Identification of a novel A4GALT exon reveals the genetic basis of the P1/P2 histo-blood groups

Britt Thuresson,1,2 Julia S. Westman,2 and Martin L. Olsson1,2

1Department of Clinical Immunology and Transfusion Medicine, University and Regional Laboratories, Lund, Sweden; and 2Division of Hematology and Transfusion Medicine, Department of Laboratory Medicine, Lund University, Lund, Sweden

The A4GALT locus encodes a glycosyltransferase that synthesizes the terminal Galα1-4Gal of the Pk (Gb3/CD77) glycosphingolipid, important in transfusion medicine, obstetrics, and pathogen susceptibility. Critical nucleotide changes in A4GALT not only abolish Pk formation but also another Galα1-4Gal-defined antigen, P1, which belongs to the only blood group system for which the responsible locus remains undefined. Since known A4GALT polymorphisms do not explain the P1−Pk+ phenotype, P2, we set out to elucidate the genetic basis of P1/P2. Despite marked differences (P1 > P2) in A4GALT transcript levels in blood, luciferase experiments showed no difference between P1/P2-related promoter sequences. Investigation of A4GALT mRNA in cultured human bone marrow cells revealed novel transcripts containing only the noncoding exon 1 and a sequence (here termed exon 2a) from intron 1. These 5′-capped transcripts include poly-A tails and 3 polymorphic sites, one of which was P1/P2-specific among > 200 donors and opens a short reading frame in P2 alleles. We exploited these data to devise the first genotyping assays to predict P1 status. P1/P2 genotypes correlated with both transcript levels and P1/Pk expression on red cells. Thus, P1 zygozyosity partially explains the well-known interindividual variation in P1 strength. Future investigations need to focus on regulatory mechanisms underling P1 synthesis. (Blood. 2011; 117(2):678-687)
Ceramide is abbreviated to Cer. Names of the involved glycosyltransferases are given, and genes known to underlie expression of blood group antigens are given in parentheses. Blood group antigens are written in bold, as are the enzyme activities involved in P1 and Pk synthesis, the 2 antigens most important for this study (also highlighted by black frames and thicker arrows).

The blood group system is still unknown but strikingly, the P1 antigen is lacking in individuals with the p phenotype. This implies a common enzymatic background for Pk and P1. However, RBCs of the relatively common P1-negative (P2) phenotype type positive for Pk, which shows that the explanation is not straightforward, and this puzzle has remained unresolved for decades. In addition, the well-known phenomenon of variable P1 expression on RBCs has been suggested to depend on zygosity for the P1 trait, but this theory awaits confirmation. Until now, a polymorphic genetic symbol used follow the recommendations of Varki et al;50 ie, glucose (•), N-acetylglucosamine (□), galactose (○), N-acetylgalactosamine (□), fucose (△), and sialic acid (▲).

Figure 1. Scheme representing the biosynthesis of Pk, P and P1 antigens and some other related structures, such as the blood group H, A, and B antigens. The symbols used follow the recommendations of Varki et al.154 is, glucose (•), N-acetylglucosamine (□), galactose (○), N-acetylgalactosamine (□), fucose (△), and sialic acid (▲). Ceramide is abbreviated to Cer. Names of the involved glycosyltransferases are given, and genes known to underlie expression of blood group antigens are given in parentheses. Blood group antigens are written in bold, as are the enzyme activities involved in P1 and Pk synthesis, the 2 antigens most important for this study (also highlighted by black frames and thicker arrows).

Table 1. Blood group antigens, carbohydrate structures, and null phenotypes with relevance for this study

<table>
<thead>
<tr>
<th>Blood group antigen</th>
<th>Alternative name(s)</th>
<th>Carbohydrate structure</th>
<th>Null phenotype</th>
</tr>
</thead>
<tbody>
<tr>
<td>P</td>
<td>CDH, Glb, LacCer</td>
<td>Galβ1→4Glcβ1→1-Cer</td>
<td></td>
</tr>
<tr>
<td>Pk</td>
<td>CTH, Glb, CD77</td>
<td>Galα1→4Galβ1→1-Cer</td>
<td>p</td>
</tr>
<tr>
<td>P</td>
<td>Globose, Glb</td>
<td>GalNAα1→3Galβ1→4Galβ1→1-Cer</td>
<td>P1,3,P2,4</td>
</tr>
<tr>
<td>P</td>
<td>Paragloboside, nLc</td>
<td>Galβ1→4GlcNAcβ1→3Galβ1→4Glcβ1→1-Cer</td>
<td></td>
</tr>
<tr>
<td>P1</td>
<td>nLc</td>
<td>Galα1→4Galβ1→4GlcNAcβ1→3Galβ1→4Glcβ1→1-Cer</td>
<td>P2</td>
</tr>
</tbody>
</table>

— indicates that this structure does not correlate to a blood group antigen or null phenotype.

Methods

Samples and nucleic acid preparation

Bone marrow (n = 3) and blood (n = 205) from apparently healthy donors were obtained following informed consent in accordance with the Declaration of Helsinki. Approvals from The Regional Ethics Review Board at Lund University were obtained for bone marrow collection and genetic blood group analysis on blood samples from blood donors. RBCs were collected and stored in the CellStab low-ionic strength preservative solution.
**Table 2. Oligonucleotide primers used in this study**

<table>
<thead>
<tr>
<th>Primer</th>
<th>Sequence (5' → 3')</th>
<th>Application</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pk 150R</td>
<td>CTGGCAGGTTATAGGCTGC</td>
<td></td>
</tr>
<tr>
<td>Pk 47R</td>
<td>GTGCAAGACCACCCTGCTGTGT</td>
<td></td>
</tr>
<tr>
<td>Pk 2a-90R</td>
<td>TCTCGAATCTGACTTTACAA</td>
<td></td>
</tr>
<tr>
<td>Pk 2a-R-con</td>
<td>CAAGTGTGCGAACTACGACG</td>
<td></td>
</tr>
<tr>
<td>Pk ex1-5F</td>
<td>CCGGGCAGCGCCGTGTC</td>
<td>3' RACE</td>
</tr>
<tr>
<td>Pk 69F</td>
<td>CACCGGTCTGACACTGG</td>
<td>3' RACE</td>
</tr>
<tr>
<td>Pk 110-F</td>
<td>TATGCTACTGCAAGTCAGTGTGT</td>
<td>3' RACE</td>
</tr>
<tr>
<td>Pk –1070F</td>
<td>TATTAACCGCGTACGCACTCTGAGTGGGAGATGAC</td>
<td>Luciferase</td>
</tr>
<tr>
<td>Pk –577F</td>
<td>TATTAACCGCGTCTCTATACCTGAGAAATACCCA</td>
<td>Luciferase</td>
</tr>
<tr>
<td>Pk –382F</td>
<td>TATTAACCGCGTCCCCAGGGGTAGACACGTC</td>
<td>Luciferase</td>
</tr>
<tr>
<td>Pk –316F</td>
<td>TATTAACGCAAGGTTTCTTTTTCATCGATCCAGT</td>
<td>Luciferase</td>
</tr>
<tr>
<td>Pk –240F</td>
<td>TATTAACGCGTGAAGGCAGTGTCCACGC</td>
<td>Luciferase</td>
</tr>
<tr>
<td>Pk –136F</td>
<td>TATTAACGCGTACCCCGCAGGGTAGTGCTG</td>
<td>Luciferase</td>
</tr>
<tr>
<td>Pk –6R</td>
<td>TATTAACATCTCTAGCTGTTCCT</td>
<td>Luciferase</td>
</tr>
<tr>
<td>Pk P1P2-F</td>
<td>TGGTAAATCTTACCGAAGAACGAGTAGG</td>
<td>PCR-ASP</td>
</tr>
<tr>
<td>Pk P1m-R*</td>
<td>CCAAAGTGCTGGAATGACAGCA</td>
<td>PCR-ASP</td>
</tr>
<tr>
<td>Pk P2m-R*</td>
<td>CCAAATGCGAGGTCTACAGCAGA</td>
<td>PCR-ASP</td>
</tr>
<tr>
<td>JK ASP-CF</td>
<td>GACGTGCCATAGATGTGTC</td>
<td>PCR-ASP</td>
</tr>
<tr>
<td>JK ASP-CR</td>
<td>GAGCGAGGAGTGGTTGCC</td>
<td>PCR-ASP</td>
</tr>
<tr>
<td>Pk II 2648R</td>
<td>ATGACATAGCAGAGAGCCAAGA</td>
<td>1 seq</td>
</tr>
<tr>
<td>Pk II 2145F</td>
<td>TGAATTAACGCGAAGAACGAGTAGG</td>
<td>1 seq, PCR-RFLP</td>
</tr>
<tr>
<td>Pk 2a-240-R</td>
<td>CTCCCGCCAGCTGCAGGTCT</td>
<td>PCR-RFLP</td>
</tr>
</tbody>
</table>

*Mutations were introduced in the –3 position in both reverse primers to increase allele specificity.

(DiaMed AG) for phenotyping and flow cytometry. DNA was isolated using QIAamp DNA Blood Mini kit (QIAGEN GmbH). RNA was extracted from Buffy coat using Trizol reagent (Invitrogen). RNA from cultured cells was extracted using QIAshredder and RNeasy Mini kit (QIAGEN). RNA was extracted from cultured cells using TRIzol reagent (Invitrogen). RNA from cultured cells was extracted using QIAamp DNA Blood Mini kit (QIAGEN GmbH). RNA was extracted from cultured cells using QIAamp DNA Blood Mini kit (QIAGEN GmbH).

**Real-time PCR and data analysis**

Quantitative polymerase chain reaction (PCR) was performed on 3 μL of cDNA with TaqMan probes and the 7500 Real-Time PCR System (Applied Biosystems), according to the manufacturer’s instructions. Data were analyzed using Sequence Detection software Version 1.3.1 (Applied Biosystems). Enzyme-coding A4GALT transcripts were detected with a TaqMan Gene Expression Assay (Hs00213726_m1; Applied Biosystems), binding to exon 2-3 boundary. Transcript target quantifications were normalized to 18S ribosomal RNA (assay Hs99999901_s1). All samples were run in triplicate. The sample with the lowest cycle threshold value was used as a calibrator. We considered as positive the results from any sample with at least 2 detected (cycle threshold > 40) values within the triplicate.

**RACE**

Messenger RNA was isolated from total RNA extracted from bone marrow cells cultured toward erythropoietic maturation as described previously by an mRNA Isolation kit (Roche Diagnostics). Rapid amplification of cDNA ends (RACE) was performed with the FirstChoice RLM-RACE kit (Ambion) according to the manufacturer’s recommendations. In the 5’-RACE, cDNA was synthesized with random primers provided with the kit. Gene-specific primers Pk-150R and Pk-47R were used for PCR amplification together with the 5’-RACE primers provided in the kit. Primers Pk 2a-90R and Pk-2a-R-con were used to define the 5’ end of the transcripts including the new exon 2a. For the 3’-RACE, PCR was performed with primers Pk-ex1-5F, Pk 69F, and Pk-110-F, together with the 3’-primers included in the kit. Primer sequences are shown in Table 2.

**P1/P2 phenotyping**

All samples were phenotyped for the P1 antigen using commercially available anti-P1 reagents according to routine blood banking procedures. The anti-P1 used varied over time (because this was done as part of routine practice), but all were Conformité Européenne (CE)-labeled reagents approved for clinical use on the European market.

A 15-donor cohort of samples was investigated with 3 different antisera (Table 3). Agglutinates were scored visually, and reaction strength was assigned as negative or positive, from weak (−) to the strongest (+4) according to immunohematologic practice. RBCs of known P1/P2 phenotypes were used as controls.

**Sequencing**

The novel exon 2a was amplified and sequenced. The buffered amplification contained 2 μmol of each dNTP (Applied Biosystems), 4 pmol forward primer Pk II 2145F, 4 pmol reverse primer Pk II 2648R (Table 2), 100 ng DNA, and 0.5 U Taq Gold Polymerase (Applied Biosystems). Reactions were executed in the GeneAmp PCR System 2700.

PCR conditions were 96°C for 7 minutes, then 30 cycles of 94°C for 30 seconds, 62°C for 30 seconds, 72°C for 20 seconds. The same primers

**Table 3. Antibodies used in this study**

<table>
<thead>
<tr>
<th>Specificity (name)</th>
<th>Clonality</th>
<th>Class</th>
<th>Species</th>
<th>Clone</th>
<th>Manufacturer</th>
<th>Application (fluorescence)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Anti-P1 (ImmuClone)</td>
<td>Monoclonal</td>
<td>IgM</td>
<td>Human</td>
<td>P9NIL100</td>
<td>Immucor</td>
<td>Hemagglutination, FCM/primary</td>
</tr>
<tr>
<td>Anti-P1 (Seraclon)</td>
<td>Monoclonal</td>
<td>IgM</td>
<td>Mouse</td>
<td>650</td>
<td>Biotest</td>
<td>Hemagglutination</td>
</tr>
<tr>
<td>Anti-P1</td>
<td>Polyclonal</td>
<td>IgM</td>
<td>Goat</td>
<td>—</td>
<td>Immucor</td>
<td>Hemagglutination</td>
</tr>
<tr>
<td>Anti-CD77</td>
<td>Polyclonal</td>
<td>IgM</td>
<td>Rat</td>
<td>38-13</td>
<td>Immunotech</td>
<td>FCM/primary</td>
</tr>
<tr>
<td>Anti-human IgM</td>
<td>Polyclonal</td>
<td>F(ab’)2</td>
<td>Rabbit</td>
<td>—</td>
<td>DakoCytometry</td>
<td>FCM/secondary (FITC)</td>
</tr>
<tr>
<td>Anti-rat IgM</td>
<td>Polyclonal</td>
<td>F(ab’)2</td>
<td>Goat</td>
<td>—</td>
<td>Beckman Coulter</td>
<td>FCM/secondary (PE)</td>
</tr>
</tbody>
</table>

IgM indicates immunoglobulin M; FCM, flow cytometry; FITC, fluorescein isothiocyanate; PE, phycoerythrin; and —, not applicable.
were used for sequencing. All amplification products were separated by high-voltage electrophoresis on 3% agarose gels (SeaKem; FMC Bioproducts) stained with ethidium bromide (0.56 mg/L gel; Sigma-Aldrich). Products were purified using the QIAquick gel extraction kit (QIAGEN), sequenced with the BigDye Terminator kit v1.1 (Applied Biosystems), and analyzed on a 3130 Avant/Genetic Analyzer (Applied Biosystems).

**P1/P2 genotyping methods**

**PCR-ASP.** A PCR-ASP (allele-specific primer) assay was designed for P1/P2 genotyping. Buffered amplification mixture with a total volume of 11 μL included 2 nmol of each dNTP, 4 pmol PkP1P2-F, 4 pmol PkP1m-R or PkP2m-R, 0.8 pmol each forward and reverse control primers JK-ASP-CF and JK-ASP-CR (Table 2), 100 ng DNA, and 0.5 U Taq Gold Polymerase. Amplification was performed as above with addition of an elongation step for 1 minute at 72°C. Amplicons were visualized on 3% agarose gels.

**PCR-RFLP.** Primers Pk2145F and Pk2a-240-R were used to amplify a PCR fragment using the above conditions. The amplified fragment was digested with NalIII (New England Biolabs) for 1 hour at 37°C and digestion products separated and stained on 4% agarose gels.

**Allelic discrimination.** A custom-made TaqMan SNP Genotyping Assay (Applied Biosystems) was run according to the manufacturer’s instructions on the 7500 Sequence Detection System (Applied Biosystems).

**Luciferase assay**

Constructs for the luciferase assay were made from 2 promoter variants containing 5 polymorphic sites. PCR fragments of 6 different sizes from each variant were amplified with primers listed in Table 2. PCR conditions: 200 ng genomic DNA were mixed in a final volume of 25 μL containing 3 mM MgCl₂, 0.4 mM dNTP, 1 × guanine-cytosine (GC)-rich buffer with DMSO (dimethyl sulfoxide), 0.5M GC-rich resolution solution, 0.4 μM each forward primer and Pk+6R for all fragments, and 2 μM GC-rich enzyme (Roche). Thermal cycling was undertaken in the GeneAmp PCR system 2700: initial denaturation at 96°C for 3 minutes was followed by 10 cycles at 94°C for 15 seconds, 63°C for 30 seconds, 68°C for 2 minutes, then 25 cycles at 94°C for 15 seconds, 60°C for 30 seconds, and 68°C for 2 minutes. Amplicons were separated and eluted as above. The fragments were digested with MluI/HindIII and cloned into the pGL3 basic vector (Clontech). pGL3 promoter vector (Clontech) was used as positive control and pGL3 basic vector (Clontech) as negative control. Constructs were introduced into 8 × 10⁵ Ramos cells by electroporation using a Gene Pulser (Bio-Rad Laboratories) with electrical settings of 320 V/960 μF. After incubation for 16 hours at 37°C, luminescence was measured using the Dual-Luciferase Reporter Assay System (Promega) and a Glomax 20/20 luminometer (Promega) according to the manufacturer instructions. Values of Firefly luciferase were normalized to the values of Renilla luciferase, which was used as an internal control of transfection efficiency.

**Flow cytometry**

Washed RBCs (approximately 0.5 × 10⁶) were suspended in 50 μL of phosphate-buffered saline (PBS) in 96-well plates (NUNC) and fixed with 0.07% gluteraldehyde for 10 minutes at room temperature (RT) to prevent agglutination. After incubation, the plate was centrifuged for 1 minute at 350 × g, the supernatant discarded, and RBCs resuspended in PBS. The primary antibody was incubated with the RBCs for 10 minutes at RT, followed by 35 minutes at 4°C. RBCs were washed twice with PBS and incubated with secondary antibody for 10 minutes at RT. The antibodies used are described in Table 3. All incubation steps were performed in the dark on a rotary mixer. Data were collected with a calibrated FACSscan flow cytometer (BD Biosciences) and analyzed using Cell Quest software v3.1f (BD Biosciences). PP1Pk-negative (p phenotype) RBCs were used as negative control and P1/P₁ RBCs as positive control.

**Statistical analysis**

Independent 2-sample t test assuming equal variance and 2-tailed distribution was used to determine the significance. The genotype groups (P1/P1, P1/P₂, or P2/P2) were compared with each other using the XLSTAT 2009 (Addinsoft) data analyzer. Data were considered statistically significant with respect to the following criteria: *P < .05, **P < .01, ***P < .001.

**Results**

**Semi-quantification of A4GALT transcripts from P₁ and P₂ phenotypes**

Enzyme-encoding A4GALT transcript levels were initially measured in 10 random blood samples with (P₁) or without (P₂) the P1 antigen on RBCs. As shown in Figure 2A, the A4GALT transcript levels were approximately 30 times higher in the 5 P₁ + samples compared with the P₁– samples (Figure 2B is discussed below). A substantial variation in transcript levels among P₁ samples was noted, however. To investigate whether the marked difference in transcript levels between P₁ and P₂ individuals is due to previously described variations in the 5’-upstream sequence, a functional study of the proposed promoter region was undertaken.

**Qualitative characterization of A4GALT transcripts**

Before creating constructs for the Luciferase assay, 5′/3′-RACE analysis was performed to define the transcription start sites and ends of the A4GALT transcripts. Four different A4GALT transcripts were detected in the Ramos cell line used for these experiments. The originally described transcript with 3 exons (transcript I in

---

**Figure 2. Quantification of A4GALT transcripts.** (A) Transcript levels measured by TaqMan gene expression assay in 5 P₁ and 5 P₂ individuals shows a significant difference. (B) A4GALT transcript levels depend on P₁/P₁ genotype. Gene expression levels in peripheral blood were determined with the TaqMan assay in 15 samples of the genotypes P₁P₁ (n = 5), P₁P₂ (n = 5), and P₂P₂ (n = 5). Target quantities were normalized to 18S ribosomal RNA. Both graphs show the mean values and error bars represent SEM values. The y-axis represents the percentage of the highest value obtained. Significance levels are shown as asterisks above the bars.
Figure 3A) was found with 2 different transcription start sites, in which the length of exon 1 was either 30 or 60 bp. Another transcript (transcript II in Figure 3A) lacked part of exon 2 and most of exon 3, whereas a third transcript (transcript III in Figure 3A) lacked the whole of exon 2 and most of exon 3. These transcripts were sequenced in P1 and P2 individuals, but no polymorphisms were found compared with the genomic consensus sequence. The fourth transcript (designated transcript IV in Figure 3) only consisted of exon 1, a sequence from intron 1 and a poly-A tail, while exons 2 and 3 were missing. This transcript was also found following RACE analysis on cultured human bone marrow cells from donors with different P1/P2 phenotypes and is described in more detail below. Sequences of the identified transcripts and their characteristics were deposited in GenBank (supplemental Figure 1, available on the Blood Web site; see the Supplemental Materials link at the top of the online article).

Two variants of the A4GALT promoter are equally active

The sequence 5′ upstream of exon 1 in A4GALT was found to harbor strong promoter activity (Figure 4). The 2 shortest deletion constructs (136 and 240 bp long) used to demonstrate this gave approximately 50% higher luciferase levels than the positive control including the strong simian virus 40 (SV40) promoter, whereas sequences longer than 240 bp showed slightly lower expression levels compared with the control. Thus, none of 5 polymorphic sites in the promoter appeared to influence expression levels.

Discovery of SNPs in a novel A4GALT transcript

The intronic 289-290-bp sequence spliced to exon 1 in one of the new transcript variants (IV) described above is in fact an alternative exon 2 and was therefore designated exon 2a. The novel exon was sequenced in 95 samples from individuals with the P1 and P2 phenotypes. Three polymorphisms were identified at nucleotide (nt) positions 42C/T, 122T/G, and 135C/delC (where nt 1 is the first residue in exon 2a). Remarkably, nt 42 predicted P1/P2 status, whereas the alterations at positions 122 and 135 showed no allele specificity. All P2 samples were homozygous for 42T; P1 samples were either homozygous for C or heterozygous.

\[ \text{P}^1 {\text{C}}/\text{P}^2 {\text{T}} \]

The C > T substitution at nt 42 introduces a potential start codon in the \( P^2 \), which gives rise to a short hypothetical open reading frame (ORF) of 28 amino acids (Figure 3B). One of the other polymorphic sites (122T/G in exon 2a) tentatively changes the last residue in the potential ORF from Gly28 to Trp, thus resulting in 2 variants of the \( P^2 \)-related ORF. 122G is found in 64% of alleles in the \( P^2P^2 \) genotype group, but none of the alleles among \( P^1P^2 \) individuals.

**P*/P* genotype screening**

To investigate the correlation between exon 2a polymorphism and \( P^1/P^2 \) phenotypes, more samples were genotyped. Three different genotyping methods were designed and evaluated based on the above findings: PCR-ASP, PCR-RFLP, and AD by a SNP genotyping assay. All 3 assays showed specific and easily interpretable typing patterns (Figure 5) compared with sequence data. It was...
concluded that all 3 methods could be used for screening purposes as outlined below.

A total of 208 donor samples, including the ones previously sequenced, were P1/P2-typed by serology and at least one of the newly developed genotyping methods. The results of this screening are summarized in Table 4. Full concordance between phenotype and genotype was observed in 207 samples, and only in one case was a discrepancy noted (see Table 4 for details).

Effects of P1/P2 zygosity on A4GALT, P1, and Pk expression levels

Fresh blood samples from selected donors were collected and genotyped with the PCR-ASP method until 5 of each of the 3 different genotypes P1P1, P1P2, and P2P2 were available for in-depth studies of the correlation between zygosity for P1/P2 and 3 expression parameters.

P1 expression measured by serologic testing. Hemagglutination was performed with 2 monoclonal and one polyclonal anti-P1.

Table 4. Summary of the screening results from the P1/P2 genotyping of 208 blood donors at position 42 in the novel A4GALT exon 2a

<table>
<thead>
<tr>
<th>Phenotype</th>
<th>Polymorphic site</th>
<th>C/C</th>
<th>C/T</th>
<th>T/T</th>
<th>Total (n = 208)</th>
</tr>
</thead>
<tbody>
<tr>
<td>P1</td>
<td></td>
<td>57</td>
<td>93</td>
<td>1</td>
<td>151</td>
</tr>
<tr>
<td>P2</td>
<td></td>
<td>0</td>
<td>0</td>
<td>57</td>
<td>57</td>
</tr>
</tbody>
</table>

*This sample was very weakly positive for P1 antigen by serologic routine methods, but the donor was unavailable for repeat sampling to investigate the discrepant result. However, a flow cytometric semiquantification of P1 antigen on his RBCs was performed and confirmed the serologic typing, although the P1 level was the lowest encountered in the study (data not shown).
All \( P^2/P^2 \) samples were negative with the monoclonal reagents, and the \( P^1/P^1 \) samples were strongly positive, whereas \( P^1/P^2 \) heterozygous samples showed weaker serologic reactions than homozygous samples. The polyclonal antibody gave similar reaction patterns but displayed a broader variation, and in 2 cases, \( P^2/P^2 \) samples actually gave weakly false positive reactions (Figure 6).

**A4GALT transcription levels.** TaqMan analysis of enzyme-encoding A4GALT transcripts (transcript variant 1 in Figure 3A) showed significantly higher levels in the \( P^1/P^1 \) samples (approximately 28-fold \( P^2/P^2 \) level) than both \( P^1/P^2 \) and \( P^2/P^2 \) samples (Figure 2B), thus confirming the initial finding in random samples defined by serology only (Figure 2A). There was also a significant (4-fold) difference in transcript levels between \( P^1/P^2 \) and \( P^2/P^2 \) samples (Figure 2B).

**Antigen expression measured by flow cytometry.** Flow cytometric analysis with monoclonal anti-P1 confirmed the serologic analysis and showed the same pattern as gene expression with significantly lower expression of P1 antigen on RBCs with \( P^1/P^1 \) genotype than on \( P^1/P^2 \) cells. \( P^2/P^2 \) cells appeared to have similar expression as the single example of cells with p phenotype (ie, negative) and showing only background levels of fluorescence (Figure 7). In addition, analysis of \( P^2 \) antigen expression demonstrated that the \( P^2 \) levels on RBCs from \( P^1/P^2 \)-heterozygous donors were lower than \( P^1/P^2 \)-homozygous, although the difference was not statistically significant. The \( P^2/P^2 \)-homozygous cells displayed more \( P^2 \) expression than cells with p phenotype but significantly lower than \( P^1/P^1 \) cells (Figure 7).

## Discussion

The lack of knowledge regarding the genetic basis for the P1 antigen of the P blood group system has prevented DNA-based typing. We have now identified a polymorphic site in the \( P^6 \)-synthase-encoding A4GALT that correlates to the \( P_1/P_2 \) histoblood group phenotypes. This discovery makes it possible for the first time to predict the presence or absence of P1 antigen on RBCs by testing DNA. The findings presented here also definitively link the 2 Gal\( \alpha_4\)Gal-terminating glycosphingolipid antigens, P1 and \( P^6 \), to each other genetically. The 3 long-standing hypotheses about the genetic relationship between the P1 and \( P^6 \) antigens outlined in the introduction can now be evaluated in a new light. As indicated by our discovery of a \( P_1/P_1 \)-predictive SNP in the novel exon 2a of A4GALT and also suggested by previous studies, it is highly likely that the \( \alpha_4\)GalT encoded by A4GALT, synthesizes both antigens. Accordingly, the concept of a closely linked homologous galactosyltransferase gene, which encodes for P1 synthesis only while the A4GALT product makes \( P^6 \), can be ruled out. Thus, it is logical that genetic alterations found previously in exon 3 cause loss of both P1 and \( P^6 \) antigens in the p phenotype. In fact, our data favor the hypothesis of one gene with 3 types of alleles at a single locus: one that makes P1 and \( P^6 \), another that makes \( P^6 \) only, and a third that makes neither. The only difference, of course, is that 26 \( p \) alleles (data from dbRBC) have been identified since the cloning of A4GALT a decade ago. Therefore, it would be expected that both \( P^1 \) and \( P^2 \) alleles (defined by the newly discovered exon 2a SNP) can be inactivated by a range of different crucial nucleotide changes to become \( p \) alleles. This is indeed the case, as evidenced by our recent finding that the 2 main \( p \) alleles in Swedish individuals, 548T>A and 560G>A, are based on the \( P^1 \) and \( P^2 \) allelic backbones, respectively (B.T., J.S.W., Å. Helberg, M.L.O., unpublished data, August 2010).

The third hypothesis proposes that a regulating factor encoded by a gene at a chromosomal location close to A4GALT binds to the \( P^6 \)-synthesizing \( \alpha_4\)GalT. This would modify the enzyme’s acceptor preference from lactosylceramide only (\( P_2 \) phenotype), to lactosylceramide and paragloboside, which would allow synthesis of both P1 and \( P^6 \) (Figure 1). This model has an appealing analogy in the 4-\( \beta \)-galactosyltransferase, which can bind \( \alpha\)-lactalbumin made in mammary tissue so that the enzyme’s acceptor specificity changes from N-acetylgalactosamine toward glucose and permits synthesis of lactose disaccharides during lactation.

Interestingly, we identified a short ORF for a 28-amino acid peptide in the new A4GALT transcripts described here that is unrelated to the \( \alpha_4\)GalT ORF. However, this does not occur in \( P^1 \) as hypothesized but only in \( P^2 \). In addition to this qualitative change, we also observed striking quantitative differences in transcript levels (Figure 2) and antigen expression (Figures 6-7) between P1 and P2 individuals. Consequently, it is tempting to conclude that the original regulator theory...
is invalid. Instead, an alternative hypothesis on the same theme may be formulated: a P2-related molecule (genomic DNA sequence, the new transcript, or a peptide) may down-regulate transcription at the A4GALT locus so that lower amounts of enzyme-encoding mRNA (transcript variant I in Figure 3A) and hence less α4GalT will be produced in the presence of P2. According to this new hypothesis, Pk can be synthesized even at low enzyme levels because lactosylceramide is the favored acceptor, while paragloboside can only be used to make P1 when more enzyme is present. If the regulatory effect of the P2-related molecule is dose-dependent, this may mean that more P1 and Pk antigens would be synthesized in P2-homozygous individuals than in heterozygotes, which is what our data indicate.

Okuda et al studied regulation of A4GALT and found 3 potential binding sites for the Sp1 transcription factor in the promoter but none of these sites involved positions −551_−550insC or −160A>G.35 We measured promoter activity in 2 common promoter sequences, differing at 5 polymorphic sites including these 2 SNPs. However, no differences in the promoter activity could be detected (Figure 4). In theory, the absence of different transcription levels between these promoter variants could be due to lack of transacting factors important for distinguishing between P1 and P2 in the cell line used. It may therefore be necessary to use cell lines of different origins and P1/P2 phenotype to rule out any functional difference between the 2 A4GALT promoter variants. However, it should be kept in mind that none of the genetic differences observed in this region correlated fully with P1/P2 phenotypes.33 Okuda et al investigated the regulation of A4GALT to determine if E coli responsible for hemolytic uremic syndrome can up-regulate Pk expression to aggravate the disease,35 a topic also relevant to other infections, such as urinary tract infections and HIV, and of particular interest since we show here that Pk levels, at least in RBCs, vary according to the zygosity of P1.

We detected significantly more P1 and Pk antigens on RBCs from donors homozygous for P1 than those heterozygous. Thus, zygosity may at least in part explain the well-known but poorly understood variation in P1 strength on RBCs between individuals. It has been proposed previously29 that zygosity for the P1 trait may underlie the strong (scored as 3 to 4+) medium, or weak (scored as 1 to 3+ or even w+) agglutination with anti-P1, and this study indicates that this is indeed the case. Genotyping for P1/P2 may therefore be of value for those reference laboratories producing in-house test RBCs. As previously proposed for other blood groups,41 it is important to characterize the zygosity of test RBC donors so that detection of irregular blood group antibodies pretransfusion is performed at the highest level of safety. Although anti-P1 does not generally cause hemolytic transfusion reactions, accurate identification of anti-P1 permits exclusion of other specificities and selection of P1− cells for further serologic analysis or cross-match–compatible units.

In addition to the quantitative aspects (ie, lower transcript and antigen levels when P2 is present in single or double dose), we identified several different splicing variants among A4GALT-mRNA. The latter is not unexpected, especially in immortalized cell lines. Two of these transcripts (transcript variants II/III in Figure 3A) were not further investigated here due to lack of P1/P2-specific polymorphisms even if they were present also in human mRNA. It has been suggested that A4GALT may have an upstream exon 1 with an alternative promoter,42 but this was deduced from incomplete transcripts in melanoma and breast cancer cells. All transcripts detected in the current study had approximately the same transcription start point and appeared to
use the same promoter. The polymorphic transcript (transcript variant IV) was both found in cell lines and human bone marrow independent of P1/P2 genotype but was undetectable in peripheral blood. Attempts to design a real-time PCR assay for its quantification failed due to unspecific binding (data not shown), possibly due to the Alu-derived exon 2a, see below.

Interestingly, the P2-derived ORF of exon 2a is located within an Alu sequence. These approximately 300-bp elements constitute approximately 10% of the human genome, are only found in primates,41 but are thought not to be translated. Comparisons were made to 34 Alu subfamilies and the highest alignment scores were observed for AluSx/AluSq, but neither subfamily had a thymidine at the position corresponding to the P1/P2-specific SNP. Basic Local Alignment Search Tool analysis against several protein databases identified numerous human proteins that are highly homologous to the candidate peptide (data not shown). It has been suggested that Alu motifs are involved in regulation of gene expression,42 but the impact of this is unclear as is the potential for translated Alu sequences to act as transacting factors.

Our data establish the previously suspected genetic linkage between the P1 and P2 antigens, which calls for a blood group terminologic change. Based on abstract presentations of these data at recent scientific congresses,45,46 we proposed to the ISBT Working Party on Red Cell Immunogenetics and Blood Group Terminology that the P2 antigen should be removed from the GLOB collection (ISBT no. 209) and instead join the P1 antigen in the P blood group system (ISBT no. 003). Our proposal also included that the new and more appropriate name for this system should be P1PK to reflect the antigens included. This change makes to 34

In summary, the identification of a novel A4GALT transcript with a P1- versus P2-defining SNP has eventually resulted in the possibility of determining these phenotypes genetically. This tool has potential value not only in transfusion medicine, a field where large scale, may also be of interest for DNA-based prediction of susceptibility to infections caused by pathogens for which P1 and P2 glycosphingolipid levels are of importance. More work is required to understand the mechanisms by which the newly found SNP exerts its function and how the enzyme that synthesizes the P1/P2 antigens is regulated. Furthermore, any clue toward a possible function of Alu motifs in the human genome has implications beyond blood group expression.

Acknowledgments

We thank Dr Åsa Hellberg and Dr Alan Chester for fruitful discussions and critical review of data. We also thank Annika Hult for technical assistance with flow cytometric analysis, Dr Magnus Jöud for Alu database searches, and Dr Fredrik Svennelid for assistance with genotyping. Dr Jill Storry is thanked for constructive review of the manuscript. Dr Ed Nudelman is acknowledged for verifying by enzyme-linked immunosorbent assay that the monoclonal anti-P1 used in the study does not cross-react with commercially available pure P/Gal3/CD77 glycolipid.

This work was supported by the Swedish Research Council (project no. 71X-14251), the Medical Faculty at Lund University, governmental ALF research grants, and the Skåne county council’s Research and Development foundation, Sweden.

Authorship

Contribution: B.T. and J.S.W. conducted experiments; B.T., J.S.W., and M.L.O. analyzed and interpreted data; and B.T. and M.L.O. designed the study and wrote the paper.

Conflict-of-interest disclosure: B.T. and M.L.O. have applied for intellectual property protection of methods described in the paper. J.S.W. declares no competing financial interests.

Correspondence: Martin L. Olsson, Division of Hematology and Transfusion Medicine, Department of Laboratory Medicine, Lund University, BMC C14, SE-221 84 Lund, Sweden; e-mail: Martin.L.Olsson@med.lu.se.

References


21. Cederberg B. Population studies in northern


33. Hellberg A, Chester MA, Olsson ML. Two previously proposed P1/P2 differentiating and nine novel polymorphisms at the A4GALT (Pk) locus do not correlate with the presence of the P1 blood group antigen. BMC Genet. 2005;6:49.

34. Tilley L, Green C, Daniels G. Sequence variation in the 5’ untranslated region of the human A4GALT gene is associated with, but does not define, the P1 blood-group polymorphism. Vox Sang. 2006;90(3):198-203.


37. Edvardsson L, Dykes J, Olsson ML, Olsson ML, Olofsson T. Sequence variation at the A4GALT (Pk) locus: the missing link between Pk and P1 Histo-blood group antigens [abstract]. Transfusion. 2009;49(3S):48A.


45. Thuresson B, Westman JS, Olsson ML. Genetic typing of the P1/P2 histo-blood groups made possible by identification of a novel A4GALT exon [abstract]. Transfusion. 2009;49(3S):48A.


