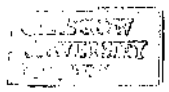
A MOLECULAR BIOLOGICAL INVESTIGATION OF THE KELL BLOOD GROUP SYSTEM

By MARGO TAYLOR MURPHY
BSc. (Hons) (Glasgow)

A thesis presented for the degree of Doctor of Philosophy at the University of Glasgow.

September 1995



ABSTRACT

Molecular biology techniques are becoming increasingly widely used in the field of transfusion science. The investigation of the Kell blood group system, using a selection of these techniques, is described in this thesis.

The first part of the investigation involved the synthesis of *KEL* cDNA from total RNA by reverse transcription and its subsequent amplification using the polymerase chain reaction (PCR). The resulting material was cloned into a plasmid and the cloned PCR products were then used for *in situ* hybridization studies to refine the location of the *KEL* gene on human chromosome 7 to 7q33-q35. This was a more precise location than had been previously published.

Kell null is a rare phenotype within the Kell blood group system, where none of the Kell system antigens can be detected on the red cell surface. Southern blot investigation of total genomic DNA from one such individual revealed the presence of a *KEL* gene in a similar form to that present in normal Kell phenotype donors. These results indicate that this example of the null phenotype did not arise as a result of a major gene deletion. This suggests that the development of Kell system antigens, in this particular case, is blocked during transcription or translation of the protein.

Direct sequence analysis of PCR products derived from individuals of predetermined Kell phenotype allowed the identification of a single C->T nucleotide substitution at position 701. This is responsible for the difference in expression of k (Cellano) and K (Kell) antigens on the red cell surface. This provided independent confirmation of the results published by another group of workers while this section of the work was in progress. The C->T nucleotide substitution creates a *Bsm* I restriction enzyme site when the sequence encoding the K antigen (the T nucleotide) is present. This enabled the development of a PCR-based genotyping assay for Kell (K1 or K) and Cellano (K2 or k). The assay involves the co-amplification from a total genomic DNA target of two PCR products, one which spans the K/k polymorphic site and a second which spans a conserved *Bsm* I site and which acts as an internal control for the assay. Several samples of KK, Kk and kk genomic DNA were analysed by this assay and definitive genotyping results were obtained for all samples. This should prove to be a useful tool for antenatal genotype determination in cases

where haemolytic disease of the newborn is implicated, and could be incorporated into a battery of PCR tests to establish a panel of typed donors.

The use of molecular biology techniques in the investigation of the Kell blood group system has allowed new information to be obtained regarding the molecular basis for the expression of antigens on the red cell surface. This should be a suitable foundation for the continuation of the work, allowing a more detailed investigation of the system as a whole, defining the basis for the other polymorphisms and enabling the development of diagnostic tests for the other allellic antigen sets.

DECLARATION

I hereby declare that this thesis has been composed by myself and has not been submitted previously to any other University. Unless otherwise stated, all work discussed was performed by the author. All sources of information are acknowledged by references and quotations distinguished by quotation marks.

Margo T. Murphy

ACKNOWLEDGEMENTS

Firstly, I would like to thank the Glasgow and West of Scotland Blood Transfusion Service for giving me the opportunity to undertake this project and for continued financial support. Thanks also go to Dr John Miles, who supervised the project at its outset, to my supervisors Dr John Goddard and Dr Robin Fraser, and to Dr Graham Templeton for their continued support and valuable advice throughout the project.

A special thanks goes to all of the donors who gave their blood willingly, especially the Kell null and KK donors who each made a very special effort and whose help was invaluable. Dr Myrtle Peterkin of the Glasgow and West of Scotland Blood Transfusion Service played a key role in the organisation of suitable donors for the project and Dr Callum Livingstone of the Biochemistry Department, University of Glasgow preformed numerous, painless venepunctures for which I am grateful.

I would like to acknowledge the useful collaboration which was established with Dr Elizabeth Boyd and Mrs Norma Morrison of the Duncan Guthrie Institute for Medical Genetics, Yorkhill, Glasgow and with Mr Gary Mallinson and Mr Peter Martin of the International Blood Group Reference Laboratory, Southmead, Bristol. These workers gave of their time and experience for my benefit and I appreciate everything I have learned as a result of these collaborations.

Special thanks go to Dr Andrew Munro for advice, discussion and guidance throughout the project and for constructive criticism of my thesis.

This thesis is dedicated to my parents.

CONTENTS

FIGURES AND TABLES

TABLES

ABBREVIATIONS

A	adenine
A_{260}	absorbance at 260 nanometres
AET	2-aminoethylisothiouronium bromide
AGPC	sodium acetate/guanidinium thiocyanate/phenol/chloroform
AMV	avian myeloblastosis virus
Asn	asparagine
ASPCR	allele specific polymerase chain reaction
C	cytosine
CCM	chemical cleavage of mismatches
CGD	chronic granulomatous disease
dATP	deoxyadenosine triphosphate
dCTP	deoxycytosine triphosphate
ddNTP	dideoxynucleotide triphosphate
DEPC	diethyl pyrocarbonate
dGTP	deoxyguanosine triphosphate
dITP	deoxyinosine triphosphate
dNTP	deoxynucleotide triphosphate
DNA	deoxyribonucleic acid
DTT	dithiothreitol
dTTP	deoxythymidine triphosphate
E. coli	*Escherichia coli*
FITC	fluoroisothiocyanate
Fuc	fucose
G	guanine
GalNAc	N-acetyl galactosamine
HDN	haemolytic disease of the newborn
IPTG	isopropylthio-β-D-galactoside
kb	kilobase
kDa	kiloDalton
MAIEA	monoclonal antibody-specific immobilisation of erythrocyte antigens
Met	methionine
MoMuLV	Moloney Murine Leukaemia Virus
MOPS	3-(N-morpholino) propanesulphonic acid
MPC	magnetic particle concentrator
mRNA	messenger ribonucleic acid

NaOAc	sodium acetate
OPC	oligonucleotide purification column
PEG	polyethylene glycol
PIP	prolactin inducible protein
poly A	polyadenylated
psi	pounds per square inch
RFLP	restriction fragment length polymorphism
Rh	Rhesus
RT-PCR	reverse transcription polymerase chain reaction
SDS PAGE	sodium dodecyl sulphate polyacrylamide gel electrophoresis
SSCP	single strand conformation polymorphism
T	thymine
Taq	*Thermus aquaticus*
TEAA	triethyl ammonium acetate
TEMED	tetramethylethylene diamine
Thr	threonine
TM	trademark
TNS	tri-isopropyl naphthalene sulphonic acid
X-gal	5-bromo-4-chloro-3-indolyl-β-D-galactoside

All other abbreviations used in the text are "accepted" abbreviations in accordance with the guidelines documented in the Biochemical Journal 305, 1-15(1995)

SECTION 1

Introduction to the application of molecular biology in transfusion science

SECTION 1 Introduction to the application of molecular biology in transfusion science

1.1 Introduction to the development of a membrane protein

Since the middle of this century, when DNA was identified as the genetic material and its structure determined, progress in molecular biology has provided a theoretical basis and methodologies for the understanding of blood group systems in terms of molecular genetics. The basic steps involved in production of red cell antigens is summarised in Figure 1.1. DNA encoding the proteins which carry blood group antigens is transcribed to RNA. This is then modified by capping of the 5' end, polyadenylation of the 3' end and removal of non-coding introns. The mRNA is then transported to the cell cytoplasm where it is translated to form protein which carries the antigens. This is then inserted into the red cell membrane. As a result, blood group antigens are expressed on the surface of the red cell. It is likely that the proteins carrying red cell antigenic determinants may associate together in the red cell membrane but precise details of the nature of these possible interactions remain to be described (Anstee, 1995).

1.2 Outline of molecular biology techniques used in transfusion science

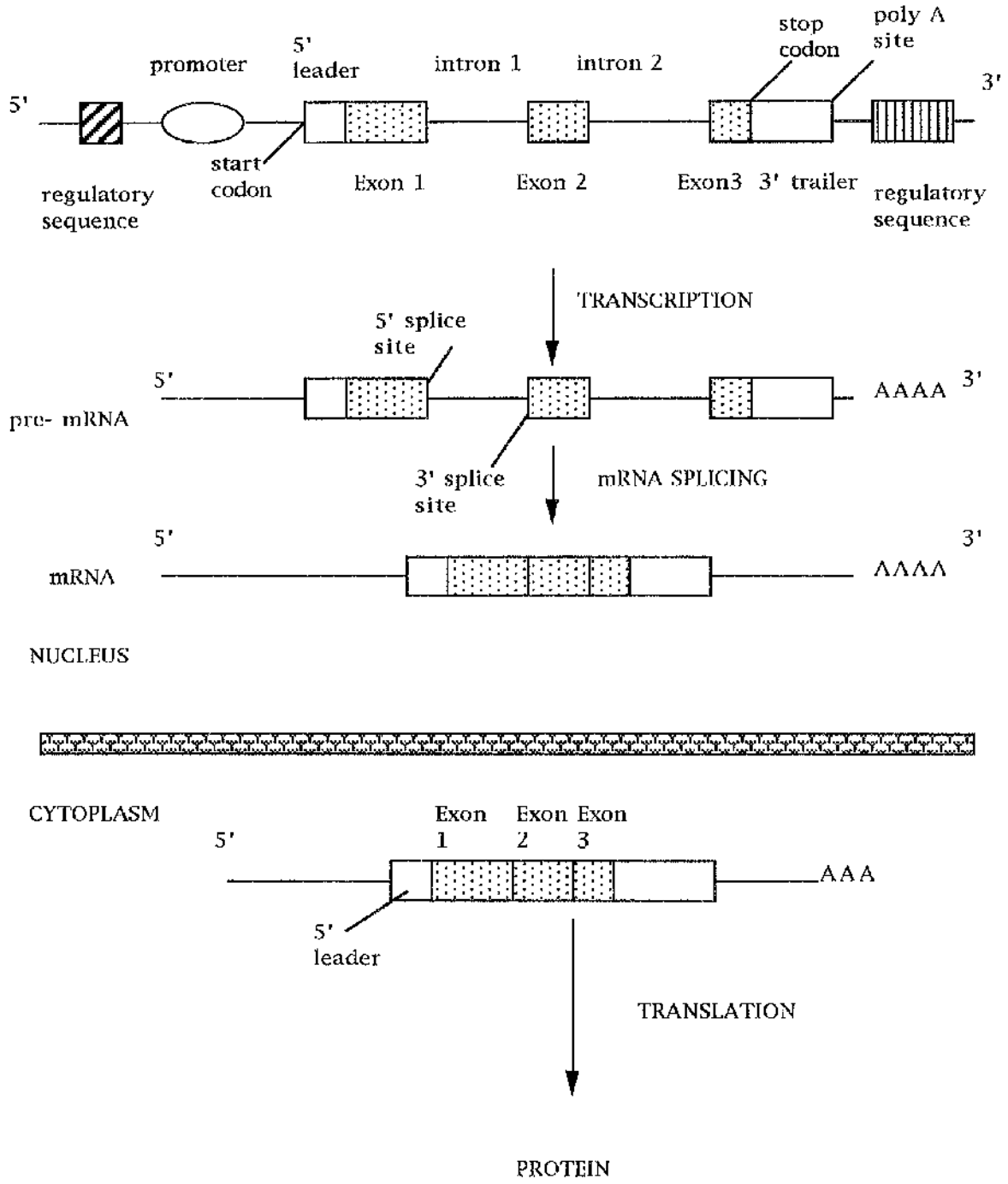
Investigation of the ABO, Rhesus, Kell, Kidd, Duffy and MNSs blood group systems using methods which relate to the genetic material is already well established. Several techniques will be described below and the the results which relate to individual blood group systems will be discussed in the subsequent subsections.

1.2.1 Use of complementary DNA libraries to isolate blood group cDNAs

One approach which can be adopted when locating a gene encoding a particular blood group system requires that the protein is purified and a partial amino acid sequence determined. A degenerate nucleic acid sequence can be derived from the amino acid sequence and used to design an oligonucleotide probe corresponding to this section of the gene. The oligonucleotide probe is used to screen a cDNA library for the presence of that particular sequence.

Figure 1.1 Diagram of the development of a membrane protein from the gene encoding it

The DNA located on a chromosome is termed genomic DNA. The DNA sequences encoding proteins are contained in exons and these are separated by non-coding introns. A leader sequence precedes the first exon. The entire sequence from leader sequence to the polyadenylation site are transcribed into pre-messenger RNA (pre-mRNA). A promoter region immediately upstream from the leader sequence, controlled by regulatory sequences located on either side of the structural gene, is required for the binding of polymerase enzymes which catalyse transcription. Introns are removed during the splicing of pre-mRNA into mature mRNA. This is then exported to the cytoplasm where the sequence is translated into protein. Translation begins at a specific start codon and ends at a stop codon. The leader sequence is translated but is later removed prior to the insertion of the protein into the cell membrane.


5'
promoter
5' leader
intron 1
intron 2
stop codon
poly A site
3'
regulatory sequence
start codon
Exon 1
Exon 2
Exon3
3' trailer
regulatory sequence
TRANSCRIPTION
5' splice site
5'
AAAA
3'
pre- mRNA
3' splice site
mRNA SPLICING
5'
3'
mRNA
AAAA
NUCLEUS
CYTOPLASM
Exon 1
Exon 2
Exon 3
5'
AAA
5' leader
TRANSLATION
PROTEIN

A cDNA library is constructed by isolating mRNA from a particlar source or cell type. The mRNA is then reverse transcribed using reverse transcriptase (a retroviral enzyme which is capable of reversing the normal progression of DNA to mRNA) and an oligo dT primer complementary to the polyadenylated 3' tail present on almost all eukaryotic mRNAs (section 2.5). This gives rise to RNA-DNA hybrids containing DNA which is complementary to the mRNA used as the target. This is representative of the entire polyadenylated RNA population. As the mRNA does not contain any introns, these cDNAs also lack intron sequences. The resulting cDNA is then cloned into vectors and used to transform bacterial cells for storage. This is referred to as a cDNA library. Human libraries can be produced in house (Sambrook *et al.*, 1989) or purchased commercially. If the human cDNA library is from a commercial source, then it is unlikely that the complete red cell phenotype of the donors cells will be known. This is likely to produce problems at a later stage in the investigation as the reference sequence obtained would be of unknown phenotype and would complicate the comparison of polymorphic sequences. The library is screened with the oligonucleotide probe which will only hybridize to the clones containing the complementary cDNA sequence. Hybridization is detected by autoradiography if the probe was radiolabelled. Alternatively, the positive clones can be identified by the incorporation of a fluorescent or coloured tag into the probe. The positive clones are then visualised. The probe should only hybridize to the complementary sequences in the library and this leads to the isolation of a specific gene. The DNA for the gene under investigation is then sequenced (section 1.2.4).

1.2.2 The polymerase chain reaction

Once the DNA sequence for any particular gene has been determined (section 1.2.4), two synthetic complementary oligonucleotide primers can be synthesised. These primers are short, specific strings of nucleotides (17 to 35 nucleotides in length) which are completely homologous to the cDNA sequence. One of the primers is specific for the 3' end of the region which is to be amplified and the other is specific for the sequence of the 5' end. The oligonucleotide primers anneal to opposite strands of the DNA. These can be used to amplify the sequence by the use of a technique known as the polymerase chain reaction (PCR) (Saiki *et al.*, 1985) (section 2.7). With the knowledge of the gene sequence and having synthesised appropriate

oligonucleotide primers, it is feasible to produce and investigate cDNA variants from chosen individuals. To do this, RNA is isolated from a cell sample and this is used as the target material for reverse transcription PCR (RT-PCR). This is a technique whereby total RNA is purified from the sample and a DNA copy is made using a reverse transcriptase enzyme (Varmus, 1987) (section 2.5). This forms a double stranded RNA/cDNA duplex molecule which is heat denatured to release single stranded cDNA. The resulting single stranded cDNA can then be amplified by PCR (Saiki *et al.*, 1985). In PCR DNA is amplified exponentially by successive rounds of denaturation, primer annealing and primer extension in the presence of a thermostable DNA polymerase enzyme. The cDNA, in this case, would be added to a mixture of free dNTPs, buffer containing magnesium ions, a heat stable DNA polymerase and the two specifically designed oligonucleotide primers. The reaction mixture is heated to approximately 94°C to separate the cDNA strand from the original mRNA template. It is then cooled to 50-65°C to allow the primers to anneal to the cDNA target. The next step is heating to 72°C (the optimum temperature for *Taq* DNA polymerase) to allow the extension of the primer to form a complementary strand. This completes the first round of the PCR and results in the production of a double stranded product which acts as the target in the next round. Since the products of each cycle act as the target for the next round, it then follows that each cycle represents a doubling in the number of product molecules. Therefore, 20 rounds of PCR represents an increase in the number of specific sequences by 10^6 (2^{20}) assuming 100% efficiency (Erlich, 1989). This method can be adapted to give the co-amplification of more than one product by the incorporation of more than two PCR primers.

1.2.3 Cloning of PCR products for use as probes in *in situ* hybridization

Subsequent to amplification by PCR, the cDNA is cloned in a vector such as a bacterial plasmid (Sambrook *et al.*, 1989). The vector is circular DNA and must be cut to form a gap at which a cDNA fragment can be inserted. This is accomplished by the use of one or more restriction enzymes, which recognize short, specific sequences on the vector and will cut the DNA only at these points (section 2.12-2.14). The amplified cDNA is then cut with the same restriction enzymes before it is added into the plasmid. The plasmid and amplified fragment are joined together by the use of a DNA ligase enzyme. This joins the complementary ends of the cut plasmid and the insert to re-form a circular

recombinant DNA containing the insert. The recombinant plasmid clone is then used to transform a strain of *E. coli* cells. The recombinant replicates and is reproduced as the bacterial cells grow. The transformants can be stored long term in glycerol at low temperature (-70°C). The cloned fragment can be isolated from the bacterial cells when required by carrying out a plasmid preparation (Sambrook *et al.*, 1989) (section 2.19). The isolated plasmid clone may be used to probe normal human chromosome spreads in order to identify the chromosomal location of the corresponding gene (section 2.21). This can be achieved by using methods such as nick translation where the probe is labelled with either a radioisotope or with biotin (Rigby *et al.*, 1977). In this process, an enzyme is used to introduce a single stranded nick in the double stranded plasmid DNA. Labelled dNTPs are then introduced into the reaction with a polymerase enzyme which incorporates labelled nucleotides in order to repair the nick. Thus, the probe becomes labelled. This probe can then be used for *in situ* hybridization where it binds specifically to normal male chromosome spreads (Cox *et al.*, 1984). Male chromosome spreads are used as both X and Y chromosomes are present. This is an assay where a probe is labelled with biotin and biotin/FITC avidin is added as a means of visualising the results. The detection process is dependent on the presence of a fluorescent dot on the chromosome at the point of hybridization of the probe. Under optimised conditions, the probe will only hybridize to the specific area on the chromosome where the exact complementary sequence is present. Bearing in mind that sequences tends to be highly individual for each gene, it follows that the precise location of the gene on a particular chromosome can be defined. Using this technique and others such as linkage analysis, the chromosomal locations of several of the known blood group genes have been identified to date (sections 1.4 to 1.7).

1.2.4 Sanger dideoxy sequencing

Nucleotide sequences of cloned PCR amplified fragments can be determined. Comparison of several sequences from samples derived from different, known phenotype individuals may allow the identification of the nucleotide differences which account for the expression of particular antigens within a blood group system. Cloned cDNA can be isolated from the *E. coli* cells, into which the plasmid clone has been transformed (section 1.2.3), in single stranded form by the use of an M13 helper phage. This allows preferential secretion of single stranded vector DNA over the phage single stranded DNA

when the plasmid incorporates the appropriate phage regions which allow its recognition by the M13. This helper phage is M13 mutated in its own replicative recognition sequence and, as a result, it is unable to reproduce itself efficiently. The strand which is recovered is dependent on which of the two plasmid strands the M13 recognition region is placed (Sambrook *et al.*, 1989). The single stranded DNA can then be sequenced using the Sanger dideoxy method (Sanger *et al.*, 1977) (section 2.20). Dideoxy sequencing involves annealing of a sequencing primer to the DNA template and extension of the primer by a DNA polymerase. Synthesis is initiated at the 3' end of the oligonucleotide primer. The synthesis reaction is terminated by the incorporation of a nucleotide analogue that will not support continued DNA elongation. The chain terminating nucleotide analogues are dideoxynucleoside triphosphates (ddNTPs) which lack the 3'-OH group necessary for chain elongation. When mixtures of dNTPs and one of the four ddNTPs are used, enzyme catalysed polymerisation will be terminated in a fraction of the population of growing chains at each site where the ddNTP is incorporated. Four separate reactions, each with a different ddNTP, give complete sequence information. A radioactively labelled dNTP is also included in the synthesis reaction. The labelled chains of various length are visualised by autoradiography after separation by high resolution polyacrylamide gel electrophoresis. Several alterations have been made to the sequencing technique to allow it more flexibility. For example, a protocol has been developed to allow the direct sequencing of PCR products without the necessity for cloning (Erlich, 1991). In this situation, one of the PCR primers is added in great excess to the second primer and, as a result, single stranded DNA is amplified in a non-logarithmic fashion. This single stranded DNA can be used directly for sequencing using the dideoxy method. This is a useful alternative technique for two reasons. Firstly, it eliminates the necessity for cloning each PCR fragment before sequencing and secondly, it allows sequencing of a population of PCR products. Therefore, *Taq* DNA polymerase misincorporations can be distinguished from true sequence differences. This is an important point as *Taq* DNA polymerase misincorporations occur at a rate of 1 in 400 nucleotides during PCR (Saiki *et al.*, 1988, Paabo *et al.*, 1988, Dunning *et al.*, 1988). A misincorporation is where the *Taq* DNA polymerase inserts the "wrong" nucleotide at points along the DNA, since it does not have a 3' -> 5' proof reading facility, unlike certain other polymerases such as Klenow fragment and T4 DNA polymerase. Determination of the sequence of a cloned PCR product is liable to give errors, as each clone represents a single original PCR product which may have an incorrect nucleotide at 1/400 positions. Direct sequencing of a mixture

of PCR products negates the possibility of reading an incorrect sequence, since misincorporations are random and likely to be present in the same position in only a small proportion of the total PCR amplified molecules.

A second useful modification to traditional dideoxy sequencing is the use of automated sequencing (sections 2.20.6 and 6). In this method, each of the four ddNTPs has a different dye label. All four reactions are carried out in the one tube and the gel is scanned using a laser. The dye labels can be identified by the laser and the sequence is automatically read. This technique eliminates the necessity for the use of radioisotopes and, under suitable conditions, allows a greater length of sequence to be determined. However, it is expensive and requires high quality DNA to obtain good sequence.

1.2.5 Restriction fragment length polymorphism analysis

Another method for the identification of genetic differences between individuals is a technique known as restriction fragment length polymorphism (RFLP) analysis (Botstein *et al.*, 1980). This initially involved the analysis of total genomic DNA, but has been adapted for use with PCR products. The most common application of this technique involves the amplification of small stretches of genomic DNA by the polymerase chain reaction. The resulting products are then cut into smaller fragments by digestion with a restriction enzyme specific for the sequence of interest. This may be one which cuts in a region encoding a single red cell antigen. The digested material gives a series of smaller fragments which can be identified by gel electrophoresis. The resulting pattern of restriction fragments may be diagnostic for a particular phenotype e.g. that of a red cell antigen of interest. Therefore, the blood donor genotype may be determined by analysis of the banding pattern obtained by RFLP analysis (section 7). This technique can also be carried out directly on genomic DNA, but the use of PCR products gives clearer results, particularly when using RFLP analysis as a diagnostic test in a clinical situation. The use of this method has been described for the investigation of several blood group systems (Zelinski *et al.*, 1991 c).

1.2.6 Southern blotting using total genomic DNA

This method allows the comparison of genes from normal and null or variant phenotypes. It gives comparative information regarding similarities and

differences between genomic DNA derived from such samples. Southern blotting can be used to determine the presence and the molecular size of fragments of a specific gene, and also allows genetic characterisation (Southern, 1975). This method involves the digestion of total genomic DNA with various restriction enzymes. The digested DNA is then resolved by gel electrophoresis and blotted onto a nylon or nitrocellulose membrane. A DNA probe or a PCR product containing part or all of the gene from a known phenotype is then radiolabelled and used to hybridize to the blot. The blot is washed and exposed to X-ray film, giving rise to a series of bands on the film. If major differences in banding pattern are present, this may provide an indication that there are differences in the sequences of the two genes. They may be as little as one or two nucleotides which alter a restriction enzyme site, or could indicate major differences in long stretches of the sequence. If only a small number of nucleotides differ between the probe and target sequences, and these do not alter the restriction enzyme sites, then detection by Southern blotting becomes more difficult. Further investigation must then be carried out and sequencing of the cDNA is essential in order to identify these variations.

1.3 Introduction to molecular biological research on blood group systems

History of research into blood groups spans approximately 90 years and consists mainly of the results of serological techniques, such as haemagglutination, which were used to define blood group antigens by their complementary antibodies. Family studies were also used to learn about the inheritance of such antigens. Red cell antigens are structures exposed on the red cell surface which, under certain circumstances, can provoke the production of a specific circulating antibody in an individual's bloodstream. 191 of these structures are known to comprise 23 recognised blood group systems (Anstee, 1995).

Several blood group systems are currently under investigation using molecular biology techniques. The results from these investigations have revealed important information about the ABO, Rhesus, Kell, Kidd, Duffy and MNSs blood group systems. These findings include information regarding gene sequence and the resulting protein structure, the chromosomal location of the gene and the genetic basis for the presence or absence of particular antigens on the red cell surface. The red cell phenotype can be predicted by use of PCR and associated techniques and this should prove a useful tool to ensure provision of

compatible blood in cases where a rare phenotype donor has multiple clinically significant antibodies present in the circulation. This approach would be slower than traditional serological methods, but it would be useful where reagents are difficult to obtain.

Prior to a detailed discussion on the Kell blood group system, a brief review of important aspects of ABO, Rh, MNSs and some minor blood group systems is presented.

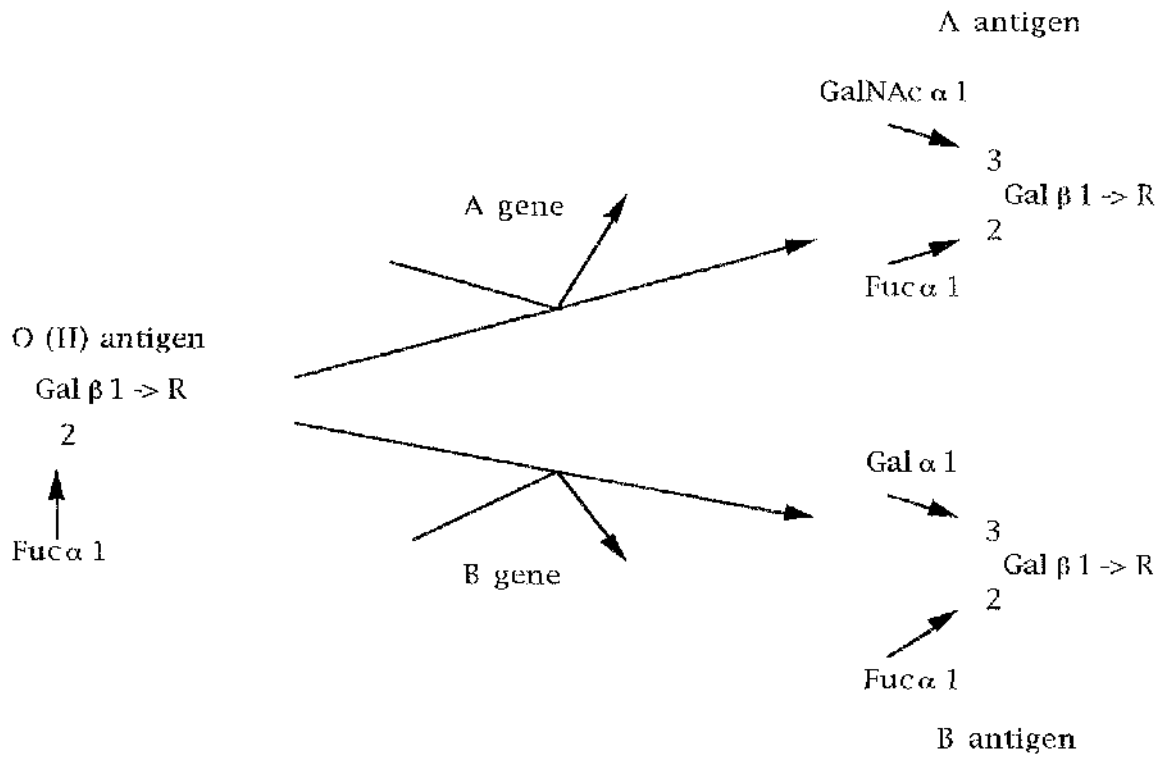
1.4 The ABO blood group system

The ABO blood group system, discovered by Landsteiner in 1900, was the first alloantigen system recognized on erythrocytes and is the most important system in transfusion medicine (Landsteiner, 1900). Between 1950 and 1970 the antigens were characterised chemically and were shown to be oligosaccharide in nature (Figure 1.2) (reviewed in Watkins, 1980). When the biochemical pathways for the formation of these antigens were defined it became clear that A and B antigens are derived from a precursor, H substance, produced by the attachment of a common L-fucose residue to a carbohydrate chain with N-acetylgalactosamine for the A antigen and galactose for the B antigen. In both antigens, an α-1-2-fucosyltransferase adds a fucose residue to the β-galactose of a preformed glycoprotein chain (Clausen et al., 1989). Individuals who are of group A express α1->3 N-acetylgalactosaminyltransferase and group B individuals express α1->3 galactosyltransferase. Individuals who group as AB express both transferases and group O individuals express neither (i.e. the H substance remains unchanged). A and B transferases have overlapping substrate specificities.

Some of the original molecular biology work on the ABO blood group system was carried out by Yamamoto *et al.* (Yamamoto *et al.*, 1990 a). This work was the first evidence of the molecular basis for the ABO phenotypes. The isolation of a soluble form of the A transferase from human lung and intestinal mucosa was reported (Clausen *et al.*, 1990, Yamamoto *et al.*, 1990 b). A partial amino acid sequence was determined and this facilitated the design of oligonucleotide primers which were used for RT-PCR. The resulting cDNA, which encodes the A transferase, was cloned and sequenced (Yamamoto *et al.*, 1990 a). The gene was cloned from cDNA libraries from individuals of known phenotype and nucleotide substitutions which determine the differences between A, B and O

Figure 1.2 Biochemical relationship between A, B and H antigens

Diagram illustrating the addition of sugar molecules to a precursor glycoprotein (H substance) to allow the expression of A or B blood group antigens. The group A glycosyltransferase gene product catalyses the addition of N-acetylgalactosamine and the B gene product catalyses the addition of galactose to position 3 of the β-1-galactose molecule on the H substance.


A antigen
GalNAc α 1
3
Gal β 1 -> R
2
Fuc α 1
A gene
O (II) antigen
Gal β 1 -> R
2
Fuc α 1
Gal α 1
3
Gal β 1 -> R
2
Fuc α 1
B gene
B antigen

phenotypes were identified (Yamamoto *et al.*, 1990 a). The difference between the A and B phenotype can be accounted for by several nucleotide differences which give rise to four consistent amino acid substitutions at residues 176, 235, 266 and 268. The H gene has also been isolated (Ernst *et al.*, 1989).

Southern blot analysis of genomic DNA (section 1.2.6) allowed the identification of a single nucleotide deletion in the coding region at amino acid residue 258 which gives rise to a shift in reading frame. This deletion accounts for the expression of the O phenotype and is likely to lead to translation of an enzymatically inactive protein (Yamamoto *et al.*, 1990 a).

Three of the four nucleotide substitutions and the deletion at amino acid residue 258 generated allele specific restriction enzyme sites (Yamamoto *et al.*, 1990 a). These sites were used as a diagnostic test for *ABO* genotype determination using the RFLP technique (section 1.2.5) as particular enzymes are diagnostic for only one individual genotype. These enzymes will only cut the PCR products where the specific enzyme site exists and this is directly dependent on the donor genotype.

The existence of subgroups within the ABO system raises the question of the basis for these differences in light of the results which arise from the genotyping experiments. It has been suggested that the A_1 and A_2 subgroups, identified by the use of either antibodies or lectins (Race *et al.*, 1975), arise as a result of the expression of α1->3 N-acetylgalactosaminyltransferases with different kinetic properties (Yamamoto et *al.*, 1990 b). This would result in quantitative differences in the A carbohydrate antigens (Engelmann *et al.*, 1993). More recently, the nucleotide sequences which account for the A^2 allele (Yamamoto *et al.*, 1992) and some of the A^3 and B^3 alleles (Yamamoto *et al.*, 1993 a) have been determined. Characterisation of one A^x and one $B^{(A)}$ allele have been described by analysis of their genomic DNA using PCR, subcloning and DNA sequencing (Yamamoto *et al.*, 1993 a). A_x individuals do not secrete A substance in their saliva and they exhibit very low levels of α1->3 N-acetylgalactosaminyltransferase activity. In group $B^{(A)}$ individuals, the B transferase has the ability to transfer some GalNAc as well as galactose to the H substance and, as a result, some A antigen also appears on the cell surface.

The report of Yamamoto *et al.* outlines the nucleotide differences which account for the expression of these particular phenotypes (Yamamoto *et al.*, 1993 a). The

A_X phenotype appears to arise as a result of a single nucleotide change in the A^1 nucleotide sequence at position 646 (T->A). The $B^{(A)}$ phenotype can be accounted for by two changes in the *B* nucleotide sequence at positions 657 and 703 (T->C and A->G respectively). The A_{703}->G mutation results in a serine -> glycine substitution at amino acid residue 235 and this coincides with the second of four amino acid substitutions which discriminate A and B transferases. The glycine residue in the $B^{(A)}$ individual is identical to the residue present at that position in group A individuals. Other forms of the *O* allele have also been reported (Yamamoto *et al.*, 1993 b)

An allele specific PCR (ASPCR) method has been described which can be used directly for the *ABO* genotyping of individuals (Ugozzoli *et al.*, 1992). The PCR oligonucleotide primer design is crucial for the success of this technique. Four sets of oligonucleotide primers, each of which is specific for a different set of *ABO* alleles, are incorporated into a single PCR reaction and the resulting products are analysed by gel electrophoresis. This allows PCR to be carried out in an allele specific manner and has been reported to be 100% reliable (Wu *et al.*, 1989). The ASPCR technique is dependent on the PCR primers being an exact match to the target sequence at their 3' end. A single nucleotide mismatch at that point will prevent extension of the primers and, as a result, will prevent the amplification of PCR products. It is essential that each primer pair used is designed to generate a product that differs sufficiently in size from the other products simultaneously generated, such that all can be separated clearly by gel electrophoresis. The advantage of using this technique is that allelic genotypes can be determined in a single reaction.

The application of modern molecular biology techniques has led to major advances in the comprehension and diagnosis of genetic variation in the ABO blood group system. The ABO system is an excellent example of the progress feasible with these methods for understanding the molecular genetic basis of other blood group systems.

1.5 The Rhesus blood group system

More molecular biology studies have been performed on the Rhesus blood group system than on any other. This is not surprising, as the Rhesus system is one of the most complex blood group systems known. It comprises at least 47 distinct antigens (Anstee *et al.*, 1994), and the antigens C, c E, e and D are

polymorphic. The Rhesus D antigen was discovered more than 50 years ago (Levine *et al.*, 1939). This system is extremely important in transfusion medicine, as the associated antibodies can be responsible for severe cases of haemolytic disease of the newborn and can cause fatal transfusion reactions.

1.5.1 Genetic analysis of *Rhesus* genes

Genetic analyses have demonstrated that the D, C/c and E/e antigens are inherited as a single entity, but it was initially unclear whether these antigens were the product of a single locus, or were encoded by genes at several sub-loci. The *Rh* locus has been assigned to the short arm of chromosome 1 (McGuire *et al.*, 1988). A more precise location of 1p34.3-1p36.1 was determined by the use of *in situ* hybridization with metaphase chromosomes (Cherif-Zahar *et al.*, 1991). In 1990, Avent *et al.* reported the cloning of the cDNA encoding a 30 kDa erythrocyte membrane protein associated with Rhesus blood group antigen expression (Avent *et al.*, 1990).

1.5.2 Rhesus polypeptides

The structure and sequence of the protein carrying Rhesus antigens was poorly understood before the cDNA encoding the 30 kDa Rhesus polypeptide was cloned. It had been suggested that the Rhesus polypeptide might be involved in maintaining the asymmetric distribution of phospholipids in the erythrocyte membrane. Immunoprecipitation studies had identified this unglycosylated 30 kDa membrane polypeptide as a Rhesus antigen carrying protein (Moore *et al.*, 1982, Gahmberg, 1982 and 1983). The Rhesus antigens are known to be sulphydryl containing polypeptides (Green, 1967 and 1983). Anti-c and anti-E antibodies have been used to immunoprecipitate related proteins (Moore *et al.*, 1982). Polypeptides which carry the Rhesus D and C/c, E/e antigens were used to establish a partial amino acid sequence by a protein sequencing technique (Bloy *et al.*, 1988). cDNA library methods were then used, as described in section 1.2.1, to isolate clones containing *Rhesus* related sequences. When Rhesus polypeptides were precipitated individually, using anti-c, anti-D or anti-E it was found that, although the polypeptides are highly related, each is a distinct protein (Blanchard *et al.*, 1988).

1.5.3 Molecular biological analysis of the Rhesus blood group system

The molecular cloning of the human *Rhesus* gene (leading to the derivation of the protein sequence of the human Rhesus polypeptide) was reported in 1990 (Cherif-Zahar *et al.*, 1990). Again, the cloning work involved the screening of cDNA libraries, using PCR products derived from the 5'-terminal end of the *Rhesus* gene as probes to identify the *Rhesus* cDNA in the library. This encodes a conserved region in the Rhesus proteins. These workers reported the size of the entire *Rhesus* structural gene as being 1384 nucleotides. Translation of the open reading frame indicated that the Rhesus protein is composed of 416 amino acids, with a predicted molecular mass of the protein as 45.5 kDa. SDS PAGE analysis of the protein predicts only 30 kDa to 32 kDa. These results suggest that either the Rhesus protein behaves in a peculiar manner on SDS PAGE, or that the mRNA encodes a precursor substance which is subsequently cleaved at the C-terminal end (the extracellular portion of the protein).

Three separate Rhesus polypeptides can be expressed on any one cell. One carries the D antigen and the other two carry the C/c and E/e antigens (Umenishi *et al.*, 1994). The sequence of these three polypeptides is identical up to the 41st amino acid residue, after which they start to diverge. The Rhesus C, D and E polypeptides are closely related but distinct proteins and it is likely that two separate, but closely linked, loci exist which encode the C/c, E/e and D antigens (Le Van Kim *et al.*, 1992 a). RFLP analysis of the genomic DNA from Rh D positive and Rh D negative individuals indicated that Rh D positive individuals have two Rh polypeptide genes whereas the Rh D negative individuals have only one gene (Mouro *et al.*, 1993, Umenishi *et al.*, 1994 b). These results suggest that the Rhesus D negative phenotype may arise as a result of a *RhD* gene deletion, as had been indicated by previous work (Colin *et al.*, 1991). This reinforces the idea that two separate loci exist within the Rhesus system. One and two dimensional peptide maps of the Rhesus polypeptides show that they are structurally distinct (Bloy *et al.*, 1988, Suyama *et al.*, 1988, Blanchard *et al.*, 1988). Southern blot analyses have shown that the *Rh* locus carried in the genome of RhD positive individuals is composed of two strongly related genes, whereas similar investigation of RhD negative individuals reveals the presence of only one of these genes (Colin *et al.*, 1991). These results support the peptide mapping and RFLP analysis results and agree with the hypothesis that Rhesus D antigen is coded for by one gene and Rhesus C/c and E/e antigens are coded for

by a second gene (Bloy *et al.*, 1988, Suyama *et al.*, 1988, Blanchard *et al.*, 1988). Alternative splicing then gives rise to the synthesis of the polypeptides already described. The report of Umenishi *et al.* shows the presence of bands on Southern blots of genomic DNA digested with various restriction enzymes and probed with fragments specific for the 3' and 5' ends of the *Rh* sequence (Umenishi *et al.*, 1994 a). Their experiments demonstrated the presence of a fragment for the Rh c phenotype, irrespective of the RhD status, a new development in the investigation of the *C/c* sequences. Further evidence was published simultaneously which showed sequence differences in the genes encoding the C and the c antigens (Hyland *et al.*, 1994 a). These authors also reported that the gene sequences in individuals who are c+, E+ are different in RhD positive and RhD negative samples.

The gene encoding a 50 kDa membrane glycoprotein which is associated with Rhesus blood group antigen expression has been isolated and cloned (Ridgwell *et al.*, 1992). This protein coprecipitates with the 30 kDa unglycosylated Rhesus protein on immunprecipitation with anti-Rh antibodies. The gene was mapped to 6p21-qter, a completely different chromosomal location to that determined for the gene encoding the 30 kDa protein. This position has been defined as the *Rh50* glycoprotein locus. That is, this locus encodes a polypeptide of predicted molecular mass 50 kDa. This polypeptide has been shown to be present on Rh null U+ cells (where U is an antigen of the MNSs blood group system, unrelated to Rhesus) and, therefore, does not account for the polymorphic variation of the Rhesus system, although it is a related polypeptide. These results indicate that two separate genes encode the 30 kDa and 50 kDa Rhesus related polypeptides. This 50 kDa protein is glycosylated, unlike the 30 kDa polypeptide (Moore *et al.*, 1982). It is not yet clear whether the expression of Rh antigens requires the association of the 50 and 30 kDa polypeptides or whether the 30 kDa polypeptide can give rise to antigen expression independently.

The isolation of three *Rh* related cDNAs from a human bone marrow library has been reported (Le Van Kim *et al.*, 1992 a). The authors propose the existence of Rhesus protein isoforms which differ from those whose genes had been cloned previously. The differences include peptide deletions and the generation of a new C-terminal polypeptide caused by frameshift mutations. All of the isoforms which had been described until this point arose from differential splicing of transcripts from the same *Rh* gene. The results reported by this group support the idea that two genes reside at the *Rhesus* locus and that these

direct the synthesis of three protein species which may relate to different Rhesus antigenic variants (Figure 1.3).

1.5.4 Rhesus null phenotypes

RFLP analysis suggests that the two samples of amorph type Rhesus null blood (where the phenotype arises as result of both parents being either homozygous or heterozygous for the Rhesus null condition) arose as a result of two independent mutations (Carritt *et al.*, 1993). This would account for the extremely low incidence of this phenotype, as these mutations would have to occur on both alleles for the phenotype to manifest itself. This suggests that the Rh null phenotype does not arise as a result of a gross gene deletion and also shows that this phenotype can have different Rhesus genetic backgrounds. The second, and more common, type of Rh null phenotype, is the "regulator" type, which arises as a result of the presence of a double dose of a repressor gene ($X^{O}r$) at an autosomal locus unlinked to the *Rh* locus. A recent report which has investigated the genes of both Rh null types suggests that Rh null regulator types are unable to express Rh antigens, but can pass on functional *Rh* genes to the next generation (Cherif-Zahar *et al.*, 1993). The Rh null regulator phenotype can apparently arise in DD, Dd or dd individuals and the repression of antigen expression is therefore independent of the *RhD* genotype. This was confirmed by Southern blot analysis, which showed identical banding pattern between normal and Rh null regulator individuals (Cartron *et al.*, 1993).

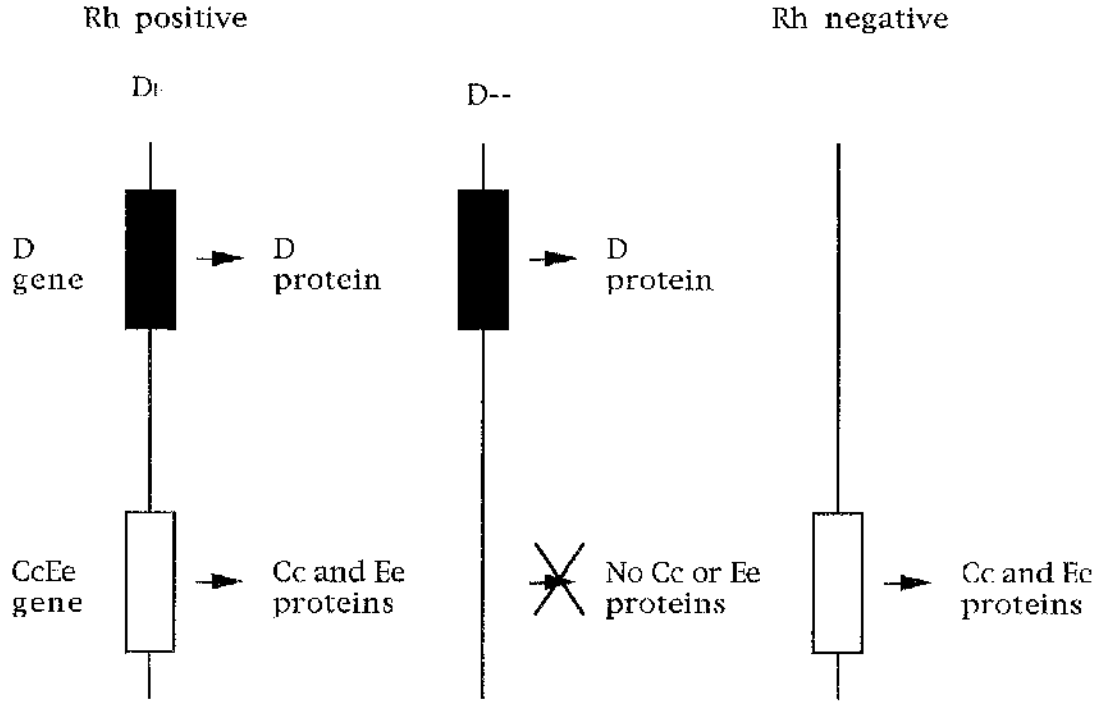
Similarly, results suggest that, in the Rh null amorph types investigated, the phenotype did not arise as a result of a major deletion of the *Rh* locus. Cloning and sequencing of *in vitro* amplified products also showed that the amorph Rh null does not arise through gene rearrangement or point mutation. One explanation may be that a transcriptional inhibitor blocks the expression of the Rh polypeptides.

1.5.5 Expression of Rhesus polypeptides *in vitro* and *in vivo*

Suyama *et al.* described the *in vitro* expression of the 32 kDa Rhesus polypeptide (Suyama *et al.*, 1993). This was achieved by inserting previously cloned *Rh* cDNA into a pGEM vector. This was transcribed/translated in a reticulocyte lysate system and resulted in the production of proteins of 30 kDa (minor) and 32 kDa (major) as determined by SDS PAGE. Expression was also

Figure 1.3 The arrangement of the *Rhesus* genes

Schematic diagram of the structure of the *Rh* locus. Two *Rh* genes (one which encodes the D protein and the second which encodes the CcEe proteins) are present in Rhesus positive individuals. Only one gene, encoding the CcEe proteins, is present in Rhesus negative individuals.


Rh positive
Rh negative
D+
D--
D gene
D protein
D protein
CcEe gene
Cc and Ee proteins
No Cc or Ee proteins
Cc and Ee proteins

1p34-36

achieved successfully in a Chinese hamster ovary cell system. Immunoprecipitation experiments using anti-Rh specific rabbit antibodies also confirmed the Rhesus specificity of these polypeptides.

Northern blot analysis, which can be used to detect the expression of *Rhesus* genes in specific tissues, has shown that expression is likely to be restricted to cells of erythroid lineage (Cherif-Zahar *et al.*, 1990). This may be a useful tool for the investigation of the biosynthesis of the Rhesus antigens during the development of the red cells.

1.5.6 *Rhesus* genotyping

Genotyping, with respect to the Rhesus blood group system, has important implications in the field of transfusion science. The determination of the foetal genotype is critical in cases where HDN may arise. Therefore, the development of a PCR-based genotyping test which can utilize small cell numbers, such as chorionic-villus samples or amniotic fluid samples, is a major step forward. This type of assay was described in 1993 (Bennet *et al.*, 1993). The Bennet method was used to determine the foetal genotype and, as a result, positive decisions could be made regarding the possibility of further immunisation in previously sensitised mothers. Clinical decisions regarding induction of birth could also be taken based on the results obtained. As described above, the Rh blood group antigens are carried by at least three homologous, but distinct, membrane proteins. Two of the proteins have serologically distinguishable isoforms, that is C, c and E, e. The third protein, which carries the D antigen mosaic, has no defined isoform. The *Rhesus* gene locus resides on chromosome 1 at position 1p34-p36 and this consists of two adjacent homologous structural genes designated *RhCcEe* and *RhD* (Colin *et al.*, 1991). The sequences of the coding regions of these two genes are 96% identical, suggesting that these may have arisen through the duplication of a single ancestral gene. The *RhCcEe* gene encodes the C/c and E/e proteins and the isoforms are most likely to arise as a result of alternative splicing of a primary transcript. *E* and *e* alleles vary by only one nucleotide, whereas the *C* and *c* alleles vary by as many as six nucleotides which result in four amino acid substitutions (Mouro *et al.*, 1993, Simsek *et al.*, 1994 a). The *RhD* gene encodes the D protein and is absent on both chromosomes of RhD negative individuals. PCR assays have been described which co-amplify a 136 base pair fragment common to both the *RhCcEe* and the *RhD* genes, and a 186 base pair fragment which is specific for the 3'

untranslated region of the *RhD* gene (Bennet *et al.*, 1993). Gel analysis of the PCR products would therefore reveal the presence or absence of the *RhD* gene depending on the presence of double or single bands on the gel. The report suggests this assay is extremely fast, reliable and reproducible. Contamination of the sample by the RhD negative mother's cells would not pose a problem in this type of test, as only the foetal cells could give rise to a PCR product specific for the *RhD* gene. However, contamination of the sample by the individual undertaking the procedure must be avoided at all cost, as this could give rise to a false positive result due to the sensitivity of the PCR technique. Appropriate controls must be included to ensure that false negative results are not obtained as a result of failure of the PCR itself. One report has been published which details problems in the application of the Bennet PCR method for *Rh* genotyping (Simsek *et al.*, 1994 b). Other groups published work simultaneously which detail similar PCR based diagnostic tests for *Rhesus* genotype determination (Wolter *et al.*, 1993, Lo *et al.*, 1994, Simsek *et al.*, 1995).

1.5.7 Organisation of *Rhesus* genes

A report published in 1993 describes the amplification of *Rhesus* cDNAs from cultured progenitor erythroid cells by PCR (Kajii *et al.*, 1993 b). These PCR products were then sequenced directly. Two Rhesus polypeptide cDNAs were isolated and the sequences of these were determined. One differs from the published sequence by a single nucleotide substitution, whereas the second contains 41 nucleotide substitutions resulting in 31 amino acid changes. This illustrates the degree of polymorphism which exists within the Rhesus blood group system.

The structural organization of the *CcEe* gene has recently been established. This consists of 10 exons spread over 75 kb of DNA. It is likely that the transcripts of the *CcEe* gene are processed by alternative splicing of exons 4-8 to produce the isoforms already described (Cherif-Zahar *et al.*, 1994 a).

Clearly, a great deal of information has come to light as a result of the molecular biological investigation of the Rhesus blood group system. Some of these findings have enabled the development of diagnostic assays for *Rhesus* genotyping which are superior to the traditional serological techniques available. The methods already described in this section have major clinical

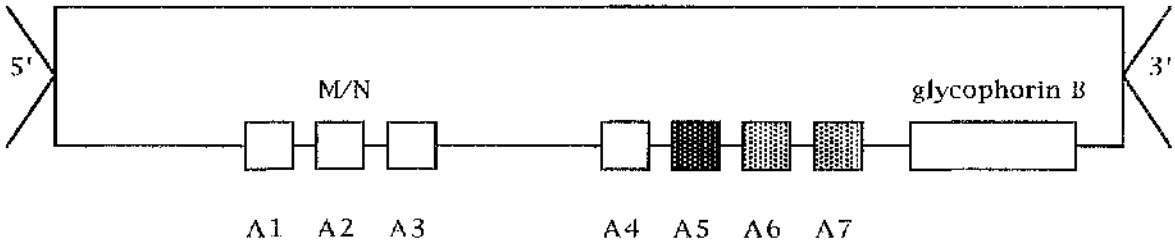
implications, particularly in ante-natal work, and should prove valuable tools for the improvement of patient care.

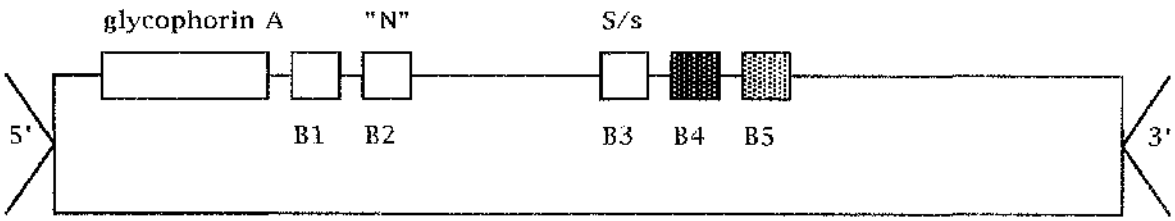
1.6 The MNSs blood group system

The antigens of the MNSs blood group system were among the first to be recognized after the ABO antigens were discovered (Landsteiner *et al.*, 1927). They are carried on red cell membrane glycophorins A and B (α and β). Glycophorins A and B are sialoglycoproteins carrying the MN and the Ss antigens respectively (Figure 1.4). The primary sequences for these proteins have been determined (Tomita *et al.*, 1978, Blanchard *et al.*, 1987 b). The expression of M or N antigen is determined by the amino acid residues at codons 1 and 5 (serine and glycine for M and leucine and glutamic acid for N). Glycophorin B also carries N antigen as it has leucine and glutamic acid residues at positions 1 and 5. The expression of S and s antigens on glycophorin B is determined by the presence of methionine or threonine, respectively, at amino acid position 29. Glycophorin A is a transmembrane protein of 131 amino acids and the predicted molecular mass is 36 kDa. This was the first membrane protein for which the amino acid sequence was determined. Glycophorin B is likely to have arisen by gene duplication from a parent gene for Glycophorin A as the amino acid sequences show a high degree of homology. This protein is 72 amino acids in length and has a predicted molecular mass of 20 kDa. The nucleotide sequences of isolated cDNA clones for each of the two proteins have confirmed the primary sequences previously obtained (Siebert *et al.*, 1986 b, Siebert *et al.*, 1987, Rahuel *et al.*, 1988, Tate *et al.*, 1988). The chromosomal location of the gene encoding Glycophorin A was determined by *in situ* hybridization (Rahuel *et al.*, 1988). This was found to be located on chromosome 4q28-q31. The isolation of cDNA clones for both of these glycophorins has been reported (Tate *et al.*, 1988). The resulting nucleotide sequences have been determined and the results revealed that both proteins have virtually identical N-terminal sequences. cDNA clones encoding these two glycophorins exhibit strong sequence homology, giving evidence that the two proteins are coded for by separate but evolutionarily related genes (Siebert *et al.*, 1986 a). Both glycophorins are heavily glycosylated and this would explain the high apparent molecular weight of these small proteins.

Figure 1.4 The arrangement of the *MNSs* gene

Diagrammatic representation of the genes encoding glycophorins A and B. Exons labelled A refer to glycophorin A and those labelled B refer to glycophorin B. Exons which encode red cell antigens have the antigen name above the corresponding exon. Exons encoding extracellular sequence are unshaded, exons encoding transmembrane sequence are dark shaded and exons encoding intracellular sequence are lightly shaded.


5'
M/N
glycophorin B
3'
A1
A2
A3
A4
A5
A6
A7


glycophorin A
"N"
S/s
5'
B1
B2
B3
B4
B5
3'

Work has been published which demonstrates that red cells of the phenotype S-s-U- (which lack Glycophorin B) result from a structural deletion of part of the gene encoding Glycophorin B (Huang *et al.*, 1987).

Miltenberger antigens are variant hybrid phenotypes within the MNSs blood group system, the antigens of which are carried on glycophorins A and B. Seven classes of Miltenberger antigens have been described (Dahr, 1992). Miltenberger antigen classes III, IV and VI result from the production of a peculiar glycophorin B protein. The primary structure of this protein has been reported to show similarity to that of glycophorin A (King *et al.*, 1989). However, more detailed primary structure analysis is required to allow full comparison between the two proteins.

The use of DNA sequence analysis and of cDNA probes for the investigation of the genes encoding Glycophorin A and B has revealed some important information regarding the molecules which carry Stones antigenic determinants. Stones is a rare antigen of the MNSs blood group system, the presence of which can cause suppression or enhancement of M antigen but not of N, S or s antigens. Stones antigen arises from a crossover of the two genes encoding Glycophorins A and B, and the point of crossover has been defined (Blanchard *et al.*, 1987 a, Dahr *et al.*, 1987, Blumenfield *et al.*, 1987). Dantu is another antigen within the MNSs blood group system which, like Stones, arises as a result of a hybrid involving the gene encoding Glycophorin B (Huang *et al.*, 1988). The crossovers which give rise to the Stones and Dantu antigens are most likely to arise from chromosomal misalignment between normal genes encoding Glycophorins A and B at meiosis (Anstee *et al.*, 1982).

1.7 Other blood group systems investigated using molecular biology techniques

The Kidd blood group antigens are carried on a 45 kDa protein. cDNA fragments have been isolated which encode these antigens and the cDNA length is known to be 2.3 kilobases. (Allen *et al.*, 1988). Little more is known about the genetic basis of this blood group system to date. The Kidd blood group and urea transport function of red blood cells have recently been shown to be carried on the same protein (Olives *et al.*, 1995)

The cDNA encoding the Duffy glycoprotein was isolated recently from a bone marrow cDNA library, as already described in section 1.2.1 (Chaudhuri *et al.*, 1993). The cDNA isolated was 1267 nucleotides in length, with an open reading frame encoding a protein of 338 amino acids and apparent molecular weight 35.7 kDa.

Several other blood group systems and individual antigens have been investigated using molecular biological techniques, although not to the extent of the blood group systems detailed above. These include Gerbich antigens (carried on Glycophorins C and D) (Colin *et al.*, 1986, Colin *et al.*, 1989, High *et al.*, 1987) and Cromer blood group antigens which are composed of a 25-26 kDa protein encoded by a genetic locus at chromosome 1q3.2 (Lublin *et al.*, 1987, Stafford *et al.*, 1988). Chido and Rogers antigens, which are complement component related, have also been studied. These antigens are encoded by two subloci located in the major histocompatibility complex class III region on the short arm of chromosome 6 (Belt *et al.*, 1984, Yu *et al.*, 1986).

1.8 The Kell blood group system

1.8.1 Introduction to the Kell blood group system

The Kell blood group system is the third most clinically significant blood group system with respect to transfusion. Molecular biological investigation of this blood group system has revealed some interesting and useful information and these findings will be covered in detail in the remainder of this section.

During the early use of the antiglobulin test, which revolutionised transfusion science, Coombs *et al.* discovered an IgG indirect agglutinating antibody which defined the antigen Kell (K1) (Coombs *et al.*, 1946). The antithetical antigen Cellano (K2) was discovered three years later (Levine *et al.*, 1949). The discovery of the antithetical pairs Kp^a (K3 or Penney) and Kp^b (K4 or Rautenberg), and Js^a (K6 or Sutter) and Js^b (K7 or Matthews) marked the recognition of the Kell blood group system as a complex polymorphism of red cell antigens (Allen *et al.*, 1957). The name of this system is derived from the individual who produced the first recognised sample of the antibody, but numerical notation has now been adopted to allow the designation of newly recognised antigen specificities to the Kell blood group system.

1.8.2 Kell system antigens

The Kell blood group system to date consists of at least 21 antigens, K1 to K24 (K8, K9 and K15 now being obsolete) (Marsh *et al.*, 1993). Ten of these antigens are arranged in five antithetical sets with opposing high frequency and low frequency antigens. These sets include K1 and K2; K3 and K4; K6 and K7; and K11 and K17 (Marsh *et al.*, 1990). The fifth set consists of a low frequency antigen K24 which is thought to be the antithetical partner of the high frequency antigen K14 (Eicher *et al.*, 1985). Inheritance of these antigens is autosomal and codominant. A series of high frequency antigens (K12, K13, K18, K19 and K22) has been recognised to be present on all red cells, except K_O cells (section 1.8.5). Due to the high frequency of these antigens, family studies which would demonstrate the control of their production by genes at the *KEL* locus are particularly difficult to perform. These antigens have therefore been labelled as Kell related or para-Kell and biochemical studies have shown them to be markers on the Kell glycoprotein (Marsh *et al.*, 1987).

1.8.3 Immunogenicity of Kell system antigens

The antigens of the Kell blood group system are particularly immunogenic. K1 is second only to Rhesus D in its immunising potential. Approximately 5% of K:-1 people will form an antibody specific for the K1 antigen if transfused with one unit of K:1 red cells (Marsh *et al.*, 1993). Antibodies to Kell antigens have been implicated in cases of HDN and can occasionally cause severe reactions if incompatible blood is transfused. Anti-K1 antibodies are mainly red cell stimulated immune antibodies which tend to be of the IgG1 isotype (Marsh *et al.*, 1990). Transfusion is far more likely to stimulate anti-K1 production than is a foetomaternal bleed, but the consequences of transfusing a K:-1 individual with K:1 red cells are not serious enough to justify selection and transfusion of K:-1 red cells to such individuals. On occasion, anti-K1 may be found in the serum of a person who has no history of possible exposure to K:1 red cells. The reasons for the occurrence of such antibodies will be discussed in section 1.8.15.

1.8.4 Phenotypes and their recognition

In order to understand fully a blood group system, it is important to have a knowledge of the antigenic complexity, the genetic background, the biochemical nature of the antigenic determinants and the membrane structure which carries them. The roles that the structures play in the organisation and function of the red cell membrane, and the molecular basis for normal and variant phenotypes are also important factors to consider.

As already discussed, there are a number of common Kell-associated antigens which have been defined at the phenotypic level, but not yet at the genotypic level. These are referred to as the para-Kell antigens and their phenotypic association with the Kell blood group system has been determined in the following ways:-

A

The antigens in question are absent from K_O cells. K_O is a null phenotype cell which lacks all known Kell system antigens. However, it does express Kx antigen. This will be discussed in section 1.8.5. The first indication that an antibody defines a Kell system-associated antigen is usually that the antibody reacts with all red cell samples with the exception of K_O phenotype cells and the antibody maker's own red cells.

B

This red cell phenotype will be discussed in more detail in section 1.8.6, but it is sufficient to say at this point that Kell blood group system antigens are expressed at low levels on M^cLeod phenotype red cells. This phenotype is named after the propositus.

C

Red cells that lack very common Kell system associated antigens often have a very weak expression of K4 (Kp^b).

Table 1.1 gives a summary of the known Kell blood group system antigens.

From Table 1.1 it can be seen that, of the antithetical groupings already described, one allele tends to have a greatly increased incidence over its antithetical partner. It is clear that a number of loci exist which are either

Table 1.1 Notations and frequencies of Kell system antigens

NAME	LETTER	NUMBER	ANTIGEN FREQUENCY(%)
KELL	K	K1	9.0
CELLANO	k	K2	99.8
PENNEY	Kp^a	K3	2.0
RAUTENBERG	Kp^b	K4	> 99.9
PELTZ	Ku	K5	> 99.9
SUTTER	Js^a	K6	< 1.0 in whites, 19.5 in blacks
MATTHEWS	Js^b	K7	> 99.9 in whites, 99.9 in blacks
KARHULA	Ul^a	K10	2.6 in Finns, < 0.1 in others
COTE		K11	> 99.9
BOCKMAN		K12	> 99.9
SGRO		K13	> 99.9
SANTINI		K14	> 99.9
	k-like	K16	99.8
WEEKS		K17	0.3
MARSHALL		K18	> 99.9
SUBLETT		K19	> 99.9
	Km	K20	> 99.9
LEVAY	Kp^c	K21	< 0.1
IKAR		K22	> 99.9
		K23	< 0.1
Cls		K24	< 2.0

Data from Redman *et al.*, 1993

very closely linked or, more likely, are subloci within a single genetic locus at which sets of allelic genes of the Kell blood group system reside (section 1.8.12).

The antigens K12, K13, K14, K18, K19 and K22 have been shown to be associated with the Kell blood group system at a phenotypic level, but not yet at a genetic level (Marsh *et al.*, 1987). These antigens are absent from cells which lack the antigens of the known antithetical groupings already described. However, the genetic control of the production of some of these antigens may involve genes that are positioned near the Kell complex locus. It is possible that these genes are not expressed in the K^O diploid individual.

In 1976, Marsh listed K16 as being an antigen similar to K2, which is found on K:2 cells but which is absent from K:-2 cells (Marsh *et al.*, 1976). The possible qualitative difference between K2 and K16 is apparent when investigating McLeod phenotype red cells which exhibit a weakened form of the K2 antigen. This is discussed in section 1.8.6.

1.8.5 The Kell Null phenotype

K_O is a red cell phenotype where all Kell complex antigens are absent, with the exception of Kx. Although Kx is related to Kell, it has been shown to be the product of an allele outwith the *KEL* gene complex locus (Ho *et al.*, 1992). The K_O phenotype occurs when two K^O genes are inherited at the Kell complex locus. K^O is a silent allele at this locus and inheritance of two of these alleles results in red cells with the K:-1, -2, -3, -4 etc. phenotype. Ascertainment of the K_O phenotype is often by the detection of the "total-Kell" antibody anti-K^u (anti-K5). This may be produced following immunisation by transfusion or pregnancy (Corcoran *et al.*, 1961).

1.8.6 The McLeod phenotype

The McLeod phenotype is associated with the Kell blood group system. In this phenotype, the expression of the Kell complex antigens is markedly weaker than in normal red cells (Allen *et al.*, 1961). McLeod cells were firstly thought to be of the K_O phenotype, but weak (w) expression of some Kell system antigens was later observed. However, detection of these weakly expressed antigens required adsorption/elution studies. The first McLeod cells studied were of the K:-1, 2w, -3, 4w phenotype. Subsequently, it was shown that red cells

of the McLeod phenotype also have weak expression of Ku (K5) and Js^b (K7) (Stroup *et al.*, 1965, Nunn *et al.*, 1966). McLeod red cells lack the otherwise ubiquitous red cell Kx antigen (Marsh *et al.*, 1975). Red cells which lack the Kx antigen are acanthocytic in appearance, i.e. irregular in size and shape with numerous protrusions extending from the cells. They are cleared from the circulation faster than normal and the person with McLeod phenotype is likely to have anaemia, reduced haptoglobin levels, increased bilirubin levels and splenomegaly.

An antibody which reacted with all normal Kell system phenotypes and K_O cells was discovered in a donor serum sample (Claas) (van der Hart *et al.*, 1968). However, it did not react with the red cells of the donor or with those of the original McLeod propositus. This antibody was named anti-KL (anti-K9). At first it was unclear how a Kell system antibody could react with K_O cells, since these were known to lack all Kell system antigens. Seven years later, it was reported that McLeod phenotype red cells lack the antigen Kx and that McLeod phenotype individuals can be immunised and can produce anti-Kx (Marsh *et al.*, 1975). K_O red cells express large amounts of the Kx antigen, while individuals with two fully functional genes at the *KEL* complex locus produce Kx in smaller quantities. From these data, it became clear that anti-KL and anti-Kx were not of the same specificity, as anti-KL was shown to react equally well with all cells except those of the McLeod phenotype, regardless of the presence or absence of the K^O gene. Using K_O cells and the adsorption/elution technique it was possible to separate anti-Kx from anti-KL (Marsh *et al.*, 1975). A second antibody, anti-Km (anti-K20) was not separable from anti-KL by adsorption with K_O red cells (Marsh, 1979). This antibody defines an antigen which is present on all normal Kell system phenotype cells, but which is absent from cells of the K_O and McLeod phenotypes. This suggests that production of Km is controlled by a gene at the *KEL* complex locus. Anti-KL was therefore regarded as a mixture of anti-Kx and anti-Km which could be separated by differential adsorption.

The increased amount of Kx antigen on the cells of K^O homozygous or heterozygous individuals, by comparison with cells from individuals who do not inherit a K^O gene, allows the determination of K^O in the heterozygous state. This can be achieved by titration with anti-Kx against the red cell samples.

Comparison of antigen expression in K_O and M^{C}Leod phenotype red cells has demonstrated the weakened expression of Kell system antigens on M^{C}Leod phenotype red cells, compared with the complete lack of expression of all Kell system antigens on K_O cells (Issitt, 1985).

1.8.7 Inheritance of the M^{C}Leod phenotype

An important aspect of the M^{C}Leod phenotype is that it is confined to males. This indicates that it has an X-linked mode of inheritance (Marsh *et al.*, 1975). The gene locus may be referred to as *Xk* (Bertelson *et al.*, 1988, Ho et al., 1992) and the normal allele that orders synthesis of the Kx antigen is *X'k*. This allele is absent from M^{C}Leod phenotype individuals. Female carriers will bear both normal and M^{C}Leod phenotype sons, and both normal and carrier daughters depending on the maternal X chromosome inherited (Marsh *et al.*, 1990). It is likely that the *KEL* gene complex locus of M^{C}Leod subjects is normal, as weak Kell system antigens can be detected on M^{C}Leod phenotype cells. The modified expression of the *KEL* complex antigens correlates with the inheritance of a variant gene at the *Xk* locus (Marsh *et al.*, 1990). It is unlikely that a specific inhibitory effect leads to the modification of the Kell system antigens, but the gene present at the *Xk* locus may affect the production of the normal Kell protein or its insertion into the red cell membrane. Lack of a normal allele at the *Xk* locus may result in incomplete biosynthesis and processing of Kell protein, but would still allow the production of Kx carrying protein. The synthesis of this Kx protein may be at an increased level in K_O individuals, as there is a decreased requirement for the cell to produce the Kell protein at normal levels. This would explain why K_O cells are Kx positive despite the fact that the individual has two silent genes at the *KEL* complex locus (Marsh *et al.*, 1990).

Determination of the expression of the M^{C}Leod phenotype may be by the detection of anti-Km or anti-Kx, or by the detection of markedly weakened red cell surface antigens of the Kell blood group system during typing tests. A strong indicator to the M^{C}Leod phenotype is the detection of acanthocytic red cell morphology in a blood smear. It is important to note that the M^{C}Leod phenotype has been linked with diseases such as X-linked chronic granulomatous disease (Van der Hart *et al.*, 1968). This will be discussed in greater detail in section 1.8.18.

1.8.8 Position effects in the Kell system

When discussing the Kell blood group system it is necessary to consider the effect of the presence of one antigen on others within the system, i.e. position effects. This is related to the inheritance of parental gene complexes: *cis* is when a gene complex encoding two particular antigens is inherited from one parent; *trans* relates to the inheritance of a gene complex encoding one of the antigens from one parent and a gene complex encoding the second antigen from the other parent. When considering possible *cis* and *trans* position effects within the *KEL* complex locus, two sets of findings give important information regarding the production of K2, K3, K4 and K7. These findings suggest a similar situation to that found in the Rhesus system, where *C* in the *trans* position to *D* may result in the D^u phenotype. That is, a weakening of the D antigen may occur where *C* is found in the *trans* position (Ceppellini *et al.*, 1955). In 1958 Allen *et al.* reported a marked weakening of the K2 antigen on K:-1, 2, 3, -4 red cells and that the expression of K2 on K:-1, 2, 3, 4 red cells was less than on K:-1, 2, -3, 4 cells, when tested in parallel (Allen *et al.*, 1958). This could be explained by assuming that *K3* in the *trans* or *cis* position to *K2* results in a weakening of the K2 antigen. Similar studies have also reported that *K3* in *trans* or *cis* to *K7* lowers expression of the latter gene (Issitt, 1985).

Evidence was published which reinforced an earlier observation that K^o in *trans* to a *K3* containing gene complex depresses expression of the genes of that complex (Ford *et al.*, 1977). These workers tested cells of individuals who were genetically *K2, K3, K7/K2, K3, K7* and *K2, K4, K7/K2, K4, K7* in order to demonstrate that this effect was due to K^o and not *K3*. In contrast to other reported results, these studies showed no effect of *K3* on *K7* or *K2* expression (Issitt, 1985).

Brocteur *et al.* worked on a family which appeared to be heterozygous for K^o and found that K^o in *trans* to *K3* resulted in reduced production of K3. However, when K^o was in *trans* to *K4*, the antigen production attained normal levels (Brocteur *et al.*, 1975). Several groups found, from family studies, that when K^o was paired with the *K2, K3, K7* complex the red cells had reduced levels of K2, K3, K5, K7 and K13 (Kaita *et al.*, 1959, Kout, 1975, Marsh *et al.*, 1975). When K^o was in *trans* to the *K2, K4, K7* complex, there was no evidence of reduction in any of the antigen stengths. These results suggest that *K3* in *trans* or *cis* to *K2* and *K7* sometimes causes a reduction in the amount of K2 and K7 produced. K^o in

trans to a gene complex containing *K3* always causes a reduction in the expression of most or all of the antigens encoded by that gene complex. K^O in *trans* to the *K2, K4, K7* complex has no effect on the genes at that complex. No information is yet available as to the possible *trans* effect of K^O on the *K1, K4, K7* or *K2, K4, K6* complexes.

Additional position effects become obvious when some of the very common Kell system related antigens are taken into consideration. For example, cells which are K:-13 are reputed to have weakened expression of K2, K4, K7, K11, K14 and K18 (Marsh *et al.*, 1974, Marsh, 1976)

1.8.9 Linkage analysis within the Kell blood group system

As already described, there are a series of allelic genes in the Kell blood group system. These sets of alleles are similar to C and c, and E and e of the Rhesus system. As with the Rhesus genes, those in the Kell system show very close linkage and crossing over has not yet been detected (Issitt, 1985). Several theoretically possible gene complexes have not been encountered in the Kell blood group system. In each of the known allele sets, one allele is always very much less frequent than the other. For example, high frequency antigens include K2, K4, K7 and K11 whereas K1, K3, K6, K10 and K17 are much less frequent. Since K10 is rare, it is likely that its allelic partner (when found), will be very common.

The most common combination of alleles is *K: 2, 4, 7, 11.* Gene combinations which code for one of the less frequent antigens encode only a single antigen. That is, a gene complex which codes for the production of any two of the antigens K1, K3, K6, K10 and K17 has not been described. It has been suggested that the complex *K2, K4, K7* represents the original *KEL* gene and that *K1, K3* and *K6* each represent mutations (Chown, 1964). The existence of two of the less frequent genes would represent a double mutation and would therefore be extremely rare. Sufficient family studies have already been performed to expect at least one of the unidentified gene complexes to have been discovered. The reason they have not may be that when more than one of the genes *K1, K3, K6, K10* or *K17* are present on a single chromosome, only one of them can be expressed as a recognisable antigenic determinant. Alternatively, the genetic material may be unable to cope with two mutations at the same locus. Blood samples showing expression of two of the three antigens K1, K3 and K6 have

been reported, but on every occasion these antigens were the products of genes on different alleles.

Absolute linkage provided the key to the designation of Ul^a (K10) and Wk^a (K17) to the Kell system (Furuhjelm *et al.*, 1969). It was found that K:1, 2, 10 samples were always a product of the genotype *K1/K2, K10*. That is, only one of *K1* or *K10* can be inherited from one parent. Therefore, it is possible that *K1* is at one sublocus on the chromosome and *K2* and *K10* are at another. If *K10* had not been at the *KEL* complex locus and, if independent segregation of *KEL* genes occurred, then a normal Mendelian distribution would be expected. However, 100% of the samples tested were either *K:1, -10* or *K:-1, 10*. Family studies revealed that *K17* is at the *KEL* complex locus and travels with *K2* (Strange *et al.*, 1974).

1.8.10 Biochemistry of Kell system antigens

Early evidence suggested that both protein and sugar are involved in the antigenic activity of the Kell blood group system. Branch *et al.* concluded from their results that two distinct disulphide bonds are required for maintenance of Kell blood group antigens (Branch *et al.*, 1983). Kell protein on red cells is in a folded configuration held in position by intra- chain disulphide bonds. It is probable that both the folding of the protein and the lipid environment of the red cell membrane are important in maintaining the correct conformation for the presentation of Kell antigenic determinants on the cell surface.

Isolation of Kell protein has been achieved by sensitisation of red cells expressing "normal" Kell system antigens with Kell related antibodies (Redman *et al.*, 1984, Wallas *et al.*, 1986). These sensitised red cells were solubilised and the antigen-antibody complexes removed from the mixture by affinity chromatography using Protein-A or a second antibody bound to a gel filtration matrix. This process is referred to as immunoprecipitation and the eluted immune complexes can then be resolved using sodium dodecyl sulphate polyacrylamide gel electrophoresis (SDS PAGE). Under reducing conditions, this technique generates bands which relate to the heavy and light chains of the antibody and a Kell antigen-carrying 93 kDa protein, after appropriate staining procedures have been performed. Several different antibody specificities have been used to isolate this protein (anti-K1, K2, K4, K5, K7, K12, K13, K14, K19, K22

and K23) (Redman *et al.*, 1984, Redman *et al.*, 1986, Wallas *et al.*, 1986, Marsh *et al.*, 1987). This suggests that almost all of the high incidence Kell related antigens are carried on the same 93 kDa protein, indicating that they may all be products of the *KEL* gene. This 93 kDa protein was found to stain by the periodic acid-Schiff technique, demonstrating the glycoprotein nature of the antigens (Branch *et al.*, 1983, Redman *et al.*, 1984).

Treatment of the isolated protein with N-glycanase resulted in a reduction in apparent molecular mass from 93 kDa to 78-80 kDa. This suggests that five or six N-linked oligosaccharide side chains are attached to the cell surface exposed portion of the protein, assuming that one oligosaccharide side chain has an average molecular mass of 2.5 kDa (Redman *et al.*, 1986).

When the immune complexes were separated under non-reducing conditions, two protein bands of estimated molecular mass 85 kDa and 115 kDa were detected. These data suggest that the Kell protein may be complexed with one or more other membrane constituents on the red cell.

In one study, quantitative experiments using anti-K7, anti-K22 and a mixture of both were performed. The results indicated that both antigen specificities are present on the same molecule (Redman *et al.*, 1984). Another report showed that anti-K7 and anti-K14 could only immunoprecipitate a detectable radioactive band at 93 kDa from K:7, 14 cells and not from K_O cells, confirming the correlation between this protein band and the Kell complex protein (Redman *et al.*, 1986). It was also demonstrated that chymotryptic and tryptic peptide maps of the Kell protein and Band 3 protein were markedly different (Redman *et al.*, 1986). This strongly suggests that the Kell protein and Band 3 protein, the major red cell membrane protein, are different proteins despite their similarity in size. Further confirmation of this was obtained when Kell protein was analysed using a sheep antibody to a 43 kDa fragment of human Band 3. The antibody reacted only with Band 3 protein and not with Kell protein by Western blot analysis. Another important factor is that Band 3, but not the Kell protein, is phosphorylated when cells are incubated with ^{32}P ATP (Redman *et al.*, 1986).

It has recently been suggested that it is unlikely the oligosaccharide side chains play a major part in the determination of the specificity of any of the Kell system antigens, although they may have some involvement (Marsh *et al.*, 1990, Murphy *et al.*, 1993). The conformation of the antigen molecule on the

membrane is important in maintaining its antigenicity. This is reinforced by the fact that Kell protein, isolated under reducing conditions, does not react with antibodies directed against Kell system antigens when the protein is bound to nitrocellulose paper and analysed by the Western blot technique (Redman *et al.*, 1986).

Immunoprecipitation experiments using anti-Kx antisera, resulted in the isolation of a protein of estimated mass 37 kDa which carries the Kx antigen (Marsh *et al.*, 1990). Quantitative radiolabelling experiments reinforce the evidence that K_O red cells yield high levels of Kx, but M^cLeod phenotype red cells yield none (Redman *et al.*, 1986) (sections 1.8.5-1.8.7).

1.8.11 Interaction of genes at the *KEL* and *M^cLeod* loci

As discussed earlier (section 1.8.5), the Kell phenotype detectable on the surface of red blood cells may be a result of an interaction between products of two independent genes; an autosomal gene located on chromosome 7q33 and an X-linked gene on the short arm of the X-chromosome at Xp21 (Ho *et al.*, 1992, Lee *et al.*, 1993). The 93 kDa Kell protein is encoded by the autosomal gene while the X-linked gene encodes a 37 kDa protein carrying the Kx antigen (Redman *et al.*, 1988). This X-linked gene has been referred to as *Xk*. The role of the *Xk* gene in relation to Kell antigen expression is unknown, but it is already well established that the absence of the Kx antigen is linked to a marked weakening of normal Kell system antigens. It is possible that an interaction of the two gene products is required for the two proteins to assume the correct location in the plasma membrane. Lack of the Kx protein in the red cell membrane presents an acanthocytic cell morphology (section 1.8.6).

1.8.12 The *KEL* locus

Recent work has probed more deeply into the genetic locus encoding Kell system antigens. Zelinski *et al.* have published work which shows genetic linkage between the Kell blood group system and prolactin inducible protein (*PIP*) loci, and have provisionally assigned *KEL* to chromosome 7 (Zelinski *et al.*, 1991 a). They suggest a similar assignment for *KEL* as for *PIP* at 7q32-q36. These results were established through family studies. This work disproved an earlier assignment of *KEL* to chromosome 5 (Leppert *et al.*, 1987). It has also been suggested that the location of the *KEL* locus on chromosome 7 is distal to the

cystic fibrosis locus (Purohit *et al.*, 1992). Data which has arisen as a result of this project defined the location of the gene encoding the Kell protein more closely on chromosome 7 to 7q33-q35 (Murphy *et al.*, 1993). A simultaneous publication defined the location as 7q33 (Lee *et al.*, 1993).

Genetic assignment of *KEL* depends on the availability of a Kell specific DNA probe for hybridization to human chromosomes. With foreknowledge of the cDNA sequence of *KEL* (and hence the Kell amino acid sequence) a suitable oligonucleotide *KEL* probe was designed for hybridization studies and used to isolate the full length *KEL* from a human cDNA library (e.g. bone marrow λgt 10) (Lee *et al.*, 1991). PCR may also be used for amplification of *KEL* specific sequences from human cDNA libraries, or from cDNA manufactured from RNA isolated from individuals of known Kell phenotype. The application and usefulness of PCR techniques in transfusion science are discussed in section 1.2.2.

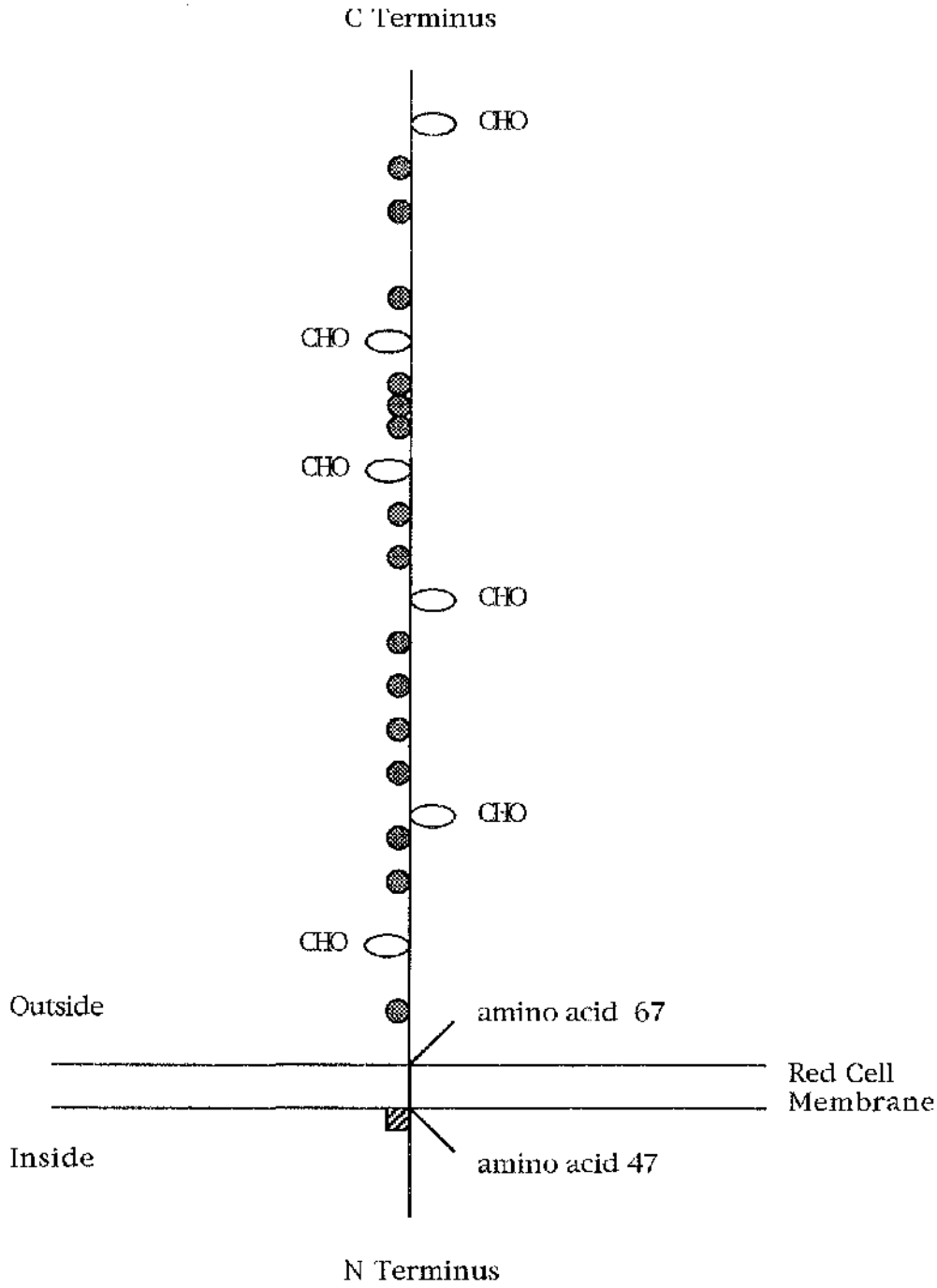
The reported results of hydropathy analysis of the predicted amino acid sequence of the Kell protein suggest that it has one membrane spanning region from amino acids 48-67 (Lee *et al.*, 1991) (Figure 1.5). The 47 amino acids at the N-terminal end are hydrophilic and are predicted to be cytoplasmic. The N-terminal end is blocked by pyroglutamic acid and cannot be sequenced for this reason. Immediately adjacent to the membrane spanning region, the predicted cytoplasmic domain is marked by a pair of arginine residues and there is a negative net charge difference between residues on the outer and inner (cytoplasmic) side of the putative membrane spanning region. These workers confirmed that there are six N-glycosylation sites on the protein.

The putative extracellular domain also has a peculiar amino acid sequence which allows the binding of zinc atoms i.e. a "zinc finger" (Lutz *et al.*, 1992). This suggests that the Kell glycoprotein may be distantly related to the zinc metalloprotease family. In the extracellular segment of the protein, at amino acid residues 581-612, two overlapping leucine-zipper like regions have been identified (Redman *et al.*, 1993).

A recent report describes the organisation of the *KEL* gene. The complete gene spans 21.5 kb, containing 19 exons (or coding regions). These 19 exons relate to the 731 amino acid protein sequence (Lee *et al.*, 1995 b). The intervening introns (or non-coding regions) must constitute about 19 kb of gene sequence

Figure 1.5 Diagrammatic representation of the Kell protein

Diagrammatic representation of the Kell protein illustrating the presence of carbohydrate side chains, cysteine residues, and arginine residues which mark the start of the membrane spanning domain. The protein is 732 amino acids in length. The intracellular domain comprises the first 47 amino acids, amino acids 48 to 67 constitute the single membrane spanning region and the remainder of the amino acid sequence is extracellular. The N terminal end is blocked by pyroglutamic acid (Lee *et al.*, 1991).



denotes cysteine residue

CHO denotes carbohydrate side chain

denotes pair of arginine residues

as the *KEL* cDNA is known to be approximately 2.4 kb in length. The putative membrane spanning region of the protein is encoded in exon 3 and the zinc binding site region is encoded in exon 16.

1.8.13 Kell antigen copy number

Kell system antigens have been studied quantitatively using I^{125} labelled anti-K1. It was estimated that K:1, -2 cells have approximately 6000 K1 sites per cell and K:1, 2 cells have approximately 3500 K1 sites per cell (Hughes-Jones *et al.*, 1971). The number of K2 sites on both homozygous and heterozygous cells is thought to be between 2000 and 5000 sites per cell. Kell system antigens have fewer antigen sites on the red cell membrane than those of any other major blood group system for which data are available. This is not reflected in the immunogenicity of the antigens themselves, as the corresponding antibodies give strong agglutination and are often of high titre. Thus, Kell blood group antigens are extremely immunogenic given their low copy number.

1.8.14 Chemical modification of Kell system antigens

It is possible to use chemicals to denature the Kell antigenic structures and thus modify the cells with respect to antigen expression. It has been shown that both incubation of cells in formaldehyde solution and treatment of cells with sulphydryl compounds cause inactivation of Kell system antigens (Branch *et al.*, 1983). Treatment of intact cells with 2-aminoethylisothiouronium bromide (AET) effects the complete inactivation of all known Kell system antigenic determinants, making the cells non-reactive with antibodies which define these antigens (Advani *et al.*, 1982). Artificial production of K_O cells is thus achieved. AET treatment also increases the level of Kx antigen on the cells (Issitt, 1985). These findings support the theory that Kx may be the membrane protein to which Kell system antigens are attached. In cells other than K_O, the presence of Kell system antigens may block the access of anti-Kx to the Kx determinant, resulting in a weak haemagglutination reaction. As the AET removes the Kell system antigens the Kx antigens may become more accessible to anti-Kx and the reaction strength would increase accordingly. The ability to create K_O cells artificially is of great advantage when studying Kell system related autoantibodies, which cause warm autoimmune haemolytic anaemia. AET, as expected, removes weak Kell system antigens from M^cLeod phenotype red cells. It has also been shown to inactivate the Km antigen, which is lacking

from McLeod cells but not from any other cells. This finding was evidence that the *Km* gene resides at the *KEL* complex locus, unlike the *Kx* gene.

AET can be extremely useful when assignment of new low frequency antigens to the Kell system is an issue. If the "new" antigen can be inactivated by AET, then this provides an indication that the antigen may reside within the Kell blood group system.

Another reagent, ZZAP, which is a mixture of dithiothreitol (DTT) and papain, also destroys Kell system antigens on intact red blood cells (Branch *et al.*, 1982). Again, as the Kell system antigens are removed, the amount of detectable Kx antigen increases. ZZAP, unlike AET, does not denature the Km antigen. When the constituents of ZZAP are used separately, the Kell antigen is inactivated only when the cells are incubated first with DTT and then with papain (Savalonis *et al.*, 1988). This may suggest that the outermost structure of the protein responsible for antigenic activity is maintained by disulphide bonds which conceal the papain sensitive sites. Therefore, DTT must be used first to cleave these bonds before papain can act on the protein responsible for Kell activity. It has been suggested that K6 and K7 antigens are different from other Kell system antigens on the basis of their sensitivity to DTT treatment (Branch *et al.*, 1983). These two antigens are denatured by 1 to 2mM DTT, whereas other Kell system antigens require 200mM DTT for complete denaturation. These results indicate that K6 and K7 antigens may be located on a different antigenic domain.

1.8.15 Kell antigens in other species

Kell "like" antigens are not exclusive to human red blood cells, although they have been shown to be absent from human platelets, lymphocytes, monocytes and granulocytes (Jaber *et al.*, 1991). One recent report details Northern blot analysis of various human tissues using a *KEL* specific probe and this suggests that *KEL* mRNA is present only in erythroid tissues (Lee *et al.*, 1993). Chimpanzee red blood cells have a Kell system very similar to that of humans with respect to antigenic specificity. The most significant difference between the two is the size of the membrane glycoprotein carrying the antigenic determinants. The human determinants are carried on a 93 kDa protein whereas the chimpanzee version is carried on a protein of estimated molecular mass of 97 kDa (Brendel *et al.*, 1985). This difference is due to increased glycosylation of

the chimpanzee Kell protein. It has also been noted that chimpanzee Band 3 protein migrates with a slightly higher predicted molecular mass than human Band 3 by SDS-PAGE. This suggests there may also be heavier glycosylation of this protein in the primates. There are no reports of Kell-like red cell antigens in other vertebrate species.

Bacteria have the ability to express antigenic determinants which mimic Kell blood group system antigens. This was discovered when individuals were shown to produce anti-K1 antibodies with no evidence of previous transfusion or pregnancy. It was found that immunisation was by way of a Kell-like antigen on *E. coli* 0125:K70 (B15) (Marsh *et al.*, 1990). It has been reported that *E. coli* 0125:B15 subtype 12808 has specific K1-like activity when tested by haemagglutination inhibition (Savalonis *et al.*, 1988). Springer *et al.* had previously shown that microorganisms may express antigens similar to that of human blood groups and can stimulate the formation of naturally occurring alloantibodies (Springer *et al.*, 1961). The absence of K1-like activity in all but this single *E. coli* subtype relates to the rarity of naturally occurring non-immune anti-K1 antibodies. Savalonis *et al.* showed that this antigenic determinant can only be detected on disrupted organisms (Savalonis *et al.*, 1988). Therefore, processing of an intact organism by phagocytosis, and its subsequent destruction is likely to be a necessary stage in the formation of naturally occurring antibodies.

1.8.16 Kell related antibodies *in vivo*

The production of antibodies is essential for the definition and identification of Kell antigens. Anti-K1 is a persistent antibody produced mainly in response to an antigen positive blood transfusion or foetomaternal bleed. Antibodies to other Kell system antigens are less common due to the very high or low incidence of these antigens. Anti-K3 and anti-K6 are made less frequently than anti-K1, purely as a result of the low frequency of the antigens. Anti-K2, anti-K4 and anti-K7 are also rare antibodies. When anti-K2, anti-K4 and anti-K7 do arise, provision of compatible blood for transfusion can be difficult and suitable frozen blood may be required. There is relatively little information available regarding the activity and clinical significance of alloimmune anti-K11, -K12, -K13, -K14, -K18, -K19 or -K22. Therefore, the decision to transfuse compatible blood may be more difficult. Antibodies specific for some of the common Kell system antigens can occasionally be autoimmune in nature and can cause *in*

vivo red cell destruction and haemolytic anaemia. Anti-K1 can cause severe HDN, but the disease is generally mild compared to that caused by anti-Rhesus D. HDN can also be caused by anti-K3, anti-K4, anti-K6 and anti-K7. A case resulting from the presence of anti-K5 merited exchange transfusion due to its severity (Hardy *et al.*, 1981). Since most of the Kell system antibodies defined to date have a considerable ability to clear antigen positive red cells from the circulation, transfusion of antigen negative blood to immunised patients is practised by most clinicians. The expression of Kell system antigens starts about week 10 of foetal life and also begins early in red cell maturation, at the early erythroblast stage in adults (Marsh *et al.*, 1990).

1.8.17 Kell related antibodies *in vitro*

Monoclonal antibodies have been described which are specific for the Kell system antigens K1, K2 and K14 (Parsons *et al.*, 1982, Sonneborn *et al.*, 1983, Nichols *et al.*, 1987, Jaber *et al.*, 1989, Jaber *et al.*, 1991). These are mainly mouse monoclonal antibodies produced by fusion of the spleen cells from hyperimmunised mice to a mouse myeloma cell line. The immunogens in every case were native or papain-treated intact red blood cells. A human anti-K1 monoclonal antibody has been described (Jaber *et al.*, 1989). This was produced by transforming a B-lymphocyte cell line secreting anti-K1 with Epstein Barr Virus. This allows continuous culture of these cells *in vitro* and they can be fused to a suitable cell line to increase stability and to facilitate continuous culture or storage.

The use of such monoclonal antibodies has proved beneficial in the investigation of the spatial organisation of Kell system antigens on the protein, by use of a technique known as the MAIEA assay (monoclonal antibody-specific immobilisation of erythrocyte antigens) (Petty *et al.*, 1994). Monoclonals of predetermined specificities are used in a competitive binding assay with human polyclonal antibodies. The results of this type of assay yield valuable infomation regarding the relative positions of Kell system antigens on the folded protein. The MAIEA assay was used to determine the close proximity of the K12 antigen to the K/k and Js^a/Js^b antigens on the Kell protein (Reid *et al.*, 1995).

1.8.18 Disease associations and the Kell system

As stated briefly earlier, disease associations exist with respect to some Kell related phenotypes. The most extensively documented association is McLeod phenotype, the intricacies of which have already been discussed, and X-linked chronic granulomatous disease (CGD). Van der Hart *et al.* revealed a recurrence of severe bacterial infections in people of McLeod phenotype and it was later shown that there is an association between McLeod phenotype and X-linked CGD (Van der Hart *et al.*, 1968). In X-linked CGD, the patient's granulocytes can phagocytose bacteria but cannot kill them. This results when the patient's phagocytes fail to generate superoxide and other activated oxygen derivatives on ingestion of microbes. A consistent finding is the absence of the haem spectrum derived from a cytochrome *b* (Royer-Pokora *et al.*, 1986, Thrasher *et al.*, 1994). The way in which this association presents itself is as follows. The Kx antigen is present on both red cells and granulocytes. When Kx is present on the red cells, autosomal Kell system genes are expressed normally. However, the absence of the Kx antigen on red cells causes a weakening of the red cell surface Kell system antigens, resulting in McLeod phenotype. When Kx is present on granulocytes they function normally, but when Kx is missing they cannot destroy bacteria and X-linked CGD may occur. It should be noted that X-linked CGD and McLeod phenotype is not an absolute relationship. A condition known as chronic granulomatous disease which is not X-linked has also been identified. In this case, the granulocytes carry normal levels of the Kx antigen and the inheritance of the disorder involves a recessive autosomal gene. There is no linkage between McLeod phenotype and this form of the disease. Samples of blood from unrelated McLeod individuals have been shown to exhibit a chromosomal deletion at about 500 kb towards the pter side of the CGD locus on the X-chromosome at Xp21 and it is assumed that the *Xk* locus is within this deletion (Francke *et al.*, 1985, Ho *et al.*, 1992). Other associations between McLeod phenotype and disease states have been identified. These include diseases such as retinitis pigmentosa, Duchenne type muscular dystrophy, cardiomyopathy and neurological disorders. The likely reason for these links is that the disease loci are found on the X-chromosome in the vicinity of the McLeod locus.

1.8.19 Weakened Kell system antigens

Temporary weakening of all the Kell system antigens can be an acquired condition and may result in the antigens being almost undetectable by

haemagglutination tests. As a result, the cells may be wrongly typed as K_O. Each case of this condition has been associated with the production of a Kell system specific autoantibody (Seyfried *et al.*, 1972, Beck *et al.*, 1979, Marsh *et al.*, 1979, Manny *et al.*, 1983, Brendel *et al.*, 1985, Marsh *et al.*, 1990). Some of these autoantibodies have broad specificity, reacting with all except K_O cells. The only way to ascertain serologically that cells from these individuals are not the product of a K^OK^O genotype is to detect this autoantibody in the serum. Weakening of Kell antigens is rare, but when it does arise most of the detectable autoantibodies are anti-K1-like in specificity and have been produced without any known antigenic stimulus. The mechanism whereby the antigen strength is reduced is unknown. One unlikely possibility is that a *de novo* modification occurs during the biosynthesis of the Kell protein, or during its incorporation into the red cell membrane. Alternatively, it may be a result of antigen alteration on the red cell surface by an extracellular enzyme, possibly of microbial origin, such as a glycosidase, a protease or a reductase.

Two cases have been reported which give a clue to the weakening of Kell system antigens (Venglen-Tyler *et al.*, 1987). In both cases, the patient was transfused with red cells which had the same Kell phenotype but a different Rhesus phenotype to the patient's own cells. Samples of blood were taken at intervals over several weeks and the red cells were separated by differential agglutination using Rhesus antisera. This established that the transfused red cells retained their Rhesus antigens, but that there was a progressive decrease in the immunoreactivity of the Kell system antigens present. This must reflect the activity of an agent present in the patient's plasma. Five months after the initial tests, the Kell related antibody which had been present in the patients' plasma had disappeared and the Kell system antigens on the red cells returned to normal strength.

In a situation where the patient has anaemia and has an autoantibody which is specific for a high incidence antigen within the Kell blood group system, transfusion of incompatible blood can result in a severe haemolytic reaction. This may result indirectly from the weakening of the patient's own Kell system antigens in order to protect them from the haemolytic consequences of the serum autoantibody. Therefore, it is important to transfuse compatible red cells to patients in this situation. Finding compatible blood for transfusion in such cases may prove difficult and the patient may have to be transfused with K_O

blood. Fortunately, the majority of cases of this type do not cause accelerated *in vivo* red cell destruction and, therefore, do not require transfusion.

1.8.20 Another rare phenotype in the Kell system

There is an additional rare phenotype in which the Kell antigens have permanently weak expression. Initially this may resemble K_O, but adsorption/elution studies have shown the presence of weak K5 and other Kell system antigens. Some individuals of this phenotype can produce an alloantibody which resembles anti-K5 as a result of transfusion or pregnancy. This is misleading when identifying the red cell phenotype and may point towards the K_O phenotype, as these cells also have an enhanced reaction with anti-Kx. This type of weakening of Kell system antigens persists over several years and appears to be inherited, although the mechanism is unknown. This phenotype is referred to as Kmod (Marsh *et al.*, 1987).

1.8.21 Conclusions on the Kell blood group system

A great deal has yet to be learned about the genetics of the Kell system and the biochemistry of the antigens themselves. It is hoped that more murine or human monoclonal antibodies will be produced which can be used to define more precisely the Kell system antigenic structures and which could be used as valuable blood grouping reagents. It is very likely that techniques such as PCR may be used more widely to investigate new antigens and to shed light on linkage disequilibria present within the Kell blood group system. PCR may be particularly useful for investigation of variations at the genetic level between different serologically defined antigens within the Kell blood group system. The application of novel techniques such as *in situ* hybridization, reverse transcription PCR and immunofluorescence microscopy will hopefully provide novel genetic and immunological data on the Kell system in the next few years.

1.9 Summary

Much important information has been gathered in recent years regarding the basis of red cell blood group system variation. These findings could have come to light only by the use of molecular biological techniques. These developments will allow more detailed investigation of rare samples than would otherwise be possible through use of conventional techniques. Molecular biological

techniques are, therefore, extremely important in the field of transfusion science, both for the continued investigation of red cell antigen genes and for the development of diagnostic assays with direct patient care implications.

SECTION 2

Materials and methods

SECTION 2 Materials and methods

2.1 List of suppliers

Unless otherwise stated, all chemicals were Analar grade and supplied by BDH Chemicals or Fisons Scientific. All radiochemicals were purchased from Amersham International plc. Growth media were obtained from Difco Laboratories. Where chemicals or equipment were obtained from other sources, this is indicated in the text. A list of suppliers of special reagents is given below:-

Aldrich Chemical Co., Gillingham, Dorset.
Amersham International plc, Amersham, Bucks.
Applied Biosystems (Perkin Elmer), Warrington, Cheshire.
BDH Chemicals, Poole, Dorset.
Biogenesis Ltd., Bournemouth, England.
Boehringer Corporation (London) Ltd., Lewes, East Sussex.
Clontech, Cambridge BioSciense, Cambridge.
Cruachem Ltd., West of Scotland Science Park, Glasgow.
Difco Laboratories, Detroit, USA.
Dynal, New Ferry, Wirral.
Imperial Cancer Reasearch Fund, South Mimms, Potters Bar, Herts.
Fisons Scientific, Loughborough, Leicestershire.
Fluka Chemicals Ltd., Gillingham, Dorset.
Genosys Biotechnologies, Inc., Cambridge.
Gibco/BRL (Bethesda Research Laboratories) Ltd., Paisley.
Hybaid Ltd., Teddington, Middlesex.
Kodak Ltd., Kirby, Liverpool.
Northumbria Biologicals Ltd., Cramlington, Northumberland.
Pharmacia Ltd., Milton Keynes.
Promega Ltd., Southampton.
Scotlab Bioscience, Coatbridge, Scotland.
Sigma Chemical Co., Poole, Dorset.
Stratagene Ltd., Cambridge.
United States Biochemicals, Amersham International plc, Bucks.
Whatman International Ltd., Maidstone

2.2 Media and buffers

All media and solutions used for the growth of microorganisms or the handling of nucleic acids were sterilised by autoclaving for 20 minutes at 15 pounds per square inch or by filter sterilisation.

BACTERIAL GROWTH MEDIA :

LB medium (per litre)	10g Bacto tryptone 5g yeast extract 10g NaCl
LB plates (per litre)	1 litre LB medium 15g Bacto agar
LBA plates (per litre)	996ml LB medium 15g Bacto agar 4ml ampicillin (25mg/ml stock solution)
2x YT (per litre)	16g Bacto tryptone 10g yeast extract 5g NaCl
Minimal glucose plates (per litre)	15g minimal agar in 900ml H_2O 100ml 10 x M9 salt 1ml 0.1M $CaCl_2$ 10ml 20% (w/v) glucose 1ml 1M $MgSO_4$ 1ml 1M thiamine HCl
5x M9 salts (per litre)	64g $Na_2HPO_4.7H_2O$ 15g KH_2PO_4 2.5g NaCl 5.0g NH_4Cl

BUFFERS

10x TBE	0.89M Tris-HCl (pH 8.3) 0.89M boric acid 0.025M Na_2EDTA
TE	10mM Tris-HCl (pH 7.6) 1mM Na_2EDTA
TBE Loading Dyes	50% v/v glycerol in TE (pH 7.6) 0.25% w/v bromophenol blue 0.25% w/v xylene cyanol FF
20x SSC	3M NaCl 0.3M sodium citrate (pH 7.0)
100x Denhardt's solution	10g Ficoll (type 400) 10g polyvinylpyrrolidone 10g Bovine serum albumin (fraction V) H_2O to 500ml
solution D	4M guanidinium thiocyanate 25mM sodium citrate pH 7.0 0.5% SDS 0.1M 2-mercaptoethanol
5x formaldehyde gel running buffer	0.1M MOPS (pH 7.0) 40mM NaOAc 5 mM Na_2EDTA (pH 8.0) to 1 litre with DEPC treated H_2O
DEPC-treated formaldehyde gel loading buffer	50% glycerol in 1mM Na_2EDTA (pH 8.0) 0.25% w/v bromophenol blue 0.25% w/v xylene cyanol FF

STE buffer	10mM Tris-HCl (pH 7.5) 1mM Na_2EDTA 150mM NaCl
5x TAE buffer	0.2 M Tris/acetate 5 mM Na_2EDTA (pH 8.0)
TAE loading dyes	10% w/v PEG 6000 0.25% w/v bromophenol blue 0.25% w/v xylene cyanol FF 0.1% w/v orange G
10x ligase buffer	300mM Tris-HCl (pH 7.8) 100mM $MgCl_2$ 100mM DTT 10mM ATP
solution A	50mM glucose 25mM Tris-HCl (pH8.0) 10mM Na_2EDTA
solution B	1% w/v SDS 0.2M NaOH
buffer P1	50mM Tris-HCl (pH 8.0) 10mM Na_2EDTA
buffer P2	200mM NaOH 1% w/v SDS
buffer P3	2.55M KOAc (pH 4.8)
buffer QBT	750mM NaCl 50mM MOPS (pH 7.0) 15% ethanol 0.15% Triton X100

buffer QC	1.0M NaCl 50mM MOPS (pH 8.2) 15% ethanol
2x BW buffer	10mM Tris-HCl (pH 7.5) 1mM Na_2EDTA 2.0M NaCl
TTL buffer	300mM Tris-HCl (pH 8.0) 0.3% Tween 20 6M LiCl
TT buffer	100mM Tris-HCl (pH 8.0) 0.1% Tween 20
solution A4	0.2mM dATP, dCTP and dGTP in 500mM Tris-HCl (pH 7.8) 50mM $MgCl_2$ 100mM 2-mercaptoethanol 100µg/µl bovine serum albumin
solution C	0.4 Units/µl *E. coli* DNA polymerase I 40pg/µl DNA *Pol* I/DNAase I 50mM Tris-HCl (pH 7.5) 5mM magnesium acetate 1mM 2-mercaptoethanol 0.1mM phenylmethylsulphonyl fluoride 50% v/v glycerol 100µg/µl bovine serum albumin
solution D	300mM Na_2EDTA (pH 8.0)
lysis buffer	155mM NH_4Cl 10mM NH_4HCO_3 0.1mM Na_2EDTA (pH 7.4)

blood lysis buffer	320mM sucrose 10mM Tris-HCl (pH 7.5) 5mM $MgCl_2$ 1% Triton X100
nuclear lysis buffer	10mM Tris-HCl (pH 8.2) 400mM NaCl 2mM Na_2EDTA
Megaprime labelling buffer solution 2 (Manufacturer's recipe)	dATP, dGTP and dTTP in Tris-HCl (pH 7.5) 2-mercaptoethanol $MgCl_2$
solution 3	1 Unit/μl DNA polymerase I Klenow fragment in 50mM potassium phosphate (pH 6.5) 10mM 2-mercaptoethanol 50% glycerol

2.3 *E. coli* strains, plasmid vectors and their storage

The *E. coli* strain used was stored in the short term (for up to one month) at 4°C on tightly sealed LB plates. Longer term storage was in 50% (v/v) glycerol at -70°C. The following *E. coli* strain was used in this study.

Strain used as host to plasmids pTZ18U and pTZ18R

XL1-Blue *sup E* 44, *hsdR* 17, *rec* A1, *end* A1, *gyr* A46, *thi*, *rel* A1, *lac*⁻ F'[*pro* AB, *lacI* ^{q}, *lacZ* ΔM15tn*10(tet^{r})*] genotype (Sambrook *et al.*, 1988)

DNA vectors, recombinants and oligonucleotides used in this study.

DNA vectors were stored in sterile TE (pH7.6) (section 2.2) at -20°C and oligonucleotides were stored in sterile distilled water at -20°C. DNA stored in this way remains stable indefinitely, providing there is no nuclease contamination.

pTZ18U and pTZ18R plasmid vectors.

The plasmid vectors pTZ18U and pTZ18R (United States Biochemicals Corp.) were used for the subcloning of PCR products described in this study. All of the restriction enzyme digested fragments were inserted into the multiple cloning site. Maps of pTZ18U and pTZ18R plasmids are shown in Figures 2.1 and 2.2 respectively.

2.4 Preparation and analysis of total RNA from whole blood

All procedures involving the isolation or manipulation of RNA were undertaken with care in order to prevent contamination of samples with RNase enzymes (gloves were worn at all times). Contamination with these enzymes would lead to the degradation of the RNA contained in the samples. All glassware and plastics used for these procedures were treated with a 0.1% (v/v) solution of DEPC for 1 hour at 37°C prior to sterilisation by autoclaving for 20 minutes at 15 psi. DEPC is an inhibitor of RNase, but is toxic and must be handled with care.

Figure 2.1 pTZ18U plasmid map

This vector contains a T7 RNA polymerase promoter within a lac Z' gene, adjacent to the polylinker sequence of pUC18. It also contains the intergenic (origin) region of phage f1, oriented in the "U" direction, and a β-lactamase gene which confers ampicillin resistance. T7 RNA polymerase initiates transcription and proceeds through the multiple cloning site. Infection with helper phage will produce single stranded copies of DNA. Fragments of *KEL* cDNA were cloned into the polylinker region of this vector at the *Bam* HI and *Sal* I sites.

pTZ18U

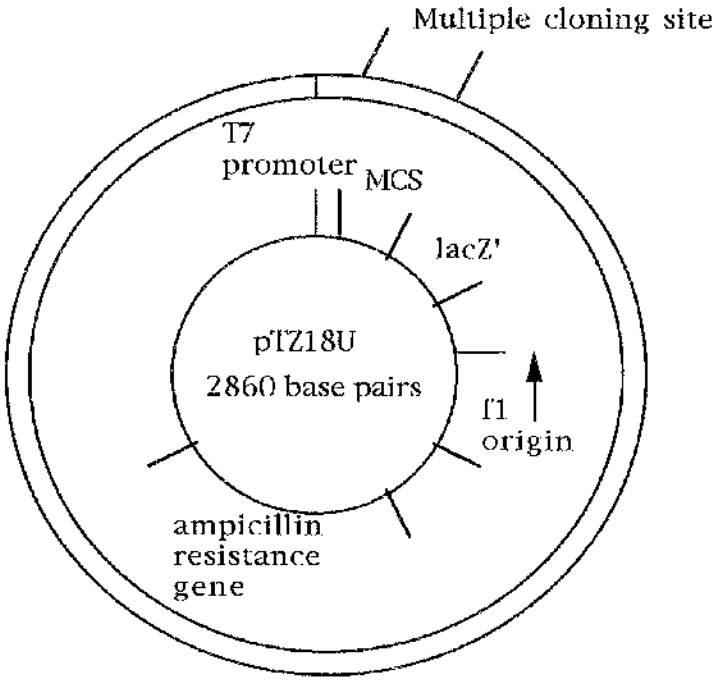
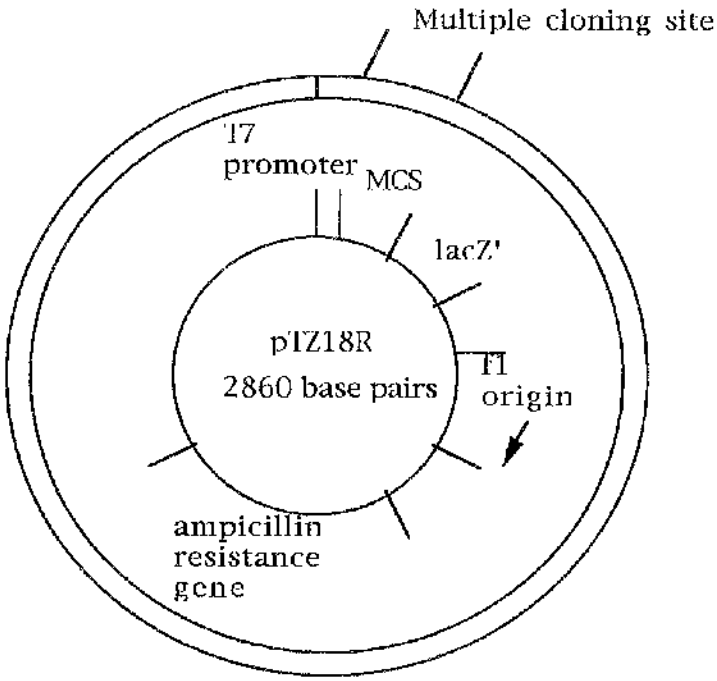

Multiple cloning site
T7
promoter
MCS
lacZ'
pTZ18U
2860 base pairs
f1
origin
ampicillin
resistance
gene

Figure 2.2 **pTZ18R plasmid map**

This vector contains a T7 RNA polymerase promoter within a lac Z' gene, adjacent to the polylinker sequence of pUC18. It also contains the intergenic (origin) region of phage f1, oriented in the "R" direction, and a β-lactamase gene which confers ampicillin resistance. T7 RNA polymerase initiates transcription and proceeds through the multiple cloning site. Infection with helper phage will produce single stranded copies of DNA with the inverse complement. Fragments of *KEL* cDNA were cloned into the polylinker region of this vector at the *Bam* HI and *Sal* I sites.

pTZ18R


Multiple cloning site
T7
promoter
MCS
lacZ'
pTZ18R
2860 base pairs
f1
origin
ampicillin
resistance
gene

2.4.1 Isolation of total RNA from blood - the AGPC method
Chomczynski *et al.*, 1987

Red cells from whole blood samples were washed x3 in cold isotonic saline (0.95% NaCl w/v). 5ml aliquots of washed cells were homogenised on ice in a glass/teflon homogeniser, then two volumes of solution D (section 2.2) were added and the mixture homogenised a second time at room temperature. Homogenates were transferred to sterile DEPC-treated glass universal tubes prior to the sequential addition of 0.2 volumes of 2M sodium acetate (pH 4.0), 2 volumes of Tris-HCl equilibrated phenol (pH>7.6) (Fisons) and 0.4 volumes of chloroform: isoamyl alcohol (49:1 v/v), mixing by inversion after the addition of each reagent. The samples were vortex mixed for 10 seconds before chilling on ice for 15 minutes. They were then centrifuged at 10000g for 20 minutes at 4°C. The aqueous RNA-containing phases were transferred to fresh, sterile DEPC treated 30ml Corex (TM) centrifuge tubes prior to the addition of two volumes of isopropanol. The residual interphases and phenol phases containing DNA and proteins, respectively, were discarded. The samples were centrifuged at 10000g for 20 minutes at 4°C. The supernatant was discarded carefully and the RNA pellet dissolved in 100µl of solution D. 10µl of 3M NaOAc (pH 4.8) were then added and the samples were mixed thoroughly. Two volumes of ice cold ethanol were added and the samples were mixed before RNA precipitation at -70°C for one hour. The samples were centrifuged at 10000g for 20 minutes at 4°C before draining off the supernatant. The RNA pellets were washed with 1ml of cold ethanol prior to drying under a vacuum. They were then dissolved in approximately 100µl of sterile DEPC treated distilled water prior to storage at -70°C. The concentration of RNA in each sample was estimated by optical density measurement where 1 A_{260} unit indicates a concentration of 40µg of RNA/ml.

2.4.2 Agarose/formaldehyde gel electrophoresis of total RNA
Sambrook *et al.*, 1989

1.5% agarose denaturing gels were prepared as follows:-

0.75g of agarose (BRL-Ultra pure agarose) was placed in a sterile 250ml conical flask. 30ml of distilled water were added and this was heated in a microwave at half power until the agarose was completely dissolved. Additional water was added to give a final volume of 31.1ml. 10ml of 5x formaldehyde gel running

buffer (section 2.2) were added and the mixture was allowed to cool to 60°C before the addition of 8.9ml of 12.3M formaldehyde solution, giving a final concentration of 2.2M formaldehyde. The total volume prior to casting the gel was 50ml.

RNA samples were prepared as follows:-

4.5µl of RNA solution (containing 10-30µg of RNA) were placed in a sterile DEPC treated Eppendorf tube. 2.0µl of 5x formaldehyde gel running buffer (section 2.2), 3.5µl of 12.3M formaldehyde and 10µl of dimethylformamide were added. The samples were then mixed and incubated for 15 minutes at 65°C, before chilling on ice. 2µl of sterile, DEPC-treated formaldehyde gel loading buffer (section 2.2) were added. Prior to loading, the gel was pre-run for 5 minutes at 5 volts/cm. The samples were then loaded and the gel was run at 5-8 volts/cm until the leading dye front (bromophenol blue) had travelled 2/3 the length of the gel. The gel was then stained in 1x formaldehyde gel running buffer containing 0.5µg/ml ethidium bromide for 30-45 minutes before destaining in distilled water and viewing on a transilluminator.

2.5 First strand cDNA synthesis

First strand cDNA synthesis was achieved by reverse transcription of poly-A RNA contained in total RNA samples. This involved the use of two different reverse transcription kits. Synthesis of first strand cDNA was monitored to ensure the effectiveness of one of the kits.

2.5.1 Synthesis of first strand cDNA from total RNA using MoMuLV reverse transcriptase

First strand cDNA synthesis using a Pharmacia first strand cDNA synthesis kit

In each case 5µg of total RNA in a volume of 20µl were used in 33µl reactions. The recommended protocol was as follows:-
The diluted RNA was heated to 65°C for 10 minutes before chilling on ice. 1µl of bulk first strand reaction mix was then added. The total reaction mixture contained MoMuLV reverse transcriptase, RNAguard (RNase inhibitor), 0.08mg/ml RNase/DNAse free bovine serum albumin and 1.8mM each dATP, dCTP, dGTP and dTTP in aqueous buffer (45mM Tris-HCl pH 8.3, 68mM KCl, 9mM

$MgCl_2$). DTT at a final concentration of 15mM and 0.2µg of *Not* I-d$(T)_{18}$ bifunctional primer were added and the sample was mixed prior to incubation for one hour at 37°C. The *Not* I-d$(T)_{18}$ bifunctional primer is a hybrid primer composed of a string of dT residues and an anchor domain which contains a *Not* I restriction site. The completed reaction mixture was heated to 90°C for 5 minutes to denature the RNA-cDNA duplex and inactivate the reverse transcriptase prior to storage at -20°C for subsequent use in the polymerase chain reaction (section 2.7.1). This method was also used successfully for specific synthesis of first strand *KEL* cDNA using 40pmol of a 31mer oligonucleotide specific for the 3' end of the published *KEL* nucleotide sequence.

2.5.2 Synthesis of first strand cDNA from total RNA using AMV reverse transcriptase

First strand cDNA was also synthesed using a Promega first strand cDNA synthesis kit.

The reverse transcription reaction was prepared by placing 1µg of substrate RNA in a sterile Eppendorf tube. Reaction constituents were then added to give final concentrations as follows:-
5mM $MgCl_2$, 10mM Tris-HCl (pH 8.0), 50mM KCl, 0.1% Triton X100, 1mM each dNTP, 1 unit/µl rRNasin, 0.75 Units/µl AMV reverse transcriptase, 0.05µg/µl oligo$(dT)_{15}$ primer and RNase free H_2O to a final volume of 20µl.

The mixture was incubated at 42°C for 1 hour, then heated to 99°C for 5 minutes. The samples were immediately cooled on ice prior to use in PCR (section 2.7.1).

2.5.3 Monitoring reverse transcription

The synthesis of first strand cDNA using method 2.5.1 was monitored in order to determine the approximate length of the first strand cDNA produced. This was carried out according to manufacturer's instructions. The protocol was as follows:-
1µg of RNA standard or 5µg of sample RNA was dissolved in 20µl of RNase free H_2O and was incubated at 65°C for 10 minutes prior to chilling on ice. Reaction constituents were added as detailed in section 2.5.1 then 20µCi of [α-^{32}P] dCTP (3000Ci/mmol) were added and the sample was incubated for 1 hour at 37°C. The

final volume of the sample was adjusted to 100µl by the addition of 65µl of STE buffer (section 2.2). 100µl of Tris-HCl equilibrated phenol (pH>7.6) (Fisons) were added and the sample was vortexed prior to centrifugation for one minute at 12000g. This step was repeated using 100µl of chloroform:isoamyl alcohol (49:1 v/v). The upper aqueous phase was transferred to a fresh Eppendorf tube for cDNA precipitation by the method described in section 2.10. The sample was redissolved in 10µl of TE buffer (section 2.2). 2µl of this were electrophoresed on a 1% agarose/TBE gel (section 2.8). The gel was then placed on a glass plate, covered with Saran wrap(TM), clamped down and exposed to X-ray film (Amersham Hyperfilm) at -70°C for one week. The gel sandwich was wrapped in lead to prevent radiation leakage.

2.6 Oligonucleotide preparation

Oligonucleotides were synthesised by the phosphite-triester method (Atkinson *et al.*, 1984) and were purchased from the Department of Biochemistry, University of Glasgow; Pharmacia; Cruachem and from the Department of Biochemistry, University of Bristol. Biotinylated oligonucleotides for use in the polymerase chain reaction and for subsequent direct sequencing of the PCR products on Streptavidin coated Dynabeads(TM) (section 2.20.5) were synthesized at the Beatson Institute for Cancer Research, Glasgow, and at Genosys Inc. in Cambridge. Some of the oligonucleotide samples were supplied ready for use, but others required purification prior to use. Details of individual oligonucleotide primers used in this project can be found in sections 3, 4, 6 and 7.

2.6.1 Ethanol precipitation of oligonucleotide primers
Sambrook *et al.*, 1989

The 500µl oligonucleotide sample was dried under vacuum in an Eppendorf tube prior to dissolving in 500µl of sterile distilled water. 1/10th of the volume of 3M sodium acetate (pH 4.8) and 2 volumes of sterile filtered cold ethanol were added and the mixture placed at -70°C for 15 minutes. The sample was centrifuged at 12000g for 10 minutes before carefully draining the supernatant and washing the pellet with 1ml of cold ethanol. The pellet was dried under a vacuum prior to dissolving in 50µl of sterile distilled water. The concentration was estimated by measurement of the optical density at 260nm where an A_{260} reading of 1 represents an oligonucleotide concentration of 33µg/ml.

2.6.2 Purification of biotinylated oligonucleotide primers using the oligonucleotide purification column (OPC) desalting method

Biotinylated oligonucleotide primers obtained from the Beatson Institute for Cancer Research were incubated overnight at 55°C. 300µl of each oligonucleotide were dried under vacuum prior to resuspension in 3ml of 0.1M TEAA and storage on ice. 5ml of 100% acetonitrile followed by 5ml of 2M TEAA were passed through the OPC to waste. The 3ml of oligonucleotide mixture was passed through the OPC at a rate of 40µl/second. The filtrate was collected and passed through the column a second time. 15ml of 0.1M TEAA were passed through the OPC to waste. The pure oligonucleotide primer was eluted dropwise with 3 aliquots of 1ml 100% acetonitrile and the eluate was collected in Eppendorf tubes and dried under vacuum. Each oligonucleotide was dissolved in 500µl sterile distilled water. 0.5µl and 1.0µl samples of each oligonucleotide were analysed on a 15% PAGE (section 2.8) to assess biotinylation. The biotinylated primer migrates slightly slower than the unbiotinylated form and these can be seen as two distinct bands on the gel following staining with ethidium bromide.

2.7 Polymerase chain reaction

The polymerase chain reaction is an extremely versatile technique for the generation of copies of a target DNA molecule. The products of PCR have many uses. This subsection describes a variety of ways in which PCR has been applied in the course of this project.

2.7.1 PCR using first strand cDNA as a target

PCR amplification of cDNA involves the enzymatic synthesis of specific cDNA target sequences using two oligonucleotide primers which flank the cDNA region of interest and hybridize to opposite strands. The "upstream" PCR primer should be complementary to the 3' flanking region of the first strand cDNA. This oligonucleotide will prime synthesis of the second strand cDNA. The "downstream" primer, which is complementary to the 3' flanking region of the second strand cDNA, is used to prime synthesis of the first strand cDNA. Amplification of the target sequence is achieved by repeating cycles of primer annealing, extension and denaturation of the extended duplex. This results in

an increase in the amount of product by 2^n after n cycles have been completed, assuming the amplification is 100% effective.

The first strand cDNA samples described in section 2.5.1 were used as target material for PCR. The final amounts or concentrations of PCR constituents were as follows-
5µl, 10µl or 33µl of first strand cDNA reaction product, 56mM KCl, 14.5mM Tris-HCl (pH 8.4), 2.0mM $MgCl_2$, 200µM of each dNTP, 40pmol each primer, 2.5 units *Taq* DNA polymerase and sterile distilled water to 100µl .

The first strand cDNA samples described in section 2.5.2 were used as target material for PCR. The final amounts or concentrations of PCR constituents were as follows-
20µl first strand cDNA reaction product, 50mM KCl, 10mM Tris-HCl (pH 8.8), 1.5mM $MgCl_2$, 200µM of each dNTP, 40pmol each primer, 2.5 units *Taq* DNA polymerase and sterile distilled water to 100µl .

The samples were overlaid with 100µl of mineral oil prior to amplification in a DNA Thermal Cycler (TM) (Perkin Elmer). After heating to 95°C for 3 minutes to denature the RNA-cDNA duplex, they were subjected to 30 rounds of 95°C for 1 minute, 55°C for 1.5 minutes and 72°C for 1 minute. The samples were subsequently heated to 72°C for 7 minutes to allow complete extension of chains, then cooled to 4°C for storage. 10µl of the total PCR products were analysed on a 1% agarose/TBE gel with ethidium bromide staining (section 2.8).

2.7.2 PCR using first strand cDNA as a target with biotinylated oligonucleotide primers

PCR reactions were carried out using a first strand cDNA as target and one biotinylated and one unbiotinylated primer. 40pmol of each primer were added to the 100µl reaction. The concentrations of reaction constituents were as detailed above for the AMV based reverse transcription and subsequent PCR (sections 2.5.2 and 2.7.1). The final PCR product was analysed by agarose gel electrophoresis (2.8).

2.7.3 Asymmetric PCR

Asymmetric PCR reactions were performed using only one oligonucleotide primer. The primer concentration used was either 50ng or 1μg per 100μl reaction with 50ng of target DNA [either gel purified PCR product (section 2.9) or double stranded DNA from plasmid clones (section 2.19.1)].

The final concentration of other reaction constituents were as detailed in 2.7.2. The reaction mixtures were heated to 94°C for 3 minutes to denature the double stranded template. The samples were then subjected to 45 cycles of 94°C for 1 minute, 55°C for 1 minute and 72°C for 20 seconds. Another 2.5 Units of *Taq* DNA polymerase were added and the samples were subjected to 10 more cycles of heating and cooling prior to a final extension at 72°C for 7 minutes. The samples were cooled to 4°C and 1/10th of the reaction products were analysed by agarose gel electrophoresis (section 2.8).

2.7.4 PCR using genomic DNA as a target

PCRs were performed using 1μg of total genomic DNA as a target and 40pmol of each primer per 100μl reaction. Preparation of total genomic DNA is described in sections 2.22.1 and 2.22.2. The concentrations of other reaction constituents were as detailed in 2.7.2. The reaction mixtures were thermally cycled using the parameters described in section 2.7.1 but using 65°C instead of 55°C for the annealing step. The products were analysed on 15% polyacrylamide gels or by agarose gel electrophoresis (section 2.8).

2.7.5 PCR using somatic cell hybrid DNA as a target

PCR reactions with somatic cell hybrid DNA as a target were performed using 0.5-2.5μg of the supplied target DNA (I. C. R. F.) with 1μg of each of the appropriate oligonucleotide primers and other constituent concentrations as described (section 2.7.2). The samples were thermally cycled by heating the samples to 94°C for 3 minutes, then subjecting them to 30 cycles of 94°C for 1 minute, 65°C for 1 minute and 72°C for 1.5 minutes. A final extension at 72°C for 7 minutes was performed prior to cooling to 4°C prior to gel analysis (section 2.8).

2.8 Analysis of DNA by gel electrophoresis

DNA of size greater than 150bp can easily be analysed on an agarose gel; whereas smaller fragments (90bp or less) are best seen on a high percentage polyacrylamide gel, where separation is more efficient.

Sample Preparation

DNA in a volume of 10µl was added to an Eppendorf tube. 4µl of TE buffer (section 2.2) and 2µl of TBE loading dyes (section 2.2) were added. The sample was then spun down briefly at 12000g to collect all fluid and the samples were loaded onto the gel immediately.

Agarose gel electrophoresis

Recipe for a 1% agarose/TBE gel- 100ml gel

1g of agarose (BRL Ultrapure agarose) was added to 90ml of distilled water. It was then placed in a microwave and heated on half power until the agarose was completely dissolved. The mixture was allowed to cool prior to the addition of 10ml of 10x TBE (section 2.2). This was mixed thoroughly and ethidium bromide solution was added to a final concentration of 0.5µg/ml. The gel was then poured into a casting tray and allowed to polymerise before use.
(N.B. the concentration of agarose in the gel was increased up to 2% to provide better resolution of fragments smaller than 0.4 kb, or decreased to 0.7% to allow the separation and analysis of fragments larger than 4 kb).

Gel electrophoresis

The gel was submerged in 1x TBE buffer containing 0.5µg/ml ethidium bromide (section 2.2). The gel was loaded and electrophoresed at 8-10 volts/cm until the leading dye front (bromophenol blue) had travelled 2/3 of the length of the gel. The DNA bands were visualised by UV transillumination and photographed.

Polyacrylamide gel electrophoresis (PAGE)

Samples for analysis by PAGE were prepared as for agarose gel analysis.

Gel Preparation- 15% Polyacrylamide Gel

For an 80ml gel the following constituents were thoroughly mixed prior to pouring the gel:-
40ml 30% acrylamide solution (acrylamide: bisacrylamide 29: 1), 8ml 10x TBE, 32ml water, 500µl 10% (w/v) ammonium persulphate solution, 80µl TEMED.
The gel was poured and allowed to polymerise for at least 30 minutes prior to use.

Gel electrophoresis

1x TBE was used as the running buffer (section 2.2). The samples were loaded onto the gel and were electrophoresed at 8 volts/cm until the leading dye front (bromophenol blue) reached the end of the gel. The gel was then stained for one hour in 1x TBE containing 0.5µg/ml ethidium bromide, prior to visualisation of DNA by UV transillumination and photography.

2.9 Gel purification of DNA

This was achieved by electrophoresis of the DNA sample through a 0.7% low melting point agarose/TAE gel. A 100ml gel was prepared by melting 0.7g of low melting point agarose (Gibco BRL LMP agarose) in 80ml of water. 20ml of 5x TAE buffer (section 2.2) were added along with ethidium bromide to a final concentration of 0.5µg/ml. The samples were prepared by the addition of 2µl of TAE loading dyes (section 2.2) to a maximum sample volume of 28µl. The gel was then loaded and electrophoresed at 4°C at 8 volts/cm until the leading dye front (bromophenol blue) had travelled 2/3 of the total length of the gel. After electrophoresis, the required bands were excised (under brief UV transillumination) and each band was placed in a pre-weighed Eppendorf tube (weighed accurately to within 1mg). The tubes and gel were then reweighed to estimate the gel volume assuming 0.1g of gel is equivalent to 100µl of liquid. The gel was melted at 65°C and 1/9th volume of 5M NaCl was added. The samples were mixed thoroughly and returned to 65°C to ensure the gel was properly melted. 0.5 volumes of Tris-HCl equilibrated phenol (pH>7.6) (Fisons) at 65°C were added

and the sample was vortexed immediately for at least 2 minutes. The mixture was centrifuged for 10 minutes at 12000g, the aqueous phase was transferred into a fresh Eppendorf tube and the phenol extraction step was repeated. The aqueous phase was removed and extracted with 0.5 volumes of chloroform: isoamyl alcohol (49:1 v/v) prior to centrifugation as before. The DNA was then ethanol precipitated (section 2.10).

2.10 Ethanol precipitation of DNA

0.1 volume of 3M NaOAc (pH 4.8) and 2 volumes of sterile filtered cold ethanol were added to the DNA sample. This was placed at -20°C for 2 hours to allow precipitation of the DNA. The sample was centrifuged for 10 minutes at 12000g at 4°C and the pellet was drained and washed with 1ml of sterile filtered ice cold ethanol before vacuum drying and resuspension in sterile TE buffer (section 2.2). TE buffer (section 2.2) was added to an aliquot of the sample to a maximum final volume of 28µl. 2µl of TBE loading dyes (section 2.2) were added and the sample was electrophoresed on a 1% agarose/TBE gel (section 2.8) to ensure DNA had been recovered. The amount of DNA recovered was estimated by comparison of band fluorescence with a known standard.

2.11 Removal of oligonucleotide primers and *Taq* DNA polymerase from PCR products

The removal of excess oligonucleotide primers and unused *Taq* DNA polymerase is essential if a PCR product is to be used as a target for further amplification, or for direct sequencing of the product. One method for their removal involves the use of Chroma spin$^{(1M)}$ gel filtration columns (Clontech). Various columns are available and the one chosen was dependent on the size of the fragment which was to be recovered, and the size of the contaminants which were to be removed. The protocol for use of these columns is as follows.

The Chromaspin column matrix was thoroughly resuspended, the caps were removed and the column drained by gravity flow into a microfuge tube. It was then centrifuged for 3 minutes at 700g. The spin step was repeated after discarding the filtrate. The PCR product (75-100µl) was carefully loaded onto the column and spun at 700g for 5 minutes. The eluate contained the purified PCR product. A sample of the purified DNA was analysed on a 1% agarose/TBE gel (section 2.8) to confirm the removal of the primers. The concentration of

purified product was calculated by measuring the absorbance at 260nm (A_{260} = 1 indicates an approximate concentration of double stranded DNA of 50μg/ml).

2.12 Restriction endonuclease digestion of PCR products

Restriction enzymes can be used to cleave double stranded DNA at pre-defined sites, thus allowing positive identification of products of the expected size where the sequence is already known. One Unit of restriction enzyme activity is defined as the quantity of enzyme required for the complete digestion of 1μg of λ DNA in one hour under the correct buffer and temperature conditions for the enzyme. In practice, several fold excess of enzyme was normally used to ensure complete digestion of the sample.

The restriction digest reaction:-

The required quantity of DNA for digestion (100-500ng of PCR product or up to 1μg of genomic DNA) was added to an Eppendorf tube in a volume of less than 17.5μl. 2μl of 10x restriction enzyme buffer (this buffer is specific for each restriction enzyme) were added with the particular restriction enzyme (5 Units for PCR products or up to 60 Units for genomic DNA) and distilled water to give a final volume of 20μl. The mixture was incubated at the appropriate temperature for the enzyme for two hours prior to the addition of 2μl of TBE loading dyes (section 2.2) and analysis on a 2% agarose/TBE gel (section 2.8).

The size of each band generated by restriction enzyme digestion was estimated by measuring distances of migration relative to DNA markers of known size (1kb DNA ladder-Gibco BRL Inc.) and interpolating the distance on a graph of $\log_{10}$ marker size versus distance migrated for these. The distance migrated by bands which result from the experimental digests can be measured and compared to this standard curve to allow size estimation. Restriction digests can be used to excise fragments which have been cloned into a plasmid via restriction enzyme sites. Total genomic DNA and plasmid DNA can also be digested by this method using larger relative volumes.

2.13 Purification of PCR products for cloning into plasmid vectors

PCR products were ethanol precipitated as described in 2.10. The samples were resuspended in 30µl of TE (pH 7.6) (section 2.2). 2µl of the resuspended material was analysed on a 1% agarose/TBE gel (section 2.8) after the addition of 18µl TE buffer (section 2.2) and 2µl of TBE loading dyes (section 2.2). The remainder of the samples were then gel purified on a 0.7% low melting point agarose gel (section 2.9).

2.14 Ligation of enzyme digested PCR products into plasmids

Restriction enzyme digests of 500ng of each plasmid vector were carried out as described in section 2.12.

The ligation reaction mixtures were as follows:-

50ng of digested plasmid vector and 50ng of similarly digested PCR products were mixed in a sterile Eppendorf tube. 2µl of 10x ligase buffer (section 2.2) and 0.5µl T4 DNA ligase enzyme (Promega) were added and the total volume was made up to 20µl with sterile distilled water. The mixtures were centrifuged briefly to collect the fluid prior to incubation at room temperature for 18-24 hours to allow ligation to take place. Controls included digested plasmid with no PCR product.

2.15 Preparation of L-AMP plates with X-gal and IPTG

To prepare twenty plates of 90mm diameter, 500ml of L-broth were prepared (section 2.2). 7.5g Bacto agar were then added. This was autoclaved for 15-20 minutes and the broth was allowed to cool to 60°C prior to the addition of ampicillin to a final concentration of 50µg/ml. X-gal (Promega) and IPTG (Promega) were added, each to a final concentration of 50µg/ml. The broth was mixed thoroughly prior to pouring 25ml into each plastic dish. The surface of the agar was flamed before replacing the lid. The agar was allowed to set at room temperature. The plates were then stored at 4°C and dried at 37°C before use.

2.16 Preparation of competent *E. coli* cells and transformation with plasmid DNA

Sambrook *et al.*, 1989

A single colony of XL1-Blue *E. coli* cells (Stratagene) was picked from a glucose/minimal medium plate (section 2.2) and incubated at 37°C overnight with vigorous shaking in 5ml of 2x YT medium (section 2.2), containing 10μg/ml tetracycline. 40ml of 2x YT medium (in a sterile 250ml flask) were inoculated with 400μl of the overnight culture. The mixture was returned to a shaking incubator for 2-4 hours until the A_{550} was 0.6-0.7. The culture was then centrifuged at 1500g for 5 minutes at 4°C. The cells were kept on ice from this point onwards. The supernatant was discarded and the cells resuspended in 20ml of sterile, ice cold 50mM $CaCl_2$. They were then left on ice for 20 minutes prior to centrifugation at 1500g for 2 minutes at 4°C. The supernatant was discarded and the cells resuspended in 4ml of cold, sterile 50mM $CaCl_2$. The competent cells could be used directly for transformation. However, the efficiency of transformation obtained was increased if the cells were stored on ice for several hours before use. After 24 hours the efficiency of transformation begins to decrease. Using a cold, sterile pipette tip, 200μl of the competent cells were transferred to each of the required number of sterile Eppendorf tubes. 1μl or 4μl of the ligation mixture (section 2.14) were added to each tube. The contents of the tubes were mixed by gentle swirling and stored on ice for 30 minutes. Controls were included i.e. competent cells with no plasmid and competent cells with undigested plasmid. The mixtures were placed in a 42°C water bath for 90 seconds without shaking. The tubes were then transferred rapidly to an ice bath and chilled for 1-2 minutes. 800μl of 2x YT medium were added prior to incubation at 37°C for 45 minutes, to allow the bacteria to recover and to express the antibiotic resistance gene encoded on the plasmid. The tubes were centrifuged briefly in order to pellet the cells and 800μl of the supernatant were discarded. The samples were mixed briefly on a vortex mixer to resuspend the cells prior to transferring them to L-amp plates containing X-gal and IPTG (section 2.15). The cells were spread evenly over the surface of the plate using a cell spreader. The plates were then inverted and incubated at 37°C. Colonies usually appeared after 12-16 hours.

2.17 Culture and purification of transformed cells

White colonies from the plates of transformed cells were purified by scraping them from the plates using a sterile wire loop, streaking onto L-amp plates and growing at 37°C overnight. The plates can then be stored for a period of up to 4 weeks at 4°C.

2.18 Storage of transformed cells in glycerol
Sambrook *et al.*, 1989

5ml of L-broth (section 2.2) containing ampicillin (50μg/ml) was inoculated with a colony of the appropriate transformed cells from a plate (section 2.16) and cultured overnight at 37°C with mixing at 250 rpm. 0.15ml of sterile glycerol was added to 0.85ml of *E. coli* overnight culture. The sample was vortex mixed to ensure that the glycerol was evenly dispersed. The culture was transferred to a labelled storage tube and frozen in ethanol/dry ice or in liquid nitrogen prior to long term storage at -70°C

To recover the bacteria, the frozen surface of the culture was scraped with a sterile inoculating needle and the bacteria that adhered to the needle were streaked immediately onto the surface of an LB agar plate (section 2.2) containing the appropriate antibiotics. The plate was then incubated overnight at 37°C.

2.19 Plasmid preparation

Plasmid preparations can be carried out on small, medium or large scale. This sub-section describes the techniques used for small and medium scale preparations.

2.19.1 Plasmid preparation - miniprep
Birnboim *et al.*, 1979

1.5ml of a 5ml overnight culture of transformed cells in L-broth plus ampicillin was transferred to a sterile Eppendorf tube and the cells were pelleted by centrifugation at 12000g for 10 minutes. They were resuspended in 0.1ml of solution A (section 2.2) then left on ice for 30 minutes before the addition of 200μl of solution B (section 2.2). They were then incubated on ice for 5 minutes.

When the mixture became clear, 150µl of 3.0M NaOAc (pH 4.8) (section 2.2) were added and the sample left on ice for 30 minutes to precipitate chromosomal DNA and proteins. The mixture was centrifuged at 12000g for 10 minutes before transferring the plasmid containing supernatant to a fresh Eppendorf. 1ml of cold ethanol was added in order to precipitate the nucleic acids and the tube was stored at -20°C for 30 minutes. The nucleic acids were then pelleted by centrifugation at 12000g for 10 minutes, rinsed with ice cold ethanol and dried under a vacuum. The sample was then redissolved in 25µl of TE buffer (section 2.2). Aliquots were analysed by gel electrophoresis (section 2.8) to confirm the presence of plasmid. The samples can be digested with restriction enzymes (section 2.12) to ensure the presence of insert in the plasmid.

2.19.2 Plasmid preparation - midiprep

In order to carry out chromosome mapping experiments by *in situ* hybridization, a midiprep of plasmid clone was prepared and purified using a Qiagen 100 column (Qiagen Inc.). This was carried out according to the manufacturer's recommendations (Qiagenologist 3rd Edn.):-

A 20ml overnight culture of *E. coli* containing the plasmid clone was grown in T broth (section 2.2) containing 50µg/ml ampicillin. The overnight culture was centrifuged at 1400g for 10 minutes at 4°C and the supernatant was discarded. The bacterial pellet was resuspended in 4ml of buffer P1 (section 2.2). 4ml of buffer P2 (section 2.2) were then added and the sample was mixed by inversion prior to incubation at room temperature for 5 minutes. 4ml of buffer P3 (section 2.2) were added and the sample was mixed immediately by inversion. The sample was then centrifuged at 30000g for 30 minutes at 4°C. The supernatant was removed promptly and centrifuged again at 30000g for 20 minutes at 4°C in order to obtain a particle free lysate. A Qiagen tip 100 (Qiagen Inc.) was equilibrated with 3ml of buffer QBT (section 2.2) and allowed to empty by gravity flow. The particle free lysate was then loaded and allowed to enter the resin by gravity flow. The tip was washed with 10ml of buffer QC (section 2.2). The DNA in the eluate was then precipitated with 0.7 volumes of isopropanol (equilibrated to room temperature) and pelleted by centrifugation at 4°C for 30 minutes at 30000g. It was then washed with 70% ethanol, air dried and redissolved in 100µl TE (pH 7.6) (section 2.2).

2.20 Sanger dideoxy sequencing

Several adaptations have been made to the original protocol for Sanger dideoxy sequencing to afford it more flexibility with respect to recently developed molecular biology techniques (Sanger *et al.*, 1977). In this sub-section several, of these variations used in this research are described.

2.20.1 Preparation of single stranded DNA from XL1-Blue transformants for use in sequencing
Blondel *et al.*, 1991

50µl of 2x YT medium (section 2.2) in a sterile Eppendorf tube were inoculated with a colony from a plate of the appropriate phagemid transformant. This was vortexed briefly to resuspend the cells. 1µl of high titre phage stock (> 10^{10} plaque forming units/ml M13K07 helper phage-Promega) was added to the cells and these were incubated for 15 minutes at room temperature. 500µl of 2x YT medium containing ampicillin (150µg/ml) were added and the mixture incubated at 37°C with shaking (250 rpm) for 60-75 minutes. 200µl of the cell mixture were transferred to 4ml of 2x YT medium containing kanamycin (75µg/ml) and ampicillin (150µg/ml). The tube was incubated for 18-21 hours at 37°C with shaking (250 rpm). After incubation, the cells were precipitated by centrifugation. The supernatant was transferred to a fresh tube and the phagemid was precipitated from the supernatant by the addition of 200µl 20% PEG 6000/2.5M NaCl per 1.3ml of supernatant with incubation for 2-3 hours at room temperature. The DNA was sedimented by centrifugation at 12000g for 10 minutes and, after discarding the supernatant and washing in ethanol, the DNA was dried under a vacuum and resuspended in 100µl of TE buffer (pH 7.6) (section 2.2). The DNA was purified by extraction with 50µl of Tris-HCl equilibrated phenol (pH>7.6) (Fisons) followed by 50µl of chloroform: isoamyl alcohol (49:1 v/v). It was then ethanol precipitated (section 2.10). The pellet was then dissolved in 50µl of TE (pH 7.6) (section 2.2). 5µl of this were examined on a 1% agarose/TBE gel (section 2.8).

2.20.2 Dideoxy sequencing of single stranded DNA using Sequenase Version 2.0 (TM)

Preparation of sequencing gel:-

42g of urea were added to 15ml of 40% stock acrylamide solution (acrylamide: bisacrylamide 19: 1) (NBL). 10ml of 10x TBE buffer (section 2.2) were added and the total volume was made up to 100ml with water. 600µl of freshly prepared ammonium persulphate solution (10% w/v) were added with 100µl TEMED (BDH). The solution was mixed thoroughly before pouring the gel. The gel was then allowed to polymerise for at least 30 minutes before use.

Preparation of samples for sequencing was carried out in accordance with manufacturer's instructions.

Annealing mix-
5µl (3-5µg) of single stranded DNA prep
2µl H_2O
2µl of Sequenase 5x buffer (200mM Tris-HCl pH 7.5, 100mM $MgCl_2$, 250mM NaCl)
1µl of universal primer or specific sequencing primer (0.5pmol per reaction)

This mixture was placed in a water bath at 65°C for 5 minutes. The bath was then switched off and allowed to cool until it had reached 40°C. The samples were then stored on ice. While the annealing reaction was cooling, 2.5µl of each of the four termination mixtures (each containing 80µM each dNTP, 8µM of the specific ddNTP and 50mM NaCl) were added to appropriately labelled tubes. These were placed at 37°C to allow the mixture to warm. The labelling mix was diluted 5 fold in sterile distilled water to provide the working concentration and stored on ice until use. The Sequenase enzyme (Version 2.0) was diluted 8 fold in enzyme dilution buffer (10mM Tris-HCl pH 7.5, 5mM DTT, 0.5mg/ml BSA) and stored on ice until required.

Reagents were added to the annealed DNA mixture in the following order:-
1µl 0.1M DTT, 2µl diluted labelling mix, 0.5µl [α ^{35}S] dATP 600Ci/mmol, 2µl diluted Sequenase Version 2.0.
The contents of the tubes were then mixed and incubated for 5 minutes at room temperature.

Termination Reaction

3.5µl of labelling reaction were transferred to each of the four pre-warmed termination mixture tubes. These were incubated at 37°C for 5 minutes after brief centrifugation to mix the fluids. The reactions were stopped by the addition of 4µl of stop solution (95% formamide, 20mM Na_2EDTA, 0.05% w/v bromophenol blue, 0.05% w/v xylene cyanol FF) and brief centrifugation to mix the fluids at the base of the Eppendorf tubes. The samples were placed at 95°C for 2 minutes before loading 2µl-4µl of each sample onto the gel to separate the DNA double strands.

Denaturing Gel Electrophoresis

Before loading the samples, the gel was pre-run for approximately one hour at 50W to allow the gel to heat. Heating the gel is important in order to keep the DNA denatured and facilitates better resolution of DNA single strands. The running buffer was 1x TBE buffer (section 2.2). The gel was loaded and electrophoresed for 2 hours or 5 hours before fixing for 20 minutes in 10% acetic acid/10% methanol. The gel was then removed from the fixing solution, drained and attached to Whatman 3MM paper prior to drying under vacuum for approximately one hour with heating (80°C). Once the gel was completely dry it was exposed to X-ray film (Amersham Hyperfilm) for up to one week before developing.

2.20.3 Dideoxy sequencing of cloned, double stranded DNA using Sequenase Version 2.0 (TM)

Double stranded DNA was prepared by the Qiagen tip 100 midiprep method (section 2.19.2). A sample of this double stranded DNA was used for sequencing using the Sequenase Version 2.0 sequencing kit. The protocol followed was from the "sequencing double stranded DNA" section of the Sequenase booklet. 5ng of sample DNA were made up to a volume of 10µl with sterile distilled water. 10µl of 0.4M NaOH/0.4M Na_2EDTA were then added. This mixture was incubated at 37°C for 30 minutes prior to neutralisation with 2µl of 3M NaOAc (pH 4.8). The DNA was precipitated at -70°C for 30 minutes after the addition of 50µl of cold ethanol. It was sedimented by centrifugation, drained and rinsed with 70% ethanol prior to drying under a vacuum. The pellet was then dissolved in 7µl of sterile distilled water and 2µl Sequenase reaction buffer plus 0.5pmol of

Universal or Reverse Universal sequencing primer were added. The subsequent steps were exactly the same as those for sequencing single stranded DNA using the Sequenase kit (section 2.20.2).

2.20.4 Δ*Taq* DNA polymerase cycle sequencing of double stranded DNA

The delta *Taq* cycle sequencing kit (USB) protocol was used without modification. This was as follows:-
delta *Taq* DNA polymerase Version 2.0 was diluted 8 fold in ice cold enzyme dilution buffer (10mM Tris-HCl pH 8.0, 1mM 2-mercaptoethanol, 0.5% Tween-20, 0.5% Nonidet P-40) and stored on ice. A single labelling reaction was then performed for each sample by the addition of 0.5pmol Universal primer to 0.1μg plasmid DNA in a sterile Eppendorf tube. 2μl of reaction buffer (260mM Tris-HCl pH 9.5, 65mM $MgCl_2$), 1μl of dGTP cycle mix, 1μl dCTP cycle mix, 0.5μl of [α-^{35}S] dATP 600Ci/mmol, 2μl of delta *Taq* Version 2.0 (diluted 1/8) and 5μl of H_2O. The samples were overlaid with 15μl of mineral oil and were cycled 45 times at 95°C for 15 seconds and 60°C for 30 seconds.

Termination Reactions

4μl of each of ddGTP, ddCTP, ddATP and ddTTP (15μM each dNTP and 300μM the specific ddNTP) were added to appropriately labelled tubes. These were stored until the labelling reactions were complete. 3.5μl of labelling mix were then transferred to each of the four tubes containing termination mixes. The solutions were then mixed and overlaid with 100μl mineral oil. The samples were then subjected to 45 cycles of 95°C for 30 seconds and 72°C for 2 minutes. When cycling was complete, 4μl of stop solution (95% formamide, 20mM Na_2EDTA, 0.05% bromophenol blue, 0.05% xylene cyanol FF) were added to each termination reaction. The samples were then heat denatured and electrophoresed as described in section 2.20.2. The resulting gel was exposed to X-ray film (Amersham Hyperfilm) for four days before developing.

2.20.5 Sequencing of Dynabead immobilised biotinylated PCR products

PCRs were carried out as described in section 2.7.2 and they were then used in direct sequencing. PCR products were ethanol precipitated (section 2.8) and

resuspended in 80µl of TE (pH 7.6) (section 2.2). 40µl of this product were then used to sensitise 20µl of Dynabead suspension (Dynal Inc.) according to the Dynal protocol as follows:-
A tube containing 20µl of Dynabeads M-280 Streptavidin$^{(TM)}$ (10µg/µl) was placed in the magnetic particle concentrator (MPC) and the supernatant removed with a pipette while keeping the tube in the magnet. The beads were resuspended in 20µl of 1x BW buffer (section 2.2) and mixed gently. The supernatant was removed as before. 40µl of the prewashed beads were added to 40µl of the PCR amplified product and incubated for 15 minutes at room temperature with mixing. The tube was placed in the MPC and the supernatant removed. The beads were washed with 40µl of 1x BW buffer and resuspended in 8µl of freshly prepared 0.1M NaOH. This was incubated at room temperature for 10 minutes. The tube was placed in the MPC and the NaOH removed. The beads were washed once with 50µl 0.1M NaOH, once with 40µl 1x BW buffer and once with 50µl TE buffer (section 2.2). The supernatant was removed and the beads were resuspended in 17µl H_2O. 8.5µl of this were used for sequencing by the Sequenase method (section 2.20.1).

The NaOH-denatured, unbiotinylated complementary strand was ethanol precipitated (section 2.10) and resuspended in 10µl of TE (pH 7.6) (section 2.2). This was analysed on a 1% agarose/TBE gel (section 2.8) to confirm that dissociated single strands had been recovered.

2.20.6 Automated sequencing of DNA

This technique involved the use of the *Taq* DyeDeoxy Terminator Cycle Sequencing Kit (Applied Biosystems Inc.) and the Applied Biosystems Model 373A automated sequencer. Automated sequencing was performed according to the manufacturer's instructions. Reaction mixes containing target DNA and sequencing primer in the quantities shown below were prepared in a total volume of 10.5µl.

single stranded DNA target
5µl of 100ng/µl single stranded DNA + 0.8pmol primer

double stranded DNA target
5µl of 200-250ng/µl double stranded DNA + 3.2pmol primer

PCR product target
40ng of PCR product + 3.2pmol primer

9.5µl of reaction mix were then added to the sample. The reaction mix is supplied by ABI (Applied Biosystems Inc.) and sufficient mix for one sample contains 4µl 5x TACS buffer (400mM Tris-HCl pH 9.0, 10mM $MgCl_2$, 100mM $(NH_4)_2SO_4$), 1µl dNTP mix (750 µM dITP, 150µM dATP, 150µM dTTP, 150µM dCTP), 1µl of each dyedeoxy terminator (ddATP, ddGTP, ddTTP, ddTP) and 0.5µl AmpliTaq DNA polymerase (8U/µl). The 20µl sample was then overlaid with 40µl of mineral oil and placed in a pre-heated thermal cycler at 96°C. The samples were then thermally cycled using 25 rounds of 96°C for 30 seconds, 50°C for 15 seconds and 60°C for 4 minutes. The samples were then cooled rapidly to 4°C.

After cycling, 80µl of H_2O were added to the sample prior to centrifugation at 12000g for 10 minutes. The sample was removed from under the oil and 2 phenol/chloroform extractions were carried out to remove excess dye. The sample was then ethanol precipitated (section 2.10) and dried under a vacuum prior to resuspension in 4µl of deionised formamide and heating to 95°C for 2 minutes. The sample was then loaded onto an appropriately prepared sequencing gel (section 2.20.2).

2.20.7 Use of sensitized Dynabeads for automated sequencing

This was carried out according to the Applied Biosystems User Bulletin number 21. 20µl of Dynabeads M-280 Streptavidin (Dynal) were transferred to an Eppendorf tube and placed in the MPC. The supernatant was removed and the beads were washed in 1 volume of TTL buffer (section 2.2). They were then resuspended in 1 volume TTL buffer. The biotinylated PCR products (section 2.7.2) were bound onto the beads by adding 20µl of the washed beads to each 40µl of PCR product, followed by incubation for 15 minutes at 37°C. The beads were then washed twice in 100µl of TT buffer (section 2.2). The liquid was removed, the beads resuspended in 100µl 0.15M NaOH and incubated for 5 minutes at room temperature to denature the double stranded PCR product. The supernatant was removed. The beads were then washed briefly in 0.15M NaOH and neutralised by washing twice in 100µl TT buffer. The beads were then resuspended in 20µl of distilled water and 10µl sensitised Dynabeads were mixed with 1.5µl of 0.5pmol/µl sequencing primer. This was then used for automated sequencing as described in section 2.20.6.

2.21 *in situ* hybridization

This procedure was carried out at the Duncan Guthrie Institute for Medical Genetics, Yorkhill, Glasgow.

Probe labelling by nick translation (BRL nick translation kit)

The following reagents were added sequentially to an Eppendorf tube:- 5µl BRL kit solution A4 (section 2.2), 1µg of DNA to be labelled, 2.5µl 0.4mM Biotin-11-dUTP and water to make the volume to 45µl. 5µl of solution C (section 2.2) were added and this was mixed and incubated for 90 minutes at 15°C. The reaction was stopped by the addition of 5µl of solution D (section 2.2). The probe DNA was then precipitated by the addition of 4.6µl 3M sodium acetate (pH 5.2), 1µl 20µg/µl glycogen and 122µl iced ethanol, and the sample was vortex mixed. This was then centrifuged for 30 minutes at 12000g. The supernatant was removed and the pellet dried under vacuum. 10µl TE buffer (section 2.2) were added to give a final probe concentration of 100ng/µl. The sample was then vortex mixed and centrifuged briefly, before incubating at room temperature for 2 hours. Occasional vortexing and centrifugation were performed to ensure solution of the pellet, prior to storage at -20°C.

Chromosome banding

Chromosome metaphase spreads were prepared on glass microscope slides by standard cytogenetic methods. The chromosome spreads were treated with 1% Lipsol for 10-15 seconds and rinsed with saline (0.95% w/v). 1: 4 (v:v) Leishmann's stain in pH 6.8 buffer was added for 2.5 minutes, then the samples were rinsed with pH 6.8 buffer. After banding, the slides were mounted in buffer, blotted dry and photographed. Well-banded and -spread metaphases within 20 Vernier scale units along the slide length were chosen for hybridization (this represents an area of approximately $2cm^2$). These areas on the slides were marked with a diamond pencil. The slides were soaked in buffer until the coverslip could be removed and they were then destained in a 50%, 70%, 90%, 100% ethanol series, and air dried.

Slide preparation (Cambio chromosome painting kit)

Fresh slides were prepared from 3:1 methanol:acetic acid fixed preparations and air dried. A $2cm^2$ area was selected and marked for hybridization. The chromosomes were fixed for 1 hour in 3:1 methanol:acetic acid (v/v) and air dried. They were then dehydrated by serial ethanol washing (70%, 70%, 90%, 90%, 100%) and dried overnight at 42°C. The appropriate volume of paint was denatured (15µl per slide) by incubation at 65°C for 10 minutes. The paint was transferred to a waterbath at 37°C and incubated for 1 hour. Meanwhile, the chromosomes on the slides were denatured by incubation in 70% (w/v) formamide in 2x SSC (section 2.2) at 65°C for 2 minutes. The chromosomes were then quenched in ice cold 70% ethanol and dehydrated using ethanol as above, before air drying. 15µl of probe were pipetted onto each slide. The slides were overlaid with a 22x22mm coverslip, sealed with cow gum and incubated overnight at 42°C.

Detection

500ml of wash solution A were prepared (100ml 2x SSC (section 2.2), 250µl Tween 20, 25g dried skimmed milk) and centrifuged at 400g for 10 minutes. Three Coplin jars of formamide: 2x SSC (1:1) were prewarmed at 42°C. The dried cow gum was carefully removed from the slides and they were rinsed in the first jar of 2x SSC. The slides were washed twice by incubating for 5 minutes at 42°C in each of the two jars containing formamide: 2x SSC (1:1), followed by two 5 minute washes at 42°C in each of the two jars containing 2x SSC. They were then incubated in wash solution A at 37°C for 30 minutes. Meanwhile, a 500 fold dilution of detection layer 1 (fluorescein avidin 2mg/ml (Sigma)- 0.25% in wash A) and a 250 fold dilution of detection layer 2 (biotinylated anti-avidin 2mg/ml (Sigma)- 0.25% in wash A) were prepared. These were incubated at room temperature for a minimum of 10 minutes, centrifuged for 10 minutes at 12000g and the supernatants were retained for use. After the slides had been incubated, 100µl of the diluted detection layer 1 were pipetted onto the slides and they were covered with a glass 32x22mm coverslip. These were incubated at 37°C for 15-20 minutes in a humidified box. The slides were then washed three times in wash solution A at 42°C for 5 minutes. Detection layer 2 was added and the slides were incubated at room temperature for 15-20 minutes in a humidified box. They were then washed twice in wash solution A at 42°C for 5 minutes. The slides were again incubated with diluted detection layer 1 at room temperature for 15-

20 minutes in a humidified box and were then rinsed twice in wash solution A at 42°C for 5 minutes. They were then rinsed twice in 4x SSC (section 2.2) plus 0.05% (v/v) Tween 20 for 5 minutes at room temperature. 1µl of counterstain 1 (DAPI (Sigma) 40µg/ml diluted 1/50 in Citifluor AF1) was further diluted with 9µl of distilled water. 1µl of counterstain 2 (PI (Sigma) 20µg/ml diluted 1/50 in Citifluor AF1) was further diluted with 49µl of distilled water. 6µl of diluted counterstain 1 and 6µl of diluted counterstain 2 were then added to 200µl of mountant. The slides were dehydrated using an ethanol series, air dried and mounted by adding 20µl of the mountant/counterstain mixture. They were overlaid with a 32x22mm coverslip and sealed with nail varnish. Slides can be stored in the dark at 4°C and viewed using standard epifluoresence filters for FITC and for counterstains (propidium iodide).

2.22 Preparation of total genomic DNA

2.22.1 Isolation of total genomic DNA from whole blood I Herrmann *et al.*, 1987

30ml of lysis buffer (section 2.2) were added to 10ml of anticoaggulated whole blood in a sterile 50ml plastic tube. The sample was stored on ice for 15 minutes, with occasional gentle shaking. The mixture was then centrifuged for 10 minutes at 650g in a table top centrifuge. The supernatant was discarded, the pellet resuspended in 10ml of lysis buffer and stored on ice for 5 minutes. The centrifugation step was repeated and the resulting pellet was washed once with 5ml of lysis buffer and resuspended in 4.5ml of 75mM NaCl, 25mM Na_2EDTA (pH 8.0). To the homogeneous suspension, 0.5ml of proteinase K (10mg/ml) and 0.25ml 20% SDS were added at room temperature. The tube was mixed gently and incubated at 37°C overnight on a slowly rocking platform. This gave a clear solution after proteinase K digestion. It was transferred into a flat sided glass bottle with a tight cap and and 20ml of Tris-HCl equilibrated phenol (pH>7.6) (Fisons) was added. The bottle was placed on a rocking platform for 1-3 hours at room temperature. The mixture was centrifuged in a sterile 50ml plastic tube in a table top centrifuge for 10 minutes at 1500g to separate the phases. The lower phenol phase was removed and discarded. The DNA containing aqueous phase was transferred to a Corex (TM) tube and centrifuged at 7500g for 20 minutes at 25°C (to avoid precipitation of the SDS). Protein and undissolved materials were precipitated. The homogeneous supernatant was then dialysed twice, for 6 hours, against TE (pH 7.6) (dilution factor 1000) first at room temperature to

avoid SDS precipitation, secondly at 4°C. After dialysis, 0.8 volumes of propanol and sodium acetate (pH 6.5) to a final concentration of 0.3M were added. The tube was inverted gently and repeatedly and the precipitate of DNA was spooled out and washed sequentially in 70%, 80%, 90%, and 100% ethanol. The DNA was air dried for 5 minutes before dissolving in TE buffer (pH 7.6) (section 2.2).

2.22.2 Isolation of total genomic DNA from whole blood II
Miller *et al.*, 1988

10ml of fresh whole blood were added to 40ml of blood lysis buffer (section 2.2). The mixture was centrifuged at 1200g for 10 minutes at 4°C. The supernatant was discarded and the tube was allowed to stand for 1 minute to allow drainage of residual liquid. The cell pellet was gently resuspended in 3ml of nuclear lysis buffer (section 2.2). 200µl of 10% SDS solution and 100µl of proteinase K (10mg/ml) were added and the mixture was incubated at 37°C for 18-24 hours. After incubation, 1ml of 5M NaCl was added. The mixture was then extracted twice with an equal volume of Tris-HCl equilibrated phenol (pH>7.6) (Fisons) (v/v) and once with chloroform: isoamyl alcohol (49:1 v/v). 2.5 volumes of ice cold ethanol were then added to the supernatant and the DNA was spooled out. The DNA was then washed in 100% ethanol before drying and resuspension in 100µl TE buffer (section 2.2). The concentration of the DNA was estimated by A_{260} measurement.

2.23 Southern blotting
Southern, 1975, Sambrook *et al.*, 1989

2.23.1 Blot preparation

10µg total genomic DNA were digested overnight at 37°C with approximately 60 Units of each restriction enzyme in a final volume of 100µl (section 2.12). 10µl of the mixture were then used to digest 1µg lambda DNA (Gibco BRL) for 90 minutes at 37°C to confirm sufficient enzyme activity remained to allow complete digestion of the total genomic DNA. Confirmation of the completion of the λ DNA digestion was on a 1% agarose/TBE gel (section 2.8) after the addition of 2µl TBE loading dyes to the sample (section 2.2). The digested total genomic DNA was electrophoresed on a 1% agarose/TBE gel (section 2.7.6). 2µg equivalent of the starting DNA (i.e. 20µl of digest) were loaded in each track. After electrophoresis, the gel was soaked in 0.2M HCl for 10 minutes to nick the DNA

by depurination. The gel was washed briefly in deionised water prior to denaturing the DNA by soaking the gel for 45 minutes in 1.5M NaCl/ 0.5M NaOH with gentle aggitation. The gel was rinsed in deionised water before neutralising for 30 minutes in 1M Tris-HCl (pH 7.4), 1.5M NaCl at room temperature with mixing. The buffer was then changed to a fresh solution and the gel was soaked for a further 15 minutes. While the gel was neutralising, a piece of Whatman 3MM paper was placed over an inverted gel casting stand (the paper must overhang at both ends). This was placed in a large plastic box and 10x SSC (section 2.2) was poured into the box until the level of the liquid almost reached the top of the support. When the 3MM paper was thoroughly dampened, air bubbles were smoothed out with a glass rod. A piece of Hybond N+ (Amersham) nylon filter slightly larger than the gel was cut and dampened in deionised water. The filter was then immersed in 2x SSC (section 2.2) for at least 5 minutes. The gel was removed from the neutralisation solution and inverted. It was placed on the support and centred on the wet 3MM paper, making sure that there were no air bubbles. Parafilm $^{(TM)}$ was placed around the gel to prevent liquid flowing directly from the reservoir to the paper towels placed on top of the gel. The wet nylon filter was placed on top of the gel making sure there were no air bubbles. Two pieces of 3MM paper, cut to the same size as the gel, were soaked in 2x SSC (section 2.2) and placed on top of the wet filter. A stack of paper towels (5-8cm high) were then placed on top. A glass plate was placed on top of the towels and weighed down. This was left for 18-24 hours. At the end of this time, the paper towels and upper 3MM paper were removed. The filter was then removed and the position of the gel slots was marked with a pencil. The filter was rinsed in 5x SSC (section 2.2) for a maximum of 1 minute. The membrane was then transferred onto dry 3MM paper to allow it to air dry prior to storage at -20^{o}C in an air tight bag.

2.23.2 Probe preparation

Qiagen plasmid minipreps (section 2.19.2) were used for preparation of a labelled DNA fragment for probing the blot described in 2.23.1. The fragment to be used as the probe was extracted from a plasmid clone by restriction enzyme digestion using *Bam* HI and *Sal* I (section 2.12). The resulting fragment was gel purified on a 0.7% agarose gel (section 2.13). The gel band containing the probe fragment was labelled directly in the gel using the Megaprime DNA labelling system (Amersham Inc.). The protocol for the labelling step was as follows:-

The excised gel band was placed in a preweighed Eppendorf tube and water was added at a ratio of 3ml H_2O per gram of gel. This was placed at 100°C for 5 minutes in order to melt the gel and denature the double stranded DNA. 25ng of the probe were then transferred to a fresh Eppendorf tube in a maximum volume of 25µl. 5µl of the supplied primer (random nonamer primers in aqueous solution - details of concentration not supplied) were then added and the volume was made up to 33µl by the addition of sterile distilled water. This was then denatured for 5 minutes at 100°C. 10µl of Megaprime labelling buffer solution 2 (section 2.2), 50µCi of ^{32}P dCTP (3000Ci/mmol) and 2µl of enzyme solution 3 (section 2.2) were then added, giving a final volume of 50µl. This labelling reaction was incubated at 37°C for 30 minutes. The reaction was stopped by the addition of 5µl 0.2M Na_2EDTA and 20µg of sonicated salmon sperm carrier DNA in 10mM Tris-HCl (pH8.0), 1mM Na_2EDTA were then added. The mixture was heated to 100°C for 5 minutes and then used for hybridization to the blot.

The incorporation of radiolabelled dCTP in the probe was measured by placing 0.5µl of the labelling reaction onto a strip of Whatman 3MM paper. The postion of the sample was marked and the lower end of the strip was placed in 0.3M ammonium formate (pH8.0). The ammonium formate was then allowed to diffuse to the upper end of the strip. The strip was then cut in half and the upper and lower portions were placed in separate scintillation vials and counted using the Cerenkov channel of a scintillation counter. The sum of counts from the upper and lower portions was calculated, and then the percentage incorporation was calculated by dividing the number of counts from the bottom portion by the total counts and multiplying by 100.

2.23.3 Probing the blot

Sambrook *et al.*, 1989

Blots and probes were prepared as described in the previous two subsections. The probes were then hybridized to the blots as follows:-
5ml of pre-hybridization solution were prepared (1.25ml 20x SSC (section 2.2), 0.25ml 100x Denhardt's solution (section 2.2), 0.25ml 10% SDS). 0.5mg of sonicated salmon sperm DNA were heat denatured at 100°C for 5 minutes. This was chilled on ice before it was added to the pre-hybridization solution in a hybidization bottle. The filter was placed in the bottle and prehybridized in a Hybaid mini oven at 65°C for one hour. The probe (section 2.23.2) was heated to

100°C for 5 minutes. This was then added to the bottle and hybridization was allowed to proceed at 65°C for 18 hours with constant rotation. Following incubation, the filter was washed twice in 2x SSC (section 2.2) containing 0.1% SDS at room temperature for 10 minutes. It was then washed once in 1x SSC containing 0.1% SDS for 15 minutes at 65°C with shaking and once in 0.1x SSC containing 0.1% SDS at 65°C for 10 minutes. The filter was then wrapped in Saran wrap and exposed to X-ray film (Amersham Hyperfilm) for three days prior to developing.

SECTION 3

Preparation, amplification and cloning of *KEL* cDNA

SECTION 3 Preparation, amplification and cloning of *KEL* cDNA

3.1 Introduction to the preparation of cloned *KEL* cDNA

In an attempt to elucidate some of the intricacies of the expression of Kell system antigens, various molecular biology approaches were employed. The initial steps taken in this investigation involved the synthesis of cDNA from RNA isolated from donor samples of known Kell phenotype. The resulting material was used for the majority of the work carried out thereafter. This section describes the isolation of total RNA from donor blood samples; the use of this target material for the synthesis of cDNA; the amplification of the cDNA by the polymerase chain reaction and the cloning of the PCR products for sequencing, and for analysis by other techniques which will be described and discussed in subsequent results sections.

3.2 Isolation of total RNA from whole blood

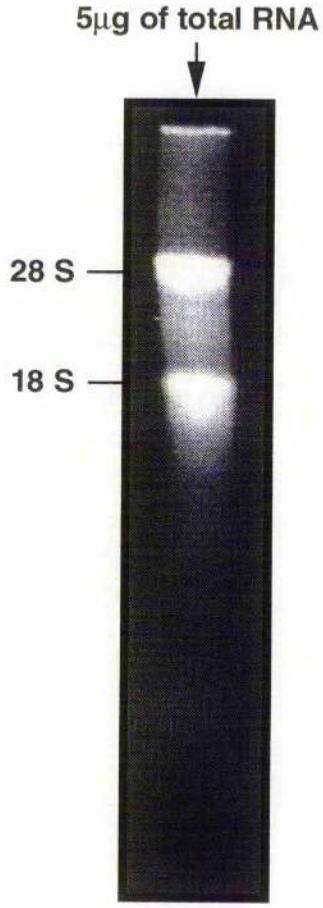
Total RNA was isolated from OKk whole blood as described in section 2.4.1. This phenotype was chosen as the donor was available to donate samples at regular intervals, although a homozygous donor may have been preferable for cloning of PCR products to ensure that the gene cloned was of known type. This has been discussed in section 8. The total RNA obtained was analysed on a formaldehyde denaturing gel (section 2.4.2) to ensure that the quality was suitable for use as a target in reverse transcription reactions (Figure 3.1). The presence of the two ribosomal RNA bands (18S and 28S) demonstrated the suitability of the sample for subsequent reverse transcription, as this showed that the RNA had not degraded to a significant extent during preparation.

3.3 Reverse transcription of mRNA

The resulting total RNA was used as a target in reverse transcription, employing either of the two techniques described in sections 2.5.1 and 2.5.2. In both of these methods, an oligo dT primer was annealed to the poly-A mRNA fraction contained in the total RNA and, by use of either Avian myeloblastosis virus (AMV) reverse transcriptase or Moloney murine leukaemia virus (MoMuLV) reverse transcriptase, the annealed primer was extended to produce a complementary first strand cDNA. From the work carried out with both of

Figure 3.1 Agarose/formaldehyde gel analysis of total RNA

Figure showing a 5µg sample of total RNA electrophoresed on a 1.5% agarose/formaldehyde gel. The 18S (1.9 kb) and 28S (5.1 kb) bands of ribosomal RNA are marked.


5μg of total RNA
28 S
18 S

these reverse transcription kits, it became clear that AMV reverse transcriptase was more suitable.

The reason for this was that cDNA molecules could be generated from both the 3' and 5' halves of the *KEL* gene using AMV reverse transcriptase but only the 3' half length fragment could be reverse transcribed using MoMuLV reverse transcriptase. The most likely explanation for this is that, as AMV reverse transcriptase can be used at higher temperatures (42°C compared with 37°C for MoMuLV reverse transcriptase), secondary structure is reduced within the target molecule and the probability of generating longer cDNA molecules is increased. This would suggest that there is sufficient secondary structure within the 5' half of the *KEL* mRNA to inhibit reverse transcription by MoMuLV reverse transcriptase (section 8).

3.4 Generation of PCR products encoding the Kell protein

The first strand cDNA described above was used directly as a target in the polymerase chain reaction (section 2.7.1). The fragments amplified from the Kk first strand cDNA corresponded to two approximately equally sized fragments (1157bp and 1208bp) which are referred to as the 3' and 5' halves (Figure 3.2). The oligonucleotide primers used for these amplifications are detailed in Table 3.1. These two fragments have a central 87bp overlap (Figure 3.2). The PCR products generated were analysed on a 1% agarose/TBE gel (section 2.8) allowing the direct analysis of each PCR product with respect to both yield and size of the amplified fragment. PCR products of good quality were subjected to further analysis by restriction enzyme digestion and cloning/DNA sequencing.

3.5 Analysis of PCR products encoding segments of the Kell protein

In order to confirm that the amplified PCR products corresponded to those of *KEL* cDNA, the presence of restriction sites predicted from the published sequence was investigated (Lee *et al.*, 1991). The PCR products were gel purified on a 0.7% low melting point agarose gel (section 2.9). The gel bands containing the putative *KEL* fragments were excised and DNA was then extracted from the gel bands (section 2.9). The resulting samples were then digested with *Bgl* II for the 3' half or *Pst* I for the 5' half, generating two and three fragments respectively. The digests were analysed on 1% agarose/TBE gel (Figure 3.3).

Table 3.1 Oligonucleotide primers used for the amplification of *KEL* cDNA by PCR

PRIMER NUMBER	SEQUENCE	POSITION
Primer 48	ccgtcgacGAAGGATTCTGGAGCCACAGAAG *Sal* I	93-115
Primer 54	ccggatccCTTTCTGCGTGCCTCCTGGAATT *Bam* HI	1284-1262
Primer 49	ccgtcgacACATGATCTTAGGGCTGGTGGTG *Sal* I	1214-1236
Primer 51	ccggatccATTTCTGTGCTGTGGCATCTTTG *Bam* HI	2354-2332

The primer pairs used in PCR were primer 48 plus primer 54 and primer 49 plus primer 51. These amplified the 5' and 3' *KEL* cDNA halves respectively. All oligonucleotide primers had restriction enzyme sites incorporated in the 5' termini. These sites contain the six nucleotide recognition sequence for *Bam* HI or *Sal* I and an extra two nucleotides to optimise the chance of the restriction site being generated fully in PCR. The eight extra nucleotides are shown in lower case letters. Position refers to the *KEL* cDNA nucleotide sequence numbered as previously reported by Lee *et al.*, 1991.

Figure 3.2 Map of *KEL* fragments generated by PCR

Illustration of the two half length fragments of *KEL* cDNA generated by PCR. Oligonucleotide primers and their directions are shown by arrows. The central 87bp overlap is illustrated by the shaded box. PCR fragments are illustrated by hatched boxes and are labelled with their corresponding sizes.

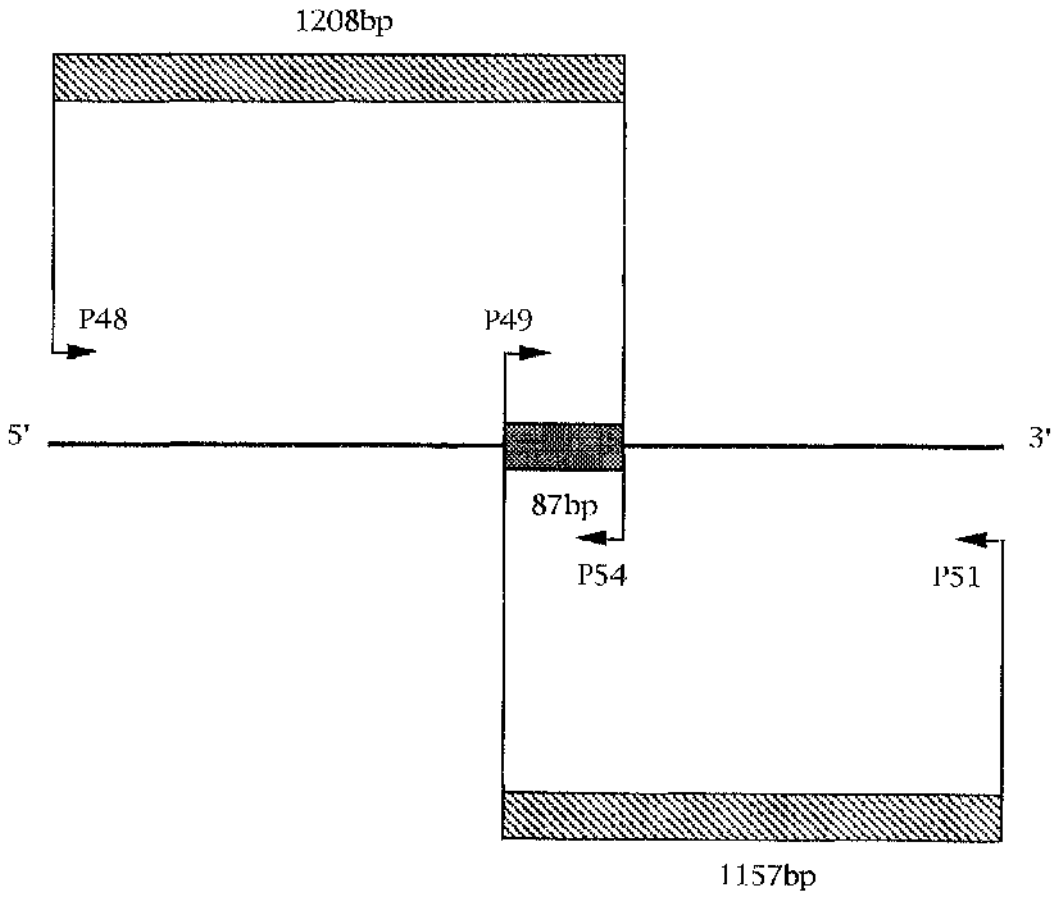
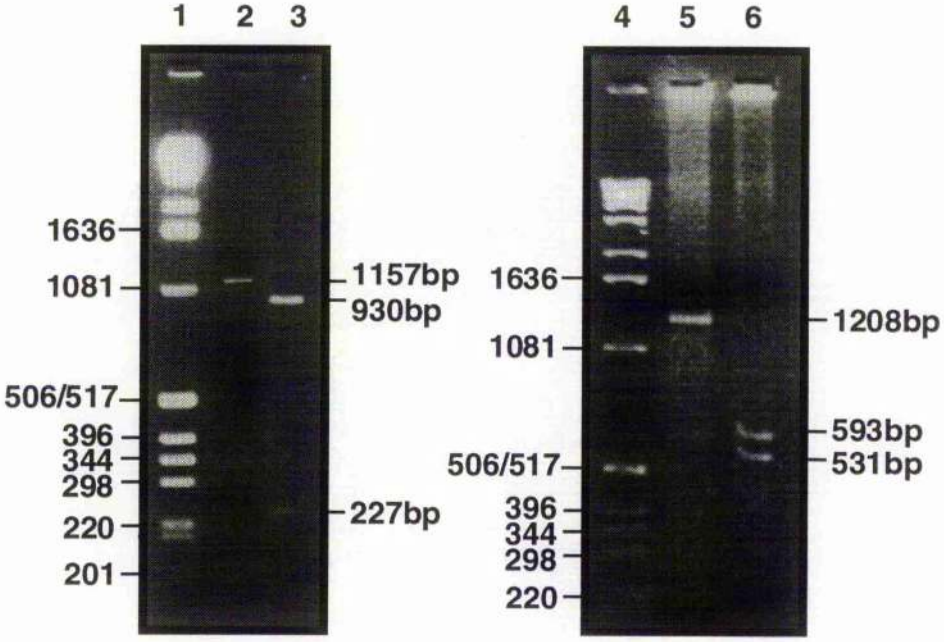

1208bp
P48
P49
5'
3'
87bp
P54
P51
1157bp

Figure 3.3 Gel analysis of restriction enzyme digests of PCR products

Figure illustrating agarose gel analysis of restriction enzyme digests of putative *KEL* cDNA fragments. Tracks 1 and 4 are 1 kb DNA ladder. Track 2 is undigested 3' half length fragment and track 3 is *Bgl* II digested 3' half length fragment. Track 5 is undigested 5' half length fragment and track 6 is *Pst* I digested 5' half length fragment. Marker and fragment sizes are labelled.


1
2
3
1636
1081
506/517
396
344
298
220
201
1157bp
930bp
227bp
4
5
6
1636
1081
506/517
396
344
298
220
1208bp
593bp
531bp

The resulting fragments were compared with size markers and the approximate sizes were calculated by interpolation on a graph of $\log_{10}$ apparent molecular weight v's distance migrated for the size markers. This gave good correlation between the calculated and expected sizes (Table 3.2).

The restriction enzymes used were chosen according to a computer search of the restriction enzyme sites present within the published *KEL* cDNA sequence. Digestion with these enzymes yields fragments of sufficiently large and varied size to give a diagnostic test for the origin of the PCR product. When the estimated fragment sizes correlated closely with the expected sizes, allowing for small errors which arise from the measurement of the extent of band migration of samples and standards, this was a good indication that the PCR product was from the desired source. It is unlikely that material derived from another source would give exactly the same banding pattern when digested with the enzymes chosen.

3.6 Cloning of PCR products encoding segments of the Kell protein

After confirmation of the identity of PCR products by restriction enzyme digestion, they were cloned into the multiple cloning site of the plasmid vectors pTZ18U and pTZ18R (section 2.14). This was achieved by use of the *Bam* HI and *Sal* I restriction enzyme sites incorporated into the PCR product through their presence at the 5' termini of the oligonucleotide primers used in the PCR. The two plasmids were digested with *Bam* HI and *Sal* I, sites for both of which are present in the multiple cloning site of the plasmids (Figures 2.1 and 2.2). The PCR products were similarly digested and then the fragments were ligated into the plasmids using T4 DNA ligase enzyme (section 2.14). The ligation mixes were used to transform competent XL1-blue *E. coli* cells (section 2.16). These transformants were plated out and grown on L-broth agar plates containing ampicillin, X-gal and IPTG (section 2.15). Only colonies of cells which contain plasmid will grow, as the presence of the plasmid imparts antibiotic resistance to the bacterial host cell due to the production of β-lactamase. The cells which contain plasmid lacking insert appear blue due to the formation of β-galactosidase via α-complementation. Cells which contain recombinant plasmid do not produce β-galactosidase and appear white. It was therefore possible to isolate and purify transformant cells containing the cloned product. In order to

Table 3.2 Calculation of restriction enzyme digest fragment sizes

FRAGMENT DIGESTED	EXPECTED FRAGMENT SIZE (bp)	ESTIMATED FRAGMENT SIZE (bp)
3' half + *Bgl* II	227	240+/-24
	930	1050+/-105
5' half + *Pst* I	84	-
	531	540+/-54
	593	630+/-63

The 84bp fragment size could not be estimated by this method due to the poor resolution and staining of this small fragment on a 1% agarose/TBE gel.
A 10% error margin has been allowed in the estimation of fragment size.

confirm that the selected plasmids actually contained the required insert, each plasmid was isolated (section 2.19.1) and fractions of the samples were digested with *Bam* HI and *Sal* I. The digests were then analysed on a 1% agarose/TBE gel (section 2.8). Typical results are shown in Figure 3.4. The plasmids which contained inserts generated two bands, one corresponding to the insert and which was the same size as the original cloned fragment, and the second corresponding to the linearised plasmid. In the case where a PCR product had not been inserted into the plasmid and where the colony appeared white, only one band of linearised plasmid was evident on the gel. Such apparent plasmid tranformants may arise as a result of failure of the *lac* system through mutation in plasmid or chromosomal genes. It should be noted that, if the plasmid was incompletely digested, then extra bands of uncut circular plasmid DNA should also be apparent on the gel. Transformants containing plasmids with the appropriate insert were then stored in glycerol at -70°C for future use (section 2.18).

3.7 Partial sequencing of cloned PCR products

Clones of each of pTZ18U containing 3' half and pTZ18U containing 5' half were used to produce single stranded DNA through infection of the cells with M13KO7 helper phage. This allows preferential secretion of single stranded vector DNA for use in sequencing reactions (section 2.20.1). The plasmid pTZ18U yields the positive strand of the DNA when infected with helper phage. The single stranded DNA was used directly in Sanger dideoxy sequencing (section 2.20.2) using Universal sequencing primer, or a *KEL* sequence specific primer. Approximately 200 nucleotides of each cloned fragment were sequenced (Figure 3.5). This confirmed that the cloned materials were of *KEL* origin, as the experimental sequence was the same as the published sequence with the exception of a single nucleotide at position 1391 where a C on the published sequence was replaced by a T in the experimental sequence (Figure 3.5).

3.8 Discussion

The initial step in this project involved the isolation of total RNA from samples of whole blood. Several methods were assessed before the AGPC method (section 2.4.1) was chosen for routine use. Other techniques and the reasons for this choice are discussed more fully in section 8. The AGPC method is a single step procedure incorporating guanidinium thiocyanate (a strong inhibitor of

Figure 3.4 Gel analysis of restriction enzyme digested transformant plasmid preparations

Figure illustrating agarose gel analysis of intact pTZ18U plasmid (track 2), *Bam* HI/*Sal* I digested pTZ18U plasmid (track 3) and *Bam* HI/*Sal* I digest of pTZ18U transformant containing the *KEL* 3' half length fragment (track 4). 1 kb DNA ladder is included as a size marker (track 1).

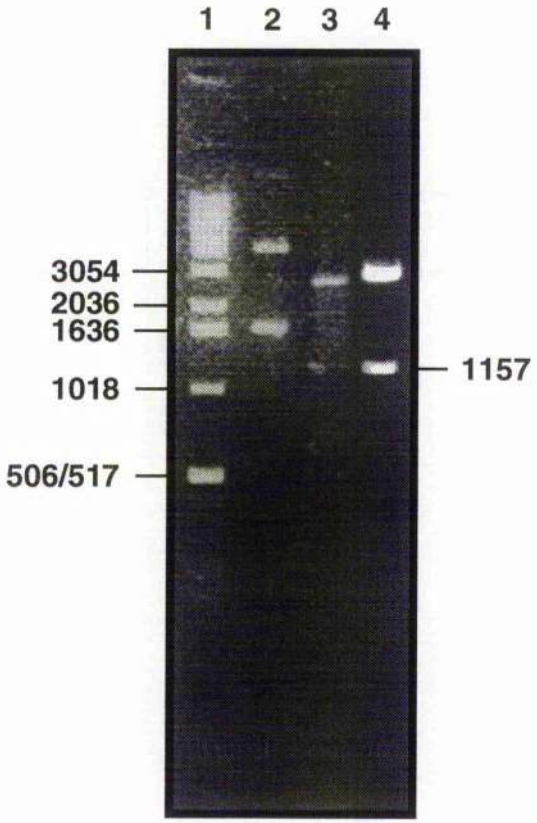

1
2
3
4
3054
2036
1636
1018
506/517
1157

Figure 3.5 Partial sequencing of cloned 3' half length *KEL* PCR products

Partial sequencing of 3' half length *KEL* cDNA fragment cloned in pTZ18U. Positions refer to the *KEL* cDNA nucleotide sequence numbered as previously reported by Lee *et al.*, 1991. The sequence data relating to this figure is detailed below.

Published sequence

```
1214 ACATGATCTTAGGGCTGGTGGTGACCCTTTCTCCAGCCCTGGACAGT 1260
     ----------------------------------------------
1214 ACATGATCTTAGGGCTGGTGGTGACCCTTTCTCCAGCCCTGGACAGT 1260
```

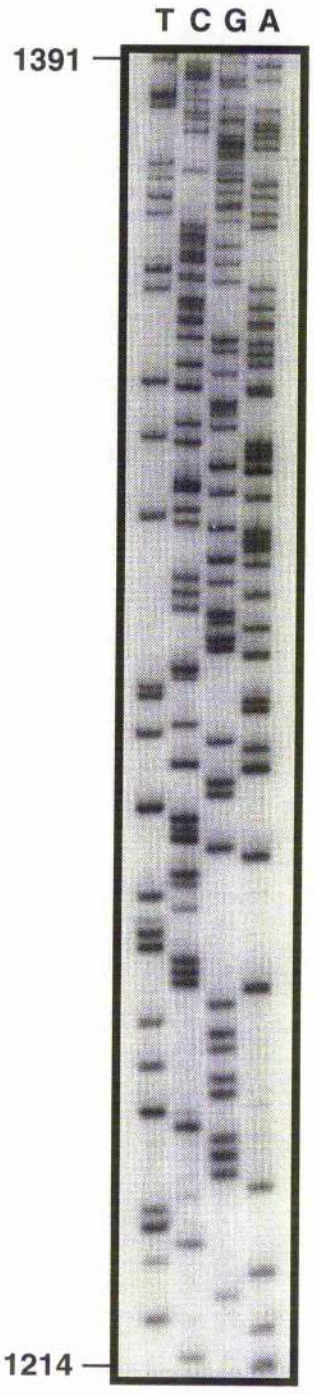
Experimental sequence

```
1261 CAATTCCAGGAGGCACGCAGAAAGCTCAGCCAGAAACTGCGGGAACT 1307
     ----------------------------------------------
1261 CAATTCCAGGAGGCACGCAGAAAGCTCAGCCAGAAACTGCGGGAACT 1307


1308 GACAGAGCAACCACCCATGCCTGCCCGCCCACGATGGATGAAGTGCG 1354
     ----------------------------------------------
1308 GACAGAGCAACCACCCATGCCTGCCCGCCCACGATGGATGAAGTGCG 1354


1355 TGGAGGAGACAGGCACGTTCTTCGAGCCCACGCTGGC 1391
     ---------------------------------*
1355 TGGAGGAGACAGGCACGTTCTTCGAGCCCACGCTGGT 1391
```

* denotes discrepancy between published and experimental sequences.


T C G A
1391
1214

RNase) with sodium acetate, phenol and chloroform to allow the isolation of total RNA from the sample. This technique gave high yields of total RNA (approximately 2μg per ml of whole blood used). Problems were encountered, with respect to RNase contamination, at a later stage in this project and these are discussed in section 8.

The generation of product by PCR using first strand cDNA as a target proved a reliable and reproducible technique, and the resulting products were generally of high yield (approximately 3-5μg per 100μl reaction) and good quality with no non-specific bands being generated as a result of mispriming or poor target quality.

Cloning of *KEL* PCR products into suitable plasmids was prerequisite for the work described in sections 4 and 5. However, this is not the most appropriate approach for the sequence analysis of *KEL* PCR products. The reason for this is that, during the cloning process, a single PCR generated molecule is incorporated into each individual plasmid clone. This gives rise to problems as it is impossible to determine whether differences between the experimental and published sequences are true sequence differences or have arisen as a result of *Taq* DNA polymerase misincorporations. This is apparent from the one difference determined for the partial sequence described above, where there is a C -> T nucleotide substitution at position 1391 of the 3' half length fragment. This would result in an amino acid change from alanine to valine. Valine is a small nonpolar hydrophobic amino acid, similar to alanine in its characteristics. One possible explanation for this difference is that a natural mutation has resulted in a conservative amino acid substitution in the sequenced sample. This mutation may not alter structure/function to any great extent. A second possible explanation is that lack of fidelity of *Taq* DNA polymerase has resulted in an error in the sequence. It is well documented that *Taq* DNA polymerase enzyme, when used under normal PCR conditions, will give rise to misincorporations at a rate of one in every 400 nucleotides (Saiki *et al.*, 1988). As a result, one might expect around 3 misincorporations in each of the two half fragments, based on the length of the two fragments being amplified. In order to determine the true source of the differences, many individual clones must be sequenced. Therefore, it is more appropriate to sequence directly a population of PCR products, where the average of the sequence results is determined. This greatly reduces the problem of determining true sequence differences. This is discussed in more detail in section 8.

SECTION 4

Locating the *KEL* gene by analysis of somatic cell hybrid PCR products and by *in situ* hybridization

SECTION 4 Locating the *KEL* gene by analysis of somatic cell hybrid PCR products and by *in situ* hybridization

4.1 Introduction to PCRs using somatic cell hybrid DNA targets

The polymerase chain reaction is an extremely versatile technique and, in the work described in this section, it has been used for the amplification of an 87bp product from somatic cell hybrid DNA samples. The somatic cell hybrids used are cells which contain a combination of chromosomes from human and another mammalian source. DNA preparations were performed on a variety of different somatic cell hybrid samples, each with different human chromosome complements. These were used to determine which chromosome carries the *KEL* gene. Some of the DNA samples used contained chromosomes which had been translocated i.e. some of the individual chromosomes are partly human and partly rodent. In these cases the junction between the human and rodent fragments had been defined. In order to generate PCR products using this type of target material, oligonucleotide primers were used which were likely to amplify a fragment contained within a single exon (Table 4.1). In the absence of genomic sequence data, this was predicted by comparison of the *KEL* cDNA sequence with the gene structure of CALLA neutral endopeptidase. These two genes have been shown to have sequence homology thus giving a guide as to where introns might arise in the *KEL* gene (D'Adamio *et al.*, 1989). PCR products could only be amplified where the part of the human chromosome carrying the *KEL* gene was present. The experimental results can be easily interpreted, as only the samples containing the appropriate gene will give rise to a PCR product. The use of somatic cell DNA samples containing translocations allow refinement of the chromosomal location, as product will not be generated if the part of the human chromosome carrying the gene is missing.

An 87 base pair fragment was generated using these primers (Figure 3.2). A product of 71 nucleotides would be expected based on the annealing positions of the oligonucleotide primers to the published *KEL* cDNA sequence (1214 and 1284). The extra 16 nucleotides arise from the incorporation of the additional 8 nucleotide residues at the 5' end of each oligonucleotide primer.

Table 4.1 Oligonucleotide primers used for the amplification of an 87bp fragment from somatic cell hybrid DNAs by PCR

PRIMER NUMBER	SEQUENCE	POSITION
Primer 49	ccgtcgacACATGATCTTAGGGCTGGTGGTG	1214-1236
Primer 54	ccggatccCTTTCTGCGTGCCTCCTGGAATT	1284-1262

The oligonucleotide primers had restriction enzyme sites incorporated in the 5' termini as they were initially designed for the generation of PCR products for ligation into plasmids (section 3.4). The nucleotides relating to the restriction enzyme sites are in lower case letters. Position refers to the *KEL* cDNA nucleotide sequence numbered as previously reported by Lee *et al.*, 1991.

4.2 Basis for *in situ* hybridization studies

The *in situ* hybridization work for this section of the project was performed at the Duncan Guthrie Institute of Medical Genetics, Yorkhill, Glasgow with the assistance of Mrs Norma Morrison.

Previous published work had provided provisional assignment of the *KEL* gene to chromosome 7q32-q36 (Zelinski *et al.*, 1991 a). This assignment was based on the use of antigenic variation as a marker. Genetic analysis had demonstrated that the *KEL* gene is closely linked to the prolactin-inducible protein (PIP) locus, allowing assignment of the *KEL* locus to chromosome 7 (Zelinski *et al.*, 1991 a). This section describes the use of *in situ* hybridization analysis in conjunction with PCR amplification of a *KEL* specific product from somatic cell hybrid DNA to refine the location of the *KEL* gene on chromosome 7.

in situ hybridization involves the binding of a sequence specific labelled probe to locate a particular gene sequence on human metaphase chromosomes, as described more fully in sections 1.2.3 and 2.21.

4.3 Analysis of somatic cell hybrid PCR products

PCR reactions were performed using a panel of somatic cell hybrid DNA samples which was kindly donated by Dr Nigel K. Spurr, Imperial Cancer Research Fund, Human Genetic Resources Laboratory, Clare Hall Laboratories, Blanche Lane, South Mimms, Herts.

PCR reactions were carried out as described in section 2.7.5. The resulting PCR products were analysed on a 15% polyacrylamide gel (section 2.8) (Figure 4.1). The results obtained were then compared with the chromosomal composition of the somatic cell hybrid DNA samples used as templates (Table 4.2). Table 4.2 details the human chromosomal content of the cell lines used as targets. From these data, the chromosome containing the target sequence for amplification of the 87bp product was determined. These results illustrate that PCR products are present in lanes 1, 5, 6, 7, 10, 11, 12 and 18. Products were only generated from somatic cell hybrid DNA samples which contain human chromosome 7 (Table 4.2). This is discussed in section 4.5.

Figure 4.1 Polyacrylamide gel analysis of somatic cell hybrid DNA PCR products

Polyacrylamide gel analysis of the 87bp product amplified by PCR from human-rodet somatic cell hybrids containing the human *KEL* sequence. Target DNA samples were from (1) cloned *KEL* cDNA, (2) mouse, (3) rat, (4) hamster, (5) human, (6) AMIR2VIII, (7) CTP412E3, (8) DUR4.3, (9) FG10, (10) 3W4C15, (11) MOG2, (12) Horp27-RC114, (13) SIF4A31, (14) F4SC13C112, (15) TWIN19/D12, (16) FIR5R3, (17) FIR5, (18) CLONE21E.

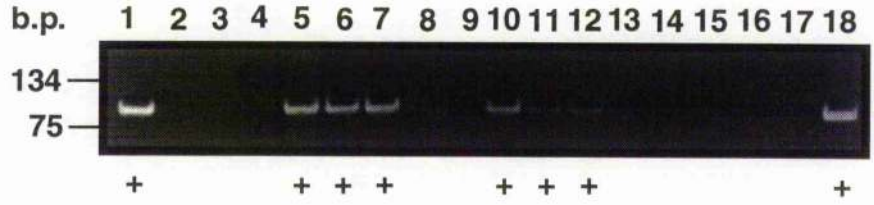

b.p.
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18
134
75
+ + + + + + + +

Table 4.2 Segregation of *KEL* in relation to human chromosomal content in 13 human-rodent somatic cell hybrids

+ = human chromosome present

- = human chromosome not detected

tr = presence of chromosome at levels < 10%

nt = not tested

HYBRID	PCR Product	CHROMOSOME NUMBER																						
		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	X
AMIR 2 VIII	+	+	+	+	+	tr	tr	tr	tr	-	+	+	-	-	+	+	-	+	+	+	+	+	+	a
CTP41E3	+	-	+	tr	-	-	+	+	-	-	-	-	-	-	+	-	-	-	-	-	-	-	-	+
3W4C15	+	-	+	-	-	-	-	+	-	-	+	+	+	-	+	+	-	+	-	-	-	+	-	+
MOG2	+	+	-	+	+	+	+	+	nt	+	nt	+	+	nt	+	+	+	+	+	+	-	+	-	+
Horp 27RC14	+	-	+	-	+	-	-	+	-	-	-	+	+	-	+	+	-	-	-	-	-	+	+	-
CLONE 21E	+	-	-	-	-	-	-	+	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
DUR 4.3	-	-	-	+	-	+	-	-	-	-	+	+	+	+	+	b	-	+	+	-	+	+	-	b
FG10	-	-	+	-	-	+	-	-	+	-	+	-	-	-	-	+	-	-	+	-	-	+	-	+
SIF4A31	-	-	-	+	+	-	+	-	-	-	-	-	-	-	+	-	-	+	-	-	-	-	-	+
F4Sc13C112	-	c	-	-	-	-	-	-	-	-	-	-	-	-	+	-	-	-	-	-	-	-	-	-
TWIN 19/D12	-	+	-	+	-	-	+	-	+	-	-	-	+	-	+	-	+	+	+	-	+	+	+	-
FIR 5R3	-	-	-	-	-	-	-	-	-	-	-	-	-	-	+	-	-	-	tr	-	-	-	-	-
FIR 5	-	-	-	-	-	-	-	d	-	-	-	-	-	-	+	-	-	-	tr	-	-	-	-	+

a X/Y translocation Xqter - p22.3 : Yq1.1 - qter

b X/15 translocation Xq11 - qter : 15p11 - qter

c Contains only 1p

d X/7 translocation Xqter - q13 : 7pter - q22

4.4 *in situ* hybridization using a *KEL* specific probe

The cloned 3' half length *KEL* fragment described in section 3 was used as a probe for *in situ* hybridization experiments with normal male metaphase chromosome spreads, according to the method in section 2.21. The probe was biotinylated by nick translation (section 1.2.3) and used for hybridization to banded metaphase chromosome spreads, as described previously (section 2.21). The use of nick translation allowed the generation of a biotinylated probe containing a random population of sequences of varying length, which constitute the entire *KEL* cDNA. Twenty four normal human male metaphases were scored following hybridization and the positions of 118 signals recorded. A highly significant ($P < 0.005$) 37 signals were located on chromosome 7, with 22 comprising a signal peak on 7q33-35. A second hybridization was performed and the signal distribution at 7q33-35 was noted on 32 chromosomes 7, all of which were sufficiently intense to permit assignment of signals to a single band. These experiments showed clear hybridization of the probe to the chromosome and allowed the refinement of the mapping of the *KEL* gene to chromosome 7q34 (Figures 4.2 and 4.3).

4.5 Discussion

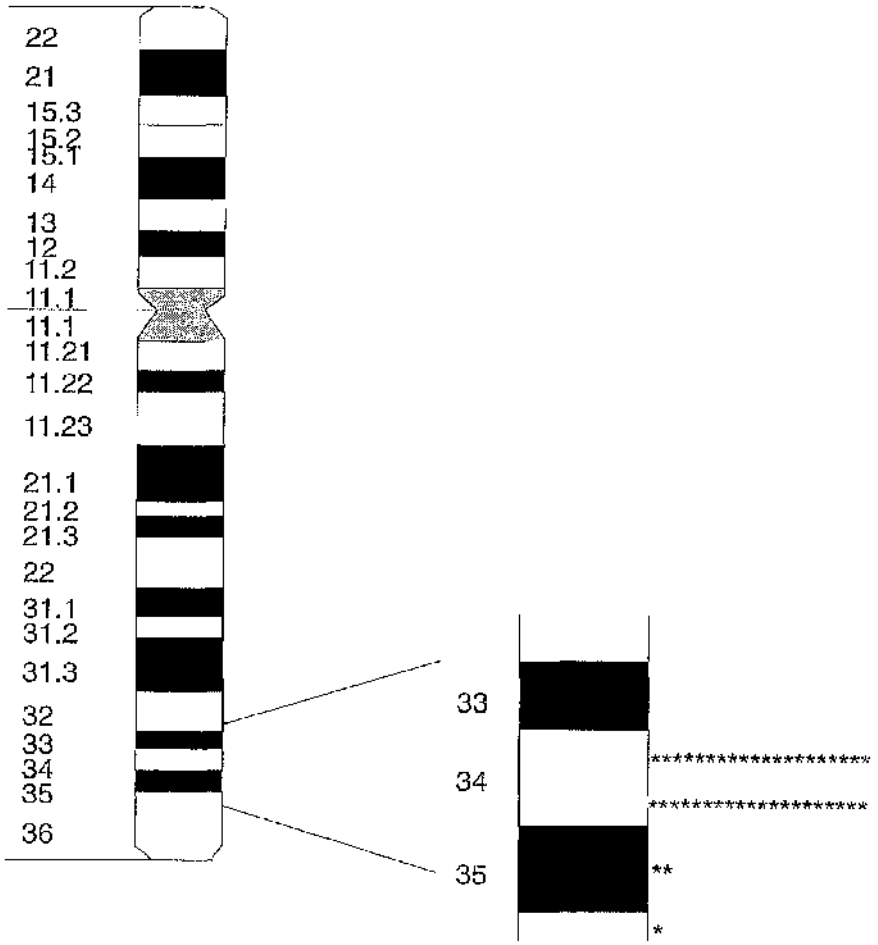
From the results of the PCR experiments (Table 4.2) it was clear that only the somatic cell hybrids which contained intact human chromosome 7 yielded a PCR product, indicating that the gene encoding the Kell blood group polypeptide is located on chromosome 7. One of the somatic cell hybrid samples used contained a translocation in chromosome 7. This was FIR5 which contains a X/7 translocation (X:7) Xqter-q13:7pter-q22 (Hobart *et al.*, 1981). The 87bp *KEL* fragment was not amplified from this sample (Figure 4.1), indicating that the location of the *KEL* gene is 7q22-qter. These findings were reinforced by the *in situ* hybridization results which allowed the position of the gene to be localised to 7q34. This allowed further refinement of the *KEL* chromosomal location from that which had been published previously (Zelinski *et al.*, 1991 a). The initial mapping results published by Zelinski *et al.* were confirmed by the linkage of the *KEL* locus to the cystic fibrosis locus (Purohit *et al.*, 1992). The results obtained by these experiments place the *KEL* locus distal to the cystic fibrosis locus (7q31-32) and proximal to the T cell receptor beta cluster located in 7q35 (Barker *et al.*, 1984).

Figure 4.2 Hybridization of a biotinylated probe to human chromosome 7 *in situ*

Figure illustrating hybridization of a *KEL* specific biotinylated probe to normal male metaphase chromosome spreads. The bright spots represent the position at which the probe has hybridized to both alleles of an individual chromosome. These can be seen on two chromosomes; one to the top right of the chromosome spread and a second in the centre of the spread.

Figure 4.3 **Signal distribution at 7q33-7q35 on 32 copies of chromosome 7 following hybridization to the *KEL* cDNA probe**

Diagrammatic representation of signals detected on 32 chromosomes 7 after hybridization of a biotinylated *KEL* specific probe to normal male metaphase chromosome spreads. Dots indicate the chromosomal positions of each signal.


22
21
15.3
15.2
15.1
14
13
12
11.2
11.1
11.1
11.21
11.22
11.23
21.1
21.2
21.3
22
31.1
31.2
31.3
32
33
34
35
36
33
34
35
****
****
**
*

7

This assignment, using a cDNA probe which encodes the Kell polypeptide, is to the same region of chromosome 7 as indicated by genetic analysis of the antigenic variation associated with the Kell blood group system (Zelinski *et al.*, 1991 b). The co-localisation of the antigenic determinant and the gene locus suggests that the Kell antigenic determinants are part of the polypeptide chain and not the associated sugar molecules, as both localisations are to the *KEL* locus and not to a separate glycosyl transferase gene locus.

In studies performed by Lee *et al.*, the *KEL* locus was mapped to chromosome 7q33 (Lee *et al.*, 1993). This mapping was achieved by hybridization of a radiolabelled *KEL* specific cDNA probe to somatic cell hybrid DNA panels. Genomic clones from a human placental DNA genomic library were used in the isolation of the probes for these *in situ* hybridization experiments. The results obtained located the *KEL* gene to chromosome 7q33. These studies were published simultaneously with the author's work described in this section and the results obtained were virtually identical (Murphy *et al.*, 1993, Lee *et al.*, 1993).

The results of the work described in this section and their implications will be discussed more fully in section 8.

SECTION 5

Southern blot analysis of total genomic DNA with *KEL* specific probes

SECTION 5 Southern blot analysis of total genomic DNA with *KEL* specific probes

5.1 Introduction to the genetic basis for the Kell null phenotype

In a few rare individuals, the antigens of the Kell blood group system are completely undetectable by serological techniques. This phenotype is referred to as the Kell null (K_O) phenotype. Despite the apparent lack of antigen expression in these individuals, work has been published which gives evidence for the presence of Kell protein on the surface of K_O erythrocytes (Merry *et al.*, 1984). However, the genetic basis for this phenotype has not been established. The results described in this section provide evidence that the K_O phenotype, in one particular case, does not arise as a result of a complete or a major gene deletion.

5.2 Analysis of total genomic DNA by Southern blotting.

A sample of whole blood was obtained from one K_O individual and total genomic DNA was prepared by the method described in section 2.22.1. Southern blotting (section 2.23) was then used to compare this sample with samples of genomic DNA derived from normal Kell phenotype individuals. The probes used in this investigation were the two approximately equally sized cloned Kk PCR products, termed the *KEL* cDNA 3' and 5' halves, described in section 3. These two cloned products span the complete coding region of the published cDNA sequence (Lee *et al.*, 1991). The probes were radiolabelled by the random priming technique, using the Megaprime (TM) labelling system described in section 2.23.2. Both undigested and restriction enzyme digested total genomic DNA were electophoresed and Southern blots were prepared as described in section 2.23.1. The resulting blots were probed using the method described in section 2.23.3. Table 5.1 shows the positions of the sites for the restriction enzymes used in this section on the published *KEL* cDNA sequence.

5.2.1 Southern blots of K_O and Kk total genomic DNA digested with *Hpa* II and *Msp* I

The initial results obtained by probing Southern blots of *Hpa* II and *Msp* I digested K_O and Kk total genomic DNA with the 3' and 5' half length fragments

Table 5.1 Positions of restriction enzyme sites present in the *KEL* cDNA sequence

RESTRICTION ENZYME	POSITION OF SITE IN cDNA SEQUENCE
Bam HI	no sites present in cDNA sequence
Sal I	no sites present in cDNA sequence
Pst I	651, 1208, 1550, 1640, 2034, 2063
Hpa II/*Msp* I	378, 1461, 1659
Eco RI	531

Position refers to the *KEL* cDNA nucleotide sequence numbered as previously reported by Lee *et al.*, 1991.
The positions and numbers of restriction enzyme sites for these enzymes within the intron sequences in *KEL* total genomic DNA sequence are unknown.

are shown in Figure 5.1. These indicated that similar banding patterns were obtained for both K_O and Kk samples, but that two extra bands were present in the K_O samples when the 3' probe was used.

5.2.2 Hybridization of the *KEL* 3' probe to blots of undigested K_O and Kk total genomic DNA

In order to establish that the samples of total genomic DNA used for these blots were of good quality and not degraded, undigested genomic DNA from each of the samples was electrophoresed and blotted and the blots were probed with the 3' and 5' fragments. The band sizes were estimated from their relative distances of migration on the gel compared to those of DNA fragments of known size (Figure 5.2).

The results showed that bands were detected when a blot of undigested total genomic DNA from the K_O and Kk donors was probed with the labelled 3' half length fragment. The band sizes obtained from this blot are detailed on Figure 5.2. It is clear that the K_O sample has 4 bands present and the Kk sample has only two bands present, these being equivalent in size to two of the bands present in the K_O sample. These results indicate that random degradation of the total genomic DNA samples is unlikely to account for the extra bands, as hybridzation of a probe to degraded material would not give rise to a sharp band. If these arose as a result of contamination, then a different contaminant would have to be present in each of the samples to give rise to the different banding patterns observed.

5.2.3 Southern blots of several samples of total genomic DNA probed with the 3' *KEL* half length fragment

In an attempt to identify the source of the extra bands which were present on probing blots of the undigested K_O and Kk samples, several samples of DNA from Kk and kk donors were prepared by three individuals using two different techniques in three separate laboratories. Undigested total genomic DNA prepared from the two donors already tested (one sample from the K_O donor and three different samples from the Kk donor) were used in conjuction with samples from two kk donors in the preparation of Southern blots. Unfortunately, repeat samples of blood from the K_O donor could not be obtained at this point. Blots were prepared from a total of seven samples. These were then

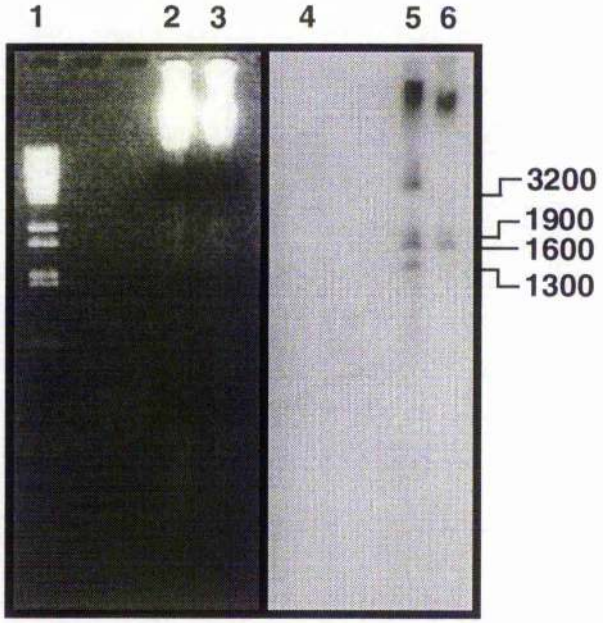
Figure 5.1 Southern blot of *Hpa* II and *Msp* I digested genomic DNA probed with 3' and 5' half length *KEL* fragments

Southern blots of Kk and K_O total genomic DNA digested with *Hpa* II and *Msp* I and probed with the *KEL* 3' and 5' half length fragments. Tracks 1 and 5 are Kk genomic DNA digested with *Hpa* II, tracks 2 and 6 are Kk genomic DNA digested with *Msp* I, tracks 3 and 7 are K_O genomic DNA digested with *Hpa* II and tracks 4 and 8 are K_O genomic DNA digested with *Msp* I. Tracks 1 to 4 were probed with the 3' half fragment and tracks 5 to 8 were probed with the 5' half fragment.


Kk̄
Ko
Kk̄
Ko
1 2 3 4 5 6 7 8
Extra bands in Ko
3' probe
5' probe

Figure 5.2 Southern blot of K_O and Kk undigested genomic DNA probed with the 3' *KEL* fragment

Track 1 is 1 kb DNA ladder and tracks 2 and 3 are Ko and Kk undigested total genomic DNA electrophoresed on a 1% agarose/TBE gel prior to blotting. These samples can be seen to be undegraded. Tracks 4, 5 and 6 are the blotted samples from tracks 1, 2 and 3 probed with the 3' *KEL* fragment.


1
2
3
4
5
6
3200
1900
1600
1300

probed with the 3' half length fragment (Figure 5.3). The 3' *KEL* fragment was removed from the blot prior to hybridization of the 5' *KEL* fragment (Figure 5.4). The results obtained illustrated the fact that the extra bands present in the K_O sample were also detected in samples of DNA prepared from kk individuals. A DNA size marker was electrophoresed on the gel and a track containing gel loading dye and TE buffer was also included to establish that these extra bands had not arisen as a result of a contaminant in the gel loading mixture (Figure 5.4).

The results obtained from these blots indicate that the bands which were detected did not originate from either the loading dyes or the TE buffer used for loading the samples onto the gel, as no bands were detected in this lane of the blot when probed with either the 3' or 5' *KEL* fragment.

When the blot was probed with the 3' fragment (Figure 5.3), the K_O sample gave rise to the same banding pattern as shown in Figure 5.2, but a small (480bp) band was also detected. It is likely that this band was too faint to be detected on the autoradiograph shown in Figure 5.2. The original Kk sample gave the same two bands as those shown in Figure 5.2. The remaining five samples each gave a band of 1900bp. This band was not present in either of the two original samples. It is important to note that two of these five samples originated from the same Kk donor as the Kk sample used in the previous experiments. Two of the kk samples, each from different donors, had faint bands at 480bp (this band was present in the K_O sample). This band was not present in the other sample derived from the same donor.

The results of probing with the 5' fragment showed that the K_O sample had three bands present, only one of which relates to a band present with the 3' fragment (1700bp) (Figure 5.4). The original Kk sample gave a single band at 1700bp. The remaining five samples showed no banding pattern with this probe.

The results from two blots described above (Figures 5.3 and 5.4) rule out the presence of a common contaminant which gives rise to the presence of bands on a blot prepared from samples of undigested total genomic DNA. However, they do not resolve the situation in relation to the source of these bands, as the results obtained from different samples prepared from the same donor are in disagreement.

Figure 5.3 Southern blot of samples of undigested genomic DNA probed with the 3' *KEL* fragment

Track 1 is undigested K_O genomic DNA, Tracks 2, 3 and 5 are samples of undigested Kk genomic DNA from the same donor, tracks 4 and 6 are undigested kk genomic DNA from one donor and track 7 is undigested genomic DNA from a second kk donor. The blot was probed with the 3' *KEL* fragment.

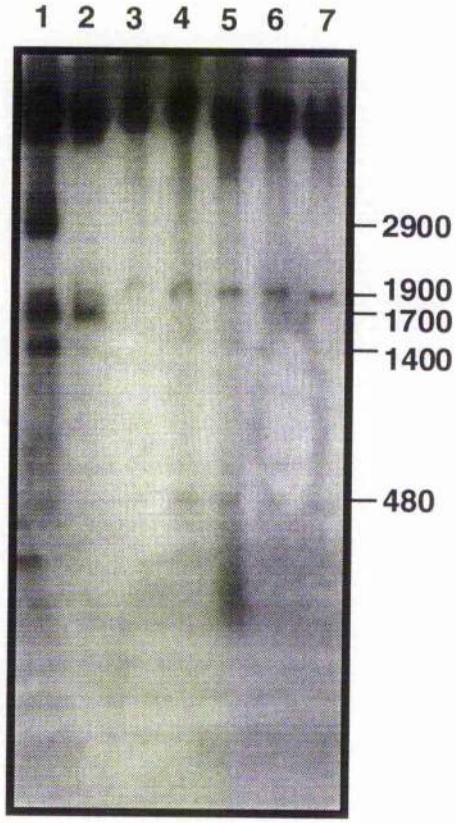
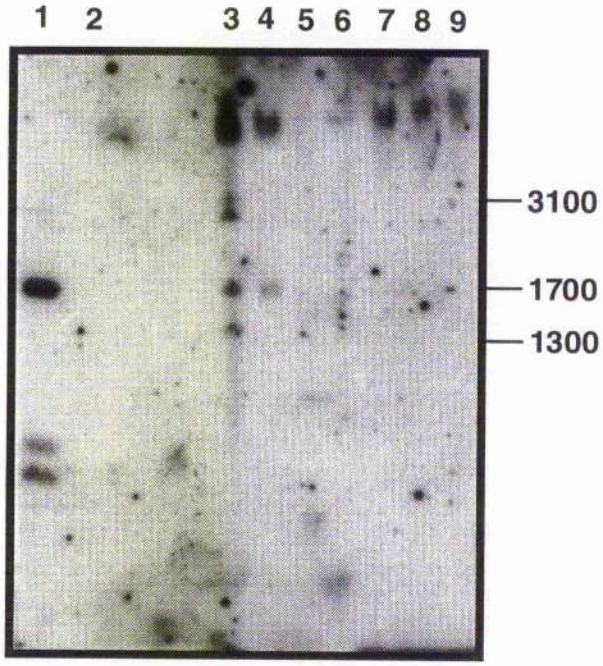

1 2 3 4 5 6 7
2900
1900
1700
1400
480

Figure 5.4 Southern blot of samples of undigested genomic DNA probed with the 5' *KEL* fragment

Track 1 is 1 kb DNA ladder and track 2 contains only loading dyes and TE buffer. Track 3 is undigested K_O genomic DNA, Tracks 4, 5 and 7 are samples of undigested Kk genomic DNA from the same donor, tracks 6 and 8 are undigested kk genomic DNA from one donor and track 9 is undigested genomic DNA from a second kk donor. The blot was probed with the 5' *KEL* fragment.


1 2 3 4 5 6 7 8 9
3100
1700
1300

5.2.4 Hybridization of parts of the 3' *KEL* fragment to blots of K_O and Kk total genomic DNA

In an attempt to clarify the situation the 3' probe was divided into five smaller fragments by digestion with *Pst* I (Figure 5.5). Three of these fragments were sufficiently large to be isolated by gel purification (section 2.9). These were the 291bp, 336bp and 394bp fragments. The two smaller fragments (29bp and 90bp) are too small to be conveniently isolated by gel purification.

Each of the three larger fragments were isolated, radiolabelled and used as probes with blots of *Hpa* II and *Msp* I digested total genomic DNA from K_O, Kk and kk samples (one sample of each). It was hoped that this approach would highlight which part of the 3' fragment gave rise to each band on the blot, aiding the interpretation of the structure of the K_O gene with respect to normal *KEL* genes. Four blots were prepared, each with a size marker and the samples described above. The blots were then probed with the intact 3' fragment and each of the three smaller fragments of the 3' half (Figure 5.6).

The results obtained from these blots showed that two bands (660-750bp and 3200-3800bp) were present in the K_O/*Hpa* II digested sample with all four probes. An extra band of 5000-5500bp was present in this sample when probed with the 336bp, the 394bp probe and the intact 3' probe. A single band of around 3500bp (3200-3800bp) was detected in the K_O/*Msp* I sample with all four probes. One band was detected in the Kk/*Hpa* II digest and this was present with the 336bp, the 394bp and the intact 3' probe, but was absent with the 291bp probe. A second band of 660bp was also detected in this sample when the blot was probed with the 394bp probe. These results have failed to clarify the situation as bands which were detected in the initial experiments (Figure 5.1) were not detected in a repeat of this experiment using the intact 3' fragment (Figure 5.6 D).

5.2.5 Southern blots of total genomic DNA digested with a variety of restriction enzymes

The banding patterns obtained from these initial experiments indicated that this particular K_O individual is likely to have a *KEL* gene similar to that present in the Kk individual. In an attempt to clarify the structure of the *KEL* gene in the K_O individual, further restriction enzyme digests using *Eco* RI, *Bam* HI, *Sal* I, *Pst* I and combinations of *Eco* RI/ *Msp* I and *Eco* RI/ *Hpa* II were performed and

Figure 5.5 Map of fragments generated on *Pst* I digestion of the *KEI.* 3' half length fragment

Diagrammatic representation of the 3' half length *KEI.* cDNA fragment indicating the positions of *Pst* I sites within the sequence and the fragment sizes which would be generated on digestion with *Pst* I. The positions of the *Pst* I restriction sites are according to the numbering as previously defined by Lee *et al.*, 1991.

3' *KEL* FRAGMENT

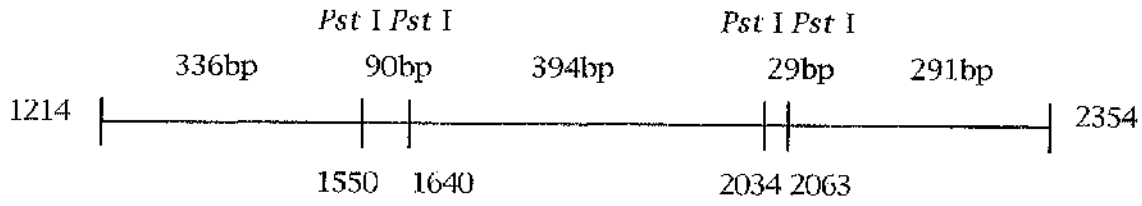
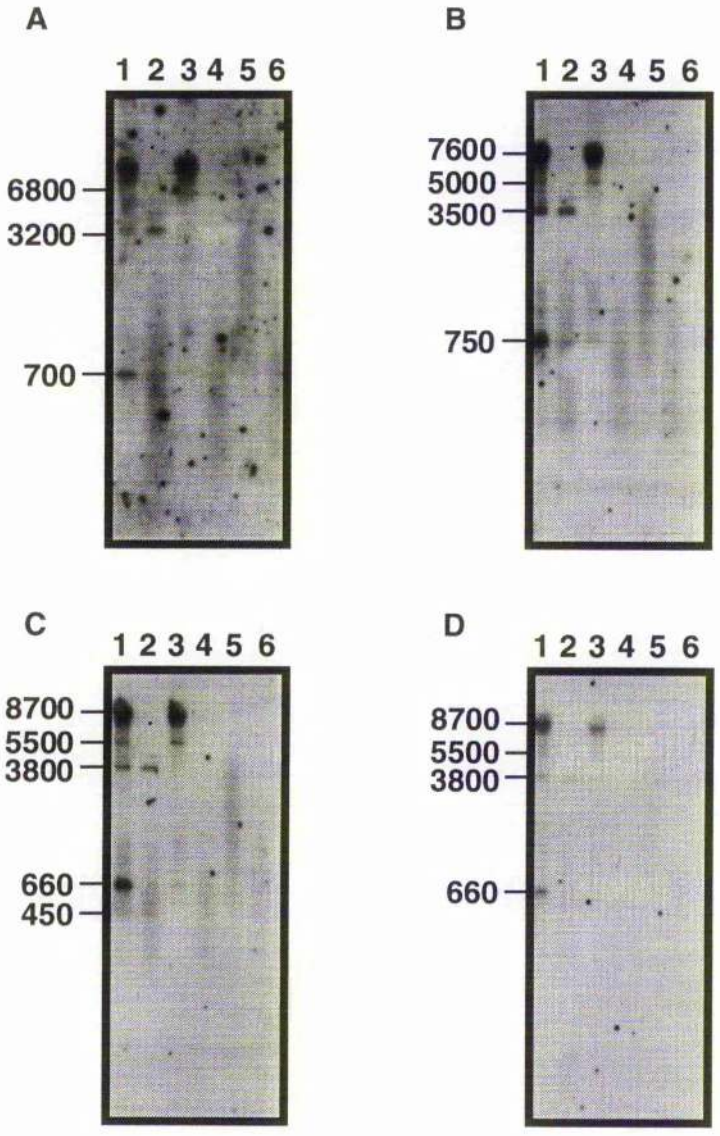


Figure 5.6 Southern blots of *Msp* I and *Hpa* II digested genomic DNA probed with fragments of the *KEL* 3' half

Four blots were prepared as follows; track 1 was *Hpa* II digested K_O genomic DNA, track 2 was *Msp* I digested K_O genomic DNA, track 3 was *Hpa* II digested Kk genomic DNA, track 4 was *Msp* I digested Kk genomic DNA, track 5 was *Hpa* II digested kk genomic DNA and track 6 was *Msp* I digested kk genomic DNA.
Blot A was probed with the *Pst* I 291bp fragment, blot B was probed with the *Pst* I 336bp fragment, blot C was probed with the *Pst* I 394bp fragment and blot D was probed with the intact *KEL* 3' fragment.


A
1 2 3 4 5 6
6800
3200
700
B
1 2 3 4 5 6
7600
5000
3500
750
C
1 2 3 4 5 6
8700
5500
3800
660
450
D
1 2 3 4 5 6
8700
5500
3800
660

used in the preparation of Southern blots. The blots were then probed with the 3' and 5' *KEL* fragments. The results obtained with the 3' probe are shown in Figure 5.7.

The results obtained from this blot could not be easily interpreted as numerous bands were generated. The results of the 3' probed blot indicated that the K_O sample has extra bands compared with the Kk sample with all digests analysed. The bands present in the Kk sample are always present in the K_O sample with the exception of a 5200bp band present when the Kk genomic DNA was digested with *Bam* HI. The results obtained when the same blot, after removal of the 3' probe by stripping with SDS, was probed with the labelled 5' half fragment were very similar to those obtained with the 3' fragment. Again, all the bands present in the Kk sample were present in the K_O sample, with the exception of the 5200bp band in the *Bam* HI digest. On comparison of the results of probing blots of undigested and digested K_O total genomic DNA (Figures 5.2 and 5.7) it was apparent that a band of 3200bp is common to every sample, implying that this band lacks sites for *Eco* RI, *Msp* I, *Hpa* II, *Sal* I and *Pst* I.

The results obtained from these blots reinforce the initial findings that the K_O gene, in this case, is similar to that present in an individual expressing normal Kell system antigens on the red cell surface.

5.3 Overall results obtained from Southern blots

The results of the work described in this section suggest that the *KEL* gene is present in a recognisable form in this particular K_O individual, although precise information regarding the gene arrangement has not been established. It is likely that this example of the K_O phenotype does not arise as a result of a major or complete gene deletion since similar banding patterns can be detected when Southern blots of the total genomic DNA prepared from the individual were performed using *KEL* sequence specific labelled probes. Had a major or complete gene deletion been responsible for the presentation of this phenotype, then few or no bands would be expected to arise as a result of this line of investigation.

The remainder of this section will deal with the problems encountered with interpretation of the results obtained.

Figure 5.7 Southern blot of digested genomic DNA probed with 3' half length *KEL* fragment

Tracks 1, 3, 5, 7, 9 and 11 are restriction enzyme digests of Ko genomic DNA and tracks 2, 4, 6, 8, 10 and 12 are restriction enzyme digests of Kk genomic DNA. The samples in each track have been digested as follows:-

Tracks 1 and 2 - *Eco* RI

Tracks 3 and 4 - *Eco* RI and *Msp* I

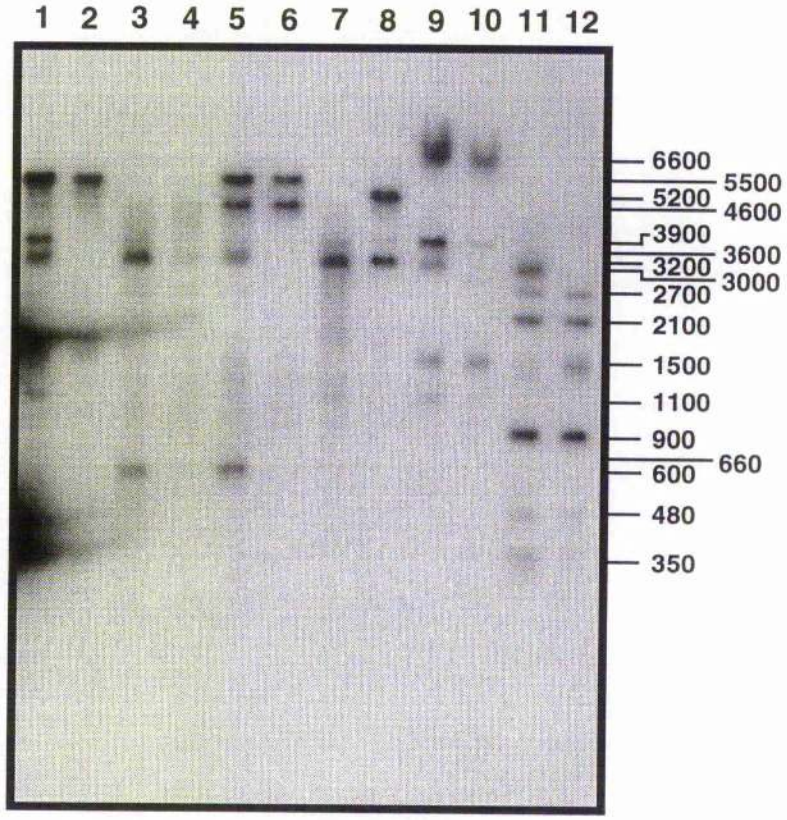
Tracks 5 and 6 - *Eco* RI and *Hpa* II

Tracks 7 and 8 - *Bam* HI

Tracks 9 and 10 - *Sal* I

Tracks 11 and 12 - *Pst* I

The blot was probed with intact *KEL* 3' fragment.


1 2 3 4 5 6 7 8 9 10 11 12
6600
5500
5200
4600
3900
3600
3200
3000
2700
2100
1500
1100
900
660
600
480
350

Firstly, the results which came to light by the use of undigested total genomic DNA in Southern blots, include the following-

1

A band of 1900bp appears in samples of undigested total genomic DNA, prepared under various conditions from different blood samples, when probed with the intact 3' fragment (Figure 5.2). This band does not appear when the 5' probe is used, suggesting that this band is an authentic band which relates to the 3' half of the cDNA sequence.

2

A 480bp band is present in undigested total genomic DNA from the two kk samples and the K_O sample when these are probed with the intact 3' probe (Figures 5.3 and 5.4). This band is not apparent in any of the three Kk samples prepared from the same donor. This suggests the presence of a *KEL* specific sequence common to kk and this K_O individual which appears to be absent in the Kk donor.

3

A 1700bp band is present in the original Kk sample and the K_O sample of undigested total genomic DNA when probed with either the 3' or the 5' fragments (Figure 5.3). This band was not detected in subsequent samples prepared from the same Kk donor, or in any of the kk samples.

4

Five bands of estimated sizes 2900bp, 1900bp, 1700bp, 1400bp and 480bp were detected on probing a blot of undigested total genomic DNA from the K_O donor with the 3' fragment (Figure 5.3). Bands of 3100bp, 1700bp and 1300bp were detected on probing a blot of the same sample with the 5' fragment. It is possible that, allowing for errors in determination of fragment sizes, these bands are the same as the 2900bp, 1700bp and 1400bp bands detected with the 3' fragment. The 1900bp band detected with the 3' fragment was shown to be specific for the 3' fragment as discussed in 1 above. None of the band sizes obtained from these blots relate to the size of the 3' or 5' PCR products cloned in pTZ plasmids (i.e. 2860bp of plasmid + 1157bp of 5' PCR product = 4016bp or + 1208bp of 3' PCR product = 4068bp). This discounts contamination of the sample with cloned *KEL* cDNA.

Secondly, the results obtained from probing blots of total genomic DNA digested with restriction enzymes indicate the following-

1

Eco RI digests of the Kk sample yield three bands when probed with the 3' fragment. The K_O sample gives five bands, three of which relate to those detected in the Kk sample. This suggests sequence differences between the two samples in the genomic DNA which relate to the 3' half of the *KEL* cDNA.

2

Digestion of the Kk sample with *Eco* RI and *Msp* I and probing with the 3' fragment revealed a complete absence of bands (Figure 5.7). When the K_O sample was treated in the same manner, two bands of 3200bp and 660bp were present. The 3200bp band corresponds to one of the extra bands present on probing a blot of undigested K_O genomic DNA with the 3' *KEL* fragment (Figure 5.2). The sizes of the two bands present in the *Eco* RI/*Msp* I digest add up to give a band size which is the same as one of the bands present in the K_O/*Msp* I digest (1300bp). This suggests that *Eco* RI cuts between two *Msp* I sites in the vicinity of the chromosomal *KEL* gene to generate the 660bp band present on probing the Southern blot.

3

When the original sample of Kk total genomic DNA was digested with *Msp* I and probed with the 3' or 5' fragments, four bands were evident in each case. When K_O DNA was treated in the same way, 6 and 4 bands were present respectively (Figure 5.1). These digests were repeated, but on this occasion no bands were present in the Kk sample with either probe. However, one band was present (3800bp) in the K_O sample with the 3' probe (Figure 5.6 D). This band may relate to the large extra band seen on the first occasion using this probe.

4

When Kk total genomic DNA was digested with *Eco* RI and *Hpa* II, two bands (5500bp and 4600bp) were detected with the 3' fragment (Figure 5.3). When the K_O sample was treated in the same way, four bands were present (5500bp, 4600bp, 3200bp and 660bp). This shows that the two bands present in the Kk sample are also present in the K_O sample, but that there are two bands present in the K_O sample which are peculiar to that sample.

5

On digestion of Kk total genomic DNA with *Hpa* II, high molecular weight material (8700bp) and a discrete band (5500bp) were detected with the 3' fragment (Figure 5.6 D). A band of 5500bp was also detected when the sample was digested with a combination of *Eco* RI and *Hpa* II (Figure 5.7). This indicates that there is no *Eco* RI site within the fragment generated by *Hpa* II digestion. When K_O total genomic DNA was digested with *Hpa* II three bands were generated (5500bp, 3800bp and 660bp) (Figure 5.6 D). The 5500bp and 660bp bands were present in the *Eco* RI/*Hpa* II digest but two additional bands (4600bp and 3200bp) were also present in the double digest. These two bands may arise from the partial digestion of the 5500bp and 3800bp fragments by *Eco* RI to give 4600bp and 3200bp fragments plus smaller undetectable fragments. This would represent cleavage of at least two sites by *Eco* RI. The results obtained with the K_O sample (*Eco* RI cuts at one site inside the *Hpa* II fragment) reinforce the findings from the *Msp* I blot i.e. the *Eco* RI cuts once inside the *Msp* I fragment. This is not surprising as *Hpa* II and *Msp* I have the same sequence recognition site but vary in their sensitivity to DNA methylation status, suggesting that the *KEL* gene is methylated.

6

Digestion of Kk total genomic DNA with *Bam* HI and probing of the blot with the 3' fragment gave rise to 2 bands of 5200bp and 3200bp (Figure 5.7). When the K_O sample was treated in the same manner, two bands of 3200bp and 1100bp could be detected. The 3200bp band is common to both samples and the 5200bp band may represent incomplete digestion of the total genomic DNA. The presence of the 1100bp band in the K_O sample may represent a sequence difference between the Kk and K_O samples as *Bam* HI is known not to cut the coding sequence of the 3' half of the *KEL* cDNA. However, the introduction of a *Bam* HI site in the K_O sequence would allow the enzyme to cut the DNA and produce a smaller fragment. An alternative explanation to this is that the intron sequences in the two donors are sufficiently different for a *Bam* HI site to be present in K_O but absent from Kk.

7

Kk total genomic DNA was digested with *Sal* I and the resulting blot was probed with the 3' fragment (Figure 5.7). This gave rise to a high molecular weight band which may arise as a result from the incomplete digestion of the sample, a faint band of 3900bp and a band of 1500bp. When K_O DNA was treated in the

same way, five bands were generated. The first was the high molecular weight band described for the Kk sample and four other bands of 3900bp, 3200bp, 1500bp and 1100bp. These results show that the K_O sample contains two extra bands which were not detected in the Kk sample. It is known that there is no *Sal* I recognition sequence in the 3' half of the *KEL* cDNA sequence therefore, extra specific bands would have to arise by alterations of the K_O coding sequence to generate the restriction enzyme site or by the presence of *Sal* I sites in the intron sequence which are absent from the Kk intron sequence.

8

Digestion of Kk total genomic DNA with *Pst* I gave rise to five bands on probing the blot with the 3' probe (2700bp, 2100bp, 1500bp, 900bp and 480bp). Analysis of the K_O sample by the same method resulted in the detection of 7 bands. Five of these bands were present in the Kk sample and the remaining two bands were of 3200bp and 350bp. The presence of so many bands in these samples is not surprising as the 3' cDNA will generate 5 fragments on digestion with this enzyme.

Thirdly, the results which came to light as a result of probing blots with short length fragments generated by the digestion of the 3' probe with *Pst* I (section 5.2.4 and Figure 5.6) were as follows-

1

Probing a blot of Kk DNA digested with *Msp* I with each of the short length probes failed to yield any bands. Investigation of the K_O DNA in the same way gave one band with each probe i.e. 3200bp with the 291bp probe, 3500bp with the 336bp probe and 3800bp with the 394bp probe. The 3800bp band is the only band present when *Msp* I digested Kk DNA was probed with the intact 3' probe. It is possible that, given errors in estimation of band size, that the bands detected with all three of the short probes are the same band.

2

When Kk and K_O total genomic DNA were digested with *Hpa* II and the resulting blot was probed with each of the three short length probes, the following results were obtained (Figure 5.6 and Table 5.2):-

Table 5.2 Bands detected by probing Southern blots of *Hpa* II digested K_O and Kk genomic DNA with small fragments of the *KEL* 3' half

PROBE	SIZE OF BAND DETECTED	
	K_O	Kk
291bp	>6000bp	>6000bp
	3200bp*	
	700bp*	
336bp	5000bp	5000bp
	3500bp*	750bp
	750bp	
394bp	5500bp	5500bp
	3800bp*	660bp
	660pp	
	450bp*	

From these results it can be seen that extra bands are present in all K_O samples when compared with those obtained from the Kk sample. These band sizes are marked with an asterisk.

5.4 Discussion

The results detailed above have shown that a *KEL* gene is present in the K_O individual tested but have failed to yield clear explanations of the gene structure in this particular case of the Kell null phenotype. There are at least five possible explanations for the differences between the K_O sample and other samples investigated which have been detailed in section 5.3. These have been discussed below.

1 Contamination

It is unlikely that the results detailed earlier in this section have arisen as a result of contamination of the samples. There is no published data which details sequence homology between the *KEL* cDNA sequence and any human gene contaminants which would be likely to arise in a laboratory situation, and contamination of the samples with cloned *KEL* material has already been discounted (section 5.3).

2 Degradation of the total genomic DNA

If degradation of the sample had occurred, then a smear would be seen on the gel and the autoradiogram when the blot was probed. There was no evidence of such degradation (Figure 5.2). It is highly unlikely that the DNA would degrade by cleaving at exact points along its length to generate fragments of precise sizes which would yield clear bands on probing the blot. Bands of the same size were detected in both the K_O and Kk samples when probed with the 3' probe. It is unlikely that degradation would occur to exactly the same extent in two independent samples.

3 Sequence duplication in the K_O gene

Extra bands appeared in the K_O sample when the DNA was digested with a variety of enzymes and the resulting blots were probed with all of the fragments prepared. In most cases, the bands present in the Kk sample, under the same conditions, were also present in the K_O sample. This indicates that the sequences of both these samples are similar, but that the extra bands present in the K_O sample must have considerable sequence homology with the probe. This suggests that parts of the *KEL* gene may have been duplicated in the K_O individual. This would give rise to a banding pattern the same as the "normal" phenotype sample with extra bands present which would arise from the duplicated sections of the sequence. However, an internal sequence duplication

would alter the sizes of the restriction enzyme fragments generated on digestion. This would result in a significant alteration in the banding pattern obtained when the probed blots from normal and the K_O sample were compared. This was shown not to be the case. Therefore, any gene duplication which has occurred would have to be in the terminal portions of the gene.

4 Deletions in the K_O sequence

Equivalent sized bands appear in both normal and K_O samples in all the Southern blots prepared. This is contradictory to a partial gene deletion having occurred, as a deletion would truncate the gene and consequently give rise to different banding patterns when compared with a normal phenotype sample by Southern blot analysis.

5 Gene rearrangement

Extra bands occur in the K_O sample when compared with a Kk sample by Southern blot analysis. Sequence homology between the probe and the DNA relating to the bands detected is prerequisite for the detection of these extra bands. This may indicate that the K_O gene is rearranged in such a way that the sequence is similar to the "normal" sequence, but is sufficiently different that Kell protein cannot be expressed on the red cell surface in an antigenic form. However, gene rearrangments would give rise to different banding patterns when compared with a sample from a normal Kell phenotype donor.

5.5 Conclusions

One important piece of information which has come to light as a result of this work is that this particular K_O individual appears to have a *KEL* gene which is similar to the gene present in normal Kell phenotype individuals. This implies that this particular example of the Kell null phenotype is unlikely to have arisen from a major gene deletion or from a complete absence of the *KEL* gene. Lee *et al.* have reported the sequencing of the *KEL* gene of two K_O individuals (Lee *et al.*, 1995 a). Their results show no differences, in the coding sequence, between the normal and null individuals. This suggests that the K_O phenotype may arise as a result of the action of a modifying gene at a later stage in the development of the protein and its insertion into the red cell membrane; from the lack of post-translational modification required for the protein to be inserted into the cell membrane and expressed in an antigenic form; or from

the presence of mutated upstream regulators which are repressing the *KEL* gene.

In summary, therefore, the results obtained from the work described in this section give rise to many questions regarding intra- and inter-donor differences. Many of these questions remain unanswered as continuing this line of investigation would have proved detrimental to the progress of other aspects of the project. However, the important fact which has come to light as a result of Southern blot investigation is that this particular example of the K_O phenotype does not appear to arise from major alterations in the *KEL* gene stucture.

SECTION 6

Automated direct sequencing of *KEL* PCR products

SECTION 6 Automated direct sequencing of *KEL* PCR products

6.1 Introduction to direct sequencing of PCR products

Sequencing of DNA can be approached from several different angles, some of which involve the use of cloned cDNA target in either single stranded or double stranded forms, as discussed in sections 1.2.4 and 3.7. This approach allows the determination of the DNA sequence derived from a single cloned DNA molecule. Misincorporation by *Taq* DNA polymerase may generate products which differ from that of the original target molecule. As a result, misincorporations arising from *Taq* DNA polymerase infidelity can only be distinguished from true differences by sequencing material from several different clones. To eliminate such errors, material from several different clones must be sequenced. This is both time consuming and expensive and, therefore, a more suitable alternative is the direct sequencing of uncloned target material whereby the determined sequence is that of a large population of molecules. Several direct sequencing methods are currently available and some of these were used in the course of this project. These techniques and their associated problems will be discussed in section 6.4. Having assessed these, it was decided that the technique of choice was the automated direct sequencing of PCR products using the dye labelled dideoxy terminator cycle sequencing method described in section 2.20.6.

The direct automated sequencing of PCR products was achieved with the help of Mr Peter Martin and Mr Gary Mallinson of the International Blood Group Reference Laboratory (IBGRL), Southmead, Bristol.

6.2 Introduction to automated sequencing

The *Taq* dyedeoxy terminator automated sequencing method (Applied Biosystems Inc.) was chosen for the direct sequencing of *KEL*-related PCR products. The reason for this choice was that, in view of the size of the *KEL* cDNA and the fact that sequence information for both sense and antisense strands of several individuals of known Kell phenotype was required, longer stretches of information could be obtained from each reaction than would have been obtained from sequencing by other techniques. Automated sequencing does not involve the use of radioisotopes, which is also beneficial from a safety point of view. This method is based on the Sanger dideoxy sequencing method, but

incorporates four dye labelled dideoxy nucleoside triphosphates, each having a different coloured tag, which effect chain termination. Target DNA and sequencing primer were added together in a single tube in the presence of each of the four dye labelled dideoxy terminators, a dNTP mix (containing dITP, dATP, dCTP and dTTP), *Taq* DNA polymerase and terminator ammonium cycle sequencing buffer (supplied by the manufacturer). The reaction mixture was then heated and cooled in the same manner as PCR to allow the annealing and extension of the sequencing primer (section 2.20.6). This generated chains of different lengths, each of which end in a dye labelled ddNTP. The resulting products were then electrophoresed in a single track of a denaturing polyacrylamide gel. A laser scanned the gel on a fixed axis and, as the dye tags passed, the sequence was recorded on a computer system. This type of analysis generates a series of coloured peaks, each peak relating to the dye tag carried by the terminal nucleotide. The order of the coloured peaks then generates the sequence information. Sequences of up to 600 nucleotides can be generated from one single reaction by this technique. This compares favourably to manual methods which require four gel lanes for each sample (one for each of the individual terminators). This can generate approximately 400 nucleotides of sequence using aliquots of the samples electrophoresed for 2 and 5 hours respectively. In practice, rather fewer nucleotides can be resolved in the manual technique.

6.3 Automated sequencing of PCR products and analysis of results

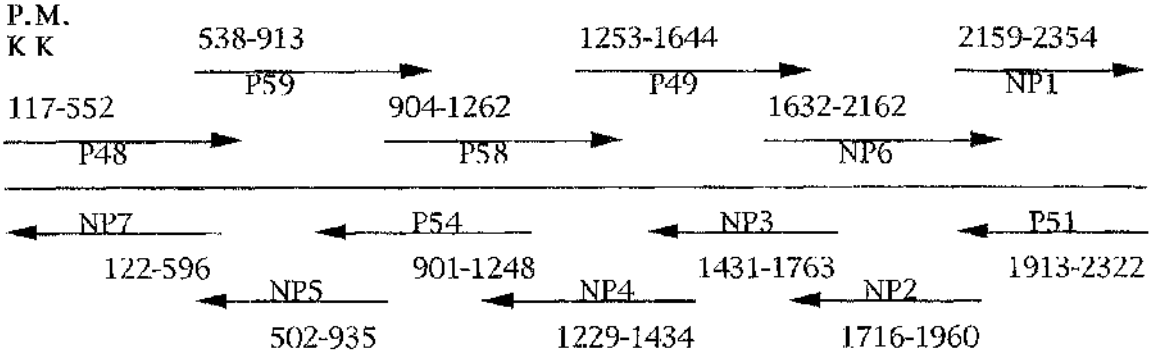
The method used for the direct sequencing of *KEL*-related PCR products was the Applied Biosystems *Taq* DyeDeoxy Terminator Cycle Sequencing Kit with the Applied Biosystems Model 373A automated sequencer according to the protocol described in section 2.20.6. All of the PCR products sequenced were generated by RT-PCR (sections 2.5.2 and 2.7.1) from individuals of known K/k phenotype and all were Js(a-b+), Kp(a-b+). The status of other Kell blood group system antigens was unknown.

Samples from two KK and two kk donors were completely sequenced on both strands in order to determine differences which may account for the expression of the K or k antigens (Figure 6.1 and appendix). The sequencing primers used for this purpose are detailed in Table 6.1.

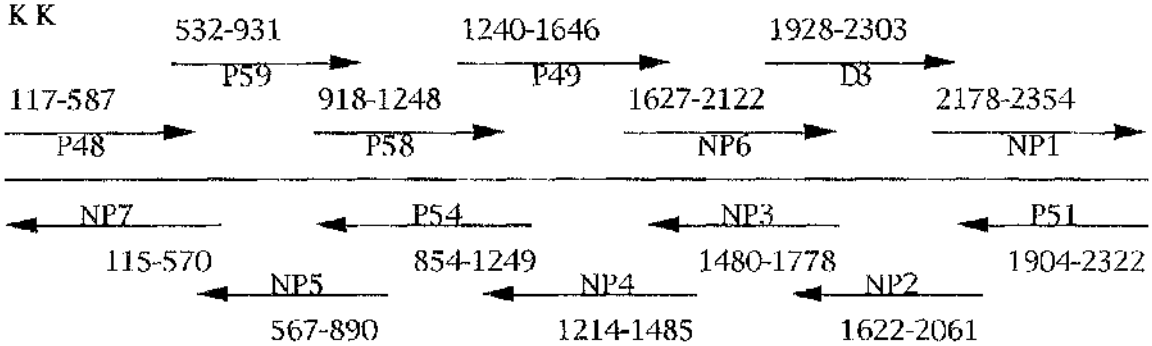
Figure 6.1 Complete sequence determination for the coding region of *KEL* cDNAs

Diagrammatic representation of the complete sequencing of both strands of PCR products derived from two KK and two kk donors. The donor's initials and Kell phenotype are noted above the relevant diagram. Arrows represent the direction of the stretch of sequence obtained from each individual reaction. The first and last nucleotide positions of each sequence are noted. The coding sequence of *KEL* cDNA stretches from nucleotide positions 124 to 2322. Position refers to cDNA sequence numbered as described by Lee *et al.*, 1991. Primers used for sequencing are noted above or below the relevant arrow and details of these can be found in Table 6.1.

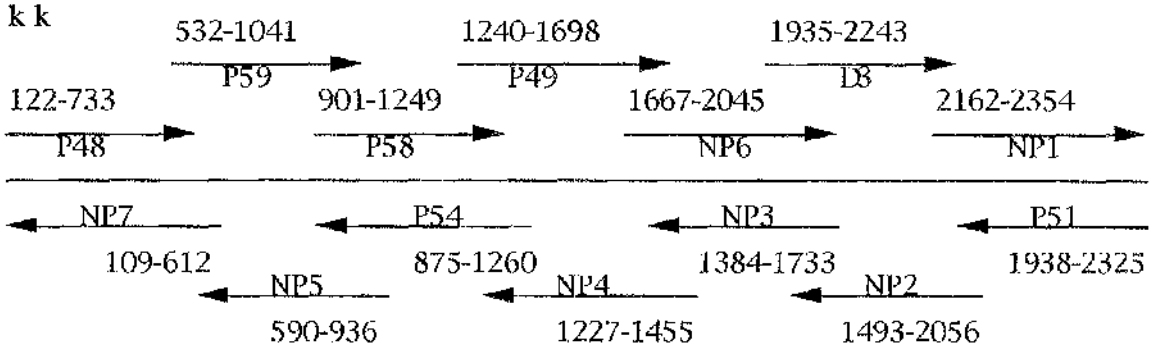
P.M.
K K



A.T.
K K



G.T.
k k



T.C.
k k

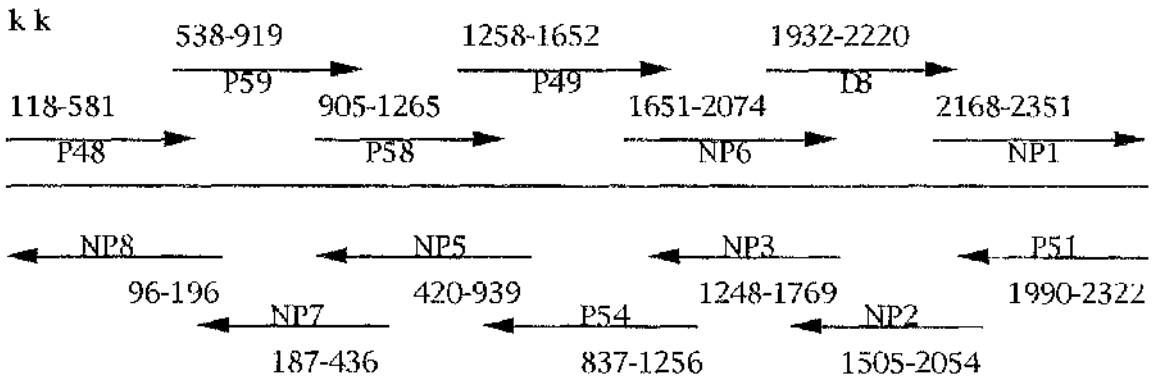


Table 6.1 Oligonucleotide primers used for sequencing

PRIMER NUMBER	SEQUENCE	POSITION
P48	ccgtcgacGAAGGATTCTGGAGCCACAGAAG	93-115
P59	ccgtcgacAAACCGACTTCGGAGAATACTGG	501-523
P58	ccgtcgacCGGGAATACCTGACTTACCTGAA	865-887
P49	ccgtcgacACATGATCTTAGGGCTGGTGGTG	1214-1236
NP6	GCCCGACAAGAATACAACGAT	1594-1614
D8	CACGAGCTGTTGCACATCTTC	1864-1884
NP1	GGACCTCAGCCCCCAGCAGAT	2118-2138
P51	ccggatccATTTCTGTGCTGTGGCATCTTTG	2354-2333
NP2	GGTGCCGTAACAGCCTCTTGC	2092-2072
NP3	AGAATGGGGGTTGGAGGAGTC	1807-1787
NP4	TTCTGAGGCGAGTGATGAGGG	1489-1469
P54	ccggatccCTTTCTGCGTGCCTCCTGGAATT	1284-1262
NP5	GCCGTGAAGTGATGGAGATTG	967-947
NP7	TAACTTGTCTGAGGGGACCAG	640-620
NP8	TCCACGGGCAGCCTCTCTTCT	233-213

Lower case letters denote the presence of an extra eight residues which incorporate a restriction enzyme site. These oligonucleotides were originally designed for PCR work and subsequent cloning of the products (section 3). Position refers to the *KEL* cDNA nucleotide sequence numbered as previously reported by Lee et *al.*, 1991.

The sequence of each of the four different *KEL* cDNAs was determined on both strands with overlaps. The extent of sequence determination from each primer for each cDNA is shown in Figure 6.1. The complete sequence data obtained for each sample is detailed in the appendix.

During the course of this section of work, when only part of the sequence analysis required for the comparison of *KEL* cDNAs encoding the K and k antigens was completed, Lee *et al.* published evidence indicating that the sequence difference between k and K is a C->T substitution at position 701; where C gives rise to the k antigen and T gives rise to the K antigen on the red cell surface (Lee *et al.*, 1995 a). The project was completed to provide an independent determination of any sequence difference.

The results obtained from these sequencing experiments confirm that a single C->T substitution is highly likely to be responsible for the difference in expression of k and K antigens as no other differences specific for either of the two KK or the two kk sequences were apparent. However, several differences were detected within individuals. These presented as non-complementary nucleotides on opposite strands of *KEL* cDNA from a single individual. For example, one strand may have a G at a particular position and the same position on the opposite strand may sequence as a T. Closer analysis of the chromatograms resulting from the sequencing experiments failed to give agreement between the two strands. These differences are detailed in Table 6.4. It is important to note that two of the differences detected on only one strand from one individual matched with the same differences from a different individual i.e. the nucleotide at position 1000 on the positive strand of both A.T.(KK) and G.T.(kk) is A in both cases, whereas the published sequence has a C at this position. The same two individuals also have a T at position 1818 instead of the published C. Both of these sequence differences conserve the amino acid residues i.e. the amino acid encoded at position 1000 is arginine in both cases and the amino acid encoded at position 1818 is glycine in both cases. The other nucleotide differences detailed in Table 6.2 appear to be specific to each particular donor and may have occurred as a result of *Taq* DNA polymerase infidelity early in the initial PCR reactions used to generate the material for sequencing. In view of the fact that various individual PCR reaction products were used to complete the sequence data for any donor, it is unlikely that product from a single PCR reaction was used for sequencing the same region of both strands.

Table 6.2 Inter-strand sequence differences within individuals

Donor	Nucleotide Position	Published Nucleotide	Strand Sequenced	Nucleotide Detected
P.M.	2000	T	+	A
P.M.	1977	T	-	C
P.M.	1873	T	-	A
P.M.	1460	T	-	G
P.M.	1227	G	-	A
A.T.	1000	C	+	A
A.T.	1818	C	+	T
A.T.	2102	C	+	T
A.T.	1548	G	-	C
A.T.	1576	G	-	A
A.T.	1211	G	-	T
G.T.	908	G	+	A
G.T.	1000	C	+	A
G.T.	1186	C	+	T
G.T.	1349	A	+	T
G.T.	1626	A	+	T
G.T.	1673	C	+	G
G.T.	1818	C	+	T
G.T.	1723	G	-	C
G.T.	1725	G	-	C
G.T.	891	G	-	C
G.T.	168	C	-	G
T.C.	2252	C	+	T
T.C.	2310	C	+	T
T.C.	1305	A	-	G
T.C.	146	A	-	T

Nucleotides detected on the negative strand sequence are given as reverse complement nucleotides.

PCR products from four KK individuals, two Kk individuals and five kk individuals were sequenced in the region around residue 701. Both sense and antisense strands from each individual were sequenced. Primers P59 and NP5 were used in these experiments and the sequences of the oligonucleotide primers used are shown in Table 6.1. The sequencing results obtained are detailed in Table 6.3.

The results of sequencing clearly show the C->T change which is responsible for the expression of the K and/or k antigen on the red cell surface. In the cases where PCR products from heterozygous individuals have been sequenced, it is clear that the nucleotide which has been detected at position 701 is T. This sequence codes for the expression of the K antigen. The reason that T is clearly detected and C is undetected in heterozygotes is because the C peaks generated by automated sequencing are normally of lower intensity than any of the peaks generated by the other three nucleotides. The T signal dominates in the chromatogram, but a smaller C signal is visible beneath the larger peak. The peaks would be required to have similar intensities for the sequencer to assign the nucloetide as N, which means that it is unable to define the precise nucleotide present (it could be either of two nucleotides based on the peaks detected). The presence of both peaks at this postion strongly suggest that the T sequence is present in a proportion of the PCR product molecules and the C sequence is present in the remainder of the PCR product molecules. Homozygous samples have only a single clear peak of either C or T at this position. This has been illustrated in Figure 6.2.

6.4 Assessment of direct sequencing techniques

As mentioned above, several different techniques were assessed in order to establish the most suitable method for the direct sequencing of PCR products. All of these techniques had associated problems and, on the basis of the initial results obtained, automated direct sequencing was determined to be the method of choice. The other methods which were assessed are detailed below and some of the problems and the reasons for their occurrence have been discussed.

Table 6.3 **Sequencing of sense and antisense strands of PCR products generated from known Kell phenotype individuals in the K/k polymorphic region.**

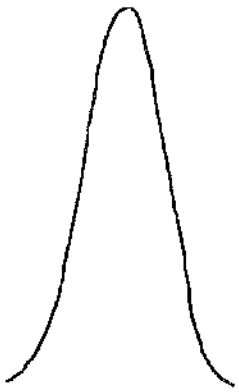
DONOR	PHENOTYPE	SEQUENCE AND POSITION			
		681	691	701	711
				*	
Published sequence		TTCCTTAAAC TTTAACCGAA		C GCTGAGACTTCTGATGAGT	
PM	KK	TTCCTTAAAC TTTAACCGAA		T GCTGAGACTTCTGATGAGT	
AT	KK	TTCCTTAAAC TTTAACCGAA		T GCTGAGACTTCTGATGAGT	
ST	KK	TTCCTTAAAC TTTAACCGAA		T GCTGAGACTTCTGATGAGT	
KM	KK	TTCCTTAAAC TTTAACCGAA		T GCTGAGACTTCTGATGAGT	
MM	Kk	TTCCTTAAAC TTTAACCGAA		T GCTGAGACTTCTGATGAGT	
AM	Kk	TTCCTTAAAC TTTAACCGAA		T GCTGAGACTTCTGATGAGT	
GT	kk	TTCCTTAAAC TTTAACCGAA		C GCTGAGACTTCTGATGAGT	
TC	kk	TTCCTTAAAC TTTAACCGAA		C GCTGAGACTTCTGATGAGT	
GI	kk	TTCCTTAAAC TTTAACCGAA		C GCTGAGACTTCTGATGAGT	
SM	kk	TTCCTTAAAC TTTAACCGAA		C GCTGAGACTTCTGATGAGT	
AA	kk	TTCCTTAAAC TTTAACCGAA		C GCTGAGACTTCTGATGAGT	

Sequences given are for the sense strand only. Results obtained for the antisense strands were the exact reverse complement sequences to those shown above for each donor sample. Primers P59 and NP5 were used to sequence the sense and antisense strands respectively in the polymorphic region (Table 6.1). The K/k polymorphism is present at position 701 and this is marked with an asterisk.

Figure 6.2 **Sequencing peaks generated from the automated sequencing of PCR products derived from homozygous and heterozygous individuals.**

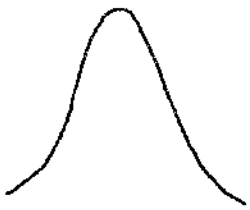
Diagrammatic representation of relative T and C peak heights obtained by automated sequencing of PCR products. 1 indicates a single T peak from a homozygous KK sample where T is the only nucleotide present at position 701. 2 indicates a single C peak from a homozygous kk sample where C is the only nucleotide present at position 701. 3 indicates T plus C peaks from a heterozygous Kk sample where both C and T are present at position 701 and appear as a strong T peak accompanied by a less intense, underlying C peak.

T PEAK



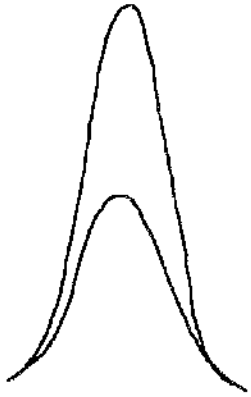
1

C PEAK



2

T PEAK WITH UNDERLYING C PEAK



3

1

The first method investigated was *Taq* DNA polymerase cycle sequencing of double stranded DNA (section 2.20.4). This method was carrried out in two steps, the first of which was the labelling step in which the primer was extended using three of the four deoxynucleoside triphosphates including one α-labelled dNTP. The mixture was then thermally cycled to produce a labelled, extended primer, the length of which is dependent on the template sequence. In the second step, further amounts of all four of the deoxynucleoside triphosphates were added along with a chain terminating dideoxynucleoside triphosphate and the reaction was again thermally cycled. DNA synthesis during this step continues until all growing chains are terminated. After electrophoresis of the products on a denaturing polyacrylamide gel and exposure of the gel to X-ray film, the sequence can be determined. This method was attempted several times using *KEL* PCR products as the double stranded DNA target material. The results obtained on all occasions were very poor. The autoradiographs which resulted from these experiments completely lacked any banding pattern even after long exposure of the gel to the film. This suggested that the sequencing primer was not binding to the target to allow extension and termination of chains incorporating the radioisotope. Several different PCR products of varying quantities and different sequencing primers were used in attempts to solve this problem. The sequencing primers were analysed by polyacrylamide gel electrophoresis and were found to give a single band, indicating that there was no primer degradation which could give rise to problems. The cycling parameters were also varied to try to optimise the conditions for use of the technique with the *KEL* PCR product targets, but none of these approaches proved successful. It was decided that this method was unlikely to be suitable for routine use.

2

The second approach investigated was the use of asymmetric PCR (section 2.20.2) to produce predominantly single stranded PCR products for direct sequencing by the Sequenase Version 2.0 (TM) method. Linear (not logarithmic) amplification of PCR products was achieved by addition of one of the two oligonucleotide primers in large excess. As a result, large quantities of one of the two strands of the PCR template were preferentially produced. The resulting material was analysed on an agarose gel to confirm the presence of single stranded product and this was gel purified to ensure that the target material for sequencing was of suitable quality. Single stranded material migrates faster

than double stranded material of equivalent size on agarose gel electrophoresis. A variety of concentrations of target material were then used for sequencing. The samples were electrophoresed on a denaturing polyacrylamide gel and, after fixing and drying, this was exposed to X-ray film. The results obtained showed that, although the control sample supplied with the kit gave rise to a faint banding pattern, no banding pattern was observed with the test samples, even when the gel had been exposed to the X-ray film for prolonged periods of time. Again, this indicated a problem with either the target material or the annealing and extension of the primer. The target material used in these experiments had been prepared by PCR and the products had been purified to remove contaminants which could interfere with the sequencing reaction. This reduces the possibility of the problem having arisen from the template material. The primers used were sequence specific primers which were either those used in the original PCR for the generation of the template or were internal primers which had been shown to be of good quality by polyacrylamide gel analysis as discussed above. This also reduces the possibility of failure of the sequencing reactions occurring as a result of poor quality primers. This technique was also considered as unsuitable for the direct sequencing of PCR products.

3

The third approach which was investigated was the binding of PCR products to Streptavidin-coated Dynabeads (TM) and subsequent direct sequencing of the bound product (section 2.20.5). The use of these beads eliminated both the necessity for centrifugation steps and the need for ethanol precipitation or phenol extraction of the template material. This technique involved amplification of the DNA target by PCR using two primers, one of which had been biotinylated at the 5' end during synthesis. The degree of biotinylation of each primer was assessed by polyacrylamide gel analysis where the biotinylated fraction of the primer migrates more slowly than the unbiotinylated fraction. Not all primer molecules become biotinylated during their synthesis. This generates a PCR product which has a biotin label on one strand and the PCR product can be bound onto streptavidin coated magnetic beads by a biotin/streptavidin linkage, which is a very stable bond and cannot easily be denatured. Alkaline denaturation was used to separate the unbiotinylated strand from the bound, biotinylated strand. The bound strand was then used for sequencing using the Sequenase Version 2.0 (TM) method for single stranded targets. This method guarantees removal of enzymes,

oligonucleotide primers and unincorporated nucleotides from the target, leaving highly purified DNA for sequencing. These experiments produced autoradiographs which had single bands at the wells where the samples had been loaded. This suggested that there had been some incorporation of the isotope into the sample, but that the presence of the magnetic beads had prevented the migration of the samples into the gel. It is possible that the heat denaturation step prior to loading the sample had been ineffective in releasing the radiolabelled strand from the template strand which was bound to the bead. However, this is unlikely to prevent generation of readable sequence on the gel, as the sequencing products should be free in solution. The bound material only serves as a template for their generation. A possible explanation for these results is that the sequencing reaction had failed and that residual radioactivity present in the sample was retained in the well during electrophoresis. However, this is unlikely. The biotinylated PCR product was analysed by agarose gel electrophoresis prior to bead sensitisation and appeared normal. The unbound strand from the bead sensitisation was recovered, ethanol precipitated and analysed by agarose gel electrophoresis. The results obtained from this showed that the single stranded DNA could be recovered. This indicates that the biotinylated single strand remained attached to the beads and, therefore, a target was present for sequencing. The quality of data obtained suggested that this technique would be unsuitable for the direct sequencing of PCR products.

4

The fourth method used was the automated sequencing of the Dynabead (TM) immobilised material described above (section 2.20.7). This was a logical progression from the work which had been carried out, as an automated sequencer had been purchased by the Biochemistry Department, University of Glasgow at this time. In this case, internal sequencing primers were designed which would anneal to the region 20 nucleotides downstream from the annealing site of the unbiotinylated PCR primer. The reason these new primers were synthesised was that incomplete chain synthesis during PCR may inhibit the binding of the PCR oligonucleotide primers and, consequently, the sequencing of PCR-generated material. The biotinylated, immobilised product was then used as a target with the Applied Biosystems *Taq* DyeDeoxy Terminator Cycle Sequencing Kit and the Applied Biosystems Model 373A automated sequencer. PCR products were generated using one normal and one biotinylated primer and these were used to sensitise Dynabeads as before. Aliquots of these sensitised beads were then used directly with the *Taq* DyeDeoxy Terminator

Cycle Sequencing Kit and the resulting mixtures were loaded onto a denaturing polyacrylamide gel. This was electrophoresed and analysed on the Applied Biosystems Model 373A automated sequencer. The results obtained from this were variable. In a few instances, short stretches of sequence were obtained. However, these bore no relation to the published *KEL* cDNA sequence. The reasons for the occurrence of these results is unknown. In most cases no sequence information showing similarity to the published *KEL* cDNA sequence was obtained. The reasons for this are likely to be the same as those discussed in 3 above.

5

The final method, and the method which was chosen for use, was the automated direct sequencing of PCR products (section 2.20.6). Initial results obtained from this were poor but, after discussion of the protocol with Mr Gary Mallinson of IBGRL, the optimum conditions for the direct sequencing of PCR products were established and sequencing was performed to give the results detailed above. This work was carried out with the help of Mr Peter Martin of IBGRL. A fault in the sensitivity of the sequencer in Glasgow could not be resolved during the project, necessitating the use of the IBGRL facility.

6.5 Conclusions

The results detailed above are clear evidence to support the published data that the K/k polymorphism is coded for by a single C->T substitution at position 701 (Lee *et al.*, 1995 a). No other sequence differences were detected when PCR products from two KK and two kk individuals were sequenced completely on both strands, although inter-strand differences were detected within individuals. These were likely to have arisen as a result of infidelity of the *Taq* DNA polymerase, as discussed in section 6.3. The results obtained from the work described in this section have proved to be clear cut and do not give rise to any questions regarding the basis for the K/k polymorphism.

SECTION 7

Development of a PCR-based assay for *KEL* genotyping

SECTION 7 Development of a PCR-based assay for *KEL* genotyping

7.1 Development of a PCR-based *KEL* genotyping assay

The introduction of the PCR technique marked an important turning point in transfusion science. It has allowed the development of diagnostic assays for the determination of blood group phenotypes by investigating directly the genetic material. This is of benefit in cases where rare antisera are unavailable or where only small numbers of cells are available for typing, such as foetal samples being investigated where HDN is implicated. The development of such a technique for Kell and Cellano typing is described in this section. The assay is based on the fact that the polymorphism which determines Kell and Cellano arises from a single nucleotide sequence difference, creating a *Bsm* I restriction enzyme site (section 6 and Lee *et al.*, 1995 a). This site is generated only when the T (Kell) nucleotide is present at position 701 (section 6). The target material used in this investigation was total genomic DNA which can be isolated rapidly from small samples and in microgram quantities sufficient for use in PCR. The resulting PCR products can then be digested with *Bsm* I restriction enzyme and analysed by gel electrophoresis.

7.2 Rationale of the assay

The development of a PCR-based genotyping assay for determination of K/k status was an important aspect of this project. It has allowed the establishment of an alternative to traditional serological techniques whereby small samples of blood, such as ante-natal samples, could be used for the purposes of unambiguous Kell typing.

7.2.1 Preparation of the target DNA

Several methods for the preparation of total genomic DNA were assessed in the course of this work. The criteria on which the assessment was based were the yield of total genomic DNA obtained per millilitre of whole blood; the amount of time required to complete the preparation; and the quality of the target material obtained with respect to generation of PCR products. From initial experiments, it was evident that the method which fulfilled all of the criteria was the technique described in section 2.22.2 (Miller *et al.*, 1988). This method was used in the preparation of all of target material used in this section.

7.2.2 Selection and design of oligonucleotide primers

Three factors were taken into account when designing the oligonucleotide primers for use in this assay. Firstly, primers were selected which would anneal to exon sequences within the *KEL* gene. The reason for this was that the exon sequences are less likely to exhibit sequence variation between individuals than are intron sequences. Therefore, choosing primers which anneal to exon sequences maximises the likelihood of the primers annealing to all the total genomic DNA targets used. The second factor was the expected sizes of the PCR products which would be generated using the chosen oligonucleotide primers. This was important as two products were co-amplified, each of which was to be cleaved by digestion with *Bsm* I, and had to be resolved by gel electrophoresis without ambiguity (section 7.2.3). The third factor which was taken into account was the nucleotide compostion of each primer. The A+T : G+C ratio was carefully chosen to ensure all four primers used in the reaction had similar melting points to optimise the chance of co-amplification of the two PCR products (Table 7.1).

7.2.3 Restriction enzyme digestion of PCR products to allow genotype determination

The presence of the gene encoding the Kell antigen implies a T residue at nucleotide postiton 701 (section 6, Lee *et al.*, 1995 a). This gives rise to a *Bsm* I restriction site which is absent from the gene encoding the Cellano antigen. This sequence difference formed the basis of the assay, as *Bsm* I digestion of a PCR product which spans nucleotide 701 would allow the determination of the presence of gene sequences encoding K and/or k antigens. Two oligonucleotide primer pairs were designed, one which spans the K/k polymorphic site, and a second which spans a separate *Bsm* I site (the presence of which is independent of K/k status). The amplification of the conserved *Bsm* I site and its subsequent digestion acts as a control for the test (section 7.3).

7.2.4 Analysis of restriction enzyme digests by gel electrophoresis

The digested PCR products described in section 7.2.3 had to be sufficiently different in size to be resolved by gel analysis. The primers chosen allowed the co-amplification of two fragments of approximate sizes 600bp and 860bp.

Table 7.1 Oligonucleotide primers used in the *KEL* genotyping assay

PRIMER NUMBER	SEQUENCE	POSITION
D1	AGGTCCAGAATTCCTGGCACC	524-544
D2	CTGGATGACTGGTGTGTGTGG	795-775
D3	CACGAGCTGTTGCACATCTTC	1864-1884
D4	AAGAAGATCTGCTGGGGGCTG	2144-2124

Position refers to the *KEL* cDNA nucleotide sequence numbered as previously reported (Lee *et al.*, 1991).

One set of primers (D1 + D2) was designed to span the K/k polymorphic site at 701 and the second set (D3 + D4) amplifies a region containing a conserved *Bsm* I site (Figure 7.1).

Digestion of each of these fragments by *Bsm* I generated two fragments. The 600bp fragment gave rise to 500bp and 100bp digestion products, and the 860bp fragment gave rise to 395bp and 465bp digestion products. These fragments could be resolved on a 2% agarose gel (section 2.8). This allowed a rapid analysis of results, a factor which is prerequisite in any transfusion science diagnostic test.

7.3 Results of PCR-based *KEL* genotyping

Total genomic DNA was isolated from samples of anticoagulated whole blood as described in section 2.22.2. This method was used because the final product can be obtained within 24 hours and the resulting material is of suitable quality and adequate yield for use in a diagnostic assay as discussed in section 7.2.1. The resulting DNA samples were resuspended in TE buffer (section 2.2), the A_{260} measurements were taken and the DNA concentrations estimated assuming that an A_{260} of 1 is equivalent to 50μg/ml of total genomic DNA (Sambrook *et al.*, 1988). 1μg of this DNA was then used as a target for PCR, following the protocol described in section 2.7.4, with the exception that 60°C was used as the annealing temperature. Two sets of PCR oligonucleotide primers were used in a single reaction as discussed in section 7.2.2 (Table 7.1). It is possible that the PCR products generated could contain *Taq* DNA polymerase misincorporations which would alter *Bsm* I restriction enzyme sites thus preventing digestion. The use of 1μg of target material ensures that, even if a misincorporation occurs in an early round of the PCR, then this will only represent a small proportion of the overall product and should not lead to misinterpretation of the donor's genotype.

The products amplified span nucleotide residues 524-795 (K/k polymorphism is at nucleotide position 701) and 1864-2144 on the *KEL* cDNA sequence. However, intron sequences are also amplified when the oligonucleotide primers are used with total genomic DNA as a target. These increase the sizes of the final PCR product to approximately 600bp and 860bp respectively (Table 7.2 and Figure 7.1).

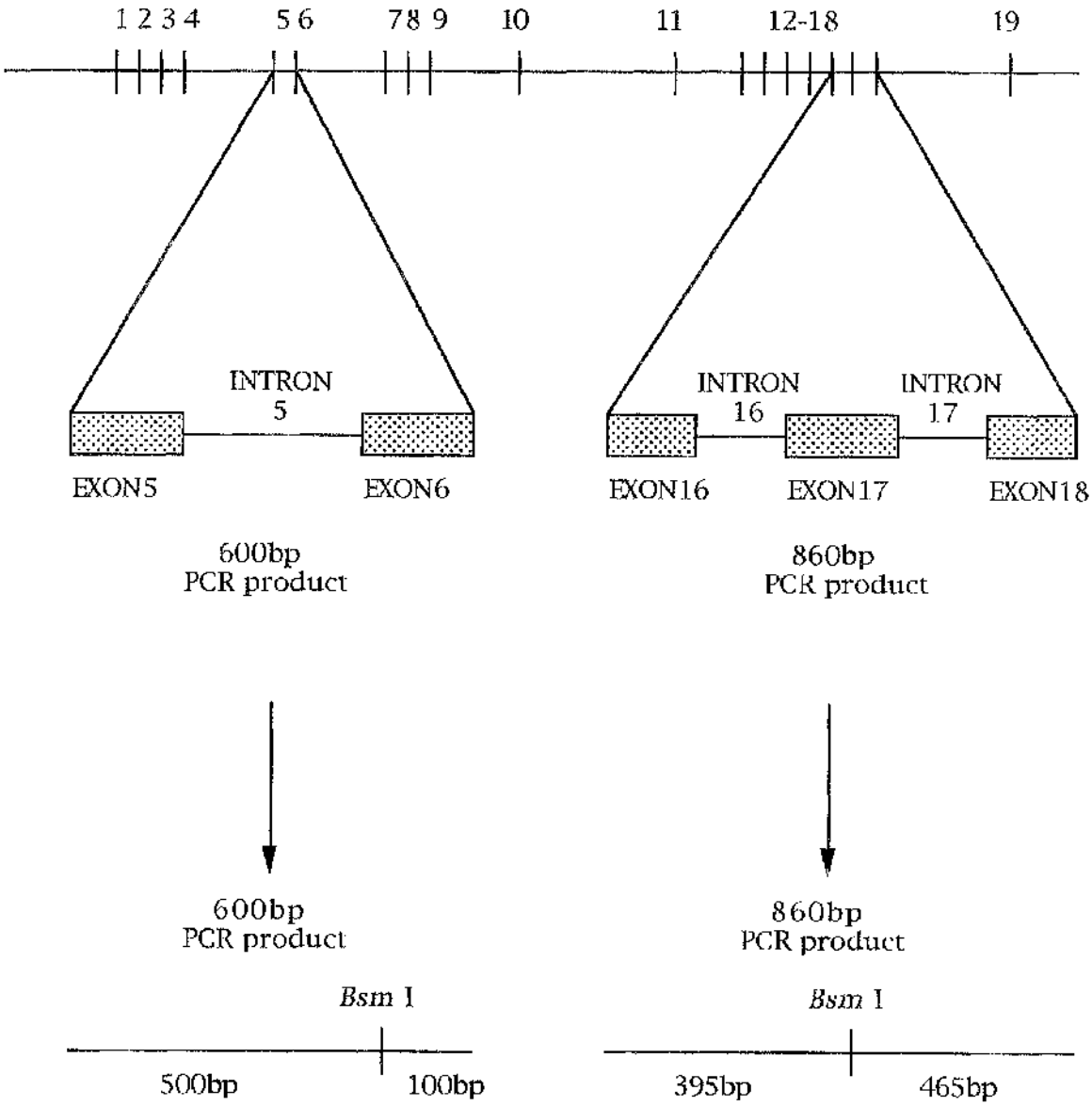
Table 7.2 Sizes of *KEL* fragments generated by PCR

PCR FRAGMENT	INTRON SITE	APPROX. INTRON SIZE	APPROX. PCR FRAGMENT SIZE
524-795	648	330bp	600bp
1864-2144	1894	230bp	860bp
1864-2144	2064	350bp	860bp

Numbering is based on the *KEL* cDNA nucleotide sequence as previously reported by Lee *et al.*, 1991. Intron positions, derived from Lee *et al.* 1995 b, are corrected from those published to conform to the numbering published by Lee *et al.* 1991. Intron lengths are as stated by Lee *et al.*, 1995 b.

Figure 7.1 Diagrammatic representation of the *KEL* gene and the regions amplified by PCR

Diagram of the *KEL* gene. Exons are numbered 1 to 19. PCR products amplified using primer pairs D1+D2 and D3+D4 (Table 7.1) are the 600bp and 860bp fragments. The 600bp fragment spans exons 5 and 6, and intron 5. The 860bp fragment spans parts of exons 16 and 18, and all of exon 17 and introns 16 and 17. The positions of *Bsm* I sites within the two PCR fragments and the expected sizes of the digestion products are indicated.


1 2 3 4
5 6
7 8 9
10
11
12-18
19
INTRON 5
INTRON 16
INTRON 17
EXON5
EXON6
EXON16
EXON17
EXON18
600bp PCR product
860bp PCR product
600bp PCR product
860bp PCR product
Bsm I
Bsm I
500bp
100bp
395bp
465bp

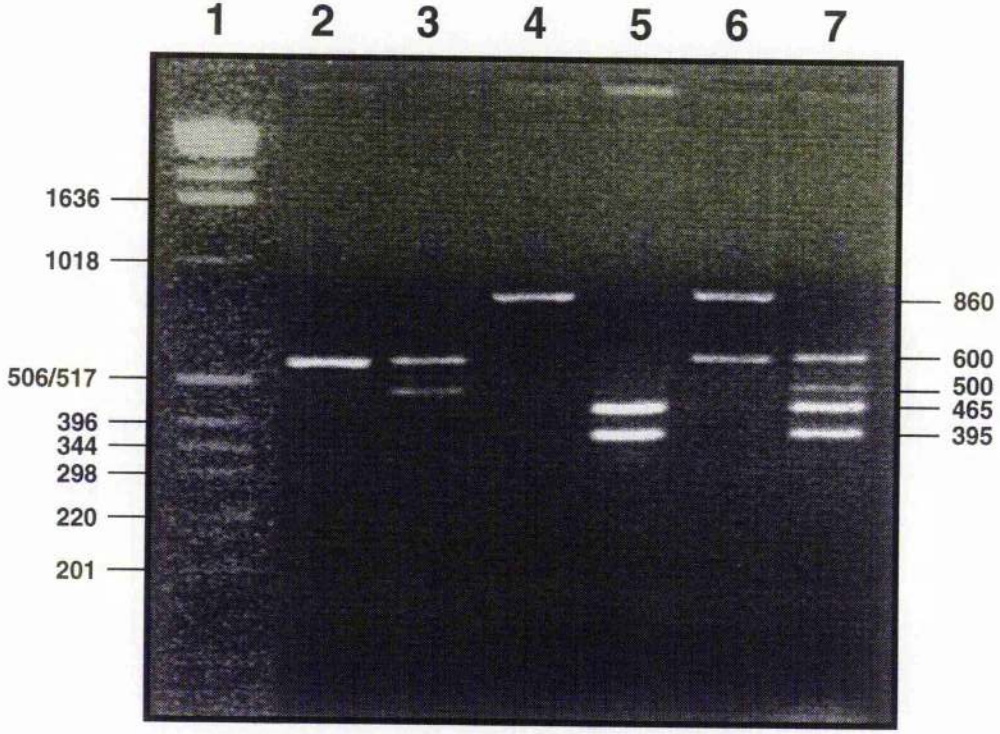
The PCR products, which were generated using total genomic DNA samples from a control Kk sample and the two primer pairs, were digested with *Bsm* I for 1 hour at 55°C. 2.5 Units of enzyme were used in 20µl digests (section 2.12). The total digests were electrophoresed on a 2% agarose/TBE gel (section 2.8). Figure 7.2 illustrates the results expected on *Bsm* I digestion of each of the fragments generated from PCR using the oligonucleotide primers detailed above and a control Kk sample.

The inclusion of the 860bp internal control serves two purposes. The first is that the 860bp product always contains a *Bsm* I site irrespective of the Kell status of the donor and, hence, it monitors the activity of the *Bsm* I restriction enzyme. Such a control is essential as a 600bp fragment from the high incidence kk phenotype lacks the *Bsm* I site and, as a result, will remain undigested. Therefore, in most cases the final product of the assay will be the intact 600bp fragment resulting from genomic DNA of the kk phenotype. The same pattern would be detected for KK and Kk samples if inactive *Bsm* I had been used for the digest and this would result in the mis-diagnosis of the Kell status of the donor. Failure of digestion of the 860bp product would indicate a problem with the enzyme and the test could be repeated. The second purpose is to illustrate that the digest has reached completion i.e. that all possible *Bsm* I sites have been digested by the enzyme. If insufficient enzyme had been added or the activity of the enzyme was significantly reduced and the reaction has not gone to completion, then the banding pattern obtained from a partially digested KK sample would resemble that from a completely digested Kk sample. The incorporation of the control product would show the presence of some residual 860bp band in this situation. This would reduce the possibility of misinterpretation of genotype as a result of incomplete digestion.

Samples of total genomic DNA were prepared from each of five kk, Kk and KK donors. 1µg of each of these samples was used in PCR as described above. 10µl of each of the 100µl PCR reaction products were then digested as detailed above and the digested material was analysed by gel electrophoresis. 10µl of each of the undigested PCR products were electrophoresed for comparison. PCR controls included one reaction containing all constituents with the exception of target DNA and a second reaction containing all constituents with the exception of oligonucleotide primers. These controls eliminate the possibility of non-specific products being amplified as a result of contamination of any of the reaction constituents. 1µg of a 1 kb DNA ladder (Gibco BRL) was used as a size

Figure 7.2 Agarose gel analysis of restriction enzyme digests of *KEL* PCR products amplified from genomic DNA from a Kk donor

2% agarose gel of undigested and *Bsm* I restriction enzyme digested PCR products. Track 1 contains 1 kb DNA ladder size marker, track 2 contains undigested 600bp PCR product, track 3 contains *Bsm* I digest of 600bp PCR product, track 4 contains undigested 860bp PCR product, track 5 contains *Bsm* I digest of 860bp PCR product, track 6 contains undigested 600bp plus 860bp PCR products and track 7 contains *Bsm* I digest of 600bp plus 860bp PCR products.


1
2
3
4
5
6
7
1636
1018
506/517
396
344
298
220
201
860
600
500
465
395

marker and a digest of the two fragments amplified from a known Kk control sample was included to allow direct comparison with bands generated in the test samples (Figure 7.3 a,b,c).

The results shown in Figure 7.3 are evidence that this assay yields unambiguous K/k genotyping results. This could be used confidently in a routine situation as the results obtained are easily interpreted and the controls included in the assay would give strong indicators to the source of any problems which may arise through its use.

7.4 Discussion

Other methods which could have been used to obtain genotype information include RFLP analysis using genomic DNA. This is a method whereby the DNA is directly digested with a restriction enzyme and the digested material is analysed by Southern blotting. This is a very time consuming technique which does not always give unambiguous results. The assay described in this section is based on RFLP analysis of PCR products and this yields more precise information than RFLP on genomic DNA. Another method which could have been adopted was the use of mismatch analysis where the strands of double stranded control PCR products are alkali-denatured and are mixed with similarly treated test sample. The strands are then allowed to reanneal to form homoduplex and some heteroduplex molecules. These are then gel electrophoresed and differential migration of the two types of molecules allows genotype determination. This method may not always yield definitive information depending on the resolution obtained in different gel analyses. Genotyping can also be performed using allele specific primer PCR (section 1.4). This method involves the synthesis of three oligonucleotide primers, one of which anneals at a position which is remote from the polymorphic site and two which are identical in sequence with the exception of the extreme 3' nucleotide. One of the two oligonucleotide primers has the 3' terminal nucleotide specific for one genotype and the other has the 3' terminal nucleotide specific for the opposing genotype. Two separate PCR reactions are then performed, each containing the non genotype specific oligonucleotide primer and either of the two genotype specific oligonucleotide primers. The PCR primer will be extended only when the 3' terminal nucleotide is complementary to the target sequence. The presence or absence of detectable PCR product by gel analysis then allows genotype determination. This method would have been a suitable alternative

Figure 7.3 Agarose gel analysis of co-amplified diagnostic PCR products digested with *Bsm* I

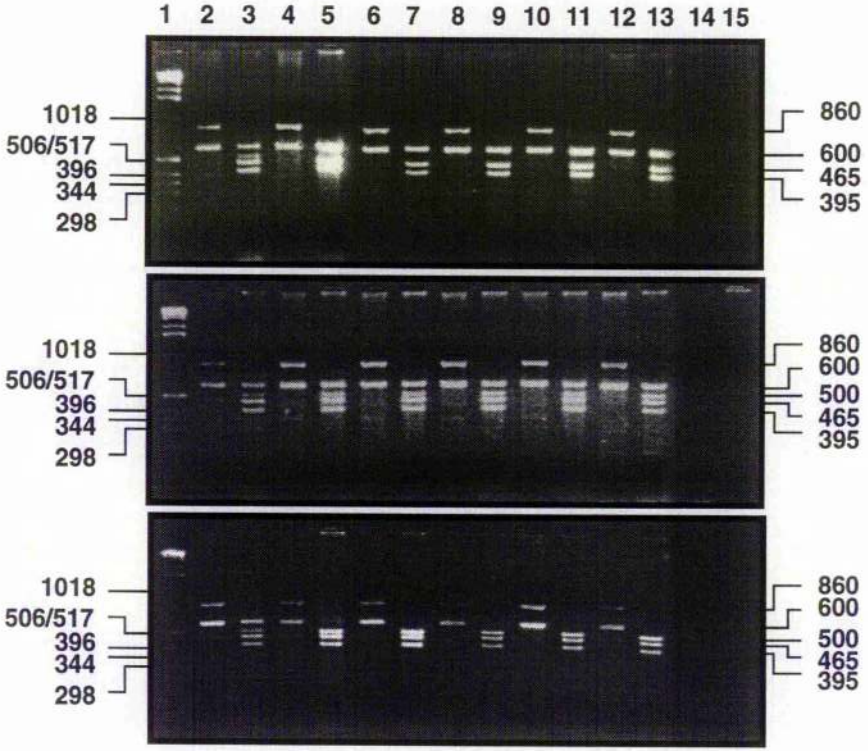
The top figure (Figure 7.3 a) illustrates the results obtained using the diagnostic assay with samples derived from five independent kk donor samples.

The middle figure (Figure 7.3 b) illustrates the results obtained using the diagnostic assay with samples derived from five independent Kk donor samples.

The bottom figure (Figure 7.3 c) illustrates the results obtained using the diagnostic assay with samples derived from five independent KK donor samples.

The gels shown in all three figures were as follows:-
Track 1 contains 1 kb DNA ladder marker, track 2 contains undigested Kk control PCR products, track 3 contains *Bsm* I digested Kk control PCR products. Tracks 4, 6, 8, 10 and 12 contain undigested PCR products from five different test samples, tracks 5, 7, 9, 11 and 13 contain the corresponding *Bsm* I digested PCR products from these five test samples. Track 14 contains the PCR control which had no DNA target added and track 15 contains the PCR control which had no oligonucleotide primers added.

The sizes of the markers and of the diagnostic DNA fragments are shown.


1 2 3 4 5 6 7 8 9 10 11 12 13 14 15
1018
506/517
396
344
298
860
600
465
395
1018
506/517
396
344
298
860
600
500
465
395
1018
506/517
396
344
298
860
600
500
465
395

technique to the assay described in this section. In view of the problems which may arise in the use of the first two alternative techniques discussed above, the PCR-based RFLP assay was the method of choice for genotype determination in donor samples.

The results described in earlier parts of this section deal purely with the investigation of KK, Kk and kk known phenotype samples. The reason for using known phenotype donors was that the likelihood of obtaining sufficient KK and Kk samples from a small random selection of donor samples is extremely small in view of the low incidences of these phenotypes. Therefore, in order to demonstrate that the assay works equally well for all three of these common phenotypes, samples of pre-determined Kell status had to be used. Samples of more unusual phenotype blood, such as K_O and Kmod, have not been assayed by this method. This will be discussed in section 8.

The assay described in this section has been shown to give unambiguous results for K/k genotyping when using total genomic DNA as a target. This would be particularly useful for ante-natal K/k typing in cases where HDN is implicated. The assay is rapid and reasonable numbers of samples could be dealt with at any one time. Total genomic DNA can be prepared from blood samples and stored in a stable form for long periods of time. This facilitates the planning of work where K/k typing is not to be performed with any degree of urgency, or where repeat tests may be required at a later date. This method of Kell status determination is extremely sensitive and reduces the possibility of mis-typing where weak reactions are obtained by traditional serological methods. The interpretation of the results obtained using the PCR-based genotyping method is straightforward and leaves little space for error, making this a very user-friendly technique.

The development of this genotyping assay illustrates the importance of the use of molecular biology techniques in the field of transfusion science. It is a rapid technique which yields conclusive data regarding the Kell genotype of the donor, enabling clinical decissions to be made with a great degree of confidence. A battery of similar assays for other blood group systems could allow the establishment of a panel of fully typed donors who could be called on in emergencies. Therefore, molecular biology methods are being established as important techniques with respect to patient care.

SECTION 8

General discussion

SECTION 8 General discussion

8.1 Amplification and cloning of *KEL* PCR products

8.1.1 Isolation of total RNA from blood

The results discussed in section 3 described the isolation of total RNA from samples of whole blood and the amplification and cloning of *KEL* PCR products from a Kk donor. Several methods for the isolation of total RNA from whole blood were investigated before the AGPC method, described in section 2.4.1, was chosen for use (Chomczynski *et al.*, 1987). The other techniques which were assessed included two methods reported by Wai Kan *et al.* and Temple *et al.* (Wai Kan *et al.*, 1975, Temple *et al.*, 1977). The Temple method is based on the use of a buffer which will lyse the red cell population, leaving the leukocytes intact. The RNA is then precipitated from the lysate at low pH and purified using phenol/chloroform. The Wai Kan method again involves lysis, this time of the total cell population, and low pH precipitation. The resulting precipitate is treated with a buffer containing Tris-HCl (pH 8.0), NaCl, SDS and Na_2EDTA and the RNA is isolated by phenol/chloroform extraction. Three kits were also assessed for their ability to yield reasonable quantities of total RNA of a suitable quality for use in RT-PCR. These included the Nucleon kit (Scotlab), the RNAzol B kit (Biogenesis) and the TRIzol kit (Gibco BRL). All three methods were based on the method of Chomczynski *et al.*, but they are designed for use with small sample volumes. All of these methods proved less successful than the AGPC method in the isolation of microgram quantites of total RNA suitable for use in subsequent work, as they allow the processing of only microlitre quantities of whole blood and thus yield only small amounts of RNA. This was unsuitable for the initial developmental stages of the project, where optimum conditions for subsequent experimental procedures were to be established. It was concluded that the AGPC method was the one of choice in view of the fact that up to 60ml of whole blood could be processed at any one time. This allowed the isolation of microgram quantities of total RNA which could then be stored at -70°C for use when required.

8.1.2 Reverse transcription of total RNA

Total RNA was used as a target in reverse transcription and, as described in section 3.3, two different approaches were adopted. The reason for this was that

full length first strand *KEL* cDNA could not be obtained using MoMuLV reverse transcriptase. The most likely explanation for this is that the *KEL* RNA target material contains some secondary structure in its central region which prevents reverse transcription through this area with MoMuLV reverse transcriptase. When AMV reverse transcriptase was used, transcripts of adequate length were obtained. The use of AMV reverse transcriptase allows the reverse transcription reaction to be performed at a higher temperature (42°C instead of 37°C with MoMuLV reverse transcriptase). The use of a higher temperature reduces the secondary structure and, in this case, may have eliminated the problem.

8.1.3 Choice of donor for initial work

The reason for using a heterozygous donor for the initial stages of the project was that the donor was willing to donate blood samples whenever required. This allowed all the developmental work to be carried out on samples from a single donor and eliminated differences which may arise between donors. There are both problems and advantages with using a heterozygous donor. The problems arise mainly as a result of cloning individual PCR products. The target material used for the PCR reaction would have contained both K and k sequences and, when the resulting products were cloned, it would be impossible to determine which of the K or k sequences was present in each individual clone. In order to distinguish between the two, many different clones would have to be completely sequenced to distinguish true sequence differences from differences which had arisen as a result of *Taq* DNA polymerase infidelity (section 8.4.1). This was the main reason for the choice of direct sequencing of PCR products later in the project. This technique reduces these problems, as a population of PCR products is sequenced at any one time.

An advantage of using a heterozygous donor for the initial work is that, in hybridization studies for either *in situ* work or probing of Southern blots, material from different clones could be used as probes. This would have been important if the K/k polymorphism arose as a result of major sequence differences. Under such circumstances, the K sequence probe may not hybridize to a k sequence target, and vice versa. Using clones derived from a heterozygous individual increases the chances of hybridization of at least a proportion of probes to the target sequence.

The use of a heterozygous donor for the initial work has not posed any problems with respect to the progress of this project and it has proved useful with respect to obtaining regular samples.

8.1.4 General points

The cloned material allowed the work described in sections 4 and 5 to be performed and it also allowed RT-PCR conditions to be established. This facilitated experiments described in sections 6 and 7. It is clear that the work described in section 3 was fundamental to the progression of this project.

8.2 *KEL* gene localisation by *in situ* hybridization

The results obtained from the *in situ* hybridization experiments were based on the hybridization of the 3' half of the *KEL* cDNA to the gene locus. The cloned 5' half was also used as a probe, but the signals obtained were extremely weak and no definitive information could be established as a result. The evidence obtained from use of the 3' half probe was sufficiently strong to allow definite mapping of the genetic locus on chromosome 7.

The information which arose from this part of the project gave clear assignment of the *KEL* locus to chromosome 7q33-q35. This refined previous assignments and the results were reinforced by a report simultaneously published (Lee et *al.*, 1993). The same basic techniques were used for both investigations, but this project involved the use of intact human chromosomes to determine the site of probe hybridization and confirmation of the assignment was by PCR on somatic cell hybrid DNA. Lee *et al.* used a radiolabelled *KEL*-specific probe to hybridize to somatic cell DNA panels (Lee *et al.*, 1993). The fact that the two independent sets of mapping work gave the same result is a good illustration of the use of different routes to arrive at the same final conclusion.

This section of work provided conclusive results on the location of the *KEL* gene on chromosome 7. These correlated well with previously published work by other groups (Zelinski *et al.*, 1991 a, Purohit *et al.*, 1992) and with the data published by Lee *et al.* (Lee *et al.* 1993). *KEL* is not the only blood group gene which has been assigned to chromosome 7. Others known to be present on chromosome 7 include Colton and Cartwright (Tsui *et al.*, 1991).

8.3 The basis of the K_O phenotype

The results obtained from the Southern blots have already been discussed in detail at the end of section 5. As reported, these results did not permit positive conclusions regarding the basis of the K_O phenotype. The only conclusion which could be drawn was that this particular example of K_O does not appear to arise as a result of a major gene deletion or a complete absence of the gene. This has been reported by two groups of workers where RFLP analysis and DNA sequencing, respectively, showed no differences between K_O and normal Kell phenotype individuals (Lutz *et al.*, 1992, Lee *et al.*, 1995 a). Sequencing of the cDNA from several K_O individuals is the most appropriate way of determining the differences between the K^O and normal *KEL* genes, and this has shown that no differences are present in the coding sequences of two K_O individuals (Lee *et al.*, 1995 a). This leaves at least three possible explanations for the lack of Kell antigens in these individuals. The first is that errors in, or failure of, transcription do not allow the development of a normal antigenic Kell protein on the red cell membrane. This theory could be investigated by analysing K^O mRNA. If this was shown to be different from mRNA from normal phenotype individuals, then a transcription error is a possible explanation. However, it has recently been reported that two K_O individuals contain *KEL* mRNA in peripheral blood and that the sequence of this is identical to the mRNA coding sequence obtained from individuals with normal Kell phenotypes, discounting this explanation (Lee *et al.*, 1995 a). A second explanation is that a separate gene has a post-transcriptional effect on the development of the Kell protein and its insertion into the red cell membrane. This would be similar to the repressor type of Rh null (section 1.5), where a double dose of a repressor gene is inherited, preventing the expression of Rhesus antigens. A third possible explanation is that the expression of all Kell antigens is dependent on glycosylation. It would then follow that the K_O phenotype could arise as a result of a fault in glycosyl transferase enzymes, which would result in a lack of glycosylation of the Kell protein and, hence, lack of expression of Kell antigens. Two groups of workers have reported that Kell and Cellano antigen expression is not dependent on glycosylation, but this will be discussed further in section 8.4.3 (Marsh *et al.*, 1990, Murphy *et al.*, 1993).

Due to the conflicting results which arose from Southern blotting studies, it was decided that, in the interest of the progression of the project, this should be

abandoned in favour of the DNA sequencing experiments. These experiments yielded important information regarding the basis of the K/k polymorphism.

8.4 Automated sequencing of *KEL* cDNAs

8.4.1 Problems caused by *Taq* DNA polymerase infidelity in point mutation analysis

Taq DNA polymerase infidelity is an important factor which must be taken into account when sequencing PCR products, either directly or after cloning. The reason for this is that *Taq* DNA polymerase does not exhibit the proof-reading 3'->5' exonuclease activity present in other DNA polymerase enzymes, such as *E. coli* DNA polymerase I and T4 DNA polymerase (Tindall *et al.*, 1988). These workers reported that *Taq* DNA polymerase inserts a single wrong nucleotide at a rate of 1 for each 9000 nucleotides polymerised in a PCR reaction. Saiki *et al.* report work in which they observed a cumulative error frequency of about 0.25% after 30 cycles of PCR (Saiki *et al.*, 1988). It follows that, when amplifying two approximately equal sized cDNA *KEL* fragments of around 1200bp, then three misincorporations might be expected to arise in each fragment. This has important implications when sequence differences between individuals are to be established. As already mentioned in section 6, sequencing of individually cloned PCR products would not allow determination of true sequence differences from *Taq* DNA polymerase misincorporations. Direct sequencing of PCR products should eliminate, or at least substantially reduce this problem, as a population of PCR products is sequenced at any one time. Many copies of target material are present in the first round of a PCR and this means that, even if the misincorporation has occurred in the first PCR cycle, then it will arise in a tiny proportion of the product population. As a result, the possibility of the "wrong" nucleotide being assigned as the true sequence will be reduced on interpretation of the results of direct sequencing. This is particularly important when novel sequence determination is being undertaken. It is likely that, when misincorporation does occur, that it will occur at different positions in each PCR reaction. Therefore, comparison of the sequencing results from more than one PCR reaction derived from any one individual should enable the determination of any misincorporation sites. This is the major factor for the decision to choose direct sequencing of PCR products for the investigation of the genetic basis for the K/k polymorphism in this project.

8.4.2 Assessment of techniques for mismatch analysis

In order to establish a suitable protocol for identification of sequence differences between PCR products derived from different individuals, several techniques were assessed. These have been discussed in some detail in section 6.4. This took a substantial amount of time, as techniques were repeated and adapted several times in order to try to improve the results obtained. In view of the fact that most of these methods gave very poor results, alternative techniques which could reduce the amount of sequencing required were investigated. These included single strand conformation polymorphism analysis (SSCP) (Kovar *et al.*, 1991, Hayashi, 1992, Fan *et al.*, 1993, Hongyo *et al.*, 1993, Sekiya, 1993), denaturing gradient gel electrophoresis (Cariello *et al.*, 1993, Ganguly *et al.*, 1993) and the chemical cleavage of mismatches (CCM) technique (Roberts *et al.*, 1989, Dr S. Bidichandani-Personal Communication). After discussion with workers at the Duncan Guthrie Institute of Medical Genetics (Yorkhill, Glasgow), where these techniques are used routinely, it was decided that the most appropriate method for the purpose of determining sequence differences between the cDNAs encoding Kell and Cellano antigens was chemical cleavage of mismatches. This would pinpoint the region in which a mismatch has occurred between the K and k sequences. The other two methods were discounted as they may detect only 50% of mismatches and this could lead to conflicting results between samples. Use of the CCM technique should reduce the amount of sequencing required to determine the K/k polymorphism, as it allows the determination of the points at which there are differences between a reference and a target sequence. Once the positions of the differences have been established, a region of only 100-200 nucleotides around this point can be sequenced to define the nucleotide substitution. This method uses some of the chemical cleavage reagents employed in Maxam and Gilbert sequencing (Maxam *et al.*, 1977). A reference cDNA, kk in this case, was 5' end radiolabelled. Thereafter, the double stranded reference molecule was alkali-denatured and mixed with a similarly denatured, unlabelled test molecule. The strands were allowed to renature, generating homoduplex molecules of reference and test samples, and a proportion of heteroduplex molecules consisting of one labelled reference strand and one test strand. The resulting mismatched nucleotides were then chemically modified using piperidine and the modified sites were cleaved using osmium tetroxide. The sample was electrophoresed on a denaturing polyacrylamide gel along with size markers. With knowledge of which end of the reference sequence had been labelled, and by comparison of

the band size with the size markers, the region containing the mismatch can be defined. The appropriate region can be sequenced using a sequence specific oligonucleotide primer. This is a particularly unpleasant method with respect to Health and Safety procedures, as use of most of the chemicals requires extreme care. This method was attempted on several occasions, but the results obtained gave either no bands on the autoradiograph or a ladder of bands which should, in principle, never arise. Only a single band should be detected per track, as scission of the heteroduplex should release one radiolabelled and one unlabelled fragment. This technique was abandoned in favour of direct automated sequencing of PCR products derived from individuals of known K/k phenotype.

8.4.3 Implications of automated sequencing results

The results described in section 6 provide evidence that a single C->T substitution at nucleotide position 701 is responsible for the difference in expression of the k and K antigens. This work was in progress when the report of Lee *et al.* was published (Lee *et al.*, 1995 a). The results of this project have given independent confirmation of the polymorphism. The intron-exon boundaries and exon sizes within the *KEL* gene have also been published and this has shown that nucleotide 701 is present in exon 6 of the 19 exons which constitute the *KEL* gene (Lee *et al.*, 1995 b). This single nucleotide substitution encodes a threonine to methionine change, which is likely to prevent N-glycosylation of the asparagine residue at a consensus N-glycosylation motif (Asn X Thr->Met) and creates a *Bsm* I restriction enzyme site in the K sequence. Lee *et al.* report that this change in glycosylation reduces the apparent molecular mass of the Kell protein from 93 kDa to around 90.5 kDa and, as a result, suggest that the change in glycosylation of the protein, resulting from the presence of the cDNA sequence encoding the Kell antigen, may be responsible for the difference in antigen expression (Lee *et al.*, 1995 a). However, work presented in this thesis and data published by Marsh *et al.* point strongly to a polypeptide-related basis for differences in antigen expression (Marsh *et al.*, 1990, Murphy *et al.*, 1993). This has been discussed in section 8.3 .

8.5 PCR-based diagnostic assay

8.5.1 Reasons for the development of a PCR-based diagnostic assay

The development of a PCR-based diagnostic assay was an important aspect of the project in relation to patient care. The assay development proved to be straightforward and has provided a method which will prove useful in a variety of situations, including antenatal *KEL* genotyping and Cellano typing in situations where reagents may be expensive. This assay can also be used to confirm donor samples which type as KK by traditional serological techniques, in order to establish a panel of KK donors which can be called upon to donate in situations where substantial volumes of KK blood are required at short notice. The assay is rapid, easy to apply and yields results which can be readily interpreted. It also has inbuilt controls which should help pinpoint any problems, should they arise.

8.5.2 Problems associated with the use of a PCR-based diagnostic assay

Possible problems with this technique may arise when individuals expressing the K_O or Kmod phenotypes are investigated. Neither of these phenotypes have been assessed in the course of the development of this assay, but results reported by Lee *et al.* show that the coding sequence of the *KEL* gene in the K_O individuals studied had complete homology with that of kk phenotype individuals (Lee *et al.*, 1995 a). It follows, therefore, that these rare individuals would type as kk by this assay. This would not be important for a recipient of this unit of blood, but could have serious implications for the K_O person who was typed in this manner and was subsequently transfused with kk blood.

8.6 Major problems encountered in the course of the project

In the latter stages of the project the use of the AGPC method for the isolation of total RNA became problematic due to apparent RNase contamination. Normal A_{260} values were obtained, but gel analysis of the RNA showed a bright patch near the bottom of the gel, with no detectable RNA smear or ribosomal RNA bands. The source of the contamination could not be determined despite attempts to prepare RNA in three different laboratories using freshly prepared reagents and DEPC treated/heat treated equipment. Use of the other methods

described in section 8.1.1 failed to solve the problem, but sufficient undegraded total RNA was isolated from the contaminated samples to allow completion of the project.

The second problem resulted from the purchase of a Perkin Elmer 9600 DNA Thermal Cycler. This cycler was purchased as it was the only model which had been licensed for diagnostic purposes. Several months were spent attempting to optimise the conditions for use of this piece of equipment. Every aspect of the PCR reaction was varied, but the largest fragment which could be amplified was approximately 800bp. This was at least 300bp smaller than either of the *KEL*-related products which could be amplified routinely using a basic model Perkin Elmer DNA Thermal Cycler. This problem could not be solved by product specialists and, as a result, the remainder of the PCR reactions for this project were performed on the basic Perkin Elmer model used successfully in earlier work.

The third major problem encountered involved the use of the Applied Biosystems automated sequencer in the Department of Biochemistry at the University of Glasgow. A sensitivity problem appeared to be the cause of failure to obtain good quality data from this sequencer. Duplicate samples sequenced at IBGRL in Bristol gave good sequence data, but the Glasgow-based machine failed to yield any sequence information due to poor maintenance and a malfunctioning laser. The cycle sequencing kit, the thermal cycler used for the sequencing reactions and the gel constituents were each varied on different occasions. However, no sequence data could be obtained. A proportion of the cycle sequencing kit used in Glasgow was also used in Bristol and good data was obtained. It was concluded that the problem with the Glasgow-based machine was one of detection sensitivity. The reason for this was that the signal strengths for each of the four peaks were very low. Poor signal intensity was also indicated by the fact that the gel image which could be visualised on a computer screen was dark. The kit control sample gave short stretches of sequence in each case (200-250 nucleotides). The sequence obtained for the control was correct, but the length of the sequence obtained was much shorter than expected (by approximately 400 nucleotides). Direct sequencing of PCR products by this method uses only small amounts of specific target and should, therefore, give good data, as there is no other target sequence present for possible mispriming to occur. The fact that the results obtained in Glasgow were so poor is a strong indication to a mechanical fault in the machine. The low

peak intensity suggests that the signal detected by the laser is too weak and that, as a result, the information transferred to the computer is poor.

8.7 Major findings arising from this project

The results obtained from this project have illustrated how molecular biology techniques can be used to investigate the genetic basis of a blood group system. The project has allowed refinement of the location of the *KEL* gene on chromosome 7, it has independently confirmed the genetic basis for the difference between K and k antigens on the red cell surface and it has allowed the development of a PCR-based diagnostic assay for K/k genotyping. Another aspect of the Kell blood group system which was investigated was the genetic basis for the K_O phenotype. Although the results obtained from this section of the work proved difficult to interpret, they indicated that the particular example of K_O under investigation carried a *KEL* gene which was at least similar to *KEL* genes from normal phenotype individuals. This is in agreement with work published by two other groups, as discussed in section 8.5.

8.8 Future work

The work carried out in the course of this project has provided a good basis for the continuation of the investigation of the Kell blood group system. It constitutes the initial step which may lead to the determination of the genetic basis for the expression of other antigens within the Kell blood group system, particularly the Js^a/Js^b and the Kp^a/Kp^b allelic sets. The role of glycosylation with respect to antigen expression, although reported to be independent of the expression of the K and k antigens (Marsh *et al.*, 1990, Murphy *et al.*, 1993), is still to be conclusively established. A recent report gives evidence that the apparent mass of the Kell protein may be reduced in KK individuals, suggesting that glycosylation may be partly responsible for the differences in K/k antigenicity (Lee *et al.*, 1995 b). The genetic basis for the K_O and Kmod phenotypes must be determined and, in view of the fact that K_O individuals appear to be no different from normal Kell phenotype individuals with respect to coding DNA sequence (Lee *et al.*, 1995 a), there may be other genes involved which affect the expression of Kell blood group antigens in these individuals. This is the case with repressor type Rh null in the Rhesus blood group system (section 1.5).

The use of the diagnostic assay described in this project must be investigated with respect to amniotic cell samples and chorionic villus samples. The use of such samples has been reported for the PCR-based determination of Rhesus phenotype (Fisk *et al.*, 1994). Its validity with respect to rare phenotype samples such as K_O must be determined before this technique can be used routinely.

In an attempt to produce a suitable immunogen for the production of antibodies against Kell system antigens, the protein could be expressed. The expression of the Rh 32 kDa polypeptide has already been described (Suyama *et al.*, 1993). In this case, the cDNA encoding the polypeptide was inserted into a pGEM vector (Stratagene Inc.) and transcription/translation was performed in a reticulocyte lysate system. Transient expression of the polypeptide in COS-1 (Chinese Hamster Ovary Cells) was also reported by this group. The type of system chosen for expression of the Kell protein must be one which will allow glycosylation to take place. This excludes systems such as Xenopus oocytes, where there is no glycosylation of the translated protein. The expression of the Kell protein would be of great benefit, as some of the Kell system related alloimmune polyclonal antibodies are difficult to obtain in substantial quantity by virtue of the incidence of antigen expression. One report of expression of the Kell protein in COS-1 cells was recently reported, indicating that significant progress could be made with this line of investigation (Russo *et al.*, 1994). There is, therefore, sufficient scope for the continuation of this project, as a great deal of information is yet to be uncovered.

SECTION 9

Publications arising from the project

PUBLICATIONS

1

Regional chromosomal assignment of the Kell blood group locus (*KEL*) to chromosome 7q33-q35 by fluorescence *in situ* hybridization: evidence for the polypeptide nature of antigenic variation.

Margo T. Murphy, Norma Morrison, John S. Miles, Robin H. Fraser, Nigel K. Spurr, Elizabeth Boyd.

Human Genetics (1993) 91,585-588.

2

Independent confirmation that a single nucleotide substitution is responsible for the expression of Kell and Cellano antigens on the red cell surface.

Margo T. Murphy, Robin H. Fraser and John P. Goddard.

Submitted to Transfusion Medicine in September, 1995.

3

Development of a PCR based diagnostic assay for the determination of Kell genotype in donor blood samples.

Margo T. Murphy, Robin H. Fraser and John P. Goddard.

Submitted to Transfusion Medicine in September, 1995.

Hum Genet (1993) 91:585–588

human genetics

Regional chromosomal assignment of the Kell blood group locus *(KEL)* to chromosome 7q33–q35 by fluorescence in situ hybridization: evidence for the polypeptide nature of antigenic variation

Margo T. Murphy[1,2], Norma Morrison[3], John S. Miles[1], Robin H. Fraser[2], Nigel K. Spurr[4], Elizabeth Boyd[3]

[1] Department of Biochemistry, University of Glasgow, Glasgow G12 8QQ, UK
[2] Glasgow and West of Scotland Blood Transfusion Service, Law Hospital, Carluke, Lanarkshire ML5 8ES, UK
[3] Duncan Guthrie Institute of Medical Genetics, Yorkhill, Glasgow G3 8SJ, UK
[4] Imperial Cancer Research Fund, Human Genetic Resources Laboratory, Clare Hall Laboratories, Blanche Lane, South Mimms, Herts EN6 3LD, UK

Received: 4 September 1992

Abstract. The gene encoding the Kell blood group polypeptide has been localized to chromosome 7q33–35 by in situ hybridization using a biotinylated 1.1-kb DNA fragment containing the 3′ half of the human cDNA. This assignment is in accord with genetic localization using antigenic variation as a marker, and strongly suggests that Kell antigenic determinants are part of the polypeptide chain rather than the associated sugar molecules.

Introduction

The Kell blood group protein is a 93-kDa membrane glycoprotein found in human erythrocytes (Redman et al. 1986). Antigens associated with this protein are particularly immunogenic: antibodies to Kell system antigens have been implicated in several cases of haemolytic disease of the newborn and can cause severe transfusion reactions if incompatible red cells are introduced (Marsh and Redman 1990). The antigenic complexity displayed by the Kell blood group system is thought to be due to variation in the polypeptide chain as opposed to differences in glycosylation of the protein (Marsh and Redman 1990), although this has not been well defined.

The complementary DNA (cDNA) encoding the Kell blood group polypeptide has been cloned and its sequence shows homology to a family of zinc metalloglycoproteins with neutral endopeptidase activity (Lee et al. 1991). Using antigenic variation as a marker, genetic analysis has shown that the Kell blood group system is tightly linked to the prolactin-inducible protein (PIP) locus giving a provisional assignment of *KEL* to chromosome 7, probably in the region 7q32–q36 (Zelinski et al. 1991a, b). In order to determine genetically whether the Kell blood group polypeptide itself is the source of the antigenic variation, we have produced a probe, corresponding to the 3′ half of the Kell cDNA, and used it to map the Kell blood group polypeptide locus to chromosome 7q33–35 by in situ hybridization. The co-localization of the antigenic determinant and polypeptide loci strongly supports the suggestion that Kell antigenic determinants are part of the polypeptide chain rather than the associated sugar molecules.

Correspondence to: E. Boyd

Materials and methods

Human-rodent somatic cell hybrids

The human chromosome content of the somatic cell hybrids used in this study is shown in Table 1.

Chromosome mapping using the polymerase chain reaction (PCR)

Comparison of the Kell cDNA sequence with the gene structure of CALLA neutral endopeptidase (D'Adamio et al. 1989) suggested that positions 1214–1284 are within an exon. Two oligonucleotide primers P49, CCGTCGACACATGATCTTAGGGCTGGTGGTG (hybridizing to positions 1214–1233) and P54, CCGGATCCCT-TTCTGCGTGCCTCCTGGAATT (positions 1284–1262) amplified the same 87-bp DNA fragment in a PCR (Saiki et al. 1988) using cloned Kell cDNA (see below) and human genomic DNA, but not rodent genomic DNA, as targets (Fig. 1). DNA (0.5–2.5 μg) from the series of human-rodent somatic cell hybrids was amplified in 100 μl mixtures containing 1 μg of each primer, 0.2 m*M* each dNTP, 2.5 U *Taq* DNA polymerase and reaction buffer supplied by the manufacturer. The reactions were incubated for 3 min

at 94°C followed by 30 cycles of 94°C for 1 min, 65°C for 1 min, and 72°C for 1 min 30 s.

Chromosome preparation

Using standard cytogenetic methods, metaphase chromosome spreads were obtained from two human males of normal karyotype. Chromosome spreads in marked slide areas were banded, photographed and destained prior to hybridization as previously described (Garson et al. 1987; Boyd et al. 1989).

DNA probe

The PCR was used to generate a Kell DNA probe from reverse-transcribed mRNA derived from whole blood. The 1.1-kb DNA fragment, corresponding to positions 1214–2354 in the cDNA (Lee et al. 1991), was cloned into pTZ18U and its identity confirmed by sequencing.

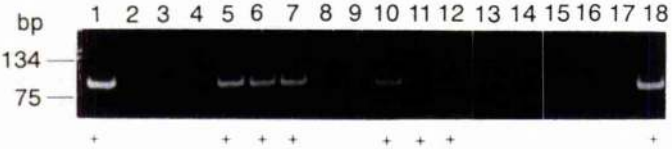


Fig. 1. Polyacrylamide gel analysis of the 87-bp product amplified by the polmerase chain reaction (PCR) from human-rodent somatic cell hybrids containing the human Kell sequence. Target DNA samples were from (1) cloned Kell cDNA, (2) mouse, (3) rat, (4) hamster, (5) human, (6) AMIR2VIII, (7) CTP412E3, (8) DUR4.3, (9) FG10, (10) 3W4C15, (11) MOG2, (12) Horp27-RC114, (13) SIF4A31, (14) F4SC13CI12, (15) TWIN19/D12, (16) FIR5R3, (17) FIR5, (18) CLONE21E

In situ hybridization

The probe was labelled by nick translation with biotin-11-dU (Sigma) and used at a concentration of 10–30 ng/µl for fluor cence in situ hybridization according to the procedure described Pinkel et al. (1986). Modifications detailed by Carter et al. (19 were employed but the prehybridization step was omitted for single-copy sequence probe. Hybridization was detected by in bation with avidin-fluorescein isothiocyanate (avidin-FITC, V tor Laboratories) and the signal amplified twice by two furt incubations with biotinylated anti-avidin (Vector Laboratori followed by avidin-FITC. Chromosomes were counterstained mounting the slides in antifade AF1 (Citifluor) containing µg/ml 4,6-diamidino-2-phenylindole (DAPI) and 0.4 µg/ml p pidium iodide and examined using a Zeiss Axioplan fluoresce microscope. FITC and propidium iodide were excited at 490 (Zeiss filter combination 9). Hybridization signals appear as yell green spots against the red propidium iodide counterstain. Pre ously photographed, DAPI-stained cells were relocated using Z filter combination 1. Signals visualized on the post-hybridizat metaphases using filter set 9 were marked on the photographs pre-hybridization, banded metaphases. The distribution of bridization in chromosome spreads was analysed statistically us the χ^2 test.

Results and discussion

Only those DNA samples from somatic cell hybrids co taining intact chromosome 7 generated the 87-bp K DNA fragment in the PCR indicating that the gene enco ing the Kell blood group polypeptide (*KEL*) is loca on chromosome 7 (Fig. 1, Table 1). The somatic cell h brid FIR5 contains an X/7 translocation (X:7) Xqt q13:7pter–q22 (Hobart et al. 1981) and the 87-bp hum

Table 1. Segregation of *KEL* in relation to human chromosomal content in 13 human-rodent somatic cell hybrids. +, human chromosome present; –, human chromosome not detected; tr, presence of chromosome at levels <10%; nt, not tested. Numbers correspond to the following references: (1) Jones et al. (1976), (2) Nabholz et al. (1969), (3) Povey et al. (1980), (4) Van Heyningen et (1975), (5) Croce and Koprowski (1974), (6) Solomon et (1976), (7) Kielty et al. (1982), (8) Edwards et al. (1985), (9) H terkamp et al. (1982), (10) Phillips et al. (1985), (11) Hobart et (1981)

Hybrid	Ref.	PCR product	Chromosome number																					
			1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22
AMIR2VIII		+	+	+	+	+	tr	tr	tr	tr	–	+	+	–	–	+	+	–	+	+	+	+	+	+
CTP41E3	1	+	–	+	tr	–	–	+	+	–	–	–	–	–	–	+	–	–	–	–	–	–	–	–
3W4C15	2	+	–	+	–	–	–	–	+	–	–	+	+	+	–	+	+	–	+	–	–	–	+	–
MOG2	3	+	+	–	+	+	+	+	+	nt	+	nt	+	+	nt	+	+	+	+	+	+	–	+	–
Horp27RC14	4	+	–	+	–	+	–	–	+	–	–	–	+	+	–	+	+	–	–	–	–	–	+	+
CLONE21E	5	+	–	–	–	–	–	–	+	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–
DUR4.3	6	–	–	–	+	–	+	–	–	–	–	+	+	+	+	+	[b]	–	+	+	–	+	+	–
FG10	7	–	–	+	–	–	+	–	–	+	–	+	–	–	–	–	+	–	–	+	–	–	+	–
SIF4A31	8	–	–	–	+	+	–	+	–	–	–	–	–	–	–	+	–	–	+	–	–	–	–	–
F4Sc13CI12	9	–	[c]	–	–	–	–	–	–	–	–	–	–	–	–	+	–	–	–	–	–	–	–	–
TWIN19/D12	10	–	+	–	+	–	–	+	–	+	–	–	–	+	–	+	–	+	+	+	–	+	+	+
FIR5R3	11	–	–	–	–	–	–	–	–	–	–	–	–	–	–	+	–	–	–	tr	–	–	–	–
FIR5	11	–	–	–	–	–	–	–	[d]	–	–	–	–	–	–	+	–	–	–	tr	–	–	–	–
Kell/intact chromosome discordance		+–	4	2	3	3	4	3	0	4	5	3	2	3	5	1	2	5	3	4	4	5	2	4
		–+	1	1	3	1	2	2	0	2	0	2	1	2	1	6	1	1	3	5	0	2	3	1

[a] X/Y translocation Xqter–p22.3:Yq1.1–qter; [b] X/15 translocation Xq11–qter:15p11–qter; [c] contains only 1p; [d] X/7 translocat Xqter–q13:7pter–q22

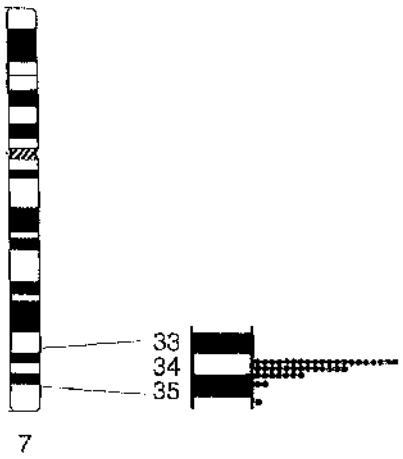


Fig. 2. Signal distribution over 7q33–35 on 32 copies of chromosome 7 following hybridization to the Kell cDNA probe

Kell sequence is not amplified from this DNA by the PCR (Fig. 1). This indicates that the location of the Kell gene is 7q22–qter. To localize the region further, in situ hybridization was carried out. Twenty-four normal human male metaphases were scored following hybridization, and the positions of 118 signals recorded. Of these, a highly significant ($P < 0.005$) 37 signals were located on chromosome 7 with 22 comprising a signal peak on 7q33–35. A second hybridization was performed and the signal distribution at 7q33–35 noted on 32 chromosomes 7 which were sufficiently long to permit assignment of signals to a single band (Fig. 2).

The Kell blood group locus was provisionally mapped to the region 7q32–36 through linkage to the PIP locus (Zelinski et al. 1991a,b) and this has recently been confirmed by the demonstration of linkage to the cystic fibrosis locus (Purohit et al. 1992). The results presented here refine the position of the Kell locus to 7q34, placing it distal to the cystic fibrosis locus (7q31–32), and proximal to the T cell receptor beta cluster located in 7q35 (Barker et al. 1984).

The Kell blood group protein is not related to a human band 3-like protein whose gene is located at 7q35–36 (Palumbo et al. 1986). Because of the absence of many markers in the region, the localization of *KEL* to 7q33–35, probably 7q34, using a cDNA probe should prove useful for further genetic analysis in this region.

This assignment using the cDNA probe, which defines the gene encoding the Kell polypeptide, is to the same region of chromosome 7 as indicated by the genetic analysis of the antigenic variation associated with the Kell blood group system. This co-localisation of the antigenic determinant and the polypeptide loci gives strong genetic support to the suggestion that the Kell antigenic determinants are part of the polypeptide chain as opposed to the associated sugar molecules. This finding has important implications for the further analysis of the genetic basis of antigenic variation in the Kell blood group system.

Acknowledgments. We would like to thank Lesley Rooke for her help in preparing the somatic cell hybrid DNA.

References

Barker PE, Ruddle FH, Royer HD, Acuto O, Reinherz EL (1984) Chromosomal location of human T cell receptor gene Ti beta. Science 226:348–349

Boyd E, Theriault A, Goddard JP, Kalaitsidaki M, Spathas DH, Connor JM (1989) Chromosomal assignment of a glutamic acid transfer RNA ($tRNA^{Glu}$) gene to 1p36. Hum Genet 81:153–156

Carter NP, Ferguson-Smith MA, Perryman MT, Telenius H, Pelmear AH, Leversha MA, Glancy MT, Wood SL, Cook K, Dyson HM, Ferguson-Smith ME, Willatt LR (1992) Reverse chromosome painting: a method for the rapid analysis of aberrant chromosomes in clinical cytogenetics. J Med Genet 29: 299–307

Croce C, Koprowski HM (1974) Somatic cell hybrids between mouse peritoneal macrophages and SV-40 transformed human cells. I. Positive control of the transformed phenotype by the human chromosome 7 carrying SV-40 genome. J Exp Med 140: 1221–1229

D'Adamio L, Shipp MA, Masteller EL, Reinherz EL (1989) Organization of the gene encoding common acute lymphoblastic leukemia antigen (neutral endopeptidase 24.11): multiple miniexons and separate 5′ untranslated regions. Proc Natl Acad Sci USA 86:7103–7107

Edwards YH, Parkar M, Povey S, West LF, Parrington JM, Solomon E (1985) Human myosin heavy chain genes assigned to chromosome 17 using a cDNA clone as a probe. Ann Hum Genet 49:101–109

Garson JA, Berghe JA van den, Kemshead JT (1987) Novel non-isotopic in situ hybridization technique detects small (1 kb) unique sequences in routinely G-banded human chromosomes: fine mapping of N-myc and beta-NGF genes. Nucleic Acids Res 15:4761–4770

Heisterkamp N, Groffen J, Stephenson JR, Spurr NK, Goodfellow PN, Solomon E, Carritt B, Bodmer WF (1982) Chromosomal localization of human cellular homologues to two viral oncogenes. Nature 299:747–749

Hobart MJ, Rabbitts TH, Goodfellow PN, Solomon E, Chambers S, Spurr N, Povey S (1981) Immunoglobulin heavy chain genes in humans are located on chromosome 14. Ann Hum Genet 45: 331–335

Jones EA, Goodfellow PN, Kennett RH, Bodmer WF (1976) The independent expression of HLA and ß2 microglobulin on mouse-human hybrids. Somatic Cell Genet 2:483–496

Kielty CM, Povey S, Hopkinson DA (1982) Regulation of expression of liver-specific enzymes. III. Further analysis of a series of rat hepatoma × human somatic cell hybrids. Ann Hum Genet 46:307–327

Lee S, Zambas ED, Marsh WL, Redman CM (1991) Molecular cloning and primary structure of Kell blood group protein. Proc Natl Acad Sci USA 88:6353–6357

Marsh WL, Redman CM (1990) The Kell blood group system: a review. Transfusion 30:158–167

Nabholz M, Miggiano V, Bodmer WF (1969) Genetic analysis with human-mouse somatic cell hybrids. Nature 223:358–363

Palumbo AP, Isobe M, Huebner K, Shane S, Rovera G, Demuth D, Curtis PJ, Ballantine M, Croce CM, Showe LC (1986) Chromosomal localization of a human band 3-like gene to region 7q35→7q36. Am J Hum Genet 39:307–316

Phillips IR, Shephard EA, Povey S, Davis MB_a, Kelsey G, Monteiro M, West LF, Cowell F (1985) A cytochrome P-450 gene family mapped to human chromosome 19. Ann Hum Genet 49: 267–274

Pinkel D, Straume T, Gray JW (1986) Cytogenetic analysis using quantitative, high sensitivity, fluorescence hybridization. Proc Natl Acad Sci USA 83:2934–2938

Povey S, Jeremiah SJ, Barker RF, Hopkinson DA, Robson EB, Cook PJL, Solomon E, Bobrow M, Carritt B, Buckton KE (1980) Assignment of the human locus determining phosphoglycolate phosphatase (PGP) to chromosome 16. Ann Hum Genet 43:241–248

Purohit KR, Weber JL, Ward LJ, Keats BJB (1992) The Kell blood group locus is close to the cystic fibrosis locus on chromosome 7. Hum Genet 89:457–458

Redman CM, Avellino G, Pfeffer SR, Mukherjee TK, Nichols M, Rubinstein P, Marsh WL (1986) Kell blood group antigens are part of a 93000-dalton red cell membrane protein. J Biol Chem 261:9521–9525

Saiki RK, Gelfand DH, Stoffel S, Scharf SJ, Higuchi R, Horn GT, Mullis KB, Erlich HA (1988) Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase. Science 239:487–491

Solomon E, Bobrow M, Goodfellow PN, Bodmer WF, Swallow DM, Povey S, Nuel B (1976) Human gene mapping using an X/autosome translocation. Somatic Cell Genet 2:125–140

Van Heyningen V, Bobrow L, Bodmer WF, Gardiner SE, Povey Hopkinson DA (1975) Chromosome assignment of some man loci: mitochondrial malate dehydrogenase to 7, mann phosphate isomerase and pyruvate kinase to 15, and proba esterase D to 13. Ann Hum Genet 38:295–303

Zelinski T, Coghlan G, Myal Y, Shiu RPC, Philipps S, White Lewis M (1991a) Genetic linkage between the Kell blood gr system and prolactin-inducible protein loci: provisional assi ment of *KEL* to chromosome 7. Ann Hum Genet 55:137–14

Zelinski TA, Coghlan GE, Myal Y, White LJ, Phillips SE (199 Assignment of the Kell blood group locus to chromosome Cytogenet Cell Genet 58:1927

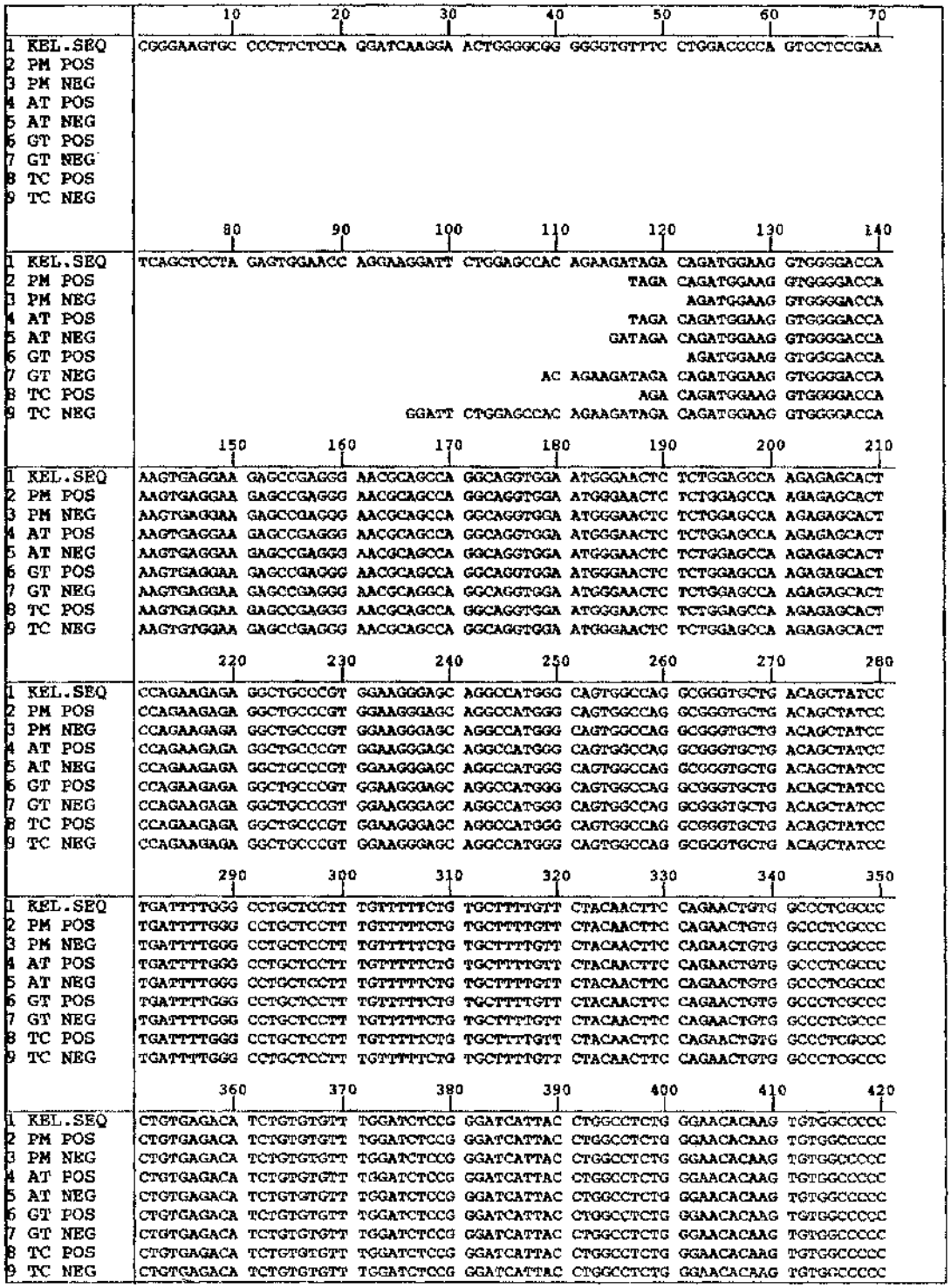
SECTION 10

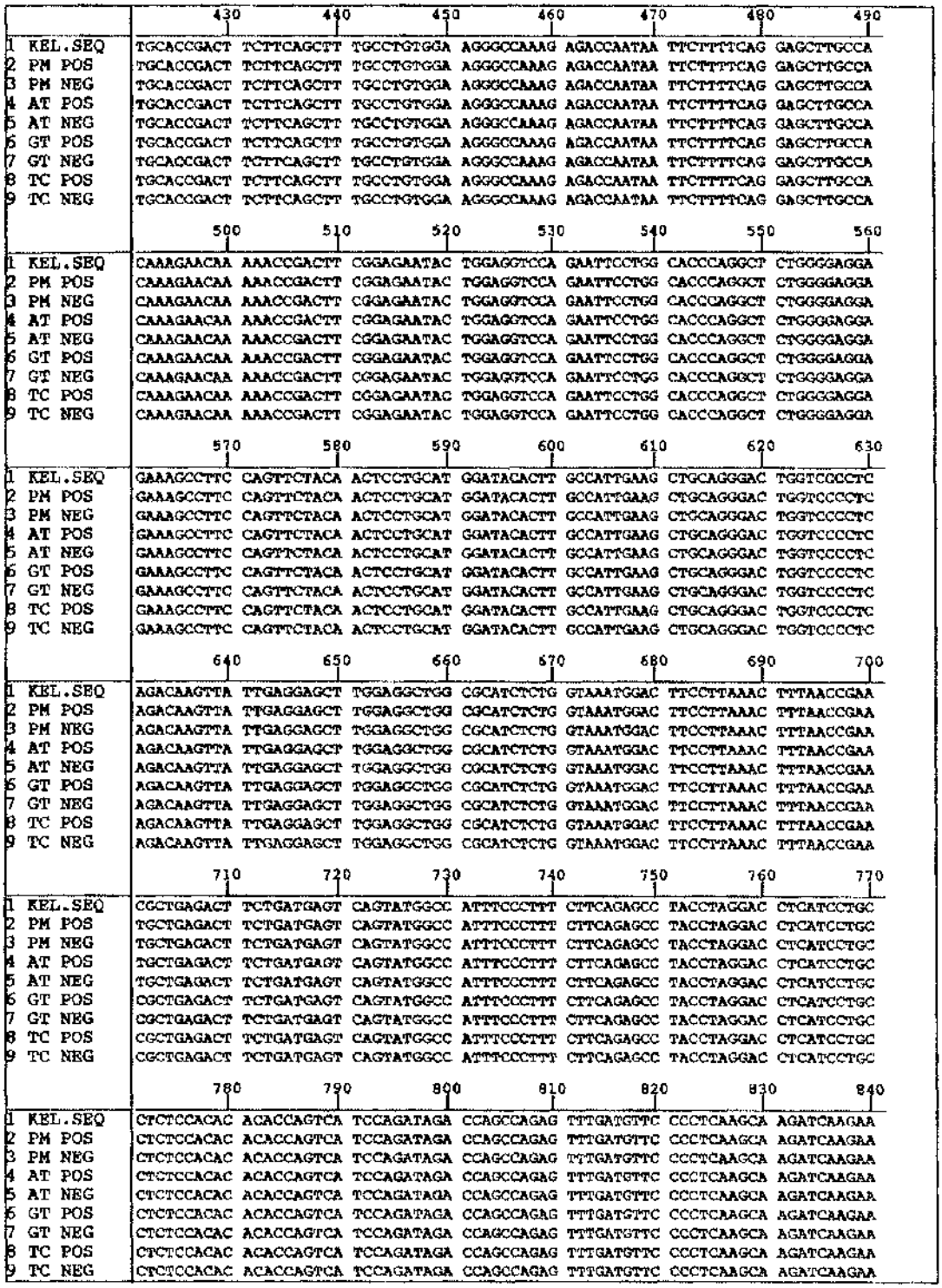
Appendix

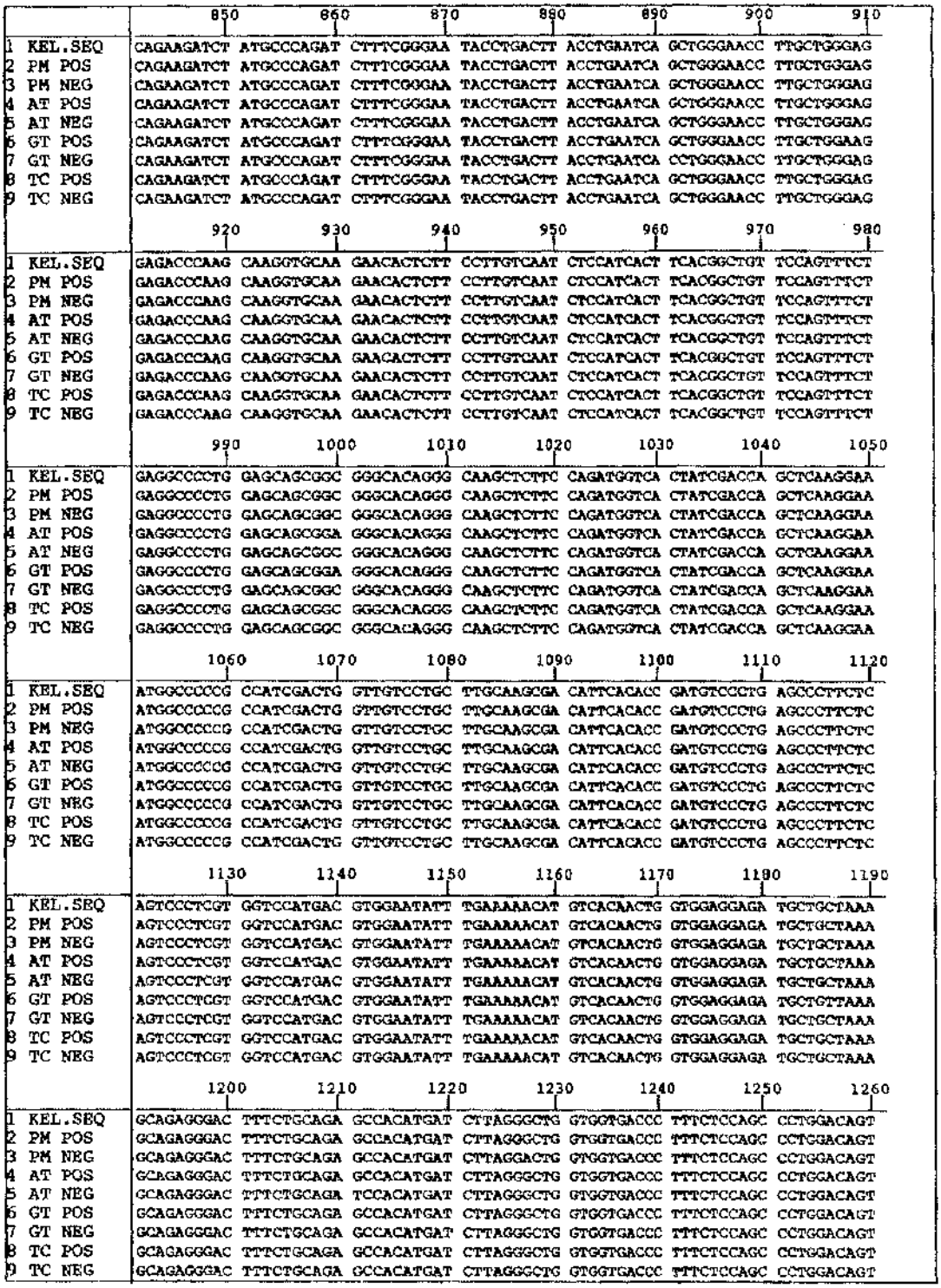
Appendix Direct automated sequencing of *KEL* PCR products

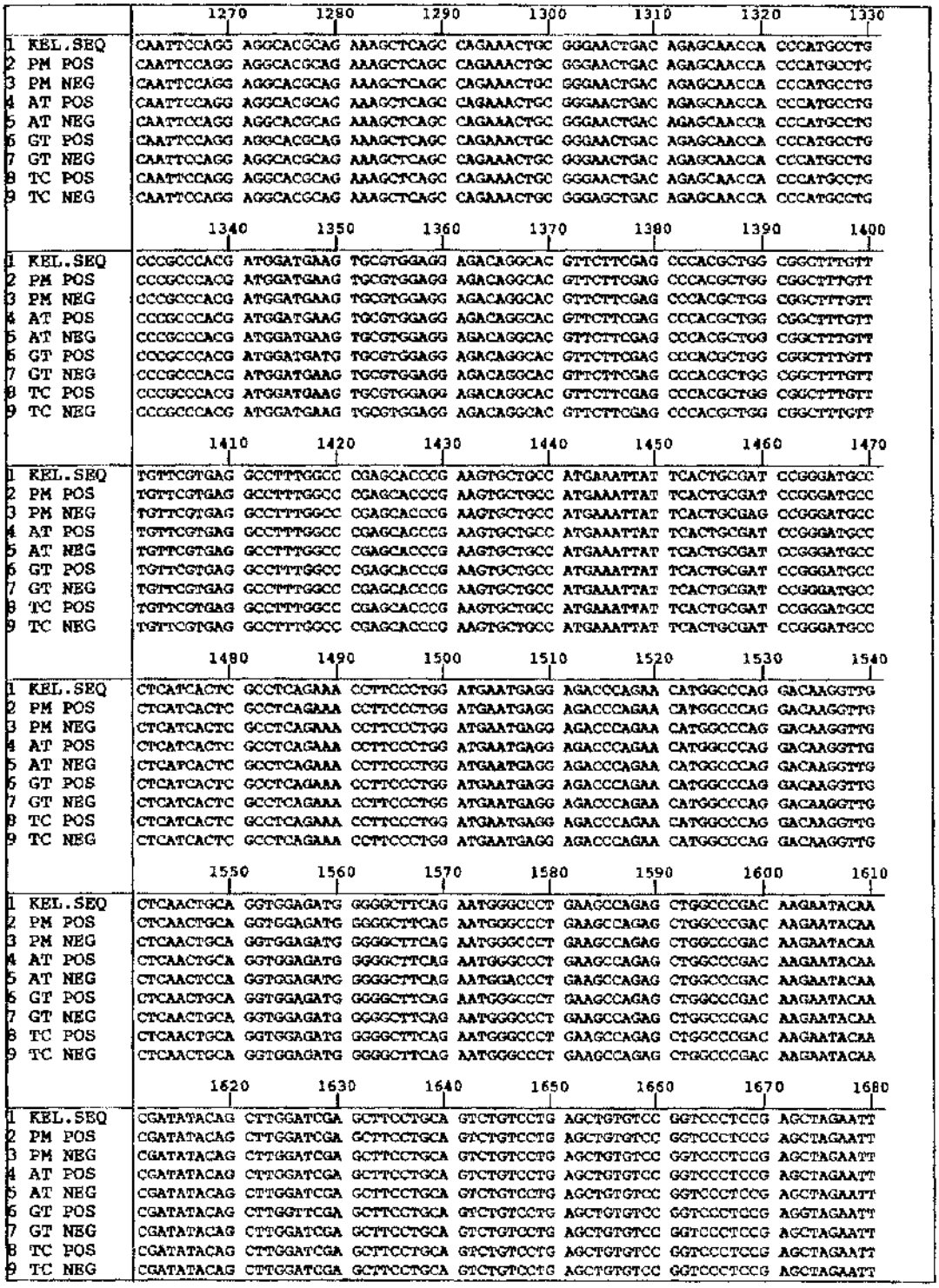
The data shown in this appendix are those obtained from the direct automated sequencing of *KEL* PCR products. Contiguous sequences for the coding region of both positive and negative strands of four individuals are shown. The coding region of the *KEL* cDNA starts at nucleotide 124 and finishes at nucleotide 2322. The published *KEL* cDNA sequence (KEL.SEQ) as reported by Lee *et al.*, 1991 has been included for reference. All negative strand data are given as the reverse complement sequence. The four donors were as follows:-

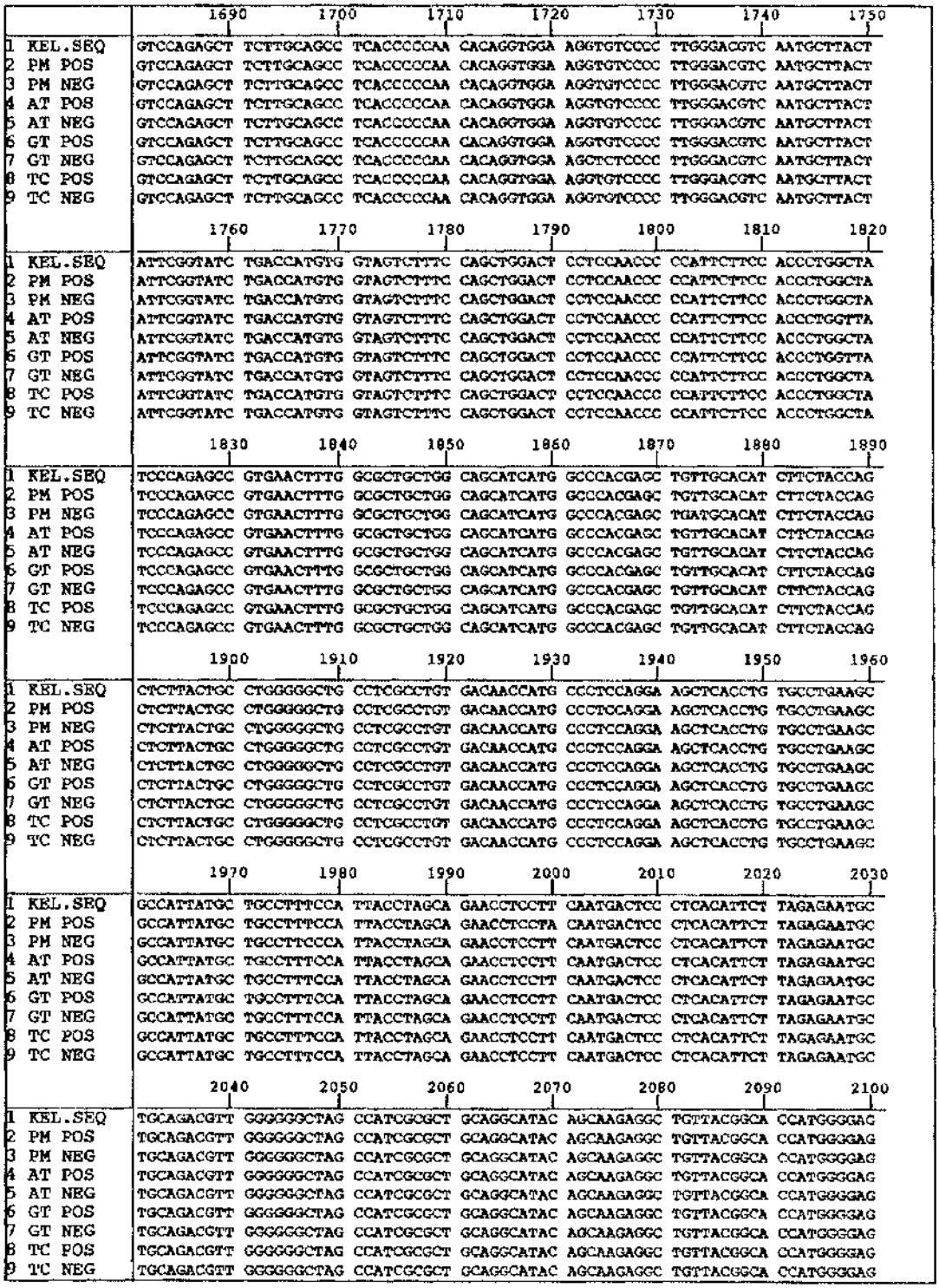
PM KK donor
AT KK donor
GT kk donor
TC kk donor

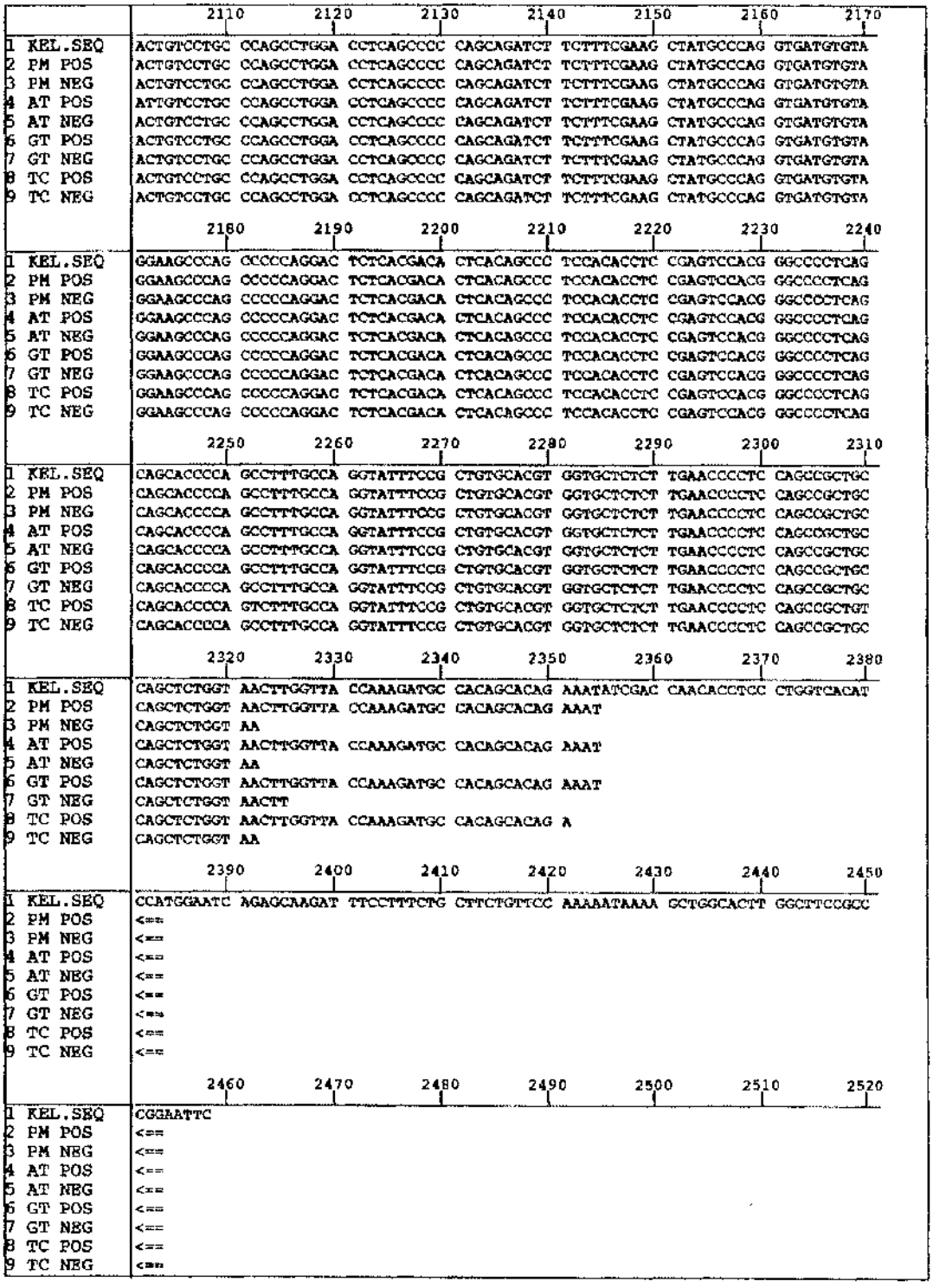

10 20 30 40 50 60 70
1 KEL.SEQ CGGGAAGTGC CCCTTCTCCA GGATCAAGGA ACTGGGGCGG GGGGTGTTTC CTGGACCCCA GTCCTCCGAA
2 PM POS
3 PM NEG
4 AT POS
5 AT NEG
6 GT POS
7 GT NEG
8 TC POS
9 TC NEG
80 90 100 110 120 130 140
1 KEL.SEQ TCAGCTCCTA GAGTGGAACC AGGAAGGATT CTGGAGCCAC AGAAGATAGA CAGATGGAAG GTGGGGACCA
2 PM POS TAGA CAGATGGAAG GTGGGGACCA
3 PM NEG AGATGGAAG GTGGGGACCA
4 AT POS TAGA CAGATGGAAG GTGGGGACCA
5 AT NEG GATAGA CAGATGGAAG GTGGGGACCA
6 GT POS AGATGGAAG GTGGGGACCA
7 GT NEG AC AGAAGATAGA CAGATGGAAG GTGGGGACCA
8 TC POS AGA CAGATGGAAG GTGGGGACCA
9 TC NEG GGATT CTGGAGCCAC AGAAGATAGA CAGATGGAAG GTGGGGACCA
150 160 170 180 190 200 210
1 KEL.SEQ AAGTGAGGAA GAGCCGAGGG AACGCAGCCA GGCAGGTGGA ATGGGAACTC TCTGGAGCCA AGAGAGCACT
2 PM POS AAGTGAGGAA GAGCCGAGGG AACGCAGCCA GGCAGGTGGA ATGGGAACTC TCTGGAGCCA AGAGAGCACT
3 PM NEG AAGTGAGGAA GAGCCGAGGG AACGCAGCCA GGCAGGTGGA ATGGGAACTC TCTGGAGCCA AGAGAGCACT
4 AT POS AAGTGAGGAA GAGCCGAGGG AACGCAGCCA GGCAGGTGGA ATGGGAACTC TCTGGAGCCA AGAGAGCACT
5 AT NEG AAGTGAGGAA GAGCCGAGGG AACGCAGCCA GGCAGGTGGA ATGGGAACTC TCTGGAGCCA AGAGAGCACT
6 GT POS AAGTGAGGAA GAGCCGAGGG AACGCAGCCA GGCAGGTGGA ATGGGAACTC TCTGGAGCCA AGAGAGCACT
7 GT NEG AAGTGAGGAA GAGCCGAGGG AACGCAGGCA GGCAGGTGGA ATGGGAACTC TCTGGAGCCA AGAGAGCACT
8 TC POS AAGTGAGGAA GAGCCGAGGG AACGCAGCCA GGCAGGTGGA ATGGGAACTC TCTGGAGCCA AGAGAGCACT
9 TC NEG AAGTGTGGAA GAGCCGAGGG AACGCAGCCA GGCAGGTGGA ATGGGAACTC TCTGGAGCCA AGAGAGCACT
220 230 240 250 260 270 280
1 KEL.SEQ CCAGAAGAGA GGCTGCCCGT GGAAGGGAGC AGGCCATGGG CAGTGGCCAG GCGGGTGCTG ACAGCTATCC
2 PM POS CCAGAAGAGA GGCTGCCCGT GGAAGGGAGC AGGCCATGGG CAGTGGCCAG GCGGGTGCTG ACAGCTATCC
3 PM NEG CCAGAAGAGA GGCTGCCCGT GGAAGGGAGC AGGCCATGGG CAGTGGCCAG GCGGGTGCTG ACAGCTATCC
4 AT POS CCAGAAGAGA GGCTGCCCGT GGAAGGGAGC AGGCCATGGG CAGTGGCCAG GCGGGTGCTG ACAGCTATCC
5 AT NEG CCAGAAGAGA GGCTGCCCGT GGAAGGGAGC AGGCCATGGG CAGTGGCCAG GCGGGTGCTG ACAGCTATCC
6 GT POS CCAGAAGAGA GGCTGCCCGT GGAAGGGAGC AGGCCATGGG CAGTGGCCAG GCGGGTGCTG ACAGCTATCC
7 GT NEG CCAGAAGAGA GGCTGCCCGT GGAAGGGAGC AGGCCATGGG CAGTGGCCAG GCGGGTGCTG ACAGCTATCC
8 TC POS CCAGAAGAGA GGCTGCCCGT GGAAGGGAGC AGGCCATGGG CAGTGGCCAG GCGGGTGCTG ACAGCTATCC
9 TC NEG CCAGAAGAGA GGCTGCCCGT GGAAGGGAGC AGGCCATGGG CAGTGGCCAG GCGGGTGCTG ACAGCTATCC
290 300 310 320 330 340 350
1 KEL.SEQ TGATTTTGGG CCTGCTCCTT TGTTTTTCTG TGCTTTTGTT CTACAACTTC CAGAACTGTG GCCCTCGCCC
2 PM POS TGATTTTGGG CCTGCTCCTT TGTTTTTCTG TGCTTTTGTT CTACAACTTC CAGAACTGTG GCCCTCGCCC
3 PM NEG TGATTTTGGG CCTGCTCCTT TGTTTTTCTG TGCTTTTGTT CTACAACTTC CAGAACTGTG GCCCTCGCCC
4 AT POS TGATTTTGGG CCTGCTCCTT TGTTTTTCTG TGCTTTTGTT CTACAACTTC CAGAACTGTG GCCCTCGCCC
5 AT NEG TGATTTTGGG CCTGCTCCTT TGTTTTTCTG TGCTTTTGTT CTACAACTTC CAGAACTGTG GCCCTCGCCC
6 GT POS TGATTTTGGG CCTGCTCCTT TGTTTTTCTG TGCTTTTGTT CTACAACTTC CAGAACTGTG GCCCTCGCCC
7 GT NEG TGATTTTGGG CCTGCTCCTT TGTTTTTCTG TGCTTTTGTT CTACAACTTC CAGAACTGTG GCCCTCGCCC
8 TC POS TGATTTTGGG CCTGCTCCTT TGTTTTTCTG TGCTTTTGTT CTACAACTTC CAGAACTGTG GCCCTCGCCC
9 TC NEG TGATTTTGGG CCTGCTCCTT TGTTTTTCTG TGCTTTTGTT CTACAACTTC CAGAACTGTG GCCCTCGCCC
360 370 380 390 400 410 420
1 KEL.SEQ CTGTGAGACA TCTGTGTGTT TGGATCTCCG GGATCATTAC CTGGCCTCTG GGAACACAAG TGTGGCCCCC
2 PM POS CTGTGAGACA TCTGTGTGTT TGGATCTCCG GGATCATTAC CTGGCCTCTG GGAACACAAG TGTGGCCCCC
3 PM NEG CTGTGAGACA TCTGTGTGTT TGGATCTCCG GGATCATTAC CTGGCCTCTG GGAACACAAG TGTGGCCCCC
4 AT POS CTGTGAGACA TCTGTGTGTT TGGATCTCCG GGATCATTAC CTGGCCTCTG GGAACACAAG TGTGGCCCCC
5 AT NEG CTGTGAGACA TCTGTGTGTT TGGATCTCCG GGATCATTAC CTGGCCTCTG GGAACACAAG TGTGGCCCCC
6 GT POS CTGTGAGACA TCTGTGTGTT TGGATCTCCG GGATCATTAC CTGGCCTCTG GGAACACAAG TGTGGCCCCC
7 GT NEG CTGTGAGACA TCTGTGTGTT TGGATCTCCG GGATCATTAC CTGGCCTCTG GGAACACAAG TGTGGCCCCC
8 TC POS CTGTGAGACA TCTGTGTGTT TGGATCTCCG GGATCATTAC CTGGCCTCTG GGAACACAAG TGTGGCCCCC
9 TC NEG CTGTGAGACA TCTGTGTGTT TGGATCTCCG GGATCATTAC CTGGCCTCTG GGAACACAAG TGTGGCCCCC


430 440 450 460 470 480 490
1 KEL.SEQ TGCACCGACT TCTTCAGCTT TGCCTGTGGA AGGGCCAAAG AGACCAATAA TTCTTTTCAG GAGCTTGCCA
2 PM POS TGCACCGACT TCTTCAGCTT TGCCTGTGGA AGGGCCAAAG AGACCAATAA TTCTTTTCAG GAGCTTGCCA
3 PM NEG TGCACCGACT TCTTCAGCTT TGCCTGTGGA AGGGCCAAAG AGACCAATAA TTCTTTTCAG GAGCTTGCCA
4 AT POS TGCACCGACT TCTTCAGCTT TGCCTGTGGA AGGGCCAAAG AGACCAATAA TTCTTTTCAG GAGCTTGCCA
5 AT NEG TGCACCGACT TCTTCAGCTT TGCCTGTGGA AGGGCCAAAG AGACCAATAA TTCTTTTCAG GAGCTTGCCA
6 GT POS TGCACCGACT TCTTCAGCTT TGCCTGTGGA AGGGCCAAAG AGACCAATAA TTCTTTTCAG GAGCTTGCCA
7 GT NEG TGCACCGACT TCTTCAGCTT TGCCTGTGGA AGGGCCAAAG AGACCAATAA TTCTTTTCAG GAGCTTGCCA
8 TC POS TGCACCGACT TCTTCAGCTT TGCCTGTGGA AGGGCCAAAG AGACCAATAA TTCTTTTCAG GAGCTTGCCA
9 TC NEG TGCACCGACT TCTTCAGCTT TGCCTGTGGA AGGGCCAAAG AGACCAATAA TTCTTTTCAG GAGCTTGCCA
500 510 520 530 540 550 560
1 KEL.SEQ CAAAGAACAA AAACCGACTT CGGAGAATAC TGGAGGTCCA GAATTCCTGG CACCCAGGCT CTGGGGAGGA
2 PM POS CAAAGAACAA AAACCGACTT CGGAGAATAC TGGAGGTCCA GAATTCCTGG CACCCAGGCT CTGGGGAGGA
3 PM NEG CAAAGAACAA AAACCGACTT CGGAGAATAC TGGAGGTCCA GAATTCCTGG CACCCAGGCT CTGGGGAGGA
4 AT POS CAAAGAACAA AAACCGACTT CGGAGAATAC TGGAGGTCCA GAATTCCTGG CACCCAGGCT CTGGGGAGGA
5 AT NEG CAAAGAACAA AAACCGACTT CGGAGAATAC TGGAGGTCCA GAATTCCTGG CACCCAGGCT CTGGGGAGGA
6 GT POS CAAAGAACAA AAACCGACTT CGGAGAATAC TGGAGGTCCA GAATTCCTGG CACCCAGGCT CTGGGGAGGA
7 GT NEG CAAAGAACAA AAACCGACTT CGGAGAATAC TGGAGGTCCA GAATTCCTGG CACCCAGGCT CTGGGGAGGA
8 TC POS CAAAGAACAA AAACCGACTT CGGAGAATAC TGGAGGTCCA GAATTCCTGG CACCCAGGCT CTGGGGAGGA
9 TC NEG CAAAGAACAA AAACCGACTT CGGAGAATAC TGGAGGTCCA GAATTCCTGG CACCCAGGCT CTGGGGAGGA
570 580 590 600 610 620 630
1 KEL.SEQ GAAAGCCTTC CAGTTCTACA ACTCCTGCAT GGATACACTT GCCATTGAAG CTGCAGGGAC TGGTCCCCTC
2 PM POS GAAAGCCTTC CAGTTCTACA ACTCCTGCAT GGATACACTT GCCATTGAAG CTGCAGGGAC TGGTCCCCTC
3 PM NEG GAAAGCCTTC CAGTTCTACA ACTCCTGCAT GGATACACTT GCCATTGAAG CTGCAGGGAC TGGTCCCCTC
4 AT POS GAAAGCCTTC CAGTTCTACA ACTCCTGCAT GGATACACTT GCCATTGAAG CTGCAGGGAC TGGTCCCCTC
5 AT NEG GAAAGCCTTC CAGTTCTACA ACTCCTGCAT GGATACACTT GCCATTGAAG CTGCAGGGAC TGGTCCCCTC
6 GT POS GAAAGCCTTC CAGTTCTACA ACTCCTGCAT GGATACACTT GCCATTGAAG CTGCAGGGAC TGGTCCCCTC
7 GT NEG GAAAGCCTTC CAGTTCTACA ACTCCTGCAT GGATACACTT GCCATTGAAG CTGCAGGGAC TGGTCCCCTC
8 TC POS GAAAGCCTTC CAGTTCTACA ACTCCTGCAT GGATACACTT GCCATTGAAG CTGCAGGGAC TGGTCCCCTC
9 TC NEG GAAAGCCTTC CAGTTCTACA ACTCCTGCAT GGATACACTT GCCATTGAAG CTGCAGGGAC TGGTCCCCTC
640 650 660 670 680 690 700
1 KEL.SEQ AGACAAGTTA TTGAGGAGCT TGGAGGCTGG CGCATCTCTG GTAAATGGAC TTCCTTAAAC TTTAACCGAA
2 PM POS AGACAAGTTA TTGAGGAGCT TGGAGGCTGG CGCATCTCTG GTAAATGGAC TTCCTTAAAC TTTAACCGAA
3 PM NEG AGACAAGTTA TTGAGGAGCT TGGAGGCTGG CGCATCTCTG GTAAATGGAC TTCCTTAAAC TTTAACCGAA
4 AT POS AGACAAGTTA TTGAGGAGCT TGGAGGCTGG CGCATCTCTG GTAAATGGAC TTCCTTAAAC TTTAACCGAA
5 AT NEG AGACAAGTTA TTGAGGAGCT TGGAGGCTGG CGCATCTCTG GTAAATGGAC TTCCTTAAAC TTTAACCGAA
6 GT POS AGACAAGTTA TTGAGGAGCT TGGAGGCTGG CGCATCTCTG GTAAATGGAC TTCCTTAAAC TTTAACCGAA
7 GT NEG AGACAAGTTA TTGAGGAGCT TGGAGGCTGG CGCATCTCTG GTAAATGGAC TTCCTTAAAC TTTAACCGAA
8 TC POS AGACAAGTTA TTGAGGAGCT TGGAGGCTGG CGCATCTCTG GTAAATGGAC TTCCTTAAAC TTTAACCGAA
9 TC NEG AGACAAGTTA TTGAGGAGCT TGGAGGCTGG CGCATCTCTG GTAAATGGAC TTCCTTAAAC TTTAACCGAA
710 720 730 740 750 760 770
1 KEL.SEQ CGCTGAGACT TCTGATGAGT CAGTATGGCC ATTTCCCTTT CTTCAGAGCC TACCTAGGAC CTCATCCTGC
2 PM POS TGCTGAGACT TCTGATGAGT CAGTATGGCC ATTTCCCTTT CTTCAGAGCC TACCTAGGAC CTCATCCTGC
3 PM NEG TGCTGAGACT TCTGATGAGT CAGTATGGCC ATTTCCCTTT CTTCAGAGCC TACCTAGGAC CTCATCCTGC
4 AT POS TGCTGAGACT TCTGATGAGT CAGTATGGCC ATTTCCCTTT CTTCAGAGCC TACCTAGGAC CTCATCCTGC
5 AT NEG TGCTGAGACT TCTGATGAGT CAGTATGGCC ATTTCCCTTT CTTCAGAGCC TACCTAGGAC CTCATCCTGC
6 GT POS CGCTGAGACT TCTGATGAGT CAGTATGGCC ATTTCCCTTT CTTCAGAGCC TACCTAGGAC CTCATCCTGC
7 GT NEG CGCTGAGACT TCTGATGAGT CAGTATGGCC ATTTCCCTTT CTTCAGAGCC TACCTAGGAC CTCATCCTGC
8 TC POS CGCTGAGACT TCTGATGAGT CAGTATGGCC ATTTCCCTTT CTTCAGAGCC TACCTAGGAC CTCATCCTGC
9 TC NEG CGCTGAGACT TCTGATGAGT CAGTATGGCC ATTTCCCTTT CTTCAGAGCC TACCTAGGAC CTCATCCTGC
780 790 800 810 820 830 840
1 KEL.SEQ CTCTCCACAC ACACCAGTCA TCCAGATAGA CCAGCCAGAG TTTGATGTTC CCCTCAAGCA AGATCAAGAA
2 PM POS CTCTCCACAC ACACCAGTCA TCCAGATAGA CCAGCCAGAG TTTGATGTTC CCCTCAAGCA AGATCAAGAA
3 PM NEG CTCTCCACAC ACACCAGTCA TCCAGATAGA CCAGCCAGAG TTTGATGTTC CCCTCAAGCA AGATCAAGAA
4 AT POS CTCTCCACAC ACACCAGTCA TCCAGATAGA CCAGCCAGAG TTTGATGTTC CCCTCAAGCA AGATCAAGAA
5 AT NEG CTCTCCACAC ACACCAGTCA TCCAGATAGA CCAGCCAGAG TTTGATGTTC CCCTCAAGCA AGATCAAGAA
6 GT POS CTCTCCACAC ACACCAGTCA TCCAGATAGA CCAGCCAGAG TTTGATGTTC CCCTCAAGCA AGATCAAGAA
7 GT NEG CTCTCCACAC ACACCAGTCA TCCAGATAGA CCAGCCAGAG TTTGATGTTC CCCTCAAGCA AGATCAAGAA
8 TC POS CTCTCCACAC ACACCAGTCA TCCAGATAGA CCAGCCAGAG TTTGATGTTC CCCTCAAGCA AGATCAAGAA
9 TC NEG CTCTCCACAC ACACCAGTCA TCCAGATAGA CCAGCCAGAG TTTGATGTTC CCCTCAAGCA AGATCAAGAA


850 860 870 880 890 900 910
1 KEL.SEQ CAGAAGATCT ATGCCCAGAT CTTTCGGGAA TACCTGACTT ACCTGAATCA GCTGGGAACC TTGCTGGGAG
2 PM POS CAGAAGATCT ATGCCCAGAT CTTTCGGGAA TACCTGACTT ACCTGAATCA GCTGGGAACC TTGCTGGGAG
3 PM NEG CAGAAGATCT ATGCCCAGAT CTTTCGGGAA TACCTGACTT ACCTGAATCA GCTGGGAACC TTGCTGGGAG
4 AT POS CAGAAGATCT ATGCCCAGAT CTTTCGGGAA TACCTGACTT ACCTGAATCA GCTGGGAACC TTGCTGGGAG
5 AT NEG CAGAAGATCT ATGCCCAGAT CTTTCGGGAA TACCTGACTT ACCTGAATCA GCTGGGAACC TTGCTGGGAG
6 GT POS CAGAAGATCT ATGCCCAGAT CTTTCGGGAA TACCTGACTT ACCTGAATCA GCTGGGAACC TTGCTGGAAG
7 GT NEG CAGAAGATCT ATGCCCAGAT CTTTCGGGAA TACCTGACTT ACCTGAATCA CCTGGGAACC TTGCTGGGAG
8 TC POS CAGAAGATCT ATGCCCAGAT CTTTCGGGAA TACCTGACTT ACCTGAATCA GCTGGGAACC TTGCTGGGAG
9 TC NEG CAGAAGATCT ATGCCCAGAT CTTTCGGGAA TACCTGACTT ACCTGAATCA GCTGGGAACC TTGCTGGGAG
920 930 940 950 960 970 980
1 KEL.SEQ GAGACCCAAG CAAGGTGCAA GAACACTCTT CCTTGTCAAT CTCCATCACT TCACGGCTGT TCCAGTTTCT
2 PM POS GAGACCCAAG CAAGGTGCAA GAACACTCTT CCTTGTCAAT CTCCATCACT TCACGGCTGT TCCAGTTTCT
3 PM NEG GAGACCCAAG CAAGGTGCAA GAACACTCTT CCTTGTCAAT CTCCATCACT TCACGGCTGT TCCAGTTTCT
4 AT POS GAGACCCAAG CAAGGTGCAA GAACACTCTT CCTTGTCAAT CTCCATCACT TCACGGCTGT TCCAGTTTCT
5 AT NEG GAGACCCAAG CAAGGTGCAA GAACACTCTT CCTTGTCAAT CTCCATCACT TCACGGCTGT TCCAGTTTCT
6 GT POS GAGACCCAAG CAAGGTGCAA GAACACTCTT CCTTGTCAAT CTCCATCACT TCACGGCTGT TCCAGTTTCT
7 GT NEG GAGACCCAAG CAAGGTGCAA GAACACTCTT CCTTGTCAAT CTCCATCACT TCACGGCTGT TCCAGTTTCT
8 TC POS GAGACCCAAG CAAGGTGCAA GAACACTCTT CCTTGTCAAT CTCCATCACT TCACGGCTGT TCCAGTTTCT
9 TC NEG GAGACCCAAG CAAGGTGCAA GAACACTCTT CCTTGTCAAT CTCCATCACT TCACGGCTGT TCCAGTTTCT
990 1000 1010 1020 1030 1040 1050
1 KEL.SEQ GAGGCCCCTG GAGCAGCGGC GGGCACAGGG CAAGCTCTTC CAGATGGTCA CTATCGACCA GCTCAAGGAA
2 PM POS GAGGCCCCTG GAGCAGCGGC GGGCACAGGG CAAGCTCTTC CAGATGGTCA CTATCGACCA GCTCAAGGAA
3 PM NEG GAGGCCCCTG GAGCAGCGGC GGGCACAGGG CAAGCTCTTC CAGATGGTCA CTATCGACCA GCTCAAGGAA
4 AT POS GAGGCCCCTG GAGCAGCGGA GGGCACAGGG CAAGCTCTTC CAGATGGTCA CTATCGACCA GCTCAAGGAA
5 AT NEG GAGGCCCCTG GAGCAGCGGC GGGCACAGGG CAAGCTCTTC CAGATGGTCA CTATCGACCA GCTCAAGGAA
6 GT POS GAGGCCCCTG GAGCAGCGGA GGGCACAGGG CAAGCTCTTC CAGATGGTCA CTATCGACCA GCTCAAGGAA
7 GT NEG GAGGCCCCTG GAGCAGCGGC GGGCACAGGG CAAGCTCTTC CAGATGGTCA CTATCGACCA GCTCAAGGAA
8 TC POS GAGGCCCCTG GAGCAGCGGC GGGCACAGGG CAAGCTCTTC CAGATGGTCA CTATCGACCA GCTCAAGGAA
9 TC NEG GAGGCCCCTG GAGCAGCGGC GGGCACAGGG CAAGCTCTTC CAGATGGTCA CTATCGACCA GCTCAAGGAA
1060 1070 1080 1090 1100 1110 1120
1 KEL.SEQ ATGGCCCCCG CCATCGACTG GTTGTCCTGC TTGCAAGCGA CATTCACACC GATGTCCCTG AGCCCTTCTC
2 PM POS ATGGCCCCCG CCATCGACTG GTTGTCCTGC TTGCAAGCGA CATTCACACC GATGTCCCTG AGCCCTTCTC
3 PM NEG ATGGCCCCCG CCATCGACTG GTTGTCCTGC TTGCAAGCGA CATTCACACC GATGTCCCTG AGCCCTTCTC
4 AT POS ATGGCCCCCG CCATCGACTG GTTGTCCTGC TTGCAAGCGA CATTCACACC GATGTCCCTG AGCCCTTCTC
5 AT NEG ATGGCCCCCG CCATCGACTG GTTGTCCTGC TTGCAAGCGA CATTCACACC GATGTCCCTG AGCCCTTCTC
6 GT POS ATGGCCCCCG CCATCGACTG GTTGTCCTGC TTGCAAGCGA CATTCACACC GATGTCCCTG AGCCCTTCTC
7 GT NEG ATGGCCCCCG CCATCGACTG GTTGTCCTGC TTGCAAGCGA CATTCACACC GATGTCCCTG AGCCCTTCTC
8 TC POS ATGGCCCCCG CCATCGACTG GTTGTCCTGC TTGCAAGCGA CATTCACACC GATGTCCCTG AGCCCTTCTC
9 TC NEG ATGGCCCCCG CCATCGACTG GTTGTCCTGC TTGCAAGCGA CATTCACACC GATGTCCCTG AGCCCTTCTC
1130 1140 1150 1160 1170 1180 1190
1 KEL.SEQ AGTCCCTCGT GGTCCATGAC GTGGAATATT TGAAAAACAT GTCACAACTG GTGGAGGAGA TGCTGCTAAA
2 PM POS AGTCCCTCGT GGTCCATGAC GTGGAATATT TGAAAAACAT GTCACAACTG GTGGAGGAGA TGCTGCTAAA
3 PM NEG AGTCCCTCGT GGTCCATGAC GTGGAATATT TGAAAAACAT GTCACAACTG GTGGAGGAGA TGCTGCTAAA
4 AT POS AGTCCCTCGT GGTCCATGAC GTGGAATATT TGAAAAACAT GTCACAACTG GTGGAGGAGA TGCTGCTAAA
5 AT NEG AGTCCCTCGT GGTCCATGAC GTGGAATATT TGAAAAACAT GTCACAACTG GTGGAGGAGA TGCTGCTAAA
6 GT POS AGTCCCTCGT GGTCCATGAC GTGGAATATT TGAAAAACAT GTCACAACTG GTGGAGGAGA TGCTGCTAAA
7 GT NEG AGTCCCTCGT GGTCCATGAC GTGGAATATT TGAAAAACAT GTCACAACTG GTGGAGGAGA TGCTGCTAAA
8 TC POS AGTCCCTCGT GGTCCATGAC GTGGAATATT TGAAAAACAT GTCACAACTG GTGGAGGAGA TGCTGCTAAA
9 TC NEG AGTCCCTCGT GGTCCATGAC GTGGAATATT TGAAAAACAT GTCACAACTG GTGGAGGAGA TGCTGCTAAA
1200 1210 1220 1230 1240 1250 1260
1 KEL.SEQ GCAGAGGGAC TTTCTGCAGA GCCACATGAT CTTAGGGCTG GTGGTGACCC TTTCTCCAGC CCTGGACAGT
2 PM POS GCAGAGGGAC TTTCTGCAGA GCCACATGAT CTTAGGGCTG GTGGTGACCC TTTCTCCAGC CCTGGACAGT
3 PM NEG GCAGAGGGAC TTTCTGCAGA GCCACATGAT CTTAGGACTG GTGGTGACCC TTTCTCCAGC CCTGGACAGT
4 AT POS GCAGAGGGAC TTTCTGCAGA GCCACATGAT CTTAGGGCTG GTGGTGACCC TTTCTCCAGC CCTGGACAGT
5 AT NEG GCAGAGGGAC TTTCTGCAGA TCCACATGAT CTTAGGGCTG GTGGTGACCC TTTCTCCAGC CCTGGACAGT
6 GT POS GCAGAGGGAC TTTCTGCAGA GCCACATGAT CTTAGGGCTG GTGGTGACCC TTTCTCCAGC CCTGGACAGT
7 GT NEG GCAGAGGGAC TTTCTGCAGA GCCACATGAT CTTAGGGCTG GTGGTGACCC TTTCTCCAGC CCTGGACAGT
8 TC POS GCAGAGGGAC TTTCTGCAGA GCCACATGAT CTTAGGGCTG GTGGTGACCC TTTCTCCAGC CCTGGACAGT
9 TC NEG GCAGAGGGAC TTTCTGCAGA GCCACATGAT CTTAGGGCTG GTGGTGACCC TTTCTCCAGC CCTGGACAGT


1270 1280 1290 1300 1310 1320 1330
1 KEL.SEQ CAATTCCAGG AGGCACGCAG AAAGCTCAGC CAGAAACTGC GGGAACTGAC AGAGCAACCA CCCATGCCTG
2 PM POS CAATTCCAGG AGGCACGCAG AAAGCTCAGC CAGAAACTGC GGGAACTGAC AGAGCAACCA CCCATGCCTG
3 PM NEG CAATTCCAGG AGGCACGCAG AAAGCTCAGC CAGAAACTGC GGGAACTGAC AGAGCAACCA CCCATGCCTG
4 AT POS CAATTCCAGG AGGCACGCAG AAAGCTCAGC CAGAAACTGC GGGAACTGAC AGAGCAACCA CCCATGCCTG
5 AT NEG CAATTCCAGG AGGCACGCAG AAAGCTCAGC CAGAAACTGC GGGAACTGAC AGAGCAACCA CCCATGCCTG
6 GT POS CAATTCCAGG AGGCACGCAG AAAGCTCAGC CAGAAACTGC GGGAACTGAC AGAGCAACCA CCCATGCCTG
7 GT NEG CAATTCCAGG AGGCACGCAG AAAGCTCAGC CAGAAACTGC GGGAACTGAC AGAGCAACCA CCCATGCCTG
8 TC POS CAATTCCAGG AGGCACGCAG AAAGCTCAGC CAGAAACTGC GGGAACTGAC AGAGCAACCA CCCATGCCTG
9 TC NEG CAATTCCAGG AGGCACGCAG AAAGCTCAGC CAGAAACTGC GGGAGCTGAC AGAGCAACCA CCCATGCCTG
1340 1350 1360 1370 1380 1390 1400
1 KEL.SEQ CCCGCCCACG ATGGATGAAG TGCGTGGAGG AGACAGGCAC GTTCTTCGAG CCCACGCTGG CGGCTTTGTT
2 PM POS CCCGCCCACG ATGGATGAAG TGCGTGGAGG AGACAGGCAC GTTCTTCGAG CCCACGCTGG CGGCTTTGTT
3 PM NEG CCCGCCCACG ATGGATGAAG TGCGTGGAGG AGACAGGCAC GTTCTTCGAG CCCACGCTGG CGGCTTTGTT
4 AT POS CCCGCCCACG ATGGATGAAG TGCGTGGAGG AGACAGGCAC GTTCTTCGAG CCCACGCTGG CGGCTTTGTT
5 AT NEG CCCGCCCACG ATGGATGAAG TGCGTGGAGG AGACAGGCAC GTTCTTCGAG CCCACGCTGG CGGCTTTGTT
6 GT POS CCCGCCCACG ATGGATGATG TGCGTGGAGG AGACAGGCAC GTTCTTCGAG CCCACGCTGG CGGCTTTGTT
7 GT NEG CCCGCCCACG ATGGATGAAG TGCGTGGAGG AGACAGGCAC GTTCTTCGAG CCCACGCTGG CGGCTTTGTT
8 TC POS CCCGCCCACG ATGGATGAAG TGCGTGGAGG AGACAGGCAC GTTCTTCGAG CCCACGCTGG CGGCTTTGTT
9 TC NEG CCCGCCCACG ATGGATGAAG TGCGTGGAGG AGACAGGCAC GTTCTTCGAG CCCACGCTGG CGGCTTTGTT
1410 1420 1430 1440 1450 1460 1470
1 KEL.SEQ TGTTCGTGAG GCCTTTGGCC CGAGCACCCG AAGTGCTGCC ATGAAATTAT TCACTGCGAT CCGGGATGCC
2 PM POS TGTTCGTGAG GCCTTTGGCC CGAGCACCCG AAGTGCTGCC ATGAAATTAT TCACTGCGAT CCGGGATGCC
3 PM NEG TGTTCGTGAG GCCTTTGGCC CGAGCACCCG AAGTGCTGCC ATGAAATTAT TCACTGCGAG CCGGGATGCC
4 AT POS TGTTCGTGAG GCCTTTGGCC CGAGCACCCG AAGTGCTGCC ATGAAATTAT TCACTGCGAT CCGGGATGCC
5 AT NEG TGTTCGTGAG GCCTTTGGCC CGAGCACCCG AAGTGCTGCC ATGAAATTAT TCACTGCGAT CCGGGATGCC
6 GT POS TGTTCGTGAG GCCTTTGGCC CGAGCACCCG AAGTGCTGCC ATGAAATTAT TCACTGCGAT CCGGGATGCC
7 GT NEG TGTTCGTGAG GCCTTTGGCC CGAGCACCCG AAGTGCTGCC ATGAAATTAT TCACTGCGAT CCGGGATGCC
8 TC POS TGTTCGTGAG GCCTTTGGCC CGAGCACCCG AAGTGCTGCC ATGAAATTAT TCACTGCGAT CCGGGATGCC
9 TC NEG TGTTCGTGAG GCCTTTGGCC CGAGCACCCG AAGTGCTGCC ATGAAATTAT TCACTGCGAT CCGGGATGCC
1480 1490 1500 1510 1520 1530 1540
1 KEL.SEQ CTCATCACTC GCCTCAGAAA CCTTCCCTGG ATGAATGAGG AGACCCAGAA CATGGCCCAG GACAAGGTTG
2 PM POS CTCATCACTC GCCTCAGAAA CCTTCCCTGG ATGAATGAGG AGACCCAGAA CATGGCCCAG GACAAGGTTG
3 PM NEG CTCATCACTC GCCTCAGAAA CCTTCCCTGG ATGAATGAGG AGACCCAGAA CATGGCCCAG GACAAGGTTG
4 AT POS CTCATCACTC GCCTCAGAAA CCTTCCCTGG ATGAATGAGG AGACCCAGAA CATGGCCCAG GACAAGGTTG
5 AT NEG CTCATCACTC GCCTCAGAAA CCTTCCCTGG ATGAATGAGG AGACCCAGAA CATGGCCCAG GACAAGGTTG
6 GT POS CTCATCACTC GCCTCAGAAA CCTTCCCTGG ATGAATGAGG AGACCCAGAA CATGGCCCAG GACAAGGTTG
7 GT NEG CTCATCACTC GCCTCAGAAA CCTTCCCTGG ATGAATGAGG AGACCCAGAA CATGGCCCAG GACAAGGTTG
8 TC POS CTCATCACTC GCCTCAGAAA CCTTCCCTGG ATGAATGAGG AGACCCAGAA CATGGCCCAG GACAAGGTTG
9 TC NEG CTCATCACTC GCCTCAGAAA CCTTCCCTGG ATGAATGAGG AGACCCAGAA CATGGCCCAG GACAAGGTTG
1550 1560 1570 1580 1590 1600 1610
1 KEL.SEQ CTCAACTGCA GGTGGAGATG GGGGCTTCAG AATGGGCCCT GAAGCCAGAG CTGGCCCGAC AAGAATACAA
2 PM POS CTCAACTGCA GGTGGAGATG GGGGCTTCAG AATGGGCCCT GAAGCCAGAG CTGGCCCGAC AAGAATACAA
3 PM NEG CTCAACTGCA GGTGGAGATG GGGGCTTCAG AATGGGCCCT GAAGCCAGAG CTGGCCCGAC AAGAATACAA
4 AT POS CTCAACTGCA GGTGGAGATG GGGGCTTCAG AATGGGCCCT GAAGCCAGAG CTGGCCCGAC AAGAATACAA
5 AT NEG CTCAACTCCA GGTGGAGATG GGGGCTTCAG AATGGACCCT GAAGCCAGAG CTGGCCCGAC AAGAATACAA
6 GT POS CTCAACTGCA GGTGGAGATG GGGGCTTCAG AATGGGCCCT GAAGCCAGAG CTGGCCCGAC AAGAATACAA
7 GT NEG CTCAACTGCA GGTGGAGATG GGGGCTTCAG AATGGGCCCT GAAGCCAGAG CTGGCCCGAC AAGAATACAA
8 TC POS CTCAACTGCA GGTGGAGATG GGGGCTTCAG AATGGGCCCT GAAGCCAGAG CTGGCCCGAC AAGAATACAA
9 TC NEG CTCAACTGCA GGTGGAGATG GGGGCTTCAG AATGGGCCCT GAAGCCAGAG CTGGCCCGAC AAGAATACAA
1620 1630 1640 1650 1660 1670 1680
1 KEL.SEQ CGATATACAG CTTGGATCGA GCTTCCTGCA GTCTGTCCTG AGCTGTGTCC GGTCCCTCCG AGCTAGAATT
2 PM POS CGATATACAG CTTGGATCGA GCTTCCTGCA GTCTGTCCTG AGCTGTGTCC GGTCCCTCCG AGCTAGAATT
3 PM NEG CGATATACAG CTTGGATCGA GCTTCCTGCA GTCTGTCCTG AGCTGTGTCC GGTCCCTCCG AGCTAGAATT
4 AT POS CGATATACAG CTTGGATCGA GCTTCCTGCA GTCTGTCCTG AGCTGTGTCC GGTCCCTCCG AGCTAGAATT
5 AT NEG CGATATACAG CTTGGATCGA GCTTCCTGCA GTCTGTCCTG AGCTGTGTCC GGTCCCTCCG AGCTAGAATT
6 GT POS CGATATACAG CTTGGTTCGA GCTTCCTGCA GTCTGTCCTG AGCTGTGTCC GGTCCCTCCG AGGTAGAATT
7 GT NEG CGATATACAG CTTGGATCGA GCTTCCTGCA GTCTGTCCTG AGCTGTGTCC GGTCCCTCCG AGCTAGAATT
8 TC POS CGATATACAG CTTGGATCGA GCTTCCTGCA GTCTGTCCTG AGCTGTGTCC GGTCCCTCCG AGCTAGAATT
9 TC NEG CGATATACAG CTTGGATCGA GCTTCCTGCA GTCTGTCCTG AGCTGTGTCC GGTCCCTCCG AGCTAGAATT


1690 1700 1710 1720 1730 1740 1750
1 KEL.SEQ GTCCAGAGCT TCTTGCAGCC TCACCCCCAA CACAGGTGGA AGGTGTCCCC TTGGGACGTC AATGCTTACT
2 PM POS GTCCAGAGCT TCTTGCAGCC TCACCCCCAA CACAGGTGGA AGGTGTCCCC TTGGGACGTC AATGCTTACT
3 PM NEG GTCCAGAGCT TCTTGCAGCC TCACCCCCAA CACAGGTGGA AGGTGTCCCC TTGGGACGTC AATGCTTACT
4 AT POS GTCCAGAGCT TCTTGCAGCC TCACCCCCAA CACAGGTGGA AGGTGTCCCC TTGGGACGTC AATGCTTACT
5 AT NEG GTCCAGAGCT TCTTGCAGCC TCACCCCCAA CACAGGTGGA AGGTGTCCCC TTGGGACGTC AATGCTTACT
6 GT POS GTCCAGAGCT TCTTGCAGCC TCACCCCCAA CACAGGTGGA AGGTGTCCCC TTGGGACGTC AATGCTTACT
7 GT NEG GTCCAGAGCT TCTTGCAGCC TCACCCCCAA CACAGGTGGA AGCTCTCCCC TTGGGACGTC AATGCTTACT
8 TC POS GTCCAGAGCT TCTTGCAGCC TCACCCCCAA CACAGGTGGA AGGTGTCCCC TTGGGACGTC AATGCTTACT
9 TC NEG GTCCAGAGCT TCTTGCAGCC TCACCCCCAA CACAGGTGGA AGGTGTCCCC TTGGGACGTC AATGCTTACT
1760 1770 1780 1790 1800 1810 1820
1 KEL.SEQ ATTCGGTATC TGACCATGTG GTAGTCTTTC CAGCTGGACT CCTCCAACCC CCATTCTTCC ACCCTGGCTA
2 PM POS ATTCGGTATC TGACCATGTG GTAGTCTTTC CAGCTGGACT CCTCCAACCC CCATTCTTCC ACCCTGGCTA
3 PM NEG ATTCGGTATC TGACCATGTG GTAGTCTTTC CAGCTGGACT CCTCCAACCC CCATTCTTCC ACCCTGGCTA
4 AT POS ATTCGGTATC TGACCATGTG GTAGTCTTTC CAGCTGGACT CCTCCAACCC CCATTCTTCC ACCCTGGTTA
5 AT NEG ATTCGGTATC TGACCATGTG GTAGTCTTTC CAGCTGGACT CCTCCAACCC CCATTCTTCC ACCCTGGCTA
6 GT POS ATTCGGTATC TGACCATGTG GTAGTCTTTC CAGCTGGACT CCTCCAACCC CCATTCTTCC ACCCTGGTTA
7 GT NEG ATTCGGTATC TGACCATGTG GTAGTCTTTC CAGCTGGACT CCTCCAACCC CCATTCTTCC ACCCTGGCTA
8 TC POS ATTCGGTATC TGACCATGTG GTAGTCTTTC CAGCTGGACT CCTCCAACCC CCATTCTTCC ACCCTGGCTA
9 TC NEG ATTCGGTATC TGACCATGTG GTAGTCTTTC CAGCTGGACT CCTCCAACCC CCATTCTTCC ACCCTGGCTA
1830 1840 1850 1860 1870 1880 1890
1 KEL.SEQ TCCCAGAGCC GTGAACTTTG GCGCTGCTGG CAGCATCATG GCCCACGAGC TGTTGCACAT CTTCTACCAG
2 PM POS TCCCAGAGCC GTGAACTTTG GCGCTGCTGG CAGCATCATG GCCCACGAGC TGTTGCACAT CTTCTACCAG
3 PM NEG TCCCAGAGCC GTGAACTTTG GCGCTGCTGG CAGCATCATG GCCCACGAGC TGATGCACAT CTTCTACCAG
4 AT POS TCCCAGAGCC GTGAACTTTG GCGCTGCTGG CAGCATCATG GCCCACGAGC TGTTGCACAT CTTCTACCAG
5 AT NEG TCCCAGAGCC GTGAACTTTG GCGCTGCTGG CAGCATCATG GCCCACGAGC TGTTGCACAT CTTCTACCAG
6 GT POS TCCCAGAGCC GTGAACTTTG GCGCTGCTGG CAGCATCATG GCCCACGAGC TGTTGCACAT CTTCTACCAG
7 GT NEG TCCCAGAGCC GTGAACTTTG GCGCTGCTGG CAGCATCATG GCCCACGAGC TGTTGCACAT CTTCTACCAG
8 TC POS TCCCAGAGCC GTGAACTTTG GCGCTGCTGG CAGCATCATG GCCCACGAGC TGTTGCACAT CTTCTACCAG
9 TC NEG TCCCAGAGCC GTGAACTTTG GCGCTGCTGG CAGCATCATG GCCCACGAGC TGTTGCACAT CTTCTACCAG
1900 1910 1920 1930 1940 1950 1960
1 KEL.SEQ CTCTTACTGC CTGGGGGCTG CCTCGCCTGT GACAACCATG CCCTCCAGGA AGCTCACCTG TGCCTGAAGC
2 PM POS CTCTTACTGC CTGGGGGCTG CCTCGCCTGT GACAACCATG CCCTCCAGGA AGCTCACCTG TGCCTGAAGC
3 PM NEG CTCTTACTGC CTGGGGGCTG CCTCGCCTGT GACAACCATG CCCTCCAGGA AGCTCACCTG TGCCTGAAGC
4 AT POS CTCTTACTGC CTGGGGGCTG CCTCGCCTGT GACAACCATG CCCTCCAGGA AGCTCACCTG TGCCTGAAGC
5 AT NEG CTCTTACTGC CTGGGGGCTG CCTCGCCTGT GACAACCATG CCCTCCAGGA AGCTCACCTG TGCCTGAAGC
6 GT POS CTCTTACTGC CTGGGGGCTG CCTCGCCTGT GACAACCATG CCCTCCAGGA AGCTCACCTG TGCCTGAAGC
7 GT NEG CTCTTACTGC CTGGGGGCTG CCTCGCCTGT GACAACCATG CCCTCCAGGA AGCTCACCTG TGCCTGAAGC
8 TC POS CTCTTACTGC CTGGGGGCTG CCTCGCCTGT GACAACCATG CCCTCCAGGA AGCTCACCTG TGCCTGAAGC
9 TC NEG CTCTTACTGC CTGGGGGCTG CCTCGCCTGT GACAACCATG CCCTCCAGGA AGCTCACCTG TGCCTGAAGC
1970 1980 1990 2000 2010 2020 2030
1 KEL.SEQ GCCATTATGC TGCCTTTCCA TTACCTAGCA GAACCTCCTT CAATGACTCC CTCACATTCT TAGAGAATGC
2 PM POS GCCATTATGC TGCCTTTCCA TTACCTAGCA GAACCTCCTA CAATGACTCC CTCACATTCT TAGAGAATGC
3 PM NEG GCCATTATGC TGCCTTTCCA TTACCTAGCA GAACCTCCTT CAATGACTCC CTCACATTCT TAGAGAATGC
4 AT POS GCCATTATGC TGCCTTTCCA TTACCTAGCA GAACCTCCTT CAATGACTCC CTCACATTCT TAGAGAATGC
5 AT NEG GCCATTATGC TGCCTTTCCA TTACCTAGCA GAACCTCCTT CAATGACTCC CTCACATTCT TAGAGAATGC
6 GT POS GCCATTATGC TGCCTTTCCA TTACCTAGCA GAACCTCCTT CAATGACTCC CTCACATTCT TAGAGAATGC
7 GT NEG GCCATTATGC TGCCTTTCCA TTACCTAGCA GAACCTCCTT CAATGACTCC CTCACATTCT TAGAGAATGC
8 TC POS GCCATTATGC TGCCTTTCCA TTACCTAGCA GAACCTCCTT CAATGACTCC CTCACATTCT TAGAGAATGC
9 TC NEG GCCATTATGC TGCCTTTCCA TTACCTAGCA GAACCTCCTT CAATGACTCC CTCACATTCT TAGAGAATGC
2040 2050 2060 2070 2080 2090 2100
1 KEL.SEQ TGCAGACGTT GGGGGGCTAG CCATCGCGCT GCAGGCATAC AGCAAGAGGC TGTTACGGCA CCATGGGGAG
2 PM POS TGCAGACGTT GGGGGGCTAG CCATCGCGCT GCAGGCATAC AGCAAGAGGC TGTTACGGCA CCATGGGGAG
3 PM NEG TGCAGACGTT GGGGGGCTAG CCATCGCGCT GCAGGCATAC AGCAAGAGGC TGTTACGGCA CCATGGGGAG
4 AT POS TGCAGACGTT GGGGGGCTAG CCATCGCGCT GCAGGCATAC AGCAAGAGGC TGTTACGGCA CCATGGGGAG
5 AT NEG TGCAGACGTT GGGGGGCTAG CCATCGCGCT GCAGGCATAC AGCAAGAGGC TGTTACGGCA CCATGGGGAG
6 GT POS TGCAGACGTT GGGGGGCTAG CCATCGCGCT GCAGGCATAC AGCAAGAGGC TGTTACGGCA CCATGGGGAG
7 GT NEG TGCAGACGTT GGGGGGCTAG CCATCGCGCT GCAGGCATAC AGCAAGAGGC TGTTACGGCA CCATGGGGAG
8 TC POS TGCAGACGTT GGGGGGCTAG CCATCGCGCT GCAGGCATAC AGCAAGAGGC TGTTACGGCA CCATGGGGAG
9 TC NEG TGCAGACGTT GGGGGGCTAG CCATCGCGCT GCAGGCATAC AGCAAGAGGC TGTTACGGCA CCATGGGGAG


2110 2120 2130 2140 2150 2160 2170
1 KEL.SEQ ACTGTCCTGC CCAGCCTGGA CCTCAGCCCC CAGCAGATCT TCTTTCGAAG CTATGCCCAG GTGATGTGTA
2 PM POS ACTGTCCTGC CCAGCCTGGA CCTCAGCCCC CAGCAGATCT TCTTTCGAAG CTATGCCCAG GTGATGTGTA
3 PM NEG ACTGTCCTGC CCAGCCTGGA CCTCAGCCCC CAGCAGATCT TCTTTCGAAG CTATGCCCAG GTGATGTGTA
4 AT POS ATTGTCCTGC CCAGCCTGGA CCTCAGCCCC CAGCAGATCT TCTTTCGAAG CTATGCCCAG GTGATGTGTA
5 AT NEG ACTGTCCTGC CCAGCCTGGA CCTCAGCCCC CAGCAGATCT TCTTTCGAAG CTATGCCCAG GTGATGTGTA
6 GT POS ACTGTCCTGC CCAGCCTGGA CCTCAGCCCC CAGCAGATCT TCTTTCGAAG CTATGCCCAG GTGATGTGTA
7 GT NEG ACTGTCCTGC CCAGCCTGGA CCTCAGCCCC CAGCAGATCT TCTTTCGAAG CTATGCCCAG GTGATGTGTA
8 TC POS ACTGTCCTGC CCAGCCTGGA CCTCAGCCCC CAGCAGATCT TCTTTCGAAG CTATGCCCAG GTGATGTGTA
9 TC NEG ACTGTCCTGC CCAGCCTGGA CCTCAGCCCC CAGCAGATCT TCTTTCGAAG CTATGCCCAG GTGATGTGTA
2180 2190 2200 2210 2220 2230 2240
1 KEL.SEQ GGAAGCCCAG CCCCCAGGAC TCTCACGACA CTCACAGCCC TCCACACCTC CGAGTCCACG GGCCCCTCAG
2 PM POS GGAAGCCCAG CCCCCAGGAC TCTCACGACA CTCACAGCCC TCCACACCTC CGAGTCCACG GGCCCCTCAG
3 PM NEG GGAAGCCCAG CCCCCAGGAC TCTCACGACA CTCACAGCCC TCCACACCTC CGAGTCCACG GGCCCCTCAG
4 AT POS GGAAGCCCAG CCCCCAGGAC TCTCACGACA CTCACAGCCC TCCACACCTC CGAGTCCACG GGCCCCTCAG
5 AT NEG GGAAGCCCAG CCCCCAGGAC TCTCACGACA CTCACAGCCC TCCACACCTC CGAGTCCACG GGCCCCTCAG
6 GT POS GGAAGCCCAG CCCCCAGGAC TCTCACGACA CTCACAGCCC TCCACACCTC CGAGTCCACG GGCCCCTCAG
7 GT NEG GGAAGCCCAG CCCCCAGGAC TCTCACGACA CTCACAGCCC TCCACACCTC CGAGTCCACG GGCCCCTCAG
8 TC POS GGAAGCCCAG CCCCCAGGAC TCTCACGACA CTCACAGCCC TCCACACCTC CGAGTCCACG GGCCCCTCAG
9 TC NEG GGAAGCCCAG CCCCCAGGAC TCTCACGACA CTCACAGCCC TCCACACCTC CGAGTCCACG GGCCCCTCAG
2250 2260 2270 2280 2290 2300 2310
1 KEL.SEQ CAGCACCCCA GCCTTTGCCA GGTATTTCCG CTGTGCACGT GGTGCTCTCT TGAACCCCTC CAGCCGCTGC
2 PM POS CAGCACCCCA GCCTTTGCCA GGTATTTCCG CTGTGCACGT GGTGCTCTCT TGAACCCCTC CAGCCGCTGC
3 PM NEG CAGCACCCCA GCCTTTGCCA GGTATTTCCG CTGTGCACGT GGTGCTCTCT TGAACCCCTC CAGCCGCTGC
4 AT POS CAGCACCCCA GCCTTTGCCA GGTATTTCCG CTGTGCACGT GGTGCTCTCT TGAACCCCTC CAGCCGCTGC
5 AT NEG CAGCACCCCA GCCTTTGCCA GGTATTTCCG CTGTGCACGT GGTGCTCTCT TGAACCCCTC CAGCCGCTGC
6 GT POS CAGCACCCCA GCCTTTGCCA GGTATTTCCG CTGTGCACGT GGTGCTCTCT TGAACCCCTC CAGCCGCTGC
7 GT NEG CAGCACCCCA GCCTTTGCCA GGTATTTCCG CTGTGCACGT GGTGCTCTCT TGAACCCCTC CAGCCGCTGC
8 TC POS CAGCACCCCA GTCTTTGCCA GGTATTTCCG CTGTGCACGT GGTGCTCTCT TGAACCCCTC CAGCCGCTGT
9 TC NEG CAGCACCCCA GCCTTTGCCA GGTATTTCCG CTGTGCACGT GGTGCTCTCT TGAACCCCTC CAGCCGCTGC
2320 2330 2340 2350 2360 2370 2380
1 KEL.SEQ CAGCTCTGGT AACTTGGTTA CCAAAGATGC CACAGCACAG AAATATCGAC CAACACCTCC CTGGTCACAT
2 PM POS CAGCTCTGGT AACTTGGTTA CCAAAGATGC CACAGCACAG AAAT
3 PM NEG CAGCTCTGGT AA
4 AT POS CAGCTCTGGT AACTTGGTTA CCAAAGATGC CACAGCACAG AAAT
5 AT NEG CAGCTCTGGT AA
6 GT POS CAGCTCTGGT AACTTGGTTA CCAAAGATGC CACAGCACAG AAAT
7 GT NEG CAGCTCTGGT AACTT
8 TC POS CAGCTCTGGT AACTTGGTTA CCAAAGATGC CACAGCACAG A
9 TC NEG CAGCTCTGGT AA
2390 2400 2410 2420 2430 2440 2450
1 KEL.SEQ CCATGGAATC AGAGCAAGAT TTCCTTTCTG CTTCTGTTCC AAAAATAAAA GCTGGCACTT GGCTTCCGCC
2 PM POS <==
3 PM NEG <==
4 AT POS <==
5 AT NEG <==
6 GT POS <==
7 GT NEG <==
8 TC POS <==
9 TC NEG <==
2460 2470 2480 2490 2500 2510 2520
1 KEL.SEQ CGGAATTC
2 PM POS <==
3 PM NEG <==
4 AT POS <==
5 AT NEG <==
6 GT POS <==
7 GT NEG <==
8 TC POS <==
9 TC NEG <==

SECTION 11

References

REFERENCES

1) Allen, F. H. and Lewis, S. J. (1957)
Vox Sang. 2, 81-87.
Kp^a (Penney), a new antigen in the Kell blood group system.

2) Allen, F. H. , Lewis, S. J. and Fudenberg, H. (1958)
Vox Sang. 3, 1-13.
Studies of anti-Kp^b, a new antibody in the Kell blood group system.

3) Allen, F. H., Krabbe, S. M. R. and Corcoran, P. A. (1961)
Vox Sang. 6, 555-560.
A new phenotype (M^cLeod) in the Kell blood group system.

4) Allen, J. R., Mitas, M., Malone, L., Brunner-Jackson, B., Wimalasena, D. S. and Gargus, J. J. (1988 a)
Am. J. Hum. Genet. 43, A1
Inborn error in human urea transport: analysis of the cDNA encoding the putative transporter.

5) Advani, H., Zamor, J., Judd, W. J., Johnson, C. L. and Marsh, W. L. (1982)
British Journal of Haematology 51, 107-115.
Inactivation of Kell blood group antigens by 2-aminoethylisothiouronium bromide.

6) Anstee, D. J., Mawby, W. J. and Tanner, M. J. (1982) in
Membranes and Transport: a critical review
Plenum Press, New York.

7) Anstee, D. J. (1990)
Vox Sang. 58, 1-20.
Blood group-active surface molecules of the human red blood cell.

8) Anstee, D. J. and Mallinson, G. (1994)
Vox Sang. 67, Suppl. 3, 1-6.
The biochemistry of blood group antigens - some recent advances.

9) Anstee, D. J. (1995)
Transfusion Medicine 5, 1-13.
Blood group antigens defined by the amino acid sequence of red cell surface proteins.

10) Atkinson, T. and Smith, M. (1984)
Oligonucleotides synthesis, a practical approach.
IRL Press.

11) Avent, N. D., Ridgwell, K., Tanner, M. J. A. and Anstee, D. J. (1990)
Biochem. J. 271, 821-825.
Cloning of a 30 kDa erythrocyte membrane protein associated with the Rh (Rhesus) blood group antigen expression.

12) Beck, M. L., Marsh, W. L., Pierce, S. R., DiNapoli, J., Oyen, R. and Nichols, M. E. (1979)
Transfusion 19, 197-202.
Auto-anti-Kp^b associated with weakened antigenicity in the Kell blood group system: a second example.

13) Belt, K. T., Carroll, M. C. and Porter, R. R. (1984)
Cell 36, 907-914.
The structural basis of the multiple forms of human complement component C4.

14) Bennet, P. R., Le Van Kim, C., Colin, Y., Warwick, R. M., Cherif-Zahar, B., Fisk, N. M. and Cartron, J-P. (1993)
New England Journal of Medicine 329, 607-610.
Prenatal determination of foetal RhD type by DNA amplification.

15) Bennett, P. R., Overton, T. G., Lighten, A. D. and Fisk, N. M. (1995)
The Lancet 345, 661-662.
Rhesus D typing.

16) Bertelson, C. J., Pogo, A. O., Chaudhuri, A., Marsh, W. L., Redman, C. M., Banjeree, D., Symmans, W. A., Simon, T., Frey, D. and Kunkel, L. M. (1988)
Am. J. Hum. Genet. 42,703-711.
Localization of the McLeod locus (XK) within XP21 by deletion analysis.

17) Birnboim, H. C. and Doly, J. (1979)
Nucleic Acids Research 7,1513-1523.
A rapid alkaline extraction procedure for screening recombinant plasmid DNA.

18) Blanchard, D., Dahr, W., Beyreuther, K., Moulds, J. and Cartron, J-P. (1987 a)
Eur. J. Biochem. 167,361-366.
Hybrid glycophorins from human erythrocyte membranes.

19) Blanchard, D., Dahr, W., Hummel, M., Latron, F., Beyreuther, K. and Cartron, J-P. (1987 b)
The Journal of Biological Chemistry 262,5808-5811.
Glycophorins B and C from human erythrocyte membranes.

20) Blanchard, D., Bloy, C., Hermand, P., Cartron, J-P., Saboori, A. M., Smith, B. L. and Agre, P. (1988)
Blood 72,1424-1427.
Two-dimensional iodopeptide mapping demonstrates that erythrocyte Rh D, c, and E polypeptides are structurally homologous but nonidentical.

21) Blondel, A. and Thillet, J. (1991)
Nucleic Acids Research 19,181.
A fast and convenient way to produce single stranded DNA from a phagemid.

22) Bloy, C., Blanchard, D., Dahr, W., Beyreuther, K., Salmon, C. and Cartron, J-P. (1988)
Blood 72,661-666.
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.

23) Blumenfield, O. O., Smith, A. J. and Moulds, J. J. (1987)
The Journal of Biological Chemistry 262,11864-11870.
Membrane glycophorins of Dantu blood group erythrocytes.

24) Botstein, D., White, R. L., Skolnick, M. and Davis, R. W. (1980)
Am. J. Hum. Genet. 32,314-331
Construction of a genetic linkage map in man using restriction fragment length polymorphisms.

25) Branch, D. R. and Petz, L. D. (1982)
Am. J. Clin. Pathol. 78,161-167.
A new reagent (ZZAP) having multiple applications in immunohematology.

26) Branch, D. R., Muensch, H. A., Hian, A. L. S. S. and Petz, L. D. (1983)
British Journal of Haematology 54,573-578.
Disulfide bonds are a requirement for Kell and Cartwright (Yt^a) blood group antigen integrity.

27) Brendel, W. L., Issitt, P. D., Moore, R. E., Lenes, B. A., Zeller, D. J., Tse, T. P. and Atkin, N. (1985)
Biotest Bulletin 2,201-206.
Temporary reduction of red cell Kell system antigen expression and transient production of anti-Kp^b in a surgical patient.

28) Brocteur. J. from Issitt. P. D. and Issitt C. H. (1981)
Applied blood group serology (second edition)
Biological Corporation of America.

29) Cariello, N. F. and Skopek, T. R. (1993)
Mutation Research 288, 103-112.
Mutational analysis using denaturing gradient gel electrophoresis and PCR.

30) Carritt, B., Blunt, T., Avent, N., Daniels, G. and Steers, F. (1993)
Ann. Hum. Genet. 57, 273-279.
Rh null phenotypes are not due to a gross deletion and can occur on different Rh genetic backgrounds.

31) Cartron, J-P. and Agre, P. (1993)
Seminars in Hematology 30, 193-208.
Rh blood group antigens : protein and gene structure.

32) Cartron, J-P., Le Van Kim, C., Cherif-Zahar, B., Mouro, I., Rouillac, C. and Colin, Y. (1994)
Biochem. J. 306, 877-878.
The two gene model of the RH blood-group locus.

33) Ceppellini, R., Dunn, L. C. and Turri, M. (1955)
Proc. Natl. Acad. Sci. USA 41, 283-288.
An interaction between alleles at the Rh locus in man which weakens the reactivity of the Rh_o factor (D^u).

34) Chaudhuri, A., Polyakova, J., Zbrzezna, V., Williams, K., Gulati, S. and Pogo, O. (1993)
Proc. Natl. Acad. Sci. USA 90, 10793-10797.
Cloning of glycoprotein D cDNA, which encodes the major subunit of the Duffy blood group system and the receptor for the *Plasmodium vivax* malaria parasite.

35) Cherif-Zahar, B., Bloy, C., Le Van Kim, C., Blanchard, D., Bailly, P., Hermand, P., Salmon, C., Cartron, J-P. and Colin, Y. (1990)
Proc. Natl. Acad. Sci. USA 87, 6243-6247.
Molecular cloning and protein structure of a human blood group Rh polypeptide.

36) Cherif-Zahar, B., Mattei, M. G., Le Van Kim, C., Bailly, P., Cartron, J-P. and Colin, Y. (1991)
Hum. Genet. 86,398-400.
Localization of the human Rh blood group gene structure to chromosome region 1p34.3-1p36.1 by in situ hybridization.

37) Cherif-Zahar, B., Raynal, V., Le Van Kim, C., D'Ambrosio, A-M., Bailly, P., Cartron, J-P. and Colin, Y. (1993)
Blood 82,656-662.
Structure and expression of the RH locus in the Rhesus deficiency syndrome.

38) Cherif-Zahar, B., Le Van Kim, C., Rouillac, C., Raynal, V., Cartron, J-P. and Colin, Y. (1994 a)
Genomics 19, 68-74.
Organization of the gene (RHCE) encoding the human blood group Rh CcEe antigens and characterization of the promoter region.

39) Cherif-Zahar, B., Raynal, V., D'Ambrosio, A.-M., Cartron, J-P. and Colin, Y. (1994 b)
Blood 84,4354-4360.
Molecular analysis of the structure and expression of the RH locus in individuals with D--, Dc-, and DC^{w}- gene complexes.

40) Chomczynski, P. and Sacchi, N. (1987)
Analytical Biochemistry 162,156-159.
Single-step method of RNA isolation by acid guanidinium thiocyanate-phenol-chloroform extraction.

41) Chown, B (1964)
13th John G. Gibson II Lecture.
Columbia Presbyterian Medical Centre, New York.

42) Clausen, H. and Hakomori, S. (1989)
Vox Sang. 56, 1-20.
ABH and related histo-blood group antigens: Immunochemical differences in carrier isotypes and their distribution.

43) Clausen, H., White, T., Takio, K., Titani, K., Stroud, M., Holmes, E., Karkov, J., Thim, L. and Hakomori, S-I. (1990)
J. Biol. Chem. 265,1139-1145.
Isolation to homogeneity and partial characterization of a histo-blood group A defined Fucα1->2Galα1->3-N-acetylgalactosaminyltransferase from human lung tissue.

44) Colin, Y., Rahuel, C., London, J., Romeo, P-H., d'Auriol, L., Galibert, F. and Cartron, J-P. (1986)
The Journal of Biological Chemistry 261,229-233.
Isolation of cDNA clones and complete amino acid sequence of human erythrocyte Glycophorin C.

45) Colin, Y., Le Van Kim, C., Tsapis, A., Clerget, M., d'Auriol, L., London, J., Galibert, F. and Cartron, J-P. (1989)
The Journal of Biological Chemistry 264,3773-3780.
Human erythrocyte glycophorin C.

46) Colin, Y., Cherif-Zahar, B., Le Van Kim, C., Raynal, V., Van Huffel, V. and Cartron, J-P. (1991)
Blood 78,2747-2752.
Genetic basis of the RhD-positive and RhD-negative blood group polymorphism as determined by Southern analysis.

47) Colin, Y., Cherif-Zahar, B., Le Van Kim, C., Mouro, I. and Cartron, J-P. (1993)
J. Med. Primatol. 22, 36-43.
Recent advances in molecular and genetic analysis of Rh blood group structures.

48) Colin, Y., Bailly, P. and Cartron, J. P. (1994)
Vox Sang. 67, Suppl. 3, 67-72.
Molecular genetic basis of Rh and LW blood groups.

49) Coombs, R. R., Mourant, A. E. and Race, R. R. (1946)
The Lancet 1,264-266.
In-vivo isosensitisation of red cells in babies with haemolytic disease.

50) Corcoran, P. A., Allen, F. H., Lewis, M. and Chown, B. (1961)
Transfusion 1,181-183.
A new antibody, anti-Ku (anti-Peltz), in the Kell blood group system.

51) Cox, K. H., DeLeon, D. V., Angerer, L. M. and Angerer, R. C. (1984)
Developmental Biology 101,485-502.
Detection of mRNAs in sea urchin embryos by *in situ* hybridization using asymmetric RNA probes.

52) D'Adamio, L., Shipp, M. A., Masteller, E. L. and Reinherz, E. L. (1989)
Proc. Natl. Acad. Sci. USA 86,7103-7107.
Organization of the gene encoding common accute lymphoblastic leukemia antigen (neutral endopeptidase 24.11): multiple mini-exons and separate 5' untranslated regions.

53) Dahr, W., Beyreuther, K. and Moulds, J. J. (1987)
Eur. J. Biochem. 166, 27-30.
Structural analysis of the major human erythrocyte membrane sialoglycoprotein from Miltenberger class VII cells.

54) Dahr, W. (1992)
Vox Sang. 62,129-135.
Miltenberger subsystem of the MNSs blood group system.

55) Dunning, A. M., Talmud, P. and Humphries, S. E. (1988)
Nucleic Acids Research 16,10393.
Errors in the polymerase chain reaction.

56) Eicher, C., Kirkley, K., Porter, M. and Kao, Y. (1985)
Transfusion 25,448.
A new low frequency antigen in the Kell system: K24 (Cls).

57) Engelmann, B. and Schumacher, U. (1993)
Comp. Biochem. Physiol. 105A,197-203.
The emerging role of ABH blood group antigens as modulators of cell membrane function.

58) Erlich, H. A. (1989)
PCR technology- principles and applications for DNA amplification.
Stockton Press, New York, USA.

59) Erlich, H. A., Gelfand, D. and Sminsky, J. J. (1991)
Science 252,1643-1651.
Recent advances in the polymerase chain reaction.

60) Ernst, L. K., Rajan, V. P., Larsen, R. D., Ruff, M. M. and Lowe, J. B. (1989)
The Journal of Biological Chemistry 264,3436-3447.
Stable expression of blood group H determinants and GDP-L-fucose: β-D-galactoside 2-α-L-fucosyltransferase in mouse cells after transfection with human DNA.

61) Fan, E., Levin, D. B., Glickman, B. W. and Logan, D. M. (1993)
Mutation Research 288, 85-92.
Limitations in the use of SSCP analysis.

62) Fisk, N. M., Bennett, P., Warwick, R. M., Letsky, E. A., Welch, R., Vaughan, J. I. and Moore, G. (1994)
American Journal of Obstetrics and Gynecology 171, 50-54.
Clinical utility of fetal RhD typing in alloimmunized pregnancies by means of polymerase chain reaction on amniocytes or chorionic villi.

63) Ford, D. S., Knight, A. E. and Smith, F. (1977)
Vox Sang. 32,220-223.
A further example of Kp^a/K^o exhibiting depression of some Kell group antigens.

64) Franke, U., Ochs, H. D., de Martinville, B., Giacalone, J., Lindgren, V., Disteche, C., Pagon, R. A., Hofker, M. H., van Ommen, G-J. B., Pearson, P. L. and Wedgwood, R. J. (1985)
Am. J. Hum. Genet. 37,250-267.
Minor Xp21 chromosome deletion in a male associated with expression of Duchenne muscular dystrophy, chronic granulomatous disease, retinitis pigmentosa and M^cLeod syndrome.

65) Furuhjelm, U., Nevanlinna, H. R., Nurkka, R., Gavin, J. and Sanger, R. (1969)
Vox Sang. 16,496-499.
Evidence that the antigen Ul^a is controlled from the Kell complex locus.

66) Gahmberg, C. G. (1982)
FEBS Letters 140, 93-97.
Molecular identification of the human Rh_O(D) antigen.

67) Gahmberg, C. G. (1983)
The EMBO Journal 2,223-227.
Molecular characterization of the human red cell Rh_O(D) antigen.

68) Ganguly, A., Rock, M. J. and Prockop, D. J. (1993)
Proc. Natl. Acad. Sci. USA 90,10325-10329.
Conformation-sensitive gel electrophoresis for rapid detection of single-base differences in double-stranded PCR products and DNA fragments: evidence for solvent-induced bends in DNA heteroduplexes.

69) Green, F. A. (1967)
Immunochemistry 4,247-257.
Erythrocyte membrane sulfhydryl groups and Rh antigen activity.

70) Green, F. A. (1983)
Molecular Immunology 20,769-775.
The mode of attenuation of erythrocye membrane Rh_O(D) antigen activity by 5,5'-dithiobis-(2-nitrobenzoic acid) and protection against loss of activity by bound anti-Rh_O(D) antibody.

71) Hardy, J. and Napier, J. A. F. (1981)
British Journal of Obstetrics and Gynaecology 88,91-100.
Red cell antibodies detected in antenatal tests on Rhesus positive women in South and mid Wales, 1948-1978.

72) Hayashi, K. (1992)
G.A.T.A. 9, 73-79.
PCR-SSCP: a method for detection of mutations.

73) Herrmann, B. G. and Frischauf, A-M. (1987)
Methods in Enzymology 152, 180-183.
Isolation of genomic DNA.

74) High, S. and Tanner, M. J. A. (1987)
Biochem. J. 243, 277-280.
Human erythrocyte membrane sialoglycoprotein β.

75) Ho, M. F., Monaco, A. P., Blonden, L. A. J., van Ommen, G. J. B., Affara, N. A., Ferguson-Smith, M. A. and Lehrach, H. (1992)
Am. J. Hum. Genet. 50, 317-330.
Fine mapping of the McLeod locus (XK) to a 150-380kb region in Xp21.

76) Hongyo, T., Buzard, G. S., Calvert, R. J. and Weghorst, C. M. (1993)
Nucleic Acids Research 21, 3637-3642.
'Cold SSCP' : a simple, rapid and nonradioactive method for optimized single-strand conformation polymorphism analyses.

77) Huang, C-H., Johe, K., Moulds, J. J., Siebert, P. D., Fukuda, M. and Blumenfeld, O. O. (1987)
Blood 70, 1830-1835.
δ glycophorin (glycophorin B) gene deletion in two individuals homozygous for the S-s-U- blood group phenotype.

78) Huang, C. H. and Blumenfield, O. O. (1988)
Proc. Natl. Acad.Sci. USA 85, 9640-9644.
Characterization of a genomic hybrid specifying the human erythrocyte antigen Dantu: Dantu gene is duplicated and linked to a δ glycophorin gene deletion.

79) Hughes-Jones, N. C., Gardner, B. and Lincoln, P. J. (1971)
Vox Sang. 21,210-216.
Observation of the number of available c, D, and E antigen sites on red cells.

80) Hyland, C. A., Wolter, L. C., Liew, Y.-W. and Saul, A. (1994 a)
Blood 83,566-572.
A Southern analysis of Rh blood group genes: association between restriction fragment length polymorphism patterns and Rh serotypes.

81) Hyland, C. A., Wolter, L. C. and Saul, A. (1994 b)
Blood 84,321-324.
Three unrelated RhD gene polymorphisms identified among blood donors with Rhesus CCee (r'r') phenotypes.

82) Issitt, P. D. (1985)
Applied blood group serology (3rd edition)
Montgomery Scientific Publications, Miami.

83) Jaber, A., Blanchard, D., Goossens, D., Bloy, C., Lambin, P., Rouger, P., Salmon, C. and Cartron, J. P. (1989)
Blood 73,1597-1602.
Characterization of the blood group Kell (K1) antigen with a human monoclonal antibody.

84) Jaber, A., Loirat, M-J., Willem, C., Bloy, C., Cartron, J-P. and Blanchard, D. (1991)
British Journal of Haematology 79,311-315.
Characterization of murine monoclonal antibodies directed against the Kell blood group glycoprotein.

85) Kaita, H., Lewis, M., Chown, B. and Gard, E. (1959)
Nature 183,1586.
A further example of the Kell blood group phenotype K-k-Kp(a-b-).

86) Kajii, E., Umenishi, F. and Ikemoto, S. (1993 a)
Vox Sang. 64, 196.
A cDNA clone encoding an Rh polypeptide detected in RhD-negative erythroid cells.

87) Kajii, E., Umenishi, F., Iwamoto, S. and Ikemoto, S. (1993 b)
Hum. Genet. 91, 157-162.
Isolation of a new cDNA clone encoding an Rh polypeptide associated with the Rh blood group system.

88) Kajii, E., Umenishi, F., Nakauchi, H. and Ikemoto, S. (1994)
Biochemical and Biophysical Research Communications 202, 1497-1504.
Expression of RH blood group gene transcripts in human leukocytes.

89) King, M. J., Poole, J. and Anstee, D. J. (1989)
Transfusion 29, 106-112.
An application of immunoblotting in the classification of the Miltenberger series of blood group antigens.

90) Kout, M. from Race, R. R. and Sanger, R. (1976)
Blood groups in man (6th edition)
Blackwell press. Oxford.

91) Kovar, H., Jug, G., Auer, H., Skern, T. and Blaas, D. (1991)
Nucleic Acids Research 19, 3507-3510.
Two dimensional single-strand conformation polymorphism analysis: a useful tool for the detection of mutations in long DNA fragments.

92) Landsteiner, K. (1900)
Translation from Dunsford Memorial (1972) 1, 11-17.
Blood Bank Center, U S Army Medical Research Laboratory, Fort Knox, Kentuky.
Agglutination phenomena in normal human blood.

93) Landsteiner, K. and Levine, P. (1927)
Proc. Soc. Exp. Biol. and Med. 24,941-942.
Further observations on individual differences of human blood.

94) Lee, S., Zambas, E. D., Marsh, W. L. and Redman, C. M. (1991)
Proc. Natl. Acad. Sci. USA 88,6353-6357.
Molecular cloning and primary structure of the Kell blood group protein.

95) Lee, S., Zambas, E. D., Marsh, W. L. and Redman, C. M. (1993)
Blood 81,2804-2809.
The human Kell blood group gene maps to 7q33 and its expression is restricted to erythroid cells.

96) Lee, S., Wu, X., Reid, M., Zelinski, T. and Redman, C. (1995 a)
Blood 85,912-916.
Molecular basis of the Kell (K1) phenotype.

97) Lee, S., Zambas, E., Green, E. D. and Redman, C. (1995 b)
Blood 85,1364-1370.
Organization of the gene encoding the human Kell blood group protein.

98) Leppert, M., Wasmuth, J., Overhauser, J., Dobbs, M., Nakamura, Y., Lathrop, M., O'Connell, P., Cartwright, P., Lalouel, J. M. and White, R. (1987)
Cytogenet. Cell. Genet. 46,649.
A primary genetic linkage map of chromosome 5.

99) Le Van Kim, C., Cherif-Zahar, B., Raynal, V., Mouro, I., Lopez, M., Cartron, J.-P. and Colin, Y. (1992 a)
Blood 80,1074-1078.
Multiple Rh messenger RNA isoforms are produced by alternative splicing.

100) Le Van Kim, C., Mouro, I., Cherif-Zahar, B., Raynal, V., Cherrier, C., Cartron, J.-P. and Colin, Y. (1992 b)
Proc. Natl. Acad. Sci. USA 89,10925-10929.
Molecular cloning and primary structure of the human blood group RhD polypeptide.

101) Levine, P., Newark, N. J. and Stetson, R. E. (1939)
JAMA 113,126-127.
An unusual case of intra-group agglutination.

102) Levine, P., Backer, M., Wigod, M. and Ponder, R. (1949)
Science 109,464-467.
A new human hereditary blood property (Cellano) present in 99.8% of all bloods.

103) Lo, Y-M. D., Noakes, L., Bowell, P. J., Fleming, K. A. and Wainscoat, J. S. (1994)
British Journal of Haematology 87,658-660.
Detection of fetal RhD sequence from peripheral blood of sensitized RhD-negative pregnant women.

104) Lublin, A. M., Lemons, R. S. and Le Beau, M. M. (1987)
J. Exp. Med. 165,1731-1736.
The gene encoding decay-accelerating factor (DAF) is located in the complement-regulatory locus on the long arm of chromosome 1.

105) Lutz, P. and Dzik, W.H. (1992)
Transfusion 32,467-483.
Molecular biology of red cell blood group genes.

106) McGuire, M., Smith, B. L. and Agre, P. (1988)
Blood 72,287-293.
Distinct variants of erythrocyte protein 4.1 inherited in linkage with elliptocytosis and Rh type in three white families.

107) Manny, N., Levene, C., Sela, R., Johnson, C. L., Mueller, K. A. and Marsh, W. L. (1983)
Vox Sang. 45, 252-256.
Autoimmunity in the Kell blood groups: auto-anti-Kp^b in a Kp(a+b-) patient.

108) Manny, N., Levene, C., Harel, N., Schneiderman, S., Simon, A. and Daniels, G. L. (1985)
Vox Sang. 49, 135-137.
The second sample of anti-K22 and a family genetically informative for K and K22.

109) Marsh, W. L., Jensen, L., Oyen, R., Stroup, M., Gellerman, M., McMahon, F. J. and Tsitsera, H. (1974)
Vox Sang. 26, 34-40.
Anti-K13 and the K:-13 phenotype.

110) Marsh, W. L., Oyen, R., Nichols, M. E. and Allen, F. H. (1975)
British Journal of Haematology 29, 247-262.
Chronic granulomatous disease and the Kell blood groups.

111) Marsh, W. L., Oyen, R. and Nichols, M. E. (1976)
Vox Sang. 31, 356-362.
Kx antigen, the McLeod phenotype, and chronic granulomatous disease: further studies.

112) Marsh, W. L. (1979)
Vox Sang. 36, 375-376.
'Letter to the Editor'

113) Marsh, W. L., Redman, C. M., Kessler, L. A., DiNapoli, J., Scarborough, A. L., Philipps, A. G. and Mody, K. M. (1987)
Transfusion 27, 36-40.
K23 a low incidence antigen in the Kell blood group system identified by biochemical characterization.

114) Marsh, W. L. and Redman, C. M. (1990)
Transfusion 30,158-167.
The Kell blood group system: a review.

115) Marsh, W.L. (1992)
Transfusion 32,98-101.
Molecular biology of blood groups: cloning the Kell gene.

116) Marsh, W. L. and Redman, C. M. (1993)
Immunohematology 9, 1-6.
Review: developments in the Kell blood group system.

117) Maxam, A. M. and Gilbert, W. (1977)
Proc. Natl. Acad. Sci. USA 74,560-564.
A new method for sequencing DNA.

118) Merry, A. H., Thomson, E. E., Anstee, D. J. and Stratton, F. (1984)
Immunology 51,793-800.
The quantification of erythrocyte antigen sites with monoclonal antibodies.

119) Miller, S. A., Dykes, D. D. and Polesky, H. F. (1988)
Nucleic Acids Reasearch 16,1215.
A simple salting out procedure for extracting DNA from human nucleated cells.

120) Moore, S., Woodrow, C. F. and McLelland, D. B. L. (1982)
Nature 295,529-531.
Isolation of membrane components associated with human red cell antigens Rh (D), (c), (E) and Fy^a.

121) Mouro, I., Colin, Y., Cherif-Zahar, B., Cartron, J-P. and Le Van Kim, C. (1993)
Nature Genetics 5, 62-65.
Molecular genetic basis of the human Rhesus blood group system.

122) Murphy, M. T., Morrison, N., Miles, J. S., Fraser, R. H., Spurr, N. K. and Boyd, E. (1993)
Hum. Genet. 91,585-588.
Regional chromosomal assignment of the Kell blood group locus (KEL) to chromosome 7q33-q35 by fluorescence *in situ* hybridization: evidence for the polypeptide nature of antigenic variation.

123) Nichols, M. E., Rosenfield, R. E. and Rubenstein, P. (1987)
Vox Sang. 52,231-235.
Monoclonal anti-K14 and anti-K2.

124) Nunn, H. D., Giles, C. M. and Dormandy, K. M. (1966)
Vox Sang. 11,611-619.
A second example of anti-Ku in a patient who has a rare Kell phenotype, K^{o}.

125) Olives, B., Mattei, M. G., Huet, M., Neau, P., Martial, S., Cartron, J-P. and Bailly, P. (1995)
Journal of Biological Chemistry 270,15607-15610.
Kidd blood group and urea transport functionof human erythrocytes are carried by the same protein.

126) Paabo, S. and Wilson, A. C. (1988)
Nature 334,387-388.
Polymerase chain reaction reveals cloning artefacts.

127) Parsons, S. F., Judson, P. A. and Anstee, D. J. (1982)
Journal of Immunogenetics 9,377-380.
BRIC 18: a monoclonal antibody with a specificity related to the Kell blood group system.

128) Petty, A. C., Daniels, G. L. and Tippett, P. (1994)
Vox Sang. 66,216-224.
Application of the MAIEA assay to the Kell blood group system.

129) Purohit, K. R., Weber, J. L., Ward, L. J. and Keats, B. J. B. (1992)
Human Genetics 89,457-458.
The Kell blood group locus is close to the cystic fibrosis locus on chromosome 7.

130) Race, R. R. and Sanger, R. (1975)
Blood groups in man (6th edition).
Blackwell Scientific Publications, Oxford.

131) Rahuel, C., London, J., d'Auriol, L., Mattei, M-G., Tournamille, C., Skrzynia, C., Lebouc, Y., Galibert, F. and Cartron, J-P. (1988)
Eur. J. Biochem. 172,147-153.
Characterization of cDNA clones for human glycophorin A.

132) Redman, C. M., Marsh, W. L., Mueller, K. A., Avellino, G. P. and Johnson, C. L. (1984)
Transfusion 24,176-178.
Isolation of Kell-active protein from the red cell membrane.

133) Redman, C. M., Avellino, G., Pfeffer, S. R., Mukherjee, T. K., Nichols, M., Rubenstein, P. and Marsh, W. L. (1986)
Journal of Biological Chemistry 261,9521-9525.
Kell blood group antigens are part of a 93 000-Dalton red cell membrane protein.

134) Redman, C. M., Marsh, W. L., Scarborough, A., Johnson, C. L., Rabin, B. I. and Overbeeke, M. (1988)
British Journal of Haematology 68,131-136.
Biochemical studies on McLeod phenotype red cells and isolation of Kx antigen.

135) Redman, C. M. and Marsh, W. L. (1993)
Seminars in Hematology 30,209-218.
The Kell blood group system and the McLeod phenotype.

136) Reid, M. E., Oyen, R., Redman, C. M., Gillespie, G., Jones, J. and Eckrich, R. (1995)
Vox Sang. 68, 40-45.
K12 is located on the Kell blood group protein in proximity to K/k and Js^a/Js^b.

137) Ridgwell, K., Spurr, N. K., Laguda, B., MacGeoch, C., Avent, N. D. and Tanner, M. J. A. (1992)
Biochem. J. 287, 223-228.
Isolation of cDNA clones for a 50kDa glycoprotein of the human erythrocyte membrane associated with Rh (Rhesus) blood group antigen expression.

138) Rigby, P. W. J., Dieckmann, M., Rhodes, C. and Berg, P. (1977)
J. Mol. Biol. 113, 237-251.
Labelling deoxyribonucleic acid to a high specific activity *in vitro* by nick translation with DNA polymerase I.

139) Roberts, R. G., Montandon, A. J., Bobrow, M. and Bentley, D. R. (1989)
Nucleic Acids Research 17, 5961-5971.
Detection of novel genetic markers by mismatch analysis.

140) Royer-Pokora, B., Kunkel, L. M., Monaco, A. P., Goff, S. C., Newburger, P. E., Baehner, R. L., Cole, F. S., Curnutte, J. T. and Orkin, S. H. (1986)
Nature 322, 32-38.
Cloning the gene for an inherited human disorder- chronic granulomatous disease- on the basis of its chromosomal location.

141) Russo, D. C. W., Lee, S., Reid, M. and Redman, C. M. (1994)
Blood 84, 3518-3523.
Topology of Kell blood group protein and the expression of multiple antigens by transfected cells.

142) Saiki, R. K., Scharf, S., Faloona, F., Mullis, K. B., Horn, G. T., Erlich, H. A. and Arnheim, N. (1985)
Science 230, 1350-1354.
Enzymatic amplification of β-globin genomic sequences and restriction site analysis for diagnosis of sickle cell anemia.

143) Saiki, R. K., Gelfand, D. H., Stoffel, S., Scharf, S. J., Higuchi, R., Horn, G. T., Mullis, K. B. and Erlich, H. A. (1988)
Science (January), 487-491.
Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase.

144) Sambrook, J., Fritsch, E. F. and Maniatis, T. (1989)
Molecular cloning- a laboratory manual (2nd edition)
Cold Spring Harbor Laboratory Press.

145) Sanger, F., Nicklen, S. and Coulson, A. R. (1977)
Proc. Natl. Acad. Sci. USA 74,5463-5467.
DNA sequencing with chain terminating inhibitors.

146) Savalonis, J. M., Kalish, R. I., Cummings, E. A., Ryan, R. W. and Alosi, R. (1988)
Transfusion 28,229-232.
Kell blood group activity of gram-negative bacteria.

147) Sekiya, T. (1993)
Mutation Research 288, 79-83.
Detection of mutant sequences by single-strand conformation polymorphism analysis.

148) Seyfried, H., Gorska, B., Maj, S., Sylwestrowicz, T., Giles, C. M. and Goldsmith, K. L. G. (1972)
Vox Sang. 23,528-536.
Apparent depression of antigens of the Kell blood group system associated with autoimmune acquired haemolytic anaemia.

149) Siebert, P. D. and Fukuda, M. (1986 a)
The Journal of Biological Chemistry 261,12433-12436.
Human glycophorin A and B are encoded by separate, single copy genes coordinately regulated by a tumor-promoting phorbol ester.

150) Siebert, P. D. and Fukuda, M. (1986 b)
Proc. Natl. Acad. Sci. USA 83,1665-1669.
Isolation and characterization of human glycophorin A cDNA clones by a synthetic oligonucleotide approach: nucleotide sequence and mRNA structure.

151) Siebert, P. D. and Fukuda, M. (1987)
Proc. Natl. Acad. Sci. USA 84,6735-6739.
Molecular cloning of a human glycophorin B cDNA: nucleotide seuqnce and genomic relationship to glycophorin A.

152) Simsek, S., de Jong, C. A. M., Cuijpers, H. Th. M., Bleeker, P. M. M., Westers, T. M., Overbeeke, M. A. M., Goldschmeding, R., van der Schoot, C. E. and von dem Borne, A. E. G. Kr. (1994 a)
Vox Sang. 67,203-209.
Sequence analysis of cDNA derived from reticulocyte mRNAs coding for Rh polypeptides and demonstration of E/e and C/c polymorphisms.

153) Simsek, S., Bleeker, P. M. M. and von dem Borne, A. E. G. Kr. (1994 b)
The New England Journal of Medicine 330,795-796.
Prenatal determination of fetal RhD type.

154) Simsek, S., Faas, B. H. W., Bleeker, P. M. M., Overbeeke, M. A. M., Cuijpers, H. Th. M., van der Schoot, C. E. and von dem Borne, A. E. G. Kr. (1995)
Blood 85,2975-2980.
Rapid Rh D genotyping by polymerase chain reaction-based amplification of DNA.

155) Sonneborn, H. H., Uthemann, H. and Pfeffer, A. (1983)
Biotest Bulletin 4,328-330.
Monoclonal antibody specific for human blood group k (Cellano).

156) Southern, E. M. (1975)
J. Mol. Biol. 98,503-517.
Detection of specific sequences among DNA fragments separated by gel electrophoresis.

157) Springer, G. F., Williamson, P. and Brandes, W. C. (1961)
Journal of Experimental Medicine 113,1077-1093.
Blood group activity of Gram-negative bacteria.

158) Stafford, H. A., Tykocinski, M. L., Lublin, D. M., Holers, V. M., Rosse, W. F., Atkinson, J. P. and Medof, M. E. (1988)
Proc. Natl. Acad. Sci. USA 85, 880-884.
Normal polymorphic variations and transcription of the decay accelerating factor gene in paroxysmal nocturnal hemoglobinuria cell.

159) Strange, J. J., Kenworthy, R. J., Webb, A. J. and Giles, C. M. (1974)
Vox Sang. 27, 81-86.
Wk^a (Weeks), a new antigen in the Kell blood group system.

160) Stroup, M., MacIlroy, M., Walker, R. and Aydelotte, J. V. (1965)
Transfusion 5, 309-314.
Evidence that Sutter belongs to the Kell blood group system.

161) Suyama, K. and Goldstein, J. (1988)
Journal of Cell Biology 107, 561a.
The presence of Rh antigens on different polypeptide chains.

162) Suyama, K., Roy, S., Lunn, R. and Goldstein, J. (1993)
Blood 82, 1006-1009.
Expression of the 32-Kd polypeptide of the Rh antigen.

163) Tang, L. L., Redman, C. M., Williams, D. and Marsh, W. L. (1981)
Vox Sang. 40, 17-26.
Biochemical studies on McLeod phenotype erythrocytes.

164) Tate, C. G. and Tanner, M. J. A. (1988)
Biochem. J. 254, 743-750.
Isolation of cDNA clones for human erythrocyte membrane sialoglycoproteins α and δ.

165) Temple, G. F., Chang, J. C. and Wai Kan, Y. (1977)
Proc. Natl. Acad. Sci. USA 74,3047-3051.
Authentic β-globin mRNA sequences in homozygous β^0-thalassemia.

166) Thrasher, A. J., Keep, N. H., Wientjes, F. and Segal, A. W. (1994)
Biochimica et Biophysica Acta 1227, 1-24.
Chronic granulomatous disease.

167) Tindall, K. R. and Kunkel, T. A. (1988)
Biochemistry 27,6008-6013.
Fidelity of DNA synthesis by the *Thermus aquaticus* DNA polymerase.

168) Tomita, M., Furthmayr, H. and Marchesi, V. T. (1978)
Biochemistry 17,4756-4770.
Primary structure of human erythrocyte glycophorin A. Isolation and characterization of peptides and complete amino acid sequence.

169) Tsui, L. C. and Farrall, M. (1991)
Cytogenet. Cell. Genet. 58,337.
Report of the committee on the genetic constitution of chromosome 7.

170) Ugozzoli, L. and Wallace, R. B. (1992)
Genomics 12,670-674.
Application of an allele-specific polymerase chain reaction to the direct determination of ABO blood group genotypes.

171) Umenishi, F., Kajii, E. and Ikemoto, S. (1994 a)
Biochem. J. 299,203-206.
A new genomic polymorphism of Rh-polypeptide genes.

172) Umenishi, F., Kajii, E. and Ikemoto, S. (1994 b)
Biochem. J. 299,207-211.
Molecular analysis of Rh polypeptides in a family with RhD-positive and RhD-negative phenotypes.

173) Van der Hart, M., Szaloky, A. and van Loghem, J. J. (1968)
Vox Sang. 15,456-458.
A "new" antibody associated with the Kell blood group system.

174) Varmus, H. (1987)
Scientific American (September), 48-54.
Reverse transcription.

175) Venglen-Tyler, V., Gonzalez, B., Garratty, G., Kruppe, C., Johnson, C. L. and Mueller, K. A. (1987)
British Journal of Haematology 65,231-234.
Acquired loss of red cell Kell antigens.

176) Wai Kan, Y., Holland, J. P., Dozy, A. M. and Varmus, H. E. (1975)
Proc. Natl. Acad. Sci. USA 72,5140-5144.
Demonstration of non-functional β-globin mRNA in homozygous β^{o}-thalassemia.

177) Wallas, C., Simon, R., Sharpe, M. A. and Byler, C. (1986)
Transfusion 26,173-176.
Isolation of a Kell-reactive protein from red cell membranes.

178) Watkins, W. (1987)
Biochem. Soc. Trans. 15,620-624.
Biochemical genetics of blood group antigens: retrospect and prospect.

179) Wolter, L. C. and Hyland, C. A. (1993)
Blood 82,1682-1683.
Rhesus D genotyping using polymerase chain reaction.

180) Wu, D. Y., Ugozzoli, L., Pal, B. K. and Wallace, R. B. (1989)
Proc. Natl. Acad. Sci. USA 86,2757-2760.
Allele-specific enzymatic amplification of β-globin DNA for diagnosis of sickle cell anemia.

181) Yamamoto, F., Clausen, H., White, T., Marken, J. and Hakomori, S. (1990 a)
Nature 345, 229-233.
Molecular genetic basis of the histo-blood group ABO system.

182) Yamamoto, F., Marken, J., Tsuji, T., White, T., Clausen, H. and Hakomori, S. (1990 b)
The Journal of Biological Chemistry 265, 1146-1151.
Cloning and characterization of DNA complementary to human UDP-galNAc:fuc α 1->2 gal α 1->3 galNAc transferase (histo-blood group A transferase) mRNA.

183) Yamamoto, F., McNeill, P. D. and Hakomori, S. (1992)
Biochemical and Biophysical Research Communications 187, 366-374.
Human histo-blood group A^2 transferase coded by A^2 allele, one of the A subtypes, is characterized by a single base deletion in the coding sequence, which results in an additional domain at the carboxyl terminal.

184) Yamamoto, F., McNeill, P. D., Yamamoto, M., Hakomori, S. and Harris, T. (1993 a)
Vox Sang. 64, 171-174.
Molecular genetic analysis of the ABO blood group system: 3. Ax and B(A) alleles.

185) Yamamoto, F., McNeill, P. D., Yamamoto, M., Hakomori, S., Bromilow, I. M. and Duguid, J. K. M. (1993 b)
Vox Sang. 64, 175-178.
Molecular genetic analysis of the ABO blood group system: 4. Another type of O allele.

186) Yu, C. Y., Belt, K. T, Giles, C. M., Campbell, R. D. and Porter, R. R. (1986)
The EMBO Journal 5, 2873-2881.
Structural basis of the polymorphism of human complement components C4A and C4B: gene size, reactivity and antigenicity.

187) Zelinski, T., Coghlan, G., Myal, Y., Shiu, R. P. C., Philipps, S., White, L. and Lewis, M. (1991 a)
Ann. Hum. Genet. 55,137-140.
Genetic linkage between the Kell blood group system and prolactin-inducible protein loci: provisional assignment of *KEL* to chromosome 7.

188) Zelinski, T. A., Coghlan, G. E., Myal, Y., White, L. J. and Phillips, S. E. (1991 b)
Cytogenet. Cell. Genet. 58,1924.
Assignment of the Kell blood group locus to chromosome 7q.

189) Zelinski, T. (1991 c)
Transfusion 31,762-770.
The use of DNA restriction fragment length polymorphisms in conjunction with blood group serology.