

Sequence-Based Typing of Human Blood Groups

Axel Seltsam^a Andrea Doescher^b

^a Institute Springe

^b Institute Oldenburg, Blood Center of the German Red Cross Chapters of NSTOB, Germany

Key Words

Blood group · Sequencing · ABO · Rhesus · Genotyping

Summary

In the last two decades, all but one of the genes encoding the 30 blood group systems present on red blood cells have been identified. This body of knowledge has permitted the application of molecular techniques to characterize the common blood group antigens and to elucidate the background for some of the variant phenotypes. DNA sequencing methodology was developed in the late 1970s and has become one of the most widely used techniques in molecular biology. In the field of immunohematology, this method is currently used by specialized laboratories to elucidate the molecular basis of unusual blood group phenotypes that cannot be defined by serology and genotyping. Because of the heterogeneity of the blood groups on both the antigen and the genetic level, special knowledge of the biology of blood group systems is needed to design sequencing strategies and interpret sequence data. This review summarizes the technical and immunohematologic expertise that is required when applying sequence-based typing for characterization of human blood groups.

Schlüsselwörter

Blutgruppe · Sequenzierung · ABO · Rhesus · Genotypisierung

Zusammenfassung

Die Klonierung der Gene fast aller Blutgruppensysteme (29 von insgesamt 30) in den letzten 20 Jahren war die Grundlage dafür, dass viele häufige und auch seltene Blutgruppenmerkmale molekularbiologisch abgeklärt werden konnten. Dabei spielte die Technologie der Sequenzierung, die sich seit ihrer Entwicklung in den späten 1970er Jahren zu einer zentralen Methode in der Molekularbiologie entwickelt hat und eine Analyse von Gensequenzen auf Einzelnukleotidbasis ermöglicht, eine entscheidende Rolle. In der Immunhämatologie hat sich die Sequenzierung als Referenzmethode etabliert, die vor allem dann zum Einsatz kommt, wenn ungewöhnliche Blutgruppeneigenschaften serologisch und mit Hilfe von anderen Genotypisierungsverfahren nicht abgeklärt werden können. Da sich die verschiedenen Blutgruppen in ihrer molekularen Struktur und genetischen Grundlage stark unterscheiden, setzten die Konzeption von Sequenzierungsstrategien und die Auswertung der Sequenzdaten ein Spezialwissen über die Biologie der Blutgruppensysteme voraus, welches in der Regel nur in Referenzlaboratorien verfügbar ist. Diese Übersichtsarbeit befasst sich mit dem technischen und immunhämatologischen Know-how, das notwendig ist, wenn Sequenzierungsverfahren für die Charakterisierung von menschlichen Blutgruppenmerkmalen eingesetzt werden.

Introduction

In the last 20 years, the field of immunohematology has faced a rapid development and application of molecular biologic techniques. Numerous studies have analyzed blood samples

from people with known antigen profiles and identified the molecular basis associated with many red blood cell (RBC) antigens. In the process of gathering more information on the molecular basis of RBC phenotypes it became obvious that

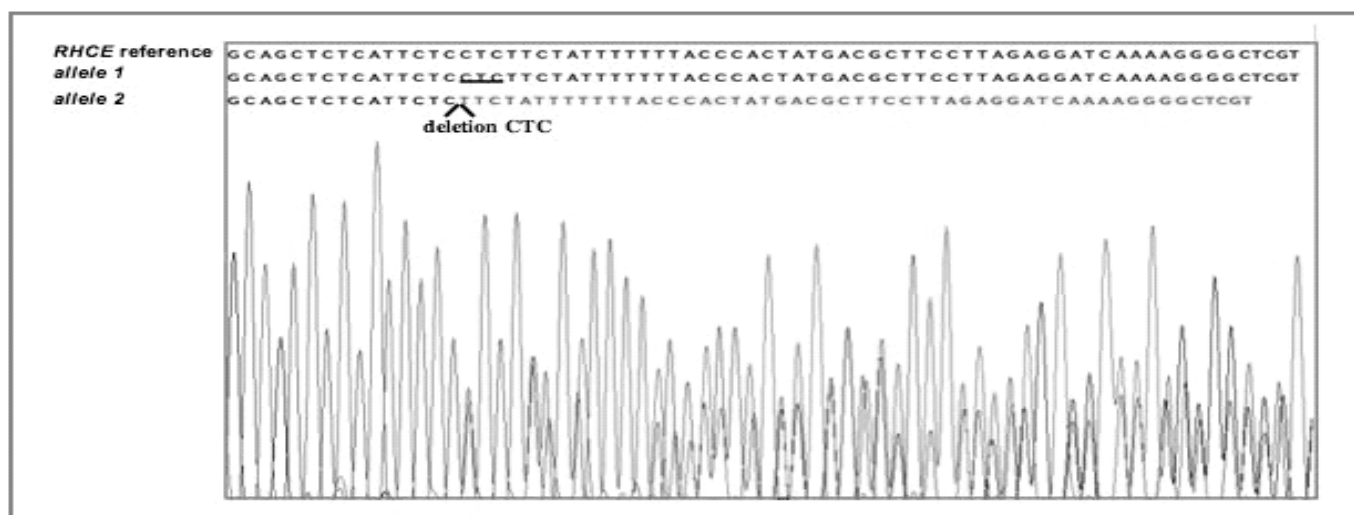


Fig. 1. Electropherogram and superimposed sequences of *RHCE* exon 1: top line representing the *RHCE* reference sequence, in comparison with a wild type allele (line 2) and an in-frame triplet deletion at position 79–81 (line 3).

the blood group systems are based on a great genetic diversity. According to the Blood Group Antigen Gene Mutation (BGMUT; www.ncbi.nlm.nih.gov/projects/gv/mhc/xslcgi.cgi?cmd=bgmur/resource) database of variations in the genes that encode antigens of blood group systems, more than 1,000 alleles of 40 genes are currently known [1]. The characterization of genes and determination of the molecular basis of antigens and phenotypes have made it possible to use PCR to amplify parts of DNA of interest in order to detect alleles encoding blood groups. While there are many molecular events that give rise to blood group antigens and phenotypes, the majority of genetically defined blood group antigens are the consequence of a single nucleotide polymorphism (SNP). Thus, simple DNA-based assays can be used to identify SNPs in genes encoding blood groups. Innumerable genotyping assays have been described to detect specific blood group SNPs, including PCR-RFLP (RFLP = restriction fragment length polymorphism), allele-specific PCR, real-time quantitative PCR, and high-throughput bead technology. However, all these genotyping methods are designed to detect and identify known alleles. In cases where new nucleotide mutations or new combinations of known mutations occur, a nucleotide-by-nucleotide analysis of the respective gene becomes necessary to determine the genetic sequence. Nucleic acid sequencing is a technique that allows determination of the order of the four different nucleotides in a DNA molecule [2]. It is considered the reference method in molecular biology. The nucleic acid sequence data obtained by this method can provide detailed information on any mutation present, on the organization of genes and on the protein product of the DNA/RNA analyzed.

While nucleic acid sequencing has already become an indispensable tool in blood group research, it is also increasingly used in routine blood grouping to finally clarify the genetic

background of variant blood group phenotypes when genotyping assays do not provide conclusive results. The importance of nucleic acid sequencing as reference method for blood group determination will grow with the increasing application of molecular methods in immunohematology. Therefore, it becomes more and more relevant for immunohematologists to become familiar with the principle and the areas of application of this method.

Principle of Sequence-Based Typing

The introduction of the PCR technique [3] presented the opportunity for the generation of target-specific DNA fragments and the easy detection of known polymorphisms in different genes, for example via single specific primer PCR (SSP-PCR). For the investigation of suspected mutations, DNA sequencing provides a powerful tool. A variety of sequencing methods have been described, mainly based on chain-terminating dideoxynucleotide sequencing published by Sanger et al. [2] and Rapley (ed.) [4]. While the original methods were time-consuming and often involved radioisotopes, the introduction of commercially available dye-labeled terminators has increased the reliability for cycle sequencing of PCR products. Each terminator (dideoxynucleotide triphosphates; ddNTPs) is labeled with a different fluorescent dye and attaches covalently to the amplicon during the extension reaction. The resulting products are separated with capillary electrophoresis and detected with laser technology. During this process an electropherogram is generated, which graphically represents the sequence of the PCR product (fig. 1).

There are principally two ways to sequence DNA: direct sequencing and sequencing of cloned DNA segments. Direct sequencing means that the PCR products obtained by ampli-

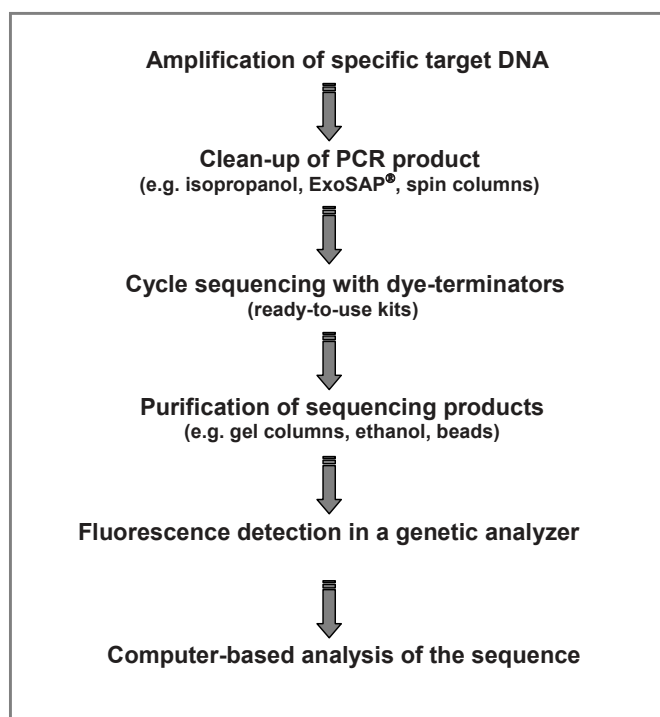


Fig. 2. Schematically depiction of the performance of a sequencing experiment.

fication of a certain DNA part are directly sequenced without first cloning the fragment. Figure 2 schematically shows the general procedure of a direct sequencing experiment.

In contrast to direct sequencing, sequencing of cloned DNA refers to the procedure in which a DNA sequence is amplified by genetic engineering techniques and then used as template for nucleic acid sequencing. The major advantages over direct sequencing are that the cloned DNA can be propagated indefinitely and that individual haplotypes are separated which can subsequently be sequenced. The drawbacks are that cloning is labor-intensive and time-consuming and requires special know-how and laboratory equipment. In addition, it can sometimes be difficult to discriminate true mutations from those artificially introduced into the DNA during the amplification process in bacteria. Nevertheless, sequencing of cloned DNA is particularly helpful in cases where haplotypes cannot otherwise be separated or a chimerism is suspected, suggesting the presence of more than two haplotypes.

Indications of Sequence-Based Typing in Immunohematology

The classical method of testing for blood group antigens and antibodies is hemagglutination. This technique is simple and inexpensive and, when done correctly, has a specificity and sensitivity that is appropriate for the clinical care of the vast majority of patients. However, in some aspects, hemaggluti-

nation has limitations. For example, it cannot indicate *RHD* zygosity in D-positive individuals precisely and it cannot be relied upon to type patients and donors who have a positive direct antiglobulin test and who have recently received transfusions. DNA-based blood group typing assays have proven extremely useful in those areas where hemagglutination is of limited value. There is a growing list of clinical applications of DNA analysis for blood group antigens [5]. At present, methods for blood group genotyping are predominantly used as a supplement to hemagglutination. In clinical laboratories it is becoming more and more common practice to follow a graduated operating procedure that provides for hemagglutination followed by DNA analysis when a blood sample shows an unusual blood group phenotype or contains difficult-to-identify RBC antibodies. While this step-wise approach proves successful in many cases, its limitation is documented by the fact that there are consistently a number of blood samples showing unusual serologic characteristics in the results of the genotyping assays that do not match with the corresponding phenotype. In these situations, extended DNA analysis by sequence-based typing is the most reliable way to define the genetic variations that may have caused the unusual blood group phenotypes. Consequently, modern blood group diagnostics is based on a three-step algorithm including serology, genotyping, and sequencing (fig. 3).

For correct determination of variant blood groups it will be crucial for the immunohematologist to know when sequencing is indicated. Figure 3 shows a flow chart demonstrating which findings make further DNA analysis by nucleic acid sequencing necessary. Initially, the genotyping result for a blood sample has to be compared with the possible combinations of known alleles and correlated with the serology-defined blood group. If the genotyping result is either incompatible with known allele combinations or inconsistent with the RBC phenotype, the genotype has subsequently to be determined by sequence analysis for further clarification. In other words, sequencing has always to be taken into account if genotyping does not provide plausible results that fit the phenotype.

Sequencing Strategies

When designing a sequencing strategy, the following questions have to be addressed:

- i) Which part(s) of the gene of interest are the most relevant (coding vs. non-coding regions) and the most informative (e.g. exons encoding for the extracellular part of the protein)?
- ii) Are there any homologous genomic segments that could interfere with specificity of the applied PCR reactions and interpretation of the sequencing data?
- iii) Are advanced sequencing approaches (e.g. RNA and haplotype-specific sequencing) required for resolution of the genotype?

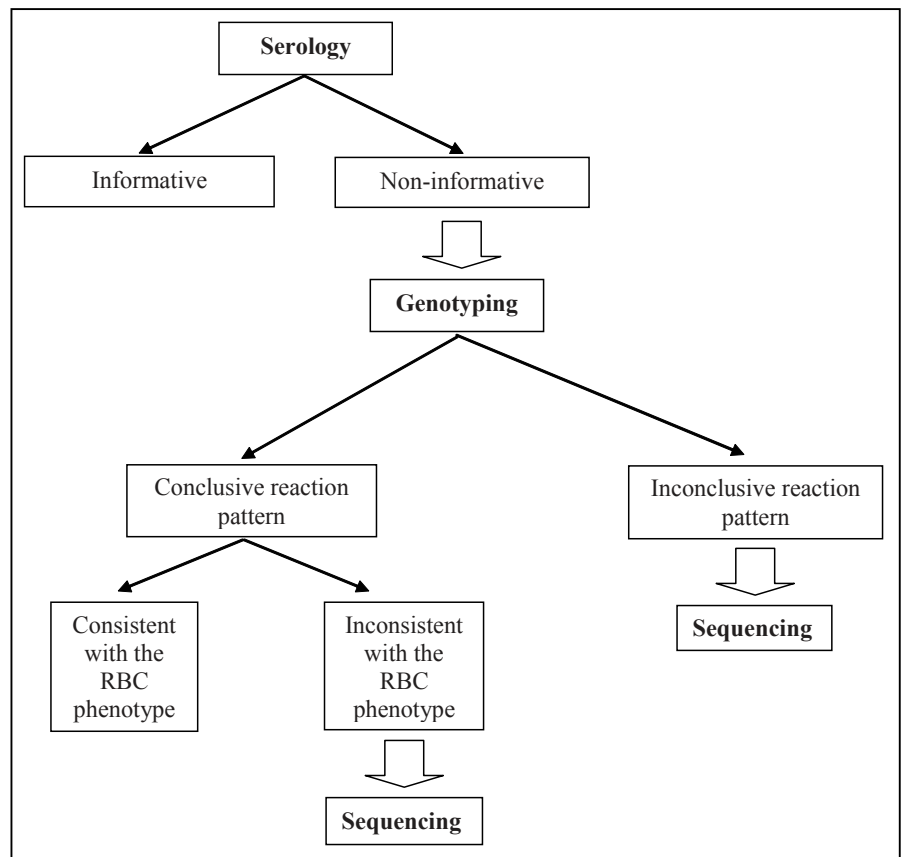


Fig. 3. Algorithm for determination of unusual blood groups. The flow chart illustrates when sequence-based typing of blood samples with unusual phenotypes is indicated. Blood samples with informative serologic results are subjected to genotyping. The genotyping result of a blood sample is then compared with the possible combinations of known alleles. In cases where the genotyping results are incompatible with known allele combinations (inconclusive reaction pattern), the genotype is subsequently determined by sequence analysis. Blood samples with genotyping results that are compatible with known allele combination (conclusive reaction pattern) are only further subjected to sequence analysis if their allele combinations do not correlate with the serology-defined blood group phenotype.

Depending on the structure and allelic repertoire of the gene of interest, some or all of these questions may become relevant for blood group sequencing.

Coding and Non-Coding Regions

A gene is a portion of DNA that contains both ‘coding’ sequences that determine what the gene does, and ‘non-coding’ sequences that determine when the gene is active (expressed) [6]. The number of exons differs between the blood group genes, e.g. *ABO* has 7 exons, *RHD* has 10, and *KEL* has 19 exons. The length of the intermediate non-coding regions shows a wide range and significantly contributes to the total length of the gene. Thus, it depends on the gene structure and the lengths of the intron sequences whether the complete blood group gene can be amplified by a single PCR reaction or whether a set of PCR reactions is needed to amplify different gene segments. Most sequencing strategies focus on the coding parts and the adjacent splice sites of the blood group genes because mutations in these regions can directly affect the associated blood group phenotypes. The *FY(DARC)* gene, for example, comprises 1,558 base pairs and can therefore easily be amplified by a single PCR and afterwards sequenced using three or four primers. In contrast, the *RHD* gene comprises more than 55,000 base pairs while its coding region consists of only 1,254 (<2%) nucleotides. *RHD* has 10 exons and the size of the intervening introns ranges from 426 base pairs to 11.8

kilo base pairs. Because of the considerable lengths of most of the introns, most investigators prefer an exon-wise amplification and sequencing approach to determine the *RHD* genotype.

Homologous Genes

The two *RH* genes, *RHD* and *RHCE* are highly homologous and share more than 90% identity [7]. Because the orientation of both genes is opposite, the molecular basis for gene conversion in *cis* is given [8]. The proposed mechanism is the formation of a hairpin at the chromosomal level, resulting in the alignment of two homologous gene segments in identical orientation. Resolving the hairpin leads to different hybrid genes often recognized in *RHD* and *RHCE*, for example D category VI type 2 (*RHD-CE(4-6)-D*) or *RHCE-D(5)-CE*. Also parts of *RH* exons can exchange with their D or CE-specific counterpart (e.g. D category Va type 1 or E Variant type III). So far, a deletion of one *RHCE* gene like *RHD* in D-negatives is not known [7]. Therefore, hybrid genes may sometimes remain unrecognized by exon-wise sequencing of the *RHCE* gene when a normal *RHCE* in *trans* is present. The primer design for PCR and/or cycle sequencing requires the differentiation between these two genes [9]. A similar situation is given for the MNS blood group where three homologous genes with an identity of 95% and higher (*GYP A*, *GYP B*, *GYP E*) are

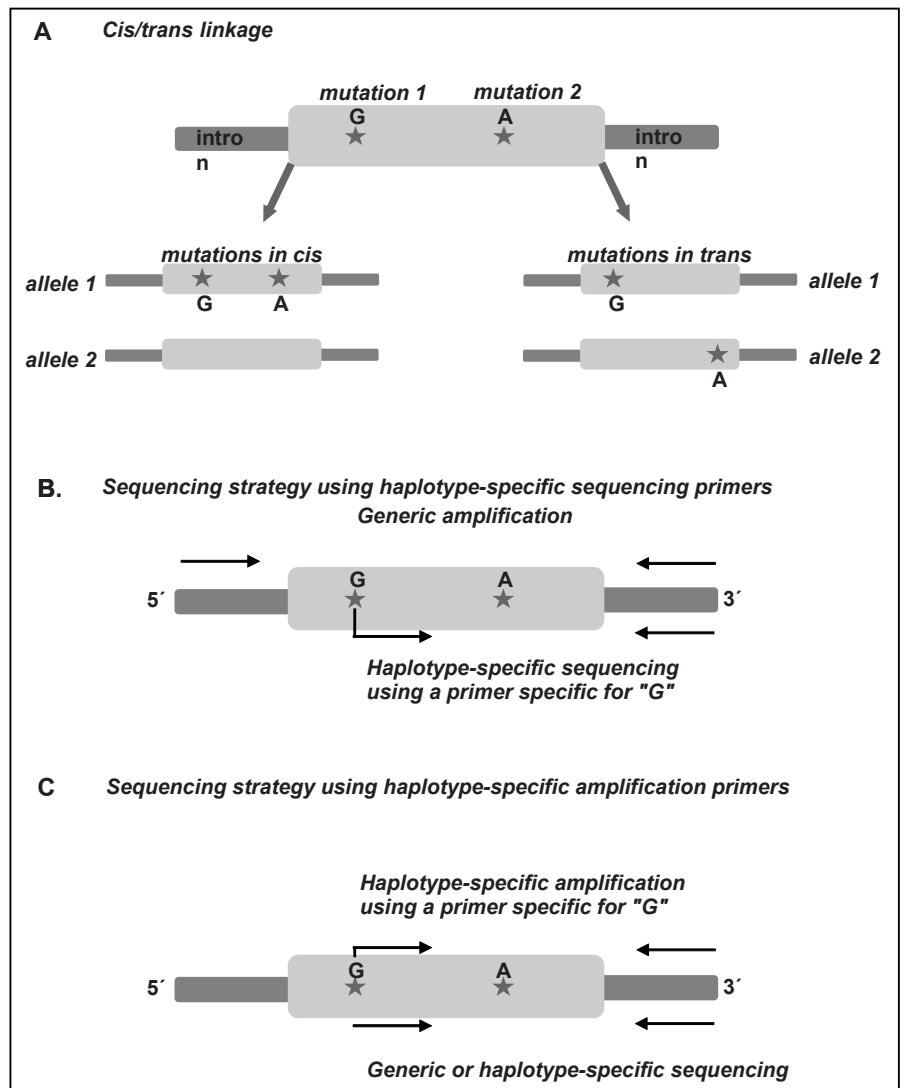


Fig. 4. A The localization of polymorphisms in *cis* and *trans* configuration and **B, C** the sequencing strategy for the resolution of a *cis/trans* linkage.

present. As described for *RH*, gene conversion occurs in *GYP A* and *GYP B* [10]. Thus, amplification and sequencing of homologous genes require special attention to a possible misinterpretation of data and a profound knowledge of the methods in use.

Cis/Trans Linkage of Polymorphisms

For correct definition of the genotype it is necessary to determine whether nucleotide mutations are located on the same allele (*cis* configuration) or different alleles (*trans* configuration) (fig. 4A). There are principally two strategies to determine the *cis/trans* linkage of mutations in a certain gene segment by direct sequencing. One strategy is to simultaneously amplify both alleles of the template (generic amplification) and use allele-specific primers for sequencing (fig. 4B). The alternative strategy is to generate haplotype-specific templates using allele-specific primer pairs; these templates can then be sequenced using generic or allele-specific primers (fig. 4C) [11]. Separation of haplotypes by cloning and subsequent

DNA sequencing provides another opportunity to determine the *cis/trans* linkage of polymorphisms. Haplotype determination can also be accomplished by the bead technology (so far only available for HLA genotyping). The principle of this method is the haplotype-specific extraction of DNA with specific oligonucleotides bound to electromagnetic beads (QIAGEN, Hilden, Germany).

Sequencing of RNA

Ribonucleic acid or RNA is a nucleic acid polymer consisting of nucleotide monomers that plays several important roles in the processes that translate genetic information from DNA into protein products; RNA acts as a messenger between DNA and the protein synthesis complexes known as ribosomes. RNA is less stable in the cell and experimentally also more prone to nuclease attack. As RNA is generated by transcription from DNA, the information is already present in the cell's DNA. However, it is sometimes desirable to sequence RNA molecules. In particular, in eukary-

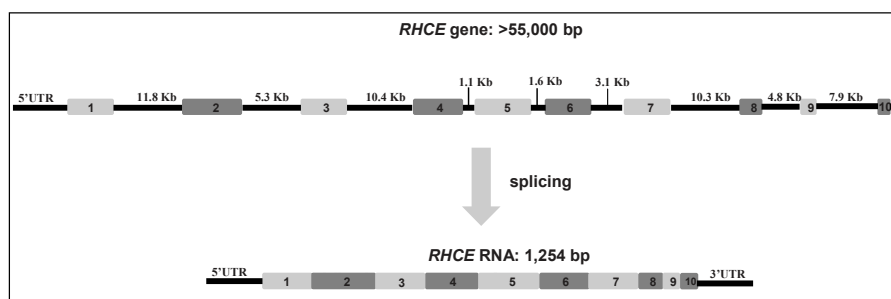


Fig. 5. Removal of introns during the splicing process of the *RHCE* transcript.

Table 1. Gene, cytogenetic location and reference sequences of blood group systems routinely sequenced at BSD-NSTOB

Blood group	Gene symbol	Number of exons	Reference sequence
ABO	<i>ABO</i>	7	HSA536140
Rh	<i>RHD, RHCE</i>	10	BN000065
Kell	<i>KEL</i>	19	AY228336
Duffy	<i>FY</i>	1	NM002036
Kidd	<i>SLC14A1</i>	8	NC000018
MNS	<i>GYPA</i>	7	NM0029099
	<i>GYPB</i>	5	J02982
	<i>GYPE</i>	4	AY297569
Lutheran	<i>LU</i>	15	NC000019
Colton	<i>AQPI</i>	4	NC000007
Landsteiner-Wiener	<i>ICAM4</i>	3	DQ011692
Rh-associated glycoprotein	<i>RHAG</i>	10	NC000006

otes RNA molecules are not necessarily co-linear with their DNA template, as introns are excised. To sequence RNA, the usual method is first to reverse transcribe the sample to generate DNA fragments. These can then be sequenced as described above [12, 13]. Figure 5 shows the organizations of the *RHCE* gene and its mRNA transcript. Similar to *RHD*, the intron sequences are up to 11 kilo base pairs, whereas the length of the RNA as a product of transcription and splicing has only 1,254 base pairs. Analogously to genomic DNA sequencing, the *cis/trans* linkage of mutations can be determined by using allele-specific amplification or sequencing primers.

A main problem of handling RNA is its instability in blood samples [14]. Because RNA can be rapidly degraded by RNases, storage and preparation of RNA require special care. The simplest (and cheapest) method to keep RNA in blood samples stable is freezing of the sample immediately after collection (below -30°C). In addition, some companies offer special blood vacutainer or stabilizers for storage of RNA samples. Several methods for manual and automated RNA extraction from blood samples exist. Independent of the procedure, a possible contamination with DNA should be proofed. An easy control is the amplification of RNA with intron-specific primer pairs; a detectable PCR product is an indicator for a contamination with DNA.

Data Analysis

Several computer programs are available for the analysis of sequencing data and comparison to known reference sequences. In our institute we use the Sequencing Analysis Software v. 5.2[®] (Applied Biosystems, Foster City, CA, USA) for analysis of raw data and either SeqScape[®] (Applied Biosystems) or MacVector[®] (Accelrys Inc., San Diego, CA, USA) to compare the sequence of a sample with a known reference.

The exactness of the alignment and notification of nucleotide positions depends on the settings of the programs. The identification of heterozygous positions, especially if the second peak is much lower than the first, can lead to an oversight of single mutations. To avoid this error, we regularly perform a manual examination of the sequence. For the comparison with reference sequences the NCBI databank (National Center for Biotechnology Information; www.ncbi.nlm.nih.gov) provides a huge number of sequences. If the sequencing for a gene is performed for the first time, we compare our data with several references to avoid a misinterpretation because of sequencing errors. An example can be given for the *RHD* gene: The alignment with Acc.No. L08429 shows a missense mutation at position 1,035 while the comparison with another reference or with a control sample reveals a wild-type sequence. The gene name, numbers of exons, and reference sequence

for the blood group systems which are routinely sequenced in our institute (except *ABO*) are given in table 1.

A deletion or insertion of one or more nucleotides can cause problems in computer-based analysis of sequencing data if a generic sequencing strategy was used. The missing/additional nucleotides causes a frame shift in one haplotype and leads to a superimposition of the haplotype sequences 3' from this mutation (fig. 1), and manual analysis is required. Databases are available listing nucleotide changes so far detected in blood group genes. For most of the blood group systems the BGMUT database is a useful source for information about known variants. For the variations of the *RHD* gene, however, the 'Rhesus Base' on the server of the university of Ulm (www.uni-ulm.de/~fwagner/RH/RB) is a regularly updated database with information about phenotype and origin of the mutation as well as several links to publications. Unfortunately, the database does not include the *RHCE* gene, which is highly homologous to the *RHD* gene and also shows a great number of known and newly detected variations [15, 16].

To make a new mutation public the 'EMBL database' (European Molecular Biology Laboratory; www.ebi.ac.uk/embl) offers a submission entry in which all information about the source of the sample, the blood group system, and the kind of variation is stored.

Interpretation of Sequence Data

Nucleic acid sequencing provides detailed information on the genetic sequence by determining the order of the four different nucleotides in a DNA molecule. The advantage of this method compared to other DNA-based typing approaches is that not only known but also novel nucleotide mutations can be detected. However, it is a general limitation of nucleic acid sequencing that the genetic data obtained are restricted to those parts of the DNA for which a certain sequencing strategy was designed. Therefore, immunohematologists who interpret sequencing results need to have a profound knowledge of the molecular basis of blood groups as well as of the specific sequencing strategy used.

It is known from numerous publications that the phenotypical relevance of a certain mutation depends on its type and location within the blood group gene. In general, mutations within coding and regulatory regions are more likely to have an impact on the blood group phenotype than mutations located within the non-coding gene segments. In addition, multiple base exchanges as well as single sequence mutations significantly affecting transcription or translation of the encoded protein (splice site mutation, early stop codon) normally have a greater influence on blood group antigen formation than nucleotide mutations resulting in a single amino acid exchange only. Nevertheless, the real impact on the phenotype of a given sequence variation can only be assessed if the biology of the respective blood group is also considered.

If the antigens of a blood group system reside on a (glyco)protein, their occurrence depends on the conformation and surface expression level of the protein. The protein conformation in turn is based on a number of different factors: the membrane protein type (single-pass, multi-pass, glycosylphosphatidylinositol(GPI)-linked, passively absorbed), the structure of the extracellular domain(s), and the degree of glycosylation. Dependence of the phenotypical relevance of an amino acid exchange on the topology of the protein and the site where it occurs in the molecule is nicely exemplified by the multi-pass protein RhD. Amino acid substitutions in weak RhD are predominantly located in the intracellular or transmembraneous protein segments and those found in partial RhD phenotypes in the extracellular protein segments [17]. Whereas weak RhD samples are characterized by a reduced RhD antigen density, partial RhDs have altered RhD proteins that differ sufficiently from normal RhD to allow allo-anti-D production. These findings strongly suggest that the changes in shape and expression level of RhD caused by amino acid substitutions follow a role that is strongly associated with the domains defined by the three-dimensional structure of the protein.

Different types of mutations in blood group genes were shown to cause blood group null phenotypes. In 2001 [8] and 2004 [18] two large studies on the presence of *RHD* in RhD-negative Europeans were published which found large hybrid genes (e.g. *RHD-CE(2-9)-D*), disrupted start or stop codons, splice site mutations, and single base exchanges in *RHD* to be associated with truncated or non-expressed Rh proteins. *RHD ψ* and *Cde^s* [19] are the leading *RHD* null alleles in populations of African origin and K409K in Asians [8, 20, 21]. Non-expressed genes also exist in other blood group systems. In the *FY* gene a mutation in the GATA-1 binding motif causes Fy(a- b-) phenotypes, frequently observed in individuals of African origin [22]. Similar to *RHD*, the *KEL* gene shows a broad spectrum of null alleles detected in a large survey in Austria [23]. So far, 17 different non-expressed *KEL* genes are known with either splice site mutations in introns or premature stop codons caused by different single mutations.

Many blood group proteins are expressed in close association with (e.g. RhD in the Rhesus complex) or are covalently linked (Kell and Kx) [24] to other membrane proteins. This explains why abnormal expression of a neighboring protein or mutations in those parts of the blood group protein which are in direct contact with other RBC proteins can cause unusual blood group phenotypes. For example, cell surface expression of RhD blood group polypeptide is posttranscriptionally regulated by the RhAG glycoprotein; in most cases, the lack of Rhesus in Rh(null) red cells is associated with mutations in the *RHAG* gene [25]. Thus, for correlation of the genetic with the serologic findings a basic understanding of the molecular mechanisms associated with the intracellular transport pathways of membrane proteins and the requirements for cell surface expression is needed.

In those cases in which the blood group antigens of interest are carried by carbohydrate chains, it is critical for assessing the phenotype from the genotype to know that the monosaccharide residues are sequentially assembled by the action of specific glycosyltransferases and that the sequence in sugar chains is determined by the substrate specificity of the respective glycosyltransferases involved. The AB determinants of the ABO blood group, for example, are characterized by a single monosaccharide residue, and their biosynthesis is based on the action of the *A* or *B* allele products, α 1.3-N-acetylgalactosaminyl- or α 1.3-galactosyltransferase [26]. The H structure, which is the determinant of blood group O and the substrate for the A and B transferases, is in turn the result of the action of the *H* gene-encoded α 1.2-fucosyltransferase. This proves that the genes of sugar-based blood group systems are not responsible for the synthesis of the entire blood group substance (as is the case for the blood group proteins); rather, they are responsible only for the transfer of the respective immunodominant monosaccharide to a suitable acceptor substance.

The difference in enzyme specificity between the highly homologous A and B transferases is merely based on amino acid exchanges at three positions of the catalytic cleft of the AB enzyme. However, additional amino acid variations in this region can change the specificity of the glycosyltransferase or produce enzymes with dual specificity generating AB phenotypes (e.g. *cis*-AB) [27]. More frequently, amino acid exchanges in the catalytic domain of the ABO transferase are associated with a reduction or complete loss of enzymatic activity (e.g. A_{el} blood group) [10]. According to the principle that the assembly of the sugar chains is based on the sequential action of different glycosyltransferases, the synthesis of an oligosaccharide chain and its antigenic epitopes is interrupted when one step in the biosynthesis sequence is blocked. This regularly happens in O individuals who express only non-functional ABO transferases incapable of converting the H determinants. Another example is the H-deficient phenotype (Bombay phenotype), in which the formation of H-specific structures is blocked because of a catalytically inactive H transferase, thus no A or B substance can be produced [28]. Overall, the fact that the impact of genetic variations on antigen formation is indirect and mediated by enzymes makes phenotype-genotype correlation in sugar-based blood group systems very complex. The recent finding that aberrant trafficking of variant A and B transferases can be involved in the formation of weak ABO phenotypes further emphasizes the need for special molecular biologic knowledge when dealing with unusual blood group phenotypes [29].

Sequence databases like BGMUT are extremely helpful in validation and interpretation of sequencing results when known nucleotide mutations are detected. However, in those cases in which novel mutations are found it is up to the immunohematologist to check for plausibility of these findings and to discriminate real sequence variations from technical artifacts. An even more challenging situation for the investigator is when no sequence variations are found in a sample with an unusual phenotype. Differential diagnosis in this case includes phenotypically relevant mutations in those parts of the gene that are not covered by the sequencing strategy used, antigen diminution due to epigenomic factors like DNA methylation, and antigen alteration associated with various medical conditions [30, 31]. At this point, reference laboratories with the relevant scientific expertise and an extended spectrum of methods should be engaged to assist in the diagnosis of unusual blood groups.

Perspectives

The increasing use of molecular genetic typing methods in immunohematology will further augment the importance of sequence-based typing as a reference method for genetic determination of unusual human blood group phenotypes. In the last few years, research on the genetic background of the blood group systems revealed that some of the systems, particularly ABO and Rhesus, show a great allelic diversity similar to that observed for HLA. Because only selected populations have been studied so far, it can be expected that the nucleotide sequence database on blood groups will continuously increase in the future. Since traditional genotyping methods are based on detection of known nucleotide mutations, both the increasing number of alleles and the fact that allele frequencies vastly differ between populations generally limit their possible applications. Nucleic acid sequencing is a powerful tool and provides by far the most detailed analysis of a particular piece of DNA. New high-throughput technologies for DNA sequencing combined with powerful computer-based data analysis have opened the avenue for rapid and efficient large-scale typing of donors and patients. Therefore, sequence-based typing strategies including whole gene sequencing hold the potential to become the standard for blood group genotyping.

Disclosure

The authors declared no conflict of interest.

References

- 1 Blumenfeld OO, Patnaik SK: Allelic genes of blood group antigens: a source of human mutations and cSNPs documented in the Blood Group Antigen Gene Mutation Database. *Hum Mutat* 2004;23: 8–16.
- 2 Sanger F, Nicklen S, Coulson AR: DNA sequencing with chain-terminating inhibitors. *Proc Natl Acad Sci U S A* 1977;74:5463–5467.
- 3 Mullis KB, Falcone FA: Specific synthesis of DNA in vitro via a polymerase catalysed chain reaction. *Methods Enzymol* 1987;155:335–350.
- 4 Rapley RE (ed): *PCR Sequencing Protocols*. Totowa, Humana Press, 1996
- 5 Reid ME: Applications of DNA-based assays in blood group antigen and antibody identification. *Transfusion* 2003;43:1748–1757.
- 6 Lewin B: *Genes*. New York, Oxford University Press, 1998.
- 7 Wagner FF, Flegel WA: RHD gene deletion occurred in the Rhesus box. *Blood* 2000;95:3662–3668.
- 8 Wagner F, Frohmajer A, Flegel WA: RHD positive haplotypes in D negative Europeans. *BMC Genet* 2001;2:10.
- 9 Wagner FF, Ladewig B, Flegel WA: The RHCE allele ceRT: D epitope 6 expression does not require D-specific amino acids. *Transfusion* 2003;43: 1248–1254.
- 10 Daniels G: *Human Blood Groups*. Oxford, Blackwell Scientific, 2002.
- 11 Seltsam A, Hallensleben M, Eiz-Vesper B, Lenhard V, Heymann G, Blasczyk R: A weak blood group A phenotype caused by a new mutation at the ABO locus. *Transfusion* 2002;42:294–301.
- 12 Szekely M: Rapid RNA sequencing. *Nature* 1977;269:754–755.
- 13 Wang Z, Gerstein M, Snyder M: RNA-Seq: a revolutionary tool for transcriptomics. *Nat Rev Genet* 2009;10:57–63.
- 14 Farrel RE: *RNA Methodologies: A Laboratory Guide for Isolation and Characterization*. San Diego, Academic Press, 1993, pp 47.
- 15 Doescher A, Vogt C, Bittner R, Gerdes I, Petershofen EK, Wagner FF: RHCE alleles detected in automated RH blood grouping in Northern Germany. *Transfusion* (in press).
- 16 Bugert P, Scharberg E, Geisen C, von Zabern I, WA F: RhCE protein variants in South-Western Germany detected by serological routine testing. *Transfusion* (in press).
- 17 Flegel WA, Wagner FF: Molecular biology of partial D and weak D: implications for blood bank practice. *Clin Lab* 2002;48:53–59.
- 18 Gassner C, Doescher A, Dove T, Rozman P, Eicher NI, Leger TJ, Lukin S, Garritsen H, Kleinrath T, Egger B, Kormoczi GF, Kilga-Nogler S, Schoenitzer D, Petershofen EK: RHD presence among RhD-negative, RhC/RhE positives – a central European multi-center study. *Transfusion* 2004;45:527–538.
- 19 Singleton B, Green C, Avent N, Martin P, Smart E, Daka A, Narter-Olaga E, Hawthorne L, Daniels G: The presence of an RHD pseudogene containing a 37 base pair duplication and a nonsense mutation in africans with the Rh D-negative blood group phenotype. *Blood* 2000;95:12–18.
- 20 Shao C, Maas J, Su Y, Koehler M, Legler T: Molecular background of RhD-positive, D-negative, D(el) and weak D phenotypes in Chinese. *Vox Sang* 2002;83:156–161.
- 21 Kim J, Kim S, Kim C-A, Yon G, Park S: Molecular characterization of D-Korean persons: development of a diagnostic strategy. *Transfusion* 2005;45:345–352.
- 22 Tournamille C, Collin Y, Cartron JP, Le Van Kim C: Disruption of a GATA motif in the Duffy gene promoter abolishes erythroid gene expression in Duffy-negative individuals. *Nat Genet* 1995;10: 224–228.
- 23 Körmöczy G, Wagner T, Jungbauer C, Vadon M, Ahrens N, Moll W, Mühlbacher A, Özgül-Gülce S, Kleinrath T, Kilga-Nogler S, Schönitzer D, Gassner C: Genetic diversity of KELnull and KELel: a nationwide Austrian survey. *Transfusion* 2007;47:703–714.
- 24 Yazdanbakhsh K, Lee S, Yu Q, Reid ME: Identification of a defect in the intracellular trafficking of a Kell blood group variant. *Blood* 1999;94:310–318.
- 25 Mouro-Chanteloup I, D'Ambrosio AM, Gane P, Le Van Kim C, Raynal V, Dhermy D, Cartron JP, Colin Y: Cell-surface expression of RhD blood group polypeptide is posttranscriptionally regulated by the RhAG glycoprotein. *Blood* 2002;100:1038–1047.
- 26 Yamamoto F, Hakomori S: Sugar-nucleotide donor specificity of histo-blood group A and B transferases is based on amino acid substitutions. *J Biol Chem* 1990;265:19257–19262.
- 27 Yamamoto F, McNeill PD, Kominato Y, Yamamoto M, Hakomori S, Ishimoto S, Nishida S, Shima M, Fujimura Y: Molecular genetic analysis of the ABO blood group system: 2. cis-AB alleles. *Vox Sang* 1993;64:120–123.
- 28 Kelly RJ, Ernst LK, Larsen RD, Bryant JG, Robinson JS, Lowe JB: Molecular basis for H blood group deficiency in Bombay (Oh) and para-Bombay individuals. *Proc Natl Acad Sci U S A* 1994;91: 5843–5847.
- 29 Seltsam A, Gruger D, Just B, Figueiredo C, Gupta CD, Deluca DS, Blasczyk R: Aberrant intracellular trafficking of a variant B glycosyltransferase. *Transfusion* 2008;48:1898–1905.
- 30 Kominato Y, Hata Y, Takizawa H, Tsuchiya T, Tsukada J, Yamamoto F: Expression of human histo-blood group ABO genes is dependent upon DNA methylation of the promoter region. *J Biol Chem* 1999;274:37240–37250.
- 31 Bianco T, Farmer BJ, Sage RE, Dobrovic A: Loss of red cell A, B, and H antigens is frequent in myeloid malignancies. *Blood* 2001;97:3633–3639.