JOURNAL OF VIROLOGY, Feb. 1984, p. 615–620
0022-538X/84/020615-06$02.00/0

Vol. 49, No. 2

Nucleotide Sequence Analysis of the Long Terminal Repeat of Integrated Bovine Leukemia Provirus DNA and of Adjacent Viral and Host Sequences

DOMINIQUE COUEZ,[1]* JACQUELINE DESCHAMPS,[1,2] RICHARD KETTMANN,[1,2] ROBERT M. STEPHENS,[3] RAYMOND V. GILDEN,[3] AND ARSENE BURNY[1,2]

Département de Biologie Moléculaire, Université Libre de Bruxelles, 1640 Rhode-St-Genèse,[1] and Faculté des Sciences Agronomiques de l'Etat, 5800 Gembloux,[2] Belgium, and National Cancer Institute, Frederick Cancer Research Facility, Frederick, Maryland 21701[3]

Received 6 April 1983/Accepted 7 October 1983

The nucleotide sequence of the 3′ long terminal repeat and adjacent viral and host sequences was determined for a bovine leukemia provirus cloned from a bovine tumor. The long terminal repeat was found to comprise 535 nucleotides and to harbor at both ends an imperfect inverted repeat of 7 bases. Promoter-like sequences (Hogness box and CAT box), an mRNA capping site, and a core enhancer-related sequence were tentatively located. No kinship was detected between this bovine leukemia proviral fragment and other retroviral long terminal repeats, including that of human T-cell leukemia virus.

Retroviruses provide useful systems to study the integration of foreign DNA into eucaryotic genomes and the regulation of expression of the viral genes (2, 48). The provirus, the integrated form of the virus genome, contains the viral genes flanked by long terminal repeat (LTR) sequences covalently linked to cell DNA. The integration event causes the loss of (usually) 2 base pairs (bp) at the distal end of each LTR and the appearance of inverse repeats of a few bp in cell DNA (8, 17, 39, 44, 45). Encoded in the LTRs are signal sequences for regulating the transcription of viral genes and for processing the transcripts by the addition of polyadenylic acid-containing tails. LTRs also contain potential Z-DNA-forming sequences and core enhancer sequences thought to be part of enhancer elements (27, 47). All these regulatory structures play a role in activating the expression of viral genes. They also activate the expression of a cellular gene in leukosis induced by avian leukosis virus. In some tumors, avian leukosis virus integrates immediately upstream of the c-*myc* gene and activates it by downstream promotion (14, 26). In other tumors, provirus integration exerts a regional effect, regardless of orientation and position (28, 30). Downstream promotion makes use of the LTR promoter; the regional activation suggests that the viral DNA contains sequences that enhance expression from heterologous promoters. An enhancing activity has been recently identified in the Rous sarcoma virus promoter (21). It has been localized partly (88 bp) in the 5′ end of the LTR within the U_3 region and partly (55 bp) in the viral sequences flanking the 5′ boundary of the LTR from circular DNA. In this case, the enhancing effect is only present at the 3′ end of naturally acquired proviruses.

The availability of DNA clones of the bovine leukemia virus (BLV) genome has allowed investigations of the mode of action of the virus in inducing the tumor phase of enzootic bovine leukosis (5, 19, 20). The provirus found in the tumor cell is in a repressed state, integrated in different chromosomes in different tumors (D. Gregoire, D. Couez, J. Deschamps, S. Heuertz, M.-C. Hors-Cayla, J. Szpirer, C. Szpirer, A. Burny, G. Huez, and R. Kettmann, submitted for publication). Possible explanations of the BLV mode of action include (i) cell transformation by a viral product at a particularly critical stage of cell differenciation or (ii) enhancement of expression of one or several cellular oncogenes. The sites of BLV provirus integration in bovine tumors being multiple (20), it can be hypothesized that the provirus integrates at a crucial spot in host DNA and is transposed later on to a secondary site, where it is found in the tumor.

Two sets of data motivated our in-depth studies of the BLV provirus structure: (i) The presence at rather high frequency (25% of total cases) of proviruses with deletions in tumor tissue. The systematic occurrence of the deletion in the 5′ moiety of the provirus was indicative of the putatively important role played by the 3′ side of the molecule in the neoplastic process. (ii) The discovery of the human T-cell leukemia virus (HTLV) (31, 32) and of its apparent uniqueness among the Retroviridae, except for a distinct but clear-cut relatedness to BLV (29).

Since the U_3 region of the LTR has been repeatedly (3, 34, 43) suspected of playing a key role in leukemogenesis by a number of leukemia viruses, we first focused our attention on the detailed structure of the proviral LTR cloned from a tumor tissue. This work was in progress when Seiki et al. (37) published the LTR sequence of HTLV, thus allowing a direct comparison between corresponding regions of the related proviral agents.

Localization of LTR sequences within integrated BLV DNA and sequencing strategy. An *Eco*RI tumor DNA fragment designated T15-4, containing the right viral LTR with flanking cellular and viral sequences, was previously cloned in Charon 21 A phage (7, 22). The fragment was recovered from an *Eco*RI digest of the recombinant phage after agarose gel electrophoresis and electroelution. An 8.3-kilobase (kb) *Sac*I fragment containing the BLV information (5), except for 104 bases missing because of a second *Sac*I site in the LTR, was subcloned into the *Pst*I site of pBR322 by the method of Villa-Komaroff et al. (46). Restriction of the cloned BLV DNA with *Pst*I generated four fragments of 1.2, 0.5, 1.4, and 5.2 kb (Fig. 1B). The 5′ LTR-*gag* gene junction fragment (1.2 kb) was recovered as described above for the *Eco*RI DNA fragment.

* Corresponding author.

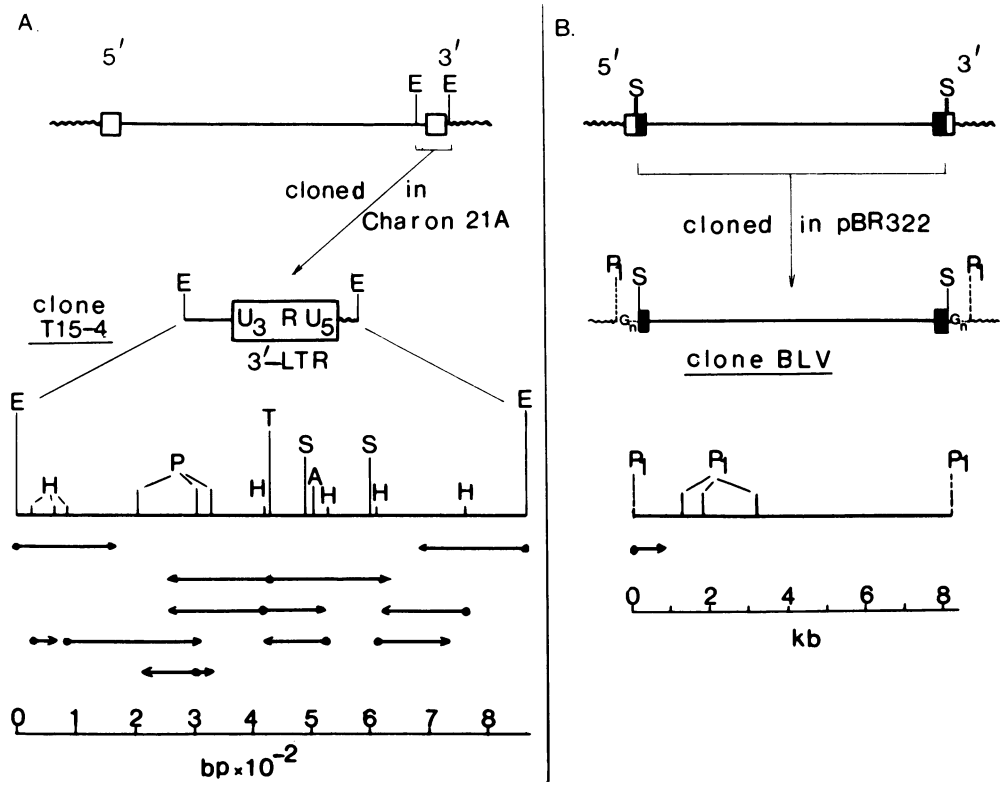

FIG. 1. Molecular cloning of BLV LTRs and strategy for determining their nucleotide sequence. Box, LTR DNA; straight line, viral DNA; wavy line, cellular DNA; A, *Aval*; E, *Eco*RI; H, *Hpa*II; P, *Pvu*II; P_1, *Pst*I; S, *Sac*I; T, *Taq*I. (A) *Eco*RI tumor DNA fragment designated T15-4, containing the 3' LTR with flanking viral and cellular sequences, was cloned in Charon 21A. This fragment was isolated from a *Eco*RI digest of the recombinant phage, and its single restriction map was established. The T15-4 fragment was then cut with *Taq*I, *Hpa*II, and *Pvu*II and was subcloned in M13mp7, mp8, or mp9. (B) The 8.3-kb *Sac*I fragment containing all the BLV information was cloned into the *Pst*I site of pBR322. This cloned BLV DNA was cut with *Pst*I, and the 1.2-kb *Pst*I fragment which bears the U_5 of the 5' LTR was subcloned in M13mp8. These different subclones were sequenced by the method of Sanger et al. (36). Horizontal arrows indicate the direction and extent of sequence determination obtained by cleaving the DNA with the indicated endonuclease.

The map of the *Eco*RI fragment T15-4 was based on restriction endonuclease cleavage products obtained and visualized either on agarose gels after staining with ethidium bromide or by autoradiography after labeling of 5' ends with [γ-^{32}P]ATP and T4 polynucleotide kinase.

To use a shotgun approach as described by Messing et al. (24, 25), the T15-4 fragment was cut with *Taq*I, *Hpa*II, or *Pvu*II and inserted into the *Acc*I or *Sma*I site of the coliphage M13mp7, mp8, or mp9. After cloning, these sets of small fragments yielded banks of recombinant phages in which the T15-4 fragment was represented piecemeal with the two complementary DNA strands present in separate clones. The nucleotide sequences of the cloned single-stranded DNAs were determined by the method of Sanger et al. (36).

Figure 1A provides a summary of the strategy used to determine the entire sequence. Each piece of DNA was sequenced at least twice, and the sequences were read by two different investigators. The entire primary structure was constructed by compiling the data obtained and aligning the successive fragments by overlaps and complementarities (Fig. 2A).

The 1.2-kb *Pst*I fragment derived from cloned BLV DNA was subcloned into the *Pst*I site of phage M13mp8, and sequence analysis was limited to the terminal regions.

The 8.3-kb *Sac*I fragment was also subcloned into a modified pUC8 vector. This clone has been used for complete sequencing of the T15 proviral genome, except for a small *Sac*I fragment (104 bases) located within the LTR. Sequencing was performed in both directions over the 500-bp regions at both ends of the large fragment by the method of Maxam and Gilbert (23). This sequence was then compared with that obtained independently from the *Eco*RI fragment.

The complete sequence of the T15-4 fragment and the 5' LTR-*gag* junction fragment is shown in Figure 2. The plus strand (same sense as viral RNA) is presented.

The sites for a number of restriction endonucleases are presented in Table 1. It should be noticed that the *Taq*I sites at positions −467 and −416 were apparently not accessible to the enzyme.

The strategy used to delineate LTR boundaries rested upon sequencing of the entire 3' LTR and the right-hand side of the 5' LTR. The point of sequence divergence identified the limit between the LTR and *gag* regions (for the 5' LTR; Fig. 3B) and between the LTR and cellular genomic DNA (for the 3' LTR; Fig. 3A). Such a transition point was found corresponding to the residues numbered 321 in Fig. 2. Residue 321 of the 5' LTR (Fig. 2B) and the stretch of 18 bases (TGGGGGCTCGTCCGGGAT) complementary to the 3' sequence of $tRNA^{Pro}$ (13) were separated only by one T residue. Divergence of the T15-4 and 5' LTR-*gag* sequences, together with the position for accommodation of initiator $tRNA^{Pro}$, was consistent with the idea that the sequence

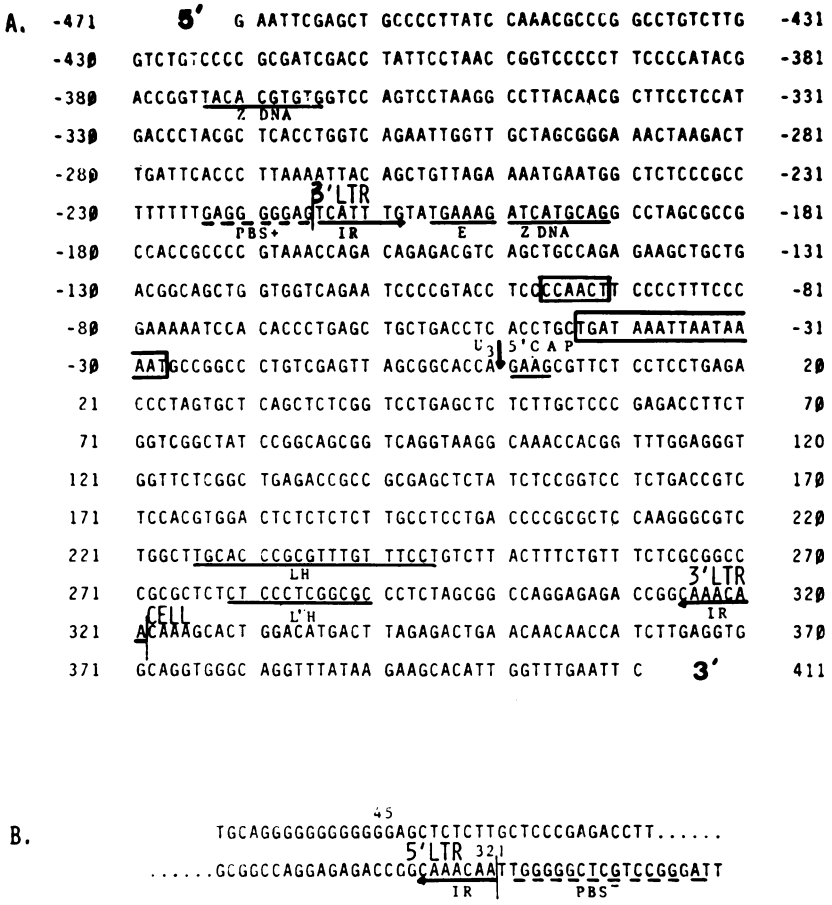

FIG. 2. (A) Nucleotide sequence of the integrated 3′ LTR with adjacent viral and cellular DNA in T15-4 fragment. (B) Partial nucleotide sequence of the 5′ LTR in 1.2-kb *Pst*I fragment. The sequence of the plus strand (same sense as viral RNA) is presented. Position 1 in the sequence corresponds to the 5′ end of the RNA genome, the capping site (↓). Important features of the BLV LTR sequence: IR, inverted repeat; Z-DNA, zone of Z-DNA; E, enhancer core sequence; U_3, unique sequence derived from the 3′ end of genomic RNA; PSB+, plus-strand binding site; PBS−, $tRNA^{Pro}$ binding site; LH, RU_5 locally homologous regions to U_3 regions of the simian sarcoma virus LTR. The possible promoter sequences Hogness and CAT boxes are shown in boxes.

. . .CAAACAA included the expected inverted repeat known to exist at both ends of LTRs. Indeed, the search for the inverted repeat of CAAACAA pointed to sequence TCATTTG, numbered −215 to −209 in Fig. 2A, although these proposed inverted repeats did not exactly match (they are imperfect inverted repeats owing to two AC pairs). Moreover, the left-hand boundary of all 3′ LTRs studied so far is adjacent to a purine tract, thought to act as a plus-strand primer-binding site (3, 38, 40).

In the Rous sarcoma virus system, at least 9 and at most 29 of the nucleotides of this region adjacent to U_3 were demonstrated to be necessary for growth, even in the presence of a helper virus. It was concluded that the polypurine tract provided a *cis*-acting function necessary for retrovirus replication (40). The purine array GAGGGGAG from residues −224 to −216 is an obvious candidate for such a function and confirms the strong conservative behavior of this sequence in evolution.

The single T residue mentioned above as being located between the inverted repeat and the $tRNA^{Pro}$ binding site was lacking on the right side of the T15-4 fragment. A few bp, usually two, are eliminated from the ends of both LTRs at some point before or during integration (8, 39, 44, 45). In the present case, only one T-A bp was apparently involved in this process. Based upon the observations and assumptions detailed above (Fig. 2), it was concluded that the BLV LTR extended from the nucleotide T labeled −215 to the nucleotide A labeled +321, totaling 535 residues.

TABLE 1. Restriction endonuclease sites of T15-4 DNA

Enzyme	Sequence	Positions
*Alu*I	AGCT	−465, −261, −150, −125, −63, +32, +46, +144
*Ava*I	CCCGAG	+58
*Dde*I	CTNAG	−357, −289, −66, +15, +29, +43, +130, +339
*Eco*RI	GAATTC	−472, +406
*Hae*III	GGCC	−442, −353, −192, −24, +267, +299
*Hpa*II	CCGG	−402, −380, −24, +80, +154, +311
*Hin*fI	GANTC	−280, −219, −113, +179
*Pvu*I	CGATCG	−420
*Pvu*II	CAGCTG	−262, −151, −126
*Sac*I	GAGCTC	+45, +143
*Sau*3A	GATC	−419
*Sau*96I	GGNCC	−400, −366, −24, +39, +156, +267
*Taq*I[a]	TCGA	−468, −417, −18

[a] Two *Taq*I sites (−468, −417) were resistant to endonuclease digestion.

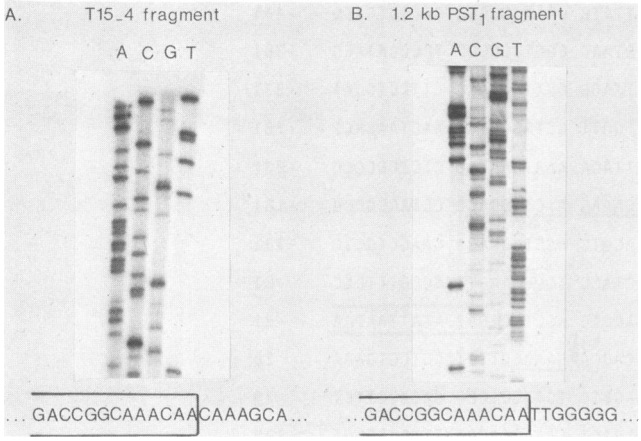

FIG. 3. Autoradiograms of sequencing gels illustrating the point of sequence divergence among the two fragments analyzed. (A) T15-4 fragment containing the 3′ LTR with an adjacent cellular sequence; (B) *Pst*I fragment of 1.2 kb bearing the U_5 of the 5′ LTR, and the nucleotide sequence deduced from the autoradiograms. The sequences commonly found in A and B accompanied by samples from each of the reactions (lanes designated ACGT) were fractionated by electrophoresis in a 6% polyacrylamide gel containing 7 M urea. After electrophoresis at 1,500 V for 1.5 or 3 h, the gel was removed for autoradiography.

Potential transcription regulatory sites in the BLV LTR. Once the boundaries of the BLV LTR were delineated and the $tRNA^{Pro}$ initiator site was located, it became possible to locate the putative viral RNA capping site at the GAAG sequence and to number the left-side G of this sequence as residue number 1. The assignment was made possible by taking advantage of the size of BLV strong-stop cDNA (320 ± 1 nucleotides [37]) for BLV produced by the chronically infected fetal lamb kidney cell line.

Once presently known LTRs are all arranged according to the same general scheme, one should find upstream from the capping site the two transcriptional regulatory entities called the TATA (or Hogness) box (11, 12) and the CAT box (10). The TATA box element has been tentatively identified within the AT-rich region extending between positions −44 and −28; the 5′ side of this sequence, TGATAAA, is the closest to the standard recognized sequence, NTATAAAN, and runs from −44 to −38. The sequence CCAACT occurred between positions −97 and −92 and was taken as the putative CAT box. The positions of all three transcriptional signals relative to each other are in accordance with what is known in other virus systems (9, 17, 35, 41, 44).

An obvious candidate for the polyadenylation signal (8, 26) would be the AATAAA sequence extending between positions −35 and −30, nearly adjacent to the presumed TATA box. No experimental attempt has been made so far to locate the adenylation site; CA is the consensus sequence (33). In HTLV LTRs, the polyadenylation site has been tentatively assigned to a TA or a TG pair but not to a CA (37). A number of CA and TG were found in the region where a polyadenylation site was theoretically expected to be located. Accurate determination of this site is mandatory before the R region can be exactly delineated.

A computer search was carried out to identify DNA segments of alternating purine-pyrimidine sequences considered to be potential Z-DNA-forming regions and key structures of transcription enhancers (27). Such structures were found as pairs with nucleotide sequences between them varying in length from 50 to 80 bp. They seem to exist in many retrovirus LTRs (27), and in some systems, at least, the enhancers involve viral sequences at the 3′ end of proviral DNA but outside the LTR (21). Our data point to the sequence TACACGTGTGGT (−374 to −365) as the best candidate region with Z-DNA-forming potential. However, this DNA segment is located outside the LTR. In the LTR itself, the computer analysis showed one 8-bp segment, ATCATGCA (−200 to −192), with one bp out of alternation. A 6-bp segment, CACGTG, was found at positions 388 to 393 in cell DNA. Another segment of interest is the core enhancer element with the consensus sequence 5′-$^{A}_{C}$GGAAGTGA$^{A}_{C}$-3′ (15) or 5′-TGGAAAG-3′ (47), in which one critical replacement appears to be the first G. The BLV LTR contains the segment 5′-TGAAAG-3′ at positions −206 to −201, a 6-bp sequence consistent with the consensus sequences mentioned above. However, the assumptions made here are hypothetical, and the homologies observed with known core sequences of enhancer elements may be coincidental. The role of these sequences in the activation of transcription should now be explicitly tested.

Protein-coding potential. As illustrated in Fig. 4, there is no large open reading frame in the BLV LTR. The largest one exists in frame A and extends over a region of about 200 nucleotides, but it does not contain any initiation codon ATG.

Comparison of the BLV LTR with the LTRs from other mammalian retroviruses. The overall size of the BLV LTR,

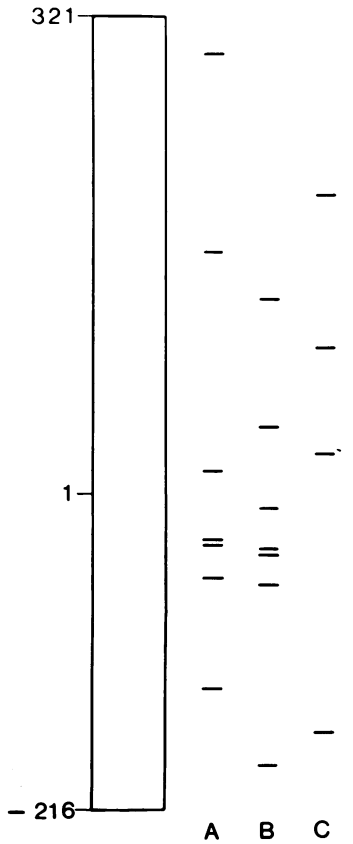

FIG. 4. Reading frame analysis of the BLV 3′ LTR. To the right of the LTR box, horizontal lines represent the position of translation termination codons (TAA, TGA, TAG) found in the three different reading frames designated A, B, and C.

535 nucleotides, and its right-hand boundary, immediately adjacent to a 18-bp sequence complementary to the 3′ end of $tRNA^{Pro}$ (13), are peculiarities shared by LTRs in mammalian viruses. The primer tRNA used by BLV had not been identified yet. The presence of the $tRNA^{Pro}$ complementary sequence strongly suggests that $tRNA^{Pro}$ is used in BLV retrotranscription, as in Moloney murine leukemia and sarcoma viruses (8, 44, 45), spleen necrosis virus (38), baboon endogenous virus (41), simian sarcoma virus (6), and HTLV (37). Thus, all analyzed mammalian type C viruses use $tRNA^{Pro}$ as primer molecules, whereas avian viruses specialize in using $tRNA^{Trp}$ (4, 42). The only known exception to this rule is avian spleen necrosis virus, which also uses $tRNA^{Pro}$. Interestingly enough, spleen necrosis virus displays homologies with mammalian retroviruses (1, 16), and thus fits within the frame of the proposition of Devare et al. (6) that the $tRNA^{Pro}$ binding site might be a marker for the evolutionary relationship among retroviruses.

A computer-assisted search by the method of Kaneshisa (18) was used to look for homologous subsequences in BLV LTR and LTRs of Moloney leukemia virus, simian sarcoma virus, Rauscher leukemia virus, gibbon ape leukemia virus, baboon endogenous virus, and HTLV. Except for two locally homologous regions, TGCACCCGCGTTTG TTTCCT and CTCC · CTCGGCGC, found at successive positions in SSV-U_3 (positions 313 and 350 [6]) and BLV-RU_5 (positions 226 and 279), no significant homology was revealed between BLV LTR and those cited above, including HTLV. It thus seems that, aside from a number of practically invariant structures and signals, LTRs fulfill their biological functions while being permissive to genetic drift. We did not, indeed, expect many similarities to appear between BLV and murine LTRs since only a very distant relationship seems to exist between the bovine virus and the murine group. Somewhat to our surprise, we reached the same conclusion when comparing BLV and HTLV LTR sequences. The human virus LTR is significantly longer (37) (754 bp), with no apparent homology to BLV. It should thus be concluded that similarities between BLV and HTLV found in several structural proteins (27) had a more stringent biological function to fulfill than the overall base sequence of LTR besides the basic, well-conserved signals like inverted repeats, CAT and TATA boxes, the capping site, and promoter binding sites for retrotranscription.

We thank M. Heilporn and D. Denicourt for help in some experiments. We also thank R. Herzog and C. Wilmart for their kind help with computer analyses.

This work was supported by the Fonds Cancérologique of the Caisse Générale d'Epargne et de Retraite and the Fonds National Belge de la Recherche Scientifique (crédit aux chercheurs). D.C. received a fellowship from the Institut pour la Recherche Scientifique dans l'Industrie et l'Agriculture. J.D. is an Assistant of the Fonds Cancérologique de la Caisse Générale d'Epargne et de Retraite. R.K. is a Chercheur Qualifié of the Fonds National de la Recherche Scientifique.

ADDENDUM IN PROOF

A recent publication by Tsimanis et al. (Nucleic Acids Res. **11**:6079–6087, 1983) enables us to locate the polyadenylic acid addition site at base C, position 234. It follows that the R region of BLV LTR is 234 bp long, a value very similar to that found for the HTLV LTR (R = 228 bp), thus reemphasizing the similarities that exist between both viruses.

LITERATURE CITED

1. **Barbacid, M., E. Hunter, and S. A. Aaronson.** 1979. Avian reticuloendotheliosis viruses: evolutionary linkage with mammalian type C retroviruses. J. Virol. **30**:508–514.
2. **Bishop, J. M.** 1978. Retroviruses. Annu. Rev. Biochem. **47**:35–88.
3. **Czernilofsky, A. P., W. De Lorbe, R. Swanstrom, H. E. Varmus, J. M. Bishop, E. Fischer, and H. M. Goodman.** 1980. The nucleotide sequence of an untranslated but conserved domain at the 3′ end of the avian sarcoma virus genome. Nucleic Acids Res. **8**:2967–2984.
4. **Dahlberg, J. E., R. C. Sawyer, J. M. Taylor, A. J. Faras, W. E. Levinson, H. M. Goodman, and J. M. Bishop.** 1974. Transcription of DNA from the 70*S* RNA of Rous sarcoma virus: identification of a specific 4*S* which serves as primer. J. Virol. **13**:1126–1133.
5. **Deschamps, J., R. Kettmann, and A. Burny.** 1981. Experiments with cloned complete tumor-derived bovine leukemia virus information prove that the virus is totally exogenous to its target animal species. J. Virol. **40**:605–609.
6. **Devare, S. G., E. P. Reddy, J. D. Law, and S. A. Aaronson.** 1982. Nucleotide sequence analysis of the long terminal repeat of integrated simian sarcoma virus: evolutionary relationship with other mammalian retroviral long terminal repeats. J. Virol. **42**:1108–1113.
7. **de Wet, J. R., D. L. Daniels, J. L. Schroeder, B. G. Williams, K. Denniston-Thompson, D. D. Moore, and F. R. Blattner.** 1980. Restriction maps for twenty-one Charon vector phages. J. Virol. **33**:401–410.
8. **Dhar, R., W. L. McClements, L. W. Enquist, and G. F. Vande Woude.** 1980. Nucleotide sequences of integrated Moloney sarcoma provirus long terminal repeats and their host and viral junctions. Proc. Natl. Acad. Sci. U.S.A. **77**:3937–3941.
9. **Donehower, L. A., A. L. Huang, and G. L. Hager.** 1981. Regulatory and coding potential of the mouse mammary tumor virus long terminal redundancy. J. Virol. **37**:226–238.
10. **Efstratiadis, A., J. W. Posakony, T. Maniatis, R. W. Lawn, C. O. Connell, R. A. Spritz, J. K. Deriel, B. G. Forget, S. M. Weissman, J. L. Slightom, A. E. Blechl, O. Smithies, F. E. Baralle, C. C. Shoulders, and N. J. Proudfoot.** 1980. The structure and evolution of the human globin gene family. Cell **21**:653–668.
11. **Gannon, F., K. O'Hare, F. Perrin, J. P. Le Pennec, C. Benoist, M. Cochett, R. Breathnach, A. Royal, A. Garapin, B. Cami, and P. Chambon.** 1979. Organisation and sequences at the 5′ end of a cloned complete ovalbumin gene. Nature (London) **278**:428–433.
12. **Gilboa, E., S. Goff, A. Shields, F. Yoshimura, S. Mitra, and D. Baltimore.** 1979. In vitro synthesis of a 9 kbp terminally redundant DNA carrying the infectivity of Moloney murine leukemia virus. Cell **16**:863–874.
13. **Harada, F. G., G. Peters, and J. E. Dahlberg.** 1979. The primer tRNA for Moloney murine leukemia virus DNA synthesis. Nucleotide sequence and aminoacylation of $tRNA^{Pro}$. J. Biol. Chem. **254**:10979–10985.
14. **Hayward, W. S., B. G. Neel, and S. M. Astrin.** 1981. Activation of a cellular onc gene by promoter insertion in ALV-induced lymphoid leukosis. Nature (London) **290**:475–480.
15. **Hearing, P., and T. Shenk.** 1983. The adenovirus type 5 E1A transcriptional control region contains a duplicated enhancer element. Cell **33**:695–703.
16. **Hunter, E., A. S. Bhown, and J. C. Bennett.** 1978. Amino-terminal amino acid sequence of the major structural polypeptides of avian retroviruses: sequence homology between reticuloendotheliosis virus p30 and p30s of mammalian retroviruses. Proc. Natl. Acad. Sci. U.S.A. **75**:2708–2712.
17. **Ju, G., and A. M. Skalka.** 1980. Nucleotide sequence analysis of the long terminal repeat (LTR) of avian retroviruses: structural similarities with transposable elements. Cell **22**:379–386.
18. **Kanehisa, M. I.** 1982. Los Alamos sequence analysis package for nucleic acids and proteins. Nucleic Acids Res. **10**:183–195.
19. **Kettmann, R., J. Deschamps, Y. Cleuter, D. Couez, A. Burny, and G. Marbaix.** 1982. Leukemogenesis by bovine leukemia

virus: proviral DNA integration and lack of RNA expression of viral long terminal repeat and 3′ proximate cellular sequences. Proc. Natl. Acad. Sci. U.S.A. **79**:2465–2469.

20. **Kettmann, R., J. Deschamps, D. Couez, J.-J. Claustriaux, R. Palm, and A. Burny.** 1983. Chromosome integration domain for bovine leukemia provirus in tumors. J. Virol. **47**:146–150.
21. **Luciw, P. A., J. M. Bishop, H. E. Varmus, and M. R. Capecchi.** 1983. Location and function of retroviral and SV 40 sequences that enhance biochemical transformation after microinjection of DNA. Cell **33**:705–716.
22. **Maniatis, T., R. C. Hardison, E. Lacy, J. Lauer, C. O'Connell, D. Quon, D. K. Sim, and A. Efstratiadis.** 1978. The isolation of structural genes from libraries of eucaryotic DNA. Cell **15**:687–701.
23. **Maxam, A. M., and W. Gilbert.** 1977. A new method for sequencing DNA. Proc. Natl. Acad. Sci. U.S.A. **74**:560–564.
24. **Messing, J., R. Crea, and P. H. Seeburg.** 1981. A system for shotgun DNA sequencing. Nucleic Acids Res. **9**:309–320.
25. **Messing, J., B. Gronenborn, B. Muller-Hill, and P. H. Hofschneider.** 1977. Filamentous coliphage M13 as a cloning vehicle: insertion of a Hind III fragment of the lac regulatory region in M13 replicative form in vitro. Proc. Natl. Acad. Sci. U.S.A. **74**:3642–3646.
26. **Neel, B. G., W. S. Hayward, H. L. Robinson, J. Fang, and S. M. Astrin.** 1981. Avian leukosis virus-induced tumors have common proviral integration sites and synthesize discrete new RNAs: oncogenesis by promoter insertion. Cell **23**:323–334.
27. **Nordheim, A., and A. Rich.** 1983. Negatively supercoiled simian virus 40 DNA contains Z-DNA segments within transcriptional enhancer sequences. Nature (London) **303**:674–679.
28. **Nusse, R., and H. E. Varmus.** 1982. Many tumors induced by the mouse mammary tumor virus contain a provirus integrated in the same region of the host genome. Cell **31**:99–109.
29. **Oroszlan, S., M. G. Sarngadharan, T. D. Copeland, V. S. Kalyanaraman, R. V. Gilden, and R. C. Gallo.** 1982. Primary structure analysis of the major internal protein p24 of human type C T-cell leukemia virus. Proc. Natl. Acad. Sci. U.S.A. **79**:1291–1294.
30. **Payne, G. S., J. M. Bishop, and H. E. Varmus.** 1982. Multiple arrangements of viral DNA in an activated host oncogene in bursal lymphomas. Nature (London) **295**:209–213.
31. **Poiesz, B. J., F. W. Ruscetti, A. F. Gazdar, P. A. Bunn, J. D. Minna, and R. C. Gallo.** 1980. Detection and isolation of type-C retrovirus particles from fresh and cultured lymphocytes of a patient with cutaneous T-cell lymphoma. Proc. Natl. Acad. Sci. U.S.A. **77**:7415–7419.
32. **Poiesz, B. J., F. W. Ruscetti, M. S. Reitz, V. S. Kalyanaraman, and R. C. Gallo.** 1981. Isolation of a new type-C retrovirus (HTLV) in primary uncultured cells of a patient with Sezary T-cell leukemia. Nature (London) **294**:268–271.
33. **Proudfoot, N. J., and G. G. Brownlee.** 1974. Sequence at the 3′ end of the globin mRNA shows homology with immunoglobulin light chain mRNA. Nature (London) **252**:359–362.
34. **Robinson, H. L., B. M. Blais, P. L. Tsichlis, and J. M. Coffin.** 1982. At least two regions of the viral genome determine the oncogenic potential of avian leukosis viruses. Proc. Natl. Acad. Sci. U.S.A. **79**:1225–1229.
35. **Rushlow, K. E., J. A. Lautenberger, E. P. Reddy, L. M. Souza, M. A. Baluda, J. G. Chirikjian, and T. S. Papas.** 1982. Nucleotide sequence analysis of the long terminal repeat of avian myeloblastosis virus and adjacent host sequences. J. Virol. **42**:840–846.
36. **Sanger, P., S. Nicklen, and R. Coulson.** 1977. DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. U.S.A. **74**:5463–5467.
37. **Seiki, M., S. Hattori, and M. Yoshida.** 1982. Human adult T-cell leukemia virus: molecular cloning of the provirus DNA and the unique terminal structure. Proc. Natl. Acad. Sci. U.S.A. **79**: 6899–6902.
38. **Shimotohno, K., S. Mizutani, and H. M. Temin.** 1980. Sequence of retrovirus provirus resembles that of bacterial transposable elements. Nature (London) **285**:550–554.
39. **Shimotohno, K., and H. M. Temin.** 1980. No apparent nucleotide sequence specificity in cellular DNA juxtaposed to retrovirus proviruses. Proc. Natl. Acad. Sci. U.S.A. **77**:7357–7361.
40. **Sorge, J., and S. H. Hughes.** 1982. Polypurine tract adjacent to the U_3 region of the Rous sarcoma virus genome provides a *cis*-acting function. J. Virol. **43**:482–488.
41. **Tamura, T., M. Noda, and T. Takano.** 1981. Structure of the baboon endogenous virus genome: nucleotide sequence of the LTR. Nucleic Acids Res. **9**:6615–6625.
42. **Taylor, J. M.** 1977. An analysis of the role of tRNA species as primers for the transcription into DNA of RNA tumor virus genomes. Biochim. Biophys. Acta **473**:57–71.
43. **Tsichlis, P. N., and J. M. Coffin.** 1980. Recombinants between endogenous and exogenous avian tumor viruses: role of the C region and other portions of the genome in the control of replication and transformation. J. Virol. **33**:238–249.
44. **Van Beveren, C., J. G. Goddard, A. Berns, and I. Verma.** 1980. Structure of Moloney murine leukemia viral DNA: nucleotide sequence of the 5′ long terminal repeat and adjacent cellular sequences. Proc. Natl. Acad. Sci. U.S.A. **77**:3307–3311.
45. **Van Beveren, C., E. Rands, S. K. Chattopadhyay, D. R. Lowy, and I. M. Verma.** 1982. Long terminal repeat of murine retroviral DNAs: sequence analysis, host-proviral junctions, and preintegration site. J. Virol. **41**:542–556.
46. **Villa-Komaroff, L., A. Efstratiadis, S. Broome, P. Lomedico, R. Tizard, S. P. Naker, W. L. Chick, and W. Gilbert.** 1978. A bacterial clone synthesizing proinsulin. Proc. Natl. Acad. Sci. U.S.A. **75**:3727–3731.
47. **Weiher, H., M. König, and P. Gruss.** 1983. Multiple point mutations affecting the simian virus 40 enhancer. Science **219**:626–631.
48. **Weiss, R., N. Teich, H. Varmus, and J. Coffin (ed.).** 1982. RNA tumor viruses, 2d ed. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.