Proc. Natl. Acad. Sci. USA
Vol. 86, pp. 4619–4623, June 1989
Genetics

Structural organization of glycophorin A and B genes: Glycophorin B gene evolved by homologous recombination at *Alu* repeat sequences

(gene duplication/exon–intron junction mutation/erythrocyte membrane)

SHINICHI KUDO AND MINORU FUKUDA

La Jolla Cancer Research Foundation, Cancer Research Center, 10901 North Torrey Pines Road, La Jolla, CA 92037

Communicated by Yuet W. Kan, March 30, 1989 (received for review January 13, 1989)

ABSTRACT **Glycophorins A (GPA) and B (GPB) are two major sialoglycoproteins of the human erythrocyte membrane. Here we present a comparison of the genomic structures of GPA and GPB developed by analyzing DNA clones isolated from a K562 genomic library. Nucleotide sequences of exon–intron junctions and 5′ and 3′ flanking sequences revealed that the GPA and GPB genes consist of 7 and 5 exons, respectively, and both genes have >95% identical sequence from the 5′ flanking region to the region ≈1 kilobase downstream from the exon encoding the transmembrane regions. In this homologous part of the genes, GPB lacks one exon due to a point mutation at the 5′ splicing site of the third intron, which inactivates the 5′ cleavage event of splicing and leads to ligation of the second to the fourth exon. Following these very homologous sequences, the genomic sequences for GPA and GPB diverge significantly and no homology can be detected in their 3′ end sequences. The transition site from homologous to nonhomologous sequences can be localized within *Alu* repeat sequences. The analysis of the *Alu* sequences and their flanking direct repeat sequences suggest that an ancestral genomic structure has been maintained in the GPA gene, whereas the GPB gene has arisen 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.**

Glycophorins A (GPA) and B (GPB) are major membrane sialoglycoproteins of human erythrocytes. GPA carries the M or N blood group antigen, which is determined by amino acids at residues 1 and 5. GPB carries N blood group antigen as well as S or s blood group antigen, which is determined by the amino acid residue at position 29 (for review, see ref. 1). The amino acid sequences of GPA and GPB and their cDNA sequences have been determined (2–8). Comparison of these sequences shows that GPA and GPB share highly homologous amino acid sequences, particularly in the first 26 N-terminal residues, which are completely identical when group N GPA and GPB are compared (5, 6). The nucleotide sequences derived from cDNA clones encoding these regions are almost identical (5, 8). Another highly homologous region is observed in the transmembrane domain. On the other hand, a segment (amino acid residues 27–58) of GPA is absent from GPB, and GPA has an additional 30 amino acid residues at the C terminus. Furthermore, cDNA sequences show that the 3′ untranslated region of GPA is quite different from that of GPB (5, 7, 8). Nevertheless, both genes have structures similar enough to assume they derived from a common ancestral gene. It also has been shown that the expression of GPA and GPB are regulated in a similar manner. Both GPA and GPB are specifically expressed on erythroid cell lineage (9, 10) and both are decreased substantially when erythroleukemic K562 and HEL cells are treated with tumor-promoting phorbol esters (4, 5).

We have recently shown by genomic DNA hybridization with oligonucleotide probes that GPA and GPB are encoded by separate and distinct genes (4, 5). In an effort to further delineate the genomic structure, gene evolution, and regulated expression of the glycophorins, we have isolated genomic clones that span >30 kilobases (kb) each of the GPA and GPB genes from a K562 genomic library. The nucleotide sequencing results presented here indicate that the GPB gene evolved from an ancestral gene by homologous recombination through *Alu* repeats during or after gene duplication and by nucleotide substitutions, including those at splicing sites.*

EXPERIMENTAL PROCEDURE

Materials. Restriction endonucleases, T4 DNA ligase, T4 kinase, and DNA polymerase I (Klenow fragment) were purchased from Bethesda Research Laboratories. Calf intestine alkaline phosphatase was from Boehringer Mannheim. Phage vector Lambda FIX, Gigapack *in vitro* packaging kit, and Bluescript plasmid vector were obtained from Stratagene. *Escherichia coli* MB406 was obtained from Promega. Radiolabeled nucleotides were supplied by New England Nuclear. The oligolabeling kit was purchased from Pharmacia. Oligonucleotides were synthesized on an Applied Biosystems synthesizer and purified as described (5). Sequenase sequencing kit was obtained from United States Biochemical.

Cell Culture and Genomic DNA Isolation. Human erythroleukemia cell line K562 was cultured and maintained in RPMI 1640 medium supplemented with 10% fetal calf serum. High molecular weight DNA was isolated from K562 cells with $NaDodSO_4$/proteinase K treatment following phenol/chloroform extraction as described (11).

Construction and Screening of Genomic DNA Libraries. K562 genomic DNA was partially digested with restriction endonuclease *Mbo* I and the resulting cohesive ends were modified so as to fill in the first two nucleotides, using the Klenow fragment of *E. coli* DNA polymerase and dATP and dGTP as substrates. Resultant genomic fragments were cloned without size selection into the phage vector Lambda FIX (Stratagene), which had been treated as follows. Lambda FIX DNA was completely digested with *Xho* I endonuclease and the cohesive ends were partially filled in with Klenow fragment and dCTP and TTP to make the termini complementary to those of the genomic fragments. These partial fill-in procedures inhibit the self-ligation of insert or vector. After ligation of genomic fragment to the vector with T4 DNA ligase, DNAs were packaged *in vitro* using Gigapack Gold (Stratagene). The library was plated on *E. coli* MB406 and

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

Abbreviations: GPA and GPB, glycophorins A and B.
*The sequence reported in this paper has been deposited in the GenBank data base (accession nos. M24123–M24137).

screened by the *in situ* hybridization method described elsewhere (12). Two oligonucleotides, LP and GC, were synthesized and labeled by T4 kinase and [γ-^{32}P]ATP and used for hybridization as described (5). LP (37-mer) corresponds to the nucleotide sequence in the leader peptide coding region, starting with the methionine codon, whereas GC (27-mer) corresponds to the nucleotide sequence encoding amino acid residue 18 (lysine) to residue 26 (asparagine). Screening was repeated until all plaques showed positive signals. Recombinant phage DNAs were purified by the method of Leder *et al.* (13) and identified and characterized by Southern blot (14) hybridization analysis. The cDNA probes were made by the random-oligonucleotide primer-extension method (15).

Sequence Determination. DNA inserts in phage DNA were subcloned into Bluescript vectors (Stratagene). Plasmids were purified by the alkaline lysis method followed by CsCl ultracentrifugation (16). Double-stranded DNA clones were sequenced by the dideoxynucleotide chain-termination method (17) using synthetic oligonucleotides as primers and a Sequenase sequencing kit. Sequences obtained by these methods were analyzed by Beckman Microgenie software.

RESULTS AND DISCUSSION

Isolation of GPA and GPB Genomic Clones. A genomic DNA library was constructed in phage vector Lambda FIX from total DNA of human erythroleukemia cell line K562. We chose this cell line because we had already obtained glycophorin cDNAs from this cell line and partially analyzed genomic structures using cDNA sequences (3–5). Both GPA and GPB genes in K562 cells were shown to have the same structures as those of normal human leukocytes by Southern hybridization analysis (unpublished data). Furthermore, the nucleotide sequences of these cDNAs are identical to those isolated from normal human reticulocyte cDNA libraries (7, 8). First, $\approx 10^6$ Lambda FIX recombinants were screened in duplicate with the 37-mer oligonucleotide probe LP, whose sequence corresponds to the leader peptide of glycophorin and is identical in both GPA and GPB cDNA sequences. After purification of phage DNAs from 16 positive clones, Southern hybridization with *Bgl* II digests of these DNAs was performed with the same oligonucleotide probe. Previous results (data not shown) suggested that three different sizes of *Bgl* II fragments, 2.4, 2.6, and 3.8 kb, could be detected by Southern hybridization with this probe to total K562 genomic DNA. The isolated clones were therefore divided into three groups according to the size of hybridizing fragments. One clone was chosen from each group and the hybridizing fragment was subcloned into the *Bam*HI site of Bluescript for DNA sequencing. Clone LP1-3 of the 2.4-kb positive group contains the same 5′ sequence upstream from the leader peptide coding region as that of GPA cDNA, which was isolated in our laboratory and characterized as group M type by its predicted amino acid sequence (3, 5). On the other hand, clone LP1-8 (2.6-kb positive group) contains the same 5′ upstream sequence as that of GPB cDNA isolated by Tate and Tanner (8) and characterized as group N type. Therefore, clones LP1-3 and LP1-8 were judged to be derived from the GPA and GPB genes, respectively (Fig. 1). The clone LP1-2 of the 3.8-kb positive group contains a 5′ upstream sequence different in two nucleotides from those of the glycophorin cDNA sequences available so far and was not investigated further. In these clones the exon encoding two-thirds of the leader peptide was identified. But the following exons were not found in these clones by Southern analysis with glycophorin cDNA probes.

To obtain clones carrying the downstream genomic region, we screened the same library with the GC oligonucleotide probe, whose sequence is also perfectly shared in GPA and GPB cDNA sequences (5). A total of 25 clones were isolated and analyzed by Southern hybridization. Our previous results showed that a 2.0-kb fragment is reactive with the GC oligonucleotide probe in *Eco*RI-digested total genomic DNA (4, 5). *Eco*RI fragments with the same size were detected in two clones, C2 and C25. Clone C2 has four *Eco*RI fragments, 2.0, 3.5, 2.8, and 1.25 kb, which are also detected in total genomic DNA hybridization by GPA cDNA (4). On the other hand, clone C25 has three *Eco*RI fragments, 2.0, 3.5, and 4.2 kb, which are detected in total genomic DNA by GPB cDNA (HGpB-1; ref. 5) (unpublished results). Clones C2 and C25 were therefore expected to be derived from the GPA and GPB genes, respectively (Fig. 1). *Eco*RI and/or *Sal* I fragments from each genomic insert were subcloned into Bluescript and their nucleotide sequences were determined.

Organization of the GPA Gene. To understand the organization of the GPA gene, the first attempt was made to define the transcription initiation site by S1 nuclease mapping and primer-extension analysis. For S1 nuclease mapping, a *Bgl* I/*Fok* I-digested fragment was used, whereas LP oligonucle-

FIG. 1. Structure of GPA and GPB genes. The gene organization was constructed by combining two different groups of genomic clones, which represent the extreme 5′ end and the rest of the genes. Each set of three lines (from top to bottom) shows *Eco*RI sites with sizes between sites in kb, genomic structures, and genomic clones in which phage flanking sequences are indicated by zigzags. For brevity, the restriction sites are shown only for *Eco*RI. Exons are indicated by vertical lines or solid boxes, according to their size. The recombination site (or breakpoint) is indicated by BRP. Exon 1 and exon 2 are separated by a >30-kbp intron.

otide (see above) was used for primer extension. The results of both experiments agree well within a few bases, and adenosine, which is 55 nucleotides upstream from the adenosine for the initiaion methionine, was identified to be the transcription initiation site (data not shown). It is noteworthy that this initiation site is the same as the uppermost 5′ end of GPA cDNA recently identified by Tate and Tanner (8).

Nucleotide sequences were determined around exon–intron junctions and at 5′ and 3′ flanking regions using synthetic oligonucleotides based on known GPA cDNA sequences as primers. Each boundary was confirmed by comparison of obtained sequences to cDNA sequences. These results are shown in Fig. 2 and reveal the following features. Two-thirds of the leader peptide is encoded in the first exon and one-third is in the second exon. These two exons are interrupted by a large intron of >30 kb. The length of the first intron was estimated by isolation of genomic clones from a K562 cell cosmid library. One cosmid clone, isolated by LP oligonucleotide as a probe, encompasses close to 30 kb downstream from exon 1 but was found not to contain exon 2 (unpublished results). Further studies are necessary to define completely intron 1. Blood groups M and N are specified by both the first and fifth amino acid residues of mature GPA protein, and these are encoded in the second exon. In clone C2, the first codon specifies serine and the fifth glycine, indicating that this clone was derived from the group M GPA gene. The second exon is spliced to the third exon, which is unique in the GPA gene. The third exon is followed by a 0.9-kb intron and is spliced to the fourth exon. In this exon, the genomic DNA sequences diverge between GPA and GPB genes when compared with the first and second exons. The fourth exon was spliced to the fifth exon, which contains sequences encoding the transmembrane domain. The fifth exon was spliced to the sixth, a unique exon in the GPA gene, interrupted by a 2.1-kb intron (Fig. 2*B*). Finally, this exon is spliced to the exon encoding the 3′ untranslated region, with removal of a 3.2-kb intron to form the mature mRNA. All the 5′ donor and 3′ acceptor sites conform with the exon---/gt---intron---ag/---exon consensus sequence (18). Also, there are branch points in every intron that are necessary for lariat formation of precursor mRNA (19, 20), positioned within 37 base pairs (bp) from the 3′ splicing sites, as indicated by asterisks in Fig. 2.

Organization of the GPB Gene. Nucleotide sequences of the GPB gene are found to be very similar to those of the GPA gene up to the region ≈1 kb downstream from exon 5. Splicing sites between the first and second exon are the same sites as found in the GPA gene. The first nucleotide of the second exon is changed from that of the GPA gene, explaining the difference in the amino acid at position −7 in the leader peptide. The first and fifth amino acid residues of a mature GPB protein encoded in the second exon are leucine and glutamic acid, respectively, showing that this clone was certainly derived from a GPB gene (Fig. 2*A*).

By analysis of GPB cDNA, the GPB gene lacks exon 3 (5). However, nucleotide sequence determination of GPB genomic DNA revealed the existence of sequences quite similar to exon 3 and the adjacent introns of the GPA gene. There are two distinct points of change in this region, which are likely to affect splicing, compared with that of the GPA gene. The GPA gene has two possible branch sites in the second intron, with the sequence of the 5′ branch site perfectly conforming

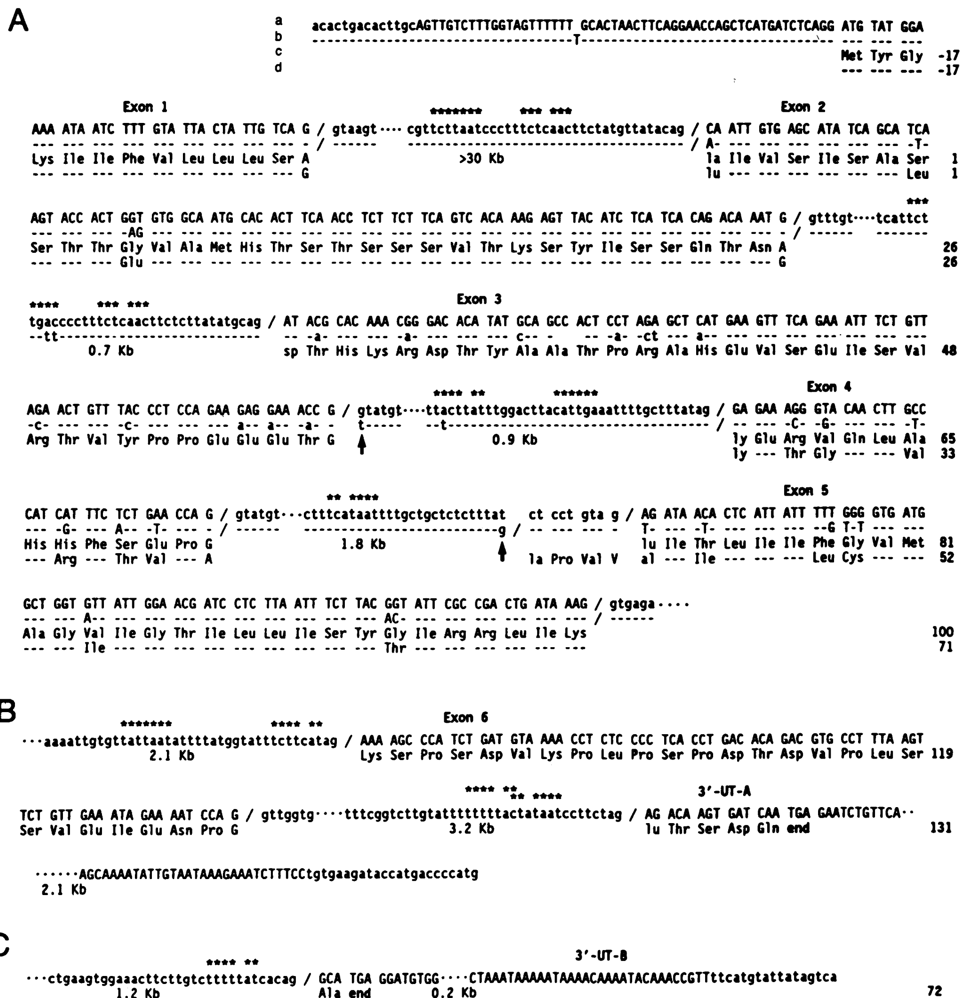

FIG. 2. Comparison of the nucleotide sequences of GPA and GPB genes. The sequences of all exons, exon–intron junctions, 5′ and 3′ flanking regions, and the amino acid sequences encoded by exons are described. (*A*) The sequences in which GPA and GPB genes are almost identical. Arrows indicate two point mutations in GPB genes, which result in different exons. Each set of four lines (from top to bottom) shows the nucleotide sequences of GPA and GPB (a and b) followed by amino acid sequences (c and d). Identical sequences are shown by dashes. (*B* and *C*) The 3′ end sequence, which is unique to the GPA (*B*) or the GPB (*C*) gene. The sequences in *B* and *C* are those directly following the transition site from homologous to nonhomologous sequences (BRP in Fig. 1). The splicing branch points are indicated by asterisks. Amino acid residues are numbered at the far right.

with the consensus sequence (YNYTRAY) (Y, pyrimidine; R, purine) (19). But the GPB gene has substitutions in this site, changing TCTTGAC to TCTTGTT. The second significant difference in this region is that the 5′ splice site of intron 3 is disrupted in the GPB gene. The change from CG/gtat to cgttat resulted in the loss of the 5′ splice cleavage site (arrow in Fig. 2*A*). Because of this disruption in the 5′ splice site, the third exon of the GPB gene is skipped and direct ligation between the second and fourth exon apparently takes place. Two mechanisms are possible for explaining these results. One is that the mutation in the branch site of intron 2 abolishes the function of the 3′ splice site so that intron 2, mutated exon 3, and intron 3 work as a single intron. In this case, the presumptive branch point, tctcaac in intron 2, does not work as a branch point, but the presumptive branch point in intron 3 must work as a branch point. The second and most interesting possibility is that the splicing machinery moves from the 3′ to 5′ end of precursor RNA so that the 5′ cleavage of intron 3 as well as the 3′ cleavage of intron 2 are ignored. In relation to this result, it is noteworthy that one type of β-thalassemia was produced by a similar splicing pattern due to a mutation at the 5′ splice site (21). Since no mutation was detected at the upstream splice sites, the results strongly favor the 3′ to 5′ movement of splicing machinery.

Compared with GPA, GPB has an additional three amino acid residues in the transmembrane domain encoded by exon 5. This difference is due to a change at the 3′ splice site generated by a single base pair substitution (T → G), resulting in the formation of a ttag/CT sequence from ttatct (arrow in Fig. 2*A*). In addition, the first nucleotide in exon 5 of the GPA gene is substituted from adenine to thymidine in the GPB gene. Therefore, a 3′ splice site 9 bp upstream is used, and an additional three amino acids are encoded by 9 nucleotides of in-frame intron sequence (Fig. 2*A*).

Comparison of cDNA sequences of GPA and GPB shows that the GPB gene lacks exon 6. To know whether there is a region corresponding to this exon in the GPB gene, Southern hybridization of clone C25 was performed with a synthetic oligonucleotide used as probe. The oligonucleotide GA was synthesized according to the sequence of GPA exon 6 (see ref. 5). Hybridization was not detected, even under conditions of low stringency (data not shown). Sequencing studies also confirmed that GPA and GPB genes have quite different structures downstream from exon 5 and showed that exon 5 is spliced to the exon encoding the 3′ untranslated region, which is unique to the GPB gene (Fig. 2*C*).

Transition Sites in GPA and GPB Genes. To find the sites of transition from homologous to nonhomologous sequences in the GPA and GPB genes, we sequenced from the 5′ end of a 2.8-kb subclone of the GPA gene and a 4.2-kb subclone of the GPB gene, which are adjacent to exon 5 (Fig. 1). Highly homologous sequences continued until 350 bp distant from an upstream *Eco*RI site. Within this region, computer search analyses revealed the existence of human *Alu* repeat sequences. The *Alu* sequences lie in inverted orientation and are composed of 304 bp in GPA and 294 bp in GPB genes, respectively (Fig. 3*A*). The difference in size is mainly due to a variation in the length of the 5′ poly(T) stretch. In general, *Alu* sequences are flanked on both sides by direct repeats of 7–20 bp, suggesting structures similar to mobile genetic elements (22, 23). In fact, the transposition of *Alu* element led to inactivation of the target gene when cells were fused with UV-irradiated cells (24). The GPA gene has direct repeats consisting of 18 bp with 3-bp mismatches having 83.3% homology. In contrast, the direct repeats in the GPB gene contain only 65.5% homology (Fig. 3*B*). This latter value is significantly low, because the average homology estimated by Schmid and Shen using a data base is 88% (25). On the other hand, the upstream element is well conserved between GPA and GPB genes with only a 1-bp mismatch. Therefore,

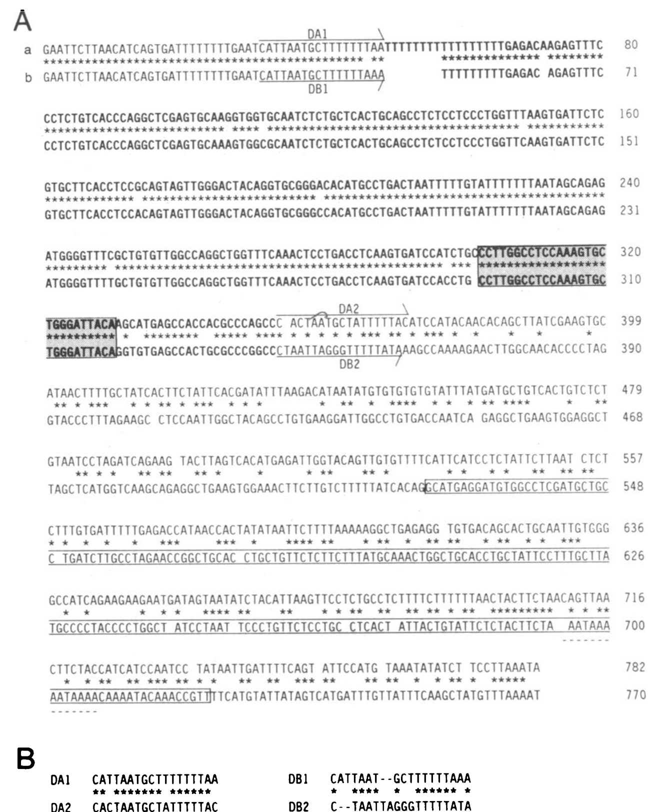

FIG. 3. (*A*) Nucleotide sequences around the transition site from homologous to nonhomologous sequences in GPA and GPB genes. Each set of two lines shows the nucleotide sequences of GPA (a) and GPB (b) genes. Identical sequences are indicated by asterisks and *Alu* repeats are boldface (nucleotides 50–353 or 343). The exon for the 3′ untranslated region of the GPB gene is boxed and the polyadenylylation signal is underlined by dashes. Arrows indicate the direct repeats flanking *Alu* sequences. The presumptive recombination site is indicated by a shaded box. These sequences represent the 5′ ends of 2.8-kb or 4.2-kb segments (see Fig. 1). (*B*) Comparison of the direct repeats flanking *Alu* sequences. DA1 and DA2, GPA direct repeats; DB1 and DB2, GPB direct repeats, which are indicated in *A*.

it is thought that the upstream repeat sequences of GPA and GPB genes were derived from the same origin, but the downstream repeat sequences were derived from different origins. Since the direct repeats were conserved in the GPA gene, the original genomic structure around the *Alu* repeat is represented by the GPA gene. It is likely, therefore, that the GPB gene acquired a different *Alu* sequence in the downstream sequence. Although it is difficult to identify the exact recombination point in *Alu* repeats, the homology between GPA and GPB sequences is lost within the last 23 bp of the 3′ end of *Alu* repeats having 6-bp mismatches. These results strongly suggest that recombination took place somewhere upstream of the 3′ end of the *Alu* sequences, as shown in Fig. 3*A*. This recombination site in *Alu* sequences lies between A and B promoter sequences for RNA polymerase III (32), as has been observed in the majority of recombination sites in *Alu* sequences (26–28). The results strongly suggest that recombination might be enhanced by the interaction between the transcription apparatus and these control regions. Downstream from the *Alu* sequences, there is no significant homology between GPA and GPB genes, indicating that homologous recombination in this gene segment was driven by only one *Alu* repeat.

Based on these results, it can be postulated that the GPB gene was produced by duplication of the original gene, diversion by recombination through *Alu* repeats, and point

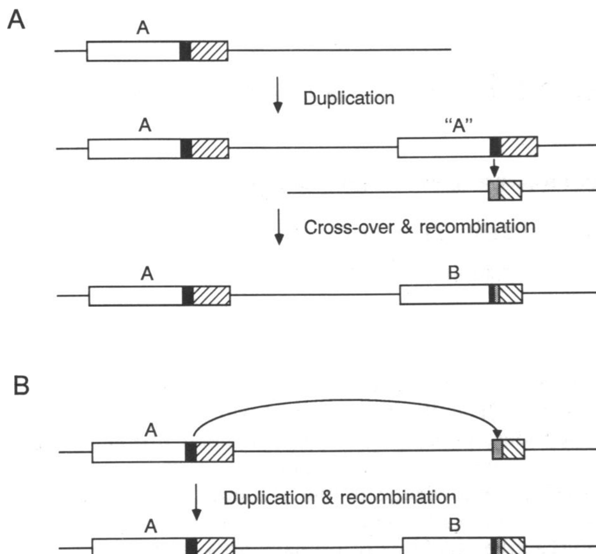

FIG. 4. Two possible mechanisms for evolution of GPA and GPB genes. (*A*) The original gene was first duplicated and the duplicated gene acquired the GPB-specific 3′ end by cross-over at *Alu* repeats. (*B*) Duplication of the original gene took place at *Alu* sequences to generate the GPB-specific 3′ end in one step. In both cases, the original gene is a direct precursor to GPA genes. *Alu* sequences in GPA and GPB genes are indicated by a solid box, and solid and stippled boxes, respectively. Hatched boxes indicate 3′ end sequences unique to GPA or GPB genes, while open boxes indicate the homologous sequences shared by GPA and GPB genes.

mutation of the duplicated gene. By analyzing the sequences in the homologous region, it is possible to estimate the period after the gene duplication. Intron sequences flanking exon 1 to exon 5 between the GPA and GPB genes have 95.5% identity in a total 4-kb sequence determined so far. These regions are expected to evolve at the rate for neutral evolution. It was therefore estimated that the duplication of the ancestral gene took place about 9–35 million years ago, according to two different divergence rates at 0.5% per million years for the average mammalian rate (29) or at 0.13% per million years for the average primate rate (30).

It is possible that *Alu* repeats were directly involved during duplication (Fig. 4*B*) or during recombination upon cross-over after the gene was duplicated (Fig. 4*A*). In both cases, homologous recombination through *Alu* repeats resulted in a 3′ end sequence unique to the GPB gene and the loss of exon 6. The original gene, on the other hand, directly evolved into the GPA gene. Most of the human population has the GPB gene. It is significant that homologous recombination of *Alu* repeats was involved in evolution of a gene existing generally in humans. It has been proposed that *Alu*-related repeated sequences function as origins for chromosomal replication (31). If this is the case, it also explains why *Alu* repeats may serve as recombination sites. It is therefore possible that *Alu*–*Alu* homologous recombination plays a role in generating a number of genes.

We thank Gizella S. Nagy and Todd Dubnicoff for excellent technical assistance, Dr. Mark Williams for critical reading of the manuscript, and Tami Clevenger and Kaarin Soma for secretarial assistance. This work was supported by Grant R01 CA33000 awarded by the National Cancer Institute.

1. Dahr, W. (1986) in *Recent Advances in Blood Group Biochemistry*, eds. Vengelen-Tyler, V. & Judd, W. J. (American Association of Blood Banks, Arlington, VA), pp. 23–65.
2. Tomita, M., Furthmayr, H. & Marchesi, V. T. (1978) *Biochemistry* **17,** 4756–4770.
3. Siebert, P. D. & Fukuda, M. (1986) *Proc. Natl. Acad. Sci. USA* **83,** 1665–1669.
4. Siebert, P. D. & Fukuda, M. (1986) *J. Biol. Chem.* **261,** 12433–12436.
5. Siebert, P. D. & Fukuda, M. (1987) *Proc. Natl. Acad. Sci. USA* **84,** 6735–6739.
6. Blanchard, D., Dahr, W., Hummel, M., Latron, F., Beyreuther, K. & Cartron, J. P. (1987) *J. Biol. Chem.* **262,** 5808–5811.
7. Rahuel, C., London, J., d'Auriol, L., Mattei, M. G., Tournamallie, C., Skrzynia, C., Lebouc, Y., Galibert, F. & Cartron, J. P. (1988) *Eur. J. Biochem.* **172,** 147–156.
8. Tate, C. G. & Tanner, M. J. A. (1988) *Biochem. J.* **254,** 743–750.
9. Gahmberg, C. G., Jokinen, M. & Andersson, L. C. (1978) *Blood* **52,** 379–387.
10. Fukuda, M. & Fukuda, M. N. (1984) in *The Biology of Glycoproteins*, ed. Ivatt, R. J. (Plenum, New York), pp. 183–234.
11. Friedrich, T. O. & Lehman, J. M. (1981) *Virology* **110,** 159–161.
12. Woods, D. E., Markham, A. F., Ricker, A. T., Goldberger, G. & Colton, H. R. (1982) *Proc. Natl. Acad. Sci. USA* **79,** 5661–5665.
13. Leder, P., Tiemeier, D. & Enquist, L. (1977) *Science* **196,** 175–177.
14. Southern, E. M. (1975) *J. Mol. Biol.* **98,** 503–517.
15. Feinberg, A. & Vogelstein, B. (1984) *Anal. Biochem.* **137,** 266–267.
16. Maniatis, T., Fritsch, E. F. & Sambrook, J. (1982) *Molecular Cloning: A Laboratory Manual* (Cold Spring Harbor Lab., Cold Spring Harbor, NY).
17. Sanger, F., Nicklen, S. & Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. USA* **74,** 5463–5467.
18. Breathnach, R. & Chambon, P. (1981) *Annu. Rev. Biochem.* **50,** 349–383.
19. Ruskin, B., Krainer, A. R., Maniatis, T. & Green, M. R. (1984) *Cell* **38,** 317–331.
20. Padgett, R. A., Grabowski, P. J., Konarska, M. M., Seiler, S. & Sharp, P. A. (1986) *Annu. Rev. Biochem.* **55,** 1119–1150.
21. Treisman, R., Proudfoot, N. J., Shandler, M. & Maniatis, T. (1982) *Cell* **29,** 903–911.
22. Schmid, C. W. & Jelinek, W. R. (1982) *Science* **216,** 1065–1070.
23. Calabretta, B., Robberson, D. L., Barrera-Saldana, H. A., Lambrou, T. P. & Saunders, G. F. (1982) *Nature (London)* **296,** 219–225.
24. Lin, C. S., Goldthwait, D. A. & Samols, D. (1988) *Cell* **54,** 153–159.
25. Schmid, C. W. & Shen, C. K. J. (1985) in *Molecular Evolutionary Genetics*, ed. MacIntyre, R. J. (Plenum, New York), pp. 323–358.
26. Lehrman, M. A., Goldstein, J. L., Russel, D. W. & Brown, M. S. (1987) *Cell* **48,** 827–835.
27. Nicholls, R. D., Fishel-Ghodsian, N. & Higgs, D. R. (1987) *Cell* **49,** 369–378.
28. Rouyer, F., Simmler, M. C., Page, D. C. & Weissenbach, J. (1987) *Cell* **51,** 417–425.
29. Hayashida, H. & Miyata, T. (1983) *Proc. Natl. Acad. Sci. USA* **80,** 2671–2675.
30. Britten, R. J. (1986) *Science* **231,** 1393–1398.
31. Chambers, J. C., Watanabe, S. & Taylor, J. H. (1982) *Proc. Natl. Acad. Sci. USA* **79,** 5572–5576.
32. Paollela, G., Lucero, M. A., Murphy, M. H. & Baralle, F. E. (1983) *EMBO J.* **2,** 691–696.