Glycophorin He(St*) of the Human Red Blood Cell Membrane Is Encoded by a Complex Hybrid Gene Resulting From Two Recombinational Events

By C.H. Huang, C. Lomas, G. Daniels, and O.O. Blumenfeld

A complex glycophorin (GP) variant of the human red blood cell membrane exhibiting both He and St* antigens was characterized at the molecular level. Restriction mapping identified two novel MspI fragments derived from the 5' and 3' portions of the GPHe(St*)-gene, respectively. Genomic DNA, including exons II-V and their splice junctions, was amplified by polymerase chain reaction, and the nucleotide sequences were determined. Comparison with the GPA and GPB sequences showed the presence in GPHe(St*) of multiple recombination breakpoints. In the 5' region of the variant gene, a sequence covering a portion of exon II to intron 2 had been transferred from GPA to GPB, resulting in a B-A-B hybrid structure. Such a gene conversion-like event introduced a number of templated and untemplated nucleotide replacements and was the direct cause for the expression of the He antigen. In the 3' region of the variant gene, an unequal crossover from GPB to GPA took place in the third intron at a recombination site apparently identical to that observed in the B-A hybrid GPSt* type A gene. These results indicated that GPHe(St*) occurs as a B-A-B-A hybrid gene, most likely originating from a two-step mechanism of homologous recombination. Transcript analysis showed the maturation from the GPHe(St*) pre-mRNA of two shortened mRNAs of which the exon III-deleted species encodes both the He and St* antigens.

© 1994 by The American Society of Hematology.

MATERIALS AND METHODS

Blood sample and serologic testing. Peripheral blood samples from normal human donors were used as controls. The variant blood was obtained from a black proband who was the first example known to express both the He and St* antigens on the erythrocyte surface. Serum typing showed that the proband's RBCs were M N + S - s + as well as He- and St(a+).

Restriction endonuclease mapping. High molecular weight genomic DNA was isolated from the leukocyte nuclear pellets and used for restriction analyses as described. Blots were hybridized with glycophorin cDNA or intron (IVS2 and IVS3) probes that were labeled with α-32P-JCTP (Amersham Corp., Arlington Heights, IL) by the random primer extension method.

Amplification of genomic DNA. Single-copy polymerase chain reaction (PCR) procedure was used for amplification of genomic DNAs. Templates for amplification of GPHe(St*)-gene sequences were the single copy 9.2-kb and 6.5-kb MspI fragments (Fig 1) electrophoresed from agarose gels as described. A genomic clone (λ hGpA 4-3) and the 8.0-kb SacI fragment from normal DNA were used as control templates for amplification of corresponding segments of GPA and GPB genes, respectively. The location of oligonucleotides (a to j) used as PCR primers is indicated in Fig 1; their nucleotide sequences were as described in previous publications. All the primers were synthesized on a 380A DNA synthesizer (Applied Biosystems, Foster City, CA) and purified by 15% polyacrylamide gel electrophoresis (PAGE) with 7.0 mol/L urea. Cycling was performed with Tag DNA polymerase. (AmpliTaq; Cetus Co, Norwalk, CT) in an automated DNA thermal cycler (Ernipro, San Diego, CA) using conditions described previously. The products were purified on native 5% PAGE, and the amplified

From the Lindsley F. Kimball Research Institute, New York Blood Center, New York, NY; the Department of Biochemistry, Albert Einstein College of Medicine, Bronx, NY; Gamma Biologicals, Inc, Houston, TX; and the Medical Research Council, Blood Group Unit, London, UK.

Submitted October 4, 1993; accepted February 7, 1994.

Supported by Grant No. RO1 GM16389 from the National Institutes of Health, Bethesda, MD.

Address reprint requests to Olga O. Blumenfeld, PhD, Department of Biochemistry, Albert Einstein College of Medicine, 1300 Morris Park Ave, Bronx, NY 10461 or to Cheng-Han Huang, MD, PhD, Lindsley F. Kimball Research Institute, 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.

© 1994 by The American Society of Hematology.
fragments were eluted from the gel in TE buffer. The supernatants were purified by 5% PAGE and eluted from gels. Aliquots of the eluates were used directly for nucleotide sequencing.

**Synthesis and amplification of glycoporphin cDNAs.** A modification of the procedure of Goossens and Kan was used to isolate total RNA from the hemolysate of a fresh sample of the proband's blood. Reverse transcription was performed in a volume of 20 μL using reverse transcriptase of avian myeloblastosis virus (Promega, Madison, WI). 2 to 5 μg of total RNA, and 250 ng of antisense primer spanning the region immediately downstream from the GPD termination codon (5'GTAAGAAAAATATCCTTGTATTAC-3'). For cDNA amplification, this primer was coupled with an upstream oligonucleotide defining a portion of the leader sequence in exon I of glycoporphin genes (5'GTATGGAAAAATATCCTTGTATTAC-3'). Amplification conditions were as described. 

**Direct DNA sequencing.** The PCR-amplified fragments were directly sequenced on both strands by the dideoxy chain termination method using T7 DNA polymerase (Sequenase; US Biochemicals, Cleveland, OH). Synthetic primers were 5'-end labeled with γ-ATP (DuPont-New England Nuclear, Boston, MA). Conditions for primer labeling, sequencing reaction, and denaturing PAGE were the same as described previously.

**Immunoblot analysis of RBC glycoporphins.** RBC lysates from a normal control and He + St(a+) proband were separated by sodium dodecyl sulfate (SDS)-15% PAGE. Proteins were electroblotted onto the nitrocellulose filter, and glycoporphins were visualized by a panel of antibodies, as described.

**RESULTS**

Restriction digestion with Mspl showed gross alterations in the GPHe(St') gene caused by the introduction of a unique Mspl site in Fig 1, left). The He(St') genomic DNA produced two novel fragments of 9.2 kb and 6.5 kb in addition to those bands from the GPA, GPB, and GPE genes. A polymorphic band of 7.3 kb originating from the GPE gene was also seen. The unique Mspl site was mapped to the second intron by separate hybridization with the IVS probes (Fig 1, right; gels not shown). Therefore, the 9.2-kb and 6.5-kb fragments represented the 5' and 3' portions of the GPHe(St') gene, respectively. Because the normal bands showed no change of intensity, the occurrence of the new fragments indicated an expansion in the MNS locus of a hybrid glycoporphin gene. It should be noted that an apparently identical 9.2-kb band also occurs in the GPHe(P2) gene.

Regarding the 6.5-kb band, double digestion with Mspl and Pvu II did not change its apparent size (Fig 1), but resulted in a pattern common to all the genes specifying St' glycoporphin. Taken together, the data suggested that the GPHe(St') gene is a complex genetic variant whose 5' and 3' regions were similar to GPHe and GPS' genes, respectively.

Amplification and nucleotide sequence of GPHe(St') gene. As expected, alterations in the GPHe(St') gene should be confined to the genomic region encompassing the extracellular domain. The 9.2-kb and 6.5-kb single-copy fragments were used as templates for PCR amplification for they together span the sequence from intron I to intron 6 (Fig 1, right). Electrophoresis on agarose gels indicated that each of the five amplified segments was of the predicted size (not shown). After secondary amplification and template purification, all the GPHe(St') gene segments were sequenced to define the location of recombination breakpoints; (2) exclude possible mutations at various splice junctions; and (3) verify the presence of GPA-like sequences downstream from the most distal breakpoint.

Comparison of the nucleotide (nt) sequences indicates the presence of two recombination sites in the GPHe(St') gene (Fig 2). The first site resides in the 5' region across exon II-intron 2 in which a short sequence of GPA had replaced the homologous counterpart of GPB to form a B-A-B configuration. As shown, the proximal B-A breakpoint was located in exon II and defined by nt -20 and 13 (Fig 2A). The sequence upstream of nt -20 was GPB-like, whereas that downstream of nt 13 was GPA-like. This GPA-like sequence was apparently derived from the M allele of GPA because its fifth codon (nt 13-15) is GGT, which specifies a glycine...
Fig 2. Nucleotide sequences surrounding the gene conversion site (A) and unequal crossing-over site (B). (A) The sequencing gel panel showing six untemplated nucleotide changes in the proximal B-A breakpoint region. The GPHe(St') gene is aligned with GPA and GPB to show the converted sequence. An octameric repeat that encloses all the nucleotide changes is denoted by two arrows. The exon-intron junction is marked by an arrowhead. Identical nucleotides are denoted by dashes. The first nucleotide of the coding sequence for mature glycophorin is numbered 1. Solid and open bars denote the GPB- and GPA-like sequences, respectively. (B) The nucleotide sequence encompassing the unequal crossing-over point of the GPHe(St') gene. The 5' splice site (G → T) mutation at the first nucleotide of intron 3 is indicated by a vertical arrow. A unique HindIII site in the GPA gene is shown. The sequence transition from GPB to GPA is illustrated. The inverted repeat (arrows), the A phage attB-like element (boxed), and a polypurine stretch (boxed) in the crossover site are shown.

For mature glycophorin (Fig 2A; see also Fig 5). The distal A-B breakpoint was located in intron 2, and the GPB-like sequence resumed via a shared region at nt 41 downstream from exon II (data not shown). Such a segmental DNA transfer was most likely driven by the mechanism of gene conversion, as observed in the formation of other B-A-B glycophorin hybrid genes.23,24

The DNA sequence also shows that the converted region is clustered with multiple untemplated nucleotide replacements that differ from either parent gene. Concerning the location and nature of these alterations, one took place at the 3' end of intron 1 involving a T → C transition; five others occurred in exon II and included both transitions and transversions (Fig 2A). Of the exonic nucleotide variations, two account for silent substitutions in the third position of serine codons and three result in changes of codons 1 and 4 for mature glycophorin. The latter replacements have not only abolished the coding sequence determining the common M or N blood group polymorphism but created a composite sequence that elicits the expression of the He antigen (Dahr et al25 and see below). Such untemplated variations were apparently induced during repair synthesis of the heteroduplex DNA26,27 because they were also found in the conversion of GPHe(P2) and its related genes.13,15

The second site of recombination resides in intron 3 and points to an unequal crossover from the GPB gene to the GPA gene (Fig 2B). Comparison with the parent sequences shows that this portion of the GPHe(St') gene is arranged in a B-A configuration, and the crossing-over point lies between nt 110 and 246. Significantly, the same sequence has been mapped as the crossing-over site for the B-A hybrid GPSt type A gene.13 Although the sequence upstream of nt 110, including the mutant donor splice site, was GPB-like, the pseudoexon sequence of GPHe(St') differed from that of GPB by one G deletion and one A → G transition (Fig 2B). The sequence downstream from nt 246 was almost identical to the GPA gene. Further sequencing of exons IV and V confirmed that the remainder of GPHe(St') was composed of the GPA sequences (data not shown). Together, the observed sequence data is consistent with the occurrence of a complex B-A-B-A hybrid gene.

Surrounding the two recombination sites, several genomic
motifs were identified. In the first site, a perfect inverted repeat was found to flank the upstream breakpoint region and enclose all the templated and untemplated nucleotide changes (Fig 2A). The converted region is A,T-rich and exhibits strong local strand asymmetry. The structural elements that characterize the second recombination site have been described previously (Fig 2B).

Composition and nucleotide sequence of GPHe(St*) transcripts. The composition of glycophorin gene transcripts in the He+St(a+) erythroid cells was analyzed by reverse transcription of total reticulocyte RNA and PCR amplification. Figure 3 shows the resulting glycophorin cDNA products resolved by agarose gel electrophoresis. In the control, one major band (478 bp) was apparent for GPA cDNA. In the He(St*), three cDNA bands of decreasing size (478, 382, and 343 bp) were seen (lane 2); they were re-electrophoresed after purification and secondary amplification (lanes 3 through 5).

Figure 5 shows the nucleotide and deduced amino acid sequences of the two shortened GPHe(St*) cDNA species. The 382-bp cDNA product encodes a mature glycophorin of 99 amino acid residues typical of GPsSt*, except for a stretch of the first five residues at the extreme amino terminus (Fig 5, upper section). In this cDNA form, the 92-bp se-

![Fig 3. Amplification of the glycophorin cDNA products by RT-PCR. Total RNA from peripheral reticulocytes was reverse transcribed into cDNA and then amplified by PCR. The amplification products were analyzed by agarose gel electrophoresis and stained with ethidium bromide. 4x174 DNA size markers are shown at left (lane M). In the control, a major band (478 bp) expected for GPA was present (lane 1). In He(St*), three cDNA bands of decreasing size (478, 382, and 343 bp) were seen (lane 2); they were re-electrophoresed after purification and secondary amplification (lanes 3 through 5).](image-url)
Fig 5. Nucleotide and deduced amino acid sequences of GPHe(St) cDNAs. GPHe(St') (II-IV), the exon III-deleted 382-bp cDNA species; GPHe(St) (II-V), the 343-bp cDNA species in which both exons III and IV are deleted. The exon-exon junctions are indicated by arrowheads. Deletion of exon sequences is denoted by a dashed line. Sequences for the He and St' epitopes are underlined. A comparison of the M-, N-, and He-defining sequences is shown at bottom.

The nucleotide sequence also shows that both exons III and IV could be excised from the GPHe(St') primary transcript to yield a further shortened mRNA species. Thus, the 343-bp cDNA product should retain an exon II-V junction and encode a mature glycoporphin of 86 amino acid residues (Fig 5, lower). This protein would contain an He-like sequence at its amino terminus but lack the St'-specific sequence because of the deletion of exon IV. In addition, an amino acid change from G to E at position 27 was apparent as a result of aberrant splicing.

In this study, a complex hybrid gene encoding GPHe(St') of the human erythrocyte membrane was characterized. We presented evidence that the GPHe(St') gene retains multiple recombinational breakpoints and occurs as a B-A-B-A fusion hybrid. To our knowledge, this represents the first human structural gene with three breakpoints that can be clearly defined in a stretch of 1-kb genomic sequence. As shown by DNA sequencing, the B-A-B breakpoints reside in the 5' region across exon 11-intron 2 in which a number of templated and untemplated nucleotide replacements characteristic of a GPHe gene are clustered. The distal B-A crossing-over point lies in the third intron at a recombination site.

Immunoblot analysis of glycoporphins from He + St(a+) erythrocytes. Compared with normal blots, the He(St') blots showed the presence of one major and one minor shortened glycoporphin whose molecular sizes were between GPA and GPB (Fig 6). The major species was GPHe(St') (II-IV) encoded by the exon III-deleted mRNA whose mobility was typical of all St' glycoporphins. The electrophoretic behavior of the minor species was as expected for GPHe(St') (II-V). Both variant proteins contained the carboxy terminal region of GPA, as shown by a positive reaction with α PC antiserum (Fig 6). Although the minor glycoporphin retains a novel junction sequence, its antigenic specificity is not yet established. An M-specific glycoporphin variant of similar size was identified in the Mz(Sta) erythrocytes.

**DISCUSSION**

In this study, a complex hybrid gene encoding GPHe(St') of the human erythrocyte membrane was characterized. We presented evidence that the GPHe(St') gene retains multiple recombinational breakpoints and occurs as a B-A-B-A fusion hybrid. To our knowledge, this represents the first human structural gene with three breakpoints that can be clearly defined in a stretch of 1-kb genomic sequence. As shown by DNA sequencing, the B-A-B breakpoints reside in the 5' region across exon 11-intron 2 in which a number of templated and untemplated nucleotide replacements characteristic of a GPHe gene are clustered. The distal B-A crossing-over point lies in the third intron at a recombination site.
Fig 6. Immunoblot analysis of glycophorin from the He+St(a+) erythrocytes. Nitrocellulose blots of normal and He+St(a+) RBC lysates were probed with polyclonal antibodies against aGP (the GPA antiserum) and aPC (an antiserum to the carboxy terminal residues 82-131 of GPA). Molecular weight markers and band origins are denoted. A2 and A1, GPA dimer and monomer; B2 and B1, GPB dimer and monomer; AHS, GPA and GPHe(St*) heterodimer; HS, and HS1, GPHe(St*) dimer and monomer; HS(II-V)1, the dimer band of the shortened glycophorin lacking the internal sequence encoded by exons III and IV.

apparently identical to that observed in the B-A hybrid GPSt type A gene. Thus, the GPHe(St*) gene can be viewed as a composite of the GPHe and GPSt genes.

By cDNA analysis we demonstrated the presence in the He + St(a+) reticulocytes of three mRNA species, one of which codes for GPA and two, of smaller size, that arise from the GPHe(St*) pre-mRNA. Direct sequencing of the two shortened cDNA species showed they were products of exon skipping. The exon III-deleted form contained sequences encoding both He and St* antigens, whereas the further truncated form possessed the He-defining sequence only. These aberrant splicing events removed from the GPHe(St*) primary transcript two breakpoints, one in intron 2 and the other in intron 3. Therefore, the mature mRNAs are in the form of B-A hybrid, although their cognate gene is arranged in a B-A-B-A configuration.

With respect to recombinational diversification of glycophorin genes, the exchange of homologous domains and subsequent fusion of parent strands generally proceeded with high fidelity.9,12,23,24 The identification of six untemplated replacements in the 5' region of GPHe(St*) raises an important question regarding the molecular basis for their origin. Several lines of evidence favor the hypothesis that these replacements have arisen from heteroduplex repair-induced alterations on gene conversion rather than by the mechanism of independent spontaneous point mutations. (1) As the intrinsic feature of gene conversion, the formation and repair synthesis of heteroduplex DNA is not an error-free process;29 occasionally, untemplated mutation(s) may be introduced into the cognate sequence if the template-directed mismatch repair has failed.29 (2) The untemplated variations, as such, do not occur at random; instead, they tend to be tightly associated with or clustered around the conversion site. This
observation has also been noted in the generation of GPHe(P2), GPIMl, and GPMLX genes. Similar untemplated mutational events have been known to take place in the Ig and major histocompatibility complex gene systems. In summary, although this kind of untemplated changes is largely unpredictable in contrast to templated sequence transfers, it provides a novel means for further diversification of glycophorin antigens.

As noted, the St* epitope is specified by the sequence of the exon II-exon IV junction. Whereas exon IV must be contributed by the GPA gene, exon II could be derived from either normal or variant genes. Thus, any mechanism leading to the creation of such a junction sequence should result in the St* antigen. In fact, previous studies have identified multiple St* genetic isoforms, which include GPSt*, GPM*St*), and GPErik(St*) genes. GPM*St* is a B-A hybrid present in three types (designated A, B, and C) that differ in the location of unequal recombination sites. GPM*St* is an A-B-A hybrid product formed via the mechanism of gene conversion. In contrast, GPM*St* carries a spontaneous G → A mutation in the terminal nucleotide position of GPA exon III that affects pre-mRNA splicing. The identification of GPHe(St*) as a B-A-B-A hybrid further pinpoints the molecular heterogeneity underlying the St* phenotypic expression. Although its molecular basis is different, GPHe(St*) is similar to other hybrids in having acquired the same defective donor splice site from the GPB gene. Moreover, it shares with others, including GPErik(St*), a common scheme of pre-mRNA splicing. That leads to the maturation of an exon III-deleted mRNA and its subsequent translation into an St*-bearing glycophorin. Figure 7 compares the protein structures of GPHe(St*), GPHe, and other St* isoforms so far characterized.

Previous studies have shown that the majority of glycophorin variants are hybrids resulting from either gene conversion or unequal homologous recombination. In the case of gene conversion, all the recombinant products are simple convertants, each being flanked by two breakpoints. The GPHe(St*) gene is the first convertant associated with an additional crossover event. Thus, the process of its formation embodies the features of two mechanisms for homologous recombination. In Figure 8, a model is presented to account for the B-A-B-A configuration observed in the GPHe(St*) gene. Whether the gene conversion and unequal crossing-over events are associated or are independent of each other cannot be defined unequivocally. However, based on the following observations, we consider that the mechanisms leading to GPHe(St*) occurred in two separate recombinational events. (1) As two pre-existing variants, both GPHe and GPSt* type A genes have been encountered in the black population. (2) The GPHe(St*) gene is a compound variant of GPHe (B-A-B hybrid) and GPSV (B-A) hybrid.
hybrid product of genome expansion, whereas gene conversion-associated crossover would not necessarily cause a change in gene number. In conclusion, if the two proposed recombination events took place independently, the GPHe(St') gene would, in fact, represent a very rare genetic variant.

ACKNOWLEDGMENT

We thank Ying Chen for her excellent technical assistance and Emma Fay Ricksy for her expertise in the preparation of this manuscript. We also acknowledge the Oligonucleotide Synthesis Facility supported by the Cancer Center Grant SP30 CA 13330.

REFERENCES

10. Huang C-H, Blumenfeld O0: Identification of recombination events resulting in three hybrid genes encoding human MIV, MiV(JL), and St' glycoporphins. Blood 77:1813, 1991
14. Lomas C, Daniels GL, Green CA, Kelly L, Moulds M: The first He(+) St(a+) donor; immunoblotting determined the presence of He and St' on one novel glycoporphin molecule. Transfusion 32:548, 1992
20. Adronov E, Monnig D, Ricksy EM: For personal use only.
Glycophorin He(Sta) of the human red blood cell membrane is encoded by a complex hybrid gene resulting from two recombinational events

CH Huang, C Lomas, G Daniels and OO Blumenfeld