SfaNI Polymorphism Distinguishes the Alleles of the Glycophorin A Locus That Determine the MN Blood Group

S.H.	Sharon H.	Leos ^a
W.L.	William L.	Bigbee ^b
R.H.	Ronald H.	Jensen ^c
S.G.	Stephen G.	Grant ^b

^aBiomedical Sciences Division, Lawrence Livermore National Laboratory, Livermore, Calif., ^bCenter for Environmental and Occupational Health and Toxicology, University of Pittsburgh, and Molecular Carcinogenesis Program, University of Pittsburgh Cancer Institute, Pittsburgh, Pa., ^cDepartment of Laboratory Medicine, University of California School of Medicine, San Francisco, Calif., USA

Dr. Stephen G. Grant, Center for Environmental and Occupational Health and Toxicology, University of Pittsburgh, 260 Kappa Drive, RIDC Park, Pittsburgh, PA 15238 (USA)

The GPA gene is a single-copy autosomal gene coding for the major erythroid cell-specific sialoglycoprotein, glycophorin A. This locus is part of a family of related genes, including GPB and GPE, that have been localized to 4q28-31 [1]. The human GPA gene has been cloned and extensively characterized to reveal a mammalian gene of typical size and complexity consisting of seven coding exons and six introns extending over 40 kb. The gene is transcribed and processed to yield mRNA species of 1.0, 1.7, and 2.2 kb with identical coding regions specifying a cleavable hydrophobic leader peptide in addition to the 131 amino acids of the mature protein, differing only in the extent of the 3' untranslated region. GPA expression is detectable at the early erythroblast stage of erythropoiesis and is complete at the normoblast stage of differentiation. Bone marrow and peripheral blood reticulocytes as well as mature erythrocytes maintain a stable level of $0.5\text{--}1.0 \times 10^6$ GPA molecules on the cell surface. Two common GPA alleles are present in the human population, coding for proteins that differ at the first and fifth amino acids of the mature protein. These two allelic forms of GPA are expressed codominantly in heterozygous GPA^{Mm} individuals and serve to determine the MN blood group. These two alleles are approximately equally represented in the America population, with allele frequencies estimated to be 0.53:0.47 M:N for Caucasians and 0.49:0.51 for African-Americans [2]. The nucleotide sequences of the cloned GPAM and GPAN cDNAs have been determined [3-5], and, as shown in figure 1, the protein polymorphisms are attributable to

M allele N allele

ATA He

ATA

lie

TCA Ser

TCA
 Ser
 Sfa
 NI
 AGT ACC Ser Thr
 AGT ACC Ser Thr
 ACT
 Thr
 ACT
 Thr
 G|G-t|GTG GCA Gly Val| Ala
 glA-GIgTG GCA Glu Val Ala

GCA
 TfÇiA

Ala
 GCA
 Ala
 Ser
 t[t]a
 Leu

Fig. 1. Nucleotide and amino acid sequence of the GPA gene and protein in the region specifying the M and N alleles. Nucleotide and amino acid differences between the two alleles are indicated

with bold type and the nucleotides boxed. The recognition sequence of the SfaNI restriction endonuclease (shaded nucleotides) spans the last codon for the leader peptide and the first codon for the mature protein of the M allele. This sequence is disrupted in the N allele by the C→T base substitution responsible for the Ser→Leu difference at codon 1. Also indicated are the two base changes responsible for the Gly→Glu difference at codon 5 of the mature protein. three nucleotide substitutions, one in the middle of codon 1 and two clustered at the end of codon 5. These three base pair differences have been used previously to design allele-specific oligonucleotides for the GPAM and GPAN alleles [6, 7]. Computer analysis of these sequences revealed five restriction enzymes capable of unambiguously distinguishing these alleles, four of which are frequently cutting enzymes with 4-bp recognition sequences [8], as well as a unique SfaNI site in the first codon of the M allele which is disrupted by the single base change in the corresponding region of the N allele. Since SfaNI has a

KARGER

E-Mail karger@karger.ch Fax + 41 61 306 12 34 <http://www.karger.ch>

©1997 S.KargerAG, Basel

This article is also accessible online at: <http://BioMedNet.com/karger>

Fig. 2. Southern analysis of DNA from individuals of known GPA phenotype. Genomic DNA prepared from lymphocytes was digested with the restriction enzyme SfaNI, electrophoresed through a 1% agarose gel, blotted, and hybridized with radiolabelled GPA cDNA probe. A similar pattern of hybridizing fragments is detected in all 3 individuals, with the exception of (a) the indicated 3.0-kb fragment strongly hybridizing in the lane containing DNA from a homozygous GPAN/N individual (NN), at reduced intensity in DNA from a GPAN/WN heterozygote (MN), and missing altogether in a GPAM/M homozygote (MM) and (b) a 0.3-kb band, detectable upon longer exposure, with the reciprocal pattern of appearance, identifying it as the remaining hybridizing product of M-specific cleavage by SfaNI. Flanking marker lanes contain both λ DNA digested with HindIII and ΦX174 DNA digested with HaeIII, with fragment sizes indicated to the left in kilobases.

restriction recognition sequence of 5 bp, and the average fragment length for this enzyme in mammalian genomic DNA has been calculated as 1.2 kb [9], digestion with this enzyme should reveal a restriction fragment length polymorphism (RFLP) in human genomic DNA easily detectable by Southern analysis.

As demonstrated in figure 2, genomic DNA from an individual homozygous for the GPAN allele exhibited a unique 3.0-kb fragment upon digestion with SfaNI that is missing in an individual homozygous for the GPAM allele and reduced in intensity (40% by densitometric scan) in a Qpju/N heterozygote when hybridized to a cloned GPA cDNA [3]. This pattern is consistent with specific digestion of this fragment at the unique SfaNI site present in the GPAM allele. Indeed, a unique DNA fragment of approximately 0.3 kb can be detected in the DNA from the GPAM individual after long exposure. To ensure that these restriction patterns are associated with the known GPAM and GPAN alleles, and not merely random associations within the 3 unrelated individuals studied in figure 2, a total of 30 analyses, ten each from individuals of putative GPAM, GPA^Δ, and GPAM/N genotype, were performed, with complete concordance observed between restriction pattern and phenotype (data not shown).

Since the SfaNI site is associated with the allelic polymorphism located 27 bp from the 5' end of the 99-bp GPA exon 2, it can be deduced that flanking SfaNI sites common to both alleles occur approximately 2.7 kb upstream of the polymorphic site, within intron 1, which is known to extend for over 30 kb, and 0.3 kb downstream, placing it within the 0.7-kb intron 2 (fig. 3). Thus, the 3.0-kb genomic fragment observed from the GPAN allele hybridizes to the homologous 99 bp, corresponding to the intact exon 2 sequence in the cDNA probe. The 2.7-kb fragment present in the GPAM allele has only 27 bp of homology with the cDNA and is not detectable at our level of stringency. If the distances and directions to these flanking sites were reversed, the 2.7-kb GPAM fragment would be easily detectable, since it would then overlap with coding sequences from exons 3 and 4, as well as the remaining 72 bp in exon 2.

The SfaNI polymorphism described in this report differs from most such RFLPs in several important respects. Whereas RFLPs are normally found by screening ge-

52

Acta Haematol 1997;98:51-53

Leos/Bigbee/Jensen/Grant

M/N

M M/N

> 30

II III

0.1

IU

1.8

u

1.0

Fig. 3. Genomic structure of approximately 8 kb of the human GPA gene in the region of the MN polymorphism, with exons numbered and the sizes of introns and exons indicated in kilobases. The nucleotide differences responsible for the MN protein polymorphism and the M-specific SfaNI restriction site occur in exon 2, as indicated by the dashed arrow. The positions of the common flanking SfaNI

sites in introns 1 and 2 are indicated by solid arrows. Were the distances to these flanking sites reversed, a 2.7-kb SfaNI restriction fragment would be easily detected in the M allele using our cDNA probe, through hybridization to exons 3 and 4, as well as the remaining 70% of exon 2.

omic DNA by Southern analysis with a random battery of enzymes, the SfaNI polymorphism described above was predicted by sequence analysis. In addition, this DNA polymorphism is associated with a well-characterized, medically relevant protein polymorphism. Indeed, unlike a previously reported RFLP shown to be in linkage disequilibrium with the MN blood group [10], this SfaNI RFLP is intrinsically linked with the MN phenotype. This polymorphism should be useful in forensic studies to match blood typing data with solid tissue samples. The system also lends itself well to polymerase chain reaction adaptation [8], with restriction enzyme recognition sites specific for each of the single base pair differences distinguishing the M and N alleles in codon 1, as well as restriction sites specific for each of the 2-bp differences observed in codon 5.

Acknowledgements

This work was performed under the auspices of the US Department of Energy by Lawrence Livermore National Laboratory under contract W-7405-ENG-48. We thank M. Fukuda of the La Jolla Cancer Research Institute for providing us with the GPA cDNA clone pH2L, P.J.E. Quintana, B.R. DuPont, and J.J. Latimer for computer analysis of GPA sequences, L.N. Mascio for computer analysis of Southern band intensities, and A.G. Gorvad for technical assistance.

References

- Divelbiss J, Shiang R, German J, Moore J, Murray JC, Patil SR: Refinement of the physical location of glycophorin A and β fibrinogen using in situ hybridization and RFLP analysis. *Cytogenet Cell Genet* 1989;51:991. Issitt PD: The MN Blood Group System. Cincinnati, Montgomery, 1981. Siebert PD, Fukuda M: Isolation and characterization of human glycophorin A cDNA clones by a synthetic oligonucleotide approach: Nucleotide sequence and mRNA structure. *Proc Natl Acad Sci USA* 1986;83:1665-1669. Rahuel C, London J, d'Auriol L, Mattei M-G, Tournamille C, Skrzynia C, Lebouc Y, Galibert F, Cartron J-P: Characterization of cDNA clones for human glycophorin A: Use for gene localization and for analysis of normal or glycophorin-A-deficient (Finnish type) genomic DNA. *Eur J Biochem* 1988;172:147-153.
- Tate CG, Tanner MJA: Isolation of cDNA clones for human erythrocyte membrane sialoglycoproteins alpha and delta. *Biochem J* 1988;254:743-750.
- Corfield VA, Moolman JC, Martell R, Brink PA: Polymerase chain reaction-based detection of MN blood group-specific sequences in the human genome. *Transfusion* 1993;33:119-124.
- Huang CH, Guizzo ML, McCreary J, Leigh EM, Blumenfeld OO: Typing of MNSs blood group specific sequences in the human genome and characterization of a restriction fragment tightly linked to S-s-alleles. *Blood* 1991;77:381-386.
- DuPont BR, Grant SG, Oto SH, Bigbee WL, Jensen RH, Langlois RG: Molecular characterization of glycophorin A transcripts in human erythroid cells using RT-PCR, allele-specific restriction, and sequencing. *Vox Sang* 1995;68: 121-129.
- Drmanac R, Petrovic N, Glisin V, Crkvenja-kov R: A calculation of fragment lengths obtainable from human DNA with 78 restriction enzymes: An aid for cloning and mapping. *Nucleic Acids Res* 1986;14:4691-4692.
- 10 Murray JC, Buetow KH, Ferrell RE, Siebert PD, Fukuda M: An RFLP for glycophorin A (MN) is in linkage disequilibrium with MN and Ss. *Cytogenet Cell Genet* 1988;47:149-151.

SfaNI Restriction Defines MN Blood Type
Acta Haematol 1997;98:51-53