A Southern Analysis of Rh Blood Group Genes: Association Between Restriction Fragment Length Polymorphism Patterns and Rh Serotypes

By C.A. Hyland, L.C. Wolter, Y.-W. Liew, and A. Saul

Polymorphisms within the Rh blood group system have been defined by serologic agglutination methods, but have not yet been defined at the DNA level. Two closely related genes associated with the Rh D antigen and with the Rh C/c and E/e antigens have been cloned. We used a Southern analysis incorporating probes to the 5' and 3' regions of the Rh C,E gene and D gene to identify polymorphisms associated with Rh C/c and E/e antigens, respectively. The D gene dosage could be determined by comparing the relative intensities of the D bands with bands from the 5' and 3' region of the Rh C,E gene. The concordance between restriction fragment length polymorphism (RFLP) patterns and serologic phenotypes for 102 randomly selected blood donors was 100% for C, c, and D, 94.8% for c, and 94.3% for E. The data are consistent with the sequences encoding the C/c epitopes residing on the 5' side of those for the E/e epitopes. All samples discordant for the 3' probe and E had the C,E (r') serotype. These data show that the gene coding for the C,E serotype is different in Rh-positive and -negative individuals. The study demonstrates that Rh DNA typing, including D gene dosage measurements and Rh gene haplotyping, may supplement traditional serotyping methods in transfusion medicine. © 1994 by The American Society of Hematology.

The major antigens contained within the highly polymorphic human Rh blood group system are the Rh D and the anti-allelic allelic products C,c and E,e. All of the Rh antigens, particularly Rh D, are clinically significant because of their importance in the pathogenesis of hemolytic disease of the newborn, incompatible transfusion reactions, and autoimmune hemolytic anemia. Rh antigens are generally defined using human polyclonal or monoclonal antisera in a red blood cell agglutination test system. The zygosity state for C,c and E,e can usually be obtained by direct testing for each allelic product. Where only one antigen is found serologically, it is assumed that the genotype is homozygous with two copies of the gene for that antigen. Serologic methods do not show whether one or two D genes are present in Rh D-positive individuals, as there is no alternative allelic product. The zygosity state for Rh D individuals is either deduced on a probability basis from gene frequency data for defined populations or inferred from family studies.

The Rh C, D, and E antigens appear to reside on three distinct polypeptide chains (Mr 28,000 to 34,000) that share common N-terminal amino acid sequences and a high degree of homology. Full cDNA sequences have been reported for two Rh genes, one of which encodes the Rh D-associated polypeptide, and the other, the C,c- and E,e-associated peptides. Differential splicing at intron/exon boundaries along the C,E gene may give rise to the different Rh polypeptides. These genes are on the same locus on chromosome 1 to which the CDE Rh antigens have previously been assigned. Nucleotide sequences for the two genes and predicted protein sequences show a 96.5% and 91.6% homology, respectively.

Southern analysis has demonstrated that the D gene is absent in Rh (D)-negative individuals. This explains the inability of serologic methods to identify an alternate allelic product for D. No DNA polymorphisms associated with the C,c or E,e antigens have been reported, although sequence variations were reported among cDNA clones derived from the C,E gene. One variation was a nonconservative G-to-C change at base 48, counting from the initiation ATG codon, which removes a Hae III restriction site.

In this study, we report on the use of probes to the 5' and 3' region of the Rh C,E gene and the closely related D gene in a Southern analysis. Using DNA restricted with one enzyme, Msp I, these probes identified gene fragments associated with the Rh D, C,c, and E,e antigens. Using DNA restricted with Hae III, the 5' probe identified a polymorphism associated with C,c expression. The concordance between these restriction fragment length polymorphism (RFLP) patterns and Rh serologic phenotypes was examined in a study of 102 randomly selected, volunteer blood donors.

Materials and Methods

DNA preparation. Citrate-phosphate-dextrose (CPD)-anticoagulated blood was collected from volunteer blood donors (n = 9) who belonged to an Rh-typing panel and from donors selected at random (n = 102). A further two r/r donors were used to check discordant RFLP patterns. Genomic DNA was isolated from peripheral blood leukocytes as described by Miller et al.

Serology. Phenotyping sera were in-house-prepared polyclonal anti-c, anti-E, and anti-c, polyvalent anti-C (Biostek, Frankfurt, Germany); and monoclonal IgM anti-D (Immunor; Churchill Diagnostics, Sydney, Australia). Red blood cells were papain-treated for all tests except for anti-D. Hemagglutination tests were performed at 37°C for 30 minutes. Samples that were negative with the IgM anti-D were tested further using monoclonal IgG anti-D (in-house) in the indirect anti-human globulin (AHG) test to detect weak D expression. Appropriate positive and negative samples were run in parallel with all tests.

Preparation of DNA probes. The 5' region of the Cc,De gene was amplified from genomic DNA, prepared from a donor with Rh CcDe phenotype, using the inverted polymerase chain reaction.

From www.bloodjournal.org by guest on July 15, 2017. For personal use only.
SOUTHERN ANALYSIS OF RH BLOOD GROUP GENES

A

Fig 1. Nucleotide sequence for the 5' and 3' regions amplified from the Rh C,c/E,e-associated gene. (A) 5' region amplified from genomic DNA using inverted PCR technique.14 Sequence lodged with GenBank, accession number M84983. (B) 3' region amplified from cDNA prepared from bone marrow extract. Bases numbered from ATG initiation codon of the published Rh sequence.10,11 Bases underlined indicate position of primers used in PCR.

B

For personal use only.on July 15, 2017. For personal use only.
cent dye–labeled M13 sequencing primers, following the Applied Biosystems sequencing protocol. Sequences were read by a model 370A Applied Biosystems Sequencer (Foster City, CA).

Southern analysis. Human genomic DNA (10 μg) was digested with restriction enzymes (10 U/μg), precipitated with ethanol, washed, dried, and resuspended in water. DNA fragments were separated by electrophoresis on a 1.0% agarose (Seakem; FMC Bioproducts) gel at 30 volts for 20 hours. HindIII-digested λ-phage and HaeIII-digested φX174 DNA served as molecular weight markers. DNA was transferred onto Zeta probe GT membrane (Bio-Rad, Richmond, CA) using PosiBlot apparatus (Stratagene) and 0.4 mol/L NaOH as transfer buffer. After transfer, membranes were washed in 2 × SSPE (0.18 mol/L NaCl, 0.1 mol/L NaH2PO4, and 0.01 mol/L EDTA at pH 7.4) and prehybridized for 1 hour in Church and Gilbert hybridization solution.14 The PCR products were labeled with (α-32P)deoxycytidine triphosphate (dCTP) in 10 PCR cycles using radioactive label instead of dCTP and 100 ng of DNA template, added to the membrane, and hybridized for 16 hours at 65°C. Membranes were then washed twice with SSPE for 5 minutes and exposed for 7 days to a Kodak X-Omat AR5 film (Eastman Kodak, Rochester, NY) with intensifying screen at −80°C or exposed for 2 days on a PhosphorImager Screen (Molecular Dynamics, Sunnyvale, CA).

Hybridization patterns were interpreted for coded samples by two independent persons. The banding patterns, including estimation of D zygosity, were readily interpreted by eye. Selected samples were verified by a quantitative densitometry measurement of radioactivity in each band using a PhosphorImager. Samples that gave an RFLP pattern discordant to that expected from the Rh serotype were retested.

RESULTS

Sequence of PCR products amplified from 5' and 3' regions of an Rh gene. A 490-base product was amplified from the 5' region of the Rh C,c and E,e gene, as shown in Fig 1A. This sequence extended 117 bases downstream from the translation initiation ATG codon into the first exon, and exhibited 100% homology with published sequences.10,11 The remaining 373 bases extended into the upstream region for which 44 and 30 bases had previously been reported.10,11 The published DNA sequence for Rh D gene,7 which extends 16 bases upstream, is identical to the C/c/E/e sequence in the region common to both the published sequences10,11 and the amplified 490-base product. This latter PCR product therefore cannot be assigned to either the D or C/c/E/e gene with certainty.

A 130-base product was amplified from the 3' region of cDNA transcribed from mRNA related to the Rh C,c and E,e gene, as shown in Fig 1B. This sequence corresponds precisely with published sequences and lies within the last three exons of the gene.10,11,13 The sequence for the Rh D gene differs by two bases in this region,4 and the amplified product can therefore be assigned to the C/c/E/e gene.

Southern analysis using the Rh 5' and 3' probes. An RFLP pattern that associated closely with C or c serologic phenotype was observed when genomic DNA was cut with the restriction enzyme Msp I and hybridized to the 5' probe in a Southern analysis. Figure 2A shows that the c gene in ccee and ccDDEE phenotypes gives two fragments of 2.5 and 1.1 kb, whereas the C gene in a CCDDDee phenotype gives two fragments of 2.6 and 1.1 kb. An Msp I restriction site lies within the sequence spanned by the 490-base probe (at nucleotide 27, Fig 1A), which explains the appearance of two bands for each gene. The probe also binds to two
fragments of 2.45 and 2.2 kb derived from the closely related Rh D gene, that are present in the Rh D-positive phenotypes but absent in the D-negative type. The 2.5-kb c and 2.45-kb D bands overlap and cannot be resolved into two discrete bands when both are present.

The 5' probe also identified an RFLP associated with C and c using Hae III-restricted DNA. The Rh C homozygous phenotype gave bands of 600, 320, and 175 bases, whereas c homozygous phenotypes show only the 600 and 175 bases. Cc heterozygous phenotypes showed the 600-, 320-, and 175-base bands; however, the ratio of the 320-band intensity to the 175 band was less than 1:1, whereas for CC types it was greater than 1:1 (not shown).

An RFLP associated with E or e was identified for Msp I-restricted DNA using the 3' probe. Figure 2B shows that this probe gave two fragments associated with the e gene of 2.8 and 1.85 kb (lanes with ccee and CCDDDee samples) and two bands for the E gene of 2.8 and 1.92 kb (ccDDee lane). No Msp I site lies within the cDNA probe sequence, and the two bands are attributed to an Msp I site lying within an intron between one of the last three exons. The 3' probe also identified a 2.0-kb fragment from the closely related D gene (Fig 2B).

The 5' and 3' probes were combined in subsequent hybridization procedures using DNA restricted with Msp I to produce the composite hybridization pattern shown in Fig 3. The number and type of Rh genes present was deduced by comparing the position and relative intensity of bands. For example, when the D gene is present, the genes associated with C or c can be identified by comparing the relative intensity of the C band with that of the c/D band. This is about 1:2 or less for one C gene (Fig 3, lanes 1 and 2) and 1:1 for two C genes (Fig 3, lane 5). The number of D genes can be determined by comparing the intensity of the 5' D gene fragment (middle D band, Fig 3) with the 5' constant 1.1-kb C/c band (bottom band, Fig 3) and by comparing the 3' D gene fragment (lowest D band, Fig 3) with the E/e gene fragments.

Concordance between RFLP patterns and serologic phenotypes. Rh DNA types for 102 randomly selected blood donors were determined from the bands obtained on Southern blots using the composite 5' and 3' probes. Concordances between genes assigned from RFLP patterns and the serologic phenotypes are shown in Table 1. Concordances were 100% for samples with C, D, and e serotypes. The c and E concordances were 94.8% (109/115) and 94.3% (33/35), respectively. Discordant results were obtained for eight samples. Six samples were discrepant because a restriction fragment was assigned to the C gene where the c antigen was detected serologically. Two discordant negative bands were an RFLP associated with the E gene where the E antigen was present. Where discordant results were found between the assigned RFLP pattern and observed serotype, the differences were clear. Figure 4 lane 1 shows a cEe serotype that showed an RFLP assigned to the ccee type; lane 3 shows a ccee serotype that has a Ccee RFLP type. Similarly, lanes 7 and 9 show two samples with Cc serotypes but with CC RFLP band types. In all cases, samples giving the discordant results were retested for both serology and RFLP patterns and gave the same result.

The concordance between the Hae III RFLP patterns observed with the 5' probe and C,c serotypes was also examined. Results were identical to the Msp I RFLP types, with the same six samples being discordant through an RFLP pattern, suggesting that the C gene was present where the c antigen was observed serologically.

This DNA typing by RFLP analysis identified two donors who each carried a modified Rh gene. One donor typed serologically as CcDee, but the RFLP pattern indicated that the 3' region for one C,c,E,e gene was altered (Fig 4, lane 5). The intensity of the e-associated fragment was reduced and a new fragment appeared (second bottom band indicated with an arrow, Fig 4, lane 5). The other donor exhibited a partial deletion for the D gene in the 3' region. The 5' probe
Table 1. Concordance Between Rh Genes Assigned by RFLP Analysis and Serologic Phenotypes for 102 Volunteer Blood Donors

<table>
<thead>
<tr>
<th>Rh Phenotypes (predicted genotype)†</th>
<th>Alleles Detected by Southern Blotting*</th>
</tr>
</thead>
<tbody>
<tr>
<td>CcDe (cDe/ce) n = 35</td>
<td>C  c  D  E  e</td>
</tr>
<tr>
<td>CcDe (cDe/ce) n = 20</td>
<td>38/35 32/35 38/35 — 68/70</td>
</tr>
<tr>
<td>ccee (ce/ce) n = 1</td>
<td>40/40 27/28 14/14 14/14 14/14</td>
</tr>
<tr>
<td>ccee (ce/ce) n = 1</td>
<td>1/0 27/28 0/0 — 28/28</td>
</tr>
<tr>
<td>CcDe (cDe/cDe) n = 14</td>
<td>14/14 14/14 28/28 14/14 14/14</td>
</tr>
<tr>
<td>CcDe (cDe/cDe) n = 3</td>
<td>14/14 14/14 28/28 14/14 14/14</td>
</tr>
<tr>
<td>ccDe (ce/ce) n = 1</td>
<td>0/0 6/6 5/6 5/6 1/0</td>
</tr>
<tr>
<td>ccDe (ce/ce) n = 1</td>
<td>1/0 1/2 1/1 — 2/2</td>
</tr>
<tr>
<td>Total n = 102</td>
<td></td>
</tr>
</tbody>
</table>

* Alleles detected compared with the number expected from Rh phenotyping.
† Rh phenotyping gives no indication of D gene dosage, but predicts most probable genotype based on gene frequencies in defined populations (here considered white).
‡ One donor carried an allele with a mutation and/or deletion at the 3' region of the C/e gene.
§ One donor carried a deletion in the 3' region of the D gene; 5' region appeared normal.

showed that two D gene fragments were present, but the 3' probe showed only one D gene (Fig 4, lane 10). Two further r"r" (ccEE) samples were tested in addition to the 102 random samples. Both gave an apparent RFLP type of ccee (identical to Fig 4, lane 1). Therefore, all five unambiguous cE (r") haplotypes defined by serology gave a ccee type RFLP. All 18 unambiguous cDE (R2) haplotypes typed as E by using the 3' probe RFLP analysis. The probability of this distribution occurring by chance is less than 0.00001. One other sample was serotyped as ccDEE (ie, probable R2R2), but analysis by RFLP showed that only one D gene was present and the RFLP type was ccDeE. This is consistent with a serotype of R2r" showing the same discordant RFLP result with the putative r" as observed in unambiguous cases.

**DISCUSSION**

This study identifies a close association between the 5' RFLP and the C/c polymorphism and between the 3' RFLP and the E/e polymorphism. By contrast, there was no association between the 3' RFLP and C/c (eg, 20 out of 20 CCDeE serotypes had the 1.85-kb form of the 3' RFLP, and 14 out of 14 ccee serotypes also had the 1.85-kb form of the 3' RFLP). Similarly, there was no association between the 5' RFLP and E/e (eg, three out of three ccDDeE showed the 2.5-kb form of the 5' RFLP, and 13 out of 14 ccee serotypes also had only the 2.5-kb form). These data are consistent with the sequences coding for the C/e epitopes lying on the 5' side of the sequences for the E/e epitopes. Le Van Kim et al 3 suggested that the E,e epitopes are encoded by the full-length mRNA sequence from the CE gene, and that the
C,c epitopes may reside on the shorter mRNA isoforms that result from differential intron/exon splicing of the gene. Localization of the C,c epitopes on the 5' side of the E,e is consistent with this suggestion, since the shorter mRNA isoforms carry deletions at the 3' region. It is also consistent with serologic studies that place D, C, and E along the Rh locus in that respective order.1

However, it appears unlikely that the C/c epitopes lie within the 5' probe region. Sequencing data shows a G-to-C polymorphism at base 48 in the coding region that is part of an Hae III site that also failed to have a perfect association with C/c polymorphism. Sequencing data (not shown) failed to find any other coding change on the genes coding for C or c. However, we can deduce that the published full-length cDNA clones10,11 are derived from the gene coding for the c epitotype, as c types have the Hae III site (and hence lack the 320 RFLP band).

Analysis of the results showed that where true Rh genotypes (ie, haplotype combinations) could be determined from a combination of serotypes and D gene dosage, four discrepant predictions based on serotyping occurred. Three CDE/cE predictions were in fact CDE/cDe, and one CDE/cDE prediction was in fact cDE/cE. This highlights the advantage of direct measurement of the D gene dosage over serologic techniques. Two different internal checks were used to measure the D gene dosage; first, by comparing the relative intensities of the D- and C/c-associated fragments identified by the 5' probe, and second, by comparing the D- and E/e-associated fragments identified by the 3' probe. The method assumes that two C,c/E,e genes are present, but by probing in different regions the method would generally prove valid for all but a complete deletion of the CE gene. Precise Rh D gene dosage measurements will have significant clinical value in antenatal counseling where, for example, it is desirable to know the Rh D zygosity state for the husband of a Rh D-negative mother who had been sensitized against Rh D from a previous pregnancy.

The D gene measurement using an internal check did identify one case with a partial deletion in the 3' region of the D gene. It is not known whether this affected the phenotypic expression, since the donor had an apparent CcDEe (R2,R2) serotype and a second D gene. All Rh D-negative samples examined here lacked the D gene (within the regions probed), which is consistent with the reported genetic basis of red blood cells typing as Rh D-negative.12 However, the partial 3' deletion detected may belong to a second, less common class of Rh D-negative types (C.A. Hyland, manuscript submitted). Alternatively, the modified D gene may be associated with one of the clinically important D variant types, and family studies could be informative in resolving these possibilities. The RFLP analysis also identified one donor with an unusual E/e fragment in the 3' region caused either by a mutation removing the Msp I site or by a partial deletioa. This donor carried C,c antigens but only e specificity, so it is not known whether the mutation affected the E/e phenotype. Again, family studies or quantitative studies on the level of e antigen expression on the red blood cells of this donor could be informative.

More detailed analysis of the results from randomly selected donors showed that both discordant results between E/e serotypes and the 3' probe could be associated with a CE serotype (ie, an r+ typing as ce). This suggested that a hitherto unrecognized variant of the Rh-negative haplotype exists. These observations were further investigated by examining two more cceE (r+/r−) samples. All four haplotypes typed as ccce by RFLP using the 3' probe. These show that the cE genes of the cE (r+) and cDE (R2) haplotypes are not identical (P < .00001).

In conclusion, the RFLP analysis identified polymorphisms associated with the C,c and E,e epitopes, respectively, determined the D gene dosage, and in two cases (r+ and R2) identified variants associated with Rh haplotypes. Neither of these latter two tests are possible using routine serologic methods on individual samples. In the future, these molecular biology-based techniques will supplement traditional Rh serotyping within transfusion medicine and will probably identify further heterogeneity in the Rh antigen complex.

REFERENCES
6. Bloy C, Blanchard D, Dah W, Beyreuther K, Salmon C, Cartron JP: Determination of the N-terminal sequence of human red cell Rh(D) polypeptide and demonstration that the Rh(D), (c), and (E) antigens are carried by distinct polypeptide chains. Blood 72: 661, 1988
7. Saboori AM, Smith BL, Agre P. Polymorphism in the M, M2, 32,000 Rh protein purified from Rh(D)-positive and -negative erythrocytes. Proc Natl Acad Sci USA 85:4042, 1988
8. Avent ND, Ridgwell K, Mawby WJ, Tanner MJA, Anstee DJ, Kampel B. Protein-sequence studies on Rh-related polypeptides suggest the presence of at least two groups of proteins which associate in the human red cell membrane. Biochem J 256:1043, 1988


A southern analysis of Rh blood group genes: association between restriction fragment length polymorphism patterns and Rh serotypes

CA Hyland, LC Wolter, YW Liew and A Saul