LUND UNIVERSITY

Studies on the Polymorphism and Transcriptional Regulation of the ABO and P1PK Histo-blood Group Genes

Thuresson, Britt

2011

Link to publication

Citation for published version (APA):
Thuresson, B. (2011). *Studies on the Polymorphism and Transcriptional Regulation of the ABO and P1PK Histo-blood Group Genes*. [Doctoral Thesis (compilation), Division of Hematology and Transfusion Medicine].

Total number of authors:
1

General rights
Unless other specific re-use rights are stated the following general rights apply:
Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.
• Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
• You may not further distribute the material or use it for any profit-making activity or commercial gain
• You may freely distribute the URL identifying the publication in the public portal

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

Take down policy
If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

LUND UNIVERSITY

PO Box 117
221 00 Lund
+46 46-222 00 00

Studies on the Polymorphism and Transcriptional Regulation of the ABO and P1PK Histo-blood Group Genes

Doctoral thesis
by

Britt Thuresson

LUND
UNIVERSITY

Britt Thuresson
Division of Hematology and Transfusion Medicine
Department of Laboratory Medicine
Lund University
Lund, Sweden
e-mail: britt.thuresson@med.lu.se

Printed by Media-Tryck, Lund, Sweden 2011
© Britt Thuresson 2011

ISSN 1652-8220
ISBN 978-91-86671-84-6
Lund University, Faculty of Medicine Doctoral Dissertation Series 2011:36

Till min familj,

Put up in a place
where it's easy to see
the cryptic admonishment
T. T. T.

When you feel how depressingly
slowly you climb,
it's well to remember that
Things Take Time.

Kumbel Kumbell

Abstract

Antigens of the clinically important ABO and P1PK blood group systems are carbohydrate structures. Thus the underlying genes do not encode antigens directly but glycosyltransferases that add specific sugar molecules to selected precursor chains. The aim of this study was to investigate the transcriptional regulation of *ABO* and *A4GALT*, with an emphasis on interindividual differences.

The up-and downstream regions of the major *ABO* alleles were sequenced to identify allele-specific motifs. Transcript levels were evaluated in both peripheral blood and bone marrow samples.

Considerable allelic variation was observed between the common *ABO* alleles although this did not appear to influence transcript levels. Strikingly, however transcripts from the two major *A* alleles, A^1 and A^2 were undetectable in peripheral blood while, *B*/*O* transcripts were readily found. Consequently all alleles were transcribed in bone marrow cells undergoing erythroid culture.

In a second study, a novel *ABO* hybrid allele with an anomalous enhancer region was characterized in a A_1B_{weak} sample. Contrary to current beliefs, the number minisatellite 43-bp elements in the enhancer region did not correlate with *ABO* transcript levels in these two studies.

A4GALT gene transcripts were measured in P_1 and P_2 phenotype samples, regarding transcript levels and RBC surface antigen expression.

Through the discovery of a novel *A4GALT* transcript containing a previously unrecognized and polymorphic *A4GALT* exon, the long-suspected link between the P1 and P^k antigens was established. A P^1/P^2 polymorphism was confirmed useful for genotyping in >200 donors. The P^2 allele was shown to lower *A4GALT* transcript levels as well as P1 and P^k antigen expression. Accordingly P^1/P^2 zygosity appears to explain the well-known but poorly understood variability in P1 strength on erythrocytes. Based on these studies, the P^k antigen has now joined P1 in the former P blood group system, appropriately re-designated P1PK.

In summary, these studies of *ABO* and *A4GALT* transcription have resulted in significant discoveries towards increased understanding of two clinically important blood group systems. The mechanisms underlying 1) the apparent absence of A transcripts in peripheral blood, and 2) how the P^2-specific polymorphism down-regulates *A4GALT* transcripts remains to be explained.

Key words: Blood group, *ABO*, *A4GALT*, P1PK, Gene regulation, Enhancer

Contents

Abstract 5

Contents 7

Abbreviations 9

List of papers 10

Introduction 11

Blood groups 11
Antigens and antibodies 11
Structure and function 11
Carbohydrate blood groups 12
Terminology 12
Disease associations 15

Glycosyltransferases 16

Genetics 18
Polymorphisms 18
Gene regulation 19
Transcription factors 19
Alterations in non-blood-group genes affecting blood group expression 20
Alu repeats 21

Erythropoiesis 21

The ABO blood group system 23

Antigens and phenotypes 23

Biochemistry 24

Genetics 25
Polymorphisms in the *ABO* gene 26
Hybrid genes 27
Transcriptional regulation of the *ABO* gene 27

Tissue distribution 28

The P1PK blood group system 29

Antigens and phenotypes 29

Other related antigens 30
Biochemistry 31
Genetics 32
Transcriptional regulation of the *A4GALT* gene 33
Tissue distribution 34
Summary and discussion of the present investigation 35
General aims 35
Commentary on the materials and methods used 35
Paper I 40
Paper II 41
Additional work on *ABO* 42
Paper III 43
Additional work on *A4GALT* 45
Conclusions 57
Future perspectives 59
Svensk sammanfattning 61
Acknowledgements 65
References 67

Abbreviations

aa	Amino acid
AD	Allelic discrimination
ACD	Acid citrate dextrose
ASP	Allele-specific primer
bp	Base pair
EDTA	Ethylenediaminetetraacetic acid
EKLF	Erythroid Kruppel-Like Factor (also KLF1)
EMSA	Electrophoretic Mobility Shift Assay
EPO	Erythropoietin
GFP	Green fluorescent protein
GT	Glycosyltransferase
HDFN	Haemolytic disease of foetus and the newborn
HRP	Horseradish peroxidase
HSC	Haematopoietic stem cells
HTR	Haemolytic transfusion reaction
ISBT	International Society of Blood Transfusion
ORF	Open reading frame
PCR	Polymerase chain reaction
RACE	Rapid Amplification of cDNA Ends
RBC	Red blood cell
RQ-PCR	Real-time quantitative polymerase chain reaction
SNP	Single nucleotide polymorphism
TF	Transcription factor
UTR	Untranslated region

List of papers

This thesis is based on the following papers, which are referred to in the text by Roman numerals.

I. **Thuresson B**, Chester MA, Storry JR and Olsson ML.
ABO transcript levels in peripheral blood and erythropoietic culture show different allele-related patterns independent of the CBF/NF-Y enhancer motif and multiple novel allele-specific variations in the 5'- and 3'-noncoding regions.
Transfusion, 2008 Mar; 48(3):493-504. Epub 2007 Dec 7.

II. **Thuresson B**, Hosseini-Maaf B, Hult AK, Hustinx H, Chester MA and Olsson ML.
A novel B^{weak} hybrid allele lacks three enhancer repeats but generates normal *ABO* transcript levels.
Vox Sanguinis, 2011; in press.

III. **Thuresson B**, Westman JS and Olsson ML.
Identification of a novel *A4GALT* exon reveals the genetic basis of the P_1/P_2 histo-blood groups.
Blood, 2011 Jan 13; 117(2):678-687. Epub 2010 Oct 22.

Papers have been printed with permission from copyright owners:
Papers I and II: © 1999-2010 John Wiley & Sons, Inc.
Papers III: This research was originally published in *Blood*. © the American Society of Hematology.

Introduction

Blood groups

Antigens and antibodies

Blood group antigens are polymorphic variations of red blood cell (RBC) surface molecules that can give rise to antibody production in individuals lacking the corresponding molecule or epitope residing on it. Blood group antibodies can be produced during pregnancy or after transfusion and transplantation. Naturally-occurring antibodies, usually IgM, are often observed against carbohydrate blood group antigens.[1,2] Allo-antibodies can cause haemolytic transfusion reactions (HTR), haemolytic disease of the foetus and newborn (HDFN) and in some cases spontaneous abortions. It is therefore of major importance to be able to type both blood donors and transfusion recipients as well as pregnant women correctly.

A blood group system is comprised of one or more antigens encoded by a single gene or a cluster of two or three closely linked homologous genes. Currently there are 30 blood group systems characterized, and seven blood group collections for which insufficient information is available.

Structure and function

Blood group antigens can be carried either on proteins or on carbohydrate moieties that can be found on both glycoproteins and glycolipids. This thesis will focus on selected carbohydrate blood groups. Monosaccharides in the form of nucleotide-sugars are added as donor substrate by a glycosyltransferase to a precursor in the form of an oligosaccharide chain serving as acceptor substrate. It is the gene encoding the transferase that is considered the blood group gene. Six out of the 30 blood groups systems (ABO, P1PK, LE, H, I and GLOB) are known to be carbohydrate-based. Some blood group systems are restricted to RBCs, whilst the carbohydrate-based systems are often referred to as histo-blood group systems since their antigens can be found on many other cell types.[3]

There is a vast variety of functions for the molecules of the different blood group systems, as protein molecules involved in transport, cell-cell or cell-matrix interaction, complement regulation, enzymatic activity and structural RBC integrity. Carbohydrates form the glycocalyx of the cell and are also utilized as involuntary receptors for pathogens, see below. In some cases the function is still unknown (Figure 1).[4] The function of the blood group molecule is not always known to be affected by the genetic polymorphism, although in the case of null

alleles where the entire molecule is missing from the cell the effects range from apparently none to severe dysfunction of the RBC and perhaps the whole organism.

Adhesion and receptor molecules	Glycocalyx	Transporters and channels
Landsteiner-Wiener Xg Duffy Lutheran Indian Scianna Raph John Milton Hagen Ok	ABO P1PK Lewis Globoside H I	Rh RH-associated glycophorin Kidd Diego Colton Gill Kx
Glycocalyx and cell structure	**Complement regulation**	**Enzymes**
MNS Gerbich	Chido/Rogers Cromer Knops	Kell Yt Dombrock

Figure 1. Functional aspects of the different blood group systems.

Carbohydrate blood groups

Six of the blood group systems marked with grey in Table 1, are carbohydrate structures. These antigen structures often resemble carbohydrates on bacterial surfaces, and the naturally-occurring antibodies are thought to be the result of a gastrointestinal exposure to bacteria in the normal gut flora carrying the structures absent in host RBCs.[2] For instance if a person´s RBCs express the A antigen of the ABO system but lack the B antigen, antibodies against the B structure are formed.

Terminology

Historically blood groups have often been named after the discoverer or the individual that had made the antibody. In 1980 the International Society of Blood Transfusion (ISBT) organized a committee, now named the ISBT Working Party on Red Cell Immunogenetics and Blood Group Terminology, to classify the blood group antigens in a more systematic way, and to devise a genetically based numerical terminology for RBC surface antigens. Changes in terminology are

decided and published in connection with the ISBT meetings.[5] Blood group antigens can be categorized in systems (Table 1), collections (200 series), low-prevalence-antigens (700 series) or high-prevalence antigens (901 series). For a blood group antigen to be included in a blood group system the gene or genes encoding the involved molecules must have been identified. Antigens, which are serologically, biochemically or genetically related, but does not yet fit the criteria for system status, are joined in collections. Low-prevalence antigens are antigens with an incidence of less than 1% while high-prevalence antigens must have an incidence greater than 90%. Antigens have been given a six-digit number, where the first three represent the blood group system, collection or series, and the last three the antigen. Each blood group system is also represented with a one- to four-letter symbol. Emerging antigens have not yet been designated to any of the classifications above and need to be further investigated. Guidelines for classification of blood group antigens are published on the ISBT website (www. blood.co.uk/ibgrl). Information about the allelic variation of the blood group antigens can be found in the Blood Group Antigen Gene Mutation Database (dbRBC at www.ncbi.nlm.nih.gov).[6]

Table 1. Blood group systems according to ISBT classification guidelines. Carbohydrate blood group systems are marked in grey.

No.	System name	System symbol	Gene name(s)*	Antigens	Chromosomal location	CD number
001	ABO	ABO	*ABO*	4	9q34.2	
002	MNS	MNS	*GYPA, GYPB, GYPE*	46	4q31.21	CD235
003	P1PK	P1PK	*A4GALT*	2	22q13.2	CD77
004	Rh	RH	*RHD, RHCE*	59	1p36.11	CD240
005	Lutheran	LU	*LU*	22	19q13.32	CD239
006	Kell	KEL	*KEL*	35	7q34	CD238
007	Lewis	LE	*FUT3*	6	19p13.3	
008	Duffy	FY	*DARC*	6	1q23.2	CD234
009	Kidd	JK	*SLC14A1*	3	18q12.3	
010	Diego	DI	*SLC4A1*	22	17q21.31	CD233
011	Yt	YT	*ACHE*	2	7q22.1	
012	Xg	XG	*XG, MIC2*	2	Xp22.33	CD99
013	Scianna	SC	*ERMAP*	7	1p34.2	
014	Dombrock	DO	*ART4*	4	12p12.3	CD297
015	Colton	CO	*AQP1*	7	7p14.3	
016	Landsteiner-Wiener	LW	*ICAM4*	7	19p13.2	CD242
017	Chido/Rodgers	CH/RG	*C4A, C4B*	9	6p21.3	
018	H	H	*FUT1*	1	19q13.33	CD173
019	Kx	XK	*XK*	1	Xp21.1	
020	Gerbich	GE	*GYPC*	12	2q14.3	CD236
021	Cromer	CROM	*CD55*	16	1q32.2	CD55
022	Knops	KN	*CR1*	9	1q32.2	CD35
023	Indian	IN	*CD44*	4	11p13	CD44
024	Ok	OK	*BSG*	3	19p13.3	CD147
025	Raph	RAPH	*CD151*	1	11p15.5	CD151
026	John Milton Hagen	JMH	*SEMA7A*	6	15q24.1	CD108
027	I	I	*GCNT2*	1	6p24.2	
028	Globoside	GLOB	*B3GALNT1*	1	3q25.1	
029	Gill	GIL	*AQP3*	1	9p13.3	
030	Rh-associated glycoprotein	RHAG	*RHAG*	3	6p21-qter	CD241

Disease associations

Antibodies against blood group antigens can cause haemolytic transfusion reactions, haemolytic disease of the foetus and newborn, autoimmune haemolytic anemia, graft rejection and spontaneous abortion. Malignant diseases have been linked to ABO type and some antigens, notably carbohydrate antigens have been revealed to be associated with parasites, bacterial and viral infections (Table 2).[7-22]

Table 2. Examples of microorganisms using blood group antigens as receptors.

Microorganism	Antigen	Disease	Ref.
Parasites			
Plasmodium falciparum	ABO	Malaria	7, 8
Plasmodium vivax	Duffy	Malaria	9, 10
Bacteria			
Escherichia coli	P1, P^k	UTI (Urinary tract infection)	11,12,13
Escherichia coli	Glob	UTI	14
Escherichia coli	Dr^a (Cromer)	UTI	11
Helicobacter pylori	Le^x	Gastritis	15
Haemophilus influenzae	Lu(a-b-), AnWj-	URI	16
Mycoplasma pneumoniae	I	Pneumonia	17,18
Virus			
Parvovirus B19	P	Fifth disease	19, 24
HIV	P^k	AIDS	20
Norovirus	Le^b	Gastroenteritis	21, 22

E. coli adheres to cells with P1, P^k, P, LKE antigens but not to cells with p phenotype.[12-14,23] The P antigen is a known receptor for Parvovirus B19 causing fifth disease.[19,24] P^k is known to be a ligand for verotoxins (Shiga toxins and VT1/2).[25] A study has shown that the rosetting parasite *Plasmodium falciparum*, which is the most virulent strain of malaria parasites, form larger and stronger rosettes in A and AB than in O individuals, leading to more severe form of cerebral malaria.[7,26] *Helicobacter pylori,* uses H antigen (the O blood group) according to some investigators, but also Lewis antigens as receptors. Recent studies have also found that the P^k antigen is involved in HIV infection and that high amounts of P^k antigen should decrease the risk for infection while lack of P^k as in the p phenotype increases the risk.[27] Multiple other pathogen associations and other disease linkage have been described but will not be dealt with in detail here.[28,29]

Glycosyltransferases

Oligosaccharide structures are derived from formation of glycosidic linkages (Figure 2) between activated donor substrate, usually in the form of nucleoside diphosphate sugars, and an acceptor substrate in the form of an oligosaccharide but can also be a monosaccharide, lipid, protein, nucleic acid or other small molecules. Glycosyltransferases (GT) are classified in families based on amino acid (aa) similarities and there are currently a total of 92 families (www.cazy.org). GTs account for 1-2% of the gene products of an organism, which is similar for all organisms whether bacterial or eukaryote. There are about 230 GT genes in humans.[30]

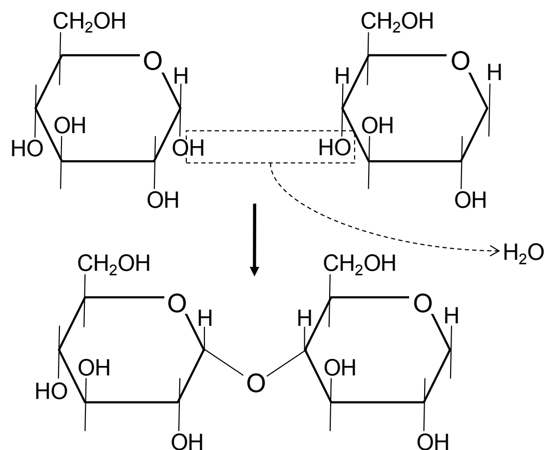


Figure 2. Schematic picture of a 1-4 glycosidic linkage, where two monomers are joined with a covalent bond, releasing a water molecule in the process.

Carbohydrate blood group antigens like H, A, B and P1 are based on paragloboside chains, while P^k, P and LKE are based on lactosylceramide chains. A schematic picture of these oligosaccharide chains is shown in Figure 3.

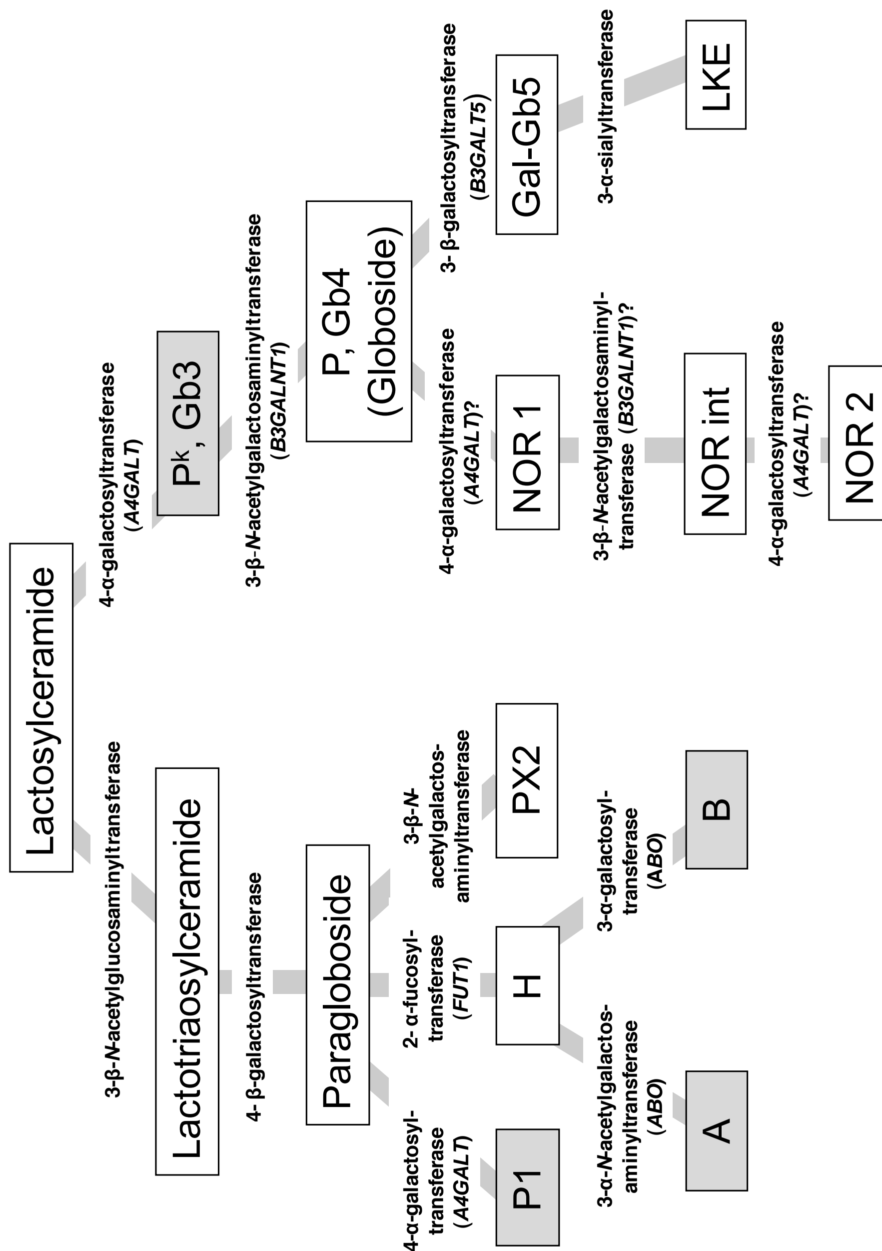


Figure 3. Oligosaccharide chains harbouring the P^k, P1, A and B antigens (grey boxes), as well as other related structure (white boxes). If the responsible glycosyltransferase gene is known, its name is given in brackets below the implicated enzyme.

Genetics

Carbohydrate blood group system genes differ from protein system genes in the way that they encode an enzyme synthesizing the antigen rather than the actual antigen itself. Mutations in a glycosyltransferase gene that alters the aa sequence of the enzyme may change its affinity for either the acceptor or the donor molecule. In case of the *ABO* gene, four amino acids differ between the A and the B enzyme and this alters the donor substrate from *N*-acetylgalactosamine to galactose, and the most common cause of blood group O is a deletion of a nucleotide, 261delG, that truncates the enzyme so that most of the catalytic domain is lost, leaving an nonfunctional protein.[31]

Polymorphisms

Many carbohydrate blood group genes are highly polymorphic and the expression of the antigens varies depending on mutations in the coding region or in introns affecting splice sites. The number of alleles with normal, weak or no expression within the different carbohydrate blood group systems, are listed in Table 3. Data have been collected from the ISBT website (www. blood.co.uk/ibgrl), the Blood Group Antigen Gene Mutation Database (dbRBC at www.ncbi.nlm.nih.gov)[6] and a review article by Storry & Olsson.[32] The numbers may have changed since the data were collected during March 2011.

Table 3. Number of known exons and alleles within the carbohydrate blood group systems

System symbol	Gene	Exons	Antigens	Normal expression	Weak expression	No expression
ABO	*ABO*	7	A, B	26	97	58
P1PK	*P1PK*	4	P1, P^k	3*	0	26**
LE	*FUT3*	3	Le^a, Le^b	12	0	38
H	*FUT1*	4	H type 2	2	23	17
H	*FUT2*	2	H type 1	12	3	18
I	*GCNT2*	3	I	1	3	6
GLOB	*B3GALNT1*	5	P	1	0	8

* P^1/P^2 alleles
** *p* alleles

Gene regulation

Protein synthesis begins with transcription and translation of genes, and this is a highly regulated process. DNA is compactly folded together with histones in the chromosomes in a structure called chromatin. For the RNA polymerase to be able to bind promoter regions to initiate transcription, the gene has to be accessible by unwinding of the double-stranded DNA. Histone acetyltransferase enzymes dissociate DNA from the histone complex allowing transcription to proceed.

RNA polymerase recognizes specific sequences in the promoter region such as the TATA box or CpG islands to which it can bind together with a protein complex and thereby initiate transcription.

Transcription can be up-regulated by transcription factors (TF) binding to specific sequences (*trans*-activation). Enhancer regions that can be located at a distance from the promoter region can fold back to the promoter region and enhance transcription (*cis*-active elements). Repressors are proteins or protein complexes that down-regulate transcription. Another mechanism for down-regulating gene expression is methylation of CpG islands within the promoter region.[33,34] As soon as the RNA polymerase starts the synthesis, the hnRNA produced is being modified. First the 5′end is "capped" to protect the RNA from degradation. When the polymerase reaches the polyadenylation signal at the end of the transcript a poly(A) tail is added.[35] The final modification is the out-splicing of the introns from the transcripts. A fully processed mRNA includes a 5′ cap, 5′ untranslated region (UTR), coding region, 3′ UTR, and poly(A) tail. The 5′and 3′UTR can function as post-transcriptional regulatory regions binding regulatory elements or making loops which prevent ribosomes to bind and start translation.

Transcription factors

Proteins needed for initiation of transcription, are defined as a TF. These can act by binding directly to the DNA strand or by binding to other factors or to the RNA polymerase. The promoter is recognized by TFs rather than the polymerase in eukaryotes.[36] Many *cis*-acting sites, that are parts of promoters or enhancers, are targets for TFs. General factors are needed for initiation at all promoters, forming a complex round the transcription start point (TSP). Upstream factors recognize short specific DNA sequences, located upstream of the transcription start point, e.g. Sp1 which binds GC-boxes. These factors are ubiquitous and can bind to any promoter that has the appropriate binding site. Inducible factors function the same way as upstream factors but have a regulatory role and the binding sites are

called response elements. They are synthesized in specific tissues at specific times. TFs often work as complexes of two or more proteins acting together.[37]

Alterations in non-blood-group genes affecting blood group expression

Variation in antigen expression is not only dependent on polymorphisms in the genes but can also depend on mutations in regulatory regions, binding sites for transcriptions factors but also on mutations in transcription factor genes lowering the expression level of these proteins. Several examples have been found and some of them related to blood group antigen expression are mentioned in the following. It should be noted that some regulators also affect blood group expression in a non-transcriptional way as will be apparent below.

Mutations in the Golgi GDP-fucose transporter gene (*GFTP*) result in a defect of fucose metabolism leading to Leukocyte Adhesion Deficiency type II (LAD II) causing recurrent bacterial episodes during the first years of life, dysmorphism, short status, and mental retardation. These patients have the Bombay blood group phenotype, lacking the H antigen and thereby also A and/or B antigens, and they lack sialyl-Lewis X (sLeX) ligand for the selectins on the leukocyte surface, a finding that explains the susceptibility to infections.[38-40]

Mutations in the *EKLF* gene can result in Congenital Dyserythropoietic Anemia (CDA) where no CD44 is expressed.[41,42] Normally CD44 carries the Indian blood group antigens, so in these cases the RBC phenotype was In(a–b–) as well as Co(a–b–), antigens carried on Aquaporin1. Thus, antigens encoded by at least two independent blood group genes were affected. In another unusual phenotype In(Lu), the Lutheran antigens are missing.[43] In(Lu) has been described as a rare unlinked suppressor of the Lutheran antigens, and was first reported in 1961[44] and subsequently in several other families[45-48] where no inherited linkage to the Lutheran gene was found. The RBCs of these individuals also showed reduced expression of the P1, I and AnWj antigens. Using elution/adsorption, low expression of the Lutheran antigens were found, explaining why no antibody against LU antigens has been detected in the In(Lu) phenotype, as opposed to the recessively inherited Lu(a–b–) phenotype that is due to mutations in the *LU* gene.[49] In 2008 Singleton *et al* found that heterozygous mutations in the *EKLF* gene were found in most (21 of 24) samples from In(Lu) individuals.[50] Another Lu_{null} phenotype, the rare X-linked Lu(a–b–) status is caused by a mutation in the stop codon of the *GATA-1* gene, resulting in 41 additional aa.[51]

The Tn syndrome is a disease where subpopulations of blood cells carry an exposed Tn antigen.[52] The rarely exposed Tn antigen is an incompletely glycosylated membrane glycoprotein, and most sera contain anti-Tn. A mutation in the gene for the C1GalT1-specific chaperone 1 (*C1GALT1C1*), a molecular chaperone (cosmc) required for expression of T-synthase, the enzyme that glycosylates the Tn antigen, has been associated with this rare phenotype.[53]
Other phenotypes suggested to be due to a regulator or inhibitor gene are the A_y subgroup of the ABO system and the In(Jk) phenotype of the JK system but in these cases the genes responsible have not yet been identified.

Alu *repeats*

Alu repeats are short (~300 bp) repeated sequences dispersed throughout the entire genome, mostly in introns, 3´untranslated regions of genes and intergenic regions.[54,55] They account for about 10% of the total genome and consist of two similar monomers joined with an A-rich linker. *Alu* repeats are only found in primates and are derived from cytoplasmic 7SL RNA which is a signal recognition particle. The *Alu* sequences are polymorphic, have no open reading frame (ORF) and are normally not expressed but can be inserted in functional genes and interrupt their expression or change the translated aa sequence.[56] There is also a high proportion of CpG islands which are easily methylated and may therefore influence gene regulation. Since the *Alu* is so common and also can be found in reverse directions it can form loops influencing gene regulation. Inverted *Alu* repeats in the 3´untranslated region may influence trafficking of gene products.[57] *Alu*-containing exons are mostly alternatively spliced.[58] A polymorphism within an *Alu* element in the promoter region of myeloperoxidase has been associated with atherosclerosis and Alzheimers disease.[59] A novel polymorphic transcript of the *A4GALT* gene that we described in Paper III, consists mostly of such an *Alu* repeat.

Erythropoiesis

Haematopoiesis which takes place in the bone marrow, but in the liver during the foetal period, is the proliferation of haematopoietic stem cells (HSC) to all the mature cells found in peripheral blood.[60] During haematopoiesis the HSC will first differentiate to the common lymphoid progenitors (CLP) and common myeloid progenitors (CMP). CMPs will further differentiate to granulocyte/monocyte progenitors (GMP) and megakaryocyte/erythrocyte progenitors (MEP). Megakaryocyte progenitors will then mature to platelets while

erythrocyte progenitors will evolve through several stages to reticulocytes and then to mature RBCs. Thus erythropoiesis is the cell development from the hematopoietic stem cell in the bone marrow to a mature RBC in peripheral blood. (Figure 4) The different steps are controlled by the influence of general and lineage specific transcription factors and cytokines.[61] TFs like SCL, GATA-2, LMO2 and AML-1 are expressed during the differentiation from HSC to CMP. When the erythro-specific TF GATA-1 is up-regulated GATA-2 is down-regulated and both of these TFs use the same binding site. Other erythroid-specific TFs like FOG-1 (Friend of GATA-1), EKLF and NF-E2 are essential for development of mature RBCs.

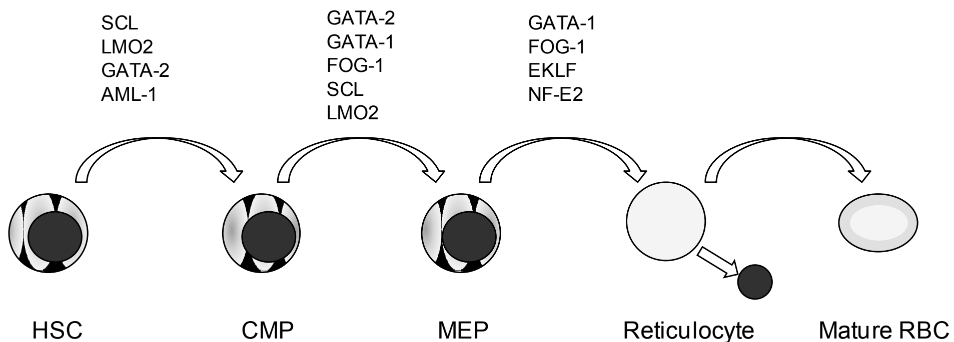


Figure 4. Transcriptional regulation during early erythropoietic differentiation from the hematopoietic stem cell (HSC) to the mature RBC, showing the most important transcription factors involved in the different stages of lineage commitment.

The ABO blood group system

In 1900 Landsteiner discovered the first blood group system, ABO,[62] which made it possible to transfuse blood between individuals safely. This is the clinically most important blood group system in transfusion medicine and in solid organ transplantation, due to the naturally-occurring antibodies made against all ABO antigens not carried on the individual´s own RBCs. Since the ABO blood group is typically inherited in a straightforward Mendelian manner it has also been important in paternity investigations. The majority of the ABO antigens on RBCs are carried on glycoproteins, whilst approximately 10% are carried on glycolipids.[3]

Antigens and phenotypes

A and B are the two major antigens belonging to the ABO blood group system, although two other antigens, A_1 and A,B,[63-65] have also been acknowledged by the ISBT. A and B are inherited in a co-dominant manner, i.e. they can both be expressed if both *A* and *B* alleles are present. Both A and B antigens are created by adding a hexose to the H antigen and thereby covering its antigenicity, but if the *ABO* gene is inactive the H antigen will be more exposed and the phenotype O. The most common phenotypes involved in the ABO system are listed in Table 4. In addition, there are several phenotypes with weak A and/or B expression, mostly due to alterations in the *ABO* gene. Inactivating mutations in the *FUT1* and *FUT2* genes lead to the Bombay phenotype in which the H antigen is missing and thereby also the A and B antigens, independent of the *ABO* genotype.[66] The Bombay phenotype (O_h) is rare but important in transfusion medicine since it is hard to find compatible blood for patients with this phenotype. Para-Bombay phenotypes arise when *FUT1* is inactivated whilst *FUT2* is intact, or if the *FUT1*-derived fucosyltransferase is weakened, independent of *FUT2* status.[66] The A_h and B_h phenotypes are H-negative but very weakly positive for A and B, respectively, and occur in the presence of *A* or *B* genes on the para-Bombay genetic background.

Table 4. Phenotypes and antibodies involved in the ABO blood group system

Phenotype	Frequency (Sweden)	Antigens present on RBC	Antibodies in serum
A_1	35%	A	Anti-B
A_2	9%	A	Anti-B, Anti-A1*
B	12%	B	Anti-A
A_1B	4,5%	A, B	none
A_2B	1,5%	A, B	none
O	38%	H	Anti-A Anti-B
Bombay	rare	none	Anti-A Anti-B Anti-H

* not always present

Biochemistry

The biosynthetic pathways of the ABO system were clarified by Morgan & Watkins and Kabat.[67,68] The precursor of the A and B antigens is the H antigen found in blood group O.

There are many different kinds of precursor chains, but the most important are types 1-4. Type1 chains are synthesized in non-erythroid cells and are found with secreted ABO antigens in plasma and other body fluids. Type2 is mainly found on RBCs, while Type3 harbour repetitive A-associated chains on RBC glycolipids.[69,70]

FUT1 and *FUT2* are genes that encode two highly homologous 2-α-L-fucosyltransferases that catalyse the addition of a L-fucose molecule to the galactose on Type2 and Type1 precursor chains respectively.[66] The fucose is necessary for the 3-α-*N*-acetylgalactosaminyltransferase (A) or the 3-α-galactosyltransferase (B) to be able to add the respective sugars to the precursor chain and both donor sugars are bound with the same α1,3-glycoside linkage (Figure 5).

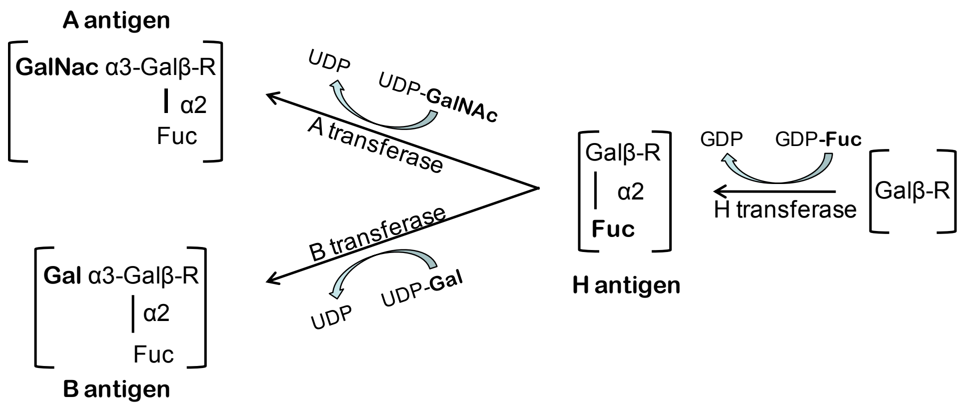


Figure 5. Schematic view of the biochemical synthesis of the H, A and B antigens

Genetics

The *ABO* gene was cloned in 1990[71,72] and it was clarified that the same locus encodes both the blood group A and B transferases. The gene has seven exons (Figure 6), and the ORF includes the first six exons and most of exon seven.[73,74] The gene was first assigned to the long arm of chromosome 9q34 in 1976[75] and was later confirmed by fluorescent in situ hybridisation in 1995.[73] An alternative exon 1 (exon 1a)[76,77] was described in 2001 but appears to be of only limited use, at least in CD34-positive cells where exon 1a-containing transcripts only constitute ~2% of all transcripts.

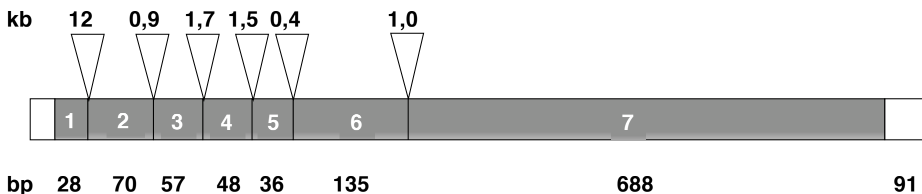


Figure 6. Genomic view of the exon (in bp) and 91bp and intron sizes (in kb) of the *ABO* gene. Short 5´UTR which can vary insize and a 91bp 3´UTR.

The blood group allele A^1 [A101] is considered the consensus gene to which all other alleles are compared when SNPs and mutations are annotated. *B* [B01] alleles differ from A^1 by seven nucleotide substitutions in the coding region of which four lead to aa changes.[31,72] These in turn result in the change of enzyme

substrate preference from *N*-acetylgalactosamine to galactose that can be added to the H antigen acceptor. The O^1 [O01] allele, which results in a null phenotype where addition of donor sugar to the acceptor substrate cannot take place, differs only from the A^1 allele by a single nucleotide deletion at position 261 (delG). This frameshifting polymorphism in exon 6, causes a premature stop codon and consequently a non-functional truncated protein.[31] O^{1v}[O02] alleles[78] have the same deletion and nine additional mutations in exons 2-7. Another null allele named O^2[O03][79] does not have this nucleotide 261 deletion but has a crucial substitution at nucleotide 802 from G to A which changes aa 268 located in the catalytic cleft of the enzyme from glycine to arginine. This aa substitution effectively blocks the donor sugar´s access to the enzymatically active site and prevents synthesis of A/B antigens, although it has been questioned if the resulting enzyme is completely inactive or not.[80] Beside this critical mutation it also has four other nucleotide substitutions within the ORF. One of those mutations, 53G>T will be discussed in Paper II. The A^2 [A201] allele has a nucleotide substitution 467C>T and a deletion of one of three cytosines at positions 1059-1061 which prolongs the reading frame, extending the enzyme by 21 aa. In addition to these major alleles many other polymorphisms have been found in close to 200 alleles, many of which are associated with weak A or B antigen expression, see Table 3. Apart from missense mutations in the coding region affecting the affinity for the donor or acceptor sugar, mutations can also affect splice sites,[81] Golgi localization[82] or the initiation codon,[83,84] all influencing the enzyme level or efficiency. While the N-terminus and the transmembrane region of the enzyme seem to influence the Golgi localisation,[85,86] the C-terminus affects the binding to the H antigen acceptor.[87] It is still not known if the N-terminally truncated ABO transferases exist, but alleles with mutations abolishing the start codon surprisingly seems able to produce A antigen[77] which implies at least one escape mechanism. In addition to inherited weak A and B loss of A antigen expression can be seen in some patients with haematological disease, mostly associated with acute and chronic myeloid leukaemia[88] and also sometimes during pregnancy.[89]

Polymorphisms in the ABO *gene*

Besides the polymorphisms in the coding region of the gene sometimes causing weak or even complete lack of antigen expression, the intronic regions of the gene have been investigated for the most common and important alleles. Sequence variations in intron 6 was published in 1998[90] and for introns 2-5 in 2003.[91,92] The large intron 1 (>12 kb) has not yet been published for different alleles. The

enhancer region has also been sequenced and variations linked to the common alleles,[93,94] but systematic detection of allelic polymorphisms in the up-and downstream regulatory regions was first published in Paper I. The knowledge of allele associated polymorphisms is of use to identify hybrid alleles, to perform allele-specific amplification of particular regions of the gene and also in phylogenetic studies. In addition, there still ABO phenotypes (including the common A subgroup A_3) which are not yet explained and alterations in non-exonic parts of the gene should be considered in such cases.

Hybrid genes

Most cases of weak antigen expression are due to simple nucleotide substitutions in the gene but in some cases these substitutions are not just single base exchanges but a recombination of two alleles forming a hybrid of two *ABO* genes.[81,91,95,96] Sometimes a part of one allele is substituted in another allele which gives two breakpoints, a change often due to gene conversion. Since the *ABO* gene is so polymorphic and these variations in both exons, introns and up- and downstream regions of the gene have been mapped, it is possible to narrow in on the probable breakpoint, at least between some alleles. These breakpoints can occur both in exons and introns. One such hybrid formation between an O^2 allele and a *B* allele giving rise to weakened B expression is described in Paper II.

Transcriptional regulation of the ABO *gene*

Most studies on *ABO* gene regulation have been performed in gastric cancer cell lines and are not necessarily true for haematopoietic cells since many regulatory factors are tissue-specific. As is the case for many other glycosyltransferase genes,[97-99] no TATA box is apparent in the *ABO* promoter. Instead the RNA polymerase binds to the CpG island, located close to the translation start site making the 5´UTR short, varying from 45 to 12 bp.[74] An alternative exon 1 (designated 1a) was noted in about 2% of the transcripts from cultured CD34-positive cells.[76] Although these transcripts lack the usual translation start site in exon 1 they could produce a functional enzyme, probably by using an in-frame ATG further into the gene.[77] However, there is yet no proof of such N-truncated ABO glycosyltransferases existing naturally.

The transcription factor Sp1 has been shown to bind to the *ABO* promoter and when the transcription factor´s binding site was mutated the transcript levels decreased in a haematopoietic model system.[100] An enhancer region located almost 4 kb upstream of exon 1[101] has four repetitive 43-bp units in alleles A^2, O^1, O^{1v} and *B* while A^1 and O^2 only have one 43-bp unit (Figure 7).[93,94]

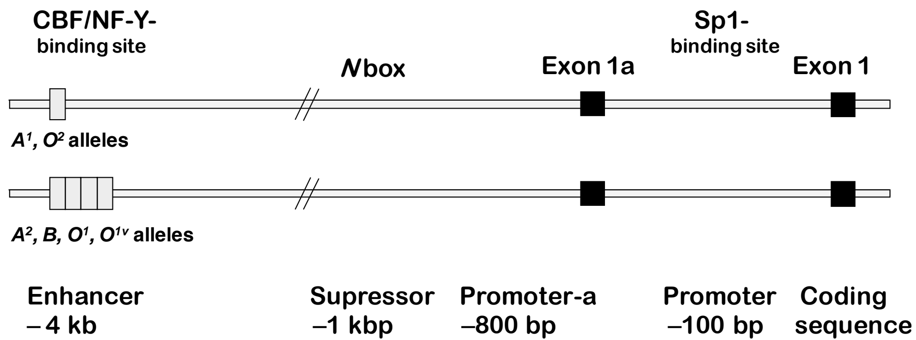


Figure 7. Schematic view of the *ABO* promoter region of different alleles, The enhancer repeat units are shown as grey boxes and exons as black boxes. Transcription factors binding sites and the regulatory N-box region are marked above the figure and sizes of regulatory regions below.

According to Luciferase assay results with constructs containing one to four repeat units, the 4-repeat unit gives a more than 100-fold increase in transcripts than an enhancer with only one unit, at least in a gastric cancer cell line.[102] It has also been proposed that some ABO subgroups may depend on alterations in the repetitive enhancer.[103] Transcription factor CBF/NF-Y binds to each repeat unit and increases transcription.[101] The gene is also repressed by an N-box located about 1 kb upstream of exon 1.[104]

Tissue distribution

The ABO blood group antigens are actually histo-blood group antigens which means they are not restricted to RBCs but also present on many cell types.[105] Thus, the term histo-blood group system has become increasingly used for some of the carbohydrate-based blood group systems (both expressions are used in this thesis). Lymphocytes from secretors express A and B antigens but lymphocytes from non-secretors do not, since the antigens are absorbed from the plasma.[106] Granulocytes and monocytes do not appear to express ABO antigens at all.[107] Platelets have both intrinsic (type 2) and absorbed ABO antigens (type 1).[108] Other example of cells expressing ABO antigens are skin, vascular endothelium, renal tissues, digestive and respiratory epithelium.[109] Because of the naturally-occurring antibodies, ABO compatibility is therefore important to take into consideration not only in transfusion medicine but also in solid organ transplantation.

The P1PK blood group system

The P1PK blood group system has a complicated nomenclature history. Landsteiner and Levine discovered a new blood group antigen in 1927[110] which is now named P1. They also found that this antigen was absent in about 20% of the European population. Originally both the antigen and the blood group system were called P. In 1951 an antibody specificity was found and named anti-Tj^a.[111] The corresponding antigen was called Tj^a and the blood group system designated Jay. We now know that this antibody can in fact be divided into three antibody specificities aimed at the P1, P^k and P antigens and that the patient had the p phenotype. The P^k antigen was discovered in 1955 by Sanger [112] and she proposed that the previously named P antigen should be called P1. In 1959 Matson *et al.* found yet another antigen which was and still is named P.[113] Both the P^k and P antigens were later assigned to the Globoside collection and accordingly did not make it to blood group system status. The common denominator for these three antigens was that they were all missing in individuals with the p phenotype. When the genetic background for human globoside deficiency, and thereby the reason for lack of the P antigen was established by Hellberg *et al* in 2002,[114] this antigen was transferred to a novel blood group system which was named GLOB. In 2010 the connection between the P1 and the P^k antigens was clarified (see Paper III), and the P^k antigen thereby moved from the Globoside collection to the P blood group system which was renamed P1PK.

Antigens and phenotypes

At the moment, the P1PK blood group system includes two antigens P1 and P^k, and the phenotypes P_1 and P_2 (also named P1–) and the null phenotype p. The P antigen in the GLOB blood group system is also involved in these phenotypes, since P^k is the precursor of the P antigen (Table 5). Crucial mutations in the *A4GALT* gene, responsible for the synthesis of the P1 and P^k antigens abolish not only P1 and P^k but also the P antigen and thus result in the p phenotype. Mutations in the *B3GALNT1* gene which eliminate the P antigen, expose more of the underlying P^k antigen, and this will give rise to the phenotypes P_1^k or P_2^k depending on the *A4GALT* genotype (P1+ or –). Alloanti-P1 is typically a weak IgM antibody which rarely causes transfusion reactions.[115,116] Anti-PP1P^k found in individuals with the p phenotype is a potent antibody which can cause severe haemolytic reactions.[111] Also, spontaneous abortions are more frequent in women

with the p, P^1_k and P^2_k phenotypes.[117-120] The P_1 phenotype frequency varies from 90% in Africans, 80 % in Europeans and 20% in Asian populations.[121]

Table 5. Phenotypes and antibodies involved in the P1PK and GLOB blood group systems

Phenotype	Frequency	Antigens present on RBC	Antibodies in serum
P_1	20-90%*	P1, P^k, P	none
P_2	10-80%*	Pk, P	Anti-P1**
p	rare	none	Anti-$PP1P^k$
P_1^k	rare	P1, P^k	Anti-P
P_2^k	rare	P^k	Anti-PP1

* Varies between populations
**Not always detectable/present

Other related antigens

LKE and PX2 are two antigens in the GLOB collection and are associated with the P1, P^k and P antigens. LKE is the terminal antigen, a sialic acid added by a 3-α-sialyltransferase, onto the globoseries oligosaccharide chain (Figure 3).[122] It is a high-prevalence antigen (99%), and the LKE negative phenotype is associated with high P^k expression.[123] Interestingly, the gene encoding the LKE-specific 3-α-sialyltransferase has not yet identified. However, the LKE weak phenotype has been associated with changes in the galactosylgloboside (Gb5)-synthesizing enzyme B3GALT5. [124]

The x2 glycolipid is a β3GalNAc added to the paragloboside precursor chain and is an antigen that is more common in individuals with the p phenotype.[125] It was recently discovered that anti-P in P_1^k individuals reacts against p phenotype RBCs. Serological investigations in Lund and Paris have shown that these antisera contain additional antibodies, one specificity of which is aimed at the x2 glycolipid. It was therefore decided that this high-prevalence antigen (probably only missing on P^k phenotype RBCs) should join LKE in the Globoside collection and be designated PX2.[32] It is still unclear if the P synthase is responsible not only for synthesis of P antigen but also of PX2.

The NOR antigen is a low incidence polyagglutinating antigen, only found in two families, one in the USA[126] and one in Poland.[127] The antigen to which most people have naturally-occurring antibodies reacts with most donor plasma, but

since pre-transfusion crossmatching is performed between patient´s plasma and donor RBCs this does not pose a problem for transfusion of NOR-positive individuals. However, NOR-positive individuals cannot be blood donors because they will regularly give rise to positive crossmatch reactions. Recent studies on the Polish family with inherited NOR antigens showed that a 631C>G mutation in the *A4GALT* gene, predicted to cause a Gln211Glu aa substitution in the 4-α-galactosyltransferase, is associated with the NOR antigen expression.[128] The NOR antigen is a α4-galactose added to the P antigen as shown in Figure 3. All family members expressing the NOR antigen were heterozygous for this mutation while all of the NOR-negative members lacked the mutation. If expression studies can show that a mutated enzyme can give rise to NOR antigens on the surface of RBCs, it is likely that this antigen will be included in the P1PK blood group system as the third *A4GALT*-related antigen.

Biochemistry

All three antigens, P1, P^k and P, are carbohydrate residues located on glycolipids in the RBC membrane. The biochemistry was explained by work of Morgan and Watkins[129] in the 1960s and by Marcus[130] in the 1970s. Ceramide is the lipid tail to which one glucose and one galactose are coupled to form the lactosylceramide base, which is the most common precursor for glycosphingolipids in mammals.[131] Addition of an *N*-acetylglucosamine and another galactose forms paragloboside which is the precursor chain for the P1 antigen that is formed by addition of a galactose residue by 4-α-galactosyltransferase. The P^k antigen is also formed by addition of a galactose but added directly to the lactosylceramide precursor by the *A4GALT*-derived 4-α-galactosyltransferase (Figure 8). The P antigen is created by addition of a β3GalNAc added by a 3-β-galactosaminyltransferase to P^k but this antigen will not be discussed further in this thesis. There have been several hints in the literature that both P1 and P^k are synthesised by the 4-α-galactosyltransferase encoded by the *A4GALT* gene.[132,133] This theory was substantiated by the genetic findings presented in Paper III.

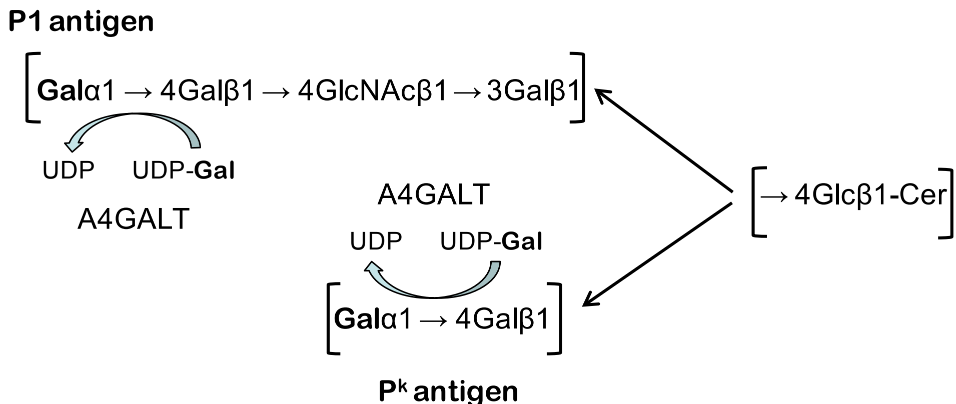


Figure 8. Schematic view of the biochemical synthesis of the P^k and P1 antigens

Genetics

The *A4GALT* gene is located on chromosome 22 and encodes a 4-α-galactosyltransferase which adds galactose to the lactosylceramide creating the P^k antigen also known as Gb3 or CD77. This gene was cloned in 2000[134-136] but although the gene behind the P^k antigen was identified, P^k was not considered part of a blood group system but was kept in the Globoside collection since its relation to the (former) P blood group system (now P1PK) was not clear. The gene has three exons, with the entire coding region (1062 bp) in exon 3 (Figure 9).

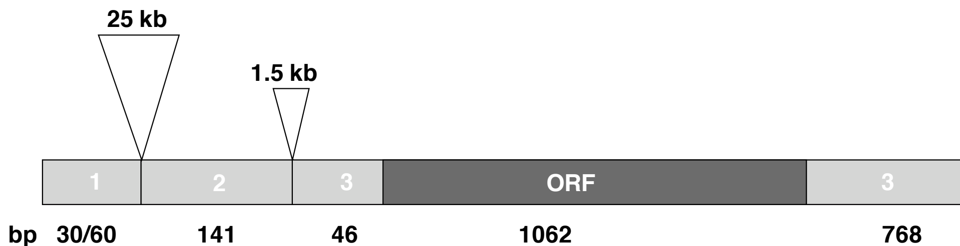


Figure 9. Schematic picture of the *A4GALT* gene. Exons are shown as grey boxes with white numbering. The Open reading frame (ORF) is shown as a black box. The exon sizes and sizes of the ORF and UTR in exon 3 are marked below (in bp) and intron sizes above the figure (in kb).

Mutations in the gene that stop the production of a functional enzyme cause the p phenotype, which not only lacks the P^k, P and LKE blood group antigens but also P1.[134,137,138] This suggested that *A4GALT* might be involved in the synthesis of P1 antigen although no polymorphisms in the gene had so far been able to explain the

P_1/P_2 phenotypes.[134] Alternative theories included involvement of other tightly linked α4GalT genes (cf. *FUT1/FUT2*) or nearby genes encoding protein modifiers that could alter the α4GalT characteristics. This was based on an earlier pedigree study indicating a similar chromosomal location at 22q for the P1 trait[139] as was later shown for the P^k-synthesizing *A4GALT* gene.[134,140]

Transcriptional regulation of the A4GALT *gene*

The regulation of the *A4GALT* gene has not been the subject of many investigations despite the multiple clinical associations found for the secondary gene product, the P^k antigen and possibly also P1. In most cases the p (null) phenotypes have been explained by mutations in the coding region abolishing the enzymatic activity.[134,137,138,140] Constructs containing the coding sequence of exon 3 could add a galactose to change a ceramide dihexoside (CDH) to a ceramide trihexoside (CTH) and also express P^k on Namalwa cells, a cell line which normally does not express P^k.[134] However, the cloned *A4GALT* gene could not add galactose to the paragloboside acceptor to produce P1 antigen in High Five cells[134]. Another paper reported that in studies on cells transfected with *A4GALT*, they found high expression of P^k antigen but also low amounts of P1 antigen on the cell surface was observed, but P1 antigen was mainly found in the cytoplasm.[132] P_2 cells showed low expression of P1 by flow cytometry, when treated with papain or pronase, which could mean that P_2 cells are not negative but have much less antigens which cannot be detected with the standard methods. Using reverse transcriptase PCR these investigators found lower levels of *A4GALT* mRNA in P_2 samples than in P_1. They made attempts to correlate polymorphism in the assumed promoter region, to the P_1/P_2 phenotypes and suggested that polymorphisms at positions –550insC, –164C>T and –160A>G may be responsible for the P_2 phenotype (Figure 10).

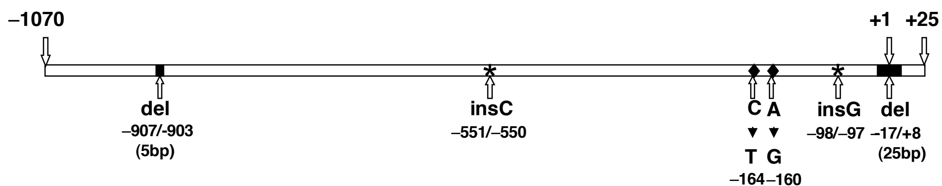


Figure 10. Schematic view of 1 kb of the *A4GALT* promoter region with polymorphic sites marked below and the start of exon 1 marked as +1 above.

However, these findings could neither be confirmed in transcription assays by the same authors, nor in two independent European studies.[141,142] In the latter two

studies, it can be speculated that the failure to correlate the proposed promoter polymorphisms with the P_1/P_2 phenotypes could perhaps be due to the fact that only 20 percent of the Asian population (as opposed to 80% among Caucasians) has the P_1 phenotype and only 7 P_1 samples were analyzed in the Japanese study.[132] Although the Swedish investigators also found several other polymorphisms in the promoter region as well as the 3´region downstream of the gene,[141] none could explain the enigmatic P_1/P_2 phenotypes. Okuda *et al.* who studied the promoter region described that it lacks a TATA box but is GC-rich and has three binding sites for the transcription factor Sp1.[143] Expression studies with Luciferase assays showed that the GC-rich region is essential for gene expression and also that the promoter region has domains repressing expression.

Tissue distribution

Both P^k and P1 antigens are expressed on other haematopoietic cells.[144] P^k is also expressed on smooth muscle cells of the digestive tract and urogenital system[145] and in kidney where high expression of P^k has been associated with haemolytic uremic disease (HUS).[146] Both antigens are also expressed on urogenital epithelial cells where it is the cellular receptor for P-fimbriated *E.coli*.[147] This is also the basis for impaired binding of such pathogens to mucosa from individuals with the p phenotype.[148]

Summary and discussion of the present investigation

General aims

The ABO blood group system

- To clarify allelic differences in gene expression and gene regulation by investigating the qualitative and quantitative aspects of *ABO* transcripts with RACE technique, fragment analysis and by measuring transcript levels and allelic polymorphisms in the up- and downstream regions.

The P1PK blood group system

- To study the regulatory mechanisms of the *A4GALT* gene and to investigate whether any of these could clarify the unexplained question regarding the P_1/P_2 RBC phenotypes. The ultimate goal here was to identify the genetic basis of the P blood group system to be able to predict P1 antigen expression by analysis at the DNA level and explain the interindividual strength in P1 reactivity on RBCs.

Commentary on the materials and methods used

Details on manufacturers and products used for the published experiments described below can be found in the original papers.

Blood, bone marrow and cell lines

In all papers peripheral blood from apparently healthy donors drawn into either EDTA or ACD anticoagulant solutions was used. Anonymized donors were used for most of the experiments, but in some cases donors with known ABO or P1PK phenotypes were required. The only patient studied was one with A_1B_{weak} phenotype investigated in Paper II, for which samples were referred specifically. Blood samples were used to investigate cell surface expression on RBCs and for DNA or RNA extraction from the leucocytes.
The bone marrow samples used in Paper I and for the additional work on P1PK, were taken from anonymous voluntary donors and was used for CD34+ cell isolations which were then cultured towards erythroid maturation. The purpose was to investigate the *ABO* and *A4GALT* gene expression during erythropoiesis. Both blood and bone marrow donations were obtained following national rules and

guidelines and with informed consent and approval from the Regional Ethics Review Board at Lund University when required.
For the RACE analysis in Paper I the erythroblast cell line K562 (blood group O) and the gastric cancer cell line MKN 45 (blood group A) were used. The Ramos cell line was used for RACE analysis to determine the transcription start site of the *A4GALT* gene and also to investigate transcript variants in Paper III. Ramos cells were also used in the Luciferase assay to measure promoter activity in Paper III. In Paper II, the cervical cancer HeLa cell line was used for transfection of a normal gene and its mutated counterpart to measure expression with flow cytometry.

PCR and sequencing

PCR was used in all papers for amplification of DNA fragments for sequencing. For shorter fragments (≤ 2,000bp) a single PCR run of 30 cycles was used, but for longer fragments (e.g. 2,000-4,000 bp), nested or semi-nested amplifications were optimized. Promoter sequences in both the *ABO* and the *A4GALT* gene are very GC-rich and a special polymerase kit (GC-rich PCR system, Roche) was used to amplify these regions. All DNA sequencing was performed using the BigDye terminator kit and run on the Avant 3130. Sequences were analysed with the SeqEd analysis program. Different genotyping methods were set up in Paper III:
PCR-ASP is a genotyping method that specifically amplifies sequences, even if they only differ by as little as one nucleotide substitution. Primers are designed so that the polymorphic nucleotide position is placed at the 3′end of the primers, one primer for each variant. These primers are used in separate reactions together with a common reverse or forward primer. If the sample is homozygous for either allele, only one of the two amplifications will be positive and if the sample is heterozygous both amplifications will be positive. For optimization of such an assay, heterozygous samples are used as the positive control and samples homozygous for the opposite allele are used as a negative control. In addition a water control is used to exclude contamination and an independent primer pair amplifies an invariable control fragment from another gene to prove the presence of DNA and functional amplification conditions in each tube.
PCR-RFLP can be used if the polymorphic site is recognized by a restriction enzyme (RE). A PCR product is amplified and then digested with an RE that cuts one of the amplified alleles into at least two parts so that band of different sizes can be visualized on the detection gel, while the other allele gives only one band. It is beneficial to design the PCR primers so that the bands will be of different sizes. For this type of assay, controls typically include all three genotypes and a water control. For diagnostic assays, it is often recommended that an additional

(possibly artificially introduced) RE site is used as an in-tube control to ensure that the enzyme works under sufficient conditions to be able to cut.
The allelic discrimination (AD) method is performed with two primers labelled with different fluorophores that can distinguish between the two allelic variants. After amplification the products are run on a RQ-PCR machine that measures the amount of fluorescence at the two different wavelengths. Controls of different genotypes as well as water controls are required.

Transcript quantification

Real-time Quantitative PCR was used in all papers to measure mRNA transcript levels. Since the amount of cDNA varies depending on both the amount of mRNA in the total RNA used, and also depending on the efficiency of the cDNA synthesis, a housekeeping gene that has a stable expression in all cells is used to be able to compare different samples and normalize the obtained values. The housekeeping genes used in this thesis are either *18S* or *β-actin*. C_T (cycle threshold) is the cycle number in the PCR run at which the amount of fluorescence reaches a certain threshold value. The highest value was used as a calibrator and was set to 100%, to which all other values were compared.

Bone marrow culture

In Paper I and in the additional work on P1PK, CD34+ cells were selected from the mononuclear fraction of fresh primary human bone marrow using magnetic beads. The cells were then cultured in StemSpam®SFEM solution with addition of erythropoietin (EPO), interleukin 3, hydrocortisone, granulocyte/macrophage colony-stimulating factor and 25% fetal calf serum towards erythroid lineage for eighteen days. Cells were harvested every third day, one aliquot frozen for Western blot analysis, another aliquot collected in stabilising buffer for RNA preparation and subsequent cDNA synthesis. Cells from selected bone marrows were prepared by cytospin and stained with May-Grünewald dye to evaluate the differentiation of the cultured cells. The yield of CD34+ cells from the bone marrows differed and therefore the number of cells possible to collect at each time point varied. For quantification of transcripts, these differences were normalized as described above.

RACE

Rapid Amplification of cDNA Ends was used in Papers I and III to determine both the transcription start site (5´RACE) and also the end of the transcript (3´RACE). The mRNA was isolated from total RNA and analysed with a FirstChoice™

RLM-RACE kit. This kit selects for capped transcripts to ensure that only full-length mRNA will be amplified. The mRNA is first treated with Calf Intestine Phosphatase (CIP) to remove phosphate groups from degraded mRNA. rRNA, tRNA and DNA. The mRNA is then treated with Tobacco Acid Pyrophosphatase (TAP) to remove the cap from the 5′end and an adaptor is then ligated to the decapped mRNA, and cDNA is synthesised and amplified. 3′RACE is performed using an adaptor starting with a poly(T) sequence which anneals to the poly(A) tail of the mRNA. The adaptor sequence is used with a forward gene-specific primer to amplify the 3′ sequence. This kit differs from many others by detecting only capped mRNAs.

Transfection

The transfection used in Paper II was performed with the Lipofectamine™ 2000 kit. A normal and an *in vitro* mutated *B* gene were cloned and co-transfected with a green fluorescent protein (GFP) vector into HeLa cells. The latter was used to be able to assess the transfection efficacy. A plasmid (pcDNA3.1) used by other investigators previously was chosen for the *ABO* inserts. Cells were cultured for 48 hours before antigen expression was measured by flow cytometry. HeLa cells were used since they express the H antigen but not endogenous A or B antigens. HeLa cells are also adherent cells which are easier to transfect than suspension cells. They have been used extensively in previous studies[149-151] to evaluate the effect of alterations in the *ABO* gene and were therefore chosen so that comparison could be made with previous studies.

Flow cytometry

The flow cytometric analysis used in Papers II and III was run on a FACScan and analysed with Cell Quest software v3.1f . B antigen expression on the RBCs from the A_1B_{weak} sample was analysed using anti-B clone 9621A8 in Paper II. When detecting expression on the transfected HeLa cells, another primary antibody, anti-B Diaclon LB-2 was used to be able to detect small differences in expression.
In Paper III, P1 expression was detected with a monoclonal anti-P1 and P^k expression with monoclonal anti-CD77.

Luciferase assay

In Paper III, promoter activity of the *A4GALT* gene was analysed with the Dual-Luciferase™ Reporter Assay System and a Glomax 20/20 luminometer.
Fragments corresponding to various lengths of the proximal promoter region were cloned into an expression vector containing the firefly luciferase gene.

Values of firefly luciferase were normalized to the values of *Renilla* luciferase, which was used as an internal control of transfection efficiency.

Methylation detection

Methylation of CpG islands is a mechanism for epigenetic influence on gene expression. Methylation is the addition of a methyl group to the carbon-5 position of the cytosine in a dinucleotide CpG. When treated with bisulfite the normal cytosine will convert to a uracil and will thereby be read as a T when sequenced, while methylated cytosines will be left unaffected and will still be read as a C. Bisulfite-treated *ABO* promoter DNA was sequenced and compared to untreated DNA in the additional work on the *ABO* gene.

Western blot

Western blotting was used in the additional work on the *A4GALT* gene. Approximately 1 million cells were harvested from the bone marrow culture from one of each of the three different genotypes every third day, from days 3 to 18. The cells were sonicated, solubilised and run on a 4-12% Tris-Bis polyacrylamide gradient gel. Separated proteins were transferred by electrophoretic blotting to a Hybond-N membrane. The membrane was hybridized with different primary antibodies and horse radish peroxidase (HRP) labelled secondary antibodies. Antibody binding was detected with enhanced chemiluminescence (ECL) and visualised on ECL film.

EMSA (Electrophoretic mobility shift assay)

EMSA was performed in the additional work on the *A4GALT* gene. In EMSA, proteins that bind to DNA sequences are detected. Short DNA sequences and proteins are incubated under optimal buffer, time and temperature conditions and are then run on a polyacrylamide gel. The separation of the DNA fragments on the gel will then be detected and if the protein has bound to the DNA fragment it will migrate slower on the gel than the fragment without any protein – accordingly, a *shift* in its mobility can be detected. If a mixture of proteins are used the binding protein can be identified by including specific antibodies in the binding reaction incubation stage, which will retard the migration even more – giving a *supershift*. There are different techniques for this and we have used one where the gel is incubated with SYBR®Green dye and the fragments are then visualized with UV.

Paper I

Published data on transcript levels in different *ABO* genotypes were not available so we wanted to investigate if we could detect any variation. We collected peripheral blood to investigate the transcript levels from three donors each homozygous for A^1 [A101], A^2 [A201], O^1 [O01], O^{1v} [O01] and *B* [B01]. Two heterozygous donors with the O^2 [O03] allele (A^1O^2 and O^1O^2) and three *AB* samples were also examined. Although the O^2 allele is less common its principally interesting since it is not inactivated by the usual 261delG as in O^1 and O^{1v} alleles, but instead a mutation 802G>A which prevents the enzyme from using the sugar donor; additionally it has two other mutations; 53G>T and 220C>T. Transcripts were quantified with two *ABO*-specific TaqMan assays, ABO 3-4 and ABO 6-7, binding to two different exon boundaries to be able to detect transcript variants. Surprisingly, no transcript could be detected in samples from the homozygous A^1 or A^2 donors, while all *O* and *B* transcripts were readily detectable. No significant difference was seen in expression levels within the *O* and *B* samples. The regulatory region upstream of the gene has been investigated in expression analysis but no consideration has been taken to the possible nucleotide variation between different alleles, since no data on this had been published. To be able to correlate weak antigen expression with mutation in the regulatory regions, data on the normal sequence of these regions are necessary. In order to see if any differing regulatory motifs could explain this finding, 4,000 bp upstream and 1,800 bp downstream of the gene were sequenced in different alleles. DNA from two individuals homozygous for each of the five major alleles A^1, A^2, O^1, O^{1v} and *B* were collected and we also had access to two homozygous O^2 [O03] DNA samples. Although several novel allele-specific and allele-related polymorphisms were found, none of these could be coupled to the lack of transcripts in the *A* alleles. The knowledge of these polymorphic sites is very useful when allele-specific amplification of the promoter region is needed, and to detect and define crossing-over points in hybrid alleles. The previously investigated enhancer region, which has four 43-bp repeats in A^2, O^1, O^{1v} and *B* alleles while A^1 and O^2 lacks three of those repeats, was verified in this investigation. An *in vitro* study of expression variation with different numbers of enhancer repeats had shown that insertion of the 4-repeat units gave a 100-fold higher expression than one repeat. This did not fit with our results from normal peripheral blood, which showed no transcripts in either A^1 (1 x 43-bp unit) or A^2 (4 x 43-bp unit) samples. In addition, samples heterozygous A^1 and O^2 did not have lower transcript levels than other samples heterozygous for *A* and other *O* alleles. The fact that we did not detect any

A transcript in peripheral blood was intriguing and this had also been found in another study from Asia where A^1B samples were investigated.[152] Taken together, this made us decide to investigate the transcript levels in the bone marrow. CD34+ cells isolated from four healthy bone marrow donors were cultured towards erythroid lineage and cells harvested continuously during the 15 days of culture until all the cells had differentiated into erythrocytes. Transcript levels were measured with real-time PCR, but since all bone marrow donors were heterozygous for *A*, *B* and *O* alleles, a TaqMan assay discriminating *A/B* and O^1/O^{1v} transcripts regarding the 261G/261delG position was designed. In all cell cultures, *A* transcripts were readily detected and in higher amounts than the *O* transcripts. *B* transcripts were found at similar or lower levels compared to the *O* transcripts. The enhancer region in the four marrow donors was sequenced and the expected numbers of repeats were found in three samples but one donor with the genotype A^2O^1 had only three 43-bp units in the A^2 allele. However, this did however not lower the transcript levels compared to the other A^2O^1 donor with normal number (4 x 43-bp unit) of repeats. Furthermore, there was no difference between A^1 (4 x 43-bp) and A^2 (3 or 4 x 43-bp) donors. Interestingly, our data indicates that the number of enhancer repeats does not influence the transcript levels in peripheral blood or bone marrow, which does not support previous investigations in non-haematopoietic cell lines.

Paper II

In this paper we investigated the *ABO* gene sequence of an A_1B_{weak} sample sent to us from Switzerland. Serologically there was a moderately weakened B antigen expression on the RBCs with no anti-B in serum. This sample was genotyped as *AB* but the enhancer region lacked three of the four repeats expected in a normal *B* allele. Sequencing of the exons showed a normal *B* allele sequence except for a substitution G>T at position 53 in exon 2 which normally is found only in O^2 alleles. Further sequencing of the promoter region and the introns, except most of intron 1, confirmed that this allele was indeed a hybrid between an O^2 and a *B* allele with the breakpoint in intron 4. It has been suggested in several papers that the number of repeat enhancer regions relates to transcript levels,[102,103] although we could not find any support for this concept in our work in Paper I. We therefore measured the transcript levels in fresh samples from the A_1B_{weak} donor and compared with AB and B control donors with normal B antigen expression on the RBCs. To minimize irrelevant factors influencing the result, all samples were

drawn at the same time and transported under the same conditions. We did not find lower transcript levels in the index sample compared to the eight controls. If anything, the transcript levels appeared to be slightly higher. All control samples had normal enhancer regions. The only difference in the coding region in the B^{weak} allele compared to a normal *B* was the 53 G>T substitution in exon 2 leading to an aa change from arginine to a leucine (Arg18Leu). To investigate if this nucleotide and also aa substitution in the N-terminal part of the enzyme may influence the antigen expression, we then transfected HeLa cells with a normal *B* allele construct and also a *B* allele where the 53 G>T mutation had been introduced, and measured the antigen expression on the transfected cells by flow cytometry. There was a slightly lower (14%), B antigen expression on the cells transfected with the mutant allele. This minor decrease of antigen expression mirrors the moderately low expression on the patient´s RBCs but the difference is too small to confirm that the mutation is solely responsible for this weak B phenotype even if it was statistically significant. Furthermore, the mean fluorescent intensity of the two transfected cell populations did not differ in the same way. Instead, the mutated transfectants had a slightly higher fluorescence level. Thus the effect of Arg18Leu on enzyme activity is not clear. Even so, we have shown another scenario, where the number of enhancer repeats does not seem to affect the transcription levels of the *ABO* gene, thus questioning the importance of the enhancer for ABO expression on RBCs.

Additional work on *ABO*

Methylation

Methylation status of the proximal promoter (–246 bp to +31bp) of the *ABO* gene in normal cells and malignant bone marrow, and in cell lines K562 and Jurkat was investigated by sequencing of bisulfite-treated DNA (unpublished work). Normal cells showed no methylation in the region examined while most malignant bone marrow cells appeared to be methylated at some positions but no specific pattern could be defined. The Jurkat cell line was completely methylated while K562 was not methylated at all. According to a more recent paper[27] a region from –1,500 to –1,000 bp showed methylation in normal cultured cells, so this area needs to be further investigated to see if methylation influences the regulatory mechanism in normal cells.

Expression in different blood cells

Since the TaqMan assays performed on mRNA from peripheral blood showed that *A* transcripts could not be detected, the question arose which haematopoietic cells actually express *ABO* transcripts. Initially, reticulocytes, young RBCs and possibly to some extent platelets had been hypothesized but not proven to be main contributors since they are the only blood cells known to express endogenous A and B antigens at reasonably high levels. Different blood cells (B cells, T cells monocytes, neutrophils and platelets) were sorted by flow cytometry. RNA was prepared from each population containing between 50,000 and 100,000 cells, and cDNA synthesized with random hexamer primers and/or *ABO*-specific primers. Preliminary data indicate that platelets contained low amounts of *ABO* transcripts while the other cells showed none or inconclusive results. Very low amounts of transcripts in the other sorted cells preparations may be explained by contaminating platelets sticking to the other cells making it difficult to get pure cell populations by sorting. When examining RNA preparations from leucocyte-depleted platelet concentrates taken from donors with A and O phenotypes, *ABO* transcripts could be detected in both cases.

Paper III

In an effort to elucidate the correlation between the P1 antigen and the *A4GALT* gene we started by collecting blood samples from individuals with P_1 and P_2 phenotypes. *A4GALT* transcripts were measured with RQ-PCR and the levels in P_2 samples were significantly lower than in the P_1 samples. We also found a broad variation within the investigated P_1 samples. Several attempts have been made previously to identify the genetic variation which distinguishes between P^1 and P^2 alleles within the *A4GALT* gene´s three exons and the regulatory regions up- and down-stream of the gene, but none with satisfactory correlation has been found.[132,141,142] To see if the difference in transcript levels could depend on any of the many polymorphisms in the promoter region,[132,141] A Luciferase assay measuring transcriptional activity was performed. We transfected the Ramos cell line with cloned promoter regions of various sizes from two promoter sequences with polymorphisms associated with,but not fully correlated to the P_1/P_2 phenotypes. The transcriptional activity was shown to be high in the proximal promoter, and a possible repressor site was identified, but no difference was seen between the two variants with different promoter sequences. While investigating the transcription start point with RACE technique we discovered a short novel

transcript containing exon 1 and a short part of the 23 kb long intron 1 but lacking exons 2 and 3 (Figure 11).

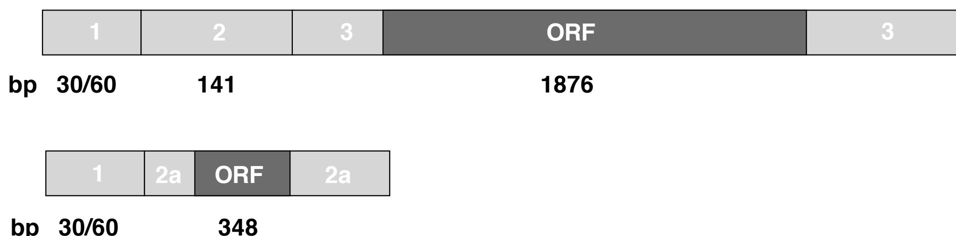


Figure 11. Two variants of *A4GALT* transcripts. Exons are shown as grey boxes with white numbering. Open reading frames (ORF) are shown as black boxes. The exon sizes are marked below (in bp).

The new exon was named 2a, since it appeared to take the place of exon 2. In this exon three polymorphic sites were found, two of them with incomplete correlation to the P_1/P_2 phenotypes and one of them with a preliminary correlation in the few samples first investigated. The latter polymorphism was a C in the P^1 alleles which was substituted to a T in the P^2 alleles. This pattern was confirmed in >200 blood donors with varying P_1/P_2 phenotypes and all P_2 donors were homozygous for T. Interestingly this SNP opens a short reading frame in the P^2 alleles, which potentially could be translated to a 28-aa peptide. Three different methods for P^1/P^2 genotyping were set up of which all were specific and a useful tool for investigating larger number of samples and to distinguish between homo- and heterozygous P_1 samples. A cohort of selected samples, five of each of the three different genotypes P^1/P^1, P^1/P^2 and P^2/P^2 were further investigated comparing different antisera used in the blood group immunology laboratory, transcript levels and RBC antigen expression of both P^k and P1 measured by flow cytometry. All assays showed that P1 and P^k expression as well as *A4GALT* transcript levels were negatively influenced by the presence of a P^2 allele. Homozygous P^1/P^1 samples could be separated from heterozygous P^1/P^2 samples with all methods, only the P^k antigen did not show significantly lower expression but only a trend. In summary, this study showed that polymorphisms in the *A4GALT* gene are correlated with expression levels of both P1 and P^k antigens and that it is now possible to genotype for the P_1/P_2 status. This provides the first direct linkage between these two antigens which for long have been suspected to be more than biochemically related. It was also the final proof required by the ISBT to transfer the P^k antigen

to the P blood group system and change the name of the latter to P1PK, thereby reflecting the antigens encoded by the *A4GALT* gene.

Additional work on *A4GALT*

Based on our findings in Paper III where we clarified that the α4Gal-T enzyme is likely to synthesize both P1 and P^k antigens and that transcript levels in peripheral blood varied depending on the P^1/P^2 genotype, our aim was to explore the underlying mechanism behind the P_1/P_2 phenotypes. The novel transcript we found with the P^1/P^2-distinguishing C/T polymorphism in exon 2a had a short ORF in the P^2 allele, but not in the P^1 allele. We formed different hypotheses regarding how the *A4GALT* transcription level could be decreased: as discussed in Paper III the effect could be on either DNA, RNA or the protein levels. At the DNA level the sequence of exon 2a could be a binding site important for regulation of the gene and the polymorphism could influence this binding. At the RNA level the exon 2a sequence which is mostly an *Alu* repeat could form a loop binding to other inverse *Alu* repeats in the hnRNA and influence splicing of the gene. At the protein level this new ORF could be translated and the small protein from the P^2 alleles could work as a regulatory peptide. We decided to begin investigating the third hypothesis by trying to investigate if this ORF was translated and if it thereby would produce a small protein of 28 aa, which we hypothetically called P2RP – P2-Related Protein. We conducted a database search to investigate if any similar protein had been described. Since there was no antibody available for detection of the protein, a synthetic peptide was ordered (Innovagen, Lund, Sweden) and this was used to immunize two rabbits (Figure12), animals A and C (Agrisera, Vännäs, Sweden).

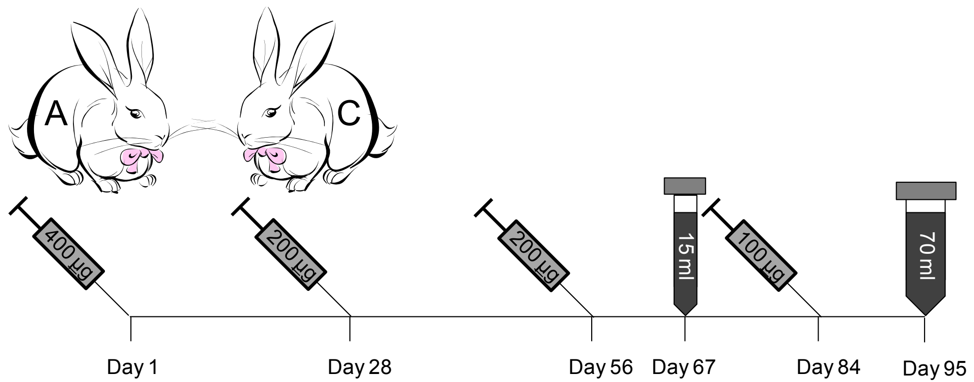


Figure 12. Schedule for immunization of two rabbits with a synthetically produced P2RP peptide. Two rabbits were immunized four times with 400, 200 and 100 µg peptide according to the time line and blood was drawn on day 67 and a final bleeding on day 95.

To investigate the mechanism behind the P_1/P_2 phenotypes, peripheral blood was not a suitable source. Instead we cultured bone marrow towards erythroid differentiation. While the P2RP hypothesis was appealing, a complementary theory was that the EKLF transcription factor was involved in regulation of the *A4GALT* gene. This idea was based on the fact that Singleton *et al.* recently described that mutations in the *EKLF* gene is responsible for the In(Lu) phenotype in which the expression of the Lutherans antigen is inhibited. However, other blood group antigens are also affected by lowered antigen expression.[50] The P1 antigen is one of those antigens that are down-regulated.[46] We therefore began by investigating if there are EKLF binding sites in the promoter region of the *A4GALT* gene and this was done by the Electrophoresis Mobility Shift Assay (EMSA).

Database searches

We also performed a database search using NCBI protein Basic Local Alignment Search Tool (BLASTP) and did not find any identical aa sequence, but surprisingly revealed a number of proteins with significant homology to P2RP. The ones found with the highest homology were mostly described as hypothetical proteins, mostly larger than our protein but the homologous sequence was often found in the beginning or at the end of a larger protein. However, one almost identical protein (29 aa) was described in the *Macaque* monkey. A problem with the P2RP sequence is that almost the entire novel exon consists of an *Alu*-repeat (Figure 13), and that the whole ORF is located within this. An *Alu* repeat is a

sequence dispersed throughout the genome in primates, although it is not supposed to have ORFs[54] and the ATG found in *P²* alleles is quite unusual among *Alu* families.

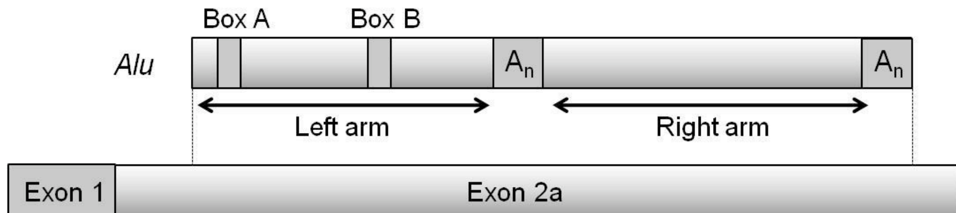


Figure 13. Schematic picture of an *Alu* repeat above with its left and right arms followed by A-rich sequences. Two Polymerase III promoter sites, boxes A and B, are located in the left arm. Below is the novel *A4GALT* transcript with exon 2a that mainly consists of an *Alu* repeat.

Bone marrow culture

Primary bone marrow cells from seven donors with genotypes P^1/P^1 (n=2), P^1/P^2 (n=3) and P^2/P^2 (n=2) were cultured towards erythroid lineage as previously described (Paper I). Cells were harvested with three days interval from day 3 to day 18. One marrow was also harvested at day 0. These cells were divided into aliquots for RNA preparation, protein detection with Western Blot and cytospin preparations to study morphology.

Cell differentiation

100,000 cells from bone marrow culture were spun to glass slides and stained with May-Grünewald technique. Figure 14 shows differentiation stages at day 3 and 15.

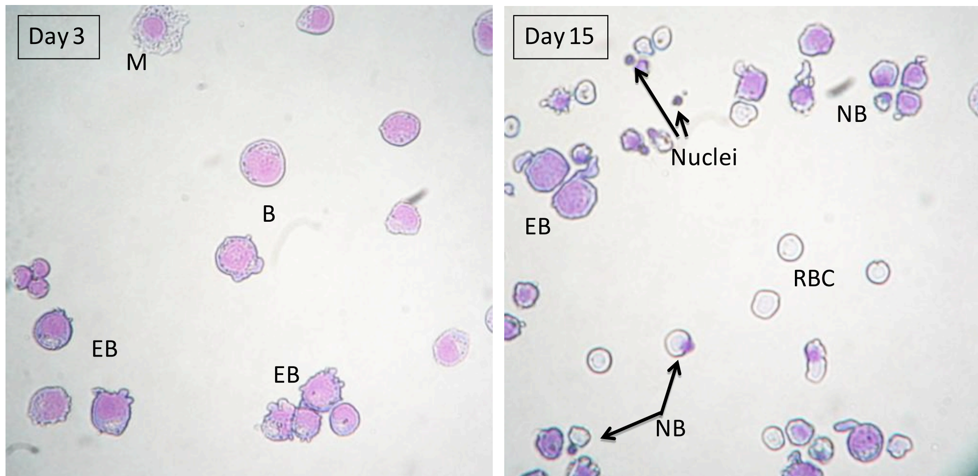
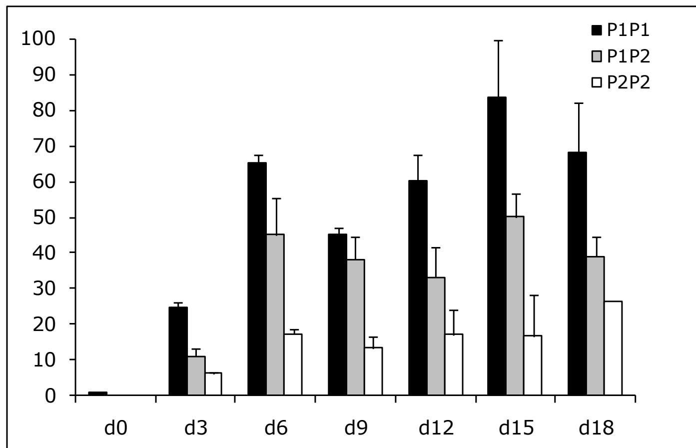


Figure 14. May-Grünewald stained cells from the culture of bone marrow cells into erythropoietic lineage. Cells harvested on days 3 and 15. Day 3 shows immature cells with early signs of erythropoietic differentiation (EB, erythroblast), early blast cells without definitive signs of differentiation (B), and some macrophages (M). Day 15 shows a high proportion of mature RBC and normoblasts at final maturation (NB), some in the process of expelling the nucleus (NB with arrows), and some naked nuclei. A few still immature erythroblasts are also present (EB).

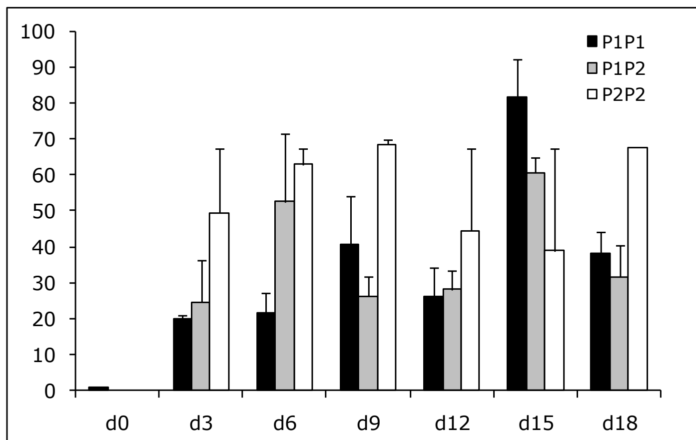
Transcript levels

Cells were harvested from all seven marrows, and RNA and cDNA was synthesized at the six different time points of the culture. The cDNAs were analyzed by RQ-PCR using commercially available assays for detection of *A4GALT* and *β-actin* transcripts. The *β-actin* assay was used as a calibrator when calculating the transcript levels. To detect the novel transcript (designated transcript IV in Paper III), we ordered a custom-made assay from Applied Biosystems. Taking into consideration that the sequence in this transcript contains an *Alu* repeat which is very common in the genome, amplification products were cloned and sequenced to confirm their identity. All clones had the expected sequence including the polymorphism separating the P^1/P^2 genotypes. *A4GALT* (exons 1+2+3) transcript levels showed the same pattern as in peripheral blood, namely that the P^1/P^1 samples had the highest levels. P^1/P^2 samples had intermediate levels and P^2/P^2 samples had the lowest levels, see Figure 15. The transcript levels rose from day 3 to day 6 and then stayed on similar levels. When

looking at the levels of the new transcript (exons 1+2a), the same pattern could not be seen. The levels were the same on all days, except for the P^1/P^1 samples on day 15 had higher levels than on the other days. The levels also seem to differ less between the different genotypes but transcripts from the P^2/P^2 samples were higher than the others at all times except day 15.



Exons 2+3



Exons 1+2a

Figure 15. *A4GALT* transcript detection during bone marrow cultures. The ordinary transcript encoding the α4GalT enzyme was detected with the *A4GALT* probe (designated Exons 2+3), while the novel short transcript including the exon 2a was detected with a separate assay (Exons 1+2a). cDNA from cells of the three different genotypes and from culture days 0-18 were analysed with both assays. Both graphs show the mean values and error bars represent S.E.M. values.

Western Blot

Approximately 1 million cells were harvested from the bone marrow culture from one of each of the three different genotypes at days 6 and 12. Membranes were hybridized with antibodies against α4Gal-T, EKLF, GAPDH and the antibodies from the two immunized rabbits, P2RP-A and P2RP-C (Table 6).

Table 6. Antibodies used in Western blot analysis.

Specificity	Clonality	Species	Manufacturer	Secondary antibody
α4Gal-T	Poly	Rabbit	Atlas	Goat α-Rabbit HRP
P2RP-A	Poly	Rabbit	Agrisera	Goat α-Rabbit HRP
P2RP-C	Poly	Rabbit	Agrisera	Goat α-Rabbit HRP
EKLF	Poly	Rabbit	Santa Cruz	Goat α-Rabbit HRP
GAPDH	Mono	Mouse	Santa Cruz	Goat α-Mouse HRP

The GAPDH antibody showed the expected band of ~40 kD and that the loading was not equal: Day 6 in all three samples had more proteins than day 12. The EKLF immunostain showed bands of an unspecific band of ~40 kD, the specific bands of ~65 kD and also a weaker specific band of about ~55 kD.
The α4GalT antibody gave quite weak bands of ~65 and ~60 kD, but surprisingly also a weak binding to P2RP. The two rabbit antisera gave different results. Rabbit A gave bands at ~65 and ~50 kD and very weak binding to P2RP, while rabbit C gave strong binding to P2RP and a weak band at ~60 kD and a stronger band at ~40 kD (Figure 16).

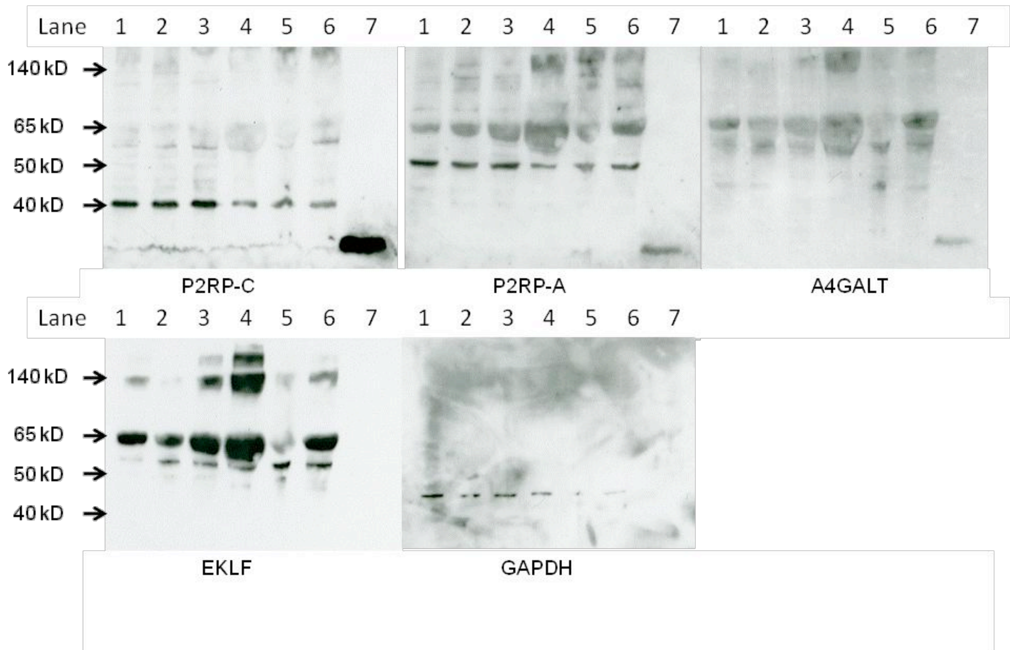


Figure 16. Solubilized cells were loaded in the 4-12% Tris-Bis gel. Lane 1: P^1P^1 day6, Lane 2: P^1P^2 day6, Lane 3: P^2P^2 day6, Lane 4: P^1P^1 day12, Lane 5: P^1P^2 day12, Lane 6: P^2P^2 day12 and Lane 7: P2RP. Blotted membranes immunostained with anti-P2RP A and C, anti-α4GalT, anti-EKLF and anti-GAPDH.

EMSA

One hypothesis was that the P2RP (if translated) in some way would interact with EKLF or its targets and repress transcription. This could work in different ways, either that the P2RP could bind to the EKLF binding site and prevent EKLF binding, or the P2RP could bind directly to EKLF and block its binding to the promoter (Figure 17). All of this is of course under the assumption that P2RP exists as a translated product.

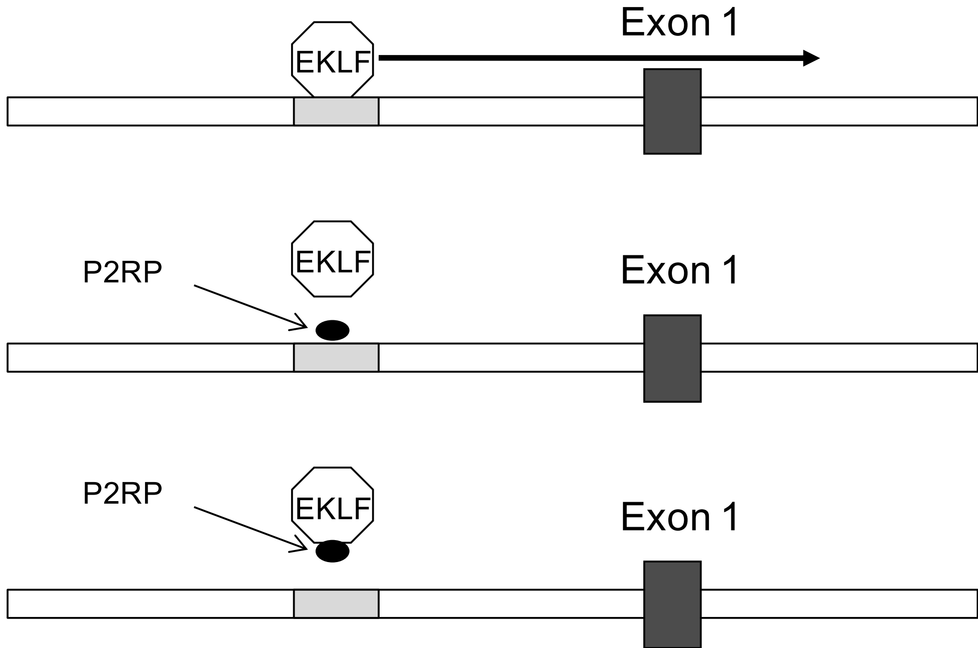


Figure 17. Two possible mechanisms for the hypothetical P2RP peptide (black oval) to influence the EKLF (octagonal) binding to its promoter binding site (grey rectangle). Exon 1 is shown as a black rectangle and the transcriptional direction represented by the black horizontal arrow.

Seven oligonucleotides A-G, ranging from 29 to 35 bp covering a total of 221 bp in the proximal promoter region including the two transcription start points (TSPs), were ordered. Four of these oligonucleotides (A;B;D;E) had potential binding sites for EKLF, where A had the strongest K_a value (personal communication from Dr. Nick Burton, School of biochemistry, University of Bristol, UK) (Figure 18).

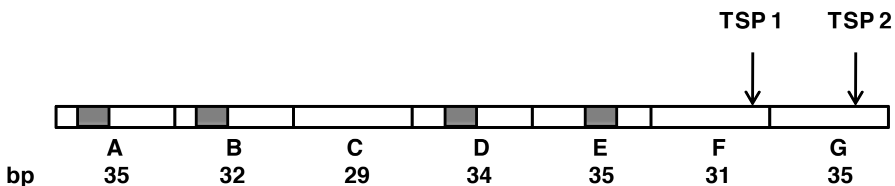


Figure 18. Schematic view of the seven promoter oligonucleotides A-G used for EMSA analysis. Possible EKLF binding sites are marked in grey. The two detected transcription start points are marked with arrows.

The purified EKLF and a mutated form of EKLF (unable to bind its DNA target) used in the EMSA were kindly provided by Dr. Nick Burton and Prof. David

Anstee at the Bristol Institute of Transfusion Science, UK. An oligonucleotide sequence from the β-globin promoter containing an EKLF binding site was used as a positive control (K). Fragments A-G + K were incubated with EKLF before loading on a Tris-borate-EDTA (TBE) gel in lanes 3-8, and stained with SYBR®Green to visualize the bands. Fragments A, B, D and E showed a clear mobility shift as expected (Figure 19). However, also the F and to a lower degree fragment G showed a weaker shift indicating that some unspecific binding also occurred under the conditions chosen (after optimization of incubation times, temperature and buffer composition). The β-globin control gave a positive signal but showed a weaker shift than was expected.

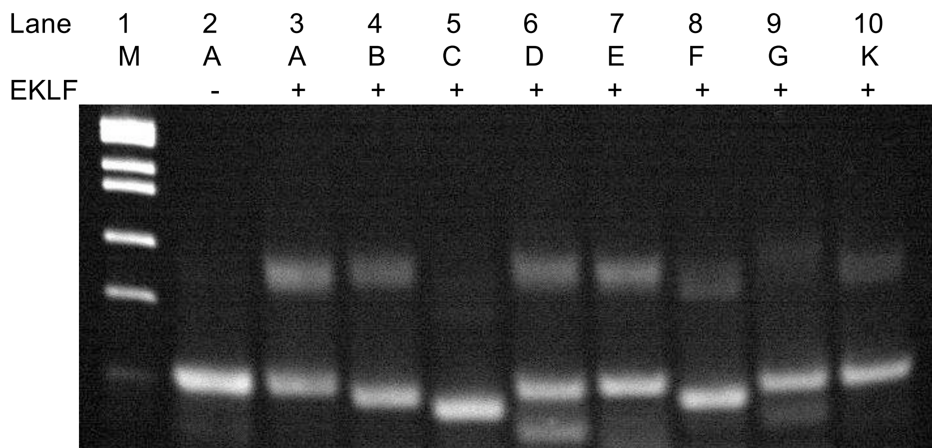


Figure 19. PhiX 174 was used as size marker (M). Fragment A was loaded without addition of EKLF in lane 2 and represent the native size of the A fragment. Fragments A-G, + K incubated with EKLF were loaded in lanes 3-10.

Summary of the results from additional studies on *A4GALT* regulation

Towards the end of the Ph.D. studies, different hypotheses regarding the mechanism behind the P_1/P_2 phenotypes were explored. If the P2RP indeed exists as a translated product, there are several different ways in which it could affect expression. It could bind to either the DNA of the promoter region, to EKLF or α4GalT or any other yet unindentified factor (Figure 20).

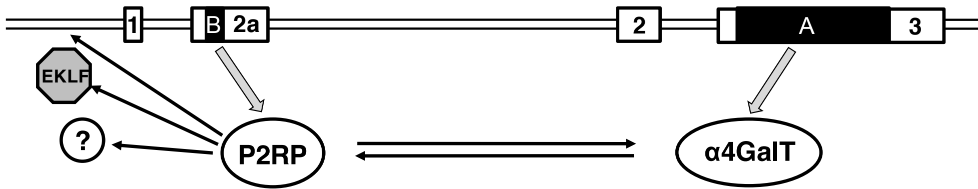


Figure 20. Possible ways for the hypothetical P2RP peptide to interact with the *A4GALT* gene. Exons are shown as white boxes with the ORFs marked in black. The translational products are shown in ovals and their potential interactions are marked with arrows.

Our ongoing work on elucidation of the P_1/P_2 phenotype question has so far shown that the *A4GALT* transcript levels in bone marrow reflects the situation in peripheral blood where the presence of a P^2 allele is correlated to lower transcript levels. We found the exon 2a-containing transcript to be present at relatively stable levels in bone marrow during erythroid culture for 18 days, but here the transcript levels were not too different between genotypes, although there was a tendency for more transcripts when P^2 was present. It could be noted that although the levels were quite high even on day 18 we still have not been able to detect the transcript in peripheral blood (data not shown). These data indicate that the down-regulation of *A4GALT* transcripts is an early event and that the regulator responsible for this may be expressed early during erythropoiesis.

In order to find out if P2RP exists as a translated product two rabbits were immunized with the synthetic protein as mentioned above. The polyclonal antisera from the two animals gave very different results in the Western blot analysis on extracts from the cultured bone marrow cells. Rabbit C reacted strongly with the synthesized P2RP protein and also with a protein of ~40 kD. Rabbit A reacted only weakly with the P2RP, but stronger with a ~50kD protein.

According to the ELISA test made by the company that immunized the rabbits, Rabbit C gave a ten-fold higher response than rabbit A. These data indicate that rabbit C had reacted more strongly and with a more specific response to the small protein. So far we have not been able to visualize any binding to a free protein of about 3 kD, neither in these cells (Figure 15) and nor in the Western immunoblots of cell line extracts representing different P^1/P^2 genotypes (data not shown). Interestingly, we observed binding to larger proteins of various sizes. This could potentially mean crossreactivity against unrelated and irrelevant epitopes but could also represent the larger proteins found to carry *Alu*-related sequences by BLAST analysis. On the other hand, there was no difference between extracts from cells with P_1 and P_2 phenotypes, indicating that the larger bands detected were independent of the ORF found in P^2 alleles. Thus, these polyclonal antisera seem

to react also with other proteins than P2RP. If the P2RP interacts with another protein and the polyclonal rabbit antibodies would recognize the complex, we would have expected to see a difference in individuals or cell lines depending on P_1/P_2 status. The polyclonal antibody against α4GalT gave a very weak reaction with two proteins of about 65kD and 50 kD. No difference in quantity was seen between P_1 and P_2 as otherwise might expected, based on the lower transcript levels detected when P^2 alleles are present. This antiserum needs further optimization for the Western blot analysis and a positive control to verify its specificity. One of the more striking findings was that it also reacted weakly with P2RP on the gel. We also tested a monoclonal anti-α4GalT that did not react with the cell extracts at all and also gave a very high background level despite multiple cycles of optimization. We used an antibody against EKLF and found that the expression of this erythroid transcription factor seems similar on days 6 and 12 and that some variation is seen between the P^1/P^2 genotypes.
To investigate if the EKLF binds to the promoter of the *A4GALT* gene, EMSA was performed. We were able to document binding to the four oligonucleotides that harbour potential binding sites for EKLF, but also weaker apparently unspecific binding to one of the oligonucleotides without a known binding site. These data taken together, suggest that EKLF may well be a transcriptional regulator of the *A4GALT* gene, confirming data that heterozygosity for mutations in the *EKLF* gene cause the In(Lu) phenotype, which lowers not only the Lutheran glycoprotein but also several other blood group antigens including P1. However, further detailed analysis is needed. The EMSA method we used does not allow for the use of competitor DNA (confirming the specificity of the reaction), so alternative methods using labelled oligonucleotides are required, to verify that the findings represent a specific binding of EKLF to the *A4GALT* promoter region. We also intend to test if the P2RP interacts with the oligonucleotides or the EKLF protein itself.
In conclusion, we have not been able to confirm that P2RP exists, at least not according to the sensitivity limitations of the techniques used until now. Instead we have confirmed that the novel transcripts containing exon 2a are present during erythropoiesis. We also found strong indications for EKLF binding to the *A4GALT* promoter, which may explain down-regulation of P1 in the In(Lu) phenotype and EKLF therefore constitutes a potential target for how the newly found P_1 vs. P_2-defining SNP may regulate *A4GALT* transcription. The surprising finding that the polyclonal anti-α4GalT reacts against the synthetic P2RP peptide, has given rise to speculation that the P2RP may bind to the enzyme itself, in accordance with the

old regulator hypothesis.[153] However, such an interaction does not appear to explain the *P*^{1}/*P*^{2}-genotype-dependent difference in *A4GALT* transcript levels observed. Thus, by solving one long-standing mystery (i.e. if the P^k and P1 are linked to the same locus or not), another set of enigmatic questions have arisen.

Conclusions

The studies in this thesis on regulation of the *ABO* gene have led to the following conclusions:

Several polymorphisms related to and/or associated with the most common *ABO* alleles were identified by sequencing of the 5′-and 3′-regulatory regions of the gene. These data are useful for identification of hybrid alleles, allele-specific amplification of the regulatory regions and also for further studies on gene regulation in a field where knowledge about normal polymorphic variation is necessary to be able to identify mutations associated with disease or weak expression.

A^1 and A^2 transcripts were found in bone marrow but not in peripheral blood while *B*, O^1, O^{1v}, and O^2 transcripts could be detected in both sources. The explanation underlying this surprising finding will need further studies since none of above polymorphisms, or others previously found, appear to be common to the *A* alleles.

No correlation could be established between *ABO* transcript levels and the number of enhancer repeats, neither in bone marrow nor in peripheral blood.

A new allele was found, a hybrid between an O^2 and a *B* allele was identified in a patient with A_1B_{weak} phenotype. Compared to a normal *B* allele, the promoter region had only one enhancer repeat, and the only difference in the coding region was a 53G>T substitution. No difference in transcript levels was found compared to normally B-expressing blood donors, confirming our previous data that the number of enhancer repeats does not seem to influence transcript levels in blood.

The studies on the *A4GALT* gene have led to the following conclusions:

A novel transcript from the *A4GALT* gene revealed a polymorphism in a previously unrecognized exon 2a that completely correlates with and can predict the P_1/P_2 phenotypes. This finding has solved the long-standing question about the connection between the P1 and P^k antigens, and resulted in a change of nomenclature so that the P^k antigen has now been moved from the GLOB collection to the P1PK blood group system (previously known as the P blood group system).

With the new possibility to genotype for P^1 and P^2 alleles to predict phenotype, heterozygous P^1P^2 samples were found to have lower transcript levels as well as cell surface P1 and P^k antigen expression on RBCs than P^1-homozygous samples. The well-known variation in P1 expression levels can now at least partially be explained by the zygosity of P^1 alleles.

In the unpublished work we found that transcription factor EKLF can bind to sequences in the *A4GALT* proximal promoter. This may explain why the P1 antigen expression is lowered in individuals with the In(Lu) phenotype, which has been found to be caused by mutations in the *EKLF* gene. However, the mechanism by which the P^2 allele is correlated with lowered transcript and antigen levels awaits further elucidation.

Future perspectives

Even if the work presented in this thesis has answered some of the questions asked when the studies started, the new information provided has also raised several new ones.

The fact that *A* transcripts, as opposed to *B* and *O*, cannot easily be detected in peripheral blood, requires further investigation to clarify the underlying mechanism. ABO antigens on lymphocytes are supposed to be absorbed from the plasma but this has not been studied at the mRNA level. The presence of *ABO* transcripts in different subpopulations of blood cells needs to be examined to understand if *ABO* mRNA is derived from erythroid cells and platelets only or if other cell types add to the complexity of the tissue. Some efforts have been made to sort blood cells in a flow cytometer, but platelets tend to stick to other cell types and made it difficult to get pure subpopulations. Thus it is likely to be important to remove platelets before using methods for cell separation. It would be interesting to investigate at what differentiation stage the *A* transcripts disappear, and the mechanism behind it.

Elucidation of the mechanism behind the P_1 and P_2 phenotypes is a major priority, now that the genetic basis is identified. The three hypotheses on how the SNP may change the phenotype (DNA, RNA or protein level) must still be tested and rejected or accepted. The hypothetical *Alu*-related regulatory peptide, P2RP, has not been detected so far. The polyclonal rabbit antibodies produced against the synthetic peptide recognized the synthetic peptide but did not detect a protein of the expected size (~3 kD) in the cells tested, although this may well be technically challenging. An aim is to clarify if the larger bands detected by the antisera are due to unspecific (*Alu*-unrelated) binding of the antibody or if it stains P2RP interacting with other proteins carrying *Alu*-related sequences, the presence of which were indicated by the protein database search. Yet another, perhaps even more exciting, alternative is if anti-P2RP reacts against other proteins to which the P2RP has bound. This would imply that the antisera must have been raised against P2RP interacting with the target already in the rabbit *in vivo*. These possibilities could be further explored if chromatin immunoprecipitation (ChIP) was applied to study protein-protein interactions.

Another way to investigate whether P2RP exists is to perform expression studies with transfection of cloned variants of the different mRNA transcripts into suitable cell lines with the necessary phenotype. Both cells with the P_1, P_2 phenotypes, and cells lacking functional *A4GALT* gene expression are needed for these

experiments. Simple overexpression of α4GalT may turn P_2 cells into P_1 if they have the appropriate acceptor, paragloboside.

The finding of EKLF binding sites in the *A4GALT* promoter and confirmation of EKLF binding *in vitro* to the promoter region also call for more experiments. EMSA studies with both specific and unspecific competitors to verify the binding are planned, and so are super-shift assays with anti-EKLF to confirm EKLF binding. In addition, different variants of protein pre-incubation experiments between P2RP and possible partners are ongoing. Since *Alu*-repeats have been shown to regulate gene expression, experiments to address the potential role of *Alu* sequence in both DNA, RNA and protein form are also under way. It would be of great interest beyond transfusion medicine to define a clearer role for these repetitive sequences in gene regulation, using *A4GALT* as a model system.

Svensk sammanfattning

Generna är den del av vår arvsmassa (DNA) som innehåller information om hur äggviteämnen (proteiner) ska se ut. Detta kan ske genom att DNA kopieras (transkriberas) till ett stort antal liknande molekyler, kallade RNA. Dessa översätts (translateras) till de livsviktiga proteiner som gör att våra celler kan överleva och fungera. Alla gener är inte aktiva samtidigt och vissa är bara aktiva i en typ av celler. DNA består av byggstenar (nukleotider) av fyra olika typer: adenin (A), cytosin (C), guanidin (G) och thymidin(T) som kopplas samman i långa kedjor i en viss sekvens. Nukleotiderna i två kedjor binder till varandra så att A alltid binder till T och C binder till G. De två strängarna vrids till en spiralform. För upptäckten av detta 1953 fick forskarna Francis Crick och James Watson Nobelpriset i medicin 1962. En gensekvens är oftast uppdelad i ett antal kortare viktiga sekvenser som kallas exoner och den mindre informativa sekvensen mellan dessa kallas introner. När genen kopierats till RNA klipps intronerna bort, vilket kallas splicing och gör så att det till slut bara återstår de delar som skall översättas till ett protein. Det på så sätt modifierade RNA:t kallas budbärar-RNA (mRNA). Det som gör att en gen kan transkriberas är att en speciell sorts proteiner, transkriptionsfaktorer, binder till en DNA sträng före den första exonen. Inbindning sker i det område som kallas promotorsekvensen. Det finns olika typer av transkriptionsfaktorer: en del finns i alla celler medan andra bara finns i vissa typer av celler. När dessa faktorer bundit till promotorn kan enzymet RNA-polymeras, som bygger RNA-strängen, binda och därmed startar transkriptionen.

Målet med mitt doktorandarbete har varit att studera dessa promotorsekvenser i två gener som kodar för två olika enzym, som båda kallas glykosyltransferaser och kopplar enstaka sockermolekyler till kolhydratkedjor. Dessa enzymer är mycket specifika och kan oftast bara koppla en viss typ av sockermolekyl till en annan specifik sockermolekyl i slutet av kolhydratkedjan. Det finns därför väldigt många olika glykosyltransferaser. De gener som har studerats här ger upphov till kolhydratbaserade blodgruppsmolekyler.

Blodgrupper är molekylmarkörer (antigen) som finns på ytan av de röda blodkropparna. En del finns också på andra typer av celler i kroppen. Det som gör att de betecknas som blodgruppsantigen är att de kan ge upphov till en slags immunreaktion som kallas antikroppsbildning hos personer som inte själv har den aktuella blodgruppen, men som har exponerats för den genom exempelvis en blodtransfusion eller vid graviditet. Blodgruppsantigener kan utgöras av proteiner

eller kolhydrater, och när det gäller kolhydratantigen har man i allmänhet naturligt förekommande antikroppar mot de antigen man själv saknar, d.v.s. även utan att ha fått blod eller varit gravid. Det viktigaste blodgruppssystemet är ABO-systemet där de naturliga antikropparna är mycket starka. Det finns två viktiga antigener i detta system, A och B, och om man har A-antigenet så har man antikroppar mot B-antigenet i plasman och tvärtom. Om man har både A- och B-antigenet (blodgrupp AB) saknar man ABO-antikroppar. Är man istället blodgrupp O har man varken A- eller B-antigen men väl antikroppar mot både A och B.
ABO-systemet upptäcktes av österrikaren Karl Landsteiner 1901 och genom att blodgruppera både patient och blodgivare är det möjligt att säkert transfundera blod mellan individer. Eftersom ABO-antigenen även finns på andra typer av celler så måste man ta hänsyn till dessa vid organtransplantationer.
Man ärver en gen från vardera föräldern och har alltså två kopior (alleler) av varje gen. *ABO*-genen skiljer sig mellan de som är blodgrupp A och B genom att endast ett fåtal nukleotider i gensekvensen är olika. Dessa ger upphov till att några av aminosyrorna som bygger upp enzymet är utbytta och detta gör i sin tur att enzymet bygger på en galaktosmolekyl (Gal) vid blodgrupp B istället för en *N*-acetylgalaktosamin (GalNac) vid blodgrupp A på kolhydratkedjan. I de flesta fall när en person har blodgrupp O, beror det på att en nukleotid saknas i genen vilket resulterar i att enzymet ej fungerar. Det finns också en mängd förändringar i gensekvensen som kan leda till att enzymet fungerar sämre och ger färre antigen på cellytan. Ibland kan det verka som att en person är blodgrupp O trots att den har en muterad *A*- eller *B*-gen. Genom att studera generna förutom de röda blodkropparna kan en säkrare bestämning göras av vilken blodgrupp en individ har.
Alla blodceller bildas i benmärgen. Där finns stamceller som förutom att de kan bilda nya stamceller också kan utvecklas till alla de celltyper som finns i blodplasman, d.v.s. vita blodkroppar (leukocyter), röda blodkroppar (erytrocyter) och blodplättar (trombocyter). Utvecklingen från stamcell till en mogen röd blodkropp kallas erytropoes.

Artikel I

Mutationer förkommer också i promotorregionen och i det första arbetet har jag kartlagt hur DNA-sekvenser i promotorn varierar mellan olika blodgruppsgener i ABO-systemet. Ett speciellt område som kan påverka genuttrycket, en så kallad enhancer-region, studerades. I vissa alleler finns det fyra repeterade sekvenser bestående av 43 nukleotider vardera medan andra alleler saknar tre av dessa repetitioner. Det finns beskrivet att fyra repeterade sekvenser ger mer än hundra

gånger högre uttryck av RNA jämfört med om man bara har en 43-nukleotidssekvens. Därför undersökte jag hur mycket mRNA som fanns i blodet hos individer med olika *ABO*-alleler. Märkligt nog kunde inget mRNA detekteras från personer med de två vanligaste *A*-varianterna som finns (A^1 och A^2) trots att A^2-genen har fyra repetitioner och A^1 bara en. Däremot fanns det gott om mRNA från både *B*-gener och från olika *O*-gener, även en lite ovanligare variant som bara har en enhancer-enhet. Genom att isolera stamceller från benmärg och sedan odla dem under sådana förhållanden att de utvecklades till erytrocyter kunde celler från olika utvecklingsfaser studeras och mängden *ABO*-mRNA mätas. I dessa celler fanns mRNA i högre nivå i celler från givare med blodgrupp A, och ingen skillnad mellan A^1 och A^2, än från de som var blodgrupp B och O. Dessa data visar alltså att antalet enhancerrepetitioner inte påverkar mRNA-nivån, varken i blod eller benmärg. Däremot hittades ingen förklaring till varför inget mRNA från *A*-generna kunde detekteras i blodet.

Artikel II

I arbete två undersöktes blod från en schweizisk patient som hade ett svagt uttryck av B-antigen på sina röda blodkroppar. DNA-sekvensen för både promotor, exoner och introner undersöktes och det visade sig att genen var en sammankoppling av två olika *ABO*-gener (s.k. hybridgen). Den första delen kom från en ovanlig *O*-gen (O^2), medan resten av genen var en vanlig *B*-gen. Denna kombination innebar att en aminosyra var annorlunda jämfört med det vanliga B-enzymet. För att se om detta påverkade antigenuttrycket sattes (transfekterades) den muterade genen in i en celltyp som hade blodgrupp O för att se om den kunde uttrycka B-antigen på cellytan. Samma sak gjordes med en normal *B*-gen och sedan jämfördes uttrycket på cellerna. Den muterade genen gav ett något lägre (14%) uttryck. Eftersom O^2-genen bara har en enhancer enhet och *B*-genen fyra, så hade den aktuella hybridgenen bara en enhet. För att se om detta påverkade mängden mRNA som transkriberas från genen undersöktes nivåerna i detta prov och jämfördes med åtta andra prov från givare med normalt uttryck av B-antigen. Nivån av mRNA som kunde detekteras i provet med hybridgenen var inte lägre än i kontrollerna, vilket bekräftar studierna i artikel I, d.v.s. att antalet enheter i enhancerregionen inte påverkar mängden RNA-transkript.

Artikel III

Vår forskningsgrupp har under nästan 10 år letat efter den genetiska orsaken att man får kolhydratblodgrupperna P_1 och P_2, som varit kända ända sedan 1927. Faktum är att P_1/P_2 tillhör det enda av 30 blodgruppssystem för vilket man ej

kände till vilken gen som styrde blodgruppsuttrycket. I mitt sista delarbete lyckades jag glädjande nog lösa denna gamla gåta efter många års arbete. Det handlar om två icke-ABO-blodgruppsantigen, P1 och P^k som liknar blodgrupp B något eftersom de båda har galaktos längst ut i sockerkedjan, dock bundet på ett annat sätt. Det har länge pågått en diskussion ifall det är samma enzym och därmed samma gen som ligger bakom uppkomsten av båda antigenen. Anledningen till misstanken att det skulle vara samma gen är att om det blir mutationer i genen som styr P^k-uttrycket (kallad *A4GALT*-genen) så inaktiveras enzymet varvid båda antigenen saknas. Däremot kan man sakna P1-antigenet utan att sakna P^k, och detta har ingen kunnat förklara tidigare. När transkriptnivåerna från *A4GALT*-genen undersöktes i blodet hos personer som var P_1 eller P_2 (som alltså saknar P1) fann vi mycket lägre nivåer hos P_2-individerna. Vid undersökning av *A4GALT*-transkript från benmärg, hittades en ny variant av transkriptet med en sekvens som kom från intronen mellan exon 1 och 2 i *A4GALT*-genen. I denna sekvens upptäcktes en mutation som kunde relateras till om en person var P_1 eller P_2. Detta kunde nu bekräftas på över 200 blodgivarprover. Därmed fanns också den efterlängtade kopplingen mellan P1, P^k antigenen och *A4GALT*-genen. När det nu gick att bestämma om en person hade P^1- eller P^2-alleler (eller båda), kunde uttrycket av antigen på cellytan korreleras till dessa. Det visade sig att uttrycket av både transkript och P1- respektive P^k -antigen på röda blodkroppar var lägre om man hade en P^2-allel. Om en person hade två P^2-alleler saknade man P1 helt och hade mindre mängd P^k antigen. Sammanfattningsvis är det härmed klarlagt att *A4GALT*-genen ger upphov till både P1 och P^k antigen och arbetet kommer att fortsätta med att försöka förklara mekanismen bakom detta.

Acknowledgements

I want to express my gratitude to everyone who has helped me throughout the past eight years. Thanks to:

Martin L Olsson, my supervisor for giving me the opportunity to work on and write this thesis despite my age, for all support especially with the writing and patiently trying to teach me grammar, both Swedish and English
Urban Gullberg my co-supervisor for guidance in the field of gene regulation and for providing opportunity to culture cells
Jill Storry for always being so helpful and supportive and for driving us round in South Africa, and of course for helping me with my manuscripts
Åsa Hellberg for teaching me everything about P1, P^k and P, and for being such a nice room mate
Elisabet Sjöberg Wester for all nice research meetings in Tosteberga, for being such a nice travel companion and friend
Bahram Hosseini-Maaf for all support and wild discussions on everything from research to flowers and politics
Annika Hult for having patience with me trying to use the FACScan and for guiding me around Berlin
Julia Westman, Josefina Horvat-Dykes, Jannica Johannesson Samuelsson for support and for making the research group so nice
Alan Chester for guiding me and the group in the scrubby field of glycobiology
Tor Olofsson for helping me with the cell morphology picture and for recommending me to the employment on the Blood Centre
Gunilla Hochbergs and Li Zhou for being the best of colleagues and friends
All friends and colleagues at the Blood Centre (I know it is not called that any more, but..)
Karina Vidovic for always being there answering all my questions on cell culturing
Ann-Maj Persson and Bodil Rosberg for helping me troubleshooting the Western blots
May-Louise Andersson for always helping me with administrative stuff
All friends and colleagues at BMC C14 for making it such a nice place to work
Eva Nilsson, Ellinor Johnsson, Asta Lundqvist, Ingrid Gärtner for being fantastic colleagues in the past
Nick Burton and David Anstee for providing the EKLF protein and guidance in the EMSA work
Ingrid Thörn and Carina Wasslavik for all support in the clinical world

My friends since childhood, Lisbeth and Gunilla, for all good times and bad.
Kerstin and Stene for distant support in all fields possible

But most of all thanks to my family
My lovely children, Nina for all your concern and Kent who also brought me new family members Harmony, Jessica and my grandson Max
Monica, Tom and Margita for all the support you have been giving me, especially this last year

And of cause
Bert without your help and support this would not have been possible, and I have years of cleaning the house to do to compensate you

References

1. Mollison PL, Engelfriet CP, Contreras M. Blood Transfusion in Clinical Medicine. London, U.K.: Blackwell Scientific, 1997.
2. Springer GF, Horton RE. Blood group isoantibody stimulation in man by feeding blood group-active bacteria. J. Clin. Invest. 1969;48: 1280-91.
3. Daniels G. Human Blood Groups. Oxford, U.K.: Blackwell Scientific, 2002.
4. Storry JR, Olsson ML. Genetic basis of blood group diversity. Br. J. Haematol. 2004;126: 759-71.
5. Storry JR, Castilho L, Daniels G, Flegel WA, Garratty G, Francis CL, Moulds JM, Moulds JJ, Olsson ML, Poole J, Reid ME, Rouger P, van der Schoot E, Scott M, Smart E, Tani Y, Yu LC, Wendel S, Westhoff C, Yahalom V, Zelinski T. International Society of Blood Transfusion Working Party on red cell immunogenetics and blood group terminology: Berlin report. Vox Sang. 2011, In press.
6. 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.
7. Rowe JA, Handel IG, Thera MA, Deans AM, Lyke KE, Kone A, Diallo DA, Raza A, Kai O, Marsh K, Plowe CV, Doumbo OK, Moulds JM. Blood group O protects against severe Plasmodium falciparum malaria through the mechanism of reduced rosetting. Proc.Natl.Acad.Sci.U.S.A. 2007;104: 17471-6.
8. Cserti CM, Dzik WH. The ABO blood group system and Plasmodium falciparum malaria. Blood. 2007;110: 2250-8.
9. Miller RW. The discipline of epidemiology. Science. 1995;269: 1327.
10. Zimmerman PA, Woolley I, Masinde GL, Miller SM, McNamara DT, Hazlett F, Mgone CS, Alpers MP, Genton B, Boatin BA, Kazura JW. Emergence of FY*A(null) in a Plasmodium vivax-endemic region of Papua New Guinea. Proc.Natl.Acad.Sci.U.S.A. 1999;96: 13973-7.
11. Moulds JM, Nowicki S, Moulds JJ, Nowicki BJ. Human blood groups: incidental receptors for viruses and bacteria. Transfusion. 1996;36: 362-74.
12. 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.
13. Lomberg H, Jodal U, Eden CS, Leffler H, Samuelsson B. P1 blood group and urinary tract infection. Lancet. 1981;1: 551-2.
14. 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.

15. Vandenbroucke-Grauls CM, Appelmelk BJ. Helicobacter pylori LPS: molecular mimicry with the host and role in autoimmunity. Ital. J. Gastroenterol. Hepatol. 1998;30 Suppl 3: S259-60.
16. van Alphen L, Poole J, Geelen L, Zanen HC. The erythrocyte and epithelial cell receptors for Haemophilus influenzae are expressed independently. Infect. Immun. 1987;55: 2355-8.
17. Marsh WL. Anti-i: a cold antibody defining the Ii relationship in human red cells. Br. J. Haematol. 1961;7: 200-9.
18. Hakomori S. Blood group ABH and Ii antigens of human erythrocytes: chemistry, polymorphism, and their developmental change. Semin. Hematol. 1981;18: 39-62.
19. Brown KE, Anderson SM, Young NS. Erythrocyte P antigen: cellular receptor for B19 parvovirus. Science. 1993;262: 114-7.
20. Lund N, Olsson ML, Ramkumar S, Sakac D, Yahalom V, Levene C, Hellberg A, Ma XZ, Binnington B, Jung D, Lingwood CA, Branch DR. The human P(k) histo-blood group antigen provides protection against HIV-1 infection. Blood. 2009;113: 4980-91.
21. Thorven M, Grahn A, Hedlund KO, Johansson H, Wahlfrid C, Larson G, Svensson L. A homozygous nonsense mutation (428G-->A) in the human secretor (FUT2) gene provides resistance to symptomatic norovirus (GGII) infections. J. Virol. 2005;79: 15351-5.
22. Vildevall M, Grahn A, Oliver SL, Bridger JC, Charpilienne A, Poncet D, Larson G, Svensson L. Human antibody responses to bovine (Newbury-2) norovirus (GIII.2) and association to histo-blood group antigens. J. Med. Virol. 2010;82: 1241-6.
23. Moulds JM, Moulds JJ. Blood group associations with parasites, bacteria, and viruses. Transfus. Med. Rev. 2000;14: 302-11.
24. Young NS, Brown KE. Parvovirus B19. N. Engl. J. Med. 2004;350: 586-97.
25. O'Brien AD, Holmes RK. Shiga and Shiga-like toxins. Microbiol. Rev. 1987;51: 206-20.
26. Fry AE, Griffiths MJ, Auburn S, Diakite M, Forton JT, Green A, Richardson A, Wilson J, Jallow M, Sisay-Joof F, Pinder M, Peshu N, Williams TN, Marsh K, Molyneux ME, Taylor TE, Rockett KA, Kwiatkowski DP. Common variation in the ABO glycosyltransferase is associated with susceptibility to severe Plasmodium falciparum malaria. Hum. Mol. Genet. 2008;17: 567-76.
27. Lund N, Sakac D, Lingwood C, Levene C, Hellberg A, Olsson ML, Branch DR. Pk blood group antigen: A natural resistance factor to HIV infection (abstract). Vox Sang. 2006;91: 23.
28. Garratty G. Blood groups and disease: A historical perspective. Transfus. Med. Rev. 2000;14: 291-301.
29. Anstee DJ. The relationship between blood groups and disease. Blood. 2010;115: 4635-43.

30. Lairson LL, Henrissat B, Davies GJ, Withers SG. Glycosyltransferases: structures, functions, and mechanisms. Annu. Rev. Biochem. 2008;77: 521-55.
31. Yamamoto F, Clausen H, White T, Marken J, Hakomori S. Molecular genetic basis of the histo-blood group ABO system. Nature. 1990;345: 229-33.
32. Storry JR, Olsson ML. The ABO blood group system revisited: a review and update. Immunohematology. 2009;25: 48-59.
33. Li E, Beard C, Jaenisch R. Role for DNA methylation in genomic imprinting. Nature. 1993;366: 362-5.
34. Jaenisch R. DNA methylation and imprinting: why bother? Trends Genet. 1997;13: 323-9.
35. Shatkin AJ, Manley JL. The ends of the affair: capping and polyadenylation. Nat. Struct. Biol. 2000;7: 838-42.
36. Matsui T, Segall J, Weil PA, Roeder RG. Multiple factors required for accurate initiation of transcription by purified RNA polymerase II. J. Biol. Chem. 1980;255: 11992-6.
37. Myers LC, Kornberg RD. Mediator of transcriptional regulation. Annu. Rev. Biochem. 2000;69: 729-49.
38. Frydman M, Etzioni A, Eidlitz-Markus T, Avidor I, Varsano I, Shechter Y, Orlin JB, Gershoni-Baruch R. Rambam-Hasharon syndrome of psychomotor retardation, short stature, defective neutrophil motility, and Bombay phenotype. Am. J. Med. Genet. 1992;44: 297-302.
39. Luhn K, Wild MK, Eckhardt M, Gerardy-Schahn R, Vestweber D. The gene defective in leukocyte adhesion deficiency II encodes a putative GDP-fucose transporter. Nat. Genet. 2001;28: 69-72.
40. Gazit Y, Mory A, Etzioni A, Frydman M, Scheuerman O, Gershoni-Baruch R, Garty BZ. Leukocyte adhesion deficiency type II: long-term follow-up and review of the literature. J. Clin. Immunol. 2010;30: 308-13.
41. Parsons SF, Jones J, Anstee DJ, Judson PA, Gardner B, Wiener E, Poole J, Illum N, Wickramasinghe SN. A novel form of congenital dyserythropoietic anemia associated with deficiency of erythroid CD44 and a unique blood group phenotype [In(a-b-), Co(a-b-)]. Blood. 1994;83: 860-8.
42. Singleton BK, Fairweather VS, Lau W, Parsons SF, Burton NM, Frayne J, Brady RL, Anstee DJ. A Novel EKLF Mutation in a Patient with Dyserythropoietic Anemia: The First Association of EKLF with Disease in Man (abstract). Blood. 2009;114: 162.
43. Arnaud L, Helias V, Menanteau C, Peyrard T, Lucien N, Ripoche P, Lapegue R, Pham BN, Le Pennec PY, Moulds JJ, Cartron JP. A functional AQP1 allele producing a Co(a-b-) phenotype revises and extends the Colton blood group system. Transfusion. 2010;50: 2106-16.
44. Crawford MN, Greenwalt TJ, Sasaki T, Tippett P, Sanger R, Race RR. The Phenotype Lu(a−b−) Together with Unconventional Kidd Groups in One Family. Transfusion. 1961;1: 226-32.

45. Contreras M, Tippett P. The Lu(a-b-) syndrome and an apparent upset of P1 inheritance. Vox Sang. 1974;27: 369-71.
46. 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.
47. Gibson T. Two kindred with the rare dominant inhibitor of the lutheran and p1 red cell antigens. Hum. Hered. 1976;26: 171-4.
48. Shaw MA, Leak MR, Daniels GL, Tippett P. The rare Lutheran blood group phenotype Lu(a-b-): a genetic study. Ann. Hum. Genet. 1984;48: 229-37.
49. Daniels G. Lutheran. Immunohematology. 2009;25: 152-9.
50. Singleton BK, Burton NM, Green C, Brady RL, Anstee DJ. Mutations in EKLF/KLF1 form the molecular basis of the rare blood group In(Lu) phenotype. Blood. 2008;112: 2081-8.
51. Singleton BK, Roxby D, Stirling J, Spring FA, Wilson C, Poole J, Anstee DJ. A Novel GATA1 Mutation (Ter414Arg) in a Family with the Rare X-Linked Blood Group Lu(a-b-) Phenotype (abstract). Blood. 2009;114: 1979.
52. Moreau R, Dausset J, Bernard J, Moullec J. [Acquired hemolytic anemia with polyagglutinability of erythrocytes by a new factor present in normal blood]. Bull. Mem. Soc. Med. Hop. Paris 1957;73: 569-87.
53. Ju T, Cummings RD. Protein glycosylation: chaperone mutation in Tn syndrome. Nature. 2005;437: 1252.
54. Batzer MA, Deininger PL. Alu repeats and human genomic diversity. Nat. Rev. Genet. 2002;3: 370-9.
55. Rowold DJ, Herrera RJ. Alu elements and the human genome. Genetica. 2000;108: 57-72.
56. Hasler J, Samuelsson T, Strub K. Useful 'junk': Alu RNAs in the human transcriptome. Cell. Mol. Life Sci. 2007;64: 1793-800.
57. Chen LL, DeCerbo JN, Carmichael GG. Alu element-mediated gene silencing. EMBO J. 2008;27: 1694-705.
58. Sorek R, Ast G, Graur D. Alu-containing exons are alternatively spliced. Genome Res. 2002;12: 1060-7.
59. Kumar AP, Piedrafita FJ, Reynolds WF. Peroxisome proliferator-activated receptor gamma ligands regulate myeloperoxidase expression in macrophages by an estrogen-dependent mechanism involving the -463GA promoter polymorphism. J. Biol. Chem. 2004;279: 8300-15.
60. Kondo M, Wagers AJ, Manz MG, Prohaska SS, Scherer DC, Beilhack GF, Shizuru JA, Weissman IL. Biology of hematopoietic stem cells and progenitors: implications for clinical application. Annu. Rev. Immunol. 2003;21: 759-806.
61. Kerenyi MA, Orkin SH. Networking erythropoiesis. J. Exp. Med. 2010;207: 2537-41.
62. Landsteiner K. Über Agglutinationserscheinungen normalen Menschlichen Blutes. Wien. Klin. Wochenschr. 1901;14: 1132-4.

63. Clausen H, Levery SB, Dabelsteen E, Hakomori S. Blood group ABH antigens: a new series of blood group A-associated structures (genetic regulation and tissue distribution). Transpl. Proc. 1987;19: 4408-12.
64. Korchagina EY, Pochechueva TV, Obukhova PS, Formanovsky AA, Imberty A, Rieben R, Bovin NV. Design of the blood group AB glycotope. Glycoconj. J. 2005;22: 127-33.
65. Svensson L, Bindila L, Angstrom J, Samuelsson BE, Breimer ME, Rydberg L, Henry SM. The structural basis of blood group A-related glycolipids in an A3 red cell phenotype and a potential explanation to a serological phenomenon. Glycobiology. 2011;21: 162-74.
66. Oriol R, Candelier JJ, Mollicone R. Molecular genetics of H. Vox. Sang. 2000;78 Suppl 2: 105-8.
67. Kabat EA. Blood group substances. Their chemistry and immunohistochemistry. New York: Academic Press, 1956.
68. Watkins WM. Biochemistry and genetics of the ABO, Lewis and P blood group systems. In: Harris H, Hirschhorn K, eds. Advances in Human Genetics. New York: Plenum Press, 1980:1-136.
69. Clausen H, Levery SB, Nudelman E, Tsuchiya S, Hakomori S. Repetitive A epitope (type 3 chain A) defined by blood group A 1 - specific monoclonal antibody TH-1: chemical basis of qualitative A 1 and A 2 distinction. Proc. Natl. Acad. Sci. U.S.A. 1985;82: 1199-203.
70. Clausen H, Holmes E, Hakomori S. Novel blood group H glycolipid antigens exclusively expressed in blood group A and AB erythrocytes (type 3 chain H). II. Differential conversion of different H substrates by A1 and A2 enzymes, and type 3 chain H expression in relation to secretor status. J. Biol. Chem. 1986;261: 1388-92.
71. Clausen H, White T, Takio K, Titani K, Stroud M, Holmes E, Karkov J, Thim L, Hakomori S. Isolation to homogeneity and partial characterization of a histo-blood group A defined Fuc-alpha1-2Gal alpha1-3-N-acetylgalactosaminyltransferase from human lung tissue. J. Biol. Chem. 1990;265: 1139-45.
72. Yamamoto F, Marken J, Tsuji T, White T, Clausen H, Hakomori S. 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.
73. Bennett EP, Steffensen R, Clausen H, Weghuis DO, van Kessel AG. Genomic cloning of the human histo-blood group ABO locus. Biochem. Biophys. Res. Commun. 1995;206: 318-25.
74. Yamamoto F, McNeill PD, Hakomori S. Genomic organization of human histo-blood group ABO genes. Glycobiology. 1995;5: 51-8.
75. Ferguson-Smith MA, Aitken DA, Turleau C, de Grouchy J. Localisation of the human ABO: Np-1: AK-1 linkage group by regional assignment of AK-1 to 9q34. Hum. Genet. 1976;34: 35-43.
76. Kominato Y, Hata Y, Takizawa H, Matsumoto K, Yasui K, Tsukada J, Yamamoto F. Alternative promoter identified between a hypermethylated

upstream region of repetitive elements and a CpG island in human ABO histo-blood group genes. J. Biol.Chem. 2002;277: 37936-48.

77. Hata Y, Kominato Y, Takizawa H, Tabata S, Michino J, Nishino K, Yasumura S, Yamamoto F. Transcription starting from an alternative promoter leads to the expression of the human ABO histo-blood group antigen. Transfusion. 2003;43: 656-62.
78. Olsson ML, Chester MA. Frequent occurrence of a variant O1 gene at the blood group ABO locus. Vox. Sang. 1996;70: 26-30.
79. Yamamoto F, McNeill PD, Yamamoto M, Hakomori S, Bromilow IM, Duguid JK. Molecular genetic analysis of the ABO blood group system: 4. Another type of O allele. Vox. Sang. 1993;64: 175-8.
80. Yazer MH, Olsson ML. The O2 allele: questioning the phenotypic definition of an ABO allele. Immunohematology. 2008;24: 138-47.
81. Hosseini-Maaf B, Smart E, Chester MA, Olsson ML. 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.
82. Seltsam A, Gruger D, Just B, Figueiredo C, Gupta CD, Deluca DS, Blasczyk R. Aberrant intracellular trafficking of a variant B glycosyltransferase. Transfusion. 2008;48: 1898-905.
83. Hult AK, Yazer MH, Jorgensen R, Hellberg A, Hustinx H, Peyrard T, Palcic MM, Olsson ML. Weak A phenotypes associated with novel ABO alleles carrying the A2-related 1061C deletion and various missense substitutions. Transfusion. 2010;50: 1471-86.
84. Seltsam A, Das GC, Bade-Doeding C, Blasczyk R. A weak blood group A phenotype caused by a translation-initiator mutation in the ABO gene. Transfusion. 2006;46: 434-40.
85. Paulson JC, Colley KJ. Glycosyltransferases. Structure, localization, and control of cell type- specific glycosylation. J. Biol. Chem. 1989;264: 17615-8.
86. Clausen H, Bennett EP, Dabelsteen E. Carbohydrates of the cell surface: molecular aspects of glycosyltransferases and their genes. APMIS Suppl. 1992;27: 9-17.
87. Yamamoto F, McNeill PD, Hakomori S. Human histo-blood group A2 transferase coded by A2 allele, one of the A subtypes, is characterized by a single base deletion in the coding sequence, which results in an additional domain at the carboxyl terminal Biochem. Biophys. Res. Commun. 1992;187: 366-74.
88. Bianco T, Farmer BJ, Sage RE, Dobrovic A. Loss of red cell A, B, and H antigens is frequent in myeloid malignancies. Blood. 2001;97: 3633-9.
89. Olsson ML, Irshaid NM, Hosseini-Maaf B, Hellberg A, Moulds MK, Sareneva H, Chester MA. 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.

90. Olsson ML, Chester MA. Heterogeneity of the blood group A x allele: genetic recombination of common alleles can result in the A x phenotype Transfus. Med. 1998;8: 231-8.
91. Hosseini-Maaf B, Hellberg A, Rodrigues MJ, Chester MA, Olsson ML. ABO exon and intron analysis in individuals with the AweakB phenotype reveals a novel O1v-A2 hybrid allele that causes four missense mutations in the A transferase. BMC Genet. 2003;4: 17.
92. Seltsam A, Hallensleben M, Kollmann A, Blasczyk R. The nature of diversity and diversification at the ABO locus. Blood. 2003;102: 3035-42.
93. Irshaid NM, Chester MA, Olsson ML. Allele-related variation in minisatellite repeats involved in the transcription of the blood group ABO gene. Transfus. Med. 1999;9: 219-26.
94. Shimada I, Kominato Y, Hata N, Takizawa H. DNA polymorphisms in the 5'-flanking sequence of human ABO blood group genes and their association with the alleles for the common ABO phenotypes. Leg.Med.(Tokyo). 1999;1: 217-25.
95. Sun CF, Chen DP, Tseng CP, Wang WT, Liu JP. Identification of a novel A1v-O1v hybrid allele with G829A mutation in a chimeric individual of AelBel phenotype. Transfusion. 2006;46: 780-9.
96. Suzuki K. ABO blood group alleles and genetic recombination. Leg. Med. (Tokyo). 2005;7: 205-12.
97. Barre L, Venkatesan N, Magdalou J, Netter P, Fournel-Gigleux S, Ouzzine M. Evidence of calcium-dependent pathway in the regulation of human beta1,3-glucuronosyltransferase-1 (GlcAT-I) gene expression: a key enzyme in proteoglycan synthesis. FASEB J. 2006;20: 1692-4.
98. Koda Y, Soejima M, Kimura H. Changing transcription start sites in H-type alpha(1,2)fucosyltransferase gene (FUT1) during differentiation of the human erythroid lineage. Eur. J. Biochem. 1998;256: 379-87.
99. Sato T, Furukawa K. Transcriptional regulation of the human beta-1,4-galactosyltransferase V gene in cancer cells: essential role of transcription factor Sp1. J. Biol. Chem. 2004;279: 39574-83.
100. Hata Y, Kominato Y, Yamamoto FI, Takizawa H. Characterization of the human ABO gene promoter in erythroid cell lineage. Vox Sang. 2002;82: 39-46.
101. Kominato Y, Tsuchiya T, Hata N, Takizawa H, Yamamoto F. Transcription of human ABO histo-blood group genes is dependent upon binding of transcription factor CBF/NF-Y to minisatellite sequence. J. Biol. Chem. 1997;272: 25890-8.
102. Yu LC, Chang CY, Twu YC, Lin M. Human histo-blood group ABO glycosyltransferase genes: different enhancer structures with different transcriptional activities. Biochem. Biophys. Res. Commun. 2000;273: 459-66.
103. Seltsam A, Wagner FF, Grüger D, Gupta CD, Bade-Doeding C, Blasczyk R. Weak blood group B phenotypes may be caused by variations in the

CCAAT-binding factor/NF-Y enhancer region of the ABO gene. Transfusion. 2007;47: 2330-5.
104. Kominato Y, Hata Y, Matsui K, Takizawa H, Tsukada J, Nakajima T, Kaneko Y, Kishi K. Transcriptional regulation of the human ABO histo-blood group genes is dependent on the N box upstream of the proximal promoter. Transfusion. 2004;44: 1741-9.
105. Clausen H, Hakomori S. ABH and related histo-blood group antigens; immunochemical differences in carrier isotypes and their distribution. Vox Sang. 1989;56: 1-20.
106. Rachkewich RA, Crookston MC, Tilley CA, Wherrett JR. Evidence that blood group A antigen on lymphocytes is derived from the plasma. J. Immunogenet. 1978;5: 25-9.
107. Dunstan RA. Status of major red cell blood group antigens on neutrophils, lymphocytes and monocytes. Br. J. Haematol. 1986;62: 301-9.
108. Mollicone R, Caillard T, Le Pendu J, Francois A, Sansonetti N, Villarroya H, Oriol R. Expression of ABH and X (Lex) antigens on platelets and lymphocytes. Blood. 1988;71: 1113-9.
109. Ravn V, Dabelsteen E. Tissue distribution of histo-blood group antigens. APMIS. 2000;108: 1-28.
110. Landsteiner K, Levine P. Further observations on individual differences of human blood. Proc. Soc. Exp. Biol. 1927;24: 941-2.
111. Levine P, Bobbitt OB, Waller RK, Kuhmichel A. Isoimmunization by a new blood factor in tumor cells. Proc. Soc. Exp. Biol. Med. 1951;77: 403-5.
112. Sanger R. An association between the P and Jay systems of blood groups. Nature. 1955;176: 1163-4.
113. Matson GA, Swanson J, Noades J, Sanger R, Race RR. A "new" antigen and antibody belonging to the P blood group system. Am. J. Hum. Genet. 1959;11: 26-34.
114. Hellberg, Poole J, Olsson ML. Molecular basis of the globoside-deficient Pk blood group phenotype. Identification of four inactivating mutations in the UDP-N-acetylgalactosamine:globotriaosylceramide 3-beta-N-acetyl-galactosaminyltransferase gene. J. Biol. Chem. 2002;277: 29455-9.
115. Chandeysson PL, Flye MW, Simpkins SM, Holland PV. Delayed hemolytic transfusion reaction caused by anti-P1 antibody. Transfusion. 1981;21: 77-82.
116. Cheng MS. Potent anti-P1 following blood transfusions. Transfusion. 1984;24: 183.
117. Cantin G, Lyonnais J. Anti-PP1Pk and early abortion. Transfusion. 1983;23: 350-1.
118. Lindstrom K, Dem Borne AE, Breimer ME, Cedergren B, Okubo Y, Rydberg L, Teneberg S, Samuelsson BE. 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.

119. Shirey RS, Ness PM, Kickler TS, Rock JA, Callan NA, Schlaff WD, Niebyl J. The association of anti-P and early abortion. Transfusion. 1987;27: 189-91.
120. Yoshida H, Ito K, Emi N, Kanzaki H, Matsuura S. A new therapeutic antibody removal method using antigen-positive red cells. II. Application to a P-incompatible pregnant woman. Vox Sang. 1984;47: 216-23.
121. Mourant AE, Kope AC, Domaniewska K. The distribution of human blood groups and other polymorphisms: Oxford University Press, 1976.
122. Tippett P, Sanger R, Race RR, Swanson J, Busch S. An agglutinin associated with the P and the ABO blood group systems. Vox Sang. 1965;10: 269-80.
123. Cooling LL, Kelly K. Inverse expression of P(k) and Luke blood group antigens on human RBCs. Transfusion. 2001;41: 898-907.
124. 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 (abstract). Transfusion. 2002;42: 9S.
125. Thorn JJ, Levery SB, Salyan ME, Stroud MR, Cedergren B, Nilsson B, Hakomori S, Clausen H. Structural characterization of x2 glycosphingolipid, its extended form, and its sialosyl derivatives: accumulation associated with the rare blood group p phenotype. Biochemistry. 1992;31: 6509-17.
126. Harris PA, Roman GK, Moulds JJ, Bird GW, Shah NG. An inherited RBC characteristic, NOR, resulting in erythrocyte polyagglutination. Vox Sang. 1982;42: 134-40.
127. Kusnierz-Alejska G, Duk M, Storry JR, Reid ME, Wiecek B, Seyfried H, Lisowska E. NOR polyagglutination and Sta glycophorin in one family: relation of NOR polyagglutination to terminal alpha-galactose residues and abnormal glycolipids. Transfusion. 1999;39: 32-8.
128. Suchanowska A, Czerwinski M, Smolarek D, Majorczyk E, Duk M, Lisowska E, Grodecka M, Wasniowska K (abstract). A possible genetic background of NOR polyagglutination. Vox Sang. 2010;99: 333.
129. Morgan WT, Watkins WM. Blood group P1 substance. I. Chemical properties. Bibl. Haematol. 1964;19: 225-9.
130. Marcus DM. Isolation of a substance with blood-group P1 activity for human erythrocyte stroma. Transfusion. 1971;11: 16-8.
131. Sandhoff K, Kolter T. Biosynthesis and degradation of mammalian glycosphingolipids. Philos. Trans. R. Soc. Lond. B. Biol. Sci. 2003;358: 847-61.
132. Iwamura K, Furukawa K, Uchikawa M, Sojka BN, Kojima Y, Wiels J, Shiku H, Urano T. 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.
133. Suzuki N, Yamamoto K. Molecular cloning of pigeon UDP-galactose:beta-D-galactoside alpha1,4-galactosyltransferase and UDP-

galactose:beta-D-galactoside beta1,4-galactosyltransferase, two novel enzymes catalyzing the formation of Gal alpha1-4Gal beta1-4Gal beta1-4GlcNAc sequence. J. Biol. Chem. 2010;285: 5178-87.
134. Steffensen R, Carlier K, Wiels J, Levery SB, Stroud M, Cedergren B, Nilsson SB, Bennett EP, Jersild C, Clausen H. Cloning and expression of the histo-blood group Pk UDP-galactose: Gal-beta1-4Glc-beta1-Cer alpha1,4-galactosyltransferase. Molecular genetic basis of the p phenotype. J. Biol. Chem. 2000;275: 16723-9.
135. Kojima Y, Fukumoto S, Furukawa K, Okajima T, Wiels J, Yokoyama K, Suzuki Y, Urano T, Ohta M. Molecular cloning of globotriaosylceramide/CD77 synthase, a glycosyltransferase that initiates the synthesis of globo series glycosphingolipids. J. Biol.Chem. 2000;275: 15152-6.
136. Keusch JJ, Manzella SM, Nyame KA, Cummings RD, Baenziger JU. 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.
137. Hellberg Å, Steffensen R, Yahalom V, Sojka BN, Heier HE, Levene C, Poole J, Olsson ML. Additional molecular bases of the clinically important p blood group phenotype. Transfusion. 2003;43: 899-907.
138. Hellberg A, Schmidt-Melbye AC, Reid ME, Olsson ML. Expression of a novel missense mutation found in the A4GALT gene of Amish individuals with the p phenotype. Transfusion. 2008;48: 479-87.
139. McAlpine PJ, Kaita H, Lewis M. Is the DIA1 locus linked to the P blood group locus? Cytogenet. Cell Genet. 1978;22: 629-32.
140. Furukawa K, Iwamura K, Uchikawa M, Sojka BN, Wiels J, Okajima T, Urano T. Molecular basis for the p phenotype. Identification of distinct and multiple mutations in the alpha1,4-galactosyltransferase gene in Swedish and Japanese individuals. J. Biol. Chem. 2000;275: 37752-6.
141. Hellberg A, Chester MA, Olsson ML. Two previously proposed P1/P2-differentiating and nine novel polymorphisms at the A4GALT (Pk) locus do not correlate with the presence of the P1 blood group antigen. BMC Genet. 2005;6:49.: 49.
142. 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.
143. Okuda T, Nakayama K. Identification and characterization of the human Gb3/CD77 synthase gene promoter. Glycobiology. 2008;18: 1028-35.
144. Ingólfsdottir R, Hult A, Olsson ML. Flow cytometric assessment of carbohydrate histo-blood group expression on leucocyte subpopulation (M.Sc. thesis of R. Ingólfsdottir's). Faculty of Medicine, University of Iceland, 2009.

145. Kasai K, Galton J, Terasaki PI, Wakisaka A, Kawahara M, Root T, Hakomori SI. Tissue distribution of the Pk antigen as determined by a monoclonal antibody. J. Immunogenet. 1985;12: 213-20.
146. Lingwood CA. Verotoxins and their glycolipid receptors. Adv. Lipid. Res. 1993;25: 189-211.
147. Kallenius G, Mollby R, Svenson SB, Winberg J, Hultberg H. Identification of a carbohydrate receptor recognized by uropathogenic Escherichia coli. Infection. 1980;8 Suppl 3: 288-93.
148. Kallenius G, Svenson S, Mollby R, Cedergren B, Hultberg H, Winberg J. Structure of carbohydrate part of receptor on human uroepithelial cells for pyelonephritogenic Escherichia coli. Lancet. 1981;2: 604-6.
149. Seltsam A, Hallensleben M, Eiz-Vesper B, Lenhard V, Heymann G, Blasczyk R. A weak blood group A phenotype caused by a new mutation at the ABO locus. Transfusion. 2002;42: 294-301.
150. Seltsam A, Blasczyk R. Missense mutations outside the catalytic domain of the ABO glycosyltransferase can cause weak blood group A and B phenotypes. Transfusion. 2005;45: 1663-9.
151. Yamamoto F, Yamamoto M, Blancher A. Generation of histo-blood group B transferase by replacing the N-acetyl-D-galactosamine recognition domain of human A transferase with the galactose-recognition domain of evolutionarily related murine alpha1,3-galactosyltransferase. Transfusion. 2010;50: 622-30.
152. Twu YC, Hsieh CY, Yu LC. Expression of the histo-blood group B gene predominates in AB-genotype cells. Transfusion. 2006;46: 1988-96.
153. Naiki M, Marcus DM. An immunochemical study of the human blood group P1, P, and PK glycosphingolipid antigens. Biochemistry. 1975;14: 4837-41.