Volume 8 Number 24 1980 Nucleic Acids Research

A family of long reiterated DNA sequences, one copy of which is next to the human beta globin gene

J.W.Adams*, R.E.Kaufman*, P.J.Kretschmer+, M.Harrison+ and A.W.Nienhuis

*Clinical Hematology Branch, and +Laboratory of Molecular Hematology, National Heart, Lung and Blood Institute, National Institutes of Health, Bethesda, MD 20205, USA

Received 23 September 1980

ABSTRACT

An unusually long repeated DNA sequence was identified in cloned DNA, three kb 3' to the human β-globin gene. Other members of this repeated sequence family were isolated from a human genomic DNA library and characterized by Southern blotting techniques, electron microscopy, and solution hybridization. The copy located next to the β-globin gene was found to be 6.4 +/- 0.2 kb long and continuous over that length. This repeated sequence family comprises about 1% of the human genome and contains 3000-4800 copies of moderate sequence divergence which are interspersed with other less-highly repeated DNA. The 6.4 kb repeated unit does not appear to be composed of any smaller tandemly repeated subunits, nor is it expressed at a high level in bone marrow cell RNA.

INTRODUCTION

The genomes of eukaryotes are composed of 20-60% repetitive DNA which is divided into the highly repetitive (about 1,000,000 copies) and moderately repetitive (generally 100,000 copies or less) sequences (1). Moderately repetitive DNA exists in two different patterns termed short-period interspersion, or Xenopus-type, and long-period interspersion, or Drosophila-type (2). In short-period interspersion, which occurs in mammals, most vertebrates and sea urchins, the predominant repeated DNA sequences are about 300 base pairs (bp) long and interspersed with unique sequence DNA (3). A minority of the repetitive DNA sequences (6% in Xenopus) are longer than 1.5 kilobase (kb). These long repetitive DNA sequences show less divergence than the short DNA sequences (4) but may contain some components in common with the short repeats (5-9).

Analysis of the human genome indicates that most moderately repetitive DNA conforms to the short period interspersion pattern (10,11); short repetitive DNA sequences and inverted repeats are present which are transcribed into hnRNA (12-14). A highly repeated 300 bp family, 60% of the copies of which contain an identically positioned site for the restriction endonuclease, Alu

I, has been shown to be interspersed with unique sequence DNA (15,16). Also a different short tandemly arranged 340 bp repeated DNA family has been localized to the centromeres of human chromosomes and its copies are homologous to a repeated DNA family in monkeys (17-20). A family of longer repetitive DNA sequences which appear as 3.4 kb fragments after Eco RI, Hae III, or Eco RII digestion of DNA from human males has been localized exclusively to the Y chromosome. Copies of this family consist of three 800 bp repeated DNA sequences which are unique to male DNA separated by 250 bp units of moderately repetitive DNA sequences also found in female DNA (21).

Numerous studies of repetitive DNA families using solution hybridization techniques form the basis of present knowledge of these sequences but have so far failed to establish their significance. By using recombinant DNA and molecular cloning techniques, specific repetitive DNA sequences may be isolated and characterized and their relationship to specific structural genes defined. Recently, repetitive DNA sequences interspersed among the rabbit β-like globin genes have been characterized by this strategy. Shen and Maniatis (22) have studied five families of repetitive DNA sequences, individual copies of which are interspersed among four rabbit β-like globin genes as direct or inverted repeats. Each of the repeated DNA sequences are less than 0.5 kb except for one pair of 1.4 kb which brackets the entire four gene complex. In human DNA, a repeated family of DNA sequences has also been found among the β-like globin genes (23,24); individual copies are transcribed *in vitro* by RNA polymerase III to yield either a 515 or 575 bp RNA (24).

No examples of repeated DNA sequences longer than 1.5 kb have been identified next to specific mammalian genes and hence these longer repeat families remain less well characterized. Discovery of a repeated DNA sequence longer than 3.2 kb, one copy of which is located beginning 3 kb 3' to the human β globin gene (25) provided an opportunity to characterize this long repeated DNA family from the human genome. Hence we have determined the length, sequence homology, number of copies, and interspersion pattern of this family of long moderately repetitive DNA sequences.

MATERIALS AND METHODS

Materials Restriction endonucleases were obtained from Bethesda Research Laboratories and Boehringer Mannheim, [^{32}P]dCTP and [^{32}P]dGTP from Amersham Searle, and DNA polymerase I from Boehringer Mannheim. The genomic DNA used for Southern blots was obtained from patients with sickle cell anemia or β thalassemia. The DNA used to prepare the DNA fragment library was obtained

from a female patient homozygous for β^+ thalassemia (25). E. coli DP50 SupF, bacteriophage Charon 4A, and associated materials necessary for verification of EK2 status were obtained from F. Blattner and maintained as described (26). E. coli χ1776 was provided by R. Curtiss III (27). λHβG1 (28) was kindly provided by T. Maniatis.

Cloning of DNA Into Plasmids and Bacteriophage Methods for maintenance and growth of bacterial strains and for amplification and extraction of plasmid DNA from E. coli strains χ1776 and C600 and extraction of bacteriophage DNA from high titer CsCl stocks were as described (29). The human Charon 4A genomic library was prepared by the procedure described in detail by Maniatis et al. (30), using the in vitro packaging system of Blattner et al. (26), except that high molecular weight DNA, isolated from splenic tissue, was partially digested with Eco RI. Fragments of 15-20 kb in length were isolated by sucrose gradient centrifugation and ligated to the arms of Charon 4A as previously described (25,29). A library of 10^6 recombinant bacteriophage was amplified and then screened by the method of Benton and Davis (31). Specific DNA fragments were subcloned into pBR322 using standard techniques (32). All cloning experiments were conducted in compliance with the NIH Guidelines for Recombinant DNA Research as approved by the NIH Biohazard Committee under a Memorandum of Understanding and Agreement.

Restriction Endonuclease and Southern Blot Analysis Enzyme digestions were performed according to conditions specified by the vendor. Nick-translation was performed as previously described (33,34). Specific activities ranged from 10^7 to 10^8 cpm/μg. RNA and cDNA were prepared exactly as described (35,36). Southern blot analysis (37) was by standard techniques and the hybridization and washing conditions were exactly those of Jeffreys and Flavell (38).

Spot Hybridization DNA concentrations were measured by absorbance at 260 nm. Aliquots of the DNA samples in 10 mM Tris pH 7.5, 1 mM EDTA, were denatured by heating at 100°C for 10 minutes then applied in 5 μl to a nitrocellulose filter, allowed to dry thoroughly at room temperature and then baked at 80°C for 2 hours prior to hybridization under routine conditions (38).

Solution Hybridization These analyses were performed as described (35) in varying concentrations of formamide containing 0.5 M NaCl at 50°C with nick-translated DNA fragments as tracers. DNAs were sheared to 600-800 bp lengths by sonication. Formation of duplexes was measured by digestion with single-strand specific S_1 nuclease (39,35).

Electron Microscopic Heteroduplex Analysis Equal amounts of DNA from two bacteriophage (at a total final concentration of 5 μg/ml) were denatured and reannealed 45-60 minutes in 47.6% formamide exactly as previously described (40). Spreading conditions, electron microscopy, photography, and measurement of heteroduplex molecules will be described in detail elsewhere (41).

RESULTS

A Repeated DNA Sequence 3' to the Human Beta Globin Gene. We used Southern blot analysis to screen several DNA fragments from near the human δ and β globin genes for the presence of repeated sequences (Fig. 1). Two Pst I fragments and two Eco RI fragments were isolated from cloned DNA, nick-translated and used as probes against Pst I digested or Eco RI digested human DNA, respectively. The 3.1 kb Pst I fragment hybridized only to a single band in the Southern blot analysis, suggesting that it contained only unique sequence. The 2.1 kb Pst I fragment hybridized weakly to a continuum of bands, and the 3.6 kb Eco RI fragment hybridized intensely to a similar continuum, suggesting that these fragments contained repeated sequences. In contrast, the 3.2 kb Eco RI fragment from 3' to the β globin gene, hereafter referred to as the 3.2 kb β-3' Eco RI fragment, when annealed to Eco RI digested genomic DNA showed an intense continuum above 3.2 kb, an intense band at 3.2 kb, and a series of discrete bands below 3.2 kb. One of several possible interpretations of this pattern is that the 3.2 kb β-3' Eco RI fragment is contained within one copy of a family of long repeated DNA sequences which extends beyond the Eco RI sites in both directions. Because of this unusual pattern we decided to investigate this repeated sequence family more closely.

Isolation and Restriction Endonuclease Mapping of Different Members of the Long Repeated DNA Family. The 3.2 kb β-3' Eco RI fragment was used to screen a human genomic DNA library by the plaque hybridization method (31). A positive signal on radioautography was obtained for 2-3% of the plaques. Five of the more strongly hybridizing of these bacteriophage were isolated, and the DNA extracted and analyzed by restriction endonucleases and Southern blotting, using probes from the β-globin gene 3' flanking region (Fig. 2). Three of the clones, λJA64, 67, and 76, contain a 3.2 kb Eco RI fragment while λJA79 contains Eco RI fragments of 2.35 and .95 kb (which sum to 3.2 within the accuracy of our technique) all of which anneal to the radiolabeled 3.2 kb β-3' Eco RI fragment. λJA70 contains a larger Eco RI fragment which hybridizes to this probe. In addition, four of the five clones were found to

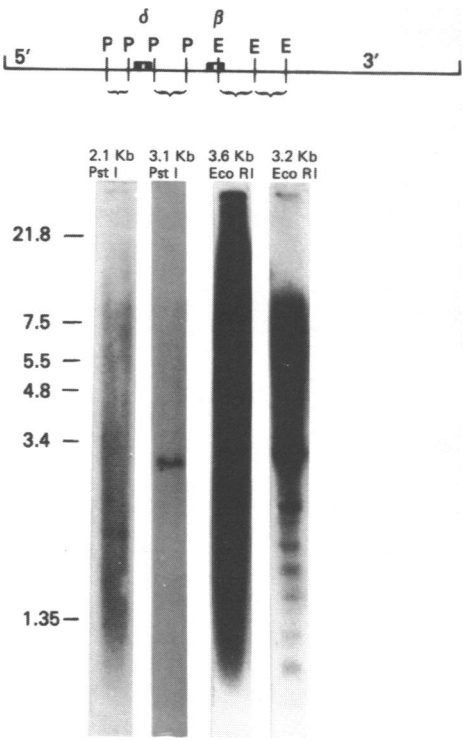

Figure 1. Southern blot analysis of human DNA. At the top are shown the origins of the fragments recovered from recombinant plasmids, nick-translated, and used as probes against Southern blots of human genomic DNA, 15 µg per lane. For the two Pst I fragment probes the genomic DNA was digested with Pst I, and for the two Eco RI probes, with Eco RI. Electrophoresis was in a 1% agarose gel. Specific activities of the probes varied between 10^7 - 10^8 cpm/µg while radioautography exposure varied from 1-5 days. Size markers are from Eco RI digestion of lambda DNA and Hae III digestion of Ø174 DNA. The two Pst I fragments were subcloned from λHβG1 (28), and the two Eco RI fragments were subcloned from λTβG41, a recombinant bacteriophage containing the 3' end of the β-globin gene and 17 kb of 3' flanking DNA (25).

contain a 1.4 kb Eco RI-Bam HI fragment which hybridized to the 1.4 kb Eco RI-Bam HI fragment which lies just 3' to the 3.2 kb β-3' Eco RI fragment in the β globin gene region. Restriction endonuclease maps were derived for each of the human DNA segments present in the five clones (Fig. 3). The 1.4 kb Bam HI-Eco RI fragment, mentioned above, was used to orient tentatively the five recombinants relative to the repeated sequence in the β globin gene 3' flanking region. These orientations were confirmed by electron microscopy (described

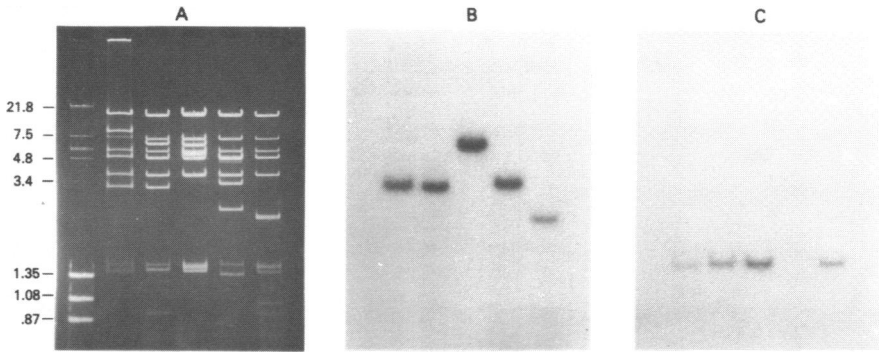

Figure 2. Four µg each of DNA from λJA64, 67, 70, 76, and 79 were digested with Eco RI and Bam HI, electrophoresed on 1% agarose, and stained with ethidium bromide (panel A). This gel, and a duplicate, were transferred by the Southern technique to nitrocellulose filters. These filters were annealed to different nick-translated probes, then washed and exposed to X-ray film to produce the autoradiographs shown. Panel B: 3.2 kb β-3' Eco RI fragment probe. Panel C: 1.4 kb Bam HI-Eco RI fragment probe from just 3' to the 3.2 kb β-3' Eco RI fragment in the β globin gene region (see map in Fig. 3). Molecular weight size markers (in kb) are indicated.

below). The fact that this 1.4 kb Bam HI-Eco RI fragment is present and adjacent to the sequences which hybridize to the 3.2 kb β-3' Eco RI probe in four of the five recombinants suggested that these individual copies of the repeated DNA sequence are at least 3.2 plus 1.4 or 4.6 kb long, as was confirmed by electron microscopy.

Length of the Individual Copies of the Repeated DNA Sequence. Study of DNA heteroduplexes using the electron microscope established the length of the copy of this repeated DNA sequence which is near the β globin gene. Note that DNA inserted into λJA64, 67, and 79, and λTβG41 are oriented with the 3' end of their copy of the repeated sequence (relative to the β globin gene region) next to the left (long) arm of the bacteriophage vector while the human DNA in λJA70 and 76 are in the opposite orientation (Fig. 3). A total of 17 heteroduplexes of 4 types were examined. A heteroduplex between λJA64 and λTβG41 (Figs. 4 and 5) shows a loop of single-stranded DNA (from λTβG41) next to the right arm of the vector, then a 5.94 $\pm$ 0.14 kb (95% conf. limit) duplex, followed by separation of the two DNA strands which come back together as the left arm of the vector. This heteroduplex demonstrates the 3' end of the homology of the copies of the repeated DNA sequences in these two clones. The heteroduplex between λJA64 and λJA67 (Fig. 5) shows two single-stranded DNA loops, one at either end of a 4.58 $\pm$ 0.04 kb duplex. This shows that

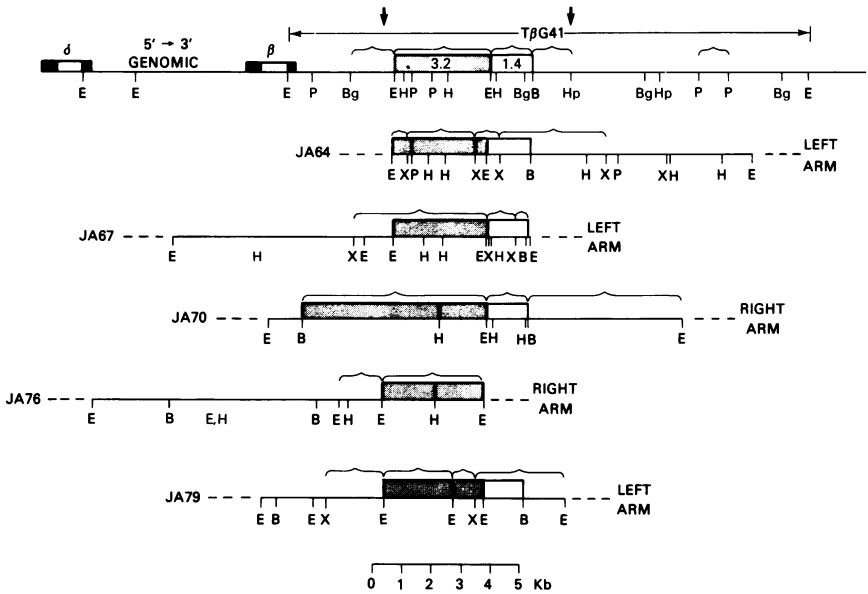

Figure 3. Restriction endonuclease maps of several recombinant bacteriophage which contain copies of the long repeated DNA sequence. At the top is a map of the 3' flanking DNA of the β globin gene, with restriction sites as previously published (25) except that the Bgl II and Hpa I sites within the 3.2 kb Eco RI fragment are not shown. The regions which are contained in λTβG41, the 3.2 kb β-3' Eco RI fragment, and the 1.4 kb Bam HI-Eco RI fragment, are indicated. Below are shown the restriction endonuclease maps for the clones λJA64, 67, 70, 76, and 79. The fragments which hybridized to the 3.2 kb β-3' Eco RI fragment probe and those which hybridized to the 1.4 kb Bam HI-Eco RI fragment are indicated by the shaded and the clear boxes respectively. The 5' and 3' borders of the 6.4 kb repeated sequence as determined by electron microscopy (see text below) are shown by heavy arrows. The fragments which hybridized to a nick-translated genomic DNA probe are indicated by brackets (see text and Fig. 6). The restriction sites are labelled as follows: E is Eco RI, X is Xba I, P is Pst I, H is Hind III, B is Bam HI, Hp is Hpa I, Bg is Bgl II. The Eco RI and Hind III sites in the 5' end of λJA76, and the Hind III sites in the 5' end of λJA67, could not be unambigously located and so are not precisely marked. λJA70 and 76 have not been mapped for Xba I, λJA79 has not been mapped for Hind III, and only λJA64 has been mapped for Pst I.

these copies of the repeated DNA sequence extend continuously over the 4.6 kb between the 5' end of the λJA64 insert and the 3' end of the λJA67 insert. The heteroduplex between λJA67 and λTβG41 (Fig. 5) shows the right arm, then a division of the strands which rejoin to form a 5.04 $\pm$ 0.12 kb duplex, then a loop of single-stranded DNA (from λTβG41) and finally the left arm of the

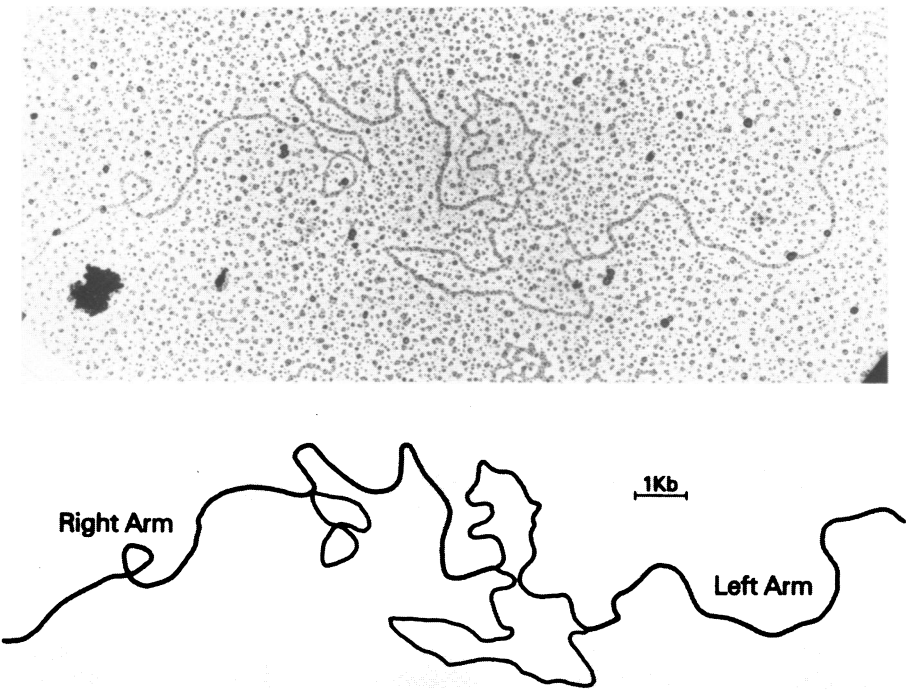

Figure 4. Heteroduplex between λJA64 and λTβG41. This heteroduplex shows the right arm of the vector, a loop-out of single-stranded DNA (from λTβG41), a 5.9 kb duplex region and then division of the two DNA strands which rejoin for the left arm. This molecule demonstrates the 3' end of these copies of the repeated DNA sequence. A circular ØX174 size marker (5,386 bp, double-stranded) is seen at the upper right. Magnification = 46,000 x.

bacteriophage vector. This demonstrates the position of the 5' end of the homology of the copies of the repeated sequence in these two clones. The total length of the copy in λTβG41 is therefore 5.94 $\pm$ 5.04 - 4.58, or 6.4 $\pm$ 0.2 kb (95% conf. limit). Figure 5 illustrates the heteroduplex between λJA70 and λJA76, which has a central duplex region of 3.6 kb, showing that the end of the homology of the copies of the repeated DNA sequence in these two clones is in the same location as that between λJA67 and λTβG41. We must emphasize that these data indicate the length of the member of this long repetitive DNA family located 3' to the β globin gene; other members of the family might be somewhat longer or shorter although the data in Fig. 1 supports an interpretation that most copies are at least 3.2 kb long. For convenience we refer subsequently to the family of 6.4 kb repetitive DNA sequences.

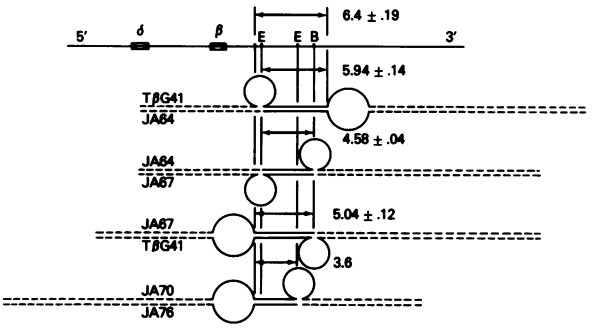

Figure 5. Length of the repeated DNA sequence. This diagram summarizes the heteroduplex data showing the location and size of the copy of the repeated DNA sequence located 3' to the human β globin gene. At the top is a diagram of the human δ and β globin genes and 3' flanking DNA. The Eco RI (E) and Bam HI (B) restriction endonuclease sites falling within the repeated sequence are shown. Below are shown diagrammatically the four heteroduplex experiments, arrayed so that the homologous DNA sequences are aligned vertically. The human DNA is drawn with solid lines and the bacteriophage arms with dashed lines. A total of 17 heteroduplex molecules were examined. The lengths of the central duplex regions and 95% confidence limits are given for the four experiments. (We do not have enough examples of JA70 vs 76 to compute confidence limits.) The total length of the copy of the repeated sequence next to the globin gene is given by (5.94 + 5.04) - 4.58 = 6.4 +/- 0.2 kb.

Interspersion Pattern of the Repeated Sequence. The restriction endonuclease maps and Southern hybridization patterns of the five recombinant DNAs suggest that these repeated sequences are not arrayed in tandem in the genome. To confirm this and to investigate the DNA which flanks the repeated sequences, we used nick-translated genomic DNA to probe Southern blots of our cloned DNAs. Recombinant plasmids which contain DNA 3' to the human β-globin gene were digested with Bgl II, Eco RI, Pst I, Bam HI, and Hpa I (Fig. 6), λJA64 and 67 were digested with Xba I, λJA70 and 76 with Eco RI and Bam HI (double digestions), and λJA79 with Eco RI and Xba I (double digestion) (data not shown). Four μg of restricted DNA was electrophoresed in each lane, and the resulting nitrocellulose filters were hybridized to 0.5 μg of nick-translated genomic DNA (sp. act. 1.6 x 10^7 cpm per μg) in a volume of 25 ml, and autoradiographed one week. The intensity of an autoradiographic band in this experiment should be roughly proportional to the prevalence in the genome of sequences homologous to that fragment (22), and fragments containing only unique sequence or repeated sequences of low reiteration number produce too

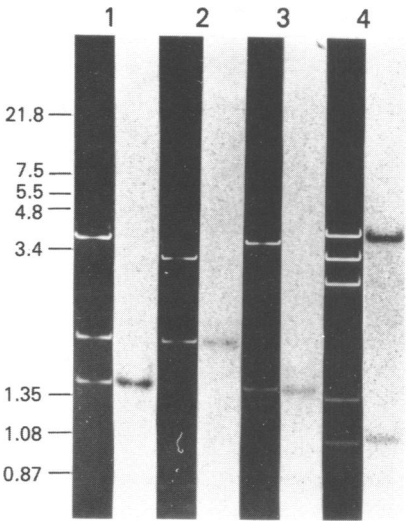

Figure 6. Repetitive sequence map 3' to the β-globin gene. Five μg each of recombinant pBR322 plasmids containing the DNA 3' to the β globin gene (25) were digested with restriction endonucleases, transferred to nitrocellulose, and allowed to anneal with nick-translated genomic DNA as described in the text. For each pair the left lane is the ethidium bromide stain of the agarose gel, and the right lane is the autoradiogram. 1) pRK12 (contains the 3.6 kb 3' β-globin gene Eco RI fragment) digested with Eco RI and Bgl II. 2) pRK11 (contains the 3.2 kb Eco RI fragment) digested with Eco RI and Pst I. Two fragments smaller than 1 kb were not retained by the nitrocellulose. 3) pRK19 (contains the 1.4 kb Eco RI-Bam HI fragment) digested with Eco RI and Bam HI. 4) pRK20 (contains the 9.5 kb Bam HI-Eco RI fragment) digested with Pst I and Hpa I. Molecular weight size markers are shown in kb. These results and the results of similar experiments with λJA64-79 (data not shown) are diagrammed in Fig. 3, where the fragments that hybridized to the probe are identified with brackets.

weak a signal to be detected under our conditions. Figures 3 and 6 illustrate that, except for a 1 kb Pst I fragment which lies 4.2 kb to the right of the 3' edge of the large repeated sequence in λTβG41, the only fragments in the recombinant DNAs which hybridized detectably to the genomic DNA probe are those fragments which contain part of the 6.4 kb repeated sequence. All other fragments hybridized at least 10 fold less intensely. Hence these individual copies of the 6.4 kb sequence neighbor other less highly repeated, or possibly unique, DNA sequences. The 1 kb Pst I fragment which hybridized strongly to the probe must contain a repeated sequence, which we have not yet characterized.

Internal Sequence Complexity of the 6.4 kb Repeated Sequence. Although without complete nucleotide sequence data it is not possible to rule out short sequences which are homologous to each other within the large repeated unit, our data show that the 6.4 kb unit is not reducible to homologous tandem subunits. The Southern blotting data show that neither the 3.2 kb Eco RI fragment nor the 1.4 kb Bam HI-Eco RI fragment, which make up 72% of the length of the repeated unit, hybridize either to each other nor to the fragments containing the other 28%. Further evidence that most of the 6.4 kb sequence occurs only once in the repeated unit is provided by the fact that no more than one type of heteroduplex molecule was seen for each hybridization experiment studied by electron microscopy. No intramolecular stem and loop structures were observed involving any two regions which were both within the 6.4 kb repeated unit.

Number and Homology of Copies in the 6.4 kb Family Solution hybridization studies were done with nick-translated 3.2 kb β-3' Eco RI fragments as tracer and fragmented total genomic DNA as driver. Under conditions of relatively high stringency (11°C below the Tm) reannealing occurs at a $Cot_{1/2}$ of about 0.25 - 0.4 (Fig.7). Since unique sequence DNA reanneals at a $Cot_{1/2}$ of about 1200 (42 and data not shown), we estimate that there are 3000 to 4800

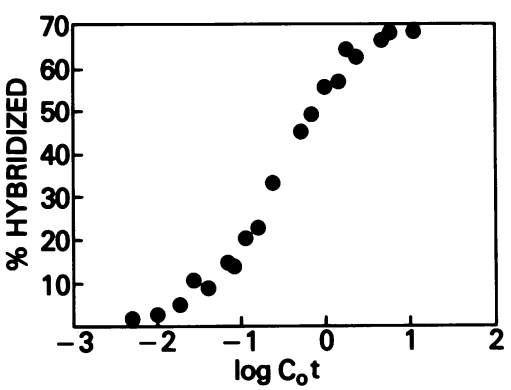

Figure 7. Analysis of the frequency of repetition of the 3.2 kb β-3' Eco RI sequence. Total human DNA, sonicated to yield 600-800 bp fragments, was incubated in capillaries at a concentration of 21 μg/ml with nick-translated 3.2 kb β-3' Eco RI tracer (400-800 bp fragments) in 40% formamide, 25 mM Hepes pH 7.0, 2.5 mM EDTA, 0.5 M NaCl at 50°C for times varying from two minutes to two days. Duplex formation was assayed by S_1 nuclease digestion and is plotted against the logarithm of Cot (concentration of nucleotides x time, in mole x sec/L). We estimate from this curve that the renaturation reaches 50% of endpoint at Cot = 0.25 - 0.4.

copies of the 6.4 kb repeated DNA sequence in the human genome.

A 6.4 kb DNA sequence repeated approximately 4000 times would be 1% of the 2.5×10^9 bp in the haploid human genome. This surprisingly high percentage is approximately confirmed by the spot hybridization study (Fig.8) where sequential dilutions of genomic DNA and the 3.2 kb β-3' Eco RI fragment were spotted onto a nitrocellulose filter and allowed to anneal to radiolabeled 3.2 kb β-3' Eco-RI fragment. The results suggest that the probe anneals to 1-2% of human genomic DNA.

Duplexes formed between the 3.2 kb β-3' Eco RI fragment (tracer) and genomic DNA (driver) melted in a slightly flattened curve with a Tm 3°C below that of perfectly matched duplexes of similar length, suggesting a moderate sequence mismatch between different copies of the 6.4 kb repeated DNA sequence family (Fig. 9).

Apparent Lack or Low Level of Expression of the 6.4 kb Family in RNA. A mixed radioactive cDNA was prepared from bone marrow poly-adenylated RNA, and annealed to a Southern blot of DNA from recombinants which contain a copy of the 6.4 kb family. No fragment which contained only DNA derived from a 6.4 kb repeat hybridized to this probe, though annealing to other fragments, and to the cloned β globin gene, was seen. When the 3.2 kb β-3' Eco RI fragment probe was annealed to a Northern transfer (43) of total bone marrow RNA onto aminobenzoxymethylcellulose paper, no discrete bands were observed. In solution hybridization studies which were sensitive enough to detect about 200 complete transcript molecules per cell (the sensitivity was limited by trace DNA contamination of the RNA), we were unable to detect RNA sequences complementary to the 3.2 kb β-3' Eco RI fragment probe in bone marrow cells. Hence we do not feel that all or most of the 6.4 kb repeated sequence is highly expressed into bone marrow cell RNA, though these studies

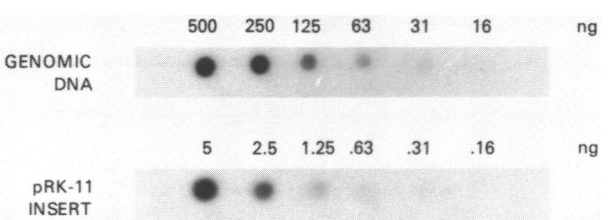

Figure 8. Spot hybridization analysis to determine reiteration frequency. Serial dilutions of heat-denatured genomic DNA and the 3.2 kb β-3' Eco RI fragment were applied, in 5 µl aliquots, containing 500 to 16 or 5 to 0.16 ng of DNA respectively, to a nitrocellulose filter and analyzed as described in the text. The fragment size of the nick-translated probe was 400-800 bp.

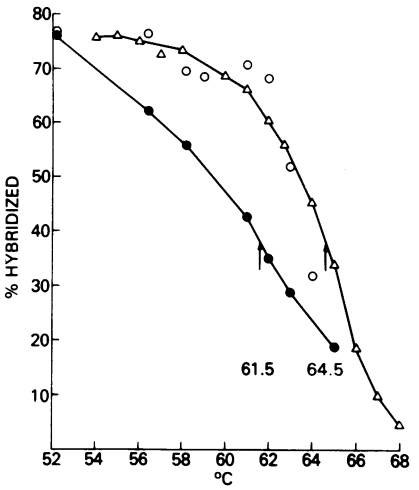

Figure 9. Thermal stability of duplexes formed between genomic DNA and the 3.2 kb β-3' Eco RI fragment. The melting curves of duplexes formed at 50°C are shown: reannealed tracer - 3.2 kb β-3' Eco RI fragment (0); genomic DNA and 3.2 kb β-3' Eco RI fragment (●); reannealed radiolabeled 3.1 kb Pst I fragment from between the δ and β globin genes in λHβG1 (Δ). All melt curves were done in 40% formamide and 0.5 M NaCl as previously described (35). From these curves we estimate the Tm for the genomic repeated sequences as 61.5°C and for the cloned DNA fragments as 64.5°C.

do not rule out low level expression or expression of a small part of individual copies of the 6.4 kb repeat family.

DISCUSSION

We have identified a repeated DNA sequence which begins 3 kb 3' to the human β-globin gene, is 6.4 kb in length, is reiterated 3000-4800 times in the human genome, and comprises about 1% of total human DNA. It is interspersed in the genome with other less highly repeated, or possibly unique, sequences. The 6.4 kb repeated unit does not appear to be highly repetitious internally, as determined by Southern blotting and inspection of heteroduplexes. Sequence polymorphism between the individual copies of this repeated sequence was revealed both by restriction endonuclease maps and thermal stability studies. Our estimate of the length of the repeated unit by electron microscopy is supported by restriction endonuclease mapping data and by Southern blotting experiments using a probe made from genomic DNA. Our estimate of the reiteration frequency by Cot curve analysis is also confirmed by filter

hybridization studies. This particular repeated sequence family is of interest because it is longer than the majority of human repeated sequences, which have an average length of 300 bp (10, 11).

This repeated sequence can be distinguished from the multigene families for histones, rRNA, tRNA, and 5S RNA (44-47) by its reiteration frequency, restriction endonuclease map, and lack of high-level expression into RNA. Preliminary data of ours show that DNAs from several primate species hybridize only slightly less intensely to the 3.2 kb β-3' Eco RI fragment probe than does human DNA under conditions where calf DNA failed to hybridize (Adams, J.W., Scott, A.F., Smith, K.D., unpublished data). Hence it appears that the origin of this repeated sequence family pre-dates the divergence of the primate species.

As to the function or genesis of this sequence we can make only vague hypotheses. The fact that it is not expressed into RNA, at least in bone marrow cells, at levels proportionate to its reiteration frequency, suggests that it does not code for a protein or major nuclear RNA in this tissue. However, there may be a low-level transcript which has some functional role, or there may be transcription in some other tissue. Alternatively this sequence may be a binding site for a chromosomal protein, or serve as a signal for chromosomal folding. As such it could conceivably have some role in the regulation of expression of the β-globin or other nearby genes. The interspersion of this sequence among other DNA is consistant with but not by itself supportive of such a role. Finally it is possible that this repeated sequence has no function relevant to the organism, but is carried in the genome in an essentially parasitic fashion (48).

The mechanism of sequence replication and distribution for this and other eukaryotic repeated sequences is unknown. Transposable elements have been described, though, in yeast (49), maize (50), and Drosophila (51, 52), as well as in prokaryotes, and possibly a repeated sequence such as the one we have identified may be able to facilitate its own duplication and transposition in human chromosomes. This could be tested by using a unique sequence DNA probe from next to one copy of this repeated sequence to screen by Southern blotting the DNA from humans of different racial backgrounds, looking for genomes in which that copy of the repeated sequence was missing or in a different location.

ACKNOWLEDGEMENT

We wish to thank Dr. Tom Maniatis for the recombinant λHβG1 and Mrs. Linda Garza for assistance in preparation of manuscript.

REFERENCES

1. Britten, R.J., and Kohne, D.E. (1968) Science 161, 529-540.
2. Crain, W.R., Eden, F.C., Pearson, W.R., Davidson, E.H., and Britten, R.J. (1976) Chromosoma 56, 309-326.
3. Davidson, E.H., Graham, D.E., Neufeld, B.R., Chamberlin, M.E., Amenson, C.S., Hough, R.R., and Britten, R.J. (1973) Cold Spring Harbor Symp. Quant. Biol. 38, 295-301.
4. Britten, R.J., Graham, D.E., Eden, F.C., Painchaud, D.M., and Davidson, E. H. (1976) J. Mol. Evol. 9, 1-23.
5. Eden, F.C., Graham, D.E., Davidson, E.H., and Britten, R. J. (1977) Nuc. Acid Res. 4, 1553-1567.
6. Chaudhari, N., and Craig, S.P. (1979) Proc. Natl. Acad. Sci. USA 76, 6101-6105.
7. Wu, J., Pearson, W.R., Posakony, J.W., and Bonner, J. (977) Proc. Natl. Acad. Sci. USA 74, 4382-4386.
8. Craig, S.P., Chaudhari, N. and Steinert, M. (1979) Biochim. Biophys. Acta 565, 33-50.
9. Chaudhari, N., and Craig, S.P. (1979) Biochim. Biophys. Acta 562, 438-452.
10. Schmid, C.W., and Deininger, P.L. (1975) Cell 6, 345-358.
11. Deininger, P.L., and Schmid, C.W. (1976) J. Mol. Biol. 106, 773-790.
12. Jelinek, W., Evan, R., Wilson, M., Salditt-Georgieff, M., and Darnell, J.E. (1978) Biochemistry 17, 2776-2783.
13. Robertson, H.D., Dickson, E., and Jelinek, W. (1977) J. Mol. Biol. 115, 571-589.
14. Jelinek, W.R. (1977) J. Mol. Biol. 115, 591-601.
15. Houck, C.M., Rinehart, F.P. and Schmid, C.W. (1979) J. Mol. Biol. 132, 289-306.
16. Jelinek, W.R., Toomey, T.P., Leinwand, L., Duncan, C.H., Biro, P.A., Choudary, P.V., Weissman, S.M., Rubin, C.M., Houck, C. M., Deininger P.L., Schmid, C.W. (1980) Proc. Natl. Acad. Sci. USA 77, 1398-1402.
17. Manuelidis, L. (1976) Nuc. Acids Res. 3, 3063-3076.
18. Manuelidis, L. (1978) Chromosoma 66, 1-21.
19. Manuelidis, L. (1978) Chromosoma 66, 23-32.
20. Manuelidis, L. and Wu, J.C. (1978) Nature 276, 92-94.
21. Kunkel, L.M., Smith, K.D., and Boyer, S.H. (1979) Biochemistry 18, 3343-3353.
22. Shen, C. J., and Maniatis, T. (1980) Cell 19, 379-391.
23. Fritsch, E.F., Lawn, R.M., Maniatis, T. (1980) Cell 19, 959-972.
24. Duncan, C., Biro, P.A., Choudary, P.V., Elder, J.T., Wang, R.R.C., Forget, B.G., DeRiel, J.K., and Weissman, S.M. (1979) Proc. Natl. Acad. Sci. USA 76, 5095-5099.
25. Kaufman, R.E., Kretschmer, P.J., Adams, J.W., Coon, H.C., Anderson, W.F., and Nienhuis, A.W. (1980) Proc. Natl. Acad. Sci. USA 77, 4229-4233.
26. Blattner, F.R., Blechl, A.E., Denniston-Thompson, K., Faber, H.E. Richards, J.E. Slighton, J.L., Tucker, P.W., and Smithies, O., (1978) Science 202, 1279-1284.
27. Curtiss, R. III, Pereira, D.A., Hsu, J.C., Hull, S.C., Clark, J.E., Maturin, L.J. Sr., Goldschmidt, R., Moody, R., Inoue, M., and Alexander, L. (1977) In Recombinant Molecules: Impact on Science and Society, Beers R.F., Jr. and Bassett, E.S., eds. (Raven Press, New York) 45-56.
28. Lawn, R.M., Fritsch, E.F., Parker, R.C., Blake, G., and Maniatis, T., (1978) Cell 15, 1157-1174.
29. Kretschmer, P.J., Kaufman, R.E., Coon, H.C., Chen, M., Geist, C.E., and Nienhuis, A.W. (1980) J. Biol. Chem. 255, 3204-3211.

30. Maniatis, T., Hardison, R.C., Lacey, E., Lauer, J., O'Connell, C., Quon, D., Sim, G.K. and Efstratiadis, A. (1978) Cell 15, 687-701.
31. Benton, W.F., and Davis R.W. (1977) Science 196, 180-182.
32. Ullrich, A., Shine, J., Chirgwin, J., Pictet, R., Tischer, E., Rutter, W. J. and Goodman, H.M. (1977) Science 196, 1313-1319.
33. Maniatis, T., Jeffrey, A., and Kleid, D.G., (1975) Proc. Natl. Acad. Sci., USA 72, 1184-1188.
34. Benz, E.J. Jr., Kretschmer, P.J., Geist, C.E., Kantor, J.A., Turner, P. A., and Nienhuis, A.W. (1979) J. Biol. Chem. 254, 6880-6888.
35. Benz, E.J., Jr., Geist, C.E., Steggles, A.W., Barker, J.E., and Nienhuis, A.W. (1977) J. Biol. Chem. 252, 1908-1916.
36. Young, N.S., Benz, E.J. Jr., Kantor, J.A., Kretschmer, P., and Nienhuis, A.W., (1978) Proc. Natl. Acad. Sci. USA 75, 5884-5888.
37. Southern, E.M. (1975) J. Mol. Biol. 98, 503-517.
38. Jeffreys, A.J. and Flavell, R.A. (1977) Cell 12, 429-439.
39. Housman, D., Skoultchi, A., and Forget, B.G. (1974) Ann. N. Y. Acad. Sci. 241, 280-289.
40. Davis, R.W., Simon, M., and Davidson, N. (1971) Meth. Enzymol. 21, 413-428.
41. Kretschmer, P.J., Coon, H.C., Kantor, J.A., Davis, A., Harrison, M., and Nienhuis, A.W. (manuscript submitted).
42. Deisseroth, A., Nienhuis, A.W., Turner, P., Velez, R., Anderson, W.F., Ruddle, F., Lawrence, J., Creagan, R., Kucherlapati, R. (1977) Cell 12 (1), 205-218.
43. Alwine, J.C., Kemp, D.J., Stark, G.R. (1977) Proc. Natl. Acad. Sci., USA 74, 5350-5354.
44. Wilson, M.C., and Melli, M. (1977) J. Mol. Biol. 110, 511-535.
45. Wellauer, P.K., and Dawid, I.B. (1979) J. Mol. Biol. 128, 289-303.
46. Wilson, G.N., Hollar, B.A., Waterson, J.R., and Schmickel, R.D. (1978) Proc. Natl. Acad. Sci. USA 75, 5367-5371.
47. Hatlen, L.E., and Attardi, G. (1970) J. Mol. Biol. 56, 535-554.
48. Doolittle, W.F., Sapienza, C. (1980) Nature, 284, 601-603.
49. Farabaugh, P.J., Fink, G.R. (1980) Nature, 286, 352-355.
50. Nevers, P. and Saedler, H. (1977) Nature 268, 109-115.
51. Young, M.W. (1979) Proc. Natl. Acad. Sci. USA 76, 6274-6278.
52. Potter, S.S., Brorein, W.J. Jr., Dunsmuir, P. and Rubin, G.M. (1979) Cell 17, 415-427.