2127

0041-1337/00/7009-1275/0
TRANSPLANTATION
Copyright © 2000 by Lippincott Williams & Wilkins, Inc.

Vol. 70, 1275–1283, No. 9, November 15, 2000
Printed in U.S.A.

Transplantation®

RAPID COMMUNICATIONS

ISOLATION OF THE REGULATORY REGIONS AND GENOMIC ORGANIZATION OF THE PORCINE α1,3-GALACTOSYLTRANSFERASE GENE[1]

CHIHIRO KOIKE,[2,3,5] ROBERT P. FRIDAY,[3] IZUMI NAKASHIMA,[4] PATRIZIA LUPPI,[3] JOHN J. FUNG,[2] ABDUL S. RAO,[2] THOMAS E. STARZL,[2] AND MASSIMO TRUCCO[3]

Thomas E. Starzl Transplantation Institute and Department of Surgery, University of Pittsburgh, Pittsburgh, PA 15213; Division of Immunogenetics, Department of Pediatrics, Children's Hospital of Pittsburgh, University of Pittsburgh, Pittsburgh, PA 15213; and Department of Immunology, Nagoya University School of Medicine, 65 Tsurumai, Showa, Nagoya, 466 Japan

Background. **α1,3-galactosyltransferase (α1,3GT) is an enzyme that produces carbohydrate chains termed αGal epitopes found in most mammals, although some species of higher primates, including human, are notable exceptions. The evolutionary origin of the lost α1,3GT enzyme activity is not yet known, although it has been suggested that the promoter activity of this gene in the ancestors of higher primates was inactivated.**

Methods. **We used 5′-or 3′-RACE, GenomeWalking, reverse transcriptase polymerase chain reaction (RT-PCR) and dual Luciferase reporter assay for identification of the full-length cDNA, which includes the transcription initiation site and the promoter region of porcine α1,3GT gene.**

Results. **The region around exon 1 is guanine and cytosine (GC)-rich (about 70%), comprising a CpG island spanning more than 1.5 kbp. The 5′-flanking region of exon 1 contains multiple transcription factor consensus motifs, including GC-box, SP1, AP2, and GATA-box sites, in the absence of TATA or CAAT-box sequences. The entire gene consists of three 5′ noncoding and six coding region exons spanning more than 52 kbp. Detailed analysis of α1,3GT transcripts revealed two major alternative splicing patterns in the 5′-untranslated region (5′-UTR) and evidence for minor splicing activity that occurs in a tissue-specific manner. Interspecies comparison of 5′-UTR shows minimal homology between porcine and murine sequences except for exon 2, which suggests that the regulatory regions differ among species.**

Conclusions. **These observations have important implications for experiments involving genetic manipulation of the α1,3GT gene in transgenic animals in terms of promoter utilization, and particularly in genetically engineering cells for the animal cloning technology by nuclear transfer.**

α1,3-galactosyltransferase (α1,3GT) is a member of the family of Golgi membrane-bound enzymes known as glycosyltransferases. In general, these enzymes catalyze the transfer of a monosaccharide sugar residue (i.e., galactose) from a particular covalent sugar-nucleotide complex (i.e., uridine diphosphate–galactose, UDP-Gal) to an acceptor molecule, either a glycoprotein or a glycolipid. α1,3GT specifically catalyzes the following reaction:

$$\text{UDP-Gal} + \text{Gal-}\beta\text{1,4-GlcNAc-R} \rightarrow \text{Gal-}\alpha\text{1,3-Gal-}\beta\text{1,4-GlcNac-R} + \text{UDP} \quad (1)$$

where Gal-β1,4-GlcNAc-R is galactose in β1,4 glycosidic linkage with N-acetylglucosamine attached to either a glycoprotein or glycolipid (R). Importantly, the addition of a galactose residue in α1,3 glycosidic linkage to the Gal-β1,4-GlcNAc-R substrate creates what is known immunologically as an αGal epitope.

Patterns of αGal epitope expression on the surfaces of cells from different species correlates with the genetic expression of a functional α1,3GT enzyme. Evolutionarily, αGal epitopes seem to have first appeared in mammals (notably absent from reptiles, amphibians, birds and fish) but eventually disappeared in higher primate species (Catarrhines), including apes, Old World monkeys and humans (*1*, *2*). A major consequence of α1,3GT functional disruption is the loss of immunological tolerance to αGal epitopes. Normal human and higher primate subjects have significant titers of preformed anti-αGal immunoglobulins in serum (*3*), such that cells or tissues derived from αGal-positive mammals are hyperacutely rejected when they are introduced into αGal-negative species (*4*).

Efforts to explain the absent expression of α1,3GT in hu-

[1] Supported in part by the Imanaga Fund, Nagoya, Japan; the McCune Foundation, Pittsburgh, PA; the National Institute of Health grants DK 29961, R01 AI/DK 38899, AI 40329, DK 49615, R01 DK 54232; and the Juvenile Diabetes Foundation 4–1999-807.

[2] Thomas E. Starzl Transplantation Institute and Department of Surgery, University of Pittsburgh.

[3] Division of Immunogenetics, Department of Pediatrics, Children's Hospital of Pittsburgh, University of Pittsburgh.

[4] Department of Immunology, Nagoya University School of Medicine.

[5] Address correspondence to: Chihiro Koike, MD, PhD, Thomas E. Starzl Transplantation Institute, University of Pittsburgh, 3601 Fifth Avenue, Pittsburgh, PA 15213.

mans and other higher primates have identified two homologous gene sequences on human chromosomes 9 (*GGTA1*) and 12 (*GGTA1P*) (*5*), the latter of which is a processed pseudogene that cannot function due to multiple coding region mutations. Transcripts from the chromosome 9 homologue, *GGTA1*, have not been identified from normal human tissues, implying alteration of sequence control elements vital to the appropriate regulation of α1,3GT gene transcription. It has been suggested, however, that transcriptional activation of α1,3GT may occur within human cells during certain disease states, including neoplastic and autoimmune processes (*6*, *7*). Cell surface expression of αGal epitopes in this manner generally enhances the immunogenicity of the expressing tissue.

The normal lack of immune tolerance to αGal epitopes in humans precludes the use of non-human donors as sources of organs and tissues for transplantation into human subjects (xenotransplantation) because of the hyperacute rejection phenomenon. However, available molecular biological tools have made it possible to genetically manipulate organisms to modulate the expression of genes that unfavorably influence the outcome of xenotransplantation experiments. As a further step toward developing a strategy to circumvent the hyperacute rejection that targets αGal epitopes, we have set out to fully characterize the transcriptional regulation of the porcine α1,3GT gene responsible for the creation of αGal epitopes on porcine tissues. Deciphering the regulation of this gene in relevant porcine tissues and identifying the genomic sequences responsible for its expression are expected to enhance our ability to modify the immune response to αGal in model systems.

This report includes an extensive analysis of porcine α1,3GT transcripts from multiple porcine tissues, as well as a comparison with similar data for transcripts found in other species. Although other investigators have obtained α1,3GT cDNA clones from murine (*8*, *9*), bovine (*10*), and porcine (*11–14*) tissues, none of these clones represents the full-length α1,3GT gene transcript, including the transcription initiation site, nor do they point to the genomic sequences likely responsible for transcript expression from this genetic locus.

Using a combination of RT-PCR experiments, 5′-RACE, GenomeWalking, and long PCR, we have defined the exon-intron boundaries and suggestive tissue-specific expression patterns for alternative transcripts of porcine α1,3GT and have identified the likely transcription initiation site and regulatory elements controlling expression of this gene from porcine genomic DNA.

MATERIALS AND METHODS

Cells and tissues. Porcine fetal tissues, including aorta, brain, and liver, were obtained from a local slaughterhouse. Samples to be used later for isolation of DNA or RNA were flash frozen in liquid nitrogen, whereas aortic tissue was treated with collagenase in phosphate-buffered saline and pig aortic endothelial cells (PAEC) were isolated. PAEC were maintained in Dulbecco's modified Eagle medium (DMEM, Gibco, Grand Island, NY), 10,000 U of heparin sodium (Elkinns-Sinn, Inc., Cherry Hill, NJ), 15 mg endothelium growth supplement (Collaborative Biomedical Products, Inc., Bedford, MA), L-glutamine, and penicillin-streptomycin. Culture flasks were kept loosely capped in a 37°C incubator with an atmosphere of 5% CO_2.

Isolation of nucleic acids. To isolate porcine genomic DNA, PAEC were grown to confluence in tissue culture flasks, trypsinized briefly at 37°C, and pelleted by centrifugation. High molecular weight porcine DNA was recovered using a standard protocol involving phenol-chloroform extraction, overnight incubation with RNase A, isopropanol precipitation, and spooling of precipitated DNA.

Total RNA was extracted from fetal tissue samples and cultured PAEC using Trizol reagent (Gibco) according to the manufacturer's instructions. For experiments in which polyadenylated (poly A^+) RNA was used, poly A^+ RNA was separated from total RNA using the Dynabeads mRNA Purification Kit (Dynal, Oslo, Norway) in accord with the protocol provided. Total yield of poly A^+ RNA ranged from 1–5% of total RNA.

Genome walking and long PCR amplification of genomic DNA. A porcine GenomeWalker library was constructed using the Universal GenomeWalker Library Kit (Clontech, Palo Alto, CA). Briefly, five aliquots of porcine genomic DNA were separately digested with a single blunt-cutting restriction endonuclease (*Dra*I, *Eco*RV, *Pvu*II, *Sca*I, or *Stu*I). After phenol-chloroform extraction, ethanol precipitation and resuspension of the restricted fragments, a portion of each digested aliquot was used in separate ligation reactions with the GenomeWalker adapters provided with the kit. This process created five "libraries" for use in the PCR-based cloning strategy of GenomeWalking. A primary PCR amplification of each library using a gene-specific primer (GSP1) and an adapter sequence-specific primer (AP1) was followed by a secondary reaction in which a nested gene-specific primer (GSP2) and a nested adapter primer (AP2) were used to amplify the diluted primary PCR reaction product. Either eLONGase or TaKaRa LA *Taq* (Takara Shuzo Co., Ltd., Shiga, Japan) enzyme was used for PCR in all GenomeWalker experiments as well as for direct long PCR of genomic DNA. The thermal cycling conditions recommended by the manufacturer were employed in all GW-PCR experiments on a Perkin Elmer Gene Amp System 9600 or 9700 thermocycler.

5′- or 3′-RACE analyses. To identify the 5′ and 3′ ends of α1,3GT gene transcripts, 5′- and 3′-RACE procedures were performed using the Marathon cDNA Amplification Kit (Clontech) with PAEC poly A^+ RNA as template. First strand cDNA synthesis from 1 μg of poly A^+ RNA was accomplished using 20 U of AMV-RT and 1 pmol of the supplied cDNA Synthesis Primer by incubating at 48°C for 2 hr. Second strand cDNA synthesis involved incubating the entire first strand reaction with a supplied enzyme cocktail composed of RNase H, *Escherichia coli* DNA polymerase I, and *E. coli* DNA ligase at 16°C for 1.5 hr. After blunting of the double-stranded cDNA ends by T4 DNA polymerase, the supplied Marathon cDNA Adapters were ligated to an aliquot of purified, double-stranded cDNA. Dilution of the adapter-ligated product in 10 mM tricine-KOH/0.1 mM EDTA buffer provided with the kit readied the cDNA for PCR amplification. To obtain the 5′- and 3′-most sequences of α1,3GT transcripts, PCR reactions using the Marathon cDNA Adapter and Nested Adapter primers (mAP1 and mAP2) were paired with gene-specific and nested gene-specific primers (mGSP1 and mGSP2) designed from the coding region. To facilitate amplification through the GC-rich portions of the 5′-untranslated sequences, the GC buffer supplied with the LA *Taq* enzyme was included in the PCR mix. 5′- and 3′-RACE products were subcloned and sequenced as described below.

RT-PCR. First strand cDNA template for RT-PCR from PAEC or tissue total RNA was generated with the SuperScript Preamplification System (Gibco) according to the manufacturer's protocol. Briefly, 5 μg of total RNA were reverse transcribed using SuperScript II and the oligo(dT) primers provided, followed by digestion of the RT reaction with RNase H. Amplification of 1–2 μL of this first strand cDNA reaction allowed for analysis of α1,3GT transcript splicing patterns using primers specific for exons 1, 2, 4 and 9 of the α1,3GT gene.

Subcloning and sequencing of amplified products. PCR products amplified from genomic DNA, Gene Walker-PCR (Clontech), RT-PCR, and 5′- or 3′-RACE were gel-purified using the Qiagen Gel Extraction Kit (Qiagen, Valencia, CA), if necessary, then subcloned into the pCR II vector provided with the Original TA Cloning Kit (Invitrogen, Carlsbad, CA). Plasmid DNA minipreps of pCR II-li-

gated inserts were prepared with the QIAprep Spin Miniprep Kit (Qiagen) as directed. Automated fluorescent sequencing of cloned inserts was performed using an ABI 377 Automated DNA Sequence Analyzer (Applied Biosystems, Inc., Foster City, CA) with either the dRhodamine or BigDye Terminator Cycle Sequencing Kits (Applied Biosystems) primed with T7 and SP6 promoter primers or primers designed from internal insert sequences.

Primer synthesis. All oligonucleotides used as primers in the various PCR-based methods were synthesized on an ABI 394 DNA Synthesizer (Applied Biosystems, Inc., Foster City, CA) using solid phase synthesis and phosphoramidite nucleoside chemistry, unless otherwise stated.

Transcription factor binding site analysis of 5′-flanking sequences. To assess the 5′ flanking region of the porcine α1,3GT gene for potential regulatory sequences, genomic DNA sequences spanning this region were analyzed using the MatInspector v2.2 software program available online (http://genomatix.gsf.de). A core similarity of ≥0.900 on the MatInspector report was considered to be significant for matches of query sequences with known transcription factor recognition sequences. The positions of significant sequence matches within the 5′ flanking region of α1,3GT are indicated in Figure 3.

Promoter activity assay. To assess the 5′-flanking region of the porcine α1,3GT gene for specific promoter activity, a luciferase reporter gene construct was assembled from the cloned porcine 5′ flanking region of genomic DNA and the promoterless pGL3 Basic vector containing the firefly (FF) luciferase cDNA downstream of a multiple cloning site (MCS). Briefly, a fragment of 1910 bp upstream of a *Bss*HII restriction site in the α1,3GT exon 1 was cloned into the MCS immediately upstream of the luciferase gene coding region to generate a construct termed pGL3-ZX. This insert is comprised of 1550 bp of the upstream sequence along with 360 bp of exon 1. Transient transfection of pGL3-ZX into PAEC using Lipofectamine reagent (Gibco) followed by lysis of the transfected cells allowed for an assessment of FF luciferase gene expression driven by the insert DNA sequence. For each assay, 2×10^5 PAEC/well of a six-well plate were set up in triplicate for each transfected vector. Strength of pGL3-ZX promoter activity was measured by comparison to basal luciferase expression from the promoterless pGL3-Basic vector transfected in the same assay. To monitor transfection efficiency, a coreporter vector encoding Renilla luciferase driven by the herpes simplex virus thymidine kinase promoter (pRL-TK) was included at a ratio of 1:10 of coreporter plasmid to experimental promoter construct (or pGL3 Basic) in the transfection mixture. Lysis of cells was achieved by addition of Passive Lysis Buffer (Promega, Madison, WI) to each well followed by scraping of the attached PAEC with a cell scraper. The lysates were then subjected to two freeze-thaw cycles and stored at −70°C until luciferase activity was measured.

Measurements of FF and Renilla luciferase activity of each lysate was measured sequentially in a Monolight 2010 luminometer using reagents from the Dual-Luciferase Reporter Assay System (Promega) according to the protocol. To compare inter-construct FF luciferase activity values, the raw data relative light unit (RLU) readings were corrected by normalizing each sample according to transfection efficiency. The highest measured Renilla luciferase RLU measurement was used to normalize all values in a given experiment according to the following formula: (highest R-RLU/sample R-RLU) x sample FF RLU=normalized FF luciferase activity. The normalized triplicate values for each construct were then averaged to arrive at a relative measure of luciferase activity. Comparison of these average values to the average FF luciferase activity for the pGL3-Basic vector resulted in a fold-increase measure of promoter activity over the promoterless control vector.

RESULTS

Genomic Organization of Porcine α 1,3GT

Working from the published sequences of α1,3GT cDNA clones and sequence obtained from our own 5′-RACE analysis (see below), we were able to determine the complete genomic organization for the gene. First, to delineate the 5′ UTR exon-intron boundaries, estimations of exon size were made based on comparison of porcine cDNA sequences to those of murine α1,3GT (*9*). Gene-specific primers (GSPs) for use in GenomeWalker PCR (GW-PCR) experiments were designed based on these estimations. The sequences of crucial primers used in all analyses appear in Table 1, along with information regarding the location of each primer's target sequence and the type of experiment(s) in which each was used. Automated fluorescent sequencing of GW-PCR clones extending from exon sequences into the upstream or downstream intron verified the exact positions of exon-intron boundaries. All three of the 5′ UTR exons followed the AG-GT rule for splice acceptor-donor participation with one another, as well as for the splicing of either exon 1 or 2 to coding region exon 4, which contains the translation start codon. Exon 3 has not been observed in any tissue examined. Exon-intron boundaries for the coding region exons were also confirmed by long PCR or GW-PCR, and also found to adhere to the AG-GT rule. In our investigations, exon 5 has not been observed in PAEC. The sequences of all exon-intron bound-

TABLE 1. Sequences of oligonucleotides used as PCR primers

Name	Sequence	Origin	Direction	Purpose
4U1	5′-CTGTTGATGTATTCCCAAAACACAACCATTACAGT-3′	Exon 4	Antisense	GW, 5′-RACE
4U2	5′-AGACAAGCAGCATTGACAGAACCACTC-3′	Exon 4	Antisense	GW, 5′-RACE
2U1	5′-CTCATCCTCTGCTTCTCTCCCCCCA-3′	Exon 2	Antisense	GW
2U2	5′-CCCCCCAGAGTAAAAGGCGAAACAAG-3′	Exon 2	Antisense	GW
2D1	5′-AACGCAGCACCTTCCCTTCCTCCCA-3′	Exon 2	Sense	GW
2D2	5′-CTTGTTTCGCCTTTTACTCTGGGGG-3′	Exon 2	Sense	GW
1D1	5′-GCCACTGTTCCCTCAGCCGAGGA-3′	Exon 1	Sense	GW, RT-PCR
1D2	5′-CGAGCGCACCCAGCTTCTGCCGAT-3′	Exon 1	Sense	GW, RT-PCR
1U1	5′-TGCGCTCGGGGATGGCCCTCTCCT-3′	Exon 1	Antisense	GW
1U2	5′-GGCGTCCTCGGCTGAGGGAACAGT-3′	Exon 1	Antisense	GW
2AD	5′-CAGAACAACTTCTGAAGCCTAAAGGATG-3′	Exon 2A	Sense	RT-PCR
Xk	5′-CAAATGGTGGATCGGACCTCCCAGGCT-3′	Exon 1	Sense	RT-PCR
Xh	5′-AGTACTGGGTGATAGACCCCACTCCAC-3′	Exon 1	Sense	RT-PCR
Xg	5′-GCGCAGGGCTCCGGGGCCCCTCCCT-3′	Exon 1	Sense	RT-PCR
9A	5′-CTGGGATTATCATATAGGCATGTCTGT-3′	Exon 9	Sense	GW, 3′-RACE
9B	5′-AGAGTATTACTCTGGCTACTTCTCCAG-3′	Exon 9	Sense	GW, 3′-RACE

aries have been deposited in the GenBank database under the accession numbers indicated in Table 2.

After the identification of exon-intron boundaries using GW-PCR, additional GenomeWalker experiments were used to obtain contiguous intron sequence between adjacent exons at the α1,3GT locus. Internal sequence from these bridging long PCR clones was then used to design primers for confirmation of the genomic alignment by amplifying back to the exons from which GenomeWalker experiments originated. The entire set of PCR-derived genomic subclones allowed for the construction of a sequence contig spanning the 5′ UTR exons of the α1,3GT gene, as illustrated in Figure 1. Also depicted are long PCR-amplified clones spanning the downstream coding region exons, allowing for the determination of all intervening intron lengths. Only intron 1 has not been completely spanned due to GW-PCR difficulties posed by the presence of short interspersed element (SINE) repeats in this region. Nevertheless, we estimate the genomic interval between exons 1 and 2A to be on the order of >10 kb based on sequenced data.

Variable Exon Usage in the 5′ UTR

Identification of "major" and "minor" splice patterns. Disparity among α1,3GT cDNA clone sequences published by other investigators has created some confusion regarding the true length and sequence of the 5′ UTR. To help clarify the situation, we conducted 5′-RACE and RT-PCR experiments using RNA isolated from either cultured PAEC or fetal brain and liver tissue. Our 5′-RACE results are summarized and compared with previously reported cDNA clones in Figure 2. As indicated, four basic UTR exon splice patterns have been detected. Patterns A and B are regarded as "major" UTR splice variants, while patterns C and D are designated "minor," because the former patterns are observed in every tissue examined although the latter two may or may not appear depending on the tissue examined. RT-PCR analysis of RNA from PAEC identified patterns A and B but not pattern C or D, although RNA from fetal brain or liver tissue exhibited the C and D splice patterns as well as the major patterns A and B.

Difference between patterns C and D. Splice patterns C and D are observed in cDNA clones from fetal brain or liver when RT-PCR is performed using a sense primer specific for exon 2A and an antisense primer based on the sequence of exon 4. Based on the number of clones isolated, pattern C appears to predominate over pattern D when both are present (in fetal liver and brain). This observation indicates that the region between exon 2A and exon 2 may be retained in some tissues and that the particular region that Katayama et al. postulated to be a potential transcription initiation site (TIS) locates within this retained intron (see Fig. 2, pattern D). As these authors derived their data from a 5′-RACE experiment that was not specifically tailored to the GC richness of this region, it is possible that the 5′ end of their clones represents a truncated RT extension product and does not truly represent a TIS for α1,3GT expression (*14*).

Katayama's clone D. As for the nucleotide base composition of our RACE, GW-PCR, and RT-PCR products, there is general agreement with previously published data, save one important exception. Comparison of the sequence of exon 1 derived from our clones with Katayama's clone D (termed clone *KD*) demonstrates a perfect match of the 23 bases at the 3′ end of the exon. However, there exists a great discrepancy between our exon 1 sequence and the clone *KD* which has 135 additional bp at the 5′-end (dashed box in Fig. 2). Because it was possible that this region may represent an additional UTR exon of the α1,3GT gene, we performed RT-PCR using the sense primer 5′-AGCCAGCTGGACAAGCTGACGGTCACCT-3′ based on the 135-bp region and an antisense primer (4U1 or 4U2) with cDNA derived from PAEC as well as porcine fetal brain or liver tissue. However, repeated reactions failed to generate any visible bands, indicating that there existed no conclusive evidence to suggest this possibility. Prompted by the disagreement between clone *KD* and our exon 1 sequence, we designed a second sense primer 5′-TGCAGCTGGAGAGCTGCGGATGAAGCTT-3′ based on the 135-bp sequence downstream of the first primer and used the two primers in GW-PCR experiments with our porcine GenomeWalker libraries. Single major bands were obtained from two of the libraries. These products were cloned and subject to sequence analysis. GenBank BLAST searches with these sequences revealed striking homology to exons 2, 3, and 4 of the bovine invariant chain (*Ii*) gene, as well as the intervening intron sequences (see GenBank Accession numbers D83961 and D83962). Based on close inspection and comparison of the sequences from porcine and bovine DNA, we estimated the boundaries of porcine invariant chain gene exons 2, 3, and 4 and designed an antisense primer 5′-CAGGAGCAGGTGCATCACGTGGTCCTG-3′ specific for our estimated porcine exon 4. Use of this primer in combination with the *Ii* gene sense primers in RT-PCR of PAEC or porcine fetal brain or liver cDNA did produce bands corresponding to

TABLE 2. Exon/intron organization of the porcine α1,3GT gene

Exon	Exon size (bp)	Amino acids encoded	GC-content in exon (%)	Intron size (bp)	GenBank accession no.
1	554		71	>10 k	AF221510
2A	81		40	427	AF221511
2	112		56	15 k	AF221511
4	89	38	38	6.3 k	AF221512
5	36	12	42	1.9 k	AF221513
6	63	21	54	549	AF221514
7	105	35	45	6.7 k	AF221515
8	138	46	48	6.2 k	AF221516
9	2586	230	41		AF221517
Total	3766	371	47	>52 k	

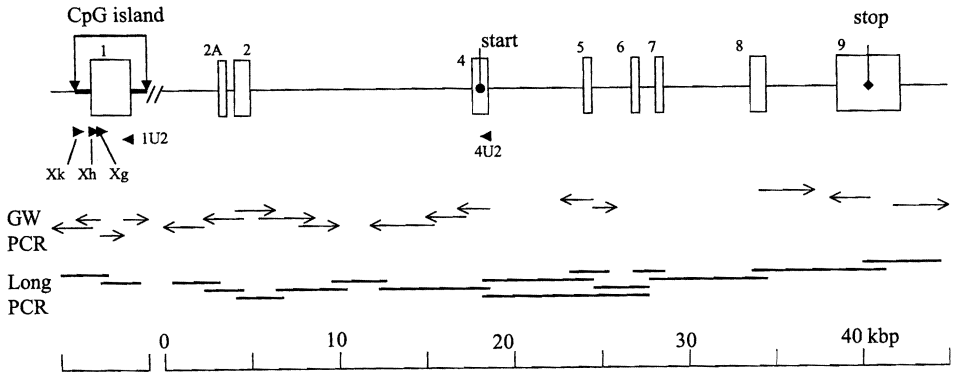

FIGURE 1. Schematic genomic organization of the porcine α1,3GT gene. Numbered boxes indicate the approximate location of exons within the locus, but do not accurately reflect the length of each exon. A set of overlapping subgenomic clones derived from GenomeWalking (horizontal arrows) or long PCR experiments (thick horizontal lines) is arrayed as a contig below the gene locus representation. The translation start and stop codons are located in exons 4 and 9, respectively. The symbols Xk, Xh, Xg, 1U2, or 4U2 indicate the location of primers used in 5′-RACE or RT-PCR experiments to confirm the transcription initiation site. For specific information regarding exon and intron lengths, refer to Table 2.

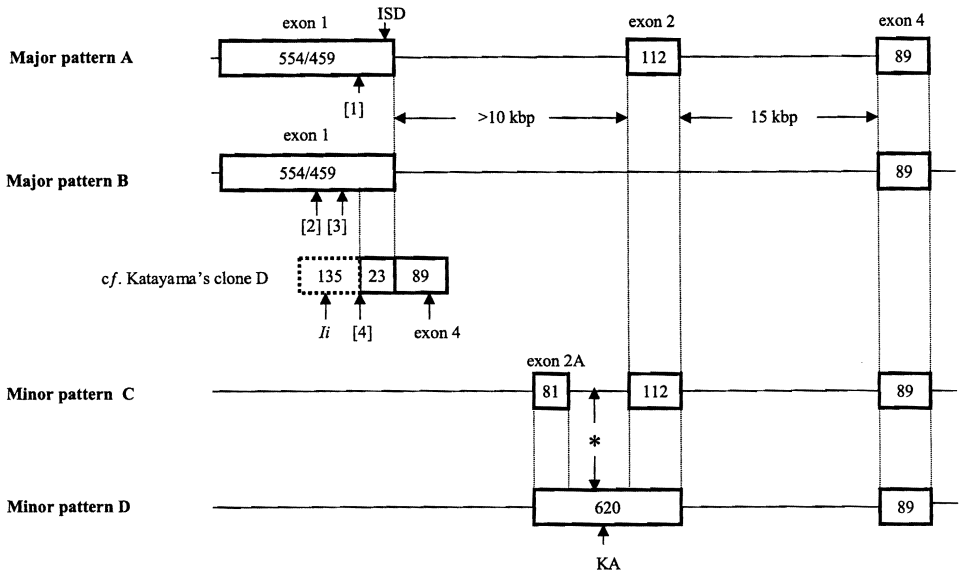

FIGURE 2. 5′-UTR splicing patterns of porcine α1,3GT. Exons are represented by numbered open boxes, indicating the length of each exon in base pairs. Splice patterns A and B are considered to be "major" splice patterns, as they are detected more frequently than the "minor" patterns C and D in all of the porcine tissues examined (*exception*: in PAEC patterns C and D are not observed). As a reference for this and other figures, the 3′-most end of exon 1 is denoted as nucleotide position +1, with ascending numbering extending in the 5′ direction. ISD indicates the position of an internal splice donor site that truncates exon 1 by 41 bp relative to other transcripts detected. Bracketed numbers correspond to the 5′ ends of transcripts detected by other investigators as follows: [2]=Sandrin's clone and [1], [3], [4]=Katayama's clones B, E, and D, respectively. Under major pattern B the symbol Ii indicates the 135-bp region of Katayama's clone D believed to be derived from the porcine invariant chain gene (see *Results*). Minor patterns C and D differ by the presence of a retained intron marked by the asterisk. The transcription initiation site suggested by Katayama et al. for their clone A (14) falls within this interval and is indicated by the symbol KA.

appropriately spliced porcine *Ii* gene transcripts involving our estimated exons 2, 3, and 4. Sequence analysis of these products confirmed their identity. Comparative analysis of the bovine and porcine *Ii* exon sequences revealed high homology: 87%.

Further inspection of the junction between these 135 bases from the presumed porcine *Ii* gene and the 23 bp of α1,3GT gene sequence suggests the absence of appropriate RNA splicing signals for adherence to the AG-GT rule. Genomic sequence continuing from the 3′-end of the 135 bases in our GW-PCR clone reads CGCCCC, which is in obvious disagreement with the accepted splice donor consensus: GT (*15*). We conclude from our results that the 135-bp region in question is actually derived from genomic sequences for porcine invariant chain and that the 5′ end of clone *KD* is an artifact of cDNA synthesis or aberrant RNA splicing, bearing no relation to the normal splicing of porcine α1,3GT gene.

Importantly, our RT-PCR and RACE analyses of the

α1,3GT gene 5′-UTR help to clarify some previously confusing results regarding porcine α1,3GT transcript processing, setting the stage for a complete understanding of this gene's transcriptional regulation

5′-Flanking Region and the Location of Transcription Initiation

Two potential sites for the initiation of porcine α1,3GT transcription were suggested by our 5′-RACE data, corresponding to distances of 554 and 459 bp from the 3′ end of exon 1. To analyze these results further, we designed three oligonucleotide primers (Xk, Xh, and Xg from Table 1) based on a GW-PCR clone extending beyond the sites indicated by our RACE results. Viability of these oligonucleotides as functional PCR primers for complex templates was confirmed from parallel PCR reactions in which each was paired with a downstream genomic sequence-based primer (1U2 from Table 1). All three reactions produced strong bands of the expected size from porcine genomic DNA, the identities of which were confirmed by subcloning and sequencing of the PCR products (expected lengths: Xk/1U2=554 bp, Xg/1U2=492 bp, Xh/1U2=465 bp). To test for the presence of these oligonucleotide sequences in mRNA transcripts, a primer based on the sequence of exon 4 (primer 4Ul) was paired with either Xk, Xh, or Xg in parallel RT-PCR reactions using cDNA prepared from PAEC total RNA. As shown in Figure 3A, only Xh and Xg produced visible PCR products, the two visible bands in each lane corresponding to amplified transcripts of both major splice patterns A (Xg/4U2=718 bp, Xh/4U2=691 bp) and B (Xg/4U2=606 bp, Xh/4U2=579 bp). Subcloning and sequencing of the RT-PCR products confirmed these results. From these experiments, we conclude that an initiation site for α1,3GT transcription is located in the region between the two primers Xk and Xh in the porcine genomic sequence, and that the nucleotide 554 bp upstream from the 3′-end of exon 1 derived from 5′-RACE is most likely the transcription initiation site.

Importantly, it has been identified that exon 1 is larger that previously estimated (>450 bp versus 120 bp) and that the region around exon 1 spanning about 1.5 kbp comprises so-called CpG island since this particular region is GC-rich (>70%), a characteristic commonly described for the regulatory regions of many "housekeeping genes" (see Fig. 1).

Promoter Activity

The sequence of the immediate 5′ flanking region of the porcine α1,3GT gene encompassed by our GW-PCR clone and the entirety of exon 1 are detailed in Figure 3B. The position of primers used to localize the transcription initiation site are indicated within this sequence, and the two transcript 5′ ends mapped by 5′-RACE are indicated by the numbers (*1*) and (*2*). Closer inspection and comparison of the sequence to the TRANSFAC database by the MatInspector v2.2 transcription factor binding site program reveals a richness of potential transcription factor consensus motifs. These include multiple SP1/GC-box (consensus: KGGGCGGRRY), AP2 (consensus: MKCCCSCNGGCG), and GATA-box (consensus: WGATAR) binding sites in the immediate vicinity of the 5′ end of exon 1 and within the sequence of exon 1 itself.

As shown in Figure 4, transfection of a luciferase reporter construct pGL3-ZX into PAEC produced an 81-fold increase in luciferase activity over the promoter-less pGL3-Basic vector. Based on this result, we conclude that the sequence of this region encompasses the porcine α1,3GT gene promoter controlling transcription initiation from exon 1. Because of the richness of potential transcription factor binding motifs located within the boundaries of exon 1, we believe that alternative transcription initiation and/or multiple promoters in the region may account for variations in the 5′ end of exon 1-containing α1,3GT transcripts. A detailed deletion mutation analysis intended to identify the crucial sequence elements controlling α1,3GT gene is currently underway.

3′ UTR and Polyadenylation Site

To identify the 3′ end of α1,3GT transcripts, we performed 3′-RACE and GW-PCR experiments using primers designed from exon 9 of the cDNA sequence (9A and 9B from Table 1). Analysis of 3′-RACE clones suggests that no other 3′ UTR exons exist and that the length of the 3′ UTR from the stop codon to the poly A site is 1892 bp. Thus, the overall length of exon 9 from its 5′ end to the poly A site is 2586 bp, comprising greater than two-thirds the length of the longest potential and fully processed α1,3GT mRNA transcript.

Comparison of α1,3GT Gene 5′ UTRs from Different Species

Comparison of porcine exon 2 with the 5′-UTR of bovine cDNA revealed that the region immediately upstream of the start codon, 112 bp in length, is highly homologous (83.7%, data not shown), although the homology between porcine and murine sequences was a more modest 71.7% (*9*). Similarly, porcine exon 1 and the initial 248 bp of the bovine cDNA sequence seem to be closely related with a nucleotide homology of 80% (*9*), whereas porcine exon 1 exhibits little or no homology to the sequence of murine exon 1. The sequence of porcine exon 2A appears to be unique, in that a corresponding homologous sequence has not been found in either murine or bovine cDNA sequences thus far isolated in our laboratory. In addition, a sequence homologous to murine exon 3 has not appeared in porcine PAEC or porcine fetal brain RT-PCR.

These interspecies comparison data suggest some degree of sequence conservation within the 5′ UTR of distantly related mammalian species, as well as preservation species-specific exon sequences that are still important to the expression of α1,3GT gene transcription.

DISCUSSION

Genomic organization of porcine α 1,3GT gene and evidence for tissue-specific regulation of α 1,3GT expression. In this study, we have mainly focused on the expression of α1,3GT in PAEC for the following reasons: 1) endothelial cell expression of αGal epitopes provides the molecular target for antibody binding and formation of membranous immunocomplexes, 2) genetic information derived from a single cell type repertoire is helpful in analyzing splicing patterns on other tissues or organs which have a variety of cell types, 3) the genetic engineering of porcine tissue is a major focus of research in xenotransplantation. Our results indicate that a1,3GT transcripts in PAEC are alternatively spliced to yield two major patterns in the 5′-UTR but do not exhibit the minor splicing patterns. Other tissues examined also show the two major splicing patterns in the 5′-UTR but may or

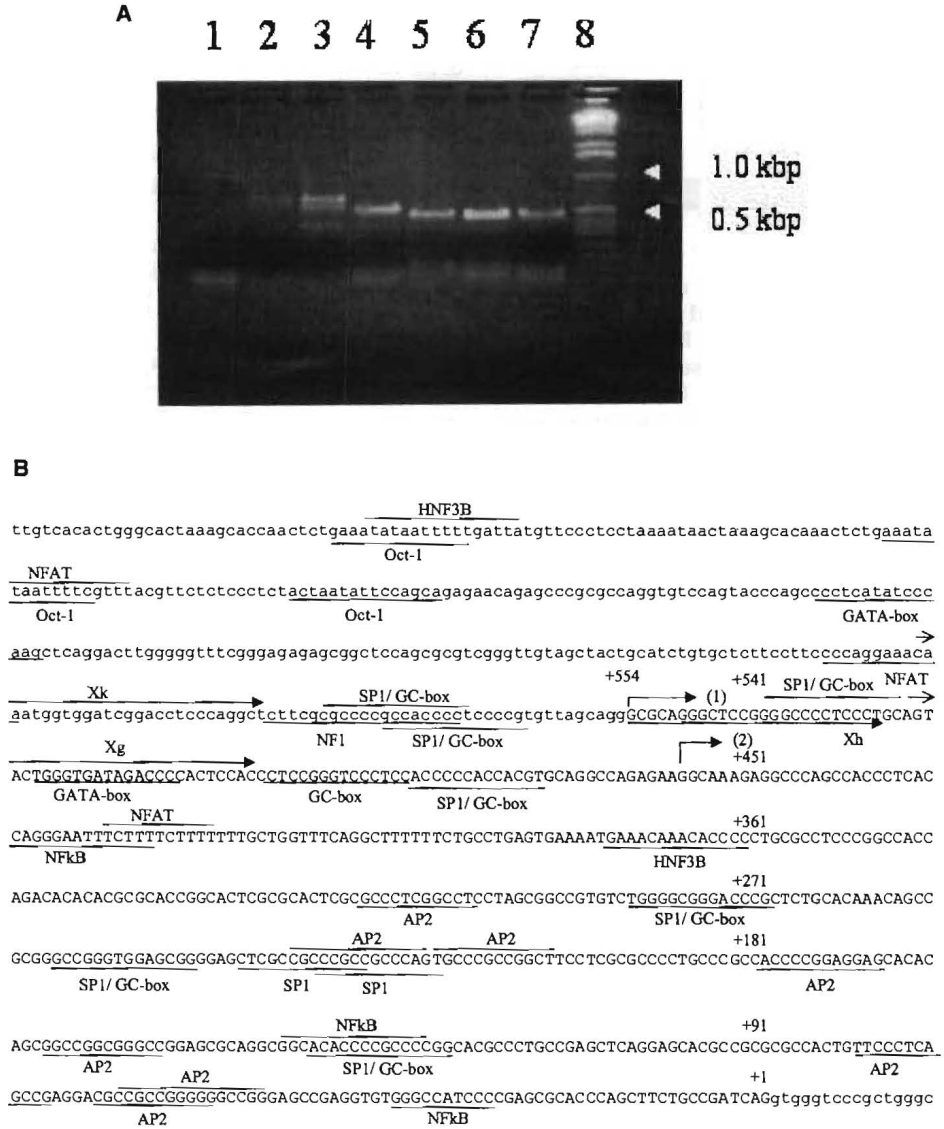

FIGURE 3. Definition of the α1,3GT 5′-flanking region. A, Electrophoresis of RT-PCR products. Localization of the transcription initiation site was performed by analyzing RT-PCR products produced by amplification of cDNA derived from PAEC. Products of RT-PCR amplification with the antisense primer 4U2 specific for exon 4 and sense primers Xk, Xh, and Xg specific for sequence around the TIS appear in lanes 1, 2, and 3, respectively. Exact sizes for the alternative splice pattern products are indicated in the text (see *Results*). Lanes 4, 5, and 6 represent the PCR reaction products obtained by amplification of porcine genomic DNA using primers Xk, Xh, and Xg, respectively, paired with the antisense primer 1U2 designed from the sequence of the downstream intron. In lane 7, a second RT-PCR reaction using Xg and 1U2 with PAEC cDNA is included. Lane 8 contains a DNA ladder with relevant band sizes indicated. In lane 1 no visible band was observed, although lanes 2 to 7 contain the expected size band(s). B, Sequence of the porcine α1,3GT 5′-flanking region and exon 1-intron boundary. Nucleotides comprising exon 1 are shown in all capital letters although nucleotides of the 5′-flanking sequence and the downstream intron are shown in lower case letters. The 3′ end nucleotide of exon 1 is numbered +1. The genomic fragment included in the luciferase reporter assay construct pGL3-ZX spans the upstream nucleotides +2000 to +90. Underlining indicates potential transcription factor binding sites identified MatInspector v2.2 analysis of the sequence against the TRANSFAC database. Bent-forward arrows identified with [1] or [2] indicate potential transcription initiation sites (TIS) as identified by 5′-RACE. Long horizontal arrows with the labels Xk, Xh, or Xg indicate PCR primers used to localize the TIS.

may not have the minor patterns. The occurrence of the minor splice patterns C and D involving exons 2A and 2 warrant additional studies of a possible alternative promoter upstream of exon 2A.

Comparison of the genomic organization with other glycosyltransferases. Of the estimated 100 glycosyltransferases identified on the basis of substrate specificity and glycosidic linkage formed, only about 10 enzymes have been cloned to date. From a comparison of the primary structures of the cloned genes of glycosyltransferases, two patterns have emerged. One is that of a gene spanning over six coding exons, although the other is expressed from a single coding exon. Examples of the former are the β1,4-galactosyltransferase gene (β1,4GT) (*16*), the blood group ABO transferase

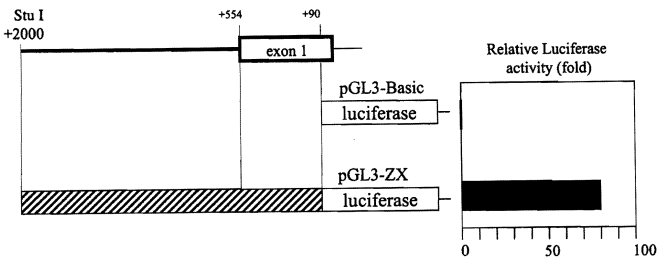

FIGURE 4. Functional assay for promoter activity of the 5′ flanking region. The 5′-flanking region and exon 1 of porcine α1,3GT are indicated by a bold horizontal line and bold-outlined open box. The segment of genomic sequence tested in the luciferase reporter assay is indicated below by a cross-hatched bar fused just upstream of the FF luciferase gene contained in the pGL3-Basic vector. The pGL3-ZX construct demonstrated an 81-fold increase in luciferase activity from transfection of PAEC as compared to the promoterless pGL3-Basic vector as shown in the bar graph at right. Normalization of raw data values to sample Renilla luciferase activity (cotransfectant for transfection efficiency) was performed before comparing triplicate-average values for the different firefly luciferase reporter constructs.

genes (ABO-T) (*17*), and the α2,6-sialyltransferase gene (α2,6ST) (*18*). An example of the latter is the α1,2-fucosyltransferase (α1,2FT) gene, which spans less than 3 kbp including its intron (*19–21*). The multi-coding exon type of glycosyltransferase normally spans more than 40 kbp of genomic sequence, typified by the ABO-T genes that are comprised of 7 coding exons spanning more than 90 kbp. The α1,3GT gene falls into the same general pattern, having three 5′-untranslated and six coding exons distributed over more than 52 kbp.

Information regarding the regulatory regions of these genes is more limited. The regulatory region of the murine β1,4GT is reported to be GC-rich and without a TATA-box (*16*). Comparison of the 5′-UTR regions between murine β1,4GT and the porcine α1,3GT gene reveals that they are not homologous. Enzymatically, α1,2FT, α2,6ST, or α1,3GT gene share a commonality in that each utilizes the same precursor: *N*-acetyllactosamine, the final modification product of α1,4GT. However, although the α2,6ST gene does contain a TATA-box in rat (*20*) as well as in human (*21*), the α1,2FT gene and α1,3GT gene have the same features in terms of GC-richness and lack of a TATA-box in genomic DNA. The biological significance for these differences requires further investigation.

Interspecies comparison in the regulatory region. Interspecies comparisons among 5′-UTRs of porcine and murine α1,3GT cDNA sequences indicates that only minimal homology exists between the transcripts, with the exception of exon 2. In depth studies of cross-species α1,3GT promoter activity, including those of the murine, porcine, and bovine genes are planned. It is hoped that interspecies comparisons of α1,3GT promoter activity will shed light on the possible mechanisms for inactivation of human α1,3GT gene expression once the genomic region of the presumed human α1,3GT homologue, *GGTA1* on chromosome 9, has been completely sequenced. To date we have successfully cloned the 5′-flanking region of murine exon 1 and, in pilot transfection studies, have detected some promoter activity for the murine sequence in a luciferase reporter assay (data not shown). Surprisingly the murine 5′-flanking sequence exhibits little homology to that of porcine α1,3GT. Most notably the porcine sequence extending from the 5′ flanking region through the entirety of exon 1 fulfills the criteria for a putative CpG island (*22*). A similar sequence composition is not observed at the murine locus. This finding may have important implications for the study of α1,3GT gene regulation because, in general, vertebrate genes with CpG islands in their promoter regions have a potential to show developmental or cell-stage specific regulation depending on their level of methylation. Taken together, these observations suggest that the regulation of α1,3GT gene expression may differ between the two species even though the enzyme performs an identical function. This possibility will be of special concern to the field of xenotransplantation, as the extension of studies in the mouse will generally be undertaken with the hopes of extending experimental findings to human or other model animal systems. Based on the future realistic possibility of transplantation from pigs to humans, we believe that careful study of the porcine regulatory region of α1,3GT will lead to more fruitful research avenues involving genetic manipulation of porcine cells or tissues to be used for xenotransplantation experiments or therapies.

Acknowledgments. The authors thank Drs. Reiji Kannagi, Nozomu Hiraiwa, Annalisa Licata, and William Rudert for kind discussions; Therese Libert, Jennifer Profozich, Suzanne Bertera, Yuki Ito, Shigeko Nakano, Natsuko Noguchi, Mr. Adam Shaefer, Drs. Lee Yi Horng, and Tang Jian-Lin, Drs. Satoshi Hayashi and Fumiko Akutsu of the Itochu Feed Mill Co. Ltd., and Dr. Hiroto Yamakawa of the Nisshin Flour Milling Co. Ltd. for their technical assistance; and Kyoko Maruyama, Atsuko Nakamura and Terry Mangan for secretarial assistance.

REFERENCES

1. Galili U, Shohet SB, Kobrin E, Stults CLM, Macher BA. Man, apes, and Old World monkeys differ from other mammals in the expression of α-galactosyl epitopes on nucleated cells. J Biol Chem 1988; 263: 17755.
2. Galili U, Swanson K. Gene sequences suggest inactivation of alpha-1,3-galactosyltransferase in catarrhines after the divergence of apes from monkeys. Proc Natl Acad Sci USA 1991; 88: 7401.
3. Galili U, Clark MR, Shohet SB, Buehler J, Macher BA. Evolutionary relationship between the natural anti-Gal antibody and the Galα1,3-Gal epitope in primates. Proc Natl Acad Sci USA 1987; 84: 1369.
4. Cooper DKC, Koren E, Oliol R. Genetically engineered pigs. Lancet 1993; 1: 682.
5. Joziasse DH, Shaper JH, Jabs EW, Shaper NL. Characterization of an α1,3-galactosyltransferase homologue on human chromosome 12 that is organized as a processed pseudogene. J Biol Chem 1991; 266: 6991.
6. Etienne-Decerf J, Malaise M, Mahieu P, Winand R. Elevated anti-alpha-galactosyl antibody titres. A marker of progression in autoimmune thyroid disorders and in endocrine ophthalmopathy? Acta Endocrinol 1987; 115: 67.
7. Winand RJ, Devigne JW, Meurisse M, Galili U. Specific stimulation of Graves' disease thyrocytes by the natural anti-Gal antibody from normal and autologous serum. J Immunol 1994; 153: 1386.
8. Larsen RD, Rajan VP, Ruff MM, Kukowska-Latallo J, Cummings RD, Lowe JB. Isolation of a cDNA encoding a murine UDP-galactose: β-D-galactosyl-1,4-N-acetyl-D-glucosaminide

α-1,3-galactosyltransferase: expression cloning by gene transfer. Proc Natl Acad Sci USA 1989; 86: 8227.
9. Joziasse DH, Shaper NL, Kim D, Van den Eijinden DH, Shaper JH. Murine α1,3-galactosyltransferase. J Biol Chem 1992; 267: 5534.
10. Joziasse DH, Shaper JH, Van den Eijinden DH, Van Tunen AJ, Shaper NL. Bovine α1,3-galactosyltransferase: isolation and characterization of a cDNA clone. J Biol Chem 1989; 264: 14290.
11. Sandrin MS, Dekowski PL, Henning MM, Mouhtouris E, McKenzie IFC. Characterization of cDNA clones for porcine α(1,3)galactosyl transferase: the enzyme generating the GalαGal epitope. Xenotransplantation 1994; 1: 81.
12. Starahan K, Gu F, Preece AF, Gustavsson I, Andersson L, Gustafsson K. cDNA sequence and chromosome localization of pig α1,3 galactosyltransferase. Immunogenetics 1995; 41: 101.
13. Vanhove B, Goret F, Soulillou JP, Pourcel C. Porcine α1,3-galactosyltransferase: tissue-specific and regulated expression of splicing isoforms. Biochim Biophys Acta 1997; 1356: 1.
14. Katayama A, Ogawa H, Kadomatsu K, et al. Porcine α-1,3-galactosyltransferase; full length cDNA cloning, genomic organization, and analysis of splicing variants. Glycoconj J 1998; 16: 583.
15. Shapiro MB, Senapathy P. RNA splice junction of different classes of eukaryotes: sequence statics and functional implications in gene expression. Nucleic Acids Res 1987; 15: 7155.
16. Shaper NL, Harduin-Lepers A, Shaper JH. Male germ cell expression of murine β4-galactosyltransferase. J Biol Chem 1994; 269: 25165.
17. Yamamoto F, McNeill PD, Hakomori S. Genomic organization of human histo-blood group ABO genes. Glycobiology 1995; 5: 51.
18. Svensson EC, Soreghan B, Paulson JC. Organization of β4-galactoside α2,6-sialyltransferase gene. J Biol Chem 1990; 265: 20863.
19. Soejima M, Koda Y, Wang B, Kimura H. Functional analysis of the 5′-flanking region of FTA for expression of rat GDP-L-fucose:β-D-galactoside 2-α-L-fucosyltransferase. Eur J Biochem 1999; 266: 274.
20. Loa NW, Lau JTY. Transcription of b-galactoside α2,6-sialyltransferase gene in B lymphocytes is directed by a separate and distinct promoter. Glycobiology 1996; 6: 271.
21. Svensson EC, Conley P, Paulson JC. Regulated expression of α2,6-sialyltransferase by the liver-enriched transcription factors HNF-1, DBP, and LAP. J Biol Chem 1992; 267: 3466.
22. Jones PA. The DNA methylation paradox. Trends Genet 1999; 15 (1): 34.

Received 29 March 2000.
Accepted 24 July 2000.