Identification of a Novel Human Glycophorin, Glycophorin E, by Isolation of Genomic Clones and Complementary DNA Clones Utilizing Polymerase Chain Reaction*

Shinichi Kudo and Minoru Fukuda

From the La Jolla Cancer Research Foundation, Cancer Research Center, La Jolla, California 92037

In our previous report, we described the structural organizations of glycophorin A and B genes (Kudo, S., and Fukuda, M. (1989) Proc. Natl. Acad. Sci. U. S. A. 86, 4619–4623). During the course of isolation of these genomic clones, we also obtained genomic clones encoding a novel glycophorin. This novel glycophorin, termed glycophorin E (GPE), has a similar genomic structure to that of the GPB gene, and its nucleotide sequence is almost identical to that of the GPB gene. These sequences include a region downstream of an Alu repeat sequence, which has been suggested to be a site for homologous recombination in the GPB gene during or after gene duplication. However, the predicted GPE amino acid sequence specifies blood group M, in contrast to GPB which carries blood group N. Polymerase chain reaction was employed to analyze the transcript of this gene, and its cDNA sequence revealed that the novel glycophorin gene encodes 78 amino acids, including a 19-residue leader peptide. Comparison of genomic and complementary DNAs demonstrates that this gene consists of four exons, and point mutations at sites corresponding to the 5'-splicing sites of intron 3 and intron 4 of the GPA gene lead to the joining of the exon 2 to potential exon 5. Interestingly, an insertion of 24 nucleotides coding for eight amino acid residues in-frame was found in exon 5. The predicted amino acid sequence within this exon indicates that it has a hydrophobic character, suggesting the possible expression of GPE as a membrane protein. Northern blot analysis demonstrated that this novel glycophorin gene is expressed in an erythroid-specific manner and coordinately down-regulated together with GPA and GPB genes by a tumor-promoting phorbol ester. During evolution, this gene might have derived from an ancestral gene common to the GPB gene and cDNA sequence of GPC have been determined and revealed that the presumptive product carries GPA and GPB (4, 5) and their cDNA sequences (6–9) showed a similarity between GPA and GPB, indicating that these genes derived from a common ancestral gene. We have recently investigated the genomic organization of GPA and GPB to understand their precise structural relationship and evolutionary relatedness (10); both genes span more than 30 kb and have quite similar sequences (more than 95% identity in sequenced regions) from the 5'-flanking region to the repeat sequence of an Alu family about 1 kilobase downstream from the exon encoding the transmembrane domain. From the analysis of the Alu sequences and their adjacent direct repeats, the GPA gene is suggested to have arisen by homologous recombination at Alu sequences during or after gene duplication, explaining the difference between GPA and GPB in the 3'-region (10). On the other hand, GPA and GPB contain the blood group Gerbich (Ge) antigen. The amino acid sequence and cDNA sequence of GPE have been determined and revealed no significant homology with GPA and GPB (5, 11). GPA is assumed to be derived from GPA, and these two proteins are quite different from GPA and GPB.

During the course of isolation of GPA and GPB genomic clones from a K562 genomic library, we have obtained a novel glycophorin genomic clone. This gene has a structure similar to that of the GPB gene, but the presumptive product carries the M blood group antigen based on cDNA cloning of adopting PCR products. We propose that this gene be called glycophorin E gene, and we report here the characterization of this gene.

EXPERIMENTAL PROCEDURES

Cell Culture—Human erythroleukemic cell line K562 and human promyelocytic cell line HL-60 were grown in suspension culture in RPMI 1640 medium supplemented with 2 mM glutamine and 10% fetal bovine serum in a humidified 5% CO2 atmosphere. Treatment of cells with the tumor promoter 12-O-tetradecanoylphorbol 13-acetate (TPA) was conducted as described previously (12).

Isolation of Nucleic Acids—Total genomic DNA was isolated from K562 cells by overnight incubation at 37 °C with 0.4% NaDodSO4 and protease K (1 mg/ml) followed by phenol/chloroform extraction as described (13). Total RNA was prepared from cultured cells by a modification of the method of Chirgwin et al. (14).

The abbreviations used are: GPA, -B, -C, -D, and -E, glycophorins A, B, C, D, and E; kb, kilobase(s); TPA, 12-O-tetradecanoylphorbol 13-acetate; INS, insertion sequence specific to GPE; PCR, polymerase chain reaction; NaDodSO4, sodium dodecyl sulfate; bp, base pair(s).
was selected on oligo(dT)-cellulose (15). Screening of Genomic DNA Library—K562 genomic DNA library was constructed in the Lambda FIX phage vector (Stratagene) as described previously (10). The library was screened by using two synthetic oligonucleotides as probes. LP oligonucleotide (37-mer) corresponds to the nucleotide sequence in the leader peptide coding region, starting with the initiating methionine codon, and GC oligonucleotide (27-mer) corresponds to the nucleotide sequence encoding amino acid residues 18 (lysine) to 26 (asparagine). Oligonucleotides were labeled at the 5'-end with [γ-32P]ATP and T4 polynucleotide kinase (16, 17). Recombinant phage DNAs were purified by the method of Leder et al. (18) and identified and characterized by Southern blot (19) hybridization analysis. DNA fragments of interest were subcloned into EcoRI or EcoRV/SalI sites of BlueScript vectors (Stratagene).

Complementary DNA Sequencing by Polymerase Chain Reaction—The oligo(dT)-primed, complementary DNA strand was synthesized from 0.5 μg of K562 poly(A)+ RNA in a 20-μl reaction mixture containing 50 mM Tris-HCl, pH 8.3, 3 mM MgCl2, 75 mM KCl, 10 mM dithiothreitol, 0.5 mM each of dATP, dGTP, dCTP, and dTTP, and 0.5 μg of oligo(dT), incubated with 200 units of Moloney murine leukemia virus reverse transcriptase (Bethesda Research Laboratories) at 37 °C for 1 h. For the polymerase chain reaction (20), 2 μl out of the above reaction mixture was added to a final volume of 50 μl of 10 mM Tris-HCl, pH 8.3, 50 mM KCl, 1.5 mM MgCl2, 0.1% (w/v) gelatin, 200 μM each of dATP, dGTP, dCTP, and dTTP, 1 μM each primer, and 1 unit of Thermus aquaticus (Taq) DNA polymerase. Amplification of cDNA was repeated 35 times under the conditions of denaturation at 94 °C, annealing for 2 min at 55 °C, and extension for 2 min at 72 °C. After 30 cycles of amplification for 5 min at 72 °C on a Perkin-Elmer-Cetus DNA thermal cycler. A 5-μl aliquot of this reaction was cloned in a 3% NuSieve GTG agarose gel (FMC Corp.), and amplified fragments were analyzed by Southern blot (19) hybridization using GPB cDNA (HGpB-1, Ref. 7) and INS oligonucleotide, corresponding to a 24-bp insertion sequence found in the exon encoding a potential transmembrane domain in the novel glycophorin gene (see below), as probes. The cDNA probes were made by the random oligonucleotide primer-extension method (21). Then PCR products were treated with the Klönov fragment of E. coli DNA polymerase and cloned into the Smal site of pUC13 vector by blunt-end ligation. Recombinant DNAs were transformed into E. coli XL1-Blue (Stratagene) and plated on a lawn of competent XL1-Blue cells. After transformation, transformants were screened with two synthetic oligonucleotides as probes. LP oligonucleotide probe according to the method described previously (22). Positive clones were isolated from the original oligonucleotide filter and plated on a nitrocellulose filter. Colonies on the replica plates were incubated at 37 °C for 2 h and subjected to colony hybridization using an INS oligonucleotide probe according to the method described previously (22). Positive clones were isolated from the original oligonucleotide filter and plated on a lawn of competent XL1-Blue (Stratagene) and plated on a lawn of competent XL1-Blue cells. After transformation, transformants were screened with two synthetic oligonucleotides as probes. LP oligonucleotide probe according to the method described previously (22). Positive clones were isolated from the original oligonucleotide filter and plated on a lawn of competent XL1-Blue (Stratagene) and plated on a lawn of competent XL1-Blue cells. After transformation, transformants were screened with two synthetic oligonucleotides as probes. LP oligonucleotide probe according to the method described previously (22). Positive clones were isolated from the original oligonucleotide filter and plated on a lawn of competent XL1-Blue (Stratagene) and plated on a lawn of competent XL1-Blue cells. After transformation, transformants were screened with two synthetic oligonucleotides as probes. LP oligonucleotide probe according to the method described previously (22). Positive clones were isolated from the original oligonucleotide filter and plated on a lawn of competent XL1-Blue (Stratagene) and plated on a lawn of competent XL1-Blue cells. After transformation, transformants were screened with two synthetic oligonucleotides as probes. LP oligonucleotide probe according to the method described previously (22). Positive clones were isolated from the original oligonucleotide filter and plated on a lawn of competent XL1-Blue (Stratagene) and plated on a lawn of competent XL1-Blue cells. After transformation, transformants were screened with two synthetic oligonucleotides as probes. LP oligonucleotide probe according to the method described previously (22). Positive clones were isolated from the original oligonucleotide filter and plated on a lawn of competent XL1-Blue (Stratagene) and plated on a lawn of competent XL1-Blue cells. After transformation, transformants were screened with two synthetic oligonucleotides as probes. LP oligonucleotide probe according to the method described previously (22). Positive clones were isolated from the original oligonucleotide filter and plated on a lawn of competent XL1-Blue (Stratagene) and plated on a lawn of competent XL1-Blue cells. After transformation, transformants were screened with two synthetic oligonucleotides as probes. LP oligonucleotide probe according to the method described previously (22). Positive clones were isolated from the original oligonucleotide filter and plated on a lawn of competent XL1-Blue (Stratagene) and plated on a lawn of competent XL1-Blue cells. After transformation, transformants were screened with two synthetic oligonucleotides as probes. LP oligonucleotide probe according to the method described previously (22). Positive clones were isolated from the original oligonucleotide filter and plated on a lawn of competent XL1-Blue (Stratagene) and plated on a lawn of competent XL1-Blue cells. After transformation, transformants were screened with two synthetic oligonucleotides as probes. LP oligonucleotide probe according to the method described previously (22). Positive clones were isolated from the original oligonucleotide filter and plated on a lawn of competent XL1-Blue (Stratagene) and plated on a lawn of competent XL1-Blue cells. After transformation, transformants were screened with two synthetic oligonucleotides as probes. LP oligonucleotide probe according to the method described previously (22). Positive clones were isolated from the original oligonucleotide filter and plated on a lawn of competent XL1-Blue (Stratagene) and plated on a lawn of competent XL1-Blue cells. After transformation, transformants were screened with two synthetic oligonucleotides as probes. LP oligonucleotide probe according to the method described previously (22). Positive clones were isolated from the original oligonucleotide filter and plated on a lawn of competent XL1-Blue (Stratagene) and plated on a lawn of competent XL1-Blue cells. After transformation, transformants were screened with two synthetic oligonucleotides as probes. LP oligonucleotide probe according to the method described previously (22). Positive clones were isolated from the original oligonucleotide filter and plated on a lawn of competent XL1-Blue (Stratagene) and plated on a lawn of competent XL1-Blue cells. After transformation, transformants were screened with two synthetic oligonucleotides as probes. LP oligonucleotide probe according to the method described previously (22).

RESULTS AND DISCUSSION

Isolation of Genomic Clones Encoding Novel Glycophorin—A K562 genomic DNA library was constructed as described previously (10). A total of 107 recombinant phages were screened with two synthetic oligonucleotides as probes. LP and GC oligonucleotides, which correspond to nucleotide sequences in the leader peptide and the amino-terminal region of mature glycophorin, respectively, are perfectly shared in GPA and GPB genes. As a result, we isolated 16 LP-positive clones and 25 GC-positive clones. In our previous study, a 2.0-kb fragment was detected by hybridization of EcoRI-digested total K562 genomic DNA with the GC oligonucleotide probe. Two clones, C2 and C25, having a GC-reactive fragment of the same size, were isolated and found to be derived from GPA and GPB genes, respectively. In addition, we also obtained one clone, C4, whose reactive EcoRI fragment with GC accompanied the short arm of Lambda FIX vector DNA. When this was digested further with SalI, whose recognition site in the vector flanks the cloning site, a 0.8-kb fragment was generated that was GC-reactive. EcoRI digestion of this cloned DNA produced three other fragments of the same sizes (3.5, 0.7, and 4.2 kb) found in the GPB genomic clone C25. At first, this clone seemed to contain a truncated GC-reactive EcoRI fragment and following 3′-region of GPB gene. However, DNA sequencing of the 0.8-kb GC-reactive fragment revealed that the first and fifth amino acids for a putative mature protein are serine and glycine, respectively, suggesting that this glycophorin gene product carries the group M antigen even though it has structural similarity to GPA gene. Therefore we decided to analyze this genomic clone in more detail. As in the cases of GPA and GPB genes, this novel glycophorin gene has a large first intron and was isolated as two separate clones. Previously we divided LP-positive genomic clones having 5′-sequences into three groups according to the size of BglII fragments (2.4, 2.6, and 3.8 kb) which hybridize to the LP oligonucleotide probe. Clones LP1-3 (2.4 kb) and LP1-8 (2.6 kb) of two positive groups were assigned to the GPA and GPB genes, respectively, by comparison with CDNA sequences (10). Clone LP1-2 of the 3.8-kb positive group could not be assigned to either gene because of the discrepancy in nucleotide sequences at 5′-untranslated region. In the present study, we found that this 3.8-kb fragment was derived from a novel glycophorin gene, as shown below. These results enabled us to construct a physical map of this novel gene, using one 3.8-kb LP-positive clone, LP1-2, and clone 4 that includes the rest of the exons (Fig. 1).

Organization of Novel Glycophorin Gene—Predicted exon-intron boundaries and 5′- and 3′-flanking regions were sequenced by using synthetic oligonucleotides based on GPB CDNA sequences (7) as primers. The sequences of the opposite strands were then determined with newly synthesized oligonucleotides based on the obtained sequencing data. These results revealed the following features. The nucleotide sequences of this novel gene are aligned well with that of the GPB gene over the entire genomic region, as shown in Fig. 2. Thus three glycophorin genes have almost identical sequences from their 5′-flanking regions to the Alu sequences about 1 kb downstream from the exon encoding the transmembrane domain. In particular, the sequences in exon 1 and 2, and around their exon-intron boundaries are highly conserved among them. In general, 5′-splicing and 3′-splicing sites have consensus sequence, exon-·gt·-intron-·ag/-exon (24).

In our previous study we revealed that within this homologous genomic region GPB gene has a point mutation at the 5′-
Fig. 1. Comparison of the organization of GPA, GPB, and GPE genes. The genomic organization in each gene was constructed by two separate clones, covering exon-1 and following exons, respectively. The first intron is suggested by the previous study to be more than 30 kb in size (10). In each set of two lines, the upper line represents the EcoRI restriction pattern with distances between sites in kilobases, and the lower line represents the genomic structure. Exons are expressed by vertical bars or filled boxes in proportion to sizes. For convenience, the exon number in the GPB and GPE genes followed that of the GPA gene except for the last exon having the 3'-untranslated (UT) region. The transition site or break point is indicated by BRP.

splicing site of the third intron and results in the joining of exon 2 to exon 4 (10).

Similar point mutations thought to effect the splicing event are observed at two 5'-splicing sites in GPE gene. One is positioned at the same 5'-splicing site of the potential third intron, as found in GPB gene. A change from guanine to adenine is noticed at the first nucleotide of this intron as shown in Fig. 2. Another change is observed at 5'-splicing site of the potential fourth intron, changing from guanine to adenine. Since these changes are different from that of GPB gene, it is judged that these mutations have taken place independently. These mutations abolish the 5'-splicing events in both intron 3 and intron 4, which were confirmed by the isolation of cDNA clone of GPB transcript (see below). The cDNA sequence showed that exon 3 and 4 are spliced out and direct ligation of exon 2 to exon 5 takes place for this mRNA from GPE gene. It is worth noting that these mutations do not give rise to utilization of any crypptic splicing sites nor larger transcripts containing third and/or fourth intron sequences. Intriguingly, similar splicing patterns were noticed in vitro study of β-globin gene from β-thalassemia (25) and the mutation of the 5'-splicing site in rabbit β-globin gene (26). In both cases a point mutation at the 5'-splicing site of the second intron resulted in production of transcript having ligation of the first exon to the third exon, and the other aberrant transcripts. These results strongly suggest that the sequence of the 5'-splicing site is closely involved in the upstream splicing event of these transcripts.

In GPB gene, the 9-bp preceding 3'-splice site from that of GPA gene in exon 5 was produced by the nucleotide substitution in intron. However, the GPA gene has the same 3'-splicing site as that of GPA gene. Therefore, it is judged that this substitution in GPB gene also occurred independently after gene duplication. As another interesting observation, GPA gene was found to have an insertion of 24 nucleotides (INS) in the exon corresponding to exon 5 of GPA (Fig. 2). The origin of this sequence is not known. However, the homology with this sequence is observed at the adjacent 24-bp downstream sequence in this gene and moreover at corresponding regions in both GPA and GPB genes (see Fig. 3). It is of interest to note that the homology between the insertion sequence and corresponding GPA or GPB sequence is higher than that between the insertion and adjacent sequences in GPE gene (Fig. 3).

The presence of this third glycoporphin gene has been recognized in previous studies. When genomic DNA from En(a−) cells, which lack glycoporphin A gene, was analyzed, a 6.4-kb fragment obtained by HindIII and a 1.9-kb fragment obtained by PvuI digestion, were detected as additional bands by cDNA probes encoding the 5'-end of GPA or GPB transcript (27). These fragments could be also detected in DNA from individuals lacking the GPB gene (28). Therefore these restriction fragments were judged to be derived from other than glycoporphin A or B gene, which shares a strong homology with the GPA or GPB gene at 5'-end of the gene. The present study demonstrates that this novel gene is actually similar to the GPB gene in gross organization, yet distinctly different from the GPB gene in various structural features, as shown above. Our studies also demonstrated that a pair of alleles exists in GPE genes which differ in the first nucleotide of exon 2 (see below). Furthermore, a genomic clone encoding GPE was isolated by PCR from genomic DNA of leukocytes of an individual having St' variant glycoporphin.2 These combined results strongly suggest that each pair of haploid chromosomes contain one copy of GPE gene. It will be essential to elucidate the chromosomal localization of GPE gene.

cDNA Amplification by Polymerase Chain Reaction—In order to know whether this novel glycoporphin gene is transcribed or not, we employed PCR for cDNA isolation. Oligo(dT)-primed cDNA templates were made from K562 poly(A)+ RNA and they were amplified by PCR with two oligonucleotides, which were synthesized according to the genomic DNA sequence. The sequence used for the 5'-primer was positioned 20 base pairs upstream from the predicted translation initiation codon and is perfectly shared among GPA, GPB, and putative GPE genes. Between this sequence and the translation initiation codon there are two nucleotides in the putative GPE sequence, which differ from the GPA or GPB gene. At the other end, the 3'-primer sequence was chosen to contain the last exon of GPB gene, excluding direct repeat sequence (see below), and was also perfectly shared between GPB and GPE genes (see Fig. 5).

PCR products were found to have sizes expected from the analysis of GPB and GPE genomic DNA sequences (Fig. 4). In addition, the upper band, predicted to correspond to GPB transcript, reacted strongly with GPB cDNA whereas the lower band reacted with INS oligonucleotide, indicating its origin as GPE transcript (Fig. 4, lanes 2 and 3). From the intensity of the bands stained with ethidium bromide, GPE

**Fig. 2. Nucleotide-sequences comparison of GPA, GPB, and GPE genes.** The sequences of all exons, exon-intron junctions, and 5′- and 3′-flanking regions are aligned among three genes. A, the sequence in which GPA, GPB, and GPE genes are almost identical. Each set of three lines (from top to bottom) shows the nucleotide sequences of GPA (M-phenotype), GPB, and GPE genes, respectively. Arrows (filled and open) and the triangle indicate the point mutations affecting the splicing pattern of GPB and GPE transcripts (see text). A 24-bp insertion sequence in the GPE is indicated by a shaded box. The following 24-bp sequences, which are homologous to the insertion sequence among GPA, GPB, and GPE genes, are shown by a box. The transcription initiation site of the GPE gene was assumed to be the same position as shown for the GPA gene (10). The sequences coding the mature RNA sequences are written in capital letters, whereas 5′-upstream, 3′-downstream, and intron sequences are written in small letters. Nucleotides identical to GPA sequences are shown by dashes. B, the 3′ unique sequence in GPA gene. C, the 3′ sequence is almost identical between GPB and GPE genes. Nucleotides identical to GPB sequences are shown by dashes. The sequences in B and C are those directly following the transition site from homologous to non-homologous sequence (BRP in Fig. 1). The splicing branch points are indicated by asterisks.
gene can be expected to be transcribed slightly less than GPB gene (Fig. 4, lane 1). PCR cDNA products were subcloned into pUC13 vector, and novel glycophorin-specific cDNA clones were isolated by colony hybridization with the unique INS oligonucleotide as a probe. One clone, HGpE-7, contains cDNA starting at the fifth nucleotide of the 5'-primer extending to the 17th nucleotide of the 3'-primer, encompassing four exons sequences (see the sequence between asterisks in Fig. 5). These exons correspond to exons 1, 2, and 5 of the GPA gene, and the last exon of the GPB gene, but sequences corresponding to exons 3 and 4 of GPA were not included. The 5'-untranslated region of this cDNA has a sequence identical to that of the 3%kb LP-positive genomic clone, with two base pairs different from the GPA and GPB genes (Fig. 5). These results indicated that the 3%kb LP-positive clone LPl-2 was derived from the GPE gene and that clone C4 is a physical map of the GPE gene as shown in Fig. 1.

During these experiments, we have isolated other GPE cDNA clones by PCR and found that these clones have a different residue at the first nucleotide of exon 2 (Fig. 5), a guanine residue instead of adenine. These results strongly suggest that this GPE gene exists as two alleles.

Comparison of cDNA sequences among GPA, GPB, and GPE indicates that the sequence encoding the transmembrane portion of GPE diverges significantly from that of GPA and GPB. Both the insertion sequence and frequent point mutations lie in this portion, coded by corresponding exon 5 of GPA. One of the changes from cytosine to thymine at nucleotide 272 introduces a termination codon (TGA) at that position, 5 codons preceding than that of GPB. On the other hand, the sequence corresponding to the last exon of the GPE gene is almost identical in GPE gene, indicating that the signal for poly(A) addition at the same site as GPB gene is most likely utilized in GPE gene (Fig. 5). Northern analyses indicate that mRNA for GPE has almost identical size as mRNA for GPB, supporting the above conclusion (see below).

Within the last exon in both GPE and GPB genes, we detected a direct repeat, consisting of 34-bp elements (see the horizontal arrows in Fig. 5). This duplication in the 3'-untranslated region must have taken place a long time ago, since the two sequences have diverged significantly from each other.

**Fig. 3. Analysis of 24-bp insertion sequence.** a, homology between the 24-bp insertion sequence and the adjacent following sequence in the GPE gene. b, homologies of the 24-bp insertion sequence of GPE gene with the corresponding sequences in GPA and GPB genes. c, homologies of an adjacent following sequence of GPE gene with the corresponding sequences in GPA and GPB genes. d, homology of corresponding sequences between GPA and GPB genes. The nucleotides of GPE are numbered as described in Fig. 5. The location of the insertion sequence is shown by a shaded box and the adjacent or corresponding sequences are boxed in Fig. 2.

**Fig. 4. Southern blot analysis of PCR products of GPB and GPE cDNAs.** Complementary DNA was synthesized from K562 poly(A) mRNA and amplified by PCR as described under "Experimental Procedures." Two oligonucleotide primers (18L and 241R) used for PCR are shown in Fig. 5. PCR product was fractionated by 3% NuSieve GTG agarose gel and stained with ethidium bromide (lane 1). Southern blot (19) hybridization was performed according to the method described by Maniatis et al. (19), with INS oligonucleotide probe (lane 2) and GPB cDNA (HGpB-1; Ref. 7) probe (lane 3). Markers at the left show positions and sizes in base pairs of HaeIII-digested λx174RF DNA.

**Predicted Amino Acid Sequence of GPE—**The nucleotide sequences of both genomic clones and cDNA clones indicate that the GPE gene encodes a GPE protein of 78 amino acids, including a putative leader peptide consisting of 19 amino acids (Fig. 5). The presence of a basic amino acid close to the initiation methionine (29), and the amino acid residues at the third and the first position from the COOH terminus of the putative leader peptide fit very well with the so-called (−3, −1) rule for signal peptide cleavage (30). The predicted amino acid residue at −7 is glutamic acid or glycine in each allele, the former being the same as that of GPB. The first and fifth amino acids of the mature GPE protein are serine and glycine, indicating this protein expresses the blood group M type. Because of the lack of exons corresponding to exon 3 and exon 4 of GPA gene, the extracellular domain is coded only by exon 2 and is shorter than GPA or GPB (Fig. 6). GPB expresses S or s blood group antigen determined by the amino acid residue at 29. Since the sequences corresponding to the residues 27–42 of GPB are encoded by exon 4, which is absent in GPE, GPE protein does not express this blood group antigen. For the same reason, N-glycan is absent, since joining
of exon to exon 5 does not result in the formation of the Asn-X-Ser(Thr) sequence for N-glycan attachment site which is present in GPA protein (4, 6).

More notable differences can be observed in the putative transmembrane portion encoded by exon 5. The insertion sequence (INS) in exon 5 encodes an additional 8 amino acids (Asn-Trp-Trp-Ala-Met-Ala-Arg-Val) not found in GPA or GPB. The hydropathy profiles of the predicted amino acid sequences of exon 5 in GPA, GPB, and GPE are shown in Fig. 7. By having insertions and substitutions of amino acids, the middle part of the putative transmembrane portion in GPE has reduced hydrophobicity and separated this domain into two hydrophobic regions. However, the carboxyl-terminal region of this domain consists of 20 hydrophobic amino acids, whose size is long enough to span a membrane (29, 31), and one basic amino acid is present at the carboxyl-terminal end, which could be situated in the cytoplasmic side (Figs. 5 and 7). Since many of transmembrane proteins contain two or three basic amino acids at the cytoplasmic face of the membrane, as shown in GPA and GPB (4-7) (see Fig. 7), it is not known whether only one basic amino acid would be enough to act with the hydrophobic domain to form a transmembrane configuration. It is possible that both hydrophobic domains interact with the membrane, with the second hydrophobic portion being oriented toward the extracellular domain, making the carboxyl-terminal arginine exposed to extracellular environment. If this is the case, it would be likely that GPE is less strongly bound to the lipid bilayer than GPA or GPB. Nevertheless, it is highly possible that some of the hitherto unidentified sialoglycoproteins of human erythrocytes (see Ref. 5) may belong to GPE. Further studies are necessary to clarify how efficiently synthesized GPE is incorporated to the plasma membrane and how the two hydrophobic portions are oriented.
Novel Human Glycophorin Gene

Transcriptional Regulation of Novel Glycophorin Gene—To investigate transcriptional regulation of the GPE gene, Northern blot hybridization of mRNA was performed with different probes. First GPE cDNA was used as probe, and detected a 0.5–0.6-kb, strong, broad band, and 2.7-, 1.7-, and 1.0-kb minor but distinct bands on a K562 RNA blot (Fig. 8). The size of the smallest band is consistent with the results of genomic and cDNA analyses. The same filter was rehybridized with a cDNA probe specific to the 3' untranslated sequence of GPB. This probe detected a similar band of 0.5–0.6 kb. All of the bands detected by GPE cDNA were cross-hybridized with GPA cDNA probe, detecting 2.7, 1.7, and 1.0 kb for GPA mRNA, 0.6–0.7 kb for GPB and/or GPE mRNA (data not shown, see also Ref. 7).

The signals of all bands were remarkably decreased in the RNA prepared from TPA-treated K562 cells, indicating that the transcription of GPE gene and GPA and GPB genes are coordinately down-regulated by TPA treatment of K562 cells (Fig. 8). As in the cases of GPA and GPB, GPE mRNA was not detected in HL-60 human promyelocytic leukemia cells. These results strongly suggest that the GPE gene has quite similar promoter and enhancer functions as GPA and GPB genes. In fact, the nucleotide sequencing data show that the 5'-flanking sequences of the GPE and GPB genes are 98.5% identical until 750 bp upstream from the transcription initiation site (data not shown).

Evolution of Glycophorin Gene—In our previous study (10), we found that GPA and GPB genes have almost identical sequences from 5'-flanking sequence to the region about 1 kilobase downstream from exon 5. The transition site from homologous to non-homologous regions in GPA and GPB genes was positioned at a sequence of an Alu family (32). In general, Alu sequences are flanked on both sides by direct repeats of 7-20 bp, suggesting structures similar to mobile genetic elements (32,33). By precise analysis of Alu sequences and their flanking direct repeats, it was suggested that the GPB gene arose from the acquisition of 3'-sequences different from those of the GPA gene by homologous recombination at the Alu repeats during or after gene duplication (10).

Fig. 9 shows the comparison of the nucleotide sequences surrounding this Alu repeat obtained from GPA, GPB, and GPE genomic clones. The results can be summarized as follows. (a) The sequences upstream from this Alu repeat are very homologous among three genes, whereas the sequences...
Fig. 9. **Alu** family sequence at the transition site and the following 3′-sequence in GPA, GPB, and GPE genes. (a) (upper), each set of three lines shows the nucleotide sequences among GPA (A), GPB (B), and GPE (E) genes. Identical nucleotides between GPA and GPE genes are indicated by asterisks, and the nucleotides of the GPE gene identical to those of GPB gene are indicated by dashes. Alu sequences are presented by bold letters. Flanking direct repeats are indicated by arrows. The presumptive recombination site is indicated by a shaded box. Boxed region shows the polyadenylation signal. The numbering starts from the 5′-EcoRI site.

Fig. 10. **Ancestral gene** was first duplicated and one of the duplicated genes directly evolved into the GPA gene. The other gene acquired an unique 3′-end sequence by homologous recombination through Alu repeats, and then the resulting gene was duplicated to produce GPB and GPE genes.
much less homology (Fig. 9B and Ref. 10). Since the direct repeats are conserved in GPA gene, the original genomic structure around the Alu repeat is represented by GPA gene. These results clearly indicate two sequential duplications of glycophorin genes. First, the ancestral gene was duplicated and one of them became a direct precursor to GPA gene. The other gene acquired a 3′-end sequence unique to GPB or GPE gene, probably by cross-over (see below). This gene then duplicated to produce GPR and GPE genes (Fig. 10).

Recently we performed the Southern hybridization of K562 genomic DNA with the HaeIII DNA fragment containing only the 3′-untranslated region of the GPB gene. Several restriction enzymes were utilized in this experiment and in some digestions more than two hybridizing bands were detected. For instance EcoRI digestion gave two bands, a 4.2-kb strong band and a 16-kb faint band.3 The 4.2-kb band certainly corresponds to the genuine EcoRI fragment observed in GPB and GPE genomic clones, whereas the 16-kb fragment must be produced from an unrelated gene. Interestingly, in primates the same EcoRI hybridizing pattern was observed until gorilla. It is thus likely that the 16-kb fragment represents a precursor gene that donated the 3′-end of GPB and GPE genes. Studies are now in progress to identify this precursor gene. The results strongly suggest that the ancestral gene to GPB and GPE acquired its unique 3′-end sequence probably by cross-over, since this donating gene still exists after GPB and GPE formation.

We estimated using the intron sequences that gene duplication of ancestral gene of GPA and GPB took place 9–35 million years ago (10). Our studies on primate glycophorins confirm this conclusion and it was estimated that this first duplication took place about 8 16 million years ago.2 GPB and GPE genes have evolved from their common ancestral gene by gene duplication after the Alu-Alu homologous recombination, showing quite identical 3′-regions (Fig. 9). The difference in sequence of this region between GPB and GPE genes have been produced after gene duplication. Eight nucleotide mismatches can be observed in Alu sequences between nucleotide 50–346 of GPB and 50–348 of GPE genes (Fig. 9). Alu sequences have been shown to evolve at the rate of neutral evolution, providing a suitable evolutionary clock (35). If we apply the same rate for this divergence (36,37), we can estimate that this second gene duplication took place 5–21 million years ago. In addition, it is possible that gene conversion took place, since GPA (M-type) and GPE genes share the same sequence in exon 2. As a result of these events at least three closely related glycophorin genes exist on haploid genomes, probably constituting a gene family. Further studies to elucidate the structural organizations of the GPA, GPB, and GPE genes will be important to understand how this family of genes evolved through duplication, recombination, and nucleotide substitution.

Acknowledgments—We thank Todd Dubinoff and Gizella S. Nagy for excellent technical assistance, John Knight for oligonucleotide synthesis, Dr. Mark Williams for critical reading of the manuscript, and Henny Bierhuizen for secretarial assistance.

REFERENCES

Identification of a novel human glycophorin, glycophorin E, by isolation of genomic clones and complementary DNA clones utilizing polymerase chain reaction.

S Kudo and M Fukuda


Access the most updated version of this article at http://www.jbc.org/content/265/2/1102

Alerts:
- When this article is cited
- When a correction for this article is posted

Click here to choose from all of JBC's e-mail alerts

This article cites 0 references, 0 of which can be accessed free at http://www.jbc.org/content/265/2/1102.full.html#ref-list-1