Volume 13 Number 18 1985 Nucleic Acids Research

A Chinese $^{G}\gamma+(^{A}\gamma\delta\beta)^{o}$ thalassemia deletion: comparison to other deletions in the human β-globin gene cluster and sequence analysis of the breakpoints

Dixie L.Mager, Paula S.Henthorn* and Oliver Smithies

Laboratory of Genetics, University of Wisconsin, Madison, WI 53706, USA

Received 24 June 1985; Revised 19 August 1985; Accepted 23 August 1985

ABSTRACT

A clone was isolated that contains the deletion junction region from an individual with a deletion associated with Chinese $^{G}\gamma+(^{A}\gamma\delta\beta)^{o}$ thalassemia. A clone containing the normal DNA corresponding to the 3' breakpoint of this deletion was also isolated. Portions of these two clones were sequenced and compared to the region in the $^{A}\gamma$-globin gene where the 5' breakpoint occurs. This comparison reveals that the breakage and reunion event was nonhomologous and that it probably involved the insertion of 36-41 bases of DNA belonging to the L1 (KpnI) family of repetitive DNA. Genomic mapping revealed that the DNA on the 3' side of this deletion is closely linked in normal DNA to the 3' breakpoints of two different large deletions that are associated with hereditary persistence of fetal hemoglobin (HPFH). We cloned and mapped 35 kbp of normal DNA from this region (greater than 45 kbp downstream of the human ß-globin gene) that contains the 3' breakpoints of the Chinese thalassemia and the two HPFH deletions. An endogenous retrovirus-like element and several other repetitive sequences are located within this region. We show that the Chinese thalassemia deletion is greater than 80 kbp in length and differs in size from the two HPFH deletions by less than 6%. We also show that the Chinese thalassemia deletion is at least 40 kbp larger than several other deletions associated with a very similar phenotype.

INTRODUCTION

The human ß-globin gene cluster is one of the most thoroughly studied regions of DNA in any mammal. Numerous naturally-occurring mutations affecting expression of the adult ß-globin gene have been described and characterized at the DNA sequence level (1, 2). The majority of these mutations are single nucleotide changes, but some are due to large deletions or rearrangements of DNA. Some of the deletion mutations are associated with changes in the expression of genes in the ß-globin cluster that are not easily interpretable. Most of the deletions are associated with thalassemia of varying degrees of severity, but some give rise to the virtually asymptomatic condition of hereditary persistence of fetal hemoglobin (HPFH) in which the fetal γ-globin genes remain active in adult life (3). Two hypotheses have been suggested to account for this persistence of fetal chain synthesis: 1,

the deletions remove sequences within the ß-globin gene cluster that normally suppress expression of the fetal globin genes in adults (1, 4); or 2, the deletions juxtapose sequences next to the fetal genes that act to increase their expression (1, 5). We have studied deletions associated with abnormal globin gene expression in an attempt to distinguish between these two possibilities and to gain insight into the normal process of hemoglobin switching during development.

Naturally-occurring deletions are also useful in the investigation of the nature of recombinational events that occur within mammalian genomes. Close to 20 different deletions that remove part or all of the ß-globin gene or affect its expression in some way have been identified (1). Two of these, the deletions associated with Hemoglobins Lepore and Kenya, appear to be the result of homologous recombination between closely related linked genes (6, 7). DNA sequence analysis has revealed that four of the deletions associated with different types of ß thalassemia resulted from nonhomologous DNA exchanges (8-10). We have continued to analyze other ß-globin gene deletions with the aim of learning more about the mechanisms involved in these DNA rearrangements.

We report here the cloning of the deletion junction region in a case of $^{G}\gamma^{+}(^{A}\gamma\delta\beta)^{o}$ thalassemia found in China (11, 12). The DNA sequence of the deletion junction was determined and compared to the two normal sequences involved in its formation. The sequence features revealed by this analysis are discussed in terms of the recombination event that led to the deletion. In addition, the normal DNA encompassing the 3' breakpoints of the Chinese thalassemia deletion and two deletions associated with HPFH was isolated and characterized. The extent of the Chinese $^{G}\gamma^{+}(^{A}\gamma\delta\beta)^{o}$ thalassemia deletion was compared to the extents of other deletions associated with $^{G}\gamma^{+}(^{A}\gamma\delta\beta)^{o}$ thalassemia and with HPFH. This comparison is discussed in relation to the mechanisms of generating large deletions and in relation to the hypotheses regarding the persistence of fetal globin gene expression caused by gene deletion.

MATERIALS AND METHODS

Sources

The normal human DNA used in all cloning and mapping experiments was isolated from cultured fibroblasts from a normal female abortus (13). DNA from the individual (AH) heterozygous for the Chinese thalassemia deletion was provided by Dr. David Weatherall.

Probes and Hybridizations

The γIVS2 probe (box 1 in Figure 1c) is a 457 bp BamHI-PvuII fragment from the large intervening sequence of the Aγ-globin gene. Probes 2 and 3 (boxes 2 and 3 in Figure 1c) are respectively a 470 bp RsaI-HindIII fragment and a 1.4 kbp EcoRI-StuI fragment from the 3' side of the Chinese thalassemia deletion junction. The 3' HPFH-2 probe mentioned in the text but not shown in Figure 1 is a 1.08 kbp SstI-BamHI fragment (called probe W in ref. 8) from the 3' side of the HPFH-2 deletion junction.

The L1 (KpnI) repeat family probes, pCa4A-10 and pCa4A-12, are two halves of a long primate L1 family member and were provided by Dr. Maxine Singer. The Alu family probe was isolated in our laboratory and contains the Alu repeat found immediately 5' to the Gγ-globin gene.

Mapping of genomic DNA and hybridizations were performed as described previously (8).

Cloning

Clone ChT-1, mentioned briefly elsewhere (14), was obtained from a library constructed by isolating, by preparative agarose gel electrophoresis, 5 to 7 kbp fragments from a complete BamHI-BglII digest of DNA from an individual (AH) with Chinese thalassemia and ligating these fragments into BamHI arms of DNA from λ phage Charon 27 (15). The ligated DNA was then packaged *in vitro* (16) and the resulting phages screened without amplification. Clone B1H2-1 has been described previously (8) and contains a 14 kbp deletion-spanning BamHI fragment from an HPFH-2 homozygous cell line.

Phages N3'ß-2 and N3'ßdel-3 were obtained from a library containing 8-10 kbp fragments of a complete HindIII digest of normal DNA ligated into HindIII arms of λ phage Charon 30 (15). Phages N3'ß-1 and N3'ß-4 were isolated from a Charon 32 (17) partial EcoRI library constructed by Dr. Jerry Slightom from normal DNA. Phage N3'ß-5 was isolated from a Charon 30 partial Sau3A library constructed in our laboratory by Dr. Elio Vanin from normal DNA. Probe 2 (Figure 1c) was used to obtain phages N3'ß-1 and N3'ß-2. Phages N3'ß-4 and N3'ß-5 were obtained using the 1.08 SstI-BamHI fragment from phage B1H2-1. Phages N3'ßdel-3 and N3'ß-4 were referred to previously as N30Hdel-1 and N32E-14 respectively (14).

DNA Sequence Analysis

All DNA sequences were determined by the method of Maxam and Gilbert (18). More than 90% of the sequences were determined from both DNA strands and the remainder were sequenced at least twice. DNA sequences were analysed

using software provided by the University of Wisconsin Genetics Computer Group (19).

RESULTS

Isolation of a Clone Spanning the Chinese Thalassemia Deletion Junction

From genomic mapping and solution hybridization analysis, Jones et al. (12) determined that the deletion associated with Chinese $^{G}\gamma^{+}(^{A}\gamma\delta\beta)^{o}$ thalassemia (Ch-thal) has its 5' endpoint in the $^{A}\gamma$-globin gene and extends in the 3' direction deleting the ψß-, δ- and ß-globin genes. (The terms 5' and 3' are assigned relative to the direction of globin gene transcription.) The 5' breakpoint of the Ch-thal deletion was localized to between the XbaI and HpaI sites in the large intervening sequence of the $^{A}\gamma$ globin gene (12). A restriction enzyme map of the affected chromosome in the region of the deletion junction is shown in Figure 1b where it is compared with a map of the normal fetal globin gene locus shown in Figure 1a. The sites 3' to the deletion junction were determined by mapping DNA from an individual (AH) heterozygous for the Chinese thalassemia deletion (12, unpublished data). The map of the DNA on the 3' side of the deletion is not similar to any mapped region downstream of the ß-globin gene (1, 20, unpublished data) suggesting that the 3' deletion breakpoint is in a region not previously cloned.

To characterize the Ch-thal deletion junction, a phage clone containing the 6 kbp BamHI-BglII fragment that spans the deletion junction was isolated and analyzed. A restriction enzyme map of the insert of this clone, ChT-1, is shown in Figure 1c and agrees with the genomic map of DNA from the patient AH (Figure 1b). The first mapped restriction enzyme site on the 3' side of the deletion junction in the cloned DNA is a RsaI site (starred in Figure 1c). Thus the deletion junction occurs within the PvuII-RsaI fragment of ChT-1, a distance of approximately 280 bp. The 1.2 kbp BamHI-EcoRI fragment from phage ChT-1 that spans the deletion junction was subcloned into a plasmid and the DNA sequence of the 280 bp PvuII-RsaI fragment was determined.

Isolation of a Normal Clone Containing the 3' Deletion Breakpoint

To clone the normal DNA surrounding the Ch-thal 3' breakpoint, two fragments containing single copy sequences from the 3' side of the deletion junction (boxes 2 and 3 in Figure 1c) were isolated from subclones of phage ChT-1. These fragments were used as probes to screen phage libraries of normal human DNA for clones containing the 3' breakpoint region. One clone containing the region was obtained when a Charon 32 phage library of partially digested EcoRI fragments was screened with probe 2. This clone, N3'ß-1, contains a 13 kbp insert and the region hybridizing to probe 2 was localized

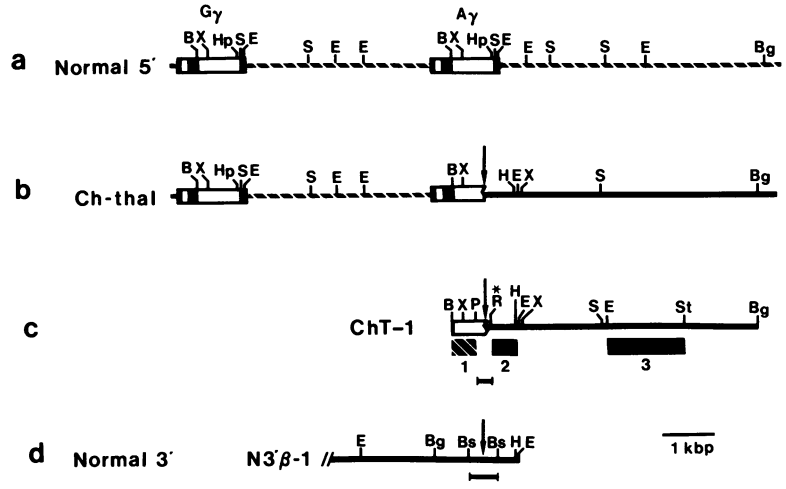

Figure 1. a) Restriction enzyme map of the human fetal globin gene region. The exons and introns of the $^{G}\gamma$-and $^{A}\gamma$-globin genes are shown by solid and open boxes respectively. b) Restriction map of the Chinese thalassemia deletion junction region (12, unpublished data). The sites 3' to the deletion junction (marked by a vertical arrow) were derived by mapping the genomic DNA of an individual heterozygous for the Ch-thal deletion. c) Restriction map of the phage clone ChT-1. Box 1 represents the γIVS2 probe used to isolate the clone. Boxes 2 and 3 show probes isolated from ChT-1. The bar shows the 280 bp PvuII-RsaI fragment that was sequenced. d) Map of the 3' portion of phage N3'ß-1. The bar indicates the 540 bp BstNI fragment that was sequenced. Dashed lines represent normal DNA from around the 5' deletion breakpoint and solid lines represent normal DNA from around the 3' breakpoint. B, BamHI; Bg, BglII; Bs, BstNI; E, EcoRI; H, HindIII;, Hp, HpaI; R, RsaI; S, SstI; St, StuI; X, XbaI. Not all BstNI, HindIII, RsaI, PvuII and StuI sites are shown.

to one end of this insert. Figure 1d is a map of this region of the clone which also includes the 3' breakpoint of the Ch-thal deletion (marked with an arrow). The right-most 1.6 kbp BglII-EcoRI fragment from N3'ß-1 was subcloned and the DNA sequence of a 540 bp BstNI fragment containing the 3' breakpoint was determined.

Comparison of the Sequences Involved in Deletion Formation

Figure 2a compares the DNA sequence of the Ch-thal deletion junction region and of the regions containing the 5' and 3' breakpoints in normal DNA. The normal 5' DNA sequence is from the large intervening sequence of the Aγ-globin gene and the normal 3' DNA sequence is from phage N3'ß-1. A comparison of the normal 5' and deletion-spanning sequences shows that the 5' breakpoint in the normal $^{A}\gamma$-globin gene is between positions 192 and 193 of

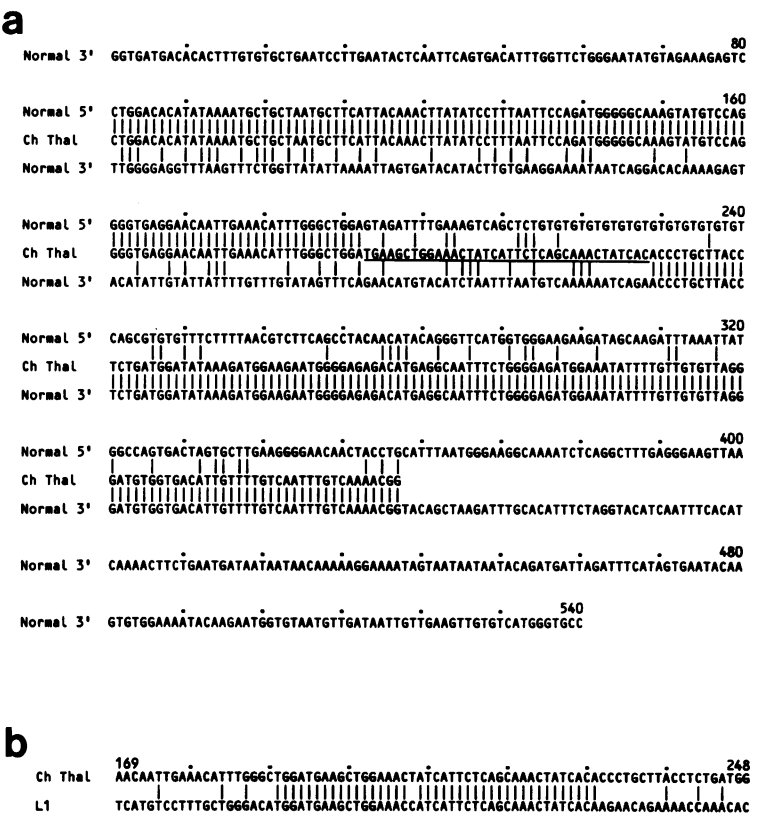

Figure 2. a) Comparisons of sequences across the deletion junction of the Ch-thal deletion and of the corresponding normal 5' and 3' DNAs. The normal 5' sequence is from Slightom et al. (13) and includes their residues 929-1268 of the 165.24 $^{A}\gamma$-globin gene sequence. The Ch-thal sequence is that of the 280 bp PvuII-RsaI fragment from phage ChT-1. The normal 3' sequence is that of the 540 bp BstNI fragment from phage N3'ß-1. The apparent insertion of DNA at the deletion junction is underlined. Vertical lines connect identical bases and dots are placed at 10 bp intervals.

b) Comparison of the Ch-thal sequence (bases 169-248 of part a) to residues 5685-5764 of the L1 family consensus sequence (M. Singer, personal communication). This region of the consensus long L1 family repeat was compiled from 8 different sequences.

Figure 2a because that is where the homology between the two sequences ceases. The most distinctive sequence feature near the 5' breakpoint in normal DNA is a region of simple sequence consisting primarily of the dinucleotide TG that is located 22 bp downstream of the 5' breakpoint. The 3' breakpoint in normal DNA is between positions 228 and 229 in Figure 2a. No distinctive sequence features are present in this region.

Alignment of all three sequences shows that there are 36 nucleotides in the Ch-thal deletion-spanning sequence (underlined in Figure 2a) that are not found in either of the two normal DNA sequences. A search for the possible origin of this 36 bp segment revealed that it is a member of the L1 (KpnI) family of repetitive DNA. Full length members of this family are over 6 kbp long but the family members are highly heterogeneous in size with most being truncated, usually from the 5' end (for review see 21). A compilation or consensus sequence of a full length primate L1 family member has been assembled from many L1 sequences by Singer and co-workers. Figure 2b shows the alignment of the deletion-spanning sequence to an 80 bp portion of the L1 consensus sequence. A 41 bp sequence (the 36 bp segment plus 4 bases on its 5' side, TGGA, and an A residue on its 3' side) is homologous to the L1 sequence with only one mismatch. On either side of these 41 bp there is no significant homology between the deletion-spanning sequence and any L1 family sequences. Therefore a 36 to 41 bp L1 family repetitive sequence appears to have been inserted at the Ch-thal deletion junction.

The normal DNA in the vicinity of the 5' and 3' breakpoints was examined for local sequence features such as stretches of homology or direct or inverted repeats that might have facilitated the deletion event. When the sequences were aligned as in Figure 2a, no homology between the normal 5' and 3' DNAs was found. These sequences also showed no homology when the apparent 36 bp L1 family insertion in the deletion-spanning sequence was removed and the sequences realigned (data not shown). No significant direct repeats could be detected in the normal DNAs or the deletion-spanning sequence. In addition, when 60 bp sequences immediately surrounding each breakpoint were analyzed for possible secondary structure, no structures were found that were significantly more stable than those found in several other 60 bp regions of the sequences shown in Figure 2a. Thus the only unusual sequence feature detected in the normal 5' and 3' DNAs is the simple sequence "TG" repeat located near the 5' breakpoint.

Linkage of the 3' Breakpoints of the Chinese Thalassemia and HPFH-2 Deletions

Because the Ch-thal deletion extends an unknown distance 3' to the ß-globin gene into a region not previously characterized, we used probes 2 and 3 (Figure 1c) to map the normal DNA encompassing the 3' breakpoint. We have also mapped the normal DNA surrounding the 3' breakpoint of a deletion, HPFH-2, that is associated with hereditary persistence of fetal hemoglobin (22, 23). This deletion and different deletion of similar size associated with HPFH-1 (5, 8, 23, 24) also extend an unknown distance 3' to the ß-globin

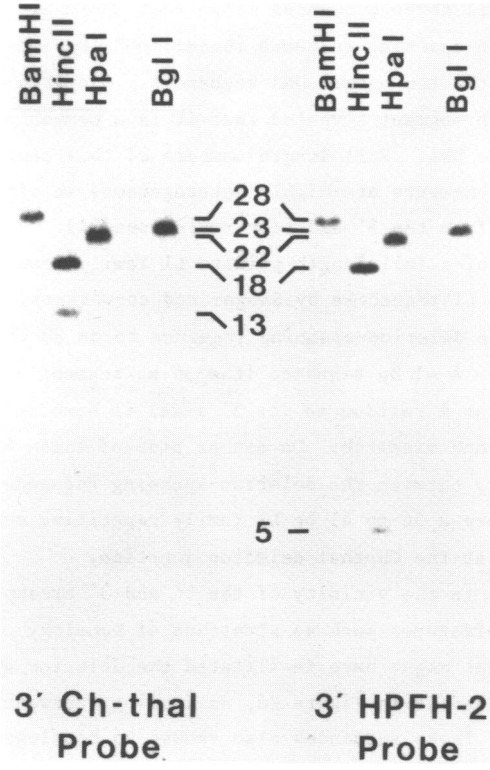

Figure 3. Autoradiographs of Southern transfers of normal human DNA hybridized to the 3' Ch-thal probe (box 2 in Figure 1c) and the 3' HPFH-2 probe W. Both sets of digests were electrophoresed on the same gel.

gene. The probe used for mapping the normal DNA encompassing the 3' breakpoint of the HPFH-2 deletion is from the 3' side of the HPFH-2 deletion junction (called probe W in ref. 8). Figure 3 shows that probe 2 (see Figure 1c) and the 3' HPFH-2 probe W hybridize to the same-sized BamHI, HpaI, BglI and HincII restriction fragments. The two probes therefore must hybridize to DNA sequences within 18 kbp of each other since the smallest common fragment (HincII) is 18 kbp in length. In the HincII digests, probe 2 also hybridizes to a 13 kbp fragment and the 3' HPFH-2 probe W to a 5 kbp fragment (Figure 3). These fragments are formed when a HincII site located between the regions hybridizing to the two probes is cleaved. This HincII site does not cleave completely in DNA samples derived from fibroblasts and this results in the presence of a partially digested 18 kbp HincII fragment. This genomic mapping analysis demonstrates that the 3' breakpoints of the Ch-thal and HPFH-2

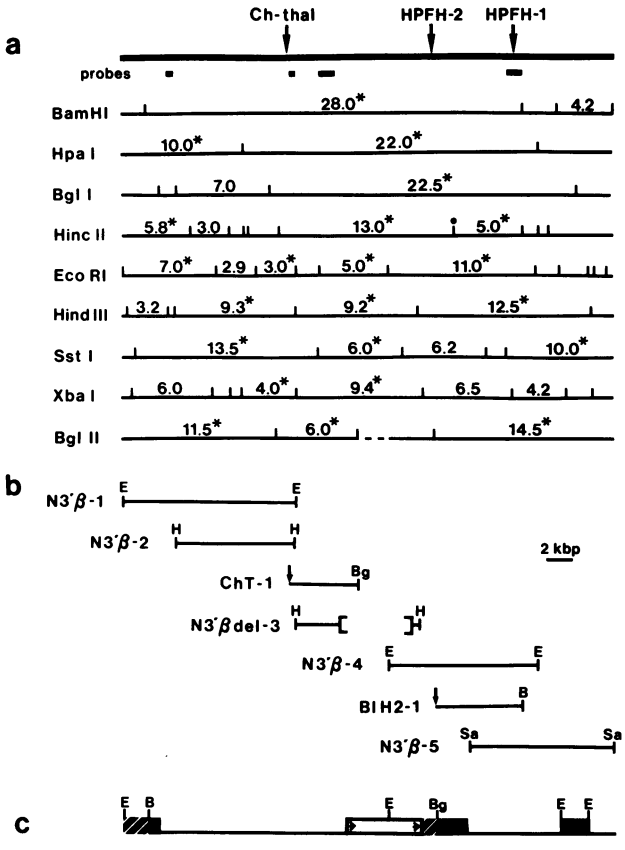

Figure 4. a) Restriction map of the normal DNA that encompasses the 3' endpoints of the Ch-thal, HPFH-2 and HPFH-1 deletions. Numbers are sizes of fragments in kbp and sizes smaller than 2.5 kbp are not shown. The HincII site that is not completely cleaved is indicated by a dot. The dashed line shows a region that is not cloned and in which possible BglII sites could not be mapped. The fragments marked by asterisks were mapped in genomic DNA using the probes shown as bars in the upper part of the figure. The left most probe is a 500 bp HindIII fragment isolated from phage N3'ß-1. The middle two probes are fragments 2 and 3 from ChT-1 (see Figure 1c). The rightmost probe is a 1.08 kbp SstI-BamHI fragment isolated from clone B1H2-1 and is the 3' HPFH-2 probe W used in Figure 3.

b) Bacteriophage clones encompassing the region shown in part a. For the deletion-spanning clones ChT-1 and B1H2-1, only the portion of the clone that is 3' to the deletion junction (marked by an arrow) is shown. The isolation of the phage clones is described in Materials and Methods. Sa, Sau 3A; all other enzyme abbreviations are as in Figure 1. Only the restriction sites that mark the ends of the clones are shown.

c) Regions of repetitive DNA. The dark and hatched boxes represent restriction fragments from the clones in part b that hybridize to Alu family or L1 family repeat probes respectively. The open box shows the location of the repetitive retrovirus-like element described previously (14). Horizontal arrows represent that element's long terminal repeats or LTRs.

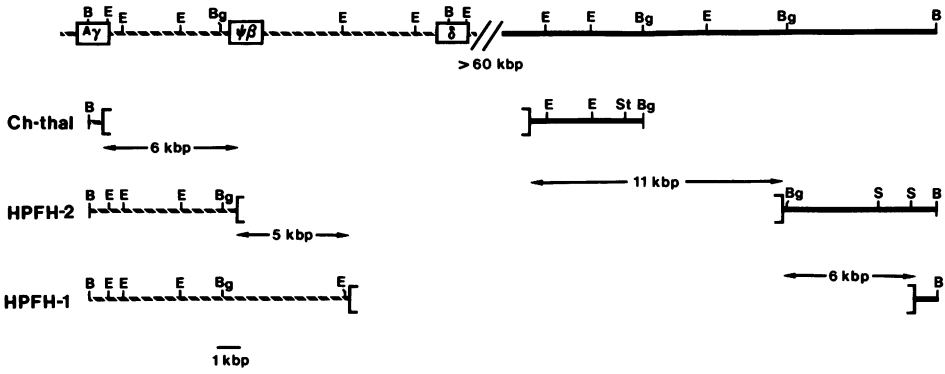

Figure 5. Restriction maps and alignment of the deletion junction regions of the Ch-thal, HPFH-2 and HPFH-1 deletions. The dashed portion of the top line shows the normal 5' DNA including the $^{A}\gamma$-, ψß- and δ-globin genes. The map of the normal 3' DNA (the solid portion of the top line) was derived from Figure 4a. The regions between the brackets represent the deleted DNA. The Ch-thal and HPFH-2 maps are derived from clones ChT-1 (Figure 1c) and B1H2-1(8, unpublished data) respectively. The map of HPFH-1 is from previous work (5, 8).

deletions, two deletions of unknown length, are relatively close to each other.

The Normal DNA Surrounding the 3' Breakpoints of the Ch-thal, HPFH-2 and HPFH-1 Deletions

Because the 3' Ch-thal and HPFH-2 probes hybridize to closely linked regions of DNA, they were used to isolate several overlapping phage clones that encompass the region of the 3' breakpoints of the Ch-thal, HPFH-2, and HPFH-1 deletions. The restriction map derived from these clones is shown in Figure 4a and the location of each clone is shown in Figure 4b. All of these clones were isolated from libraries of normal human DNA except for ChT-1 and B1H2-1 which are the deletion-spanning clones of the Ch-thal and HPFH-2 affected chromosomes respectively. B1H2-1 contains a 14 kbp BamHI fragment from an HPFH-2 homozygote (8). All restriction sites shown in Figure 4a have been mapped in cloned DNA and all fragments that have been sized by genomic mapping (marked by asterisks) agree with the map derived from the phage clones.

The locations of several different repeats in the region are shown in Figure 4c. The open box containing two arrows represents a repetitive retrovirus-like element, RTVL-H, found in the region (14). This element appears to cause some DNA cloning instability and is at least partially

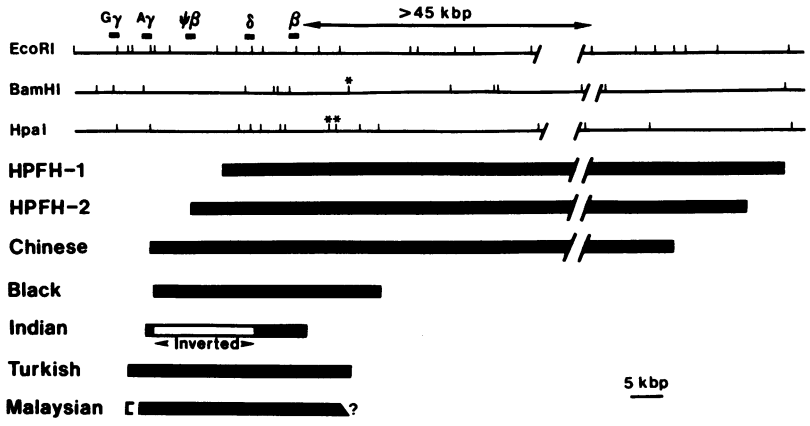

Figure 6. Extents of several deletions in the ß-globin gene cluster. The top portion, showing the EcoRI, BamHI and HpaI restriction maps of the ß-globin cluster and regions downstream was compiled from mapping either cloned or genomic DNA and is from published maps (1, 20), Henthorn et al., manuscript in preparation, unpublished results of F. Grosveld and Figure 4a. The regions between the slanted lines are gaps of unknown length that have not been cloned or mapped. The BamHI site immediately 5' to the gap in the BamHI restriction map was determined by genomic mapping and is located within the gaps in the EcoRI and HpaI maps. Sites marked with an asterisk are polymorphic (see ref. 1). The top three black bars show the minimum extent of the HPFH-1, HPFH-2 and Chinese $^{G}\gamma^{+}(^{A}\gamma\delta\beta)^{o}$ thalassemia deletions. The other bars show the extents of four different $^{G}\gamma^{+}(^{A}\gamma\delta\beta)^{o}$ thalassemia deletions. The references for these cases are: Black (26); Indian (27, 10); Turkish, (28, 5, Henthorn et al., manuscript in preparation); Malaysian, (29).

responsible for our inability to isolate the DNA between clones ChT-1 and N3'ß-4 (see Figure 4b). Clone N3'ßdel-3 was expected to contain this DNA on a 9.2 kbp HindIII fragment but was found to have undergone a deletion through homologous recombination during cloning between the long terminal repeats (LTRs) of the retrovirus-like element (unpublished data).

Alu family (25) and L1 (KpnI) family (21) repeats are also found in the cloned region and their locations are shown in Figure 4c. The locations of the two Alu repeats in the 3' portion of the map agree with those found by Tuan et al. (5) in the same region. The DNA immediately 3' to the retrovirus-like element hybridizes weakly to an L1 family probe (Figure 4c). In addition, DNA sequence analysis has revealed that the DNA immediately 5' to the RTVL-H element is also homologous to L1 family repetitive DNA but this homology is not sufficient to be detected in hybridization experiments (14).

Comparison of the Ch-thal Deletion to Other ß-Globin Gene Deletions

Earlier work demonstrated that the deletions associated with HPFH-1 and 2

have their 3' breakpoints approximately the same distance apart and in the same relative order as their 5' breakpoints (5, 8). This arrangement of breakpoints indicates that the two HPFH deletions differ in length by only 1 kbp. This is illustrated in Figure 5 which shows the regions surrounding the deletions associated with Ch-thal, HPFH-2 and HPFH-1. The 5' breakpoint of the Ch-thal deletion is 6 kbp to the left of the 5' breakpoint of HPFH-2 and the 3' breakpoints of the two deletions are 11 kbp apart in the same relative order. Thus the Ch-thal deletion is 5 kbp (11 kbp-6 kbp) shorter than the HPFH-2 deletion.

Although the absolute lengths of the Ch-thal, HPFH-2 and HPFH-1 deletions are not known, minimum estimates of their sizes can be made. By genomic mapping Tuan et al. (5) estimated the minimum size of the HPFH deletions to be 70 kbp. With the additional mapping data now available, we can revise this estimate. Figure 6 shows a restriction map and the minimum extents of the Ch-thal, HPFH-1 and HPFH-2 deletions. By adding distances of restriction fragments we have determined that the EcoRI site immediately 3' to the region of unknown length is greater than 45 kbp 3' to the ß-globin gene (Figure 6). This indicates that the two HPFH deletions are greater than 85 kbp in length and the Ch-thal deletion is at least 80 kbp long. Thus the Ch-thal deletion differs in length from the HPFH deletions by less than 6%.

The Ch-thal deletion can also be compared to other deletions associated with $^{G}\gamma^{+}(^{A}\gamma\delta ß)^{o}$ thalassemia. Figure 6 compares the extent of the Ch-thal deletion to the extents of four different deletions associated with a similar phenotype. Although the 5' breakpoints of all five deletions are relatively close to each other, localization of the 3' breakpoint of the Ch-thal deletion reveals that it is at least 40 kbp longer than the Black, Indian and Turkish $^{G}\gamma^{+}(^{A}\gamma\delta ß)^{o}$ thalassemia deletions.

DISCUSSION

Recombinational events in higher eukaryotes are poorly understood. In this study we have shown that the Chinese $^{G}\gamma^{+}(^{A}\gamma\delta ß)^{o}$ thalassemia deletion, like four other ß thalassemia deletions analyzed previously (8-10), was generated by a nonhomologous breakage and reunion event. The 5' breakpoint of the Ch-thal deletion is located 22 bp upstream of the "TG" repeat in the $^{A}\gamma$-globin gene. Stretches of alternating purine and pyrimidine residues have the potential to form Z-DNA (30) and there is some evidence that such regions may be hot spots for recombination (13, 31). Thus it is possible that the $(TG)_n$ sequence found near the 5' breakpoint of the Ch-thal deletion could have been involved in the recombination event.

The Ch-thal deletion junction is of interest because it contains a 36-41 bp segment of L1 (KpnI) family repetitive DNA. This segment is not derived from either the normal 5' or 3' sequences. There are several explanations for the occurrence of this apparent insertion. First, this segment could have been inserted as part of the deletion-forming event, possibly during the repair process. Second, this sequence could be the remnant of a larger L1 repeat that was inserted in the normal DNA at the 5' or 3' breakpoint prior to the deletion and then mostly removed during the generation of the Ch-thal deletion. Finally, it is possible that the Ch-thal deletion was formed by two smaller deletions each with an endpoint in the same region of an L1 family repeat located somewhere between the final 5' and 3' breakpoints. The 36-41 bp segment would represent the only undeleted portion of the L1 repeat. Consistent with this third explanation is the observation that the 3' breakpoints of at least four other ß-globin gene deletions are located in L1 family repetitive DNA (Henthorn et al., manuscript in preparation). The data presented here cannot distinguish between these possibilities but the first explanation, that this sequence was inserted at the time of the deletion, receives some support from the finding of other examples of deletions in the ß-globin gene cluster that appear to have sequences inserted at the deletion junctions (9, 10, Henthorn et al., manuscript in preparation). In fact, recombination junctions containing insertions of DNA have been found during the analysis of several types of DNA rearrangements (32-35). In most cases the origin of this "extra" DNA is unknown but these many examples suggest that the insertion of such segments during the rejoining of broken ends of DNA is not uncommon.

It was reported previously that four large deletions in the human ß-globin gene cluster, two associated with $(\gamma\delta\beta)^o$ thalassemia (8) and two with HPFH (5, 8), form different sets of similar-sized deletions. In this study we have shown that the deletion associated with Chinese $^G\gamma+(^A\gamma\delta\beta)^o$ thalassemia differs in size from the two deletions associated with HPFH by less than 6%. In addition, the 5' breakpoints of the three deletions are closely linked, as are the 3' breakpoints, and all the breakpoints are in the same relative orders (see Figures 5 and 6). Thus the Ch-thal deletion and the two HPFH deletions are grouped both in terms of their sizes and in the relative positions of their breakpoints. This finding is consistent with our previously described hypothesis (8) that some deletions may be close in size because they represent deletions of similarly-sized loops of DNA anchored to the nuclear matrix.

We find it intriguing that the 3' breakpoints of the Ch-thal, HPFH-2 and HPFH-1 deletions are located relatively close to an endogenous retrovirus-like element (Figure 4). This retrovirus-like sequence may facilitate nearby recombinational events, possibly by changing the chromatin structure of the region. Alternatively, the presence of the retrovirus-like element may indicate that this is a region that is particularly susceptible to breakages or integrations of foreign DNA. The possibility that deletion breakpoints may tend to occur in such susceptible regions is supported by our finding (Henthorn et al., manuscript in preparation) that the 3' breakpoint of a deletion associated with an Indian HPFH (36) occurs in a repetitive sequence that shows evidence of genomic instability. Localization and analysis of other deletion breakpoints is likely to provide further clues to the mechanisms involved in chromosomal rearrangements in mammalian cells.

Studies of the extents of deletions associated with HPFH or thalassemia have resulted in the formulation of two main hypotheses to explain the high expression of the fetal globin genes in adults (for review see 1). The first hypothesis initially proposed by Huisman (4) suggested that a region between the $^{A}\gamma$- and δ-globin genes regulates expression of the fetal γ-globin genes. This hypothesis assumes that deletions including this region remove important control sequences that normally cause the shutdown of the fetal genes in adults. The second hypothesis suggests that DNA sequences located far downstream of the normal β-globin gene cluster, but brought near to the fetal γ-globin genes by the deletions, are important in regulating the expression of the genes. This second hypothesis was suggested by Forget and others (1, 5) mainly because of the finding that the 3' breakpoints of the HPFH-1 and 2 deletions are only 6 kbp apart so that in both instances the same region of new DNA is brought close to the fetal genes. The supposition is that the presence of this new DNA in some way increases expression of the γ-globin genes.

With regard to these two hypotheses, the Chinese thalassemia deletion is of interest because it is similar in length to two deletions associated with the different phenotype of HPFH. However, although these three deletions are close in size, they are not very useful for distinguishing between the two models because of important differences in their 5' and 3' extents. The most obvious difference at the 5' end is that the Ch-thal deletion removes part of the Aγ-globin gene while the HPFH deletions leave both fetal genes intact and functional. This difference is the simplest explanation for the thalassemia phenotype associated with the Chinese deletion. However, at the 3' end the

retrovirus-like element is brought close to the γ-globin genes by the Ch-thal deletion but is deleted in HPFH-1 and 2. Thus the different 3' extents of the deletions may also play a role in determining phenotype.

The Ch-thal deletion can be grouped with four other $^{G}\gamma+(A\gamma\delta\beta)^{o}$ thalassemia deletions on the basis of associated phenotype and this group is more useful for testing the two hypotheses. All five of these deletions remove part or all of the Aγ-globin gene but leave the Gγ-globin gene intact (Figure 6). Trent et al. (29) compared the hematological data in heterozygotes affected with the Chinese, Indian, Turkish and Malaysian forms of $^{G}\gamma+(A\gamma\delta\beta)^{o}$ thalassemia and found them to be remarkably similar. The hematological findings in heterozygotes for the more recently described Black thalassemia deletion shown in Figure 6 are also very similar to the hematological findings in the other four cases (26). This phenotypic similarity coupled with the finding of different genomic maps of the 3' DNA led Trent et al. (29) to suggest that the similar phenotypes of the $^{G}\gamma+(A\gamma\delta\beta)^{o}$ thalassemias result from removal of the same critical sequences from between the Aγ- and δ-globin genes and not from the juxtaposition of new DNA on the 3' side. Figure 6 shows that the 3' extents of all but the Malaysian deletion have now been more precisely defined. The 3' breakpoint of the Chinese thalassemia deletion is more than 40 kbp from the 3' breakpoints of the Black, Indian and Turkish deletions. Thus these data support the first hypothesis that it is the 5' extent of the deletions that may be most important in determining the phenotype in these affected individuals.

The large variations in size of the $^{G}\gamma+(^{A}\gamma\delta\beta)^{o}$ thalassemia deletions shown in Figure 6 suggest that the new DNA brought into the β-globin gene cluster by these particular deletions may not affect the $^{G}\gamma$-globin gene expression. However, there is at least one deletion, a Dutch $(\gamma\delta\beta)^{o}$ thalassemia (8, 37), in which the features of the DNA brought into the cluster appear to affect β-globin gene expression. This deletion leaves the β-globin gene intact but inactive and there is evidence that the DNA normally found almost 100 kbp upstream and brought close to the β-globin gene by the deletion is in an inactive configuration (38). Thus there may be several ways in which different deletions can increase or decrease the expression of a nearby gene. Continued analysis of the β-globin gene deletions should lead to a better understanding of the variety of mechanisms that control gene expression in higher eukaryotes.

ACKNOWLEDGEMENTS

We thank Dr. D. Weatherall for providing DNA from a patient with Chinese thalassemia. We are grateful to Dr. J. Ross, Dr. R. Spritz and members of our laboratory for their comments on this manuscript. We also thank Paul Breyer and Natalie Borenstein for technical assistance and Frances Mann, Carol Jokinen and Sheryl Logan for typing. This work was supported by National Research Service Award GM09261 to D.M. and by grants GM20069 and AM20120 to O.S. from the National Institutes of Health. This is paper no. 2827 from the Laboratory of Genetics, University of Wisconsin-Madison.

* Current address: Department of Human Genetics, University of Pennsylvania Philadelphia, PA 19104

REFERENCES

1. Collins, F.S. and Weissman, S.M. (1984) Prog. Nucleic Acid Res. Molec. Biol. 31, 315-458.
2. Orkin, S.H., Antonarakis, S.E. and Kazazian, H.H. (1983) Prog. Hematol. 13, 49-73.
3. Weatherall, D.J. and Clegg, J.B. (1981) The Thalassaemia Syndromes, 3rd edn. Blackwell Sci., Oxford.
4. Huisman, T.H.J., Schroeder, W.A., Efremov, G.D., Duma, H., Mladenovskil, B., Hyman, C.B., Rachmilewitz, E.A., Bouver, N., Miller, A., Brodie, A., Shelton, J.R., Shelton, J.B. and Apell, G. (1974) Ann. N.Y. Acad. Sci. 232, 107-124.
5. Tuan, D., Feingold, E., Newman, M., Weissman, S.H. and Forget, B.G. (1983) Proc. Natl. Acad. Sci. USA 80, 6937-6941.
6. Flavell, R.A., Kooter, J.M., DeBoer, E., Little, P.F.R. and Williamson, R. (1978) Cell 15, 25-41.
7. Ojwang, P.J., Nakatsuji, T., Gardiner, M.B., Reese, A.L., Gilman, J.G. and Huisman, T.H.J. (1983) Hemoglobin 7, 115-123.
8. Vanin, E.F., Henthorn, P.S., Kioussis, D., Grosveld, F. and Smithies, O. (1983) Cell 35, 701-709.
9. Spritz, R.A. and Orkin, S.H. (1982) Nucleic Acids Res. 10, 8025-8029.
10. Jennings, M.W., Jones, R.W., Wood, W.G. and Weatherall, D.J. (1985) Nucleic Acids Res. 13, 2897-2906.
11. Mann, J.R., MacNeish, A.S., Bannister, D., Clegg, J.B., Wood, W.G. and Weatherall, D.J. (1972) Br. J. Haematol. 23, 393-402.
12. Jones, R.W., Old, J.M., Trent, R.J., Clegg, J.B. and Weatherall, D.J. (1981) Nucleic Acids Res. 9, 6813-6825.
13. Slightom, J.L., Blechl, A.E. and Smithies, O. (1980) Cell 21, 627-638.
14. Mager, D.L. and Henthorn, P.S. (1984) Proc. Natl. Acad. Sci. USA 81, 7510-7514.
15. Rimm, D.L. Horness, D., Kucera, J. and Blattner, F.R. (1980) Gene 12, 301-309.
16. Hohn, B. (1979) Methods Enzymol. 68, 299-309.
17. Leonen, W.A.M. and Blattner, F.R. (1983) Gene 26, 171-179.
18. Maxam, A.M. and Gilbert, W. (1980) Methods Enzymol. 65, 499-560.
19. Devereux, J., Haeberli, P. and Smithies, O. (1984) Nucleic Acids Res. 12, 387-395.
20. Grosveld, F.G., Lund, T., Murray, E.J., Mellor, A.L., Dahl, H.H.M. and Flavell, R.A. (1982) Nucleic Acids Res. 10, 6715-6732.

21. Singer, M.F. and Skowronski, J. (1985) Trends Biochem. Sci. 10, 119-122.
22. Bernards, R. and Flavell, R.A. (1980) Nucleic Acids Res. 8, 1521-1534.
23. Tuan, D., Murnane, M.J., deRiel, J.K. and Forget, B.G. (1980) Nature 285, 335-337.
24. Fritsch, E.F., Lawn, R.M. and Maniatis, T. (1979) Nature 279, 598-603.
25. Schmid, C.W. and Jelinek, W.R. (1982) Science 216, 1065-1070.
26. Henthorn, P.S., Smithies, O., Nakatsuji, T., Felice, A.E., Gardiner, M.B., Reese, A.L. and Huisman, T.H.J. (1985) Br. J. Haematol. 59, 343-356.
27. Jones, R.W., Old, J.M., Trent, R.J., Clegg, J.B. and Weatherall, D.J. (1981) Nature 291, 39-44.
28. Orkin, S.H., Alter, B.P. and Altay, C. (1979) J. Clin. Invest. 64, 866-869.
29. Trent, R.J., Jones, R.W., Clegg, J.B., Weatherall, D.J., Davidson, R. and Wood, W.G. (1984) Br. J. Haematol. 57, 279-289.
30. Wang, A.H.J., Quigley, G.J., Kolpak, F.J., Crawford, J.L., vanBoom, J.H., vander Marel, G. and Rich, A. (1979) Nature 287, 680-686.
31. Rogers, J. (1983) Nature 305, 101-102.
32. Botchan, M., Stringer, J., Mitchison, T. and Sambrook, J. (1980) Cell 20, 143-152.
33. Anderson, R.A., Kato, S. and Camerini-Otero, D. (1984) Proc. Natl. Acad. Sci. USA 81, 206-210.
34. Gerondakis, S., Cory, S. and Adams, J.M. (1984) Cell 36, 973-982.
35. Piccoli, S.P., Caimi, P.G. and Cole, M.D. (1984) Nature 310, 327-330.
36. Kutlar, A., Gardiner, M.B., Headlee, M.G., Reese, A.L., Cleek, M.P., Nagle, S., Sukumaran, P.K. and Huisman, T.H.J. (1984) Biochem. Genet. 22, 21-35.
37. Van der Ploeg, L.H.T., Konings, A., Oort, M., Roos, D., Bernini, L. and Flavell, R.A. (1980) Nature 283, 637-642.
38. Kioussis, D., Vanin, E., deLange, T., Flavell, R.A. and Grosveld, F.G. (1983) Nature 306, 662-666.