Proc. Natl. Acad. Sci. USA
Vol. 81, pp. 7510–7514, December 1984
Genetics

Identification of a retrovirus-like repetitive element in human DNA

(human endogenous retrovirus/repetitive DNA/long terminal repeats/globin gene deletions)

DIXIE L. MAGER AND PAULA S. HENTHORN

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

Communicated by Oliver Smithies, August 6, 1984

ABSTRACT We describe a 5- to 6-kilobase-pair repetitive family in human DNA. One member of this family is linked to the β-globin gene cluster and is close to the 3′ breakpoints of three different naturally occurring deletions involving this gene cluster. Sequence analysis indicates that this element includes terminal direct repeats of 415 base pairs that exhibit the features of long terminal repeats (LTRs) of retroviruses. A potential histidine tRNA primer binding site occurs just 3′ to the 5′ direct repeat. This retrovirus-like element interrupts a member of the *Kpn* I family of repeated DNA and is bracketed by a 5-base-pair directly repeated sequence. When attempts are made to clone the element in bacteriophage, homologous recombination between the LTR-like sequences is very frequently observed. Copy number estimates by two methods indicate that the element is repeated 800–1000 times in the human genome. We term this *Homo sapiens* family of retrovirus-like elements having a histidine tRNA primer binding site the *hs*RTVL-H family.

Endogenous retroviral sequences have been the subject of considerable research over many years (1, 2), and their possible functions have been the cause of much speculation. Retroviral proviruses have been found in the normal DNA of a variety of species, but those found in chickens and mice are by far the best characterized (1, 3). Some endogenous proviruses can produce virus but others, such as the multicopy VL30 elements (4–6) and intracisternal A-type particles (IAPs) (7–11) in mice, have not been shown to produce infectious virions and hence are usually called retrovirus-like elements.

There have been many attempts to identify endogenous proviruses in human DNA (12). The approach used most recently to search for such sequences is that of DNA hybridization of known retroviral sequences to normal human DNA. Using this method, several laboratories have detected and cloned human retrovirus-like DNA (13–16). The best characterized of these elements are homologous to Moloney leukemia virus, contain long terminal repeats (LTRs), and are present in multiple copies, some of which are transcribed (17–19).

We report here the identification of a different human multicopy endogenous retrovirus-like element. Unlike the sequences previously described, this element was not detected by nucleic acid hybridization to viral probes. Rather it was discovered during attempts to clone and analyze a region of DNA involved in three naturally occurring deletions in the human β-globin gene cluster. These three deletions are those that cause hereditary persistence of fetal hemoglobin (HPFH) type 1 and type 2 (20, 21), and a particular type of thalassemia found in a Chinese individual, Ch-$^{G}\gamma^{+}(^{A}\gamma\delta\beta)^{0}$ thalassemia (22). The 5′ breakpoints of these three deletions all lie between the fetal $^{A}\gamma$ gene and the adult δ gene. Their 3′ breakpoints, although relatively close to each other, are located at least 60 kilobase pairs (kbp) downstream of the adult β-globin gene (refs. 20 and 21; unpublished data). (The terms 5′ and 3′ are assigned with respect to globin transcription.) It is in the region of the 3′ breakpoints of these deletions that we have found the retrovirus-like element.

MATERIALS AND METHODS

Sources. Human DNA was isolated from cultured normal human embryonic fibroblasts (563 from Robert DeMars). The *Kpn* I family probes, pCa4A-10 and pCa4A-12, represent two halves of a long *Kpn* I family member and were kindly provided by Maxine Singer.

Phage Cloning. The *Xba* I library was constructed by isolating the 9- to 18-kbp fragments of a complete *Xba* I digest of human DNA by preparative agarose gel electrophoresis and ligating these fragments into *Xba* I arms of DNA from λ phage Charon 35 (23). The ligated DNA was then packaged *in vitro* (24). The size-selected *Hin*dIII, *Eco*RI, and *Sst* I libraries mentioned in the text were constructed by digesting human DNA to completion with the appropriate restriction enzyme, isolating the 8- to 10-kbp *Hin*dIII fragments, the 4- to 6-kbp *Eco*RI fragments, and the 5- to 7-kbp *Sst* I fragments and ligating them into DNA of phages Charon 30 (25), Charon 3AΔ*lac* (26), or Charon 35, respectively. In all cases the phages packaged *in vitro* were screened without amplification.

Hybridizations. All probes (1–8 in Fig. 1) were DNA fragments isolated from plasmid subclones of the region. Unless otherwise noted, all hybridizations were done as previously described (21) under standard conditions: 68°C and at salt concentrations of 3× standard saline/citrate (1× is 0.15 M NaCl/0.015 M sodium citrate) for genomic DNA and 6× standard saline/citrate for cloned DNA. When probe 1 was used, a final post-hybridization wash at 68°C in 0.6× standard saline/citrate was done. For low-stringency hybridizations, the temperature was lowered to 45°C and post-hybridization washes were at successively higher temperatures with 6°C increments up to 63°C.

For the dot blotting experiments, human DNA fragments and linearized plasmid DNA containing the probe sequence were denatured for 10 min in 0.4 M NaOH. Serial dilutions of these DNAs in 1 M ammonium acetate/0.2 M NaOH/0.1 mg of tRNA per ml were then dotted onto nitrocellulose, using a 96-well microsample filtration manifold. Hybridizations were done under standard conditions.

DNA Sequencing and Analysis. DNA sequencing was by the Maxam–Gilbert method (27) with modifications (28). Most of the region shown in Fig. 3 was sequenced on both strands and the remainder (bases 10–55 and 370–415 of the 5′ sequence and bases 45–90 and 358–415 of the 3′ sequence) was sequenced three times on the same strand.

Abbreviations: bp, base pair(s); HPFH, hereditary persistence of fetal hemoglobin; *hs*, *Homo sapiens*; RTVL-H, retrovirus-like element with a histidine tRNA primer binding site; LTR, long terminal repeat.

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

RESULTS

A Repetitive Element Linked to the Human β-Globin Gene Cluster. We have been investigating a region of DNA at least 60 kbp 3′ to the adult human β-globin gene. This region has been shown to contain the 3′ breakpoints of two deletions of unknown size that cause HPFH (20, 21). Genomic mapping data have indicated that the 3′ endpoint of a third deletion causing Ch-$^{G}\gamma^{+}(^{A}\gamma\delta\beta)^{0}$ thalassemia (22) is close to the endpoints of the HPFH deletions (unpublished data). During attempts to confirm this relationship by DNA cloning experiments, we found that we could not obtain clones that completely spanned the region between the 3′ endpoint of the Chinese thalassemia deletion and the endpoints of the two other deletions. A restriction enzyme map of the region of interest, including the uncloned region, is shown in Fig. 1. Bacteriophage clones isolated from the region are also shown. During our cloning experiments and analysis of the region, we isolated several DNA fragments, indicated in Fig. 1 as probes 1–8, to use as probes. With the exception of probe 1, all fragments contained repetitive DNA that reduced their usefulness as unique probes for the region. However, the characterization of these repetitive probes led to the identification of the repetitive element, RTVL-H.

The initial indication of a repetitive sequence in the region was seen when Southern transfers of genomic DNA were hybridized to probe 3. Fig. 2 shows that prominent bands too strong for single-copy DNA were observed. The most striking of these was an *Eco*RI band of 950 bp. Bands as large as 5.8 kbp (*Hin*dIII) and 5.9 kbp (*Bgl* II) were also seen. No prominent bands greater than 6 kbp were evident. Hybridization of probe 4 to genomic DNA gave the same patterns as when probe 3 was used (data not shown). Probe 5, isolated from phage N32E14, unexpectedly hybridized to the phage ChT-1 (see Fig. 1), as did probe 7 (weakly) (data not shown). However, probes 3, 4, and 8 did not. Restriction enzyme mapping localized the hybridizing region in ChT-1 to a 500-bp *Stu* I/*Mst* II fragment (Fig. 1). DNA sequence analysis, presented below, shows that the regions corresponding to probe 5 and to the *Stu* I/*Mst* II fragment of ChT-1 form direct repeats of 415 bp (indicated by the arrows in Fig. 1).

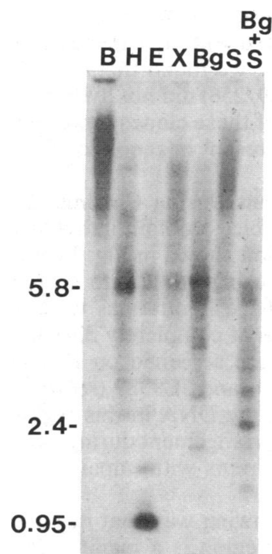

FIG. 2. Autoradiograph of a Southern transfer of normal human DNA digested with various restriction enzymes and hybridized to probe 3. The enzyme abbreviations above each lane are the same as in Fig. 1. After hybridization, the filters were washed in 3× standard saline/citrate for 3 hr at 68°C and exposed to x-ray film with an intensifying screen for 2.5 hr.

Mapping studies show that the direct repeats are 5.5 to 6.0 kbp apart. These results indicate that a repetitive element of close to 6 kbp is located in the region and that this element includes flanking direct repeats.

Other Copies of the Repetitive Element. To gain further evidence that the DNA on both sides of the uncloned region shown in Fig. 1 is part of the same repetitive unit, we screened a human library containing 9- to 18-kbp *Xba* I inserts in Charon 35 with probes 2, 3, and 6. This phage library was plated on a *recA*$^{-}$ derivative of *Escherichia coli* strain K802 and nitrocellulose replicas of each plate, containing a total of 2.4 × 10^{4} plaques, were hybridized to the three probes. The results showed several interesting correlations. First, we found that 137 phages hybridized to either probe 2 or 3, and 71 of these hybridized to both probes. This demonstrates that DNA sequences homologous to these two probes are part of a repetitive unit. Second, we found that all 137 phages hybridizing to probes 2 or 3 also hybridized to the probe for the direct repeat (probe 6), indicating that se-

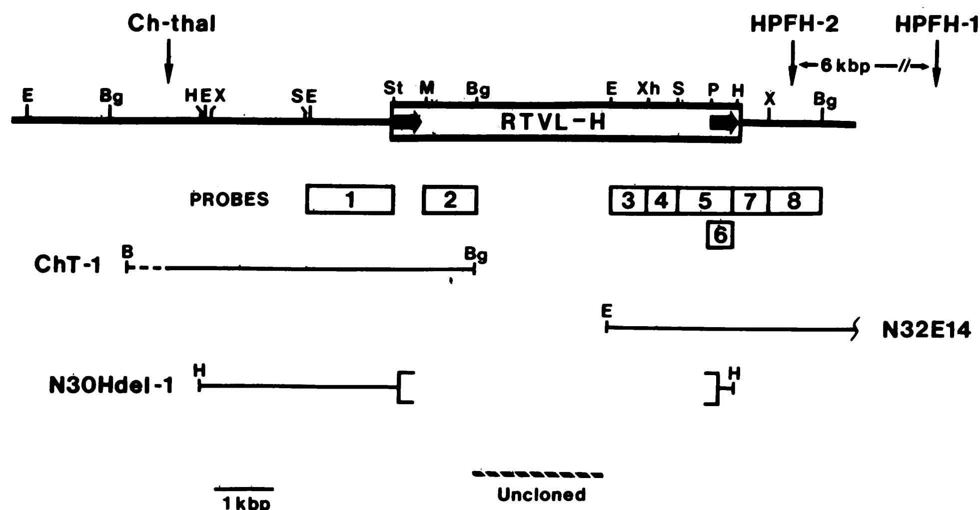

FIG. 1. Restriction enzyme map of a region of DNA ≥60 kbp 3′ to the human β-globin gene. The map was derived from genomic mapping and maps of phage clones. The insert of phage ChT-1 was isolated from an individual with Ch-$^{G}\gamma^{+}(^{A}\gamma\delta\beta)^{0}$ thalassemia and includes some DNA, shown by the broken line, from the 5′ side of the deletion junction in the $^{A}\gamma$ globin gene. The 11-kbp *Eco*RI insert of phage N32E14 was isolated from a normal individual. Phage N30Hdel-1 is described later in the text. The large open box and the horizontal arrows within it show the location of the repetitive element and its direct repeats described in the text. The small numbered boxes show the locations of sequences used to make probes 1–8. Fragments 1 and 2 were isolated from phage ChT-1 and fragments 3–8 were isolated from phage N32E14. The vertical arrows mark the locations of the 3′ breakpoints of the deletions associated with Ch-$^{G}\gamma^{+}(^{A}\gamma\delta\beta)^{0}$ thalassemia, HPFH-1, and HPFH-2 (refs. 20–22; unpublished data). B, *Bam*HI; Bg, *Bgl* II; E, *Eco*RI; H, *Hin*dIII; M, *Mst* II; P, *Pst* I; S, *Sst* I; St, *Stu* I; X, *Xba* I; Xh, *Xho* I. Only those *Mst* II, *Pst* I, and *Stu* I sites that are the ends of DNA fragments used as probes are shown.

quences from *between* the direct repeats illustrated in Fig. 1 (probes 2 and 3) are always associated with a sequence homologous to the direct repeat. Third, we found that 46% of the clones hybridizing to probe 6 (119/256) did not hybridize to either probe 2 or 3. Whether any of these clones represent solitary copies of the direct repeat present in genomic DNA remains to be determined.

Instability of the Repetitive Element During Cloning. As was mentioned above, we were unable to obtain individual clones that completely span the region shown in Fig. 1. Numerous cloning strategies included using probe 1 to screen partial *Sau*3A and partial *Eco*RI phage libraries as well as libraries enriched for particular sizes of completely digested DNA fragments. These libraries were screened on *E. coli* strains K802 (*recA*$^+$), DH-1 (*recA*$^-$), and JC9387 (*recBC*$^-$, *sbcB*$^-$) (29) without success. Thus, the DNA in this region may be susceptible to deletion or rearrangement during cloning experiments even in bacterial strains with impaired recombinational systems.

We were able to show in the following way that deletion between the direct repeats in this region is a major factor contributing to this failure to obtain unaltered clones. Using probe 1, we screened 1.2×10^6 phages from a Charon 30 library containing 8- to 10-kbp *Hin*dIII fragments of completely digested DNA for the 9.2-kbp *Hin*dIII fragment spanning the uncloned region (see Fig. 1). (This fragment had been sized by genomic mapping with probe 1.) We obtained three clones, typified by phage N30Hdel-1, that have insert lengths of 3.4 kbp. This observation suggests that the cloned 9.2-kbp genomic fragment has suffered a deletion of 5.8 kbp of DNA between the direct repeats during growth in *E. coli*. We have confirmed this interpretation by DNA sequence analysis of N30Hdel-1 (unpublished data). Deletions of DNA between the direct repeats also occur in clones of other copies of the repetitive element (unpublished observations). Once the direct repeats had been identified as factors in cloning the region we constructed libraries to screen for the 6-kbp *Sst* I and the 5-kbp *Eco*RI fragments from the region (Fig. 1) on the assumption that they might be easier to clone because they do not include both direct repeats. However, we were unable to isolate either fragment.

Nucleotide Sequences of the Direct Repeats. The sequences of the direct repeats shown in Fig. 1 are presented in Fig. 3*a*. The 5′ and 3′ copies of the direct repeats are 95% identical over 415 bp. (5′ and 3′ are assigned relative to the direction of globin transcription.) These direct repeats contain several features characteristic of LTRs of retroviruses (reviewed in ref. 31) and we will subsequently refer to them as LTRs. The putative regulatory sequences T-A-T-A-A-A and A-A-T-A-A-A are found in the 3′ half of most LTRs and are found in our LTRs at positions 279 and 351 (underlined in Fig. 3*a*). In mammalian LTRs, the poly(A) addition site (C-A) is typically located 15–21 bp 3′ to the sequence A-A-T-A-A-A (31, 32). In our LTRs the dinucleotide C-A is found twice within this distance at positions 372 and 374 (Fig. 3*a*). All integrated LTRs of retroviruses known to be still infectious that have been sequenced begin with T-G and end with C-A, with these nucleotides being part of a short inverted repeat. Our repeats begin with T-G but end with A-A. These dinucleotides are part of a pair of 5-bp inverted repeats with one mismatch. Deviation from the "T-G–C-A" rule applicable to infectious retroviruses has also been seen in the LTRs of a previously described class of human retrovirus-like sequences (19). Several murine retroviruses contain a duplicated sequence in the 5′ half (within U3) of the LTR (32). Our LTRs contain repeats of 31 bp (29/31 matches) in the 5′ portion (overlined in Fig. 3*a*). There are also two different 10-bp sequences, present in two copies.

Several features of the sequences flanking the LTRs are characteristic of integrated proviruses and these are shown in Fig. 3*b*. When retroviruses integrate, a host cell sequence of 4–6 bp is duplicated and flanks the integrated provirus. We observe the sequence C-T-G-A-A just 5′ to the 5′ copy of the LTR and just 3′ to the 3′ copy (Fig. 3*b*). An unintegrated LTR begins with A-A and ends with T-T (31). During integration, the A-A preceding the 5′ LTR and the T-T following the 3′ LTR are lost. This leaves an unpaired A-A in front of the 3′ copy and an unpaired T-T following the 5′ copy. These dinucleotides are seen in the expected positions in our sequence (Fig. 3*b*). For reasons that become apparent below, we feel that the A-A preceding the 5′ copy of the LTR is part of the cellular direct repeat, C-T-G-A-A, and not part of the LTR.

One of the most distinctive features of retroviruses is the occurrence just outside the 3′ boundary of the 5′ LTR (after the T-T) of a sequence, usually 18 bp in length, that is complementary to the 3′ end of a tRNA. This sequence binds to the 3′ end of a specific tRNA which then serves as a primer for the synthesis of the first (minus) strand of viral DNA. At the expected position in our sequence we find an 18-bp sequence that is complementary (17/18 matches) to the 3′ end of a human histidine tRNA (Fib. 3*b*). Another feature of retroviruses is a short (10–20 nucleotides) polypurine tract that occurs just 5′ to the 3′ LTR and appears to serve as a primer binding site for synthesis of plus-strand DNA. In our case, we have a polypurine stretch of 13 nucleotides that is interrupted by a single C (Fig. 3*b*).

These numerous sequence features strongly suggest that the 5- to 6-kbp repetitive unit described here is an endogenous *Homo sapiens* retrovirus-like (*hs*RTVL) element. In view of its histidine tRNA binding site, we propose to designate it as *hs*RTVL-H or as RTVL-H when the human context is self-evident.

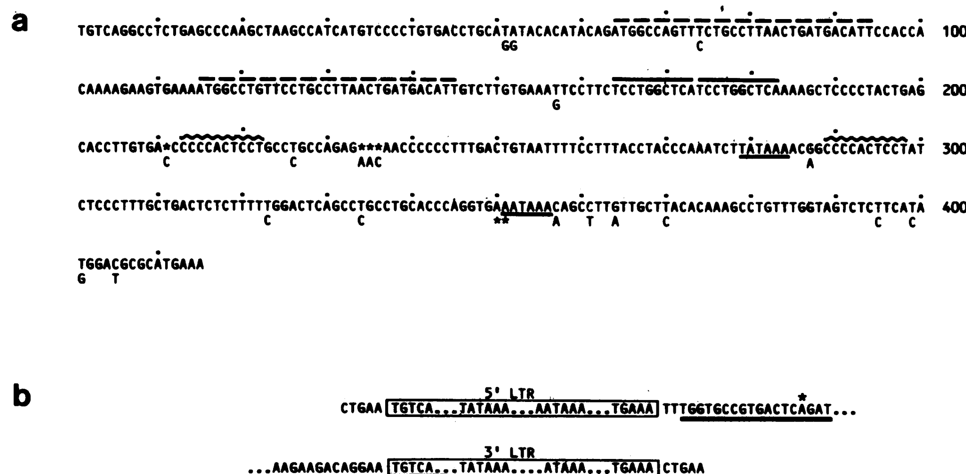

FIG. 3. (*a*) Nucleotide sequence of the direct repeats shown in Fig. 1. The entire sequence of the 5′ repeat is shown on the top line. Dots indicate every 10th nucleotide. Positions where the 3′ repeat differs are shown below each line. Gaps introduced for alignment are indicated with asterisks. The solid, broken, and wavy overlines show the three short directly repeated sequences within the LTR. Two possible control signals are underlined. (*b*) Retrovirus-like features associated with the direct repeats. The boxes represent the direct repeats. The underlined 18-bp sequence is complementary to the 3′ end of a human histidine tRNA (30). The one mismatch with the tRNA is marked with an asterisk.

The RTVL-H Element Interrupts a *Kpn* I Repeat. In the course of analyzing the DNA around the deletion breakpoint of HPFH-2 (Fig. 1), we observed that the DNA 3′ to the RTVL-H element shown in Fig. 1 is part of the *Kpn* I family of repetitive DNA (unpublished data). *Kpn* I family members are heterogeneous in length and occur on the order of 10^4 times throughout the human genome (reviewed in ref. 33). Full-length *Kpn* I repeats in humans contain a conserved *Hin*dIII fragment of 1.9 kbp (34). When the sequence of one of these repeated *Hin*dIII fragments (35) is compared to sequences around the RTVL-H element a striking result is seen (Fig. 4). Although there is considerable sequence divergence, homology between the *Kpn* I repeat sequence and sequences on both sides of the RTVL-H element is evident from the broken line extending diagonally through the plot. This homology is abruptly interrupted by the retrovirus-like element at position 988 of the 1.9-kbp *Hin*dIII repeat. The pentanucleotide C-T-G-A-A that is repeated on either side of the RTVL-H element is present as a single copy in the *Kpn* I repeat sequence at the point of interruption. This observation is further evidence that the sequence C-T-G-A-A is the genomic sequence that was duplicated upon integration of the RTVL-H element.

The possible association of other RTVL-H elements with *Kpn* I family members was checked by screening phages from the Charon 35 *Xba* I library with both the RTVL-H LTR probe (probe 6) and *Kpn* I family probes. Under our hybridization conditions, 9% (18/196) of the LTR-containing phages hybridized to *Kpn* I family probes, but approximately 15% of all phages hybridized to the same *Kpn* I family probes. Thus, the 9% level is probably due to chance associations between the two repeat families. Because the *Kpn* I family sequences adjacent to the RTVL-H element may be somewhat divergent from the *Kpn* I family probes used, we also hybridized probes 1 and 8 (the *Kpn* I family sequences flanking the RTVL-H element) at low stringency to over 50 clones containing RTVL-H LTR sequences as judged by hybridization to probe 6. No evidence was found that the RTVL-H LTRs are generally linked to these particular *Kpn* I family sequences.

Although a portion of fragment 1 is homologous to a *Kpn* I family sequence, this homology does not prevent the use of this fragment as a single-copy probe when a stringent post-hybridization wash is employed (see *Materials and Methods*).

Number of RTVL-H Elements in the Human Genome. We estimated the copy number of the RTVL-H element by two methods. Dot blotting experiments compared the hybridization of probe 6 (LTR) or probe 3 (interior sequence) with different amounts of total human DNA to their hybridization with different amounts of the DNA sequences corresponding to the probes. The copy numbers obtained from two separate dot blotting experiments were 1050 and 1300 for interior sequences (probe 3) and 2600 and 3000 for the LTR sequence (probe 6). We also estimated copy number by measuring the fraction of phage plaques of the human *Xba* I library that hybridized to the two probes and calculating from this the number that would hybridize in 2×10^5 phages (the number of phages with an average insert of 14 kbp needed to cover the whole genome of 2.9×10^9 bp). Probe 3 hybridized to 0.43% of the phages and probe 6 hybridized to 1.06%. These results indicate a copy number of 860 for interior sequences and 2100 for LTR sequences. The plaque hybridization estimate for RTVL-H LTR sequences could be low because all positive phages were scored as containing one LTR even though many might have contained two. The plaque hybridization estimate for interior sequences might also be low due to the tendency of RTVL-H elements to delete between LTRs in phage. However, because the plaque hybridization estimate agrees reasonably well with the dot blotting estimate we probably have not missed a large proportion of elements by using this *unamplified* phage library.

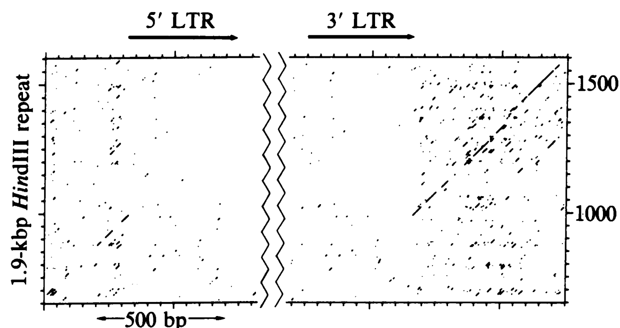

FIG. 4. Comparison of a *Kpn* I repeat sequence to the sequences surrounding and including the LTRs. The sequences were compared by a dot matrix program (36), and each dot represents 13 or more identities per 20 nucleotides. The vertical axis corresponds to a portion of the human 1.9-kbp *Hin*dIII repeat (a *Kpn* I family member) (35) and the numbering is as in ref. 35. The horizontal axis corresponds to the sequences of both LTRs, indicated by horizontal arrows, and their flanking sequences. Approximately 5 kbp of DNA sequence between the LTRs is not shown.

DISCUSSION

We have described a 5- to 6-kbp repetitive sequence in human DNA that has several properties of a retrovirus-like element. This element has terminal LTR-like direct repeats and contains a polypurine tract and a potential histidine tRNA primer binding site in the expected positions. The possible use of histidine tRNA as a primer has not been reported previously, as other retroviruses and retrovirus-like elements have primer binding sites that are homologous to proline, tryptophan, lysine, phenylalanine, glycine, or glutamic acid tRNAs (8, 10, 11, 19, 32). For this reason we have designated the element as RTVL-H, H representing histidine. This designation distinguishes the family from the human retrovirus-like family described previously (19).

We have screened 25 unrelated humans (16 White, 6 Black, and 3 Oriental) by genomic mapping for the presence of the element in the position shown in Fig. 1, and all 50 of their chromosomes 11 contained the RTVL-H element in that location. Thus, this copy appears to have integrated into the genome in its present location prior to the divergence of human races. This conclusion is consistent with the finding of 4.5% differences (19/415) between the 5′ and 3′ LTR. If the LTRs have diverged at a rate similar to that of noncoding DNA [estimated at 1% differences per 2 million years (38)], then this RTVL-H element may have integrated approximately 9 million years ago.

The retrovirus-like element described here is unstable in phage when propagated in a variety of strains of *E. coli*. This is perhaps not unexpected because difficulty in cloning some other retroviral sequences in phage because of homologous recombination between the LTRs has been reported (39–41). Deletions due to recombination between LTRs have not been reported when proviruses are grown in plasmid vectors on $recA^-$ hosts (42). Preliminary efforts to use plasmid vectors to obtain stable copies of the RTVL-H element have thus far been unsuccessful (unpublished observations).

We performed computer searches for homology between the RTVL-H LTR sequence and LTRs of known retroviruses, but we found no strong homology. This search included all the LTR sequences in the GenBank and European Molecular Biology Laboratory sequence libraries as of April 1984 as well as others more recently published. Among those examined were the LTRs of human T-cell leukemia virus type I (43) and type II (44) and the multicopy mouse intracisternal

A-type particles (IAPs) (9–11) and VL30 elements (5, 6). We also found no homology between the RTVL-H LTR and the LTRs of the human multicopy retrovirus-like elements described by Martin and co-workers (19) or the human low-copy element described by Bonner *et al.* (14). The 31-bp repeat within the RTVL-H LTR (Fig. 3*a*) was also compared specifically to the 60- to 100-bp repeats found in many mouse LTRs (31, 45). The longest stretch of identity found was between the 31-bp repeat and the Moloney sarcoma virus 72-bp repeat (45). They share the sequence G-T-T-C-C-T-G-C-C, found in positions 71–79 and 122–130 in the RTVL-H LTR. Although the significance of this small region of homology is difficult to assess, it is interesting that this 9-bp sequence occurs just 3′ to the "core" enhancer sequence, T-G-G-T-A-A-G, in the Moloney sarcoma virus 72-bp repeat (37).

Interestingly, both the RTVL-H element (Fig. 2) and the human endogenous retrovirus described by Martin's group (13, 17) contain an internal conserved *Eco*RI fragment of about 1 kbp. This similarity, however, is probably coincidental, because hybridization analysis has indicated that the internal sequences of these two classes of elements are not closely related (unpublished data; M. A. Martin, personal communication).

The RTVL-H element near the β-globin gene cluster that we analyzed here is positioned within a *Kpn* I repeat. We have also noticed that the 3′ flanking DNA of one of the human proviral elements described by Martin's group (clone 4-1) is part of a *Kpn* I repeat as judged by sequence data presented by that group (19). Martin and co-workers have shown that many of their retrovirus-like elements have flanking DNAs that cross-hybridize with each other. These three observations led us to analyze different clones containing RTVL-H sequences for a generalized association of the RTVL-H elements with *Kpn* I repeats. No such association was detectable by hybridization experiments.

The detection of different repetitive retrovirus-related sequences by hybridization analysis has led to the suggestion that several classes of such sequences may exist in human DNA (16, 19). The identification of the *hs*RTVL-H family supports this suggestion and raises the distinct possibility that other families of retrovirus-like elements, not necessarily detectable by hybridization to available viral probes, may exist in humans.

We are especially grateful to Dr. Oliver Smithies for invaluable advice and for the opportunity to work in his laboratory. We thank Dr. Howard Temin for helpful discussions, Dr. Maxine Singer for the *Kpn* I family probes, Dr. David Weatherall for providing DNA from a Chinese thalassemic individual, and Dr. Malcolm Martin for help in determining the relatedness between the retrovirus described by his group and the element described here. We also thank Paul Breyer and Natalie Borenstein for expert technical assistance and Frances Mann and Carol Jokinen for typing. This work was supported by National Research Service Award GM09261 to D.L.M. and by Grants GM20069 and AM20120 to Oliver Smithies from the National Institutes of Health. This is paper no. 2765 from the Laboratory of Genetics, University of Wisconsin–Madison.

1. Coffin, J. (1982) in *Molecular Biology of Tumor Viruses: RNA Tumor Viruses*, eds. Weiss, R. A., Teich, N. M., Varmus, H. & Coffin, J. M. (Cold Spring Harbor Laboratory, Cold Spring Harbor, NY), 2nd Ed., pp. 1109–1203.
2. Temin, H. M. & Engels, W. (1984) in *Evolutionary Theory: Paths Into The Future*, ed. Pollard, J. W. (Wiley, New York), pp. 173–201.
3. Teich, N. (1982) in *Molecular Biology of Tumor Viruses: RNA Tumor Viruses*, eds. Weiss, R. A., Teich, N. M., Varmus, H. & Coffin, J. M. (Cold Spring Harbor Laboratory, Cold Spring Harbor, NY), 2nd Ed., pp. 25–207.
4. Keshet, E., Shaul, Y., Kaminchik, J. & Aviv, H. (1980) *Cell* **20**, 431–439.
5. Itin, A. & Keshet, E. (1983) *J. Virol.* **47**, 656–659.
6. Hodgson, C. P., Elder, P. K., Ono, T., Foster, D. N. & Getz, M. J. (1983) *Mol. Cell. Biol.* **3**, 2221–2231.
7. Leuders, K. K. & Kuff, E. L. (1977) *Cell* **12**, 963–972.
8. Ono, M., Cole, M. D., White, A. T. & Huang, R. C. (1980) *Cell* **21**, 465–473.
9. Kuff, E. L., Feenstra, A., Lueders, K., Smith, L., Hawley, R., Hozumi, N. & Shulman, M. (1983) *Proc. Natl. Acad. Sci. USA* **80**, 1992–1996.
10. Ono, M. & Ohishi, H. (1983) *Nucleic Acids Res.* **11**, 7169–7179.
11. Ono, M., Kitasato, H., Ohishi, H. & Motobayashi-Nakajima, Y. (1984) *J. Virol.* **50**, 352–358.
12. Weiss, R. (1982) in *Molecular Biology of Tumor Viruses: RNA Tumor Viruses*, eds. Weiss, R. A., Teich, N. M., Varmus, H. & Coffin, J. M. (Cold Spring Harbor Laboratory, Cold Spring Harbor, NY), 2nd Ed., pp. 1205–1281.
13. Martin, M. A., Bryan, T., Rasheed, S. & Khan, A. S. (1981) *Proc. Natl. Acad. Sci. USA* **78**, 4892–4896.
14. Bonner, T. I., O'Connell, C. & Cohen, M. (1982) *Proc. Natl. Acad. Sci. USA* **79**, 4709–4713.
15. Callahan, R., Drohan, W., Tronick, S. & Schlom, J. (1982) *Proc. Natl. Acad. Sci. USA* **79**, 5503–5507.
16. Noda, M., Kurihara, M. & Takano, T. (1982) *Nucleic Acids Res.* **10**, 2865–2878.
17. Repaske, R., O'Neill, R. R., Steele, P. E., and Martin, M. A. (1983) *Proc. Natl. Acad. Sci. USA* **80**, 678–682.
18. Rabson, A. B., Steele, P. E., Garon, C. F. & Martin, M. A. (1983) *Nature (London)* **306**, 604–607.
19. Steele, P. E., Rabson, A. B., Bryan, T. & Martin, M. A. (1984) *Science* **225**, 943–947.
20. Tuan, D., Feingold, E., Newman, M., Weissman, S. M. & Forget, B. G. (1983) *Proc. Natl. Acad. Sci. USA* **80**, 6937–6941.
21. Vanin, E. F., Henthorn, P. S., Kioussis, D., Grosveld, F. & Smithies, O. (1983) *Cell* **35**, 701–709.
22. Jones, R. W., Old, J. M., Trent, R. J., Clegg, J. B. & Weatherall, D. J. (1981) *Nucleic Acids Res.* **9**, 6813–6825.
23. Loenen, W. A. M. & Blattner, F. R. (1983) *Gene* **26**, 171–179.
24. Hohn, B. (1979) *Methods Enzymol.* **68**, 299–309.
25. Rimm, D. L., Horness, D., Kucera, J. & Blattner, F. R. (1980) *Gene* **12**, 301–309.
26. Blattner, F. R., Blechl, A. E., Denniston-Thompson, K., Faber, H. E., Richards, J. E., Slightom, J. L., Tucker, P. W. & Smithies, O. (1978) *Science* **202**, 1279–1284.
27. Maxam, A. M. & Gilbert, W. (1980) *Methods Enzymol.* **65**, 499–560.
28. Slightom, J. L., Blechl, A. E. & Smithies, O. (1980) *Cell* **21**, 627–638.
29. Leach, D. R. F. & Stahl, F. W. (1983) *Nature (London)* **305**, 448–450.
30. Rosa, M. D., Hendrick, J. P., Lerner, M. R., Steitz, J. A. & Reichlin, M. (1983) *Nucleic Acids Res.* **11**, 853–870.
31. Varmus, H. E. (1982) *Science* **216**, 812–820.
32. Chen, H. R. & Barker, W. C. (1984) *Nucleic Acids Res.* **12**, 1767–1778.
33. Singer, M. F. (1982) *Cell* **28**, 433–434.
34. Manuelidis, L. & Biro, P. A. (1982) *Nucleic Acids Res.* **10**, 3221–3239.
35. Manuelidis, L. (1982) *Nucleic Acids Res.* **10**, 3211–3219.
36. Maizel, J. B. & Lenk, R. P. (1981) *Proc. Natl. Acad. Sci. USA* **78**, 7665–7669.
37. Weiher, H., Konig, M. & Gruss, P. (1983) *Science* **219**, 626–631.
38. Hayashida, H. & Miyata, T. (1983) *Proc. Natl. Acad. Sci. USA* **80**, 2671–2675.
39. McClements, W. L., Enquist, L. W., Oskarsson, M., Sullivan, M. & Vande Woude, G. F. (1980) *J. Virol.* **35**, 488–497.
40. O'Rear, J. J., Mizutani, S., Hoffman, G., Fiandt, M. & Temin, H. M. (1980) *Cell* **20**, 423–430.
41. Keshet, E. & Shaul Y. (1981) *Nature (London)* **289**, 83–85.
42. Varmus, H. E. (1983) in *Mobile Genetic Elements*, ed. Shapiro, J. A. (Academic, New York), pp. 411–503.
43. Seiki, M., Hattori, S. & Yoshida, M. (1982) *Proc. Natl. Acad. Sci. USA* **79**, 6899–6902.
44. Shimotohno, K., Golde, D. W., Miwa, M., Sugimura, T. & Chen, I. S. Y. (1984) *Proc. Natl. Acad. Sci. USA* **81**, 1079–1083.
45. Van Beveren, C., Rands, E., Chattopadhyaty, S. K., Lowy, D. R. & Verma, I. M. (1982) *J. Virol.* **41**, 542–556.