Molecular analysis of classical red cell markers

D.A.Hopkinson
MRC Human Biochemical Genetics Unit, London

Introduction

The application of blood groups to criminology was predicted by Landsteiner in his first major paper on the ABO polymorphism in 1902 [1]. This proved to be a far sighted observation. Since then more than 20 independent blood group systems similar to ABO have been identified by serological means [2]. The major driving force behind these discoveries was the importance of blood group antigens in blood tranfusion and in the pathogenesis of haemolytic disease of the newborn. However many of the systems were sufficiently variable in major world populations to be used as tools in forensic and anthropological investigation.

About fifty years ago the technological approaches to the analysis of red cell markers changed with the introduction of electrophoretic techniques. First, the many different forms of haemoglobin and their variants were revealed by the relatively crude techniques of zone electrophoresis using paper, cellulose acetate and agar gels as the support matrix. More sophisticated methods, such as starch gel and polyacrylamide gel electrophoresis were then introduced in the fifties and sixties and the range of genetic variation revealed by these procedures rapidly expanded to other abundant proteins in blood such as the haptoglobin and transferrin systems [refer to Sensabaugh this Symposium] and also to much less abundant proteins, enzymes, which were demonstrated as isozymes using powerful new specific detection methods [3].

The serological methods, originally devised to investigate the classical erythrocyte antigens, led on to the discovery of many other polymorphisms such as the platelet antigen systems and the highly variable histocompatibility (HLA) systems [refer to Mayr this Symposium], which have found extensive use in forensic work. The electrophoretic methods have been equally diverse and rewarding. The introduction of refinements such as isoelectric focussing were particularly important both in improving the resolution of known polymorphic markers and in identifying sub-types in classical systems. Forensic application of these classical markers was often an important driving force behind the introduction of these techniques; for example, I well remember the excitement generated in my own field by the work of forensic laboratories in the UK [4] and in Germany [5] leading

to the discovery of the phosphoglucomutase (PGM1) sub-types and the transformation of a classical 2-allele, 3-phenotype system into 4 alleles and 10 phenotypes.

This then, very briefly, is the background to the classical markers which are at the heart of our Symposium. We shall review the transformation wrought by the introduction of DNA technology which allows us to analyse these markers directly at the level of the gene. We shall discuss the merits (and demerits) of the molecular approaches which give direct information about genotype and we shall review the benefits which continue to acrue from the further analysis of these long established classical polymorphisms. In many cases predictions were made several years ago from detailed phenotypic analysis about the nature of the mutations which might be responsible for several of these polymorphic systems. In some cases the predictions are being fullfilled but in others we find, to paraphrase the cliche, that the truth of the DNA genotype is stranger than the prediction of the protein phenotype.

Table 1 lists the classical red cell polymorphisms which have been extensively studied in forensic medicine and used as tools in this arena.

Table 1: "Classical" red cell polymorphisms

1. Red cell antigens		Reference
1.1	ABO, H, Secretor & Lewis	[6-10,13,27, 40]
1.2	MNSs	[28-38, 40]
1.3	Rh	[14-19, 40]
1.4	Kell	[39, 40]
1.5	Kidd	[40]
1.6	Duffy	
2. Red cell isozymes		
2.1	Acid Phosphatase	[Dissing, this Symposium]
2.2	Phosphoglucomutase	[22-24]
2.3	Phosphogluconate dehydrogenase	[41]
2.4	Adenylate kinase	[42]
2.5	Adenosine deaminase	[43]
2.6	Glutamate pyruvate transaminase	
2.7	Esterase D	[44]
2.8	Glyoxylase	[45]
2.9	Peptidase D	[46]
2.10	Cytidine deaminase	[47]

The markers are arranged in roughly chronological order of discovery. In most cases progress has been made in cloning the gene(s) responsible for the polymorphic phenotypes and this is indicated by the references alongside. The lists are not comprehensive and similarly this review cannot be all embracing. In order to illustrate the general points for the purpose of this Symposium, this review focuses on three red cell systems - the ABO and Rh blood groups and the phosphoglucomutase (PGM1) isozyme polymorphism.

ABO Blood group system

The molecular basis of this polymorphism has been known for many years to be mediated via carbohydrate determinants and variant glycosyltransferases encoded by the ABO locus on the long arm of human chromosome 9. Classical biochemistry led to the discovery that the A enzyme was an N-acetyl-D-galactosaminyl transferase and the B encoded enzyme was a D-galactosyl transferase. All the biochemical and serological evidence indicated that the O gene was an amorph.

The molecular cloning of the ABO locus came about following the purification of the A-transferase from human lung tissue and the determination of part of the amino acid sequence. This provided the basis for the design of an oligonucleotide probe for cDNA library screening and the derivation of the complete sequence of the A-transferase mRNA [6]. Comparison of cDNA clones derived from human adenocarcinoma cell lines of known ABO phenotype showed that the DNA sequence of the B gene differs from A at 7 nucleotides which result in 4 amino acid substitutions in the catalytic domain and these account for the different sugar specificities of the A and B transferases [7,8,9]. The O gene was found to be identical to the A coding sequence except for a single base deletion at nt 261 which leads to a frame-shift, then a premature stop codon and the synthesis of a functionally inactive O transferase of only 117 amino acid residues which lacks most of the catalytic domain. Thus the O gene is indeed an amorph as predicted from the phenotypic data. The minor subtype of A, the A_2 transferase differs from A1 by a single base substitution at nt 467 giving rise to Leu for Pro in the catalytic domain at codon 156 and by a single base deletion at the C terminus which produces a frameshift and 21 extra amino acid residues [9]. Several other well known but uncommon classical types such as cis AB, A^x, $B^{(A)}$ have also recently been characterised [10].

The cloning of ABO and the identification of DNA sequence differences between the A, B and O alleles provides a means for direct genotyping. Three methods have been described all involving PCR amplification of genomic DNA. The first method involves diagnostic restriction enzyme digestion of the PCR products derived from the sites of the substitutions which distinguish A, B and O, followed by Southern blot analysis with

specific probes [8]; this is a fairly lengthy procedure. The second is a much more rapid multiplex PCR procedure [11] which takes advantage of the critical single base deletion in the O gene at nt 261 and the nucleotide substitutions at nt 526 and nt 703 which usually distinguish B from A and O. Using four different sets of primers in a single reaction, 41 individuals of known ABO type were unambiguously genotyped and no exceptional cases were encountered. The third method [12] uses denaturing gradient gel electrophoresis (DGGE) to separate a 250 nt fragment of genomic sequence which spans the site of the single nucleotide deletion (nt 261) associated with the O allele. The procedure is also very rapid but requires specialised equipment and GC clamps on the oligonucleotide primers for the PCR. However, DGGE distinguishes hitherto unrecognised polymorphisms associated with the O and B alleles. In testing, 95 European individuals, 4 different O alleles, 2 B alleles and 1 A allele were identified by DGGE and the level of recognisable heterozygosity and hence the information content of the locus as a genetic marker was raised from 3/95 (3%) to 66/95 (70%). The mutations underlying this heterogeneity have not yet been defined precisely but it seems most likely that the single base deletion characteristic of O is the major feature which separates O from A and B on the denaturing gradient gels. Combinations of point mutations in intron sequence flanking this deletion presumably lead to the four O and three B subtypes.

ABO genotyping is of course already extremely valuable in linkage analysis and gene mapping on human 9q34. Previously uninformative families, O by O matings for example can now be scored by the DGGE technique and the AO and BO heterozygotes can easily be resolved from AA and BB homozygotes respectively. Applications in the forensic field are easily envisaged in paternity work and criminal investigation.

As an aside it is also interesting to note that the gene for the H transferase, which acts on the precursor of the ABO substrate, has also been cloned and sequenced [13]. The gene, located on chromosome 19, encodes an α (1, 2) fucosyltransferase. Its isolation opens the door to the investigation of variations in the ABO system which are due to the action of this gene - in particular the Bombay and para Bombay phenotypes. Preliminary studies have already revealed that in some of these rare individuals, the H gene is disabled due to point mutations leading to small deletions and frame shifts, premature stop codons or substitutions affecting the active site.

Rh Blood group system

From the very start it was clear that Rh is a complex blood group system showing enormous diversity in human populations. Literally dozens of different phenotypes can be defined by serological means and the terminology is necessarily extremely complex. At the simplest level 3 sets of antigens have been defined. These have been given various

names but here they are referred to as CDE, with the antithetical allelic antigens cde and they are encoded by a gene complex on chromosome 1p. At first it was not clear how many genes comprised the Rh gene cluster and one very famous theory was the trinitarian structure due initially to R. A. Fisher. Biochemical evidence eventually pointed to the occurrence of 2 types of Rh polypeptide chains and Patricia Tippett took this evidence into account together with serological aspects of the system and proposed a 2 locus model (the 'D' locus and the 'Cc Ee' locus) in 1986.

The attempts to clone the Rh gene has produced similarly complex results which have until very recently been difficult to interpret. For example, 2 groups [14,15] independently isolated an identical Rh cDNA clone using oligonucleotide probes based on N-terminal amino acid sequence from purified Rh polypeptides and it was not clear whether the cDNA was derived from the D or the C/E locus. It now emerges that there are several different types of Rh cDNA but they fall into 2 main classes, which show a very high level of sequence identity and the new data are in support of the 2-locus model of the Rh blood group polymorphism [16,17,18].

The D locus is present in most individuals, it is expressed in red cells and encodes a 417 residue hydrophobic polypeptide which contains about 12 regularly spaced transmembrane domains and is firmly embedded in the red cell membrane. The polypeptide corresponds to the major D antigen responsible for the well known Rh positive blood group. Rh negative (dd) individuals are homozygous for a deletion of the D gene sequence [19].

The C/E locus encodes the E/e antigens and the C/c antigens but these polypeptides differ in length due to alternative splicing of the C/E transcripts [18]. Thus the E/e antigens are full length transcripts of 417 residues, the same length as the D polypeptide. The difference between the E and e antigens depends on a point mutation in exon 5 which substitutes Ala for Pro at residue 226. The C/c transcripts are shorter due to the splicing out of exons 4, 5 and 6 or exons 4, 5 and 8 and lead to the synthesis of polypeptides 266 or 267 residues in length respectively. These differ from each other and from the E/e polypeptides in their patterns of integration into the red cell membrane. The splicing leads to the loss of exon 5 and the E/e specificity and the exposure of N-terminal residues which are cryptic in the E and D polypeptides but confer C/c antigenicity. The differences between C and c appear to reside in four amino acid substitutions in exons 1 and 2.

These data summarise very recent findings but clearly there are now opportunities to characterise precisely the molecular events underlying the many low frequency Rh antigens and to use PCR based methods for Rh genotyping in forensic case work.

Phosphoglucomutase (PGM1) isozyme polymorphism

The PGM polymorphism was first identified 30 years ago by conventional starch gel electrophoresis as a 2-allele, 3-phenotype system with a world-wide occurrence [20]. In due course, as mentioned earlier, isoelectric focussing revealed further complexity and the PGM1* 1 and the PGM1* 2 alleles were subdivided into PGM1* 1+ and PGM1* 1- and PGM1* 2+ and PGM1* 2- respectively. Shortly afterwards it was predicted that intragenic recombination might lay at the heart of this polymorphism [21]. For example, if 1+ was the founder allele and 1- and 2+ arose from it by 2 independent point mutations, the 2- allele could arise by intragenic recombination in a 2+1- heterozygote. Recent successful attempts to clone PGM1 have allowed this hypothesis to be tested and also provide new methods for PGM1 typing which maybe valuable in the forensic field [22,23].

The human PGM1 cDNA was cloned by a more roundabout route than the ABO and Rh genes described above. The first step was to raise an anti rabbit PGM1 antibody and then use this as a reagent to screen a cDNA expression library to isolate a rabbit PGM1 cDNA. The latter was then used as a probe to isolate the human PGM1 cDNA.

As expected from what was known about the rabbit PGM1 polypeptide sequence, the human PGM1 cDNA encodes an enzyme of about 60 kDa molecule size. Structural studies have revealed that the PGM1 protein is a monomer arranged in four major domains with the active site buried at the confluence of the domains. The coding sequence is divided into 11 exons but the intron/exon junctions do not coincide with the domain boundaries [24].

We have now analysed the whole of the coding region of the human PGM1 gene by DNA sequencing in individuals of known PGM1 isozymes phenotype [23]. Only 2 mutations were found, both C - T transitions, at nt 732 in exon 4 and 1320 in exon 8. These mutations correlate exactly with the protein phenotype: in exon 4, at nt 732 individuals showing the PGM1 isozyme carry the Arg codon CGT whereas individuals showing the PGM1 2 isozyme carry the Cys codon TGT. Similarly in exon 8, at nt 1320 individuals with the PGM1 + isozyme carry the Tyr codon TAT whereas individuals with the PGM1 - isozyme carry the His codon CAT. The charge changes predicted by this amino acid substitution are entirely consistent with the isoelectric profiles of the four PGM1 isozymes. Thus there is very strong evidence to support the view that only 2 point mutations are involved in the generation of the four common alleles and that one allele must have arisen by homologous intragenic recombination between these mutation sites. Convincing further evidence to support this hypothesis has come from the molecular analysis of the PGM1* 3^+, 3^-, 7^+ and 7^- alleles [25].

The elucidation of the molecular basis of the PGM1 polymorphism provides immediate and simple methods for genotyping [23]. For example the mutation at nt 723 abolishes a restriction endonuclease recognition site for Bgl II AGATCT in allele 2, when it becomes AGATCC in allele 1; but it creates a new recognition site for the enzyme Alw I (AGATCN4 to GGATCN4 in the reverse strand). Similarly at nt 1320 the DNA sequence CATG in allele - is a recognition site for Nla III but this becomes TATG in allele + and the site is lost. Restriction enzyme digestion of PCR generated products of exons 4 and 8 will therefore easily and specifically detect the PGM1 1/2 and +/- types. The system has already been tested out on forensic blood stains and works very well [26]. The mutations can also be resolved rapidly and easily by SSCP analysis on the Phast system with simple silver staining.

Conclusions

The classical genetic polymorphisms of human red cells continue to occupy an important place in the forensic field. They are well known and understood by the forensic scientists in the laboratories and in the courts by the police, lawyers, judges and to some extent by the general public. Also they are supported by a wealth of documentary evidence from family and population studies in a vast range of different human populations from all over the world [48]. The screening methods for phenotype analysis are rapid, cheap and effective and in many cases, especially for those red cell markers which are serologically or immunologically defined, can be automated at relatively low cost.

Most of the important blood group loci and red cell isozyme systems can now be analysed directly by molecular DNA methods and the genotypes correlated precisely with the phenotypes. The molecular methods will enhance these systems enormously by providing higher levels of heterozygosity and greater accuracy in the assessment of genotypes for the common and for the rare mutations.

References

1. Landsteiner, K. (1901). Wien. Klin. Wschr. **14**, 1132.
2. Race, R.R. & Sanger, R. (1975). Blood groups in Man. 6th edit. Blackwell, Oxford.
3. Harris, H. & Hopkinson, D.A. (1976). Handbook of enzyme electrophoresis in human genetics. N. Holland.
4. Bark, J.E., Harris, M.J. & Firth, M. (1976). J. Forens. Sci. Soc. **16**, 115.
5. Kühnl, P., Schmidtmann, U. & Spielmann, W. (1977). Hum. Genet. **35**, 219.
6. Yamamoto, F. et al. (1990). J. Biol. Chem. **265**, 1146.
7. Yamamoto, F. & Hakomori, S. (1990). J. Biol. Chem. **265**, 19257.
8. Yamamoto, F. et al. (1990). Nature. **345**, 229.
9. Yamamoto, F., McNeill, P.D. & Hakomori, S. (1992). Biochem. & Biophys. Res. Comm. **187**, 366.
10. Yamamoto, F. et al. (1993). Vox Sang. **64**, 116, 120, 171.
11. Ugozzoli, L. & Wallace, R.B. (1992). Genomics. **12**, 670.
12. Johnson, P.H. & Hopkinson, D.A. (1992). Hum. Mol. Genet. **1**, 341.
13. Larsen, R.D. et al. (1990). Proc. Natl. Acad. Sci. **87**, 6674.
14. Chérif-Zahar, B. et al. (1990). Proc. Natl. Acad. Sci. **87**, 6243.
15. Avent, N.D. (1990). Biochem. J. **271**, 821.
16. Le Van Kim, C. et al. (1992). Proc. Natl. Acad. Sci. **89**, 10925.
17. Kajii, E. et al. (1993). Hum. Genet. **91**, 157.
18. Mouro, I. et al. (1993). Nature Genet. **5**, 62.

19. Colin, Y. et al. (1991). Blood. **78**, 2747.
20. Spencer, N., Hopkinson, D.A. & Harris, H. (1964). Nature. **204**, 742.
21. Carter, N.D. et al. (1979). Ann. Hum. Biol. **6**, 221.
22. Whitehouse, D.B. et al. (1992). Proc. Natl. Acad. Sci. **89**, 411.
23. March, R.E. et al. (1993). Proc. Natl. Acad. Sci. **90**, in press.
24. Putt, W. et al. (1993). Biochem J. **296**, in press.
25. Takahashi, N. & Neel, J.V. (1993). Proc. Natl. Acad. Sci. **90**, in press.
26. March, R.E. et al. (1993). Forens. Sci. Internat. (submitted for publication).
27. Kukowska-Latallo, J.F. et al. (1990). Genes Dev. **4**,1288.
28. Kudo, S. & Fukuda, M. (1989). Proc. Natl. Acad. Sci. **86**, 4619.
29. Siebert, P.D. & Fukuda, M. (1986). Proc. Natl. Acad. Sci. **83**, 1665.
30. Siebert, P.D. & Fukuda, M. (1987). Proc. Natl. Acad. Sci. **84**, 6735.
31. Rahuel, C. et al. (1989). Gene. **85**, 471.
32. Rahuel, C. et al. (1988). Eur. J. Biochem. **172**, 147.
33. Tate, C.G. & Tanner, M.J.A. (1988) Biochem. J. **254**, 743.
34. Huang, C-H. et al. (1991). Bailliere's Clin. Haematol. **4**, 821.
35. Chasis, J.A. & Mohandas, N. (1992). Blood. **80**, 1869
36. Cartron, J-P et al. (1990). In Blood Cell Biochemistry, Edit. J.R. Harris. Chap. 10, 299. Plenum
37. Cartron, J-P & Rahuel, C. (1992). Transfusion Medicine Reviews. **6**, 63.
38. Corfield, V.A. et al. (1993). Transfusion. **33**, 119.
39. Lee, S. et al. (1991). Proc. Natl. Acad. Sci. **88**,6353.
40. Lutz, P. & Dzik, W.H. (1992). Transfusion. **32**, 467.
41. Kleyn, P. et al. (1989). Cytogenet. Cell Genet. **51**, 1023.
42. Puffenberger, E.G. & Francomano, C.A. (1991). Nucleic. Acid. Res. **19**, 1161.
43. Wiginton, D.A. et al. (1986). Biochemistry. **25**, 8234.
44. Lee, E.Y-H. P. & Lee, W-H. (1986). Proc. Natl. Acad. Sci. **83**, 6337.
45. Ranganathan, et al. (1993). J. Biol. Chem. **268**, 5661.
46. Tanoue, A. et al. (1990). J. Biol. Chem. **265**, 11306.
47. Kuhn, K. et al. (1993). Biochem. Biophys. Res. Comm. **190**, 1.
48. Roychoudhury, A.K. & Nei, M. (1988). Human polymorphic genes: World distribution., Oxford Uni. Press.

Acknowledgement Support from the Medical Research Council and the Forensic Science Service for the work on PGM is gratefully recorded.