JOURNAL OF VIROLOGY, Dec. 1987, p. 4060–4066
0022-538X/87/124060-07$02.00/0

Vol. 61, No. 12

Human Endogenous Retroviruslike Genome with Type C *pol* Sequences and *gag* Sequences Related to Human T-Cell Lymphotropic Viruses

DIXIE L. MAGER* AND J. DOUGLAS FREEMAN

Terry Fox Laboratory, British Columbia Cancer Research Centre, and Department of Medical Genetics, University of British Columbia, Vancouver, British Columbia V5Z 1L3, Canada

Received 23 July 1987/Accepted 3 September 1987

We have cloned several prototypic members of the family of human endogenous retroviruslike elements having a histidine tRNA primer-binding site (RTVL-H) and have determined the nucleotide sequence of one of these clones (RTVL-H2). The RTVL-H2 sequence is 5,813 nucleotides long, with long terminal repeats of 450 nucleotides. Although this particular sequence contains no long open reading frames, computer searches have revealed several segments of amino acid homology with known retroviral gene products. In the *gag* region of RTVL-H2, there is a segment with significant homology to a region of the *gag* protein p30 of type C baboon endogenous virus. In the *pol* region of RTVL-H2, three segments similar to the Moloney leukemia virus (MLV) *pol* polyprotein were detected. These correspond to parts of the protease, reverse transcriptase, and endonuclease domains of the MLV *pol* gene. Interestingly, the last two *pol* domains are equidistant in RTVL-H2 and the type C murine retroviruslike DNA sequence (MuRRS), both having deletions of equal sizes relative to the MLV *pol* gene. One other segment similar to a retroviral gene product was identified in the RTVL-H2 *gag* region. This segment has 55 to 60% amino acid homology to a 50-amino-acid region of the *gag* nucleic acid-binding proteins encoded by human T-cell lymphotropic viruses types I and II and bovine leukemia virus. Thus, the RTVL-H2 genome harbors sequences related to evolutionarily distant retroviruses.

It is becoming evident that the human genome contains a complex variety of endogenous retrovirus-related sequences. Single or low-copy-number human elements have been described (1, 11, 18), as well as several distinct multicopy families of retroviruslike sequences (2, 6, 12, 13, 15, 16, 20, 24). Two of the best characterized of these human families are the type C genomes found in 35 to 50 copies (15, 24, 34) and the 50- to 100-copy sequences homologous to the mammalian type A, B, and D virus superfamily (2, 6, 16, 20). A 10,000-copy human transposonlike structure with long terminal repeats (LTRs) but no homology to other retroviruses has also been reported (21). The possible contribution of these sequences to human development, genetic variability, or disease is unknown.

We have previously identified a retroviruslike structure in human DNA because of its proximity to the human β-globin gene cluster (13). Sequences related to this element are present in approximately 1,000 copies per haploid genome. The element contains proviruslike LTRs with potential promoter and polyadenylation signals and a potential primer-binding site (PBS) homologous to histidine tRNA (13). This family of sequences was thus termed RTVL-H (i.e., retroviruslike and with PBSs homologous to histidine tRNA).

In our previous study, cloning difficulties prevented isolation of the entire element downstream of the β-globin gene. We have observed that bacteriophage clones containing RTVL-H elements are unstable and tend to suffer deletions during phage replication. In particular, deletions involving homologous recombination between the two LTRs are frequently observed (13; unpublished observations). Several prototypic RTVL-H sequences have now, however, been isolated from a phage library (13) containing 9- to 18-kilobase *Xba*I restriction fragments of human DNA. This library was used for the following reasons. (i) Southern blotting experiments on total human DNA indicate that there are no conserved *Xba*I sites within the RTVL-H sequence (unpublished data). (ii) Many of the recombinant phage DNA molecules (those with inserts of ≤14 kilobases) would be too small to be packaged if the 5.7-kilobase RTVL-H sequence was deleted because of recombination between its LTRs.

The restriction enzyme maps of three phages, XV-3, XV-7, and XV-10, which contain RTVL-H sequences are shown in Fig. 1, along with a map of phages containing portions of the 3'β RTVL-H (RTVL-H1) element isolated previously (13). LTR and interior sequences were localized by hybridization to RTVL-H1 probes. A highly conserved *Stu*I site at the 5' end of each LTR was also used to locate these sequences (Fig. 1). The maps of these clones are similar to a consensus or composite RTVL-H map derived from Southern blotting experiments on total human DNA with various RTVL-H-specific probes (unpublished data). Because phage clone XV-10 appears to contain a representative RTVL-H element (designated here as RTVL-H2), it was chosen for DNA sequence analysis.

The phage DNA from clone XV-10 was restricted with *Eco*RI, and the fragments were cloned into plasmid vectors. Further mapping of these plasmids identified appropriate smaller fragments that were cloned into pUC18 or pUC19 (40). These plasmid inserts were then sequenced by the dideoxy chain termination method (26) modified for use with double-stranded plasmid templates (10). Exonuclease III was used to generate sets of overlapping deletion derivatives of the plasmids (8).

The DNA sequence of the RTVL-H2 element is shown in Fig. 2. This proviruslike structure (including both LTRs) is 5,813 base pairs (bp) in length and is bounded by the 5-bp cellular repeat ATGAG. The 5' LTR is 450 bp, and the 3'

* Corresponding author.

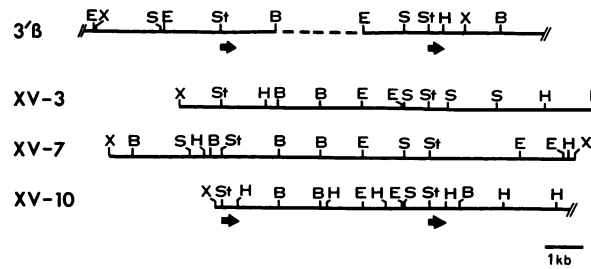

FIG. 1. Restriction enzyme maps of phages containing RTVL-H sequences. The map of the element found 3′ to the β-globin gene is from references 13 and 14. The dashed line represents sequences that were not cloned. XV-3, XV-7, and XV-10 are three cloned *Xba*I fragments containing intact RTVL-H elements. Arrows marking the location of the LTRs are shown only for 3′β and XV-10. B, *Bgl*II; E, *Eco*RI; H, *Hin*dIII; S, *Sst*I; St, *Stu*I; X, *Xba*I. Only those *Stu*I sites that occur at the 5′ end of each LTR are shown.

LTR is 451 bp long. The two LTRs differ at 18 positions (4% mismatch). Each LTR has correctly positioned potential promoter sequences and polyadenylation signals (indicated in Fig. 2) and has imperfect (83% identical) 47-bp repeats in its putative U_3 region (underlined in Fig. 2).

Just inside the 5′ LTR, at positions 453 to 470, is the potential tRNA PBS. In the 3′β (RTVL-H1) clone, this was identified as being homologous to the 3′ end of histidine tRNA (17 of 18 matches) (13). The RTVL-H2 PBS is also homologous (16 of 18 matches) to human histidine tRNA. A computer search confirmed that the RTVL-H2 PBS is more closely related to $tRNA^{His}$ than to any other tRNA species. DNA sequence analysis of the PBS portion of phage clones XV-3 and XV-7 has shown that these sequences are also homologous to histidine tRNA (data not shown). As in RTVL-H1 (13), RTVL-H2 has the expected polypurine tract immediately 5′ to the 3′ LTR (positions 5353 to 5365) which in other retroviruses serves as a PBS for synthesis of plus-strand viral DNA.

A complex set of direct repeats of variable length is found within RTVL-H2 between nucleotides 900 and 1250. Three tandem copies of the longest repeat in the region (44 bp) are overlined in Fig. 2 and span positions 1062 to 1193. These repeats are 95% homologous, each differing by two nucleotides or fewer from the consensus. No other significant direct or inverted repeats are found in RTVL-H2 (excluding the LTRs).

Translation of the RTVL-H2 sequence reveals no open reading frames longer than 600 bp in any of the six possible coding frames. Nonetheless, computer searches, using software provided by the University of Wisconsin Genetics Computer Group (7), have revealed regions of homology to other retroviral genomes. Two regions of homology to *gag* protein sequences were detected by searching the National Biomedical Research Foundation protein sequence data base release of December 1986 (release 11). The most 5′ region, corresponding to nucleotides 1977 to 2153, is shown as box A in Fig. 2. This sequence codes for a stretch of 59 amino acids which has 46.5% homology (and 52% nucleotide homology) to the type C baboon endogenous virus (BaEV) *gag* polyprotein residues 367 to 424 (35). One "X" has been inserted in the RTVL-H2 sequence at position 2115 in Fig. 2 to maintain the reading frame with homology to BaEV. The homologous regions are aligned in Fig. 3A. This region of the BaEV *gag* polyprotein corresponds to part of the core shell protein p30. Comparison of the RTVL-H2 amino acid segment with the Moloney murine leukemia virus (MLV) p30 sequence (31) and the equivalent region in a type C human endogenous retroviral element, termed 4.1, described previously (24) gives lower degrees of homology of 30 and 28%, respectively.

The other region of homology of the RTVL-H2 sequence to retroviral *gag* sequences spans nucleotides 2318 to 2458 (box B in Fig. 2). The amino acid translation of this sequence has greater than 50% homology to a portion of the *gag* nucleic acid-binding proteins of human T-cell lymphotropic virus types I (HTLV-I) (29) and II (HTLV-II) (30) and bovine leukemia virus (BLV) (25). Figure 3B shows a comparison of the HTLV-I and BLV *gag* segments to the RTVL-H2 sequence. The percent homologies of RTVL-H2 to HTLV-I, HTLV-II, and BLV are 55, 55, and 61%, respectively. At the DNA sequence level, the percent homologies of this RTVL-H2 segment to HTLV-I and BLV are 59 and 65%, respectively. This *gag* region contains the highly conserved motif $CX_2CX_4HX_4C$, invariant among retroviruses (5) and thought to be involved in binding of this protein to the retroviral genome. This sequence occurs once in the mammalian type C retroviruses such as MLV and BaEV but occurs twice in the HTLV-type genomes; avian type C viruses (e.g., Rous sarcoma virus; 28); type A (intracisternal A-type particle sequences; 19), B (mouse mammary tumor virus; 9), and D (simian retrovirus1; 22) viruses; and lentiviruses such as human immunodeficiency virus (23, 38). The RTVL-H2 region contains two copies of this conserved sequence in a location very similar to that found in other retroviruses (5), being just upstream of the protease domain (see below). The first copy has a stop codon in place of the final cysteine, but this can be accounted for by a single nucleotide substitution at position 2386 in Fig. 2. A pyrimidine instead of adenine at this position would result in a cysteine codon. The second copy of the motif is also imperfect in that it has three amino acids instead of four between the second cysteine and the conserved histidine (Fig. 3B). Both the RTVL-H segment and the HTLV-type nucleic acid-binding proteins have an abundance of proline residues which contribute to the high degree of homology between them. The proline content of nucleic acid-binding proteins of most other retroviruses, such as MLV (9%), mouse mammary tumor virus (7%), and human immunodeficiency virus (11%), is much lower than that found in HTLV-I (23.5%) or BLV (26%).

The three forward amino acid translations of the RTVL-H2 sequence were next compared to all retroviral *pol* protein sequences in the National Biomedical Research Foundation data base. Three significant regions of homology were detected by this analysis. The first region of homology spans RTVL-H2 nucleotides 2595 to 2864 (box C in Fig. 2) which corresponds to 90 amino acids. This region has 41.6% amino acid homology and 46% DNA sequence homology to the beginning of the MLV *pol* polyprotein. A comparison of the two amino acid sequences is shown in Fig. 3C. This region of the MLV *pol* gene, amino acid residues 10 to 98, is part of the protease domain of the *pol* polyprotein (31). The second region of homology spans RTVL-H nucleotides 2992 to 3474 and is indicated as box D in Fig. 2. This region has 50.3% amino acid homology and 50.7% DNA sequence homology to the MLV *pol* polyprotein residues 161 to 322 (31). This region corresponds to part of the reverse transcriptase domain which is well conserved between retroviruses of different types (3, 37). Figure 4A is a comparison of a portion of this RTVL-H2 amino acid sequence with the corresponding regions from several different retroviruses. The third region of homology spans RTVL-H2 nucleotides 4266 to


A
B
C
D
E

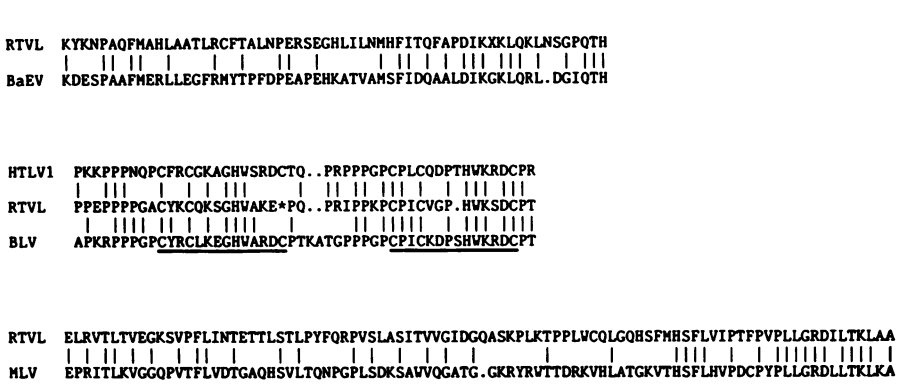

FIG. 3. (A) Amino acid homology of RTVL-H2 to the BaEV *gag* polyprotein. The RTVL-H2 region is the translation of nucleotides 1977 to 2153. The "X" in the amino acid sequence results from the single base inserted at nucleotide position 2115. The BaEV region corresponds to *gag* amino acid positions 367 to 424 (35). (B) Amino acid homology of RTVL-H2 to HTLV-I and BLV. The RTVL-H2 region shown is the translation of nucleotides 2318 to 2458. The asterisk represents a termination codon. The HTLV-I segment is part of the protein p15 and corresponds to *gag* positions 348 to 395 (29). The BLV region is part of the protein p12 and corresponds to *gag* positions 338 to 387 (25). The conserved cysteine-rich segments are underlined. (C) Amino acid homology of RTVL-H2 to the protease domain of the MLV *pol* polyprotein. The RTVL-H2 region is the translation of nucleotides 2595 to 2864. The MLV segment corresponds to *pol* positions 10 to 98. Dots are gaps introduced to improve alignment, and vertical lines mark amino acid identities with RTVL-H2.

4730 (box E in Fig. 2). This RTVL-H2 segment has 40% amino acid homology and 48.3% DNA sequence homology to residues 906 to 1065 of the MLV *pol* polyprotein (31). (It should be noted that two "X"s have been inserted in the RTVL-H2 nucleotide sequence at positions 4329 to 4330 in Fig. 2 to maintain the correct reading frame.) This region of the MLV *pol* gene corresponds to the conserved endonuclease domain (3, 37). Figure 4B compares part of this RTVL-H2 amino acid sequence with the corresponding endonuclease region from several other retroviruses. These comparisons demonstrate that the *pol* region of the RTVL-H2 genome is most similar to the *pol* genes found in mammalian type C retroviruses.

The sequence corresponding to most of the endonuclease domain shown in Fig. 4B is also available from the RTVL-H1 3′β clone (unpublished data). Positions where the translation of RTVL-H1 differs from that of RTVL-H2 are shown on the top line of Fig. 4B. The two sequences are 92% homologous at the nucleotide sequence level and differ at 15 of 106 amino acid positions. This comparison also shows that the frameshift at amino acid position 22 and three termination codons at positions 28, 35, and 91 are found in both RTVL-H sequences.

The amino acid sequence of the RTVL-H2 *pol* region is compared with the MLV *pol* polyprotein sequence by a dot matrix analysis in Fig. 5A. The three regions of homology discussed above are clearly evident. This figure also illustrates that the RTVL-H2 *pol* gene is shorter than that of MLV. There is a deletion in the RTVL-H2 sequence of 960 bp (320 amino acids) between the reverse transcriptase and endonuclease domains relative to the MLV *pol* gene.

The conserved reverse transcriptase and endonuclease domains of the RTVL-H2 sequence, shown in Fig. 4, were also compared with the equivalent regions in other type C repetitive retroviruslike genomes described previously. Specifically, we compared the RTVL-H2 sequence with the human sequence described by Repaske et al. (designated 4.1) (24) and with the mouse retroviral element described by Schmidt et al. (murine retroviruslike DNA sequence [MuRRS]) (27). The RTVL-H2 reverse transcriptase domain has 44.1% amino acid and 50.5% nucleotide sequence homology to the human 4.1 element and 51.5% amino acid and 53.4% DNA sequence homology to the MuRRS genome. The RTVL-H2 endonuclease domain has 45.7% amino acid and 51.7% DNA sequence homology to the 4.1 genome and 35.1% amino acid and 50.6% nucleotide homology to MuRRS. Thus, these comparisons reveal similar levels of homology in both regions of the RTVL-H sequence to MLV, 4.1, and MuRRS.

We have also noticed an interesting similarity between the RTVL-H and MuRRS *pol* regions; namely, their two conserved *pol* domains are separated by almost exactly the same number of nucleotides (1208 versus 1209) and therefore have the same-sized deletion with respect to the MLV *pol* gene. This finding is shown graphically in Fig. 5B, which is a dot matrix comparison of the RTVL-H2 and MuRRS *pol* regions at the amino acid level. No other significant regions of similarity beyond those shown in Fig. 5B were found between the RTVL-H2 and MuRRS sequences. Nevertheless, this finding, and the fact that the two elements are very close in overall size, suggest that the RTVL-H and MuRRS families may have had a common ancestor.

FIG. 2. Nucleotide sequence of RTVL-H2. All of the sequence shown was determined on both DNA strands. To confirm the *Eco*RI cloning site at position 3626, a 1.0 kilobase *Nco*I/*Mst*II fragment containing this site was isolated from phage XV-10 DNA and cloned into pUC19. The *Eco*RI site was sequenced by using an oligonucleotide primer complementary to positions 3637 to 3656. The other *Eco*RI site within the sequence at position 4675 falls within a region of homology with retroviral *pol* genes (see text). The numbering of the 5813-bp sequence is altered because of the insertion of one space at position 2115 and two spaces at positions 4329 to 4330 (see text). The pentanucleotide direct repeat that flanks the sequence is shown in small letters. The beginning and end of each LTR is indicated by short vertical arrows. The potential PBS and the polypurine tract are underlined. Features within the LTRs are marked only for the 5′ LTR. These features are the TATA sequence and potential polyadenylation signal (boxed) and 47-bp direct repeats shown by dashed underlines. The 44-bp direct repeats in the interior of the sequence are indicated by overlines. The five regions of homology to other retroviruses discussed in the text are boxed and lettered A through E.

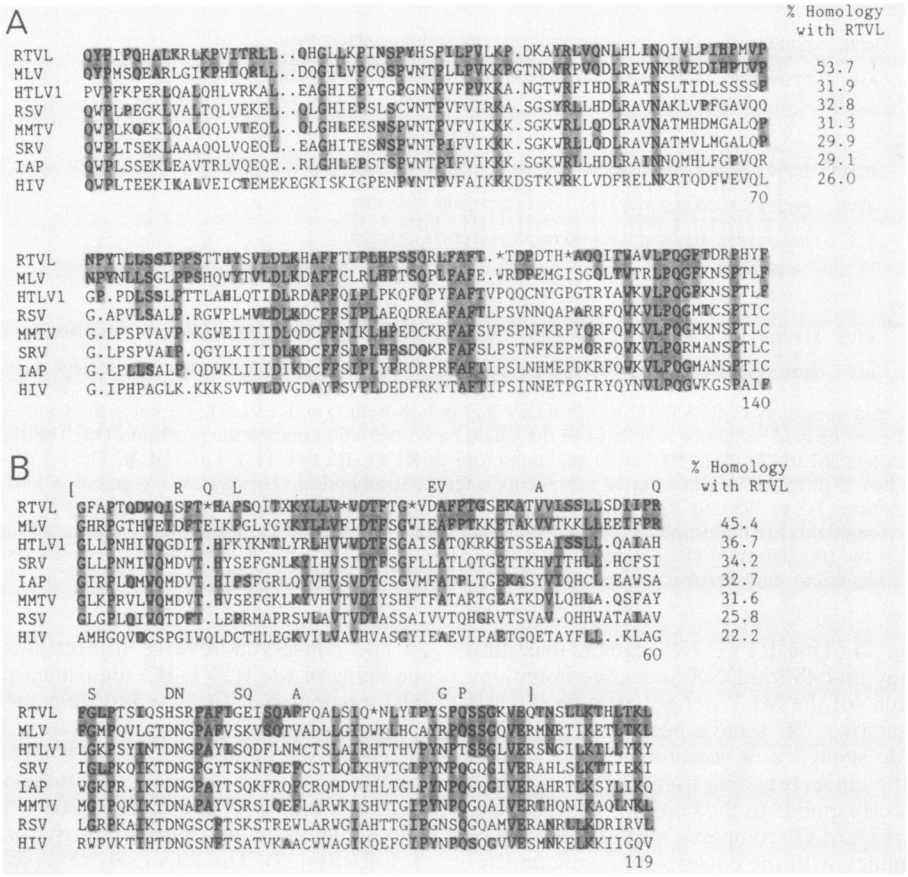

FIG. 4. (A) Comparison of RTVL-H2 with the reverse transcriptase regions of other retroviruses. The RTVL-H2 sequence is the translation of nucleotide positions 3158 to 3465. Sources for the other sequences are MLV (31), HTLV-I (29), Rous sarcoma virus (RSV) (28), mouse mammary tumor virus (MMTV) (17), simian retrovirus 1 (SRV) (22), intracisternal A-type particle sequence (IAP) (19), and human immunodeficiency virus (HIV) (23). The values at the end of the first row are the percent homology of each retroviral amino acid sequence to RTVL-H2 for the 140-amino-acid segment shown. Shaded positions are identical to those found in RTVL-H2. (B) Comparison of RTVL-H2 with the endonuclease domains of other retroviruses. The RTVL-H2 sequence is the translation of nucleotide positions 4266 to 4622. Other sequences are from the same sources as in Fig. 4A. The line above the RTVL-H2 sequence is the amino acid sequence derived from a different RTVL-H sequence (RTVL-H1). Only positions that differ from RTVL-H2 are shown. The brackets show the region for which RTVL-H1 sequence is available.

The RTVL-H2 DNA sequence was also compared with the sequences of a human mouse mammary tumor virus-related genome (20), a human transposon-like sequence (21), and a mouse ETn element (33), and no significant homology was detected.

The RTVL-H2 sequence, in particular the segment between the *pol* region and the 3′ LTR, was also analyzed for homology to retroviral *env* genes. These computer searches revealed no evidence for homology of RTVL-H2 to any known *env* sequence. In addition, no particularly long stretch of hydrophobic amino acids was found which might serve as an *env*-like transmembrane domain.

This study demonstrates that the RTVL-H family of human sequences is unique in at least three respects. First, although the RTVL-H2 sequence contains regions of similarity to other retroviruses, it is not closely related to any known retroviruslike sequence. Second, all four RTVL-H elements that have been examined have a potential histidine-tRNA PBS. No other retrovirus or retroviruslike sequence has been reported to have a PBS homologous to histidine tRNA (for a review, see reference 4). A third unique feature of the RTVL-H2 sequence is that it contains separate regions of homology to both the mammalian type C viruses, typified by MLV, and to the group of viruses consisting of HTLV-I, HTLV-II, and BLV. Although the viruses in the latter group are classified morphologically as type C (36), sequence analysis of these viruses (25, 29, 30) and phylogenetic trees of retroviruses (for an example, see reference 3) indicate that they fall into a separate class, termed type E by Sagata et al. (25).

The RTVL-H2 sequence presented here has no long open reading frames and, therefore, could not code for functional gene products. If we assume that this sequence integrated with identical LTRs at some time in the past, then the present 4% difference between LTRs suggests that integration probably occurred several million years ago. Thus, it is not unexpected that numerous termination codons may have been generated by mutation. However, Fig. 4B shows that two RTVL-H sequences have an identical frameshift and three of the same stops codons in the conserved endonuclease domain. This finding suggests that at least some RTVL-H elements may have been generated from a founder sequence containing a nonfunctional *pol* gene. If this is so, it is possible that other RTVL-H sequences may have provided

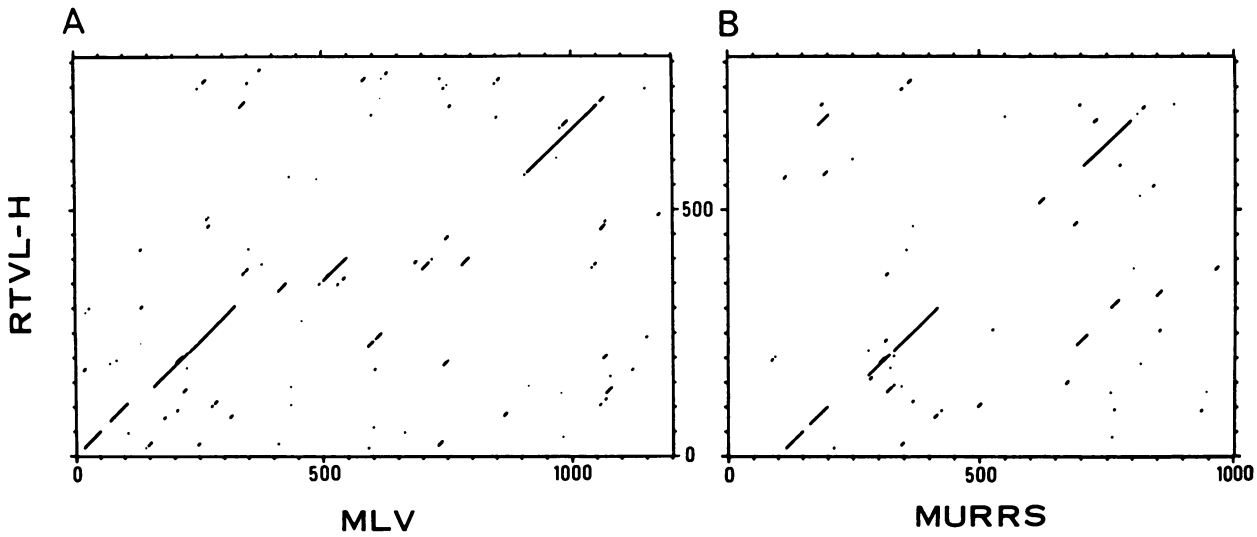

FIG. 5. Comparison of the RTVL-H2 amino acid sequence with the *pol* polyprotein of MLV (A) and the equivalent region in MuRRS (B). The sequences were compared by a dot matrix program, and each dot represents eight or more identities per 30 amino acids. The vertical axes for A and B correspond to the translation of RTVL-H2 nucleotide sequence positions 2568 to 4994. For this analysis, two bases were inserted after RTVL-H2 position 2911 and one base was inserted after position 4242 to maintain the same reading frame. The horizontal axis of part A is the MLV *pol* polyprotein (31). The horizontal axis of part B corresponds to the translation of MuRRS nucleotide sequence positions 2134 to 5150 (27). One and two bases were inserted after MuRRS positions 4312 and 4511, respectively, to maintain the correct reading frame.

the necessary functions in *trans* for reverse transcription and integration as a viruslike structure.

Alternatively, RTVL-H sequences, including H1 and H2, could have been amplified and dispersed by mechanisms other than viral reverse transcription and reintegration. This appears to have happened with the human type C (genome 4.1) family (34) and the endogenous feline leukemia viruses (32), because different family members have conserved flanking DNA. We presently have no evidence for conserved flanking DNA in the RTVL-H family. Hybridization experiments with probes flanking RTVL-H1 did not reveal a general association of these sequences to other RTVL-H elements (13). The maps in Fig. 1 also show no evidence for common restriction enzyme sites flanking the four RTVL-H genomes. In addition, the nucleotide sequence of the integration site of RTVL-H2 is not homologous to that of RTVL-H1 (data not shown). However, we cannot rule out the possibility that some RTVL-H members may have been amplified by mechanisms involving flanking DNA.

The high copy number of RTVL-H elements in the genome and the observation that RTVL-H1 is located close to the breakpoints of three naturally occurring deletions in the β-globin gene cluster (13, 14) suggest that these sequences may be involved in genetic rearrangements. In addition, recombinational events may have occurred between RTVL-H sequences and other viral genomes. For instance, it is possible that a type E virus ancestor recombined with an RTVL-H ancestor in the *gag* region encompassing the cysteine-rich motifs. The origin of the type E viruses is not clear, although there is some indication that HTLV-I and related viruses originated relatively recently in African primates (39). The findings presented here raise the possibility that endogenous RTVL-H family sequences may have contributed to the evolution of these infectious viruses.

We thank K. Humphries and F. Takei for helpful suggestions during the writing of this manuscript. We also thank S. Hayley for manuscript preparation.

This work was supported by grants from the National Cancer Institute of Canada and the British Columbia Health Care Research Foundation. Core support was provided by the Cancer Control Agency of British Columbia and the British Columbia Cancer Foundation. D.L.M. is a National Cancer Institute Group Research Scientist.

LITERATURE CITED

1. **Bonner, T. I., C. O'Connell, and M. Cohen.** 1982. Cloned endogenous retroviral sequences from human DNA. Proc. Natl. Acad. Sci. USA **79**:4709–4713.
2. **Callahan, R., I.-M. Chiu, J. F. H. Wong, S. R. Tronick, B. A. Roe, S. A. Aaronson, and J. Schlom.** 1985. A new class of endogenous human retroviral genomes. Science **228**:1208–1211.
3. **Chiu, I.-M., A. Yaniv, J. E. Dahlberg, A. Gazit, S. F. Skuntz, S. R. Tronick, and S. A. Aaronson.** 1985. Nucleotide sequence evidence for relationship of AIDS retrovirus to lentiviruses. Nature (London) **317**:366–368.
4. **Colicelli, J., and S. P. Goff.** 1986. Isolation of a recombinant murine leukemia virus utilizing a new primer tRNA. J. Virol. **57**:37–45.
5. **Covey, S. N.** 1986. Amino acid sequence homology in *gag* region of reverse transcribing elements and the coat protein gene of cauliflower mosaic virus. Nucleic Acids Res. **14**:623–633.
6. **Deen, K. C., and R. W. Sweet.** 1986. Murine mammary tumor virus *pol*-related sequences in human DNA: characterization and sequence comparison with the complete murine mammary tumor virus *pol* gene. J. Virol. **57**:422–432.
7. **Devereux, J., P. Haeberli, and O. Smithies.** 1984. A comprehensive set of sequence analysis programs for the VAX. Nucleic Acids Res. **12**:387–395.
8. **Henikoff, S.** 1984. Unidirectional digestion with exonuclease III creates targeted breakpoints for DNA sequencing. Gene **28**: 351–359.
9. **Fasel, N., K. Pearson, E. Buetti, and H. Diggelmann.** 1982. The region of mouse mammary tumor virus DNA containing the long terminal repeat includes a long coding sequence and signals for hormonally regulated transcription. EMBO J. **1**:3–7.
10. **Korneluk, R. G., F. Quan, and R. A. Gravel.** 1985. Rapid and reliable dideoxy sequencing of double-stranded DNA. Gene **40**:317–323.

11. **Leib-Mösch, C., R. Brack, T. Werner, V. Erfle, and R. Hehlmann.** 1986. Isolation of an SSAV-related endogenous sequence from human DNA. Virology **155**:666–677.
12. **Maeda, N.** 1985. Nucleotide sequence of the haptoglobin and haptoglobin-related gene pair. J. Biol. Chem. **260**:6698–6709.
13. **Mager, D. L., and P. S. Henthorn.** 1984. Identification of a retrovirus-like repetitive element in human DNA. Proc. Natl. Acad. Sci. USA **81**:7510–7514.
14. **Mager, D. L., P. S. Henthorn, and O. Smithies.** 1985. 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. Nucleic Acids Res. **13**:6559–6575.
15. **Martin, M. A., T. Bryan, S. Rasheed, and A. S. Khan.** 1981. Identification and cloning of endogenous retroviral sequences present in human DNA. Proc. Natl. Acad. Sci. USA **78**: 4892–4896.
16. **May, F. E. B., and B. R. Westley.** 1986. Structure of a human retroviral sequence related to mouse mammary tumor virus. J. Virol. **60**:743–749.
17. **Moore, R., M. Dixon, R. Smith, G. Peters, and C. Dickson.** 1987. Complete nucleotide sequence of a milk-transmitted mouse mammary tumor virus: two frameshift suppression events are required for translation of *gag* and *pol*. J. Virol. **61**:480–490.
18. **O'Connell, C., S. O'Brien, W. G. Nash, and M. Cohen.** 1984. ERV3, a full-length human endogenous provirus: chromosomal localization and evolutionary relationships. Virology **138**:225–235.
19. **Ono, M., H. Toh, T. Miyata, and T. Awaya.** 1985. Nucleotide sequence of the Syrian hamster intracisternal A-particle gene: close evolutionary relationship of type A particle gene to types B and D oncovirus genes. J. Virol. **55**:387–394.
20. **Ono, M., T. Yasunaga, T. Miyata, and H. Ushikubo.** 1986. Nucleotide sequence of human endogenous retrovirus genome related to the mouse mammary tumor virus genome. J. Virol. **60**:589–598.
21. **Paulson, K. E., N. Deka, C. W. Schmid, R. Misra, C. W. Schindler, M. G. Rush, L. Kadyk, and L. Leinwand.** 1985. A transposon-like element in human DNA. Nature (London) **316**:359–361.
22. **Power, M. D., P. A. Marx, M. L. Bryant, M. B. Gardner, P. J. Barr, and P. A. Luciw.** 1986. Nucleotide sequence of SRV-1, a type D simian acquired immune deficiency syndrome retrovirus. Science **231**:1567–1572.
23. **Ratner, L., W. Haseltine, R. Patarca, K. J. Livak, B. Starcich, S. F. Josephs, E. R. Doran, J. A. Rafalski, E. A. Whitehorn, K. Baumeister, L. Ivanoff, S. R. Petteway, Jr., M. L. Pearson, J. A. Lautenberger, T. S. Papas, J. Ghrayeb, N. T. Chang, R. C. Gallo, and F. Wong-Staal.** 1985. Complete nucleotide sequence of the AIDS virus, HTLV-III. Nature (London) **313**:277–284.
24. **Repaske, R., P. E. Steele, R. R. O'Neill, A. B. Rabson, and M. A. Martin.** 1985. Nucleotide sequence of a full-length human endogenous retroviral segment. J. Virol. **54**:764–772.
25. **Sagata, N., T. Yasunaga, J. Tsuzuku-Kawamura, K. Ohishi, Y. Ogawa, and Y. Ikawa.** 1985. Complete nucleotide sequence of the genome of bovine leukemia virus: its evolutionary relationship to other retroviruses. Proc. Natl. Acad. Sci. USA **82**:677–681.
26. **Sanger, F., S. Nicklen, and A. R. Coulson.** 1977. DNA sequencing with chain terminating inhibitors. Proc. Natl. Acad. Sci. USA **74**:5463–5467.
27. **Schmidt, M., T. Wirth, B. Kröger, and I. Horak.** 1985. Structure and genomic organization of a new family of murine retrovirus-like DNA sequences (MuRRS). Nucleic Acids Res. **13**:3461–3470.
28. **Schwartz, D. E., R. Tizard, and W. Gilbert.** 1983. Nucleotide sequence of Rous sarcoma virus. Cell **32**:853–869.
29. **Seiki, M., S. Hattori, Y. Hirayama, and M. Yoshida.** 1983. Human adult T-cell leukemia virus: complete nucleotide sequence of the provirus genome integrated in leukemia cell DNA. Proc. Natl. Acad. Sci. USA **80**:3618–3622.
30. **Shimotohno, K., Y. Takahashi, N. Shimizu, T. Gojobori, D. W. Golde, I. S. Y. Chen, M. Miwa, and T. Sugimura.** 1985. Complete nucleotide sequence of an infectious clone of human T-cell leukemia virus type II: an open reading frame for the protease gene. Proc. Natl. Acad. Sci. USA **82**:3101–3105.
31. **Shinnick, T. M., R. A. Lerner, and J. G. Sutcliffe.** 1981. Nucleotide sequence of Moloney murine leukaemia virus. Nature (London) **293**:543–548.
32. **Soe, L. H., B. G. Devi, J. I. Mullins, and P. Roy-Burman.** 1983. Molecular cloning and characterization of endogenous feline leukemia virus sequences from a cat genomic library. J. Virol. **46**:829–840.
33. **Sonigo, P., S. Wain-Hobson, L. Bougueleret, P. Tiollais, F. Jacob, and P. Brûlet.** 1987. Nucleotide sequence and evolution of ETn elements. Proc. Natl. Acad. Sci. USA **84**:3768–3771.
34. **Steele, P. E., M. A. Martin, A. B. Rabson, T. Bryan, and S. J. O'Brien.** 1986. Amplification and chromosomal dispersion of human endogenous retroviral sequences. J. Virol. **59**:545–550.
35. **Tamura, T.** 1983. Provirus of m7 baboon endogenous virus: nucleotide sequence of the *gag-pol* region. J. Virol. **47**:137–145.
36. **Teich, N.** 1985. Taxonomy of retroviruses, p. 1–16. *In* R. Weiss et al. (ed.), RNA tumor viruses, vol. 2. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
37. **Toh, H., H. Hayashida, and T. Miyata.** 1983. Sequence homology between retroviral reverse transcriptase and putative polymerases of hepatitis B virus and cauliflower mosaic virus. Nature (London) **305**:827–829.
38. **Wain-Hobson, S., P. Sonigo, O. Danos, S. Cole, and M. Alizon.** 1985. Nucleotide sequence of the AIDS virus, LAV. Cell **40**:9–17.
39. **Wong-Staal, F., and R. C. Gallo.** 1985. Human T-lymphotropic retroviruses. Nature (London) **317**:395–403.
40. **Yanisch-Perron, C., J. Vieira, and J. Messing.** 1985. Improved M13 phage cloning vectors and host strains: nucleotide sequences of the M13mp18 and pUC19 vectors. Gene **33**:103–119.