

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

Hellberg, Åsa

2007

# 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

LUND
UNIVERSITY

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

ISBN 978-91-85559-15-2
Printed by Media-Tryck, Lund, Sweden 2007

*Till min familj*

# Table of contents

*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.
Molecular Basis of the Globoside-Deficient $P^k$ Blood Group Phenotype. Identification of Four Inactivating Mutations in the UDP-*N*-acetyl-galactosamine:globotriasyl-ceramide 3-β-*N*-acetylgalactosaminyltransferase Gene.
*Journal of Biological Chemistry*, 2002;277(33):29455-9.

**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.
*Transfusion,* 2003;43(7):899-907.

**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.
*British Journal of Haematology,* 2004;125(4):528-536.

**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.
*BMC Genetics,* 2005 7;6:49 (11 pages, doi: 10.1186/1471-2156-6-49).

**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*).

# Abstract

The clinically important carbohydrate P/GLOB blood group systems and collection give rise to both common ($P_1$, $P_2$) and rare (p, $P_1^k$, $P_2^k$) 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 $P^k$, P and P1 antigen expression.

Sequence analysis of the *A4GALT* and *B3GALNT1* genes proposed to synthesize the related $P^k$ ($Gb_3$) and P ($Gb_4$) antigens, respectively, were performed in p and $P^k$ 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 $P^k$-negative Namalwa cells transfected with mutated *A4GALT*-constructs showed $P^k$ expression levels comparable to negative controls. RBCs with p phenotype showed no $P^k$ and P activity while both P and $P^k$ 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 $P_1/P_2$ phenotypes. No clear-cut correlation was found and two previously proposed $P_2$-specific mutations were detected in homozygous form both in $P_1$ and $P_2$ 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 differention |
| EC | Enzyme commission |
| EHEC | Enterohemorrhagic *Escherichia coli* |
| ER | Endoplasmic reticulum |
| Gal | Galactose |
| GalNAc | *N*-acetylgalactosamine |
| GlcNAc | *N*-acetylglucossamine |
| 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 transfusion[1] 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.[7] 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.[8]

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.[8]

In addition, a database, The Blood Group Antigen Gene Mutation Database (http://www.ncbi.nlm.nih.gov/projects/mhc/xslcgi.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).[9] 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.[10] 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]

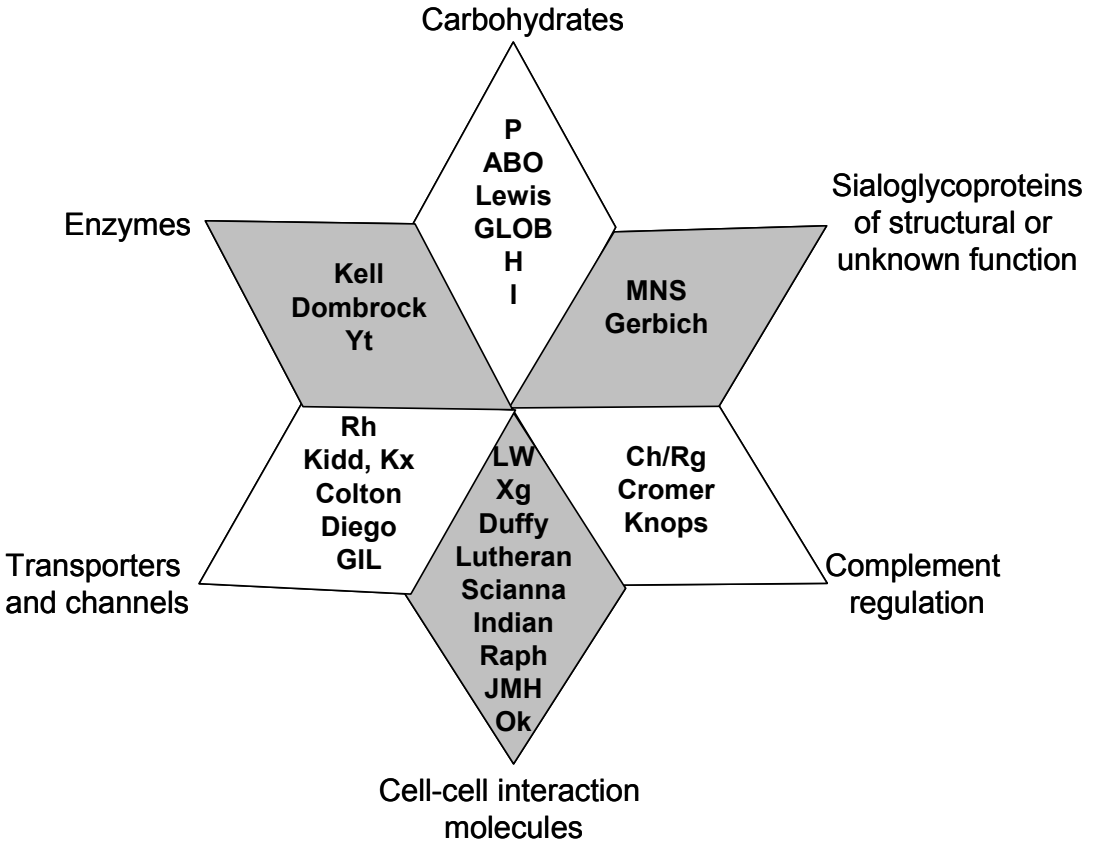



**Figure 1.**
**Picture showing the blood group systems sorted according to their proven or hypothesized functions**.
The figure is modified from an original kindly provided by JR. Storry.

***Glycosyltransferases***

There are more carbohydrates on earth than any other type of biomolecule.[14] 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[15], Paulson & Colley[16] and Kleene & Berger.[14]

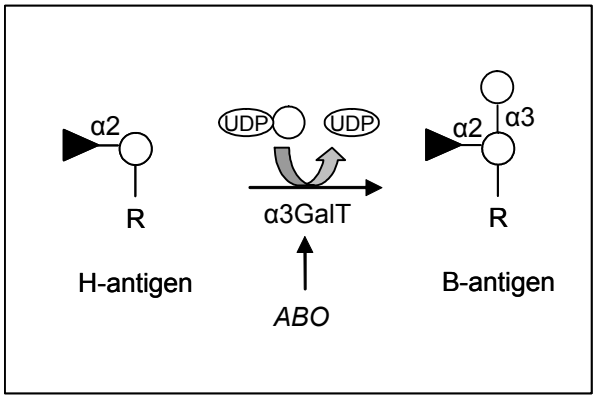



**Figure 2.**
**Schematic picture of a glycosyltransferase adding a monosaccharide to a precursor**. More precisely, the transfer of a galactose by the B transferase, 3-α-galactosyltransferase, to the precursor, H antigen, forming the B antigen.

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.[14] GTs are also classified into families by amino acid (aa) sequence similarities, listed in the CAZy database (http://afmb.cnrs-mrs.fr/CAZY).[17] X-ray

crystallography and *in silico* studies have shown that many glycosyltransferases belong to two structural fold superfamilies, GT-A and GT-B.[18] A recent report of the structure for a bacterial sialyltransferase describes a new fold family but still there are many glycosyltransferases not assigned to any of the known fold families.[18]

The expression of glycosyltransferases is highly regulated during differentiation and proliferation and also specific for different tissues and cells.[14] 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**).[16]

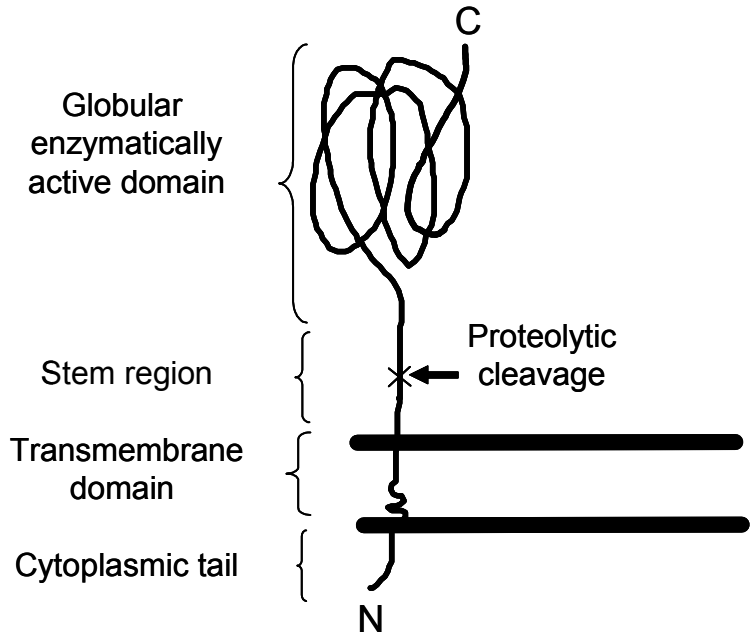



**Figure 3.**
**Golgi glycosyltransferases are type II transmembrane proteins**.
Schematic figure modified from Paulson & Colley.[16]

### *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.[19] 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,[20] the genetic code solved,[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.[22] The same year the sequence of the first human gene (for a polypeptide hormone)[23] was determined and nine years later the first gene encoding a blood group antigen was found. This gene was called *GYPA* and encodes the MN antigens in the MNS system.[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 $P^k$ and $Sd^a$ glycosyl-transferases still remain to be clarified.[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[30] where mutations in the gene give rise to two different GTs, 3-α-*N*-acetylgalactosaminyltransferase and 3-α-galactosyltransferase. These enzymes transfer *N*-acetylgalactosamine (GalNAc) or galactose (Gal), producing the A or B antigens, respectively.[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.[32,33] The molecular basis of many of the rare ABO subgroups such as $A_x$, $A_{finn}$, $A_{bantu}$, $A_{el}$, $B_3$, $B_x$, $B_{el}$, B(A) and *cis*AB has also been solved.[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.[5] In 1955 Sanger found the $P^k$ antigen,[39] and four years later Matson *et al*. found the P antigen.[40] A fourth related antigen, Luke (LKE), was described by Tippett *et al*. in 1965.[41]

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.[42] Combinations of these antigens give rise to the following phenotypes; $P_1$, $P_2$, $P_1^k$, $P_2^k$ and p. In addition, $P_1$ and $P_2$ individuals can be LKE positive or negative.

**Table 1**.
The P /GLOB blood group systems and collection with ISBT numbers in brackets.

| Antigen | Blood group system | Blood group collection |
|---|---|---|
| P1 | P (003) | - |
| P | GLOB (028) | - |
| $P^k$, LKE | - | GLOB(209) |

The frequency of the $P_1$ phenotype varies between different ethnic groups, ranging from 90% among Africans to 20% in Asians.[43] On RBCs, P1 expression changes during fetal development. The antigen is found as early as week 12 but weakens during gestation.[44] At birth the expression is low and it takes up to seven years before full expression is reached.[45] The strength of the antigen expression can differ from one person to another and it seems to be dependent on dosage.[46] 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.[47] 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,[48] 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.[49] Individuals with the LKE-negative phenotype express more $P^k$ antigen compared to individuals with LKE-positive phenotype.[50] The P antigen is well developed at birth[51] and the most abundant neutral glycolipid in the RBC membrane with $14x10^6$ antigens per cell.[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 *et al*. described individuals with either a weak P or weak $P^k$ antigen.[52,53]

The frequency of the p phenotype has been estimated at 5.8 per million in Europeans,[51] but for Swedes in Västerbotten county in Northern Sweden the number of p individuals is significantly higher (141 per million).[54] The phenotype also seems to be more common in Japan and among Amish people.[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.[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.[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.[51] Actually, DNA from the first $P^k$ individual, Mrs. Mys., is included in this study (**Paper III**).

**Table 2**.
A summary of phenotypes and possible antibodies for the P/GLOB blood groups.

| Phenotype | Frequency | Antigen present on RBC | Antibodies in serum |
|---|---|---|---|
| $P_1$ | 20-90% | P1, $P^k$, P | none |
| $P_2$ | 10-80% | $P^k$, P | Anti-P1* |
| p | rare | None | Anti-PP1$P^k$ |
| $P_1^k$ | rare | P1, $P^k$ | Anti-P |
| $P_2^k$ | rare | $P^k$ | Anti-PP1 |

* not always present/detectable.

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.[59-61]

The P and $P^k$ antibodies can cause HTR but causing HDN has not been reported. However, early spontaneous abortion has a higher frequency among women with p and $P_1^k$ /$P_2^k$ phenotypes and is a phenomenon that is most likely due to the IgG component of the anti-P attacking the placenta.[62,63] The anti-PP1$P^k$ found in individuals with the p phenotype was previously called anti-$Tj^a$, named after Mrs. Jay in whose serum this antibody specificity was first found in association with a tumor.[64]

Less than 2% of Caucasians are negative for LKE, hence this is a high-incidence antigen.[41] It is also known as the stage-specific embryonic antigen (SSEA-4).[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.[66,67] To date, this antibody specificity has not been implicated in HTR or HDN.[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*.[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.[71]

The antigens $P^k$, 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.[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.[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.[74] $P^k$ (globotriaosylceramide, Gb3, ceramide trihexoside, CD77), P ($Gb_4$, 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.[75,76] 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.[77] The biochemical, serological and clinical aspects of the P-related blood group systems have been thoroughly reviewed by Spitalnik and Spitalnik.[78]

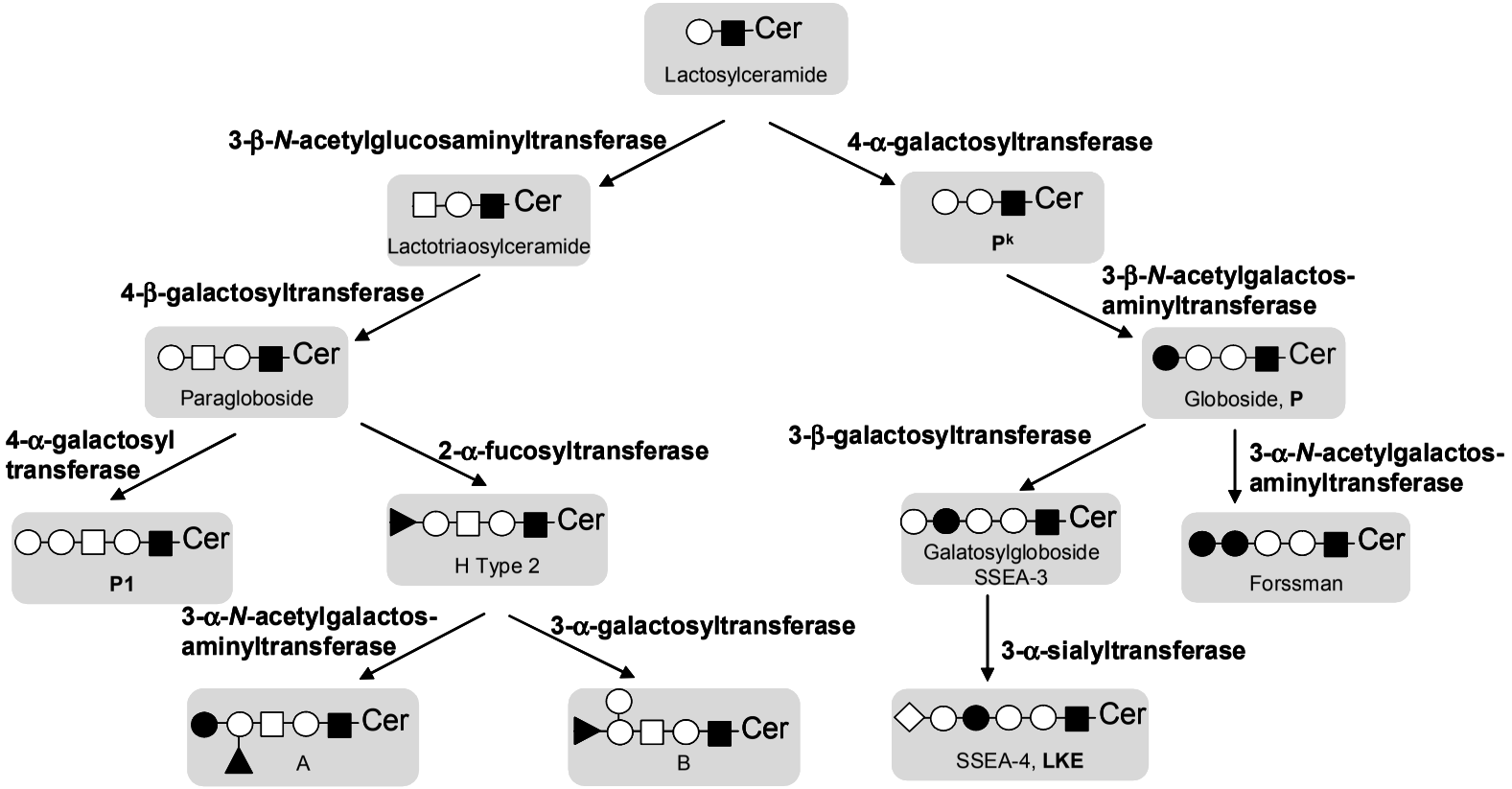



**Figure 4.**
**Scheme showing the biosynthesis of $P^k$, P and P1 antigens**.
Some other related structures, such as the blood group A and B antigens, are also included.

■ = Glucose
▶ = Fucose
o = Galactose
● = *N*-acetylgalactosamine
□ = *N*-acetylglucosamine
◊ = Sialic acid

***Genetics***

*The* $P^k$ *gene*

The gene for the 4-α-galactosyltransferase (α4Gal-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 α4Gal-T were also found.[25]

*The* P *gene*

Another gene, first cloned in 1998 as a member in the 3-β-galactosyltransferase family[83] but later shown to be a 3-β-*N*-acetylgalactosaminyltransferase, 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 (β3GalNAc-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]

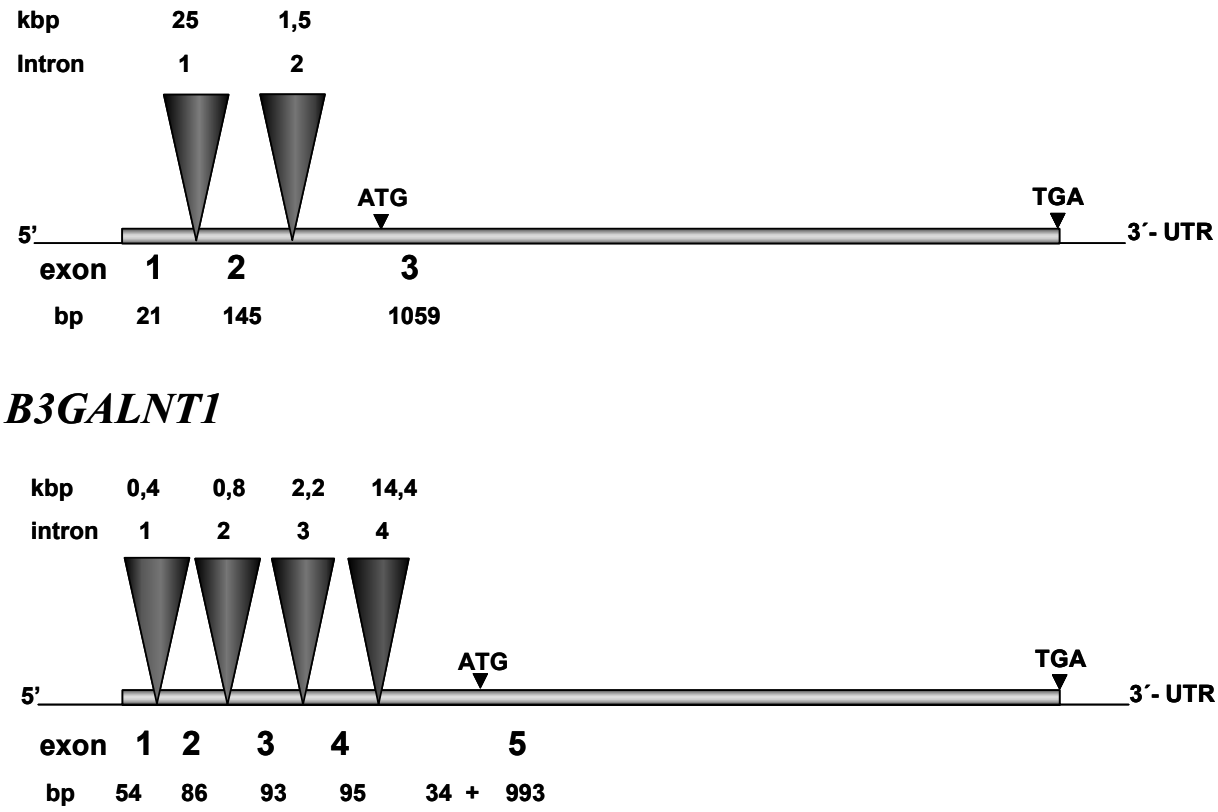



**Figure 5.**
**The genomic organisation of the $P^k$ (*A4GALT*) and *P* (*B3GALNT1*) genes** (not drawn to scale). The *A4GALT* gene is shown with three exons even if it is not certain if the gene consists of two or three exons.

*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 $P_1$ kidneys could synthesize both $P_1$ and $P^k$ while enzymes from $P_2$ kidneys only could produce $P^k$.[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.[89] However, no polymorphisms in the coding region of the *A4GALT* gene appear to explain the $P_1/P_2$ phenotypes.[25]
For synthesis of the LKE antigen, a 3-α-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-β-galactosyl-transferase gene catalysing the previous step in the biosynthetic chain.

***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.[26,90,91] Studies of mouse tissues shows expression patterns similar to humans although there are some differences.[90]
The P1 structure is found as glycolipids and/or glycoproteins in many organisms such as the nematode (*Ascaris suum)*, tapeworm (*Echinococcus granulosus)*, earthworms (*Lumbricus terrestris*), liver flukes (*Fasciola hepatica)*, bacteria (*Neisseira gonorrrhoe)* and pigeon.[66] The $P^k$ antigen is also expressed in several strains of bacteria.[92]

In humans, glycosphingolipids can be useful as surface markers of normal erythrocyte differentiation and of erythroleukemias.[93] The $P^k$, P and P1 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[61,94] but no such reports about P1 in plasma or about $P^k$, P and P1 in secretions are available.
The P1 antigen is expressed on lymphocytes, granulocytes and monocytes.[95] The $P^k$ antigen has been found on granulocytes, monocytes,[95] fibroblasts[48], platelets, smooth muscle cells of the digestive tract and urogenital system[96] and is a differentiation antigen expressed on a subset of tonsillar B cells in the germinal center.[97] High expression of $P^k$ in the kidney has been implicated in susceptibility to hemolytic uremic disease (HUS), further discussed in the *Disease associations* section below. The mechanism behind the high renal expression might be due to enhanced *A4GALT* gene transcription and reduced α-galactosidase gene transcription.[82] The $P^k$ antigen is also known as Burkitt lymphoma antigen and has been classified as CD77.[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 megakarocytes, fibroblasts but not lymphocytes and granulocytes according to van dem Borne *et al*. while Shevinsky *et al*. could not detect P on fibroblasts.[99,100] In another study P was detected in 11 of 16 investigated tissues, especially in those of mesodermal origin.[101] The P antigen can also be found in placenta.[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.[102]

Northern blot studies of human organs showed high expression of the *A4GALT* gene in kidney and heart in one study[25] whilst another one found, in addition to kidney and heart, high expression in spleen, liver, testis and placenta.[27] High expression of the *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.[84] No studies on P1 have been performed since the identity of the gene has not yet been clarified.

***Disease associations***

The $P^k$, 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.[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.[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, lyses autologous P-positive erythrocytes.[106] Some data also suggest that the $P^k$ antigen is another co-factor for the human immunodeficiency virus to enter CD4-positive cells[107] but a recent report proposes the opposite.[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 $P^k$, 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 $P^k$ as receptors.[112,113]

A disease connected to the $P^k$ antigen is Fabry disease where deficiency of the lysomal enzyme α-galactosidase A causes accumulation of sphingolipids, mainly $P^k$, in some cell types and body fluids.[114] A recent study shows that mice wih Fabry disease are protected against the verotoxin (VT) from enterohemorrhagic *Escherichia coli* (EHEC).[115] These data are surprising since $P^k$ 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 $P^k$ can work as a toxin sink, which allows the toxin to bind to $P^k$ 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 $P^k$ in immature megakaryoblasts and induces their apoptosis, leading to the restraint of platelet production in the bone marrow.[120] The $P^k$ antigen has been shown to mediate apoptotic signals following the binding of both verotoxin and anti-$P^k$ (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 $P^k$ glycolipid in their sera compared to a healthy control group.[122] These authors propose that during infection circulating VTs should bind to $P^k$ glycolipids in sera which may reduce the amount of VTs binding to the target cells. Consequently, patient with low serum levels of $P^k$ would have a higher susceptibility to EHEC infections. Another study states that only $P^k$ and not P1, as earlier believed, is the receptor for VTs, and mice without $P^k$ 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 $P^k$, P and P1 antigens) who had a gastric tumour which expressed P antigen. Levine proposed that the antibodies made against the $P^k$, P and P1 antigens prevented further growth of the tumour.[125] Expression of $P^k$ 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 $P^k$, P and P1 antigens.

| Pathogen/toxin | Disease | Antigen involved | Reference |
|---|---|---|---|
| **Virus** | | | |
| Parvovirus B19 | Fifth disease | P | 103,104 |
| HIV | AIDS | $P^k$ | 107,108 |
| **Bacteria** | | | |
| Uropathogenic *Escherichia coli* | UTI | $P^k$, P, P1 | 109-111 |
| *Streptococcus suis* | Meningitis | $P^k$, P1 | 112 |
| **Toxin/lectin from** | | | |
| *Shigella dysenteriae* (Shiga toxin) | Dysentery | $P^k$ | 118 |
| *Escherichia coli* O157 (VT1/2) | HUS, hemorrhagic colitis | $P^k$ | 129 |
| *Escherichia coli* (VT2e) | Pig edema disease | P | 130 |
| *Pseudomonas aeruginosa* (PA-IL lectin) | Opportunistic human pathogen | $P^k$, P1 | 113,131 |

# 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 ($P_1$ and $P_2$) and samples with the rare phenotypes p, $P_1^k$ and $P_2^k$ 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 $P_1^k$ 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 $P_1^k/P_2^k$ phenotypes were investigated in this study.

| **p phenotype** | n |
|---|---|
| Argentina | 1 |
| Brazil | 1 |
| England | 1 |
| France | 1 |
| Israel | |
| Arabs | 2 |
| North African Jews | 4 |
| India | 1 |
| Italy | 5 |
| Japan | 2 |
| Norway | 2 |
| Pakistan | 1 |
| Poland | 1 |
| Sweden | |
| Västerbotten | 29 |
| Skåne | 1 |
| Thailand | 2 |
| USA | |
| Amish | 19 |
| Mennonite | 1 |
| Unknown | 9 |
| Total | **83** |

| **$P_1^k/P_2^k$ phenotype** | n |
|---|---|
| Arabic origin | |
| Palestine | 1 |
| Unknown | 1* |
| Canada | 1 |
| England | 2 |
| Finland | 3* |
| France | 1* |
| Italy | 1 |
| Germany | 3 |
| Switzerland | 1 |
| USA | 1 |
| Unknown | 1 |
| Total | **16** |

* One or two of these were $P_2^k$ individuals.

To avoid duplicates in this study when two or more samples showed the same critical p or $P^k$ 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-P1PP$^k$ and p phenotype was found in 1951 in the USA.[64] Around the same time, the first Swedish individual was discovered but not until a few years later was the correct phenotype assigned.[132] 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.[54] Since then, the regional blood bank in Umeå has provided rare units to numerous patients with anti-P1PP$^k$ 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.[133] 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.[134]
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.[135] The estimated coefficient of inbreeding for the Amish population is 0.0151, which is approximately equivalent to having second cousins as parents.[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 **Paper V**.

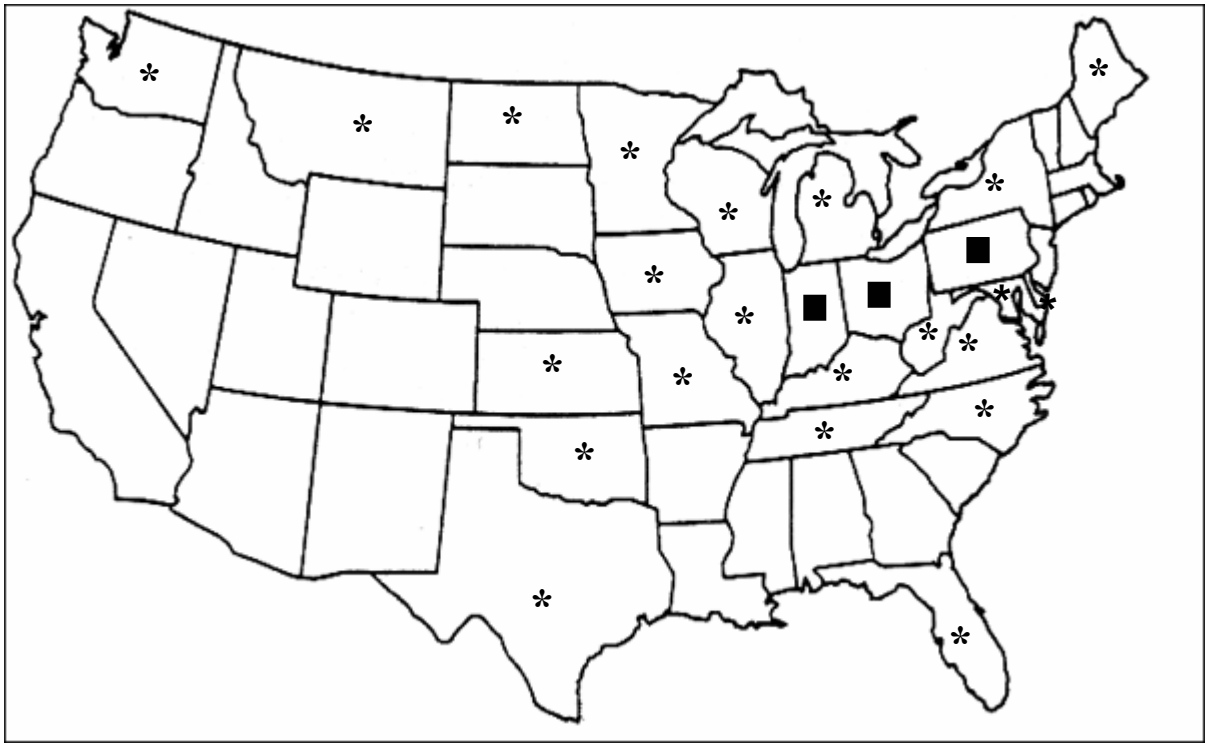


**Figure 6.**
**Map showing the distribution of Amish settlements in the United States**. Asterisks indicate states with a significant group of Amish people. The filled squares specify the three states (Ohio, Pennsylvania and Indiana) with the largest Amish populations, which corresponds to the states where the investigated samples originate from.

***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.[137] 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*

$P^k$-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 $P^k$-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 $P^k$-negative erythropoietic cell line is not known. Besides, most previous expression studies have been done in Namalwa or L cells.[25,79,80] 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 (Amaxa, 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 $P^k$ antigen on RBCs and Namalwa cells (**Paper V**).

Several monoclonal antibodies against the $P^k$ 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 $P^k$ antigen on transfected Namalwa cells and 1 hour at 4˚C for

detection of the $P^k$ 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* $P^k$ *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 $P_1^k/P_2^k$ 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 $P_1^k/P_2^k$ 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 $P_1^k$ 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 $P^k$ activity.[79] 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 $P^k$ and P antigens (**Figure 7**). As expected no expression compared to the negative controls (RBCs with $p^{548A}$ and $P_1{}^k$ phenotype, respectively) was found.

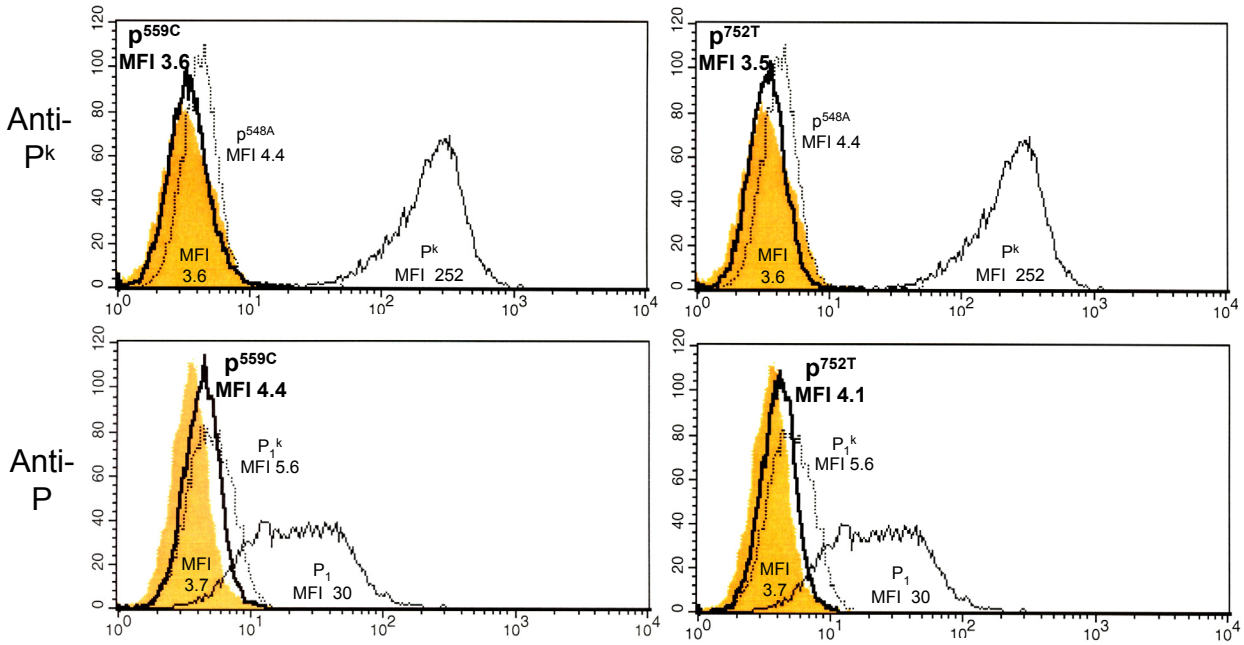



**Figure 7.**
**Cell surface expression of $P^k$ and P antigen on RBCs with the p phenotype.**
Flow cytometry histograms showing $P^k$ (top row) and P (bottom row) expression on RBCs as measured by monoclonal antibodies 5B5 and AME-2, respectively. The filled grey histograms illustrate unlabeled RBCs (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 $P_1$ phenotype, top and bottom histograms, respectively. The bold black curves show RBCs from the Thai individual ($p^{559C}$, left panel) and the Argentinian individual ($p^{752T}$, right panel). The dotted curves illustrate the negative controls; RBCs with $p^{548A}$ phenotype or RBCs with $P^k$ 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**.

| | **nt position** | **68-69** | **109** | **241-243** | **287** | **290** | **299** | **301** | **418-428** | **470-496** | **473** | **502-504** | **548** | **559** | **560** | **656** | **657** | **733-734** | **751** | **752** | **769** | **783** | **902** | **903** | **987** | **972-997** | **1026-1029** | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | **Consensus** | **G** | **A** | **TTC** | **G** | **C** | **C** | **G** | * | ** | **G** | **CCC** | **T** | **G** | **G** | **C** | **G** | **A** | **C** | **C** | **G** | **G** | **C** | **G** | **G** | | **C** | | |
| **Allele** | **Origin** | | | | | | | | | | | | | | | | | | | | | | | | | | | **Ref.** | **Acc.no.** |
| ***p(68_69insT)*** | Israel | insT | G | | | | | | | | | | | | | | | | | | | | | | A | | | **III** | AY496226 |
| ***p(241_243delTTC)-1*** | Japan/Engl | | | del | | | | | | | | | | | | | | | | | | | | | | | | [140] **II** | AF513327 |
| ***p(241_243delTTC )-2*** | Italy | | | del | | | | | | | | | | | | | | | | | | | | C | | | | **II,III** | AF513326 |
| ***p(287A)*** | Italy | | | | A | | | | | | | | | | | | | | | | | | | C | | | | **III** | AY496229 |
| ***p(290T )*** | Poland | | G | | | T | | | | | | | | | | | | | | | | | | | A | | | **II** | AF513323 |
| ***p(299T)-1*** | Amish | | | | | | T | | | | | | | | | | | | | | | | | | | | | **V** | EF192053 |
| ***p(299T)-2*** | Pakistan | | | | | | T | | | | | | | | | | | | | | | | | C | | | | **V** | EF192054 |
| ***p(301delG)*** | China | | | | | | | delG | | | | | | | | | | | | | | | | | | | | [141] | - |
| ***p(del-ins)-1*** | Asian | | | | | | | | del/ins | | | | | | | | | | | | | | | | | | | **II** | AY496232 |
| ***p(del-ins)-2*** | England | | | | | | | | | del/ins | | | | | | | | | | | | | | C | | | | **III** | AF513325 |
| ***p(473A)*** | Brazil | | | | | | | | | | A | | | | | | | | | | | | | C | | | | **III** | AY496227 |
| ***p(502_504insC)*** | Unknown | | | | | | | | | | | insC | | | | | | | | | | | | C | | | | **III** | AY496231 |
| ***p(548A)-1*** | Sweden | | | | | | | | | | | | A | | | | | | | | | | | | | | | [25,79] **II,III** | AY166861 |
| ***p(548A)-2*** | Sweden | | | | | | | | | | | | A | | | | | | | | | | | | A | | | [25] | AY166862 |
| ***p(559C)*** | Thailand | | | | | | | | | | | | | C | | | | | | | | | | | | | | **unpubl.** | EF217316 |
| ***p(560A)*** | Sweden | | | | | | | | | | | | | | A | | | | | | | | | | | | | **II** | AY166863 |
| ***p(656T)*** | France/unk | | | | | | | | | | | | | | | T | | | | | | | | | | | | **III** | AY496230 |
| ***p(657delG )*** | Israel | | | | | | | | | | | | | | | | delG | | | | | | | C | | | | **II,III** | AY166864 |
| ***p(733_734insG)*** | Norway | | | | | | | | | | | | | | | | | insG | | | | | | C | | | | **II** | AF513328 |
| ***p(751T)*** | Unknown | | | | | | | | | | | | | | | | | | T | | | | | | | | | **II** | AY496228 |
| ***p(752T)-1*** | Japan | | | | | | | | | | | | | | | | | | | T | | | | C | | | | [79] **III** | EF217317 |
| ***p(752T)-2*** | Japan | | G | | | | | | | | | | | | | | | | | T | | | | | A | | | [79] | - |
| ***p(769delG)*** | Poland | | | | | | | | | | | | | | | | | | | | delG | | | C | | | | **II** | AF513324 |
| ***p(783A)*** | Japan | | | | | | | | | | | | | | | | | | | | | A | | | | | | [79] | - |
| ***p(902delC)*** | Unknown | | G | | | | | | | | | | | | | | | | | | | | delC | | | | | **III** | AY496233 |
| ***p(972_997del)-1*** | France | | G | | | | | | | | | | | | | | | | | | | | | | | del | | **III** | AY496234 |
| ***p(972_997del)-2*** | USA | | | | | | | | | | | | | | | | | | | | | | | C | | del | | **III** | AY496234 |
| ***p(1026-1029insC)*** | Japan | | | | | | | | | | | | | | | | | | | | | | | C | | | insC | [140] **II,III** | AF513329 |
| | **aa position** | 23 | 37 | 81 | 96 | 97 | 100 | 101 | 140- | 157- | 158 | 169 | 183 | 187 | 187 | 219 | 219 | 245 | 251 | 251 | 257 | 261 | 301 | 301 | 329 | 324- | 344 | | |
| | **Consensus** | L | M | F | C | S | S | A | Q- | D- | W | Y | M | G | G | A | A | I | P | P | V | W | P | P | T | T- | T | | |
| | **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 | fs | fs | | |

* CAGATGCTCCC replaced with TGGACCTGCTGGACCTGCTGGACCTGCTGGAACA, ** ACTGGTACGCGGCCGTGCAGGGGCGCT replaced with CGTACCCGAC

**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.

| Allele | Critical aa change | Open reading frame compared to consensus (%): | |
|---|---|---|---|
| | | Intact | Total |
| 68_69insT | Leu23fs53stop | 6 | 15 |
| 418_428delins | Gln140fs218stop | 40 | 58 |
| 470_496delins | Asp157fs276stop | 44 | 80 |
| 473G>A | Trp158stop | 45 | 45 |
| 502_504insC | Tyr169fs281stop | 47 | 80 |
| 657delG | Ala219fs349stop | 62 | 99 |
| 733_734insG | Ile245fs281stop | 69 | 80 |
| 769delG | Val257fs349stop | 72 | 99 |
| 902delC | Pro301fs349stop | 85 | 99 |
| 972_997del | Thr324fs436stop | 92 | 124 |

### *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 $P^k$ phenotypes ($P_1^k$ n=1 , $P_2^k$ 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 $P^k$ 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).[42]

In **Paper III**, we continued to explore the molecular basis of the $P_1^k$ and $P_2^k$ phenotypes. Blood samples from ten $P_1^k$ and one $P_2^k$ individuals of different geographic and ethnic origins were investigated. DNA sequencing was performed following amplification of the coding regions in the *B3GALNT1* 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 $P_2^k$ phenotype, Mys.[40] This American individual of Finnish origin with the $P_2^k$ 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 $P_1^k$ 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 $P_1/P_2$ status is dependent on a different gene.

Yet another sample with the $P_1^k$ 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 $P_2^k$ 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, $P_1^k$ and common ($P_1/P_2$) 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**.

| | **nt. position** | **202** | **292-293** | **433** | **537-538** | **648** | **797** | **811** | **959** | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | **Consensus** | **C** | **A** | **C** | **A** | **A** | **A** | **G** | **G** | | |
| **Allele** | **Origin** | | | | | | | | | **Ref.** | **Acc.no.** |
| ***P^k (202T)*** | Finland | **T** | | | | | | | | **I, III** | AF494103 |
| ***P^k(292_293insA)*** | Italy | | **insA** | | | | | | | **III** | AY505344 |
| ***P^k(433T)*** | USA | | | **T** | | | | | | **III** | AY505345 |
| ***P^k(537_538insA)*** | Arabic | | | | **insA** | | | | | **I** | AF494104 |
| ***P^k(648C)*** | Canada | | | | | **C** | | | | **III** | AY505346 |
| ***P^k(797C)*** | France | | | | | | **C** | | | **I** | AF494106 |
| ***P^k(811A)*** | Europe | | | | | | | **A** | | **I** | AF494105 |
| ***P^k(959A)*** | Switzerland | | | | | | | | **A** | **III** | AY505347 |
| | **aa position** | **67** | **98** | **145** | **180** | **216** | **266** | **271** | **320** | | |
| | **Consensus** | **R** | **R** | **R** | **D** | **R** | **E** | **G** | **W** | | |
| | **Change** | **stop** | **fs** | **stop** | **fs** | **S** | **A** | **R** | **stop** | | |

**To summarize**, 16 samples with the $P^k$ phenotypes were investigated by sequencing the *B3GALNT1* gene. Eight different mutations were identified causing an inactive β3GalNAc-T leading to the lack of P antigen and the phenotypes $P_1^k$ or $P_2^k$ (**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 $P_1^k$ individuals, five of these with known origin (German n=3, English n=2) and this makes it the most common Caucasian mutation in the *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.[142] It is possible that this isolation also explain why the $P^k$ phenotype has been described as more common in Finland compared to the rest of the world[143,144] and that the 202C>T mutation might be the mutation causing a defect β3GalNAc-T in all Finnish cases. However, the five first serologically described $P^k$ families in Finland were all apparently unrelated.[144]

**Table 8.**
A summary of the translational consequences of the nonsense mutations in the *B3GALNT1* gene found in this study.

| Allele | Critical aa change | Open reading frame compared to consensus (%): | |
|---|---|---|---|
| | | Intact | Total |
| 202C>T | Arg67stop | 20 | 20 |
| 292_293insA | Arg98fs102stop | 30 | 31 |
| 433C>T | Arg145stop | 44 | 44 |
| 537_538insA | Asp180fs181stop | 54 | 55 |
| 959G>A | Trp320stop | 96 | 96 |

### *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 $P_1/P_2$ 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 $P_1/P_2$ phenotypes.[80] To investigate if this hypothesis was correct we screened $P_1$ (n=58) and $P_2$ (n=20) samples for the two mutations -160A>G and -551_-550insC, suggested to cause the null phenotype $P_2$ (**Paper IV**). The majority of the donors were of Swedish origin but a few were of Asian or African descent. Three $P_1^k$, three $P_2^k$ 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 $P_1/P_2$ phenotypes was found. Nine of these polymorphisms have not been described before. The two proposed $P_2$-specific mutations, -160A>G and -551_-550insC, were found in homozygous form both in $P_1$ and $P_2$ 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 $P_2$ samples (except one $P_2^k$), 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 commom 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.[146-149] 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, *B3GALT4* (acc.no. Y15061) and *B3GALT5* (acc.no. NM_006057) along with one 3-α-sialyltransferase gene, *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 *et al*. presented data on *B3GALT5* concerning the consequences for Lewis status and other RBC phenotypes including LKE-negative or weak.[150-152]

## Conclusion and future perspectives

In **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 (**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 $P^k$ 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 $P^k$ 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 $P_1/P_2$ phenotypes was found. However, a certain but not complete association to polymorphisms at the *A4GALT* locus was found in three separate studies,[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 $P_2$ and thus lacking the P1 antigen.

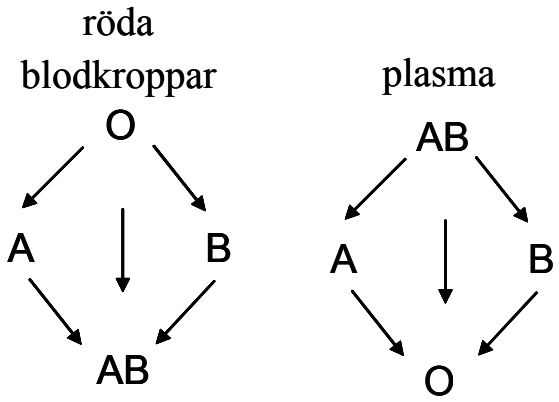
Individuals with different blood groups, especially those having null variants, that lack a certain antigen or even molecule, constitute a great potential since their cells can be used to determine the binding specificity for pathogens utilising carbohydrate and protein blood groups for invasion.

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 $P_1^k$ phenotypes, with causative mutations confirmed in the $P^k$ 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 $P^k$ inhibits HIV infection.[153] Based on this, individuals whose cells express either no or high levels of $P^k$ antigen were investigated in a HIV infection model. The lack of $P^k$ 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 $P_1^k$ individuals, may provide resistance to infection.[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 varianterna 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.

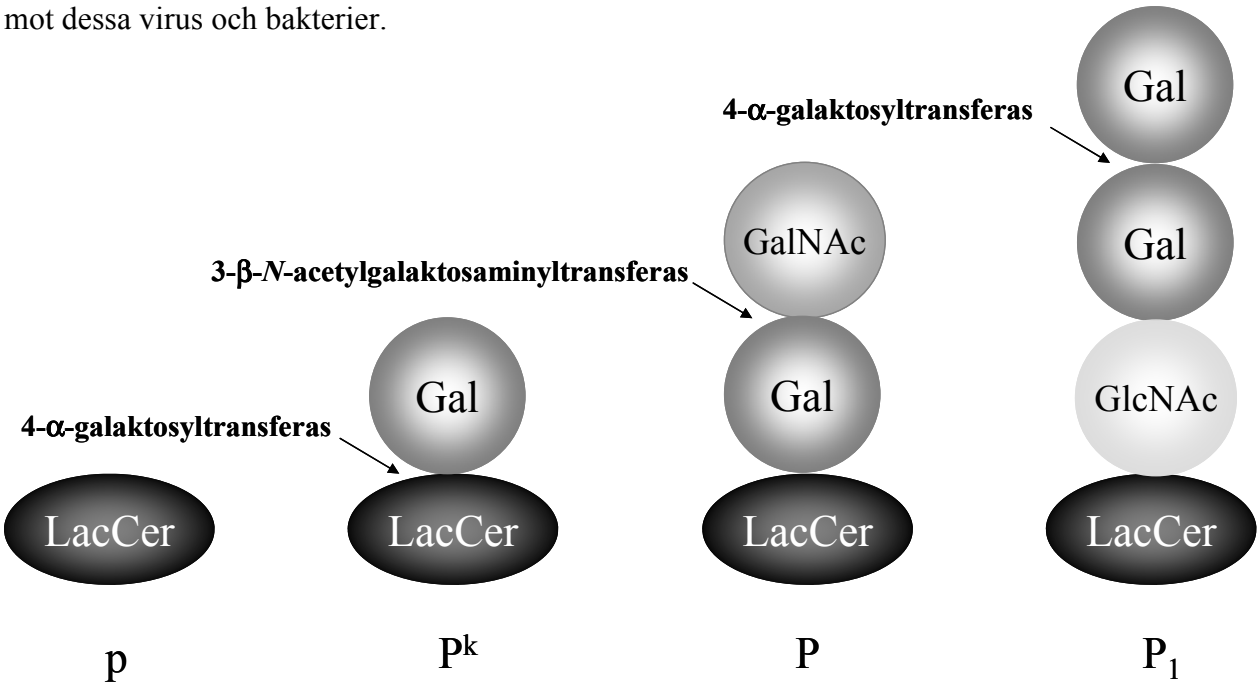



**Figur A**. Bilden visar möjliga ABO-blodgruppskombinationer vid transfusion av röda blodkroppar och plasma.

Blodgruppsantigen finns på den röda blodkroppens cellyta och även på många andra celltyper. Totalt finns det runt 270 blodgruppsantigen inordnade i blodgruppssystem, kollektioner eller serier beroende på internationellt fastställda kriterier. För att en molekyl ska klassificeras som blodgrupp krävs att den uttrycks på röda blodkroppens yta, är ärftlig, förekommer i olika varianter, samt att minst en individ har reagerat immunologiskt genom att ha bildat antikroppar mot ett antigen på molekylen i fråga. Antigenen kan utgöras av proteiner och kolhydrater eller kombinationer av de båda. Tidigare har vår kännedom om olika blodgruppsantigener baserats på serologiska och biokemiska analyser men under de senaste 20 åren har DNA-baserade analyser ökat kunskapen till att även omfatta våra blodgruppsgener. Generna fungerar som recept för cellerna så att dessa vet vilka strukturer som ska byggas. För blodgruppsantigen av protein kodar generna direkt för antigenet medan de för kolhydratantigen kodar för ett ”mellansteg”, enzymer ur gruppen glykosyltransferas.

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.

I denna avhandling har jag studerat gener och dess varianter (polymorfismer) som är ansvariga för uttrycket av tre blodgruppsantigener av kolhydratnatur, nämligen $P^k$, P och P1 (**Figur B**). Dessa utgör P- och GLOB-blodgruppssystemen och GLOB-kollektionen. Fem olika fenotyper förekommer då antigenen $P^k$, P eller P1 saknas i olika kombinationer: $P_1$, $P_2$, p, $P_1^k$ och $P_2^k$. De tre sistnämda är mycket ovanliga, enstaka per miljon människor, och gör att endast blod med samma fenotyp kan transfunderas. Hos kvinnor med fenotyperna p, $P_1^k$ och $P_2^k$ är förekomsten av återkommande missfall högre än hos andra kvinnor. Orsaken är de naturligt förekommande antikroppar som bildas hos alla individer med dessa fenotyper. P/GLOB-antigenen fungerar även som receptorer (mottagare) för vissa virus, bakterier och deras toxiner. T.ex. använder Parvovirus B19 (som orsakar barnsjukdomen femte sjukan) P-antigenet för att ta sig in i förstadiet till de röda blodkropparna samt andra celler. Ett annat exempel är vissa bakterier som bl.a. kan ge upphov till urinvägsinfektion genom att binda till $P^k$, P och P1 på urinvägsceller. Människor med vissa av ovanstående blodgrupper är resistenta mot dessa virus och bakterier.




**Figur B**. Bilden visar antigenen $P^k$, P och P1. Dessa är uppbyggda av olika sockermolekyler (galaktos- eller glukos-varianter) 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-galaktosyl-transferas medan den andra gruppen trodde att det var en 3-β-*N*-acetylgalaktosaminyl-transferasgen som kunde ge upphov till P-blodgruppsantigenet. Denna gen undersöktes 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 undersöktes, varvid fyra nya mutationer och två av de tidigare publicerade mutationerna hittades.

Genen som ger upphov till P$^k$-antigen, 4-α-galaktosyltransferasgen (*A4GALT*, *P*$^k$-genen) undersöktes i **Artikel II** och **III**. Defekter i genen orsakar p-fenotyp, dvs avsaknad av antigenen P$^k$, P och P1. Med hjälp av medarbetare runt om i världen samlade vi in 60 prover från individer med fenotypen p. Av dessa var 29 från Västerbotten, en region välkänd för sin ”höga” frekvens av denna sällsynta blodgrupp, 141 per miljon, jämfört med 5-6 per miljon i övriga Europa. Vi bekräftade tidigare fynd att p-fenotypen i Sverige beror på två olika mutationer i *P*$^k$-genen. Fynden tyder på att det inte enbart är ingifte utan kanske även någon form av selektion som ligger bakom den förhöjda frekvensen av p-fenotyp. Intressant nog hittades samma mutation också i ett prov från södra Sverige. Denna individ hade finländskt ursprung och inget känt släktskap till någon av p-individerna i Västerbotten.
Totalt fann vi tio nya mutationer i *P*$^k$-genen. Graviditetshistoriken för totalt 20 svenska och israeliska kvinnorna som ingick i studien, visade att missfallsrisken varierar mellan individer. Orsaken till detta är oklar och kräver ytterligare undersökningar.

Även Amish-folket i USA, som härstammar från södra Tyskland och Schweiz, har en förhöjd frekvens av p-individer. Genom ett samarbete med amerikanska blodcentraler fick jag möjlighet att analysera 19 prover från just Amish-individer (**Artikel V**). Alla prover innehöll samma mutation, som ger ett aminosyrabyte i ett område som är väl bevarat mellan olika glykosyltransferas, även mellan olika djurarter. Ytterligare ett prov, som skickats till oss från Norge, inkluderades i denna studie sedan det visat sig ha samma mutation. Då provet kom från en pakistansk kvinna boende i Norge var fyndet mycket överraskande.
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 $P^k$-negativ cellinje (Namalwaceller). Uttrycket på cellernas yta, efter att de märkts in med antikroppar mot $P^k$, mättes sedan med hjälp av en metod som kallas flödescytometri. Amish-individernas gen gav inget uttryck jämfört med de negativa kontrollerna. $P^k$- 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 $P^k$, eller om det är två olika genprodukter. Ännu har inte genen som tiilverkar P1 hittats men en japansk forskargrupp föreslog, i ett arbete 2003, att två polymorfismer i *A4GALT*-genens reglerande del (5'-uppströmsregion) skulle vara kopplad till förekomsten av P1-antigenet respektive avsaknad av detsamma. Sjuttioåtta svenska blodgivare undersöktes och alla $P_2$-prover hade dubbel uppsättning (homozygoti) av de s.k $P_2$-mutationerna. Dock hittades även $P_1$-prover som var homozygota för samma mutationer (**Artikel IV**). Därmed blir slutsatsen att de föreslagna $P_2$-polymorfismerna inte enbart kan vara de som styr P1/P2-blodgruppsuttrycket.

Sammantaget har den bakomliggande orsaken till ett flertal $P^k$/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 $P^k$/P/P1-antigenen och olika virus eller bakterier. I ett samarbete med en kanadensisk forskargrupp har vi, utanför denna avhandlings ramar, nyligen funnit att blodgruppen $P_1{}^k$ 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.

تو همانند خورشیدی ، هرگاه که به تو بیندیشم روزم منور است.
دوست دارم !

# References


1. Blundell J. A successful case of transfusion. Lancet 1829;i:431.
2. Landsteiner K. Über Agglutinationserscheinungen normalen menschichen Blutes. Wien Klin Wochenschr 1901;14:1132-4.
3. Landsteiner K. On individual differences in human blood. Nobel Lectures 1930;Physiology or Medicine 1922-1941:234-45.
4. Ottenberg R. Transfusion and arterial anastamosis. Am J Surg 1908;47:486.
5. Landsteiner K, Levine P. Further observations on individual differences of human blood. Proc Soc Exp Biol 1927;24:941-2.
6. Sazama K. Reports of 355 transfusion-associated deaths: 1976 through 1985. Transfusion 1990;30:583-90.
7. Rydberg L. ABO-incompatibility in solid organ transplantation. Transfus Med 2001;11:325-42.
8. Daniels GL, Fletcher A, Garratty G et al. Blood group terminology 2004: from the International Society of Blood Transfusion committee on terminology for red cell surface antigens. Vox Sang 2004;87:304-16.
9. Blumenfeld OO, Patnaik SK. Allelic genes of blood group antigens: a source of human mutations and cSNPs documented in the Blood Group Antigen Gene Mutation Database. Hum Mutat 2004;23:8-16.
10. Reid ME, Mohandas N. Red blood cell blood group antigens: structure and function. Semin Hematol 2004;41:93-117.
11. Horuk R, Chitnis CE, Darbonne WC et al. A receptor for the malarial parasite Plasmodium vivax: the erythrocyte chemokine receptor. Science 1993;261:1182-4.
12. Miller LH, Mason SJ, Clyde DF et al. The resistance factor to Plasmodium vivax in blacks. The Duffy-blood-group genotype, FyFy. N Engl J Med 1976;295:302-4.
13. Gagneux P, Varki A. Evolutionary considerations in relating oligosaccharide diversity to biological function. Glycobiology 1999;9:747-55.
14. Kleene R, Berger EG. The molecular and cell biology of glycosyltransferases. Biochim Biophys Acta 1993;1154:283-325.
15. Watkins WM. Glycosyltransferases. Early history, development and future prospects. Carbohydr Res 1986;149:1-12.
16. Paulson JC, Colley KJ. Glycosyltransferases. Structure, localization, and control of cell type- specific glycosylation. J Biol Chem 1989;264:17615-8.
17. Campbell JA, Davies GJ, Bulone V et al. A classification of nucleotide-diphospho-sugar glycosyltransferases based on amino acid sequence similarities. Biochem J 1997;326:929-39.
18. Breton C, Snajdrova L, Jeanneau C et al. Structures and mechanisms of glycosyltransferases. Glycobiology 2006;16:29R-37R.
19. Mendel G. Versuche über Pflanzen-Hybriden. Verhandlungen des naturforschenden Vereines, Abhandlungen, Brünn 1866;4:3-47.
20. Watson JD, Crick FH. Molecular structure of nucleic acids; a structure for deoxyribose nucleic acid. Nature 1953;171:737-8.
21. Matthaei J, Jones O, Martin R et al. Characteristics and composition of RNA coding units. Proc Natl Acad Sci U S A 1962;48:666-77.
22. Sanger F, Nicklen S, Coulson AR. DNA sequencing with chain-terminating inhibitors. Proc Natl Acad Sci U S A 1977;74:5463-7.
23. Seeburg PH, Shine J, Martial JA et al. Nucleotide sequence of a human gene coding for a polypeptide hormone. Trans Assoc Am Physicians 1977;90:109-16.

24. Siebert PD, Fukuda M. Isolation and characterization of human glycophorin A cDNA clones by a synthetic oligonucleotide approach: nucleotide sequence and mRNA structure. Proc Natl Acad Sci U S A 1986;83:1665-9.

25. Steffensen R, Carlier K, Wiels J et al. Cloning and expression of the histo-blood group Pk UDP-galactose: Ga1beta-4G1cbeta1-cer alpha1, 4-galactosyltransferase. Molecular genetic basis of the p phenotype. J Biol Chem 2000;275:16723-9.

26. Keusch JJ, Manzella SM, Nyame KA et al. Cloning of Gb3 synthase, the key enzyme in globo-series glycosphingolipid synthesis, predicts a family of alpha 1, 4-glycosyltransferases conserved in plants, insects, and mammals. J Biol Chem 2000;275:25315-21.

27. Kojima Y, Fukumoto S, Furukawa K et al. Molecular cloning of globotriaosylceramide/CD77 synthase, a glycosyltransferase that initiates the synthesis of globo series glycosphingolipids. J Biol Chem 2000;275:15152-6.

28. Lo PL, Cabuy E, Chiricolo M et al. Molecular cloning of the human beta1,4 N-acetylgalactosaminyltransferase responsible for the biosynthesis of the Sd(a) histo-blood group antigen: the sequence predicts a very long cytoplasmic domain. J Biochem (Tokyo) 2003;134:675-82.

29. Montiel MD, Krzewinski-Recchi MA, Delannoy P et al. Molecular cloning, gene organization and expression of the human UDP-GalNAc:Neu5Acalpha2-3Galbeta-R beta1,4-N-acetylgalactosaminyltransferase responsible for the biosynthesis of the blood group Sda/Cad antigen: evidence for an unusual extended cytoplasmic domain. Biochem J 2003;373:369-79.

30. Yamamoto F, Marken J, Tsuji T et al. Cloning and characterization of DNA complementary to human UDP-GalNAc:Fuc-alpha1-2Gal alpha1-3GalNAc transferase (histo-blood group A transferase) mRNA. J Biol Chem 1990;265:1146-51.

31. Yamamoto F, Clausen H, White T et al. Molecular genetic basis of the histo-blood group ABO system. Nature 1990;345:229-33.

32. Chester MA, Olsson ML. The ABO blood group gene: a locus of considerable genetic diversity. Transfus Med Rev 2001;15:177-200.

33. Hosseini-Maaf B, Irshaid NM, Hellberg A et al. New and unusual O alleles at the ABO locus are implicated in unexpected blood group phenotypes. Transfusion 2005;45:70-81.

34. Olsson ML, Irshaid NM, Hosseini-Maaf B et al. Genomic analysis of clinical samples with serologic ABO blood grouping discrepancies: identification of 15 novel A and B subgroup alleles. Blood 2001;98:1585-93.

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.

36. Olsson ML, Irshaid N, Kuosmanen M et al. A splice-site mutation defines the *Afinn* allele at the blood group ABO locus. Transfusion 2000;40(10S):13S.

37. Ogasawara K, Yabe R, Uchikawa M et al. Molecular genetic analysis of variant phenotypes of the ABO blood group system. Blood 1996;88:2732-7.

38. Seltsam A, Hallensleben M, Kollmann A et al. The nature of diversity and diversification at the ABO locus. Blood 2003;102:3035-42.

39. Sanger R. An association between the P and Jay systems of blood groups. Nature 1955;176:1163-4.

40. Matson GA, Swanson J, Noades J et al. A "new" antigen and antibody belonging to the P blood group system. Amer J Hum Genet 1959;11:26-34.

41. Tippett P, Sanger R, Race RR et al. An agglutinin associated with the P and the ABO blood group systems. Vox Sang 1965;10:269-80.

42. Daniels GL, Cartron JP, Fletcher A et al. International Society of Blood Transfusion Committee on terminology for red cell surface antigens: Vancouver Report. Vox Sang 2003;84:244-7.

43. Mourant AE, Kope AC, Domaniewska K. The distribution of human blood groups and other polymorphisms. Oxford University Press, 1976.

44. Ikin EW, Kay H, Playfair J et al. P1 antigen in the human fetus. Nature 1961;192:883.

45. Heiken A. Observations on the blood group receptor P1 and its development in children. Hereditas 1966;56:83-98.

46. Fisher R. The variation in the strength of the human blood group P. Heredity 1953;7:81-9.

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.

48. Fellous M, Gerbal A, Tessier C et al. Studies on the biosynthetic pathway of human P erythrocyte antigens using somatic cells in culture. Vox Sang 1974;26:518-36.

49. Fletcher KS, Bremer EG, Schwarting GA. P blood group regulation of glycosphingolipid levels in human erythrocytes. J Biol Chem 1979;254:11196-8.

50. Cooling LL, Kelly K. Inverse expression of P(k) and Luke blood group antigens on human RBCs. Transfusion 2001;41:898-907.

51. Race RR, Sanger R. The P Blood Groups.Blood Groups in Man. Oxford, UK: Blackwell Scientific Publications, 1975;6th:139-77.

52. Kundu SK, Steane SM, Bloom JE et al. Abnormal glycolipid composition of erythrocytes with a weak P antigen. Vox Sang 1978;35:160-7.

53. Kundu SK, Evans A, Rizvi J et al. A new Pk phenotype in the P blood group system. J Immunogenet 1980;7:431-9.

54. Cedergren B. Population studies in northern Sweden. IV. Frequency of the blood type p. Hereditas 1973;73:27-30.

55. Miwa S, Matuhasi T, Yasuda J. p phenotype in two successive generations of a Japanese family. Vox Sang 1974;26:565-7.

56. Obregon E, McKeever BG. Studies on offspring of pp mothers. Transfusion 1980;20:621-2.

57. Levene C, Shinar E, Yahalom V. Rare blood group p Tj(a-)in Israel 1975-1999 . Vox Sang 2000;78.

58. Nakajima H, Yokota T. Two Japanese families with Pk members. Vox Sang 1977;32:56-8.

59. Arndt PA, Garratty G, Marfoe RA et al. An acute hemolytic transfusion reaction caused by an anti-P1 that reacted at 37 degrees C. Transfusion 1998;38:373-7.

60. Chandeysson PL, Flye MW, Simpkins SM et al. Delayed hemolytic transfusion reaction caused by anti-P1 antibody. Transfusion 1981;21:77-82.

61. Issitt PD, Anstee DJ. Applied Blood Group Serology. 4th ed. Miami, FL, USA: Montgomery Scientific Publications, 1998.

62. Cantin G, Lyonnais J. Anti-PP1Pk and early abortion. Transfusion 1983;23:350-1.

63. Lindstrom K, Dem Borne AE, Breimer ME et al. Glycosphingolipid expression in spontaneously aborted fetuses and placenta from blood group p women. Evidence for placenta being the primary target for anti-Tja-antibodies. Glycoconj J 1992;9:325-9.

64. Levine P, Bobbitt OB, Waller RK et al. Isoimmunization by a new blood factor in tumor cells. Proc Soc Exp Biol Med 1951;77:403-5.

65. Tippett P, Andrews PW, Knowles BB et al. Red cell antigens P (globoside) and Luke: identification by monoclonal antibodies defining the murine stage-specific embryonic antigens -3 and -4 (SSEA-3 and SSEA-4). Vox Sang 1986;51:53-6.

66. Daniels G. Human Blood Groups. 2nd ed. Oxford, UK: Blackwell Scientific, 2002.

67. Judd WJ, Cooling LL, Dake LR et al. Clinically significant Anti-LKE: Nonreactive in prewarmed tests. Transfusion 2005;45:122A.

68. Bruce M, Watt A, Gabra GS et al. LKE red cell antigen and its relationship to P1 and Pk: serological study of a large family. Vox Sang 1988;55:237-40.

69. Moller B, Jorgensen J. Anti-LKE in a pregnant woman. Transfusion 1988;28:88.

70. Morgan WT, Watkins WM. Blood group P1 substance. I. Chemical properties. Bibl Haematol 1964;19:225-9.

71. Marcus DM. Isolation of a substance with blood-group P1 activity for human erythrocyte stroma. Transfusion 1971;11:16-8.

72. Thudichum JLW. A Treatise on the Chemical Constituents of the Brain (a facsimile edition of the original). Bailliere, Tindall and Cox, London ed. 1962.

73. Hakomori S. Structure, organization, and function of glycosphingolipids in membrane. Curr Opin Hematol 2003;10:16-24.

74. Sandhoff K, Kolter T. Biosynthesis and degradation of mammalian glycosphingolipids. Philos Trans R Soc Lond B Biol Sci 2003;358:847-61.

75. Rydberg L, Cedergren B, Breimer ME et al. Serological and immunochemical characterization of anti-PP1Pk (anti- Tja) antibodies in blood group little p individuals. Blood group A type 4 recognition due to internal binding. Mol Immunol 1992;29:1273-86.

76. Siddiqui B, Hakomori S. A revised structure for the Forssman glycolipid hapten. J Biol Chem 1971;246:5766-9.

77. Yang Z, Bergstrom J, Karlsson KA. Glycoproteins with Gal alpha 4Gal are absent from human erythrocyte membranes, indicating that glycolipids are the sole carriers of blood group P activities. J Biol Chem 1994;269:14620-4.

78. Spitalnik PF, Spitalnik SL. The P blood group system: biochemical, serological, and clinical aspects. Transfus Med Rev 1995;9:110-22.

79. Furukawa K, Iwamura K, Uchikawa M et al. Molecular basis for the p phenotype. Identification of distinct and multiple mutations in the alpha 1,4-galactosyltransferase gene in Swedish and Japanese individuals. J Biol Chem 2000;275:37752-6.

80. Iwamura K, Furukawa K, Uchikawa M et al. The blood group P1 synthase gene is identical to the Gb3/CD77 synthase gene. A clue to the solution of the P1/P2/p puzzle. J Biol Chem 2003;278:44429-38.

81. Breton C, Imberty A. Structure/function studies of glycosyltransferases. Curr Opin Struct Biol 1999;9:563-71.

82. Hughes AK, Ergonul Z, Stricklett PK et al. Molecular basis for high renal cell sensitivity to the cytotoxic effects of shigatoxin-1: upregulation of globotriaosylceramide expression. J Am Soc Nephrol 2002;13:2239-45.

83. Amado M, Almeida R, Carneiro F et al. A family of human beta3-galactosyltransferases. Characterization of four members of a UDP-galactose:beta-N-acetyl-glucosamine/beta-N-acetyl-galactosamine beta-1,3-galactosyltransferase family. J Biol Chem 1998;273:12770-8.

84. Okajima T, Nakamura Y, Uchikawa M et al. Expression cloning of human globoside synthase cDNAs. Identification of beta3Gal-T3 as UDP-*N*-acetylgalactosamine:globotriaosylceramide beta1,3-*N*-acetylgalactosaminyltransferase. J Biol Chem 2000;275:40498-503.

85. Tan JS, Mohandas N, Conboy JG. High frequency of alternative first exons in erythroid genes suggests a critical role in regulating gene function. Blood 2006;107:2557-61.

86. McAlpine PJ, Kaita H, Lewis M. Is the DIA1 locus linked to the P blood group locus? Cytogenet Cell Genet 1978;22:629-32.

87. Naiki M, Marcus DM. An immunochemical study of the human blood group P1, P, and PK glycosphingolipid antigens. Biochemistry 1975;14:4837-41.

88. Bailly P, Piller F, Gillard B et al. Biosynthesis of the blood group Pk and P1 antigens by human kidney microsomes. Carbohydr Res 1992;228:277-87.

89. Graham HA, Williams AN. A genetic model for the inheritance of the P, P1 and Pk antigens. Immunol Commun 1980;9:191-201.

90. Fujii Y, Numata S, Nakamura Y et al. Murine glycosyltransferases responsible for the expression of globo-series glycolipids: cDNA structures, mRNA expression, and distribution of their products. Glycobiology 2005;15:1257-67.

91. Zoja C, Corna D, Farina C et al. Verotoxin glycolipid receptors determine the localization of microangiopathic process in rabbits given verotoxin-1. J Lab Clin Med 1992;120:229-38.

92. Mandrell RE. Further antigenic similarities of Neisseria gonorrhoeae lipooligosaccharides and human glycosphingolipids. Infect Immun 1992;60:3017-20.

93. Suzuki A, Karol RA, Kundu SK et al. Glycosphingolipids of K562 cells: a chemical and immunological analysis. Int J Cancer 1981;28:271-6.

94. Svennerholm E, Svennerholm L. The separation of neutral blood-serum glycolipids by thin-layer chromatography. Biochim Biophys Acta 1963;70:432-41.

95. Dunstan RA. Status of major red cell blood group antigens on neutrophils, lymphocytes and monocytes. Br J Haematol 1986;62:301-9.

96. Kasai K, Galton J, Terasaki PI et al. Tissue distribution of the Pk antigen as determined by a monoclonal antibody. J Immunogenet 1985;12:213-20.

97. Mangeney M, Richard Y, Coulaud D et al. CD77: an antigen of germinal center B cells entering apoptosis. Eur J Immunol 1991;21:1131-40.

98. Knapp W, Dorken B, Rieber P et al. CD antigens 1989. Blood 1989;74:1448-50.

99. Dem Borne AE, Bos MJ, Joustra-Maas N et al. A murine monoclonal IgM antibody specific for blood group P antigen (globoside). Br J Haematol 1986;63:35-46.

100. Shevinsky LH, Knowles BB, Damjanov I et al. Monoclonal antibody to murine embryos defines a stage-specific embryonic antigen expressed on mouse embryos and human teratocarcinoma cells. Cell 1982;30:697-705.

101. Cooling LL, Koerner TA, Naides SJ. Multiple glycosphingolipids determine the tissue tropism of parvovirus B19. J Infect Dis 1995;172:1198-205.

102. Song Y, Withers DA, Hakomori S. Globoside-dependent adhesion of human embryonal carcinoma cells, based on carbohydrate-carbohydrate interaction, initiates signal transduction and induces enhanced activity of transcription factors AP1 and CREB. J Biol Chem 1998;273:2517-25.

103. Brown KE, Hibbs JR, Gallinella G et al. Resistance to parvovirus B19 infection due to lack of virus receptor (erythrocyte P antigen). N Engl J Med 1994;330:1192-6.

104. Young NS, Brown KE. Parvovirus B19. N Engl J Med 2004;350:586-97.

105. de Jong EP, de Haan TR, Kroes AC et al. Parvovirus B19 infection in pregnancy. J Clin Virol 2006;36:1-7.

106. Heddle NM. Acute paroxysmal cold hemoglobinuria. Transfus Med Rev 1989;3:219-29.

107. Fantini J, Hammache D, Pieroni G et al. Role of glycosphingolipid microdomains in CD4-dependent HIV-1 fusion. Glycoconj J 2000;17:199-204.

108. Lund N, Sakac D, Lingwood C et al. Pk blood group antigen: A natural resistance factor to HIV infection. Vox Sang 2006;91:23.

109. Moulds JM, Moulds JJ. Blood group associations with parasites, bacteria, and viruses. Transfus Med Rev 2000;14:302-11.

110. Ziegler T, Jacobsohn N, Funfstuck R. Correlation between blood group phenotype and virulence properties of Escherichia coli in patients with chronic urinary tract infection. Int J Antimicrob Agents 2004;24 Suppl 1:70-5.:70-5.

111. Lomberg H, Jodal U, Eden CS et al. P1 blood group and urinary tract infection. Lancet 1981;1:551-2.

112. Haataja S, Tikkanen K, Liukkonen J et al. Characterization of a novel bacterial adhesion specificity of Streptococcus suis recognizing blood group P receptor oligosaccharides. J Biol Chem 1993;268:4311-7.

113. Imberty A, Wimmerova M, Mitchell EP et al. Structures of the lectins from Pseudomonas aeruginosa: insight into the molecular basis for host glycan recognition. Microbes Infect 2004;6:221-8.

114. Bekri S, Lidove O, Jaussaud R et al. The role of ceramide trihexoside (globotriaosylceramide) in the diagnosis and follow-up of the efficacy of treatment of Fabry disease: a review of the literature. Cardiovasc Hematol Agents Med Chem 2006;4:289-97.

115. Cilmi SA, Karalius BJ, Choy W et al. Fabry disease in mice protects against lethal disease caused by Shiga toxin-expressing enterohemorrhagic Escherichia coli. J Infect Dis 2006;194:1135-40.

116. Lingwood CA, Law H, Richardson S et al. Glycolipid binding of purified and recombinant Escherichia coli produced verotoxin in vitro. J Biol Chem 1987;262:8834-9.

117. Smith DC, Lord JM, Roberts LM et al. Glycosphingolipids as toxin receptors. Semin Cell Dev Biol 2004;15:397-408.

118. O'Brien AD, Holmes RK. Shiga and Shiga-like toxins. Microbiol Rev 1987;51:206-20.

119. Proulx F, Seidman EG, Karpman D. Pathogenesis of Shiga toxin-associated hemolytic uremic syndrome. Pediatr Res 2001;50:163-71.

120. Furukawa K, Yokoyama K, Sato T et al. Expression of the Gb3/CD77 synthase gene in megakaryoblastic leukemia cells: implication in the sensitivity to verotoxins. J Biol Chem 2002;277:11247-54.

121. Tetaud C, Falguieres T, Carlier K et al. Two distinct Gb3/CD77 signaling pathways leading to apoptosis are triggered by anti-Gb3/CD77 mAb and verotoxin-1. J Biol Chem 2003;278:45200-8.

122. Watarai S, Yokota K, Tana et al. Relationship between susceptibility to hemolytic-uremic syndrome and levels of globotriaosylceramide in human sera. J Clin Microbiol 2001;39:798-800.

123. Okuda T, Tokuda N, Numata S et al. Targeted disruption of Gb3/CD77 synthase gene resulted in the complete deletion of globo-series glycosphingolipids and loss of sensitivity to verotoxins. J Biol Chem 2006;281:10230-5.

124. Dwek MV, Brooks SA. Harnessing changes in cellular glycosylation in new cancer treatment strategies. Curr Cancer Drug Targets 2004;4:425-42.

125. Levine P. Illegitimate blood group antigens P1, A, and MN (T) in malignancy-a possible therapeutic approach with anti-Tja, anti-A, and anti-T. Ann N Y Acad Sci 1976;277:428-35.

126. Arab S, Russel E, Chapman WB et al. Expression of the verotoxin receptor glycolipid, globotriaosylceramide, in ovarian hyperplasias. Oncol Res 1997;9:553-63.

127. Kovbasnjuk O, Mourtazina R, Baibakov B et al. The glycosphingolipid globotriaosylceramide in the metastatic transformation of colon cancer. Proc Natl Acad Sci U S A 2005;102:19087-92.

128. LaCasse EC, Bray MR, Patterson B et al. Shiga-like toxin-1 receptor on human breast cancer, lymphoma, and myeloma and absence from CD34(+) hematopoietic stem cells: implications for ex vivo tumor purging and autologous stem cell transplantation. Blood 1999;94:2901-10.

129. Lingwood CA. Verotoxins and their glycolipid receptors. Adv Lipid Res 1993;25:189-211.:189-211.

130. DeGrandis S, Law H, Brunton J et al. Globotetraosylceramide is recognized by the pig edema disease toxin. J Biol Chem 1989;264:12520-5.

131. Gilboa-Garber N, Sudakevitz D, Sheffi M et al. PA-I and PA-II lectin interactions with the ABO(H) and P blood group glycosphingolipid antigens may contribute to the broad spectrum adherence of Pseudomonas aeruginosa to human tissues in secondary infections. Glycoconj J 1994;11:414-7.

132. Cedergren B. Northern Sweden - an area with an exceptional concentration of p/T(ja-). Bibl Haematol 1971;38:162-6.

133. Bittles AH, Egerbladh I. The influence of past endogamy and consanguinity on genetic disorders in northern Sweden. Ann Hum Genet 2005;69:549-58.

134. Nordstrom S, Cedergren B. Genetic genealogical studies of 20 north Swedish families with the rare blood group p. Hum Hered 1989;39:20-5.

135. Francomano CA, McKusick VA, Biesecker LG. Medical genetic studies in the Amish: historical perspective. Am J Med Genet C Semin Med Genet 2003;121:1-4.

136. Agarwala R, Schaffer AA, Tomlin JF. Towards a complete North American Anabaptist Genealogy II: analysis of inbreeding. Hum Biol 2001;73:533-45.

137. Kozak M. At least six nucleotides preceding the AUG initiator codon enhance translation in mammalian cells. J Mol Biol 1987;%20;196:947-50.

138. Sheehan J, Pochedley M, Toy E. A retrospective study of cord blood samples from infants born to p phenotype mothers. Book of abstracts Joint Cong ISBT/AABB 1990.

139. de La Vega D, Hellberg Å, Pivetta M et al. Bases Moleculares del excepcional fenotipo de grupo sanguíneo p. A propósito de un caso. Submitted 2007.

140. Koda Y, Soejima M, Sato H et al. Three-base deletion and one-base insertion of the alpha(1,4)galactosyltransferase gene responsible for the P phenotype. Transfusion 2002;42:48-51.

141. Yan L, Zhu F, Xu X et al. Molecular basis for p blood group phenotype in China. Transfusion 2004;44:136-8.

142. Norio R. Finnish Disease Heritage I: characteristics, causes, background. Hum Genet 2003;112:441-56.

143. Kortekangas AE, Noades J, Tippett P et al. A second family with the red cell antigen Pk. Vox Sang 1959;4:337-49.:337-49.

144. Kortekangas AE, Kaarsalo E, Melartin L et al. The red cell antigen Pk and its relationship to the P system: the evidence of three more Pk families. Vox Sang 1965;10:385-404.

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.

146. Gillard BK, Jones MA, Marcus DM. Glycosphingolipids of human umbilical vein endothelial cells and smooth muscle cells. Arch Biochem Biophys 1987;256:435-45.

147. Stroud MR, Stapleton AE, Levery SB. The P histo-blood group-related glycosphingolipid sialosyl galactosyl globoside as a preferred binding receptor for uropathogenic Escherichia coli: isolation and structural characterization from human kidney. Biochemistry 1998;37:17420-8.

148. Cooling LL, Zhang D, Koerner TA. Human platelets express gangliosides with LKE activity and ABH blood group activity. Transfusion 2001;41:504-16.

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 galactosylgloboside 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.

153. Lund N, Branch DR, Mylvaganam M et al. A novel soluble mimic of the glycolipid, globotriaosyl ceramide inhibits HIV infection. AIDS 2006;20:333-43.

154. Lund N, Branch DR, Sakac D et al. Lack of susceptibility of cells from patients with Fabry disease to productive infection with R5 human immunodeficiency virus. AIDS 2005;19:1543-6.

155. Lund N, Olsson ML, Sakac D et al. Pk blood group expression provides natural protection against HIV-1 infection whilst the Pk-deficient p phenotype is hyper-susceptible. Submitted 2007.

# Appendix: Papers I-V