J Mol Evol (1995) 41:478–486

JOURNAL OF MOLECULAR EVOLUTION



# Sequence Diversification and Exon Inactivation in the Glycophorin A Gene Family from Chimpanzee to Human

**Cheng-Han Huang,[1] Shen-Si Xie,[2] W. Socha,[3] Olga O. Blumenfeld[2]**

[1] Lindsley F. Kimball Research Institute, New York Blood Center, 310 East 67th Street, New York, NY 10021, USA
[2] Department of Biochemistry, Albert Einstein College of Medicine, 1300 Morris Park Avenue, Bronx, NY 10461, USA
[3] Laboratory for Experimental Medicine and Surgery in Primates (LEMSIP), New York University School of Medicine, RR #1 Long Meadow Road, Tuxedo, NY 10987, USA

Received: 20 December 1994 / Accepted: 25 January 1995

**Abstract.** In humans, the allelic diversity of MNSs glycophorins (GP) occurs mainly through the recombinational modulation of silent exons (pseudoexons) in duplicated genes. To address the origin of such a mechanism, structures of GPA, GPB, and GPE were determined in chimpanzee, the only higher primate known to have achieved a three-gene framework as in humans. Pairwise comparison of the chimpanzee and human genes revealed a high degree of sequence identity and similar exon-intron organization. However, the chimpanzee GPA gene lacks a completely formed M- or N-defining sequence as well as a consensus sequence for the Asn-linked glycosylation. In the case of the GPB gene, exon III is expressed in the chimpanzee but silenced, as a pseudoexon, in the human. Therefore, the protein product in the chimpanzee bears a larger extracellular domain than in the human. For the GPE genes, exon III and exon IV have been inactivated by identical donor splice-site mutations in the two species. Nevertheless, the chimpanzee GPE-like mRNA appeared to be transcribed from a GPB/E composite gene containing no 24-bp insertion sequence in exon V for the transmembrane domain. These results suggest a divergent processing of exonic units from chimpanzee to human in which the inactivation of GPB exon III preserved a limited sequence repertoire for diversification of human glycophorins.

*Correspondence to:* O.O. Blumenfeld

**Key words:** Human — Chimpanzee — Glycophorins — MNSs blood-group system — Exon activation-inactivation

## Introduction

In humans, the *GYPA* locus on the long arm of chromosome 4 determines the expression of MNSs blood-group antigens on the erythrocyte membrane (Huang and Blumenfeld 1995; Cartron et al. 1990). As a small gene cluster, the locus consists of three glycophorin (GP) genes that are arranged in the physical order of 5′-GPA-GPB-GPE-3′. These genes are structurally similar and appear to have evolved from a common ancestral GPA gene through repeated *Alu-Alu* recombinations. It was proposed that these recombination events first resulted in the formation of a GPB/E progenitor and were followed by duplication, finally resulting in GPB and GPE genes (Rearden et al. 1993; Onda et al. 1993, 1994). While the GPA gene retains seven functional exons, the GPB and GPE genes both lack the exons (VI and VII) for the cytoplasmic tail and differ in the expression of exons (II-IV) for the extracellular domain. The third exon of GPB and the third and fourth exons of GPE are silenced by point mutations at the first position of the invariant GT-dinucleotide motif of their donor splice sites (Huang and Blumenfeld 1995; Cartron et al. 1990).

In the animal kingdom, only nonhuman primates, in particular the anthropoid apes or some species of Old

World monkeys, express the MNSs-related surface antigens (Socha and Moor-Jankowski 1979). Western blotting and peptide analyses have demonstrated the presence of GPA and GPB homologues in the chimpanzee (Rearden 1986; Lu et al. 1987b; Blumenfeld et al. 1983). Recent genomic studies showed that the GPA gene occurs in all anthropoid apes, whereas the GPB gene is only present in chimpanzee and gorilla but not orangutan or gibbon (Rearden et al. 1993, 1990a,b; Lu et al. 1990). The presence or absence of GPE parallels the distribution of GPB in all species except the gorilla—that is, not all gorillas retain a GPE gene. Thus, in this respect, the chimpanzee is closest to the human, having fully achieved a three-gene framework in the process of gene duplication (Rearden et al. 1993). Nevertheless, compared with the human counterparts, the structure, organization, and diversity of glycophorin genes among nonhuman primates are still largely unknown.

The three glycophorin genes occur in all segments of human populations and exhibit a large array of genetic variations. Elucidation of the molecular basis for many variant forms demonstrated that homologous recombinations between the GPA and GPB genes are the major mechanism resulting in the antigenic diversity of glycophorins (Huang and Blumenfeld 1995). The formation of various hybrid genes was shown often to involve the exchange or transfer of silent exons (pseudoexons) and modulation of donor splice sites. Therefore, novel interexon and intraexon junction sequences could be created, resulting in diversity at the level of primary protein structures. To address the origin of such a mechanism and determine whether it is unique to humans or could be predated to nonhuman primates, we characterized the transcripts and relevant genomic regions of glycophorin genes in the chimpanzee. Compared with the human counterparts, the chimpanzee glycophorins show a high degree of sequence conservation in both coding and noncoding regions. Nevertheless, the exonic sequences encoding the extracellular domain had undergone a significant divergence from chimpanzee to human. Among the major interspecies variations are the differential expression of exon III in the GPB gene and the presence or absence of a GPE transcript from a progenitor gene.

## Methods

*Blood Samples and Nucleic Acid Isolation.* Whole peripheral blood specimens from chimpanzees were freshly drawn at the Laboratory for Experimental Medicine and Surgery (LEMSIP), New York University Medical Center, Tuxedo, NY. Serologic typing was carried out with a panel of reagents specific for human M,N and simian extension of M,N, V-A-B-D blood groups (Lu et al. 1987b). To focus on interspecies differences, one chimpanzee (Pierre) was randomly chosen for detailed analysis. The blood-group phenotypes of Pierre's erythrocytes were M+, N (weak), $N^V$, and V.A. Additional chimpanzees were included when exon V genomic sequences of glycophorin genes were examined. Human peripheral bloods used as controls were freshly drawn from laboratory personnel.

Total RNAs were prepared from hemolysates of peripheral reticulocytes using a differential cell lysis method (Goossens and Kan 1981), with minor modifications (Huang et al. 1993). High-molecular-weight genomic DNAs were isolated from unlysed leukocytes by SDS solubilization and proteinase K digestion (Huang et al. 1993).

*DNA Probes and Synthetic Primers.* The cDNA probe for human GPA was 478 bp in length and was derived from reticulocyte total RNA by the reverse-transcriptase-coupled polymerase chain reaction (RT-PCR). The two genomic probes were IVS 2 for intron 2 (0.7 kb) and IVS 3 for intron 3 (0.9 kb); they were generated from a human GPA clone (Huang et al. 1991) and cross-hybridized to all three glycophorin genes with equal efficiency.

Oligonucleotides used as primers for reverse transcription and amplification of chimpanzee glycophorin cDNAs and genomic DNAs are listed in Table 1. All but four were chosen from the nucleotide sequence of human GPA, GPB, and GPE genes, and some had been used in the mapping of chimpanzee homologous genes (Lu et al. 1990). These primers were produced on an automated DNA synthesizer (Applied Biosystem) and purified by denaturing 15% polyacrylamide gel electrophoresis (PAGE).

*Southern Blot Analysis.* Chimpanzee and human genomic DNAs were digested with the restriction enzyme *Sac*I and analyzed by Southern transfer and blot hybridization as described (Huang et al. 1993). DNA probes were labeled with [$\alpha$-$^{32}$P]-dCTP by the random primer extension method (Feinberg and Vogelstein 1984).

*Synthesis and Amplification of Glycophorin cDNAs.* RT-PCR analysis of chimpanzee glycophorin cDNAs was carried out using the AMV reverse transcriptase (Promega) and 3′ primers specific for human GPA, GPB or GPE (Table 1). The procedure was as described previously (Huang et al. 1993). Two to 5 μg of total RNA mixed with 250 ng of a 3′ primer was incubated at 65°C for 5 min and then annealed on ice for 5 min. To this solution was added 2 mM dNTP, 1× reaction buffer, 20 units of RNase inhibitor, and 20 units of RT, all in a final volume of 20 μl, and the mixture was incubated at 42°C for 60–90 min. The resultant cDNA was then directly transferred to a pre-made PCR solution containing both primers and *Taq* DNA polymerase (Perkin-Elmer). Amplification was carried out for 30 cycles as follows: 94°C for 1 min, 55°C for 1.5 min, and 72°C for 1.5 min. The final chain extension step was at 72°C for 7 min.

The cDNA amplification products were analyzed by 1.8% agarose gel electrophoresis and purified by native 4.5% PAGE. The bands of interest were cut from the gel, eluted in 1 × TE buffer (10 mM Tris-HCl, 0.1 mM EDTA, pH 8.0) for 6 h, and recovered by centrifugation. When necessary, aliquots were reamplified under identical conditions using the same set of primers.

*Amplification of Glycophorin Genomic Sequences.* PCR amplification of individual glycophorin genes was achieved by selection of specific primers and/or use of unique DNA templates derived from *Sac*I digestion (Fig. 1). To generate the templates, the *Sac*I restriction fragments of chimpanzee genomic DNA were fractionated by agarose gel electrophoresis, identified by superimposable Southern blots, cut from the gel, and electroeluted, as described (Huang and Blumenfeld 1991a). Amplification focused on the genomic region encompassing the extracellular and transmembrane domains. The strategy employed is illustrated in Fig. 1. For each genomic segment, 30 cycles of PCR were carried out using conditions as described above. The amplified DNA products were analyzed by agarose gel electrophoresis and purified on native 4.5% PAGE.

**Table 1.** Synthetic oligonucleotide primers used for reverse transcription and PCR amplification

| Sequence | X-mer | Strand | Location | |
|---|---|---|---|---|
| I. cDNA synthesis and amplification | | | | |
| a 5′-GTATGGAAAAATAATCTTTGTATTAC-3′ | 26 | Sense | Exon I | (HGPA.B.E.) |
| b 5′-AGCATATCAGCATCAAGTACCACT-3′ | 24 | Sense | Exon II | (HGPA.B.E.) |
| c 5′-ATCACTTGTCTCTGGATTTTCTATTTC-3′ | 27 | Antisense | Exon VI–VII | (HGPA) |
| d 5′-TCCACATTTGGTTTGGTGAACAGATTC-3′ | 27 | Antisense | 3′-UT | (HGPA) |
| e 5′-CATAAAGAAGAGAACAGCAGGTGCAG-3′ | 24 | Antisense | 3′-UT | (HGPB) |
| f 5′-GGAGTTAGGATAGCCAAGGGTT-3′ | 22 | Antisense | 3′-UT | (HGPE) |
| g 5′-AGAATACAGTAATAGTGAGGCAG-3′ | 23 | Antisense | 3′-UT | (HGPB) |
| II. Genomic DNA amplification | | | | |
| h 5′-GTTAGAACTGTTTACCCTCCAGAAGAG-3′ | 27 | Sense | Exon III | (HGPA) |
| i 5′-ATGGACAAGTTGTCCCGTTTCTCC-3′ | 24 | Antisense | Exon IV | (HGPB) |
| j 5′-TATTAGCTCAGAGCCTCACACATT-3′ | 24 | Sense | IVS 2 | |
| k 5′-CCCTCAGTTATGAGACAATTTGCT-3′ | 24 | Sense | IVS 3 | |
| l 5′-AGAACTGTCATGAGTTACAGCTCGT-3′ | 25 | Antisense | IVS 3 | |
| m 5′-TGGTCATTTATTTCAGACTTTCAT-3′ | 24 | Sense | IVS 4 | |
| n 5′-GCAATGGATAGTTTAAAATGGAATGAC-3′ | 27 | Antisense | IVS 4 | |
| o 5′-CTGTTTCTCTTTTGAGTTTAACTG-3′ | 24 | Antisense | IVS 5 | |
| p 5′-ACTGTTTACCCTCCAGAAGAGGAA-3′ | 24 | Sense | Exon III | (CGPA) |
| q 5′-CTGGTTCAGAGAAACGATGGACAA-3′ | 24 | Antisense | Exon IV | (CGPA) |
| r 5′-ACTGTTTACCCTCCAGAAGAGGAT-3′ | 24 | Sense | Exon III | (CGPB) |
| s 5′-CTGGTCTAGAGAAAGGATGGACAG-3′ | 24 | Antisense | Exon IV | (CGPB) |

*Determination and Analysis of Nucleotide Sequences.* Nucleotide sequences of gel-purified cDNAs or genomic DNA products were directly determined on both strands using either a manual or automatic procedure. The manual procedure was a modified version of Sanger's chain termination method (Sanger et al. 1977) using Sequenase (US Biochemicals) and 5′-end $P^{32}$-labeled primer. Automatic sequencing was carried out on a DNA sequencer (model 373A, Applied Biosystems) using fluorescent dye terminators. The resultant nucleotide sequences were compiled and analyzed by a GCG computer program.

*Western Blot Analysis of Chimpanzee Glycophorins.* Membranes were prepared from chimpanzee and human erythrocytes and 2.5 to 5 μl was used for SDS-15% PAGE, as described (Lu et al. 1987a). Proteins were transferred onto nitrocellulose filter by electroblotting. Immunoblots were developed with polyclonal antisera to human glycophorins (anti-GP) and to the carboxy-terminal region (residues 82-131) of GPA (anti-PC), and visualized by autoradiography with $^{125}$I-protein A (New England Nuclear).

## Results

### *Genomic Southern Blot Analysis*

In previous studies we have shown that the *Sac*I restriction-site pattern of chimpanzee glycophorin genes is relatively invariant among individuals and is nearly identical to that of human DNA (Lu et al. 1990). Mapping with a number of gene-specific oligonucleotide probes assigned the various *Sac*I restriction fragments to GPA, GPB, and GPE.

Figure 1 shows the *Sac*I genomic blots of chimpanzee Pierre and of a human control. In the two species all but one fragment were identical in size. Compared with the human, the chimpanzee lacked the 8.0-kb band but retained a unique 4.4-kb fragment. All bands were responsive to the IVS2 and/or IVS3 probe except the 9.6-kb band encompassing the last two exons of GPA. In the chimpanzee, the IVS2 probe hybridized with the 2.7-, 4.4-, and 6.4-kb bands, whereas the IVS3 probe hybridized with the 3.1- and 6.4-kb fragments (not shown). It became evident that the chimpanzee GPB gene retains a *Sac*I cleavage site in exon III which is absent from the human GPB gene.

### *RT-PCR Analysis of Glycophorin Transcripts*

To characterize the transcripts of chimpanzee GPA, GPB, and GPE genes, cDNAs were obtained by RT-PCR amplification. To generate gene-specific products, several 3′ primers that could discriminate among the three transcripts in the human were tested and their suitability was judged by productive amplifications. Figure 2 shows a typical agarose gel electrophoresis of the primary cDNA amplification products. The use of primers "a" and "c" or "d" gave rise to a single major band expected for GPA, whereas the use of primer "a" with primers "e," "f," or "g" resulted in two cDNA species; the shorter one was comparable to human GPE and the longer one was larger than human GPB. Restriction digestion of these cDNAs with *Sac*I confirmed the occurrence of a unique *Sac*I site in the coding sequence of GPA and GPB but not in the GPE gene (not shown). The ultimate assignment of their gene identity was established by nucleotide sequence homology to the human counterparts. (See below.)

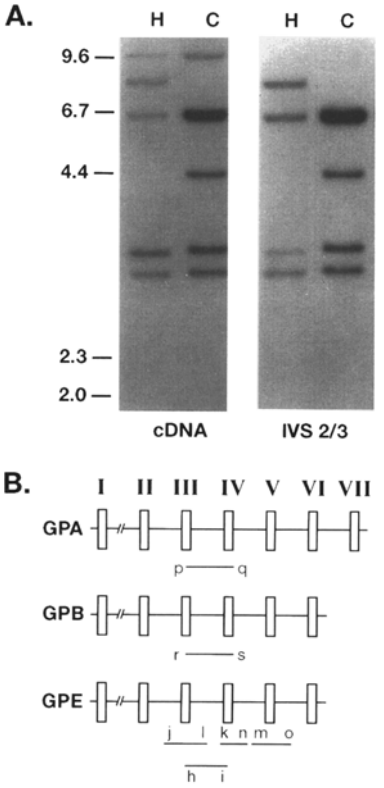



**Fig. 1.** **A** Southern blots of *Sac*I digests of human and chimpanzee genomic DNA probed with human GPA cDNA and IVS2/3. Gene origin of fragments is as follows: 9.6 kb—3′ GPA (human [H] and chimp [C] 8.0 kb—GPB (H); 6.4 kb—GPE (H, C); 4.4 kb—5′ GPB (C); 3.1 kb—middle GPA (H, C); 3′ GPB (C); 2.7 kb—5′ GPA (H, C) (Lu et al. 1990; Huang et al. 1991). **B** Schematic representation of genomic structures of GPA, GPB, and GPE and approximate location of amplified fragments (*straight lines*). Exons and introns are represented by *boxes* and *connecting lines*. Amplification primers (*small letters*) are shown in Table 1. Amplification templates were, respectively, the 3.1-kb *Sac*I fragment for amplification of exons III of GPA and GPB and the 6.4-kb *Sac*I fragment for all GPE segments.

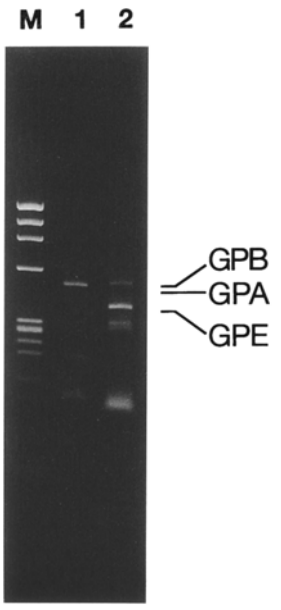



**Fig. 2.** Products of reverse transcription and PCR primary amplification for chimpanzee glycophorin genes. *M*, ψX174 DNA size markers. *Lane 1*—GPA, primers "a" and "d"; *lane 2*—GPB (483 bp) and GPE (348 bp), primers "a" and "f."

### *Nucleotide Sequences of Chimpanzee Glycophorin cDNAs*

The nucleotide and deduced amino acid sequences for chimpanzee GPA, GPB, and GPE are shown in Fig. 3. Significant features of each cDNA compared to the respective human cDNAs are as follows:

GPA

The chimpanzee and human homologues bear 97% nucleotide sequence identity and 90% amino acid identity. The number of codons and the size of corresponding exons are identical, indicating a quite similar, if not identical, exon-intron organization. The amino acid sequence divergence is mainly confined to the extracellular domain of the two proteins, while those sequences for the leader peptide, the transmembrane segment, and cytoplasmic tail are largely conserved. This observation suggested that, relative to other exons of the GPA gene, exons II–IV evolved with an increased divergence rate from chimpanzee to human.

In the human, the codons that differentiate the M and N blood-group epitopes are TCA (ser) and GGT (gly) for M, and TTA (leu) and GAG (glu) for N, at positions corresponding to residues 1 and 5 of the mature glycophorin. The corresponding codons in chimpanzee GPA—namely, TCA and GAG—should result in a hybrid M-N epitope. Significantly, the chimpanzee studied here was characterized serologically as M and N (weak). Furthermore, human GPA bears the single N-linked carbohydrate unit attached to the sequon $^{26}$Asn-Asp-Thr$^{28}$ (Huang and Blumenfeld 1995), but chimpanzee GPA lacks this consensus sequence at the corresponding position ($^{26}$Ser-Asp-Lys$^{28}$). Despite the above changes, codons 41–52 of exon III, occurring as the complex palindrome of a recombination hotspot (Huang and Blumenfeld 1991b), have been conserved with two nucleotide substitutions.

GPB

The chimpanzee and human homologues share 96% and 88% sequence identity at the nucleotide and amino acid levels, respectively. They both lack the exons for the cytoplasmic tail, in agreement with immunoblotting data. (See below.) Nevertheless, chimpanzee GPB is similar to chimpanzee GPA, retaining a hybrid M-N epitope sequence. Most importantly, it contains a functional exon III, thus encoding 104 amino acids with an extended extracellular domain. In human GPB, this exon is silenced by a G-T transversion at the GT dinucleotide of the donor splice site (Kudo and Fukuda 1989). Because of the expression of exon III and the occurrence of an A-T transversion at codon 61, chimpanzee GPB does not

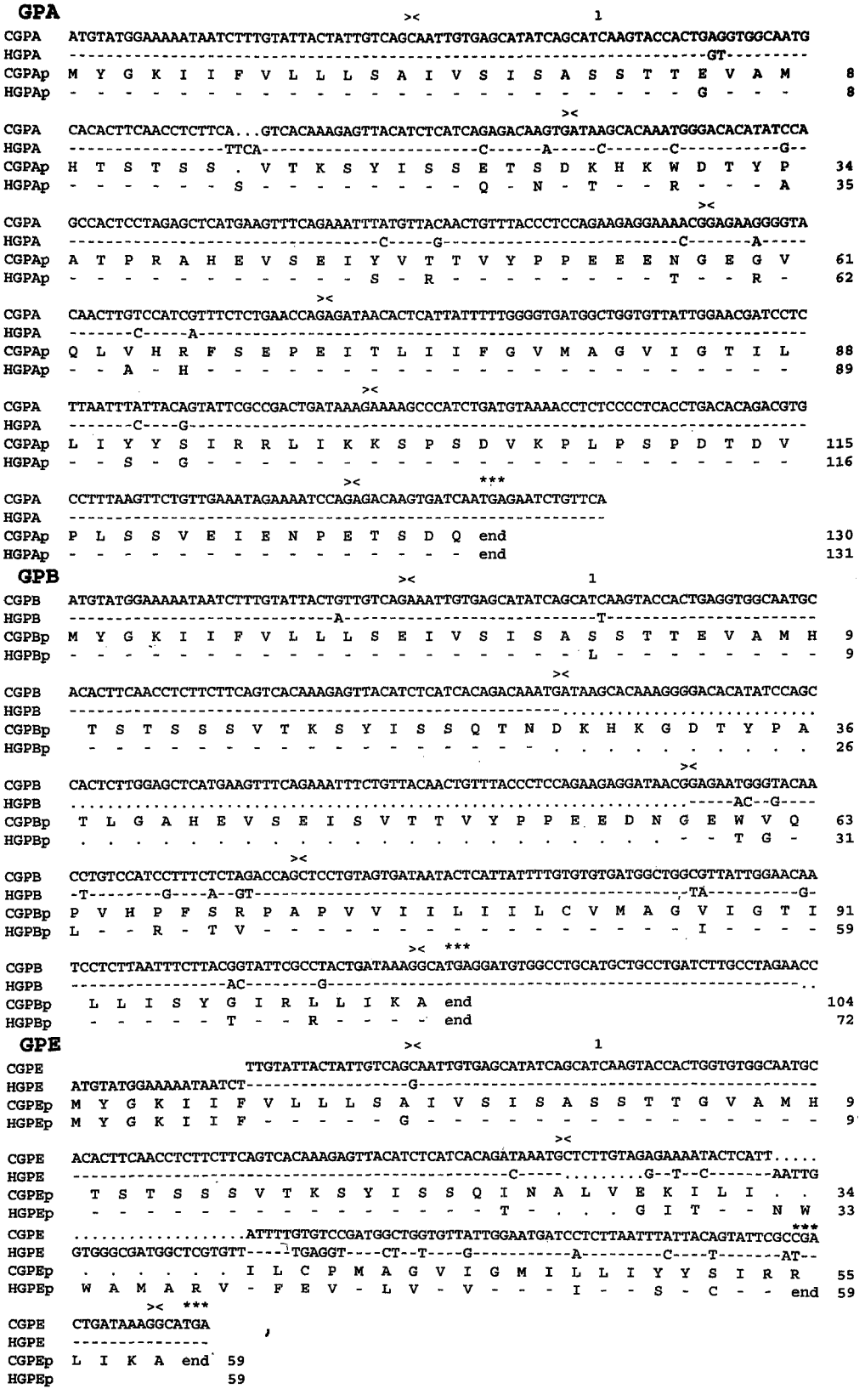



**Fig. 3.** The cDNA and deduced amino acid sequences of chimpanzee GPA, GPB, and GPE. Coding regions of corresponding exons are indicated by *brackets*. Unexpressed exon in GPB is shown as *dots*. Sequences of human glycophorins are shown for comparison. ($GPA^M$ is shown.) A dash indicates an identical nucleotide or amino acid. Codon *numbers* are indicated on the left, with codon *number* 1 indicated above the nucleotide sequence.

display the S or s blood-group antigen. Nevertheless, the complex palindrome (codons 42–53) is present in exon III of the chimpanzee, and its sequence is identical to the corresponding sequence within human pseudoexon III.

GPE

The chimpanzee GPE gene is similar to human GPE in that it contains an M-specific sequence and lacks the expression of exons III and IV. However, chimpanzee GPE carries a 9-bp in-frame insertion at the exon II–V junction but no 24-bp in-frame insertion within exon V, structural features that are shared with the GPB gene. Although chimpanzee and human GPE cDNA share 94% nucleotide identity, their amino acid identity is only 64% because of the above changes. Those results suggested that the chimpanzee GPE mRNA was derived from a GPB/E composite gene.

Patterns of Exon Inactivation and Insertion in Glycophorin Genes

To define the patterns of exon inactivation and insertions in glycophorin genes between the two species, genomic regions encompassing exons III–V were amplified and sequenced. Figure 1 shows the strategy for genomic amplification. The sizes of all chimpanzee fragments were comparable to those of the human (not shown). Sequencing of additional intronic regions revealed over 95% sequence identity of the three glycophorin genes (not shown).

Chimpanzee GPA and GPB exons III are flanked by a functional donor splicing signal G/gtatgtttct conforming to the primate splicing consensus sequence (Shapiro and Senapathy 1987). This explains the appearance of the exon III sequence in the GPB transcript. In human GPB, the G-to-T mutation at the donor splice site uniformly excludes its expression. This exon inactivation represented a major evolutionary change from chimpanzee to human.

The chimpanzee GPE-like transcript lacks exon III and exon IV sequences (Fig. 3). To determine the mechanism for this apparent exon skipping event, the nucleotide sequences encompassing the two exons were obtained (Fig. 4). Comparison with the corresponding human sequences showed remarkable sequence similarity between the two species including identical G-to-A transitions in the donor splice sites (Kudo and Fukuda 1990). These point mutations have silenced the two exons and apparently could be predated to nonhuman primates. Further, it is important to note that the complex palindrome present in exons III of human and chimpanzee GPA and GPB (Huang and Blumenfeld 1991b) also appears in pseudoexon III of GPE with two nucleotide substitutions.

To characterize the structure of GPE exon V, two primers close to the exon/intron junctions and the GPE gene-specific template (*Sac*I 6.4 kb fragment) were used for amplification. This allowed the resolution of genomic fragments with small size difference caused by the insertion sequence (Fig. 5). As shown in Fig. 5B, the nucleotide sequence of the larger fragment retained the 24-bp insertion sequence analogous to human GPE (Kudo and Fukuda 1990), whereas that of the smaller fragment retained the 9-bp insertion sequence at the 5′ end of exon V, being identical with the corresponding sequence in the chimpanzee GPE transcript (Fig. 3). The t-to-g transversion at the 3′ end of intron 4 created an acceptor splice site and thus explained the expression of the 9-bp sequence. Taken together, the results suggest that the chimpanzee examined here has two different forms of the GPE gene on homologous chromosomes, and the transcript we isolated was most likely derived from the GPB/E composite gene.

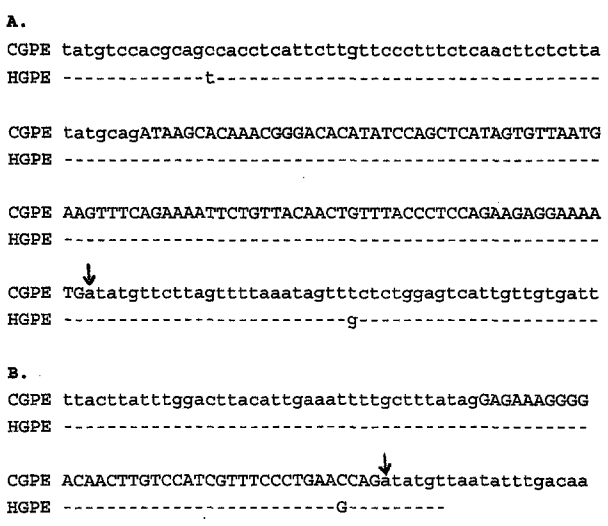
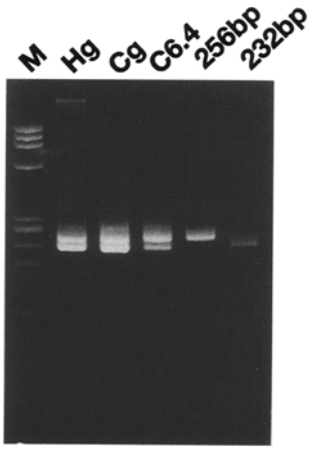



**Fig. 4.** Nucleotide sequence of unexpressed pseudoexons III (**A**) and IV (**B**) and their flanking regions in the GPE gene of human and chimpanzee. *Lower-* and *uppercases* indicate intron and exon sequences, respectively. *Arrows* point to the altered nucleotide within the 5′ splicing consensus sequence. The template for nucleotide sequencing was obtained by PCR amplification of genomic DNA with primers "k" and "n" (Table 1). The sequence for corresponding region of human GPE in **A** was obtained in this laboratory (unpublished) and for that in **B** was from Kudo and Fukuda (1990).

*Immunoblot Analysis of Chimpanzee Glycophorins*

Glycophorins of the erythrocyte membrane were examined by immunoblotting (Fig. 6). The band pattern is typical of that seen previously for chimpanzees with the V.A. Simian blood-group phenotype (pattern II) (Lu et al. 1987b) and therefore is consistent with the phenotype of the subject examined here.

The monomer of GPA shows slightly increased mobility relative to human GPA; this is consistent with a lack of N-linked glycosylation. The GPB monomer band is clearly seen showing a positive reaction with the glycophorin antiserum but, as expected, lack of reaction with the antiserum to the carboxyl-terminal peptide of GPA. In agreement with the deduced amino acid sequence (Fig. 3), chimpanzee GPB has a larger size than




```
                                                    _____insertion _____
CGPE 256bp   ctgctctctttatctcctgtagtgataacactcattattttgtgggtgatagctcgtgttatttt
CGPE 232bp   ----------g---t-----a--a--t-------.....................-----
CGPE cDNA                 ---T-----A--A--T-------.....................-----

CGPE 256bp   tgagatgacggctgttgttactggagtgatcctcttaatttgttactgtattcacctacagataa
CGPE 232bp   GTGTCCGATG----G-----T----A---------------A----A------G--G--T-----
CGPE cDNA    GTGTCCGATG----G-----T----A---------------A----A------G--G--T-----

CGPE 256bp   tggtgagaattcagtgtttaattttgctctaaataccagtgtgaacagctccaggagggttt
CGPE 232bp   A-------------t---------------t-------------------t---------a
CGPE cDNA    A-
```

**Fig. 5.** **Upper** Products of amplification of GPE exon V and its flanking regions using primers "m" and "o" (Table 1) visualized on a 4% agarose gel stained with ethidium bromide. *M,* ψX174 DNA size marker; *lanes 2–4,* primary amplification products using as templates human genomic DNA (*Hg*), chimpanzee (Pierre) genomic DNA (*Cg*), and chimpanzee (Pierre) 6.4-kb *Sac*I genomic DNA fragment (*C6.4* kb); *lanes 5 and 6* show the purified products (256 bp and 232 bp) from the latter amplification. **Lower** Comparison, in chimpanzee (Pierre), of nucleotide sequences of genomic region of GPE exon V amplified from the 6.4-kb *Sac*I fragment of genomic DNA, with the corresponding region of GPE cDNA. *Lowercase letters* are genomic sequences; *uppercase letters* represent the sequence of the corresponding region of the transcript; *dots* represent sequence deletions; the t → g transversion at the 3′ end of intron 4 resulting in a valid acceptor splice site is *boxed.* In human GPE the 5′ end of exon V starts 9 bp upstream. Insertion (~codons 32–39, Fig. 3) sequence characteristic of human GPE exon V is marked by a *horizontal straddle rule.*

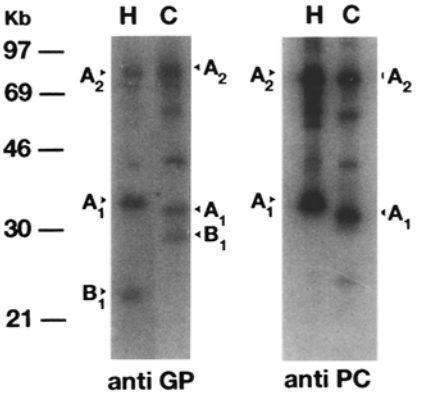



**Fig. 6.** Immunoblots of human and chimpanzee glycophorins. *H,* human; *C,* chimpanzee (Pierre). Nitrocellulose blots of red-cell membrane proteins probed with anti-GP and anti-PC. *A2* and *A1,* GPA dimer and monomer; *B2, B1,* GPB dimer and monomer.

human GPB due to the expression of the exon III sequence. No band with the mobility expected for GPE is seen in either human or chimpanzee membranes. Under the conditions used, such a band would be detected with the glycophorin antiserum if GPE were expressed in glycosylated or partially glycosylated form.

## Discussion

In the chimpanzee, GPA and GPB are expressed at the surface of the erythrocyte and bear the Simian blood-group antigens related to the human MNSs system. As in the case of the human counterpart, chimpanzee GPE, if expressed at all, does not occur in significant amounts or in an extensively glycosylated form on the cell membrane. The present study focused on the analysis of chimpanzee glycophorin transcripts and relevant genomic regions. Comparison with human glycophorins revealed major differences in the structure and expression of the three glycophorin genes and provided insights into the mechanism for their allelic diversification.

Previous studies presented evidence for a human-chimpanzee clade by analysis of the exon V insertion

sequence and the intron 5 *Alu* region in nonhuman primate glycophorin genes (Rearden et al. 1993). Results described here lend further support to such a phylogenetic relationship. Of the three glycophorin genes, GPA genes in the two species are most similar. The major variations are the presence of a hybrid M/N epitope sequence and the absence of a sequon for N-glycosylation in the chimpanzee. Concerning the M and N blood-group sequences, whether they arose from the hybrid form through gene conversion or spontaneous mutations remains to be established. It appears that *M* was formed earlier than *N* during primate evolution because a complete M-specific sequence is found in the chimpanzee GPE gene. Also, this is consistent with the observations that Old World monkey red cells react with anti-M but not anti-N reagents (Erskine and Socha 1978). Regarding N-glycosylation, only human GPA bears a single consensus sequence and thus the Asn-linked carbohydrate unit.

The chimpanzee GPB is characterized by the presence of an extended extracellular domain that retains the sequence encoded by exon III. This constitutes the major structural divergence from the human GPB gene. Difference in electrophoretic mobility was the first indication that chimpanzee GPB is larger than human GPB (Rearden 1986; Lu et al. 1987b). Here, the expression of GPB exon III was demonstrated by sequence analysis of its cognate cDNA and the donor splicing junction. Expression of exon III was noted in all chimpanzees so far examined and excludes GPB exon III as a reservoir of sequences for creation of allelic diversity. Perhaps inactive exons of GPE and, in particular, pseudoexon III, in which the complex palindrome characteristic of a hot spot of recombination was noted, could be modulated to fulfill this role. This is suggested by the existence in the chimpanzee of variant phenotypes of which some epitopes could be potentially encoded by this pseudoexon sequence, i.e., the Miltenberger (Mi)-like antigens (Socha and Moor-Jankowski 1979). More studies are needed to decipher the pattern and mechanism for intraspecies variations in the chimpanzee population.

The finding that GPB exon III is expressed in the chimpanzee but silenced in the human provides important clues as to the evolution of allelic diversity of human glycophorins. Among human glycophorin variants, many occur as hybrids of GPA and GPB involving the recurrent use of pseudoexon sequences in conjunction with the modulation of donor splice sites. Three classes of glycophorin variants—namely, MiIII, MiVI, and MiX—are notable examples that have been shown to arise by such a mechanism (Huang and Blumenfeld 1991a; Huang et al. 1992). In those variants, different patches of the silent exon became reactivated for expression by fusion to a segment of the GPA gene containing a functional donor splice site. Thus the inactivation of GPB exon III from chimpanzee to human had preserved a sequence repertoire for intraspecies variations in the human population. This evolutionary design was probably deliberate and unique to the human MNSs glycophorin gene family, for the splice site mutations, once having occurred in a structural gene, cannot be readily reverted with restoration of the wild-type.

Despite numerous attempts, the only chimpanzee GPE cDNA isolated here apparently occurs as a GPB/E composite transcript. It is similar to human GPE in two key aspects—namely, the presence of the M-specific epitope sequence and the inactivation of exons III and IV by splice site mutations identical to those present in human GPE. However, its transmembrane exon is similar to that of GPB, which contains the 9-bp rather than the 24-bp insertion. Sequence analysis indicated that GPE exon V exists in two forms in chimpanzee genomic DNA. The one that lacks the 24-bp insertion sequence is actively transcribed, whereas the other, with the 24-bp insertion sequence, probably is not. Further studies of those GPE forms in the chimpanzee and other species of anthropoid apes are required to determine whether they occur as the GPB/E progenitor gene and the GPE pseudogene, respectively.

In this study, in comparing human and chimpanzee glycophorins we focused on interspecies differences, by examining a single chimpanzee vs human. As documented previously, except for serologically documented variants, the common alleles of GPA, GPB, and GPE show insignificant variations among random humans (Race and Sanger 1975). This is in contrast to the chimpanzee, in which the existence of simian V-A-B-D blood groups related to human MN was the first indication that antigenic differences among individuals were more common than in the human. Confirmation of this intraspecies variation was obtained at the immunoblot level (Lu et al. 1987b; Rearden et al. 1993) and by restriction analyses (Lu et al. 1990; Rearden 1986). The unexpected intraspecies polymorphism has also been documented in the chimpanzee for the antigens of the Rh blood-group system (Blancher et al. 1993) and for glycophorins in other species of anthropoid apes (Lu et al. 1987b). The intraspecies variation and the interspecies differences, shown here to be confined mainly to the region encompassing the expressed exons of the extracellular domain, suggest that selective pressure may have been important in allelic diversification of certain surface proteins of the erythrocyte. If in the course of evolution a surface protein, such as glycophorin, became a ligand for a pathogenic agent, thereby facilitating its entry into the erythrocyte, amino acid changes that attenuate binding could have constituted a definite advantage. Protection against different strains of the malaria parasite could now be considered such a driving force for glycophorin in view of recent observations that human GPA and GPB can serve as receptors for *Plasmodium falciparum* (Dolan et al. 1994). Further studies will address the role of infec-

tious pathogens in shaping the intraspecies and interspecies diversity of the glycophorin A gene family.

*Acknowledgments.* We thank our colleague Dr. Sam Seifter for many helpful discussions and the critical reading of the manuscript and Ms. Emma Fay Ricksy for its preparation. This work was supported by NIH grant GM 16389. We also acknowledge the Oligonucleotide Synthesis and DNA sequencing Facility at AECOM supported by Cancer Center grant 5P30CA 13330.

## References

Blancher A, Ruffié J, Socha WW (1993) The R-C-E-F blood group system of chimpanzees: serology and genetics. J Med Primatol 22:13–18.

Blumenfeld OO, Adamany AM, Puglia KV, Socha WW (1983) The chimpanzee M blood group antigen is a variant of human M-N glycoproteins. Biochem Genet 21:333–348

Cartron J-P, Colin Y, Kudo S, Fukuda M (1990) Molecular genetics of human erythrocyte sialoglycoproteins glycophorins A, B, C, and D. In: Harris JR (ed) Blood cell biochemistry. Plenum Publishing, New York, pp 299–335

Dolan SA, Proctor JL, Alling DW, Okubo Y, Wellems TE, Miller LH (1994) Glycophorin B as an EBA-175 independent *Plasmodium falciparum* receptor of human erythrocytes. Mol Biochem Parasitol 64:55–63

Erskine AG, Socha WW (1978) The principles and practice of blood grouping, 2nd ed. CV Mosby Co, St Louis, p 207

Feinberg AP, Vogelstein B (1984) A technique for radiolabeling DNA restriction endonuclease fragments to high specific activity. Anal Biochem 137:266–267

Goossens M, Kan YW (1981) DNA analysis in the diagnosis of hemoglobin disorders. Methods Enzymol 76:805–817

Huang C-H, Blumenfeld OO (1991a) Molecular genetics of human erythrocyte MiIII and MiVI glycophorins. J Biol Chem 266:7248–7255

Huang C-H, Blumenfeld OO (1991b) Multiple origins of the human glycophorin $St^a$ gene. J Biol Chem 266:23306–23314

Huang C-H, Blumenfeld OO (1995) Molecular basis of human major blood group antigens. In: Cartron JP, Rouger P (eds) Blood cell biochemistry, vol 6. Plenum Press, pp 153–188

Huang C-H, Guizzo ML, McCreary J, Leigh EM, Blumenfeld OO (1991) Typing of MNSs blood group specific sequences in the human genome and characterization of a restriction fragment tightly linked to S-s- alleles. Blood 77:381–386

Huang C-H, Kikuchi M, McCreary J, Blumenfeld OO (1992) Gene conversion confined to a direct repeat of the acceptor splice site generates allelic diversity at human glycophorin (*GYP*) locus. J Biol Chem 267:3336–3342

Huang C-H, Reid ME, Blumenfeld OO (1993) Exon skipping caused by DNA recombination that introduces a defective donor splice into the human glycophorin A gene. J Biol Chem 268:4945–4952

Kudo S, Fukuda M (1989) Structural organization of glycophorins A and B genes: glycophorin B gene evolved by homologous recombination at Alu repeat sequence. Proc Natl Acad Sci USA 86:4619–4623

Kudo S, Fukuda M (1990) Identification of a novel human glycophorin E, by isolation of genomic clones and complementary DNA clones utilizing polymerase chain reaction. J Biol Chem 265:1102–1110

Lu Y-Q, Nichols ME, Bigbee WL, Nagel RL, Blumenfeld OO (1987a) Structural polymorphism of glycophorins demonstrated by immunoblotting techniques. Blood 69:618–624

Lu Y-Q, Liu J-F, Socha WW, Nagel RL, Blumenfeld OO (1987b) Polymorphism of glycophorins in nonhuman primate erythrocytes. Biochem Genet 25:477–491

Lu W-M, Huang C-H, Socha WW, Blumenfeld OO (1990) Polymorphisms and gross structure of glycophorin genes in common chimpanzees. Biochem Genet 28:399–413

Onda M, Kudo S, Rearden A, Mattei M-G, Fukuda M (1993) Identification of a precursor genomic segment that provided a sequence unique to glycophorin B and E genes. Proc Natl Acad Sci USA 90:7220–7224

Onda M, Kudo S, Fukuda M (1994) Genomic organization of glycophorin A gene family revealed by yeast artificial chromosomes containing human genomic DNA. J Biol Chem 269:13013–13020

Race RR, Sanger R (1975) The MNSs blood groups. In: Blood groups in man, 6th ed. Blackwell Scientific Publications, London, pp 92–138

Rearden A (1986) Evolution of glycophorin A in the hominoid primates studied with monoclonal antibodies, and description of a sialoglycoprotein analogous to human glycophorin B in chimpanzee. J Immunol 136:2504–2509

Rearden A, Phan H, Fukuda M (1990a) Multiple restriction fragment length polymorphisms associated with the Vc determinant of the MN blood group-related chimpanzee V-A-B-D systems. Biochem Genet 28:223–231

Rearden A, Phan H, Kudo S, Fukuda M (1990b) Evolution of the glycophorin gene family in the hominoid primates. Biochem Genet 28:209–222

Rearden A, Magnet A, Kudo S, Fukuda M (1993) Glycophorin B and glycophorin E genes arose from the glycophorin A ancestral gene via two duplications during primate evolution. J Biol Chem 268: 2260–2267

Sanger F, Nicklen S, Coulson AR (1977) DNA sequencing with chain-terminating inhibitors. Proc Natl Acad Sci USA 74:5463–5467

Shapiro MB, Senapathy P (1987) DNA splice junctions of different classes of eukaryotes: sequence statistics and functional implications in gene expression. Nucleic Acids Res 15:7155–7174

Socha WW, Moor-Jankowski S (1979) Blood groups of anthropoid apes and their relationship to human blood groups. J Hum Evol 8:453–465