RAPID COMMUNICATION

Genomic Organization of the Glycoprotein D Gene: Duffy Blood Group Fy\(a/Fy\) Alloantigen System Is Associated With a Polymorphism at the 44-Amino Acid Residue

By Sadahiko Iwamoto, Toshinori Omi, Eiji Kajii, and Shigenori Ikemoto

The Duffy blood group antigen has been characterized by its roles on red blood cells: as a receptor for the malarial parasites and as a promiscuous receptor for chemokine superfamily. Recently, the Duffy blood group associated glycoprotein D (gpFy) cDNA has been cloned (Chaudhuri et al: Proc Natl Acad Sci USA 90:10793, 1993). In this report we describe the organization of genomic DNA coding for the gpFy and elucidate the molecular nature of Fy\(^{a/b}\) polymorphisms. By a Southern blotting analysis probed with gpFy cDNA, gpFy gene was shown to be composed of three DNA fragments: 1.1-kb Sac I, 1.9-kb EcoRI, and their intervening 47-bp fragments. We cloned the 1.1-kb SacI and 1.9-kb EcoRI fragments by inverted polymerase chain reaction (IPCR) procedure. The promoter region of the gpFy gene was cloned by IPCR of 1.1-kb SacI fragment and the 3' flanking sequence was cloned by IPCR of 1.9 kb EcoRI fragment. The both IPCR products contained on both side the known gpFy cDNA sequence without introns, as expected. Although no TATA or CCAAT boxes are present in the promoter sequence, several transcription factor binding site motifs are contained, including AP-1, HNF-5, TCF-1, ApoE B2, W-element, H-APF-1, and Sp-1. The 3' flanking region has two additional polyadenylation signals, other than that used in the cDNA, and also has an indirect and a direct repeat sequence clustered with the 5' flanking region. These facts indicate a possibility that the gpFy gene has been evolved by multiple retrotransposition events. By comparing the coding area of the gpFy gene in 28 Duffy-positive individuals, we elucidated that one base change that results in an amino acid substitution [GA-T(Asp\(^{44})\rightarrow GGT(Gly)] is in accordance with the Fy\(^{a/Fyb}\) polymorphism. This fact proves that the gpFy cDNA and its gene described in this report encode the Duffy blood group system.

© 1995 by The American Society of Hematology.

MATERIALS AND METHODS

Cell preparation. Whole-blood samples with the phenotypes of Fy(a+b-), Fy(a-b+), Fy(a+b+), and Fy(a-b-) were collected on heparin or EDTA after informed consent. Fy(a+b+) and Fy(a-b+) Japanese samples were provided by Red Cross Bloodbank Tochigi. Total RNAs were extracted from 1 mL of packed red blood cells (RBCs) by the acid-phenol-guanidium method. 11 Genomic DNAs were separated from theuffy coat by the usual phenol/chloroform and RNaseA method. 12

Southern blot analysis. Reticuloctye poly(A)+RNA was reverse-transcribed into the first-strand cDNA with oligo(dT)16 primer and cloned Molony murine leukemia virus (MMLV) reverse transcriptase (US Biochemical Corp, Cleveland, OH). The reverse transcription product was directly amplified by polymerase chain reaction (RT-PCR) 13 with a primer pair of gpFy1 (sense, GCTTCCCCAGAGACTGTTCCTG) and gpFy1218 (antisense, ATTCAGGTTGACAGGTGGGAA). The nucleotide and amino acid numbers correspond to the sequence deposited in GenBank (accession no. U01839), and the primers were temporarily named by their 5' positions. The amplified 1,218-bp PCR product was subcloned into pGEM-T vector (Promega, Madison, WI). The subcloned full-length gpFy cDNA was amplified with a primer pair of gpFy1/735.
(antisense, GTGACAGCCAGGTCTAGTGA) (5' probe) or gpFy579 (sense, ACTTGTTGCTGGTATGCTC/AA/1/218 (3' probe). After the isolation by a low-temperature melting gel, the PCR products were labeled by the random priming method using Multiprime DNA labeling system (Amersham Japan, Tokyo) and were used for the probes of Southern blot analysis.

Ten micrograms of genomic DNA was digested with sufficient units of restriction enzyme, EcoRI or Sac I. DNA fragments were resolved by electrophoresis on a 0.8% agarose gel and were transferred to a Hybond-N+ nylon membrane (Amersham Japan). Hybridization with the (α-32P)dCTP labeled 5' probe was performed for 24 hours at 65°C in 0.5% sodium dodecyl sulfate (SDS), 0.9 mol/L NaCl, 50 mmol/L NaH2PO4, 5 mmol/L EDTA. Final wash was performed at 65°C for 45 minutes in 18 mmol/L NaCl, 1 mmol/L NaH2PO4, 0.1 mmol/L EDTA, 0.1% SDS. After autoradiography, the blot membrane was stripped in boiling 0.1% SDS and hybridized to 3' probe.

Genomic cloning. The results of Southern blot analysis and the gpFy cDNA restriction map indicate that the gpFy gene is composed of 5' 1.1-kb Sac I fragment, 3' 1.9-kb EcoRI fragment, and their intervening 47-bp fragment. The 5' region of the gpFy gene was amplified from genomic DNA that was prepared from a donor with Fy(a+b+) phenotype, using the IPCR technique described by Triglia et al16 with some modifications. One microgram of genomic DNA was digested with the restriction enzyme Sac I. The Sac I–treated DNA fragments were ligated by T4 DNA ligase (Takara, Kyoto, Japan) to form monomeric circular forms in 100 μL ligation reaction mixture, and were heat treated. To amplify the 1.1-kb Sac I fragment of gpFy gene two oligonucleotides were prepared: gpFyl08 (antisense, TCGGGTGGGAGAACAAGGTC) and gpFyl149 (sense, GAGACCTTTCGCGTGTAACCT). PCR amplification was performed in an automatic thermal cycler (Perkin Elmer-Cetus, Norwalk, CT) and in a buffer consisted of 5 mmol/L Tris (pH 9.0), 15 mmol/L Tricin (pH 8.8), 50 mmol/L KCl, 2 mmol/L MgCl2, 0.2 mmol/L of each deoxynucleotides, 0.25 mmol/L of sense and antisense primers each, and 2.5 U Taq polymerase. Each cycle consisted of 1 minute of denaturation at 94°C, annealing for 1 minute at 55°C, and extension for 3 minutes at 72°C. After completion of 35 cycles, the amplified product was ligated in pGEM-T vector.

The 3' region of gpFy gene was amplified by IPCR from 1 μg of genomic DNA fragments digested with EcoRI and successively ligated to form circular forms. To amplify the 1.9-kb EcoRI fragment of gpFy gene two oligonucleotides were prepared: gpFy735 (antisense) and gpFy1145 (sense, CCTGAGGATGGCTTCTTCTCA). The PCR product was ligated in pCR II vector (Invitrogen, San Diego, CA).

Four independent clones of each subcloned fragment were subjected to sequence analysis. The DNA sequence was determined on both strands by the dideoxy termination method using the Sequenase Version 2.0 (US Biochemical Corp, Cleveland, OH).

The gpFy cDNA sequences were amplified from untreated genomic DNAs of three common Duffy phenotypes [Fy(a+b-), Fy(a-b+), Fy(a-b+)] by usual PCR technique using four oligonucleotide primers: gpFyl, gpFy579, gpFy735, and gpFyl218. Nucleotides of the PCR-amplified fragments were sequenced by direct sequencing method on both strands with a single-stranded template amplified by asymmetric PCR according to the procedure of Gyllensköld and Erlich.17 Analysis of the nucleotide sequences was performed with the DINA-SYS software package (Hitachi Software Engineering Co, Ltd, Yokohama, Japan).

RESULTS AND DISCUSSION

Southern blotting. When the genomic DNA was cut with Sac I and hybridized to the 5' probe, 1.1-kb and 3.2-kb fragments were made. In subsequent hybridization procedure with the 3' probe, the 3.2-kb fragment reappeared but the 1.1-kb band disappeared. When the genomic DNA was cut with EcoRI, the 5' probe gave 1.9-kb and 13.5-kb fragments. The successive reproving procedure showed that the 3' probe hybridized with 1.9-kb fragment but not with 13.5-kb fragment (Fig 1). These hybridization patterns were completely reproducible in four different DNA samples. The 5' probe contains one restriction sequence for Sac I and EcoRI, respectively. Therefore, we speculated that the 1.1-kb Sac I fragment encodes 5' flanking sequence of the gpFy gene and the 1.9-kb EcoRI fragment encodes 3' flanking sequence.

In addition, the 3' probe gave a 2.9-kb faint band in Sac I–digested DNA and an 18-kb faint band in EcoRI–digested DNA (arrowhead, Fig 1). Four members of the human IL-8 receptor gene family, in which two are functioning genes and two are pseudogenes, are thought to have evolved by replication of an ancestral gene.9,10 These faint bands indicate that the human gene contains other loci closely related to the gpFy gene.
Fig 2. The gpFy gene structure. (A) Restriction sites, cloning strategy, and oligonucleotide primers are shown in a scheme. Restriction sites indicate that the gpFy gene is almost entirely composed of 1.1-kb SacI and 1.9-kb EcoRI fragments. The 1.1-kb SacI fragment was amplified by IPCR procedure with a primer pair of gpFy149/108 and the 1.9-kb EcoRI fragment was amplified with gpFy1115/735, as illustrated by shaded bars. S, SacI; B, BamHI; N, NcoI; P, PstI; E, EcoRI. (B) The complete gene sequence is shown. Nucleotide and amino acid residue numbers correspond to the sequence deposited in GenBank [accession no. U01839]. The boundaries of 57-bp direct repeats in 5' flanking region are underlined. Putative regulatory elements representing binding sites are double-underlined. Two additional polyadenylation signals are underlined boldy. A tandem dinucleotide repeat DNAs are dashed-underlined in 3' flanking region. Indirect and direct repeat clusters are indicated with solid and dashed boxes, respectively.

Genomic cloning. DNA fragment with 1.1-kb length was amplified by the IPCR procedure using a primer pair of gpFy149/108 from covalently circular SacI-digested DNA as illustrated in Fig 2A. In negative study, no PCR products were observed from untreated genomic DNA. The 5' IPCR product includes the published gpFy cDNA sequence without introns as expected; nucleotides 1 to 108 and nucleotides 149 to 206 (SacI site). The 899-bp sequence that is intervening between the known sequences encodes the upstream sequence from transcription start site that was numbered +1 in Fig 2B. The accuracy of the cloned DNA sequence was confirmed by synthesizing an oligonucleotide that was located in the upstream sequence and by PCR and subsequent direct sequencing method with gpFy108 or gpFy735 primers. A search of GenBank with the 899-bp sequence did not show any significant homologies to other genes. In addition, the boundaries of 57-bp direct repeats are located at -674 to -567, underlined in Fig 2B. Identification of putative cis-regulatory sequence elements in the 1,074 bp upstream of the translation start codon was performed by computer-associated analysis, using the transcription factor database. The promoter sequence are very pyrimidine rich, with no obvious TATA or CCAAT boxes present. There are several potential regulatory elements, double-underlined in Fig 2: AP-1, HNF-5, TCF-1, ApoE B2, W-element, H-APF-1, and Sp-1. The GATA elements are located in the transcriptional regulatory sequence of erythroid specific genes without exceptions but not located in the gpFy promoter sequence. Northern blotting analysis shown by Chaudhuri et al and Neote et al indicated that the gpFy mRNA expression was not restricted in hematopoietic organs and was widely distributed in many organs, including lung, muscle, spleen, heart, pancreas, kidney, and brain with
difference of transriptional size. The absence of GATA element in gpFy promoter sequence is compatible with the ubiquitous distribution of gpFy mRNA. Moreover, the promoter sequence lacks the TATA box, CCAAT box, G + C rich stretch, and a high proportion of CpG dinucleotides. How this promoter sequence interacts with the transcription machinery to promote initiation must be decided in the future.

From covalently self-ligated EcoRI-digested genomic DNA, 1.5-kb DNA was amplified by the IPCR procedure with a primer pair of gpFy 1145/735 (Fig 2A). This 3' IPCR product included the gpFy cDNA sequence without introns, as expected; nucleotides 1145 to 1240 and nucleotides 259 through 264 (EcoRI site) to nucleotide 735. The intervening 929-bp encodes a novel sequence. As shown in Fig 2B, other than the polyadenylation signal used in the cDNA, two additional poly(A) signals, boldly underlined in Fig 2B, were identified in downstream sequence from the transcription termination site (arrowhead). The surrounding sequence of the additional poly(A) signals was adenine rich (44%). For instance, a tandem dinucleotide repeat DNAs (GT)15 are present in nucleotides 1780 through 1809 and the sequence was polymorphic between the two alleles of studied Fy(a+b+) gene, dashed underlined in Fig 2. In addition, 89% homologous 19-bp indirect repeat (IR) (nucleotides 1421 through 1439) was clustered with the 5' upstream sequence (nucleotides −174 through −192) and 85% homologous 21-bp direct repeat (DR) (nucleotides 1590 through 1610) was identified with 5' upstream (nucleotides −416 through −399), boxed in Fig 2B. The PCR of genomic DNA in the coding area also showed that the gpFy gene was not split by introns. The intron-lacking genes are thought to be generated by reverse-transcription mechanisms of processed mRNAs and introductions of cDNA copy into the gene.19,20

The DR and IR clusters, the multiple polyadenylation signals, and the intron-lacking organization indicate a possibility that the gpFy gene has been evolved by multiple retrotransposition events from an ancestral gene.

**Nucleotide sequence analysis of Duffy allotypes.** The published gpFy cDNA sequences were amplified independently from genomes of Fy(a+b−) individuals except for two base changes. One of the base changes is located in noncoding area at nucleotide 145, which is common in the three Duffy allotypes. The other base change is located at nucleotide 306, which is specific for Fy+ allele and results in substitution of residue 44 [GAT(Asp) → GGT(Gly)] (Fig 3A). The Fy(a−b+) samples showed the identical sequence at codon 44 with the published cDNA sequence, which was cloned from BM cDNA library established from Fy(a−b+) individuals.2 The base change at nucleotide 306 creates a third recognition site in the gpFy coding area for the restriction enzyme Ban I, which cleaves at GPpyPuCC but not GAPyPuCC. The PCR-RFLP analysis was performed with Ban I in 28 genetic independent samples that includes 11 Fy(a−b−), 10 Fy(a−b+), and 7 Fy(a−b+) individuals. In all samples, PCR with a primer pair of gpFy1/735 yielded the same 735-bp products. As illustrated in Fig 3B, the Ban I cleaved two common cutting sites and the specific site from Fy(a−b−) individuals and resulted in four fragments: 306-, 213-, 107-, and 109-bp fragments, respectively, whereas the Ban I digestion of PCR products from Fy(a−b+) individuals results in three fragments: 519-, 107-, and 109-bp fragments. Serologically determined heterozygotes, Fy(a+b−), four Fy(a−b+), and three Fy(a−b+) samples are shown. The right side is a undigested PCR product from a Fy(a−b+) individual. Size markers are on the left.

**Fig 3.** (A) DNA sequence analysis of amplified gpFy genes from an Fy(a−b−) and an Fy(a−b+) individual using a primer pair of gpFy1/735. The sequence from the Fy(a−b+) individual is identical to the previously published gpFy cDNA sequence except for a base difference at nucleotide 145 (G → A), which is common in the three Duffy phenotypes. The sequence of the Fy(a−b+) individual differs from that of the Fy(a−b−) in one base at codon 44, resulting in a GAT(Asp) → GGT(Gly) polymorphism. (B) The base change of Fy+ allele results in third restriction site for Ban I, as illustrated, and yields 519-, 213-, 107-, and 109 fragments. The PCR products followed by digestion with sufficient units of Ban I were analyzed on 3% agarose gel and stained with ethidium bromide. Five Fy(a−b−), four Fy(a−b+), and three Fy(a−b+) samples are shown. The right side is an undigested PCR product from a Fy(a−b+) individual. Size markers are on the left.
tially changes the antigenicity of the first extracellular domain. Therefore, it is proved that the gpFy cDNA and its gene described in this report encode the Duffy blood group system. However, the precise epitopes for anti-Fya and anti-Fyb antiseraums must be decided by further studies—eg, with construction of synthetic peptides or recombinant proteins. Finally, the genomic organization of Duffy-associated gpFy gene described in this report will contribute to the understanding of molecular basis of fourth Duffy phenotype, Fy(a−b−), in which the gpFy mRNA is not detectable by Northern blotting from the BM.2

ACKNOWLEDGMENT

We gratefully thank S. Fukuda (Red Cross Bloodbank Tochigi) for supplying blood samples and T. Oyamada for valuable technical assistance.

REFERENCES

Genomic organization of the glycoprotein D gene: Duffy blood group Fya/Fyb alloantigen system is associated with a polymorphism at the 44-amino acid residue

S Iwamoto, T Omi, E Kajii and S Ikemoto