The genetics of the Rhesus blood group system*

Willy A. Flegel

Institut für Klinische Transfusionsmedizin und Immungenetik Ulm und Institut für Transfusionsmedizin, Universitätsklinikum Ulm, Germany

The Rhesus factor is clinically the most important protein-based blood group system. With 49 antigens so far described, it is the largest of all 29 blood group systems. The unusually large number of Rhesus antigens is attributable to its complex genetic basis. The antigens are located on two Rhesus proteins - RhD and RhCE - and are produced by differences in their protein sequences. In CD nomenclature, they are termed CD240D and CD240CE. Unlike proteins of other blood groups, Rhesus proteins are expressed only in the membranes of red blood cells and their immediate precursors.

Rhesus is second in its clinical importance only to the ABO blood group. Since the introduction of postpartum anti-D prophylaxis in the late 1960s, and combined pre- and postpartum anti-D prophylaxis in the early 1990s, the incidence of haemolytic disease in newborns due to alloimmunization has been reduced by more than 90%. Up to 1% of all pregnant women have clinically significant anti-erythrocyte antibodies.

Anti-D remains the main indication for phototherapy or exchange transfusions in newborns, and pregnant women who are D negative show an above average incidence.

The five most important Rhesus antigens are the cause of most alloimmunizations following blood transfusion. According to the German haemotherapy guidelines [Richtlinien zur Gewinnung von Blut und Blutbestandteilen und zur Anwendung von Blutprodukten], D negative transfusion recipients must always be given D negative erythrocyte products.

Since 2000, women of reproductive age and girls have also received transfusions compatible for further Rhesus antigens such as C, c, E and e in addition to the K antigen of the Kell blood group.

This procedure also applies to patients who receive regular transfusions or have immunohaematological problems, like anti-erythrocyte allo- and autoantibodies. In the case of autoantibodies, their exact specificity is not usually determined. Although one thirds of such autoantibodies are directed at Rhesus proteins, this has virtually no practical consequences for treatment.

The D antigen, discovered in 1939, was the first Rhesus antigen to be described. D positive patients were termed Rhesus-positive. In 1946, a quantitative variant with a weakly expressed D antigen was discovered and termed "D". This variant, now called "weak D", is of clinical and diagnostic importance.

Since 1953, it has been clear that there are also qualitative variants of the D antigen. Although patients with this partial D variant are positive for the D antigen, they can also form anti-D.

The genetic basis

In order to understand the genetic basis of diseases, it is important to understand individual differences in genetic variability, as well as their frequency and distribution in the population. There is usually a close correlation between the genotype and the expressed phenotype. Thus, taking a change in the RHD gene as an example, it is possible to make inferences about the expression of the RhD protein in the erythrocyte membrane. As is the case with many D variants, modified RhD protein can have important implications for transfusion related antigenicity.

The molecular basis of the RH alleles

The first Rhesus gene, the RHCE gene, was discovered in 1990. The RHD gene was found two years later, and the total deletion of this gene ascertained as the cause of the European D negative phenotype.

*Part of this review was presented by the Author during the XXXIX SIMTI Congress (Paestum, SA, 4-7 October, 2006)
More than 170 alleles have been found on the \textit{RHD} gene since. The site has still not been explored fully, even 15 years after the first \textit{RH} gene was cloned. DNB, the commonest of all European partial \textit{D} alleles, was described as recently as 2002\textsuperscript{7}.

In 2002, comparisons between the Human Genome Project and the Mammal Genome Project increased understanding of the formation of the two \textit{RH} genes on chromosome 1 (figure 1)\textsuperscript{8}.

Most mammals only have one \textit{RH} gene, whose position corresponds to the human \textit{RHCE} gene. The \textit{RHD} gene arose from the duplication of an ancestral \textit{RH} gene during mammalian evolution. An \textit{RHD} deletion occurred\textsuperscript{9} during the evolution of hominids, so that many modern humans completely lack the \textit{RHD} gene. This haplotype (glossary) is the leading cause of the \textit{D} negative phenotype worldwide.

The \textit{RH} alleles can be grouped according to their molecular structure. For the most part, these groups show point mutations (SNP, single nucleotide polymorphisms) which cause missense, nonsense, frame shift or splice site mutations (glossary). \textit{RHD-CE-D} hybrid alleles are often formed by gene conversion.

The examples of molecular changes and their effects on the \textit{D} antigen (Table I) show how the \textit{D} antigen phenotype correlates with the molecular structure.

\textbf{The molecular basis of the Rhesus phenotypes}

The two Rhesus proteins, RhD and RhCE, are very similar, differing in only 36 of the 417 amino acids, which they each comprise. Each has twelve segments within the erythrocyte membrane and six extracellular loops (figure 2). Both the amino (NH\textsubscript{3}) and the carboxyl (COOH) terminal are located within the cell.

\textbf{D negative phenotype}

The clinically essential difference between Rhesus positive and Rhesus negative hinges on the presence or

\begin{table}[h]
\centering
\begin{tabular}{|c|c|c|c|c|}
\hline
Classification of antigen change & D antigen phenotype & Molecular basis & Representative example & New rhesus antigen \\
& & Protein alteration & Description of the \textit{RHD} allele & Common name \\
\hline
Partial D & Qualitatively altered & Amino acid substitution on the external surface hybrid protein: protein segment exchange on the outer surface & Missense mutation: \textit{RHHD}(G355S) & DNB

Weak D & Quantitatively attenuated & Substitution of amino acids in the membrane or extracellularly & Missense mutation: \textit{RHHD}(V270G) & Weak \textit{D} type 1

DEL & Quantitatively markedly attenuated & Strongly reduced translation or protein at the site expression & Missense mutation: \textit{RHHD}(K409K) of splicing & \textit{RhD} (D2951) in C De n/a**

D negative & \textit{D} negative & Absent protein expression & Gene deletion: \textit{RHHD} deletion \textit{RHHD}(Y330X) & \textit{D} negative n/a**


Gene conversion: \textit{RHHD}(488 del 4) & Modifying gene: \textit{RHHD}(Y330X) & Gene conversion: \textit{RHHD}(Y330X) & \textit{RhD} & \textit{Cde}**

Antithetical RHCE protein antigen & Presence of antigen E or e & Missense mutation in amino acid position 226 codes for antigen E & Missense mutation in amino acid position 226 in \textit{RHCE} & \textit{RHCE} allele: Ala 226 codes for antigen \textit{e}, Pro 226 codes for antigen \textit{E} & n/a**

\hline
\end{tabular}
\caption{Molecular changes in \textit{RHD} alleles and their correlation with phenotypes of the \textit{D} antigen}
\end{table}

* see glossary; ** not assigned
Figure 1 - Duplication of the RH gene and deletion of the RHD gene. The ancestral condition is shown as the RH gene locus in the mouse. The single RH gene is adjacent to the three genes SMP1, P29-associated protein (P) and NPD014 (N). Duplication created a second, reversed RH gene in humans, which is located between N and SMP1. At the insertion points before and after the RHD gene is a DNA segment about 9,000 nucleotides or base pairs (bp) long. The two DNA segments flank the RHD gene and are termed the upstream or downstream Rhesus box. In the RHD positive haplotype, the RHD gene could be lost again through recombination (figure 3). The scale gives the approximate length of 50,000 nucleotides in the genomic DNA.

absence of the RhD protein in the erythrocyte membrane (D positive resp. D negative).

It is unusual for erythrocyte or other cell proteins to be lacking entirely in many humans. This particular genetic feature contributes to the strong antigenicity of the RhD protein. During duplication of the ancestral RH gene, two DNA segments were formed, known as the Rhesus box (Figure 1). The RHD deletion resulted from an unequal crossover (figure 3), which occurs when two DNA segments are highly homologous, such as those of the Rhesus box. The RHD negative haplotype commonest among Europeans is characterized by a hybrid Rhesus box. Subtle molecular differences between the various forms of the Rhesus box are used for genetic testing.

The molecular basis of D antigen variants

Aside from lack of the RhD protein, the D negative phenotype is caused mainly by a series of changes in the RhD protein, which in turn change the phenotype of the D antigen.

Depending on the phenotype and their molecular structure, these RHD alleles are classified as either partial D, weak D or DEL.

Partial D

The RhD protein traverses the erythrocyte membrane several times, leaving only part of the protein exposed at the surface (Figure 2). If an amino acid is substituted in a portion of the RhD protein which is located at the outer surface of the erythrocyte membrane, single epitopes of the D antigen can be lost or new antigens can be formed. DNB is the commonest European partial D (Table I).

D categories are a subgroup of partial D. The structure of the RH gene site facilitates gene conversions (figure 4). In the RHD gene some homologous exons of the RHCE gene will be inserted, forming a hybrid Rhesus allele which expresses a corresponding hybrid protein. This is how the D categories III to VI arose. The changes usually affect a long string of amino acids, which is always located on the erythrocyte surface.
Figure 2 - The Rhesus protein in the erythrocyte membrane. Both Rhesus proteins show 417 amino acids, shown here as circles. Mature proteins in the membrane lack the first amino acid. The amino acid substitutions which distinguish the RhD from the RhCE protein are shown in yellow, with the four amino acids which code for the C antigen in green and the one which codes for the E antigen in black. The single amino acid substitutions which code for partial D are in blue, those which code for weak D are in red. The mutations, identified by the Ulm group, are in light blue and orange.

Figure 3 - Deletion of the RHD gene. Deletion of the RHD gene resulted from recombination between an upstream and a downstream Rhesus box on two different chromosomes. This is termed an unequal crossover. When the two crossed strands separate (from A over the recombination site to B), the DNA at the RH gene site completely lacks the RHD gene (C). This haplotype (C) occurs in about 41% of the population. An individual homozygous for this haplotype (about 17% are) is D-negative.
If an amino acid substitution is located within the erythrocyte membrane or the cytoplasm, this will result in a weak D phenotype (figure 2). Integration of the RhD protein into the membrane will be hindered, leading to quantitative weakening of the D antigen. There is usually no qualitative change, and hence no anti-D immunization. The weak D type 1 is the commonest in Europe (Table I).

DEL

A particularly weakly expressed D antigen is termed DEL (earlier Del), because it could only be demonstrated using elution. In elution, antibodies are separated from erythrocytes to demonstrate them in the eluate. The molecular changes are more severe than those seen with weak D, considerably hindering but not completely preventing integration into the cell membrane. All DEL alleles are rare in Europe, but up to 30% of all apparently D negative individuals in East Asia are bearers of the DEL allele RHD(K409K)10,12.

The C/c and E/e antigens

The clinically important Rhesus antigens C, c, E and e are the result of RhCE protein changes at only five amino acid locations (figure 2). Antigens are termed antithetical if a protein can present only one of them. They are caused by protein polymorphisms. Often there are two variants of a protein, which differ at only one amino acid location, such as the Rhesus antigens E and e. RHCE alleles showing the amino acid proline at position 226 express the E antigen, whereas RHCE alleles showing the amino acid alanine at this position express the e antigen (Table I). Similar differences between two RHCE alleles account for the antithetical C and c antigens. The antigen pairs C/c and E/e are not antithetical, however, because they result from substitutions at different locations. The four possible combinations occur at different frequencies (among Europeans: Ce > ce > cE > CE) and are inherited as haplotypes.

Clinical applications

Genetic investigation, like all investigations in medicine, should only be carried out in the context of a clear aim. As far as transfusions are concerned, molecular biological techniques already are being used to provide cost-effective answers to a number of clinically important questions. Methods used include polymerase chain reaction (PCR) for gene amplification and subsequent identification by electrophoresis, nucleotide sequencing and hybridization on biochips.

Anti-D in patients

The clinical problems encountered are caused by a small number of RHD alleles. Patients usually show partial D, in some rare cases weak D immunized by normal D antigen. Since category VI (DVI) is the most important of these, the Authors recommend using monoclonal anti-D antibodies for typing, which do not react with DVI15,16.
This procedure was included in the German haemotherapy guidelines in 1996 and has not been changed since. DVI carriers are therefore deliberately typed as false negatives to prevent transfusions with D positive blood and likely anti-D immunization\textsuperscript{17}.

After these precautions were built into the German guidelines, they also were adopted by other European countries. Unlike partial D, no anti-D alloimmunization has yet been reported for weak D type 1, 2 or 3\textsuperscript{18}. From a clinical perspective, it is helpful that this involves the commonest D alleles, which make up almost 90\% of all weak D types in Germany\textsuperscript{19}, since these patients can receive D positive blood transfusions and do not require D negative products. This procedure saves up to 5\% of all D negative erythrocyte products, since they can perfectly well be replaced by D positive products\textsuperscript{11}, thus avoiding bottlenecks in the supply of D negative blood products\textsuperscript{20}.

**Pregnant women and anti-D prophylaxis**

Pregnant women with weak D types 1 to 3 can also be given D positive blood transfusions, and require no anti-D prophylaxis. Each year, a one-off genetic test helps avoid repeated administration of anti-D to 3,500 pregnant women in Germany alone (up to 5\% of all D negative pregnancies), and with it, all possible side effects of this prophylaxis, which these women do not require. Up to 5\% of all anti-D injections are therefore unnecessary.

A one-off genetic test is more cost-effective than repeated administration of anti-D products. In order to implement this approach, the guidelines for medical care during pregnancy and after birth (motherhood guidelines), issued by the Federal Committee of Physicians and Health Insurance Funds [Bundesausschuss der Ärzte und Krankenkassen], would need to be adapted accordingly\textsuperscript{21}.

All pregnant women with rare weak D types would receive the necessary prophylaxis, which they would not automatically receive under the current state of haemotherapy\textsuperscript{6} and motherhood guidelines\textsuperscript{21}.

A foetus can be shown to be D positive by demonstrating foetal DNA in the plasma of peripheral maternal blood\textsuperscript{22}. Anti-D prophylaxis is unnecessary if the foetus is D-negative. This could save about 40\% of all the anti-D prophylaxis currently given during pregnancy. This method was developed in countries bordering on Germany, where intensive efforts are under way to implement this approach to genetic diagnosis\textsuperscript{23}.

**Prenatal diagnosis**

If the foetus needs to be tested for D antigen, amniocentesis or sampling from the trophoblast is the method of choice\textsuperscript{14}. Cordocentesis is no longer performed. As already mentioned, maternal plasma may be able to be used in future.

**Having a child and anti-D antibodies**

If the father is heterozygous for the \textit{RHD} deletion, there is a 50\% chance of the foetus being D-negative, in which case the pregnancy essentially free of any haematological risk. If the father is homozygous for the \textit{RHD} gene, the foetus will definitely inherit the D antigen, which could influence the couple’s decision on whether to have a child or not.

For several decades, it was impossible to determine whether an individual is heterozygous or homozygous for \textit{RHD} because serological methods are unsuitable. With the advent of the genetic diagnosis of the hybrid \textit{Rhesus box}, however, the possibilities have been expanded considerably\textsuperscript{9}. If the father is D-positive, it is now sufficient to test him for the \textit{RHD} deletion.

**Use in other diseases**

If the standard serological methods fail, genetic diagnosis is the method of choice for a reliable blood group typing of patients after a transfusion and those with auto- or alloimmune haematological anaemias. Although transfused leucocytes can under certain circumstances persist for years, they will not interfere with routine genetic diagnosis.

**Blood donors**

Appropriate investigation for the \textit{RHD} gene can identify apparently D negative donors, who in reality are weak D or DEL, thus ensuring that their blood will be given only to D positive recipients\textsuperscript{18}. Without genetic diagnostics, D negative blood transfusion recipients will continue to be immunized by the D antigen contained in such blood\textsuperscript{24-27}.

Donors who so far were misidentified as being D negative and whose erythrocytes are D-/D+ chimeras can now be identified correctly. Lifelong chimerism can result from monochorionic twin pregnancies. Any transfusion from donor sources such as these can result in anti-D immunization, because they also contain several millilitres of erythrocytes with a perfectly normal D positive phenotype. This D positive blood can only be detected using genetic investigation, not with routine serological methods\textsuperscript{10,27}. Any case of anti-D immunization is of considerable clinical importance for girls and women of reproductive age. In the case of a D positive pregnancy,
this would be likely to result in Rhesus haemolytic disease of the newborn.

**The function of Rhesus proteins**

Most blood group proteins have a known function. While purifying human Rhesus proteins, American physician Peter Agre discovered a water transporter protein\(^2\). This discovery earned him the 2003 Nobel Prize for Chemistry. Despite intensive efforts, however, no function has been found for the RhD and RhCE proteins. Although the Rhesus associated antigen (RhAG), a Rhesus homologue contained in erythrocytes, can transport ammonium ions\(^2\), the Rhesus proteins themselves could not be shown to have any such function. One possible function under investigation involves the exchange of CO\(_2\) and even O\(_2\). Other information on the RH alleles will only be gained from the everyday clinical application of genetic diagnostics, which could thus contribute to identifying their function.

From the perspective of basic research, where transfusion medicine will continue to make a contribution, scientific work on Rhesus\(^2\) and other blood groups has been quite productive, and is anything but finished.

**Outlook**

Genetic diagnosis has been used for blood group typing in clinical transfusion medicine ever since 2000\(^11,12\). As antenatal care has shown, genetic blood group typing has led to a better quality of care, by helping to avoid potential side effects and reducing costs. This is a rare combination, and justifies the extra costs involved in optimizing care via the use of genetic diagnostic techniques. As well as improving patient care, these methods can fuel the development of new methods\(^1\), which will also be used for health care outside of Germany. European departments of transfusion medicine are leading the field in molecular blood group diagnostics and applications, and will continue to contribute to improving patient care.

**Common Genome Variability Terms**\(^4,6\)

**SNP (single nucleotide polymorphism)**

Point mutation. Variability in a nucleotide sequence due to change of a single nucleotide.

**Allele**

The expression of a coding or non-coding nucleotide sequence (the exon resp. intron of a gene) with two or more variants, often differing by only a point mutation.

**Genotype**

A pair of alleles or variants of a nucleotide sequence occurring at homologous sites on paired chromosomes.

**Haplotype**

A combination of alleles or variants of a nucleotide sequence located close together on the same chromosome, and usually inherited together.

**Missense mutation**

Amino acid substitution in a protein caused by a point mutation. It can alter the function or antigenicity of a protein.

**Nonsense mutation**

A stop codon caused by a point mutation which prematurely stops synthesis of the amino acid chain, leading to loss of protein function of its expression.

**Silent mutation**

A point mutation which does not change the amino acid at the site. Although the protein is unchanged, it still can be associated with a clinically relevant phenotype and be used diagnostically.

**Frame shift mutation**

The loss or insertion of one or two nucleotides which shifts the reading frame and prematurely stops protein synthesis (or extends it in some rare cases), resulting in loss of protein function or expression.

**Splice site mutation**

A point mutation at a splice site (the exon-intron junction), causing faulty splicing of messenger RNA (mRNA) and skipping an exon, thus changing the amino acid sequence. Leads to loss of protein function or expression.

**Gene conversion**

The non-reciprocal exchange between two or more homologous genes, whereby a certain nucleotide sequence on a gene is substituted by a sequence on another gene, which is located on the same chromosome (conversion in cis).

**Acknowledgements**

This article is published by the courtesy of Christopher Baethge, M.D., Editor in Chief of the journal *Deutsches Ärzteblatt*. This English translation was provided by the journal *Deutsches Ärzteblatt*.

**Key Word:** Rhesus, blood group, molecular diagnostic, transfusion, pregnancy.

**References**


2) Howard H, Martlew V, McFadyen I, et al. Consequences for


8) Wagner FF, Flegel WA. RHCE represents the ancestral RH position, while RHD is the duplicated gene. Blood 2002; 99: 2272-3.


20) Garratty G. Do we need to be more concerned about weak D antigens? Transfusion 2005; 45: 1547-51.


Correspondence: Prof. Dr. med. Willy A. Flegel
Institut für Transfusionsmedizin
Universitätsklinikum Ulm
Helmholtzstrasse 10 - 89081 Ulm - Germany
e-mail: willy.flegel@uni-ulm.de