Proc. Natl. Acad. Sci. USA
Vol. 81, pp. 4894–4898, August 1984
Genetics

$^{G}\gamma\beta^{+}$ Hereditary persistence of fetal hemoglobin: Cosmid cloning and identification of a specific mutation 5′ to the $^{G}\gamma$ gene

(hemoglobin switching/control of gene expression/promoter element)

FRANCIS S. COLLINS*†, CHRISTIAN J. STOECKERT, JR.*, GRAHAM R. SERJEANT‡, BERNARD G. FORGET*§, AND SHERMAN M. WEISSMAN*¶

Departments of *Human Genetics and §Internal Medicine, Yale University School of Medicine, New Haven, CT 06510; and ‡The Medical Research Council Laboratories, University of the West Indies, Kingston, Jamaica

Contributed by Sherman M. Weissman, April 20, 1984

ABSTRACT **Hereditary persistence of fetal hemoglobin (HPFH) is a benign condition in which the normal shutoff of fetal hemoglobin (Hb F) production fails to occur. In the $^{G}\gamma\beta^{+}$ type of HPFH, erythrocytes of adult heterozygotes contain ≈20% Hb F, which is almost exclusively of the $^{G}\gamma$-globin variety, without increased levels of γ-globin chains from the nearby $^{A}\gamma$-globin gene. Unlike some forms of HPFH, no major deletions in the globin gene cluster have been found by genomic blotting in the $^{G}\gamma\beta^{+}$ variety. We report here a family with this condition, from which cosmid clones of the β-globin gene cluster from the $^{G}\gamma\beta^{+}$ HPFH allele have been obtained. Sequencing around the fetal genes has identified a point mutation 202 base pairs 5′ to the $^{G}\gamma$-globin gene that is present in genomic DNA of 3/3 unrelated individuals with $^{G}\gamma\beta^{+}$ HPFH but in none of more than 100 non-HPFH individuals. Although the mutation could represent a tightly linked polymorphism, its location in a region suggested by recent data to be important in tissue-specific control of gene expression suggests the possibility that the −202 mutation accounts for the phenotype. The sequence created resembles elements of other eukaryotic promoters known to be important for efficient transcription.**

The orderly and regulated switch from fetal hemoglobin (Hb F, $\alpha_2\gamma_2$) to adult hemoglobin (Hb A, $\alpha_2\beta_2$) synthesis in man at about the time of birth (1) is of considerable current interest. From the point of view of basic molecular biology, the hemoglobin switch provides an accessible model system for study of the differential control of gene expression. Furthermore, there is good reason to suspect that the ability to reverse this switch would be therapeutically beneficial in management of the β-hemoglobinopathies, since individuals who for genetic reasons synthesize unusually large amounts of fetal hemoglobin as adults seem to be protected against the ravages of homozygous sickle cell disease (2) and β-thalassemia (3).

Such genetically controlled synthesis of fetal hemoglobin in adult life at levels greater than the normal 1% in the absence of erythropoietic stress or thalassemic imbalance of globin chain synthesis is known as hereditary persistence of fetal hemoglobin (HPFH). This is a benign but genetically heterogeneous group of conditions, which are categorized by the distribution of Hb F in erythrocytes, by the relative contributions of the duplicated fetal genes ($^{G}\gamma$ and $^{A}\gamma$), and by whether or not large deletions of the β- and δ-globin genes are found by genomic blotting (4–6).

In the so-called $^{G}\gamma\beta^{+}$ HPFH seen in Blacks, adult heterozygotes synthesize 15–25% Hb F that is almost exclusively of the $^{G}\gamma$ type; no deletions or rearrangements in the β-globin cluster have been found (7–9). No homozygotes are known, but double heterozygotes for β-chain variants show that there is continued but reduced β-chain synthesis in *cis* to the mutation. In this paper, we report the cloning of the β-globin gene cluster from the $^{G}\gamma\beta^{+}$ HPFH allele of such a double heterozygote and the demonstration of a point mutation 5′ to the $^{G}\gamma$-globin gene that is found only in this allele and may be responsible for the HPFH phenotype.

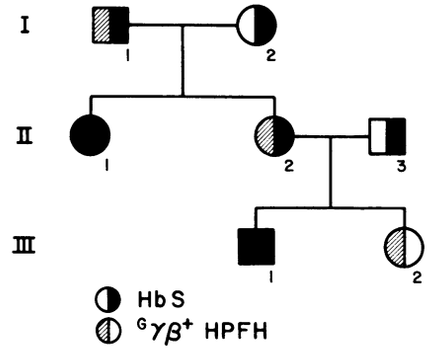

FIG. 1. Pedigree of the $^{G}\gamma\beta^{+}$ HPFH family. Cloning was from leukocyte DNA of individual II-2, who is doubly heterozygous for $^{G}\gamma\beta^{+}$ HPFH and Hb S. These must be present in *trans* since her sister and son, both of whom must share her Hb S allele, do not have unexpectedly high Hb F (see text and Table 1).

MATERIALS AND METHODS

Patients. The family we studied (see Fig. 1) was identified through a large-scale newborn screening program for sickle cell disorders in Jamaica (10). Fetal hemoglobin was determined by the method of Betke *et al.* (11), and the $^{G}\gamma/^{A}\gamma$ ratio in individual II-2 was quantitated by electrophoretic analysis (12, 13).

Cosmid Cloning. DNA from individual II-2 was prepared from peripheral blood leukocytes (14, 15), digested to completion with *Kpn* I, and size selected over a 10–40% sucrose gradient. One microgram of the 30- to 45-kilobase (kb) size-selected DNA was ligated to 1 μg of *Kpn* I-cut pFC1 (see Fig. 2), that had previously been cut with *Hin*dIII or *Sal* I and treated with phosphatase to prevent tandem ligation (16). pFC1 was prepared from the standard cosmid cloning vector pJB8 (16) by insertion of a *Kpn* I cloning site using an appropriate linker. Ligated DNA was packaged, transfected into recA⁻ HB101, and plated using standard methods (17, 18). Approximately 30,000 cosmids per μg of size-selected genomic DNA were obtained. Nitrocellulose filters were screened with a purified 1.9-kb *Bam*HI fragment containing the 5′ end of the β-globin gene, nick-translated (19) to a spe-

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: HPFH, hereditary persistence of fetal hemoglobin; kb, kilobase(s); bp, base pair(s).

†Present address: Department of Medicine, University of Michigan Medical School, Ann Arbor, MI 48109.
¶To whom reprint requests should be addressed.

Table 1. Hematologic data on the $^{G}\gamma\beta^{+}$ HPFH family

Individual	Age at determination, yr	Packed cell volume, %	Mean corpuscular volume, fl	% Hb A	% Hb S	% Hb F	% Hb A
I-1*	59	29	86	35.0	39.4	23.5	2.1
I-2†	48	28	58	70.0	26.9	0.3	2.8
II-1	18	25	87	0	90.4	6.9	2.7
II-2	23	37	88	36.3	41.2	19.9	2.6
III-1	4	24	84	0	86.4	11.1	2.5

Individual II-3 was not available for study but must be AS from the pedigree. Individual III-2 had an AF pattern on cord blood and therefore represents a simple $^{G}\gamma\beta^{+}$ HPFH heterozygote.
*The anemia in individual I-1 is attributed to chronic illness; he died shortly afterward of congestive heart failure.
†The low mean corpuscular volume and relatively low percentage Hb S in I-2 individual suggests she may be homozygous for α-thalassemia 2.

cific activity of 2×10^8 cpm/μg. Hybridizing clones were analyzed by restriction analysis; in some instances, restriction digestions were carried out with the 39-kb cosmid inserts that were prepared by *Kpn* I digestion and purification away from the vector on a 10–40% sucrose gradient.

Sequence Analysis. *Eco*RI fragments of the $^{G}\gamma\beta^{+}$ HPFH cosmids were subcloned into plasmid pBR322. These were cut with *Nco* I and subcloned into the *Hinc*II site of the single-stranded bacteriophage M13mp8. Some M13 clones were also obtained by sonication of the entire cosmid, purification of the 300- to 1000-base-pair (bp) fragments on a Sepharose 2B column, repair of the ends with S1 nuclease and the large fragment of DNA polymerase, and ligation into M13mp8, as described (20). Clones of interest were selected by hybridization to a 2.6-kb *Eco*RI $^{A}\gamma$-globin fragment, and both strands were sequenced using the Sanger dideoxy method (21).

Genomic Blotting. Five to 10 μg of genomic DNA was digested with *Apa* I + *Eco*RI, electrophoresed on 0.8% agarose, transferred to nitrocellulose filters, and hybridized with a nick-translated plasmid containing the 2.6-kb *Eco*RI $^{A}\gamma$-globin gene fragment. Standard methods were followed (22).

RESULTS

Hematologic Findings. The pedigree of the family under study is shown in Fig. 1; the pertinent hematologic data on these individuals is given in Table 1. This demonstrates that individuals I-1 and II-2 produce large amounts of Hb F. Hemoglobin electrophoretic analysis of blood from II-2 showed that the increased γ-globin chains were virtually all of the $^{G}\gamma$ type.

The levels of Hb F produced by individuals II-1 and III-1 are within the range expected for homozygous sickle cell disease at their respective ages (23, 24). This indicates that the $^{G}\gamma\beta^{+}$ HPFH gene in I-1 and II-2 must be present in *trans* to the Hb S gene. Synthesis of β^{A} globin in I-1 and II-2 occurs, but the level is reduced; in the usual AS patient, Hb A ex-

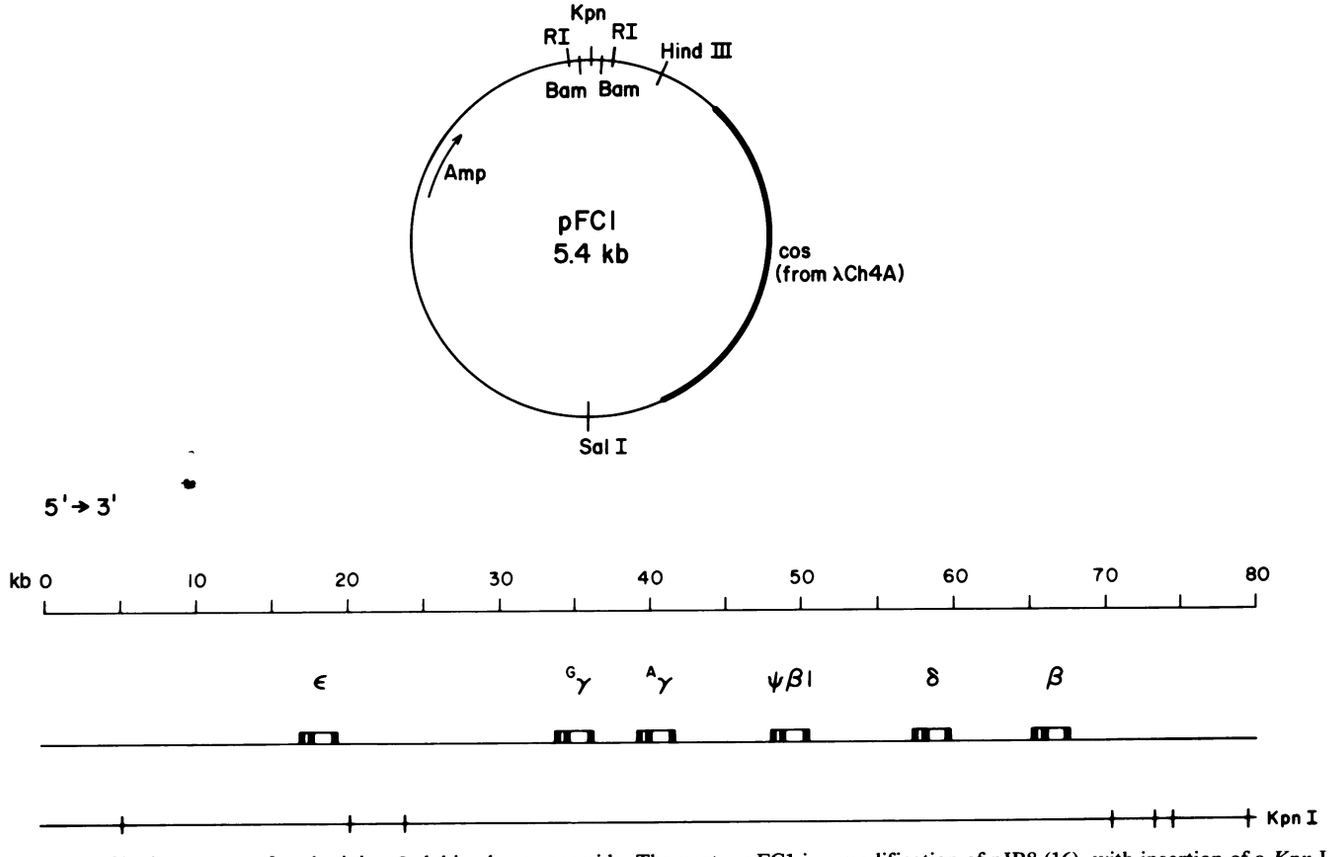

FIG. 2. Cloning strategy for obtaining β-globin cluster cosmids. The vector pFC1 is a modification of pJB8 (16), with insertion of a *Kpn* I cloning site. The presence of a 39-kb *Kpn* I fragment that includes $^{G}\gamma$-, $^{A}\gamma$-, δ-, and β-globin genes was exploited by digestion of genomic DNA to completion with *Kpn* I and size selection of the 30- to 45-kb fragments prior to ligation to the vector.

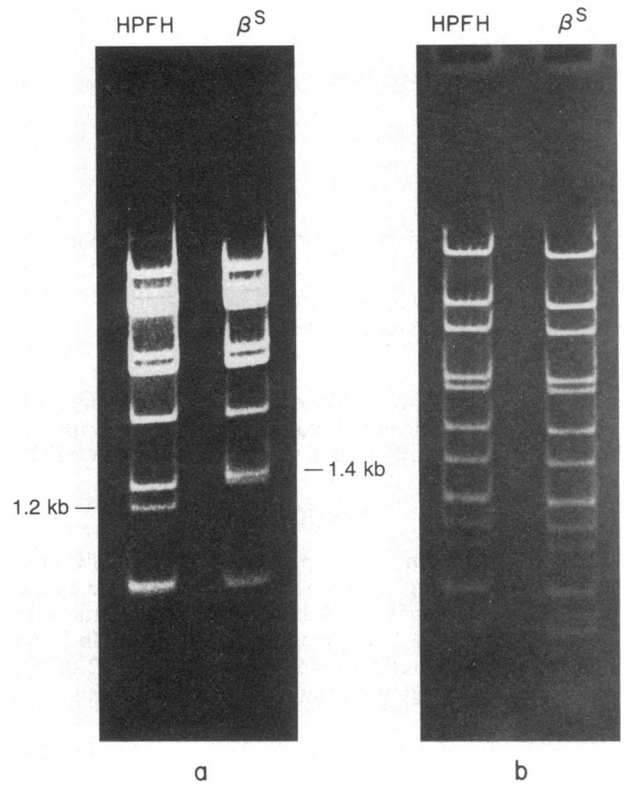

FIG. 3. Restriction digests of β-globin cluster DNA from HPFH and β^S cosmids. The 39-kb cosmid inserts were purified away from the vector by *Kpn* I digestion and sucrose gradient separation, and 1 μg of each was digested and electrophoresed. (*a*) *Mst* II digest. The β^S cosmid is recognizable by the appearance of a 1.4-kb band resulting from the sickle mutation at codon 6 of the β-globin gene. The normal pattern with a band at 1.2 kb must therefore represent the HPFH allele. (*b*) *Msp* I digest. The β^S cosmid should represent the normal restriction pattern. No differences in fragment sizes or restriction sites are apparent in the HPFH cosmid.

ceeds Hb S. This compensatory reduction of β^A-globin synthesis in *cis* to the $^G\gamma\beta^+$ HPFH mutation has been noted previously (3, 7–9) and remains unexplained.

Cosmid Cloning. Noting the existence of a 39-kb *Kpn* I fragment containing the $^G\gamma$-, $^A\gamma$-, δ-, and β-globin genes (Fig. 2), we constructed a *Kpn* I cosmid cloning vector and took advantage of the ≈10-fold enrichment of β-globin sequences by *Kpn* I digestion of genomic DNA and size selection of the 30- to 45-kb fragments. Four cosmids were obtained from 28,000 screened; results of cosmid cloning from two other individuals confirm that with this strategy approximately 1 in every 6000 cosmids will contain the β-globin cluster (unpublished data).

Identification of which cosmids represent which allele was accomplished by *Mst* II digestion. This enzyme directly recognizes the β^S-globin mutation, giving a fragment of 1.4 kilobases (kb) with β^S and 1.2 kb with β^A (25–27). This analysis is shown in Fig. 3*a*; two cosmids representing each allele were identified.

The β^S and $^G\gamma\beta^+$ HPFH cosmids were digested with several common restriction enzymes to look for any evidence of small deletions, insertions, or rearrangements that might have been missed by previous genomic blotting studies (7–9). None were found. An example is the *Msp* I digest shown in Fig. 3*b*. This also demonstrates that the C-C-G-G sites 5′ to the $^G\gamma$- and $^A\gamma$-globin genes are unaltered; thus the $^G\gamma\beta^+$ HPFH phenotype cannot be attributed to loss of one of these methylation sites.

DNA Sequencing. We then subcloned portions of the γ-globin genes and their 5′ flanking regions into bacteriophage M13mp8 and determined their sequence. Because the increased Hb F in $^G\gamma\beta^+$ HPFH is almost exclusively of the $^G\gamma$ type, the possibility of a mutation in the $^G\gamma$ regulatory region was investigated. The region from 540 bp upstream from the $^G\gamma$ cap site (−540) to codon 37 in the $^G\gamma$ second exon was sequenced, as was the region from −405 to codon 37 in the $^A\gamma$ gene. The $^A\gamma$ gene sequence exactly matched the sequence of the chromosome "B" allele from a non-HPFH patient described by Slightom *et al.* (ref. 28 and J. Slightom, personal communication). The entire γ-globin gene region sequence of the allelic chromosome "A" has been published (29). Allelic differences between the $^A\gamma$-globin genes of chromosome A and chromosome B occur at positions −369 (cytosine in A, guanosine in B), +25 in the 5′ untranslated region (adenosine in A, guanosine in B), and at position 65 in intervening sequence 1 (cytosine in A, thymidine in B).

The $^G\gamma$-globin gene sequence differs from the published A sequence (29) in five positions. Three of these must represent simple allelic differences, since they are also present in the non-HPFH chromosome B (J. Slightom, personal communication): these occur at positions −535 (guanosine in A, adenosine in B), −534 (adenosine in A, guanosine in B), and position 65 of intervening sequence 1 (cytosine in A, thymi-

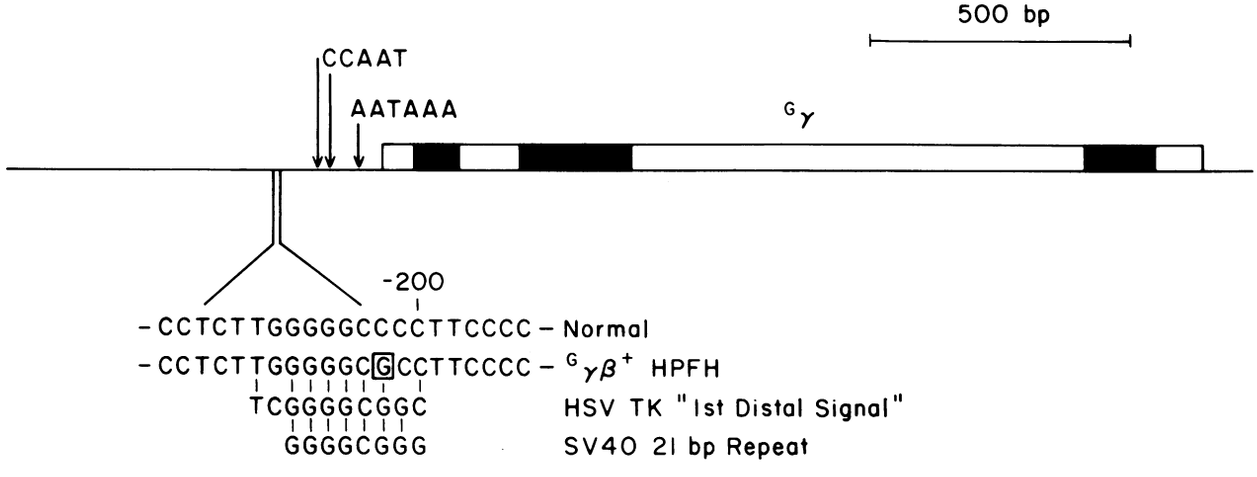

FIG. 4. Mutation 202 bp 5′ to the $^G\gamma$-globin gene cap site in $^G\gamma\beta^+$ HPFH. The regions from −540 to the second exon of the $^G\gamma$ gene and from −405 to the second exon of the $^A\gamma$ gene were sequenced. The only unique abnormality encountered in a total of 1500 bp is that shown. The sequence created has similarities to elements of the herpes thymidine kinase (HSV *TK*) first distal signal and the simian virus 40 (SV40) 21-bp repeat (here shown in inverted form), which have been shown to be important for efficient transcription.

dine in B). A fourth difference in the $^{G}\gamma$ flanking DNA is at −158, where the $^{G}\gamma\beta^{+}$ HPFH chromosome has a cytosine, whereas both A and B have a thymidine. However, both A and B have a cytosine at −158 of the normal $^{A}\gamma$-globin gene (29), which seems to rule against any functional significance of this finding. The fifth difference, guanosine instead of cytosine at −202, is *not* found in the 5′ flanking region of $^{G}\gamma$ or $^{A}\gamma$ of chromosome A or B but only in the $^{G}\gamma$-globin gene of the $^{G}\gamma\beta^{+}$ HPFH chromosome (Fig. 4).

This base difference abolishes a normal *Apa* I (G-G-G-C-C-C) site and creates an *Hha* I (G-C-G-C) site, which allowed direct confirmation that the mutation is present in both of the independently derived $^{G}\gamma\beta^{+}$ HPFH cosmids and in neither β^{S} cosmid. The analogous sequence in the 5′ flanking region of the $^{A}\gamma$ gene is unaltered in all cosmids.

Genomic Blotting. *Apa* I sites are rare in the β-globin cluster (5). Unequivocal demonstration of loss of one of these sites by genomic blotting requires double digestion with *Apa* I and another restriction enzyme (we used *Eco*RI). This was carried out and showed that, as expected, individuals I-1 and II-2 are heterozygous for loss of the $^{G}\gamma$ *Apa* I site, whereas this site was present on both chromosomes of 12 non-HPFH Black individuals (data not shown). DNAs from two other unrelated individuals with $^{G}\gamma\beta^{+}$ HPFH were subsequently subjected to the same analysis and show loss of the same *Apa* I site (R. W. Jones and D. J. Weatherall, personal communication; G. Stamatoyannopoulos and H. Kazazian, personal communication). An additional 60 β^{A} and 100 β^{S} chromosomes from non-HPFH Black individuals have also not shown the *Apa* I site mutation (C. D. Boehm and H. H. Kazazian, personal communication). The *Apa* I site is present in the $^{A}\gamma$ 5′ flanking region of all DNAs studied.

DISCUSSION

Understanding the mechanism of human hemoglobin switching has been a subject of intense interest in the past few years, particularly because of the conviction that increasing the expression of the fetal globin genes should be beneficial in treatment of the β-hemoglobinopathies (2, 3). The appearance of hypomethylation and nuclease hypersensitivity of the 5′ flanking region of the fetal globin genes has been shown to correlate with their expression (30, 31), but it remains unclear whether these changes are primary or are a result of other more fundamental and specific control mechanisms. A beginning therapeutic effort based on these observations has been made with the introduction of the hypomethylating agent 5-azacytidine (32–34), but its long-term toxicity remains unknown, as does its exact mode of action (35).

The HPFH syndromes are particularly interesting, therefore, because they offer the opportunity of identifying specific *cis*-acting DNA sequence elements that are involved in the switching mechanism. Many individuals with HPFH are found to harbor large deletions removing the δ- and β-globin genes but leaving the embryonic and fetal globin genes intact (36–40). These deletions have led to proposals regarding the possible location of control elements (39) or enhancers (40) in the globin cluster, but the fact that many involve a loss of more than 60 kb of DNA has made it difficult to use these to map all of the possible *cis*-acting elements. Some varieties of HPFH, however, do not involve detectable deletions and could well provide more subtle clues to normal developmental control of globin gene expression. Such a disorder is $^{G}\gamma\beta^{+}$ HPFH, which is present in the family described here and depicted in Fig. 1. The phenotype of the heterozygotes in this family matches closely to that described in the literature (3, 7–9): the presence of 15–25% Hb F in adult erythrocytes, which is almost exclusively $^{G}\gamma$, reduced but not absent β-globin synthesis in *cis*, and a benign course.

Cloning of the β-globin cluster from the $^{G}\gamma\beta^{+}$ HPFH allele and subsequent analysis with multiple restriction enzymes has allowed us to determine that no detectable deletions or rearrangements of more than 30–50 bp are present in the region cloned (Fig. 2), which confirms and further extends previous genomic blotting analysis (7–9). The phenotype, however, is striking in that the $^{A}\gamma$-globin gene turns off normally; this suggests the possibility of a mutation in the $^{G}\gamma$-globin control region, which was the impetus to focus our sequencing efforts on the 5′ regions of the fetal genes. The finding of a unique base change 202 bp 5′ to the $^{G}\gamma$-globin gene (Fig. 4) is thus an intriguing observation. There are reasons to suspect that this mutation might actually be the one responsible for the HPFH phenotype: (*i*) it falls in the general region (−50 to −300, see below) where other promoter elements important for differential control of gene expression have been described; (*ii*) it is the *only* mutation in the 1500 nucleotides sequenced that is not found in normal $^{G}\gamma$ or $^{A}\gamma$ DNA; (*iii*) the same *Apa* I site abolished by the mutation in our family is absent in two other $^{G}\gamma\beta^{+}$ HPFH patients not known to be related to each other or to our patient; and (*iv*) no examples of loss of this *Apa* I site have been encountered in studying more than 100 non-HPFH individuals. None of these arguments, however, completely excludes the possibility that the −202 mutation is a polymorphism in virtually complete linkage disequilibrium with the HPFH mutation.

Is it possible that the −200 area could normally have an influence on $^{G}\gamma$-globin gene expression? Dierks *et al.* (41) have studied in detail the expression of the rabbit β-globin gene incorporated in a pBR322-polyoma vector and transfected into mouse 3T6 cells. Deletions up to position −109 show no loss in transcription, whereas the duplicated C-A-C-C-C sequences from −85 to −105, the C-C-A-A-T box at −75, and the A-T-A box at −30 were all shown by site-directed mutagenesis to play important roles in efficent transcription. However, this does not rule out a more subtle role for sequences distal to −110, since these experiments were not done in erythroid cells; the need for inclusion in the vector of an enhancer sequence to obtain expression of non-α globin genes (41, 42) is further indication of the artificiality of the system. The recent identification of a tissue- and developmental-specific protein factor in chicken erythrocytes that specifically binds to a DNA region extending at least 268 bp 5′ to the chicken β-globin gene cap site (43) is evidence of the importance of the −100 to −300 region of globin genes *in vivo*. It is possible that 5′ tissue-specific promoter elements will turn out to be a common feature of specialized genes: the human insulin, rat insulin, and rat chymotrypsin genes have also been shown to possess such elements in the −170 to −300 region (44).

If the −202 $^{G}\gamma$-globin gene mutation is indeed responsible for overproduction of $^{G}\gamma$ in adult life, there are at least three possible mechanisms. First, the mutated sequence could serendipitously create a promoter-stimulating element out of previously nonfunctional DNA. It is interesting that the new sequence has the form G-G-G-G-Py-G (Fig. 4). The herpes simplex thymidine kinase promoter possesses two short inverted sequences of this form, shown by McKnight to be important for normal transcription (45, 46). The simian virus 40 early promoter similarly depends for efficient transcription on the six-times-repeated motif C-C-G-C-C-C (47, 48). In both instances, these elements function if inverted or moved upstream to about −150.

Second, the −202 mutation could augment transcription of the $^{G}\gamma$-globin gene by increasing the efficiency of a positive-regulatory sequence that normally occupies this region. The effect would have to be dramatic, however, to account for the approximately 40-fold increase in $^{G}\gamma$ expression.

Finally, the mutation might act to disable a negative regulatory sequence responsible for shutting off fetal globin syn-

thesis at the time of birth. Such a negative regulatory sequence has been described in the −107 to −210 region of the β-interferon gene (49) and conceivably could interact with a protein or RNA repressor molecule. This mechanism is a plausible way of explaining such a drastic increase in $^{G}\gamma$-globin synthesis by a single point mutation.

Note Added in Proof. The *Apa* I site 5′ to the $^{G}\gamma$ gene has now been studied by genomic blotting in three more families with $^{G}\gamma\beta^{+}$ HPFH (50). All 11 HPFH individuals in these families were found to lack the *Apa* I site.

We thank Dr. Blanche Alter for carrying out the globin chain electrophoretic analysis and Marsha Newman and Joseph Chamberlin for excellent technical assistance. F.S.C. acknowledges support from the Cooley's Anemia Foundation, Inc., and the Charles E. Culpepper Foundation. This work was supported in part by National Institutes of Health Grants GM-32156 (S.M.W. and C.J.S.) and AM-25376 (B.G.F.).

1. Bunn, H. F., Forget, B. G. & Ranney, H. M. (1977) *Human Hemoglobins* (Saunders, Philadelphia).
2. Pembrey, M. E., Wood, W. G., Weatherall, D. J. & Perrine, R. P. (1978) *Br. J. Haematol.* **40,** 415–429.
3. Wood, W. G., Weatherall, D. J., Clegg, J. B., Hamblin, T. J., Edwards, J. H. & Barlow, A. M. (1977) *Br. J. Haematol.* **36,** 461–473.
4. Wood, W. G., Clegg, J. B. & Weatherall, D. J. (1979) *Br. J. Haematol.* **43,** 509–520.
5. Collins, F. S. & Weissman, S. M. (1984) *Prog. Nucleic Acids Res. Mol. Biol.*, in press.
6. Weatherall, D. J. & Clegg, J. B. (1981) *The Thalassemia Syndromes* (Blackwell, Oxford), 3rd Ed.
7. Jones, R. W., Old, J. M., Wood, W. G., Clegg, J. B. & Weatherall, D. J. (1982) *Br. J. Haematol.* **50,** 415–422.
8. Balsley, J. F., Rappaport, E., Schwartz, E. & Surrey, S. (1982) *Blood* **59,** 828–831.
9. Farquhar, M., Gelinas, R., Tatsis, B., Murray, J., Yagi, M., Mueller, R. & Stamatoyannopoulos, G. (1983) *Am. J. Hum. Genet.* **35,** 611–620.
10. Serjeant, G. R., Grandison, Y., Lowrie, Y., Mason, K., Phillips, J., Serjeant, B. E. & Vaidya, S. (1981) *Br. J. Haematol.* **48,** 533–543.
11. Betke, K., Marti, H. R. & Schlicht, I. (1959) *Nature (London)* **185,** 1877–1878.
12. Rovera, G., Magarian, C. & Borun, T. W. (1978) *Anal. Biochem.* **85,** 506–518.
13. Alter, B. P., Goff, S. C., Efremov, G. D., Gravely, M. E. & Huisman, T. H. J. (1980) *Br. J. Haematol.* **44,** 527–534.
14. Kan, Y. W., Dozy, A., Trecartin, R. & Todd, D. (1977) *N. Engl. J. Med.* **297,** 1081–1084.
15. Blin, N. & Stafford, D. W. (1976) *Nucleic Acids Res.* **3,** 2302–2308.
16. Ish-Horowicz, D. & Burke, J. F. (1981) *Nucleic Acids Res.* **9,** 2989–2998.
17. Grosveld, F. G., Dahl, H. H., DeBoer, E. & Flavell, R. A. (1981) *Gene* **13,** 227–237.
18. Grosveld, F. G., Lund, T., Murray, E. J., Mellor, A. L., Dahl, H. H. M. & Flavell, R. A. (1982) *Nucleic Acids Res.* **10,** 6715–6732.
19. Maniatis, T., Fritsch, E. F. & Sambrook, J. (1982) *Molecular Cloning*: A Laboratory Manual (Cold Spring Harbor Laboratory, Cold Spring Harbor, NY).
20. Deininger, P. L. (1983) *Anal. Biochem.* **129,** 216–223.
21. Godson, G. N. (1983) in *Methods of DNA and RNA Sequencing*, ed. Weissman, S. M. (Praeger, New York), pp. 69–111.
22. Tuan, D., Biro, P. A., deRiel, J. K., Lazarus, H. & Forget, B. G. (1979) *Nucleic Acids Res.* **6,** 2519–2544.
23. Serjeant, G. R. (1975) *Clin. Haematol.* **4,** 109–122.
24. Mason, K. P., Grandison, Y., Hayes, R. J., Serjeant, B. E., Serjeant, G. R., Vaidya, S. & Wood, W. G. (1982) *Br. J. Haematol.* **52,** 455–463.
25. Chang, J. C. & Kan, Y. W. (1982) *N. Engl. J. Med.* **307,** 30–32.
26. Orkin, S. H., Little, P. F. R., Kazazian, H. H. & Boehm, C. D. (1982) *N. Engl. J. Med.* **307,** 32–36.
27. Wilson, J. T., Milner, P. F., Summer, M. E., Nallaseth, F. S., Fadel, H. E., Reindollar, R. H., McDonough, P. G. & Wilson, L. B. (1982) *Proc. Natl. Acad. Sci. USA* **79,** 3628–3631.
28. Slightom, J. L., Blechl, A. E. & Smithies, O. (1980) *Cell* **21,** 627–638.
29. Shen, S.-H., Slightom, J. L. & Smithies, O. (1981) *Cell* **26,** 191–203.
30. van der Ploeg, L. H. T. & Flavell, R. A. (1980) *Cell* **19,** 947–958.
31. Groudine, M., Kohwi-Shigematsu, T., Gelinas, R., Stamatoyannopoulos, G. & Papayannopoulou, T. (1983) *Proc. Natl. Acad. Sci. USA* **80,** 7551–7555.
32. Ley, T., DeSimone, J., Anagnou, N., Keller, G., Humphries, K., Turner, P., Young, N., Heller, P. & Nienhuis, A. (1982) *N. Engl. J. Med.* **307,** 1469–1475.
33. Ley, T., DeSimone, J., Noguchi, C., Turner, P., Schechter, A., Heller, P. & Nienhuis, A. (1983) *Blood* **62,** 370–380.
34. Charache, S., Dover, G., Smith, K., Talbot, C., Moyer, M. & Boyer, S. (1983) *Proc. Natl. Acad. Sci. USA* **80,** 4842–4846.
35. Kolata, G. (1984) *Science* **223,** 470–472.
36. Fritsch, E. F., Lawn, R. M. & Maniatis, T. (1979) *Nature (London)* **279,** 598–603.
37. Tuan, D., Murnane, M. J., deRiel, J. K. & Forget, B. G. (1980) *Nature (London)* **258,** 335–337.
38. Bernards, R. & Flavell, R. A. (1980) *Nucleic Acids Res.* **8,** 1521–1534.
39. Ottolenghi, S., Giglioni, B., Taramelli, R., Comi, P., Mazza, U., Saglio, G., Camaschella, C., Izzo, P., Cao, A., Galanello, R., Gimferrer, E., Baiget, M. & Gianni, A. M. (1982) *Proc. Natl. Acad. Sci. USA* **79,** 2347–2351.
40. Tuan, D., Feingold, E., Newman, M., Weissman, S. M. & Forget, B. G. (1983) *Proc. Natl. Acad. Sci. USA* **80,** 6937–6941.
41. Dierks, P., Van Ooyen, A., Cochran, M. D., Dobkin, C., Reiser, J. & Weissmann, C. (1983) *Cell* **32,** 695–706.
42. Humphries, R. K., Ley, T., Turner, P., Moulton, A. D. & Nienhuis, A. W. (1982) *Cell* **30,** 173–183.
43. Emerson, B. M. & Felsenfeld, G. (1984) *Proc. Natl. Acad. Sci. USA* **81,** 95–99.
44. Walker, M. D., Edlund, T., Boulet, A. M. & Rutter, W. J. (1983) *Nature (London)* **306,** 557–561.
45. McKnight, S. L. & Kingsbury, R. (1982) *Science* **217,** 316–325.
46. McKnight, S. L. (1982) *Cell* **31,** 355–365.
47. Everett, R. D., Baty, D. & Chambon, P. (1983) *Nucleic Acids Res.* **11,** 2447–2464.
48. Fromm, M. & Berg, P. (1982) *J. Mol. Appl. Genet.* **1,** 457–481.
49. Zinn, K., DiMaio, D. & Maniatis, T. (1983) *Cell* **34,** 865–879.
50. Gilman, J. G., Harano, T., Nakatsuji, T., Bakioglu, I., Reese, A. L., Gardiner, M. B. & Huisman, T. H. J. (1984) *Ann. NY Acad. Sci.*, in press.