The Deletion in Both Common Types of Hereditary Persistence of Fetal Hemoglobin Is Approximately 105 Kilobases

By Francis S. Collins, Jeffery L. Cole, Wendy K. Lockwood, and Michael C. Iannuzzi

The most common forms of hereditary persistence of fetal hemoglobin (HPFH) involve large deletions that remove the adult δ and β genes but leave the paired fetal genes (γ and δγ) intact. The size of these deletions has previously eluded exact definition. Using pulsed-field gel electrophoresis and the enzyme SfiI, which cuts only rarely in genomic DNA, we have constructed a large-scale restriction map of the β-globin cluster in normal and HPFH DNA. The deletions in HPFH-1, which occurs in American blacks, and in HPFH-2, which occurs in Ghanaian blacks, are found to be approximately 105 kilobases (kb) in length, though the endpoints are staggered by approximately 5 kb. The fact that two previously reported γδβ-thalassemia deletions to the 5' side of the β-globin cluster are also about 100 kb suggests a common mechanism, possibly involving the loss of a complete chromatin loop.

HEREDITARY PERSISTENCE of fetal hemoglobin (HPFH) is a clinically benign and genetically heterogeneous condition characterized by continued expression of fetal hemoglobin (HbF) into adult life, with relatively balanced globin chain synthesis.5-14 Several years ago, Southern blotting studies demonstrated that HPFH could be divided into two categories: those in which large genomic deletions removing the δ and β genes were present (deletion HPFH) and those in which the β-globin cluster appeared intact by Southern blotting (nondeletion HPFH).5,51 In many of the nondeletion forms the overproduction of HbF is observed to be primarily δγ or γ, and recent cloning studies have identified point mutations in the promoter of the overexpressed fetal gene in most of these conditions.8-14

In deletion HPFH, however, there is usually overproduction of both δγ and γ chains. The most common deletion type, denoted HPFH-1, occurs in American blacks and has a 5' breakpoint approximately 4 kilobases (kb) 5' to the δ gene (Fig 1).5,7,13,16 HPFH-2, which occurs in Ghanaian blacks, has both its 5' and 3' endpoints shifted approximately 5 kb in the 5' direction, but the total amount of DNA lost is within 1 kb of being the same as in HPFH-1.6,7,13,16 Interestingly, homozygotes for HPFH-1 and HPFH-2 have no abnormality of phenotype other than their high level of HbF, which suggests that no other essential genes lie within the deletion.

Using cloned DNA from the breakpoints of these and other deletions, several laboratories have attempted to identify the total amount of DNA deleted in HPFH-1 and HPFH-2 by genomic blotting and walking studies. The presence of many repeated and recombinogenic sequences, however, has slowed this effort. Mager et al17 recently concluded from their own walking efforts and a review of all available maps that the deletions must be at least 85 kb. We have now applied the technique of pulsed-field gel electrophoresis (PFGE),18-21 which allows the efficient separation of very large DNA molecules, to determine the actual size of these deletions and found them to be approximately 105 kb. This is remarkably close in size to the large deletions at the 5' end of the β-globin cluster that cause γδβ-thalassemia, which adds further support to a hypothesis originally proposed by Vanin et al18 that these deletions are not random but are generated by the loss of a complete chromatin loop.

MATERIALS AND METHODS

Cell lines. The source of HPFH-1 DNA was the cell line LAZ149, derived from a previously described and characterized individual homozygous for the HPFH-1 deletion7 that was originally established by Dr Herb Lazarus and kindly provided by Dr Bernard Forget. HPFH-2 DNA was obtained from cell line GM2064 (available from the Mutant Cell Repository, Camden, NJ), derived from a homozygote for HPFH-2.7 Another cell line from an HPFH homozygote, LAZ342, was obtained from Dr Stuart Orkin and was shown to be homozygous for HPFH-2 by Southern blotting (F. Collins, and J. Cole, unpublished observations) and by using the predicted fragment sizes in Table 1 of Tuan et al.15 DNA representative of the normal β-globin cluster was prepared from the lymphoblastoid cell line 31.1.0.20

DNA preparation. DNA for use in PFGE analysis must have a very high molecular weight. Accordingly, we prepared DNA in agarose gel blocks basically as described by Smith et al and Collins et al.22,23 In brief, cells were counted in a hemocytometer and then mixed with 2% low-melting-temperature (LMT) agarose (SeaPlaque, FMC Corp, Rockland, ME) and pipetted into a Lucite mold. The hardened blocks contained 2 × 107 to 2 × 108 cells in 80 μL and were then incubated for 36 hours in 500 mmol/L EDTA, 1% N-lauroylsarcosine, and 2 mg/mL protease K at 50°C to lyse the cells. Further incubations with 10 mmol/L Tris, pH 7.4, 1 mmol/L EDTA, (TE), 1 mmol/L phenylmethylsulfonylfluoride, and TE were carried out to purify the DNA. Digestions were carried out for six to eight hours at 37°C or 50°C (SfiI) by placing one block or a fraction of a block into an Eppendorf tube containing the appropriate restriction enzyme buffer in a total volume of 300 μL. We used 80 units of enzyme (all from New England Biolabs, Beverly, MA) for complete digests and 0.5 to 5 units for partial digests.

PFGE. Most gels were run by using the orthogonal field-alternation gel electrophoresis (OFAGE) modification of Schwartz and Cantor’s original apparatus.18,19 The principle involves alternation of two electric fields placed at approximately 90-degree angles.
A few gels were run on a commercially available PFGE apparatus (Pulsaphor, LKB Instruments, Houston). The running buffer was 0.4 x TBE (1 x TBE is 0.089 mol/L Tris base, 0.089 mol/L boric acid, 0.002 mol/L EDTA), and the gels were run at 14°C, 300 V, and 20 hours. Size markers were annealed λ DNA ladders, as previously described. Results were observed as DNA lengths of 254 nm UV irradiation to nick the DNA and then denatured and transferred to a nylon membrane (GeneScreen Plus, NEN, Boston) in 0.4 N NaOH.

Probes. The γ probe was a 3.5-kb HindIII fragment containing the entire γ gene. A 3D probe, kindly provided by Dr Bernard Forget, is a 1.0-kb BamHI-EcoRI fragment derived from a genomic clone containing the breakpoint in HPFH-1. A 3H probe is a 500-base pair (bp) HindIII fragment from a genomic clone kindly provided by Dr Dixie Mager that is located within the HPFH deletions, 9 kb upstream of the breakpoint of the Chinese γ-thalassemia deletion. The 3H probe is therefore approximately 27 kb 5’ to 3D and is derived from phage N317 as described by Mager et al. The 3I probe is a 680-bp FokI-BglII fragment isolated from near the breakpoint of the HPFH-3 deletion and is described by Henthorn et al. The 5’7 probe is a 3,264-bp EcoRI-HindIII fragment located 16 kb 5’ to the γ gene and was supplied by Dr Oliver Smithies. All probes were prepared by electrophoresis of the cut plasmids in LMT agarose and labeling of the desired fragment to a specific activity of >10⁹ cpm/µg by using the random hexanucleotide priming technique of Feinberg and Vogelstein.

Hybridization. Prehybridization and hybridization were performed according to the membrane manufacturer’s (GeneScreen Plus) recommendations. Blots were washed to a stringency of 0.1 x SSC (1 x SSC is 0.15 mol/L NaCl, 0.015 mol/L sodium citrate) at 65°C for 15 minutes and exposed against x-ray film for one to two days. As long as the blots were not allowed to dry, the probes could be stripped off and the blots rehybridized many times. Stripping was done by placing the blot in 0.4 N NaOH at 42°C for 30 minutes and then in 0.1 x SSC, 0.1% sodium dodecyl sulfate 0.2 mol/L Tris, pH 7.4, at 42°C for 30 minutes. Completeness of stripping was always assessed by exposure to x-ray film. A final hybridization was done with labelled λ DNA in most instances to define the size markers on the blot.

RESULTS

We performed genomic digests of lymphoblastoid cells from normal, homozygous HPFH-1, and two different homozygous HPFH-2 individuals by using a number of restriction enzymes that cut rarely in mammalian genomic DNA. These included MluI, BssHII, NotI, PvuI, SacII, and SfiI, which are all expected to produce fragments of 10 to 1,500 kb. The DNA samples were prepared and restricted in agarose gel blocks, and then the DNA was electrophoresed by PFGE using annealed λ DNA ladders as a size marker (Fig 2). The DNA was transferred to a nylon membrane and hybridized. Only SfiI gave detectable bands on the blot with any of the globin probes; all other enzymes gave a signal only at the top of the gel lane, which indicated a very large (>1,000 kb) fragment or gave no signal at all. The same blots rehybridized with nonglobin probes gave the expected single bands in these lanes (data not shown), thereby indicating that the DNA had been properly digested and transferred.

The results with SfiI, however, were useful. This enzyme recognizes the sequence GGCCNNNNNGGCC and cuts...
approximately every 250 kb in human genomic DNA. Unlike the other rare cutters tested, most SfiI sites are not expected to be blocked by methylation because the recognition site does not include a CpG dinucleotide. Probing digests of normal and two different HPFH-2 DNAs (Fig 2) with an $\alpha\gamma$ sequence detects a 140-kb band for the normal and a 75-kb band for the HPFH-2 DNAs. This blot was then stripped, exposed to x-ray film to confirm complete removal of the first probe, and rehybridized with the 3D probe shown in Fig 1, which is just distal to the HPFH-1 breakpoint. The HPFH-2 DNA reveals the same 75-kb band, which is expected because $\alpha\gamma$ and 3D are only about 10 kb apart in the HPFH deletion. In normal DNA, however, a 40-kb SfiI fragment is detected, thereby indicating that at least one SfiI site lies within the deletion region.

It is possible to locate the 5' end of the normal fragment containing the $\alpha\gamma$ gene exactly because there is a known SfiI site approximately 17 kb 5' to the $\epsilon$ gene for which an extensive 5' flanking sequence was recently reported by Li et al.\textsuperscript{26} Using a double digest of genomic DNA with EcoRI and SfiI, we were able to show with a standard Southern blot that this SfiI site cuts to completion in genomic DNA. Furthermore, a 675-bp AvaII-StuI fragment of the 5'7 probe, located just upstream of this SfiI site, hybridizes with a 10-ko SfiI fragment (data not shown). No other SfiI sites exist in the sequenced portion of the $\beta$-globin cluster.\textsuperscript{2,26}

To complete the restriction map and define the size of the HPFH deletions, it was necessary to determine whether additional SfiI fragments reside within the deletion region. The furthest available probe 3' to the $\beta$ gene, denoted 3'IH and derived from the breakpoint of the Indian type of HPFH (HPFH-3), has been determined to lie approximately 55 kb 3' to the $\alpha\gamma$ gene\textsuperscript{25} and therefore would be predicted to reside on the same 140-kb SfiI fragment as $\alpha\gamma$. This was confirmed by hybridization of a PFGE blot (data not shown). From the other direction, Mager et al. have cloned contiguous DNA sequences extending 30 kb 5' to 3D.\textsuperscript{17} Restriction digests of this cloned DNA revealed the presence of an SfiI site approximately 19 kb 5' to 3D (data not shown). A 500-bp HindIII fragment located 8 kb 5' to this SfiI site was gel purified; we denote this as H500. When this is used as a probe against normal and HPFH-2 DNA and the results compared with hybridization of the same blot with the $\gamma$ probe, it is clear that H500 and $\gamma$ are detecting the same 140-kb SfiI fragment in normal DNA (Fig 3). That the fragment detected by H500 and $\gamma$ is the same one and not a coincidental comigration of bands of equivalent size is further demonstrated by the identity of the larger bands, which result from the intentional partial digest used in this blot. The sizes of the partial-digest bands are readily explained by the position of the flanking SfiI sites. As expected, H500 gives no signal from HPFH-2 DNA.

This allows the construction of a complete SfiI map of the $\beta$-globin cluster and its 3' flanking region (Fig 4). One can calculate the size of the HPFH-2 deletion as $140 + 40 - 75 - 105$ kb. The calculation of these deletion sizes is quite sensitive to errors in the estimation of the PFGE blot bands. The use of $\lambda$ concatemers allows interpolation of sizes to an accuracy of about 5 to 10 kb so long as the DNA fragments in the genomic digest lanes are migrating faithfully. We investigated the effect of the amount of DNA in the genomic digests on their electrophoretic mobility (Fig 5). The amount of DNA in a block is calculated from the number of cells used and the presence of 6.7 pg DNA/cell. We found that migration is consistent when the amount of DNA per lane is less than 2 to 3 pg. Above this amount, nonequilibrium effects in the overloaded lane result in a shift upward of the hybridizing band, thus leading to falsely large estimates of size. The size estimates shown in Fig 4 have all been determined by using 2 pg/lane or less of DNA. Allowing for a 5- to 10-kb error in the 40-kb fragment, and another 5- to 10-kb error in the difference between the 140- and 75-kb fragments (the difference is what enters the calculation), we would estimate that the measured HPFH-2 deletion size of 105 kb should be accurate to within approximately 10 to 20 kb.

Previous Southern blotting comparisons of the deletion endpoints in HPFH-1 and HPFH-2 have demonstrated that the total amount of deleted DNA is very similar,\textsuperscript{6,7,15,16} with the HPFH-1 deletion being about 1 kb larger. Partial SfiI
Fig 4. Map of the β-globin cluster. The vertical arrows mark the location of Sfi sites. The sizes of Sfi fragments and genomic deletions are given in kilobases. Note the similarity in size of the five deletions.

digests of HPFH-1 and HPFH-2 DNA were separated by OFAGE, blotted, and probed with the γ and 3D probes (data not shown). The fragments were indistinguishable, as expected because the 1-kb difference would not be resolved with this approach.

**DISCUSSION**

The recent introduction of PFGE combined with enzymes that cut only rarely in genomic DNA allows the characterization of large regions that previously were relatively inaccessible to molecular biologic analysis.28 This technique provides a useful bridge between standard molecular genetic approaches (which are difficult to extend to sizes greater than 50 kb) and cytogenetic analysis (which does not provide resolution below about 2,000 kb). The PFGE blotting approach has recently been applied successfully to the region of the Duchenne muscular dystrophy gene29,32 and to the human major histocompatibility complex.30,31 Here we have shown the application of this technology to the human β-globin complex. An Sfi1 restriction map of the β-globin complex and its 3' flanking region has been constructed and used to determine the size of the HPFH-1 and HPFH-2 deletions without the need for time-consuming and difficult chromosome walking experiments. The map constructed here should be quite useful for analyzing other deletions.

The experiments reported here demonstrate that the deletions in HPFH-1 and HPFH-2 are approximately 105 kb. A third deletion, Chinese Gγ(γδβ)-thalassemia, has previously been shown to involve the loss of a similar length of DNA as HPFH-1 (about 6 kb less) but with the endpoints shifted relative to the HPFH deletions (Fig 4).16,17 Thus we can conclude that the Chinese deletion is approximately 100 kb in length. In the 5' direction, there are two other previously described large deletions for which the endpoints have been mapped, both of which give rise to γδβ-thalassemia. γδβ-Thalassemia 1 was described by Van der Ploeg et al33 and has its 3' breakpoint 2.5 kb 5' to the β-globin gene, and γδβ-thalassemia 2 was subsequently shown by Orkin et al34 to have its 3' breakpoint located within the β-globin gene. By cloning DNA located at the 5' breakpoints of these forms of γδβ-thalassemia, Vanin et al16 were able to demonstrate that both of these mutations were simple deletions and that the total amount of deleted DNA was very similar. The absolute size of the γδβ-thalassemia deletions was subsequently determined by Taramelli et al35 who performed a cosmid chromosome walk crossing the entire region of these two deletions and demonstrated that they were 99.4 and 95.9 kb respectively. All five of these deletions, therefore, have removed 100 ± 5 kb genomic DNA, as shown in Fig 4.
Vanin et al.,\(^\text{16}\) noting that the \(\gamma\beta\)-thalassemia deletions were of similar but at that time unknown size and that the HPFH deletions were also of similar but unknown size, have suggested a mechanism involving the loss of a complete chromosomal loop during DNA replication. What we have shown in this work is that the 5' and 3' deletions are all of approximately the same size, namely, about 100 kb. The model that we propose to explain this phenomenon, which is similar to that previously proposed, is built upon the following observations: (a) all of these deletions join nonhomologous DNA sequences,\(^\text{16}\) which mitigates against a simple unequal crossing-over mechanism; (b) an accumulating body of evidence suggests that DNA is anchored to the nuclear matrix in vivo and is arranged in loops of length 30 to 130 kb\(^\text{33-38}\); (c) a simple loss of a chromosomal loop between fixed attachment points would not account for several staggered deletions unless DNA is capable of reeling through the anchorage points; (d) this reeling phenomenon may well occur during DNA replication because evidence suggests that DNA replication and topoisomerase activity is concentrated in the matrix\(^\text{37}\); and (e) DNA replication is bidirectional in prokaryotes and is probably so in eukaryotes.\(^\text{37,38,40}\)

In Fig 6 a possible mechanism for the generation of these deletions is depicted. Two matrix attachment points, M\(_1\) and M\(_2\), are postulated to be located physically close together and generate a loop of 100 kb containing the expressed genes of the \(\beta\)-globin cluster. During DNA replication, DNA is reeled into these attachment points from both directions, and daughter strands are reeled out. So long as replication proceeds at the same rate at M\(_1\) and M\(_2\), a faulty connection of the M\(_1\) strand to the M\(_2\) strand (with breaks occurring at M\(_1\)A and M\(_2\)A or M\(_1\)B and M\(_2\)B, as marked by arrows in the

---

Fig 6. Postulated mechanism for generation of this series of \(\beta\)-globin deletions. M\(_1\) and M\(_2\) are two adjacent matrix attachment points. DNA replication occurs in a bidirectional fashion from these points, which leads to reeling in (heavy arrows) of DNA. Breaks at analogous positions on M\(_1\) and M\(_2\) followed by rejoining would then lead to deletions corresponding to the size of the original loop. Thus, in the middle panel, breakage at arrows marked \(\ldots\ldots\) would lead to Chinese \(\gamma\beta\)-thalassemia, and in the lower panel, arrows with \(\ldots\ldots\) will lead to HPFH-1. HPFH-2 and \(\gamma\delta\)-thalassemia 2 can be accounted for by the same model.
...figure) will result in the deletion of 100 kb. Examples are shown in Fig 6. Smaller or larger (up to 200 kb) deletions of varying size would be generated if breakage and rejoining occurred between the A and B attachment points at M₁ and M₂. In fact, several smaller deletions in the β-globin cluster have been described, and at least one deletion has been reported that removes the entire cluster.

The model accounts for these but also predicts the observed series of staggered deletions of approximately 100 kb. Note that M₁ and M₂ must be located approximately as drawn in Fig 6 if this loop alone is to account for both the 5′ γδβ-thalassemia and the 3′ HPFH mutations. If adjacent loops are also of 100-kb size, however, the matrix attachment points could be located anywhere in this region.

If this model is correct, one would expect to observe similar phenomena in deletions at other loci, though the size of the deletions would be expected to vary from locus to locus according to the location of nuclear matrix attachment points. Interestingly, Nicholls et al have recently performed a detailed analysis of deletions in the human α-globin locus, and a series of staggered deletions involving the loss of 20 to 30 kb has been identified.

An unexpected finding of this study was the inability to detect bands of measurable (<1,000 kb) size with a wide variety of restriction enzymes that ordinarily generate fragments of 100 to 1,000 kb in mammalian genomic DNA. All of the enzymes tested, except SfiI, have recognition sites containing at least one CG dinucleotide, and current evidence suggests that cutting by these enzymes is blocked by methylation of the cytosines. The ability to generate measurable restriction fragments with these enzymes in other parts of the genome is felt to be attributable at least in part to the presence of so-called HpaII tiny fragment islands, which are highly GC rich sequences often found in the 5′ flanking region of housekeeping genes and usually unmethylated in this setting. The fact that none of these methylation-sensitive enzymes would cut within 1,000 kb of the β-globin cluster suggests that the domain of DNA methylation around this region is very wide, at least in lymphoblastoid cells. In fact, there are MluI and BssHII sites in sequenced DNA of the β-globin cluster, and we have shown that in lymphoblastoid cells these sites do not cut at all in Southern blotting experiments (F. Collins and J. Cole, unpublished data). It would not be surprising if there are other regions of the human genome that are relatively deficient in restrictable methylation-sensitive sites, and this fact will have important consequences for general strategies to construct a large-scale restriction map of the genome.

Acknowledgment

We thank Drs Bernard Forget and Stuart Orkin for HPFH cell lines and Drs Dixie Mager, Oliver Smithies, and Bernard Forget for supplying some of the probes used in this study. We also thank Danielle Rennert for preparing the manuscript and Dr David Ginsburg for helpful comments.

References

20. Carle GF, Frank M, Olson MV: Electrophoretic separations
DELETION HPFH

of large DNA molecules by periodic inversion of the electric field. Science 232:65, 1986


27. Feinberg AP, Vogelstein B: A technique for radiolabelling DNA restriction fragments to high specific activity. Anal Biochem 137:266, 1984


33. Van der Ploeg LHT, Konings A, Oort M, Roos D, Bernini L, Flavell RA: \( \gamma-\beta \)-Thalassemia studies showing that deletion of the \( \gamma \)- and \( \delta \)-genes influences \( \beta \)-globin gene expression in man. Nature 283:637, 1980


35. Taramelli R, Kioussis D, Vanin E, Bartram K, Groffen J, Hurst J, Grosveld FG: \( \gamma-\delta \)-Thalassemias 1 and 2 are the result of a 100 kbp deletion in the human \( \beta \)-globin cluster. Nucleic Acids Res 14:7017, 1986


The deletion in both common types of hereditary persistence of fetal hemoglobin is approximately 105 kilobases

FS Collins, JL Cole, WK Lockwood and MC Iannuzzi