RAPID COMMUNICATION

The Coding Sequence of Duffy Blood Group Gene in Humans and Simians: Restriction Fragment Length Polymorphism, Antibody and Malarial Parasite Specificities, and Expression in Nonerythroid Tissues in Duffy-Negative Individuals

By Asok Chaudhuri, Julia Polyakova, Valerie Zbrzezna, and A. Oscar Pogo

The coding and untranslated flanking sequences of Duffy gene (FY) in humans and simians are in a single exon. The difference between the two codominant alleles, FY*A and FY*B, is a single change at nucleotide 306: guanine is in FY*A and adenine is in FY*B. This produces a codon change that subsequently modifies the amino acid at position 43 of gpFy, the major subunit of the Duffy blood group protein complex. The glycine at this position in antigen Fy* exchanges with aspartic acid in antigen Fy+. The guanine at nucleotide 306 creates an additional Ban I restriction site in FY*A. Ban I digestion of DNA-PCR amplified products of FY*B and FY*A yields three and four fragments, respectively. Restriction fragment length polymorphism (RFLP) studies show that Fya+b-) and Fya+b+) whites are FY homozygous, that most Fya+b-) blacks have FY*B, and most Fya+b-) blacks are FY*A/FY*B heterozygous. In the black population a silent FY*B is very common, but a silent FY*A has not been found yet. On RNA blot analysis, the gpFy cDNA clone detected mRNA in the lung, spleen, and colon but not in the bone marrow of Duffy-negative individuals. Therefore, there is no null phenotype in Fya+b-) blacks. The gpFy homology between human and chimpanzee is 99% with a single residue change at position 116 (valine to isoleucine), whereas a 94% homology is found in squirrel and rhesus monkeys, and there is a 93% homology in aotus monkey when compared with humans. The N-terminal exocellular domain of simian gpFy helps to identify a set of amino acids critical for antibody and malarial parasite specificities.

© 1995 by The American Society of Hematology.

From the Laboratory of Cell Biology, Lindsley F. Kimball Research Institute of the New York Blood Center, New York, NY. Submitted August 22, 1994; accepted October 25, 1994.

Supported by a New York Blood Center Institutional Grant. Address reprint requests to A. Oscar Pogo, MD, DMSc, Laboratory of Cell Biology, Lindsley F. Kimball Research Institute of the New York Blood Center, 310 E 67th St, New York, NY 10021.

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "advertisement" in accordance with 18 U.S.C. section 1734 solely to indicate this fact.

© 1995 by The American Society of Hematology.

0006-4971/95/8503-002S$3.00/0

Blood, Vol 85, No 3 (February 1), 1995: pp 615-621
The Fy cDNA were amplified by PCR as explained in Materials and Methods. These primers, the genomic DNA of the four Duffy phenotypes and the erythrocytes. Erythrocytes that lack gpFy have no manifested abnormalities, and individuals of Duffy-negative phenotype are healthy.

Here we report the lack of introns in FY, the RFLP of FY\(^A\) and FY\(^B\): the presence of gpFy mRNA in the lung, spleen, and colon of Duffy-negative individuals; the deduced amino acid sequences of FY in chimpanzee, aotus, squirrel, and rhesus monkeys; and the identification of antibody and malarial parasite specificities at the N-terminal exocellular domain of gpFy.

**MATERIALS AND METHODS**

Hemagglutination tests were performed on human blood according to standard procedures. Human tissues were acquired from the National Cancer Institute Cooperative Human Tissue Network (Bethesda, MD), and the blood of these tissue donors was typed in our institute. Chimpanzee (Pan troglodytes) and rhesus monkey (Macaca mulatta) blood was obtained from the Laboratory for Experimental Medicine and Surgery in Primates (LEMSIP) in Tuxedo, NY. Squirrel monkey (Saimiri sciureus) blood was procured from Hoffman-La Roche (Nutley, NJ). Aotus monkey (Aotus trivirgatus) blood was obtained from the laboratory of Dr. L. Miller at the National Institutes of Health (Bethesda, MD).

Genomic DNA and mRNA were isolated as explained.\(^7\) gpFy gene amplification was performed in a polymerase chain reaction (PCR) with 5' sense and 3' antisense primers as explained in Fig 1. For enzymatic amplification, Taq polymerase (Stratagene, La Jolla, CA) was used with the following amplification cycles: 1 minute at 94°C, 2 minutes at 65°C, and 3 minutes at 72°C for 30 cycles followed by 10 minutes elongation at 72°C. PCR products were gel purified and cloned into pBluescript-SK vector (Stratagene). Both strands were sequenced by dideoxy chain termination method with vector primers and primers designed from sequenced regions of both strands. RNA-blot analyses were performed as explained.\(^7\)

The 5' upstream and 3' downstream regions of the Duffy gene that are contiguous to the 5' and 3' untranslated flanking sequences of gpFy mRNA were cloned and sequenced. The 5' end was identified in a 1.4-kb Pst I digested fragment hybridized with the 5' end of the gpFy cDNA, and the 3' end was identified in a 1.9-kb EcoRI digested fragment hybridized with the 3' end of gpFy cDNA. To clone the 5' upstream region, genomic DNA was digested to completion with Pst I and ligated with a Pst I cassette (Pan Vera Corp, Madison, WI). PCR amplification was performed using cassette primer C1 and 3' antisense primer from Fyb cDNA clone. A second PCR amplification was performed with a nested 5' primer and Pst I cassette primer C2. To clone the 3' downstream region the same procedure was used except that an EcoRI cassette was ligated to EcoRI digested genomic DNA. The first amplification was performed with primer cassette C1 and 3' sense primer from Fyb cDNA clone, and the second amplification was performed using nested 3' primer and cassette primer C2. The amplified fragments were gel purified and sequenced as explained.\(^8\)

**RESULTS**

The coding sequence of the Duffy gene is in a single exon. gpFy mRNA consisted of 1,267 nucleotides, of which 175 and 51 untranslated nucleotides are at the 5' and 3' end, respectively.\(^8\) The 5' end of gpFy mRNA was validated by primer extension. Thus, the extended product of the antisense primer yielded a sequence of 80 nucleotides that exactly matched the predicted size at the 5' end of the Fyb cDNA clone.\(^8\)

To determine FY coding and noncoding sequences, a 5' end sense primer corresponding to nucleotides 2 to 25 and a 3' end antisense primer corresponding to nucleotides 1,211 to 1,234 of Fyb cDNA clone were synthesized (Fig 1).\(^8\) Using these primers, the genomic DNA of the four Duffy phenotypes and the Fyb cDNA were amplified by PCR as explained in Materials and Methods. The PCR products were run on a 2% agarose gel, and visualized with ethidium bromide.

![Fig 1. The size of PCR products of Fyb cDNA clone and the genomic DNA of the four Duffy phenotypes. Lanes: M, DNA size markers; 1, Fyb cDNA clone; 2, Fy(a-b-); 3, Fyb(a-b-); 4, Fyb(a+b+); and 5, Fyb(a+b-). A 5' end sense primer GCTTCCCCAGGACTGTTCCTG-TGCTG corresponding to nucleotides 1211 to 1234 of clone Fyb cDNA were chemically synthesized.\(^8\) Using these primers, the genomic DNA of the four Duffy phenotypes and the Fyb cDNA were amplified by PCR as explained in Materials and Methods. The PCR products were run on a 2% agarose gel, and visualized with ethidium bromide.](image-url)
the lung, colon, and spleen of Duffy-negative individuals synthesized the same size gpFy mRNA, albeit less abundantly, as that of Duffy-positive individuals (Fig 2). Therefore, it is reasonable to assume that Duffy-negative individuals produce the same gpFy mRNA observed in the same nonerythroid tissues of Duffy-positive individuals. FY is silent in the BM and is downregulated in the nonerythroid tissues of the Duffy-negative phenotype.

Restriction fragment length polymorphism (RFLP) of FY. A single nucleotide change was observed between the sequence of FY*A and FY*B alleles. The base at nucleotide 306 was guanine in FY*A and adenine in FY*B (Fig 3A). The difference produced a codon change that subsequently modified the residue at position 34 Glycine in FY*A changed to aspartic acid in FY*B. Furthermore, the guanine created an additional restriction site in FY. The fact that 20 FY(a+b−) and four FY(a+b+) blacks possess only FY*B indicates that a BM silent FY*A is infrequent (see Appendix).

gpFy gene and protein in nonhuman primates. gpFy gene was PCR amplified and sequenced from the genomic DNA of chimpanzee, aotus, squirrel, and rhesus monkeys (not shown). Along the entire length of the gene, there was a high degree of sequence homology between humans and simians. Simians, like humans, encoded gpFy in a single exon. The canonical ATG initiator codon was present in all the except the aotus, where it was ACG. Recognition of ACG, threonine, as the initiation codon of translation has been observed in some mouse and viral mRNAs. The protein sequence homology between humans and chimpanzees was 99% with a single residue change at position 116 (valine to isoleucine). A 94% homology was found in squirrel and rhesus monkeys, and a 93% homology was determined in aotus monkey when compared with humans (Fig 4A).

The study of the primary structure of nonhuman primates gpFy highlighted several key amino acids for antibody and merozoite specificities (Fig 4, A and B). They were clustered in two overlapping regions at the extracellular N-terminal domain. Region I, residues 22 to 40, showed the changes that affected anti-Fy6 and P vivax specificities. Amino acid substitutions and deletions that correlate with modifications in ligand-binding specificities were seen in region I of rhesus monkey. As has been shown this simian’s erythrocytes are not recognized by anti-Fy6 and are not invaded by P vivax (Fig 4B). In rhesus the phenylalanine at residue 23 was deleted and substitutions occurred at the residues 22 (aspartic acid to asparagine), 26 (glutamic acid to serine), 29 (serine to phenylalanine), 32 (glycine to aspartic acid), 33 (valine to glycine) and, finally, 40 (glycine to valine). The remaining N-terminal domain was identical to human and chimpanzee with the exception of residue 16, where glutamic acid was substituted to the homologous glutamine.

Squirrel monkey region I, when compared with rhesus monkey, showed other modifications that may affect ligand-binding specificities in a different way. Thus, residue 23 was not deleted and residue 24, glutamic acid, was not replaced by serine (Fig 4A). Residue 32 was asparaginine instead of aspartic acid in rhesus monkey, but they are homologous amino acids: residue 39 was glutamic acid for aspartic acid and 40 was isoleucine for valine. These amino acid substitutions may be significant in preserving some affinity to anti-Fy6 (Fig 4B). It has been shown that squirrel monkey erythrocytes are recognized by anti-Fy6, albeit with less affinity. Although squirrel monkey erythrocytes are invaded by P vivax merozoites, they failed, like rhesus monkey, to bind P vivax 135- to 140-kD Duffy-specific protein. Aotus monkey erythrocytes are recognized by anti-Fy6 and by P vivax merozoites, they are invaded by P vivax 135- to 140-kD Duffy-specific protein. They are invaded by P vivax merozoites but at a lower rate. Aotus monkey gpFy region I showed similar modifications

Aotus.19 Malaria is a disease caused by a parasite.20 The parasite invades red blood cells, destroys them, and multiplies inside them. This cycle is repeated every 48 to 72 hours in the presence of a suitable host. The parasite is transmitted to a new host by the bite of a mosquito that has previously fed on an infected host. The parasite enters the mosquito’s salivary glands and multiplies in the mosquito’s gut. When the mosquito bites a new host, the parasite is transmitted along with the mosquito’s saliva.

GPi and GPiR are important in the life cycle of the parasite. GPi is a ligand for the Duffy antigen, which is present on the surface of red blood cells. GPiR is a receptor for the parasite. The parasite binds to GPi using GPiR and invades the red blood cell. The parasite then multiplies inside the red blood cell and eventually causes it to rupture. This process releases the parasite into the bloodstream, where it can infect other red blood cells.

The Duffy antigen is a protein that is present on the surface of red blood cells. It is used by the parasite to attach to the cell and enter it. The Duffy antigen is also recognized by antibodies that are present in the blood of people who are infected with the parasite. These antibodies can help to prevent the parasite from attaching to the cell and entering it.

The Duffy antigen is encoded by a gene on chromosome 1. The gene is called FY and is composed of three exons. The first two exons encode the N-terminal region of the protein, which is involved in ligand binding and is recognized by antibodies. The third exon encodes the C-terminal region of the protein, which is involved in protein function and is not recognized by antibodies.

The Duffy antigen is expressed on the surface of most red blood cells in the human population. However, some people do not express the antigen on their red blood cells. These people are called Duffy-negative. Duffy-negative individuals are at risk for malaria, as they cannot be infected with the parasite. Duffy-negative people are also at risk for other diseases caused by the parasite, such as yellow fever and hepatitis B.

The Duffy antigen is also used by the parasite to attach to the cell and enter it. This is called the Duffy antigen receptor for chemokines (DARC). DARC is a protein that is present on the surface of red blood cells. It is used by the parasite to attach to the cell and enter it. The parasite then multiplies inside the red blood cell and eventually causes it to rupture. This process releases the parasite into the bloodstream, where it can infect other red blood cells.

The Duffy antigen is expressed on the surface of most red blood cells in the human population. However, some people do not express the antigen on their red blood cells. These people are called Duffy-negative. Duffy-negative individuals are at risk for malaria, as they cannot be infected with the parasite. Duffy-negative people are also at risk for other diseases caused by the parasite, such as yellow fever and hepatitis B.
as squirrel except at residue 28 where asparagine was substituted to the homologous aspartic acid and at residue 37 where phenylalanine was substituted to leucine (Fig 4A).

In humans, chimpanzees, and rhesus monkeys the first 15 residues were the same (Fig 4A). In aotus monkey residue 9 was glycine instead of alanine and in squirrel monkey essential for anti-Fy6 reactivity with synthetic peptide in an enzyme-linked immunosorbent assay (ELISA) (unpublished results, August 1994). The nucleotide number of intronless genes are the subject of debate. Introns may also represent a recent evolutionary adaptation facilitating genetic diversity, or introns may have existed in the first cells and were lost in the evolution of organisms capable of rapid DNA replication. Currently, it is difficult to assess if the loss of introns in gpFy gene provides a selective advantage.

Fig 3. (A) Nucleotide sequences of FY*A and FY*B alleles, the deduced amino acid sequences of Fy* and Fyb gpFy and RFLP. The boxed letters indicate nucleotide and amino acid difference between genes and proteins, respectively. (B) Schematic diagram of approximate size fragments generated by the BanI digestion of FY*A and FY*B alleles. (C) RFLP of non-blacks. Lanes: 1, Fy(a+b-); 2, Fy(a-b+), and 3, Fy(a+b+). Note: band 0.517 in lane 1 was not fully digested. (D) RFLP of blacks. Lanes: 1 through 4, Fy(a+b-); 5 through 8, Fy(a-b+); and 9, Fy(a+b+). (E) RFLP of Fyla-b- blacks. Lanes: 1 through 6 Fy(a-b-) (20 was the actual sample size of Fyla-b-) (unpublished results, August 1993).8

Region II, residues 39 to 62, delineated human antisera specificities. Thus, residue 43 is critical for anti-Fy* and anti-Fyb specificities. However, the study of aotus and squirrel monkeys pointed out the role of other residues in determining if erythrocytes can react with anti-Fy*. Both simians had FY*B alleles and had aspartic acid at position 43, but only aotus monkey erythrocytes reacted weakly with anti-Fyb serum.20 Substitutions seen at residue 39 (aspartic acid to glutamic acid), residue 40 (glycine to isoleucine), residue 47 (glutamic acid to alanine), residue 56 (asparagine to serine), and residue 59 (aspartic acid to asparagine) may be critical for anti-Fyb specificity. Other substitutions at residue 45 (asparagine to serine) and at residue 62 (alanine to serine) may be less critical for anti-Fyb specificity because the same substitutions were found in aotus monkey gpFy that reacted weakly with anti-Fyb (Fig 4A).

DISCUSSION

Sequence analysis of human and simian genomic DNA-PCR products indicates that the coding sequence of the Duffy gene is colinear with Duffy mRNA and is, therefore, devoid of introns. Moreover, the 5' and 3' untranslated regions also are colinear with comparable regions in gpFy mRNA. Intronless gpFy gene in humans and simians adds another case in the growing number of functional intronless mammalian genes observed. Some of them contained one or more elements found in processed pseudo-gene and others have introns in their 5' and 3' flanking regions. The purposes of intronless genes are the subject of debate. Introns may represent a recent evolutionary adaptation facilitating genetic diversity, or introns may have existed in the first cells and were lost in the evolution of organisms capable of rapid DNA replication. Currently, it is difficult to assess if the loss of introns in gpFy gene provides a selective advantage.

Fy(a-b-) individuals do not produce gpFy mRNA in the BM, and their erythrocytes lack the gpFy protein.8,7 However, as shown here, the lung, spleen, and colon of Duffy-negative individuals produce mRNA of the same size but in less quantity than those of Duffy-positive individuals. It is fair to assume that the other nonerythroid tissues of Duffy-negative individuals also produce the same size gpFy mRNA as those of Duffy-positive individuals. We do not know yet if these tissues produce gpFy or a related protein. Immunohistochemical studies on the collecting duct and endothelial capillary cells of the kidney of a Duffy-positive individual show a protein that reacts with a rabbit polyclonal anti-gpFy antibody (to be published). The endothelial cells of the postcapillary venule in human kidney has recently been shown to express a protein that reacts with anti-Fy6. Moreover, kidney membrane proteins react with anti-Fy6 (and our own unpublished observation, August 1994). The nucleotide sequence of cDNA isolated from renal tissue mRNA by reverse transcription, amplification, and cloning is identi-
Duffy frequency in the black population, the probability of

We studied genomic DNA of individuals. The mechanism must be an inherited tissue-expression in BM and downregulates specific event and, as such, may entail an element(s) or a transacting protein factor(s) binding to the element(s) to suppress gpFy gene activity.

The expression of FY in the spleen is intriguing. Adult spleen and liver in humans are not erythropoietic tissues but both generate RBCs in the fetus. The presence of gpFy mRNA in fetal liver but not in adult liver correlates with the presence of erythropoietic function in the former and the absence in the latter. Yet, this correlation does not exist in spleen: FY is active in the absence of erythropoiesis. Because spleen concentrates peripheral blood, it is possible, yet unlikely, that spleen blood is enriched in reticulocytes. However, hybridization with β-globin mRNA showed a very weak signal in spleen, indicating that gpFy mRNA was of nonerythroid origin (unpublished observation, July 1993). Like other tissues, it will be important to identify which spleen cells produce gpFy.

Our studies confirmed the presence of two FY*B haplotypes; one is active and the other is silent in BM. Presently, it is not possible to distinguish active from silent FY*B. The determination of the latter haplotype depends on the tissue-specific components that suppress the expression of FY*B. We studied genomic DNA of 20 FY(a-b-) individuals, and all were FY*B. If we assume that the frequencies of FY*A and FY*B are comparable with those of the relative FY and FY frequencies in the black population, the probability of detecting at least one BM silent FY*A haplotype is very low (P = .0001, see Appendix). Moreover, six FY(a-b+) blacks did not show a single silent FY*A. Therefore, a silent FY*A must be rare in blacks. Because FY(a-b-) individuals are highly uncommon outside the black population, it is unlikely that non-blacks have a silent FY*A or FY*B.

It is worth mentioning that of all studied cases of multiply transfused FY(a-b-) blacks, anti-Fya, not anti-Fyb, was found in the sera. This is an indication that when FY(a-b-) blacks make Duffy antibodies, they usually make anti-Fya.

As shown here, these individuals have a BM silent FY*A, but they produce gpFy of FYa phenotype in their nonerythroid tissues. For them, gpFy of this phenotype is not a foreign protein. Because a silent FY*A is uncommon in FY(a-b-) blacks, there has been no reported anti-Fyb in transfused blacks having FY(a-b-) erythrocytes. The sequence of Duffy gene in FY(a-b-) phenotype of the few non-blacks will be very illustrative. Do they have a BM silent FY*A haplotype?

The study of the Duffy gene in nonhuman primates provides clues to the origin of the Duffy gene. FY(a-b+) erythrocytes are prevalent in chimpanzee and rhesus monkeys; therefore, FY*B is their predominant allele. FY(a+b-) erythrocytes in nonhuman primates have not yet been found. There are few cases of simian FY(a-b-); however, in new world monkeys it is common. They have gpFy because their erythrocytes react with anti-Fyb. FY*A is exceedingly scarce in nonhuman primates, and it may not exist. The first Duffy allele was probably FY*B and a point mutation at nucleotide 306 produced FY*A.

The identification of antibody and P vivax specificities were obtained by studying the amino termini of gpFy in nonhuman primates. Their amino acid modifications are biochemical manifestations of genetic phenotypes, and they are exempted from false results. The study highlights a set of residues that are critical for antibody and P vivax specificities. It must be emphasized that the findings identify neither
antigenic domains nor binding sites, because the ligands recognize the overall configuration of the N-terminal domain rather than its amino acid sequence.

Two overlapping regions can be identified. Region I, residues 22 to 40, is critical for anti-Fy6 specificity and P. vivax invasion of RBCs (Fig 4A). Humans and chimpanzees share the same residues; their erythrocytes react with anti-Fy6 and are invaded by P. vivax merozoites.3,5,18,19 (Fig 4B). On the contrary, anti-Fy6 does not react with and P. vivax merozoites do not invade rhesus monkey RBCs.3,18,19 The N-terminal domain of rhesus monkey gpFy is quite different from human and chimpanzee in region I only. With the exception of a substitution at residue 16 (glutamic acid to glycine), the remaining N-terminal domain of rhesus is the same as that of human and chimpanzee (Fig 4A). Whether the glutamine at residue 16 modifies the charge on the surface of rhesus monkey gpFy in such a manner that affects ligand-binding specificities remains to be shown.

Aotus and squirrel monkeys have a similar region I, because only 3 of 18 residues are different. Whether the amino acids at region I in these simians can be correlated with anti-Fy6 reactivity and P. vivax invasion is difficult to assess with the available data. Squirrel monkey gpFy reacts with anti-Fy6 but with less affinity.23 The anti-Fy6 affinity to aotus erythrocytes remains to be determined. Squirrel monkey erythrocytes are invaded by P. vivax merozoites but the 135- to 140-kD Duffy-specific receptor protein fails to bind squirrel monkey erythrocytes.18 This observation and the fact that anti-Fy6 inhibits partially the invasion of squirrel erythrocytes by P. vivax suggest the existence of an alternative mechanism of invasion that may be not gpFy related.18

Region II is delimited by residues 39 to 62 and defines anti-Fy4 and anti-Fy6 specificities (Fig 4A). If residue 43 is glycine or aspartic acid, the antigen is Fy4 or Fy6, respectively. However, the study of aotus monkey gpFy exposed the role of residues 47, 56, and 59 in anti-Fy6 specificity. If these residues are changed to alanine, serine, and asparagine, respectively, anti-Fy6 specificity is lost. The study of aotus monkey gpFy shows that other substitutions may be less critical. Thus, substitutions at residues 39 (aspartic acid to glutamic acid), 40 (glycine to isoleucine), and 45 (asparagine to serine) decrease but do not abolish anti-Fy6 reactivity.20 It appears that substitutions at region II in aotus and squirrel monkeys may not affect anti-Fy6 reactivity because similar changes in rhesus monkey do not abolish this antibody reactivity.

Human and chimpanzee gpFy carries at the N-terminal only two potential canonical sequences for N-glycosylation to asparagine residues.3 With the exception of aotus, no modifications of glycosylation sites were observed in the other simians. In aotus the substitution at residue 28 (asparagine for aspartic acid) has excluded one glycosylation site. Whether this affects anti-Fy6 affinity remains to be evaluated. However, we have determined that in vivo deglycosylated human erythrocytes have reduced anti-Fy6 affinity.9

We have not ascertained anti-Fy3 and -Fy5 specificities, and the status of anti-Fy4 is not clear.2 The anti-Fy4 reacted with Fy(a−b−) erythrocytes and with some Fy(a+b−) or Fy(a−b+) RBCs from blacks but not with the Fy(a+b+) phenotype. The antibody is weak and probably unrelated to any Duffy antigenic determinants because Fy(a−b−) RBCs lack gpFy.

In summary, we have shown the following: (1) FY is intronless; (2) the difference between FY*A and FY*B alleles is a single base change that produces a single amino acid change at residue 43; (3) all Fy(a−b−) blacks, studied thus far, have the FY*B gene; (4) Fy(a−b−) blacks harbor a BM silent FY*B that is active in several nonerythroid tissues; (5) gpFy in chimpanzee is 99% homologous to human and rhesus monkeys; and, finally, (6) the sequence at the N-terminal domain of nonhuman primates helps to identify two overlapping regions of antibody specificities and a putative region of P. vivax of specificity.

ACKNOWLEDGMENT

We are grateful to Dr J. Adamson for his support and encouragement. We extend our appreciation to the following individuals and laboratories at the New York Blood Center: Dr C. Falk for the statistical approach to sample size estimation, the Laboratory of Immunohematology for erythrocyte typing, D.H. Lee for DNA sequencing, and Dr M. Reid for helpful discussions and reviewing this paper. We also thank Dr Louis Miller at the National Institutes of Health for discussion and review of the paper. Thanks also to Dr D. Cohn of LEMSIP for chimpanzee and rhesus blood samples, Dr C. Field of Hoffmann-La Roche (Nutley, NJ) for squirrel monkey blood samples, and S.D. Jewell and K. Sexton of the National Cancer Center, Cooperative Human Tissue Network. Finally, we extend our appreciation to H. Zhang for her technical assistance.

APPENDIX

Contributed by Dr C. Falk

Director, Laboratory of Population Genetics

The New York Blood Center

It is important to know, based on our observed sample, what we can say about the frequencies of FY*A and FY*B in Fy(a−b−) blacks. Let us first assume that the relative frequencies of the two Duffy RFLP alleles in blacks are in the same proportion as those of the corresponding antigens Fy4 and Fy6. Using the values presented in Race and Sanger (Race RR, Sanger R: Blood Groups in Man. Oxford, UK, Blackwell Scientific, 1975), the estimates would be:

\[ freq (FY*B) = p = 0.64; \]
\[ freq (FY*A) = q = 0.36. \]

We can now test the validity of assuming a frequency of FY*A as high as 0.36, even though we did not observe FY*A in a sample of 20 Fy(a−b−) blacks. This then becomes our null hypothesis and we wish to test that null hypothesis, \( q < 0.36 \). The probability of seeing N FY*B in a sample of N Fy(a−b−) blacks is equal to \( p^N \). With our sample of 20 Fy(a−b−) individuals we are, therefore, able to reject the null hypothesis with a value of \( p^N \) or 0.0001.

Because we have rejected the null hypothesis, we would now like to know "How low can our estimate of the frequency of FY*A be and still allow us to reject that estimate
with a \( p \) value of 0.05?" To answer this, we must solve the equation: \( p_0^{0.05} = 0.05 \) for \( p_0 \). This gives us \( p_0 = 0.86 \) or equivalently \( q_0 = 0.14 \). In other words, based on our sample of 20 individuals all of who are \( FY^A, FY^B \), we can, at the 5% significance level, reject any allele frequency for \( FY^A \) down to as low as 0.14.

REFERENCES

The coding sequence of Duffy blood group gene in humans and simians: restriction fragment length polymorphism, antibody and malarial parasite specificities, and expression in nonerythroid tissues in Duffy-negative individuals

A Chaudhuri, J Polyakova, V Zbrzezna and AO Pogo