Studies on the genetic basis of Pk, P and P1 blood group antigen expression

Hellberg, Åsa

2007

Citation for published version (APA):

Total number of authors:
1

General rights
Unless other specific re-use rights are stated the following general rights apply:
Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

• Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
• You may not further distribute the material or use it for any profit-making activity or commercial gain
• You may freely distribute the URL identifying the publication in the public portal

Read more about Creative commons licenses: https://creativecommons.org/licenses/

Take down policy
If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.
Studies on the Genetic Basis of P^k, P and P1 Blood Group Antigen Expression

Doctoral thesis
by
Åsa Hellberg

Division of Hematology and Transfusion Medicine
Department of Laboratory Medicine
Lund University, Sweden

With the approval of the Faculty of Medicine at Lund University,
this thesis will be defended on March 16, 2007, at 13:00 in Segerfalksalen,
Wallenberg Neurocentrum, BMC, Sölvegatan 17, Lund.

Faculty opponent:
Professor Steven Spitalnik
Department of Pathology, Division of Laboratory Medicine
College of Physicians and Surgeons of Columbia University
New York, NY
USA
Till min familj
# Table of contents

List of publications........................................................................................................6
Abstract..........................................................................................................................7
Abbreviations .................................................................................................................8

**Background**................................................................................................................9

- Blood groups............................................................................................................9
  - Historical perspective............................................................................................9
  - Antigens and antibodies.......................................................................................9
  - Nomenclature.......................................................................................................10
  - Structure and function.........................................................................................10
  - Glycosyltransferases...........................................................................................12
  - Genetics.............................................................................................................13

**The P/GLOB blood group systems and collection**.................................................15
- P<sup>k</sup>, P and P1 antigens and antibodies............................................................15
- Biochemistry.........................................................................................................17
- Genetics.................................................................................................................20
  - The P<sup>k</sup> gene...............................................................................................20
  - The P gene.........................................................................................................20
  - The unknown genes responsible for expression of P1 and LKE.........................21
- Tissue distribution..................................................................................................22
- Disease associations..............................................................................................23
  - Viruses.............................................................................................................23
  - Bacteria............................................................................................................24
  - Cancer.............................................................................................................24

**The present investigation**.......................................................................................26

- Aims of the study....................................................................................................26

**Material and methods**..........................................................................................27
- Blood samples.......................................................................................................27
  - Individuals of Swedish origin...........................................................................28
  - Individuals of Amish origin.............................................................................28
- PCR and sequencing...............................................................................................30
- Cloning..................................................................................................................31
- Cell lines and transfection....................................................................................31
- Flow cytometry......................................................................................................31

**Results and discussion**..........................................................................................32
- The A4GALT gene and P<sup>k</sup> expression............................................................32
- The B3GALNT1 gene and P expression.................................................................37
- A candidate gene for P1 expression.......................................................................40
- Additional work.......................................................................................................41

**Conclusion and future perspectives**......................................................................41

Sammanfattning på svenska..........................................................................................44

Acknowledgements....................................................................................................48

References....................................................................................................................50

Appendix: Papers I-V....................................................................................................57
Ibland tänkte han sorgset: "Varför?"
och ibland tänkte han: "Därför!"
och ibland tänkte han: "Således!"
A.A. Milne
List of publications

This thesis is based on the following papers, which are referred to in the text by their Roman numbers (I-V)

I  Hellberg Å, Poole J, Olsson ML.

II  Hellberg Å, Steffensen R, Yahalom V, Nilsson Sojka B, Heier HE, Levene C, Poole J, Olsson ML.
Additional Molecular Bases of the Clinically Important p Blood Group Phenotype.

III  Hellberg Å, Ringressi A, Yahalom V, Säfwenberg J, Reid ME, Olsson ML.
Genetic Heterogeneity at the Glycosyltransferase Loci Underlying the GLOB Blood Group System and Collection.

IV  Hellberg Å, Chester MA, Olsson ML.
Two Previously Proposed P\(_1\)/P\(_2\)-differentiating and Nine Novel Polymorphisms at the \(A4GALT\) (\(P^k\)) Locus Do Not Correlate with the Presence of the P1 Blood Group Antigen.

V  Hellberg Å, Schmidt-Melbye AC, Reid ME and Olsson ML.
Expression of a Novel Missense Mutation Found in the \(A4GALT\) gene of Amish Individuals with the p Phenotype.
(Manuscript submitted to *Transfusion*).

Paper II and III are reprinted with permission from Blackwell Publishing.
Abstract

The clinically important carbohydrate P/GLOB blood group systems and collection give rise to both common (P₁, P₂) and rare (p, P₁ᵏ, P₂ᵏ) blood group phenotypes. The associated antibodies are implicated in severe transfusion reactions and recurrent spontaneous abortions. The aim of this study was to explore the molecular genetic basis of Pᵏ, P and P₁ antigen expression.

Sequence analysis of the A4GALT and B3GALNT1 genes proposed to synthesize the related Pᵏ (Gb₃) and P (Gb₄) antigens, respectively, were performed in p and Pᵏ individuals (n=99) of different geographic/ethnic origin. A total of 24 novel mutations were identified, emphasizing the genetic heterogeneity at the glycosyltransferase loci underlying these blood groups. As a result of this study, the P antigen was assigned its own blood group system, GLOB (028), by ISBT.

Expression studies in the Pᵏ-negative Namalwa cells transfected with mutated A4GALT-constructs showed Pᵏ expression levels comparable to negative controls. RBCs with p phenotype showed no Pᵏ and P activity while both P and Pᵏ expression on RBCs varied considerably between individuals with common phenotypes.

Sixteen polymorphic sites were detected while investigating if polymorphisms in the regulatory region of the A4GALT gene might be the basis for the P₁/P₂ phenotypes. No clear-cut correlation was found and two previously proposed P₂-specific mutations were detected in homozygous form both in P₁ and P₂ donors indicating that these mutations are not the sole cause of the P₁/P₂ status. However, the correlation between the A4GALT locus and P₂ status seems to be rather strong.
Abbreviations

aa Amino acid
acc.no. Accession number
CD Cluster of differentiation
EC Enzyme commission
EHEC Enterohemorrhagic *Escherichia coli*
ER Endoplasmic reticulum
Gal Galactose
GalNAc N-acetylgalactosamine
GlcNAc N-acetylglucosamine
GT Glycosyltransferase
HDN Hemolytic disease of the newborn
HIV Human immunodeficiency virus
HTR Hemolytic transfusion reaction
HUS Hemolytic uremic disease
ISBT International Society of Blood Transfusion
LacCer Lactosylceramide
MFI Mean fluorescence intensity
NeuAc Sialic acid
PCR-ASP Polymerase chain reaction – allele-specific primer
RBC Red blood cell
SNP Single nucleotide polymorphism
SSEA Stage-specific embryonic antigen
UTI Urinary tract infection
VT Verotoxin
Background

Blood groups

Historical perspective

The interest to transfuse blood to patients suffering from various diseases goes back in history. However, for a long time the attempts to transfuse blood were mostly fatal.

Transfusions between humans as well as between different species were tried, for example sheep to human, and it was noticed that variations between blood from different persons and species existed. In 1829, Blundell performed the first successful human transfusion¹ and with today’s knowledge we understand that this success was partially achieved by pure luck. It was not until 1900 when Landsteiner made the discovery of different blood groups, that it became safer to receive blood transfusions.² Landsteiner found different agglutination patterns, A, B and C (later changed to O), when he mixed blood from different individuals. This blood group system was named ABO and Landsteiner later got the Nobel Prize for his findings.³ However, the practice to mix blood from donor and recipient before transfusion to see if clotting occurred, i.e. crossmatching, was introduced some years later by Ottenberg.⁴ The invention of appropriate devices for transfusion along with the use of anticoagulants were other factors important for the development of blood transfusions. Later came preservative solutions and the possibility to divide whole blood into components which further improved the quality of the blood and provided better treatment for the patients.

Antigens and antibodies

Blood group antigens are found on the red blood cell (RBC) surface but in many cases they exist on other cells as well and are then called histo-blood group antigens. After Landsteiner’s discovery of the ABO blood group system, it took 25 years until yet other blood group antigens were found. The MN and P1 antigens were discovered after injection of human RBCs in rabbits, which resulted in immune antibodies that could distinguish blood from different human individuals.⁵ Later, blood group antigens were discovered due to antibodies made by either transfused patients or in women who were or had been pregnant.

The two most clinically important blood group systems are ABO and Rh. Naturally-occurring antibodies against the ABO antigens not present on the individual’s own cells exist in individuals over 6-12 months of age and therefore the correct ABO blood group has to be determined before blood transfusion and organ transplantation. To avoid potentially lethal transfusion reactions ABO-compatible blood is required,⁶ whilst it appears to be possible to
transplant across the ABO-barrier if appropriate pre-treatment is undertaken. The Rh blood group system is very complex but with the RhD antigen as the main antigen. For an RhD-negative person only a small amount of RhD-positive blood is required to start the production of anti-D because of the high immunogenicity. These antibodies can cause severe transfusion reactions and hemolytic disease of the newborn (HDN). Many of the other blood group antigens can also be implicated in various degrees of hemolytic transfusion reactions (HTR) and HDN, while some appear to be of low or no clinical significance.

**Nomenclature**
The nomenclature for blood groups is administered by the International Society of Blood Transfusion (ISBT) Committee on Terminology for Red Cell Surface Antigens. A requirement for acknowledgement of blood group antigen status is that interindividual variation in these structures must be detectable with alloantibodies in human sera. A blood group system consists of one or more antigens, and is controlled by one genetic locus or by two or three closely located and homologous genes. Today, 29 blood group systems corresponding to these criteria are known. A group of antigens that does not yet fulfil the requirements for a blood group system can be assigned a blood group collection number instead. There are also series for both high (the 901 series) and low (the 700 series) incidence antigens that have no association to any other antigen or blood group system. Altogether around 270 antigens have been detected.

In addition, a database, The Blood Group Antigen Gene Mutation Database (http://www.ncbi.nlm.nih.gov/projects/mhc/xslegi.fcgi?cmd=bgmut/home) was set up under the support of Human Genome Variation Society and is now a part of dbRBC at the National Center for Biotechnology Information (NCBI). This database contains information about the blood group systems, both serological features and genetic variations. It is updated by experts in the field.

**Structure and function**
Blood group antigens are epitopes on proteins, glycoproteins and glycolipids present in the RBC membrane. The structural information about the antigens was obtained from 1950 and onwards by biochemical analysis. This information has been used while trying to designate the function of the different blood group molecules. Various functions or possible functions such as membrane transport (e.g. Kidd, Colton, Diego), chemokine receptors (Duffy), adhesion molecules (e.g. Lutheran, LW) and enzymes (e.g. Kell, Cartwright), have been
assigned to the different systems (Figure 1) but the biological significance has not been clarified in most cases. This is especially true regarding the polymorphic variation. One example is the Duffy glycoprotein, a chemokine receptor, where the two antigens Fy\(^a\) and Fy\(^b\) do not seem to affect the function while the null phenotype Fy(a-b-) has lost the capacity to bind chemokines.\(^{11}\) The Fy(a-b-) phenotype also protects against *Plasmodium vivax* infection.\(^{12}\)

Most antigens are directly encoded by a gene into membrane-associated proteins. However, for some antigens the gene encodes a glycosyltransferase (further discussed below) that catalyses the addition of a sugar onto a precursor, forming carbohydrate structures such as the A and B antigens in the ABO blood group system. Carbohydrates on the cell surface form the glycocalyx of the cell. This “sugar coat” may play many roles. One of them may be to protect the cell from mechanical damage, and others are to serve as points of attachment for other cells and to avoid microbial attacks. However, a number of bacteria, viruses and toxins use the carbohydrates for adherence and the great diversity of these structures is thought to form a basis of herd immunity.\(^{13}\)

---

**Figure 1.**

*Picture showing the blood group systems sorted according to their proven or hypothesized functions.*

The figure is modified from an original kindly provided by JR. Storry.
**Glycosyltransferases**

There are more carbohydrates on earth than any other type of biomolecule. Glucose and galactose are examples of the simplest form (monosaccharides). These can combine to give structures such as disaccharides, oligosaccharides and polysaccharides, but carbohydrates can also be attached to lipids and proteins to form glycolipids and glycoproteins. The enzymes responsible for building these structures are glycosyltransferases (GTs). Enzymes are divided into different enzyme classes (enzyme commission, EC) and GTs belong to EC 2.4. GTs transfer carbohydrates from an activated donor molecule to an acceptor, adding monosaccharides one by one to form carbohydrate chains (**Figure 2**). Most often the acceptor molecules are oligosaccharides, usually in the form of glycoconjugates but can also be a monosaccharide. To form glycolipids, GTs use ceramides as the acceptors. Some GTs utilize the amino-acid residues serine or threonine in proteins to synthesize O-linked glycoproteins; whereas asparagine is the attachment point for N-linked glycoproteins. GTs can also be glycosylated themselves but relatively little is known about the functional importance of this. The specificity for the donor sugar nucleotide, the acceptor molecule and the linkage (α or β) is generally high. A cation like manganese or magnesium is often needed by the GTs to be fully active. Glycosyltransferases and their history are nicely reviewed by Watkins, Paulson & Colley, and Kleene & Berger.

![Figure 2. Schematic picture of a glycosyltransferase adding a monosaccharide to a precursor. More precisely, the transfer of a galactose by the B transferase, 3-α-galactosyltransferase, to the precursor, H antigen, forming the B antigen.](image)

More than 240 human GTs are known that transfer a specific monosaccharide, e.g. sialyltransferases, fucosyltransferases and galactosyltransferases. Surprisingly, very little sequence identity in the catalytic domain has been found in enzymes transferring the same monosaccharide. GTs are also classified into families by amino acid (aa) sequence similarities, listed in the CAZy database (http://afmb.cnrs-mrs.fr/CAZY). X-ray
crystallography and in silico studies have shown that many glycosyltransferases belong to two structural fold superfamilies, GT-A and GT-B. A recent report of the structure for a bacterial sialytransferase describes a new fold family but still there are many glycosyltransferases not assigned to any of the known fold families. The expression of glycosyltransferases is highly regulated during differentiation and proliferation and also specific for different tissues and cells. Glycosyltransferases are localized in the membranes of the endoplasmic reticulum (ER) and Golgi apparatus. Enzymes located in Golgi share the secondary structure typical for type II transmembrane proteins, featuring a single transmembrane domain and a short amino-terminal cytoplasmic segment and a longer carboxy-terminal Golgi-residing domain (Figure 3).

\[ \text{Figure 3.} \]
\[ \text{Golgi glycosyltransferases are type II transmembrane proteins.} \]
\[ \text{Schematic figure modified from Paulson & Colley.}^{16} \]

**Genetics**

The word genetics (from Greek for to give birth) was first used in public in 1906 by Bateson at the Third International Conference on Plant Hybridization (London, England) to describe the science of heredity. However, the rules to explain the inheritance of genes was formulated by the monk Mendel already in 1866. His papers were translated by Bateson from German (for a revised English translation of the papers; see http://www.mendelweb.org/home.html). Between 1953 and 1966 some very important discoveries for genetics were published. The
structure of DNA was explained,\textsuperscript{20} the genetic code solved,\textsuperscript{21} and the processes of transcription and translation described. In 1977 Sanger (a double Nobel prize winner) reported a method to determine the sequence of DNA.\textsuperscript{22} The same year the sequence of the first human gene (for a polypeptide hormone)\textsuperscript{23} was determined and nine years later the first gene encoding a blood group antigen was found. This gene was called \textit{GYPA} and encodes the MN antigens in the MNS system.\textsuperscript{24} After years of serology and biochemistry, and to some extent genetics, it was now time for molecular biology to further add to the knowledge of blood group antigens.

As previously mentioned, 29 blood group systems exist and for these, all genes except the gene encoding the P1 antigen have been cloned and identified. For the antigens belonging to different collections and series, all genes except the genes for the P\textsuperscript{k} and Sd\textsuperscript{n} glycosyl-transferases still remain to be clarified.\textsuperscript{25-29} Actually, for most of those antigens the biochemical structure has not yet been solved.

Much work has been performed to correlate phenotype with genotype. One example is the gene encoding the ABO glycosyltransferase\textsuperscript{30} where mutations in the gene give rise to two different GTs, 3-\(\alpha\)-\(N\)-acetylgalactosaminyltransferase and 3-\(\alpha\)-galactosyltransferase. These enzymes transfer \(N\)-acetylgalactosamine (GalNAc) or galactose (Gal), producing the A or B antigens, respectively.\textsuperscript{31} The null phenotype O is in most cases caused by a single nucleotide deletion, shifting the reading frame and introducing a premature stop codon, but other mutations can also be responsible for blood group O.\textsuperscript{32,33} The molecular basis of many of the rare ABO subgroups such as \(A_x\), \(A_{\text{fimn}}\), \(A_{\text{bantu}}\), \(A_{\text{el}}\), \(B_3\), \(B_x\), \(B_{\text{el}}\), \(B(A)\) and \textit{cis}AB has also been solved.\textsuperscript{32,34-38} These weak phenotypes are generally caused by missense mutations, insertions or hybrid alleles but the sugar specificity for the transferase does not appear to be altered.
The P/GLOB blood group systems and collection

**P^k, P and P1 antigens and antibodies**

Among the carbohydrate blood group antigens are the members of the P and GLOB blood group systems and the GLOB collection, which consist of four antigens (Table 1).

The P1 antigen was discovered in 1927 by Landsteiner and Levine. In 1955 Sanger found the P^k antigen, and four years later Matson *et al.* found the P antigen. A fourth related antigen, Luke (LKE), was described by Tippett *et al.* in 1965.

According to the ISBT nomenclature, the P1 antigen belongs to the P blood group system and the P^k antigen, together with LKE, to the GLOB collection. In the light of data presented in this study, the P antigen has recently been moved from the GLOB collection to make up its own blood group system, GLOB. Combinations of these antigens give rise to the following phenotypes: P_1, P_2, P^k_1, P^k_2 and p. In addition, P_1 and P_2 individuals can be LKE positive or negative.

<table>
<thead>
<tr>
<th>Antigen</th>
<th>Blood group system</th>
<th>Blood group collection</th>
</tr>
</thead>
<tbody>
<tr>
<td>P1</td>
<td>P (003)</td>
<td>-</td>
</tr>
<tr>
<td>P</td>
<td>GLOB (028)</td>
<td>-</td>
</tr>
<tr>
<td>P^k, LKE</td>
<td></td>
<td>GLOB(209)</td>
</tr>
</tbody>
</table>

The frequency of the P_1 phenotype varies between different ethnic groups, ranging from 90% among Africans to 20% in Asians. On RBCs, P1 expression changes during fetal development. The antigen is found as early as week 12 but weakens during gestation. At birth the expression is low and it takes up to seven years before full expression is reached. The strength of the antigen expression can differ from one person to another and it seems to be dependent on dosage. The rare dominant regulator gene In(Lu), first found to inhibit the Lutheran blood group system antigens (hence the name), also prevents normal P1 expression. Its identity is still unknown, however.

The P^k antigen was first thought to be a low-frequency antigen but later it was understood that nearly all the antigens are masked by the P antigen, and thus the erroneous assumption.

RBCs from P_1 individuals express more P^k antigen compared to P_2 individuals but the amount
of P antigen is the same for both phenotypes.\textsuperscript{49} Individuals with the LKE-negative phenotype express more $P^k$ antigen compared to individuals with LKE-positive phenotype.\textsuperscript{50} The P antigen is well developed at birth\textsuperscript{51} and the most abundant neutral glycolipid in the RBC membrane with $14 \times 10^6$ antigens per cell.\textsuperscript{49} The P antigen is also called globoside, a name given because it was discovered and characterized first on RBCs (globules).

Rare null phenotypes exist for both $P^k$ and $P$; the p phenotype that lacks $P^k/P/P1$ antigens; the $P_2^k$ phenotype which lacks $P/P1$ antigens, and the $P_1^k$ phenotype that lacks the P antigen on the cell surface (Table 2). Additional phenotypes might exist since Kundu \textit{et al.} described individuals with either a weak P or weak $P^k$ antigen.\textsuperscript{52,53} The frequency of the p phenotype has been estimated at 5.8 per million in Europeans,\textsuperscript{51} but for Swedes in Västerbotten county in Northern Sweden the number of p individuals is significantly higher (141 per million).\textsuperscript{54} The phenotype also seems to be more common in Japan and among Amish people.\textsuperscript{55,56} The frequency of p in the donor population in Israel is comparable to other populations but among Jews who immigrated to Israel from North Africa the p phenotype prevalence is 10 times higher.\textsuperscript{57} The $P_1^k/P_2^k$ phenotype is even rarer than the p phenotype but also here the phenotype appears to be more common in Japan.\textsuperscript{58} The first individual described with $P^k$ phenotype was of Finnish origin and it seems that Finland has a higher incidence of the $P_1^k/P_2^k$ phenotypes.\textsuperscript{51} Actually, DNA from the first $P^k$ individual, Mrs. Mys., is included in this study (Paper III).

\begin{table}
\centering
\begin{tabular}{|c|c|c|c|}
\hline
\textbf{Phenotype} & \textbf{Frequency} & \textbf{Antigen present on RBC} & \textbf{Antibodies in serum} \\
\hline
$P_1$ & 20-90\% & $P1, P^k, P$ & none \\
$P_2$ & 10-80\% & $P^k, P$ & Anti-$P1^*$ \\
p & rare & None & Anti-$PP1P^k$ \\
$P_1^k$ & rare & $P1, P^k$ & Anti-$P$ \\
$P_2^k$ & rare & $P^k$ & Anti-$PP1$ \\
\hline
\end{tabular}
\caption{A summary of phenotypes and possible antibodies for the P/GLOB blood groups.}
\textsuperscript{*} not always present/detectable.
\end{table}
By analogy to the ABO blood group system, naturally-occurring antibodies of IgM and/or IgG classes are formed against the missing P/GLOB carbohydrate structures (Table 2). Anti-P1 is usually a weak and cold-reactive antibody not implicated in HTR or HDN. However, some P1 antibodies have been reported to react at 37°C, bind complement and cause both immediate and delayed HTRs.\textsuperscript{59-61} The P and P\textsuperscript{k} antibodies can cause HTR but causing HDN has not been reported. However, early spontaneous abortion has a higher frequency among women with p and P\textsubscript{1}P\textsubscript{1}/P\textsubscript{2}P\textsubscript{2} phenotypes and is a phenomenon that is most likely due to the IgG component of the anti-P attacking the placenta.\textsuperscript{62,63} The anti-PP1P\textsuperscript{k} found in individuals with the p phenotype was previously called anti-Tj\textsuperscript{a}, named after Mrs. Jay in whose serum this antibody specificity was first found in association with a tumor.\textsuperscript{64}

Less than 2% of Caucasians are negative for LKE, hence this is a high-incidence antigen.\textsuperscript{41} It is also known as the stage-specific embryonic antigen (SSEA-4).\textsuperscript{65} Unlike the other phenotypes in P/GLOB, naturally-occurring antibodies are not found in LKE-negative individuals. As a matter of fact, only six examples of anti-LKE have been reported.\textsuperscript{66,67} To date, this antibody specificity has not been implicated in HTR or HDN.\textsuperscript{68,69}

**Biochemistry**

The biochemistry of the P/GLOB blood groups was partially elucidated by Morgan and Watkins in the 1960s by a series of experiments on hydatid cyst fluid from sheep infected by the tapeworm *Echinococcus granulosus*.\textsuperscript{70} They purified P1-specific components and showed that a glycoprotein containing the Galα1-4Galβ1-4GlcNAc reacted as a P1 determinant. At the same time Marcus managed to extract P1 glycolipids from RBCs.\textsuperscript{71}

The antigens P\textsuperscript{h}, P and P1 are related and of carbohydrate nature. Depending on which carbohydrates are added to lactosylceramide (LacCer), different series of glycosphingolipids are formed. Glycosphingolipids were first described by Thudichum in 1884 and he named them after the inscrutable Egyptian Sphinx, since both the structure and function were unknown at the time.\textsuperscript{72} These molecules consist of a sugar moiety with a lipid tail, ceramide, and make up the outer leaflet of cell membranes together with phospholipids, cholesterol and glycerolipids. Lipids are organised in microdomains such as glycosynapses and lipid rafts.\textsuperscript{73}
All P/GLOB-related antigens are formed on the same precursor, LacCer, which is the most common precursor for glycosphingolipids in mammals and birds. P^k (globotriaosylceramide, Gb3, ceramide trihexoside, CD77), P (Gb1, globotetraosylceramide, globoside) and LKE (monosialogalactosylgloboside, SSEA-4) belong to the globo series and P1 to the neolacto/paraglobo series (Figure 4).

A 4-α-galactosyltransferase (α4Gal-T / P^k synthase) catalyzes the transfer of Gal to the Gal residue on the LacCer, producing the P^k antigen. The P antigen is created with the addition of GalNAc by a 3-β-N-acetylgalactosaminyltransferase (β3GalNAc-T1 / P synthase). In another pathway the P1 antigen is formed by three sequential glycosylation reactions, the last one performed by a 4-α-galactosyltransferase, but it is still unclear if this enzyme is identical to the one synthesizing the P^k antigen. Furthermore, other glycosyltransferases form additional antigens associated to the P/GLOB system/collection, such as Forssman, globo-H and globo-A. It has been debated whether P^k, P and P1 exist on glycoproteins in the RBC membrane and according to Yang et al. glycolipids are the sole carriers of these antigens on the RBC. The biochemical, serological and clinical aspects of the P-related blood group systems have been thoroughly reviewed by Spitalnik and Spitalnik.
Figure 4.
Scheme showing the biosynthesis of $P^k$, $P$ and $P1$ antigens.
Some other related structures, such as the blood group A and B antigens, are also included.

- ■ = Glucose
- ▼ = Fucose
- ○ = Galactose
- ● = $N$-acetylgalactosamine
- □ = $N$-acetylglucosamine
- ◊ = Sialic acid
**Genetics**

The $P^k$ gene

The gene for the 4-$\alpha$-galactosyltransferase ($\alpha 4$Gal-T / $P^k$ synthase, EC 2.4.1.228) was cloned in 2000 by three independent research groups.\(^{25-27}\) The $P^k$ gene ($A4GALT$) is located on the long arm of chromosome 22 and consists of two or three exons with the whole coding region in the last exon (**Figure 5**).\(^{79,80}\) This gene encodes a type II transmembrane glycoprotein with 353 aa and is highly conserved among species.\(^{16,26}\) The gene contains a characteristic DXD motif (aa 192-194) which is a conserved motif existing in nearly all glycosyltransferases.\(^{81}\) It has been proposed to participate in the coordination of the metal ion and in the binding of the sugar nucleotide.\(^{81}\) It has not yet been clarified what type of cation $A4GALT$ requires for its function. The gene has binding sequences for the transcription factor AP-1, 160 bp up-stream of the transcription start according to a computer search done by Hughes *et al.*\(^{82}\)

Four mutations in the $A4GALT$ gene, 548T>A (M183K), 560G>A (G187D), 752C>T (P251L), and 783G>A (W261X), were originally reported to destroy the enzyme’s activity and lead to the p phenotype.\(^{25,79}\) Two silent mutations, 903G>C (P301P) and 987G>A (T329T), as well as one missense mutation, 109A>G (M37V), with no apparent effect on the $\alpha 4$Gal-T were also found.\(^{25}\)

The $P$ gene

Another gene, first cloned in 1998 as a member in the 3-$\beta$-galactosyltransferase family\(^{83}\) but later shown to be a 3-$\beta$-$N$-acetylglactosaminytransferase, was suggested as the globoside (Gb4, P) synthase.\(^{84}\) This gene ($B3GALNT1$) is located on chromosome 3 and has at least five exons with the entire coding region in the last exon (**Figure 5**). The gene encodes a type II transmembrane glycoprotein ($\beta 3$GalNAc-T1, EC 2.4.1.79) with 331 aa and five potential $N$-glycosylation sites.\(^{84}\)

A recent paper claims that erythroid genes have a high frequency of alternative first exons and both $A4GALT$ and $B3GALT3$ (the previously used name for $B3GALNT1$) are included in that study. These genes both have two alternative first exons which according to this computational study give rise to two protein isoforms.\(^{85}\)
The unknown genes responsible for expression of P1 and LKE

The molecular genetics governing expression of the LKE and P1 antigens have not yet been clarified. The P1 gene has been mapped to the same chromosome as the A4GALT gene but to 22q11.3 instead of 22q13.2.25,86 The chromosomal location of the LKE gene is still unknown. A major enigma is why the P1 antigen is always absent in the p phenotype. Different theories exist: One model suggests that the same α4Gal-T is able to transfer galactosyl residues to both LacCer and paragloboside but in order to use the latter as the acceptor, a regulatory protein is required.87 Another hypothesis suggests that two different enzymes exist, requiring both of them to be inactivated to cause the p phenotype.87 This model is supported by a study showing that microsomal enzymes from P1 kidneys could synthesize both P1 and Pk while enzymes from P2 kidneys only could produce Pk.88 A third model proposes a single gene with three alleles, one allele coding for a α4Gal-T using LacCer and paragloboside as the possible acceptors, one allele using LacCer only and the third allele coding for an inactive form of the
transferase.\textsuperscript{89} However, no polymorphisms in the coding region of the \textit{A4GALT} gene appear to explain the P\textsubscript{1}/P\textsubscript{2} phenotypes.\textsuperscript{25}

For synthesis of the LKE antigen, a 3-\(\alpha\)-sialyltransferase gene is required to catalyse the addition of sialic acid (NeuAc) to galactosylgloboside (Gb5, SSEA-3) but it is not clear if the LKE-negative phenotype depends on a defect in that putative gene or in the 3-\(\beta\)-galactosyltransferase gene catalysing the previous step in the biosynthetic chain.

\textit{Tissue distribution}

The expression of glycosphingolipids varies between tissues and species. Expression of \(P^k\) and P antigen and their corresponding genes have been studied in several species.\textsuperscript{26,90,91} Studies of mouse tissues shows expression patterns similar to humans although there are some differences.\textsuperscript{90}

The P\textsubscript{1} structure is found as glycolipids and/or glycoproteins in many organisms such as the nematode (\textit{Ascaris suum}), tapeworm (\textit{Echinococcus granulosus}), earthworms (\textit{Lumbricus terrestris}), liver flukes (\textit{Fasciola hepatica}), bacteria (\textit{Neisseira gonorrrhoe}) and pigeon.\textsuperscript{66} The \(P^k\) antigen is also expressed in several strains of bacteria.\textsuperscript{92}

In humans, glycosphingolipids can be useful as surface markers of normal erythrocyte differentiation and of erythroleukemias.\textsuperscript{93} The \(P^k\), P and P\textsubscript{1} antigen are expressed on a number of other cells in addition to RBCs but various studies using different antibodies or methods have come to different conclusions. Both P and \(P^k\) have been detected in plasma\textsuperscript{61,94} but no such reports about P\textsubscript{1} in plasma or about \(P^k\), P and P\textsubscript{1} in secretions are available.

The P\textsubscript{1} antigen is expressed on lymphocytes, granulocytes and monocytes.\textsuperscript{95} The \(P^k\) antigen has been found on granulocytes, monocytes,\textsuperscript{95} fibroblasts\textsuperscript{48}, platelets, smooth muscle cells of the digestive tract and urogenital system\textsuperscript{96} and is a differentiation antigen expressed on a subset of tonsillar B cells in the germinal center\textsuperscript{97}. High expression of \(P^k\) in the kidney has been implicated in susceptibility to hemolytic uremic disease (HUS), further discussed in the \textit{Disease associations} section below. The mechanism behind the high renal expression might be due to enhanced \textit{A4GALT} gene transcription and reduced \(\alpha\)-galactosidase gene transcription.\textsuperscript{82} The \(P^k\) antigen is also known as Burkitt lymphoma antigen and has been classified as CD77.\textsuperscript{98} Clusters of differentiation (CD, http://www.sciencegateway.org/resources/prow/index.html) are cell surface molecules that are immunologically significant and can be determined by a monoclonal antibody.
The P antigen is expressed on megakaryocytes, fibroblasts but not lymphocytes and granulocytes according to van dem Borne et al. while Shevinsky et al. could not detect P on fibroblasts.\textsuperscript{99,100} In another study P was detected in 11 of 16 investigated tissues, especially in those of mesodermal origin.\textsuperscript{101} The P antigen can also be found in placenta.\textsuperscript{63} Furthermore, embryonal carcinoma cells express P antigen and according to Song et al. the P antigen is a possible initiator of signal transduction through AP-1 and CREB associated with cell adhesion.\textsuperscript{102}

Northern blot studies of human organs showed high expression of the \textit{A4GALT} gene in kidney and heart in one study\textsuperscript{25} whilst another one found, in addition to kidney and heart, high expression in spleen, liver, testis and placenta.\textsuperscript{27} High expression of the \textit{B3GALNT1} gene has been demonstrated in brain and heart, moderate expression in lung, placenta and testis, and low expression in kidney, liver, spleen and stomach.\textsuperscript{84} No studies on P1 have been performed since the identity of the gene has not yet been clarified.

**Disease associations**

The P\textsuperscript{k}, P and P1 antigens can act as membrane receptors for several pathogens and toxins, summarized in Table 3.

**Viruses**

Parvovirus B19, which causes the so-called fifth disease, uses erythroid precursor cells expressing the P antigen for its replication.\textsuperscript{103,104} Infection during pregnancy with B19 can give rise to fetal anemia and in some cases, fetal loss, due to the virus role as a inhibitor of hematopoiesis.\textsuperscript{105}

Paroxysmal cold hemoglobinuria, which can be seen in children following a viral infection, is most often caused by an auto-anti-P. This complement-fixing and cold-reactive antibody, also called Donath-Landsteiner, lyse autologous P-positive erythrocytes.\textsuperscript{106} Some data also suggest that the P\textsuperscript{k} antigen is another co-factor for the human immunodeficiency virus to enter CD4-positive cells\textsuperscript{107} but a recent report proposes the opposite.\textsuperscript{108} This will be further discussed in the Conclusion and future perspectives part of this thesis.
Bacteria

Uropathogenic *Escherichia coli* expressing pap-encoded adhesins binds to Pk, P and P1\(^{109-111}\) (reviewed by Spitalnik and Spitalnik\(^{78}\)) and both the *Streptococcus suis* adhesin and the PA-IL lectin from *Pseudomonas Aeruginosa* uses P1 and Pk\(^{\ast}\) as receptors.\(^{112,113}\)

A disease connected to the Pk antigen is Fabry disease where deficiency of the lysosomal enzyme α-galactosidase A causes accumulation of sphingolipids, mainly Pk, in some cell types and body fluids.\(^{114}\) A recent study shows that mice with Fabry disease are protected against the verotoxin (VT) from enterohemorrhagic *Escherichia coli* (EHEC).\(^{115}\) These data are surprising since Pk is the cellular receptor for VT\(^{116,117}\) as well as shiga toxin\(^{118}\) and therefore a higher sensitivity would be expected. The authors hypothesize that the excess Pk can work as a toxin sink, which allows the toxin to bind to Pk in tissues that normally do not have high expression. EHEC infection can induce HUS which leads to hemolytic anemia, renal failure and thrombocytopenia.\(^{119}\) According to Furukawa *et al.* the mechanism behind the thrombocytopenia might be that VT binds to Pk in immature megakaryoblasts and induces their apoptosis, leading to the restraint of platelet production in the bone marrow.\(^{120}\) The Pk antigen has been shown to mediate apoptotic signals following the binding of both verotoxin and anti-Pk (CD77 monoclonal antibody). These ligands trigger two completely different apoptotic pathways, one caspase- and mitochondria-dependent and one reactive-oxygen-species (ROS)-dependent pathway.\(^{121}\) It has also been shown that patients with HUS have lower levels of Pk glycolipid in their sera compared to a healthy control group.\(^{122}\) These authors propose that during infection circulating VTs should bind to Pk glycolipids in sera which may reduce the amount of VTs binding to the target cells. Consequently, patients with low serum levels of Pk would have a higher susceptibility to EHEC infections. Another study states that only Pk and not P1, as earlier believed, is the receptor for VTs, and mice without Pk lose sensitivity to VTs.\(^{123}\)

Cancer

Altered glycosylation patterns of glycosphingolipids such as neoexpression, underexpression or overexpression are characteristic for cancer cells.\(^{124}\) One example is the initial p individual (lacking Pk, P and P1 antigens) who had a gastric tumour which expressed P antigen. Levine proposed that the antibodies made against the Pk, P and P1 antigens prevented further growth of the tumour.\(^{125}\) Expression of Pk antigen has also been described in ovarian carcinomas, colon cancer, breast cancer and B cell lymphomas.\(^{126-128}\)
Table 3.
A selection of pathogens and their toxins with their relationship to the Pk, P and P1 antigens.

<table>
<thead>
<tr>
<th>Pathogen/toxin</th>
<th>Disease</th>
<th>Antigen involved</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Virus</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Parvovirus B19</td>
<td>Fifth disease</td>
<td>P</td>
<td>103,104</td>
</tr>
<tr>
<td>HIV</td>
<td>AIDS</td>
<td>p^k</td>
<td>107,108</td>
</tr>
<tr>
<td><strong>Bacteria</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Uropathogenic</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>Escherichia coli</em></td>
<td>UTI</td>
<td>P^k, P, P1</td>
<td>109-111</td>
</tr>
<tr>
<td><em>Streptococcus suis</em></td>
<td>Meningitis</td>
<td>P^k, P1</td>
<td>112</td>
</tr>
<tr>
<td><strong>Toxin/lectin from</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>Shigella dysenteriae</em></td>
<td>Dysentery</td>
<td>P^k</td>
<td>118</td>
</tr>
<tr>
<td>(Shiga toxin)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>Escherichia coli</em> O157 (VT1/2)</td>
<td>HUS, hemorrhagic colitis</td>
<td>P^k</td>
<td>129</td>
</tr>
<tr>
<td><em>Escherichia coli</em> (VT2e)</td>
<td>Pig edema disease</td>
<td>P</td>
<td>130</td>
</tr>
<tr>
<td><em>Pseudomonas aeruginosa</em> (PA-IL lectin)</td>
<td>Opportunistic human pathogen</td>
<td>P^k, P1</td>
<td>113,131</td>
</tr>
</tbody>
</table>
The present investigation

Aims of the study

- to study a candidate 3-β-N-acetylgalactosaminyltransferase gene, to investigate if it is responsible for synthesis of the P antigen and if mutations in this gene cause the null phenotypes P$_1^k$ and P$_2^k$ (Papers I, III and unpublished results),

- to study the $A4GALT$ gene, responsible for 4-α-galactosyltransferase expression, in individuals of different geographic and ethnic origin with the p phenotype in order to clarify the molecular genetic basis (Papers II, III, V and unpublished results),

- to investigate candidate genes/polymorphisms possibly responsible for expression of the P1 and LKE antigens (Paper IV and unpublished results).
Material and methods

Blood samples

In this study, both samples with common phenotypes (P₁ and P₂) and samples with the rare phenotypes p, P₁\(^k\) and P₂\(^k\) were investigated. The samples were obtained from the Blood Centre in Lund and from several international blood centres, for details see Papers I-V. Four additional samples are also included in this thesis, three with p phenotype (Thailand n=2, Argentina=1) and one with P₁\(^k\) phenotype (Sweden but originating from Palestine).

The geographic and ethnic origin for the individuals included in the study was diverse (Table 4). Unfortunately, for some of the samples (n=10) it was not possible to trace their exact origin.

Table 4.
Geographic and/or ethnic origin of the individuals whose samples with the p and P₁\(^k\)/P₂\(^k\) phenotypes were investigated in this study.

<table>
<thead>
<tr>
<th>p phenotype</th>
<th>n</th>
<th>P₁(^k)/P₂(^k) phenotype</th>
<th>n</th>
</tr>
</thead>
<tbody>
<tr>
<td>Argentina</td>
<td>1</td>
<td>Arabic origin</td>
<td></td>
</tr>
<tr>
<td>Brazil</td>
<td>1</td>
<td>Palestine</td>
<td>1</td>
</tr>
<tr>
<td>England</td>
<td>1</td>
<td>Unknown</td>
<td>1*</td>
</tr>
<tr>
<td>France</td>
<td>1</td>
<td>Canada</td>
<td>1</td>
</tr>
<tr>
<td>Israel</td>
<td>1</td>
<td>England</td>
<td>2</td>
</tr>
<tr>
<td>Arabs</td>
<td>2</td>
<td>Finland</td>
<td>3*</td>
</tr>
<tr>
<td>North African Jews</td>
<td>4</td>
<td>France</td>
<td>1*</td>
</tr>
<tr>
<td>India</td>
<td>1</td>
<td>Italy</td>
<td>1</td>
</tr>
<tr>
<td>Italy</td>
<td>5</td>
<td>Germany</td>
<td>3</td>
</tr>
<tr>
<td>Japan</td>
<td>2</td>
<td>Switzerland</td>
<td>1</td>
</tr>
<tr>
<td>Norway</td>
<td>2</td>
<td>USA</td>
<td>1</td>
</tr>
<tr>
<td>Pakistan</td>
<td>1</td>
<td>Unknown</td>
<td>1</td>
</tr>
<tr>
<td>Poland</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sweden</td>
<td>30</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Västerbotten</td>
<td>29</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Skåne</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Thailand</td>
<td>2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>USA</td>
<td></td>
<td>Amish</td>
<td>19</td>
</tr>
<tr>
<td>USA</td>
<td></td>
<td>Mennonite</td>
<td>1</td>
</tr>
<tr>
<td>USA</td>
<td></td>
<td>Unknown</td>
<td>9</td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td></td>
<td>83</td>
</tr>
</tbody>
</table>

* One or two of these were P₂\(^k\) individuals.
To avoid duplicates in this study when two or more samples showed the same critical p or Pk mutation and had identical ABO and Rh blood groups, we used PCR-based DNA fragment analysis of short tandem repeats spread throughout different chromosomes in the genome. Based on such analysis, four samples were excluded before summarizing the data (thus not included in Table 4). For two populations, Swedes from Västerbotten and the Amish, a larger cohort of related samples was studied with respect to mutations in the A4GALT gene; they are discussed in more detail below. Except for these groups and for two other donors with the p phenotype, who were known to be sisters (Paper III), the remaining samples were all apparently unrelated individuals.

**Individuals of Swedish origin**

The first individual in the world with anti-P1Ppk and p phenotype was found in 1951 in the USA. Around the same time, the first Swedish individual was discovered but not until a few years later was the correct phenotype assigned. In the 1960s many more Swedish cases were found. All except one were from Västerbotten county and this area became well known for its exceptionally high frequency of p individuals, 141 per million. Since then, the regional blood bank in Umeå has provided rare units to numerous patients with anti-P1Ppk worldwide over the years. The people in Västerbotten are a homogeneous group with a higher prevalence for certain genetic diseases compared to the rest of Sweden. This might be due to both founder effect and higher rates of consanguineous marriages. The genealogy of the p individuals in Västerbotten, extensively studied by Dr. Cedergren et al., showed that approximately half of the cases originate from the same ancestor in the 17th century.

In Paper II twenty-nine samples from individuals in Västerbotten with the p phenotype were investigated. Of the 28 known p families in Västerbotten, members from 20 of them were studied here. In Paper III, a Swedish individual but of Finnish origin was analysed.

**Individuals of Amish origin**

The Amish, often called the Plain People, originate from Germany and Switzerland and are descendants of Anabaptist groups formed in the early 16th century during the radical reformation. They came to North America in the beginning of the 18th century to avoid religious persecution and military service in their home countries. In 2000, Raber's Almanac estimated that there were 198,000 Old Order Amish in the USA. There are Old Order communities in 25 states (Figure 6). Ohio has the largest Amish population (55,000),
followed by Pennsylvania (47,000) and Indiana (37,000). Settlements also exist in Ontario, Canada.

Due to a high incidence of consanguineous marriages several genetic diseases are found among the Amish.\textsuperscript{135} The estimated coefficient of inbreeding for the Amish population is 0.0151, which is approximately equivalent to having second cousins as parents.\textsuperscript{136} The incidence of the p phenotype is high in some of the Amish settlements but to date, no study has been performed to find out the exact frequency or in which settlements individuals with the phenotype are found.

Nineteen samples from Amish people originating from Ohio (n=16), Pennsylvania (n=1), Indiana (n=1) and an unknown geographic area (n=1) with the p phenotype were investigated in \textit{Paper V}.

\begin{figure}
\centering
\includegraphics[width=\textwidth]{map_001.png}
\caption{Map showing the distribution of Amish settlements in the United States. Asterisks indicate states with a significant group of Amish people. The filled squares specify the three states (Ohio, Pennsylvania and Indiana) with the largest Amish populations, which corresponds to the states where the investigated samples originate from.}
\end{figure}
**PCR and sequencing**

In all papers (Papers I-V) PCR was used both to amplify general fragments of the gene of interest and to amplify a specific allele based on SNP-or allele specific primers, (PCR-ASP). Oligonucleotide primers used in the studies were synthesized by DNA Technology ApS (Aarhus, Denmark). The primers and conditions used for PCR are listed in the respective papers.

For some of the PCR-ASP designs (Papers III-V), mismatched primers were used to increase specificity. This is done by introducing a mismatched nucleotide in position 3 or 4 from the 3’-end of the primer and the best result is obtained when a pyrimidine is changed to a purine (C>A or G>T).

Parts of the 5’-end of the *A4GALT* gene have high GC contents and also contain several repetitive regions. GC pairs in the DNA spiral have three hydrogen bonds instead of two as in the AT pairs, making it more difficult to denature GC-rich areas. The templates are also more prone to renaturation and to form secondary structures within each strand before the primers have had a chance to anneal to their intended sequences. This problem can be overcome by using enzyme combinations and buffers specially designed to amplify GC-rich areas such as GC-rich PCR System, (Roche Diagnostics GmbH, Mannheim, Germany). Repetitive areas need careful primer design to avoid obtaining multiple PCR-fragments.

PCR products were excised from 3% agarose gels (Seakem, FMC Bioproducts, Rockland, ME) stained with ethidium bromide (Sigma Chemicals, St. Louis, MO) following high-voltage electrophoresis and purified using Qiaquick gel extraction kit (Qiagen GmbH, Hilden, Germany).

The Big Dye Terminator Cycle Sequencing kit (Applied Biosystems, Foster City, CA) and ABI PRISM 310/3130 Genetic Analyser (Applied Biosystems) were used for direct DNA sequencing with capillary electrophoresis and automated fluorescence-based detection according to the manufacturer’s instructions. Besides the primers used for amplification of the fragment, internal primers were used as sequencing primers. In Paper V, plasmid-specific primers were used for the cloned fragments. To avoid artefacts, sequencing was performed on both strands and using fragments obtained by independently designed PCRs.
**Cloning**

In **Paper V** cloning was performed using pcDNA3.1 Directional TOPO® Expression Kit (Invitrogen, UK) according to the manufacturer’s instructions. Briefly, PCR was performed using a forward primer including a Kozak translation initiation sequence with an ATG initiation codon to enable translation. The primer also contained the sequence CACC in the 5’-end which will base pair with a sequence in the vector used. The PCR product was transformed into *E. coli* using heat shock. The bacteria were grown overnight on agar plates with ampicillin which only permits colonies containing vector to grow. Colonies were selected and the constructs were analysed with DNA sequencing. Colonies with confirmed constructs were chosen, grown in medium overnight and plasmid DNA was purified using Qiagen Plasmid Maxi Kit (Qiagen).

**Cell lines and transfection**

Pκ-negative Namalwa cells (European Collection of Cell Cultures no. 87060801) originating from a human Burkitt’s lymphoma were used for expression studies (**Paper V**). These cells were chosen in favour of another Pκ-negative cell line, L cells (mouse fibroblasts), due to their hematopoietic background. An erythropoietic cell line would have been an even better choice but a Pκ-negative erythropoietic cell line is not known. Besides, most previous expression studies have been done in Namalwa or L cells. The Namalwa cells were grown in RPMI 1640 medium (GIBCO-BRL, France) supplemented with 10% FCS. Transient transfection of the cells with 15 µg construct and 2 µg pmaxGFP vector (Amaza, Cologne, Germany) as internal control was performed by electroporation at 0.25 V and 960 µF using GenePulser (BioRad). The Namalwa cells were resuspended in RPMI 1640 medium and grown for 48 hours before further analysis.

**Flow cytometry**

Flow cytometric analysis (FACScan, Becton Dickinson, CA, USA) was used to measure the expression of P and Pκ antigen on RBCs and Namalwa cells (**Paper V**). Several monoclonal antibodies against the Pκ antigen exist. Antibody clones 38.13 (rat IgM), 1A4 (mouse IgM), both kind gifts from J. Wiels, France and 5B5 (mouse IgM, anti-CD77, Becton Dickinson) were tested. The best result was obtained with 5B5 diluted 1:2 (final concentration in reaction well was 1:20). Incubation was performed for 10 minutes at room temperature for detection of the Pκ antigen on transfected Namalwa cells and 1 hour at 4°C for
detection of the Pk antigen on RBCs. For the P antigen no commercial antibody was available but Sanquin in the Netherlands kindly provided the monoclonal AME-2 (rat IgM). The antibody was used diluted 1:10 (final concentration in wells was 1:100) and incubated for 1 hour at 4°C.

The number of events analysed was 10,000 per gated cell population and analyses were carried out using CellQuest™ software (Becton Dickinson).

Results and discussion

The A4GALT gene and Pk expression

The sequence deposited in GenBank with accession number (acc.no.) AB041418 has been used as the reference for investigations of the A4GALT gene in Papers II, III and V.

In Paper II, 40 samples of eight different nationalities with the p phenotype were investigated by DNA sequencing of the A4GALT gene’s coding regions. A majority of the samples (n=29) was from Västerbotten county, Sweden, and we confirmed the surprising data from Furukawa et al. that two different mutations, 548T>A and 560G>A, exist in this restricted area.79 Altogether, nine different critical mutations in the A4GALT gene were encountered, five of which are novel (Table 5). The incidence of spontaneous abortions is much higher in women with p and Pk/Pk phenotype compared to women with common phenotypes.62,63 Among the Israeli (n=3) and Swedish samples, 20 were from women and we noted that whilst some had suffered multiple spontaneous abortions, others had no such events recorded in their medical histories. There were no data to suggest that the immunoglobulin class or titre would explain this difference. Other studies have also shown that some women with the p phenotype not always are prone to a higher abortion risk.58,138 The reason for this is completely unclear and merits further studies.

In Paper III, an additional 20 samples from p individuals of different geographic and ethnic origins were investigated together with Pk/Pk samples. DNA sequencing was performed following amplification of the coding regions in the A4GALT gene. Nine novel and five previously described mutations were detected (Table 5). A sample from a p individual in Skåne (southern Sweden), whose parents were both of Finnish origin, had the 548T>A mutation, which previously had only been found in the northern part of Sweden.
In order to define the molecular basis of the p phenotype in individuals of Amish origin (n=19) the coding regions of the *A4GALT* gene was analysed (**Paper V**). An additional sample was included in this study based on the results obtained. This sample was referred to our laboratory, the Nordic Reference Laboratory for Genomic Blood Typing, as a clinical investigation from Norway. The propositus was of Pakistani origin and her p phenotype was discovered during pregnancy due to an anti- PP1P\(^k\) in her serum. All samples including the Pakistani sample were homozygous for a previously not described mutation, 299C>T, changing serine to leucine in a region of the glycosyltransferase that is highly conserved in homologous genes among several species. In order to determine if the novel mutation abolishes the enzyme activity completely, expression studies were performed. Three different *A4GALT* constructs were made, containing either a consensus allele (positive control), 548T>A (negative control), or 299C>T. Namalwa cells, which do not express P\(^k\) antigen, were transfected with these constructs and P\(^k\) expression was measured by flow cytometry after labelling with monoclonal anti-CD77. The Namalwa cells with a construct containing the 299C>T showed expression levels comparable to the negative controls. P and P\(^k\) antigen expression on RBCs with P\(_1\)/P\(_2\) phenotypes were also investigated with flow cytometry to establish a baseline for common samples. The strength of P\(^k\) expression was surprisingly variable and the same was true for the P expression, although the overall P expression was much stronger, as expected. In addition, RBCs with the 299C>T mutation were analysed. RBCs with p (548T>A, known to have no P\(^k\) expression)\(^{25}\) and P\(_1\)P\(^k\) phenotype were used as controls. The RBCs with the 299C>T mutation showed no P\(^k\) and P activity compared to the negative p control. Interestingly, when looking at P\(^k\) expression, the mean fluorescence intensity (MFI) of P\(_2\) control RBCs did not differ from that of p whilst P\(_1\) RBCs was only marginally higher. According to flow cytometric analysis of P and P\(^k\) antigen levels on RBCs with the p and common phenotypes, it can therefore be proposed that analysis of P antigen levels is a better predictor for p status than P\(^k\), especially when the variation among P\(_1\)/P\(_2\) samples is taken into consideration.

Three additional samples with the p phenotype have recently been referred to us for analysis of the *A4GALT* gene. In a p individual from Argentina, with Spanish origin, 752C>T was found.\(^{139}\) This mutation has previously been described in three Japanese individuals\(^{79}\) and five Americans with unknown ethnic origins (**Paper III**). Sequencing of the *A4GALT* gene in two unrelated samples from Thailand showed that both samples were homozygous for a novel
mutation, 559G>C, that encodes an aa change from glycine to arginine at residue 187. While this mutation has not been described before it is only one nucleotide away from one of the two silencing mutations previously reported in Swedish p individuals; 560G>A, G187D. These results suggest that this region of the enzyme is critical for function. Thus, mutation of the small neutral glycine to either a positively (arginine) or negatively (aspartic acid) charged residue disrupts enzyme activity. Indeed, Furukawa et al. showed in an expression study with mouse fibroblast L cells, that constructs with the 560G>A mutation demonstrated only marginal Pk activity. In the same study a construct containing 752C>T showed no activity at all. In our study, RBCs from one of the Thai individuals (559G>C) and the Argentinian (752C>T) were subjected to flow cytometric analysis with antibodies against the Pk and P antigens (Figure 7). As expected no expression compared to the negative controls (RBCs with p548A and Pk phenotype, respectively) was found.

![Image](image_url)

**Figure 7.**

**Cell surface expression of P^k and P antigen on RBCs with the p phenotype.**

Flow cytometry histograms showing P^k (top row) and P (bottom row) expression on RBCs as measured by monoclonal antibodies 5B5 and AME-2, respectively. The filled grey histograms illustrate unlabeled RBCs (identical result compared to RBCs labelled with secondary antibody only). The thin black curves show the positive controls; RBCs with P^k phenotype or RBCs with Pk phenotype, top and bottom histograms, respectively. The bold black curves show RBCs from the Thai individual (p599C, left panel) and the Argentinian individual (p752T, right panel). The dotted curves illustrate the negative controls; RBCs with p548A phenotype or RBCs with P phenotype, in the top and bottom histograms, respectively. The MFI values for each of the samples are shown in the histograms.
Table 5. A summary of all mutations found in the *A4GALT* gene (in this study and others). Roman numerals refer to Papers II, III and V.

<table>
<thead>
<tr>
<th>nt position</th>
<th>Ref.</th>
<th>Acc.no.</th>
</tr>
</thead>
<tbody>
<tr>
<td>69</td>
<td></td>
<td></td>
</tr>
<tr>
<td>243</td>
<td></td>
<td></td>
</tr>
<tr>
<td>428</td>
<td></td>
<td></td>
</tr>
<tr>
<td>496</td>
<td></td>
<td></td>
</tr>
<tr>
<td>504</td>
<td></td>
<td></td>
</tr>
<tr>
<td>734</td>
<td></td>
<td></td>
</tr>
<tr>
<td>997</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1029</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Allele</th>
<th>Origin</th>
<th>Consensus</th>
<th>Change</th>
</tr>
</thead>
<tbody>
<tr>
<td>p(68_69insT)</td>
<td>Israel</td>
<td>insT G</td>
<td>A</td>
</tr>
<tr>
<td>p(241_243delTTC)-1</td>
<td>Japan/Engl</td>
<td>del</td>
<td></td>
</tr>
<tr>
<td>p(241_243delTTC)-2</td>
<td>Italy</td>
<td>del</td>
<td>C</td>
</tr>
<tr>
<td>p(287A)</td>
<td>Italy</td>
<td>A</td>
<td>C</td>
</tr>
<tr>
<td>p(290T)</td>
<td>Poland</td>
<td>G</td>
<td>T</td>
</tr>
<tr>
<td>p(299T)-1</td>
<td>Amish</td>
<td>T</td>
<td>A</td>
</tr>
<tr>
<td>p(299T)-2</td>
<td>Pakistan</td>
<td>T</td>
<td>C</td>
</tr>
<tr>
<td>p(301delG)</td>
<td>China</td>
<td>delG</td>
<td></td>
</tr>
<tr>
<td>p(del-ins)-1</td>
<td>Asian</td>
<td>del/ins</td>
<td></td>
</tr>
<tr>
<td>p(del-ins)-2</td>
<td>England</td>
<td>del/ins</td>
<td>C</td>
</tr>
<tr>
<td>p(357A)</td>
<td>Brazil</td>
<td>A</td>
<td>C</td>
</tr>
<tr>
<td>p(382_384delinsC)</td>
<td>Unknown</td>
<td>insC</td>
<td></td>
</tr>
<tr>
<td>p(502_504insC)</td>
<td>Unknown</td>
<td>insC</td>
<td></td>
</tr>
<tr>
<td>p(508A)</td>
<td>Sweden</td>
<td>A</td>
<td>C</td>
</tr>
<tr>
<td>p(512T)</td>
<td>Unknown</td>
<td>T</td>
<td></td>
</tr>
<tr>
<td>p(522T)</td>
<td>Japan</td>
<td>T</td>
<td>A</td>
</tr>
<tr>
<td>p(522T)-2</td>
<td>Japan</td>
<td>T</td>
<td>A</td>
</tr>
<tr>
<td>p(545A)</td>
<td>Poland</td>
<td>G</td>
<td>delG</td>
</tr>
<tr>
<td>p(73A)</td>
<td>Japan</td>
<td>A</td>
<td>C</td>
</tr>
<tr>
<td>p(922delC)</td>
<td>Unknown</td>
<td>delC</td>
<td></td>
</tr>
<tr>
<td>p(972_974del)</td>
<td>France</td>
<td>del</td>
<td></td>
</tr>
<tr>
<td>p(1026-1029insC)</td>
<td>Japan</td>
<td>insC</td>
<td></td>
</tr>
</tbody>
</table>

*a* CAGATGCTCCC replaced with TGGACCTGCTGGACCTGCTGGACCTGGAACA, ** ACTGGTACCGCGGCGTGCAGGGGCGCT replaced with CGTACCACGC

<table>
<thead>
<tr>
<th>aa position</th>
<th>Ref.</th>
<th>Acc.no.</th>
</tr>
</thead>
<tbody>
<tr>
<td>23</td>
<td></td>
<td></td>
</tr>
<tr>
<td>37</td>
<td></td>
<td></td>
</tr>
<tr>
<td>81</td>
<td></td>
<td></td>
</tr>
<tr>
<td>96</td>
<td></td>
<td></td>
</tr>
<tr>
<td>97</td>
<td></td>
<td></td>
</tr>
<tr>
<td>100</td>
<td></td>
<td></td>
</tr>
<tr>
<td>101</td>
<td></td>
<td></td>
</tr>
<tr>
<td>140</td>
<td></td>
<td></td>
</tr>
<tr>
<td>157</td>
<td></td>
<td></td>
</tr>
<tr>
<td>158</td>
<td></td>
<td></td>
</tr>
<tr>
<td>169</td>
<td></td>
<td></td>
</tr>
<tr>
<td>183</td>
<td></td>
<td></td>
</tr>
<tr>
<td>187</td>
<td></td>
<td></td>
</tr>
<tr>
<td>187</td>
<td></td>
<td></td>
</tr>
<tr>
<td>219</td>
<td></td>
<td></td>
</tr>
<tr>
<td>245</td>
<td></td>
<td></td>
</tr>
<tr>
<td>251</td>
<td></td>
<td></td>
</tr>
<tr>
<td>251</td>
<td></td>
<td></td>
</tr>
<tr>
<td>257</td>
<td></td>
<td></td>
</tr>
<tr>
<td>261</td>
<td></td>
<td></td>
</tr>
<tr>
<td>301</td>
<td></td>
<td></td>
</tr>
<tr>
<td>301</td>
<td></td>
<td></td>
</tr>
<tr>
<td>329</td>
<td></td>
<td></td>
</tr>
<tr>
<td>344</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

| Change | |
|--------|-----------------
| fs     | V del Y L L fs fs fs stop fs K R D V fs fs S L fs stop fs P T T T T |

* CAGATGCTCCC replaced with TGGACCTGCTGGACCTGCTGGACCTGGAACA, ** ACTGGTACCGCGGCGTGCAGGGGCGCT replaced with CGTACCACGC
In total, 83 samples with the p phenotype were investigated. We found 16 novel mutations and confirmed three of the first four described mutations in the A4GALT gene.25,79 Currently, 23 mutations in 28 alleles have been described by us and others (Table 5). DNA samples from 250 blood donors were screened for four of the five novel missense mutations found in this study in order to determine if they occur among individuals with common phenotypes. However, none of the tested mutations (287G>A, 656C>T, 751C>T and 299C>T) were found in the 500 alleles analysed. In addition, screening for the deletion of TTC (241_243delTTC), found in several individuals with p phenotype of different ethnic origin, revealed no such allele in any of the tested samples with common phenotype. Ten of the new mutations caused frame shifts and premature stop codons. For a summary of the translational consequences of the nonsense mutations, see Table 6.

Two samples with p phenotype investigated in this study showed no critical mutation in the coding region. However, it was not possible to amplify the putative 5’-regulatory region. This might be due to a big insertion or deletion, but in both cases it is likely to be a genetic change disrupting the 5’-regulatory region, thus implicating that this region may indeed be important for expression of P^k antigen. Further studies are needed to clarify the exact cause to why these two samples lack P^k antigen.

The genetic base, 299C>T, for the p phenotype in Amish individuals was discovered and surprisingly, the same mutation was found in an individual of Pakistani origin. Only one other mutation, 241_243delTTC, has been found in two so different populations. The 241_243delTTC mutation was found in samples from Italy, England and Japan. It is difficult to know if these mutations have arisen spontaneously in two completely different parts of the world or if they are examples of ancient mutations, arisen before the divergence of humans into current ethnic groups, although the latter appears to be the most unlikely alternative. A third hypothetical possibility, at least for the 299C>T mutation found in two apparently discrete populations, is a more recent founder gene effect in combination with settlement in two secluded areas or even based on local selective pressure.
Table 6.
A summary of the translational consequences of the nonsense mutations in the A4GALT gene found in this study.

<table>
<thead>
<tr>
<th>Allele</th>
<th>Critical aa change</th>
<th>Open reading frame compared to consensus (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Intact</td>
</tr>
<tr>
<td>68_69insT</td>
<td>Leu23fs53stop</td>
<td>6</td>
</tr>
<tr>
<td>418_428delins</td>
<td>Gln140fs218stop</td>
<td>40</td>
</tr>
<tr>
<td>470_496delins</td>
<td>Asp157fs276stop</td>
<td>44</td>
</tr>
<tr>
<td>473G&gt;A</td>
<td>Trp158stop</td>
<td>45</td>
</tr>
<tr>
<td>502_504insC</td>
<td>Tyr169fs281stop</td>
<td>47</td>
</tr>
<tr>
<td>657delG</td>
<td>Ala219fs349stop</td>
<td>62</td>
</tr>
<tr>
<td>733_734insG</td>
<td>Ile245fs281stop</td>
<td>69</td>
</tr>
<tr>
<td>769delG</td>
<td>Val257fs349stop</td>
<td>72</td>
</tr>
<tr>
<td>902delC</td>
<td>Pro301fs349stop</td>
<td>85</td>
</tr>
<tr>
<td>972_997del</td>
<td>Thr324fs436stop</td>
<td>92</td>
</tr>
</tbody>
</table>

**The B3GALNT1 gene and P expression**

Two sequences, acc.no. Y15062 and ABO50855, both deposited in GenBank, have been used as references for the B3GALNTI gene in Papers I and III.

In Paper I, the gene for a 3-β-N-acetylgalactosaminyltransferase, believed to be the gene for globoside synthase, was DNA sequenced in four individuals with Pk phenotypes (P1k n=1, P2k n=3). Four different mutations were found (Table 7), showing for the first time that crucial mutations in the globoside synthase gene cause globoside deficiency in humans and consequently the Pk blood group phenotype. As a result of this study, the P antigen was removed by the ISBT from the GLOB collection (209) and elevated to constitute a blood group system of its own, namely the GLOB blood group system (028).

In Paper III, we continued to explore the molecular basis of the P1k and P2k phenotypes. Blood samples from ten P1k and one P2k individuals of different geographic and ethnic origins were investigated. DNA sequencing was performed following amplification of the coding regions in the B3GALNTI gene. Four new and two previously described mutations were found (Table 7). One of the samples was drawn from a member of the first described family with P2k phenotype, Mys. This American individual of Finnish origin with the P2k phenotype...
had the 202C>T mutation, previously found in another Finnish person (Paper I). The same mutation was detected in an additional individual, also Finnish, but with P1^k phenotype. This is the first time identical mutations in the B3GALNT1 gene are proven to give rise to two different phenotypes making it absolutely clear that P1/P2 status is dependent on a different gene.

Yet another sample with the P1^k phenotype was investigated afterwards and therefore not included in any of the above papers. The sample was found in our laboratory due to an anti-P in the serum and the phenotype was later confirmed by the International Blood Group Reference Laboratory in Bristol. The B3GALNT1 gene was sequenced and an insertion, 537_538insA, which alters the reading frame and introduces a stop at codon 182, was found. This mutation has previously been found in an English individual of Arabic origin with the P2^k phenotype (Paper I) and here it was detected in a Palestinian woman living in Sweden, indicating that this could be a mutation more common in people of Arabic descent.

P antigen expression levels on RBCs with p, P1^k and common (P1/P2) phenotypes were measured in Paper V. The results are discussed above in the section on The A4GALT gene and P^k expression.

Table 7.
A summary of all mutations found in the B3GALNT1 gene. Roman numerals refer to Papers I and III.

<table>
<thead>
<tr>
<th>Allele</th>
<th>Origin</th>
<th>Consensus</th>
<th>nt. position</th>
<th>Ref.</th>
<th>Acc.no.</th>
</tr>
</thead>
<tbody>
<tr>
<td>P^k(202T)</td>
<td>Finland</td>
<td>C A C A A G G</td>
<td>202 - 203</td>
<td>I, III</td>
<td>AF494103</td>
</tr>
<tr>
<td>P^k(292-293insA)</td>
<td>Italy</td>
<td>insA</td>
<td>292-293</td>
<td>III</td>
<td>AY505344</td>
</tr>
<tr>
<td>P^k(433T)</td>
<td>USA</td>
<td>T</td>
<td>433</td>
<td>III</td>
<td>AY505345</td>
</tr>
<tr>
<td>P^k(537_538insA)</td>
<td>Arabic</td>
<td>insA</td>
<td>537-538</td>
<td>I</td>
<td>AF494104</td>
</tr>
<tr>
<td>P^k(648C)</td>
<td>Canada</td>
<td>C</td>
<td>648</td>
<td>III</td>
<td>AY505346</td>
</tr>
<tr>
<td>P^k(797C)</td>
<td>France</td>
<td>C</td>
<td>797</td>
<td>I</td>
<td>AF494106</td>
</tr>
<tr>
<td>P^k(811A)</td>
<td>Europe</td>
<td>A</td>
<td>811</td>
<td>I</td>
<td>AF494105</td>
</tr>
<tr>
<td>P^k(959A)</td>
<td>Switzerland</td>
<td>A</td>
<td>959</td>
<td>III</td>
<td>AY505347</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>aa position</th>
<th>67</th>
<th>98</th>
<th>145</th>
<th>180</th>
<th>216</th>
<th>266</th>
<th>271</th>
<th>320</th>
</tr>
</thead>
<tbody>
<tr>
<td>Consensus</td>
<td>R</td>
<td>R</td>
<td>R</td>
<td>D</td>
<td>R</td>
<td>E</td>
<td>G</td>
<td>W</td>
</tr>
<tr>
<td>Change</td>
<td>stop</td>
<td>fs</td>
<td>stop</td>
<td>fs</td>
<td>S</td>
<td>A</td>
<td>R</td>
<td>stop</td>
</tr>
</tbody>
</table>

38
To summarize, 16 samples with the P\textsuperscript{k} phenotypes were investigated by sequencing the \textit{B3GALNT1} gene. Eight different mutations were identified causing an inactive $\beta$3GalNAc-T leading to the lack of P antigen and the phenotypes P\textsubscript{1\textsuperscript{k}} or P\textsubscript{2\textsuperscript{k}} (Table 7). DNA samples from 220-250 blood donors were screened for three missense mutations (797A>C, 811G>A and 648A>C) found in this study (Papers I and III), in order to determine whether they occur in individuals with common phenotypes, but none of the mutations were found in any of the alleles analysed. The remaining mutations were nonsense mutations and the degree of enzyme truncation can be seen in Table 8. The mutation 811G>A was found in altogether six P\textsubscript{1\textsuperscript{k}} individuals, five of these with known origin (German n=3, English n=2) and this makes it the most common Caucasian mutation in the \textit{B3GALNT1} gene. Three samples of Finnish origin have been investigated. All were found to have the same mutation, 202C>T. Finland is known to have an overrepresentation of certain recessive disorders (Finnish Disease Heritage) due to national and regional isolation aided by a founder effect.\textsuperscript{142} It is possible that this isolation also explain why the P\textsuperscript{k} phenotype has been described as more common in Finland compared to the rest of the world\textsuperscript{143,144} and that the 202C>T mutation might be the mutation causing a defect $\beta$3GalNAc-T in all Finnish cases. However, the five first serologically described P\textsuperscript{k} families in Finland were all apparently unrelated.\textsuperscript{144}

<table>
<thead>
<tr>
<th>Allele</th>
<th>Critical aa change</th>
<th>Open reading frame compared to consensus (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Intact</td>
</tr>
<tr>
<td>202C&gt;T</td>
<td>Arg67stop</td>
<td>20</td>
</tr>
<tr>
<td>292_293insA</td>
<td>Arg98fs102stop</td>
<td>30</td>
</tr>
<tr>
<td>433C&gt;T</td>
<td>Arg145stop</td>
<td>44</td>
</tr>
<tr>
<td>537_538insA</td>
<td>Asp180fs181stop</td>
<td>54</td>
</tr>
<tr>
<td>959G&gt;A</td>
<td>Trp320stop</td>
<td>96</td>
</tr>
</tbody>
</table>

Table 8.
A summary of the translational consequences of the nonsense mutations in the \textit{B3GALNT1} gene found in this study.
A candidate gene for P1 expression

It has been suggested that the same gene is coding for both the P1 and P^k antigen but no polymorphisms in the coding region of the *A4GALT* gene appeared to explain the P1/P2 phenotypes.\(^{25}\)

Recently, Iwamura *et al.* proposed that instead transcriptional regulation, caused by two different polymorphisms in the 5′-regulatory region of the *A4GALT* gene, might be the reason for the P1/P2 phenotypes.\(^{80}\) To investigate if this hypothesis was correct we screened P1 (n=58) and P2 (n=20) samples for the two mutations -160A>G and -551_-550insC, suggested to cause the null phenotype P2 (*Paper IV*). The majority of the donors were of Swedish origin but a few were of Asian or African descent. Three P1^k, three P2^k and fifteen p samples were also included in the screening.

Furthermore, the 5′-regulatory region including the probable exon 1 and 2, and the 3′-UTR were sequenced, altogether investigating 1075 bp and 1600 bp upstream and downstream of the gene, respectively. Sixteen polymorphic sites were detected but no clear-cut correlation between any of them and the P1/P2 phenotypes was found. Nine of these polymorphisms have not been described before. The two proposed P2-specific mutations, -160A>G and -551_-550insC, were found in homozygous form both in P1 and P2 donors indicating that these mutations are not the cause of the P1/P2 status. However, since the haplotype -551_-550insC;-160G was found in all P2 samples (except one P2^k), the correlation between the *A4GALT* locus and P2 status seems to be rather strong, which is consistent with the chromosomal allocations suggested for P1 and P^k.\(^{25,86}\) Interestingly, multiple polymorphisms exist in and outside the coding region of the *A4GALT* gene. Based on linkage between these SNPs we were able to see five different haplotypic patterns common among Swedish donors. This is in sharp contrast to the *B3GALNT1* gene, in which we have found no SNPs at all, only critical but rare mutations resulting in the P^k phenotypes.

This study could not verify the suggestion by Iwamura *et al.* that the *A4GALT* gene is responsible for both P^k and P1 expression. Our results were later confirmed in a study of 50 English blood donors by Tilley *et al.*\(^{145}\)
Additional work

The LKE antigen is formed by the sequential addition of a Gal and a NeuAc to the P antigen to form galactosylgloboside and LKE, respectively. In addition to RBCs, LKE is expressed on endothelial cells, smooth muscle, kidney, platelets and mesenchymal stem cells. Since a potential LKE gene has not been mapped to a specific chromosome and it is unknown if the LKE-negative phenotype is caused by a mutation in a 3-β-galactosyltransferase or a 3-α-sialyltransferase gene, three genes were analysed. Two different 3-β-galactosyltransferase genes, \textit{B3GALT4} (acc.no. Y15061) and \textit{B3GALT5} (acc.no. NM_006057) along with one 3-α-sialyltransferase gene, \textit{SIATFL} (acc.no. AF059321), were sequenced in samples with both common and LKE-negative phenotypes. No polymorphisms explaining the LKE-negative phenotype was found in any of the investigated candidate genes. Doubts were also raised whether the samples used were really LKE-negative or if they may express the antigen weakly, and therefore the priority of this project was temporarily lowered. Thus, at this point it is unclear if the above mentioned genes or other genes/factors are responsible for the LKE expression and consequently the genetic basis for the LKE-negative phenotype remains unknown. Other investigators appear to be working with these questions but so far no breakthrough has been reported. In a series of abstracts Cooling \textit{et al.} presented data on \textit{B3GALT5} concerning the consequences for Lewis status and other RBC phenotypes including LKE-negative or weak.

Conclusion and future perspectives

In \textbf{Paper I} it was proven that a candidate gene for P expression was really responsible for the globoside-deficient $P_1^k$ and $P_2^k$ phenotypes. Based on this, a new blood group system was born.

In addition, this study showed that the genetic heterogeneity at the glycosyltransferase loci underlying two important null phenotypes of the GLOB blood group system and collection is quite remarkable (\textbf{Papers I, II, III and V}). The knowledge about all these null variants may be important for future applications in transfusion medicine such as blood group typing with DNA microarrays and when serologic tests fail or cannot be performed, for example, if fetal blood group determination is required. Furthermore, our studies showed interesting differences in the $P^k$ and $P$ genes in that the former harbours multiple additional, non-critical polymorphisms whilst the latter is virtually identical apart from the rare null mutations causing the $P_1^k/P_2^k$ phenotypes. The $p$ phenotype is more frequent than the $P^k$ phenotypes and
maybe this is a reflection of the relative frequency, at which mutations occur in these two
genes. The reason behind this difference in degree of polymorphism remains obscure.

One of the aims in this study was also to try and explain the long-standing enigma of why all
individuals with the p phenotype also lack the P1 antigen. The easiest way to explain the
simultaneous loss of Pk and P1 antigen would be that the same gene encodes a transferase able
to transfer Gal both to LacCer and to paragloboside and that a critical mutation can abolish
both Pk and P1 expression. However, the A4GALT gene, which clearly is responsible for
transfer of Gal to LacCer, was thoroughly examined both in the noncoding 5'/3'-regions and
the coding region, and no straightforward correlation to the P1/P2 phenotypes was found.
However, a certain but not complete association to polymorphisms at the A4GALT locus was
found in three separate studies,\textsuperscript{80,145} and Paper IV, which may support that this gene could be
closely linked to a functionally important second gene coding for either a chaperone or
regulatory molecule for the A4GALT gene or α4Gal-T enzyme. Yet another possibility is a
totally different galactosyltransferase but then it is more difficult to explain why p individuals
lack P1. Given that the p phenotype is geographically widespread and caused by many
different mutations in the A4GALT gene it appears unlikely that these people should all have
an additional mutation in a second gene making them P2 and thus lacking the P1 antigen.

Individuals with different blood groups, especially those having null variants, that lack a
certain antigen or even molecule, constitute a great potential since their cells can be used to
determine the binding specificity for pathogens utilising carbohydrate and protein blood
groups for invasion.

Studies of carbohydrate blood groups can lead to a better understanding of the biology of
glycosyltransferases by taking advantage of the naturally-occurring variants to explore the
structural and functional aspects of the protein. Actually, the null variants described here are
natural knock-Outs and cells from individuals with the p and P1\textsuperscript{k} phenotypes, with causative
mutations confirmed in the Pk and P genes, respectively, have recently been used in a study
involving HIV-1. In a previous study it was demonstrated that a soluble mimic of Pk inhibits
HIV infection.\textsuperscript{153} Based on this, individuals whose cells express either no or high levels of Pk
antigen were investigated in a HIV infection model. The lack of Pk antigen encountered in p
individuals increased the susceptibility to HIV-1 infection whilst over-expression of the
antigen, as seen in patients with Fabry disease and P1\textsuperscript{k} individuals, may provide resistance to
infection.\textsuperscript{108,154,155}
The results in Paper V showed that the expression of P^k and P antigens on RBCs varies significantly among donors with the P_1 and P_2 phenotype and it would be interesting to see if the susceptibility for pathogens using these antigens varies to the same degree. This requires that the antigen levels observed on RBCs also translate into varying levels on other tissue cells relevant for the infection studied, e.g. lymphocytes or uroepithelial cells. It would also be important to understand the genetics underlying interindividual variation in P^k/P expression, especially since this may constitute susceptibility markers in the general population, as opposed to the rare null individuals only.
Sammanfattning på svenska

Blodet är en livsnödvändig beståndsdel av vår kropp. Blodet består av röda blodkroppar som transporterar syre, vita blodkroppar som är en del av kroppens immunförsvar, blodplättar som hjälper till med koagulation när skador uppstår samt plasma, den proteinhaltiga vätska som beståndsdelarna flyter i. Intresset för att ge (transfundera) blod till människor vid olika sjukdomstillstånd har funnits länge. Under de senaste århundradena har både transfusioner mellan människor och mellan människor och djur prövats men oftast med dåligt resultat. Det var först när Landsteiner 1900 upptäckte att det finns olika blodgrupper som det blev säkrare att transfundera blod. De upptäckta varianteerna benämndes blodgrupp A, B och O, och ingår i det s.k. ABO-systemet. Genom att vid laboratoriet kontrollera vilken ABO-blodgrupp (och sedan i slutet av 1940-talet också Rh-grupp) patienten har och sedan ge det givarblod och plasma som passar, se Figur A, är blodtransfusioner numera en tillförlitlig process. ABO-systemet är tillsammans med Rh-systemet det kliniskt viktigaste blodgruppssystemet.

Figur A. Bilden visar möjliga ABO-blodgrupp kombinationer vid transfusion av röda blodkroppar och plasma.

Glykosyltransferaser är de enzymer som lägger ett socker till ett annat så en sockerkedja bildas. Människan har 240 olika kända varianter av glykosyltransferas, som var och en har i uppgift att använda en viss sockertyp (kallad nukleotidsocker eller substrat) på ett visst sätt till en bestämd mottagarstruktur (kallad acceptor). Variationer i en gen kan ge upphov till skillnader i antigenen, t.ex. om en individ har blodgrupp A eller B.


Figur B. Bilden visar antigenen Pk, P och P1. Dessa är upptagna av olika sockermolekyler (galaktos- eller glukos-varianten) och sitter på en grundstruktur av socker och fett (prekursor kallad laktosylceramid, LacCer) som i sin tur sitter på ytan av den röda blodkroppen (eller andra celler).
Målet med denna avhandling var att:
- Undersöka vilken gen som kodar för P-antigenet och vilka mutationer som orsakar avsaknad av P.
- Undersöka vilka mutationer i $P^k$-genen som orsakar p-fenotypen (avsaknad $P^k$) hos individer av olika geografiskt och etniskt ursprung.
- Undersöka vilka gener som är ansvariga för uttrycket av antigenen P1 och LKE.

I Artikel I undersökte jag en glykosyltransferasgen som tidigare påvisats av två olika forskningsgrupper. Den ena gruppen trodde sig ha hittat en gen för ett 3-β-galaktosyltransferas medan den andra gruppen trodde att det var en 3-β-N-acetylgalaktosaminyltransferasgen som kunde ge upphov till P-blodgruppsantigenet. Denna gen undersökt hos fyra individer med fenotyperna $P^1_k$ eller $P^2_k$, dvs de saknade P-antigenet, och ett antal normala individer. I alla $P^1_k / P^2_k$ individer fanns det mutationer i genen $B3GALT3$ (senare omdöpt till $B3GALNT1$ med ledning av mitt fynd, men här kallad $P$-genen). P-antigenet tillhörde tidigare GLOB-kollektionen eftersom den genetiska bakgrunden varit oklar men baserat på våra resultat fick P-antigenet ett eget blodgruppssystem, GLOB (ISBT nummer 028). I Artikel III fortsatte jag kartläggningen av olika mutationer i $P$-genen som kan orsaka $P^1_k / P^2_k$-fenotyp. Ytterligare 11 prover undersökte, varvid fyra nya mutationer och två av de tidigare publicerade mutationerna hittades.


För att verifiera att mutationen verkligen ger upphov till ett inaktivt glykosyltransferas, sattes genen (med och utan mutation) in i en syntetisk DNA-ring (vektor) i en Pk-negativ cellinje (Namalwaceller). Uttrycket på cellernas yta, efter att de märkts in med antikroppar mot Pk, mättes sedan med hjälp av en metod som kallas flödescytometri. Amish-individernas gen gav inget uttryck jämfört med de negativa kontrollerna. Pk- och även P-uttrycket mättes på röda blodkroppar från Amish-individer och svenska blodgivare med normala/vanliga blodgruppsfenotyper. Amish-individernas blodkroppar visade inget uttryck medan variationen hos vanliga individer var förvånansvärt stort.

Individer med p-fenotyp saknar förbryllande nog alltid P1-antigen, vilket gett upphov frågan om det är samma glykosyltransferas som kodar för både P1 och Pk, eller om det är två olika genprodukter. Ännu har inte genen som tillverkar P1 hittats men en japansk forskargrupp föreslog, i ett arbete 2003, att två polymorfismar i A4GALT-genens reglerande del (5'-uppströmsregion) skulle vara kopplad till förekomsten av P1-antigenet respektive avsaknad av detsamma. Sjutioåtta svenska blodgivare undersökte och all P2-prover hade dubbel uppsättning (homozygoti) av de s.k P2-mutationerna. Dock hittades även P1-prover som var homozygota för samma mutationer (Artikel IV). Därmed blir slutsatsen att de föreslagna P2-polymorfismerna inte enbart kan vara de som styr P1/P2-blodgruppsuttrycket.

Sammantaget har den bakomliggande orsaken till ett flertal Pk/P/P1-relaterade blodgrupper utretts. Detta gör att vi idag vet lite mer om genetiken som styr dessa kliniskt viktiga skillnader mellan olika individer. Kunskapen kan bl.a. utnyttjas för vidare utveckling av DNA-baserad blodgruppsbestämning. Informationen har också lett till fördjupad kunskap kring glykosyltransferasers funktion men mycket återstår att utreda. Det vore intressant att fortsätta studier av samspelet mellan Pk/P1-antigenen och olika virus eller bakterier. I ett samarbete med en kanadensisk forskargrupp har vi, utanför denna avhandlings ramar, nyligen funnit att blodgruppen P1 k verkar skydda mot HIV-infektion medan celler från individer med blodgruppen p är särskilt känsliga för HIV.
Acknowledgements

This thesis would not have been possible without the help and support of many people. I would like to express my gratitude to all of you, who in one or another way have been involved in this work.

Thanks to….

Professor Martin L Olsson, my supervisor, for introducing me to PCR and genes (12 years ago!), for giving me the opportunity to become part of your research group, for your never-ending enthusiasm for new projects, and for excellent guidance throughout this particular project.

Britt Thuresson, for being my mentor in the lab, teaching me everything from how to count cells to defrosting freezers, for interesting gardening conversations and very importantly, for helping me with my mathematical problems.

Elisabet Sjöberg Wester, for being a great person to collaborate with both in science and in the clinical field, for all valuable discussions (including those about house renovation), for arranging the exhausting but always fun Tosteberga-retreats, and for designing the DNA-picture on the cover of this book.

Jill Storry, for giving our lab an international atmosphere, for generously giving of your knowledge and contacts all over the world, and for being such fun company on trips.

Alan Chester, for being an invaluable guide into the complex world of carbohydrates and biochemistry, and for reviewing this manuscript.

Nidal Irshaid, for teaching me sequencing and for all stimulating discussions we had in the PCR room about religion and politics.

Forskningsgruppen, for all creative discussions, for giving great input to my presentations, and for being such great travel partners and cooks!

The people at NRLGBT, both the old and new ones, Pia, Helen, Annika, Lis and Robban, for being great people to work with.

All the colleagues at Blodcentralens lab, for your patience when I started to “escape” to the PCR room, for your encouragement and help over the years.

Gunilla Hochbergs, for many thoughtful discussions about life outside work and, together with Li and Alicja, for being such good roommates, constantly filling the candy bowl (and not complaining when it was empty after the weekend!) and for being extremely flexible with both the use of the sequencing machine and Bahram’s working hours.

LIT lab, for all nice chats, mumsiga kakor and for a fantastic celebration on my birthday.
All other personnel at the Blood Centre, for nice chats and encouraging comments.

All co-authors, in particular Anne-Christine Schmidt-Melbye, Marion Reid and Joyce Poole, for your contribution with interesting samples and expert knowledge.

Joelle Wiels, for giving me CD77-antibodies.

Charlie, Hanna and Gerd at the Clinical Chemistry department, for all your help with the sequencing and real-time PCR machines.

Karina, Susanna, Bodil, May-Louise and all the others at the Hematology department for always making me feel welcome, for always giving your time to explain and help me find things. Marcus Hansson, for giving me the GFP vector and Ellinor for finding it!

Anders Tingström, for all your good advice both about science and life, nice dinners and supportive chats during the past years.

All my friends, especially Ingela, Inga, Eva, Marie, Gigge, Maria, Sofia and Örjan for always being there when I needed encouragement and someone to talk to, and Ramtin, our phone, computer and barbeque master.

Last but not least, my family for always being there. Ett extra stort tack till mina föräldrar för att ni tagit hand om Arvin under denna sista intensiva period.

Arvin, mitt hjärtegryn, vår klump, för att du ger livet en extra dimension.

Bahram, this would not have been feasible without you, your patience and love has made it possible to complete this project. Apart from being the man in my life, you have also been a wonderful colleague and discussion partner with your sometimes wild and not so conventional ideas.

to همانند خورشیدی ؛ هرگاه که به تو بیندیشم روزم مذور است... دوست دارم!
References


35. Hosseini-Maaf B, Smart E, Chester MA et al. The Abantu phenotype in the ABO blood group system is due to a splice-site mutation in a hybrid between a new O1-like allelic lineage and the A2 allele. Vox Sang 2005;88:256-64.


47. Crawford MN, Tippett P, Sanger R. Antigens Aua, i and P1 of cells of the dominant type of Lu(a-b-). Vox Sang 1974;26:283-7.


53


145. 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;90:198-203.


149. Gang EJ, Bosnakovski D, Figueiredo CA et al. SSEA-4 identifies mesenchymal stem cells from bone marrow. Blood 2006;..

150. Cooling LL, Gu Y, Judd WJ. A Missense mutation in B3GalT5, the glycosyltransferase responsible for galatosylgloboside and Lewis c synthesis, may be associated with the LKE-weak phenotype in African Americans. Transfusion 2002;42:9S.

151. Hwang D, Cooling LL. Homozygosity for the galatosylgloboside synthase (B3GalT5)-T654 allele is associated with decreased LKE expression on LKE-weak RBC. Transfusion 2003;43:6A.

152. Cooling LL, Hwang D, Shayman JA. Upregulation of Luke (LKE) by B1,3 galactosyltransferase V (B3GalT5) and 5’ Azacytidine. Transfusion 2005;45:6A.


Appendix: Papers I-V