Proc. Natl. Acad. Sci. USA
Vol. 81, pp. 4741–4745, August 1984
Biochemistry

Bovine leukemia virus: Unique structural features of its long terminal repeats and its evolutionary relationship to human T-cell leukemia virus

(integrated provirus/secondary structure/enhancer element/retrovirus evolution)

NORIYUKI SAGATA*, TERUO YASUNAGA†, YASUKI OGAWA‡, JUNKO TSUZUKU-KAWAMURA*, AND YOJI IKAWA*

*Laboratory of Molecular Oncology and †Computation Center, The Institute of Physical and Chemical Research, Wako, Saitama 351, Japan; and ‡Department of Epizootiology, Faculty of Veterinary Medicine, Hokkaido University, Sapporo, Hokkaido 060, Japan

Communicated by Motoo Kimura, April 25, 1984

ABSTRACT **The nucleotide sequence of the long terminal repeat (LTR) of bovine leukemia virus, a unique oncogenic retrovirus of cattle, was determined. The LTR consisted of 530 base pairs (bp) with an inverted repeat of 6 bp at its 5′ and 3′ ends, flanked by a direct repeat of 6 bp of host cell origin. A $tRNA^{Pro}$ binding site for minus-strand DNA synthesis followed the 5′ LTR. The U3 region contained putative transcriptional promoters, "CAT" box and "TATA" box, but they had peculiar sequences (C-C-A-A-*C*-T and *G*-A-T-A-A-A-T). The U3 region also contained a potential enhancer element, whose sequence partially resembled those of other viral and cellular, especially of immunoglobulin, enhancers. The most striking structural feature of the LTR was an exceptionally long R region (228 bp), which separated a poly(A) addition signal (A-A-T-A-A-A) from a poly(A) site as far apart as 260 bp. The long R region was suggested to form a large stable hairpin structure on a nascent RNA chain, making the two transcription termination signals close together and thus ensuring normal termination of the chain. This structural feature of the bovine leukemia virus LTR was analogous to that of human T-cell leukemia virus LTR and, in fact, slight sequence homology (at most 50%) was observed between the R regions of these two retroviruses, indicating their evolutionary relationship. The unique structural feature of bovine leukemia virus and human T-cell leukemia virus LTRs may thus bear some relation to the biological features commonly shared by these retroviruses.**

An important structural feature of the retroviral genome is the presence of two long terminal repeats (LTRs) at the 5′ and 3′ ends of the proviral genome (1). Each LTR contains the sequences U3-R-U5, where U3 and U5 represent unique sequences derived from the 3′ and 5′ ends, respectively, of viral RNA, and R represents a short, terminally redundant sequence present at both termini of viral RNA, and it is always terminated by short inverted repeats and is flanked by direct repeats of host cell origin (2). The LTR plays a crucial role in retrovirus replication and is thus known to contain regulatory sequences necessary for transcription (1). The LTR also contains an enhancer element that greatly increases the transcriptional efficiencies of both viral and nearby cellular genes (3). Furthermore, the LTR shows significant sequence homology between those of related retroviruses, providing models for retrovirus evolution (4–7).

Bovine leukemia virus (BLV), an exogenous type C retrovirus, is a causative agent of enzootic bovine leukosis or lymphosarcoma in cattle (8). BLV differs in several biological and biochemical properties from other retroviruses. It is morphologically atypical (9), and, unlike other mammalian type C retroviruses, its reverse transcriptase has an absolute requirement for Mg^{2+} (10). It shows no extensive nucleotide or structural protein homologies with other known retroviruses (11, 12). Curiously, its productively infected cell line has rarely been established (13) and, in fact, it does not replicate efficiently in the natural target B lymphocytes (14). Finally, BLV is not integrated into a common site in the tumor cell chromosome (15, 16) and does not activate a downstream host cell gene (14). In the present study, we examined the nucleotide sequence of the cloned BLV LTR, which we expected to show some unique structural features, because of its essential role in viral replication. We found that it has several unique features with respect to its nucleotide sequence and the locations of regulatory signals for transcription and also found that BLV is distantly related only to human (or adult) T-cell leukemia virus (HTLV or ATLV).

MATERIALS AND METHODS

Molecular Cloning of BLV DNA. The isolation of a DNA clone of integrated BLV from tumor cells in λ Charon 4A and its subcloning in pBR322 have been described (17).

DNA Sequence Analysis. The nucleotide sequence was determined as described by Maxam and Gilbert (18). DNA fragments were obtained by use of various restriction endonucleases (Takara–Shuzo, Kyoto, Japan) and all of these were labeled at the 3′ end by using cordycepin 5′-[α-^{32}P]triphosphate (Amersham, 3000 Ci/mmol; 1 Ci = 37 GBq).

Computer-Aided Analysis of the Sequence Determined. Sequence homology was examined by the DNA version of the method developed for distantly related proteins (19). Secondary structures were deduced, as the forms with minimal free energies, according to the computer program developed for large RNA sequences (20). Sequence alignment of distantly related DNA sequences was performed by a modification of the method initially developed for proteins (21).

RESULTS

Localization of LTR Sequences Within Integrated BLV DNA and Sequencing Strategy. The isolation and characterization of a molecular clone (λBLV-1) of integrated BLV provirus DNA has been described (17). Only restriction sites of interest are shown in Fig. 1. The LTR sequences within the BLV genome were located on the basis of the identical restriction sites near either end of BLV DNA and the hybridization of the sequences to a BLV LTR-specific probe (17). We sequenced both the 5′ and 3′ LTRs and the host flanking sequences by the strategy shown in Fig. 1.

Structural Features Common to Other Retrovirus LTRs. The nucleotide sequences of the 5′ and 3′ LTRs and the cellular flanking regions are shown in Fig. 2*A*. The BLV LTR

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.

Abbreviations: BLV, bovine leukemia virus; bp, base pair(s); HTLV (or ATLV), human (or adult) T-cell leukemia virus; kb, kilobase(s); LTR, long terminal repeat.

consisted of 530 bp and contained an imperfect inverted repeat of 6 bp with the sequence 5′ T-G-T-A-T-G 3′ at its 5′ and 3′ termini. There was a 6-bp host sequence (G-A-C-A-G-G), directly repeated at the site of virus integration. Immediately following the 5′ LTR, the BLV genome bore a stretch of 18 bp (at positions 533–550) that was complementary to the 3′ sequence of proline tRNA, suggesting that, as with other mammalian type C viruses (2), $tRNA^{Pro}$ may serve as a primer for reverse transcription of BLV. Preceding the 3′ LTR, on the other hand, there was a polypurine tract of 9 bp (G-A-G-G-G-G-G-A-G), which might serve as a primer for plus-strand DNA synthesis (1). Thus, as schematically represented in Fig. 2*B*, the termini and boundaries of the BLV LTRs have the same structural organization as those of other known retrovirus LTRs.

Unique Structural Features of the BLV LTR Sequence. Besides the common features described above, we found several unique features in the internal sequence organization of the BLV LTR. We first attempted to identify the R region, a portion of the LTR unit (U3-R-U5) that normally begins with a guanosine (G) residue (i.e., cap site) and terminates with a dinucleotide C-A [i.e., poly(A) addition site] (1). BLV strong-stop cDNA is 320 bp long (22), predicting that the G residue at the cap site should precede the $tRNA^{Pro}$ binding site by 320 bp. We found a G residue at exactly the predicted position (position 212, Fig. 2*A*), suggesting that this G residue (or a closely located G residue at position 215) is the cap site or the beginning of the R region. On the other hand, as many as 10 C-A dinucleotides were found as possible candidates for the poly(A) site downstream of the putative cap site. Careful examination of their flanking sequences, however, revealed that only the dinucleotide C-A at positions 438–439 formed a part of 11-bp stretch (T-C-T-G-G-C-T-T-G-*C-A*) that closely matched the transcription termination signal (T-T-T-G-C-N-$\frac{G}{C}$-T-T-G-C-A) of other retroviruses (23). Furthermore, this C-A was followed by T-T-G-T (at positions 447–450), a sequence that is frequently observed immediately after the poly(A) site (1). Thus, the C-A sequence at positions 438–439 is the most likely poly(A) site, or the end of the R region. We estimated from these observations that the R region of BLV LTR is 228 bp long and hence that the U3 region is 211 bp long and the U5 region is 91 bp long (Fig. 2*B*). The R region thus constitutes 43% of the LTR

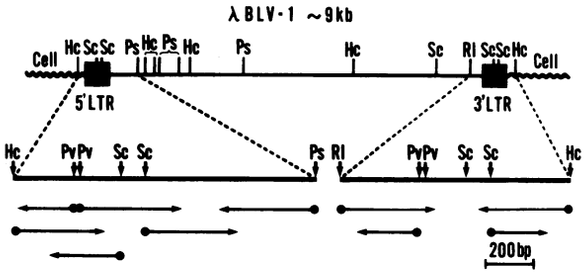

FIG. 1. Restriction enzyme map and strategy for sequencing BLV LTR. The isolation and characterization of integrated BLV provirus DNA, λBLV-1, were described (17). The 3′ end-labeled [α-^{32}P]DNA fragments were sequenced by the procedure of Maxam and Gilbert (18). RI, *Eco*RI; Sc, *Sac* I; Hc, *Hinc*II; Ps, *Pst* I; Pv, *Pvu* II. ●, Labeled end of each fragment. The extent and direction of sequencing are indicated by arrows. bp, Base pairs; kb, kilobases.

sequence, which is much more than in other retroviruses (16–79 bp, constituting 1.2–14% of the LTR; ref. 1). But surprisingly, it was identical to that of HTLV LTR (22). [It should be noted, however, that the U3 and U5 regions of BLV LTR were both much shorter than those (351 and 175 bp, respectively; ref. 22) of HTLV LTR.]

We then attempted to identify the transcription promoter signals, the CAT box (consensus; C-C-A-A-T) and TATA box (T-A-T-A-$\frac{A}{T}$-A-$\frac{T}{A}$), which are usually located in the U3 region, 70–90 and 20–30 bp, respectively, upstream from the G residue of the cap site (1). We could not find any typical signal sequences anywhere in the LTR, but closely related sequences were present 92 and 37 bp upstream from the cap site: the sequence at positions 115–120 had a single-base insertion (C-C-A-A-*C*-T) and that at positions 169–175 had a single-base substitution (*G*-A-T-A-A-A-T). Although these sequences are peculiar, they probably represent CAT and TATA boxes, judging from their locations at about the expected positions.

The canonical signal sequence (A-A-T-A-A-A) for polyadenylylation is located 10–20 bp upstream from the poly(A) site in most retroviruses (1). In the BLV LTR, this hexanucleotide was not at the expected position but as far as 260 bp upstream (positions 177–182) from the poly(A) site. If this

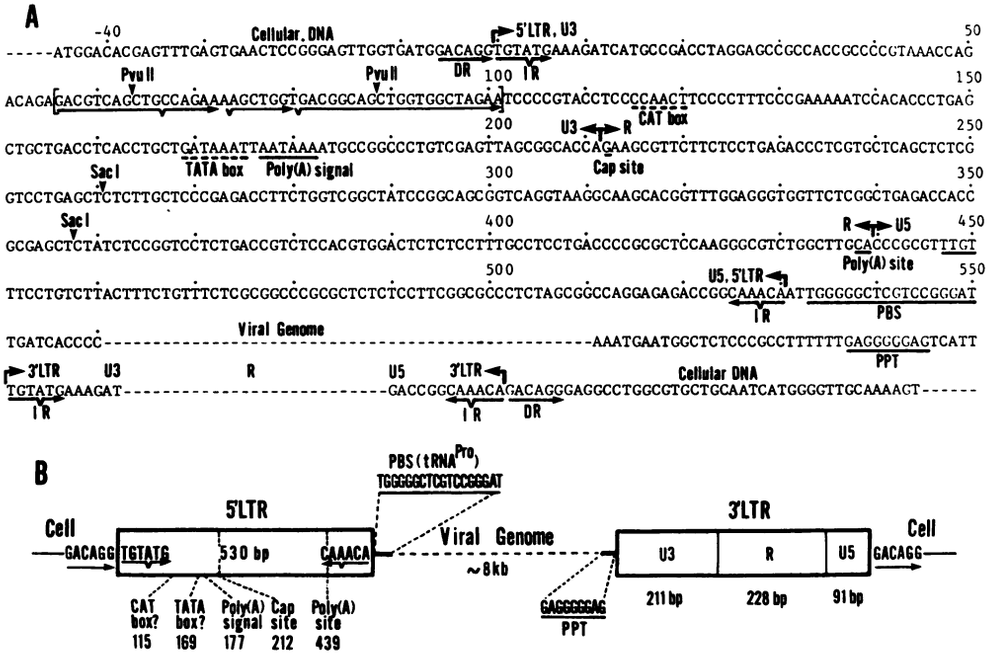

FIG. 2. Sequence (*A*) and summary of the major features (*B*) of BLV LTR and its adjacent regions in the integrated proviral genome. (*A*) The sequence of LTR and its flanking regions proceeding in the 5′ to 3′ direction with the same polarity as that of the BLV genomic RNA are shown. (*B*) Important features of the BLV LTR sequence are illustrated diagramatically. U3, sequences unique to the 3′ end of viral RNA; R, terminally redundant sequences of viral RNA; U5, sequences unique to the 5′ end of viral RNA; IR, inverted repeat; DR, direct repeat; PBS, primer binding site for minus-strand DNA synthesis; PPT, polypurine tract for plus-strand DNA synthesis. C-C-A-A-C-T at positions 115–120 and G-A-T-A-A-A-T at positions 169–175 are the most likely candidates for the "CAT" box and "TATA" box. Sequences in brackets (positions 56–101) denote a potential enhancer element.

hexanucleotide is actually a poly(A) signal, then it must normally function as part of a recognition signal for proper processing and polyadenylylation (24) of the BLV RNA transcript, irrespective of its long distance from the poly(A) site. We then performed computer-aided construction of a possible secondary structure of the 3′ terminus of the BLV RNA transcript, to determine whether the poly(A) signal-like sequence and the poly(A) site could become close together by folding of the RNA transcript. Fig. 3*A* shows the secondary structure formed by nucleotide sequences between positions 170–440. This region formed a large hairpin structure, involving the entire R region and having a high G-C pair content (71%), particularly at the base of the main stem. Consequently, the hexanucleotide A-A-T-A-A-A became located close (as close as 6 bp) to the poly(A) site. If such a secondary structure is formed *in vivo*, the A-A-T-A-A-A sequence could serve as a "normal" poly(A) addition signal on the RNA transcript.

To assess the significance of thermodynamical stability of the proposed secondary structure, we performed 300 random reshufflings of the sequence that formed the secondary structure (positions 170–440) and examined the stability of the most possible secondary structure formed by each of the 300 random sequences (20). This gave the histogram in Fig. 3*B*, showing the frequency (or number) of random sequences as a function of their thermodynamical stability. Clearly, the proposed secondary structure (arrow in Fig. 3*B*) was highly stable (−151.5 kcal/mol; 1 cal = 4.184 J) and such a stable structure was rare (only 1 of the 300 samples had slightly higher stability). Thus, the nucleotide sequences spanning the possible poly(A) signal and the poly(A) site appear to be "biologically" significant for formation of a stable secondary structure that makes the two widely separated transcription termination signals spacially close and thus functional.

Existence of a Potential Enhancer Element. Like certain DNA viruses and cellular genes, retrovirus LTRs contain enhancers or activators that greatly increase the transcriptional efficiency of viral and heterologous genes (3). They are located in the U3 region 100–250 bp upstream of the cap site (25, 26), usually as tandemly repeated sequences. The BLV LTR sequence had three tandemly repeated sequences in the U3 region 111–156 bp upstream of the cap site (positions 56–101, Fig. 2*A*); the repeated sequences were shorter than those of other retroviruses (≈70 bp; ref. 25), and the first two sequences were partial repeats of the third, relatively long sequence (Fig. 4*A*). The third repeated sequence contained an octanucleotide, G-T-G-G-C-T-A-G, which closely resembled the crucial "core" element (consensus sequence; G-T-G-G-$\frac{A}{T}$-$\frac{A}{T}$-$\frac{A}{T}$-G) commonly found in most viral enhancers (27). Fig. 4*B* shows a comparison of the core sequence and its flanking sequences of the "potential" BLV enhancer and of viral and cellular genes. With the minimal number of gaps to maximize the homology, the potential BLV enhancer showed various degrees of homology (50–70%) with those from other sources, depending on the source compared. It showed the greatest homology with the mouse immunoglobulin heavy chain gene (C_μ) enhancer, which functions in a B-cell-specific manner (32, 33).

Relationship of BLV to Other Retroviruses. To obtain a clue to the origin of BLV and its relationship to other retroviruses, we compared, by computer analysis (19), the BLV LTR sequences with those of several other retroviruses (including HTLV), transposable elements, and bovine satellite DNA bearing LTR-like sequences (34). Under conditions detecting 20-bp unit sequences with >60% homology, the two-dimensional homology matrix showed no significant homology between the BLV LTR and any other LTRs examined (data not shown). However, further analysis under less stringent conditions (50% homology detection) revealed, although over high backgrounds, a relatively long stretch (≈100 bp) of BLV sequence that showed slight homology with only HTLV LTR. This stretch was located in the R region, which was identical in size in BLV and HTLV (see above), thus permitting alignment of nucleotide matches. Fig. 5 shows the computer alignment (21) between the R re-

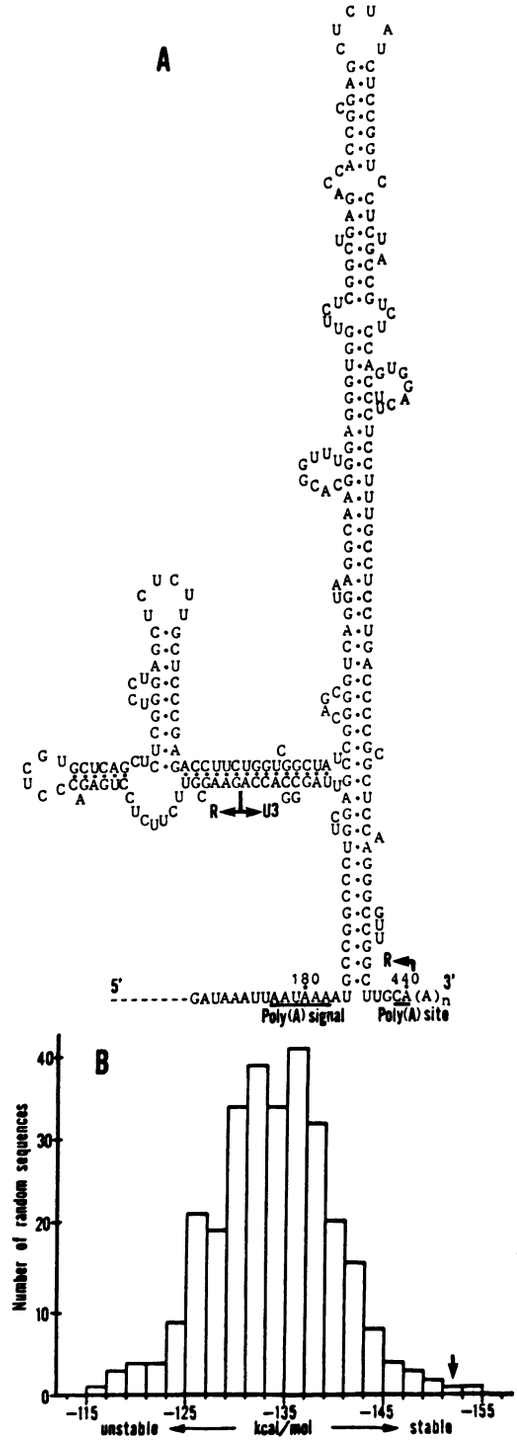

FIG. 3. Proposed secondary structure of the 3′ terminus of BLV RNA (*A*) and histogram showing the significance of its thermodynamical stability (*B*). The secondary structure is shown as it may exist in the 3′-terminal region of viral RNA (*A*), and its thermodynamical stability is assessed in comparison with the secondary structures formed by 300 reshuffled sequences (*B*). The arrow in *B* denotes the position of the proposed secondary structure.

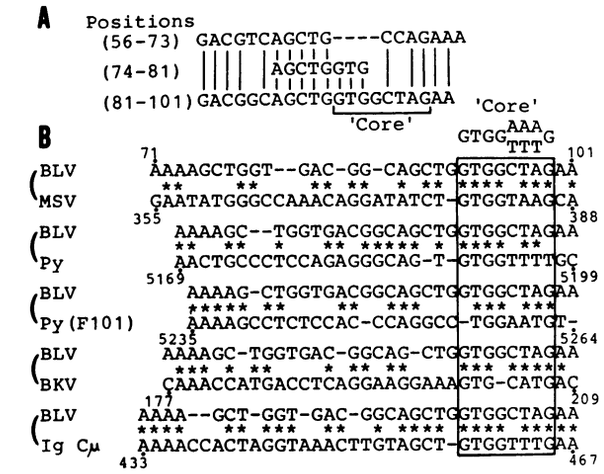

FIG. 4. Putative BLV enhancer element (*A*) and a comparison of its core and flanking sequences with those of other viral and cellular enhancer elements (*B*). See text for discussion of *A*. In *B*, sequences are aligned at the putative "core" sequences (27). The residues that match those in the putative BLV enhancer are indicated by asterisks, with the minimal number of gaps (indicated by dashes) to maximize the homology. Numbering of the sequences is according to refs. 28 (MSV), 29 (Py), 30 (PyF101), 31 (BKV), and 32 (Ig C_μ).

gions of the two retrovirus LTRs, made by introducing several gaps to maximize the homology. Overall homology (counting gaps) was as low as 48% (117 matches among 246 nucleotides). Even a stretch of 91 bp (underlined in Fig. 5), which could be aligned without gaps and was located in the same position in the R regions of BLV and HTLV, gave 46% homology. The probability that this 91-bp sequence of HTLV matches the corresponding 91-bp sequence of BLV with >46% homology was calculated to be as low as $<4 \times 10^{-5}$, indicating the significance of the homology observed. Furthermore, the last 50 nucleotides of the BLV sequence contained several stretches of nucleotides that were highly homologous with those in HTLV (Fig. 5). Thus, BLV and HTLV appear to be evolutionarily related, although their degree of relatedness is small. [We also found many, but short, stretches of homologous nucleotides in the U5 (and U3) regions of BLV and HTLV, but alignment of their matches was difficult because of their great differences in size.]

DISCUSSION

We have determined the nucleotide sequence of the LTR of BLV, finding both general and unique structural features in

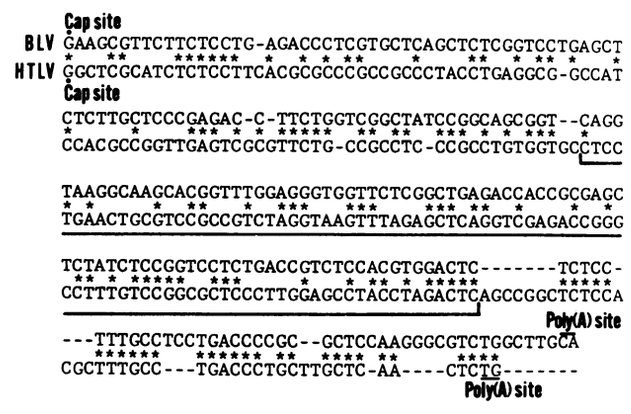

FIG. 5. Sequence alignment of the R regions of BLV and HTLV LTRs. The R region sequences of BLV are from Fig. 2*A* and those of HTLV are from ref. 22, and they are aligned by using a computer program (21). A stretch of 91 bp that was aligned without gaps is underlined.

it. The general features (1) include inverted repeats (6 bp) at both ends of the LTR, direct repeats (6 bp) of the host sequence at the provirus integration site, a primer ($tRNA^{Pro}$) binding site following the 5′ LTR, and a polypurine tract (9 bp) preceding the 3′ LTR.

The U3 region of the BLV LTR contained possible transcriptional promoters, CAT and TATA boxes at appropriate positions, although they had peculiar sequences. Such unusual sequences might affect the transcriptional efficiency of a downstream gene (35), and, if so, they could be related to the low efficiency of replication of this retrovirus (13, 14). The U3 region also contained sequences that might be an enhancer element. Although no extensive nucleotide sequence homology has been observed between various enhancers (3), the potential BLV enhancer appeared to have somewhat significant homology, within a relatively short sequence (enhancer core plus its flanking sequences, ≈30 bp), with other enhancers, especially with that of a mouse immunoglobulin gene that has recently been shown to function in a B-lymphocyte-specific manner (32, 33); the BLV enhancer may thus bear some relation to the target cell (B lymphocyte) specificity of BLV. In this connection, it is interesting to note that the U3 regions of certain murine leukemia viruses have a regulatory sequence, presumably an enhancer element, capable of determining the target cell specificity of the virus (36, 37). However, it is yet to be determined whether the putative enhancer element of BLV possesses a tissue-specific "enhancer" activity.

The most striking structural feature of the BLV LTR was an exceptionally long R region. Correlated with this long R region, the putative poly(A) signal was located in the U3 region as far as 260 bp upstream from the poly(A) site [in retroviruses and eukaryotic genes the poly(A) signal is usually only 10–20 bp upstream of the poly(A) site; ref. 1]. Judging from the primary structure, transcription of the BLV genome would thus not terminate at the correct site. To resolve this problem, we have proposed that the 3′ terminus of the BLV RNA transcript forms a large secondary structure such that the entire R region becomes put aside as a hairpin structure, thereby making the two transcription termination signals close together (as near as 6 bp). Comparison of the proposed secondary structure with as many as 300 secondary structures formed by reshuffled sequences showed that the proposed structure was thermodynamically much more stable than almost all other secondary structures derived from the random sequences, strongly suggesting its biological significance. Probably, the large hairpin structure, if it is actually present *in vivo*, is formed rapidly on the nascent RNA chain, and immediately afterwards the resulting, closely located poly(A) signal and poly(A) site would serve as a normal transcription termination signal (or, in the more strict sense, as a recognition signal for RNA cleavage possibly preceding polyadenylylation; ref. 38).

HTLV LTR has also a poly(A) signal in an unusual location and can form a similar secondary structure, although this structure has been proposed for both viral RNA and proviral DNA (41). We do not know at present why only BLV and HTLV have such an exceptionally long R region and hence, such a "round-about" mode of transcription termination. It is possible, however, that their R region bears some implication in the unique biological features of these retroviruses—i.e., lack of viral expression in the natural target cells (14, 42), low efficiency of replication in heterologous cells (13, 43), and activation of viral expression after a short culture of the target cells (8, 42).

The structural features of BLV LTR presented here enable us to speculate on the kinship of BLV and HTLV, a possibility suggested previously from the appreciable homology of the NH_2-terminal sequences of their internal protein p24 (44) and the long strong-stop cDNA in both (22). In fact,

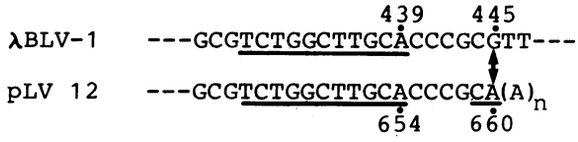

FIG. 6. Comparison of the nucleotide sequences around the poly(A) addition site in the two BLV molecular clones. The nucleotide sequences of the cDNA clone (pLV 12) are from ref. 39. The 11-bp nucleotides underlined closely match the consensus transcription termination signal of other retroviruses (23). The arrow indicates a substituted residue.

we compared the BLV LTR sequence with those of various retroviruses and found that only HTLV LTR (R region) has slight sequence homology with BLV LTR. The very low, but appreciable, sequence homology observed (at most 50%) indicates that BLV and HTLV are distantly related, presumably constituting a distinct group of mammalian retroviruses. Evolutionarily, it appears that BLV and HTLV branched away from each other much earlier than did several known murine and primate type C retroviruses, whose LTRs have >60% sequence homology with each other (4, 5).

Finally, we must refer to a paper (39) that we encountered during preparation of this manuscript, which reports the nucleotide sequence of the cloned 3'-terminal RNA region of BLV (cDNA clone). Although the BLV isolate examined (FLK-BLV; ref. 13) was different from ours and the boundary of its LTR was not clarified, the nucleotide sequences of the corresponding regions (U3 + R) are almost completely consistent (99%) with ours. However, an interesting difference is in the poly(A) addition site: as shown in Fig. 6, the G residue at position 445 in our clone is replaced by an A residue in the cDNA clone (position 660), thereby forming a C-A dinucleotide to which the poly(A) stretch is attached [in our clone, we tentatively regarded the dinucleotide C-A (positions 438–439) that preceded the G residue (or A residue in the cDNA clone) as the poly(A) site, because it constituted a consensus transcription termination signal of retroviruses (23)]. We do not know to which residue the poly(A) stretch actually attaches in our clone nor do we know whether microheterogeneity in the choice of poly(A) addition sites (40) takes place in the BLV system. However, if the poly(A) site is at the G residue in our clone (or the A residue in the cDNA clone), the poly(A) signal and the poly(A) site will be located at a more appropriate distance (12 bp apart) in the possible secondary structure that we have proposed.

Note Added in Proof. Following acceptance of this manuscript, Couez *et al.* (45) reported the LTR sequence of another BLV isolate.

We thank Dr. Masahiro Obata for kind advice in DNA sequencing and also Drs. Hiroshi Amanuma and Misao Onuma for many helpful discussions. This work was supported in part by grants from the Ministry of Education, Science and Culture, Japan, and the Tokushima Research Institute of Otuka Pharmaceutical Co., Ltd.

1. Temin, H. W. (1981) *Cell* **27**, 1–3.
2. Varmus, H. E. (1982) *Science* **216**, 812–820.
3. Khoury, G. & Gruss, P. (1983) *Cell* **33**, 313–314.
4. Lovinger, G. G., Mark, G., Todaro, G. J. & Schochetman, G. (1981) *J. Virol.* **39**, 238–245.
5. Lovinger, G. G. & Schochetman, G. (1980) *Cell* **20**, 441–449.
6. Khan, A. S. & Martin, M. A. (1983) *Proc. Natl. Acad. Sci. USA* **80**, 2699–2703.
7. Kugimiya, W., Ikenaga, H. & Saigo, K. (1983) *Proc. Natl. Acad. Sci. USA* **80**, 3193–3197.
8. Burny, A., Bruck, C., Chantrenne, H., Cleuter, Y., Dekegel, D., Ghysdael, J., Kettmann, R., Leclercq, M., Leunen, J., Mammerickx, M. & Portetelle, D. (1980) in *Viral Oncology*, ed. Klein, G. (Raven, New York), pp. 231–289.
9. Van Der Maaten, M. J., Miller, J. M. & Boothe, A. D. (1974) *J. Natl. Cancer Inst.* **52**, 491–497.
10. Gilden, R. V., Long, C. W., Hanson, M., Toni, R., Charman, H. P., Oroszlan, S., Miller, J. M. & Van Der Maaten, M. J. (1975) *J. Gen. Virol.* **29**, 305–314.
11. McDonald, H. C. & Ferrer, J. F. (1976) *J. Natl. Cancer Inst.* **57**, 875–882.
12. Deschamps, J., Kettmann, R. & Burny, A. (1981) *J. Virol.* **40**, 605–609.
13. Van Der Maaten, M. J. & Miller, J. M. (1976) *Bibl. Haematol.* **43**, 360–362.
14. Kettmann, R., Deschamps, J., Cleuter, Y., Couez, D., Burny, A. & Marbaix, G. (1982) *Proc. Natl. Acad. Sci. USA* **79**, 2465–2469.
15. Onuma, M., Sagata, N., Okada, K., Ogawa, Y., Ikawa, Y. & Oshima, K. (1982) *Microbiol. Immunol.* **26**, 813–820.
16. Kettmann, R., Deschamps, J., Couez, D., Claustriaux, J.-J., Palm, R. & Burny, A. (1983) *J. Virol.* **47**, 146–150.
17. Sagata, N., Ogawa, Y., Kawamura, J., Onuma, M., Izawa, H. & Ikawa, Y. (1983) *Gene* **26**, 1–10.
18. Maxam, A. & Gilbert, W. (1980) *Methods Enzymol.* **65**, 499–560.
19. Toh, H., Hayashida, H. & Miyata, T. (1983) *Nature (London)* **305**, 827–829.
20. Zuker, M. & Stiegler, P. (1981) *Nucleic Acids Res.* **9**, 133–148.
21. Needleman, S. B. & Wunsh, C. D. (1970) *J. Mol. Biol.* **48**, 443–453.
22. Seiki, M., Hattori, S. & Yoshida, M. (1982) *Proc. Natl. Acad. Sci. USA* **79**, 6899–6902.
23. Bonner, T. I., O'Connell, C. & Cohen, M. (1982) *Proc. Natl. Acad. Sci. USA* **79**, 4709–4713.
24. Fitzgerald, M. & Shenk, T. (1981) *Cell* **24**, 251–260.
25. Levinson, B., Khoury, G., Vande Woude, G. & Gruss, P. (1982) *Nature (London)* **295**, 568–572.
26. Luciw, P. A., Bishop, J. M., Varmus, H. E. & Capecchi, M. R. (1983) *Cell* **33**, 705–716.
27. Weiher, H., König, M. & Gruss, P. (1983) *Science* **219**, 626–631.
28. Dhar, R., McClements, W. L., Enquist, L. W. & Vande Woude, G. F. (1980) *Proc. Natl. Acad. Sci. USA* **77**, 3937–3941.
29. Tyndall, C., La Mantia, G., Thacker, C. M., Favaloro, J. & Kamen, R. (1981) *Nucleic Acids Res.* **9**, 6231–6250.
30. Fujimura, F. K., Deininger, P. L., Friedmann, T. & Linney, E. (1981) *Cell* **23**, 809–814.
31. Seif, I., Khoury, G. & Dhar, R. (1979) *Cell* **18**, 963–977.
32. Gillies, S. D., Morrison, S. L., Oi, V. T. & Tonegawa, S. (1983) *Cell* **33**, 717–728.
33. Banerji, J., Olson, L. & Schaffner, W. (1983) *Cell* **33**, 729–740.
34. Streeck, R. E. (1982) *Nature (London)* **298**, 767–769.
35. Hirose, S., Takeuchi, K. & Suzuki, Y. (1982) *Proc. Natl. Acad. Sci. USA* **79**, 7258–7262.
36. DesGroseillers, L., Rassart, E. & Jolicoeur, P. (1983) *Proc. Natl. Acad. Sci. USA* **80**, 4203–4207.
37. Chatis, P. A., Holland, C. A., Hartley, J. W., Rowe, W. P. & Hopkins, N. (1983) *Proc. Natl. Acad. Sci. USA* **80**, 4408–4411.
38. Montell, C., Fisher, E. F., Caruthers, M. H. & Berk, A. J. (1983) *Nature (London)* **305**, 600–605.
39. Tsimanis, A., Bichko, V., Dreilina, D., Meldrais, J., Lozha, V., Kukaine, R. & Gren, E. (1983) *Nucleic Acids Res.* **11**, 6079–6087.
40. Sasavage, N. L., Smith, M., Gillam, S., Woychik, R. P. & Rottman, F. M. (1982) *Proc. Natl. Acad. Sci. USA* **79**, 223–227.
41. Seiki, M., Hattori, S., Hirayama, Y. & Yoshida, M. (1983) *Proc. Natl. Acad. Sci. USA* **80**, 3618–3622.
42. Hinuma, Y., Gotoh, Y., Sugamura, K., Nagata, K., Goto, T., Nakai, M., Kamada, N., Matsumoto, T. & Kinoshita, K. (1982) *Gann* **73**, 341–344.
43. Weiss, R. (1982) in *RNA Tumor Viruses*, eds. Weiss, R., Teich, N., Varmus, H. & Coffin, J. (Cold Spring Harbor Laboratory, Cold Spring Harbor, NY), pp. 1205–1281.
44. Oroszlan, S., Sarngadharan, M. G., Copeland, T. D., Kalyanaraman, V. S., Gilden, R. V. & Gallo, R. C. (1982) *Proc. Natl. Acad. Sci. USA* **79**, 1291–1294.
45. Couez, D., Deschamps, J., Kettmann, R., Stephens, R. M., Gilden, R. V. & Burny, A. (1984) *J. Virol.* **49**, 615–620.