

Analysis of 5' Flanking Regions of the Gamma Globin Genes from Major African Haplotype Backgrounds Associated with Sickle Cell Disease

Stacy R. Month, Roger W. Wood, Panayiota T. Trifillis, Philip J. Orchowski, Bruce Sharon,* Samir K. Ballas,†
Saul Surrey, and Elias Schwartz

Division of Hematology, The Children's Hospital of Philadelphia, and the Departments of Pediatrics and Human Genetics, The University of Pennsylvania School of Medicine, Philadelphia, Pennsylvania 19104; *Division of Pediatric Hematology/Oncology, Department of Pediatrics, University of Illinois College of Medicine at Chicago, Chicago, Illinois 60612; and †Department of Medicine, Cardeza Foundation for Hematologic Research, Thomas Jefferson University, Philadelphia, Pennsylvania 19107

Abstract

There are at least three major African haplotype backgrounds on which the β^s mutation arises. Sequence changes in the immediate 5' flanking area of the γ -globin genes may account for differences in fetal hemoglobin expression among the three haplotypes. We determined the sequence from -350 to 10 bp 5' of the G γ and A γ fetal globin genes from one β^s -containing chromosome on each of the three major haplotype backgrounds. The Senegal chromosome had a T at -158 5' to the G γ gene; the Benin (BEN) chromosome had an A to G change at -309 5' to the G γ gene; and the Central African Republic (CAR) chromosome had a C to T change at -271 5' to the A γ gene. Genomic DNA from patients with sickle cell disease was analyzed using the polymerase chain reaction and radiolabeled allele-specific oligonucleotide probes. The -309 G variant 5' to the G γ gene is associated with BEN chromosomes, and the -271 T variant 5' to A γ with CAR. The -309 change was also found on β^A -containing chromosomes, while the -271 change was not. The -309 change may have predated the β^s mutation on the BEN chromosome. (*J. Clin. Invest.* 1990. 85:364-370.) fetal globin genes • haplotypes • sickle cell disease

Introduction

The β^s mutation causing sickle cell disease arose in Africa independently on at least three distinct chromosomal backgrounds or haplotypes, which are defined by the pattern of polymorphic restriction enzyme sites within and flanking the β -like globin gene cluster (1, 2). The three haplotypes are geographically distinct and have been extensively studied in Senegal, Benin, and the Central African Republic (CAR).¹

The variability in hematologic characteristics between different patients with homozygous sickle cell disease (Hb SS)

may be partially explained by the association with specific haplotypes (1, 2). Sickle cell patients from Senegal have a high hemoglobin F (Hb F) (mean 12.1%) and a low percentage of irreversibly sickled cells. Their Hb F has a high percentage of G γ chain (mean 69.7%), a finding typical of normal infants (70% G γ /30% A γ), and unlike the 40% G γ usually found after the first year of life. In contrast, the Benin patients have a low Hb F (mean 7.4%), a higher proportion of ISC, and a G γ /A γ ratio of < 50%. The CAR patients have a relatively higher level of Hb F (mean 10.4%) and a low percentage of irreversibly sickled cells, as in Senegal Hb SS patients, while they have a low percentage of G γ (mean 40%), as do Benin Hb SS patients. The CAR patients can be further divided into two subsets, one with G γ < 38% and a low level of Hb F (mean 6.3%), and the other with G γ > 38% and a higher Hb F level (mean 12.5%) (3-6).

Genetic factors linked to the β -globin gene cluster on chromosome 11 have been implicated in potentiating the high ratio of G γ to A γ (7, 8). The subhaplotype (+ - + +), referring to Hind III and Hinc II polymorphic sites within the γ - and the $\psi\beta$ -globin gene regions, is found in Senegal sickle cell patients. This subhaplotype has also been associated with a high ratio of G γ to A γ in a variety of sickle cell patients from Georgia, Eastern Saudi Arabia, and one from Jamaica with Asian-Indian ancestry, as well as in patients with β -thalassemia from Algeria and Georgia (7, 9-12). Linkage of this subhaplotype with a high ratio of G γ to A γ suggests that the area encompassing the G γ -A γ - $\psi\beta$ region may contain a critical control element.

There is ample evidence from naturally occurring and induced mutations that the 5' flanking regions of the γ -globin genes play an important role in expression. The "ATAA," "CCAAT," and "CACC" boxes are localized in this region and are known to play a critical role in expression of human globin genes (13-16). Single-base changes occurring within a few hundred bases upstream from the γ -globin genes have been associated with nondeletional forms of hereditary persistence of fetal hemoglobin (HPFH) (17, 18 and Fig. 1). Mutations at -198, -196, -175, or -117 bp from the A γ -cap site are associated with A γ - β^+ -HPFH, while changes at -202 or -175 5' to the G γ gene are associated with G γ - β^+ -HPFH (15-23). Additional changes at -202 5' to A γ and at -158 and -161 5' to G γ have also been found associated with elevated levels of A γ and G γ , respectively, in patients with mild forms of sickle cell disease (24).

Restriction enzyme analysis shows that a T residue at -158 5' to G γ , which creates an Xmn I site, is associated with a high ratio of G γ to A γ and the (+ - + +) subhaplotype in sickle cell patients from Senegal, Georgia, Turkey, Surinam, and Eastern Saudi Arabia (9, 25, 26). This Xmn I site is also associated with

Address correspondence to Dr. Elias Schwartz, Director, Division of Hematology, Children's Hospital of Philadelphia, 34th St. and Civic Center Blvd., Philadelphia, PA 19104.

Received for publication 12 June 1989.

1. Abbreviations used in this paper: BEC, buccal epithelial cells; BEN, Benin; CAR, Central African Republic; Hb F, hemoglobin F; Hb SS, homozygous sickle cell disease; HPFH, hereditary persistence of fetal hemoglobin; MCV, mean cell volume; PCR, polymerase chain reaction.

J. Clin. Invest.

© The American Society for Clinical Investigation, Inc.

0021-9738/90/02/0364/07 \$2.00

Volume 85, February 1990, 364-370

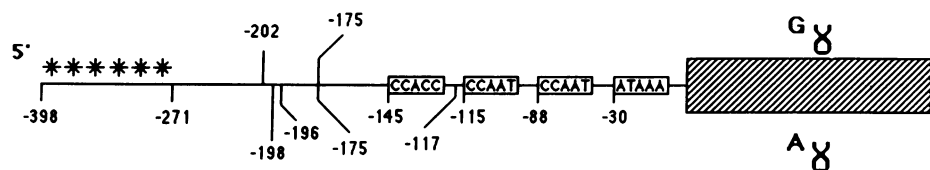


Figure 1. The immediate 5' flanking regions of G γ - and A γ -globin genes. Position of base changes associated with nondeletional forms of HPFH are indicated above line (-175 and -202) for G γ - and below (-198, -196, -175, and -117) for A γ - β^+ -HPFH. The boxes refer to the CCACC (-145), CCAAT (-115 and -88), and ATAAA (-30) boxes, while asterisks indicate the six single-base sequence differences between G γ and A γ in the 5' flanking region up to -471 bp from the cap sites.

a high G γ to A γ ratio in β -thalassemia patients from Algeria (4). Only one chromosome associated with the base change is necessary for elevated G γ to A γ ratios.

DNA sequence differences in the areas flanking the γ -globin genes might account for differences in fetal globin gene expression associated with the three sickle cell haplotypes. Therefore, we isolated fetal globin gene clones from the three major haplotypes and determined the sequence of the immediate 5' flanking regions. The association between sequence differences and β^s haplotype backgrounds was then determined by screening genomic DNA samples using dot blot analysis of polymerase chain reaction (PCR) products.

Methods

Haplotype analysis. Analysis of DNA within the β -like globin gene cluster from patients with sickle cell disease was performed as previously described (2, 27) using the following enzymes and probes: a 1.6-kb Cfo I γ -cDNA fragment isolated from JW151 (kindly supplied by B. Forget, Yale University) (28) was used to probe for Hind III polymorphisms within the fetal globin genes, and a 1.7-kb genomic Bgl II/Xba I fragment containing $\psi\beta$ isolated from pP3.9 (kindly supplied by T. Maniatis, Harvard University) (29) was used to probe for Hinc II polymorphisms within and 3' to $\psi\beta$. Some samples were further analyzed using a 1.1-kb Eco RI fragment from pRK29 (kindly supplied by R. Kaufman, University of North Carolina) (30) to probe for Hind III polymorphisms 3' to the β -globin gene, and/or a 550-bp genomic Rsa I fragment to probe for the Rsa I polymorphic site 5' to the β -globin gene (27). Data for American and Jamaican sickle cell populations show that 89% of sickle cell patients possess one of the three major African haplotypes, characterized by 11 restriction enzyme polymorphic sites, and the remaining 11% comprise minor haplotypes (2). By analyzing the Hinc II, Hind III, and Rsa I polymorphic sites, we could make haplotype assignments and assume the presence of a major haplotype with ~ 90% certainty (1, 12, 31).

Fetal globin genes on the Benin (BEN) haplotype background were obtained for DNA sequence analysis from a previously described patient who is heterozygous for G γ - β^+ -HPFH on one chromosome and the sickle mutation on the other (18, 32, 33). The Senegal (SEN) and CAR chromosome fragments were from two patients with sickle cell disease. Hb F and Hb A $_2$ levels were determined by alkali denaturation and DEAE cellulose chromatography methods, respectively (34, 35). Mean cell volume (MCV) was determined on an electronic counter (model S+3; Coulter Electronics, Inc., Hialeah, FL). Ratios of G γ to A γ were measured using reverse-phase, fast-protein liquid chromatography (Pharmacia Fine Chemicals, Piscataway, NJ). Restriction enzymes Bam HI, Bcl I, Bgl II, Hinc II, Hind III, Rsa I, Sac I, Sau 3A I, and Sst I were used under buffer conditions recommended by the supplier (Bethesda Research Laboratories, Gaithersburg, MD, or New England Biolabs, Beverly, MA).

Preparation of genomic libraries. Genomic DNA from the three patients described above was isolated from peripheral blood leukocytes (36), digested overnight with Bcl I or for varying times with Sau 3A I, and fractionated by sucrose gradient sedimentation (37). DNA fragments ~ 18 kb in size were isolated from gradients, dialyzed overnight against 10 mM Tris HCl, pH 7.8/1 mM EDTA, and the DNA was then extracted with an equal volume of *sec*-butanol, precipitated with 2.5 vol of ethanol, and then resuspended in Tris-EDTA buffer (37). The fragments were ligated to Bam HI arms of the lambda vector EMBL 3A (38), and phage were packaged and introduced into NM539 (39). Approximately 7.5×10^4 recombinants were screened by filter hybridization (40) with a γ -IVS 2 probe (supplied by O. Smithies, Department of Pathology, University of North Carolina, Chapel Hill, NC) (41), and positive clones were purified after three rounds of plaque purification. Fetal globin gene clones from the sickle cell patient heterozygous for SEN and BEN haplotypes were digested with Hind III to distinguish between the two chromosomes (1), and the SEN clone was used for further analysis.

Sequence analysis of the G γ - and A γ -globin gene promoter regions. Phage DNA encompassing the linked G γ -A γ globin genes on the three β^s haplotype backgrounds was prepared using CsCl gradients (37). DNA was digested with Bam HI for 4 h, and the 2.6-kb 5' G γ - and 4.9-kb A γ gene fragments were isolated by electroelution (International Biotechnologies, Inc., New Haven, CT) after fractionation on a 0.7% (wt/vol) agarose gel.

The 2.6-kb Bam HI fragment, extending from the 3' end of exon 2 to ~ 2,100 bp 5' to the G γ cap site, was ligated to Bam HI-digested M13mp19 (42). Ligations were performed using T4 DNA ligase at a DNA concentration of 10 ng/ μ l and a molar ratio (insert/vector) of 2:1. DNA was transfected into calcium-shocked *Escherichia coli* JM107 or HB101, and bacteria were plated in the presence of isopropyl thiogalactoside and 5-bromo-4-chloro-indolyl- β -D-galactoside to identify recombinants (43). Double-stranded DNA was purified from clear plaques (44), digested with Bgl II, and separated by electrophoresis on agarose gels to determine the orientation of the inserts.

The 4.9-kb Bam HI 5'-A γ -gene fragment was digested with Sst I and precipitated with M13mp18 or M13mp19 which had been digested with Bam HI and Sac I. The 2.7-kb Sst I-Bam HI 5' A γ -gene fragment extends from ~ 2,350 bp 5' to the A γ cap site to the 3' end of exon 2.

Single-stranded DNA preparations were made from inserts in both orientations and used as templates for sequence determination, using the dideoxy sequencing method (45-47). Synthetic oligonucleotides, one spanning -391 to -377 (ACTACAGGCCTCACT) and the other spanning 53 to 71 (CTCTGTGAAATGACCCATG), were prepared by the phosphoramidite method on an 380B DNA synthesizer (Applied Biosystems, Inc., Foster City, CA) and used as primers for sequence analysis of γ -globin genes from both loci.

PCR analysis of β^s - and β^A -containing chromosomes. Genomic DNA was isolated from peripheral blood leukocytes (35) of normal blacks and patients with sickle cell disease. Also, genomic DNA was obtained from buccal epithelial cells (BEC) after a 10-s mouthwash with 15 ml of 0.9% (wt/vol) saline (48). BEC were pelleted by centrifugation.

gation at 500 g for 10 min, resuspended in 500 μ l water, boiled for 10 min, centrifuged briefly, and an aliquot of the supernatant was used for PCR.

Amplification was as described (49), with minor modifications. Reactions in 100 μ l contained 0.5–2 μ g genomic DNA or 50 μ l BEC lysate, 50 mM KCl, 10 mM Tris · HCl, pH 8.3, 1.5 mM MgCl₂, 0.1% (wt/vol) gelatin, 200 μ M each of dATP, dGTP, dCTP, and TTP, and 30–50 pmol of the two primers (see Fig. 2). Samples were heated to 95°C for 5 min, centrifuged briefly, 2.5 U Taq polymerase (Perkin-Elmer Cetus, Norwalk, CT) was added, and the sample was covered with 100 μ l paraffin oil. PCR was accomplished using a DNA thermal cycler (Perkin-Elmer Cetus) for 30–35 cycles of amplification, with each cycle defined as follows: 93–95°C (1–2 min), 50–55°C (1–2 min), and 72°C (2–5.5 min). Samples were then incubated at 4°C, extracted once with 100 μ l chloroform, and 10 μ l (~10% total) was electrophoresed on a 2% (wt/vol) agarose gel to monitor size of the amplified product.

The sequences and positions of PCR primers (A1–A3) and radio-labeled probes (P1–P4) are shown in Fig. 2. For monitoring the G γ -309 variant, a 462-bp region was amplified using primers A1 and A3, which resulted in amplification of both G γ and A γ 5' flanking regions. Synthetic oligonucleotide probes, P1 and P2, extending from –300 to –318 5' to the G γ gene, and containing either the –309 normal (A) or –309 variant (G), respectively, were then used to screen dot blots of the A1/A3 PCR products. The probes were constructed to anneal selectively only to the G γ 5' flank. The G γ and A γ flanking sequences differ at positions –307 and –317 5' to their respective cap sites (50), and probes P1 and P2 contain only the G γ sequence at these sites, facilitating distinction between the two genes. For monitoring the A γ –271 variant, a 734-bp fragment was amplified using primers A2 and A3 (see Fig. 2), which selectively amplified only the A γ 5' flank due to sequence specificity of primer A2 for A γ (50). Oligonucleotide probes, P3 and P4, extending from –280 to –262 5' to the A γ gene and incorporating either the –271 normal (C) or the –271 variant (T),

respectively, were used to screen dot blots of the A2/A3 PCR products. It was important in this instance to amplify only the A γ 5' flank because the –271 A γ variant is identical in sequence to the normal G γ 5' flank.

Dot blots, probes, and hybridizations. Typically, 20 μ l of the PCR product was adjusted to 0.3 M NaOH in 40 μ l and incubated 3–5 min at room temperature. 40 μ l of 1.2 M Tris · HCl, pH 7.4/3 M NaCl was then added, and 40 μ l was applied immediately to a pre-wetted (with 20 \times standard saline citrate [SSC] = 3 M NaCl/0.3 M sodium citrate) nitrocellulose membrane (0.45 μ m) using a dot blot or slot blot apparatus with vacuum applied (Schleicher & Schuell, Inc., Keene, NH). Each well was rinsed with 40–50 μ l of 0.6 M Tris · HCl, pH 7.4/1.5 M NaCl and the membrane was held under vacuum for 20–30 min and then baked at 80°C for 2 h.

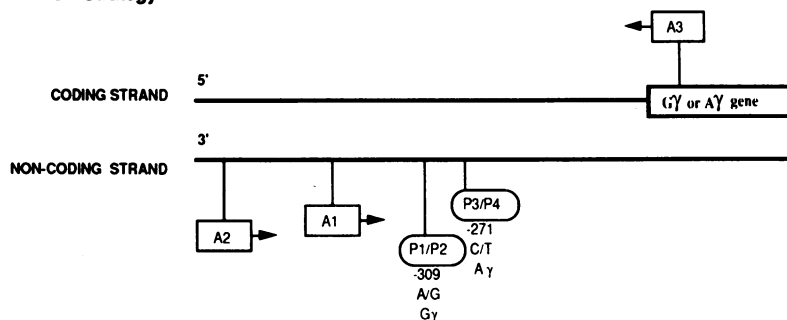
Membranes were prehybridized for 1–16 h at 37°C in 10 ml of 6 \times SSC/5 \times Denhardt's (0.1% [wt/vol] each of Ficoll, BSA, and poly[vinylpyrrolidone])/50 mM sodium phosphate, pH 6.8/0.1 mg/ml denatured salmon sperm DNA. Hybridizations were carried out at 37°C for 2–16 h in 5 ml of the same solution containing 5 \times 10⁵–1 \times 10⁶ cpm/ml of labeled oligonucleotide probe. The blots were then washed three times for 15–30 min each in 100 ml of 6 \times SSC/0.05% (wt/vol) sodium pyrophosphate, the first wash being at room temperature, the second at 2°C below the melting temperature (T_m) of the probe estimated to have the lowest T_m [T_m = 4(G + C) + 2(A + T)], and the third wash at an empirically determined temperature that adequately discriminated annealing of the two probes. Membranes were sealed in plastic wrap and autoradiographed at –70°C with two intensifying screens.

Probes were labeled at their 5' end by the following procedure: 200 ng of oligonucleotide in 3 μ l was heated to 90°C for 3 min, centrifuged briefly, 1 μ l 10 \times T4 polynucleotide kinase buffer (1 \times = 70 mM Tris · HCl, pH 7.6/0.1 M KCl/10 mM MgCl₂/5 mM DTT) was added, and the reaction was transferred to a tube containing 15–20 μ l of γ -[³²P]ATP (> 5,000 Ci/mmol; Amersham Corp., Arlington Heights,

A PCR Primers and Probes

PCR PRIMERS	A1	5' ACTACAGGCCTCACTGGAG 3'	(–391 to –373)
	A2	5' CATTAGGCTCTTATATTATG 3'	(–663 to –645)
	A3	5' CTCTGTGAAATGACCCATG 3'	(+53 to +71)
PROBES	P1	5' GGTGGGAGAGGAAACTAG 3'	(–300 to –318)
	P2	5' GGTGGGAGAGGAAACTAG 3'	(–300 to –318)
	P3	5' AGAGAAAACTGGAATGAC 3'	(–262 to –280)
	P4	5' AGAGAAAACTGGAATGAC 3'	(–262 to –280)

B PCR Strategy



C Size of PCR Product

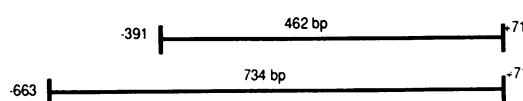


Figure 2. PCR analysis of the –309 G γ and –271 A γ variants. Sequences of PCR primers (A1–A3) and probes (P1–P4) are shown with coordinates relative to the cap site of the γ genes in A. Underlines indicate probe positions –271 or –309. Strategies for PCR using primers A1 and A3 for amplification of G γ and A γ , or primers A2 and A3 for A γ are shown with the normal (P1 and P3) and variant (P2 and P4) probes in B. Expected size of PCR products with coordinates is shown in C.

IL) that had been lyophilized to dryness in a Speed-Vac concentrator (Savant Instruments, Inc., Hicksville, NY). The tube was vortexed vigorously, centrifuged briefly, and 10 U of T4 polynucleotide kinase (Bethesda Research Laboratories) was added. The reaction was heated to 37°C for 30 min, centrifuged briefly, and cooled on ice. The labeled oligonucleotide was added to 50 mM KH₂PO₄ and then allowed to bind to 20 mg of a C-18 resin (Corasil, 37–50 µm; Waters Associates Inc., Milford, MA), which had been pretreated with 50% methanol and 50 mM KH₂PO₄. The resin was rinsed once with water, and labeled probe was eluted with 60% (vol/vol) methanol. An appropriate volume of probe was added directly to the hybridization solution.

Results

Hb F, G γ , MCV, and haplotype analyses for the three patients are shown in Table I. One patient was homozygous for β^s on the CAR haplotype (+---), the second was homozygous for β^s but heterozygous for haplotypes BEN (---+) and SEN (+---), and the third was a compound heterozygote with a nondeletional form of G γ - β^+ -HPFH on one chromosome and a β^s gene with a BEN haplotype background on the other.

The patient with one SEN chromosome expressed a high level of Hb F (12.4%) and high G γ (61%), while the one homozygous for the CAR haplotype had a lower level of Hb F (3.9%) and low G γ (36%). These results are consistent with previous data (5, 6) and within the wide range found in patients with sickle cell disease in the United States. Since the BEN chromosome came from a patient heterozygous for G γ - β^+ -HPFH, the effect of the BEN haplotype on Hb F expression could not be discerned; the patient had a high Hb F (29.5%), which was all G γ (18, 32, 33).

The DNA sequence of the immediate 5' flanking region was determined for the G γ - and A γ -globin genes on the three different haplotype backgrounds from -350 bp 5' to 10 bp 3' to the cap sites, and compared with reference fetal globin gene sequences (50) (Fig. 3). The G γ -globin gene on the BEN haplotype background had an A to G change at -309 and a C at -158, while the linked A γ gene sequence was identical to the reference gene. The reference gene from the first report of this sequence has a T at -158 in G γ , but recent data show that a C is usually present here (15, 20, 25, 26). The G γ -globin gene from the SEN haplotype background had the expected T at -158, which is associated with a high G γ to A γ ratio in a variety of patients with sickle cell disease and β -thalassemia (12, 25, 51). The A γ gene on this SEN chromosome was identical to that of the reference gene in the region of sequence analysis. The G γ gene sequence from one of the two chromosomes from the patient homozygous for β^s on the CAR haplotype background showed a C at position -158, while the linked A γ gene had a T at position -271.

We next addressed the question of whether sequence changes at -309 (A to G) 5' to the G γ gene on the BEN

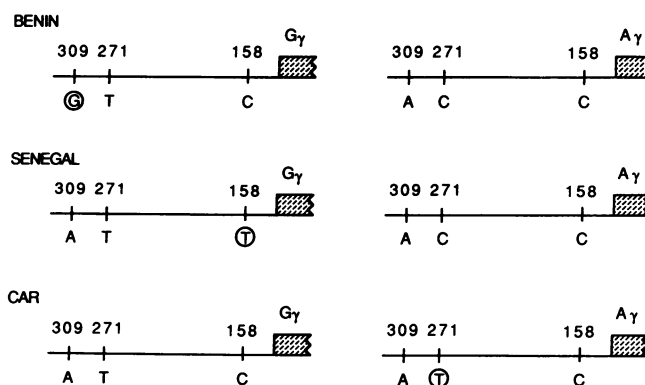


Figure 3. Base changes in the immediate 5' flanking region of the gamma globin genes on the three African β^s haplotype backgrounds. Sequence analysis spanned from -350 to 10 from the cap sites. Base changes found are circled, and the normal base found in the other two haplotypes is shown for comparison.

chromosome, and at -271 (C to T) 5' to the A γ gene on the CAR chromosome were haplotype associated and only found on β^s -containing chromosomes. Genomic DNA amplifications (PCR) using synthetic oligonucleotide probes were used to amplify and analyze the fetal globin gene 5' flanking regions (Fig. 2). A portion of the PCR product was electrophoresed on a neutral agarose gel, revealing the presence of the anticipated ~460 bp amplified product using primers A1 and A3, and a ~760-bp product using primers A2 and A3. Results of dot blot analysis of the PCR products, using selective hybridization of radiolabeled, allele-specific, synthetic oligonucleotide probes are shown in Table II.

Of the 39 patients with sickle cell disease examined for the -309 A to G change 5' to G γ , all 15 BEN homozygotes were homozygous for G; one patient homozygous for SEN and one for CAR were both homozygous for A. Of the 18 patients

Table II. Relationship between A to G Change at -309 G γ or C to T Change at -271 A γ and Haplotype in Patients with Sickle Cell Disease

Patient haplotype	Patient genotype						
	-309 G γ (A or G)				-271 A γ (C or T)		
	A:A	A:G	G:G	Total patients	C:C	C:T	T:T
BEN/BEN	—	—	15	15	14	1	—
SEN/SEN	1	—	—	1	1	—	—
CAR/CAR	1	—	—	1	—	—	2*
BEN/SEN	—	1	—	1	2*	—	—
BEN/CAR	1	11	—	12	—	10	—
SEN/CAR	2	—	—	2	—	2	—
BEN/minor	—	5	—	5	5	—	—
SEN/minor	—	—	—	—	1	—	—
CAR/minor	1	—	—	1	—	1	—
Minor/minor	—	—	1	1	1	—	—
β^A	8	7	3	18	12	—	—

* One of these patients was the source of the CAR or SEN chromosomes for which sequence data are presented in this paper.

Table I. Haplotype and Hematologic Parameters

Patient	MCV	Hb F	G γ
	μm^3	%	%
CAR/CAR	85	3.9	36
SEN/BEN	95	12.4	61
BEN/G γ - β^+ -HPFH	93	29.5	100

identified as BEN heterozygotes, 17 were heterozygous at -309 (A; G); 3 of the 4 non-BEN heterozygotes were homozygous for A. Therefore, there is a strong association between the BEN haplotype background and the -309 G γ variant in the patients with sickle cell disease that we screened; 17 homozygotes and 20 of 22 heterozygotes were consistent with this association. Two patients violated this association: one with two minor haplotypes was homozygous for G and one BEN/CAR heterozygote was homozygous for A at this site. Of the 18 normal blacks (hemoglobin A) studied, 8 were homozygous for A, 3 were homozygous for G, and 7 were heterozygous (A; G), indicating a polymorphism at the site.

Of the 40 patients with sickle cell disease examined for the -271 C to T change 5' to A γ , 2 CAR homozygotes were homozygous for T, one SEN homozygote and 14 of 15 BEN homozygotes were homozygous for C, and the remaining BEN homozygote was heterozygous (C; T). Of the 13 CAR heterozygotes, all were heterozygous (C; T). Of the 9 non-CAR heterozygotes, all were homozygous for C. Therefore, there is an association of the -271 A γ change with the CAR haplotype: 17 homozygotes and 22 heterozygotes were consistent with this association, while 1 homozygote was not. None of 12 normal blacks (Hb A) examined had a T at the A γ -271 position.

Discussion

The 5' flanking regions of G γ - and A γ -globin genes are identical out to 471 bp 5' to the cap sites, except for six base differences clustered in the region -271 to -398 bp (Fig. 1) (50). This divergent region may potentiate changes in G γ to A γ ratios and Hb F levels. The sequence changes we found on the BEN and CAR chromosomes occur in this region, and may be important in the control of fetal globin gene expression. Our PCR analysis of a variety of β^s -containing chromosomes shows that the -309 variant 5' to the G γ gene predominates on BEN, while the -271 change 5' to the A γ gene predominates on CAR. The -309 variant was also found associated with G γ genes on normal β^A -containing chromosomes (13 of 36 had the variant), while the -271 variant was not. Although these data are not sufficient to determine whether the sickle mutation or the additional G γ alteration occurred first in each instance, a possibility is that the -309 change predated the β^s mutation on the BEN haplotype since it is also common in the normal population, while the -271 change may have occurred after or concurrent with the β^s mutation on the CAR chromosome. Another possibility is that the -271 change predated the β^s mutation on CAR, but that the frequency of this haplotype in the normal population that we screened was too low to be detected. Other arguments are equally feasible on the basis of current data, and further population studies are now indicated. The findings for the -309 G γ variant with respect to Hb F production may be similar to the -158 C to T change 5' to the G γ gene on SEN chromosomes; the presence of the -158 change on normal chromosomes is associated with limited effect on fetal globin gene expression (24).

Sequence changes in the immediate 5' flanking region of the fetal globin genes could affect gene expression by altering chromatin structure, modifying DNA methylation patterns, and/or potentiating changes in critical DNA-protein interactions (17, 52, 53). A consensus core enhancer binding site

(TGGAATG) for SV40 is present in the -270-bp region 5' to the γ genes (24), so that sequence changes here might lead to altered binding of ubiquitous or erythroid-specific protein factors (54, 55) resulting in changes in expression of the fetal globin genes. The -271 (T) variant 5' to the A γ gene on CAR chromosomes might lead to altered expression of the A γ gene. The A to G change at -309 5' to the G γ gene on BEN chromosomes occurs within an area so far not described to be critical for fetal globin gene expression or protein binding.

Sequence analysis of the SEN chromosome revealed the expected T at -158 5' to the G γ -globin gene cap site, with no other base changes in the regions from the cap sites of G γ and A γ globin genes up to -350 bp. Sequence analysis of the γ -globin promoter regions from an Eastern Saudi Arabian sickle cell chromosome also showed a T at -158 5' to the G γ gene and no other base changes (51). This alteration may be important for gene expression because it is the only base change in this region on a β^s chromosome associated with a high Hb F and a high ratio of G γ to A γ . The presence of a T at -158 cannot, however, be solely responsible for the high Hb F and high G γ /A γ phenotype, since normal Eastern Saudi Arabians (genotype AA) with a T at -158 5' to the G γ -cap site and the same subhaplotype (+ - + +) do not have increased Hb F in peripheral blood or in burst-forming unit-derived cells (51). Hemolytic stress appears to be required for an effect on fetal hemoglobin expression.

The C to T change at position -271 5' to the CAR A γ gene may, in fact, be the result of a gene conversion event causing the G γ and A γ genes to resemble each other more closely (Fig. 3), perhaps conferring changes in function of the A γ gene as well. Analysis of the G γ and A γ genes on both chromosomes from one individual (41) suggested that the 5' region of the A γ -globin gene, from -50 to IVS 2, could convert by intergenic exchange, resulting in the linked G γ -A γ genes being more alike than the two A γ genes. Alternatively, this change at -271 could be a neutral polymorphism.

Results for the -271 A γ variant may be of relevance to the origin of an Rsa I polymorphism located ~ 550 bp 5' to the β -globin gene (27). Our earlier study showed association between CAR-containing β^s chromosomes and cleavage at this Rsa I site (31). We have further extended those results with analysis of additional patients in this report, and found the Rsa I cut site was also present on a minor β^s -containing haplotype found in at least five unrelated families (seven total patients with sickle cell disease). This minor haplotype (- - - - +) (Rsa I site follows the usual Hind III and Hinc II sites) differs from CAR-containing β^s chromosomes at the G γ -Hind III site (- instead of +), suggesting that either (a) the Rsa I polymorphism occurred once on CAR and then the chromosome diverged at the G γ -Hind III site, or (b) the polymorphism occurred twice, independently on CAR and on the minor haplotype. Our results favor the latter explanation, because A γ -globin genes on this minor haplotype background contain C at the -271 position instead of T, which is associated with the CAR haplotype. Analysis of haplotype frequencies and distribution of these variants in the normal black population should help delineate origins of these single-base changes.

Functional expression of isolated genes or 5' DNA regions are now required to determine the role of these changes in the expression of the fetal globin genes. Although the immediate 5' flanking area is important, there is evidence that determinants affecting Hb F expression may be in other parts of the β -globin

gene cluster on chromosome 11, or even on other chromosomes (56, 57). Further investigation including additional DNA sequence analysis and functional studies will be necessary to define these genetic determinants.

Acknowledgments

The authors wish to thank Dr. Mortimer Poncz for his advice, Dr. Eric Rappaport for preparation of oligonucleotides, Dr. Kazuhiko Adachi for analysis of G γ and A γ ratios, and Ms. Emi Asakura for her help in the preparation of the manuscript.

This work was supported in part by research grants DK-16691 and HL-28157, a Comprehensive Sickle Cell Center award (HL-38632), and a training grant (HL-07150) from the National Institutes of Health.

Note added in proof. Recently, after our manuscript was submitted, Bouhassira et al. (1989. *J. Clin. Invest.* 83:2070-2073) reported that the -271 C to T change is part of an A γ to G γ globin gene conversion event in >80% of β^s chromosomes with the Bantu (CAR) haplotype.

References

- Pagnier, J., J. G. Mears, O. Dunda-Belkhodja, K. E. Schaefer-Rego, C. Beldjord, R. L. Nagel, and D. Labie. 1984. Evidence for the multicentric origin of the sickle cell hemoglobin gene in Africa. *Proc. Natl. Acad. Sci. USA* 81:1771-1773.
- Antonarakis, S. E., C. D. Boehm, G. R. Serjeant, C. E. Theisen, G. J. Dover, and H. H. Kazazian, Jr. 1984. Origin of the β^s -globin gene in Blacks: the contribution of recurrent mutation or gene conversion or both. *Proc. Natl. Acad. Sci. USA* 81:853-856.
- Pagnier, J., V. Baudin, D. Labie, H. Wajcman, G. Jaeger, and R. Girot. 1986. Sickle cell anemia in Bantu speaking Africa. *Hemoglobin* 10:73-76.
- Labie, D., J. Pagnier, C. Lapoumeroulie, F. Rouabhi, O. Dunda-Belkhodja, P. Chardin, C. Beldjord, N. Wajcman, M. E. Fabry, and R. L. Nagel. 1985. Common haplotype dependency of high G γ -globin gene expression and high Hb F levels in β -thalassemia and sickle cell anemia patients. *Proc. Natl. Acad. Sci. USA* 82:2111-2114.
- Nagel, R. L., M. E. Fabry, J. Pagnier, I. Zohoun, H. Wajcman, V. Baudin, and D. Labie. 1985. Hematologically and genetically distinct forms of sickle cell anemia in Africa: the Senegal type and the Benin type. *N. Engl. J. Med.* 312:880-884.
- Nagel, R. L., S. K. Rao, O. Dunda-Belkhodja, M. M. Connolly, M. E. Fabry, A. Georges, R. Krishnamoorthy, and D. Labie. 1987. The hematologic characteristics of sickle cell anemia bearing the Bantu haplotype. The relationship between G γ and Hb F level. *Blood* 69:1026-1030.
- Gilman, J. G., and T. H. J. Huisman. 1984. Two independent genetic factors in the β -globin gene cluster are associated with high G γ -levels in the Hb F of SS patients. *Blood* 64:452-457.
- Papayannopoulou, T., M. Brice, and G. Stamatoyannopoulos. 1986. Analysis of human hemoglobin switching in MEL X human fetal erythroid cell hybrids. *Cell* 46:469-476.
- Miller, B. A., M. Salameh, M. Ahmed, J. Wainscoat, G. Antognetti, S. Orkin, D. J. Weatherall, and D. G. Nathan. 1986. High fetal hemoglobin production in sickle cell anemia in the eastern province of Saudi Arabia is genetically determined. *Blood* 67:1404-1410.
- Gilman, J. G., T. Harano, T. Nakatsuji, I. Bakioglu, A. L. Reese, M. B. Gardiner, and T. H. J. Huisman. 1985. The ratio of the G γ and A γ chains: variations due to anomalies at the molecular level. *Ann. NY Acad. Sci.* 445:235-247.
- Wainscoat, J. S., S. L. Thein, D. R. Higgs, J. I. Bell, D. J. Weatherall, B. H. Al-Awamy, and G. R. Serjeant. 1985. A genetic marker for elevated levels of haemoglobin F in homozygous sickle cell disease? *Br. J. Haematol.* 60:261-268.
- Hattori, P., F. Kutlar, A. Kutlar, V. C. McKie, and T. H. J. Huisman. 1986. Haplotypes of β^s chromosomes among patients with sickle cell anemia from Georgia. *Hemoglobin* 10:623-642.
- Poncz, M., M. Ballantine, D. Solowiejczyk, I. Barak, E. Schwartz, and S. Surrey. 1982. β -Thalassemia in a Kurdish Jew: single base change in the T-A-T-A box. *J. Biol. Chem.* 257:5994-5996.
- Anagnou, N. P., S. Karlsson, A. D. Moulton, G. Keller, and A. W. Nienhuis. 1986. Promoter sequences required for function of the human γ globin gene in erythroid cells. *EMBO (Eur. Mol. Biol. Organ.) J.* 5:121-126.
- Gelinas, R., B. Endlich, C. Pfeiffer, M. Yagi, and G. Stamatoyannopoulos. 1985. G to A substitution in the distal CCAAT box of the A γ -globin gene in Greek hereditary persistence of fetal hemoglobin. *Nature (Lond.)* 313:323-324.
- Collins, F. S., J. E. Metherall, M. Yamakawa, J. Pan, S. M. Weissman, and B. G. Forget. 1985. A point mutation in the A γ -globin gene promoter in Greek hereditary persistence of fetal haemoglobin. *Nature (Lond.)* 313:325-326.
- Poncz, M., P. Henthorn, C. Stoeckert, and S. Surrey. 1988. Globin gene expression in hereditary persistence of fetal hemoglobin and ($\delta\beta$) $^\circ$ thalassemia. *Oxf. Surv. Eukaryotic Genes* 5:163-203.
- Surrey, S., K. Delgrosso, P. Malladi, and E. Schwartz. 1988. A single-base change at position -175 in the 5'-flanking region of the G γ -globin gene from a Black with G γ - β^+ HPFH. *Blood* 71:807-810.
- Gigliani, B., C. Casini, R. Mantovani, S. Merli, P. Comi, S. Ottolenghi, G. Saglio, C. Camaschella, and U. Mazza. 1984. A molecular study of a family with Greek hereditary persistence of fetal hemoglobin and β thalassemia. *EMBO (Eur. Mol. Biol. Organ.) J.* 3:2641-2645.
- Tate, V. E., W. G. Wood, and D. J. Weatherall. 1986. The British form of hereditary persistence of fetal hemoglobin results from a single base mutation adjacent to a S1 hypersensitive site 5' to the A γ globin gene. *Blood* 68:1389-1393.
- Collins, F. S., C. J. Stoeckert, G. R. Serjeant, B. G. Forget, and S. M. Weissman. 1984. G γ - β^+ hereditary persistence of fetal hemoglobin: cosmid cloning and identification of a specific mutation 5' to the G γ gene. *Proc. Natl. Acad. Sci. USA* 81:4894-4898.
- Collins, F. S., C. D. Boehm, P. G. Waber, C. J. Stoeckert, S. M. Weissman, B. G. Forget, and H. H. Kazazian, Jr. 1984. Concordance of a point mutation 5' to the γ globin gene with G γ - β^+ hereditary persistence of fetal hemoglobin in the black population. *Blood* 64:1292-1296.
- Stoming, T. A., G. S. Stoming, K. D. Lanclos, Y. J. Fei, C. Altay, F. Kutlar, and T. H. J. Huisman. 1989. An A γ type of nondeletional hereditary persistence of fetal hemoglobin with a T \rightarrow C mutation at position -175 to the cap site of the A γ globin gene. *Blood* 73:329-333.
- Gilman, J. G. 1988. γ -Thalassemia and related conditions. *Hemoglobin* 12:707-716.
- Labie, D., O. Dunda-Belkhodja, F. Rouabhi, J. Pagnier, A. Ragusa, and R. L. Nagel. 1985. The -158 site 5' to the G γ gene and G γ expression. *Blood* 66:1463-1465.
- Gilman, J. G., and T. H. J. Huisman. 1985. DNA sequence variation associated with elevated fetal G γ globin production. *Blood* 66:783-787.
- Semenza, G. L., P. Malladi, S. Surrey, K. Delgrosso, M. Poncz, and E. Schwartz. 1984. Detection of a novel DNA polymorphism in the β -globin gene cluster. *J. Biol. Chem.* 259:6045-6048.
- Wilson, J. T., L. B. Wilson, J. D. De Riel, L. Villa-Komaroff, A. Efstratiadis, B. G. Forget, and S. Weissman. 1978. Insertion of synthetic copies of human globin genes into bacterial plasmids. *Nucleic Acids Res.* 5:563-581.
- Fritsch, E. F., R. M. Lawn, and T. Maniatis. 1980. Molecular cloning and characterization of the human β -like globin gene cluster. *Cell* 19:959-972.
- Adams, J. W., R. E. Kaufman, P. J. Kretschmer, M. Harrison, and A. W. Nienhuis. 1980. A family of long reiterated DNA sequences,

one copy of which is next to the human beta globin gene. *Nucleic Acids Res.* 8:6113-6128.

31. Sharon, B., M. Poncz, S. Surrey, and E. Schwartz. 1988. Non-random association of the Rsa I polymorphic site 5' to the β -globin gene with major sickle cell haplotypes. *Hemoglobin*. 12:115-124.

32. Balsley, J. F., E. Rappaport, E. Schwartz, and S. Surrey. 1982. The δ - γ - β -globin gene region in G γ - β^+ -hereditary persistence of fetal hemoglobin. *Blood*. 59:828-831.

33. Friedman, S., and E. Schwartz. 1976. Hereditary persistence of foetal haemoglobin with β -chain synthesis in cis position (G γ - β^+ -HPFH) in a negro family. *Nature (Lond.)*. 259:138-140.

34. Kim, H. C., J. Atwater, and E. Schwartz. 1983. Separation of hemoglobins. In *Hematology*. W. J. Williams, E. Beutler, A. J. Erslev, and M. A. Lichtman, editors. 3rd ed. McGraw-Hill Book Company, New York. 1611-1619.

35. Kim, H. C., and E. Schwartz. 1983. Fetal hemoglobin-alkali denaturation test. In *Hematology*. W. J. Williams, E. Beutler, A. J. Erslev, and M. A. Lichtman, editors. McGraw-Hill Book Company, New York. 1619-1620.

36. Poncz, M., D. Solowiejczyk, B. Harpel, Y. Mory, E. Schwartz, and S. Surrey. 1982. Construction of human gene libraries from small amounts of peripheral blood: analysis of beta-like globin genes. *Hemoglobin*. 6:27-36.

37. Maniatis, T., E. F. Fritsch, and J. Sambrook. 1982. *Molecular Cloning: A Laboratory Manual*. Cold Spring Harbor Laboratory, Cold Spring Harbor, NY. 1-545.

38. Frischauf, A. M., H. Lehrach, A. Poustka, and N. Murray. 1983. Lambda replacement vectors carrying polylinker sequences. *J. Mol. Biol.* 170:827-842.

39. Scalenghe, F., E. Turco, J. E. Edstrom, V. Pirrotta, and M. Melli. 1981. Microdissection and cloning of DNA from a specific region of *Drosophila melanogaster* polytene chromosomes. *Chromosoma (Berl.)*. 82:205-216.

40. Benton, W. D., and R. W. Davis. 1977. Screening lambda gt recombinant clones by hybridization to single plaques *in situ*. *Science (Wash. DC)*. 196:180-182.

41. Slightom, J. L., A. E. Blechl, and O. Smithies. 1980. Human fetal G γ - and A γ -globin genes: complete nucleotide sequences suggest that DNA can be exchanged between these duplicated genes. *Cell*. 21:627-638.

42. Messing, J., and J. Viera. 1982. A new pair of M13 vectors for selecting either DNA strand of double-digested restriction fragments. *Gene (Amst.)*. 19:269-276.

43. Messing, J., B. Gronenborn, B. Muller-Hill, and P. H. Hofschneider. 1977. Filamentous coliphage M13 as a cloning vehicle: insertion of a Hind I fragment of the lac regulatory region in M13 replicative form *in vitro*. *Proc. Natl. Acad. Sci. USA*. 74:3642-3646.

44. Riggs, M. G., and A. McLachlan. 1986. A simplified screening procedure for large numbers of plasmid mini preparations. *Biotechniques*. 4:310-311.

45. Dale, R. M. K., B. A. McClure, and J. P. Houchins. 1985. A rapid single stranded cloning strategy for producing a sequential series of overlapping clones for use in DNA sequencing: application to sequencing the corn mitochondrial 18SrDNA. *Plasmid*. 13:31-40.

46. Sanger, F., A. R. Coulson, B. G. Barrell, A. J. H. Smith, and B. A. Roe. 1980. Cloning in single-stranded bacteriophage as an aid to rapid sequencing. *J. Mol. Biol.* 143:161-178.

47. Sanger, F., S. Nicklen, and A. R. Coulson. 1977. DNA sequencing with chain-terminating inhibitors. *Proc. Natl. Acad. Sci. USA*. 74:5463-5467.

48. Lench, N., P. Stanier, and R. Williamson. 1988. Simple, non-invasive method to obtain DNA for gene analysis. *Lancet*. i:1356-1358.

49. Saiki, R. K., D. H. Gelfand, S. Stoffel, S. J. Scharf, R. Higuchi, G. T. Horn, K. B. Mullis, and H. A. Erlich. 1988. Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase. *Science (Wash. DC)*. 239:487-491.

50. Shen, S., J. L. Slightom, and O. Smithies. 1981. A history of the human fetal globin gene duplication. *Cell*. 26:191-203.

51. Miller, B. A., N. Olivieri, M. Salameh, M. Ahmed, G. Antognetti, D. G. Nathan, and S. H. Orkin. 1987. Molecular analysis of the high hemoglobin F phenotype in Saudi Arabian sickle cell anemia. *N. Engl. J. Med.* 316:244-250.

52. Groudine, M., T. Kohwi-Shigematsu, R. Gelinas, G. Stamatoyannopoulos, and T. Papayannopoulou. 1983. Human fetal to adult hemoglobin switching: changes in chromatin structure of the β -globin gene locus. *Proc. Natl. Acad. Sci. USA*. 80:7551-7555.

53. Ley, T. J., Y. L. Chiang, D. Haidaris, N. P. Anagnou, V. L. Wilson, and W. F. Anderson. 1984. DNA methylation and regulation of the human β -globin-like genes in mouse erythroleukemia cells containing human chromosome 11. *Proc. Natl. Acad. Sci. USA*. 81:6618-6622.

54. Gumucio, D. L., K. L. Rood, T. A. Gray, M. F. Riordan, C. I. Sartor, and F. S. Collins. 1988. Nuclear proteins that bind the human γ -globin gene promoter: alterations in binding produced by point mutations associated with hereditary persistence of fetal hemoglobin. *Mol. Cell. Biol.* 8:5310-5322.

55. Mantovani, R., N. Malgaretti, S. Nicolis, A. Ronchi, B. Giglioli, and S. Ottolenghi. 1988. The effects of HPFH mutations in the human γ -globin promoter on binding of ubiquitous and erythroid specific nucleic factors. *Nucleic Acids Res.* 16:7783-7797.

56. Boyer, S. H., G. J. Dover, G. R. Serjeant, K. D. Smith, S. E. Antonarakis, S. H. Embury, L. Margolet, A. N. Noyes, M. L. Boyer, and W. B. Bias. 1984. Production of F cells in sickle cell anemia. Regulation by a genetic locus or loci separate from the β -globin gene cluster. *Blood*. 64:1053-1058.

57. Milner, P. F., J. D. Leibfarth, J. Ford, B. P. Barton, H. E. Grenett, and F. A. Garbver. 1984. Increased Hb F in sickle cell anemia is determined by a factor linked to the β gene from one parent. *Blood*. 63:64-72.