Genetic Characterisation of Human ABO Blood Group Variants with a Focus on Subgroups and Hybrid Alleles

Hosseini Maaf, Bahram

2007

Link to publication

Citation for published version (APA):
Hosseini Maaf, B. (2007). Genetic Characterisation of Human ABO Blood Group Variants with a Focus on Subgroups and Hybrid Alleles Division of Hematology and Transfusion Medicine, Department of Laboratory Medicine, Lund University

General rights
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

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.
Genetic Characterisation of Human
ABO Blood Group Variants with a Focus on
Subgroups and Hybrid Alleles

Doctoral thesis
by
Bahram Hosseini-Maaf

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

With the approval of the Faculty of Medicine at Lund University,
this thesis will be defended on March 16, 2007, at 9:00 in Segerfalksalen,
Wallenberg Neurocentrum, BMC, Sölvegatan 17, Lund

Faculty opponent:
Professor Peter Påhlsson
Department of Biomedicine and Surgery
Division of Cell Biology
Linköping University Sweden
Till min familj
"the chief part of the organisation of every being is simply due to inheritance; and consequently, though each being assuredly is well fitted for its place in nature, many structures now have no direct relation to the habits of life of each species"

The Origin of Species
Charles Darwin, 1859
CONTENTS

ORIGINAL PAPERS ........................................................................................................ 6
ABSTRACT ................................................................................................................... 7
ABBREVIATIONS ....................................................................................................... 8
PROLOGUE ................................................................................................................ 9
BACKGROUND ......................................................................................................... 9

History of blood transfusion and basic principles of the ABO blood group system ...... 9
Definition of blood group antigens and antibodies .................................................. 11
Blood group terminology .......................................................................................... 12
Biochemistry of the ABO blood groups .................................................................. 15
The molecular structure of ABO glycosyltransferase ............................................. 17
Genetics of the ABO blood group system ............................................................... 19
The $ABO$ gene ........................................................................................................ 19
Serology versus molecular genetics ....................................................................... 21
Relationship between phenotype and genotype for the major alleles ................. 21
Genetic background of hybrid alleles, rare $O$ alleles and weak subgroups ............ 27

AIMS OF THE THESIS ............................................................................................... 33
MATERIAL AND METHODS ..................................................................................... 33

Blood and DNA samples .......................................................................................... 33
$ABO$ genotyping .................................................................................................... 33
Polymerase chain reaction-restriction fragment length polymorphism (PCR-RFLP) 33
Allele-specific primer PCR (PCR-ASP) .................................................................. 34
Direct DNA sequencing ............................................................................................ 34
Reverse transcriptase PCR (RT-PCR) ..................................................................... 34
Site-directed mutagenesis and cloning ................................................................. 35
Protein purification and enzyme kinetics ............................................................... 35
Computer modelling and crystallization ............................................................... 35

RESULTS AND DISCUSSION .................................................................................. 36

Single point mutations associated with A or B subgroups (Papers I and V) .......... 36
Hybrid $ABO$ alleles explain some other weak ABO subgroups (Papers I, II and V) 41
The genetic basis of the Abantu phenotype is also a hybrid allele (Paper IV) ...... 45
Novel $O$ alleles involved in unexpected blood group phenotypes (Paper III) ...... 47
Development of an improved $ABO$ genotyping assay based on PCR-ASP amplification across intron 6 (Paper VI) ................................................................. 51

CONCLUSION .......................................................................................................... 54
FUTURE PERSPECTIVES ......................................................................................... 54
SUMMARY IN SWEDISH (sammanfattning på svenska) ....................................... 56
ACKNOWLEDGEMENTS ......................................................................................... 61
REFERENCES .......................................................................................................... 63

APPENDIX: Papers I–VI .......................................................................................... 70
This thesis is based on the following papers, which will be referred to in the text by their Roman numerals (I-VI).


ABSTRACT

ABO is the most important blood group system in transfusion medicine and transplantation immunology. The ABO blood groups differ by the presence or absence of antigens on RBCs and antibodies in plasma. Accurate determination of ABO status is critical. Genomic typing can increase the precision of blood group determination in complicated cases, e.g. when variant expression of A or B antigen is encountered.

The overall aim of this study was to compare the molecular diversity of ABO alleles with various phenotypes, and to contribute to our knowledge of the ABO gene and encoded glycosyltransferases.

Novel alleles (six A\textsuperscript{weak}, eleven B\textsuperscript{weak}, seven O) were identified containing single-point mutations. Structure/function studies explained the weakening of some B subgroup glycosyltransferases. Two new hybrid A\textsuperscript{v} alleles were characterised. Analysis of introns 2-5 revealed 44 previously unknown, allele-related polymorphisms that proved valuable allelic markers. These findings enabled localisation of cross-over regions in two other new hybrids: 1) an O\textsuperscript{v} allele fused with an A\textsuperscript{2} allele, 2) the novel O\textsuperscript{bantu}-A\textsuperscript{2} combination that explained the A\textsuperscript{bantu} phenotype. Phylogenetic and population analyses indicated that O\textsuperscript{bantu} is a unique and distinct evolutionary lineage so far only found among individuals of African descent.

Of clinical importance, a new approach to ABO genotyping was developed that identifies all common alleles, most null and weak A/B subgroups as well as hybrid alleles resulting from recombinational crossing-over events.

In summary, 30 novel alleles were identified and characterized, representing 30% of all alleles reported since the start of this study in 2001.
**ABBREVIATIONS**

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>a.a.</td>
<td>Amino acid</td>
</tr>
<tr>
<td>bp</td>
<td>Base pair</td>
</tr>
<tr>
<td>kb</td>
<td>Kilobase pairs</td>
</tr>
<tr>
<td>DNA</td>
<td>Deoxyribonucleic acid</td>
</tr>
<tr>
<td>EC</td>
<td>Enzyme commission</td>
</tr>
<tr>
<td>GTA or A-enzyme</td>
<td>3-α-N-acetylgalactosaminyltransferase (EC 2.4.1.40) encoded by the blood group A gene</td>
</tr>
<tr>
<td>GTB or B-enzyme</td>
<td>3-α-galactosyltransferase (EC 2.4.1.37) encoded by the blood group B gene</td>
</tr>
<tr>
<td>nt.</td>
<td>Nucleotide</td>
</tr>
<tr>
<td>ISBT</td>
<td>International Society of Blood Transfusion</td>
</tr>
<tr>
<td>MAb</td>
<td>Monoclonal antibody</td>
</tr>
<tr>
<td>PCR</td>
<td>Polymerase chain reaction</td>
</tr>
<tr>
<td>RBC</td>
<td>Red blood cell</td>
</tr>
<tr>
<td>RNA</td>
<td>Ribonucleic acid</td>
</tr>
<tr>
<td>RFLP</td>
<td>Restriction fragment length polymorphism</td>
</tr>
<tr>
<td>ASP</td>
<td>Allele-specific primer</td>
</tr>
<tr>
<td>SNP</td>
<td>Single nucleotide polymorphism</td>
</tr>
<tr>
<td>VNTR</td>
<td>Variable number of tandem repeats</td>
</tr>
</tbody>
</table>
PROLOGUE

Elucidation of the human genome has now been completed through “the Human Genome Project” and the genetic makeup of man is known. The biologically interpretation of the information contained in the genetic sequence is an extremely complex process, resulting in the production of proteins that act as structural elements and metabolic mediators, e.g. enzymes in each cell and organ of an individual. With few exceptions, genetic variation makes every individual exceptional.

Many aspects of this universal uniqueness between individuals are central to transfusion medicine but blood group variation is probably the most important and a prerequisite for modern health care.

The need for donor-recipient matching when performing blood transfusion became apparent at early stages in the history of transfusion medicine, due to the interaction of substances on transfused red blood cells (RBCs), antigens, with other substances, antibodies, in the recipient’s plasma.

The advances in scientific knowledge have allowed characterisation of many of these antigens and antibodies. The structures of the antigens have been elucidated in many cases, some of the biochemical pathways for their biosynthesis have been determined and currently the primary sequences of the genes encoding these antigens are being revealed. The main subject of this thesis will be limited to the ABO blood groups and the following background describes chronologically the development of our knowledge about this important system from the empirical observations of transfusion reactions to the structural variations in the ABO gene itself and its products.

BACKGROUND

History of blood transfusion and basic principles of the ABO blood group system

Blood as a medical term is often related to words including hemo- or hemato- from the Greek stem "haima" for blood. Blood is a circulating tissue composed of fluid plasma and cells (RBCs, white blood cells and platelets).
Blood has been used symbolically in both spoken language as “blood money, stir up bad blood, run in the blood, blood brother, bad blood, blood feud, bloodthirsty” and in the literature in different contexts, e.g. “We few, we happy few, we band of brothers; For he today that sheds his blood with me, Shall be my brother” (Shakespeare: Henry V, act 4).

“The old concept of the movement of blood in the body was elaborated by Galen (129-200 AD). He incorrectly believed that food from the stomach was digested and taken to the liver, where it was transformed into blood, passed to heart, through arteries and veins to the tissues where it was burned up, as wood is consumed by fire” (Farr, 1980).

Different cultures put varying significance on blood. Some believed that it could be the treatment for many maladies and others were convinced that blood contained the soul of man. Ancient Egyptians used it for baths to resuscitate the dead. The first human-to-human blood transfusion recorded took place in 1492. Pope Innocent VIII sank in a coma, and the blood of three ten-year old boys was infused into the dying Pope’s veins at the suggestion of a physician. Unfortunately, both the recipient and all the donors died (Barsoum & Kleeman, 2002). There is no documentation of the fate of the physician, however.

In 1628, British physician W Harvey described the circulation of the blood in humans after having pumped water through the blood vessels of a corpse (Harvey, 1628). This observation opened a new era and some decades later, in 1667, J-B Denis (Denis, 1667) in France and R Lower (Lower, 1666) in England separately reported transfusions from animal to animal and animal to humans. However, Denis was accused of murder when a young boy died after he was transfused the blood from a lamb, and transfusion of animal blood to man was prohibited by law after the death of further patients (Farr, 1979). Nothing of importance was reported in this field until December 22 in 1818, nearly 150 years later, when J Blundell performed his first successful blood transfusion of man to man (Blundell, 1829). The tendency of blood to clot was the principal barrier to transfusions, and to storage of blood. Furthermore, serious reactions to transfusion (e.g. fever, hypotension, allergic shock, death) continued to occur until 1900 when Landsteiner published his preliminary results and the year after described the first system of blood groups, the so-called ABO system (Landsteiner, 1900, 1901). Incompatibility between donor and recipient in these blood groups could explain most of these serious transfusion reactions. Landsteiner based his observation on an unexpected interaction between serum and RBCs from himself and a few of his colleagues at the Institute
of Pathological Anatomy in the University of Vienna. He later received the Nobel Prize for his discovery (Landsteiner, 1930). This elucidation laid the basis of a new field in biology and the foundation of transfusion medicine.

This principal finding led to what has become Landsteiner’s rule, i.e. serological behaviour of the A, B and O (O was initially named type “C”) blood groups (Table 1). Subsequently, another Austrian group, Decastello and Sturli, identified the fourth group, i.e. AB, in 1902 (von Decastello & Sturli, 1902). Several other investigators independently duplicated Landsteiner’s findings. Jansky in Poland named the four blood groups I, II, III and IV based on their frequency in European populations in 1907, and Moss described in 1910 the four blood groups in the reverse order of Jansky (reviewed in Garratty et al., 2000; Farr, 1979).

**Table 1. Basic principles of the ABO system.** Relation between antigens on RBCs and antibodies in serum.

<table>
<thead>
<tr>
<th>ABO blood group</th>
<th>Red cell antigen</th>
<th>Antibodies in serum</th>
</tr>
</thead>
<tbody>
<tr>
<td>O</td>
<td>None</td>
<td>Anti-A, anti-B, anti-A,B</td>
</tr>
<tr>
<td>A</td>
<td>A</td>
<td>Anti-B</td>
</tr>
<tr>
<td>B</td>
<td>B</td>
<td>Anti-A</td>
</tr>
<tr>
<td>AB</td>
<td>A and B</td>
<td>None</td>
</tr>
</tbody>
</table>

In clinical practice, the ABO blood group system is one of the most important since the A and B epitopes may provoke a strong immune reaction. With the introduction of blood typing and crossmatching techniques, blood transfusion became not only a simple but also a much safer procedure. However, blood transfusion has continued to be associated with risks, notably the transmission of blood-borne infectious diseases such as syphilis, viral hepatitis and human immunodeficiency virus. Furthermore, although ABO typing reduced the occurrence of transfusion reactions, they still occurred, indicating the presence of other genetic differences in blood groups of importance in transfusion medicine, as well as in the later emerging field of organ transplantation.

**Definition of blood group antigens and antibodies**

The term blood group is generally based on the presence or absence of certain antigens on the RBC membrane. These are identified by characteristic agglutination reactions with specific antibodies and this field is referred to as blood group serology (Daniels, 2002).
An antigen is in fact any substance that is able to stimulate the production of an antibody under certain conditions. Some blood group antigens, e.g. in blood group systems ABO, P1, H, LE and I, are due to differences in carbohydrate structures and they are all found on glycolipids or glycoproteins on the RBC membrane. In the case of the ABO blood groups, the antigens are found on the surface of RBCs and other cells and is recoverable from body fluids and secretions. Most other blood group systems result from polymorphisms on different RBC membrane-associated proteins (Daniels, 2002).

Blood group antibodies can be divided into induced and naturally-occurring antibodies. Induced antibodies are raised by the immune system as a response to specific antigens, e.g. following transfusion, pregnancy or transplantation. Natural antibodies are also of immune origin, with the difference that the antigen eliciting their production is unknown. In the ABO system, these antibodies are sometimes referred to as isoagglutinins because they can cause blood cells from certain individuals to clump together. The two isoagglutinins, anti-A and anti-B, occur naturally in humans, contrary to most other blood groups antibodies (Erskine & Socha, 1978; Mollison et al., 1993).

Naturally-occurring antibodies can be detected in serum of persons who have never been transfused or women who never have been pregnant (Mollison et al., 1993). Natural antibodies may occur because of unidentified infections during early life or, in the case of blood group antibodies, introduction of substances that are antigenically similar to blood group substances. Interestingly, Springer et al., as early as 1959, demonstrated by an experiment with chickens bred under sterile conditions, that some of these antibodies appear following stimulation by certain bacteria, such as Escherichia coli, while the appearance of others was related to exposure to various kinds of food. The origin of natural antibodies is still a matter of debate, though.

Naturally-occurring antibodies are often IgM and sometimes also consist of an IgG component. With exception for antibodies against ABO, P^+ and P, the majority of naturally-occurring antibodies are only cold reactive and of limited clinical importance.

**Blood group terminology**
A rather incoherent terminology has evolved as more and more blood group antigens were discovered during several decades. Providing a common acceptable and reliable terminology,
which would cover both the serology and genetics, has been difficult. The International Society of Blood Transfusion (ISBT) Working Committee on Terminology for Red Cell Surface Antigens (http://www.blood.co.uk/ibgrl/) was set up in 1980 to establish and define a meaningful nomenclature for different blood groups as shown in Table 2 (Daniels et al., 2003; 2004; Storry & Olsson, 2004). Every valid blood group antigen is given a six digit identification number, according to the following categorization:
- Blood group systems; one or more antigens governed either by a single gene locus or by a complex of two or more very closely linked homologous genes with virtually no recombination occurring between them. There are 29 different systems to date, and the first three digits represent the systems (001-029).
- Blood group collections; genetically-, biochemically-, or serologically-related sets of antigens, which do not, at this time, merit system status because of insufficient data (205-211).
- A series of low frequency-antigens (the 700 series)
- A series of high frequency-antigens (the 901 series)

The symbol for a gene or cluster of genes controlling a blood group system is often the italicized symbol for the system. The ABO genotypes should consequently be written in italicised capital letters (with superscripts indicating the subgroup alleles).

However, ABO allele nomenclature poses significant problems that are still under consideration by the ISBT. In the absence of an officially agreed terminology, in this thesis, alleles are referred to by their serological activity and an alternative allele name is given in square brackets.
ISBT has recently set up a subgroup from the Terminology Working Committee mentioned above. Its task is to create a unified allele nomenclature for blood groups, (ML Olsson personal communication).
Table 2. Human blood group systems recognised by the ISBT (http://www.blood.co.uk/ibgrl/).

<table>
<thead>
<tr>
<th>System name</th>
<th>System no. ISBT</th>
<th>Symbol ISBT</th>
<th>Antigens no.</th>
<th>Gene</th>
<th>Locus</th>
<th>Epitope / carrier, or functions</th>
</tr>
</thead>
<tbody>
<tr>
<td>ABO</td>
<td>001</td>
<td>ABO</td>
<td>4*</td>
<td>ABO</td>
<td>9q34.1-q34.2</td>
<td>Carbohydrate</td>
</tr>
<tr>
<td>MNS</td>
<td>002</td>
<td>MNS</td>
<td>43</td>
<td>GYP A</td>
<td>1q28.2-q31.1</td>
<td>Glycophorin A&amp;B (CD235a-b)$^3$</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>GYP B</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>GYPE</td>
<td></td>
<td></td>
</tr>
<tr>
<td>P</td>
<td>003</td>
<td>PI</td>
<td>1</td>
<td>PI</td>
<td>22q11.2-pter</td>
<td>Carbohydrate</td>
</tr>
<tr>
<td>Rh</td>
<td>004</td>
<td>RH</td>
<td>56</td>
<td>RH</td>
<td>1q34-36.2</td>
<td>Protein, CD240D CD240E</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>RHD</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>RHE</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lutheran</td>
<td>005</td>
<td>LU</td>
<td>19</td>
<td>LU</td>
<td>19q13.2</td>
<td>Glycoprotein (IgSF$^3$), adhesion molecule, CD239</td>
</tr>
<tr>
<td>Kell</td>
<td>006</td>
<td>KEL</td>
<td>23</td>
<td>KEL</td>
<td>7q33</td>
<td>Glycoprotein, CD238</td>
</tr>
<tr>
<td>Lewis</td>
<td>007</td>
<td>LE</td>
<td>6</td>
<td>FUT3</td>
<td>19p13.3</td>
<td>Carbohydrate</td>
</tr>
<tr>
<td>Duffy</td>
<td>008</td>
<td>FY</td>
<td>6</td>
<td>FY</td>
<td>1q22-2q23</td>
<td>Glycoprotein, receptor, CD234</td>
</tr>
<tr>
<td>Kidd</td>
<td>009</td>
<td>JK</td>
<td>3</td>
<td>JK</td>
<td>18q11-1q12</td>
<td>Glycoprotein, urea transporter</td>
</tr>
<tr>
<td>Diego</td>
<td>010</td>
<td>DI</td>
<td>19</td>
<td>AE1</td>
<td>17q12-2q21</td>
<td>Glycoprotein (band 3, AE1), CD233</td>
</tr>
<tr>
<td>Yt (Cartwright)</td>
<td>011</td>
<td>YT</td>
<td>2</td>
<td>ACHE</td>
<td>7q22</td>
<td>Enzyme, acetylcholinesterase</td>
</tr>
<tr>
<td>Xg</td>
<td>012</td>
<td>XG</td>
<td>1</td>
<td>XG</td>
<td>Xp22.32</td>
<td>Glycoprotein, CD99</td>
</tr>
<tr>
<td>Scianma</td>
<td>013</td>
<td>SC</td>
<td>3</td>
<td>ERM A</td>
<td>1q36.2-p22.1</td>
<td>Glycoprotein</td>
</tr>
<tr>
<td>Dombrock</td>
<td>014</td>
<td>DO</td>
<td>5</td>
<td>DO</td>
<td>12q13.2-21.2</td>
<td>Glycoprotein, CD297</td>
</tr>
<tr>
<td>Colton</td>
<td>015</td>
<td>CO</td>
<td>3</td>
<td>AQP1</td>
<td>7q14</td>
<td>Aquaporin 1 (channel)</td>
</tr>
<tr>
<td>Landsteiner/Weiner</td>
<td>016</td>
<td>LW</td>
<td>3</td>
<td>LW</td>
<td>19p13.3</td>
<td>LW glycoprotein (IgSF$^3$) CD242</td>
</tr>
<tr>
<td>Chido-Rogers</td>
<td>017</td>
<td>CH/RG</td>
<td>9</td>
<td>C4A,C4B</td>
<td>6p21.3</td>
<td>Complement protein 4</td>
</tr>
<tr>
<td>Hh</td>
<td>018</td>
<td>H</td>
<td>1</td>
<td>FUT1</td>
<td>19q13</td>
<td>Carbohydrate, CD173</td>
</tr>
<tr>
<td>Kx</td>
<td>019</td>
<td>XK</td>
<td>1</td>
<td>XK</td>
<td>Xp21.1</td>
<td>Glycoprotein</td>
</tr>
<tr>
<td>Gerbich</td>
<td>020</td>
<td>GE</td>
<td>7</td>
<td>GYPC</td>
<td>2q14-2q21</td>
<td>Glycophorins C and D, CD236</td>
</tr>
<tr>
<td>Cromer</td>
<td>021</td>
<td>CROM</td>
<td>10</td>
<td>DAF</td>
<td>1q32</td>
<td>Glycoprotein, CD55</td>
</tr>
<tr>
<td>Knops</td>
<td>022</td>
<td>KN</td>
<td>5</td>
<td>CR1$^1$</td>
<td>1q32</td>
<td>Glycoprotein, CR1 or CD35</td>
</tr>
<tr>
<td>Indian</td>
<td>023</td>
<td>IN</td>
<td>2</td>
<td>CD44</td>
<td>11p13</td>
<td>Glycoprotein, CD44</td>
</tr>
<tr>
<td>Ok</td>
<td>024</td>
<td>OK</td>
<td>1</td>
<td>BSG</td>
<td>19p13.3</td>
<td>Glycoprotein, CD147</td>
</tr>
<tr>
<td>Raph</td>
<td>025</td>
<td>RAPH</td>
<td>1</td>
<td>MER2</td>
<td>11p15.5</td>
<td>Transmembrane glycoprotein, CD151</td>
</tr>
<tr>
<td>John Milton-Hagen</td>
<td>026</td>
<td>JMH</td>
<td>15q23-2q24</td>
<td>SEMA7A</td>
<td>15q23-2q24</td>
<td>Glycoprotein, CD108</td>
</tr>
<tr>
<td>I</td>
<td>027</td>
<td>I</td>
<td>6p24</td>
<td>GCNT2</td>
<td>6p24</td>
<td>Carbohydrate</td>
</tr>
<tr>
<td>Globoside</td>
<td>028</td>
<td>GLOB</td>
<td>3q25</td>
<td>B3GALNT3</td>
<td>3q25</td>
<td>Glycolipide</td>
</tr>
<tr>
<td>GIL</td>
<td>029</td>
<td>GIL</td>
<td>9p13</td>
<td>AQP3</td>
<td>9p13</td>
<td>Aquaporin 3 (channel)</td>
</tr>
</tbody>
</table>

* Four ABO antigens are recognised by the ISBT and noted as 001 (A), 002 (B), 003 (AB) and 004 (A1)
$^1$ Cluster of Differentiation (antigens)
$^3$ The immunoglobulin superfamily, IgSF, http://imgt.cines.fr/textes/IMGTindex/superfamily.html
$^3$ Complement component (3b/4b) receptor 1 (Knops blood group)
In 1999 a database, The Blood Group Antigen Gene Mutation Database (dbRBC), was also set up under support of Human Genom Variation Society (HGVS) lead by O Blumenfeld (Blumenfeld & Patnaik, 2004). This database contains information about variations in genes that affect blood group expression (http://www.ncbi.nlm.nih.gov/projects/mhc/xslcgi.fcgi?cmd=bgmut/home).

Biochemistry of the ABO blood groups
The biochemical basis of the ABO and H antigens is well understood due to intensive studies during the 1950s and 1960s by the pioneering work on ovarian cyst fluids (which contain large amounts of water-soluble blood-group-active glycoproteins) by Morgan & Watkins and Kabat. Subsequently, this led to the explanation of biosynthesis pathways for ABH antigens, Figure 1 (Kabat, 1956; references and review in Watkins 1980; Lowe, 1995).

The ABO antigens are not limited to erythroid tissues, but are also found in different tissues and on some epithelial cells (Oriol et al., 1992). Therefore, they can sometimes be noted as histo-blood group antigens (Clausen & Hakomori, 1989). On the other hand, they have not been observed in connective tissues, muscles, and the nervous system (Schenkel-Brunner, 2000). However, in this thesis, as in the field of transfusion medicine, these antigens will be referred as blood groups and not histo-blood groups.

The H antigen is the immediate precursor for creation of blood-group-A-and B-active structures. Two genes, \( FUT1 \) (responsible for synthesis of H antigen on type 2 precursor, \( e.g. \) on RBC) or \( FUT2 \) (responsible for synthesis of H antigen on type 1 precursor, \( e.g. \) in secretions) on chromosome 19 encode fucosyltransferases (Oriol, 1995; Lowe, 1995). These 2-\( \alpha \)-L-fucosyltransferases catalyse the transfer of L-fucose from the guanosinediphosphofucose (GDP-L-fucose) donor to the terminal galactose of at least one of the precursor types. Six precursor types have been identified but only four types (Table 3) are known to carry ABO activity (Clausen & Hakomori, 1989; Clausen et al., 1994). These substances subsequently act as acceptor substrate for the enzyme product of the \( ABO \) gene. The \( ABO \) gene product, \( i.e. \) glycosyltransferases A and B (GTA and GTB) use UDP-GalNAc and UDP-Gal, respectively, as substrates.
Table 3. H antigens precursors.

<table>
<thead>
<tr>
<th>Type</th>
<th>Terminal disaccharide structure of precursor</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Galβ1→ 3GlcNAcβ1→R (e.g. endodermal cells)</td>
</tr>
<tr>
<td>2</td>
<td>Galβ1→ 4GlcNAcβ1→R (e.g. mesodermal cells, e.g. erythrocytes)</td>
</tr>
<tr>
<td>3</td>
<td>Galβ1→ 3GalNAcα1→R (e.g. O-linked, linkage to Ser/Thr, repeating A)</td>
</tr>
<tr>
<td>4</td>
<td>Galβ1→ 3GalNAcβ1→R (e.g. Glycolipids)</td>
</tr>
</tbody>
</table>

R, indicates the carrier which can be carbohydrate, glycolipid or glycoprotein

ABO(H) antigens are found most commonly on type 1 and type 2 precursors. Type 1 chains are the main carriers of the ABH antigens in plasma and other body fluids. Type 2 precursor chains are the primary acceptors for ABO glycosyltransferases associated with the RBC membrane (Oriol, 1995). Type 3, repetitive A-associated chains are present on RBC glycolipids or secreted in blood group A individuals (Clausen et al., 1985, 1986a-b). Type 4 chain are found in small amounts on RBCs and in larger amounts in kidney cells of persons with blood group A (Breimer & Jovall, 1985). Type 4 ABH structures are only located on glycolipids (Kanagi et al., 1984).

Blood group A individuals possess a 3-α-N-acetylgalactosaminyltransferase (GTA), which catalyzes the transfer of N-acetyl-D-galactosamine from the donor UDP-GalNAc to the H precursor structure Fucα(1-2)Galβ-R to provide an A determinant GalNAcα(1-3)[Fucα(1-2)]Galβ-R. Blood group B individuals possess a 3-α-galactosyltransferase (GTB), which uses the same H structure but catalyzes the transfer of galactose from UDP-Gal to make the B determinant Galα(1-3)[Fucα(1-2)]Galβ-R. Blood group O individuals do not have any functional enzyme and AB individuals have both GTA and GTB (reviewed in Oriol, 1995; Lowe, 1995). The relationship between the A, B and H determinant structures of the ABO blood group is illustrated in Figure 1.

**Figure 1.** Schematic representation of the biosynthetic pathway for ABO(H) blood group structures shows the link between H, A and B antigens.
The existence of subgroups of A, i.e. A₁ and A₂, was first recognized by von Dungern and Hirszfeld in 1911 (von Dungern & Hirszfeld, 1911). A₁ RBCs have precursor types 1, 2, 3 and 4, while A₂ erythrocytes appear to have mainly type 1 and type 2, and even small amounts of type 3 (Clausen et al., 1985, 1986b).

The A₁ and A₂ transferases are qualitatively different with pH optima at 5.6 for A₁ and 7-8 for A₂. The activity of A₁ is higher, it has a lower $K_m$ for acceptors, and A₁ has a higher isoelectric point than the A₂ enzyme (Watkins, 1980). There is also a quantitative difference between A₁ and A₂ cells that has been illustrated by various techniques, among others, radioimmunoassay with labelled antibodies and lectins, electron microscopy. These studies showed that the number of antigen sites per RBC for A₁ is 8-12x10⁵ and for A₂ is 1-4x10⁵ (Economidou et al., 1967; Cartron et al., 1974).

Expression of ABO(H) antigens may be affected by different factors, e.g. the amount and type of precursor, place of expression and disease. The development of ABO(H) antigens is a complex process and many factors are involved and may affect the final antigen. No attempts will be made to go deeply into the biosynthesis of ABO blood groups in this thesis because of the complexity of the topic and taking into account that the scope of the study is mainly genetic.

**The molecular structure of ABO glycosyltransferase**

Glycosyltransferases (GTs), [Enzyme Commission, (EC) 2.4], make up a large family of enzymes that are involved in the biosynthesis of oligosaccharides, polysaccharides, and glycoconjugates. These enzymes show an enormous diversity and are present both in prokaryotes and eukaryotes (Breton et al., 2006). Glycosyltransferases are classified into 87 different families based on substrate/product stereochemistry according to the CAZy database (http://afmb.cnrs-mrs.fr/~pedro/CAZY/db.html) (Henrissat, 1998). Knowledge of the sequences of the ABO genes (Yamamoto et al., 1990a-b) have established that mammalian GTA (EC 2.4.1.40) and GTB (EC 2.4.1.37) are type II integral membrane proteins containing 354 amino acids (a.a.) and are localised in the lumen of the Golgi apparatus. Type II integral membrane enzymes typically have a short amino-terminal cytoplasmic tail, a hydrophobic membrane domain, a short protease-sensitive stem region and a large catalytic domain that includes the carboxy terminus (Paulson & Colley, 1989), as shown in Figure 2a. GTA and
GTB are very similar in the coding regions and the soluble enzyme can be found in serum (Schenkel-Brunner et al., 1972), urine (Chester, 1974) and milk (Ginsburg, 1972).

The high-resolution X-ray-crystal structures of the catalytic domains (a.a. residues 63-354) in the GTA and GTB enzymes were resolved by Patenaude and co-workers (Patenaude et al., 2002). Figure 2b shows a three-dimensional surface model that predicts the appearance of the active domain in humans. The glycosyltransferase structure has a central cleft where the enzymatic action takes place (Figure 2b and 2c). The enzyme contains an acceptor-recognition domain that binds H antigen and a donor recognition domain that binds UDP-GalNAc or UDP-Gal. GTA and GTB are retaining, metal-dependent family glycosyltransferases (Boix et al., 2001; Pak et al., 2006).

GTA and GTB require the metal ion Mn$^{2+}$ (manganese) for activity and have a characteristic $^{211}$DVD$^{213}$ motif that coordinates attachment of the phosphate in the donor via the metal ion (Breton et al. 2006). In vitro conversion of O cells to A by GTA as mentioned above, requires the presence of Mn$^{2+}$ ions but if Mn$^{2+}$ is changed to Mg$^{2+}$, A$_1$ glycosyltransferase remains active, but A$_2$ transferase does not (Schachter et al., 1973).

**Figure 2.** ABO glycosyltransferase. The topology of the transferase is shown in (a) as Golgi-localised membrane-bound enzyme with a cytoplasmic N-terminus and catalytic C-terminal domain (modified from Paulsen & Colley, 1989). The 3D-surface model (b) was created with the Deep View Swiss Pdb Viewer version 3.7 and the 3D-ribbon structure (c) was generated using SETOR and SetoRibbon (unpublished).
Genetics of the ABO blood group system

The genetic background for a protein polymorphism is straightforward. The occurrence of most such variations is due to a single nucleotide polymorphism (SNP) that results in a single a.a. change, which can directly affect the protein’s antigenicity if it is a blood group carrier. The genetics governing carbohydrate structures (e.g. ABO antigens) are more complex because the sequential action of a number of gene products, i.e. enzymes, is needed to generate antigens of carbohydrate nature. Any polymorphism within these genes may create extra sources of complexity.

In addition, many different genetic backgrounds may underlie one phenotype and different phenotypes appears to be possible result from the same genetic variant. This makes phenotype prediction from an ABO genotype difficult (but not impossible).

The ABO gene

The ABO blood group locus is located on the terminal portion of the long arm of chromosome 9 and has been determined through quantitative red cell adenylate kinase (AK-1) assay (Ferguson-Smith et al., 1976). This was later confirmed by linkage analysis (Allderdice et al., 1986) to position 9q34.1-q34.2.

The first successful isolation and purification of GTA, followed by cloning of cDNA representing the blood group A transferase mRNA, were completed in 1990 from human lung tissue (Clausen et al., 1990; Yamamoto et al., 1990a-b).

The coding region of the ABO gene consists of seven exons (exon 1 exists in two variants), spanning about 19.5 kb and ranging in size from 28 to 688 bp (Figure 3). The 1062 bp sequence encodes 354 a.a. corresponding to a 41 kDa protein. The ABO gene contains six introns with sizes ranging from 554 to 12982 bp.

Figure 3. Organisation of the ABO gene. The seven exons and six introns are not drawn to scale. The numerals above boxes represent the first and last nucleotides of the coding region in each exon and those below boxes show the corresponding a.a. numbers (a). The size of each intron is indicated with a thin oblique bar (a). The ABO gene drawn to scale; exons are black and introns grey (b).
The two longest exons (6 and 7) contain 823 of 1062 bp in the transcribed \( A^1 \)-mRNA. This covers 77% of the whole encoded protein or 91% of the catalytic domain of the ABO glycosyltransferase. Exons 1-5 encode the amino-terminal cytosolic domain, a transmembrane domain (a.a.17-37), the stem region, and the remaining 9% of the catalytic domain (Figure 2a) (Yamamoto et al., 1990a-b; Bennett et al., 1995; Paulson & Colley, 1989). An alternative exon 1a has been identified, located 682 bp upstream of the original exon 1. This exon was found in AC133-CD34+ cultured cells obtained from peripheral blood and represents approximately 2% of all transcripts (Kominato et al., 2002).

Our knowledge of transcriptional regulation of the \( ABO \) gene is still limited. The presumed promotor of \( ABO \) was identified by sequencing and functional testing of the 5’-upstream region and it is located between -117 and +31, relative to the upstream translation codon (Kominato et al., 2002; Hata et al., 2002). The promotor region contains CpG islands that can start transcription but neither CAAT or TATA (TATAAAA) boxes were found close to these sites (Yamamoto et al., 1995; Kominato et al., 1997). An enhancer-active minisatellite motif (CBF/NF-Y), located approximately 3.8 kbp upstream of exon 1 in the \( ABO \) gene, contains either one (alleles \( A^1 \) [A101] and \( O^2 \) [O03]) or four (\( A^2 \) [A201], \( O^1 \) [O01], \( O^1v \) [O02] and \( B \) [B101]) 43-bp repetitive units (also recognized as a variable number of tandem repeats, VNTR) and appears to play a role in expression (Kominato et al., 1997; Irshaid et al., 1999; Yu et al., 2000), Figure 4. A nucleotide substitution G>A at nt. 41 of 43 was identified in all alleles with only one repeat (Irshaid et al., 1999). This region contains a CBF/NF-Y binding site and mutations in this site appeared to decrease expression in an experimental gastric cell line system (Kominato et al., 1997) but a recent report found transcript levels to be independent of VNTR status (Thuresson et al., 2006). A zinc finger transcription factor, Sp1 or Sp1-like protein, may also play a role since there is a binding site in the proximal promoter of \( ABO \) (Hata et al., 2002, 2003), Figure 4.

Figure 4. Schematic depiction of the upstream region in different alleles of the \( ABO \) gene. The figure indicates the relative locations of exon 1, 1a and the repetitive 43 bp units with binding motifs for CBF/NF-Y in the enhancer region and for Sp1 in the proximal promoter. The figure is modified from an original kindly provided by B Thuresson.
The ABO polymorphism is a relatively recent evolutionary event. It probably developed 13 million years ago from an old ancestral gene (Martinko et al., 1993). It has been proposed that A and B antigens were established long before humans and primates diverged (Saitou et al., 1997). A high homology in nt. and a.a. sequences was detected in the ABO genes of primates and furthermore human ABO shares significant sequence homology with primates (Kominato et al., 1992). Other investigations showed, e.g. the murine gene to consist of at least six exons and encoding a cisAB-like gene with both GTA and GTB activities but unlike the separate human GTA and GTB (Yamamoto et al., 2001). The genetics of other species with special relevance for human research have been described, e.g. the porcine ABO gene (Yamamoto & Yamamoto, 2001). The ABO equivalents in other animals like cats (Griot-Wenk et al., 1993) and dogs (Yamamoto et al., 2001) have also been studied but will not be discussed further here. Although the genetic basis of many subgroups has been mapped at the molecular level, samples with unresolved structure are still encountered in a reference laboratory. Numerous point mutations resulting in a number of a.a. changes abolish or weaken the enzymatic activity of ABO glycosyltransferases.

**Serology versus molecular genetics**

**Relationship between phenotype and genotype for the major alleles**

Prior to DNA analysis of the ABO gene, serological analyses have been the only practical way to determine blood groups. In general the ABO typing process has two steps which determine either the presence or absence of A and B antigens on the RBCs and also the presence or absence of IgM anti-A and anti-B in the serum or plasma. These two parts of the ABO grouping are sometimes referred to as forward and reverse typing. Human or monoclonal murine anti-A and anti-B reagents are mixed with the RBCs in the first step. The second step involves mixing the serum with suspensions of RBCs known to be group A or B. For the blood group to be interpretable as one of the four major ABO groups A, B, O or AB, the forward and reverse typings must be concordant. Discrepancies can be due to inherited or acquired (e.g. transfusion, chimerism, disease etc) phenotype variants. Other ABO subgroups with different A and B characteristics have reduced amounts of A and B antigens on the RBC membrane.

Before this study, there were approximately 70 ABO alleles reported to the Blood Group Antigen Gene Mutation Database, “dbRBC” (Chester & Olsson, 2001). The majority of these
alleles are very uncommon; some of them have arisen by simple SNPs, whilst others were formed by recombination between two ABO alleles. Certain alleles are more common depending on which populations are involved. The frequency of ABO phenotypes varies notably between different geographic areas and ethnic groups. E.g populations with a high frequency of A phenotype are found mainly in Northern and central Europe. The A₂ phenotype is most common among the Sami people in Northern Scandinavia and the general frequency of A₂ is relatively high in Europe, Southwestern Asia and Africa. On the contrary, A₂ is very rare in native Americans and Australian Aborigines but is also rare in the Eastern Asia (Mourant & Kopec, 1976). On the other hand, the B phenotype is more common in the Far East but almost absent in Amerindians (Mourant & Kopec, 1976). Blood group O is very frequent among native American Indians. Parts of Australia and Africa also show high frequencies of blood group O (Mourant & Kopec, 1976).

This thesis will henceforth consider the A¹, A², B, O¹, O¹v and O² alleles as the six major ABO alleles. A₁ and A₂ subgroups can be distinguished serologically as shown in Table 4. RBCs from A₁ and A₂ persons both react strongly with monoclonal or human anti-A reagents in direct agglutination tests. Distinction between RBCs of these two subgroups can be made with preparations of the lectins from Dolichos biflorus seeds or from albumin glands of vineyard snails, Helix pomatia. These lectins react both with A₁ and A₂ RBCs, but at the right dilution, the lectin reagents will readily differentiate A₁ and A₁B from A₂ and A₂B. In addition, the plasma from A₁ individuals may contain Anti-H while A₂ person can make anti-A₁.

Table 4. Typical serological reactions in blood samples from persons with A₁ and A₂ phenotype.

<table>
<thead>
<tr>
<th>Phenotype</th>
<th>Forward typing</th>
<th>Reverse typing</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Anti-A</td>
<td>Anti-B</td>
</tr>
<tr>
<td>A₁</td>
<td>4⁺</td>
<td>-</td>
</tr>
<tr>
<td>A₂</td>
<td>4⁺</td>
<td>-</td>
</tr>
</tbody>
</table>

* or 1⁺ to 4⁺ representing anti-H
** or 1⁺ to 4⁺ representing anti-A₁

A³ alleles
The A³ allele represents the "wild-type" A₁ phenotype. The sequence of the A³ [A101] allele is often denoted as the "consensus sequence" in an ABO genotype context. There are variants of A³ alleles, one of which is very common in Asian populations with a 467C>T [A102]
polymorphism resulting in the substitution Pro156Leu (Yamamoto et al., 1990a). Two minor $A^I$ alleles [A103 and A104] have been described and differ as follows. The first one also has the 467C>T point mutation but an additional silent mutation 567C>T, the second allele contains a silent polymorphism (also found in $B$ and $O^2$ alleles) in nucleotide 297A>G (Ogasawara et al., 1996b).

A105 is like A102 with the same mutation in exon 7, 467C>T but analysis of intron 6 showed additional SNPs compared to A102 (Ogasawara et al., 2001). Another $A^I$ allele [A106] contains both 297A>G and 467C>T (Ogasawara et al., 2001).

$A^I$ alleles
The $A_2$ subgroup is the most common $A$ phenotype after the $A_1$ subgroup and the serology is discussed above. The main genetic difference between $A^I$ [A101] and $A^I$ alleles [A201] is one point mutation in exon 7, 467C>T (Pro156Leu), and a deletion of one of the three cytosines at nt. 1059-1061 (CCC to CC). The latter mutation results in an extension of the reading frame by 64 nucleotides. This deletion occurs in the codon before the translation stop codon (TGA), resulting in a gene product with an extra 21 a.a. at its C-terminus (Yamamoto et al., 1992). The glycosyltransferases encoded by the $A^I$ allele have lower efficiency, leading to a weaker $A$ phenotype. The enzyme activity has been suggested to be decreased by 30-50 times compared to $A_1$, at least according to immunostaining of HeLa cells transfected with A201 constructs (Yamamoto et al., 1992). Some other variants of $A^I$ alleles have subsequently been elucidated (A202-206). Three $A^I$-like alleles were shown to have three different single mutations near the 3’ end of exon 7 by Ogasawara. $A^I$-2 [A202] contains 1054C>T (Arg352Trp), $A^I$-3 [A203] with 1054C>G (Arg352Gly) (Ogasawara et al., 1996a) and $A^I$-4 [A205] with both 467C>T and an additional new mutation 1009A>G (Arg337Gly) (Ogasawara et al., 1998). A204 with four common $B$-related base substitutions (297A>G, 526C>G, 657C>T and 703G>A) seems to be a hybrid with two extra substitutions 771C>T (silent mutation) and 829G>A (Val277Met) (Ogasawara et al., 1996a). One other rare $A^I$ subgroup, $A^I$-5, [A206] carried only the single deletion (1061delC) without the usual SNP 467C>T (Olsson & Chester, 1996b; Yip, 2000).

$B$ alleles
Serologically, blood group B shows strong reaction with anti-B in forward typing. Accordingly, in reverse typing the appearance of agglutination confirms the presence of $A$
antibodies in the B plasma. High frequencies of group B are found in Central and East Asia and differ significantly compared to Europe, i.e. the frequency decreases from east to west (Mourant & Kopec, 1976). When cloning cDNA sequences from cell lines, B-1 [B101] was found in the clone FY-59-5 and showed seven single nucleotide substitutions, 297 (silent mutation), 526, 657 (silent mutation), 703, 796, 803 and 930 (silent mutation) throughout exon 6 and 7 and later one extra mutation outside the coding region at the 3’ end, at nt. 1096 (Olsson & Chester, 1995). The four a.a. substitutions governed by nt. 526C>G (Arg176Gly), 703G>A (Gly235Ser), 796C>A (Leu266Met) and 803G>C (Gly268Ala) discriminate GTA from GTB (Yamamoto & Hakomori, 1990; Yamamoto et al., 1990a). ABO glycosyltransferases can accordingly be described by using the letters A and B to illustrate the derivation of the a.a. at these four residues. GTA would be represented by AAAA indicating the presence of Arg/Gly/Leu/Gly and similarly BBBB would describe the GTB with Gly/Ser/Met/Ala at residues 176/235/266 and 268. The substitutions at positions 266 and 268 were shown to be responsible for the nucleotide/donor specificity of the transferases (Yamamoto & Hakomori, 1990) and the other residues may have a role in acceptor binding and turnover (Patenaude et al., 2002). Some other variants have been reported afterwards that differ from B-1 [B101] by lacking the 930 substitution for B-2 [B102], the 657 substitution for B-3 [B103] (Ogasawara et al., 1996b), the 526 substitution for B-4 [B107] (Ogasawara et al., 1998) and 297 for B108 (Ogasawara et al., 2001).

O alleles
The serologically-determined pattern of blood group O demonstrates the absence of A and B antigens on the RBC surface in forward blood typing. Reverse blood typing indicates the presence of both anti-A and -B in the plasma.

The first O allele [O1-1, O01] cloned was shown to be identical to the consensus A allele [A101] except for a nucleotide deletion, 261delG, in exon 6. This results in a shift in the reading frame, giving rise to a truncated protein that alters the protein sequence after a.a. 88. A stop codon halts translation after a.a. 117 and the resulting protein is enzymically inactive (Yamamoto et al., 1990a). Some other O1-1-like alleles that are characterized by the presence of the 261delG and at least one additional point mutation are presented in Table 5.

A second kind of O allele has the same inactivating deletion (261delG) as the original O allele (O1-1 [O01]), but in addition has nine point mutations spread throughout exons 3 to 7.
(Yamamoto et al., 1990a; Olsson & Chester, 1996a) and a further 13 mutations have been found amongst the intron 6 (Suzuki et al., 1997; Olsson & Chester, 1998a). This allele is referred to as $O^1_{\text{variant}} (O^{1\text{-}}-1) [O02]$ since the inactivating mutation is the same as in the original $O^1$ allele. Table 5 shows some more variants of this allele.

Other $O$ alleles not due to 261delG also exist that are caused by other inactivating mutations along the reading frame. The first allele described of this type, $O^{2\text{-}}-1 [O03]$, has a critical mutation (802G>A) which causes an a.a. change (Gly268Arg) that prevents the enzyme from utilising the nucleotide sugar donor (Yamamoto et al., 1993, 1996). Additionally, two polymorphic sites in exon 2 and 5, 53G>T and 220 C>T, respectively (Amado et al., 2000) and 297A>G in exon 6, 526 C>G in exon 7 (Yamamoto et al., 1993d) and 1096G>A (Olsson & Chester, 1995) have been found. This $O$ allele comprises approximately 2-5% of $O$ alleles in Caucasians but seems to be absent, or at least very rare, in other populations (Grunnet et al., 1994; Chester & Olsson, 2001).

It has been debated whether $O$ alleles are translated into protein at all. Even if definitive proof is lacking, the presence of $O_2$ protein has been shown by MAb staining of tissue sections (Amado et al., 2000) and a hint towards alloreactivity against polymorphic peptides derived from the N-terminal parts of $O^1$- and $O^{1\text{-}}$-translated sequences suggested in mixed lymphocyte cultures (Eiz-Vesper et al., 2005).
Table 5. Schematic comparison of known \(O\) allele mutations and deduced a.a. sequence, \([O^1-1, O01], O^{varint} (O^1-1)[O02]\) and \(O^2-1 [O03]\).

| Exon | 3  | 5  | 6  | 7  | 106 | 188 | 189 | 190 | 220 | 261 | 297 | 318 | 454 | 467 | 526 | 542 | 579 | 595 | 646 | 657 | 681 | 721 | 771 | 802 | 829 | 930 | 1054 | 1096 |
|------|----|----|----|----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| nt. position | 106 | 188 | 189 | 190 | 220 | 261 | 297 | 318 | 454 | 467 | 526 | 542 | 579 | 595 | 646 | 657 | 681 | 721 | 771 | 802 | 829 | 930 | 1054 | 1096 |
| Consensus | G   | G   | G   | G   | C   | G   | A   | C   | T   | C   | C   | G   | T   | C   | T   | C   | C   | G   | G   | G   | G   | G   | G   | G   | G   | G   | G   | G   |
| \(O^1\) alleles |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |
| \(O^1-1 [O01]\) |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |
| \(O^1-2 [O02]\) |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |
| \(O^1-3 [O03]\) |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |
| \(O^1-4\) |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |
| \(O^1-5\) |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |
| \(O^1-6 [O09]\) |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |
| \(O^1-7 [O10]\) |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |
| \(O^1-8\) |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |
| \(O^1\) variant (\(O^1v\) alleles) |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |
| \(O^1v-1 [O02]\) |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |
| \(O^1v-2\) |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |
| \(O^1v-3\) |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |
| \(O^1v-4\) |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |
| \(O^1v-5\) |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |
| \(O^1v-6\) |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |
| \(O^1v-7\) |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |
| \(O^1v-8\) |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |
| Other \(O\) alleles |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |
| \(O^2\) [O03]\* |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |
| Amino acid | 36  | 63  | 63  | 64  | 74  | 87  | 99  | 106 | 152 | 156 | 176 | 181 | 193 | 199 | 216 | 219 | 227 | 241 | 257 | 268 | 277 | 310 | 352 |
| Consensus | Val | Arg | Arg | Val | Pro | Val | Thr | Asn | Phe | Pro | Arg | Trp | Ser | Arg | Phe | His | Pro | Arg | Pro | Gly | Val | Leu | Arg |
| Change | Phe | His | Ser |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |

* Additionally, two more polymorphic sites were shown in exon 2 (53G>T) and exon 5 (220 C>T).
Genetic background of hybrid alleles, rare O alleles and weak subgroups

Hybrid alleles

From 1996 and onwards, it became increasingly clear that O alleles with the 261delG mutation could also include A2-or B-related SNPs. Mixtures of O′ and O″ sequences were also observed. The implications and consequences for ABO genotyping were serious (Olsson & Chester 2001). Some important O hybrids have been summarized in Table 6.

Table 6. Hybrid alleles reported before the started of this study. Nucleotide changes are compared with the consensus allele.

<table>
<thead>
<tr>
<th>Exon</th>
<th>2</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>nt. position</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>Ref.</td>
</tr>
<tr>
<td>Consensus</td>
<td>G</td>
<td>G</td>
<td>C</td>
<td>C</td>
<td>C</td>
<td>G</td>
</tr>
<tr>
<td>Hybrid alleles</td>
<td>0″-B</td>
<td>T</td>
<td>A</td>
<td>T</td>
<td>delG</td>
<td>G</td>
</tr>
<tr>
<td>0″-2[O′-O″]</td>
<td>delG</td>
<td>G</td>
<td>A</td>
<td>T</td>
<td>A</td>
<td>[1]</td>
</tr>
<tr>
<td>0″-3[O′-O″]</td>
<td>delG</td>
<td>G</td>
<td>A</td>
<td>A</td>
<td>T</td>
<td>A</td>
</tr>
<tr>
<td>0″-4[O′-O″]</td>
<td>delG</td>
<td>G</td>
<td>A</td>
<td>A</td>
<td>T</td>
<td>A</td>
</tr>
<tr>
<td>0″-5[O′-O″-O′]</td>
<td>delG</td>
<td>G</td>
<td>A</td>
<td>T</td>
<td>A</td>
<td>[1]</td>
</tr>
<tr>
<td>0″-6[O′-O′]</td>
<td>delG</td>
<td>T</td>
<td>aeC</td>
<td>[1] &amp; Gassner et al., 1996</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0″-7[O′]</td>
<td>delG</td>
<td>G</td>
<td>[2] &amp; Ogasawara et al., 1996</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

| Amino acid | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 | 21 | 22 | 23 | 24 | 25 | 26 | 27 | 28 | 29 | 30 | 31 |
| Consensus | V | R | R | P | V | T | P | R | P | H | P | G | P | L | G | V | L | P |  |
| Change | F | H | S | L | G | T | S | M | A | M |

Rare O alleles

O′ allele

A few other rare O alleles have subsequently been found in single samples, such as O′ [O08] that does not have 261delG but instead contains both the common A2 allele polymorphisms 467C>T and C-deletion at nt. 1059-1061 and an insertion of an extra guanosine in the 7-guanosine sequence at nt. 798-804 (Olsson & Chester, 1996c). This insertion is also responsible for A′-1 [Ael01, A109] (see later).
Other rare O alleles

Two other rare O alleles lacking 261delG were reported in the Japanese population by Ogasawara O301 [O14] has the missense mutation 893C>T (Ala268Val) on an A102 background whereas O302 [O15] has the nonsense mutation 927C>A (Tyr309Stop) on an A101 background (Ogasawara et al., 2001).

Weak subgroups

In addition to the major phenotypes characterised by either strong or absent haemagglutination with anti-A/-A1 and -B reagents, the ABO blood group system also includes phenotypes in which erythrocytes react weakly with the anti-A and -B reagents, for example A3, A*, A dist, A
d, B
d, B*, B
d, c
c-AB and B(A) (Daniels, 2002).

In some cases of weak phenotypes, it is difficult to determine clearly their specific ABO subgroup by conventional serological methods. Indeed this is sometimes a question of interpretation between laboratories and even between individuals in the same laboratory. The weak subgroups are important in more than one way. First, they risk complicating patient and donor blood group determination. At worst, the wrong group can be assigned if e.g. a weak antigen is missed. Second, they allow us to characterize and understand glycosyltransferase mechanisms by studying the results of mutations in the underlying alleles. Any defect in the ABO gene complicates ABO biosynthesis and the transcription and translation of the gene products (e.g. GTA and GTB) can fail. In addition, any factor including substrate or acceptor availability or localisation of the GTA/GTB could be the limiting factor, at least in theory. In general, SNPs in exon 7 have turned out to be significant in generating weaker phenotypes.

A3 alleles

A typical serological pattern illustrates that RBCs from individuals with the A3 phenotype agglutinate strongly with anti-A and anti-A,B but show a large number of free cells. Investigation of enzyme activity in sera of A3 individuals have shown two different results. In one study, normal A1 enzyme activity was reported and in another much lower than A1 (Cartron et al., 1978; Nakamura et al., 1989). One A3B individual had a novel point mutation, 871G>A (Asp291Asn), on the A1 [A101] background (Yamamoto et al., 1993a) and this allele was named A301. Two research groups (Olsson & Chester, 1996b; Barjas-Castro et al., 1997)
reported separately the absence of this missense mutation in seven Swedish and 11 Brazilian serologically defined A₃ samples, thus finding only consensus A¹ sequences in exons 6 and 7. The latter research group instead presented A302 which has two single point mutations, the substitution 829G>A (Val277Met) and 1061delC. In fact, A₃ appears to be more complex compared to other subgroups and the true allelic background is not yet solved.

Aᵦ alleles
These alleles are responsible for the rare subgroup Aᵦ (Fischer & Hahn, 1935) and the RBCs typically show a positive reaction with anti-A,B and anti-H but notably weaker (if any) reactions with anti-A. There are four base substitutions involved in these alleles; 646T>A (Phe216Ile), 681G>A (silent), 771C>T (silent), 829G>A (Val277Met). Aᵦ-1 [Ax01, A108] has the missense mutation 646T>A on an A¹ background (Yamamoto 1993b, Olsson & Chester 1998a). Combination of 646T>A and 681G>A on the A¹ background was demonstrated and designated as Ax04 [A113] (Ogasawara et al., 2001). Olsson & Chester presented additional alleles associated with the Aᵦ phenotype, namely Aᵦ-2 [Ax03] and Aᵦ-03 [Ax02] which contained all four defining substitutions but Aᵦ-03 had a 297A>G mutation (silent) in addition. These authors concluded that Aᵦ-2 and -3 alleles are in fact hybrid alleles between common ABO alleles (A and O1v or B/O2 and O1v) differing by their crossing-over regions in intron 6 (Olsson & Chester, 1998a).

Aₑ¹ alleles
Forward typing in this subgroup does not show any agglutination by anti-A or anti-A,B but RBCs react strongly with anti-H. The reverse typing shows that plasma lacks antibodies against A₁ and A₂ RBCs. The key analysis is a positive eluate after adsorption with human polyclonal anti-A or anti-A,B. Aₑ¹-1 [Ael01] has a single G insertion in the seven-guanosine sequence at nt. 798 to 804. This novel insertion results in a frame-shift that alters the a.a. sequences after the glycine at position 268 which is located immediately at the nucleotide-sugar binding site of the enzyme (Olsson et al., 1995a). The encoded transferase is expected to be 37 a.a. longer than the normal consensus allele, and 16 a.a. longer than A²-encoded transferase. It is surprising but very consistent in the literature that this anomalous transferase has the ability to synthesize A antigens at all (Ogasawara et al., 1996a). Experiments with gold-labelled monoclonal anti-A showed that approximately 1% of all Aₑ¹ RBC express A antigen strongly whereas the rest of the cells remain almost unlabelled (Hansen et al., 1998).
Heier’s group had hypothesized that similar findings in other subgroups could be based on recombination tendency and that eventually somatic cell clones in or outside the haematopoietic cells may result in small amounts of cells with another phenotype in the same individual, thus causing a pseudo-chimeric pattern (Heier et al. 1994).

One other allelic member in this subgroup is $A^{el}$-2 [Ael02] that contains two SNPs, 646T>A and 681G>A on the A102 background with two a.a. substitutions, Pro156Leu and Phe216Ile compared to consensus (Ogasawara et al., 1996a). This is actually very similar to an $A^e$ allele. It is probably the a.a. substitution Phe216Ile in combination with the otherwise neutral Pro156Leu that dramatically decreases the enzymatic activity of the GTA.

Other $A^{weak}$ alleles

The serological characteristics of some other minor phenotypes included among A subgroups, e.g. Aend, Afin, Ay, Afinn, Aantu and the collective description Aw are more unclear. Guiding principles exist (Table 7), but any pattern of agglutination is due to several factors, e.g. the condition of the RBCs, the antibodies used to test them and the experience of the analyst. Thus, it can be discussed if all samples categorized and reported as Aend e.g. can be differentiated from all other A subgroups.

**Table 7.** Serological reaction patterns adopted from current text books. A negative reaction is noted by 0 and positive reactions are denoted from + (very weak agglutination) to 4+ (maximal agglutination).

<table>
<thead>
<tr>
<th>Subgroup of A</th>
<th>RBC reactions with ABO substances in saliva</th>
<th>Anti-A</th>
<th>Anti-A,B</th>
<th>Anti-A1</th>
<th>Anti-H</th>
<th>Anti-A$_i$ in serum</th>
</tr>
</thead>
<tbody>
<tr>
<td>A1</td>
<td>4+</td>
<td>4+</td>
<td>4+</td>
<td>0</td>
<td>A, H</td>
<td>No</td>
</tr>
<tr>
<td>A2</td>
<td>4+</td>
<td>4+</td>
<td>0</td>
<td>4+</td>
<td>A, H</td>
<td>Sometimes</td>
</tr>
<tr>
<td>A3</td>
<td>4+</td>
<td>4+</td>
<td>2+/3+</td>
<td>2+/3+</td>
<td>A, H</td>
<td>No</td>
</tr>
<tr>
<td>A4</td>
<td>2+/+mf</td>
<td>2+/+mf</td>
<td>0</td>
<td>4+</td>
<td>A, H</td>
<td>No</td>
</tr>
<tr>
<td>A5</td>
<td>0/+</td>
<td>2+/+</td>
<td>0</td>
<td>4+</td>
<td>H</td>
<td>Often</td>
</tr>
<tr>
<td>A6</td>
<td>0§</td>
<td>0§</td>
<td>0</td>
<td>4+</td>
<td>H</td>
<td>Sometimes</td>
</tr>
<tr>
<td>A7</td>
<td>+</td>
<td>+</td>
<td>0</td>
<td>4+</td>
<td>H</td>
<td>Sometimes</td>
</tr>
<tr>
<td>Afin</td>
<td>+</td>
<td>+</td>
<td>0</td>
<td>4+</td>
<td>H</td>
<td>Yes</td>
</tr>
<tr>
<td>Afinn</td>
<td>+(+)</td>
<td>+(+)</td>
<td>0</td>
<td>4+</td>
<td>H</td>
<td>Yes</td>
</tr>
<tr>
<td>Am</td>
<td>0/+</td>
<td>0/+</td>
<td>0</td>
<td>4+</td>
<td>A, H</td>
<td>No</td>
</tr>
<tr>
<td>B</td>
<td>0§</td>
<td>0§</td>
<td>0</td>
<td>4+</td>
<td>A, H</td>
<td>No</td>
</tr>
</tbody>
</table>

§ anti-A can adsorbed and eluted from these cells despite absence of agglutination.

mf mixed field agglutination

$B^{weak}$ alleles
The above uncertainty is particularly true about B\textsubscript{weak} phenotypes which are especially difficult to classify. B variants are much more uncommon than A variants, while this may just reflect the relatively low frequency of the B blood group in many populations. Accordingly, most B subgroups so far have been characterized in Asian populations.

The molecular basis underlying many of these B subgroups has been studied and some B\textsuperscript{weak} alleles elucidated so far. Guidelines exist and are summarized in Table 8 (Daniels, 2002). Typically, these phenotypes often appear to result from missense mutations at the ABO locus causing single a.a. changes in the GTB.

Table 8. Characteristics of some more frequent, B\textsubscript{weak} phenotypes

<table>
<thead>
<tr>
<th>Subgroup of B</th>
<th>RBC reactions with ABO substance in saliva</th>
<th>Anti-B in serum</th>
<th>GTB in serum</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Anti-A</td>
<td>Anti-B</td>
<td>Anti-A,B</td>
</tr>
<tr>
<td>B</td>
<td>0</td>
<td>++++</td>
<td>+++</td>
</tr>
<tr>
<td>B\textsubscript{3}</td>
<td>0</td>
<td>mf</td>
<td>mf</td>
</tr>
<tr>
<td>B\textsubscript{x}</td>
<td>0</td>
<td>w</td>
<td>w</td>
</tr>
<tr>
<td>B\textsubscript{m}</td>
<td>0</td>
<td>*/w</td>
<td>0/w</td>
</tr>
<tr>
<td>Bel</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

mf, mixed field; w, very weak agglutination
*Anti-B may be adsorbed onto and eluted from these cells
(B\textsubscript{x}), may require inhibition of agglutination of B, cells for detection

The B\textsubscript{3} phenotype was proposed to depend on the missense mutation 1054C>T (Arg352Trp) on a B\textsubscript{1} [ B101] background and named B301 (Yamamoto \textit{et al.}, 1993a). A B\textsuperscript{1} allele [Bx01 or B104] responsible for the B\textsubscript{x} phenotype in a Japanese sample had a point mutation at nt. 871G>A (Asp291Asn) (Ogasawara \textit{et al.}, 1996a) which was also found in an A\textsubscript{3} sample (Yamamoto \textit{et al.}, 1993a).

The B\textsubscript{m} phenotype was divided into two suballeles, B\textsubscript{m}-1 and B\textsubscript{m}-2 [Bel01 and Bel02], which had substitutions at 641T>G (Met214Arg) and 669G>T (Glu223Asp), respectively (Ogasawara \textit{et al.}, 1996a). Presumably, these missense mutations reduce the enzymatic activities of the GTB. Ogasawara \textit{et al.} also found another B\textsuperscript{1} allele, B302, which differs from B consensus, by two nucleotide substitutions, 646T>A (Phe216Ile) and 657T>C (same nucleotide as A consensus). Thus, again the same mutation found in A\textsubscript{3} also appears to be responsible for a B\textsubscript{3} phenotype. It can be speculated if this allele was really a hybrid, B-O\textsuperscript{IV}-B
CisAB alleles

H Seyfried described this rare phenotype first in 1964 (Seyfried et al., 1964). It represents a very interesting phenomenon that proved that it is possible for a child with blood group O to have a parent with blood group AB. Seyfried hypothesized that both A and B determinants of the AB blood group could be located on the same chromosome. Later on, this idea confirmed in that the existence of an exceptional ABO allele encoding a glycosyltransferase is indeed able to produce both A and B enzymes at the same time. This rare blood group is characterized by the presence of A at levels comparable to the A2 phenotype, weakened B and prominent H antigen expression on RBCs. An anti-B reactive only at room temperature and lower is often present on reverse typing of cis-AB individuals and most also has a cold-reacting anti-A; (Pacuszka et al., 1975; Daniels, 2002). Cis-AB01 was sequenced and showed the substitution 803G>C (Gly268Ala) on the A1-2 [A102] background and thus can be described as AAAB (Yamamoto et al., 1993c). Another allele named cis-AB02 was discovered when a Vietnamese man who was to undergo organ transplantation showed irregular blood grouping results. The sequencing showed that his ABO genes was nearly identical to the normal B allele except for a 796A>C (Met266Leu) substitution. This enzyme can also be designated BBAB (Mifsud et al., 2000).

B(A) alleles

As the name suggests, this phenomenon was found in individuals who express normal B antigens on RBCs with anti-A in serum but whose RBCs also react with some potent monoclonal anti-A reagents (Goldstein et al., 1989). The first reported B(A) allele (Yamamoto et al., 1993b) showed that the nucleotide constellation in exon 7 was BABB, [noted in bold face: 526C>G (Arg176Gly), 703G>A (Gly235Ser), 796G>A (Leu266Met), 803G>C (Gly268Ala)]. This allele was designated B(A)01. Another B(A) allele named B(A)02, had a composition of BBBB and further showed a novel nonsynonymous substitution at 700C>G which produces an a.a. substitution, Pro234Ala. It is important to note that this substitution occurs next to the second (residue 235) of the four A- vs. B-characteristic a.a. residues just mentioned above. Accordingly, this reduced the GTB activity and enhanced the normally small amounts of GTA activity that are present in serum (Yu et al., 1999).
AIMS OF THE THESIS

The aim of this thesis was to continue earlier work at this department on the elucidation and description of the diversity of ABO alleles. Furthermore, the purpose was also to contribute to a deeper knowledge about the relationship between the structure and function of ABO subgroups.

More specifically the aims were:
- To elucidate new alleles in the ABO blood group gene in samples either with suspected A and B subgroups or other blood grouping discrepancies
- To assess the impact of polymorphisms in the ABO gene in some weak subgroup phenotypes by 3D-modelling, enzyme kinetics and crystal structure resolution of the mutant glycosyltransferase
- To develop and evaluate a novel ABO genotyping method for safer blood grouping in clinical practice

MATERIAL AND METHODS

Blood and DNA samples

Blood samples with common ABO phenotypes, undefined or rare phenotypes in Papers I-VI were obtained either from apparently healthy blood donors or in some cases patients whose samples had been referred to the Nordic Reference Laboratory for Genetic Blood Group Typing, and were from different geographic/ethnic backgrounds. This included (but was not limited to) Swedish, Swiss, South African, Brazilian, Chinese, Israeli, English, Turkish and Jordanian samples. The blood was obtained in tubes containing ethylenediaminetetraacetic acid (EDTA) or acid-citrate dextrose (ACD) as anticoagulants. DNA was isolated using a salting-out procedure described by Miller et al., 1988.

ABO genotyping

Polymerase chain reaction-restriction fragment length polymorphism (PCR-RFLP)

A PCR-RFLP method was used for the initial ABO genotype screening. The principle for this method is based on using the PCR technique to amplify exons 6 and 7. The amplified fragments cover the two major exons and constitute 91% of the catalytic domain. Subsequent digestion of the fragments simultaneously using the restriction enzymes HpaII and KpnI
results in a specific fragment pattern for the common ABO alleles and thus interpretation of at least 15 different ABO genotype can be obtained (Olsson & Chester, 1995). The limitation in this method is that it can only detect mutations at these specific points (restriction enzyme sites) in exon 6 and 7. Accordingly, the major ABO alleles (A1, A2, B, O1/O1v and O2) can be identified as well as a few subgroup alleles. As with other ABO genotyping methods, other rare or novel mutations would be missed.

**Allele-specific primer PCR (PCR-ASP)**

PCR was used throughout (Papers I-VI) both to amplify gene-specific fragments (regardless of allele) and allele-specific fragments (PCR-ASP). A method linking mutations in exons 6 and 7 by PCR with allele-specific primers (ASP) across intron 6 was used when exclusion of recombinant hybrid alleles was required (first mentioned in Paper I but mainly used in Paper VI). Oligonucleotide primers used in this thesis were synthesized by DNA Technology APS. The primers and conditions used for PCR are described in the respective articles and more extensively in Paper VI.

**Direct DNA sequencing**

In all studies PCR was used to obtain sufficient material for analysis either by gene-specific or allele-specific primers (ASP). The former primers allow amplification of all known variants of the particular gene being examined. The latter approach only allows amplification of segments having a polymorphism specific for the primer. After electrophoretic separation of the reaction products on agarose gels these were visualised by staining with ethidium bromide, excised from the gel and purified using the Qiaquick gel extraction kit (for details see Papers I-VI). The Big Dye Terminator Cycle Sequencing kit and an ABI PRISM 310 or 3130 Genetic Analyser were used for direct DNA sequencing by capillary electrophoresis and automated fluorescence-based detection according to the manufacturer’s instructions. Sequence analysis was performed with SeqEd software 1.03.

**Reverse transcriptase PCR (RT-PCR)**

In Paper IV, RNA was prepared from peripheral blood samples using both TRizol® LS Reagent and RNeasy kit. Complementary DNA (cDNA) was synthesized using random hexamers, random octamers or oligo-d(T) primers from the GeneAmp RNA PCR kit according to the manufacturer’s instructions. Two rounds of PCR amplification, *i.e.* nested
PCR, were performed improving both the specificity and the amount of required fragments for further analysis.

Site-directed mutagenesis and cloning
A Quick Change II XL Site-directed Mutagenesis Kit (Stratagene) was used to generate the required mutant enzymes. The plasmid pCWΔlac harbouring the -10GTB gene was used as a template. This sequence for -10GTB lacks the N-terminal transmembrane domain-encoding region and encodes a GTB of residues 63-354 containing the catalytic domain. The PCR products generated (Paper V) were treated with restriction enzyme DpnI to digest template methylated, non-mutated plasmid DNA. The mutated DpnI-treated fragments were cloned and transformed into E. coli XL10-Gold cells. Conformation of mutagenesis was done by sequencing as described above. Positive mutants were transformed into E. coli BL21 Gold cells for expression. These experiments were performed at the Carlsberg Laboratory, Copenhagen, Denmark.

Protein purification and enzyme kinetics
The recombinant enzyme was purified using a two-step protocol comprising ion-exchange chromatography followed by UDP-hexanolamine affinity column chromatography as has been described elsewhere (Seto et al., 1999). Enzyme kinetic studies were performed using a radiochemical assay (Palcic et al., 1988). In this assay a Sep-Pak reverse-phase cartridge is used to isolate radiolabelled reaction products that are produced when label is transferred from a radioactive donor to a hydrophobic acceptor (Palcic et al., 1988). The kinetic parameters k_cat and K_m were obtained by non-linear regression analysis of the Michaelis Menten equation with the GraphPad PRISM 3.0 program (GraphPad Software). A unit of enzyme is the amount that converts one micromole of substrate to product in one minute. Protein concentrations were estimated using the Bradford method using bovine gamma globulin as a protein standard.

Computer modelling and crystallization
The GTB mutant structures were crystallized by collaborators in Canada using conditions similar to the native GTB enzymes as described previously (Patenaude et al., 2002; Breton et al., 2006). Data were collected on a detector (Pflugrath, 1999) and X-rays were produced and subsequently the crystals were frozen. All structures were solved by using molecular
replacement techniques with wildtype GTB (Protein Data Bank accession code 1LZ7) as a starting model and were refined using the CCP4 program suite (The CCP4 program suite, 1994).

RESULTS AND DISCUSSION

The molecular genetic background of the common and a few of the variant phenotypes at the ABO blood group locus was described by Yamamoto and his colleagues in the 1990s. Until 2001, this studies was started, several research groups had reported new findings, and the number of ABO alleles registered in dbRBC* had increased to more than 70 (Chester & Olsson, 2001). However, the molecular basis for a number of variant phenotypes in the ABO blood group system had not been yet defined.

**Single point mutations associated with A or B subgroups (Papers I and V)**

To date, we have investigated the ABO gene in numerous samples of clinical or research purposes. For the studies presented in Papers I and V approximately 400 samples were included. These collected data constitute one of the most extensive sets of clinical samples of its sort and have been referred to our laboratory during the last decade. In most cases, serological discrepancies were the reasons for the referral to the Nordic Reference Laboratory for Genetic Blood Group Typing. On the basis of medical history, phenotype condition and even initial genotyping results (for some more recent samples from the referring centres), all 324 samples investigated in Paper I were categorized in different subdivisions, e.g. ABO discrepancies due to weak A or B expression in pregnancy or haemological disorders, acquired B syndrome or acquired A-like antigen, suspected chimeras or weak A/B expression without known reason. DNA was prepared and genotype determined according to published methods (Olsson & Chester, 1995; Olsson & Chester, 1996b; Hosseini-Maaf et al, 1998). Even if Paper I was a rather clinically oriented article, the main focus in this thesis is on alleles and mutations associated with ABO subgroups. The 1st and 2nd authors are both MDs who were responsible for the clinical interpretations and consequences of our findings. All samples were found in the initial ABO genotype screening by PCR-RFLP and PCR-ASP to be heterozygous for at least one A- or B-like allele. Similarly, all samples were examined at

the CBF-NF/Y enhancer regions approximately 4 kbp upstream from the translation start codon, and the PCR results were as expected without any deviation (Irshaid et al., 1999).

In total, the phenotype and genotype mis-correlation led us to sequence 42 samples (20 with suspected A subgroups and 22 with suspected B subgroups) for exons 1-7 of the ABO gene and two proposed regulatory motifs (Kominato et al, 1997). All samples were homozygous for the consensus sequence in exons 1-5, except those containing an O^v allele or A^v-3 (see below). As expected, O^v alleles were polymorphic at nt. 106 (exon 3), 188/189 (exon 4) and 220 (exon 5) (Olsson & Chester, 1996a) and the O^v alleles contained polymorphic sites at nt. 53 (exon 2) and 220 (exon 5) (Amado et al., 2000).

The summarized sequencing results of seventeen novel A and B subgroup alleles with single point mutations are shown in Table 9.

Table 9. The seventeen novel A and B subgroup alleles reported in Papers I and V. Nucleotide and a.a. changes are given in comparison to the common A or B alleles.

<table>
<thead>
<tr>
<th>Allele description</th>
<th>Nucleotide change(s)</th>
<th>Amino acid change(s)</th>
<th>Pos/ Neg</th>
<th>Invariant</th>
<th>Conserved</th>
<th>Geographic origin</th>
<th>Accession no.</th>
</tr>
</thead>
<tbody>
<tr>
<td>A^v-1</td>
<td>407C&gt;T</td>
<td>Thr136Met</td>
<td>X</td>
<td>United Kingdom</td>
<td>AF321010</td>
<td></td>
<td></td>
</tr>
<tr>
<td>A^v-2</td>
<td>350G&gt;C</td>
<td>Gly117Ala</td>
<td>X</td>
<td>Various**</td>
<td>AF324009</td>
<td></td>
<td></td>
</tr>
<tr>
<td>A^v-3</td>
<td>203G&gt;C</td>
<td>Arg68Thr</td>
<td>Pos</td>
<td>Scandinavia</td>
<td>AF324008</td>
<td></td>
<td></td>
</tr>
<tr>
<td>A^v-4</td>
<td>721C&gt;T</td>
<td>Arg241Trp</td>
<td>Pos</td>
<td>Belgium/ Germany</td>
<td>AF324011</td>
<td></td>
<td></td>
</tr>
<tr>
<td>A^v-5</td>
<td>965A&gt;G</td>
<td>Gln322Gly</td>
<td>Neg</td>
<td>Finland</td>
<td>AF324012</td>
<td></td>
<td></td>
</tr>
<tr>
<td>A^v-6</td>
<td>996G&gt;A</td>
<td>Trp332stop</td>
<td>X</td>
<td>New Zealand</td>
<td>AF324013</td>
<td></td>
<td></td>
</tr>
<tr>
<td>B^v-2</td>
<td>873C&gt;G</td>
<td>Asp291Glu</td>
<td>Neg</td>
<td>France</td>
<td>AF324018</td>
<td></td>
<td></td>
</tr>
<tr>
<td>B^v-3</td>
<td>721C&gt;T</td>
<td>Arg241Trp</td>
<td>Pos</td>
<td>Sweden</td>
<td>AF324014</td>
<td></td>
<td></td>
</tr>
<tr>
<td>B^v-4</td>
<td>548A&gt;G</td>
<td>Asp183Gly</td>
<td></td>
<td>Sweden</td>
<td>AF324016</td>
<td></td>
<td></td>
</tr>
<tr>
<td>B^v-5</td>
<td>539G&gt;A</td>
<td>Arg180His</td>
<td>Pos</td>
<td>USA</td>
<td>AF324015</td>
<td></td>
<td></td>
</tr>
<tr>
<td>B^v-6</td>
<td>1036A&gt;G</td>
<td>Lys346Glu</td>
<td>Pos</td>
<td>Finland</td>
<td>AF324019</td>
<td></td>
<td></td>
</tr>
<tr>
<td>B^v-7</td>
<td>1055G&gt;A</td>
<td>Arg352Gln</td>
<td>Pos</td>
<td>USA</td>
<td>AF324020</td>
<td></td>
<td></td>
</tr>
<tr>
<td>B^v-8</td>
<td>863T&gt;G</td>
<td>Met288Arg</td>
<td>Pos</td>
<td>Turkey</td>
<td>AF324017</td>
<td></td>
<td></td>
</tr>
<tr>
<td>B^v-15</td>
<td>565A&gt;G</td>
<td>Met189Val</td>
<td>X</td>
<td>Turkey</td>
<td>DQ646577</td>
<td></td>
<td></td>
</tr>
<tr>
<td>B^v-16</td>
<td>575T&gt;C</td>
<td>Ile192Thr</td>
<td>X</td>
<td>Pakistan</td>
<td>DQ646578</td>
<td></td>
<td></td>
</tr>
<tr>
<td>B^v-17</td>
<td>784G&gt;A</td>
<td>Asp262Asn</td>
<td>Neg</td>
<td>South Africa</td>
<td>DQ646580</td>
<td></td>
<td></td>
</tr>
<tr>
<td>B^v-18</td>
<td>802G&gt;A</td>
<td>Ala268Thr</td>
<td>X</td>
<td>France</td>
<td>DQ646581</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

* Charged residues involved in a.a. changes, Pos = positive and Neg = negative
** New Zealand, England, Australia and USA
† According to published data on nine homologous genes (Gastinel et al., 2001).
All these novel alleles carried single missense mutation except the $A'$-6 allele which carried a nonsense mutation. All mutations except one were located in exon 7 and affect the enzymatically active domain.

The mutation 203G>C provides the first example of a mutation outside exon 6-7 that is associated with phenotype change. It has been later shown by transfections to HeLa cells that this mutation appears to weaken enzyme activity (Seltsam & Blaszyk, 2005). Screening for subgroup mutations found in this study (Paper I) by PCR-ASP in samples of normal ABO phenotypes was negative except for one sample with the additional presence of the 203G>C (Arg68Thr) mutation in an $A'$ allele. The screening undertaken in Paper V was negative.

How can the effects of these mutations be explored?

The obvious question raised was how these mutations affected the enzyme structure and/or activity and furthermore the quality/quantity of antigens on RBC. There are different methods to examine this, e.g. transfection to cell lines, measuring the enzyme activity and kinetics, and using flow cytometry. An early attempt to use molecular modelling was done but not shown by Yamamoto but this was presumably limited by the available software (Yamamoto et al., 1993a). In collaboration with colleagues in Canada, a computed molecular model was used to show the presumed effect of the mutations in the 3D structure of the B glycosyltransferase.

Together with Monica Palcic’s group in Denmark, we purified the enzyme and measured the enzyme kinetics for some of the $B_{weak}$ subgroups (Paper V). Computed modelling to show the location of all these SNPs is illustrated in Figure 5. We investigated two new missense mutations found in $B_{weak}$-17 (Asp262Asn) and $B_{weak}$-19 (Phe216Ile) in Paper V. The calculations of computed modelling and kinetic constants supported each other in these cases and the predictions were good enough to understand the potential effects of SNPs in these weak B blood groups.

We also showed computed modelling of one recently published $B_{weak}$-11 (Leu232Pro) (Deng et al., 2005) in Paper V.

Of special interest was the missense mutation 721C>T (Arg241Trp), a mutation that can lead to both a weak A and B phenotype. This is the first example of the same point mutation in two different alleles affecting the expression of either antigen: another example is the 502C>G mutation which results in both weak A and B antigens (Seltsam et al., 2002; Lin et al. 2003). Later, the 539G>A mutation found here was shown also to cause both weak B and A phenotypes (Svensson et al., 2005).
In one case, we succeeded in determining the crystal structure of a subgroup glycosyltransferase, the Arg268Thr mutant enzyme in B<sup>2</sup>-18 with the 802G>A mutation. Unfortunately, the structure resolution was not clear enough to determine precisely the effect of this mutation but enzyme kinetic data indicated a dramatically decreasing $k_{cat}$ in comparison with the native GTB. This mutation is interesting, however, because the effect on GTA (Gly268Arg) is an $O$ allele ($O^O$) whilst the same mutation in $B$ (Gly268Thr in GTB) caused only moderate weakening of the phenotype.

**Figure 5.** Ribbon diagram of the general B enzyme structure showing the sites of the naturally-occurring mutations in this study. Asterisks indicate those alleles which have been described recently and shown here for the second time. Main chain atoms are shown as a thin white bar with $\beta$-pleated sheets as grey ribbons. The connectivity of the main chain across the 21- and 10-residue-disordered internal or C-terminal loop, respectively, are represented as a black dashed line. UDP-Gal modelled in the active site is shown in the enzyme cleft. The figure was modified from an original kindly provided by JA Letts.
The genetic background for the Bw phenotype appears to be very heterogeneous and usually arises through seemingly random missense mutations throughout the last two exons of the ABO gene. However, the targeted a.a. residues are well conserved during evolution and based on modelling of the resulting structural changes in the glycosyltransferase. The mutations are likely to disrupt molecular bonds of importance for enzymatic function.

Six of the novel mutations, i.e. 548A>G, 539G>A, 565A>G, 575T>C, 1036A>G and 1055G>A were situated on disordered loops. These six mutations are either within the internal disordered loop (a.a. residues 175-195) or the C-terminus (residues 344-355) which may cross the active site and are thought to be involved in catalysis, although it is unknown which a.a. residues in the loops may contact the donor substrate or be involved in proper loop functioning. However, the effect of a.a. mutations in the disordered loops of other species has been studied and it is known that corresponding disordered loops are important for catalysis in other glycosyltransferases of other species (Boix et al., 2001; Unligil et al., 2000; Morera et al., 1999 and Persson et al., 2001). Furthermore, a mutation in this region can directly impact catalytic activity. Other glycosyltransferases from both related and unrelated species also demonstrate the presence of such a disordered loop of amino acids (Yazer & Palcic, 2005).

An A<sub>x</sub> sample from New Zealand was the only sample with a nonsense mutation, 996G>A, predicting loss of 23 a.a. due to a premature stop codon (A<sup>b</sup>-6). Whether this sample is really a classical A<sub>x</sub> phenotype or differs compared to regular A<sub>x</sub> samples due to 646T>A (Phe216Ile) is unclear, but it was noted by the referring centre as a weak A<sub>x</sub>.

This study also showed ABO discrepancies associated with weakened A antigen expression in leukaemia and pregnancy, or acquired B antigens (Table 4 in Paper I). The relationship between leukaemia and weak expression of ABO antigens was first described in 1957 (van Loghem JJ et al, 1957), while the acquired B phenomenon has been described in association with bacterial infections of the gastrointestinal tract (Cameron et al, 19590).

ABO genotyping is an important complement to serology for correct determination of donor and patient ABO blood group status. As an independent tool in the clinical laboratory it can be used to confirm a weakly expressed A or B antigen in inherited subgroups (Olsson & Chester, 1996b) but also e.g. to exclude B allele markers in the acquired B antigen phenotype (Yip et al, 1996). The findings of these studies may help in improving ABO genotyping
methods to detect A and B subgroups and in solving discrepancies and detecting acquired and weak phenotypes.

**Hybrid ABO alleles explain some other weak ABO subgroups (Papers I, II and V)**

Samples from Poland and USA in Paper I were sent to Lund with phenotypic backgrounds that were suspected by the referring centres to be A\_weakB, B(A), respectively. Both samples were shown to have mutations previously characterised as associated with the A\_weak phenotype. The results of sequencing these samples (Table 10 in Paper I) showed two novel variants of the A\_ hybrid allele with the same mutations as in A\_2 [A-O\_2] and A\_3 [B-O\_3], i.e. nt. 646T>A (Phe216Ile), 681G>A, 771C>T, 829G>A (Val277Met) in exon 7. In addition, A\_3 contained a 297A>G mutation in exon 6 (Olsson & Chester, 1998a). Sequencing of intron 6 showed two novel crossing-over regions in these hybrids, designated as A\_4 and A\_5.

<table>
<thead>
<tr>
<th>Allele description</th>
<th>Nucleotide change(s)</th>
<th>Amino acid change(s)</th>
<th>Geographic origin</th>
<th>Accession no.</th>
</tr>
</thead>
<tbody>
<tr>
<td>A_4 (^\dagger)</td>
<td>646T&gt;A</td>
<td>Phe216Ile</td>
<td>Poland</td>
<td>AF324006</td>
</tr>
<tr>
<td></td>
<td>681G&gt;A</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>771C&gt;T</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>829G&gt;A</td>
<td>Val277Met</td>
<td></td>
<td></td>
</tr>
<tr>
<td>A_5 (^\dagger)</td>
<td>646T&gt;A</td>
<td>Phe216Ile</td>
<td>USA</td>
<td>AF324007</td>
</tr>
<tr>
<td></td>
<td>681G&gt;A</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>771C&gt;T</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>829G&gt;A</td>
<td>Val277Met</td>
<td></td>
<td></td>
</tr>
<tr>
<td>B_19</td>
<td>646T&gt;A</td>
<td>Phe216Ile</td>
<td>China</td>
<td>DQ646582</td>
</tr>
<tr>
<td></td>
<td>657T&gt;C</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>681G&gt;A</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

\(^\dagger\) These alleles differ by their intron 6 polymorphisms showing different crossing-over regions between A and O\_1, compared to the previously reported hybrids, namely, nt. 189-225 in intron 6 for A\_2, nt. 236-445 in intron 6 for A\_3, and between nt. 298 in exon 6 and nt. 41 in intron 6 for A\_5.

In Paper II, blood samples from five individuals with different geographic/ethnic background were analysed. These individuals were all confirmed to be of Black ethnic origin. Each donor had the unusual A\_weakB phenotype but appeared to have the A\_WeakB genotype without any new mutations in exons 6 or 7 associated with weak A or B expression.
A *BstUI*-based PCR-RFLP test (Olsson & Chester, 1996a) results in cleavage between 188C and 189G in exon 4 of all known alleles except *O*\textsuperscript{iv} (and some minor variants of *O*\textsuperscript{i} and *O*\textsuperscript{iv}), which has two mutations at these positions (188G>A and 189C>T). The fact that these five individuals already genotyped as *A*B were also heterozygous for an *O*\textsuperscript{iv}-allele-specific mutation led us to sequence all exons to find out the reason for this discrepancy.

All samples showed normal sequences in exon 1-5 as expected, apart from for the *O*\textsuperscript{iv}-related 106G>T (exon 3), 188G>A (exon 4) and 220C>T (exon 5). The additional presence of a novel missense mutation in nt. 46G>A (Ala16Thr) in exon 2 was noted, whereas 189C>T was absent, as were all *O*\textsuperscript{iv}-related mutations in exon 6 and 7 (261delG, 297A>G, 646T>A, 681G>A, 771C>T and 829G>A). Based on these results, we suspected a possible hybrid allele formed somewhere between the end of exon 5 and the beginning of exon 6. The lack of useful polymorphic markers in the *ABO* gene in this region made us sequence introns 2, 3, 4 and 5 in all major alleles *A*\textsuperscript{1}[A101], *A*\textsuperscript{2}[A201], *B* [B101], *O*\textsuperscript{1}[O\textsuperscript{1}-1, O01], *O*\textsuperscript{iv}[O02] and *O*\textsuperscript{2}[O03]) and the results revealed 7, 12, 17 and 8 polymorphic positions, respectively. Intron 6 had been analysed previously (Olsson & Chester, 1998a). Intron 1 was not examined due to its huge size (approximately 13,000 bp). At the same time, Seltsam and colleagues published findings confirming our intron data (Seltsam et al., 2003). Polymorphic sites have been identified in five of the noncoding regions (introns 2-6). Most of these were located in introns 4 and 6.

The polymorphic nt. positions are very valuable genetic markers in the *ABO* gene and allowed us to identify a cross-over event in intron 5 where the *O*\textsuperscript{iv}-related sequence was fused with an *A*\textsuperscript{2} allele to form a hybrid allele (Figure 6). This novel *O*\textsuperscript{iv}-*A*\textsuperscript{2} hybrid was identified according to *ABO* exon/intron analysis in all the five alleles studied with the AwB blood group phenotype.

Screening of more than 100 samples (Paper II) for the polymorphic positions 46G>A or 188G>A (without 189C>T) resulted in the finding of two new *O* alleles, one *O*\textsuperscript{iv} allele having 188G>A but lacking 189C>T and one otherwise normal *O*\textsuperscript{i} allele having 188G>A only. Later, an *O*\textsuperscript{iv}-like allele with the addition of 46G>A in exon 2 was identified by our laboratory in a Dutch donor of African descent during the ISBT 2006 workshop on blood group genotyping (Daniels et al., workshop report in preparation).
In Paper V, samples from a Chinese family submitted to us were subjected to an extensive serological and genetic investigation. In this family, the propositus was phenotyped as A1B3 and his mother was initially typed as group O but was later found to have very weak expression of B antigen. We found in the initial genotyping that the samples were heterozygous for a B-like allele and for either an A1 allele [A101] in the son or an O allele [O01] in the mother. Sequencing of these samples (shown in Figure 7a) revealed that the B-like allele contained a segment of O\textsuperscript{v} [O02]-related polymorphisms in the B-specific DNA fragments, nts. 646, 657 and 681, only the first of which would lead to an a.a. change, Phe216Ile. The new hybrid allele was named Bw\textsuperscript{-19} and is effectively a new hybrid allele with B and O\textsuperscript{v} characteristics (Table 3 in Paper V). It is produced by an apparent double cross-over in exon 7 (Figure 7a).

The structural consequences of the Phe216Ile change due to the 646T>A mutation in this hybrid allele has been demonstrated by molecular modelling. The result showed an empty space in the core of the glycosyltransferase that resulting from the substitution of an a.a. with a smaller side chain (Figure 4a in Paper V). This mutation occurs in a part of the enzyme that is responsible for donor binding, and a mutation that generates space in the hydrophobic core of the protein would likely destabilize the protein fold and result in lower activity. The kinetic data for the Phe216Ile mutant support the molecular modelling and phenotyping results. The crystallization was unsuccessful for this subgroup GTB, which may be due to the destabilizing effect of the mutation.
Figure 7. A schematic representation showing how the B-O1v-B hybrid may have been formed (a). DNA direct sequencing chromatogram showing exon 7 in the ABO gene of a normal B allele compared to the new hybrid allele (b).
The genetic basis of the $A_{bantu}$ phenotype is also a hybrid allele (Paper IV)

About 50 years ago, Shapiro observed the existence of a large proportion of blood group A individuals whose cells did not react with anti-$A_1$ and only weakly with anti-$A$ (Shapiro, 1951). This A subgroup, designated $A_{bantu}$, has its highest frequency in a restricted population but globally seems to be a rare allele. In about 1000 B.C., a bi-directional migration of Bantu people began that created two main language families, Eastern and Western Bantu. The former migrated to Zimbabwe, Mozambique and down into South Africa and the latter migrated into Angola and Namibia. The $A_{bantu}$ blood group has been identified in almost 4% of Black Africans of blood group A and in up to 8% of San and Khoikhoi, the ethnic groups from which the $A_{bantu}$ phenotype may have originated (Jenkins, 1974). In Paper IV, we investigated blood samples from a total of seven individuals serologically characterized as $A_{bantu}$: six from KwaZulu-Natal in South Africa and one from a person of African descent living in England. Genomic $ABO$ typing indicated that all samples were heterozygous for what appeared to be the common $A^2-I$ [A201] allele and for an $O$ allele with the 261delG mutation.

Direct sequencing of exons 1-7 was performed. All samples lacked any of the previously reported mutations associated with weak A and B phenotypes except for the silent 297A>G SNP in exon 6. A deviation found compared to the usual $A^2$ sequence was a single nt. deletion (AGGT$>$AGT) at the exon 4 and intron 4 junction, i.e. a presumed splice-site mutation. However, since it is not clear what the consequences of this mutation would be, different hypotheses concerning the potential effects on the GTA were formed. These are illustrated in Figure 8. As concluded from identical splice-site mutations in other genes, skipping of the preceding exon in the mRNA is predicted. RT-PCR analysis based on synthesis of $ABO$-cDNA from RNA extracts failed in spite of repeated attempts. We could detect only $O$-mRNA but no $A_{bantu}$-mRNA.
Figure 8. Potential consequences of the deletion at the intron 4 / exon 4 junction. Schematic representation of A\(^{-1}\) [A201] cDNA (top row) shows the characteristic sequence associated with the blood group A' allele, i.e. the 467C>T and 1061delC mutations (vertical black lines in exon 7). The latter causes an extension of the consensus (A') sequence by 21 a.a. due to frame shift.

a) Disorder in reading frame (resulting in a premature stop): The deletion results in the exclusion of one of two consecutive guanosines. If the first of the two Gs is considered to be deleted, the mutation is 203delG in exon 4. This would lead to a frame shift and result in a stop codon. This would not allow expression of an enzyme capable of creating even weak A activity but is in effect an O allele.

b) Exclusion of exon 4 from the mRNA: The deletion would act as a splice-site mutation that affect intron 4 rather than exon 4, i.e. deletion of the second guanosine. The consequence of the resulting splicing would be skipping of exon 4. Removal of the 48 nucleotides representing exon 4 does not change the reading frame but simply shorten the stem region of the GTA by 16 a.a.

An unexpected finding in this study was a novel O allele which contains the same nt. deletion (AGGT>AGT) at the exon 4/intron 4 junction and an extra mutation in 297A>G as in the A\(^{bantu}\) allele. The sequence of the A\(^{bantu}\) alleles was identical to this novel O allele sequence in introns 2-4. This sequence differs from the common A\(^{1}\) [A201] by 3, 8 and 11 polymorphic sites in introns 2, 3 and 4, respectively. On the other hand, the sequences of introns 5 and 6 of the A\(^{bantu}\) allele were identical to the common A\(^{1}\) allele but differed from this novel O allele (named O\(^{bantu}\), O54) sequence at 1 and 12 position(s), respectively. This clearly indicated that the A\(^{bantu}\) allele is a hybrid allele with a crossing-over point after nt. 1623 in intron 4 but before nt. 498 in intron 5. Interestingly, the Bantu mutation was found in 9% of O\(^{1}\) alleles in Black South Africans but not in Swedish or Jordanian blood donors, or non-O\(^{1}\) alleles.

In conclusion, a novel allele was found in all A\(^{bantu}\) samples tested, possibly causing decreased mRNA stability and/or changing the coding region. The same mutation in African O' alleles indicates that A\(^{bantu}\) has arisen by gene conversion or crossing over between O\(^{bantu}\) and A\(^{1}\) alleles. A phylogenetic analysis based on approximately 6500 bp of sequence (exons 2-7, including intervening introns) from six common alleles (A\(^{1}\), A\(^{2}\), B, O', O\(^{v}\) and O') indicated that O\(^{bantu}\) is a distinct evolutionary lineage separated from all other lineages but most closely related to O\(^{v}\).
Novel O alleles involved in unexpected blood group phenotypes (Paper III)

The sequence of the most common allele [O1^t-1, O01] is identical to the A^t sequence, apart from one nonsense mutation in exon 6, i.e. 261delG which leads to a truncated protein lacking enzymatic activity (Yamamoto & Hakomori, 1990: Yamamoto et al., 1990a). Various O alleles including hybrid O1^x-B and O1^x-A^t alleles coding for non-functional glycosyltransferases have been described by several groups but the vast majority show the 261delG mutation (for a review, see Chester & Olsson, 2001) only very few non-deletional O alleles had been described by the start of this study: O^t (Olsson & Chester, 1996c) and four Japanese alleles (Ogasawara et al., 2001), two of which are identical to the A^t hybrids discussed above. All these alleles are based on the consensus A sequence. Fundamentally, however any change which leads to the inactivation of the ABO glycosyltransferase can be considered as a lethal null mutation. In this study, we included the investigation of samples referred to our laboratory in Lund because of suspected A subgroups following discrepant serologic blood grouping results but where ABO genotyping or sequencing surprisingly found only O alleles. A correlation between these weak A phenotypes and genotypes by ABO genotyping analysis (PCR-RFLP and PCR-ASP) showed an overrepresentation of the otherwise infrequent O^t allele (Yamamoto et al., 1993d; Grunnet et al., 1994). Thus, we examined 57 existing O^t alleles in detail both from regular group O donors and from the above-mentioned samples with suspected A subgroup phenotypes. A flow chart summarizes the samples that were included in this study (Figure 1, Paper III). All 57 O^t alleles differed from the consensus allele [A101] in exons 2, 5, 6 and 7 at nt. 53, 220, 297, 526, 802 and nt. 1096 located in the 3’-untranslated region as expected (Yamamoto et al., 1993d; Olsson & Chester, 1995; Amado et al., 2000). We confirmed in this study that all O^t alleles contain the mutations 53G>T and 220C>T. This had only been described in few samples in a single paper before (Amado et al., 2000).

Two new O^t alleles were identified, having the 689G>A mutations resulting in Gly229Asp with or without 649C>T (Arg217Cys). These O^t variants were called O^t-2 [O48] and O^t-3 [O49], respectively. There was no difference in the coding sequences of O^t alleles found in normal group O donors or in samples referred to us for ABO discrepancies and suspected A subgroups. Furthermore, the same O^t allele could produce different phenotypes (O or weak A) in different members of the same family.
Yet three novel O alleles surfaced in six samples with suspected A subgroups. Further analysis including sequencing of coding regions showed that these were A-like alleles having nonsense mutations, 88dupG [O'], 322C>T [O'] or 542G>A [O'] causing premature truncation at codons 56, 107 or 181, respectively. All of these alleles would be expected to be totally incapable of producing an active enzyme. Consequently, when present with another inactive allele with 261delG they should produce a normal O phenotype. This is clearly not the case (Table 2 in Paper III). Kominato et al. suggested that perhaps A antigen expression could be the result of the use of an alternative translational start codon (Kominato et al., 2002) or that alternative splicing might explain the effect of early mutations.

In another unusual case in Paper III, we investigated an Australian blood donor who by DNA sequencing, showed homozygosity for a silent mutation (768C>A) and denoted as O26. The same mutation was described in the Basque population and proposed to be a specific Basque marker (Roubinet et al., 2001). Surprisingly our screening study showed that this mutation was quite frequent among Swedish donors and has a frequency of about 9% of all O alleles. We also examined blood samples from two other populations (Jordanian and Swiss) and the results indicate a widespread distribution of this mutation, but at lower frequencies than in Sweden. Thus, this appears to be the third most common O allele among Swedes, following O' and O'' whilst clearly more frequent than O'.

A recently-described O' variant, 488C>T (Thr163Met), denoted Aw08 (Seltsam et al., 2003) was also found in our study and classified as O'-4. Surprisingly, the original and variant O' alleles could all be associated with either O or A_Weak phenotypes. Our data (Paper III) disagree with the idea that 488C>T would result in an a.a. substitution with the ability to restore some enzymatic function as suggested by Seltsam and co-workers (Seltsam et al., 2003). Recent work from Yazer et al. indicated that they could not find any reason (based on sensitive adsorption-elution studies and flow cytometry) not to use blood from group O donors with an O' allele to recipients with anti-A (Yazer et al., 2006). Structural studies have also supported that the crucial mutation 802G>A (Gly268Arg), the most significant mutation in the O' allele, completely blocks the donor sugar recognition site (Lee et al., 2005), although extremely low GTA-like enzyme activities could be measured.

The presence of the A'-characteristic insertion at nt. 804 and the A'-characteristic polymorphisms at nt. 467 and 1061 indicated the presence of an O' allele in a Swedish blood
donor (Olsson & Chester, 1996c). In this study we observed yet another \(O^I\) allele but resulting in a different phenotype (suspected \(A_{\text{weak}}\)) than the original one (clear-cut group O). No other polymorphisms were detected after sequence analysis of exon 1-7.

**How can non-deletional \(O\) alleles result in small amounts of \(A\) antigen?**

Thus, we found that apparent \(O\) alleles lacking 261delG (non-deletional \(O\) alleles) may cause weak A expression on erythrocytes and/or inhibit anti-A production. The \(Chi\) sequence (5'-GCTGGCGG-3') motif has been described in bacteria and been shown to influence the recombination frequency in *Escherichia coli* and has also been implicated as a hot spot for recombination in human genes (Smith, 1994; Stahl & Meyers, 1995). Several investigators have published numerous examples of inheritable, recombinant \(ABO\) alleles generated during meiosis (Suzuki et al., 1997; Olsson & Chester, 1998b; Ogasawara et al., 1996a, 2001). According to the data in this study, \(O\) alleles without 261delG may cause weak A expression on RBCs in the presence of another \(O\) allele with 261delG, which in turn may result in weak anti-A in plasma. The challenge is that it does not always happen when different individuals with the same genotype are compared, even within the same family. To explain this interesting phenomenon, we hypothesized that exchange of genetic material between alleles during mitosis may result in so-called "autologous chimerism" that restores glycosyltransferase activity in some cells. The duplicated \(Chi\) sequences in the 3' end of intron 6 in the \(ABO\) gene may also increase the probability of recombination in somatic cells (during mitosis). Gene conversion/recombination events in somatic cell clones in or outside the bone marrow can then result in "autologous chimerism" which produces low levels of \(A\) antigen and inhibits the natural production of anti-A (*Figure 9*). However, we have been unable to provide experimental evidence to support this hypothesis and alternative explanations may well exist.
Figure 9. Recombination or gene conversion events that may explain how $O$ alleles can result in weak A phenotypes. Probable creation of enzymically active gene products from inactive alleles by recombination (a), gene conversion (b).
Development of an improved ABO genotyping assay based on PCR-ASP amplification across intron 6 (Paper VI)

The majority of individuals possess the standard blood group A, B, and O phenotypes. In addition, a number of A and B subgroups, B(A) and cis-AB exist. Generally, these subgroups are characterised phenotypically by the occurrence and strength of the corresponding A and B antigens on the RBCs and in secretions such as saliva.

The invention of the PCR technique and the elucidation of the sequence (Yamamoto & Kominato, 1990; Yamamoto et al., 1990a-b) and organisation (Yamamoto et al., 1995; Bennett et al., 1995) and the increasing number of alleles (now almost 170 alleles reported to dbRBC) of the ABO gene has opened another dimension that complements serological analysis.

Many alleles responsible for the rarer subgroups have been examined and an increasing number of polymorphisms have been found (Chester & Olsson, 2001; Yamamoto, 2004; Yip, 2002). Knowledge and improvement of ABO genotyping methods are very important for practical implications in transfusion medicine. A number of different screening methods have been developed and used with varying degrees of success (reviewed in Olsson & Chester, 2001). Improvement of these ABO genotyping methods for clinical use is much needed for several reasons.

Recombination and hybrid formation between common alleles in the ABO gene are recent findings that threaten to complicate ABO genotype interpretation. These hybrid alleles may lead to unexpected ABO phenotypes, including rare subgroups. Hybrids are formed by interaction between two alleles, e.g. O’ or O’iv alleles with A2 or B alleles. These phenomena are interesting and complex because they can give rise to O alleles if the nt. 261-containing exon 6 is donated by the O’ or O’iv allele. However, if nt. 261 is consensus (A, B or O2) and combined with an O allele, then there are various consequences depending on the cross-over point. Hybrid alleles obviously exhibit a whole range of serological activities. Different investigators have published genotyping methods without any attention to addressing the problem that can arise due to hybrid alleles (for a detailed description see review in Olsson & Chester, 2001). All ABO genotyping methods to date can fail in the prediction of ABO phenotype in the presence of hybrid alleles or the newly found O alleles and are potentially dangerous and difficult to use in the clinical routine laboratory because of interpretation.
problems (see example in Figure 10). In addition, all ABO genotyping methods published to date will fail in predicting the correct phenotype when A or B subgroup alleles including cisAB and B(A) are present.

We tested two versions of a commercially available ABO genotyping kit and confirmed its limitations in detecting hybrid genes and many subgroup alleles. This kit (2nd version) which is actually CE-marked predicts the common alleles and a few subgroups, but fails to define hybrid alleles and most non-deletional O alleles except O2 and O3, the latter being one of the improvements from the previous version.

**Figure 10.** Two samples with A or B phenotypes would illustrate an identical genotyping pattern in most, if not all current ABO genotyping methods used routinely today. Triangles illustrate the intron 6 with black circles where the Chi sequence may act as a hotspot. The white vertical line shows the 261delG mutation. Filled rectangles show consensus sequence shared between A1 and O1. Open rectangles represent B sequences. Thus, heterozygosity for both 261delG in exon 6 and B-SNPs in exon 7 are present in both samples, a fact that renders genotyping results potentially identical.

Over the past ten years, we have developed and expanded a PCR-ASP method that is flexible and capable of detecting an increasing number of known alleles. This method gives flexibility to different users and can be adjusted according to their requirements for detection of the vast array of ABO alleles (see Paper V1). For example, the suggested approach is flexible so that each user can add or remove primers according to the population tested or as new alleles continue to be described. In analogy with many HLA typing systems, it is also possible to choose the relevant level of resolution (standard or high) depending on the purpose of the analysis.

The goal of this part of the study was to develop and establish an uncomplicated method which can detect different alleles including hybrid, non-deletional null and subgroup alleles. The proposed method is able to detect all major and a large number of rare alleles due to its ability to cis-link the mutations in exons 6 and 7 and thus predict the ABO genotype with high
accuracy, even in the presence of hybrids or subgroup alleles with SNPs in common with the major alleles (e.g. $A^t$-$1$ to $A^t$-$5$ and $O^t$ sharing 646T>A). Most combinations of the following alleles can be discriminated: $A^t$, $A^v$, $A^l$, other $A^{weak}$, $B$, a number of $B^{weak}$ alleles and various null alleles including $O^r$, $O^v$, $O^l$, $O^x$, $O^r$, $O^f$, $O^s$ and also several cis-$AB$ and $B(A)$ variants.

In a first step, all the major alleles are defined and at the same time multiplex PCR-ASP screening for most of the known but unusual or rare hybrid, non-deletional null and A or B subgroup alleles is undertaken in four separate reactions. In most cases, this screening will turn out negative and the genotype can be defined immediately. If one of the four screening mixes are positive, the user has the choice to report the genotype at this level of resolution (e.g. $A^uO^l$, $A^lB^u$, $BO^{hybrid}$, $A^wO^{non-del}$). If, for some reason, a better definition of the genotype is desirable, a second round of high-resolution mixes is performed in which all the multiplexed primers pairs in each screening mix have been separated into a number of reactions. The result will further detail the genotype (e.g. the above-mentioned genotype examples may now instead be expressed as $A^{bantu}O^l$, $A^lB^u$-$18$, $BO^{iv-B}$, $A^2O^s$). We believe that this technique is the most cost-efficient, flexible and informative $ABO$ genotyping method reported to date. Compared to the commercial method which is based on a published PCR-ASP method originally utilizing eight PCR tubes (Gassner et al., 1996) but now upgraded to 16 tubes, the number of genotypes resolved correctly by the method presented here is superior. In principle, no $ABO$ genotyping method should be used unless hybrid alleles and known non-deletional $O$ alleles can be detected properly, especially if the method is used in a clinical reference laboratory and the results are used to predict the blood group of donors or patients.
CONCLUSIONS

The results of the studies presented as part of this Ph.D. thesis have led to a number of conclusions, some of which can be summarized as follows.

- The molecular genetic basis for ABO subgroups is very heterogeneous but single missense mutations targeting evolutionarily conserved a.a. residues is a common theme.
- Hybrids based on common alleles constitute another principal background for both A and B subgroup phenotypes.
- Some A\textsubscript{weak} phenotypes which appear to be characteristic among people of African descent are due to recombinant hybrid formation involving previously unknown O alleles, so far only found in Black individuals.
- Surprisingly, non-deletional O alleles can result in either common O or suspected A\textsubscript{weak} phenotypes but the reason for this has not yet been elucidated.
- Clarification of the relationship between mutation, structural change in the glycosyltransferase and enzymatic function can be achieved and is helpful when trying to understand ABO subgroups as well as the importance of different domains in the enzyme.
- Hybrids, non-deletional null and A or B subgroup alleles constitute major challenges for ABO genotyping.
- Improvement of the current standard of ABO genotyping was made possible by development of a new approach for allele-specific amplification across intron 6 and incorporation of multiplexed primers sets screening for rare or previously difficult-to-detect alleles.

FUTURE PERSPECTIVES

The number of SNPs known at the ABO locus has dramatically increased during the last decade. This has raised a number of new questions about how these allelic changes affect the quantity, structure, function and localisation of GTA and GTB, and subsequently the A and B antigens on RBCs. Future studies will also have to address the mechanisms by which the ABO transcription is regulated, especially in the most relevant compartment, namely the bone marrow. Until now, most work has been carried out in cell lines and most of them of gastric or colonic origin.
New imaging technologies like cellular computer tomography may help us close in on what is actually going on, in a cell biological and biochemical sense, in the Golgi apparatus where the
glycosyltransferases reside. In addition to computer-based 3D predictions and x-ray crystallography as used in this study, further improvements in software and resolution may yield further insight into the molecular enzymology of ABO. Another attractive method could be using electron microscopy, which may facilitate the understanding of what is happening with ABO antigens in different genotypes or in single erythroid progenitors and their successors including the mature RBCs.

Current and future studies will focus on applying these techniques for the characterisation of A\textsubscript{weak} and B\textsubscript{weak} samples in which the mutations are located in other, principally interesting areas of the enzyme, \textit{e.g.} close to the DXD motif (Persson \textit{et al.}, 2007) or in the disordered loops. Also, there are still samples sent to reference laboratories for which the atypical ABO phenotype has not yet been explained at the genetic level. One of the most challenging examples is the most common weak A subgroup, A\textsubscript{3}. Despite the finding of a complete A\textsuperscript{1} consensus sequence, this inheritable variant is completely different from the standard A\textsubscript{1} phenotype. One possibility in this case would be to look at mRNA stability and quantities in combination with analysis of regulatory regions outside the coding region.

Another potential aim in future studies could be to address the hypothesis in \textbf{Paper III} regarding gene conversion or recombination events in somatic cell clones in or outside the bone marrow. Our suggestion about a phenomenon of so-called "autologous chimerism" as well as other theories about how apparent O alleles predicted to result in severely truncated proteins without enzymatic activity, can sometimes give rise to weak A expression require further work. The mechanisms responsible for these apparent anomalies in correlation between serological and genetic blood group typing, is very interesting to us. As yet, we have not provided enough experimental evidence to confirm our theory, so it would be interesting to explore this hypothesis. It may be that other principally different methods are required for this goal. Preliminary work using flow cytometry or erythropoietic culture models have not yet been successful. Cloning of ABO transcripts from cell populations or even single cells may be used to identify low levels recombinant transcripts that can translate into functional GTA.

In general, the ABO blood group system continues to serve as one of the most useful models to study both human genetics, carbohydrate synthesis and glycosyltransferase functional anatomy.
SUMMARY IN SWEDISH (sammanfattning på svenska)

De första blodtransfusionerna kan härledas till 1600-talet då man försökte transfundera blod från djur till djur, djur till människa och människa till människa, vilket ofta resulterade i komplikationer och i många fall patientens död.

I början av 1900-talet upptäckte Karl Landsteiner (1868-1943) blodgruppssystemet ABO (läses A, B, noll) efter empiriska försök där han blandade blod från sig själv och sina kollegor (utanför kroppen).

Transfusionsmedicin är därmed en ung vetenskap inom den medicinska läran. Däremot har ordet blod länge använts i olika sammanhang, inte bara mediciniska, utan även i talspråk såsom t.ex. blodtörstig, blodsbröder, blodsband och blodshämnd etc.

Landsteiner beskrev principen för den antigen-antikroppsreaktion vilken numera benämns hemagglutination och som används rutinmässigt idag vid de transfusionsmedicinska laboratorierna, blodcentralerna.

Antigenen utgörs av kolhydratstrukturer på cellerna medan antikroppar är immunförsvarrets verktyg som kan attackera främmande kroppar som t.ex. bakterier eller blodgivarens blodceller.

Bilden här intill visar de möjliga ABO-blodgrupp-kombinationerna vid transfusion av röda blodkroppar (erytrocyter) och plasma.

Fenotypning innebär att man fastställer vilka antigen som förekommer hos individen i fråga, och tolkas oftast som "finns" (positivt) eller "finns inte" (negativt). Förekomsten av de vanliga ABO-fenotyperna (A₁, A₂, B, A₁B, A₂B och O) varierar liksom andra blodgrupp-fenotyper i olika populationer. Denna variation benämns polymorfism. ABO-systemets antikroppar är oftast av IgM-typ. Dessa bildas normalt (kallas även naturliga/sponata antikroppar eller isoagglutininer) mot de ABO-antigen som saknas hos individen i fråga. ABO-systemet är på grund av dessa naturligt förekommande antikroppar det kliniskt viktigaste blodgruppssystemen. För att utföra en säker blodtransfusion krävs det att ett laboratorium kontrollerar vilken ABO-blodgrupp individen tillhör och om denna har bildat antikropp mot något annat blodgruppssystem. Därefter kan passande givarblod transfunderas. Även vid transplantation av organ som njure, lever, hjärta och lungor är ABO-systemet bland det viktigaste att ta hänsyn till.

I vissa situationer, bland annat vid bestämning av årtligt svagare ABO-uttryck, s.k. subgrupper som A\textsubscript{x} och A\textsubscript{el}, förvärvade blodgruppsegenskaper (t.ex. förvärvat B-antigen, där en individ med blodgrupp A kan framstå som AB) eller vid olika sjukdomar, kan dock resultaten vara svårtolkade och framstår som svagt positiva ("finns men svagare än normalt"). Detta kan bero på någon årtlig variant av ABO-genen eller förändrat uttryck i olika typer av tumörceller och vid vissa svåra infektioner som sannolikt beror på nedsatt aktivitet i enzymet som tillverkas med ABO-genen som mall (se nedan).


Olika mutationer, d.v.s. årtliga och bestående förändringar i cellens genetiska material (DNA), har upptäckts i ABO-genen vilket i sin tur kan leda till minskning av enzymaktiviteten och följdaktigen till en minskning av antalet antigen på erytrocyter. Detta kan göra blodgruppering vansklig att tolka och risken finns att fel blodrekommendation utfärdas.

Enstaka punktmutationer i ABO-genen ansvarar för många kända fenotyper. A- och B-generna ger upphov till enzymen A- respektive B-transferas. Blodgrupp O beror däremot på dubbel uppställning av O-gener innehållande någon typ av defekt (t.ex. punktdeletion), som orsakar ett inaktivt (ofunktionellt) enzym eller t.o.m. avsaknad därav. Det finns även ett stort antal undergrupper inom A/B-blodgrupperna, t.ex. A\textsubscript{1}, A\textsubscript{x}, A\textsubscript{el}, A\textsubscript{weak}, B\textsubscript{3}, B\textsubscript{x} och B\textsubscript{weak} som uppvisar svagare reaktioner vid försök till blodgruppering. Studier avseende A- och B-specificiteten har förfinsats på senare tid genom de nya DNA-baserade teknikerna, bl.a. PCR-teknik (en "kopieringsapparat för DNA"), kloning etc. Mina studier i denna avhandling har kunnat utföras...
tack vare alla dessa framsteg inom molekylärbiologin tillsammans med upptäckten av ABO-genen.

I stort handlar avhandlingen om en fördjupad studie av bakomliggande orsaker till och konsekvenser av den överraskande mångfald som ses i ABO-genen. Länge trodde man att det fanns fyra genetiska varianter, s.k. ABO-alleler (A1, A2, B och O), men redan 1993 stod det klart att situationen var betydligt mer komplex än så. Syftet i de fem första vetenskapliga artiklarna har varit att klarrätta den molekylärgenetiska bakgrunden och blodgruppssituationer inom den svenska populationen och i andra befolkningssamhällen med fokus på svaga ABO-blodgrupper. I den sista artikeln försökte vi att utveckla en säker och kostnadseffektiv metod för att detektera alla dessa vanliga och sällsynta alleler.

En omfattande insamling av kliniska fall, remitterade till det Nordiska Referenslaboratoriet för Genomisk Blodgruppstypning vid Blodcentralen Skåne på Universitets sjukhuset i Lund, undersöktes i de olika delprojektet. I delarbete I och V identifierades olika nya punktmutationer, både s.k. missense- (som ger upphov till ett aminosyrabyte) och nonsense-mutationer (som förstör genens möjlighet att ge upphov till ett funktionellt enzym). Därmed karakteriseras den genetiska bakgrunden hos 17 nya undergrupper inom ABO-systemet.

I delarbete II beskriver en hybridallel (en blandning mellan två alleler) hos personer av afrikanskt ursprung med en variant av blodgruppen AB varvid svagt uttryck av A-antigenet på blodkropparna gjorde blodgruppering svårtolkad. Resultatet av en lång rad av undersökningar på ABO-genen visade att en fusion av en vanlig O-allel (O\(^{1v}\)) och en vanlig A\(^{2}\)-allel hade bildats. Denna nya allel verkade orsaka ett mindre effektivt enzym. För att med säkerhet kunna definiera allelen som en hybrid, undersöktes och kartlades intronerna (de icke-kodande regioner som finns mellan genens exoner, d.v.s. de kodande regionerna) och deras genetiska variation (polymorfism) hos personer med olika blodgrupper. Med denna kartläggning kunde vi bevisa var sammanblandningen (hybridformation) av dessa alleler hade skett.


I delarbete IV, rapporterade vi att den sedan flera decennier välkända ABO-undergruppen Abantu överraskande nog, även den, beror på en hybridbildning: denna gång mellan den vanliga A\(^{2}\)-allelen och en aldrig tidigare beskriven O-allel som utgör en helt egen evolutionär (historisk) utvecklingslinje inom ABO-allelfamiljen. Allelen förekommer huvudsakligen bland afrikanska folkgrupper inkl. de s.k. Bantu-folken, precis som varit känt länge i fallet Abantu och därför kallades allelen O\(^{1bantu}\). Även här spelade intron-polymorfismerna en avgörande roll för att kunna bestämma var hybridformationen skett. Medan andra svaga ABO-undergrupper är sällsynta, är Abantu den A-undergrupp som har högst frekvens i någon befolkning, upp till 8% av vissa folkgrupper i södra Afrika.

Genetiska metoder finner allt fler applikationer inom det transfusionsmedicinska fältet. Det sista delarbetet (VI), är en sammanfattnings av nästan tio års arbete (från mitt 20-poängarbete 1997 för M.Sc. fram till idag). Under många år har Martin L Olsson (min handledare), Åsa Hellberg (min käresta) och jag ständigt förbättrat och justerat en metod för att kunna detektera de allt fler vanliga och sällsynta ABO-alleler som bör kunna upptäckas (idag}
Denna analys har använts parallellt med andra metoder som publicerades från Lund 1995-6, vilket gett en god validering under kliniska betingelser. En jämförelse med en kommersiellt tillgänglig analys visade också att vår nya metod är överlägsen, framför allt vid s.k. hybrid-, ovanliga O och subgruppsallele.

Att detektera dessa komplexa allele är viktigt när blodgivare och/eller -mottagares prover ger svårtolkade resultat vid blodgruppering. Då kan DNA-baserad analys användas som ett komplement för att minska risken för falska provsvar, som skulle kunna leda till felaktiga konklusioner och transfusionsrekommandationer.

Sammanfattningsvis har studierna i denna avhandling lett till en bättre förståelse kring de funktionella och genetiska sambanden inom ABO-systemet. Det långsiktiga målet är att med någon typ av automatiserade platfformar kunna fastställa blodgruppen helt genetiskt, så som redan görs för vävnadstypen (HLA) inför transplantationer.
ACKNOWLEDGEMENTS

Although I spent many years of hard work to complete this thesis, it all depended on the fortune of having many generous friends and colleagues around me who supported me during this study. I am especially grateful to the following persons, knowing that there are many others who are not mentioned here.

First and foremost, my ever-enthusiastic supervisor Professor Martin L Olsson, for giving me the opportunity to work with such exciting projects, for guidance through scientific thinking and sharing his knowledge.

Britt Thuresson, PCR expert, for guidance and teaching me scientific methodology, especially in the field of molecular biology, and not least for taking me down to earth at occasions when too many “good ideas” were occupying my mind.

My colleagues at the Department of whom I particularly will mention Alan Chester, Jill Storry and Åsa Hellberg who always helped me with language revision and adopted a critical attitude towards the thesis.

My friend and colleague Elisabet SW, who shared with us her nice house many times during our kick-off meetings (Landet). The most enjoyable times have been those everyday-moments in the corner (dum-forum) of the lab where we could allow us to be dumb and discuss freely about different things.

Nidal and all other people in the research group who inspired me and generously provided both time for discussions (genlunch) and good advice.

I would especially like to thank my roommate and good colleague, Gunilla Hochberg for all support and patience throughout these years, which made it possible to write this thesis. Li and Alicja are thanked, for all help during this long time in the same office.

I would like to express my appreciation towards all the colleagues at the Blood Centre who helped me in all weather when I needed to be cheered up, especially Kerstin x 3 (B, H & T) and of course Elisabeth, Ingela Å, and Eva Marie.
Professor Tor Olofsson, Louise, May-Louise and Bodil in the laboratory at the Department of Hematology and Transfusion Medicine, BMC, for support and for introducing me to RNA preparation.

Professor Monica Palcic, Mattias Persson and James Letts for valuable discussions and help with enzyme purification and computer modelling at the Carlsberg laboratory in Copenhagen.

Charlie, Hannah and Gerd for helping me with sequencing at the Department of Clinical Chemistry.

All my co-authors in different parts of the world for fruitful collaboration.

The friends for reminding me about life outside the lab, Ramtin, Per, Mohammad, Mikael, Stefan (lektorn), Anna, Percy, Sören, Marika, Anica (?-alliansen), Leif, Åsa, Erik, Jytte, Ewa, Magda, Ann-sofie, Orlando, Pasha, Ola, Birger, Örjan, Hamid, Nori, Mahmood and Mattias.

My dear Family, my beloved Parents, Bahareh, Azita and specially Göran and Ingrid, for always believing in me and for being there when Åsa & I needed them.

And most of all, my friend, roommate and my love for her support and good cooperation both at work and at home. Finally the youngest one, my sunshine Arvin (Seed-Nisse) who gave me new strength every day.

This project was supported by the Medical Faculty at Lund University, the Swedish Research Council, governmental ALF research grants to Lund University Hospital, and Region Skåne, Sweden, the Claes Högman SAGMAN-stipendium, the Inga and John Hain Foundation for Medical Research, Georg Danielssons Fund for haematological disease, Tore Nilsons Foundation for Medical Research.
REFERENCES


Harvey W. *Exercitatio anatomica de moto cordis et sanguinis in animalibus.* Frankfurt, 1628.


Lower R. The success of the experiment of transfusing the blood of one animal into another. *Philos Trans R Soc* 1, 352, 1666.


Olsson ML, Chester MA. Evidence for a new type of O allele at the ABO locus, due to a combination of the A^e nucleotide deletion and the A^e nucleotide insertion. Vox Sang 71:113-117, 1996c.


Olsson ML, Chester MA. Heterogeneity of the blood group A^e allele: genetic recombination of common alleles can result in the A^e phenotype. Transfus Med 8:231-238, 1998a.


Schenkel-Brunner H. In human blood groups. Springer-Verlag, New York, 2000


Selslma A, Blaszczyk R. Missense mutations outside the catalytic domain of the ABO glycosyltransferase can cause weak blood group A and B phenotypes. Transfusion 45:1663-1669, 2005.


Thuresson B, Dykes J, Olofsson T, Olsson ML. The number of repeats in the CBF/NF-Y enhancer VNTR region does not influence the level of ABO transcripts in blood and erythropoietic culture. *Vox Sang* 91:135, 2006.


