1266
3'-UTR	266	245	49.7%	275	242	50.3%	1028

Note: The nucleotide composition of porcine *CAR* and *PXR* cDNAs was derived from BioEdit.

A. 5'-Untranslated Region

The 5'-UTR of porcine *CAR* is 216 bases in length (nucleotide positions 1-216) and contains 51.4% G+C content (Table 3.2). The 5'-UTRs in farm animals and human vary from 47 bp in chicken *CXR* to 216 bp in porcine *CAR* with average size of 136 bp (Table 3.1).

It has been known that sequence features in 5'-UTR, such as secondary structure or stem-loops, upstream open reading frame (uORF), upstream AUG, involved in post-transcriptional regulation, which control mRNA localization, stability, and translational efficiency (Pesole *et al.,* 2001; Hughes, 2006). Using analysis tools at the UTRdb (Mignone *et al.,* 2005), the 5'-UTR of porcine *CAR* (Accession no. BB408133) did not possess the known regulatory elements involved in post-transcriptional regulation such as mRNA localization, stability, and translational efficiency.

The secondary structure of the 5'-UTR of porcine *CAR* transcript was predicted using web interface of the RNAfold program from the Vienna RNA secondary structure server (Hofacker, 2003). The predicted fold energy in the 5'-UTR is -72.3

kcal/mol. The predicted secondary structure of 5'-UTR of porcine *CAR* is shown in Figure 3.2.
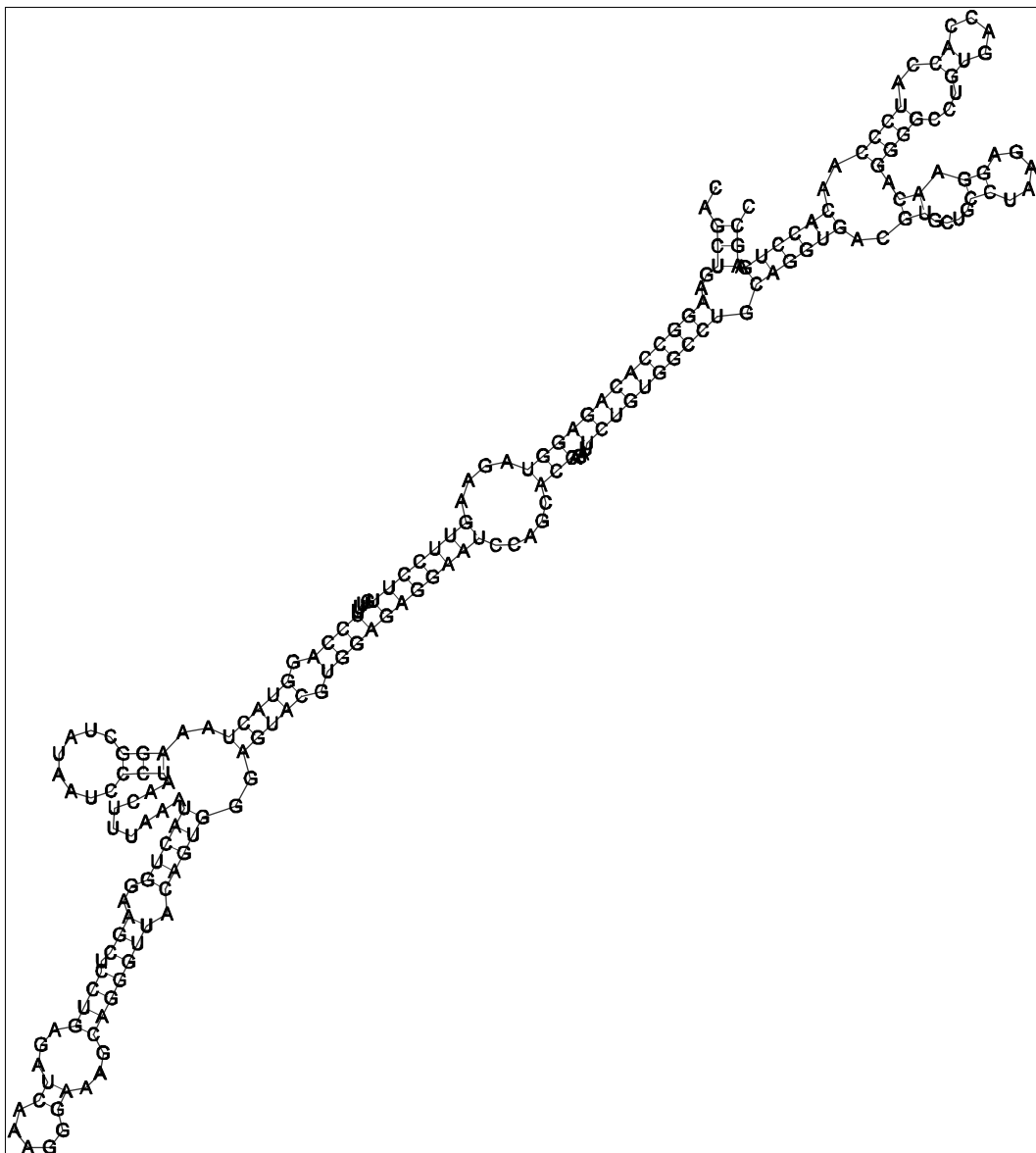
Figure 3.2. Predicted secondary structure of 5'-UTR of porcine *CAR*.

The secondary structure of 5'-UTR was predicted using the RNAfold program. The predicted fold energy of porcine *CAR* 5'-UTR is -72.3 kcal/mol.

B. Open Reading Frame

The open reading frame of full-length porcine *CAR* cDNA was deduced using the ORF Finder, and the longest ORF located in frame +1 was selected (nucleotide

positions 217-1263). The ATG start codon encoding a methionine residue and the TGA stop are located at nucleotide positions 217 and 1261, respectively (Figure 3.1). Sequence upstream of the putative start ATG codon in porcine CAR cDNA did not contain any additional in-frame start codons (upstream AUG). Eight in-frame stop codons (five TAA and three TGA stop codons) were found upstream of the predicted ATG start codon. The putative ORF contains 1047 bp and encodes 348 amino acids with a predicted molecular weight of 39809 Daltons and an isoelectric point of 8.94.

Table 3.3. shows the codon usage in the ORF of porcine *CAR* cDNA. It has 53.3% of G+C content. The GC contents in the first, second, and third codon positions of porcine *CAR* cDNA are 58.1%, 37% and 64.8%, respectively. The codons GAG (glutamic acid), UUC (phenylalanine), AUC (isoleucine), AAG (lysine), CUG and CUC (leucine), and CAG (glutamine) were frequently used in porcine *CAR* ORF.

The nucleotide sequence around putative ATG initiation codon of porcine *CAR* mRNA is according with Kozak consensus sequence (GCC<u>A/G</u>CC*ATG*<u>G</u>) (Kozak, 1996). The nucleotide flanking ATG in porcine *CAR* cDNA is GAA<u>G</u>CC*ATG*<u>G</u>. Porcine *CAR* has both purine Gs in positions -3 and +4, respectively. This classified sequence flanking porcine *CAR*'s ATG initiation codon as a strong Kozak consensus sequence. The consensus sequence flanking ATG codons in *CAR* is shown in Figure 3.3.

Table 3.3. Codon usage in the ORF of porcine *CAR* and *PXR* genes.

Amino acid	Codon	*CAR*		*PXR*	
		Number	Frequency	Number	Frequency
Alanine	GCG	3	8.60	0	0
	GCA	8	22.92	4	9.48
	GCU	7	20.06	5	11.85
	GCC	7	20.06	16	37.91
Cysteine	UGU	6	17.19	5	11.85
	UGC	7	20.06	10	23.70
Aspartic acid	GAU	6	17.19	5	11.85
	GAC	6	17.19	11	26.07
Glutamic acid	GAG	13	37.25	20	47.39
	GAA	7	20.06	13	30.81
Phenylalanine	UUU	6	17.19	7	16.59
	UUC	14	40.11	16	37.91
Glycine	GGG	7	20.06	6	14.22
	GGA	4	11.46	3	7.11
	GGU	1	2.87	7	16.59
	GGC	6	17.19	5	11.85
Histidine	CAU	8	22.92	3	7.11
	CAC	4	11.46	5	11.85
Isoleucine	AUA	3	8.60	2	4.74
	AUU	3	8.60	4	9.48
	AUC	12	34.38	17	40.28
Lysine	AAG	12	34.38	17	40.28
	AAA	6	17.19	9	21.33
Leucine	UUG	8	22.92	3	7.11
	UUA	1	2.87	2	4.74
	CUG	18	51.58	24	56.87
	CUA	3	8.60	1	2.37
	CUU	3	8.60	3	7.11
	CUC	15	42.98	10	23.70
Methionine	AUG	12	34.38	16	37.91
Asparagine	AAU	2	5.73	6	14.22
	AAC	5	14.33	8	18.96
Proline	CCG	1	2.87	2	4.74
	CCA	5	14.33	6	14.22
	CCU	6	17.19	1	2.37
	CCC	4	11.46	10	23.70
Glutamine	CAG	25	71.63	23	54.50
	CAA	6	17.19	4	9.48

Arginine	AGG	7	20.06	11	26.07
	AGA	1	2.87	1	2.37
	CGG	6	17.19	12	28.44
	CGA	5	14.33	3	7.11
	CGU	0	0	1	2.37
	CGC	4	11.46	6	14.22
Serine	AGU	3	8.60	3	7.11
	AGC	6	17.19	9	21.33
Serine	UCG	0	0	1	2.37
	UCA	2	5.73	5	11.85
	UCU	2	5.73	4	9.48
	UCC	3	8.60	6	14.22
Threonine	ACG	0	0	2	4.74
	ACA	5	14.33	5	11.85
	ACU	5	14.33	6	14.22
	ACC	6	17.19	12	28.44
Valine	GUG	6	17.19	9	21.33
	GUA	3	8.60	1	2.37
	GUU	2	5.73	2	4.74
	GUC	4	11.46	2	4.74
Tryptophan	UGG	1	2.87	3	7.11
Tyrosine	UAU	3	8.60	2	4.74
	UAC	4	11.46	6	14.22
Stop codon	UGA	1	2.87	1	2.37

Note: The frequencies of codon usage in porcine *CAR* and *PXR* ORFs were derived from SMS. Porcine *CAR* and *PXR* have 53.3% and 54.7% of GC content in its ORF, respectively. The GC contents in the first, second, and third codon positions of porcine *CAR* and *PXR* are 58.1% and 52.8%, 37% and 40.5%, and 64.8% and 70.6%, respectively.

Figure 3.3. Kozak consensus sequence around ATG start codon of *CAR* in farm animals and human.

The nucleotide flanking ATG in porcine *CAR* is GAA<u>G</u>CC*ATG*<u>G</u>. Porcine *CAR* has both purine Gs in positions -3 and +4, respectively. This classified sequence flanking porcine *CAR*'s ATG initiation codon as a strong Kozak consensus sequence. The WebLogo was used in the create diagram representing the Kozak consensus sequence.

C. 3'-Untranslated Region

The 3'-UTR of porcine *CAR* has 144 bases in length (nucleotide positions 1264-1407) with 49.3% of G+C content. The 3'-UTRs in farm animals and humans vary from 77 bp in Japanese quail *CAR* to 163 bp in bovine *CAR* with average size of 126 bp (Table 3.1). The TGA stop codon is located at nucleotide positions 1261-1263. The polyadenylation signal ATTAAA is located at nucleotide positions 1374-1379 (Figure 3.1). The secondary structure of the 3'-UTR of *CAR* transcript was predicted using the RNAfold program. The predicted fold energy in the 3'-UTR is -37.4 kcal/mol. The predicted secondary structure of porcine *CAR* 3'-UTR is shown in Figure 3.4. Using the UTRdb, the 3'-UTR of porcine *CAR* (accession no. CC456774) did not possess the known regulatory elements involved in post-transcriptional regulation such as mRNA localization, stability, and translational efficiency (Pesole *et al.,* 2001; Hughes, 2006).

.

31

Figure 3.4. Predicted secondary structure of 3'-UTR of porcine *CAR*.

The secondary structure of 3'-UTR was predicted using the RNAfold program. The predicted fold energy of porcine *CAR* 3'-UTR is -37.4 kcal/mol.

3.2.2. Nucleotide Sequence of Porcine *PXR*

The full-length cDNA sequence of porcine *PXR* contains 2567 bp (52.2% G+C content), which has a 273-bp 5'-UTR (nucleotide positions 1-273) and a 1028-bp 3'-UTR (nucleotide positions 1540-2567). It contains a 1266-bp putative coding region (nucleotide positions 274-1539) that encodes 421 amino acids (Figure 3.5). The full-

length cDNA sequence of porcine *PXR* has been submitted to the DDBJ/EMBL/GenBank databases under the accession number AB214980 (Thadtha *et al.,* 2005). Nucleotide sequence analysis revealed that full-length porcine *PXR* shares 70% identity with bovine *PXR*, 46% identity with human *PXR*, and 65% identity with mouse *PXR* (Table 3.4). The nucleotide composition of porcine *PXR* is also shown in Table 3.2.

```
   1 AACCGATGAGTAGGACAGGGAAGAGGAAGCACTGCCCTGATTTCAGCAGGAGTGGCAGCC
  61 TCTCGCCCAAACGAAGTGTTCACTCCTGAGAAAAGCAAAGGATTAAGCTAATGCTCCAAT
 121 CTTTAACAAGGCCGCTACTTCAGGGCTACTGATATTTTGATCAAAGCAGACCTCACCGAA
 181 GTCAGAGCCAAGACGTTACCACCAAGCAGTCCAAGAGACTCACAAGCAAAGCTGGAGGTG
 241 AGACCTGAAGAAAGCTGGGACCATGCTGACCTCATGCAATGCAATGAAACAGACTCCACT
                                          M  Q  C  N  E  T  D  S  T      9
 301 TCTGGAAATTCCACCACCAATGCAGATGAGGAAGATGAGGGTCCCCAGATCTGCCGTGTA
     S  G  N  S  T  T  N  A  D  E  E  D  E  G ⌐P  Q  I  C  R  V     29
 361 TGTGGGGACAAGGCCACTGGTTATCATTTCAATGTTATGACATGTGAAGGATGCAAGGGC
     C  G  D  K  A  T  G  Y  H  F  N  V  M  T  C  E  G  C  K  G     49
 421 TTTTTCAGGAGGGCCATCAAACGCAATGCCCGGCCCCGGTGCCTCTTCCGGAAGGGCGCC
     F  F  R  R  A  I  K  R  N  A  R  P  R  C  L  F  R  K  G  A     69
 481 TGCGAGATCACCCGGAAGACTCGGCGGCAGTGCCAGGCCTGCCGCCTCCGCAAGTGTCTG
     C  E  I  T  R  K  T  R  R  Q  C  Q  A  C  R  L  R  K  C  L     89
 541 GAAAGCGGCATGAGGAAGGAAATGATCATGTCAGATGCAGCTGTGGAGCAGAGGCGGGCC
     E  S  G  M  R  K  E⌐ M  I  M  S  D  A  A  V  E  Q  R  R  A    109
 601 TTGATCAGGAGGAAGAAACGAGAACAGATCGGGGCTCAGCCCCCAGGAGCCAAGGGTCTC
     L  I  R  R  K  K  R  E  Q  I  G  A  Q  P  P  G  A  K  G  L    129
 661 ACTGAAGAGCAGCGGACAATGATCAGTGAGCTGATGAACGTTCAGATGAAAACCTTTGAC
     T  E  E  Q  R  T  M  I  S  E  L  M  N  V  Q  M  K  T  F  D    149
 721 ACCACCTTCACACATTTCAAGAATTTTCGGTTACCAGAGGTGCTTAGCAGTAGCCTCGAG
     T  T  F  T  H  F  K  N  F  R  L  P  E  V  L  S  S  S  L  E    169
 781 ATTCCAGAGTGTCTGCAGACTCCGTCGTCAAGGGAAGAAGCTGCCAAGTGGAGCAAGCTC
     I  P  E  C  L  Q  T  P  S  S  R  E  E  A  A  K  W  S  K  L    189
 841 AGGGAAGATCTGTGCTCAGTGAAACTCTCTCTGCAGCTAAGGGGGGAAGATGGTAGCGTC
     R  E  D  L  C  S  V  K  L  S  L  Q  L  R  G  E  D  G  S  V    209
 901 TGGAACTACAAACCCCCAGCAGACAACAGTGGGAAAGAGATCTTTTCCCTGCTGCCCCAC
     W  N  Y  K  P  P  A  D  N  S  G  K  E  I  F  S  L  L  P  H    229
 961 ATAGCTGACATGTCAACCTACATGTTCAAAGGCATTATCAACTTTGCCAAAGTCATCTCC
     I  A  D  M  S  T  Y  M  F  K  G  I  I  N  F  A  K  V  I  S    249
1021 TACTTCAGGGACTTGCCCATTGAGGACCAGATCTCTCTGCTGAAGGGGGCCACCTTTGAG
     Y  F  R  D  L  P  I  E  D  Q  I  S  L  L  K  G  A  T  F  E    269
1081 CTGTGCCAGCTGAGATTCAACACGGTGTTCAACGCAGAGACGGGGACCTGGGAGTGTGGT
     L  C  Q  L  R  F  N  T  V  F  N  A  E  T  G  T  W  E  C  G    289
1141 CGGCTGTCCTACAGCTTGGAAGACCCCTCAGGTGGCTTCCAGCAGCTTCTCCTGCAGCCC
     R  L  S  Y  S  L  E  D  P  S  G  G  F  Q  Q  L  L  L  Q  P    309
1201 ATGCTGAAATTCCACTACATGCTGAAGAAGCTGCAGCTGCATAAGGAGGAGTATGTGCTG
     M  L  K  F  H  Y  M  L  K  K  L  Q  L  H  K  E  E  Y  V  L    329
1261 ATGCAGGCCATCTCCCTTTTCTCTCCAGACCGCCCGGGTGTGGTGCAACGCCAAGTGGTG
     M  Q  A  I  S  L  F  S  P  D  R  P  G  V  V  Q  R  Q  V  V    349
1321 GACCAGCTGCAGGAGAGGTTTGCCATTACCCTGAAGGCCTACATCGAGTGCAACCGGCCC
     D  Q  L  Q  E  R  F  A  I  T  L  K  A  Y  I  E  C  N  R  P    369
1381 CAGCCTGCCCACCGATTCCTGTTCCTGAAGATCATGGCTATGCTCACTGAGCTCCGCAGC
     Q  P  A  H  R  F  L  F  L  K  I  M  A  M  L  T  E  L  R  S    389
1441 ATCAACGCCCAACACACCCAGCGGCTGCTGCGAATCCAGGACATACACCCCTTCGCCACC
     I  N  A  Q  H  T  Q  R  L  L  R  I  Q  D  I  H  P  F  A  T    409
```

```
1501  CCACTCATGCAGGAGTTATTCAGCATCACAGAAAGCTGAACCAGGGCCCTCGGAGTCGCC
       P  L  M  Q  E  L  F  S  I  T  E  S                        421
1561  ACTCCTGGATCTAAACAGATGGAAGCAACTGACAATGCCCAAAGGTCCTGCCTCCCCAGG
1621  AAACTCAGCCATGATGATGGCTGGCTAGCATTACTCAGGAAGGGGCCATAGGTCCCCTCA
1681  GCCCCCAGTTCAGTATGTGGAAAGCCAAGCCCTAGACCACTACGTGGAGAGTATACTGGC
1741  CCATAGGTCAGTTCCAGGAGCGGCAAGGCCACCTTCCCCTTAGAAAAAGCCCTGGGTCTG
1801  GAGATTTAGTGTCCTGGTGGAAGAGGGGAAAGGGCACCTGGGGCTGGGCCATTTGAGGGT
1861  CTGTGCTCACATCCAAGTTCATTAGCTTCTTGGGTATTTTCACTGCTATGCCTAGTACCC
1921  CTGTCTCCCACTTGCTCCCCATTCCCAGCCACAGCCTCCTGCCCTGAGCTGCTCCATGAA
1981  CTCCGGGCCTTTTTCCACCAGCAGGTGCATTAATGTCTAAGATAGCCTTCTAGAGAGAGG
2041  CCAGAAGGCCCACACCAAATGTCAAAAGCTTGTCATTACCTTTCTCCATGTCTGTCCTTC
2101  CTACCTTCTGGGCATTTCATTCTGTCTCTGCATCCATTAAAACACATCATTAAGCACCAA
2161  CCATATGCAACCTGCTGTGCGTAATAAACGCTGACTCAGACATGGATCCTGACCTGCAAG
2221  GTCTGTAGTTCAGTAAAGAAAACAAAGCAGAAACACGAATAATTTGATCAAAAGGAAAAG
2281  TGATGTATGACATAAGGAGCCCAAGAAATTGATTTGCGTGGATGCTGAACTGTGCTAGGC
2341  TTCTGCATGCACCGGGACACAAGTGAGGGGATCCCTAGAACATGGCTCGTGGGAGGAAGG
2401  GTACTAGCTGCTGTGTGCATGTGTGTTCCTTGGTGCAGGTCTGCTCCACAACTGAACTGG
2461  GGCTGGGTTGCTCACGAGGCTGGAATACTGGGACTACACAGTCCACAGTTAATAAACATG
2521  TTCCTGTTTACACATCTATTTTCAAAGCTAAAAAAAAAAAAAAAAAAAA  2567
```

Figure 3.5. Nucleotide and deduced amino acid sequences of porcine *PXR*.

The numbers on the left and right indicate the nucleotide base and the deduced amino acid in porcine *PXR*, respectively. The ORF was deduced using the ORF Finder. The deduced amino acid sequence is shown in single-letter code below the nucleotide sequence. The ATG start codon (positions 274-276) and the TGA stop codon (positions 1537-1539) are printed in bold. The predicted DNA-binding domain (amino acid residues 24-96) is boxed. The putative ligand-binding domain of porcine PXR is underlined. The polyadenylation signal (AATAAA) is also underlined.

Table 3.4. Sequence identity of *PXR* cDNA sequences between farm animals and human.

Gene	mRNA			% identity	Length
	5'-UTR	ORF	3'-UTR		
Porcine *PXR*	273	1266	1028		2567
Bovine *PXR*	256	1263	831	70	2350
Human *PXR*	1839	1305	1302	46	4446
Mouse *PXR*	335	1296	909	65	2540

Note: Nucleotide sequence identities were determined using the pairwise alignment in BioEdit. The percent identity of each *PXR* mRNA sequence was compared to porcine *PXR*.

A. 5'-Untranslated Region

The porcine *PXR* has 273 bases in 5'-UTR with 50.2% G+C content (Table 3.2). The 5'-UTRs vary from 256 bp in bovine *PXR* to 1839 bp in human *PXR* with average size of 676 bp (Table 3.4). The secondary structure of the 5'-UTR of porcine *PXR* transcript was predicted using the RNAfold program. The predicted fold energy of 5'-UTR is -78.4 kcal/mol. The predicted secondary structure of porcine *PXR* 5'-UTR is shown in Figure 3.6. Using analysis tools at the UTRdb, the 5'-UTR porcine *PXR* (Accession no. BB408135) did not possess the known regulatory elements involved in post-transcriptional regulation such as mRNA localization, stability, and translational efficiency (Pesole *et al.,* 2001; Hughes, 2006).
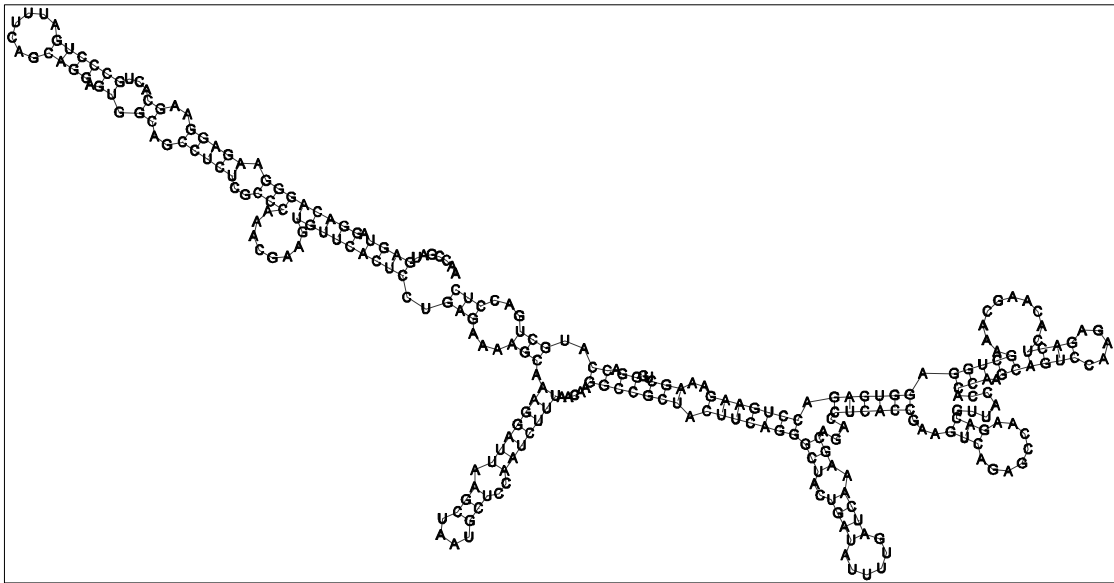
Figure 3.6. Predicted secondary structure of porcine *PXR* 5'-UTR.

The secondary structure of 5'-UTR was predicted using the RNAfold program. The predicted fold energy of porcine *PXR* 5'-UTR is -78.4 kcal/mol.

B. Open Reading Frame

The coding region of porcine *PXR* has 1266 bases with 54.7% G+C content. The ATG start codon encoding a methionine residue and the TGA stop are located at nucleotide positions 274 and 1537, respectively (Figure 3.5). The coding region of porcine *PXR* encodes 421 amino acids. Sequence upstream of the putative start ATG codon in porcine *PXR* cDNA did not contain any additional in-frame start codons. Four in-frame stop codons (each two of TAG and TAA stop codons) were found upstream of the predicted ATG start codon. Analysis of ORF indicated that porcine *PXR* has overall coding GC of 54.66%, first codon position GC of 52.84%, second codon position GC of 40.52%, and third codon position GC of 70.62%. Codon usage in porcine *PXR* is shown in Table 3.3. The codons GCC (alanine), GAG (glutamic acid), UUC (phenylalanine), AUC (isoleucine), AAG (lysine), CUG (leucine), and CAG (glutamine) were frequently used in porcine *PXR* ORF.

The nucleotide sequence around ATG initiation codon of porcine *PXR* mRNA is according with Kozak consensus sequence (GCC<u>A</u>/G<u>CC*ATG*G</u>) (Kozak, 1996). The nucleotide flanking ATG in porcine *PXR* is GAC<u>C</u>TC*ATG*<u>C</u>. Porcine *PXR* has both pyrimidine Cs in positions -3 and +4, respectively. This classified sequence flanking porcine *PXR* ATG initiation codon as a weak Kozak consensus sequence. The consensus sequence flanking ATG codons in *PXR* is shown in Figure 3.7.
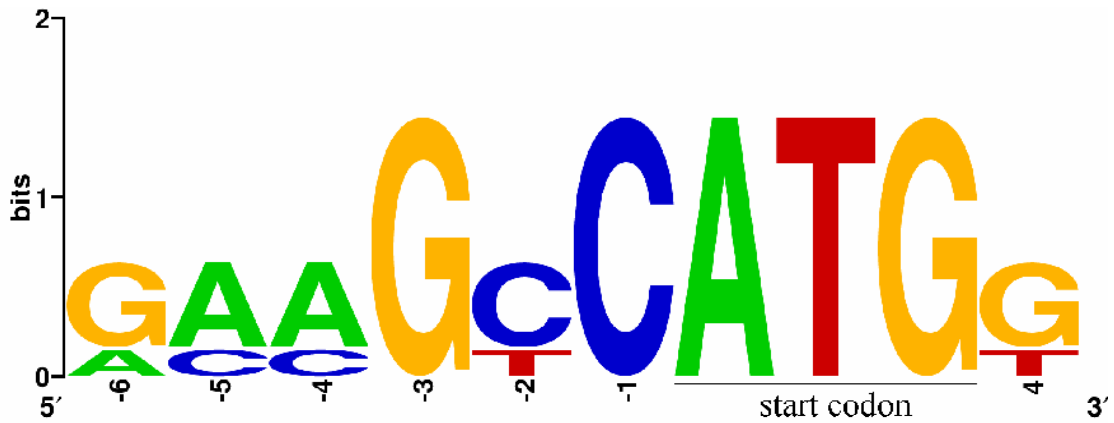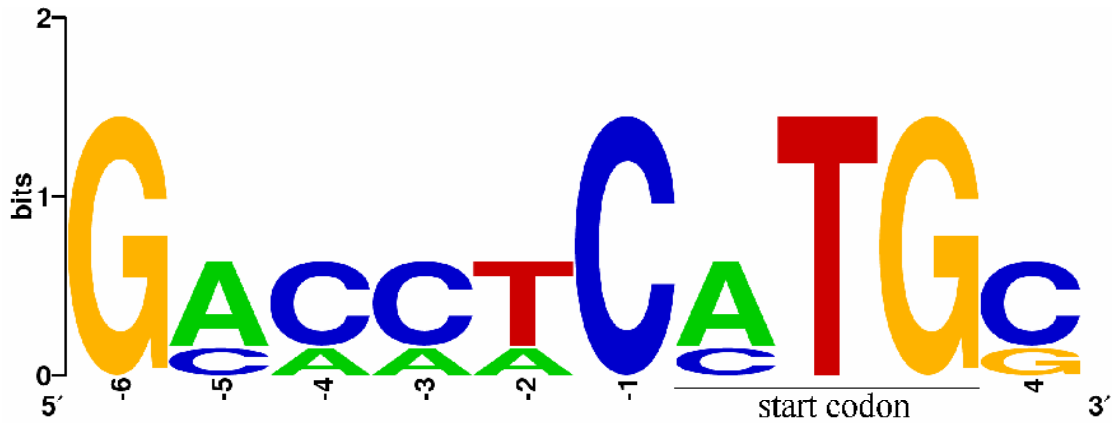
Figure 3.7. Kozak consensus sequence surrounding ATG start codon of *PXR* in farm animals and human.

The nucleotide flanking ATG in porcine *PXR* is GAC<u>C</u>TC*ATG*<u>C</u>. Porcine *PXR* has both pyrimidine Cs in positions -3 and +4, respectively. This classified sequence flanking porcine *PXR* ATG initiation codon as a weak Kozak consensus sequence. In human *PXR* the initiation codon is CTG. The WebLogo was used in the create diagram representing the Kozak consensus sequence.

C. 3'-Untranslated Region

The 3'-UTR of porcine *PXR* has 1028 bases in length (nucleotide positions 1540-2567) with 49.7% G+C content. The 3'-UTRs vary from 831 bp in bovine *PXR* to 1302 bp in human *PXR* with average size of 1018 bp (Table 3.4). The 3'-UTR is longer than 5'-UTR in *PXR* mRNA sequences. The TGA stop codon is located at nucleotide positions 1537-1539. The polyadenylation signal AATAAA is located at nucleotide positions 2511-2516 (Figure 3.5). The secondary structure of the 3'-UTR of porcine *PXR* transcript was predicted using the RNAfold program. The predicted fold energy of 3'-UTR is -357.3 kcal/mol. The predicted secondary structure of porcine *PXR* 3'-UTR is shown in Figure 3.8. Using the UTRdb, the 3'-UTR of porcine *PXR* (Accession no. CC456777) did not possess the known regulatory elements involved in post-transcriptional regulation of mRNA such as mRNA localization, stability, and translational efficiency.
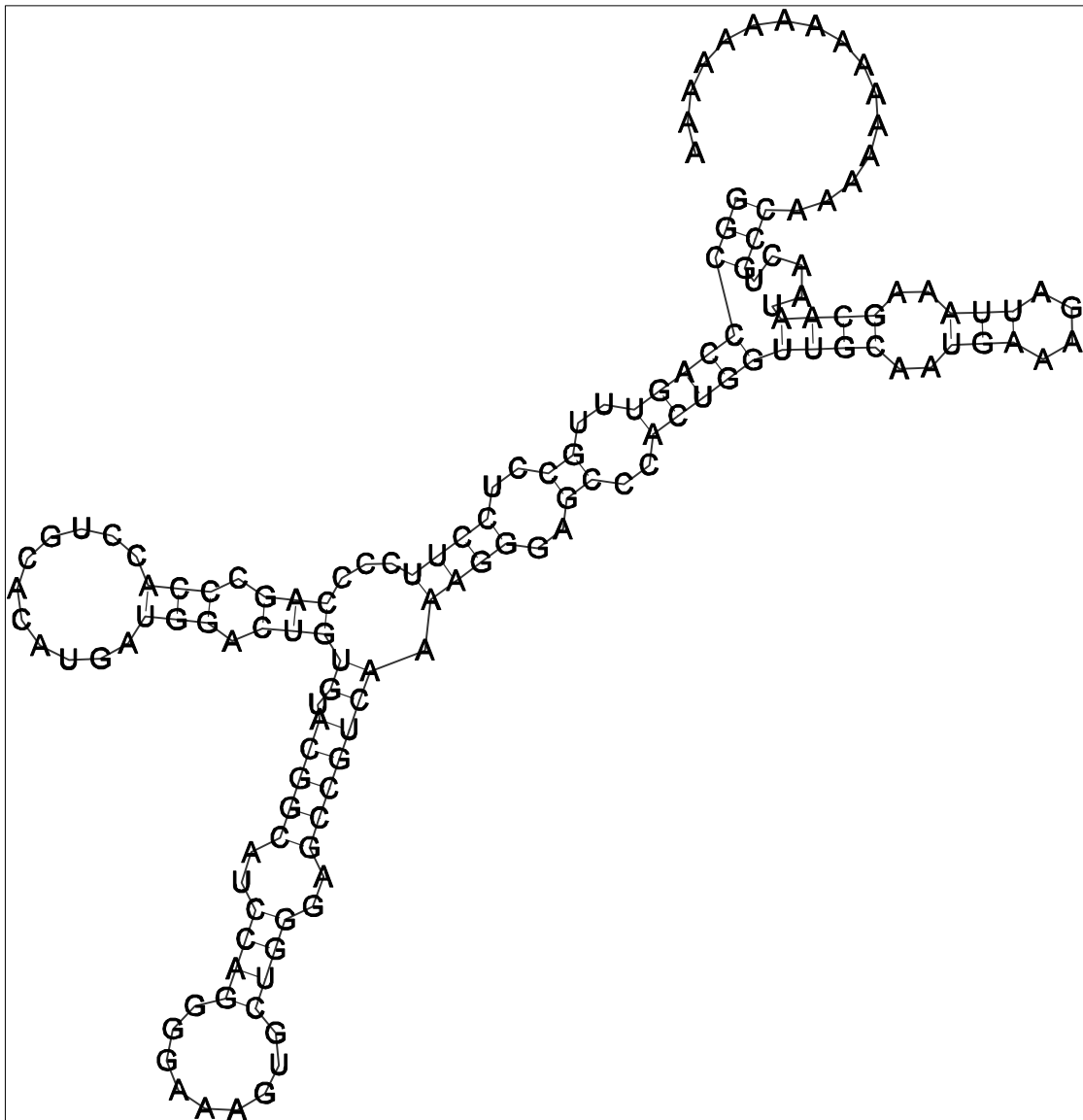
38
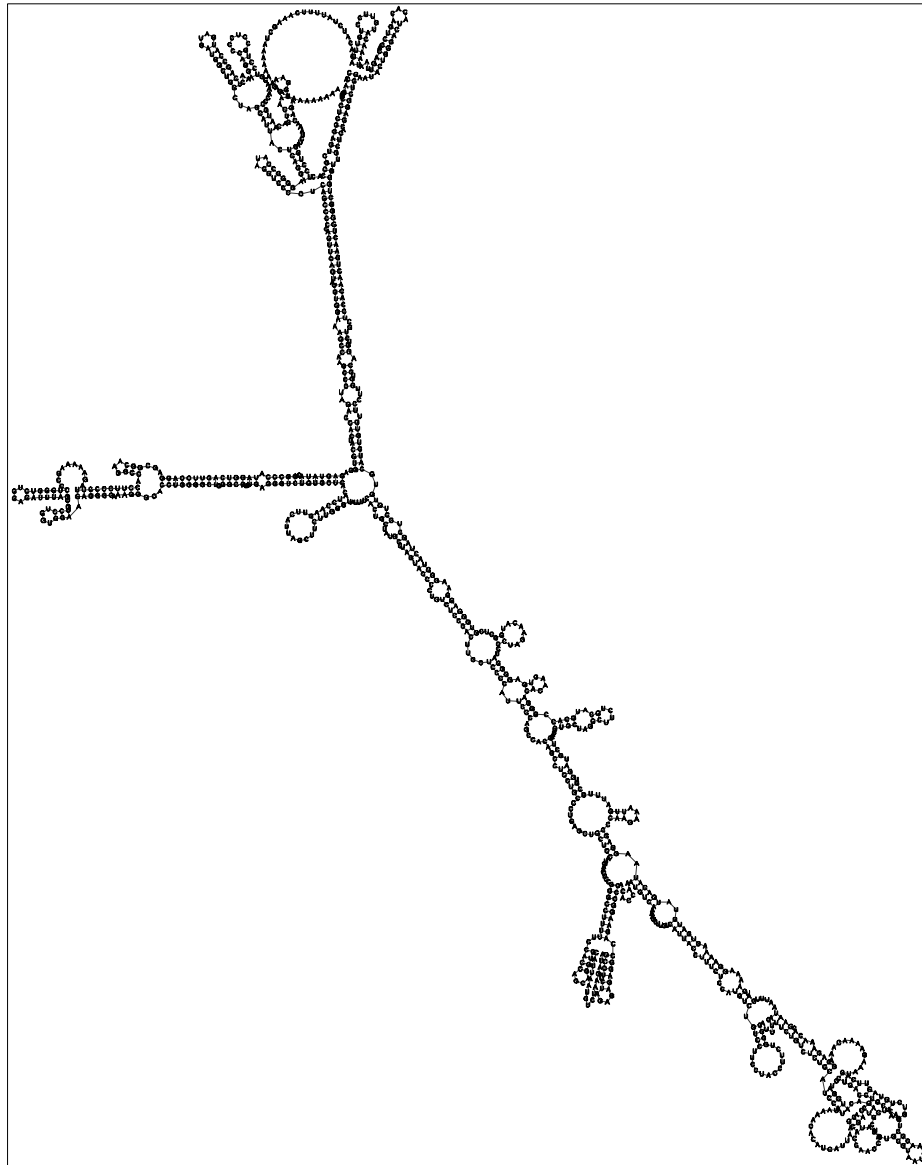
Figure 3.8. Predicted secondary structure of porcine *PXR* 3'-UTR.

The secondary structure of 5'-UTR was predicted using the RNAfold program. The predicted fold energy of porcine *PXR* 3'-UTR is -357.3 kcal/mol.

3.3. Predicted Amino Acid Sequences of Porcine CAR and PXR

3.3.1. Predicted Amino Acid Sequence of Porcine CAR

The predicted porcine CAR protein has 348 amino acid residues with a calculated molecular weight of 39809 Daltons and an isoelectric point of 8.94. The amino acid composition of porcine CAR protein is shown in Table 3.5. The accession number of porcine CAR protein is BAE54304.

Protein sequence analysis indicated that porcine CAR protein shares 84% identity with human CAR, 88% identity with bovine CAR, 32% identity with Japanese quail CAR, and 30% with chicken CXR. When compared between CAR and PXR proteins, it revealed 38% identity in both swine and humans (Table 3.6).

Table 3.5. Amino acid composition of porcine CAR and PXR proteins.

Amino Acid	CAR		PXR	
	Number	Mol (%)	Number	Mol (%)
Alanine (A)	25	7.2	25	5.94
Cysteine (C)	13	3.7	15	3.56
Aspartic acid (D)	12	3.5	16	3.80
Glutamic acid (E)	20	5.8	33	7.84
Phenylalanine (F)	20	5.8	23	5.46
Glycine (G)	18	5.2	21	4.99
Histidine (H)	12	3.5	8	1.90
Isoleucine (I)	18	5.2	23	5.46
Lysine (K)	18	5.2	26	6.18
Leucine (L)	48	13.8	43	10.21
Methionine (M)	12	3.5	16	3.80
Asparagine (N)	7	2.0	14	3.33
Proline (P)	16	4.6	19	4.51
Glutamine (Q)	31	8.9	27	6.41
Arginine (R)	23	6.6	34	8.08
Serine (S)	16	4.6	28	6.65
Threonine (T)	16	4.6	25	5.94
Valine (V)	15	4.3	14	3.33
Tryptophan (W)	1	0.3	3	0.71
Tyrosine (Y)	7	2.0	8	1.90

Note: Amino acid composition was calculated using ProParam at ExPaSy protein site. Number indicates frequency of each amino acid residue. Percentage represents calculated molecular weight of each amino acid residue in porcine CAR and PXR proteins.

Table 3.6. Percent identity in CAR proteins.

Sequence	% identity	Length
porcine CAR-bovine CAR	88	348-348
porcine CAR-chicken CXR	30	348-391
porcine CAR-quail CAR	32	348-385
porcine CAR-human CAR	84	348-348
porcine CAR-porcine PXR	38	348-421
human CAR-human PXR	38	348-434

Note: Percent identity in amino acid sequences was calculated using pairwise alignment in BioEdit with default parameters (using BLOSUM62 as the scoring matrix with open gap extension penalty of -8 and gap extension penalty of -2).

A. Domain Structure

By using the SMART analysis tool, porcine CAR protein was predicted to posses the highly conserved DBD (amino acid residues 8-79) and the moderately conserved LBD (amino acid residues 163-348) (Figure 3.1). The bipartite nuclear localization signal was also identified in the positions 77-94 (KKDMILSAEVLALRRARQ). The comparison of the structural domains of CAR receptor proteins between farm animals and human is shown in Figure 3.9. In the DBD, the porcine CAR receptor protein share 83% amino acid identity with human CAR, 97% identity with bovine CAR, 67% identity with chicken CXR, and 69% identity with Japanese quail CAR; and in the LBD, it shares 87% amino acid identity with human CAR, 85% identity with bovine CAR, and 58% identity with chicken CXR and Japanese quail CAR.

Sequence homology in the putative DBD varies from 67% identity in chicken CXR to 97% identity in cattle. The DBD is responsible for the binding of receptors to promoter of its target genes and has the P box and the zinc finger motifs (Mangelsdorf *et al.,* 1995; Giguere, 1999; Khorasanizadeh and Rastinejad, 2001; Robinson-Rechavi *et al.,* 2003). The highly conserved P-box is located at the positions 28-35 (CEGCKGFF) in the DBD and is common to most nuclear receptors (Figure 3.10). The first zinc finger motif was identified at positions 11-31 (CAVCGDRATGYHFHALTCEGC). The second zinc finger motif was identified at positions 47-66 (CPFAGSCKVNKAQRRHCPAC) (Figure 3.10). Thirteen amino acid substitutions in the DBD between porcine and human CAR were identified (Pro8Leu, Ala12Val, Arg17Gln, His24Asn, Asn40Ser, Thr43Ile, Ser44Gly, Leu45Pro, Ile46Thr, Lys54Glu, Asn56Ser, Ala58Thr, and Lys77Arg) (Figure 3.10). Eight amino acid substitutions (Pro8Leu, Ala12Val, Arg17Gln, His24Asn, Asn40Ser, Leu45Pro,

42

Asn56Ser, and Lys77Arg) are the replacement of amino acids with similar properties. Five amino acid substitutions (Thr43Ile, Ser44Gly, Ile46Thr, Lys54Glu, and Ala58Thr) are the replacement of amino acids with different properties.

Sequence alignment of the LBD for swine and human is shown in Figure 3.11. The highly conserved peptide EDQI(S/A/I/T/V)LLK is located in the LBD and is conserved in most nuclear receptors. Porcine CAR LBD has the peptide EDQISLLK at amino acid residues 188-195. The ligand-dependent activation-2 (AF-2), important in ligand and coactivator binding, was also found in the LBD (amino acid residues 341-348: PLLQEICS) (Figure 3.11). Alignment analysis revealed that porcine CAR has 24 amino acid substitutions compared with human protein in the LBD (Ile169Val, Gln170Leu, Ile172Val, Leu181Val, Met187Ile, Gln203His, Lys216Asn, His231Arg, Glu236Val, Gly244His, Lys247Gly, Arg251Lys, Met262Leu, Val263Ala, Val265Met, Lys279Asp, Pro301Arg, Ser302Arg, Leu303Pro, Lys324Glu, Glu325Ala, Trp327Gly, Asn332His, Thr338Ala) (Figure 3.11). Fourteen amino acid substitutions (Ile169Val, Ile172Val, Leu181Val, Met187Ile, Gln203His, Lys216Asn, His231Arg, Arg251Lys, Met262Leu, Val263Ala, Val265Met, Ser302Arg, Leu303Pro, and Asn332His) are the replacement of amino acids with similar properties. Other ten amino acid substitutions (Gln170Leu, Glu236Val, Gly244His, Lys247Gly, Lys279Asp, Pro301Arg, Lys324Glu, Glu325Ala, Trp327Gly, and Thr338Ala) are the replacement of amino acids with different properties.

Figure 3.9. Comparison of structural domains of porcine CAR protein.

The structural domain of porcine CAR protein was predicted by using the SMART. Percent identity in amino acid sequences was calculated using pairwise alignment in BioEdit (BLOSUM62 scoring matrix). The structural domains are indicated. Percent sequence identity is shown in each structural domain in comparison with porcine CAR. Numbers above each structural domain indicate position of amino acid in boundary region.

```
porcine CAR 8   PRNCAVCGDRATGYHFHALTCEGCKGFFRRTVNKSTSLICPFAGSCKVNKAQRRHCPACR 67
human CAR   8   L...V....Q......N................S..IGPT.......E.S.T......... 67

porcine CAR     LQKCLDAGMKKD 79
human CAR       .........R.. 79
```

Figure 3.10. Alignment of the DBD domains between porcine and human CAR proteins.

Amino acid is colored according to its property. Dots represent identical amino acid. Different amino acids between pig and human are indicated by single amino acid codes. The first zinc finger motif and the second zinc finger motif are underlined. Zinc finger motifs in porcine CAR DBD were predicted using the ScanProsite. The conserved peptide CEGCKGFF located in the highly conserved P-box is boxed. Thirteen amino acid substitutions, eight conserved substitutions and five non-conserved substitutions, in the DBD between porcine and human CAR were identified.

```
porcine CAR 163 DINTFMIQQIIKFTKDLPLFRSLPMEDQISLLKGAAVEICQIVLNTTFCLQTQKFLCGPL 222
human CAR   163 ......VL.V........V.....I................H.............N...... 222


porcine CAR     RYTIEDGAHVGFQEEFLELLFGFHKTLRRLQLQEPEYVLMVAVALFSPDRPGVTQRKEID 282
human CAR       .......R....V.......H..G...K..........LA.M.............D... 282


porcine CAR     QLQEEMALTLQSYIKGQQPSLRDRFLYAKLLGLLAELRSINKEYWYQIQNIQGLSTMMPL 342
human CAR       .................RRP.....................EA.G....H.....A.... 342


porcine CAR     LQEICS 348
human CAR       ...... 348
```

Figure 3.11. Alignment of the LBD domains between porcine and human CAR proteins.

Amino acid is colored according to its property. The conserved peptide EDQISLLK and the AF-2 (PLLQEICS) are boxed. Dots represent identical amino acids between pig and human. Different amino acids are indicated by single amino acid codes. Twenty-four amino acid substitutions, fourteen conserved substitutions and ten non-conserved substitutions, in the LBD between porcine and human CAR were identified.

B. Homology Modeling

The porcine CAR protein 3-D structure was predicted by using the SWISS Model Server (Schwede *et al.,* 2003). The porcine CAR protein (UniProtKB/TrEMBL accession no. Q2V0W3) was used as a query. The crystal structure of human CAR (PDB: 1XVP) was identified with 85% identity (Xu *et al.,* 2004). The 1XVP template was used in the 3-D homology modeling of porcine CAR. The ribbon representation of the predicted 3-D structure of porcine CAR is shown in Figure 3.12.

The secondary structure of porcine CAR LBD was predicted to enclose by helix 1 (K108-G126), helix 2 (M128-F135), helix 2' (P138-I143), helix 3 (L155-D178), helix 3' (P180-S184), helix 4 (M187-K195), helix 5 (A197-T208), helix 6 (I226-V232), helix 7 (E236-L252), helix 8 (E256-F268), helix 9 (R278-Q300), helix 10 (R306-N332), helix X or helix 11 (Q334-M339), helix AF2 (P341-S348) and 3 beta-strands (β-1, β-2, β-3: F210-L212, Q215-C219, and L222-T225, respectively)

(Figure 3.13). This secondary structure was derived from the alignment of porcine

CAR and PXR LBDs with human and mouse (Suino *et al.,* 2004).

Figure 3.12. Ribbon representation of predicted 3-D structure of porcine CAR protein.

The 3-D model was predicted using the service of the Swiss Model Repository. The crystal structure of human CAR (PDB template: 1XVP) was identified and used in the construction of porcine CAR 3-D model. The image was produced using the UCSF Chimera. The predicted secondary structural elements are indicated (H: α helix and S: β strand).

```
pCAR 106 LSKEQKALVQILLGAHTRHMGTMFDQFVQFRPPAHLFIHHQH------------------ 147
hCAR 106 LSKEQEELIRTLLGAHTRHMGTMFEQFVQFRPPAHLFIHHQP------------------ 147
pPXR 129 LTEEQRTMISELMNVQMKTFDTTFTHFKNFRLPEVLSSSLEIPECLQTPSSREEAAKWSK 188
hPXR 143 LTEEQRMMIRELMDAQMKTFDTTFSHFKNFRLPGVLSSGCELPESLQAPS-REEAAKWSQ 201
              alpha1              alpha2   alpha2'          alpha2

pCAR     ------------------------LPPLVP----ELSLLMHFADINTFMIQQIIKFTKDL 179
hCAR     ------------------------LPTLAP----VLPLVTHFADINTFMVLQVIKFTKDL 179
pPXR     LREDLCSVKLSLQLRGEDGSVWNYKPPADNSGKEIFSLLPHIADMSTYMFKGIINFAKVI 248
hPXR     VRKDLCSLKVSLQLRGEDGSVWNYKPPADSGGKEIFSLLPHMADMSTYMFKGIISFAKVI 261
                  beta1     beta1'                  alpha3

pCAR     PLFRSLPMEDQISLLKGAAVEICQIVLNTTFCLQTQKFLCGPLRYTIEDGAHVGFQEEFL 239
hCAR     PVFRSLPIEDQISLLKGAAVEICHIVLNTTFCLQTQNFLCGPLRYTIEDGARVGFQVEFL 239
pPXR     SYFRDLPIEDQISLLKGATFELCQLRFNTVFNAETGTWECGRLSYSLED-PSGGFQQLLL 307
hPXR     SYFRDLPIEDQISLLKGAAFELCQLRFNTVFNAETGTWECGRLSYCLED-TAGGFQQLLL 320
           alpha3'  alpha4      alpha5      beta2 beta3 beta4 alpha6

pCAR     ELLFGFHKTLRRLQLQEPEYVLMVAVALFSPDRPGVTQRKEIDQLQEEMALTLQSYIKGQQ 300
hCAR     ELLFHFHGTLRKLQLQEPEYVLLAAMALFSPDRPGVTQRDEIDQLQEEMALTLQSYIKGQQ 300
pPXR     QPMLKFHYMLKKLQLHKEEYVLMQAISLFSPDRPGVVQRQVVDQLQERFAITLKAYIECNR 368
hPXR     EPMLKFHYMLKKLQLHEEEYVLMQAISLFSPDRPGVLQHRVVDQLQEQFAITLKSYIECNR 381
              alpha7        alpha8              alpha9

pCAR     PSLRDRFLYAKLLGLLAELRSINKEYWYQIQNIQGLSTMM-PLLQEICS 348
hCAR     RRPRDRFLYAKLLGLLAELRSINEAYGYQIQHIQGLSAMM-PLLQEICS 348
pPXR     PQPAHRFLFLKIMAMLTELRSINAQHTQRLLRIQDIHPFATPLMQELFSITES 421
hPXR     PQPAHRFLFLKIMAMLTELRSINAQHTQRLLRIQDIHPFATPLMQELFGITGS 434
               alpha10               alpha11 alpha12/AF
```

Figure 3.13. Predicted secondary structures of porcine CAR & PXR proteins.

The secondary structures of porcine CAR and PXR proteins were adapted from Suino *et al,* 2004. The secondary structures are labeled below corresponding amino acid residues. Amino acid is colored according to its property.

3.3.2. Predicted Amino Acid Sequence of Porcine PXR

The predicted porcine PXR protein has 421 amino acid residues with a calculated molecular weight of 48455 Daltons and an isoelectric point of 8.78. The amino acid composition of porcine PXR protein is shown in Table 3.7. The accession number of porcine PXR protein is BAE54305.

Protein sequence analysis indicates that porcine PXR protein shares 86% identity with human PXR and 85% identity with bovine PXR (Table 3.7). When compared between CAR and PXR proteins, it revealed 38% identity in both swine and humans (Table 3.6).

Table 3.7. Sequence identity in PXR proteins.

Protein	Domain				% identity	Length
	A/B	DBD	D	LBD		
Porcine PXR	23	73	135	190		421
Bovine PXR	23	73	134	190	85	420
Human PXR	37	73	134	190	86	434

Note: Percent identity in amino acid sequences was calculated using pairwise alignment in BioEdit.

A. Domain Structure

Analysis of putative porcine PXR protein showed that it has a 73-residue DBD (amino acid residues 24-96), a 21-amino acid first zinc finger (positions 27-47: CRVCGDKATGYHFNVMTCEGC), a 26-aa second zinc finger (positions 63-88: CLFRKGACEITRKTRRQCQACRLRKC) (Figure 3.5), a 8-residue P-box (amino acid residues 44-51: CEGCKGFF), and a 27-aa nuclear localization signal (position 52-78: RRAIKRNARPRCLFRKGACEITRKTRR).

Sequence comparison of DBD and LBD indicates that porcine PXR DBD shares 93% identity with bovine PXR and 95% identity with human PXR, and porcine PXR LBD exhibits 89% identity with bovine PXR and 93% identity with human PXR (Figure 3.14).

The putative DBD, containing two zinc finger motifs, is conserved between human and farm animals. The highly conserved P-box (CEGCKGFF) is located at amino acid residues 44-51 in the DBD and is common to most nuclear receptors. Four amino acid substitutions in the DBD between porcine and human PXR were identified (Ile55Met69, Pro61Leu75, Leu64Pro78, and Arg94Lys108) (Figure 3.15). These amino acid substitutions are the replacement of amino acids with similar properties.

Sequence alignment of the LBD for swine and human is shown in Figure 3.16. The highly conserved peptide EDQI(S/A/I/T/V)LLK is located in the LBD and is conserved in nuclear receptors. Porcine PXR LBD has the peptide EDQISLLK at amino acid residues 257-264. The AF-2 was also found in the LBD (amino acid residues 410-417: PLMQELFS) (Figure 3.16). Alignment analysis reveals that porcine PXR has 14 amino acid substitutions compared with human PXR proteins (Asn243Ser256, Thr267Ala280, Ser294Cys307, Pro298Thr311, Ser299Ala312, Gln308Glu321, Lys324Glu337, Val344Leu357, Arg346His359, Gln347Arg360, Arg355Gln368, Ala362Ser375, Ser417Gly430, Glu420Gly433) (Figure 3.18). Seven amino acid substitutions (Asn243Ser256, Ser294Cys307, Gln308Glu321, Val344Leu357, Arg346His359, Gln347Arg360, and Arg355Gln368) are the replacement of amino acids with similar physico-chemical properties. Other seven substitutions (Thr267Ala280, Pro298Thr311, Ser299Ala312, Lys324Glu337, Ala362Ser375, Ser417Gly430, and Glu420Gly433) are the replacement of amino acids with different properties.
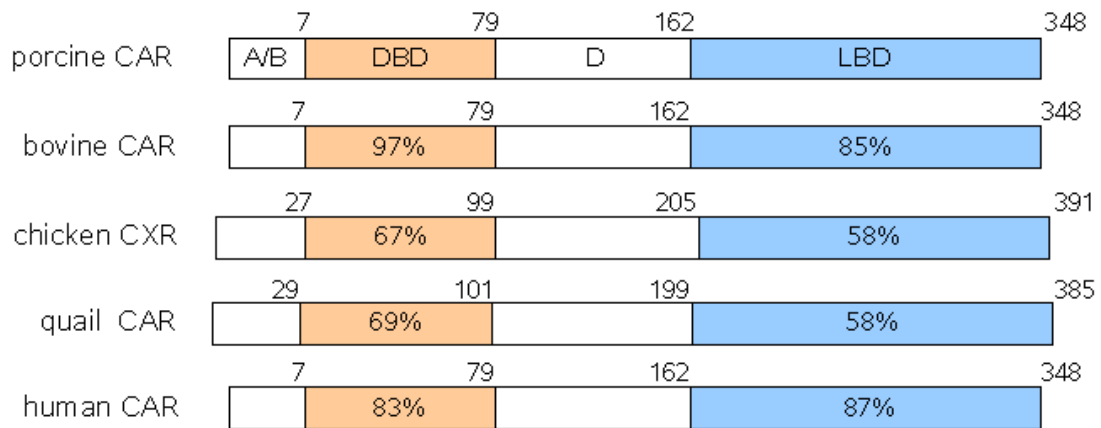
Figure 3.14. Structural domains of porcine PXR protein.

The structural domain of porcine PXR protein was predicted by using the SMART. Percent identity in amino acid sequences was calculated using pairwise alignment in BioEdit. The structural domains are indicated. Percent sequence identity is shown in each structural domain in comparison with porcine PXR. Numbers above each structural domain indicate position of amino acid in boundary region.
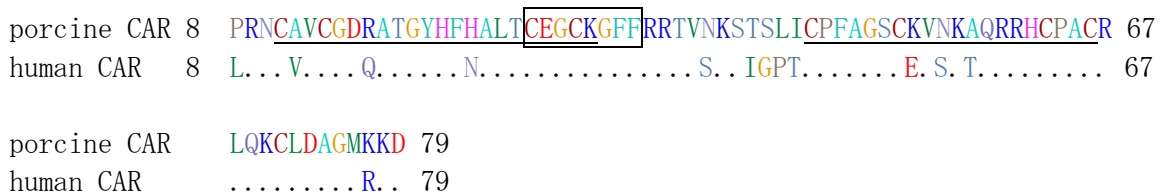
```
porcine PXR 24  PQICRVCGDKATGYHFNVMTCEGCKGFFRRAIKRNARPRCLFRKGACEITRKTRRQCQAC 83
human PXR   38  ............................M.....L..P.................. 97

porcine PXR     RLRKCLESGMRKE  96
human PXR       ..........K..  110
```

Figure 3.15. Alignment of the DBD between porcine and human PXR proteins.

Amino acid is colored according to its property. Dots represent identical amino acids. Different amino acids between porcine and human are indicated by single amino acid codes. The first zinc finger motif and the second zinc finger motif are underlined. Zinc finger motifs in porcine PXR DBD were predicted using the ScanProsite. The conserved peptide CEGCKGFF located in the highly conserved P-box is boxed. Four conserved amino acid substitutions in the DBD between porcine and human PXR were identified (Ile55Met69, Pro61Leu75, Leu64Pro78, and Arg94Lys108).

```
porcine PXR 232 DMSTYMFKGIINFAKVISYFRDLPI EDQISLLK GATFELCQLRFNTVFNAETGTWECGRL 291
human PXR   245 ...........S...................A.............................. 304


porcine PXR     SYSLEDPSGGFQQLLLQPMLKFHYMLKKLQLHKEEYVLMQAISLFSPDRPGVVQRQVVDQ 351
human PXR       ..C...TA.......E...............E....................L.HR.... 364


porcine PXR     LQERFAITLKAYIECNRPQPAHRFLFLKIMAMLTELRSINAQHTQRLLRIQDIHPFAT PL  411
human PXR       ...Q.....S.................................................... 424


porcine PXR      MQELFS ITES 421
human PXR       .....G..G.  434
```

Figure 3.16. Alignment of the LBD between porcine and human PXR proteins.

Amino acid is colored according to its property. The conserved peptide EDQISLLK and the
AF-2 (PLMQELFS) are boxed. Dots represent identical amino acids between swine and
human. Different amino acids are indicated by single amino acid codes. Fourteen amino acid
substitutions, each seven conserved and non-conserved substitutions, in the LBD between
porcine and human PXR were identified.

B. Homology Modeling

The 3-D structure of porcine PXR protein was predicted using the SWISS

Model Server (Schwede *et al.,* 2003). The porcine PXR protein (UniProtKB/TrEMBL

accession no. Q2V0W2) was used as a query. The crystal structure of human PXR

(PDB: 1NRL) was identified with 87% sequence identity (Watkins *et al.,* 2003). The

1NRL template was used in the 3-D homology modeling of porcine PXR. The

predicted 3-D structure of porcine PXR is shown in Figure 3.17.

Secondary structure of porcine PXR was derived from the comparison with

human PXR (Suino *et al.,* 2004). The secondary structure of porcine PXR LBD is

enclosed by helix 1 (E131-T147), helix 2 (R180-W186), helix 3 (L227-V247), helix 3'

(S249-D253), helix 4 (I256-K264), helix 5 (A266-T277), partial helix 7 (M310-L320),

helix 8 (K324-F336), helix 9 (R346-R368), helix 10 (R374-I404), helix AF2 (P410-

S417), and 5 beta- strands (β-1, β-1', β-2, β-3, and β-4: L198-G204, G207-K213,

F279-A281, G284-C288, and L291-S294, respectively) (Figure 3.13). The 60 amino

acid insertion in human PXR LBD (amino acid residues V177-228P), contains a novel

helix 2, β1-β1' region, was also present in porcine PXR LBD (amino acid residues
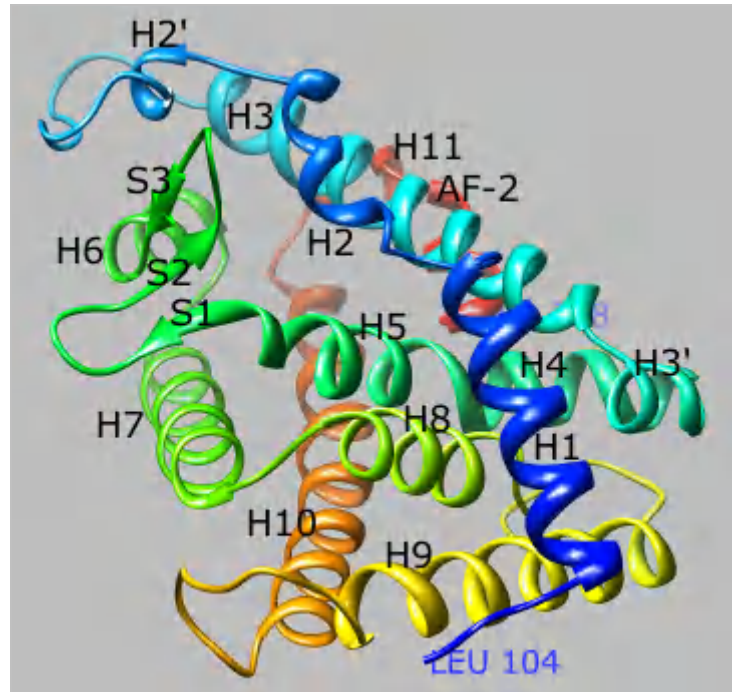
V164-215P) (Figure 3.5).
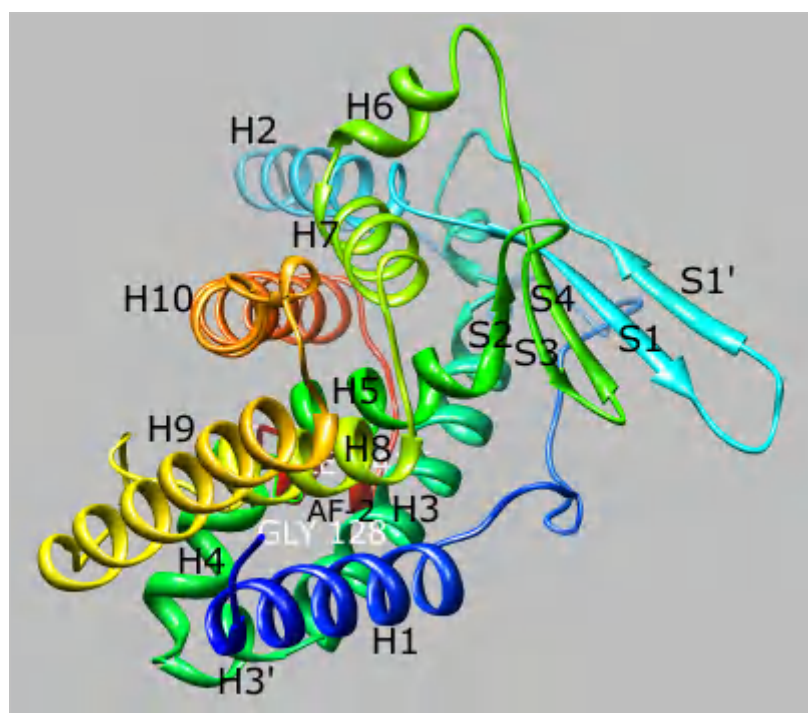
Figure 3.17. Ribbon representation of the predicted 3-D structure of porcine PXR protein.

The 3-D model was predicted using the service of the Swiss Model Repository. The crystal structure of human PXR (PDB template: 1NRL) was identified and used in the construction of porcine PXR 3-D model. The image was produced using the UCSF Chimera. The predicted secondary structural elements are indicated (H: α helix and S: β strand).

3.4. Expression Analysis

The expression pattern analysis of the porcine *CAR* and *PXR* in adult tissues was

performed using the single step RT-PCR. The mRNA expression of the porcine *CAR*

and *PXR* was detected in the liver, small intestine, and kidney, which are the same

tissues in which these genes are expressed in human and mouse (Figure 3.18). The

expression pattern of porcine *CAR* and *PXR* genes exhibited 270 bp and 500 bp DNA bands, respectively. These RT-PCR products were verified by direct sequencing.
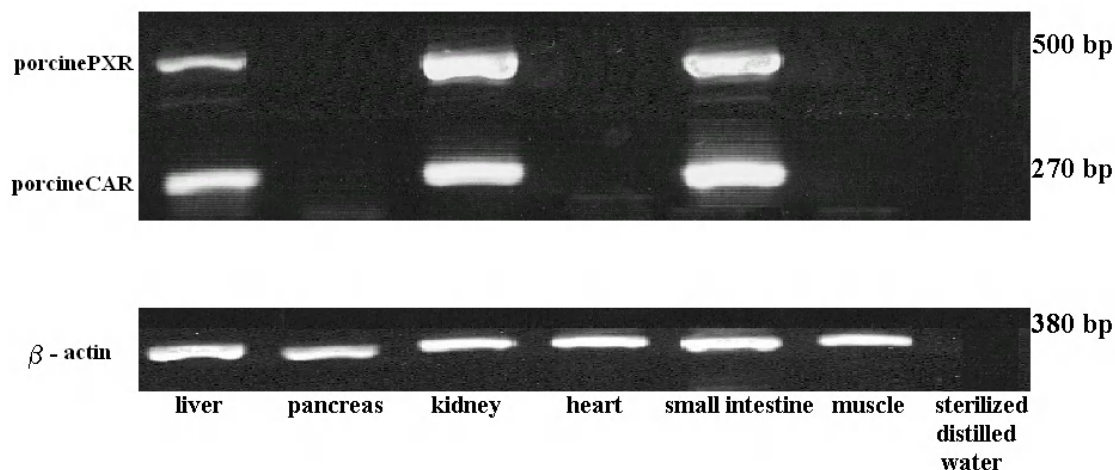
Figure 3.18. Expression analysis of porcine *CAR* and *PXR*.

Total RNA extracted from liver, pancreas, kidney, heart, small intestine, and muscle of an adult female pig was used in RT-PCR. The mRNA expression of both porcine *CAR* and *PXR* was detected in liver, kidney, and small intestine.

3.5. Prediction of Exon-Intron Structure of Porcine *CAR* and *PXR*

3.5.1. Exon-Intron Structure of Porcine *CAR*

To identify genomic DNA corresponds to porcine *CAR* mRNA sequence, we searched ongoing swine genome sequencing project using pig BLAST (http://www.ncbi.nlm.nih.gov/projects/genome/seq/BlastGen/BlastGen.cgi?taxid=982 3). Hits with threshold were not identified. Therefore, the exon-intron structure of porcine *CAR* was predicted by aligning its mRNA sequence with bovine genomic DNA (chromosome 3: 7567938-7572817) using the SIM4 program (Florea *et al.,* 1998). The porcine *CAR* gene was predicted to contain 9 exons ranging in size from 106 to 258 bp with an average of 155 bp. The longest exon is exon 9 with 258 bp. The

shortest exon is exon 8 with 106 bases. All exon-intron borders follow the consensus sequences of U2-type GT-AG splice sites that are GT at the 5' end of the intron and AG at the 3' end of the intron (Mount, 1982). The exon-intron boundaries of porcine *CAR* transcript are shown in Table 3.8. The 216-bp 5'-UTR is encoded by exon 1 and the part of exon 2. The protein-coding region of porcine *CAR* is encoded by 8 exons (exons 2-9). The ATG start codon is located in exon 2. Similar to most nuclear receptors, the DBD is encoded by two separated exons. The first and second zinc finger motifs are encoded by exon 2 and exon 3, respectively. The LBD is encoded by the part of exon 5, exons 6-8, and the part of exon 9. The TGA stop codon is located in exon 9. The 3'-UTRs of porcine *CAR* is encoded by exon 9.

Comparison of *CAR* exons revealed that the size and number of exon are highly conserved between farm animals and human, except exons that encode 5'- and 3'-UTRs. They exhibit approximately 90% identity in protein-coding exons (Figure 3.19). The exons 2 and 3, which encode DBD, are identical in size (140 bp and 131 bp, respectively) and exhibit approximately 90% identity when compared to human (Figure 3.19). The LBD-encoding exons (exons 5-9) are also conserved both in size and sequence identity but slightly lower than DBD encoding exons (~85% identity).

Table 3.8. Predicted exon-intron boundaries of porcine *CAR* transcript.

Exon	Porcine *CAR*	Bovine chr3	Size (bp)	Identity	5' end of exon	3' end of exon	Intron	Intron size	5' splice site	3' splice site
1	1-183	<89-252	183	80	CAGCTGAAGG...	...AGAGGAACAG	1	155	gtaggaaatc...	...ccaaattcag
2	184-323	408-547	140	89	GGGCCTGTGA...	...GTTTCTTCAG	2	537	gtgagaacct...	...gtctccacag
3	324-454	1085-1215	131	92	GCGAACAGTC...	...AAGAAAGACA	3	1141	gtgagttgcc...	...ccgggcacag
4	455-624	2357-2526	170	88	TGATCCTATC...	...GCAGTTCAGG	4	430	gtgagcattt...	...atctctgcag
5	625-764	2957-3096	140	87	CCTCCAGCTC...	...CCCTCTTCCG	5	222	gtacgtgacc...	...ctcctcacag
6	765-910	3319-3464	146	91	GTCCCTGCCC...	...GGAGCGCATG	6	95	gtgagatggt...	...tctcccacag
7	911-1027	3560-3676	117	91	TGGGGTTCCA...	...TTCTCTCCTG	7	181	gtgagcatcc...	...tatgttatag
8	1028-1133	3859-3964	106	86	ACCGGCCTGG...	...TCCGGGACAG	8	753	gtatggtggg...	...gaccacaaag
9	1134-1391	4718-4980	258	85	GTTTCTCTAT...	...AATAACTGCC				

Note: The exon-intron boundaries were predicted by aligning porcine *CAR* cDNA with cattle genomic DNA sequence (chromosome 3: 7567938-7572817) using SIM4 program. The poly(A) tail was excluded in the alignment. The 5'- and 3'-ends of each exon are shown in uppercase letters. The 5' and 3' splice sites of each intron are shown in lowercase letters. All splice sites are in consensus with GT-AG rule for splicing. The sequence identity between porcine *CAR* transcript and cow DNA was derived from alignment results of SIM4.
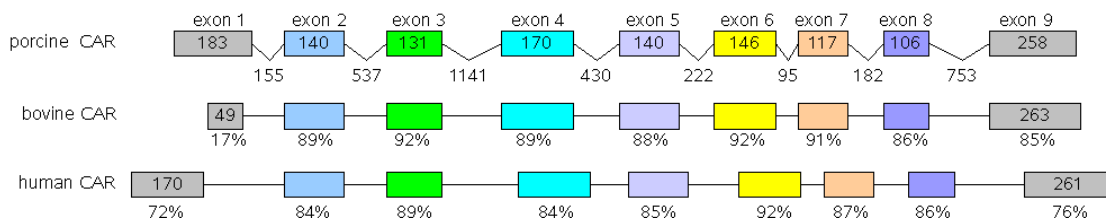
Figure 3.19. Comparison of *CAR* exons in farm animals and human.

The percent identity derived from pairwise alignment using BioEdit is shown below each exon. Boxes represent exons. Exons with similar size are colored. The size of exon is shown in the exon box. The thin lines indicate introns. The length of intron is shown only for porcine *CAR* transcript.

3.5.2. Exon-Intron Structure of Porcine *PXR*

The exon-intron structure of porcine *PXR* was predicted by aligning its mRNA sequence with porcine genomic DNA (Pollock *et al.,* 2007; DQ531175, 10997 bp DNA) using the SIM4 program (Florea *et al.,* 1998). The porcine *PXR* gene was predicted to contain 9 exons ranging in size from 106 to 1155 bp with an average of 283 bp. The longest exon is exon 9 with 1155 bp. The shortest exon is exon 8 with 106 bases. All exon-intron borders are according to the consensus sequences of GT-AG splice sites (Mount, 1982). The exon-intron boundary of porcine *PXR* transcript is shown in Table 3.9. The 273-bp 5'-UTR is encoded by exon 1 and the part of exon 2. The protein-coding region of porcine *PXR* is encoded by 8 exons (exons 2-9). The ATG start codon is located in exon 2. Similar to most nuclear receptors, the DBD is encoded by two separated exons. The first and second zinc finger motifs are encoded by exon 2 and exon 3, respectively. The LBD is encoded by the part of exon 5, exons 6-8, and the part of exon 9. The TGA stop codon is located in exon 9. The 1028-bp 3'-UTRs of porcine *PXR* is encoded by exon 9.

Comparison of *PXR* exons revealed that the size and number of exon are highly conserved when compared to humans, except exons that encode 5'- and 3'-UTRs.

They exhibit approximately 90% identity in protein-coding exons (Figure 3.20). The exon 3, which encode DBD, is identical in size (134 bp) and exhibit approximately 90% identity when compared to human (Figure 3.20). The LBD-encoding exons (exons 5-9) are also conserved both in size and sequence identity but slightly lower than the DBD-encoding exons (vary from 82% identity in exon 4 to 97% identity in exon 7). The exon 1 of human is about 2 kb and exhibits only 9% identity. In contrast to exon 1, the 3'-UTR encoding exon 9 is relative in size and exhibits 60% identity.

Table 3.9. Exon-intron structure of porcine *PXR* transcript.

Exon	Porcine *PXR*	Porcine genomic DNA (DQ531715.1)	Size (bp)	Identity	5' end of exon	3' end of exon	Intron	Intron size	5' splice site	3' splice site
1	1-207	-	207	-	AACCGATGAG...	...ACCACCAAGC	1	-	-	-
2	208-428	650-870	221	100%	AGTCCAAGAG...	...GCTTTTTCAG	2	2456	gtagagttat...	...cctctggcag
3	429-562	3327-3460	134	100%	GAGGGCCATC...	...AGGAAGGAAA	3	1296	gtgagcagca...	...ttccctgcag
4	563-750	4757-4944	188	99%	TGATCATGTC...	...GAATTTTCGG	4	948	gtatgaggtt...	...cctgtcccag
5	751-1028	5893-6170	278	100%	TTACCAGAGG...	...CCTACTTCAG	5	2034	gtaggacaca...	...tctttgccag
6	1029-1171	8205-8347	143	100%	GGACTTGCCC...	...GACCCCTCAG	6	182	gtgccctggc...	...ccacccacag
7	1172-1288	8530-8646	117	100%	GTGGCTTCCA...	...TTCTCTCCAG	7	352	gtgagagtcc...	...ttcacaccag
8	1289-1394	8999-9104	106	100%	ACCGCCCGGG...	...CTGCCCACCG	8	1147	gtgagcagca...	...tggactgcag
9	1395-2140	10252- 10997	745	100%	ATTCCTGTTC...	...TTTCAAAGCT				

Note: The exon-intron boundaries were predicted by aligning porcine *PXR* cDNA with porcine genomic DNA sequence (DQ531175, 10997 bp DNA) using SIM4 program. The poly(A) tail was excluded in the alignment. The 5'- and 3'-ends of each exon are shown in uppercase letters. The 5' and 3' splice sites of each intron are shown in lowercase letters. All splice sites are in consensus with GT-AG rule for splicing. The sequence identity between porcine PXR transcript and porcine genomic DNA was derived from alignment results of SIM4. The 207-bp sequence of porcine *PXR* could not align with DQ531715.
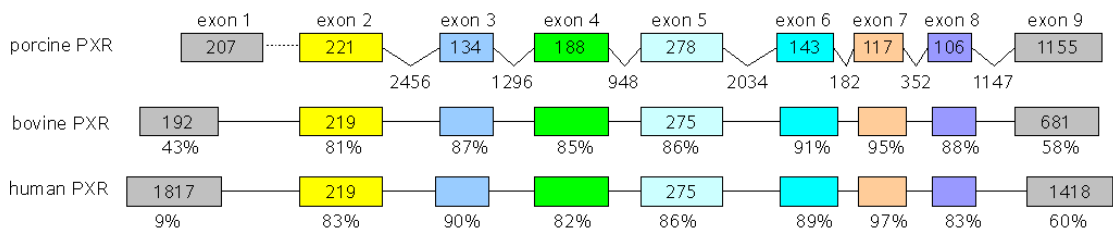
Figure 3.20. Comparison of *PXR* exons in farm animals and human.

Boxes represent exons. Exons with similar size are colored. The percent identity derived from pairwise alignment using BioEdit is shown below each exon. The size of exon is shown in the exon box. The thin lines indicate introns. The length of intron is shown only for porcine *PXR* transcript.

3.6. Discussion

The nuclear receptors CAR and PXR are related nuclear receptors that have been known as transcriptional regulators of target genes involved in the metabolism of both endogenous and exogenous substances (Honkakoski and Negishi, 2000; Wei *et al.,* 2000; Kliewer *et al.,* 2002; Timsit and Negishi, 2007). Since, full-length cDNA sequences of porcine *CAR* and *PXR* have not yet been reported. We identified clones encoding *CAR* and *PXR* deposited in PEDE database (http://pede.dna.affrc.go.jp/) and sequenced full-length of these cDNA clones. The nucleotide sequences of porcine *CAR* and *PXR* have been submitted to GenBank with accession nos. AB214979 and AB214980, respectively (Thadtha *et al.,* 2005).

The 1407-bp full-length of porcine *CAR* contains a 5'-UTR of 216 bp, a 1047-bp open reading frame encoding 348 amino acid residues, and a 3'-UTR of 144 bp. The full-length cDNA of porcine *PXR* is composed of 2567 bp, which has 273 bp of 5'-UTR, 1266 bp of ORF encoding 421 amino acid residues, and 1028 bp of 3'-UTR (Figures 3.1 and 3.5, and Tables 3.1 and 3.4). Porcine *CAR* and *PXR* share 84% and 46% identity in nucleotide sequence, and 84% and 86% identity in amino acid sequence with human, respectively (Tables 3.1, 3.4, 3.6, and 3.7).

59

Transcriptional regulation by the nuclear receptors CAR and PXR is related to the DBD, which is responsible for the recognition of specific DNA sequences in the response element of target genes, and LBD, which is responsible for the binding of receptor with ligands, interaction with coregulators, transactivation functions and a dimerization with RXR (Mangelsdorf *et al.,* 1995; Giguere, 1999; Khorasanizadeh and Rastinejad, 2001; Robinson-Rechavi *et al.,* 2003). By using SMART analysis tool, the CAR and PXR proteins were predicted to contain these functional domains (Figures 3.1 and 3.5). In the porcine CAR receptor, the first and second zinc finger motifs were predicted at positions 11 to 31 (CAVCGDRATGYHFHALTCEGC) and 47 to 66 (CPFAGSCKVNKAQRRHCPAC), respectively (Figure 3.10). It also contains the P-box (amino acid residues 28-35: CEGCKGFF) and AF-2 (amino acid residues 341-348: PLLQEICS) (Figures 3.10 and 3.11). The porcine PXR receptor was predicted to contain both the first and second zinc finger motifs at positions 27 to 47 (CRVCGDKATGYHFNVMTCEGC) and positions 63 to 88 (CLFRKGACEITRKTRRQCQACRLRKC), respectively (Figure 3.15). It also possesses the P-box (amino acid residues 44-51: CEGCKGFF) and AF-2 (amino acid residues 410-417: PLMQELFS) (Figures 3.15 and 3.16). Evolutionary studies of nuclear receptor superfamily showed that closely related nuclear receptors share a high degree of sequence identity in the conserved DBD and moderately conserved LBD. In general, evolutionarily related nuclear receptors share at least 80%-90% identity in the DBD and at least 40%-60% identity in the LBD (Nuclear Receptors Nomenclature Committee, 1999). Porcine CAR and PXR receptors are closely related to those of humans with high sequence identities in their DBDs and LBDs. The DBD and LBD of porcine CAR and PXR receptors share 83% and 87% and 95% and 93% identity with human, respectively (Figure 3.9 and 3.14). With high sequence identity

in the DBD and LBD, we conclude that porcine CAR and PXR receptors are orthologs of human CAR and PXR receptors.

The exon-intron structure of porcine *CAR* gene was determined by aligning its cDNA sequence with bovine genomic DNA (chromosome 3: 7567938-7572817), whereas porcine *PXR* gene was aligned with pig genomic DNA (Pollock *et al.,* 2007; DQ531175). The *CAR* gene is located on chromosome 1 in human and mouse, while *PXR* gene is located on chromosomes 1, 3, and 16 in cattle, human, and mouse, respectively (Zhang *et al.,* 2001). Comparison of *CAR* and *PXR* exons revealed that the size and number of exons are conserved between farm animals and humans, especially in the protein-encoding exons (Figure 3.19 and 3.20). Both porcine *CAR* and *PXR* genes were predicted to contain 9 exons separated by 8 introns. Exons 2 and 3 of porcine *CAR* gene, encoding DBD, are identical in size (140 bp and 131 bp, respectively) and exhibit ~90% identity when compared to cattle and human (Figure 3.19). The LBD-encoding exons (exons 5, 6, 7, 8, and 9) are also conserved both in size and sequence identity but slightly lower than DBD encoding exons (~85-90% identity). In porcine *PXR* gene, the DBD- and LBD-encoding exons (exons 2 to 3 and exons 5-9, respectively) also exhibit high sequence identity with cattle and human (approximately 90%) (Figure 3.20).

In humans and mice, *CAR* and *PXR* are mainly expressed in liver and intestine (Baes *et al.,* 1994; Choi *et al.,* 1997; Bertilsson *et al.,* 1998; Blumberg *et al.,* 1998; Kliewer *et al.,* 1998; Lehmann *et al.,* 1998). Liver is the major expression site of steroid and xenobiotic-metabolizing enzymes such as CYP3A4 and CYP2B. The CYP2B and CYP3A enzymes are responsible for the metabolism of steroid and thyroid hormones as well as drugs (Honkakoski and Negishi, 2000). The expression analysis of the porcine *CAR* and *PXR* genes revealed that they are expressed in liver,

kidney, and small intestine. We detected the expression of porcine *CAR* and *PXR* mRNAs in kidney. Although, *CAR* expression was detected in human and mouse kidney (Baes *et al.,* 1994; Choi *et al.,* 1997), the expression of *PXR* was not detected in human kidney using Northern analysis (Bertilsson *et al.,* 1998; Blumberg *et al.,* 1998; Lehmann *et al.,* 1998). However, weaker expression of *PXR* was found in mouse kidney (Kliewer *et al.,* 1998). These differences in tissue expression pattern among different species may indicate the species-specific expression of *CAR* and *PXR*. In addition, gene exression level summarized in UniGene database (http://www.ncbi.nlm.nih.gov/sites/entrez?db=unigene) revealed that *CAR* and *PXR* are primarily expressed in liver and intestine with difference in tissue-expression pattern in different animals. For example, *PXR* expression was found in porcine kidney (high expression level) and in bovine and human testis. Similary, the expression of *CAR* was found in porcine ovary, in mouse and human brain, and in bovine brain and testis (UniGene ID: 3009521 and 454836, 1483484 and 1413571, 193733 and 132093, 257008 and 258735, and 2939347 for CAR and PXR in swine, cattle, human, mouse, and chicken, respectively).

Together with both sequence and mRNA expression analysis, we conclude that porcine *CAR* and *PXR* are orthologous genes of human *CAR* and *PXR* and may possess similar function as found in humans.

CHAPTER 4

THE NUCLEAR RECEPTOR LXRα

4.1. Introduction

The nuclear receptor liver X receptor alpha (*LXRα*; NR1H3) is a ligand activated transcription factor that plays an important role in the regulation of genes involved in cholesterol homeostasis (Peet *et al.,* 1998b; Lu *et al.,* 2001), fatty acid biosynthesis (Repa *et al.,* 2000; Schultz *et al.,* 2000; Yoshikawa *et al.,* 2001), and glucose metabolism (Dalen *et al.,* 2003; Laffitte *et al.,* 2003). LXRα is activated by naturally occurring oxysterols, which are the cholesterol metabolites in the cholesterol biosynthetic pathway including 22 (R)-hydroxycholesterol, 24(S),25-epoxycholesterol, 24(S)-hydroxycholesterol (Lehmann *et al.,* 1997; Janowski *et al.,* 1999), and a synthetic compound T0314407, which is a selective agonist of LXRs (Schultz *et al.,* 2000). Studies in mice and rats showed that LXRα activates the expression of the cytochrome P-450 7A1 gene (CYP7A1) by binding to LXR response element (LXRE) located in the promoter region, which controls the metabolism of cholesterol into bile acids (Lehmann *et al.,* 1997; Peet *et al.,* 1998b). LXRα also controls the expression of the genes of the ATP-binding cassette (ABC) family, for example the gene ABCA1 in macrophages, which is responsible for the transport of cholesterol from peripheral tissues back to the liver for catabolism to bile acids (Venkateswaran *et al.,* 2000). Moreover, several reports indicate that LXRα functions

as a transcriptional regulator of the sterol regulatory element-binding protein 1c (SREBP-1c; Repa *et al.,* 2000; Schultz *et al.,* 2000; DeBose-Boyd *et al.,* 2001; Yoshikawa *et al.,* 2001) and the peroxisome proliferator-activated receptor gamma (PPARγ; Seo *et al.,* 2004).

Transcript variants have been identified in many members of the nuclear receptor subfamily 1 (Auerbach *et al.,* 2003; Lamba *et al.,* 2004; Kurose *et al.,* 2005). The different receptor isoforms have been found to play distinct biological roles (Robinson-Rechavi *et al.,* 2003). Transcript variants of human *LXRα* gene have been reported and are probably generated by alternative splicing and alternative promoter usage (Chen *et al.,* 2005). We identified three transcript variants of porcine *LXRα* and analyzed the mRNA expression pattern of these *LXRα* transcripts in porcine tissues.

4.2. Nucleotide Sequences of Porcine *LXRα*

The three full-length cDNAs of porcine *LXRα* were identified and sequenced from three PEDE clones using the primer walking sequencing method. We refer to these porcine *LXRα* transcripts as *LXRα-1*, *LXRα-2*, and *LXRα-3*. The porcine *LXRα-1* transcript contains a 5'-UTR of 183 bp, an open reading frame of 1344 bp, and a 3'-UTR of 152 bp (Figure 4.1). The porcine *LXRα-2* transcript is composed of a 5'-UTR of 216 bp, an open reading frame of 1344 bp, and a 3'-UTR of 138 bp (Figure 4.2). The porcine *LXRα-3* cDNA has 648 bp of 5'-UTR, 315 bp of ORF, and 359 bp of 3'-UTR (Figure 4.3). Porcine *LXRα-1* and *LXRα-2* are identical in ORFs and 3'–UTRs, but differ in the 5'-UTRs (between nucleotide positions 1-145 and 1-178) (Figures 4.4 and 4.5). The porcine *LXRα-3* represents a truncated transcript that has in-frame stop codon (TGA at nucleotide positions 961-963), resulting in LXRα protein isoform lacking the hinge D and LBD (Figure 4.3 and Table 4.4). The basic nucleotide

sequence statistics of porcine *LXRα-1* and *-2* is summarized in Table 4.1. The full-length cDNA sequences of porcine *LXRα* have been submitted to the DDBJ/EMBL/GenBank databases under the accession numbers AB254405 and AB254406 (Thadtha *et al.,* 2006).

Analysis of nucleotide sequences revealed that full-length porcine *LXRα-1* shares 93% identity with porcine *LXRα-2*, 39% identity with *LXRα-3*, 89% identity with bovine *LXRα*, 80% identity with human *LXRα*, and 34% identity with chicken *LXRα*.

```
   1 AGTCGGTCGGCTCCCGGCTCCTGGCCCCTCTCTCGCTCAGCTCCGGCTCCCTGGTGGCGG
  61 CGGACAAGGGACTGCACCATCCCCCTCTCCCAGCAAGGGGGCTCCAGAGACTGCCCAGCC
 121 AGGAAGTCTGGCGGCCTGGGCGCTCGGACAGTCCCTTGGTAATGTCCAGGGCTCCAGGAA
 181 GAGATGTCCTTGTGGGTGGAGGCCCCTGTGCCTGATGTTTCTCCTGACTCTGCAGTGGAG
      M  S  L  W  V  E  A  P  V  P  D  V  S  P  D  S  A  V  E  19
 241 CTGTGGGAGTCAGATGCACAAGATGCAAGCAGCCAGTCTCTGGGAAGCAGCAGCTGCATC
      L  W  E  S  D  A  Q  D  A  S  S  Q  S  L  G  S  S  S  C  I  39
 301 CTCAGGGAGGAATCCAGCACACCCCAGTCTGCGGGGGGGCGCTTCGAGGGTGGGGCTGGAT
      L  R  E  E  S  S  T  P  Q  S  A  G  G  A  S  R  V  G  L  D  59
 361 GCAACAGAGTCCACGGCCCTGCTTCCCGGGGTGGAGGCCTCTCCAGAGTCCACAGAGCTC
      A  T  E  S  T  A  L  L  P  G  V  E  A  S  P  E  S  T  E  L  79
 421 CGTCCACAAAAGCGGAAAAAGGGGCCAGCCCCCAAAATGCTGGGGAATGAGCTGTGCAGT
      R  P  Q  K  R  K  K  G  P  A  P  K  M  L  G  N  E  L  C  S  99
 481 GTGTGTGGGGACAAGGCCTCCGGCTTCCACTACAACGTGCTGAGCTGCGAGGGCTGCAAG
      V  C  G  D  K  A  S  G  F  H  Y  N  V  L  S  C  E  G  C  K  119
 541 GGATTCTTCCGTCGCAGTGTCATCAAAGGGGCTCGCTATGTCTGCCACAGCGGGGGGCAC
      G  F  F  R  R  S  V  I  K  G  A  R  Y  V  C  H  S  G  G  H  139
 601 TGCCCCATGGACACCTACATGCGTCGCAAGTGCCAGGAGTGCCGTCTTCGCAAGTGCCGC
      C  P  M  D  T  Y  M  R  R  K  C  Q  E  C  R  L  R  K  C  R  159
 661 CAGGCGGGCATGCGAGAGGAGTGTGTCCTGTCAGAAGAACAGATCCGCCTGAAGAAACTG
      Q  A  G  M  R  E  E  C  V  L  S  E  E  Q  I  R  L  K  K  L  179
 721 AAGCGGCAAGAGGAGGAACAGGCTCAGGCCACATCTGTGCCCCCAAGGGCTTCCTCGCCG
      K  R  Q  E  E  E  Q  A  Q  A  T  S  V  P  P  R  A  S  S  P  199
 781 CCCCAAGTCCTGCCCCAGCTTAGCCCAGAGCAGCTGGGCATGATCGAGAAGCTGGTGGCT
      P  Q  V  L  P  Q  L  S  P  E  Q  L  G  M  I  E  K  L  V  A  219
 841 GCCCAGCAGCAGTGTAACAGACGCTCCTTTTCAGACCAGCTTCGAGTCACGCCTTGGCCC
      A  Q  Q  Q  C  N  R  R  S  F  S  D  Q  L  R  V  T  P  W  P  239
 901 ATGGCACCAGATCCCCAGAGCCGGGAGGCCCGTCAGCAACGCTTTGCCCACTTCACTGAG
      M  A  P  D  P  Q  S  R  E  A  R  Q  Q  R  F  A  H  F  T  E  259
 961 CTGGCCATCGTCTCTGTGCAGGAGATCGTTGATTTTGCCAAACAGCTGCCAGGCTTCTTG
      L  A  I  V  S  V  Q  E  I  V  D  F  A  K  Q  L  P  G  F  L  279
1021 CAGCTCAGCCGGGAGGACCAGATCGCCCTCCTAAAGACCTCTGCGATTGAGGTGATGCTT
      Q  L  S  R  E  D  Q  I  A  L  L  K  T  S  A  I  E  V  M  L  299
1081 CTGGAGACATCTCGGAGGTACAACCCTGGGAGTGAGAGTATCACCTTCCTCAAGGATTTC
      L  E  T  S  R  R  Y  N  P  G  S  E  S  I  T  F  L  K  D  F  319
1141 AGTTATAATCGGGAAGACTTTGCCAAAGCAGGGCTGCAGGTGGAGTTCATCAACCCTATC
      S  Y  N  R  E  D  F  A  K  A  G  L  Q  V  E  F  I  N  P  I  339
1201 TTCGAGTTCTCCAGAGCCATGAATGAGCTGCAACTAAATGATGCTGAGTTTGCCCTGCTC
      F  E  F  S  R  A  M  N  E  L  Q  L  N  D  A  E  F  A  L  L  359
1261 ATTGCCATCAGCATCTTCTCTGCAGACCGGCCCAACGTGCAGGACCAGCTCCAGGTAGAG
      I  A  I  S  I  F  S  A  D  R  P  N  V  Q  D  Q  L  Q  V  E  379
1321 AGGCTGCAACATACATATGTGGAGGCCCTGCATGCCTACGTCTCCATCCACCACCCCCAT
      R  L  Q  H  T  Y  V  E  A  L  H  A  Y  V  S  I  H  H  P  H  399
1381 GACCGACTGATGTTCCCACGGATGCTAATGAAACTGGTGGGCCTCCGGACACTGAGCAGC
      D  R  L  M  F  P  R  M  L  M  K  L  V  G  L  R  T  L  S  S  419
1441 GTCCACTCAGAGCAAGTGTTTGCACTGCGCCTGCAGGATAAAAAGCTTCCCCCGCTGCTC
```

```
     V  H  S  E  Q  V  F  A  L  R  L  Q  D  K  K  L  P  P  L  L  439
1501 TCTGAGATCTGGGATGTGCACGAGTGACTGTTCTTCCCCCACACCGTGTCTTCTGTTTTT
     S  E  I  W  D  V  H  E                                      447
1561 TGGGGTAGATGGCTGAGGCGTGACTGAGAAGGGCAGACATTCCTGAGGGCTGGGCAAAGG
1621 ACATCCTCACGTGGCATTAAAAGAGTCAAAGGGTTGGGAAAAAAAAAAAAAAAAAAAAAA 1679
```

Figure 4.1. Nucleotide and deduced amino acid sequences of porcine *LXRα-1*.

The numbers on the left and right indicate the nucleotide base and the deduced amino acid in porcine *LXRα-1*, respectively. The ORF was deduced using the ORF Finder. The deduced amino acid sequence is shown in single-letter code below the nucleotide sequence. The ATG start codon (positions 184-186) and the TGA stop codon (positions 1525-1527) are printed in bold. The predicted DNA-binding domain (amino acid residues 95-166) is boxed. The putative ligand-binding domain of porcine LXRα-1 is underlined. The polyadenylation signal (ATTAAA) is underlined.

```
   1 ACAAGAGGCGGCTTGCCTGGACCTCCCGGGTAGGAATGGGGTCCAGGCACGGTTGCACAT
  61 GGCCTGGTCACCTAGGCCTGCTGAGGTGTGGGAGGAGGACAGTCATACCGGGGACCCTGA
 121 GCCAGCTCACCTCACAGAAGAGTGTGTTCATAGACATTCCAGACGGTCTCCAGTGTCAGG
 181 ACAGTCCCTTGGTAATGTCCAGGGCTCCAGGAAGAGATGTCCTTGTGGGTGGAGGCCCCT
                                              M  S  L  W  V  E  A  P  8
 241 GTGCCTGATGTTTCTCCTGACTCTGCAGTGGAGCTGTGGGAGTCAGATGCACAAGATGCA
      V  P  D  V  S  P  D  S  A  V  E  L  W  E  S  D  A  Q  D  A  28
 301 AGCAGCCAGTCTCTGGGAAGCAGCAGCTGCATCCTCAGGGAGGAATCCAGCACACCCCAG
      S  S  Q  S  L  G  S  S  S  C  I  L  R  E  E  S  S  T  P  Q  48
 361 TCTGCGGGGGGCGCTTCGAGGGTGGGGCTGGATGCAACAGAGTCCACGGCCCTGCTTCCC
      S  A  G  G  A  S  R  V  G  L  D  A  T  E  S  T  A  L  L  P  68
 421 GGGGTGGAGGCCTCTCCAGAGTCCACAGAGCTCCGTCCACAAAAGCGGAAAAAGGGGCCA
      G  V  E  A  S  P  E  S  T  E  L  R  P  Q  K  R  K  K  G  P  88
 481 GCCCCCAAAATGCTGGGGAATGAGCTGTGCAGTGTGTGTGGGGACAAGGCCTCCGGCTTC
      A  P  K  M  L  G  N  E  L  C  S  V  C  G  D  K  A  S  G  F  108
 541 CACTACAACGTGCTGAGCTGCGAGGGCTGCAAGGGATTCTTCCGTCGCAGTGTCATCAAA
      H  Y  N  V  L  S  C  E  G  C  K  G  F  F  R  R  S  V  I  K  128
 601 GGGGCTCGCTATGTCTGCCACAGCGGGGGCCACTGCCCCATGGACACCTACATGCGTCGC
      G  A  R  Y  V  C  H  S  G  G  H  C  P  M  D  T  Y  M  R  R  148
 661 AAGTGCCAGGAGTGCCGTCTTCGCAAGTGCCGCCAGGCGGGCATGCGAGAGGAGTGTGTC
      K  C  Q  E  C  R  L  R  K  C  R  Q  A  G  M  R  E  E  C  V  168
 721 CTGTCAGAAGAACAGATCCGCCTGAAGAAACTGAAGCGGCAAGAGGAGGAACAGGCTCAG
      L  S  E  E  Q  I  R  L  K  K  L  K  R  Q  E  E  E  Q  A  Q  188
 781 GCCACATCTGTGCCCCCAAGGGCTTCCTCGCCGCCCCAAGTCCTGCCCCAGCTTAGCCCA
      A  T  S  V  P  P  R  A  S  S  P  P  Q  V  L  P  Q  L  S  P  208
 841 GAGCAGCTGGGCATGATCGAGAAGCTGGTGGCTGCCCAGCAGCAGTGTAACAGACGCTCC
      E  Q  L  G  M  I  E  K  L  V  A  A  Q  Q  Q  C  N  R  R  S  228
 901 TTTTCAGACCAGCTTCGAGTCACGCCTTGGCCCATGGCACCAGATCCCCAGAGCCGGGAG
      F  S  D  Q  L  R  V  T  P  W  P  M  A  P  D  P  Q  S  R  E  248
 961 GCCCGTCAGCAACGCTTTGCCCACTTCACTGAGCTGGCCATCGTCTCTGTGCAGGAGATC
      A  R  Q  Q  R  F  A  H  F  T  E  L  A  I  V  S  V  Q  E  I  268
1021 GTTGATTTTGCCAAACAGCTGCCAGGCTTCTTGCAGCTCAGCCGGGAGGACCAGATCGCC
      V  D  F  A  K  Q  L  P  G  F  L  Q  L  S  R  E  D  Q  I  A  288
1081 CTCCTAAAGACCTCTGCGATTGAGGTGATGCTTCTGGAGACATCTCGGAGGTACAACCCT
      L  L  K  T  S  A  I  E  V  M  L  L  E  T  S  R  R  Y  N  P  308
1141 GGGAGTGAGAGTATCACCTTCCTCAAGGATTTCAGTTATAATCGGGAAGACTTTGCCAAA
      G  S  E  S  I  T  F  L  K  D  F  S  Y  N  R  E  D  F  A  K  328
1201 GCAGGGCTGCAGGTGGAGTTCATCAACCCTATCTTCGAGTTCTCCAGAGCCATGAATGAG
      A  G  L  Q  V  E  F  I  N  P  I  F  E  F  S  R  A  M  N  E  348
1261 CTGCAACTAAATGATGCTGAGTTTGCCCTGCTCATTGCCATCAGCATCTTCTCTGCAGAC
      L  Q  L  N  D  A  E  F  A  L  L  I  A  I  S  I  F  S  A  D  368
1321 CGGCCCAACGTGCAGGACCAGCTCCAGGTAGAGAGGCTGCAACATACATATGTGGAGGCC
      R  P  N  V  Q  D  Q  L  Q  V  E  R  L  Q  H  T  Y  V  E  A  388
1381 CTGCATGCCTACGTCTCCATCCACCACCCCCATGACCGACTGATGTTCCCACGGATGCTA
      L  H  A  Y  V  S  I  H  H  P  H  D  R  L  M  F  P  R  M  L  408
1441 ATGAAACTGGTGAGCCTCCGGACACTGAGCAGCGTCCACTCAGAGCAAGTGTTTGCACTG
      M  K  L  V  S  L  R  T  L  S  S  V  H  S  E  Q  V  F  A  L  428
```

68

```
1501  CGCCTGCAGGATAAAAAGCTTCCCCCGCTGCTCTCTGAGATCTGGGATGTGCACGAGTGA
      R   L   Q   D   K   K   L   P   P   L   L   S   E   I   W   D   V   H   E      447
1561  CTGTTCTTCCCCCACACCGTGTCTTCTGTTTTTTGGGGTAGATGGCTGAGGCGTGACTGA
1621  GAAGGGCAGACATTCCTGAGGGCTGGGCAAAGGACATCCTCACGTGGCATTAAAAGAGTC
1681  AAAGGGTTGGGAAAAAAA  1698
```

Figure 4.2. Nucleotide and deduced amino acid sequences of porcine *LXRα-2*.

Note: The numbers on the left and right indicate the nucleotide base and the deduced amino acid in porcine *LXRα-2*, respectively. The ORF was deduced using the ORF Finder. The deduced amino acid sequence is shown in single-letter code below the nucleotide sequence. The ATG start codon (positions 217-219) and the TGA stop codon (positions 1558-1560) are printed in bold. The predicted DNA-binding domain (amino acid residues 95-166) is boxed. The putative ligand-binding domain of porcine LXRα-2 is underlined. The polyadenylation signal (ATTAAA) is underlined.

```
   1 TCCCAAGAGAATGCTGGGGAGATGAACTGGGACCTGAGACCCCCCCCGTGCCTCTCTTTT
  61 GGGGGTCAGACTCTGCAGTGGAGCTGTGGGAGTCAGATGCACAAGATGCAAGCAGCCAGT
 121 CTCTGGGAAGCAGCAGCTGCATCCTCAGGGAGGAATCCAGCACACCCCAGTCTGTCGGGG
 181 GCGCTTCGAGGGTGGGGCTGGATGCAACAGAGTCCACGGCCCTGCTTCCCGGGGTGGAGG
 241 CCTCTCCAGAGTCCACAGGTGAGGAGCTGCTGGATTTGGAGGAGGTAGGGGCTCCCCCGA
 301 TCTGGGAGGGGTAACTTCAGGGCTTATAAGATCATGGTTAAGTTCAAAGGCTTTGGGGTC
 361 AACTAGACCTAAGTGATCTGAATGTAAGTTTTGTTTCTATCTTAGGCTTTTTTGAGCCTC
 421 AGTTCCTTAATGGAAATAATAAGCGTACCTATCTCTTGTAAGGACCAGTGCATATAGAG
 481 GGTTCAGACAGTACCTAGGACATAGTAGCACTTGGCATACAGGAGCTACTTGCTTAGAAT
 541 AGGGAGACTGGGACTGCATCTAATGAAGTGAATGAAGGTCACTGAGTGCTCAGAGCAGTG
 601 GCCAACTGGGCACCAAGCTTTTTCTCTGAAGAACATTAAGCCTCTTCC**ATG**TATCCAGAG
                                                        M   Y   P   E     4
 661 CTCCGTCCACAAAAGCGGAAAAAGGGGCCAGCCCCCAAAATGCTGGGGAATGAGCTGTGC
      L   R   P   Q   K   R   K   K   G   P   A   P   K   M   L   G  |N   E   L   C|    24
 721 AGTGTGTGTGGGGACAAGGCCTCCGGCTTCCACTACAACGTGCTGAGCTGCGAGGGCTGC
     |S   V   C   G   D   K   A   S   G   F   H   Y   N   V   L   S   C   E   G   C|    44
 781 AAGGGATTCTTCCGTCGCAGTGTCATCAAAGGGGCTCGCTATGTCTGCCACAGCGGGGGC
     |K   G   F   F   R   R   S   V   I   K   G   A   R   Y   V   C   H   S   G   G|    64
 841 CACTGCCCCATGGACACCTACATGCGTCGCAAGTGCCAGGAGTGCCGTCTTCGCAAGTGC
     |H   C   P   M   D   T   Y   M   R   R   K   C   Q   E   C   R   L   R   K   C|    84
 901 CGCCAGGCGGGCATGCGAGAGGAGTGTGAGTGTCTGGGAACCAGAGCTGGGGGGATAAGC
     |R   Q   A   G   M   R   E   E|  C   E   C   L   G   T   R   A   G   G   I   S    104
 961 **TGA**TAGGAAAGAAGGAATATGTGGGCGCCAGGATCTGAGGCCAGACAGGCAGCTGAGCCA
1021 ACAGGGCCTTTGTCTACTTCCCTAAGTGTGGATTAAGATCTCTTCCTTGTTGGGGGGTTCC
1081 CGTCGTGGTGCAGTGGTTAATGAATCCAAATAGGAACCATAAGGTTGCGGGTTCGATCCC
1141 TGGCCTTGCTAAGTGTGTTAAGGATCTGGCGTTGGTGAGCTGTGGTGTAGTCGCATATGC
1201 CGCTCGGATCCCACGTTGCCGTGGCTCTGGTGTAGCCAGTGACCACAGCTCTGTTTAGAC
1261 TCCTAACTTGGAACCTCCACATGCCGCGTGACCAGCCCTAAAAAGACAAAAAAAAAAAAA
1321 AA 1322
```

Figure 4.3. Nucleotide and deduced amino acid sequences of porcine *LXRα-3*.

The numbers on the left and right indicate the nucleotide base and the deduced amino acid in porcine *LXRα-3*, respectively. The ORF was deduced using the ORF Finder. The deduced amino acid sequence is shown in single-letter code below the nucleotide sequence. The ATG start codon (positions 649-651) and the TGA stop codon (positions 961-963) are printed in bold. The predicted DNA-binding domain (amino acid residues 21-92) is boxed. The truncated porcine LXRα-3 isoform lacks the D and LBD. The polyadenylation signal was not found in the 3'-UTR of porcine *LXRα-3* transcript.
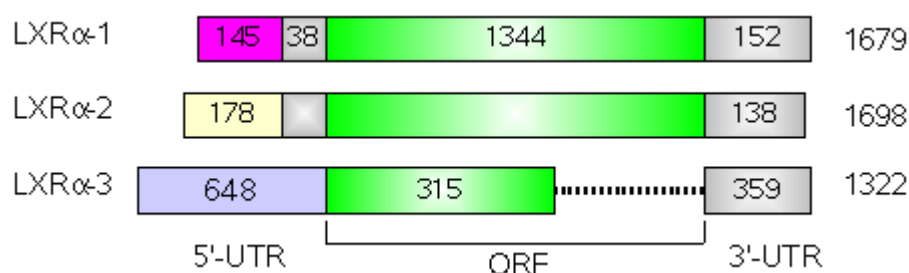
Figure 4.4. Comparison of transcripts in porcine *LXRα*.

Three porcine *LXRα* transcripts were identified and sequenced, *LXRα-1*, *LXRα-2*, and *LXRα-3*. Porcine *LXRα-1* and *LXRα-2* are identical in ORFs and 3'–UTRs, but differ in the 5'-UTRs (between nucleotide positions 1-145 and 1-178). The porcine *LXRα-3* represents a truncated transcript that has an in-frame stop codon.

Table 4.1. Nucleotide composition of porcine *LXRα* transcripts.

Sequence	Purine base			Pyrimidine base			Total (bp)
	C	G	C+G	A	T	A+T	
Full-length *LXRα-1* cDNA	490	494	58.6%	368	327	41.4%	1679
5'-UTR	67	59	68.9%	27	30	31.1%	183
CDS	395	393	58.6%	292	264	41.4%	1344
3'-UTR	28	42	46%	49	33	54%	152
Full-length *LXRα-2* cDNA	480	506	58%	375	337	42%	1698
5'-UTR	57	72	59.7%	47	40	40.3%	216
CDS	395	392	58.6%	293	264	41.4%	1344
3'-UTR	28	42	50.7%	35	33	49.3%	138

Note: The nucleotide composition of porcine *LXRα* transcripts was calculated using the BioEdit.

4.2.1. 5'-Untranslated Region

The 5'-UTR of porcine *LXRα-1*, *-2*, and *-3* contains 183 bases (68.9% G+C content), 216 bases (59.7% G+C content), and 648 bases (52.3% G+C content), respectively, (Table 4.1). Porcine *LXRα-1* and *-2* differ in the nucleotide regions of 5'-UTR (1-145 in *LXRα-1* and 1-178 in *LXRα-2*) (Figure 4.4 and 4.5). The porcine *LXRα-3* has a longer 5'-UTR. The analysis of 5'-UTR of porcine *LXRα* transcripts did not reveal the presence of known regulatory elements (e.g., upstream open reading

frame (uORF), internal ribosome entry site (IRES)) involved in post-transcriptional regulation such as mRNA localization, stability, and translational efficiency (Pesole *et al.,* 2001; Hughes, 2006). The secondary structures of the 5'-UTR of these porcine *LXRα* transcripts were predicted using web interface of the RNAfold program (Hofacker, 2003). The predicted fold energy of 5'-UTR is -80.3 kcal/mol for porcine *LXRα-1*, -89.2 kcal/mol for porcine *LXRα-2*. The predicted secondary structures of porcine *LXRα* 5'-UTRs are shown in Figures 4.6 and 4.7.

```
LXRα-1   ----------------------------AGTCGG--TCGGCTCCCGGCTCC-TGGCCCC 28
LXRα-2   ACAAGAGGCGGCTTGCCTGGACCTCCCGG-..A..AA.G..G...A...A.GG.T..A.A 59

LXRα-1   TCTCTCGCTCAGCT---CCGGCTCCCTGGTGGC-GGCGGACA-----A---GGGAC--TG 74
LXRα-2   .GG.CT.G...C..AGG..T...GAGGT....GA..A.....GTCAT.CCG.....CC.. 119

LXRα-1   CACCATCCCCCTCTCCCAGCAAGGGGGCTCCAGAGACTGCCCAGCCAGGAAGTCTGGCGG 134
LXRα-2   AG...G.T.A.-...A...A.GA.T.TG.T..T....ATT....A.-..---...----- 169

LXRα-1   CCTGGGCGCTC-GGACAGTCCCTTGGTAATGTCCAGGGCTCCAGGAAGAGATGTCCTTGT 193
LXRα-2   ..A.T.---..A................................................ 226
```

Figure 4.5. Alignment of different 5'-UTRs between porcine *LXRα-1* and *LXRα-2*.

Porcine *LXRα-1* and *-2* differ in the nucleotide regions of 5'-UTR (1-145 in *LXRα-1* and 1-178 in *LXRα-2*).

Figure 4.6. Predicted secondary structure of 5'-UTR of porcine *LXRα-1*.

The secondary structure of 5'-UTR was predicted using the RNAfold program. The predicted fold energy of porcine *LXRα-1* 5'-UTR is -80.3 kcal/mol.
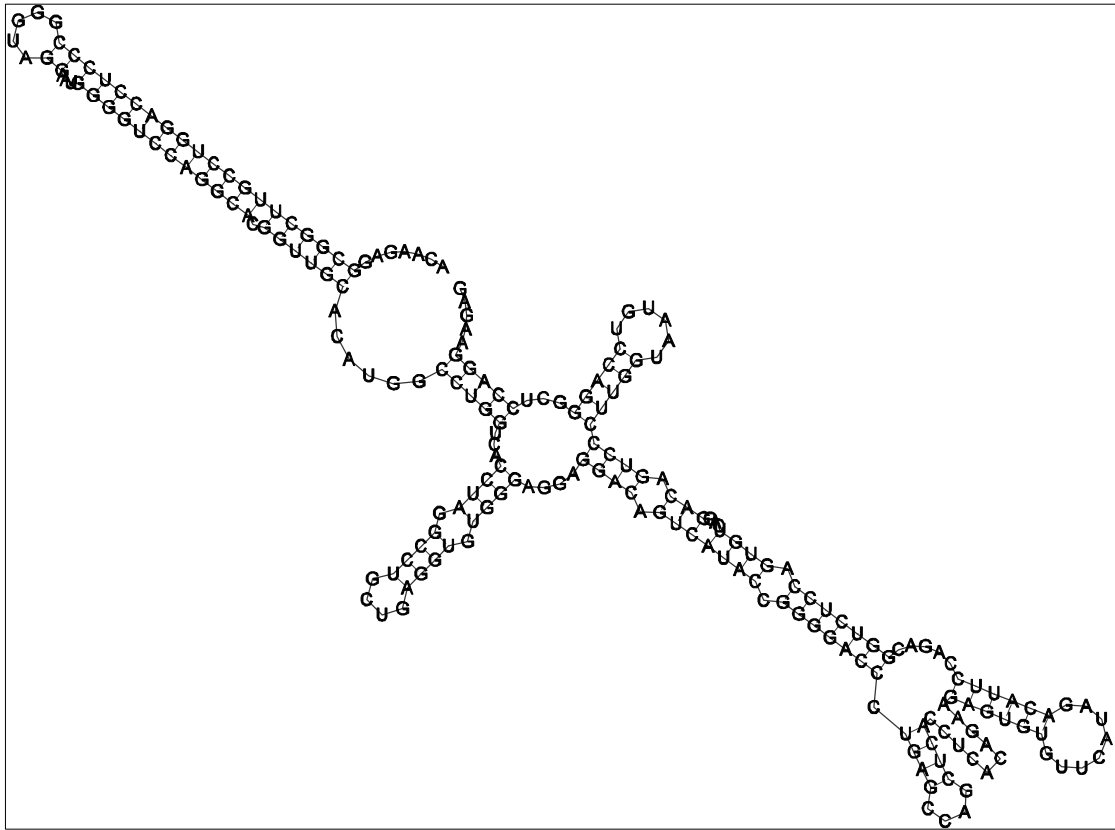
Figure 4.7. Predicted secondary structure of 5'-UTR of porcine *LXRα-2*.

The secondary structure of 5'-UTR was predicted using the RNAfold program. The predicted fold energy of porcine *LXRα-2* 5'-UTR is -89.2 kcal/mol.

4.2.2. Open Reading Frame

The open reading frame (ORF) of porcine *LXRα* cDNAs was deduced using ORF Finder. The longest ORFs were selected for these *LXRα* cDNAs. The ORF for *LXRα-1*, *LXRα-2*, and *LXRα-3* is in frame +1. The ATG start codon encoding a methionine residue is located at nucleotide position 184 for porcine *LXRα-1*, nucleotide position 217 for porcine *LXRα-2*, and nucleotide position 649 for porcine *LXRα-3*. The stop codon for all porcine *LXRα* transcripts is TGA at nucleotide position 1525, 1558, and 961 for porcine *LXRα-1*, *-2*, and *-3*, respectively (Figure 4.1, 4.2, and 4.3). Sequence upstream of the putative start ATG codon in porcine *LXRα* transcript-1 and -2 did not contain any additional in-frame start codons. There were

one (TAA stop codon) and six in-frame stop codons (three TAG, two TGA, and one TAA stop codons) upstream of the predicted ATG start codon of porcine *LXRα-1* and *-2*, respectively (Figure 4.1 and 4.2). The porcine *LXRα-1* cDNA has an open reading frame of 1344 nucleotides encoding a putative protein of 447 amino acid residues with a molecular weight of 50296 Daltons. The predicted ORF of porcine *LXRα-2* is 1344 bp and encodes 447 amino acids with a predicted molecular weight of 50328 Daltons. The ORF of porcine *LXRα-3* has a 315 bp open reading frame encoding a protein of 104 amino acid residues with a predicted molecular weight of 241 Daltons. The codon usage in the ORF of porcine *LXRα-2* is shown in Table 4.2. It has 58.6% G+C content. The GC contents in the first, second, and third codon positions of porcine *LXRα-2* are 62.7%, 42.6% and 70.3%, respectively. The codons GCC (alanine), GAG (glutamic acid), UUC (phenylalanine), AUC (isoleucine), AAG (lysine), CUG (leucine), CCC (proline), and CAG (glutamine) were frequently used in porcine *LXRα* ORF.

The nucleotide sequence around ATG initiation codon of porcine *LXRα* is according with Kozak consensus sequence (GCCA/GCCATGG) (Kozak, 1996). The nucleotide flanking ATG in porcine *LXRα-1* and *-2* is GAAGAG*ATG*T. Porcine *LXRα* has G and T in positions -3 and +4, respectively. This classified sequence flanking porcine *LXRα* ATG initiation codon as an adequate Kozak consensus sequence. In contrast, the nucleotide flanking ATG in porcine *LXRα-3* is classified as a weak Kozak consensus sequence (TCTTCC*ATG*T). The consensus sequence flanking ATG codons in *LXRα* is shown in Figure 4.8.

Table 4.2. Codon usage in porcine *LXRα-2*.

Amino acid	Codon	Number	Frequency
Alanine	GCG	3	6.7
	GCA	8	17.9
	GCU	6	13.4
	GCC	18	40.2
Cysteine	UGU	3	6.7
	UGC	9	20.1
Aspartic acid	GAU	10	22.3
	GAC	9	20.1
Glutamic acid	GAG	33	73.7
	GAA	5	11.2
Phenylalanine	UUU	6	13.4
	UUC	12	26.8
Glycine	GGG	10	22.3
	GGA	2	4.5
	GGU	0	0
	GGC	7	15.6
Histidine	CAU	3	6.7
	CAC	8	17.9
Isoleucine	AUA	0	0
	AUU	2	4.5
	AUC	14	31.3
Lysine	AAG	12	26.8
	AAA	8	17.9
Leucine	UUG	2	4.5
	UUA	0	0
	CUG	27	60.3
	CUA	3	6.7
	CUU	6	13.4
	CUC	9	20.1
Methionine	AUG	12	26.8
Asparagine	AAU	4	8.9
	AAC	5	11.2
Proline	CCG	2	4.5
	CCA	8	17.9
	CCU	6	13.4
	CCC	12	26.8
Glutamine	CAG	24	53.6
	CAA	8	17.9

Amino acid	Codon	Number	Frequency
Arginine	AGG	5	11.2
Arginine	AGA	2	4.5
	CGG	9	20.1
	CGA	3	6.7
	CGU	5	11.2
	CGC	9	20.1
Serine	AGU	5	11.2
	AGC	15	33.5
	UCG	2	4.5
	UCA	4	8.9
	UCU	11	24.6
	UCC	9	20.1
Threonine	ACG	2	4.5
	ACA	7	15.6
	ACU	1	2.2
	ACC	3	6.7
Valine	GUG	17	38
	GUA	1	2.2
	GUU	2	4.5
	GUC	8	17.9
Tryptophan	UGG	4	8.9
Tyrosine	UAU	3	6.7
	UAC	4	8.9
Stop codon	UGA	1	2.2

Note: The codon usage in porcine *LXRα-2* was derived from SMS. Porcine *LXRα-2* has 58.6% GC in its ORF. The GC contents in the first, second, and third positions of porcine *LXRα-2* were 62.7%, 42.6%, and 70.3%, respectively.
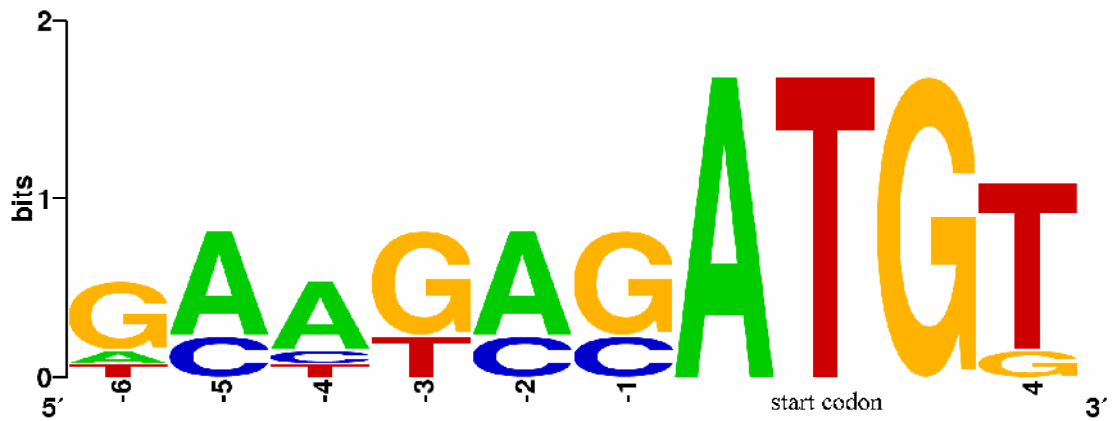
Figure 4.8. Kozak consensus sequence surrounding ATG start codon of *LXRα* in farm animals and human.

The nucleotide flanking ATG in porcine *LXRα-1* and *-2* is GAA*GAGATG*T. Porcine *LXRα* has G and T in positions -3 and +4, respectively. This classified sequence flanking porcine *LXRα* ATG initiation codon as an adequate Kozak consensus sequence. In contrast, the nucleotide flanking ATG in porcine *LXRα-3* is classified as a weak Kozak consensus sequence (TCTTCC*ATG*T). The WebLogo was used in the create diagram representing the Kozak consensus sequence.

4.2.3. 3'-Untranslated Region
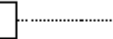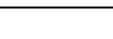
The 3'-UTRs of porcine *LXRα-1* and *-2* are identical and contains a polyadenylation signal ATTAAA at positions 1636-1641 and 1669-1674, respectively (Figures 4.1 and 4.2). In contrast, the sequence downstream TGA stop codon of porcine *LXRα-3* contains no polyadenylation signal, but has a straight of adenosine bases (15 bases of A) (Figure 4.3). The TGA termination codon is located at nucleotide position 1525, 1558, and 961 for porcine *LXRα-1*, *-2*, and *-3*, respectively. The analysis of 3'-UTRs of porcine *LXRα* transcripts did not reveal the presence of known regulatory elements (e.g., AU-rich elements (AREs)) involved in post-transcriptional regulation of mRNA such as mRNA localization, stability, and translational efficiency (Pesole *et al.,* 2001; Hughes, 2006). The secondary structures of the 3'-UTR of these porcine *LXRα* transcripts were predicted using the RNAfold program. The predicted fold energy of 3'-UTR is -35.3 kcal/mol for porcine *LXRα-1*

and porcine *LXRα-2*. The predicted secondary structure of porcine *LXRα* 3'-UTR is shown in Figure 4.9.

Figure 4.9. Predicted secondary structure of 3'-UTR of porcine *LXRα -2*.

The secondary structure of 3'-UTR was predicted using the RNAfold program. The predicted fold energy of porcine *LXRα* 3'-UTR is -35.3 kcal/mol.

4.3. Predicted Amino Acid Sequences of Porcine LXRα

Porcine LXRα-1 and -2 have identical putative proteins with 447 amino acid residues and posses a putative DBD and LBD domains typical of LXRα receptors. The porcine LXRα-3 protein has shorter amino acid with 104 residues. Porcine LXRα-3 isoform is a truncated protein lacking the hinge D and LBD. The amino acid composition of porcine LXRα proteins is shown in Table 4.3. Comparison of the deduced amino acid sequences of porcine LXRα revealed that it shares about 97% amino acid identity with bovine LXRα, 70% identity with chicken LXRα, and 92% identity with human LXRα (Table 4.4).

Table 4.3. Amino acid composition of porcine LXRα-2.

AminoAcid	Number	Mol%
Alanine (A)	35	7.8
Cysteine (C)	12	2.7
Aspartic acid (D)	19	4.3
Glutamic acid (E)	38	8.5
Phenylalanine (F)	18	4.0
Glycine (G)	19	4.3
Histidine (H)	11	2.5
Isoleucine (I)	16	3.6
Lysine (K)	20	4.5
Leucine (L)	47	10.5
Methionine (M)	12	2.7
Asparagine (N)	9	2.0
Proline (P)	28	6.3
Glutamine (Q)	32	7.2
Arginine (R)	33	7.4
Serine (S)	46	10.3
Threonine (T)	13	2.9
Valine (V)	28	6.3
Tryptophan (W)	4	0.9
Tyrosine (Y)	7	1.6

Note: Amino acid composition was calculated using ProParam. Number indicates frequency of each amino acid residue. Percentage represents calculated molecular weight of each amino acid residue in porcine LXRα-2 protein.

Table 4.4. Structure and percent sequence identity of LXRα proteins between farm animals and human.

Protein	Domain				% identity	Length
	A/B	DBD	D	LBD		
Porcine LXRα-1						447
Porcine LXRα-2					99	447
Porcine LXRα-3						104
Bovine LXRα					97	447
Chicken LXRα					70	409
Human LXRα					92	447

Note: Percent identity in amino acid sequences was calculated using pairwise alignment in the BioEdit with default parameters (BLOSUM62).

4.3.1. Domain Structure

Predicted amino acid sequences of the porcine *LXRα* gene were analyzed using SMART, and the evolutionally conserved DBD and LBD in this receptor protein were identified (Figure 4.10). Sequence comparison of DBD and LBD between porcine LXRα and its homologous gene products in human and farm animals showed that porcine LXRα-1 DBD (amino acid residues 95-166) shares 100% identity with cattle, 97% identity with human, and 86% with chicken. Porcine LXRα-1 LBD (amino acid residues 259-447) shares 100% identity with cattle and human, and 87% with chicken (Figure 4.10).

The putative DNA-binding domain, containing two zinc finger motifs (Figure 4.11), is conserved among human and farm animals. The first and second zinc fingers are predicted at amino acid residues 98 to 118 (CSVCGDKASGFHYNVLSCEGC) and amino acid residues 134 to 153 (CHSGGHCPMDTYMRRKCQEC), respectively (Figure 4.11). The highly conserved P-box is located at the positions 115-122 (CEGCKGFF) in the DBD and is common to most nuclear receptors. Two amino acid

substitutions in the DBD between porcine and human LXRαs were identified (Arg131His and Val133Ile) (Figure 4.11). These two amino acid substitutions are the replacement of amino acids with similar properties.

Sequence alignment of the LBD for swine and human is shown in Figure 4.12. The LBD of porcine LXRα-2 isoform is completely identical to human LXRα receptor. The highly conserved peptide EDQI(S/A/I/T/V)LLK is located in the LBD and is conserved in most nuclear receptors. Porcine LXRα LBD has the peptide EDQIALLK at amino acid residues 284-291. The ligand-dependent activation-2 (AF-2), important in ligand and coactivator binding, was also found in the LBD (amino acid residues 436-442: PPLLSEI) (Figure 4.12).
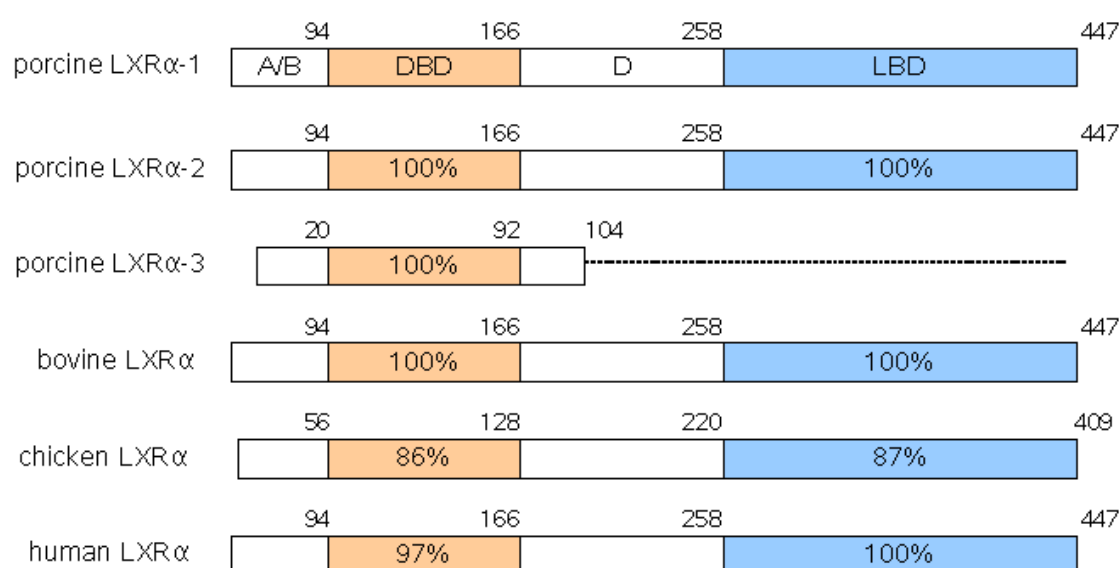
Figure 4.10. Structural domains of porcine LXRα protein.

Percent identity in amino acid sequences was calculated using pairwise alignment in BioEdit. The structural domains are indicated. Numbers inside the boxes represent percent sequence identity in comparison with porcine LXRα-1 isoform. Numbers above each structural domain indicate position of amino acid in boundary region.

```
Porcine LXRα  95 NELCSVCGDKASGFHYNVLSCEGCKGFFRRSVIKGARYVCHSGGHCPMDTYMRRKCQECR 154
Human LXRα    95 ..................................H.I..................... 154


Porcine LXRα  LRKCRQAGMREE 166
Human LXRα    ............ 166
```

Figure 4.11. Alignment of the DBD between porcine and human LXRα proteins.

Amino acid is colored according to its property. Dots represent identical amino acid. Different amino acids between pig and human are indicated by single amino acid codes. The first zinc finger motif and the second zinc finger motif are underlined. Zinc finger motifs in porcine LXRα DBD were predicted using the ScanProsite. The conserved peptide CEGCKGFF located in the highly conserved P-box is boxed. Two conserved amino acid substitutions in the DBD between porcine and human LXRα proteins were identified (Arg131His and Val133Ile).

```
Porcine LXRα-1 259 ELAIVSVQEIVDFAKQLPGFLQLSREDQIALLKTSAIEVMLLETSRRYNPGSESITFLKD 318
Porcine LXRα-2 259 ........................................................... 318
Human LXRα     259 ........................................................... 318


Porcine LXRα-1     FSYNREDFAKAGLQVEFINPIFEFSRAMNELQLNDAEFALLIAISIFSADRPNVQDQLQV 378
Porcine LXRα-2     ........................................................... 378
Human LXRα         ........................................................... 378


Porcine LXRα-1     ERLQHTYVEALHAYVSIHHPHDRLMFPRMLMKLVGLRTLSSVHSEQVFALRLQDKKLPPL 438
Porcine LXRα-2     ...............................S........................... 438
Human LXRα         ........................................................... 438


Porcine LXRα-1     LSEIWDVHE 447
Porcine LXRα-2     ......... 447
Human LXRα         ......... 447
```

Figure 4.12. Alignment of the LBD between porcine and human LXRα proteins.

Amino acid is colored according to its property. The conserved peptide EDQISLLK and the AF-2 (PPLLSEI) are boxed. Dots represent identical amino acids between swine and human. Different amino acids are indicated by single amino acid codes.

4.3.2. Homology Modeling

The 3-D structure of porcine LXRα protein was predicted using the Swiss Model Server (Schwede *et al.,* 2003). The porcine LXRα-2 isoform (UniProtKB/TrEMBL accession no. Q4TU03) was used as a query. The crystal structure of mouse LXRα (PDB: 2ACL) was identified with 98% identity (Jaye *et al.,* 2005). The 2ACL template was used in the 3-D homology modeling of porcine LXRα-2 isoform. The predicted 3-D structure of porcine LXRα-2 is shown in Figure 4.13.

Secondary structure of porcine LXRα amino acid sequence was derived from the alignment of its LBD region with the secondary structure of mouse and human LXRα proteins (protein data bank: 2ACL and 1UHL). Porcine LXRα LBD structure was predicted to enclose by helix 1 (P208-R226), helix 3 (R250-Q274), helix 4 (R283-L290), helix 5 (A294-R304), helix 6 (R323-K328), helix 7 (I336-E348), helix 8 (D353-I364), helix 9 (Q375-H396), helix 10 (L402-Q331), helix AF2 (P336-I442) and three beta-strands (β-1, β-2, β-3; Y306-N307, S312-F315, and F319-Y321, respectively) (Figure 4.14).
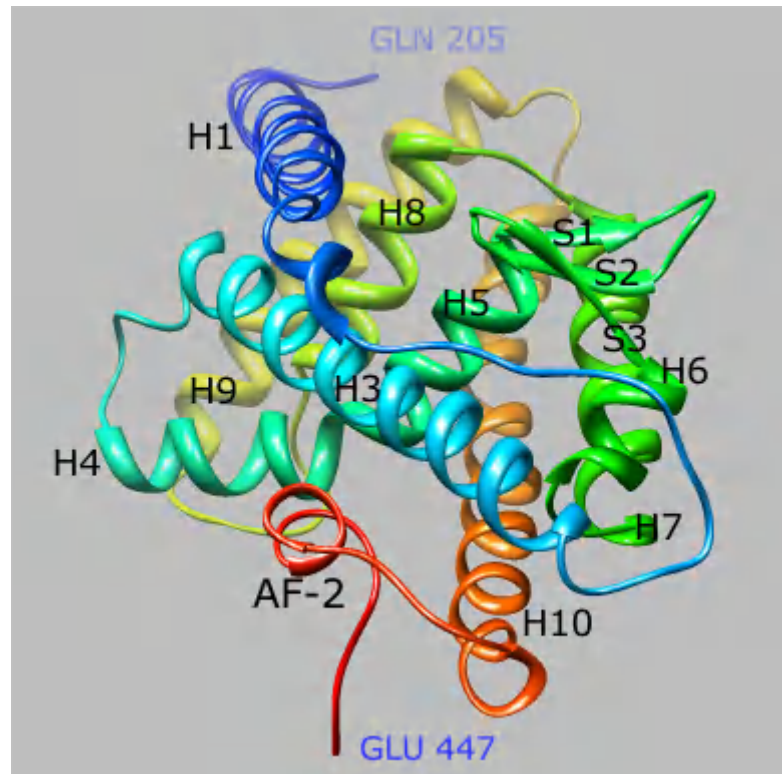
Figure 4.13. Ribbon representation of the predicted 3D structure of porcine LXRα-2 protein.

The 3-D model was predicted using the service of the Swiss Model Repository. The crystal structure of mouse LXRα (PDB template: 2ACL) was identified and used in the construction of porcine LXRα 3-D model. The image was produced using the UCSF Chimera. The predicted secondary structural elements are indicated (H: α helix and S: β strand).

```
pLXRα 205 QLSPEQLGMIEKLVAAQQQCNRRSFSDQLRVTPWPMAPDPQSREARQQRFAHFTELAIVS 264
hLXRα 205 QLSPEQLGMIEKLVAAQQQCNRRSFSDRLRVTPWPMAPDPHSREARQQRFAHFTELAIVS 264
mLXRα 203 QLSPEQLGMIEKLVAAQQQCNRRSFSDRLRVTPWPIAPDPQSREARQQRFAHFTELAIVS 262
                  alpha1                                alpha3

pLXRα     VQEIVDFAKQLPGFLQLSREDQIALLKTSAIEVMLLETSRRYNPGSESITFLKDFSYNRE 324
hLXRα     VQEIVDFAKQLPGFLQLSREDQIALLKTSAIEVMLLETSRRYNPGSESITFLKDFSYNRE 324
mLXRα     VQEIVDFAKQLPGFLQLSREDQIALLKTSAIEVMLLETSRRYNPGSESITFLKDFSYNRE 322
             alpha3          alpha4        alpha5   beta1  beta2  beta3

pLXRα     DFAKAGLQVEFINPIFEFSRAMNELQLNDAEFALLIAISIFSADRPNVQDQLQVERLQHT 384
hLXRα     DFAKAGLQVEFINPIFEFSRAMNELQLNDAEFALLIAISIFSADRPNVQDQLQVERLQHT 384
mLXRα     DFAKAGLQVEFINPIFEFSRAMNELQLNDAEFALLIAISIFSADRPNVQDQLQVERLQHT 382
           alpha6         alpha7        alpha8              alpha9

pLXRα     YVEALHAYVSIHHPHDRLMFPRMLMKLVSLRTLSSVHSEQVFALRLQDKKLPPLLSEIWDVHE 447
hLXRα     YVEALHAYVSIHHPHDRLMFPRMLMKLVSLRTLSSVHSEQVFALRLQDKKLPPLLSEIWDVHE 447
mLXRα     YVEALHAYVSINHPHDPLMFPRMLMKLVSLRTLSSVHSEQVFALRLQDKKLPPLLSEIWDVHE 445
             alpha9              alpha10                   AF−2
```

Figure 4.14. Predicted secondary structure of porcine LXRα-2 protein.

The secondary structure was derived from the 3D model of porcine LXRα-2 using the crystal structure of mouse LXRα (PDB template: 2ACL).

4.4. Expression Analysis

The expression pattern analysis of the porcine *LXRα-1* transcript and porcine *LXRα-2* transcript in adult tissues was performed using the RT-PCR. We designed primer sets that specific to each porcine *LXRα* splicing variant. These primers span between the exon-intron boundaries to prevent the generation of amplified products from genomic DNA. The expression of porcine *LXRα-1* transcript was detected in the liver, kidney, small intestine, heart, muscle, thymus, spleen, and brain, whereas porcine *LXRα-2* transcript was expressed only in the thymus and spleen (Figure 4.15 and 4.16). The expression pattern of the porcine *LXRα-1* transcript and porcine LXRα-

2 transcript exhibited a 464 bp DNA band and a 503 bp DNA band, respectively (Figure 4.15 and 4.16). These RT-PCR products were verified by direct sequencing.
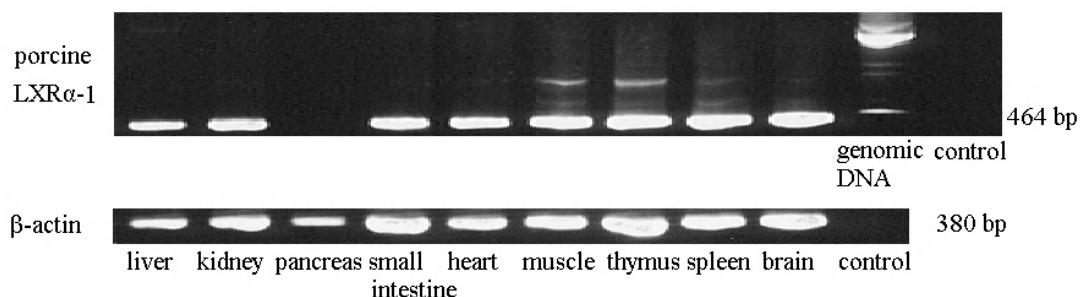
Figure 4.15. Expression analysis of porcine *LXRα-1*.

The RT-PCR was applied to detect the mRNA expression of *LXRα* from various porcine tissues. Porcine tissues including liver, kidney, pancreas, small intestine, heart, muscle, thymus, spleen, and brain were derived from a mature female pig (Landrace X Large White). The mRNA expression of porcine *LXRα-1* transcript was detected in the liver, kidney, small intestine, heart, muscle, thymus, spleen, and brain.
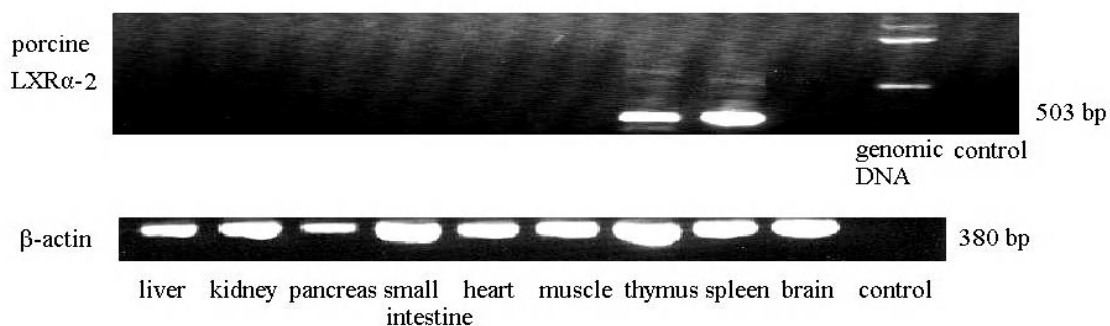
Figure 4.16. Expression analysis of porcine *LXRα-2*.

The RT-PCR was applied to detect the mRNA expression of *LXRα-2* from various porcine tissues. Porcine tissues including liver, kidney, pancreas, small intestine, heart, muscle, thymus, spleen, and brain were derived from a mature female pig (Landrace X Large White). The mRNA expression of porcine *LXRα-2* transcript was detected only in the thymus and spleen.

4.5. Predicted Exon-Intron Structure of Porcine *LXRα*

To identify genomic DNA corresponds to porcine *LXRα* mRNA sequences, we searched ongoing swine genome sequencing project using pig BLAST (http://www.ncbi.nlm.nih.gov/projects/genome/seq/BlastGen/BlastGen.cgi?taxid=982 3). Hits with threshold were not identified. Therefore, the exon-intron structure of porcine *LXRα* was predicted by aligning its mRNA sequence with bovine genomic DNA (NC_007313, chromosome 15) using the SIM4 program (Florea *et al.,* 1998) (Figure 4.17). The partial sequence of porcine *LXRα-3* was also mapped to this bovine chromosome 15. However, the splice sites did not agree with GT-AG rule. The porcine *LXRα gene* was predicted to contain 10 exons ranging in size from 80 to 278 bp. The longest exon is exon 10 with 278 bp. The shortest exon is exon 2 with 80 bases. The exon-intron boundary of porcine *LXRα* transcripts is shown in Tables 4.5 and 4.6. All exon-intron borders follow the consensus sequences of GT-AG splice sites (Mount, 1982). The 5'-UTR is encoded by exon 1 and the part of exon 2. The different sequences between transcript 1 and 2 are on exon 1. The protein-coding region of porcine *LXRα* is encoded by 9 exons (exons 2-10). The ATG start codon is located in exon 2. In contrast to porcine *CAR*, *PXR*, and *RORγ*, the DBD of porcine *LXRα* is encoded by a single exon, exon 4. The LBD is encoded by the part of exon 6, exons 7-9, and the part of exon 10. The TGA stop codon is located in exon 10. The 278-bp 3'-UTRs of porcine *LXRα* is encoded by exon 10.

Comparison of *LXRα* exons revealed that the size and number of exon are highly conserved when compared to human *LXRα*. They exhibit approximately 90% identity in protein-coding exons (Figure 4.18). In addition, exons encoding 5'- and 3'-UTR are also conserved with approximately 80% identity. The exon 4, which encodes DBD, is identical in size (267 bp) and exhibit approximately 90% identity when

compared to human (Figure 4.18). The LBD-encoding exons (exons 6-10) are also

conserved both in size and sequence identity (vary from 81% identity in exon 10 to

99% identity in exon 7). Human *LXRα* gene is located on chromosome 11 and

contains 9 exons. The beginning exon (78 bp) is similar to exon 2 in swine and cattle,

and exhibits 88% sequence identity (Figure 4.18).
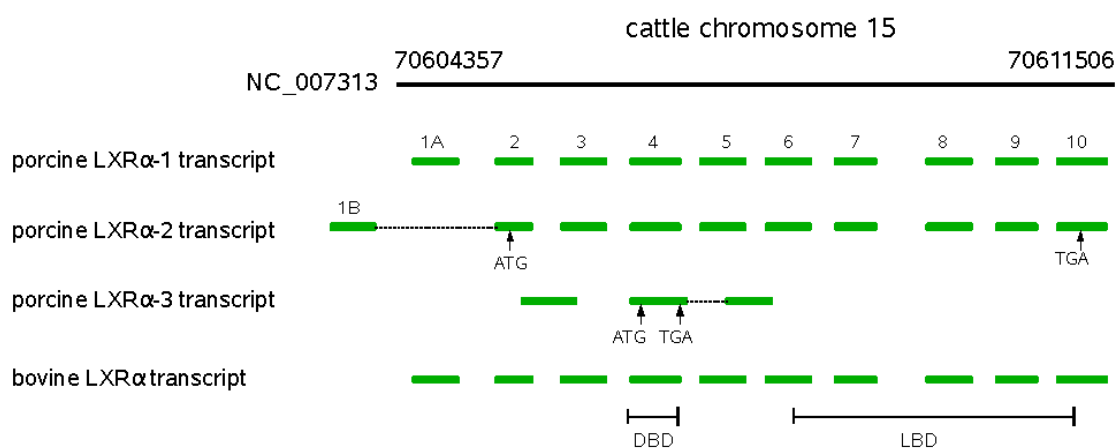
Figure 4.17. Porcine *LXRα* transcripts on bovine chromosome 15.

Porcine *LXRα* spans ~7 kb of DNA on cattle chromosome 15 (NC_007313: 70604357-70611506) and consists of potential 10 exons. The genomic region was identified using cow BLAST. The exon-intron boundaries were determined by aligning each porcine *LXRα* transcript with cow DNA sequence using SIM4 program. The continuous long line indicates bovine genomic DNA sequence with genomic region on each end. Exons are shown as short green lines. The positions of start and stop codons as well as DBD and LBD are indicated.

Table 4.5. Predicted exon-intron boundaries of porcine *LXRα-1* transcript.

Exon	*LXRα-1*	NC_007313	Size	Identity	5' end of exon	3' end of exon	Intron	Intron size	5' splice site	3' splice site
1A	1-146	2-146	146	77%	AGTCGGTCGG...	...TGGGCGCTCG	1	759	gtgggtctgc...	...cttcctgcag
2	147-226	906-985	80	95%	GACAGTCCCT...	...GTTTCTCCTG	2	527	gtaagcgtct...	...cggggttcag
3	227-415	1513-1701	189	86%	ACTCTGCAGT...	...GAGTCCACAG	3	510	gtgaggagct...	...ctgcatccag
4	416-682	2212-2478	267	95%	AGCTCCGTCC...	...CGAGAGGAGT	4	368	gtgagtgtct...	...ctctcctcag
5	683-891	2847-3055	209	92%	GTGTCCTGTC...	...TCGAGTCACG	5	107	gtacctgaga...	...ttccgggcag
6	892-1071	3163-3342	180	95%	CCTTGGCCCA...	...TGCGATTGAG	6	182	gtgactggcg...	...ttgtgtgcgg
7	1072-1171	3525-3624	100	96%	GTGATGCTTC...	...GCCAAAGCAG	7	2611	gtgaggactg...	...tcctccccag
8	1172-1285	6236-6349	114	96%	GGCTGCAGGT...	...TTCTCTGCAG	8	219	gtgtggagga...	...tgacctgtag
9	1286-1380	6569-6663	95	98%	ACCGGCCCAA...	...CCACCCCCAT	9	191	gtgagtctcc...	...cctttcctag
10	1381-1658	6855…	278	?	GACCGACTGA...	...AAGGGTTGGG				

Note: The exon-intron boundaries were predicted by aligning porcine *LXRα-1* cDNA with bovine genomic DNA sequence (NC_007313: 70604357-70611506 (7150bp)) using SIM4 program. The poly(A) tail was excluded in the alignment. The 5'- and 3'-ends of each exon are shown in uppercase letters. The 5' and 3' splice sites of each intron are shown in lowercase letters. All splice sites are in consensus with GT-AG rule for splicing. The sequence identity between porcine *LXRα-1* transcript and bovine genomic DNA was derived from alignment results of SIM4. The cDNA sequences from exon 2 to exon 10 are identical between porcine *LXRα-1* and *LXRα-2*.

Table 4.6. Predicted exon-intron boundaries of porcine *LXRα-2* transcript.

Exon	*LXRα-2*	NC_007313	Size	Identity	5' end of exon	3' end of exon	Intron	Intron size	5' splice site	3' splice site
1B	1-179	?	179	?	ACAAGAGGCG...	...CCAGTGTCAG	1	?	nnnnnnnnnn...	...cttcctgcag
2	180-259	>898-985	80	92%	GACAGTCCCT...	...GTTTCTCCTG	2	527	gtaagcgtct...	...cggggttcag
3	260-448	1513-1701	189	86%	ACTCTGCAGT...	...GAGTCCACAG	3	510	gtgaggagct...	...ctgcatccag
4	449-715	2212-2478	267	95%	AGCTCCGTCC...	...CGAGAGGAGT	4	368	gtgagtgtct...	...ctctcctcag
5	716-924	2847-3055	209	92%	GTGTCCTGTC...	...TCGAGTCACG	5	107	gtacctgaga...	...ttccgggcag
6	925-1104	3163-3342	180	95%	CCTTGGCCCA...	...TGCGATTGAG	6	182	gtgactggcg...	...ttgtgtgcgg
7	1105-1204	3525-3624	100	96%	GTGATGCTTC...	...GCCAAAGCAG	7	2611	gtgaggactg...	...tcctcccag
8	1205-1318	6236-6349	114	96%	GGCTGCAGGT...	...TTCTCTGCAG	8	219	gtgtggagga...	...tgacctgtag
9	1319-1413	6569-6663	95	98%	ACCGGCCCAA...	...CCACCCCCAT	9	191	gtgagtctcc...	...cctttcctag
10	1414-1691	6855…	278	?	GACCGACTGA...	...AAGGGTTGGG				

Note: The exon-intron boundaries were predicted by aligning porcine *LXRα-2* cDNA with bovine genomic DNA sequence (NC_007313) using the SIM4 program. The poly(A) tail was excluded in the alignment. The 5'- and 3'-ends of each exon are shown in uppercase letters. The 5' and 3' splice sites of each intron are shown in lowercase letters. The splice sites are in consensus with GT-AG rule for splicing. The sequence identity between porcine *LXRα-2* transcript and cow DNA was derived from alignment results of SIM4. The 179 bp in the 5'-UTR of porcine *LXRα-2* could not align with this bovine genomic DNA. This fragment was postulated as exon 1 that might transcribe from upstream promoter. The cDNA sequences from exon 2 to exon 10 are identical between porcine *LXRα-1* and *LXRα-2*.
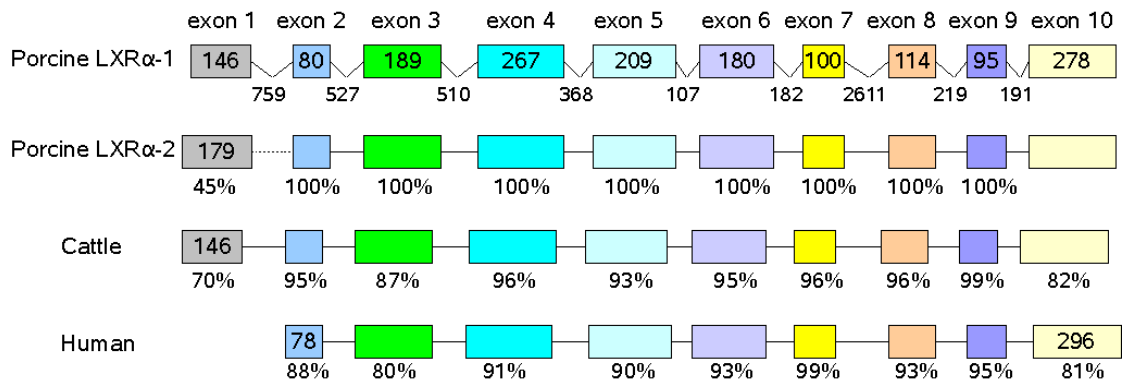
Figure 4.18. Comparison of *LXRα* exons in farm animals and human.

Boxes represent exons. Exons with similar size are colored. The percent identity derived from pairwise alignment using BioEdit is shown below each exon. The size of exon is shown in the exon box. The thin lines indicate introns. The length of intron is shown only for porcine *LXRα-1* transcript.

4.6. Discussion

The nuclear receptor LXRα is an important transcription factor of genes involved in cholesterol and glucose metabolism, and fatty acid biosynthesis (Peet *et al.,* 1998a; Schultz *et al.,* 2000; Lu *et al.,* 2001; Dalen *et al.,* 2003; Juvet *et al.,* 2003; Laffitte *et al.,* 2003; Seo *et al.,* 2004). Transcript variants of human *LXRα* gene have been reported and are probably generated by alternative splicing and alternative promoter usage (Chen *et al.,* 2005).

In the present study, we have identified and sequenced three porcine *LXRα* transcripts. The full-length cDNA sequences of porcine *LXRα* have been submitted to the DDBJ/EMBL/GenBank databases under the accession numbers AB254405 and AB254406 (Thadtha *et al.,* 2006). The porcine *LXRα-1* and *-2* are identical in ORFs and 3'-UTRs but vary in 5'-UTRs (1-145 in *LXRα-1* and 1-178 in *LXRα-2*) (Figures 4.4 and 4.5). One nucleotide substitution was found in the ORF of the transcripts 1 and 2 (G1420 in *LXRα-1* and A1453 in *LXRα-2*), resulting in one amino acid difference between porcine LXRα-1 and LXRα-2 isoforms (Glycine413Serine). The partial porcine *LXRα-3* transcript represents an incomplete transcript that has an in-frame TGA stop codon, resulting in a truncated protein lacking amino acid residues downstream from the DBD. Because porcine *LXRα-3* possesses the premature stop codon, it may be subjected to the nonsense-mediated mRNA decay (Lewis *et al.,* 2003).

Analysis of nucleotide sequences revealed that full-length porcine *LXRα-1* shares 93% identity with porcine *LXRα-2*, 39% identity with *LXRα-3*, 89% identity with bovine *LXRα*, 80% identity with human *LXRα*, and 34% identity with chicken *LXRα*. Comparison of the deduced amino acid sequences of porcine LXRα revealed

that it shares 97% amino acid identity with bovine LXRα, 70% identity with chicken LXRα, and 92% identity with human LXRα (Table 4.4).

Analysis of predicted protein sequences of porcine LXRα indicated that it contains the DBD and LBD (Figures 4.1 and 4.2). It has been established that the DBD is responsible for the binding of nuclear receptors to its target gene promoters (Mangelsdorf *et al.,* 1995; Giguere, 1999; Khorasanizadeh and Rastinejad, 2001; Robinson-Rechavi *et al.,* 2003). The zinc finger motifs are responsible for this DNA binding specificity. The porcine LXRα isoforms (isoforms 1 and 2) were predicted to possess both first and second zinc fingers (amino acid residues 98 to 118 (CSVCGDKASGFHYNVLSCEGC) and amino acid residues 134 to 153 (CHSGGHCPMDTYMRRKCQEC), respectively (Figure 4.11). They also possess the highly conserved P-box (amino acid residues 115-122: CEGCKGFF) in DBD, which is responsible for DNA recognition, and the AF-2 (amino acid residues 436-442: PPLLSEI) located in LBD, which is the binding site for coactivators (Figures 4.11 and 4.12). In general, evolutionarily related nuclear receptors share at least 80%–90% identity in the DBD and at least 40%–60% identity in the LBD (Nuclear Receptors Nomenclature Committee, 1999). The highly conserved DBD is identical in all porcine LXRα proteins and bovine LXRα and highly conserved with human LXRα (97% identity) (Figure 4.10). The LBD is completely conserved among swine, cattle and humans (100% identity). Only one glycine residue (position 413) substituted serine residue in the LBD of porcine LXRα-1 (Figure 4.12). These results suggest that porcine LXRα isoforms are likely to bind to the same or similar target gene promoters and recognize the same or similar ligands as found in human. With high degree of sequence identity in both DBD and LBD, we conclude that porcine LXRα is an ortholog of human LXRα receptor.

The exon-intron structure of porcine *LXRα* gene was determined by aligning its cDNA sequences with bovine genomic DNA (NC_007313). It was predicted to contain ten exons separated by nine introns. The *LXRα* gene is located on chromosomes 11, 15, 5, and 2 in human, cattle, chicken, and mouse, respectively (Willy *et al.,* 1995; Chen *et al.,* 2005). The exon-intron structure of porcine *LXRα* is similar to human in both number and size. They exhibit approximately 90% identity in protein-coding exons (Figure 4.18). In addition, exons encoding 5'- and 3'-UTR are also conserved with approximately 80% identity. The exon 4, encoding DBD, is identical in size (267 bp) and exhibit approximately 90% identity when compared to human (Figure 4.18). The LBD-encoding exons (exons 6-10) are also conserved both in size and sequence identity (vary from 81% identity in exon 10 to 99% identity in exon 7). However, porcine *LXRα* has one additional exon (exon 1) at the 5'-end of the mRNA. This exon is encoded for the 5'-UTR, which might result from alternative promoter usage in porcine *LXRα*. It has been known that difference in 5'-UTR indicates tissue specific expression of mRNA transcripts. Porcine *LXRα-1* and *-2* transcripts might express in different porcine tissues (see below).

To detect mRNA expression of porcine *LXRα* transcripts, we applied RT-PCR to several tissues derived from a mature crossbred female pig (Landrace X Large White). The porcine *LXRα-1* was detected in the liver, kidney, small intestine, heart, muscle, thymus, spleen, and brain, whereas porcine *LXRα-2* transcript was detected only in the thymus and spleen (Figure 4.15 and 4.16). The mRNA expression of porcine *LXRα* transcripts is similar to humans and other animals with regard to types of tissues. The expression pattern of porcine *LXRα* transcripts is similar to human *LXRα*, except only in the pancreas. As summarized in the UniGene database (http://www.ncbi.nlm.nih.gov/sites/entrez?db=unigene) (UniGene ID: 3008923,

2666506, 120536, 233900, and 261176 for swine, cattle, chicken, human, and mouse, respectively), the expression of *LXRα* in pancreas was found only in human but not in pancreas of cattle, mouse, and chicken. In addition, *LXRα* is expressed in testis and ovary in cattle and human. In this stydy, we did not apply RT-PCR with these porcine tissues. Recently, Joseph *et al.* (2004) reported the possible function of LXRα in mouse macrophages activity related with an innate immunity response to intracellular bacteria *Listeria monocytogenes* (LM). Moreover, Valledor *et al.* (2004) also reported that the activation of liver X receptors and retinoid X receptors protected mouse macrophages from apoptosis caused by infection with bacterial pathogens (e.g., *Bacillus anthracis*, *Escherichia coli*, and *Salmonella typhimurium*). With regard to swine production, Yu *et al.* (2006) reported the association between meat and fat production and polymorphisms in *LXR* genes and concluded that *LXRs* might play a role in influencing lean and fat growth in pigs.

In conclusion, we have identified and sequenced three transcripts of porcine *LXRα* cDNAs. The porcine *LXRα-1* and *LXRα-2* are likely to be alternative mRNA transcripts derived from different promoter usage. Future studies of the *LXRα* expression, especially porcine *LXRα-2* transcript, in porcine immune cells may provide new information about this important gene with regard to the immunity in swine.

CHAPTER 5

THE NUCLEAR RECEPTOR RORγ

5.1. Introduction

The retinoic acid receptor-related orphan receptor (*RORγ*; NR1F3) is a member of the nuclear receptor superfamily and has been proposed as a nuclear receptor involved in the development of T lymphocytes, lymph nodes (LNs) and Peyer's patches (PPs) in mammals (Kurebayashi *et al.,* 2000; Winoto and Littman, 2002; Eberl and Littman, 2003). Transcriptional regulation of target genes by nuclear receptors is dependent on the DNA-binding domain (DBD) that is responsible for the binding of specific DNA sequences called response elements in target genes, and the ligand-binding domain (LBD), which functions as the binding site for ligands as well as receptor dimerization and the interaction with coactivators and corepressors (Mangelsdorf *et al.,* 1995; Giguere, 1999; Robinson-Rechavi *et al.,* 2003). In vitro studies indicated that RORγ binds to a specific DNA sequence located in target gene promoter called the RORγ-response element (RORE). The binding of ROR to RORE has been shown to bind as a monomer to a single core motif AGGTCA preceded by an A/T rich region (Ortiz *et al.,* 1995), and bind to a direct repeat AGGTCA separated by four and five nucleotides (Medvedev *et al.,* 1996), which is the same response element identified as the binding site for the thyroid hormone receptors (TRs) and the retinoic acid receptors (RARs). Villey *et al.* (1999) reported that the RORγt could bind to the

T early alpha (*TEA*) promoter located upstream of the T cell receptor alpha chain (*TCR-Jα*) gene. The nuclear orphan receptor *RORγ* has been cloned from human (Hirose *et al.,* 1994) and mouse (Ortiz *et al.,* 1995; Medvedev *et al.,* 1996). He *et al.* (1998) identified a closely related isoform of *RORγ* from mouse T cell cDNA library, which they called *RORγt*. The RORγt isoform differed from RORγ in its amino acid terminal domain, possibly resulting from an alternative promoter usage. The expression of *RORγ* is detected in human and mouse skeletal muscle, liver and kidney (Hirose *et al.,* 1994; Medvedev *et al.,* 1996), mouse thymus (Ortiz *et al.,* 1995), whereas the expression of *RORγt* is specific to the thymus (He *et al.,* 1998).

Studies of *RORγ* in human and mouse suggest the potential function of these receptor isoforms in the development of T cells in thymus and the development of the lymph node (Kurebayashi *et al.,* 2000; Winoto and Littman, 2002; Eberl and Littman, 2003), which are important organs in response to diseases. Since the cDNAs encoding porcine *RORγ* have not been cloned in swine, we applied homology searching strategy and identified four cDNA clones derived from PEDE database. In this chapter, we report the identification and characterization of these four porcine *RORγ* transcripts.

5.2. Nucleotide Sequences of Porcine *RORγ*

The full-length cDNA sequences of porcine *RORγ* were identified and sequenced from four PEDE clones using the primer walking sequencing method. We refer to these mRNA transcripts as porcine *RORγ-1* transcript, porcine *RORγ-2* transcript, porcine *RORγ-3* transcript and porcine *RORγ-4* transcript. The full-length cDNA sequence of porcine *RORγ-1* transcript is 2282 base pairs. The porcine *RORγ-1* variant has a 210-bp 5'-UTR (nucleotide positions 1-210), a 1157-bp ORF (nucleotide positions 211-1767) encoding 518 amino acids, and a 515-bp 3'-UTR (nucleotide

positions 1768-2282) (Figure 5.1). The porcine *RORγ-2* transcript has 2193 bp mRNA, and contains a 184-bp 5'-UTR (nucleotide positions 1-184), a 1494-bp ORF (nucleotide positions 185-1678) encoding 497 amino acids, and a 515-bp 3'-UTR (nucleotide positions 1679-2193) (Figure 5.2). The porcine *RORγ-3* transcript has 2075 nucleotides in length, and contains a shorter 10-bp 5'-UTR, a 1221-bp ORF (nucleotide positions 11-1231) encoding 406 amino acids, and an 844-bp 3'-UTR (nucleotide positions 1232-2075) (Figure 5.3). The porcine *RORγ-4* transcript has 1229 bp in length, and contains a 211-bp 5'-UTR (nucleotide positions 1-211), a 504-bp ORF encoding 167 amino acids (nucleotide positions 212-715), and a 514-bp 3'-UTR (nucleotide positions 716-1229) (Figure 5.4). The nucleotide composition of porcine *RORγ* transcript variants is summarized in Table 5.1. For the porcine *RORγ-1* transcript, the G+C content is 59.2% for full-length cDNA, 66.7% for 5'-UTR, 59.6% for protein-coding region, and 55% for 3'-UTR. The G+C content in porcine *RORγ-2* transcript is 58.6% for full-length cDNA, 66.3% for 5'-UTR, 59% for protein-coding region, and 55% for 3'-UTR.

All four transcripts differ in the non-coding 5'-UTR, but similar in 3'-UTR. The porcine *RORγ-2* transcript is identical to transcript 1, except it lacks 26 nucleotides in 5'-UTR and 63 base pairs in the ORF. The transcript variant 3 has an in-frame TGA stop codon in exon 8, resulting in a transcript with shorter ORF and longer 3'-UTR. Transcript 4 has the nucleotide deletion in the ORF, resulting in the production of isoforms lacking AB, DBD and LBD. Sequence comparison revealed that full-length porcine *RORγ-1* shares 95% identity with porcine *RORγ-2*, 65% identity with bovine *RORγ*, 57% identity with human *RORγ-1*, 58% identity with human *RORγ-2*, and porcine *RORγ-2* transcript exhibits 69% identity with bovine *RORγ*, 59% identity with human *RORγ-1*, and 60% identity with human *RORγ-2* (Table 5.2).

```
1    GTTGGGGGGGTTGTCACTCGGCCACCTGTGTGGTGCGGAGCTTAAACCCCCCTGCCCAGA
61   AGCATTGGGGGAGAGCTAGGTGCAGAGCTGCAGGCTGAGGCTCTGCTGAGAGGGCCTCAC
121  CCCCAACCCCGCTGCCAGCTGCACCCCATCCCTGGACCACCCCCCGCTAAGAAGGACGGG
181  AGCCCAGTCGGCAAAGCCCACGGCTCAGTCATGAGAACACAAATTGAAGTGATCCCCTGC
                                          M  R  T  Q  I  E  V  I  P  C    10
241  AAAATCTGTGGGGACAAGTCGTCTGGGATCCACTACGGGGTTATCACCTGTGAGGGGTGC
     K  I  C  G  D  K  S  S  G  I  H  Y  G  V  I  T  C  E  G  C    30
301  AAGGGTTTCTTCCGCCGGAGCCAGCAGTGTAACGTGGCCTACTCCTGCACCCGTCAGCAG
     K  G  F  F  R  R  S  Q  Q  C  N  V  A  Y  S  C  T  R  Q  Q    50
361  AACTGCCCCATCGACCGCACTAGCCGAAACCGATGCCAGCACTGCCGCCTACAGAAATGC
     N  C  P  I  D  R  T  S  R  N  R  C  Q  H  C  R  L  Q  K  C    70
421  CTGGCCCTGGGCATGTCCCGAGATGGTGAGGCCATGGGGGGCAGCCCTCTGGGGCTTCCCT
     L  A  L  G  M  S  R  D  G  E  A  M  G  A  A  L  W  G  F  P    90
481  GGAGTCTCCAGAGGGGCAGCTGAGCTGCTGTGCCAGACAGGGTTAGCCTGTAAAAGTGCC
     G  V  S  R  G  A  A  E  L  L  C  Q  T  G  L  A  C  K  S  A    110
541  TTTCTGGAGCTCTGCTCTCCATCCTTCCCGCGGGGCGGGCCCCTGTCCCTTCCACTCTCAT
     F  L  E  L  C  S  P  S  F  P  R  A  G  P  C  P  F  H  S  H    130
601  CAGAGGTCTTCGACGTCCCCTCCTGCAGGGGCCCAAGGAGCAGACCCCCTCGCCTGCCCC
     Q  R  S  S  T  S  P  P  A  G  A  Q  G  A  D  P  L  A  C  P    150
661  CTGGGGCTCCCAGATGGGCAGCTGCCCCTGGGCTCCTCACCTGACCTGCCGGAGGCCTCG
     L  G  L  P  D  G  Q  L  P  L  G  S  S  P  D  L  P  E  A  S    170
721  GCCTGTCCCCCCGGTCTCCTGAGAGCCCCAAGCTGCGGGCCCTCCTACTCCAACAGCTTG
     A  C  P  P  G  L  L  R  A  P  S  C  G  P  S  Y  S  N  S  L    190
781  GTCAAGGCTGGGCTCAATGGGGCTTCTTACCACCTGGAATACAGCCCTGAGCGGGGCAAG
     V  K  A  G  L  N  G  A  S  Y  H  L  E  Y  S  P  E  R  G  K    210
841  GCTGAGGGCAGAGAGAACTTCTATGGCACAGGCAGCCAGCTGGCCCCCGACAGGTGTGGA
     A  E  G  R  E  N  F  Y  G  T  G  S  Q  L  A  P  D  R  C  G    230
901  CTCCATTTTGAGGACCCCAGGCATCCTGGGCTTGGGGAACCAGGACGGGGCCCTGACAGC
     L  H  F  E  D  P  R  H  P  G  L  G  E  P  G  R  G  P  D  S    250
961  TACTGCAGCCCCAGTTTCCGCAGTACCCCAGAGGCACCTTACGCCTCCCTGACAGAGATT
     Y  C  S  P  S  F  R  S  T  P  E  A  P  Y  A  S  L  T  E  I    270
1021 GAGCACCTGGTGCAGAATGTTTGTAAGTCTTACCGAGACACGTGTCAGCTGCGGCTGGAG
     E  H  L  V  Q  N  V  C  K  S  Y  R  D  T  C  Q  L  R  L  E    290
1081 GACCTGCTTCGGCAGCGCTCCAACATCTTCTCCCGAGAGGAGGTGGCCGGCTACCAGAGG
     D  L  L  R  Q  R  S  N  I  F  S  R  E  E  V  A  G  Y  Q  R    310
1141 AAGTCAATGTGGGAGATGTGGGAACGCTGTGCCCACCGCCTCACCGAGGCCATTCAGTAC
     K  S  M  W  E  M  W  E  R  C  A  H  R  L  T  E  A  I  Q  Y    330
1201 GTGGTGGAGTTCGCTAAGAGGCTCTCGGGCTTTATGGAGCTCTGCCAGAACGACCAGATC
     V  V  E  F  A  K  R  L  S  G  F  M  E  L  C  Q  N  D  Q  I    350
1261 GTGCTACTCAAAGCAGGAGCAATGGAAGTGGTGCTGGTCAGGATGTGCCGAGCCTACAAC
     V  L  L  K  A  G  A  M  E  V  V  L  V  R  M  C  R  A  Y  N    370
1321 GCTGACAACCACACAGTCTTTTTTGAAGGCAAATACGGTGGCATGGAGCTGTTCCGAGCC
     A  D  N  H  T  V  F  F  E  G  K  Y  G  G  M  E  L  F  R  A    390
1381 TTGGGCTGCAGCGAGCTCATCAGCTCCATCTTTGACTTCTCCCGCTCCCTGAGTGCCTTG
     L  G  C  S  E  L  I  S  S  I  F  D  F  S  R  S  L  S  A  L    410
1441 CGCTTTTCCGAGGACGAGATTGCACTCTACACAGCCCTCGTCCTCATCAATGCCAACCGG
```

```
      R   F   S   E   D   E   I   A   L   Y   T   A   L   V   L   I   N   A   N   R    430
1501  CCAGGACTCCAAGAGAAAAGGAAAGTAGAACAGCTGCAGTACAATCTGGAGCTGGCCTTT
      P   G   L   Q   E   K   R   K   V   E   Q   L   Q   Y   N   L   E   L   A   F    450
1561  CATCATCATCTCTGCAAGACCCATCGCCAAGGCATCCTGGCCAAGCTGCCACCCAAGGGG
      H   H   H   L   C   K   T   H   R   Q   G   I   L   A   K   L   P   P   K   G    470
1621  AAGCTTCGGAGCCTGTGCAGTCAGCACGTGGAAAAGCTGCAAACCTTCCAGCACCTCCAC
      K   L   R   S   L   C   S   Q   H   V   E   K   L   Q   T   F   Q   H   L   H    490
1681  CCCATCGTGGTCCAAGTTGCTTTCCCTCCACTCTATAAGGAACTCTTCAGCACTGAAATC
      P   I   V   V   Q   V   A   F   P   P   L   Y   K   E   L   F   S   T   E   I    510
1741  GAGTCATCTGATGGGCTGTCCGAG**TGA**CCTGGAAACAGGAGTCTGTTCCCTTCCCTAGAG
      E   S   S   D   G   L   S   E                                                518
1801  CCTGCTGGCTTACTTCCCCCTCGGCCTTTTCCTTTCCTGTACACCCTGGAAG**G**TGGTCTC
1861  TGCCTGTTTTTGGAGGGGGTGAGGGCGAGCAAATGCAGAGACTGATTTTCTGCCCACTGG
1921  GCTGCCAGGCAGCAGGCCAAGAGCCAGAGGGTGGGGATAGAGGCACGCTGTCTCCAGCCT
1981  CGGCTTTGTCCTGTCTCGTGTCCCATACCCTGTCACCCCCAGCTTCTGGGAGGCCTGGGG
2041  TGGGATACAAGGACCTCTAGGAGGACCTAGGTGTCCTCAGGACAGCAGGAGGATCCAGGA
2101  CACACGGGACAAATGGAACTCAACTCTGGGCTCAGAAGCTAAGAATAGGCCTTTGAAATA
2161  CCTCATTGCATTTCCCTGGGGGCTTTGGCTGGGGAGCTGCATCAAGCTCAGGGGCTGGCG
2221  GCTGGAGCCCAGAAGGAACTGTATATAAC<u>ATAAA</u>TCTGGATCTCTAAAAAAAAAAAAAAAA
2281  AA   2282
```

Figure 5.1. Nucleotide and deduced amino acid sequences of porcine *RORγ-1* transcript.

The numbers on the left and right indicate the nucleotide base and the deduced amino acid in porcine *RORγ-1*, respectively. The open reading frame (ORF) was deduced using the ORF Finder. The deduced amino acid sequence is shown in single-letter code below the nucleotide sequence. The ATG start codon (positions 211-213) and the TGA stop codon (positions 1765-1767) are printed in bold. The predicted DNA-binding domain (amino acid residues 7-78) is boxed. The putative ligand-binding domain of porcine RORγ-1 is underlined. The polyadenylation signal is underlined.

```
   1  TGTGTGGTGCGGAGCTTAAACCCCCCTGCCCAGAAGCATTGGGGGAGAGCTAGGTGCAGA
  61  GCTGCAGGCTGAGGCTCTGCTGAGAGGGCCTCACCCCCAACCCCGCTGCCAGCTGCACCC
 121  CATCCCTGGACCACCCCCCGCTAAGAAGGACGGGAGCCCAGTCGGCAAAGCCCACGGCTC
 181  AGTCATGAGAACACAAATTGAAGTGATCCCCTGCAAAATCTGTGGGGACAAGTCGTCTGG
           M  R  T  Q  I  E |V  I  P  C  K  I  C  G  D  K  S  S  G|   19
 241  GATCCACTACGGGGTTATCACCTGTGAGGGGTGCAAGGGTTTCTTCCGCCGGAGCCAGCA
          |I  H  Y  G  V  I  T  C  E  G  C  K  G  F  F  R  R  S  Q  Q|  39
 301  GTGTAACGTGGCCTACTCCTGCACCCGTCAGCAGAACTGCCCCATCGACCGCACTAGCCG
          |C  N  V  A  Y  S  C  T  R  Q  Q  N  C  P  I  D  R  T  S  R|  59
 361  AAACCGATGCCAGCACTGCCGCCTACAGAAATGCCTGGCCCTGGGCATGTCCCGAGATGC
          |N  R  C  Q  H  C  R  L  Q  K  C  L  A  L  G  M  S  R  D| A   79
 421  TGTCAAGTTTGGCCGCATGTCCAAGAAGCAAAGGGACAGCCTGCATGCTGAGGTGCAGAA
           V  K  F  G  R  M  S  K  K  Q  R  D  S  L  H  A  E  V  Q  K   99
 481  ACAGCTGCAGCAGAGGCAACAGCAGCAACGGGAACAAGCGGCCAAGTCCCCTCCTGCAGG
           Q  L  Q  Q  R  Q  Q  Q  R  E  Q  A  A  K  S  P  P  A  G  119
 541  GGCCCAAGGAGCAGACCCCCTCGCCTGCCCCCTGGGGCTCCCAGATGGGCAGCTGCCCCT
           A  Q  G  A  D  P  L  A  C  P  L  G  L  P  D  G  Q  L  P  L  139
 601  GGGCTCCTCACCTGACCTGCCGGAGGCCTCGGCCTGTCCCCCCGGTCTCCTGAGAGCCCC
           G  S  S  P  D  L  P  E  A  S  A  C  P  P  G  L  L  R  A  P  159
 661  AAGCTGCGGGCCCTCCTACTCCAACAGCTTGGTCAAGGCTGGGCTCAATGGGGCTTCTTA
           S  C  G  P  S  Y  S  N  S  L  V  K  A  G  L  N  G  A  S  Y  179
 721  CCACCTGGAATACAGCCCTGAGCGGGGCAAGGCTGAGGGCAGAGAGAACTTCTATGGCAC
           H  L  E  Y  S  P  E  R  G  K  A  E  G  R  E  N  F  Y  G  T  199
 781  AGGCAGCCAGCTGGCCCCCGACAGGTGTGGACTCCATTTTGAGGACCCCAGGCATCCTGG
           G  S  Q  L  A  P  D  R  C  G  L  H  F  E  D  P  R  H  P  G  219
 841  GCTTGGGGAACCAGGACGGGGCCCTGACAGCTACTGCAGCCCCAGTTTCCGCAGTACCCC
           L  G  E  P  G  R  G  P  D  S  Y  C  S  P  S  F  R  S  T  P  239
 901  AGAGGCACCTTACGCCTCCCTGACAGAGATTGAGCACCTGGTGCAGAATGTTTGTAAGTC
           E  A  P  Y  A  S  L  T  E  I  E  H  L  V  Q  N  V  C  K  S  259
 961  TTACCGAGACACGTGTCAGCTGCGGCTGGAGGACCTGCTTCGGCAGCGCTCCAACATCTT
           Y  R  D  T  C  Q  L  R  L  E  D  L  L  R  Q  R  S  N  I  F  279
1021  CTCCCGAGAGGAGGTGGCCGGCTACCAGAGGAAGTCAATGTGGGAGATGTGGGAACGCTG
           S  R  E  E  V  A  G  Y  Q  R  K  S  M  W  E  M  W  E  R  C  299
1081  TGCCCACCGCCTCACCGAGGCCATTCAGTACGTGGTGGAGTTCGCTAAGAGGCTCTCGGG
           A  H  R  L  T  E  A  I  Q  Y  V  V  E  F  A  K  R  L  S  G  319
1141  CTTTATGGAGCTCTGCCAGAACGACCAGATCGTGCTACTCAAAGCAGGAGCAATGGAAGT
           F  M  E  L  C  Q  N  D  Q  I  V  L  L  K  A  G  A  M  E  V  339
1201  GGTGCTGGTCAGGATGTGCCGAGCCTACAACGCTGACAACCACACAGTCTTTTTTGAAGG
           V  L  V  R  M  C  R  A  Y  N  A  D  N  H  T  V  F  F  E  G  359
1261  CAAATACGGTGGCATGGAGCTGTTCCGAGCCTTGGGCTGCAGCGAGCTCATCAGCTCCAT
           K  Y  G  G  M  E  L  F  R  A  L  G  C  S  E  L  I  S  S  I  379
1321  CTTTGACTTCTCCCGCTCCCTGAGTGCCTTGCGCTTTTCCGAGGACGAGATTGCACTCTA
           F  D  F  S  R  S  L  S  A  L  R  F  S  E  D  E  I  A  L  Y  399
1381  CACAGCCCTCGTCCTCATCAATGCCAACCGGCCAGGACTCCAAGAGAAAAGGAAAGTAGA
           T  A  L  V  L  I  N  A  N  R  P  G  L  Q  E  K  R  K  V  E  419
1441  ACAGCTGCAGTACAATCTGGAGCTGGCCTTTCATCATCATCTCTGCAAGACCCATCGCCA
```

```
          Q   L   Q   Y   N   L   E   L   A   F   H   H   H   L   C   K   T   H   R   Q   439
1501  AGGCATCCTGGCCAAGCTGCCACCCAAGGGGAAGCTTCGGAGCCTGTGCAGTCAGCACGT
          G   I   L   A   K   L   P   P   K   G   K   L   R   S   L   C   S   Q   H   V   459
1561  GGAAAAGCTGCAAACCTTCCAGCACCTCCACCCCATCGTGGTCCAAGTTGCTTTCCCTCC
          E   K   L   Q   T   F   Q   H   L   H   P   I   V   V   Q   V   A   F   P   P   479
1621  ACTCTATAAGGAACTCTTCAGCACTGAAATCGAGTCATCTGATGGGCTGTCCGAGTGACC
          L   Y   K   E   L   F   S   T   E   I   E   S   S   D   G   L   S   E       497
1681  TGGAAACAGGAGTCTGTTCCCTTCCCTAGAGCCTGCTGGCTTACTTCCCCCTCGGCCTTT
1741  TCCTTTCCTGTACACCCTGGAAGGTGGTCTCTGCCTGTTTTTGGAGGGGGTGAGGGCGAG
1801  CAAATGCAGAGACTGATTTTCTGCCCACTGGGCTGCCAGGCAGCAGGCCAAGAGCCAGAG
1861  GGTGGGGATAGAGGCACGCTGTCTCCAGCCTCGGCTTTGTCCTGTCTCGTGTCCCATACC
1921  CTGTCACCCCCAGCTTCTGGGAGGCCTGGGGTGGGATACAAGGACCTCTAGGAGGACCTA
1981  GGTGTCCTCAGGACAGCAGGAGGATCCAGGACACACGGGACAAATGGAACTCAACTCTGG
2041  GCTCAGAAGCTAAGAATAGGCCTTTGAAATACCTCATTGCATTTCCCTGGGGGCTTTGGC
2101  TGGGGAGCTGCATCAAGCTCAGGGGCTGGCGGCTGGAGCCCAGAAGGAACTGTATATAAC
2161  ATAAATCTGGATCTCTAAAAAAAAAAAAAAAAAAA      2193
```

Figure 5.2. Nucleotide and deduced amino acid sequences of porcine *RORγ-2* transcript.

The numbers on the left and right indicate the nucleotide base and the deduced amino acid in porcine *RORγ-2*, respectively. The ORF was deduced using the ORF Finder. The deduced amino acid sequence is shown in single-letter code below the nucleotide sequence. The ATG start codon (positions 185-187) and the TGA stop codon (positions 1676-1678) are printed in bold. The predicted DNA-binding domain (amino acid residues 7-78) is boxed. The putative ligand-binding domain of porcine RORγ-2 is underlined. The polyadenylation signal is underlined.

```
1      GAGTTGCCCCATGGACAGGGCCCCACAGAGACACCACCGAGCCTCGAGGGAGCTGCTGGC
               M  D  R  A  P  Q  R  H  H  R  A  S  R  E  L  L  A   17
61     TGCAAAGAAGACCCACACCTCACAAATTGAAGTGATCCCCTGCAAAATCTGTGGGGACAA
        A  K  K  T  H  T  S  Q  I  E │V  I  P  C  K  I  C  G  D  K  37
121    GTCGTCTGGGATCCACTACGGGGTTATCACCTGTGAGGGGTGCAAGGGTTTCTTCCGCCG
        S  S  G  I  H  Y  G  V  I  T  C  E  G  C  K  G  F  F  R  R  57
181    GAGCCAGCAGTGTAACGTGGCCTACTCCTGCACCCGTCAGCAGAACTGCCCCATCGACCG
        S  Q  Q  C  N  V  A  Y  S  C  T  R  Q  Q  N  C  P  I  D  R  77
241    CACTAGCCGAAACCGATGCCAGCACTGCCGCCTACAGAAATGCCTGGCCCTGGGCATGTC
        T  S  R  N  C  Q  H  C  R  L  Q  K  C  L  A  L  G  M  S  97
301    CCGAGATGCTGTCAAGTTTGGCCGCATGTCCAAGAAGCAAAGGGACAGCCTGCATGCTGA
       │R  D│ A  V  K  F  G  R  M  S  K  K  Q  R  D  S  L  H  A  E  117
361    AGTGCAGAAACAGCTGCAGCAGAGGCAACAGCAGCAACGGGAACAAGCGGCCAAGTCCCC
        V  Q  K  Q  L  Q  Q  R  Q  Q  Q  R  E  Q  A  A  K  S  P  137
421    TCCTGCAGGGGCCCAAGGAGCAGACCCCCTCGCCTGCCCCCTGGGGCTCCCAGATGGGCA
        P  A  G  A  Q  G  A  D  P  L  A  C  P  L  G  L  P  D  G  Q  157
481    GCTGCCCCTGGGCTCCTCACCTGACCTGCCGGAGGCTTCGGCCTGTCCCCCCGGTCTCCT
        L  P  L  G  S  S  P  D  L  P  E  A  S  A  C  P  P  G  L  L  177
541    GAGAGCCCCCAAGCTGCGGGCCCTCCTACTCCAACAGCTTGGTCAAGGCTGGGCTCAATGG
        R  A  P  S  C  G  P  S  Y  S  N  S  L  V  K  A  G  L  N  G  197
601    GGCTTCTTACCACCTGGAATACAGCCCTGAGAGGGGCAAGGCTGAGGGCAGAGAGAACTT
        A  S  Y  H  L  E  Y  S  P  E  R  G  K  A  E  G  R  E  N  F  217
661    CTATGGCACAGGCAGCCAGCTGGCCCCCGACAGGTGTGGACTCCGTTTTGAGGACCCCAG
        Y  G  T  G  S  Q  L  A  P  D  R  C  G  L  R  F  E  D  P  R  237
721    GCATCCTGGGCTTGGGGAACCAGGACGGGGCCCTGACAGCTACTGCAGCCCCAGTTTCCG
        H  P  G  L  G  E  P  G  R  G  P  D  S  Y  C  S  P  S  F  R  257
781    CAGTACCCCAGAGGCACCTTACGCCTCCCTGACAGAGATTGAGCACCTGGTGCAGAATGT
        S  T  P  E  A  P  Y  A  S  L  T  E  I  E  H  L  V  Q  N  V  277
841    TTGTAAGTCTTACCGAGACACGTGTCAGCTGCGGCTGGAGGACCTGCTTCGGCAGCGCTC
        C  K  S  Y  R  D  T  C  Q  L  R  L  E  D  L  L  R  Q  R  S  297
901    CAACATCTTCTCCCGAGAGGAGGTGGCCGGCTACCAGAGGAAGTCAATGTGGGAGATGTG
        N  I  F  S  R  E  E  V  A  G  Y  Q  R  K  S  M  W  E  M  W  317
961    GGAACGCTGTGCCCACCGCCTCACCGAGGCCATTCAGTACGTGGTAGAGTTCGCTAAGAG
        E  R  C  A  H̲  R̲  L̲  T̲  E̲  A̲  I̲  Q̲  Y̲  V̲  V̲  E̲  F̲  A̲  K̲  R̲  337
1021   GCTCTCGGGCTTTATGGAGCTCTGCCAGAACGACCAGATCGTGCTACTCAAAGCAGGAAC
        L̲  S̲  G̲  F̲  M̲  E̲  L̲  C̲  Q̲  N̲  D̲  Q̲  I̲  V̲  L̲  L̲  K̲  A̲  G̲  T̲  357
1081   ATGGAAGTGGTGCTGGTCAGGATGTGCCGAGCCTACAACGCTGACAACCACACAGTCTTT
        W̲  K̲  W̲  C̲  W̲  S̲  G̲  C̲  A̲  E̲  P̲  T̲  T̲  L̲  T̲  T̲  T̲  Q̲  S̲  F̲  377
1141   TTTGAAGGCAAATACGGTGGCATGGAGCTGTTCCGAACCTTGGGCTGCAGCGAGCTCATC
        L̲  K̲  A̲  N̲  T̲  V̲  A̲  W̲  S̲  C̲  S̲  E̲  P̲  W̲  A̲  A̲  A̲  S̲  S̲  S̲  397
1201   AGCTCCATCTTTGACTTCTCCCGCTCCCTGAGTGCCTTGCGCTTTTCCGAGGACGAGATT
        A̲  P̲  S̲  L̲  T̲  S̲  P̲  A̲  P̲                                      406
1261   GCACTCTACACAGCCCTCGTCCTCATCAATGCCAACCGGCCAGGACTCCAAGAGAAAAGG
1321   AAAGTAGAACAGCTGCAGTACAATCTGGAGCTGGCCTTTCATCATCATCTCTGCAAGACC
1381   CATCGCCAAGGCATCCTGGCCAAGCTGCCACCCAAGGGGAAGCTTCGGAGCCTGTGCAGT
1441   CAGCACGTGGAAAAGCTGCAAAACCTTCCAGCACCTCCACCCCATCGTGGTCCAAGTTGCT
```

```
1501  TTCCCTCCACTCTATAAGGAACTCTTCAGCACTGAAATCGAGTCATCTGATGGGCTGTCC
1561  GAGTGACCTGGAAACAGGGGTCTGTTCCCTTCCCTAGAGCCTGCTGGCTTACTTCCCCCT
1621  CGGCCTTTTCCTTTCCTGTACACCCTGGAAGGTGGTCTCTGCCTGTTTTTGGAGGGGGTG
1681  AGGGCGAGCAAATGCAGAGACTGATTTTCTGCCCACCGGGCTGCCAGGCAGCAGGCCAAG
1741  AGCCAGAGGGTGGGGATAGAGGCACGCTGTCTCCAGCCTCGGCTTTGTCCTGTCTCGTGT
1801  CCCATACCCTGTCACCCCCAGCTTCTGGGAGGCCTGGGGTGGGATACAAGGACCTCTAGG
1861  AGGACCTAGGTGTCCTCAGGACAGCAGGAGGATCCAGGACACACGGGACAAATGGAACTC
1921  AACTCTGGGCTCAGAAGCTAAGAATAGGCCTTTGAAATACCTCATTGCATTTCCCTGGGG
1981  GGCTTTGGCTGGGGAGCTGCAACAAGCTCAGGGGCTGGCGGCTGGAGCCCAGAAGGAACT
2041  GTATATAACATAAATCTGGAAAAAAAAAAAAAAAA    2075
```

Figure 5.3. Nucleotide and deduced amino acid sequences of porcine *RORγ-3* transcript.

The numbers on the left and right indicate the nucleotide base and the deduced amino acid in porcine *RORγ-3*, respectively. The ORF was deduced using the ORF Finder. The deduced amino acid sequence is shown in single-letter code below the nucleotide sequence. The ATG start codon (positions 11-213) and the TGA stop codon (positions 1229-1231) are printed in bold. The predicted DNA-binding domain (amino acid residues 28-99) is boxed. The putative ligand-binding domain of porcine RORγ-3 is underlined. The polyadenylation signal is underlined.

```
1     GGATTGAAGTGATCCCCTGCAAATCTGTGGGGACAAGTCGTCTGGGATCCACTACGGGGT
61    TATCACCTGTGAGGGGTGCAAGGGTTTCTTCCGCCGGAGCCAGCAGTGTAACGTGGCCTA
121   CTCCTGCACCCGTCAGCAGAACTGCCCCATCGACCGCACTAGCCGAAACCGATGCCAGCA
181   CTGCCGCCTACAGAAATGCCTGGCCCTGGGCATGTCCCGAGATGGAGCAATGGAAGTGGT
                                             M   S   R   D   G   A   M   E   V   V    10
241   GCTGGTCAGGATGTGCCGAGCCTACAACGCTGACAACCACACAGTCTTTTTTGAAGGCAA
       L   V   R   M   C   R   A   Y   N   A   D   N   H   T   V   F   F   E   G   K    30
301   ATACGGTGGCATGGAGCTGTTCCGAGCCTTGGGCTGCAGCGAGCTCATCAGCTCCATCTT
       Y   G   G   M   E   L   F   R   A   L   G   C   S   E   L   I   S   S   I   F    50
361   TGACTTCTCCCGCTCCCTGAGTGCCTTGCGCTTTTCCGAGGACGAGATTGCACTCTACAC
       D   F   S   R   S   L   S   A   L   R   F   S   E   D   E   I   A   L   Y   T    70
421   AGCCCTCGTCCTCATCAATGCCAACCGGCCAGGACTCCAAGAGAAAAGGAAAGTAGAACA
       A   L   V   L   I   N   A   N   R   P   G   L   Q   E   K   R   K   V   E   Q    90
481   GCTGCAGTACAATCTGGAGCTGGCCTTTCATCATCATCTCTGCAAGACCCATCGCCAAGG
       L   Q   Y   N   L   E   L   A   F   H   H   H   L   C   K   T   H   R   Q   G    110
541   CATCCTGGCCAAGCTGCCACCCAAGGGGAAGCTTCGGAGCCTGTGCAGTCAGCACGTGGA
       I   L   A   K   L   P   P   K   G   K   L   R   S   L   C   S   Q   H   V   E    130
601   AAAGCTGCAAACCTTCCAGCACCTCCACCCCATCGTGGTCCAAGTTGCTTTCCCTCCACT
       K   L   Q   T   F   Q   H   L   H   P   I   V   V   Q   V   A   F   P   P   L    150
661   CTATAAGGAACTCTTCAGCACTGAAATCGAGTCATCTGATGGGCTGTCCGAGTGACCTGG
       Y   K   E   L   F   S   T   E   I   E   S   S   D   G   L   S   E    167
721   AAACAGGAGTCTGTTCCCTTCCCTAGAGCCTGCTGGCTTACTTCCCCCTCGGCCTTTTCC
781   TTTCCTGTACACCCTGGAAGGTGGTCTCTGCCTGTTTTTGGAGGGGGTGAGGGCGAGCAA
841   ATGCAGAGACTGATTTTCTGCCCACTGGGCTGCCAGGCAGCAGGCCAAGAGCCAGAGGGT
901   GGGGATAGAGGCACGCTGTCTCCAGCCTCGGCTTTGTCCTGTCTCGTGTCCCATACCCTG
961   TCACCCCCAGCTTCTGGGAGGCCTGGGGTGGGATACAAGGACCTCTAGGAGGACCTAGGT
1021  GTCCTCAGGACAGCAGGAGGATCCAGGACACACGGGACAAATGGAACTCAACTCTGGGCT
1081  CAGAAGCTAAGAATAGGCCTTTGAAATACCTCATTGCATTTCCCTGGGGGCTTTGGCTGG
1141  GGAGCTGCATCAAGCTCAGGGGCTGGCGGCTGGAGCCCAGAAGGAACTGTATATAACATA
1201  AATCTGGATCTCTAAAAAAAAAAAAAAAAAA   1229
```

Figure 5.4. Nucleotide and deduced amino acid sequences of porcine *RORγ-4* transcript.

The numbers on the left and right indicate the nucleotide base and the deduced amino acid in porcine *RORγ-4*, respectively. The ORF was deduced using the ORF Finder. The deduced amino acid sequence is shown in single-letter code below the nucleotide sequence. The ATG start codon (positions 212-214) and the TGA stop codon (positions 713-715) are printed in bold. The putative ligand-binding domain of porcine RORγ-4 is underlined. The polyadenylation signal is underlined.

Table 5.1. Nucleotide composition of porcine *RORγ* transcript variants.

cDNA	Purine base			Pyrimidine base			Total
	C	G	C+G	A	T	A+T	(bp)
Porcine *RORγ-1*	697	654	59.2%	479	452	40.8%	2282
5'-UTR	73	67	66.7%	39	31	33.3%	210
CDS	490	438	59.6%	321	308	40.4%	1557
3'-UTR	134	149	55%	119	113	45%	515
Porcine *RORγ-2*	662	624	58.6%	488	419	41.4%	2193
5'-UTR	66	56	66.3%	37	25	33.7%	184
CDS	462	419	59%	332	281	41%	1494
3'-UTR	134	149	55%	119	113	45%	515

Note: The nucleotide composition of porcine *RORγ* transcripts was derived from BioEdit. Only transcripts that encode DBD and LBD were used in the analysis.

Table 5.2. Comparison of porcine *RORγ* transcript variants.

Note: Nucleotide sequence identities were determined using the pairwise alignment in BioEdit with default parameters. The percent identity of each transcript was compared to porcine *RORγ-1* transcript. Number in the box indicates nucleotide bases. Box with the same color represents similar nucleotides in each transcript.

5.2.1. 5'-Untranslated Region

We have isolated and sequenced four porcine *RORγ* cDNAs that have unique 5'-UTRs. The 5'-UTR of porcine *RORγ-2* is identical to transcript 1, except it lacks the first 26 base pairs (Table 5.2). Porcine *RORγ-3* and *-4* transcripts have unique 5'-UTRs (Table 5.2). We analyzed the structural and compositional features of porcine *RORγ* 5'-UTRs. Analysis of these 5'-UTRs revealed that the porcine *RORγ-1* has a 5'-UTR of 210 bases with 66.7% G+C content. The porcine *RORγ-2* transcript has 184 bases in

5'-UTR (66.3% G+C). The average length of porcine *RORγ* 5'-UTRs is 154 bp. High GC content was observed in the 5'-UTR of porcine *RORγ* transcripts (approximately 65%) (Table 5.1). In general, 3'-UTR is longer than 5'-UTR and the GC content in 5'-UTRs is higher than that of 3'-UTR.

It has been known that sequence features in 5'-UTR, such as secondary structure or stem-loops, uORFs, upstream AUG, involved in post-transcriptional regulation that control mRNA localization, stability, and translational efficiency (Pesole *et al.,* 2001; Hughes, 2006). Functional elements in 5'-UTRs of porcine *RORγ* were analyzed using dbUTRs (Mignone *et al.,* 2005). Known functional non-coding sequences were not found in the 5'-UTRs of porcine *RORγ* transcripts.

The secondary structures of the 5'-untranslated region (UTR) of porcine *RORγ* transcripts were predicted using web interface of the RNAfold program (Hofacker, 2003). The predicted fold energy of 5'-UTR is -91.6 kcal/mol for porcine RORγ-1 transcript and -74.7 kcal/mol for porcine *RORγ-2* transcript. The predicted secondary structures of porcine *RORγ* 5'-UTR are shown in Figure 5.5 and 5.6.

Figure 5.5. Predicted secondary structure of 5'-UTR of porcine *RORγ-1*.

The secondary structure of 5'-UTR was predicted using the RNAfold program. The predicted fold energy of porcine *RORγ-1* 5'-UTR is -91.6 kcal/mol.
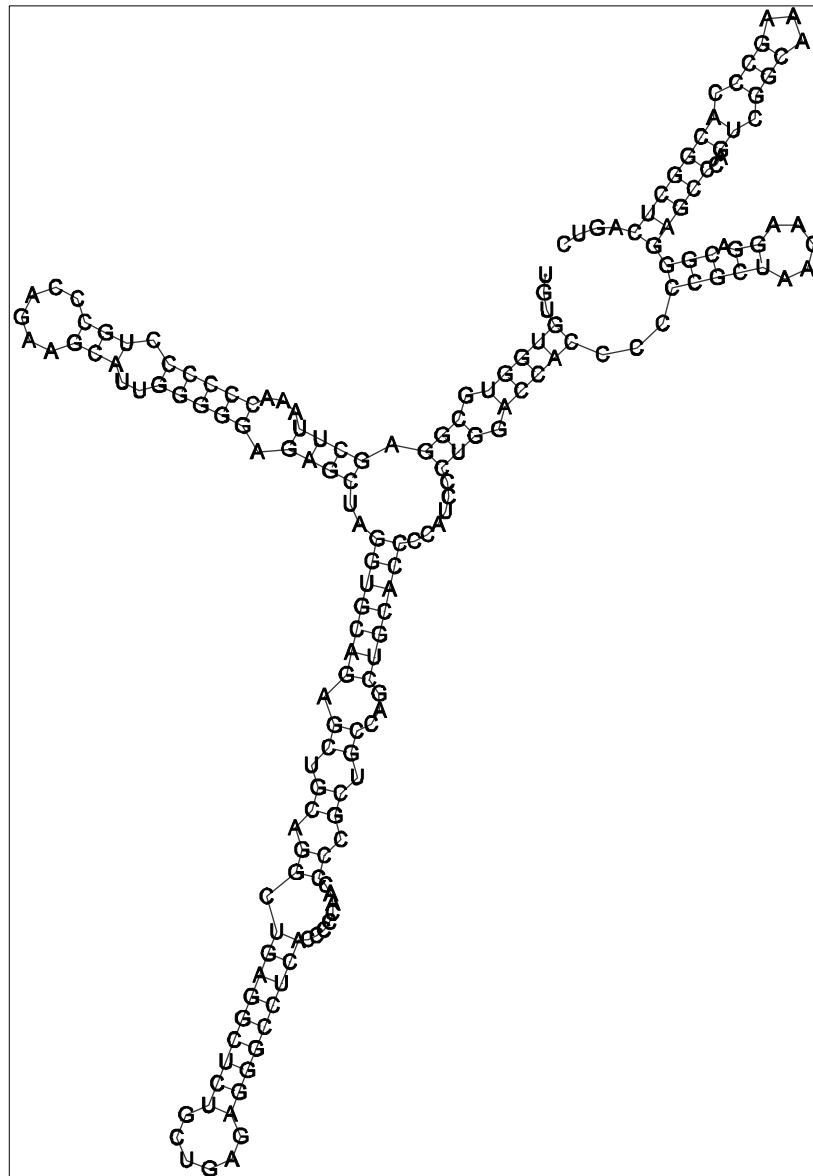
Figure 5.6. Predicted secondary structure of 5'-UTR of porcine *RORγ-2*.

The secondary structure of 5'-UTR was predicted using the RNAfold program. The predicted fold energy of porcine *RORγ-2* 5'-UTR is -74.7 kcal/mol.

5.2.2. Open Reading Frame

The ORFs of full-length porcine *RORγ* mRNAs were deduced using ORF Finder. Analysis of porcine *RORγ-1* ORF revealed that it was on the first open reading frame with the ATG start codon at positions 211-213 and the TGA stop codon at nucleotide

positions 1765-1767. There are two in-frame termination codons (TGA96-98 and TAA168-169) upstream putative ATG initiation codon. The porcine *RORγ-1* has an open reading frame of 1557 nucleotides encoding a protein sequence of 516 residues (nucleotide positions 211-1767). The porcine *RORγ-2* has 1494 bases of ORF (positions 185-1678) located in frame +2, which encodes 497 amino acids. This ORF is started with ATG codon (nucleotide positions 185-187) and stopped with TGA codon (nucleotide positions 1676-1678). There is an in-frame stop codon (TAA17-19) upstream putative ATG start codon of porcine *RORγ-2*. The porcine *RORγ-3* is translated from frame +2 with 1221 bp CDS (nucleotide positions 11-1231) encoding 406 amino acids. It begins at nucleotide position 11 (ATG) and stops at nucleotide position 1231 (TGA). The ORF of porcine *RORγ-4* is translated in frame +2 and contains 504 bp ORF, which encodes 167 amino acid residues.

The overall guanine-cytosine content in ORFs of porcine *RORγ* transcripts is 59.6% for porcine *RORγ-1*, 59% for porcine *RORγ-2*, 60.2% for porcine *RORγ-3*, and 54.6% for porcine *RORγ-4* (Table 5.1). The GC content at third codon positions is 71.3% for porcine *RORγ-1*, 71.9% for porcine *RORγ-2*, 69.5% for porcine *RORγ-3*, and 68.5% for porcine *RORγ-4*. Codon usage bias correlates with the GC content at third codon position. High GC content at the third codon positions was observed in porcine *RORγ-1* ORF (71.3%). The codon usage in porcine *RORγ* ORFs is shown in Table 5.3.

The nucleotide sequence around ATG initiation codon is according with Kozak consensus sequence (GCCA/GCC*ATG*G) (Kozak, 1996). The nucleotide sequence around ATG start codon is TCAGTC*ATG*A for porcine *RORγ-1* and *-2*, TGCCCC*ATG*G for porcine *RORγ-2*, and CTGGGC*ATG*T for porcine *RORγ-4*. This classified sequence flanking porcine *RORγ* ATG initiation codon as an adequate

Kozak consensus sequence. The consensus sequence flanking ATG codons in *RORγ* is shown in Figure 5.7.

Table 5.3. Codon usage in porcine *RORγ* ORFs.

Amino Acid	Codon	*RORγ-1*		*RORγ-2*	
		Number	Frequency	Number	Frequency
Alanine	GCG	1	1.9	1	2
	GCA	8	15.4	6	12
	GCU	7	13.5	8	16.1
	GCC	23	44.3	20	40.2
Cysteine	UGU	10	19.3	8	16.1
	UGC	17	32.8	15	30.1
Aspartic acid	GAU	3	5.8	3	6
	GAC	13	25.1	14	28.1
Glutamic acid	GAG	27	52	25	50.2
	GAA	10	19.3	11	22.1
Phenylalanine	UUU	8	15.4	8	16.1
	UUC	14	27	11	22.1
Glycine	GGG	17	32.8	14	28.1
	GGA	6	11.6	5	10
	GGU	4	7.7	3	6
	GGC	15	28.9	14	28.1
Histidine	CAU	7	13.5	7	14.1
	CAC	10	19.3	9	18.1
Isoleucine	AUA	0	0	0	0
	AUU	4	7.7	4	8
	AUC	13	25.1	13	26.1
Lysine	AAG	13	25.1	17	34.1
	AAA	7	13.5	7	14.1
Leucine	UUG	3	5.8	3	6
	UUA	1	1.9	0	0
	CUG	28	54	27	54.2
	CUA	2	3.9	2	4
	CUU	3	5.8	3	6
	CUC	20	38.5	18	36.1
Methionine	AUG	9	17.3	9	18.1
Asparagine	AAU	4	7.7	4	8.1
	AAC	10	19.3	10	20.1
Proline	CCG	2	3.9	1	2.1

Amino Acid	Codon	RORγ-1		RORγ-2	
		Number	Frequency	Number	Frequency
	CCA	8	15.4	7	14.1
	CCU	9	17.3	8	16.1
	CCC	15	28.9	13	26.1
Glutamine	CAG	21	40.5	25	50.2
	CAA	6	11.6	10	20.1
Arginine	AGG	7	13.5	8	16.1
	AGA	4	7.7	3	6
	CGG	8	15.4	8	16.1
	CGA	7	13.5	7	14.1
	CGU	1	1.9	1	2
	CGC	10	19.3	11	22.1
Serine	AGU	5	9.6	4	8
	AGC	12	23.1	13	26.1
	UCG	4	7.7	3	6
	UCA	3	5.8	3	6
	UCU	7	13.5	4	8
	UCC	16	30.8	15	30.1
Threonine	ACG	2	3.9	1	2
	ACA	6	11.6	5	10
	ACU	2	3.9	2	4
	ACC	6	11.6	6	12.1
Valine	GUG	11	21.2	12	24.1
	GUA	1	1.9	1	2
	GUU	3	5.8	3	6
	GUC	6	11.6	6	12.1
Tryptophan	UGG	3	5.8	2	4
Tyrosine	UAU	2	3.9	2	4
	UAC	14	27	14	28.1
Stop codon	UGA	1	1.9	1	2

Note: The codon usage in porcine *RORγ-1* ORF was derived from SMS. Porcine *RORγ-1* has 59.6% of GC content in its ORF. The GC content in the first, second, and third codon positions are 60.1%, 47.4%, and 71.3%, respectively. Porcine *RORγ-2* has 59% of GC content in its ORF. The GC content in the first, second, and third positions are 60.8 %, 44.2%, and 71.9%, respectively. Only transcripts that encode the complete DBD and LBD were used in the analysis.
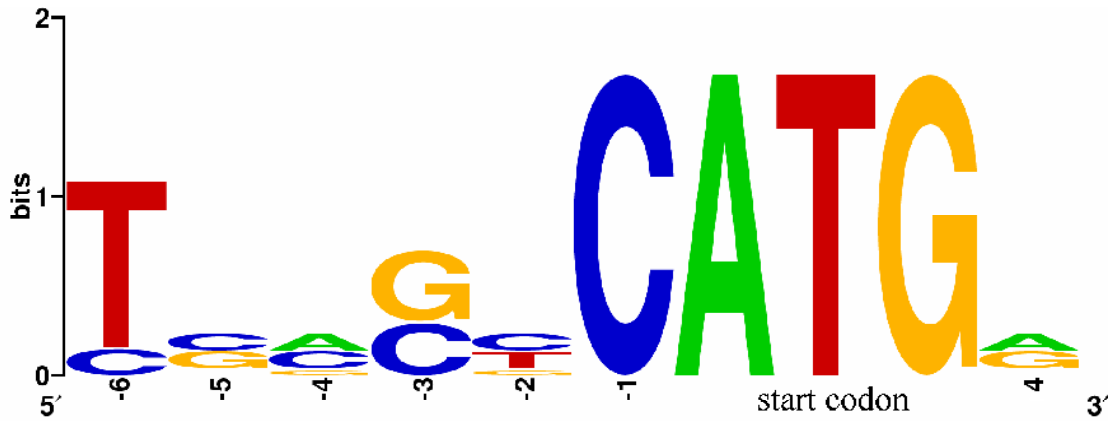
Figure 5.7. Kozak consensus sequence surrounding ATG start codon of *RORγ* in farm animals and human.

The nucleotide sequence around ATG start codon is TCA*G*TC*ATG*A for porcine *RORγ-1* and *-2*, TGC*C*CC*ATG*G for porcine *RORγ-2*, and CTG*GG*C*ATG*T for porcine *RORγ-4*. This classified sequence flanking porcine *RORγ* ATG initiation codon as an adequate Kozak consensus sequence. The WebLogo was used in the create diagram representing the Kozak consensus sequence.

5.2.3. 3'-Untranslated Region

The 3'-UTRs of porcine *RORγ-1* and *-2*, *-3*, and *-4* were 515, 844, and 514 bases in length, respectively. These 3'-UTRs are similar, except in the length of poly(A) tail. Porcine *RORγ-3* has unique 3'-UTR because of an in-frame TGA stop codon (nucleotide positions 1229-1231), resulting in the addition of 335 bases in ORF to 3'-UTR, and the ATCTCT deletion. The G+C contents in 3'-UTR of porcine *RORγ* transcripts are approximately 55% (Table 5.1).

The analysis of 3'-UTR of porcine *RORγ* transcripts did not reveal the presence of known regulatory elements (e.g., AU-rich elements (AREs)) involved in post-transcriptional regulation such as mRNA localization, stability, and translational efficiency (Pesole *et al.,* 2001; Hughes, 2006). The secondary structures of the 3'-UTR of porcine *RORγ* transcripts were predicted using the RNAfold program. The predicted fold energy of 3'-UTR is -214.7 kcal/mol for porcine *RORγ-1* and *-2*

transcripts. The predicted secondary structure of porcine *RORγ* 3'-UTR is shown in Figure 5.8.
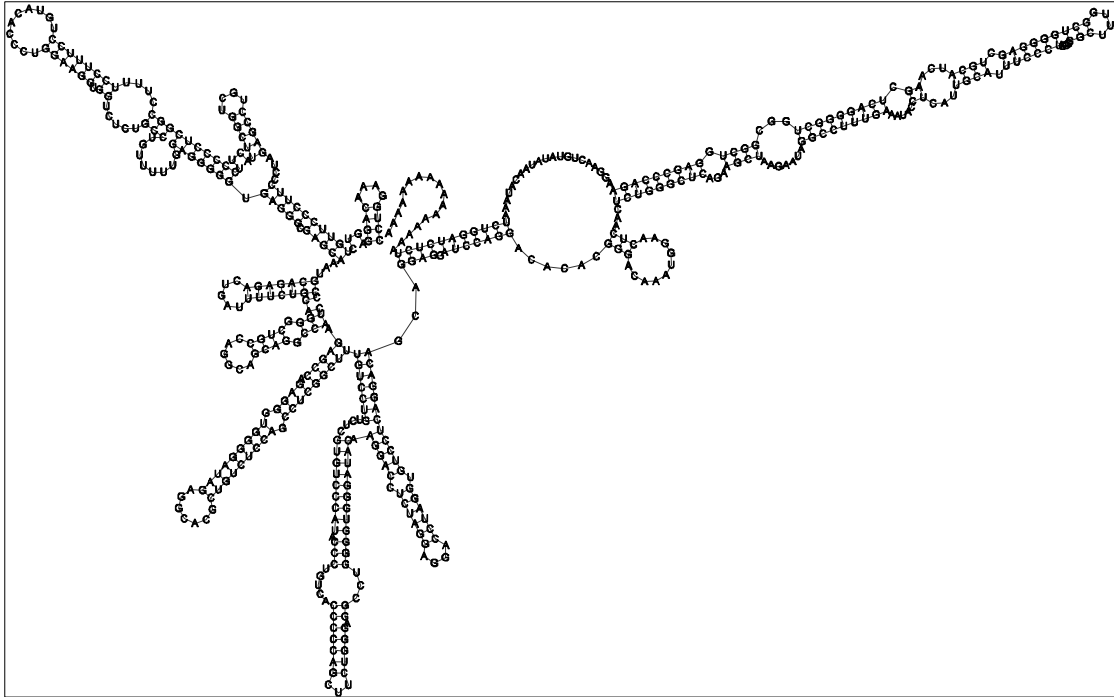
Figure 5.8. Predicted secondary structure in the 3'-UTRs of porcine *RORγ-1* and *-2*.

The secondary structure of 3'-UTR was predicted using the RNAfold program. The predicted fold energy of porcine *RORγ* 3'-UTR is -214.7 kcal/mol.

5.3. Predicted Amino Acid Sequences of Porcine RORγ

Four protein isoforms of porcine RORγ receptor were predicted from full-length cDNA clones. The predicted porcine RORγ-1 protein has 518 amino acid residues with a calculated molecular weight of 57376 Daltons and an isoelectric point of 8.07. The porcine RORγ-2 protein has 497 amino acid residues with a calculated molecular weight of 55776 Daltons and an isoelectric point of 8.60. The predicted RORγ-3 protein has 406 amino acid residues with a calculated molecular weight of 45015 Daltons and an isoelectric point of 8.90. The predicted RORγ-4 protein has 167 amino

acid residues with a calculated molecular weight of 19023 Daltons and an isoelectric point of 7.09. Amino acid composition of these porcine RORγ isoforms is shown in Table 5.4.

Comparison of these porcine RORγ isoforms revealed that porcine RORγ-1 and 2 differ in the N-terminal of the D domain (additional 21 amino acid residues in RORγ-1: GEAMGAALWGFPGVSRGAAEL) (Figures 5.9 and 5.10). The isoform 3 has an additional 21-amino acid in the A/B domain and 106-amino acid residues deletion in the LBD. The porcine RORγ isoform 4 is a shortest porcine RORγ that lacking A/B, the most of DBD, D domain, and the part of LBD (Figure 5.9).

Protein sequence analysis indicated that porcine RORγ-1 isoform shares 90% identity with RORγ isoform 2, 79% identity with bovine RORγ, approximately 80% identity with human RORγ. The porcine RORγ-2 isoform exhibits 88% identity with bovine RORγ, and ~90% identity with human RORγ (Table 5.5). Comparison with human RORγ receptors, porcine RORγ-1 and -2 exhibit higher sequence identity with human RORγ-2 isoform, which is an ortholog of mouse RORγt, than human RORγ-1. Other two isoforms, porcine RORγ-3 and -4, are unique for pig and represent protein isoforms that contain the deletion in DBD and/or LBD. Protein sequence identity for pig, cow, and human RORγ is summarized in Table 5.5.

Table 5.4. Amino acid composition of porcine RORγ proteins.

Amino acid	RORγ-1		RORγ-2	
	Number	Percentage	Number	Percentage
Alanine (A)	39	7.5%	35	7.0%
Arginine (R)	37	7.1%	38	7.6%
Asparagine (N)	14	2.7%	14	2.8%
Aspartic acid (D)	16	3.1%	17	3.4%
Cysteine (C)	27	5.2%	23	4.6%
Glutamine (Q)	27	5.2%	35	7.0%
Glutamic acid (E)	37	7.1%	36	7.2%
Glycine (G)	42	8.1%	36	7.2%
Histidine (H)	17	3.3%	16	3.2%
Isoleucine (I)	17	3.3%	17	3.4%
Leucine (L)	57	11.0%	53	10.7%
Lysine (K)	20	3.9%	24	4.8%
Methionine (M)	9	1.7%	9	1.8%
Phenylalanine (F)	22	4.2%	19	3.8%
Proline (P)	34	6.6%	29	5.8%
Serine (S)	47	9.1%	42	8.5%
Threonine (T)	16	3.1%	14	2.8%
Tryptophan (W)	3	0.6%	2	0.4%
Tyrosine (Y)	16	3.1%	16	3.2%
Valine (V)	21	4.1%	22	4.4%

Note: Amino acid composition was calculated using ProParam. Number indicates frequency of each amino acid residue. Percentage represents calculated molecular weight of each amino acid residue in porcine RORγ isoforms.
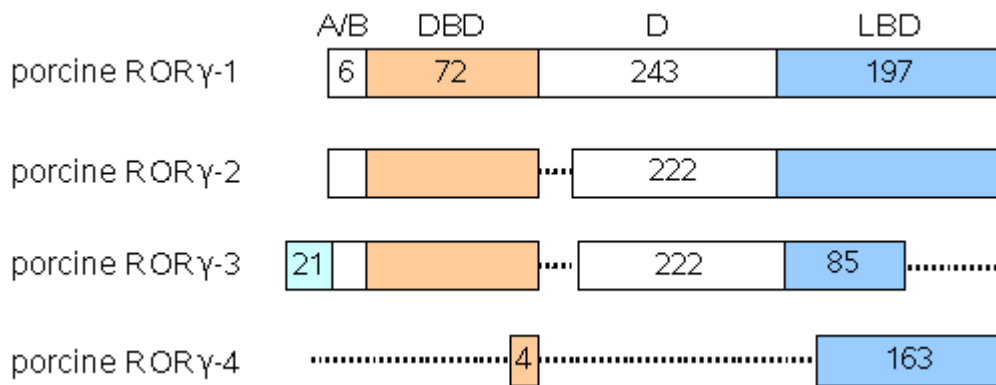
Figure 5.9. Comparison of porcine RORγ isoforms.

Four protein isoforms of porcine RORγ receptor were predicted from full-length cDNA clones.The longest isoform 1 with 518 amino acid residues is identical to isoform 2, except it has additional 21-amino acids in the D domain. The isoforms 3 and 4 are unique for pigs, which have a deletion in DBD and/or LBD. The dot line represents amino acid deletion. Four structural domains (A/B, DBD, D, and LBD) are indicated.

```
porcine ROR γ -1  79 GEAMGAALWGFPGVSRGAAELLCQTGLACKSAFLELCSPSFPRAGPCPFHSHQRSST 135
porcine ROR γ -2  79 --------------------AVKFGRMSKKQRDSLHAEVQKQLQQRQQQQREQAAK 114
porcine ROR γ -3 100 --------------------AVKFGRMSKKQRDSLHAEVQKQLQQRQQQQREQAAK 135
```

Figure 5.10. Comparison of different D domain in porcine RORγ isoforms.

The amino acids in the N-terminal of D domain (57 amino acid residues) are unique for porcine RORγ-1 isoform. It has 21 amino acids added to hinge D domain. Amino acid is colored according to its property.

Table 5.5. Sequence identity in the RORγ proteins between farm animals and humans.

Sequence	% identity	Length
porcine RORγ-1: porcine RORγ-2	90	518-497
porcine RORγ-1: bovine RORγ	79	518-506
porcine RORγ-1: human RORγ-1	77	518-518
porcine RORγ-1: human RORγ-2	81	518-497
porcine RORγ-2: bovine RORγ	88	497-506
porcine RORγ-2: human RORγ-1	87	497-518
porcine RORγ-2: human RORγ-2	91	497-497

Note: Percent identity in amino acid sequences was calculated using pairwise alignment in BioEdit with default parameters (using BLOSUM62 as the scoring matrix with open gap extension penalty of 10 and gap extension penalty of 0.5).

5.3.1. Domain Structure

Using the SMART domain analysis tool (Schultz *et al.,* 1998; Letunic *et al.,* 2004), porcine RORγ protein was predicted to possess the evolutionarily conserved DBD (amino acid residues 7-78 in RORγ-1 and -2, and 28-99 in RORγ-3) and the moderately conserved LBD (amino acid residues 322-518 in RORγ-1 and 301-497 in RORγ-2) (Figure 5.11, 5.12, and 5.13). The porcine RORγ-3 is identical to isoform 2, except 106-amino acid residues deletion in LBD. The porcine RORγ-4 is a truncate protein that has only partial LBD (Figure 5.11).

The DBD is important in the binding of RORγ to its target gene promoters. It contains two zinc finger motifs that are involved in the specificity of the DNA recognition and in dimerisation (Mangelsdorf *et al.,* 1995; Giguere, 1999; Khorasanizadeh and Rastinejad, 2001; Robinson-Rechavi *et al.,* 2003). In porcine RORγ isoforms, the first zinc finger (CKICGDKSSGIHYGVITCEGC) is located at amino acid residues 10-30 in RORγ-1 and 2 isoforms, 31-51 in RORγ-3 and the second zinc finger (CTRQQNCPIDRTSRNRCQHC) is found at residues 46-65 in RORγ-1 and 2, and 67-86 in RORγ-3. The ScanProsite was used in the prediction of zinc finger motifs. The highly conserved P-box (CEGCKGFF), which is located in the DBD and is found in most nuclear receptors, is also found in the DBD of porcine RORγ proteins (Figure 5.12). The porcine RORγ DBD is 100% and 97% identical to the bovine and human receptors, respectively (Figure 5.11). Compared to human RORγ protein, two amino acid substitutions in the DBD were identified (Gln39Arg59 and Val42Ala62) (Figure 5.12). These substitutions are the replacement of amino acids with similar properties, which do not change in solubility and/or the net charge of the amino acid. With the high percent sequence identity in the DBD between

human and swine, we speculate that porcine RORγ receptors bind and stimulate transcription of target genes similar to human RORγ receptors.

Sequence alignment of the LBD in swine and human is shown in Figure 5.13. Porcine RORγ exhibits 90% identity with cattle and 93% identity with human (Figure 5.11). The ligand-dependent activation-2 (AF-2), important in ligand and coactivator binding, was also found in the LBD (LYKELF) (Figure 5.13). Thirteen amino acid substitutions were found in the LBD of human and swine. Eight amino acid substitutions in the LBD between porcine and human RORγ LBDs were found (Arg323His, His374Arg, Arg405His, Arg411His, Asn429His, Gly461Ser, Lys482Arg, and Val496Ala). These amino acid substitutions are the replacement of amino acids with similar properties. Four amino acid substitutions (Thr485Ile, Ile510Thr, Ser513Pro, Asp514Val, and Glu518Lys) (Figure 5.13) are the replacement of amino acids with different properties.
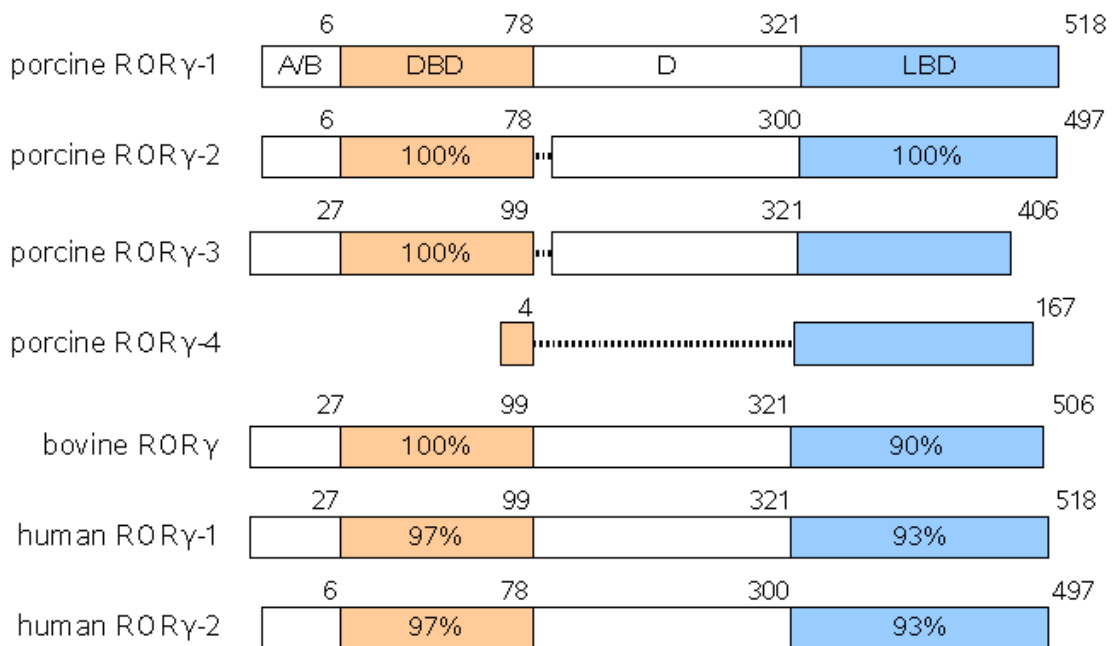
Figure 5.11. Structural domains of porcine RORγ proteins.

Percent identity in amino acid sequences was calculated using pairwise alignment in BioEdit (BLOSUM62 scoring matrix). The structural domains are indicated. Percent sequence identity is shown in each structural domain in comparison with porcine RORγ. Numbers above each structural domain indicate position of amino acid in boundary region.

```
porcine RORγ  7 VIPCKICGDKSSGIHYGVITCEGCKGFFRRSQQCNVAYSCTRQQNCPIDRTSRNRCQHCR 66
human RORγ   28 ...............................R..A........................ 87
                                                * *
porcine RORγ    LQKCLALGMSRD 78
human RORγ      ............ 99
```

Figure 5.12. Alignment of the DBD domains between porcine and human RORγ proteins.

Amino acid is colored according to its property. Dots represent identical amino acid. Different amino acids between pig and human are indicated by an asterisk. The first zinc finger motif and the second zinc finger motif are underlined. Zinc finger motifs in porcine RORγ DBD were predicted using the ScanProsite. The conserved peptide CEGCKGFF located in the highly conserved P-box is boxed. The T box (TRQQN) is unique for TOR. Two conserved amino acid substitution in LBD between porcine and human RORγ DBD were found (Gln39Arg59 and Val42Ala62).

```
porcine RORγ  322 HRLTEAIQYVVEFAKRLSGFMELCQNDQIVLLKAGAMEVVLVRMCRAYNADNHTVFFEGK 381
human   RORγ  322 .H...............................................R....... 381


porcine RORγ      YGGMELFRALGCSELISSIFDFSRSLSALRFSEDEIALYTALVLINANRPGLQEKRKVEQ 441
human   RORγ      .......................H.....H.................H............ 441


porcine RORγ      LQYNLELAFHHHLCKTHRQGILAKLPPKGKLRSLCSQHVEKLQTFQHLHPIVVQVAFPPL 501
human   RORγ –    ...................S...................R..I..........A..... 501


porcine RORγ      YKELFSTEIESSDGLSE 518
human   RORγ      ........T..PV...K 518
```
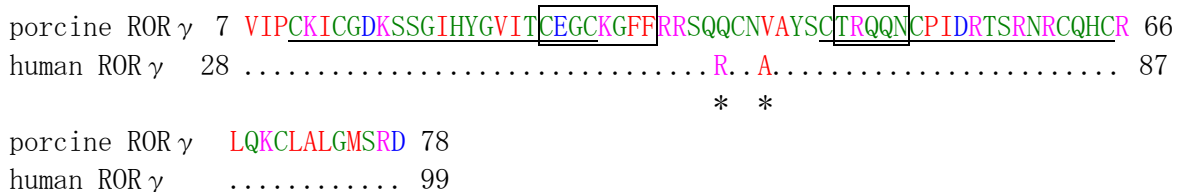
Figure 5.13. Alignment of the LBD domains between porcine and human RORγ proteins.

Amino acid is colored according to its property. The AF-2 (LYKELF) is underlined. Dots represent identical amino acids between swine and human. Different amino acids are indicated by single amino acid codes.

5.3.2. Homology Modeling

The 3-D structure of porcine RORγ protein was predicted using the web service at the Swiss Model Server. The crystal structure of RORα (PDB: 1N83) (Kallen *et al.,* 2002) was identified with 53% sequence identity. The 1N83 template was used in the 3-D homology modeling of porcine RORγ. The predicted 3-D structure of porcine RORγ is shown in Figure 5.14.

The secondary structure of porcine RORγ LBD was predicted to enclose by helix 1 (T268-D283), helix 2 (L289-R294), helix 2' (R302-R310), helix 3 (M313-K336), helix 4 (Q346-A355), helix 5 (G356-R364), helix 6 (M385-L391), helix 7 (S394-S408), helix 8 (E414-L425), helix 9 (K436-T457), helix 10 (L463-V480), helix 11 (E481-P491), helix 11' (V493-P499), helix 12 (Y502-S507) and 3 beta- strands (β-0, β-1, and β-2: Y369-N370, T375-F378, and K381-G383, respectively) (Figure 5.15).
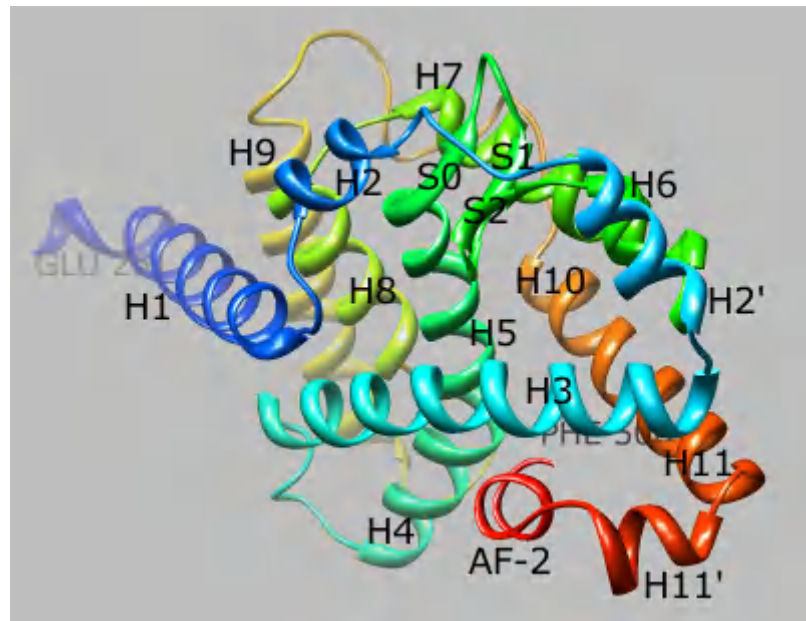
Figure 5.14. Ribbon representation of the predicted 3D structure of porcine RORγ protein.

The 3-D model was predicted using the Swiss Model Automatic Modeling Mode (First Approach mode). The crystal structure of human RORα (PDB template: 1N83) was identified and used in the construction of porcine RORγ 3-D model. The image was produced using the UCSF Chimera. The predicted secondary structural elements are indicated (H: α helix and S: β strand).
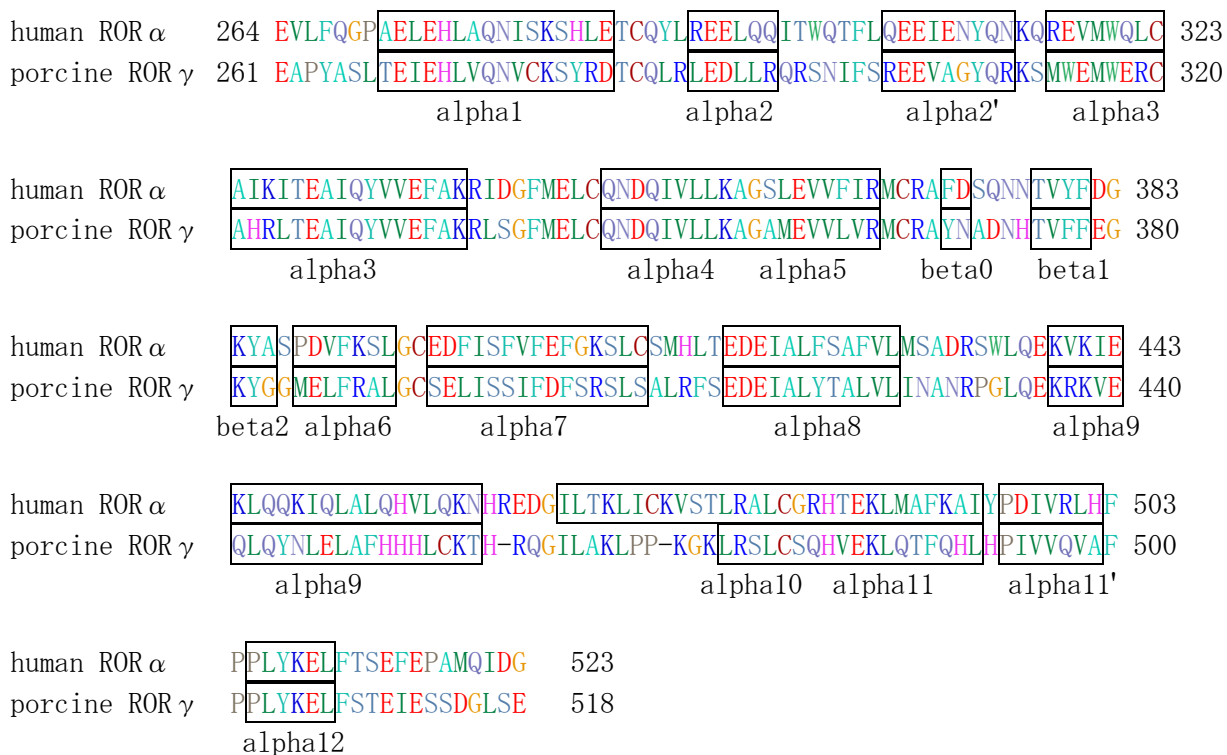
```
human RORα      264 EVLFQGP AELEHLAQNISKSHLE TCQYL REELQQ ITWQTFL QEEIENYQN KQ REVMWQLC 323
porcine RORγ    261 EAPYASL TEIEHLVQNVCKSYRD TCQLR LEDLLR QRSNIFS REEVAGYQR KS MWEMWERC 320
                         alpha1              alpha2       alpha2'      alpha3

human RORα          AIKITEAIQYVVEFAK RIDGFMELC QNDQIVLLK AGSLEVVFIR MCRA FD SQNN TVYF DG 383
porcine RORγ        AHRLTEAIQYVVEFAK RLSGFMELC QNDQIVLLK AGAMEVVLVR MCRA YN ADNH TVFF EG 380
                          alpha3                alpha4    alpha5      beta0    beta1

human RORα          KYA SPDVFKSL GC EDFISFVFEFGKSLC SMHLT EDEIALFSAFVL MSADRSWLQE KVKIE 443
porcine RORγ        KYG GMELFRAL GC SELISSIFDFSRSLS ALRFS EDEIALYTALVL INANRPGLQE KRKVE 440
                    beta2 alpha6       alpha7           alpha8               alpha9

human RORα          KLQQKIQLALQHVLQKN HREDG ILTKLICKVSTLRALCGRHTEKLMAFKAI YPDIVRLH F 503
porcine RORγ        QLQYNLELAFHHHLCKT H-RQG ILAKLPP-KGK LRSLCSQHVEKLQTFQHL HPIVVQVA F 500
                          alpha9                    alpha10  alpha11      alpha11'

human RORα          P PLYKEL FTSEFEPAMQIDG    523
porcine RORγ        P PLYKEL FSTEIESSDGLSE    518
                        alpha12
```

Figure 5.15. Predicted secondary structure of porcine RORγ protein.

The secondary structure of porcine RORγ was derived from the X-ray structure of human RORα LBD (PDB template: 1N83).

5.3.3. Expression Analysis

The expression pattern analysis of the porcine *RORγ-1* and *RORγ-2* transcripts in adult tissues was performed using the RT-PCR method. We designed primer sets that were specific to each porcine *RORγ* splice variant. These primers span between the exon-intron boundaries to prevent the generation of amplified products from genomic DNA. The expression of porcine *RORγ-1* transcript was detected in the liver, kidney, small intestine, heart, muscle, thymus, spleen, and brain, whereas the porcine *RORγ-2* transcript was expressed only in the thymus and spleen. The porcine *RORγ-1* and *-2* transcripts were expressed as DNA bands of 1800 bp and 2000 bp, respectively (Figures 5.16).
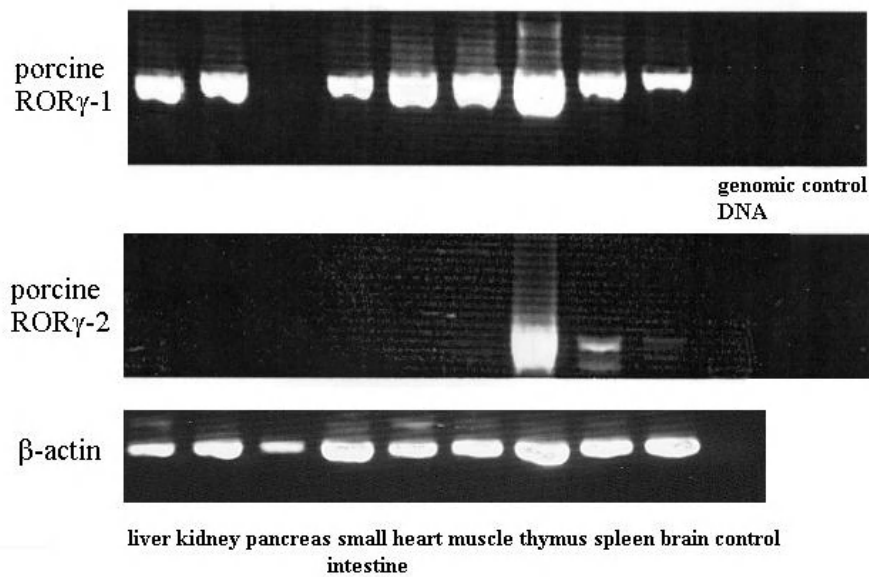
Figure 5.16. Expression analysis of porcine *RORγ* transcripts.

The porcine *RORγ-1* transcript was expressed in the liver, kidney, small intestine, heart, muscle, thymus, spleen, and brain, whereas the expression of porcine *RORγ-2* transcript was detected only in the thymus and spleen.

5.4. Prediction of Exon-Intron Structure of Porcine *RORγ*

To identify genomic DNA corresponds to porcine *RORγ* mRNA sequences, we searched ongoing swine genome sequencing project using pig blast. Hits with threshold were not identified. The entire cDNA sequence of porcine *RORγ* was searched against bovine genomic DNA. The bovine genomic *RORγ* sequence (NC_007301) was identified. This DNA sequence was used as a template in the prediction of porcine *RORγ* gene structure using the SIM4 program (Florea *et al.,* 1998) (Figure 5.17). The bovine *RORγ* gene is located on chromosome 3 with 25705 base pairs in length (region 16672946-16698650). All porcine *RORγ* transcripts were mapped into this chromosome region with possibly different in transcription start site,

resulting from multiple promoter usage and one promoter with different start site. The porcine *RORγ-1* and *-2* are possibly generated from one promoter with different transcription start site. The porcine *RORγ-3* is transcribed from separated promoter located in upstream region. The porcine *RORγ-4* is a transcript that generates from an alternative splicing of porcine *RORγ* primary transcript.

The porcine *RORγ* gene was predicted to have potential 11 exons ranging in size from 30 to 660 bp with an average size of 203 bases. The longest exon is exon 10 with 660 bases. The shortest exon is exon 1C2 with 30 bases. The predicted exon-intron boundaries follow the GT-AG consensus rule for splicing (Mount, 1982). The exon-intron boundaries of porcine *RORγ* transcripts are shown in Tables 5.6, 5.7, 5.8, and 5.9. The 5'-UTR is encoded by exon 1A (*RORγ-1* transcript), 1B (*RORγ-2* transcript), exons 1C1 (*RORγ-3* transcript), or exon 2 and the part of exon 3 (*RORγ-4* transcript). In general, the protein-coding region of porcine *RORγ* is encoded by 9 exons (exons 2-10), except one additional small coding exon (exon 1C2) with 30 base pairs in porcine *RORγ-3* transcript. The ORF of porcine *RORγ-4* is encoded by the part of exon 3, exons 7-9, and the part of exon 10. The ATG start codon is located in exon 1A (porcine *RORγ-1* transcript), 1B (*RORγ-2* transcript), 1C1 (*RORγ-3* transcript) or exon 3 (*RORγ-4* transcript). Similar to most nuclear receptors, the DBD is encoded by two separated exons, exons 2 and 3. The first and second zinc finger motifs are encoded by exon 2 and exon 3, respectively. The LBD is encoded by the part of exon 6, exon 7, exon 8, exon 9, and the part of exon 10. The TGA stop codon is located in exon 10 (transcripts 1, 2, and 4) or exon 8 (transcript 3). The 3'-UTRs of porcine *RORγ-1*, *-2* and *-4* are encoded by exon 10, whereas porcine *RORγ-3* is encoded by exons 8, 9 and 10.

Comparison of *RORγ* exons revealed that the size and number of exons are conserved between farm animals and human, especially in the protein-encoding exons (Figure 5.18). Exons 2 and 3, encoding DBD, are identical in size (86 bp and 142 bp, respectively) and exhibit approximately 95% identity when compared to cattle and human (Figure 5.18). The LBD-encoding exons (exons 6, 7, 8, and 9) are also conserved both in size and sequence identity but slightly lower than DBD-encoding exons (~90% identity). The short encoding exon 1C2, encodes A/B domain, in porcine *RORγ-3* is 97% and 100% identical to exon 2 of cow and human, respectively (Figure 5.18). However, exon 4, encoding the D domain, differs in length and exhibits ~80% identity. The 169-bp sequence (446-614) of porcine *RORγ-1* could not aligned with cow genomic DNA. Alignment of this fragment with porcine *RORγ-2* transcript (nucleotides 420-525) indicated that there are 63 nucleotides insertion, resulting in 21 amino acid residues in the hinge D domain, and this fragment is unique for *RORγ-1* transcript. Porcine *RORγ-2* transcript is 99% identical to *RORγ-3* in exon 4. Alignment of different exon 4 in porcine *RORγ* transcripts is shown Figure 5.19. In exon 5, encoding D domain, porcine *RORγ* shows 93% and 88% identity with cow and human, respectively. In contrast to protein-coding exons, exons encoding untranslated regions (UTRs) are known to have lower sequence homology. Porcine *RORγ* transcripts possess different 5'-UTR encoding exons (217 bp of exon 1A, 191 bp of exon 1B, 50 bp of exon 1C1, and 82 bp of exon 2 and part of exon 3 in transcript 1, 2, 3, and 4, respectively). Exon 1A of porcine *RORγ-1* has a unique 26 base pairs adding to its 5' end (Figure 5.20) when compared to porcine *RORγ-2* transcript. It shares 88% and 15% identity with exon 1B and 1C1 of porcine *RORγ-2* and *-3*, respectively. When compared to human, it exhibits 42% and 59% identity with human *RORγ* and *RORγt*, respectively. Porcine *RORγ-1* has 54 and 69 nucleotides longer than human *RORγ-1*

and human *RORγ-2*, respectively (Figure 5.20). Porcine *RORγ-2* has 28 and 43

nucleotides longer than human *RORγ-1* and *-2*, respectively, and exhibits only 67%

sequence identity. These low sequence identities in the 5'-UTR encoding exon are also

observed in human *RORγ*. Human *RORγ* is 50% identical to *RORγt*. These differences

in 5'-UTR may be results from alternative promoter usage and could affect post-

transcriptional regulation of mature mRNAs (mRNA stability and translation

efficiency). In contrast, the 3'-UTR of porcine *RORγ* transcripts is conserved in both

length and sequence homology. They differ in base deletion in porcine *RORγ-3*

transcript (ATCTCT deletion, one G deletion, and three nucleotide substitutions (A/G,

T/C, T/A) in porcine *RORγ-3* exon 10) and the length of poly(A) tail.
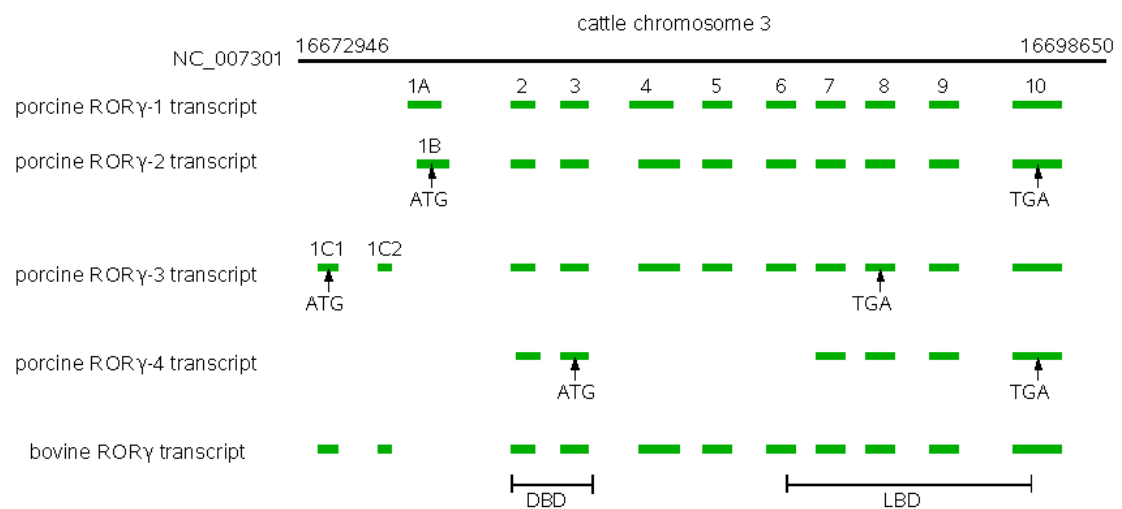
Figure 5.17. Porcine *RORγ* transcripts on bovine chromosome 3.

Porcine *RORγ* spans about 25 kb of DNA on bovine chromosome 3 (NC_007301: 16672946-16698650) and consists of 11 exons. The genomic region was identified using cow BLAST. The exon-intron boundaries were determined by aligning each porcine *RORγ* transcript with cow DNA sequence using SIM4 program. The continuous long line indicates cow DNA sequence with genomic region on each end. Exons are shown as short green lines. The positions of start and stop codons as well as DBD and LBD are indicated.

Table 5.6. Predicted exon-intron boundaries of porcine *RORγ-1* transcript.

Exon	*RORγ-1*	NC_007301	Size	Identity	5' end of exon	3' end of exon	Intron	Intron size	5' splice site	3' splice site
1A	1-217	6784-7001	217	89%	GTTGGGGGGG...	...GTCATGAGAA	1	9121	gtaagtgaat...	...cttcttccag
2	218-303	16123-16208	86	95%	CACAAATTGA...	...GGGGTGCAAG	2	284	gtgagtgctc...	...ccttcccag
3	304-445	16493-16634	142	95%	GGTTTCTTCC...	...TCCCGAGATG	3	1210	gtgaggccga...	...acaatgccta
	446-614	-	169	-	-	-	-	-	-	-
4	615-1021	18015-18421	407	89%	GTCCCCTCCT...	...ACAGAGATTG	4	332	gtaagcagct...	...ttgaccccag
5	1022-1143	18754-18875	122	92%	AGCACCTGGT...	...CCAGAGGAAG	5	844	gtgaggccag...	...tcctctgcag
6	1144-1276	19720-19852	133	96%	TCAATGTGGG...	...CTCAAAGCAG	6	163	gtgcccaggg...	...cactccccag
7	1277-1384	20016-20123	108	95%	GAGCAATGGA...	...CGAGCCTTGG	7	182	gtgaggggca...	...accctcccag
8	1385-1495	20306-20416	111	95%	GCTGCAGCGA...	...ATCAATGCCA	8	1477	gtgagtgtca...	...cctccttcag
9	1496-1605	21894-22003	110	95%	ACCGGCCAGG...	...CCTGGCCAAG	9	2951	gtaggagcag...	...ccacctccag
10	1606-2265	24955-25614	660	84%	CTGCCACCCA...	...CTGGATCTCT				

Note: The exon-intron boundaries were predicted by aligning porcine *RORγ-1* cDNA with bovine DNA sequence (NC_007313: 70604357-70611506) using SIM4 program. The poly(A) tail was excluded in the alignment. The 5'- and 3'-ends of each exon are shown in uppercase letters. The 5' and 3' splice sites of each intron are shown in lowercase letters. Splice sites are in consensus with GT-AG rule for splicing. The sequence identity between porcine *RORγ-1* transcript and bovine DNA was derived from alignment results of SIM4. The 169 bp sequence (446-614) of porcine *RORγ-1* could not align with bovine genomic DNA.

Table 5.7. Predicted exon-intron boundaries of porcine *RORγ-2* transcript.

Exon	*RORγ-2*	NC_007301	Size	Identity	5' end of exon	3' end of exon	Intron	Intron size	5' splice site	3' splice site
1B	1-191	6811-7001	191	89%	TGTGTGGTGC...	...GTCATGAGAA	1	9121	gtaagtgaat...	...cttcttccag
2	192-277	16123-16208	86	95%	CACAAATTGA...	...GGGGTGCAAG	2	284	gtgagtgctc...	...ccttccccag
3	278-419	16493-16634	142	95%	GGTTTCTTCC...	...TCCCGAGATG	3	1274	gtgaggccga...	...cctctttcag
4	420-932	17909-18421	513	90%	CTGTCAAGTT...	...ACAGAGATTG	4	332	gtaagcagct...	...ttgaccccag
5	933-1054	18754-18875	122	92%	AGCACCTGGT...	...CCAGAGGAAG	5	844	gtgaggccag...	...tcctctgcag
6	1055-1187	19720-19852	133	96%	TCAATGTGGG...	...CTCAAAGCAG	6	163	gtgcccaggg...	...cactccccag
7	1188-1295	20016-20123	108	95%	GAGCAATGGA...	...CGAGCCTTGG	7	182	gtgaggggca...	...accctcccag
8	1296-1406	20306-20416	111	95%	GCTGCAGCGA...	...ATCAATGCCA	8	1477	gtgagtgtca...	...cctccttcag
9	1407-1516	21894-22003	110	95%	ACCGGCCAGG...	...CCTGGCCAAG	9	2951	gtaggagcag...	...ccacctccag
10	1517-2176	24955-25614	660	84%	CTGCCACCCA...	...CTGGATCTCT				

Note: The exon-intron boundaries were predicted by aligning porcine *RORγ-2* cDNA with bovine DNA sequence (NC_007313) using SIM4 program. The poly(A) tail was excluded in the alignment. The 5'- and 3'-ends of each exon are shown in uppercase letters. The 5' and 3' splice sites of each intron are shown in lowercase letters. All splice sites are in consensus with GT-AG rule for splicing. The sequence identity between porcine *RORγ-2* transcript and bovine DNA was derived from alignment results of SIM4.

Table 5.8. Predicted exon-intron boundaries of porcine *RORγ-3* transcript.

Exon	*RORγ-3*	NC_007301	Size	Identity	5' end of exon	3' end of exon	Intron	Intron size	5' splice site	3' splice site
1C1	1-50	1342-1391	50	94%	GAGTTGCCCC...	...GCCTCGAGGG	1	2701	gtaagaggcc...	...ctccttacag
1C2	51-80	4093-4122	30	100%	AGCTGCTGGC...	...ACCCACACCT	2	12000	gtgagtagag...	...cttcttccag
2	81-166	16123-16208	86	95%	CACAAATTGA...	...GGGGTGCAAG	3	284	gtgagtgctc...	...ccttcccag
3	167-308	16493-16634	142	95%	GGTTTCTTCC...	...TCCCGAGATG	4	1274	gtgaggccga...	...cctctttcag
4	309-821	17909-18421	513	90%	CTGTCAAGTT...	...ACAGAGATTG	5	332	gtaagcagct...	...ttgaccccag
5	822-943	18754-18875	122	92%	AGCACCTGGT...	...CCAGAGGAAG	6	844	gtgaggccag...	...tcctctgcag
6	944-1076	19720-19852	133	95%	TCAATGTGGG...	...CTCAAAGCAG	7	163	gtgcccaggg...	...cactccccag
7	1077-1183	20016-20123	107	93%	GAACATGGAA...	...CGAACCTTGG	8	182	gtgaggggca...	...accctcccag
8	1184-1294	20306-20416	111	95%	GCTGCAGCGA...	...ATCAATGCCA	9	1477	gtgagtgtca...	...cctccttcag
9	1295-1404	21894-22003	110	95%	ACCGGCCAGG...	...CCTGGCCAAG	10	2951	gtaggagcag...	...ccacctccag
10	1405-2059	24955-25609	655	85%	CTGCCACCCA...	...ATAAATCTGG				

Note: The exon-intron boundaries were predicted by aligning porcine *RORγ-3* cDNA with bovine DNA sequence (NC_007313) using SIM4 program. The poly(A) tail was excluded in the alignment. The 5'- and 3'-ends of each exon are shown in uppercase letters. The 5' and 3' splice sites of each intron are shown in lowercase letters. All splice sites are in consensus with GT-AG rule for splicing. The sequence identity between porcine *RORγ-3* transcript and bovine DNA was derived from alignment results of SIM4. Porcine *RORγ-3* has a short exon (1C2) with 30 nucleotides that is not found in other porcine *RORγ* transcripts.

Table 5.9. Predicted exon-intron boundaries of porcine *RORγ-4* transcript.

Exon	*RORγ-4*	NC_007301	Size	Identity	5' end of exon	3' end of exon	Intron	Intron size	5' splice donor	3' splice acceptor
2	3-82	16128-16208	82	93%	GGATTGAAGT...	...GGGGTGCAAG	2	284	gtgagtgctc...	...ccttccccag
3	83-224	16493-16634	142	95%	GGTTTCTTCC...	...TCCCGAGATG	3	3381	gtgaggccga...	...cactccccag
7	225-332	20016-20123	108	95%	GAGCAATGGA...	...CGAGCCTTGG	7	182	gtgaggggca...	...accctcccag
8	333-443	20306-20416	111	95%	GCTGCAGCGA...	...ATCAATGCCA	8	1477	gtgagtgtca...	...cctccttcag
9	444-553	21894-22003	110	95%	ACCGGCCAGG...	...CCTGGCCAAG	9	2951	gtaggagcag...	...ccacctccag
10	554-1213	24955-25614	660	84.6%	CTGCCACCCA...	...CTGGATCTCT				

Note: The exon-intron boundaries were predicted by aligning porcine *RORγ-4* cDNA with bovine DNA sequence (NC_007313) using SIM4 program. The poly(A) tail was excluded in the alignment. The 5'- and 3'-ends of each exon are shown in uppercase letters. The 5' and 3' splice sites of each intron are shown in lowercase letters. All splice sites are in consensus with GT-AG rule for splicing. The sequence identity between porcine *RORγ-4* transcript and bovine DNA was derived from alignment results of SIM4. Porcine *RORγ-4* transcript represents a transcript variant in porcine *RORγ* that lacks exons 1 and 4-6.
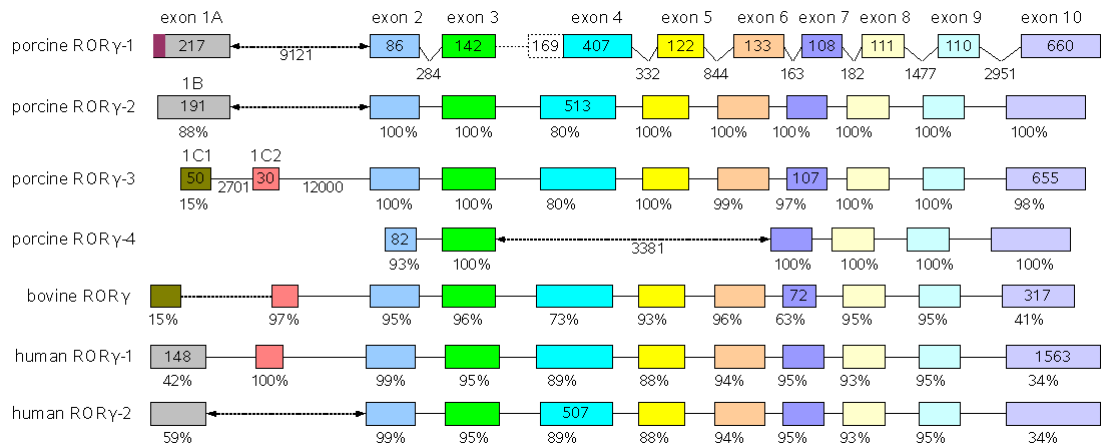
Figure 5.18. Comparison of *RORγ* exons in farm animals and human.

Boxes represent exons. Exons with similar size are colored. The percent identity derived from pairwise alignment using BioEdit is shown below each exon. The size of exon is shown in the exon box. The thin lines indicate introns. The length of intron is shown only for porcine *RORγ-1* transcript.

```
porcine RORγ-1  GTGAGGCCATGGGGGCAGCCCTCTGGGGCTTCCCTGGAGTCTCCAGAGGGGCAGCTGAGC  60
porcine RORγ-2  ------------------------------------------------.T.TCA..T     9
porcine RORγ-3  ------------------------------------------------.T.TCA..T     9
bovine  RORγ    ------------------------------------------------.T.TCA...     9
human   RORγ    ------------------------------------------------.T.TCA..T     9


porcine RORγ-1  TGCTGTGCCAGACAGGGTTAGCCTGTAAAAGTGCCTTTCTGGAGCTCTGCTCTCCATCCT  120
porcine RORγ-2  .TGGCC..AT.T.CAA.--.AG.AAAGGG.CA....GCA..CT.AGG...AGAA-.CAGC   66
porcine RORγ-3  .TGGCC..AT.T.CAA.--.AG.AAAGGG.CA....GCA..CT.AAG...AGAA-.CAGC   66
bovine  RORγ    .TGGCC..AT.T.CAA.--.AG.AAAGGG.CA....GCA..C..AGG...AGAA-.CAGC   66
human   RORγ    .CGGCC..AT.T.CAA.--.AG.A.AGGG.CA....GCA..C..AAG...AGAA-.CAGC   66


porcine RORγ-1  TCCCGCGGGCGGGCCCCTGTCCCTTCCACTCTCATCAGAGGTCTTCGACGTCCCCTCCTG  180
porcine RORγ-2  .G.A..A.A--....AA.A.---.AG.A..GGGA.CA..C..C.----.A..........  117
porcine RORγ-3  .G.A..A.A--....AA.A.---.AG.A..GGGA.CA..C..C.----.A..........  117
bovine  RORγ    .G.A..A.C--....AA.A.---.AG.A..GGGA.CA..C..C.----.A.A.......CA  117
human   RORγ    .G.A..A.C--....AA.A.---.AG.A..AGGA.C...T....----.A.A.......A.  117
```

Figure 5.19. Alignment of a 167-bp insertion in exon 4 of porcine *RORγ-1*.

Porcine *RORγ-1* transcript has a unique nucleotide sequence in exon 4. When compared to porcine *RORγ-2* transcript, porcine *RORγ-1* has a unique 63-bp insertion at the 5'-end of exon 4, resulting in 21 amino acids in the D domain. Dots represent identical nucleotide bases. Gaps are indicated by dashes.

133

```
porcine RORγ-1 GTTGGGGGGGTTGTCACTCGGCCACCTGTGTGGTGCGGAGCTTAAACCCCCCTGCCC 57
porcine RORγ-2 ------------------------.............................. 31
human RORγ-1   -----------------------------------------------------G.. 3
human RORγ-2   ----------------------------------------------------- 1

porcine RORγ-1 AGAAGCATTGGGGGAGAGCTAGGTGCAGAGCTGCAGGCTG-AGGCTCTGCTGAGAGGGCC 116
porcine RORγ-2 .......................................-................... 90
human RORγ-1   ..GT..TCCC.CCTTCCA.CCTCC..CCTC..C.CTC.CCTG...C.....CCCT.CC.T 63
human RORγ-2   -----------A....................T.......-....G............. 47

porcine RORγ-1 TCACCCCCAACCCCGCTGCCAGCTGCACCCCATCCCTGGACCACCCCCCGCTAAGAAGGA 176
porcine RORγ-2 .......................................................... 150
human RORγ-1   C.TGGG.AGC.AGG..A.....G--..GG..C.AAG...G.TG....-------.T... 113
human RORγ-2   ..G....-GC.T.T..C..............CT.............T...G....... 106

porcine RORγ-1 CGGGAGCCCA--GTCGGCAAAGCCCACGGCTCAGTCATGAGAA 217
porcine RORγ-2 ..........--............................. 191
human RORγ-1   .A..GC....-------..G..A.-.GCA.CG..C.TCAC.GG 148
human RORγ-2   .A..GAG...AG.C.....G....-.A............... 148
```

Figure 5.20. Alignment of *RORγ* exon 1 between swine and human.

In exon 1, porcine *RORγ* transcripts have longer 5'-entended region when compared to human *RORγ*. The porcine *RORγ-1* differs from that of *RORγ-2* in the first 26 nucleotide of exon 1 (GTTGGGGGGGTTGTCACTCGGCCACC). A dash represents nucleotide insertion/deletion. Conserved nucleotide is represented by a dot.

5.5. Discussion

The nuclear receptor RORγ is an important transcription factor that has been proposed to involve in the development of T lymphocytes and lymph nodes (LNs) and Peyer's patches (PPs) in mammals (Kurebayashi *et al.,* 2000; Winoto and Littman, 2002; Eberl and Littman, 2003). We have identified and sequenced four porcine *RORγ* transcripts, designated porcine *RORγ-1, RORγ-2, RORγ-3*, and *RORγ-4*. All four transcripts differ in the non-coding 5'-UTR but similar in 3'-UTR. The longest transcript 1 has an additional-26 nucleotide in 5'-UTR and an additional-63 base in the open reading frame. This 63-base insertion encodes additional 21 amino acid residues in N-terminal of the D domain (amino acid residues 79-99: GEAMGAALWGFPGVSRGAAEL) (Figure 5.9 and Figure 5.10) when compared to porcine *RORγ-2*. The transcript variant 3 has in-frame TGA stop codon (nucleotide positions 1229-1231) in exon 8, resulting in a transcript with shorter ORF and longer 3'-UTR. Transcript 4 had the nucleotide deletion in the ORF, resulting in the production of isoforms lacking the whole A/B and D domains, the most part of DBD, and the part of LBD (Figure 5.9). Because porcine *RORγ-3* possesses the premature stop codon, it may be subjected to the nonsense-mediated mRNA decay (Lewis *et al.,* 2003). It is possible that porcine RORγ-4 isoform may be unable to function.

Sequence analysis revealed that full-length porcine *RORγ-1* shares 95% identity with porcine *RORγ-2*, 57% identity with human *RORγ-1*, and 58% identity with human *RORγ-2*, and porcine *RORγ-2* transcript exhibits 59% identity with human *RORγ-1*, and 60% identity with human *RORγ-2* at nucleotide sequence level (Table 5.2). At protein sequence level, porcine RORγ-1 isoform shares 90% identity with porcine RORγ isoform 2, and ~80% identity with human RORγ. The porcine RORγ-2 isoform exhibits ~90% identity with human RORγ (Table 5.5). Comparison with

human RORγ receptors, porcine RORγ-1 and -2 exhibit higher sequence identity with human RORγ-2 isoform, which is an ortholog of mouse RORγt, than human RORγ-1. Other two isoforms, porcine RORγ-3 and -4, are unique for swine and represent protein isoforms that contain the deletion in DBD and/or LBD.

Transcriptional regulation of target genes by nuclear receptors is dependent on the DBD and LBD that have a high degree of sequence identity between closely related nuclear receptors (Mangelsdorf *et al.,* 1995; Giguere, 1999; Robinson-Rechavi *et al.,* 2003). In general, evolutionarily related nuclear receptors share at least 80%–90% identity in the DBD and at least 40%–60% identity in the LBD (Nuclear Receptors Nomenclature Committee, 1999). By using the SMART analysis tool, predicted amino acid sequences of the porcine *RORγ-1* and *-2* genes contain the evolutionarily conserved DBD and LBD (Figures 5.1 and 5.2). The DBD is important in the binding of RORγ to response element located in its target gene promoters. The first zinc finger motif (CKICGDKSSGIHYGVITCEGC) is located at amino acid residues 10-30, and the second zinc finger motif (CTRQQNCPIDRTSRNRCQHC) is found at amino acid residues 46-65 in porcine RORγ-1 and 2 isoforms. They also possess the highly conserved P-box (amino acid residues 27-34: CEGCKGFF) and the AF-2 (amino acid residues 501-506: LYKELF) located in the DBD and LBD, respectively (Figures 5.12 and 5.13). The P box is highly conserved in NRs and is responsible for DNA recognition. The ligand-dependent transcriptional activation function (AF-2) is the binding site for coactivators (Giguere, 1999; Robinson-Rechavi *et al.,* 2003). Comparison of amino acid sequences showed that the DBD of porcine RORγ receptor is highly conserved and shares 100% and 97% identity with bovine and human RORγ isoforms, respectively (Figure 5.11). Compared to human, only two amino acid substitutions in the DBD were found (Gln39Arg59 and Val42Ala62)

(Figure 5.12). These amino acid substitutions do not change in solubility and/or the net charge of the amino acid. Both porcine RORγ-1 and RORγ-2 have an identical LBD, which exhibit 100% identity with mouse RORγt (He *et al.,* 1998) and 93% identity with human RORγ (Figures 5.11 and 5.13). The mouse RORγt is an orphan nulear receptor that is generated by use of an alternative promoter usage of *RORγ* gene, resulting in a protein lacking the N-terminal 24 residues of RORγ, and primarily expressed in the immature double positive thymocytes (He *et al.,* 1998). With this high degree of sequence indentity in the DBD and LBD, we conclude that porcine *RORγ* is an orthologous gene of mouse and human *RORγ*.

The exon-intron structure of porcine *RORγ* gene was determined by aligning its cDNA sequences with bovine genomic DNA (NC_007301). It was predicted to contain 11 exons separated by 10 introns. The *RORγ* gene is located on chromosomes 3 and 1 in cattle and mouse, and human, respectively (Medvedev *et al.,* 1997). Comparison of *RORγ* exons revealed that the size and number of exons are conserved between farm animals and humans, especially in the protein-encoding exons (Figure 5.18). The exons 2 and 3, encoding DBD, are identical in size (86 bp and 142 bp, respectively) and exhibit approximately 95% identity when compared to cattle and human (Figure 5.18). The LBD-encoding exons (exons 6, 7, 8, 9, and 10) are also conserved both in size and sequence identity but slightly lower than DBD-encoding exons (approximately 90% identity).

Splicing variants, generating by alternative splicing and alternative promoter usage, have been reported in several nuclear receptors such as the *PXR*, *CAR*, and *LXRα* (Auerbach *et al.,* 2003; Lamba *et al.,* 2004; Chen *et al.,* 2005; Kurose *et al.,* 2005). The porcine *RORγ-1* has two unique DNA fragments (an additional-26 nucleotide in 5'-UTR and an additional-63 base in the ORF) which may contribute to

the different expression pattern between porcine *RORγ-1* and *-2* transcripts as found in mouse *RORγt* (He *et al.,* 1998). We applied the RT-PCR with the splice-variant specific primers to detect the presence of these 5'-transcripts in porcine tissues. The porcine *RORγ-1* variant was detected in all the tissues studied, except pancreas. In contrast, the porcine *RORγ-2* variant was specific to the thymus and spleen (Figures 5.16).

In human and mouse, *RORγ* is expressed mainly in the liver, muscle, heart, thymus, kidney and brain (Hirose *et al.,* 1994; Ortiz *et al.,* 1995; Medvedev *et al.,* 1996; He *et al.,* 1998). In mice, the expression of *RORγt* is specific to the thymus and has been postulated in the control of T lymphocytes development (He *et al.,* 1998), whereas the specific functions of RORγ are not known. The mRNA expression analysis of the porcine *RORγ-1* transcript revealed that they are expressed in the liver, kidney, small intestine, heart, muscle, thymus, spleen, and brain, whereas the porcine *RORγ-2* transcript was detected only in the thymus and spleen. The mRNA expression of porcine *RORγ* transcripts is similar to mouse and human with regard to types of tissues. However, there are some differences in expression pattern of *RORγ* in different species. We could not detect the mRNAs expression of porcine *RORγ* in pancreas but found the expression in spleen. Whereas, the expression of *RORγ* was found in pancreas in human and cattle, while no expression in spleen in human, cattle, and mouse according to the UniGene database (http://www.ncbi.nlm.nih.gov/sites/entrez?db=unigene) (UniGene ID: 1747073, 170986, and 257579 for cattle, human, and mouse, respectively). These differences in tissue expression pattern may indicate the species-specific expression of *RORγ*. Further researches are required to understand the differences in tissue expression between human and swine. With regard to functions of RORγ, recently, Kang *et al.*

(2007) reported the potential functions of RORα and RORγ receptors in the regulation of several metabolic pathways, including the metabolism of steroids, bile acids, and xenobiotics. They proposed that RORs are important in the control of metabolic homeostasis. In addition, Barendse *et al.* (2007) reported the association between *RORγ* genotypes and marbling in Australian feedlot cattle.

Based on sequence and mRNA expression analyses, we may imply specific functions of *RORγ* in porcine tissues as found in humans and mice. Future studies of the *RORγ* expression in porcine immune cells may provide new information about this important gene with regard to the development and function of immunity in pigs.

CHAPTER 6

CONCLUSION AND DISCUSSION

Because studies of nuclear receptors in farm animals, especially in swine, are very few, our objective was to identify and sequence cDNAs encoding porcine nuclear receptors *CAR*, *PXR*, *LXRα*, and *RORγ*. We searched the PEDE database (Pig EST Data Explorer) in order to identify porcine nuclear receptors *CAR*, *PXR*, *LXRα*, *RORγ* genes using human cDNA sequences as queries. In present study, we report cDNA sequence data of nuclear receptors *CAR*, *PXR*, *LXRα*, and *RORγ* in swine.

The constitutive androstane receptor (*CAR*; NR1I3) and the pregnane X receptor (*PXR*; NR1I2) are members of nuclear receptor subfamily 1 group I. Studies in human and mouse revealed that both CAR and PXR regulate the expression of several important genes that encode enzymes responsible for the metabolism of endogenous hormones and exogenous substances (Wei *et al.,* 2000; Kliewer *et al.,* 2002; Moore *et al.,* 2003; Timsit and Negishi, 2007). In addition, CAR and PXR receptors can be activated by endocrine disrupting chemicals (e.g., methoxychlor, endosulfan, dieldrin, DDT, phthalic acid, and the plasticizer nonylphenol) (Masuyama *et al.,* 2000; Mikamo *et al.,* 2003; Wyde *et al.,* 2003; Kretschmer and Baldwin, 2005; Lemaire *et al.,* 2006). These chemicals have been known to have adverse affect on health and reproduction in animals. Our sequence data showed that the DBD and LBD of porcine PXR are 95% and 93% identical to human PXR DBD and LBD, respectively. Porcine CAR is 83% and 87% identical to human CAR in the DBD and LBD, respectively (Figures

3.9 and 3.14). Moreover, the mRNA expression of both porcine *CAR* and *PXR* was detected in liver, small intestine, and kidney (Figure 3.18), which are the same tissues as found in human and mouse (Baes *et al.,* 1994; Choi *et al.,* 1997; Bertilsson *et al.,* 1998; Blumberg *et al.,* 1998; Kliewer *et al.,* 1998; Lehmann *et al.,* 1998). It is well known that these organs, especially liver, are the major expression sites of steroid and xenobiotic-metabolizing enzymes such as CYP3A4 and CYP2B, which are the target genes of both CAR and PXR receptors (Honkakoski and Negishi, 2000). Based on sequence and mRNA expression analyses, we conclude that porcine *CAR* and *PXR* are orthologous genes of human *CAR* and *PXR* genes, respectively, and are likely to play important physiological roles in the metabolism of both hormones and drugs similar to human CAR and PXR receptors.

The liver X receptor alpha (*LXRα*: NR1H3) is a member of nuclear receptor superfamily that acts as a transcriptional regulator of several important genes known to responsible for the metabolism of cholesterol, fatty acids, and glucose (Peet *et al.,* 1998b; Schultz *et al.,* 2000; Lu *et al.,* 2001; Dalen *et al.,* 2003; Juvet *et al.,* 2003; Laffitte *et al.,* 2003; Seo *et al.,* 2004). In addition, several reports show that LXRα is involved in the immune system (Joseph *et al.,* 2004; Valledor *et al.,* 2004). In the present study, we identified and sequenced three porcine *LXRα* transcripts (Figures 4.1, 4.2, and 4.3). The porcine *LXRα-1* and *-2* are identical in ORFs and 3'-UTRs, but differ in 5'-UTRs (1-145 in *LXRα-1* and 1-178 in *LXRα-2*) (Figures 4.4 and 4.5). The partial porcine *LXRα-3* transcript represents an incomplete transcript that has an in-frame TGA stop codon, resulting in a truncated protein lacking amino acid residues downstream from the DBD. Sequence analysis revealed that porcine LXRα is 97% and 100% identical to human LXRα in the DBD and LBD, respectively (Figure 4.10). The mRNA expression of porcine *LXRα-2* transcript was detected only in the thymus

and spleen, whereas the porcine *LXRα-1* transcript was detected in the liver, kidney, small intestine, heart, muscle, thymus, spleen, and brain (Figure 4.15 and 4.16). The tissue-specific expression of transcript 2 in the thymus and spleen might imply the important of this transcript in porcine immunity. Further studies are necessary in order to better understand the functional roles of porcine *LXRα-2* transcript in porcine immune system.

The nuclear receptor RORγ is an important transcription factor that has been proposed to involve in the development of T lymphocytes, lymph nodes (LNs), and Peyer's patches (PPs) in mammals (Kurebayashi *et al.,* 2000; Winoto and Littman, 2002; Eberl and Littman, 2003). We have identified and sequenced four porcine *RORγ* transcripts, designated porcine *RORγ-1*, *RORγ-2*, *RORγ-3*, and *RORγ-4* (Figures 5.1, 5.2, 5.3, and 5.4). All four transcripts are different in the non-coding 5'-UTR, but highly similar in 3'-UTR. The porcine *RORγ-1* transcript has an additional-26 nucleotide in 5'-UTR and an additional-63 base in the open reading frame. This 63-base insertion encodes additional 21 amino acid residues in N-terminal of the D domain (amino acid residues 79-99: GEAMGAALWGFPGVSRGAAEL) when compared to porcine *RORγ-2* (Figures 5.9 and 5.10). The transcript variants 3 and 4 encode porcine RORγ isoforms lacking functional DBD and/or LBD. Comparison with human RORγ receptors, porcine RORγ-1 and -2 exhibit higher sequence identity with human RORγ-2 isoform than human RORγ-1. Amino acid sequence comparison showed that the DBD and LBD of porcine RORγ are 97% and 93% identical to human RORγ DBD and LBD, respectively (Figure 5.11). The mRNA expression of porcine *RORγ-1* variant was detected in all the tissues studied, except pancreas. In contrast, the expression of porcine *RORγ-2* variant was specific to the thymus and spleen (Figure 5.16). The mRNA expression of porcine *RORγ* transcripts is similar to mice and

humans with regard to types of tissues. Based on sequence and mRNA expression analyses, we conclude that porcine *RORγ* is an orthologous gene of human *RORγ* and may possess similar function as found in human and mouse. Future studies of the *RORγ* expression in porcine immune cells, especially T lymphocytes, may provide new information about this important gene with regard to the development and function of porcine immunity.

In summary, we have identified and sequenced cDNAs encoding nuclear receptor *CAR*, *PXR*, *LXRα*, and *RORγ* in swine. Our cDNA sequence data may provide useful information for further functional analysis and the application of nuclear receptors to farm animal production.

APPENDIX A: AMINO ACID CODES

One-letter code	Three-letter code	Amino acid
A	Ala	Alanine
C	Cys	Cysteine
D	Asp	Aspartic acid
E	Glu	Glutamic acid
F	Phe	Phenylalanine
G	Gly	Glycine
H	His	Histidine
I	Ile	Isoleucine
K	Lys	Lysine
L	Leu	Leucine
M	Met	Methionine
N	Asn	Asparagine
P	Pro	Proline
Q	Gln	Glutamine
R	Arg	Arginine
S	Ser	Serine
T	Thr	Threonine
V	Val	Valine
W	Trp	Tryptophan
Y	Tyr	Tyrosine

APPENDIX B: ABBREVIATIONS

AF-1: ligand-independent activation function-1
AF-2: ligand-dependent activation-2
ARE: AU-rich element
Bcl-xL: B-cell lymphoma
bp: base pair
CAR: constitutive androstane receptor
cDNA: complementary deoxyribonucleic acid
CYP2B: cytochrome P450 B
CYP3A4: cytochrome P450 3A4
CYP7A1: cytochrome P450 7A1
DBD: DNA-binding domain
DDE: 1,1-dichloro-2,2-bis(p-chlorophenyl)ethylene
DDT: 1,1,1-trichloro-2,2-bis(p-chlorophenyl)ethane
DNA: deoxyribonucleic acid
EDC: endocrine disrupting chemical
IRES: internal ribosome entry site
LB: Luria-Bertani
LBD: ligand-binding domain
LXRE: LXR response element
LXRα: liver x receptor alpha
mRNA: messenger ribonucleic acid
NCoA-2: nuclear receptor coactivator-2
NR: nuclear receptor
NRRE: nuclear receptor response element
ORF: open reading frame
PB: phenobarbital
PCR: polymerase chain reaction
PEDE: Pig EST data explorer
PPARγ: peroxisome proliferator-activated receptor gamma
PXR: pregnane x receptor
RNA: ribonucleic acid
RORE: RORγ-response element
RORγ: retinoic acid receptor-related orphan receptor gamma
RT-PCR: reverse transcription-PCR
RXR: retinoid X receptor
SCR-1: steroid receptor coactivator-1
SREBP-1c: sterol regulatory element-binding protein 1c
TBE: Tris-Borate-EDTA
TCPOBOP: 1,4-bis[2-(3,5-dichloropyridyloxy)]benzene
TCR-Jα: T-cell receptor alpha-chain
UGT1A1: UDP-glucuronosyltransferase 1A1
uORF: upstream open reading frame
UTR: untranslated region

REFERENCES

Auerbach SS, Ramsden R, Stoner MA, Verlinde C, Hassett C, Omiecinski CJ. 2003. Alternatively spliced isoforms of the human constitutive androstane receptor. Nucleic Acids Research, 31: 3194-3207.

Baes M, Gulick T, Choi HS, Martinoli MG, Simha D, Moore DD. 1994. A new orphan member of the nuclear hormone receptor superfamily that interacts with a subset of retinoic acid response elements. Molecular and Cellular Biology, 14: 1544-1551.

Barendse W, Bunch RJ, Kijas JW, Thomas MB. 2007. The effect of genetic variation of the retinoic acid receptor-related orphan receptor C (RORC) gene on fatness in cattle. Genetics, 175: 843-853.

Bertilsson G, Heidrich J, Svensson K, Asman M, Jendeberg L, Sydow-Bäckman M, Ohlsson R, Postlind H, Blomquist P, Berkenstam A. 1998. Identification of a human nuclear receptor defines a new signaling pathway for CYP3A induction. Proceedings of the National Academy of Sciences of the United States of America, 95: 12208-12213.

Blumberg B, Sabbagh W Jr, Juguilon H, Bolado J Jr, van Meter CM, Ong ES, Evans RM. 1998. SXR, a novel steroid and xenobiotic-sensing nuclear receptor. Genes and Development, 12: 3195-3205.

Chen M, Beaven S, Tontonoz P. 2005. Identification and characterization of two alternatively spliced transcript variants of human liver X receptor alpha. Journal of Lipid Research, 46: 2570-2579.

Choi HS, Chung M, Tzameli I, Simha D, Lee YK, Seol W, Moore DD. 1997. Differential transactivation by two isoforms of the orphan nuclear hormone receptor CAR. Journal of Biological Chemistry, 272: 23565-23571.

Crooks GE, Hon G, Chandonia JM, Brenner SE. 2004. WebLogo: A sequence logo generator. Genome Research, 14: 1188-1190.

Dalen KT, Ulven SM, Bamberg K, Gustafsson JA, Nebb HI. 2003. Expression of the insulin-responsive glucose transporter GLUT4 in adipocytes is dependent on liver X receptor alpha. Journal of Biological Chemistry, 278: 48283-48291.

DeBose-Boyd RA, Ou J, Goldstein JL, Brown MS. 2001. Expression of sterol regulatory element-binding protein 1c (SREBP-1c) mRNA in rat hepatoma cells requires endogenous LXR ligands. Proceedings of the National Academy of Sciences of the United States of America, 98: 1477-1482.

Ding X, Lichti K, Staudinger JL. 2006. The mycoestrogen zearalenone induces CYP3A through activation of the pregnane X receptor. Toxicological Sciences, 91: 448-455.

Eberl G, Littman DR. 2003. The role of the nuclear hormone receptor RORγt in the development of lymph nodes and Peyer's patches. Immunological Reviews, 195: 81-90. Review.

Florea L, Hartzell G, Zhang Z, Rubin GM, Miller W. 1998. A computer program for aligning a cDNA sequence with a genomic DNA sequence. Genome Research, 8: 967-974.

Forman BM, Tzameli I, Choi HS, Chen J, Simha D, Seol W, Evans RM, Moore DD. 1998. Androstane metabolites bind to and deactivate the nuclear receptor CAR-beta. Nature, 395: 612-615.

Giguere V. 1999. Orphan nuclear receptors: from gene to function. Endocrine Reviews, 20: 689-725. Review.

Handschin C, Podvinec M, Meyer UA. 2000. CXR, a chicken xenobiotic-sensing orphan nuclear receptor, is related to both mammalian pregnane X receptor (PXR) and constitutive androstane receptor (CAR). Proceedings of National Academy of Sciences of the United States of America, 97: 10769-10774.

He YW, Deftos ML, Ojala EW, Bevan MJ. 1998. RORγt, a novel isoform of an orphan receptor, negatively regulates Fas ligand expression and IL-2 production in T cells. Immunity, 9: 797-806.

Hernandez JP, Huang W, Chapman LM, Chua S, Moore DD, Baldwin WS. 2007. The environmental estrogen, nonylphenol, activates the constitutive androstane receptor. Toxicological Sciences, 98: 416-26.

Hirose T, Smith RJ, Jetten AM. 1994. Rorγ: the third member of ror/rzr orphan receptor subfamily that is highly expressed in skeletal muscle. Biochemical and Biophysical Research Communications, 205: 1976-1983.

Hofacker IL. 2003. Vienna RNA secondary structure server. Nucleic Acids Research, 31: 3429-3431.

Honkakoski P, Negishi M. 2000. Regulation of cytochrome P450 (CYP) genes by nuclear receptors. Biochemical Journal, 347: 321-337.

Hughes TA. 2006. Regulation of gene expression by alternative untranslated regions. Trends in Genetics, 22: 119-122. Review.

Jacobs MN, Nolan GT, Hood SR. 2005. Lignans, bacteriocides and organochlorine compounds activate the human pregnane X receptor (PXR). Toxicology Appllied Pharmacology, 209: 123-133.

Janowski BA, Grogan MJ, Jones SA, Wisely GB, Kliewer SA, Corey EJ, Mangelsdorf DJ. 1999. Structural requirements of ligands for the oxysterol liver X receptors LXRα and LXRβ. Proceedings of the National Academy of Sciences of the United States of America, 96: 266-271.

Jaye MC, Krawiec JA, Campobasso N, Smallwood A, Qiu C, Lu Q, Kerrigan JJ, De Los Frailes Alvaro M, Laffitte B, Liu WS, Marino JP Jr, Meyer CR, Nichols JA, Parks DJ, Perez P, Sarov-Blat L, Seepersaud SD, Steplewski KM, Thompson SK, Wang P, Watson MA, Webb CL, Haigh D, Caravella JA, Macphee CH, Willson TM, Collins JL. 2005. Discovery of substituted maleimides as liver X receptor agonists and determination of a ligand-bound crystal structure. Journal of Medicinal Chemistry, 48: 5419-5422.

Joseph SB, Bradley MN, Castrillo A, Bruhn KW, Mak PA, Pei L, Hogenesch J, O'connell RM, Cheng G, Saez E, Miller JF, Tontonoz P. 2004. LXR-dependent gene expression is important for macrophage survival and the innate immune response. Cell, 119: 299-309.

Juvet LK, Andresen SM, Schuster GU, Dalen KT, Tobin KA, Hollung K, Haugen F, Jacinto S, Ulven SM, Bamberg K, Gustafsson JA, Nebb HI. 2003. On the role of liver X receptors in lipid accumulation in adipocytes. Molecular Endocrinology, 17: 172-182.

Kallen JA, Schlaeppi JM, Bitsch F, Geisse S, Geiser M, Delhon I, Fournier B. 2002. X-ray structure of the hRORalpha LBD at 1.63 Å: structural and functional data that cholesterol or a cholesterol derivative is the natural ligand of RORalpha. Structure, 10: 1697-1707.

Kang HS, Angers M, Beak JY, Wu X, Gimble JM, Wada T, Xie W, Collins JB, Grissom SF, Jetten AM. 2007. Gene expression profiling reveals a regulatory role for RORα and RORγ in phase I and phase II metabolism. Physiological Genomics, 31: 281-294.

Kawamoto T, Sueyoshi T, Zelko I, Moore R, Washburn K, Negishi M. 1999. Phenobarbital-responsive nuclear translocation of the receptor CAR in induction of the CYP2B gene. Molecular and Cellular Biology, 19: 6318-6322.

Khorasanizadeh S, Rastinejad F. 2001. Nuclear-receptor interactions on DNA-response elements. Trends in Biochemical Sciences, 26: 384-390. Review.

Kliewer SA, Goodwin B, Willson TM. 2002. The nuclear pregnane X receptor: a key regulator of xenobiotic metabolism. Endocrine Reviews, 23: 687-702.

Kliewer SA, Moore JT, Wade L, Staudinger JL, Watson MA, Jones SA, McKee DD, Oliver BB, Willson TM, Zetterstrom RH, Perlmann T, Lehmann JM. 1998. An orphan nuclear receptor activated by pregnanes defines a novel steroid signaling pathway. Cell, 92: 73-82.

Kozak M. 1996. Interpreting cDNA sequences: some insights from studies on translation. Mammalian Genome, 7: 563-574. Review.

Kretschmer XC, Baldwin WS. 2005. CAR and PXR: xenosensors of endocrine disrupters? Chemico-Biological Interactions, 155: 111-128. Review.

Kurebayashi S, Ueda E, Sakaue M, Patel DD, Medvedev A, Zhang F, Jetten AM. 2000. Retinoid-related orphan receptor γ (RORγ) is essential for lymphoid organogenesis and controls apoptosis during thymopoiesis. Proceedings of the National Academy of Sciences of the United States of America, 97: 10132-10137.

Kurose K, Koyano S, Ikeda S, Tohkin M, Hasegawa R, Sawada J. 2005. 5' diversity of human hepatic PXR (NR1I2) transcripts and identification of the major transcription initiation site. Molecular and Cellular Biochemistry, 273: 79-85.

Laffitte BA, Chao LC, Li J, Walczak R, Hummasti S, Joseph SB, Castrillo A, Wilpitz DC, Mangelsdorf DJ, Collins JL, Saez E, Tontonoz P. 2003. Activation of liver X receptor improves glucose tolerance through coordinate regulation of glucose metabolism in liver and adipose tissue. Proceedings of the National Academy of Sciences of the United States of America, 100: 5419-5424.

Lamba V, Yasuda K, Lamba JK, Assem M, Davila J, Strom S, Schuetz EG. 2004. PXR (NR1I2): splice variants in human tissues, including brain, and identification of neurosteroids and nicotine as PXR activators. Toxicology and Applied Pharmacology, 199: 251-265.

Lehmann JM, Kliewer SA, Moore LB, Smith-Oliver TA, Oliver BB, Su JL, Sundseth SS, Winegar DA, Blanchard DE, Spencer TA, Willson TM. 1997. Activation of the nuclear receptor LXR by oxysterols defines a new hormone response pathway. Journal of Biological Chemistry, 272: 3137-3140.

Lehmann JM, McKee DD, Watson MA, Willson TM, Moore JT, Kliewer SA. 1998. The human orphan nuclear receptor PXR is activated by compounds that regulate CYP3A4 gene expression and cause drug interactions. Journal of Clinical Investigation, 102: 1016-1023.

Lemaire G, Mnif W, Pascussi JM, Pillon A, Rabenoelina F, Fenet H, Gomez E, Casellas C, Nicolas JC, Cavailles V, Duchesne MJ, Balaguer P. 2006. Identification of new human pregnane X receptor ligands among pesticides using a stable reporter cell system. Toxicological Sciences, 91: 501-509.

Letunic I, Copley RR, Schmidt S, Ciccarelli FD, Doerks T, Schultz J, Ponting CP, Bork P. 2004. SMART 4.0: towards genomic data integration. Nucleic Acids Research, 32 (Database issue): D142-144.

Lewis BP, Green RE, Brenner SE. 2003. Evidence for the widespread coupling of alternative splicing and nonsense-mediated mRNA decay in humans. Proceedings of the National Academy of Sciences of the United States of America, 100: 189-192.

Lu TT, Repa JJ, Mangelsdorf DJ. 2001. Orphan nuclear receptors as eLiXiRs and FiXeRs of sterol metabolism. Journal of Biological Chemistry, 276: 37735-37738. Review.

Maglich JM, Sluder A, Guan X, Shi Y, McKee DD, Carrick K, Kamdar K, Willson TM, Moore JT. 2001. Comparison of complete nuclear receptor sets from the human, Caenorhabditis elegans and Drosophila genomes. Genome Biology, 2: RESEARCH0029.

Maglich JM, Stoltz CM, Goodwin B, Hawkins-Brown D, Moore JT, Kliewer SA. 2002. Nuclear pregnane x receptor and constitutive androstane receptor regulate overlapping but distinct sets of genes involved in xenobiotic detoxification. Molecular Pharmacology, 62: 638-646.

Mangelsdorf DJ, Evans RM. 1995. The RXR heterodimers and orphan receptors. Cell, 83: 841-850. Review.

Mangelsdorf DJ, Thummel C, Beato M, Herrlich P, Schutz G, Umesono K, Blumberg B, Kastner P, Mark M, Chambon P, Evans RM. 1995. The nuclear receptor superfamily: the second decade. Cell, 83: 835-839. Review.

Masuyama H, Hiramatsu Y, Kunitomi M, Kudo T, MacDonald PN. 2000. Endocrine disrupting chemicals, phthalic acid and nonylphenol, activate Pregnane X receptor-mediated transcription. Molecular Endocrinology, 14: 421-428.

Medvedev A, Chistokhina A, Hirose T, Jetten AM. 1997. Genomic structure and chromosomal mapping of the nuclear orphan receptor ROR gamma (RORC) gene. Genomics, 46: 93-102.

Medvedev A, Yan ZH, Hirose T, Giguere V, Jetten AM. 1996. Cloning of a cDNA encoding the murine orphan receptor RZR/RORγ and characterization of its response element. Gene, 181: 199-206.

Mignone F, Grillo G, Licciulli F, Iacono M, Liuni S, Kersey PJ, Duarte J, Saccone C, Pesole G. 2005. UTRdb and UTRsite: a collection of sequences and regulatory motifs of the untranslated regions of eukaryotic mRNAs. Nucleic Acids Research, 33 (Database issue): D141-146.

Mikamo E, Harada S, Nishikawa J, Nishihara T. 2003. Endocrine disruptors induce cytochrome P450 by affecting transcriptional regulation via pregnane X receptor. Toxicology and Applied Pharmacology, 193: 66-72.

Mnif W, Pascussi JM, Pillon A, Escande A, Bartegi A, Nicolas JC, Cavailles V, Duchesne MJ, Balaguer P. 2007. Estrogens and antiestrogens activate hPXR. Toxicology Letters, 170: 19-29.

Moore JT, Moore LB, Maglich JM, Kliewer SA. 2003. Functional and structural comparison of PXR and CAR. Biochimica et Biophysica Acta, 1619: 235-238.

Moore LB, Maglich JM, McKee DD, Wisely B, Willson TM, Kliewer SA, Lambert MH, Moore JT. 2002. Pregnane X receptor (PXR), constitutive androstane receptor (CAR), and benzoate X receptor (BXR) define three pharmacologically distinct classes of nuclear receptors. Molecular Endocrinology, 16: 977-986.

Moore LB, Parks DJ, Jones SA, Bledsoe RK, Consler TG, Stimmel JB, Goodwin B, Liddle C, Blanchard SG, Willson TM, Collins JL, Kliewer SA. 2000. Orphan nuclear receptors constitutive androstane receptor and pregnane X receptor share xenobiotic and steroid ligands. Journal of Biological Chemistry, 275: 15122-15127.

Mount SM. 1982. A catalogue of splice junction sequences. Nucleic Acids Research, 10: 459-472.

Nuclear Receptors Nomenclature Committee. 1999. A unified nomenclature system for the nuclear receptor superfamily. Cell, 97: 161-163.

Orans J, Teotico DG, Redinbo MR. 2005. The nuclear xenobiotic receptor pregnane X receptor: recent insights and new challenges. Molecular Endocrinology, 19: 2891-2900. Review.

Ortiz MA, Piedrafita FJ, Pfahl M, Maki R. 1995. Tor: a new orphan receptor expressed in the thymus that can modulate retinoid and thyroid hormone signals. Molecular Endocrinology, 9; 1679-1691.

Peet DJ, Janowski BA, Mangelsdorf DJ. 1998a. The LXRs: a new class of oxysterol receptors. Current Opinion in Genetics and Development, 8: 571-575. Review.

Peet DJ, Turley SD, Ma W, Janowski BA, Lobaccaro JM, Hammer RE, Mangelsdorf DJ. 1998b. Cholesterol and bile acid metabolism are impaired in mice lacking the nuclear oxysterol receptor LXRα. Cell, 93: 693-704.

Pesole G, Mignone F, Gissi C, Grillo G, Licciulli F, Liuni S. 2001. Structural and functional features of eukaryotic mRNA untranslated regions. Gene, 276: 73-81.

Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC, Ferrin TE. 2004. UCSF Chimera - A visualization system for exploratory research and analysis. Journal of Computational Chemistry, 25: 1605-1612.

Pollock CB, Rogatcheva MB, Schook LB. 2007. Comparative genomics of xenobiotic metabolism: a porcine-human PXR gene comparison. Mammalian Genome, 18: 210-219.

Repa JJ, Liang G, Ou J, Bashmakov Y, Lobaccaro JM, Shimomura I, Shan B, Brown MS, Goldstein JL, Mangelsdorf DJ. 2000. Regulation of mouse sterol regulatory element-binding protein-1c gene (SREBP-1c) by oxysterol receptors, LXRα and LXRβ. Genes and Development, 14: 2819-2830.

Robinson-Rechavi M, Escriva Garcia H, Laudet V. 2003. The nuclear receptor superfamily. Journal of Cell Science, 116: 585-586. Review.

Rozen S, Skaletsky HJ. 2000. Primer3 on the WWW for general users and for biologist programmers. In: Krawetz S, Misener S (eds) Bioinformatics

Methods and Protocols: Methods in Molecular Biology, 365-386, Humana Press, Totowa, NJ.

Schultz J, Milpetz F, Bork P, Ponting CP. 1998. SMART, a simple modular architecture research tool: identification of signaling domains. Proceedings of the National Academy of Sciences of the United States of America, 95: 5857-5864.

Schultz JR, Tu H, Luk A, Repa JJ, Medina JC, Li L, Schwendner S, Wang S, Thoolen M, Mangelsdorf DJ, Lustig KD, Shan B. 2000. Role of LXRs in control of lipogenesis. Genes and Development, 14: 2831-2838.

Schwede T, Kopp J, Guex N, Peitsch MC. 2003. SWISS-MODEL: an automated protein homology-modeling server. Nucleic Acids Research, 31: 3381-3385.

Seo JB, Moon HM, Kim WS, Lee YS, Jeong HW, Yoo EJ, Ham J, Kang H, Park MG, Steffensen KR, Stulnig TM, Gustafsson JA, Park SD, Kim JB. 2004. Activated liver X receptors stimulate adipocyte differentiation through induction of peroxisome proliferator-activated receptorγ expression. Molecular and Cellular Biology, 24: 3430-3444.

Sluder AE, Maina CV. 2001. Nuclear receptors in nematodes: themes and variations. Trends in Genetics, 17: 206-213. Review.

Stothard P. 2000. The Sequence Manipulation Suite: JavaScript programs for analyzing and formatting protein and DNA sequences. Biotechniques, 28: 1102-1104.

Suino K, Peng L, Reynolds R, Li Y, Cha JY, Repa JJ, Kliewer SA, Xu HE. 2004. The nuclear xenobiotic receptor CAR: structural determinants of constitutive activation and heterodimerization. Molecular Cell, 16: 893-905.

Thadtha P, Uenishi H, Urata H, Ishikawa M, Wada Y. 2006. Transcript variant of the porcine liver X receptor alpha (LXRα) expressed in the thymus and spleen. Journal of Animal Genetics, 34 (2): 3-9.

Thadtha P, Yamada Y, Uenishi H, Wada Y. 2005. Molecular cloning of the gene encoding pregnane X receptor (PXR; NR1I2) and the constitutive androstane receptor (CAR; NR1I3) in pigs. Journal of Animal Genetics, 33 (1 & 2): 3-10.

Timsit YE, Negishi M. 2007. CAR and PXR: the xenobiotic-sensing receptors. Steroids, 72: 231-246. Review.

Uenishi H, Eguchi T, Suzuki K, Sawazaki T, Toki D, Shinkai H, Okumura N, Hamasima N, Awata T. 2004. PEDE (Pig EST Data Explorer): construction of a database for ESTs derived from porcine full-length cDNA libraries. Nucleic Acids Research, 32 (Database issue): D484-488.

Valledor AF, Hsu LC, Ogawa S, Sawka-Verhelle D, Karin M, Glass CK. 2004. Activation of liver X receptors and retinoid X receptors prevents bacterial-

induced macrophage apoptosis. Proceedings of the National Academy of Sciences of the United States of America, 101: 17813-17818.

Venkateswaran A, Laffitte BA, Joseph SB, Mak PA, Wilpitz DC, Edwards PA, Tontonoz P. 2000. Control of cellular cholesterol efflux by the nuclear oxysterol receptor LXRα. Proceedings of the National Academy of Sciences of the United States of America, 97: 12097-12102.

Villey I, de Chasseval R, de Villartay JP. 1999. RORgammaT, a thymus-specific isoform of the orphan nuclear receptor RORgamma /TOR, is up-regulated by signaling through the pre-T cell receptor and binds to the TEA promoter. European Journal of Immunology, 29: 4072-4080.

Watkins RE, Davis-Searles PR, Lambert MH, Redinbo MR. 2003. Coactivator binding promotes the specific interaction between ligand and the pregnane X receptor. Journal of Molecular Biology, 331: 815-828.

Wei P, Zhang J, Egan-Hafley M, Liang S, Moore DD. 2000. The nuclear receptor CAR mediates specific xenobiotic induction of drug metabolism. Nature, 407: 920-923.

Willy PJ, Umesono K, Ong ES, Evans RM, Heyman RA, Mangelsdorf DJ. 1995. LXR, a nuclear receptor that defines a distinct retinoid response pathway. Genes and Development, 9: 1033-1045.

Winoto A, Littman DR. 2002. Nuclear hormone receptors in T lymphocytes. Cell, 109, Supplement 1: S57-S66. Review.

Wyde ME, Bartolucci E, Ueda A, Zhang H, Yan B, Negishi M, You L. 2003. The environmental pollutant 1,1-dichloro-2,2-bis (p-chlorophenyl)ethylene induces rat hepatic cytochrome P450 2B and 3A expression through the constitutive androstane receptor and pregnane X receptor. Molecular Pharmacology, 64: 474-481.

Xie W, Barwick JL, Simon CM, Pierce AM, Safe S, Blumberg B, Guzelian PS, Evans RM. 2000. Reciprocal activation of xenobiotic response genes by nuclear receptors SXR/PXR and CAR. Genes and Development, 14: 3014-3023.

Xie W, Yeuh MF, Radominska-Pandya A, Saini SP, Negishi Y, Bottroff BS, Cabrera GY, Tukey RH, Evans RM. 2003. Control of steroid, heme, and carcinogen metabolism by nuclear pregnane X receptor and constitutive androstane receptor. Proceedings of National Academy of Sciences of the United States of America, 100: 4150-4155.

Xu RX, Lambert MH, Wisely BB, Warren EN, Weinert EE, Waitt GM, Williams JD, Collins JL, Moore LB, Willson TM, Moore JT. 2004. A structural basis for constitutive activity in the human CAR/RXRalpha heterodimer. Molecular Cell, 16: 919-928.

Xue Y, Moore LB, Orans J, Peng L, Bencharit S, Kliewer SA, Redinbo MR. 2007. Crystal structure of the pregnane X receptor-estradiol complex provides insights into endobiotic recognition. Molecular Endocrinology, 21: 1488-1489.

Yamada Y, Piao J, Thadtha P, Wada Y. 2006. cDNA cloning of constitutive androstane receptor (CAR) gene of Japanese quail. Bullentin of the Faculty of Agriculture, Saga University, 92: 1-8.

Yoshikawa T, Shimano H, Amemiya-Kudo M, Yahagi N, Hasty AH, Matsuzaka T, Okazaki H, Tamura Y, Iizuka Y, Ohashi K, Osuga JI, Harada K, Gotoda T, Kimura S, Ishibashi S, Yamada N. 2001. Identification of liver X receptor-retinoid X receptor as an activator of the sterol regulatory element-binding protein 1c gene promoter. Molecular and Cellular Biology, 21: 2991-3000.

Yu M, Geiger B, Deeb N, Rothschild MF. 2006. Liver X receptor alpha and beta genes have the potential role on loin lean and fat content in pigs. Journal of Animal Breeding and Genetics, 123: 81-88.

Zhang J, Kuehl P, Green ED, Touchman JW, Watkins PB, Daly A, Hall SD, Maurel P, Relling M, Brimer C, Yasuda K, Wrighton SA, Hancock M, Kim RB, Strom S, Thummel K, Russell CG, Hudson JR Jr, Schuetz EG, Boguski MS. 2001. The human pregnane X receptor: genomic structure and identification and functional characterization of natural allelic variants. Pharmacogenetics, 11: 555-572.

Zhang Z, Burch PE, Cooney AJ, Lanz RB, Pereira FA, Wu J, Gibbs RA, Weinstock G, Wheeler DA. 2004. Genomic analysis of the nuclear receptor family: new insights into structure, regulation, and evolution from the rat genome. Genome Research, 14: 580-590.