Molecular Basis of Evolutionary Loss of the \( \alpha_1,3\)-Galactosyltransferase Gene in Higher Primates*


From the §Thomas E. Starzl Transplantation Institute and Department of Surgery, University of Pittsburgh and the §Division of Immunogenetics, Department of Pediatrics, Children’s Hospital of Pittsburgh, University of Pittsburgh, Pittsburgh, Pennsylvania 15213, the JDepartment of Pathology, Aichi Cancer Research Center, 81-1159 Kanokoden, Chikusa, Nagoya 465, Japan, and the **Department of Immunology, Nagoya University School of Medicine, 65 Tsurumai, Showa, Nagoya 466, Japan.

Galactose-\( \alpha_1,3\)-galactose (\( \alpha \)Gal) epitopes, the synthesis of which requires the enzyme product of \( \alpha_1,3 \)-galactosyltransferase (\( \alpha_1,3 \)GT), are sugar chains on the cell surface of most mammalian species. Notable exceptions are higher primates including Old World monkeys, apes, and humans. The \( \alpha \)Gal-negative species as well as mice with deletion of the \( \alpha_1,3 \)GT gene produce abundant anti-\( \alpha \)Gal antibodies. The evolutionary loss of \( \alpha \)Gal epitopes has been attributed to point mutations in the coding region of the gene. Because no transcripts could be found in the higher primate species with Northern blot analysis, a potential alternative explanation has been loss of upstream regulation of the gene. Here, we have demonstrated that the rhesus promoter is functional.

Five crucial mutations were delineated in the coding region of the human and rhesus and three in the upstream regulatory region of the \( \alpha_1,3 \)GT gene. Because no transcripts could be produced, we detected with sensitive PCR-based methods in the tissues of rhesus monkeys, orangutans, and humans. Five crucial mutations were delineated in the coding region of the human and rhesus and three in the orangutan, any one of which could be responsible for inactivation of the \( \alpha_1,3 \)GT gene. Two of the mutations were shared by all three higher primates. These findings, which elucidate the molecular basis for the evolutionary loss of \( \alpha \)Gal expression, may have implications in medical research.

Most mammals express the cell surface carbohydrate epitope galactose-\( \alpha_1,3 \)-galactose (\( \alpha \)Gal), with notable exceptions that include Old World monkeys, apes, and humans (1). With the loss of the \( \alpha \)Gal epitopes, the synthesis of which is dependent on enzyme product of the \( \alpha_1,3 \)-galactosyltransferase (\( \alpha_3 \)GT) gene, higher primates produce anti-\( \alpha \)Gal antibodies (2) that are responsible for the hyperacute rejection of organs transplanted from \( \alpha \)Gal-positive donors (3).

In 1989, Joziasse et al. (4) reported the sequence of full-length cDNA clone of the bovine \( \alpha_3 \)GT gene and demonstrated the presence of this gene in the DNA of human cell lines. Using an 804-bp fragment derived from bovine full-length cDNA as a probe, they also detected mRNA transcripts in bovine and marmoset (New World monkey) but not in human or African green monkey cell lines (4). The absence of \( \alpha_3 \)GT mRNA has since been widely viewed as a feature of all Old World monkeys, apes, and humans (5, 6). The molecular basis for the inactivation of the \( \alpha_3 \)GT gene in the \( \alpha \)Gal-negative species has been attributed to a mutation(s) localized to a partial sequence of exon 9 (4, 5, 7).

Stimulated by the current interest in producing transgenic pigs for clinical use as tissue and organ xenograft donors, the cDNA for the \( \alpha_3 \)GT gene of the \( \alpha \)Gal-positive pig was isolated (8), and the coding regions were characterized (9–12). Further, the full genomic organization of the porcine \( \alpha_3 \)GT gene including its upstream regulatory region, was completed (12). A CpG island (i.e. a C connected by a 3′–5′ phosphodiester bond to a G) characteristic of housekeeping genes was observed around exon 1 of the pig \( \alpha_1,3 \)GT gene (12). It is of interest that a CpG island was not present in the mouse gene (13–15), suggesting that the upstream regulatory regions might be species-specific even though the enzyme function of the \( \alpha_1,3 \)GT gene product was identical (15). If so, it would be conceivable that the loss of \( \alpha \)Gal expression in higher primates could be due to mutation in the upstream regulatory region (16).

As a further step in a phylogenetic comparison of the \( \alpha \)Gal-positive pig with the higher primates, we employed a PCR-based cloning strategy to avoid the potential artifacts of extrapolating from partial sequences and cross-species hybridization. After obtaining the full coding region of the \( \alpha_3 \)GT gene of the \( \alpha \)Gal-negative rhesus monkey, orangutan, and human, as well as the genomic organization and upstream regulatory region in the rhesus and human. In all three \( \alpha \)Gal-negative primate species, we found that the \( \alpha_3 \)GT gene produces a variety of transcripts that were confirmed in the rhesus samples with Northern blot analysis. The transcripts contained crucial mutations in addition to the one(s) described previously in exon 9 of the rhesus (6), orangutan (6), and human (5, 6). Taken together, our findings suggest substantial revisions of the characterization of the \( \alpha_1,3 \)GT gene.
that have implications for the molecular study of primate evolution and for the planning of xenotransplantation initiatives.

EXPERIMENTAL PROCEDURES

Genomic Organization of Human, Orangutan, and Rhesus Monkey 1,3GT Genes—After high molecular genomic DNA was isolated from the PBMC of multiple healthy adult male humans, two rhesus monkeys, and an orangutan, libraries for the three species were constructed using the Universal GenomeWalker™ Library Kit (CLONTECH, Palo Alto, CA). The initial gene-specific and nested primer pairs were designed from the exonic sequences obtained from the RACE experiments, primers were designed and nested primer pairs for other regions were designed from the genomic sequences from the orangutan and human gene size is calculated to be more than 50 kb.

RESULTS

Genomic Organization of the Rhesus and Human 1,3GT Gene—The full-length genomic organization determined by the combined results of GenomeWalker-PCR, RACE, and RT-PCR consisted of nine exons and eight introns (Fig. 1), corresponding to the previously delineated porcine and mouse genomic structure (8-14). However, some cryptic exons or aberrant splicing sites were observed as described below. The total rhesus and human gene size is calculated to be more than 50 kb.

Cryptic Exons—Cryptic exons were found in both the rhesus and human 1,3GT gene, e.g. between exons 2 and 4, designated 3H in human and 3R in rhesus (Fig. 2A, a and c). Exons 3H and 3R showed no homology with each other, and neither exon had any homology with the previously described mouse exon 3 (13). A cryptic human exon with two size variants, both of which had poly(A) tails, was found in 18 of 22 humans between exons 7 and 8 (Fig. 2B, a and d). At the end of the 7Ha variant, a donor splicing site was found that spliced to exon 9 (Fig. 2B, a). In rhesus, cryptic exons designated 6Ra and 6Rb were found between exons 6 and 7, and in addition, a third variant was found (6Rc) that included a nucleotide deletion of G was found only in rhesus at the border of the 7Ha variant, a donor splicing site was found that spliced to exon 9 (Fig. 2B, a). In rhesus, cryptic exons designated 6Ra and 6Rb were found between exons 6 and 7, and in addition, a third variant was found (6Rc) that included a retained intron (Fig. 2B, c).

Aberrant Splicing Sites—In the 5'-untranslated region, rhesus exon 2 is 112 bp, whereas human exon 2 is 269 bp (Fig. 3). Comparative analysis revealed that the first half (110 bp, termed 2A) of the human exon 2 is highly homologous to rhesus exon 2, and the last half (159 bp, termed 2B) is similar to the rhesus exon 2 that extends from exon 2. Although the first half (138 bp, termed 8Ra) of the rhesus exon 8 (327 bp) showed high homology with the porcine exon 8 (138 bp), the remaining half (189 bp, termed 8Rb) resembled the porcine intron 8 sequences extending from exon 8 (Figs. 2B, c, and d). A single nucleotide deletion of G was found only in rhesus at the border between the region 8Ra and 8Rb (Fig. 3, position [d]), resulting in utilization of another downstream splicing site.

Alternative mRNA Splicing of the Human and Rhesus 1,3GT Gene—A variety of alternative splicing patterns were demonstrated in the RACE and RT-PCR experiments (Fig. 2A and B). In the coding region, the predominant pattern in human consisted of exons 4 through 9. Because of the cryptic exons and aberrant splicing sites in the human and rhesus genes, the alternative splicing patterns were far more varied in these αGal-negative species than in αGal-positive murine or porcine species (8-13).

Exon-Intron Splicing Juncions of Human, Orangutan, and Rhesus 1,3GT Gene—A comparative sequence analysis of splice junctions was performed on rhesus, orangutan, and human genomic DNA (Fig. 3). The results were further compared with the porcine species sequence (GenBank™ accession numbers AF221510–AF221517) (not shown), revealing high homol-
Primate α1,3-Galactosyltransferase Gene

FIG. 1. Genomic organization on a kb scale of the α1,3GT gene of the human (above horizontal line) and rhesus (below horizontal line). The numbered boxes indicate the approximate location of exons within the locus of the two species but do not accurately reflect the length of each exon. The translation start and stop codons are located in exons 4 and 8. The horizontal thick bar above exon 9 denotes the GenBank™ data for human (J05421, 794 bp) contributed by Larsen et al. (5), and the short thick bar below exon 9 denotes the data for rhesus (M73306, 371 bp) contributed by Galili and Swanson (6). Shown below the gene locus representation is a contig of overlapping subgenomic clones derived from GenomeWalker™ libraries (horizontal arrows) or from long PCR experiments (horizontal lines).

a. Human RT-PCR
   1 M
   1638
   1557
   1370

b. Orangutan RT-PCR
   1 M
   704

c. Rhesus RT-PCR
   Human pattern A
   1 2 3H 4 5 6 7 8 9
   1 2 3 4 5 6 7 8 9
   1 2 3 4 5 6 7 8 9
   a, Human pattern A pA
   b. Orangutan pA
   c. Rhesus pA
   d. Rhesus (2) v.s. Bovine (1)

FIG. 2. A, electrophoresis of PCR (a–c) and Northern blot hybridization (d). Lane 1, electrophoresis results of RT-PCR in the human, orangutan, and rhesus cDNA libraries, respectively, using a pair of primers spanning the coding region. Lane M, DNA size marker ladder (Invitrogen, 1-kb DNA size ladder). d, Northern blot of total RNA from bovine (lane 1) or rhesus spleen tissue (lane 2) hybridized with a probe derived from rhesus exon 9 fragment. As expected, with in-species versus cross-species hybridization, gene expression was lower in the bovine than in the monkey spleen cells. B, schematic splicing variations and cryptic exons of human (a), orangutan (b), and rhesus (c) α1,3GT gene. The closed boxes denote exons in the respective species. The open boxes with arrows denote cryptic exons that were not observed in the αGal-positive species. Exon 6Rc in rhesus has exon 6, intron 6, and exon 7 (a retained intron).

Regulatory Region and 5′-Untranslated Region of the Rhesus α1,3GT Gene—GenomeWalker-PCR starting at exon 1 revealed that the rhesus and human upstream regulatory region contains several SP1/GC-box, AP2, and other consensus motifs in the absence of a TATA- or CAAT-box. The findings were
similar to those in the pig gene (12). In both the rhesus and human, as well as the pig, the region comprises a CpG island. In the upstream region between -340 and -470, the high degree of homology between the rhesus and pig, but not between the human and pig, is noteworthy (Fig. 4). The transcription initiation site has been identified in the pig (12).

Whether there are differences in gene regulation between the human, as well as the pig, the region comprises a similar to those in the pig gene (12). In both the rhesus and human, the region comprises a CpG island. In the upstream region between -340 and -470, the high degree of homology between the rhesus and pig, but not between the human and pig, is noteworthy (Fig. 4). The transcription initiation site has been identified in the pig (12). When the transcription initiation site is identified in the human and rhesus gene, the question may be clarified as to whether there are differences in gene regulation between the human and rhesus.

Rhesus α1,3GT Gene Promoter Activity—The intensity of expression promoted by the pGL3-RA construct was 18-fold greater than the pGL3-Basic control, and that of pig 81-fold higher (Fig. 5).

3'-Untranslated Region of the Rhesus α1,3GT Gene—The nucleotide divergence of the rhesus exon 9 sequence from that of the orangutan and human was found starting 1273 bp downstream from the 3'-end of exon 9 and spanning 420 bp (data not shown). A GenBank™ BLAST search revealed that this divergence had a striking homology with a retrovirus long terminal repeat sequence present in the baboon (GenBank™ accession number NC-000865; Ref. 17).

Homology in the Coding Region—We compared our results of the human, orangutan, and rhesus sequences to our αGal-positive marmoset and cebus New World monkey sequences (deposited in the GenBank™). Our marmoset sequence also was compared with an independently submitted sequence (18). Although the overall sequence from the start codon in exon 4 to the stop codon in exon 9 in marmoset is 1131 bp, the corresponding sequence of the human (from exon 4 to exon 9 through exon 8) or rhesus (from exon 4 to exon 9 through exon 8a) was 1128 or 1129 bp, respectively. Homology between human, orangutan, or rhesus to marmoset is 94.3, 94.7, or 92.4%, respectively.

Comparison of Predicted Amino Acid Sequences—There were no nonframeshift mutations in the rhesus, orangutan, or human exons 4–6. However, five crucial mutations that might account for loss of the α1,3GT expression were identified in the human and rhesus, and three were detected in the orangutan. It is noteworthy that the mutations at the position [b] and [f] were shared by all three species.

In the rhesus, at position [a] and [b] (Fig. 7), corresponding to the deduced amino acid positions 75 and 81, respectively, in marmoset, a single nucleotide deletion resulted in a frameshift. At position 96 (position [c]), a single nucleotide substitution resulted in aframeshift. As described under "Aber rant Splicing Sites," a single nucleotide deletion was detected in the orangutan, or rhesus to marmoset is 94.3, 94.7, or 92.4%, respectively.

Comparison of Predicted Amino Acid Sequences—There were no nonframeshift mutations in the rhesus, orangutan, or human exons 4–6. However, five crucial mutations that might account for loss of the α1,3GT expression were identified in the human and rhesus, and three were detected in the orangutan. It is noteworthy that the mutations at the position [b] and [f] were shared by all three species.

In the rhesus, at position [a] and [b] (Fig. 7), corresponding to the deduced amino acid positions 75 and 81, respectively, in marmoset, a single nucleotide deletion resulted in a frameshift. At position 96 (position [c]), a single nucleotide substitution resulted in aframeshift. As described under "Aber rant Splicing Sites," a single nucleotide deletion was detected in the orangutan, or rhesus to marmoset is 94.3, 94.7, or 92.4%, respectively.

Comparison of Predicted Amino Acid Sequences—There were no nonframeshift mutations in the rhesus, orangutan, or human exons 4–6. However, five crucial mutations that might account for loss of the α1,3GT expression were identified in the human and rhesus, and three were detected in the orangutan. It is noteworthy that the mutations at the position [b] and [f] were shared by all three species.

In the rhesus, at position [a] and [b] (Fig. 7), corresponding to the deduced amino acid positions 75 and 81, respectively, in marmoset, a single nucleotide deletion resulted in a frameshift. At position 96 (position [c]), a single nucleotide substitution resulted in aframeshift. As described under "Aber rant Splicing Sites," a single nucleotide deletion was detected in the orangutan, or rhesus to marmoset is 94.3, 94.7, or 92.4%, respectively.

Comparison of Predicted Amino Acid Sequences—There were no nonframeshift mutations in the rhesus, orangutan, or human exons 4–6. However, five crucial mutations that might account for loss of the α1,3GT expression were identified in the human and rhesus, and three were detected in the orangutan. It is noteworthy that the mutations at the position [b] and [f] were shared by all three species.

In the rhesus, at position [a] and [b] (Fig. 7), corresponding to the deduced amino acid positions 75 and 81, respectively, in marmoset, a single nucleotide deletion resulted in a frameshift. At position 96 (position [c]), a single nucleotide substitution resulted in aframeshift. As described under "Aber rant Splicing Sites," a single nucleotide deletion was detected in the orangutan, or rhesus to marmoset is 94.3, 94.7, or 92.4%, respectively.

Comparison of Predicted Amino Acid Sequences—There were no nonframeshift mutations in the rhesus, orangutan, or human exons 4–6. However, five crucial mutations that might account for loss of the α1,3GT expression were identified in the human and rhesus, and three were detected in the orangutan. It is noteworthy that the mutations at the position [b] and [f] were shared by all three species.

In the rhesus, at position [a] and [b] (Fig. 7), corresponding to the deduced amino acid positions 75 and 81, respectively, in marmoset, a single nucleotide deletion resulted in a frameshift. At position 96 (position [c]), a single nucleotide substitution resulted in aframeshift. As described under "Aber rant Splicing Sites," a single nucleotide deletion was detected in the orangutan, or rhesus to marmoset is 94.3, 94.7, or 92.4%, respectively.

Comparison of Predicted Amino Acid Sequences—There were no nonframeshift mutations in the rhesus, orangutan, or human exons 4–6. However, five crucial mutations that might account for loss of the α1,3GT expression were identified in the human and rhesus, and three were detected in the orangutan. It is noteworthy that the mutations at the position [b] and [f] were shared by all three species.

In the rhesus, at position [a] and [b] (Fig. 7), corresponding to the deduced amino acid positions 75 and 81, respectively, in marmoset, a single nucleotide deletion resulted in a frameshift. At position 96 (position [c]), a single nucleotide substitution resulted in aframeshift. As described under "Aber rant Splicing Sites," a single nucleotide deletion was detected in the orangutan, or rhesus to marmoset is 94.3, 94.7, or 92.4%, respectively.

Comparison of Predicted Amino Acid Sequences—There were no nonframeshift mutations in the rhesus, orangutan, or human exons 4–6. However, five crucial mutations that might account for loss of the α1,3GT expression were identified in the human and rhesus, and three were detected in the orangutan. It is noteworthy that the mutations at the position [b] and [f] were shared by all three species.

In the rhesus, at position [a] and [b] (Fig. 7), corresponding to the deduced amino acid positions 75 and 81, respectively, in marmoset, a single nucleotide deletion resulted in a frameshift. At position 96 (position [c]), a single nucleotide substitution resulted in aframeshift. As described under "Aber rant Splicing Sites," a single nucleotide deletion was detected in the orangutan, or rhesus to marmoset is 94.3, 94.7, or 92.4%, respectively.
Primate α1,3-Galactosyltransferase Gene

**Fig. 4. Exon 1 vicinity of α1,3GT gene of rhesus (R) compared with pig (P).** The rhesus exon 1 and its potential regulatory region (capital letters) are aligned with the corresponding porcine exon 1 sequence (12). The lowercase letters denote intron 1. An asterisk indicates that the nucleotide is identical between both species.

**DISCUSSION**

The αGal epitopes are chemically related to the ABO blood type antigens. In the absence of expression of A or B epitopes or of αGal epitopes, antibodies develop against the respective antigens. Phylogenetic data on the emergence of the ABO antigen system (19) and especially molecular studies on ABO transferase (20) have been valuable in determining higher primate evolution. In contrast, little comparable information is available for the evolution of the ABO-like αGal epitopes in nonhuman primates.

A single nucleotide deletion at position 48 was observed in both human and orangutan but not in rhesus. The premature stop codon at position 268 (C) was observed only in human. Thus, five crucial mutations that might account for loss of α1,3GT expression were determined in human and rhesus and three in orangutan. It is noteworthy that the mutations at position [b] and [f] were shared by the rhesus monkey, ape, and human.

**Fig. 5. Luciferase assay for promoter activity of the rhesus α1,3GT gene.**

At position 96 (position [c]), a single nucleotide substitution resulted in a premature stop codon (from TGG (Trp) to TGA). At position 268 (C), a single nucleotide substitution resulted in an additional premature stop codon (TAT (Tyr) to TAG).

81, respectively, in marmoset, a single nucleotide deletion, G, from the triplet GAA (Glu) was identified, resulting in a frameshift. At position 96 (position [c]), a single nucleotide substitution resulted in a premature stop codon (from TGG (Trp) to TGA). At position 268 (C), a single nucleotide substitution resulted in an additional premature stop codon (TAT (Tyr) to TAG). A single nucleotide deletion at position 48 was observed in both human and orangutan but not in rhesus. The premature stop codon at position 268 (C) was observed only in human. Thus, five crucial mutations that might account for loss of α1,3GT expression were determined in human and rhesus and three in orangutan. It is noteworthy that the mutations at position [b] and [f] were shared by the rhesus monkey, ape, and human.
known about the molecular basis of inactivation of the gene expressing the enzyme α1,3GT upon which synthesis of the αGal epitope depends.

The αGal system in lower mammals was not discovered until the 1980s (1, 2). Evidence of the presence of α1,3GT gene homologues in some αGal-negative higher primates was pub-
lished in 1989–1991 by three pioneer groups. Joziasse et al. (7) described an 89-bp sequence (HGT-10) in human cells that was similar to a fragment of the murine and bovine exon 4. Larsen et al. (5) characterized a 703-bp genomic sequence in human tissue with a high degree of homology to exon 9 of the mouse α1,3GT gene. Galili and Swanson (6) reported a 371-bp sequence in rhesus cells that resembled part of mouse and bovine exon 9.

All three groups concluded that point mutations in exon 9 at some time in the evolutionary past of the higher primates could be the cause of the loss of αGal epitope expression. We confirmed the point mutations that had been described earlier in exon 9 of the rhesus (6), orangutan (6), and human α1,3GT gene (5, 6), and demonstrated four additional point mutations in exons 7 and 8. The crucial mutations in the different species that result in frameshift and premature termination of translation include a deletion mutation at position [b] in exon 7 and a premature termination at position [f] in exon 9 that are shared by all three of these higher primates.

Because no α1,3GT transcripts could be found in the early investigations of the African green monkey (COS) and human cell lines (4), the possibility could not be dismissed that inactivation of the gene in the higher primates is due to mutations in the upstream regulatory regions of the α1,3GT gene (16). This potential explanation for loss of the αGal epitope was eliminated in our study using sensitive PCR methods for detection of mRNA transcripts that were not available at the time of the seminal investigations (4–7). A variety of transcripts with more than 10 different splicing patterns of the α1,3GT gene were observed in the full coding region in all rhesus, orangutan, and human tissues examined. Their presence was confirmed with Northern blot of total RNA derived from rhesus spleen.

Joziasse et al. (7, 16) estimated that inactivation of the α1,3GT gene occurred between 40 and 25 millions years ago, following the divergence between higher primates (carrhines) and the New World monkey but prior to the further divergence of the individual higher primates. The hypothesis was based on their discovery of an intron-less processed cDNA homologue (termed HGT-2) in the human genome (7, 16). In contrast, Galili and Swanson (6) postulated that inactivation occurred at a much later time and independently in Old World monkeys, apes, and humans (6). Our finding of crucial point mutations shared by the rhesus, orangutan, and human at positions [b] and [f] of exons 7 and 9, respectively, appears to be more consistent with Joziasse's hypothesis (16) that the α1,3GT gene inactivation occurred in a common ancestor than with the suggestion by Galili and Swanson (6) that it occurred independently in the three different lineages. However, the process and the timing of the inactivation remain to be determined (6, 7). The high homology of the α1,3GT gene in the αGal-negative species to the shared sequence of the αGal-positive marmoset and cebus (Fig. 7) suggests that the inactivation may have occurred more recently than previously thought.

The development of preformed antibodies directed against a wide range of microorganisms or macroparasites that express αGal-like substance has been the conventional explanation for the evolution of the αGal-negative state in higher primates (1, 2). The possibility cannot yet be excluded, however, that some other survival advantage drove the gene inactivation and that the protection from such infecting agents afforded by the "natural" anti-αGal antibodies was fortuitous. Whatever the reason, antibodies against the αGal epitope are responsible for the immediate (hyperacute) rejection by αGal-negative recipients of tissues and organs from αGal-positive donors, precluding successful clinical xenotransplantation from αGal-positive animals (3). Detailed information about the α1,3GT gene may help in mapping strategies for transgenic modification of the αGal-positive species.

If higher primates were to spontaneously recover the expression of αGal epitopes, as has been described in normal breast and MCF7 human breast carcinoma cell lines (21), the potential for the development of αGal antibody-mediated autoimmune disease is implicit. Galili and co-workers (22) observed the presence of αGal epitopes in thyroid cells of a human with autoimmune Grave's disease, but no transcripts could be detected in these cells with Northern blot. A search for α1,3GT mRNA transcripts in tissues from such "autoimmune suspect" patients may be fruitful with more sensitive PCR technology used in our study.

Acknowledgments—We thank Rickquel Tripp for technical assistance and Terry Mangan for manuscript preparation.

REFERENCES