Volume 11 Number 17 1983 Nucleic Acids Research

The structure of cloned 3'-terminal RNA region of bovine leukemia virus (BLV)

A.Tsimanis, V.Bichko, D.Dreilina*, J.Meldrais*, V.Lozha*, R.Kukaine* and E.Gren

Institute of Organic Synthesis, and *Institute of Microbiology, Latvian SSR Academy of Sciences, Riga, USSR

Received 26 July 1983; Accepted 12 August 1983

ABSTRACT

cDNA synthesized on the bovine leukemia virus RNA template has been cloned in the pBR322 Pst I site. Colony hybridization with BLV RNA fragments and oligo(dT) has revealed a clone with cDNA insert containing 660 3'-terminal nucleotides of the BLV genome. The nucleotide sequence of the insert corresponding to U3 and R regions of the long terminal repeats (LTR) of viral genome has been determined. BLV U3, like U3 of other retroviruses, presumably contains promoter. The unusually long R region (about 230 bp), a certain homology with ATLV U3-R and some other structural features allow to group BLV LTR together with ATLV LTR in a separate class of retroviral LTR.

INTRODUCTION

Long terminal repeats (LTR) of retroviruses are a typical feature of this class of oncogenic viruses playing a pivotal role in replication, integration and expression of the viral genome. In the provirus state the LTR flanking the viral genome on both sides are nearly or completely identical. The LTR of various retroviruses differ both in their length (320- 1330 bp) and in nucleotide sequence, but possess common structural (U3-R-U5, terminal IR) and functional (promoter, RNA termination site, etc.) features; for review see [1-3]. Among different retroviruses, a more or less pronounced structural homology is observed in regard of their LTR that can be divided into two classes depending on the mutual location of promoter and RNA termination site. The length of U3, R and U5 regions also belongs to the typical features of LTR. In all studied LTR the most extensive region is U3, region R being the least extensive one. ATLV closely associated with a specific type of human T-cell leukemia is an exception [4], the LTR of which is

characterized by a substantially longer R (228 bp) than commonly found (16-79 bp). In all likelihood, ATLV and closely related to it HTLV [5], whose LTR structure has not been published as yet, form a distinct class of retroviruses. BLV, a virus causing bovine B-cell leukemia also belongs to this class on the basis of preliminary data [4,5].

Studying the structure of cloned BLV cDNA corresponding to 660 nucleotides of the 3'-terminal region of the viral genome we observed a certain similarity with the appropriate ATLV RNA region supporting the placement of BLV in the class of ATLV and HTLV retroviruses.

METHODS

Preparation of BLV RNA

BLV was produced by monolayer cultures of BLV-infected foetal lamb kidney (FLK) cells [6]. BLV RNA was extracted from the viral particles purified in sucrose gradient using the phenol-chloroform procedure [7] with two cycles of oligo(dT)--cellulose chromatography [8].

Synthesis of cDNA

Single-stranded cDNA was synthesized in 0.4 ml reaction mixture containing 50 mM Tris-HCl (pH 8.3) at 42°, 10 mM $MgCl_2$, 60 mM KCl, 5 mM DTT, 1 mM each dATP, dCTP, dTTP, 0.5 mM ^{3}H-dGTP (0.8 Ci/mmol), 8 µg oligo$(dT)_{12-18}$, 1.25 µg BLV RNA and 50 U AMV reverse transcriptase for 1 hr at 42°. The reaction was stopped by adjusting the mixture to 10 mM EDTA, RNA was hydrolyzed at 60° for 20 min with 0.3 M NaOH and after neutralization the mixture was passed through a Sephadex G-50 column in 10 mM Tris-HCl pH 7.5, 0.25 M NaCl. cDNA was precipitated by the addition 2.5 volumes of ethanol.

cDNA (approximately 0.5 µg) was made double-stranded with E.coli DNA polymerase I (Klenow fragment) by the procedure described in [9]. After phenol extraction, gel filtration and ethanol precipitation the single-stranded ends and loop structures of double-stranded cDNA were digested with 100 U S1 nuclease at 37° for 30 min in a 1 ml reaction mixture containing 0.3 M NaCl, 0.3 M NaOAc pH 4.5, 0.003 M $ZnSO_4$. After phenol--chloroform extraction and gel filtration the double-stranded

cDNA was precipitated with ethanol.

Construction of Recombinant Plasmid DNA

S1 digested double-stranded cDNA was tailed with about 20 dCMP residues in a 20 µl reaction mixture, containing 0.2 M K-cacodylate pH 6.8, 0.01 mM $ZnSO_4$, 2 mM ß-mercaptoethanol, 1 mM $CoCl_2$, 0.2 mM dCTP and 100 U/ml terminal deoxynucleotidyl transferase (TdT). After 5 min at 37° the reaction was stopped and the DNA was concentrated by ethanol precipitation [10].

Plasmid pBR322 was linearized with Pst I and dGMP-tailed in a 0.125 ml reaction mixture containing 0.2 M K-cacodylate pH 6.8, 4 mM $MgCl_2$, 0.01 mM $ZnSO_4$, 2 mM ß-mercaptoethanol, 1 mM $CoCl_2$, 0.2 mM dGMP and 400 U/ml TdT. About 15 dGMP residues were added to the 3'-end after incubation at 37° for 5 min.

An equimolar amount of both tailed dsDNA and plasmid was annealed [11]. The annealed product was used to transform E.coli strain RR1 as described in [12].

Screening of Recombinant cDNA Clones

Colony hybridization was carried out according to [13,14] by using partially fragmented ^{32}P-labelled BLV RNA and ^{32}P-labelled oligo$(dT)_{12-18}$ or oligo$(dG)_{12-18}$ as a probe.

For rapid screening, plasmid DNA was prepared from 5 ml culture as described in [15].

Physical Mapping of Plasmid DNAs

Recombinant plasmid DNA and DNA fragments were digested with different combinations of restriction enzymes. DNA digests were electrophoresed on either 0.8-1.6% agarose or 3.5-20% gradient polyacrylamide gels as described in [16]. Restriction endonuclease digestion was performed under the condition recommended by suppliers. DNA restriction fragments corresponding to the ends of cloned BLV cDNA insert were determined by blot hybridization of electrophoresed DNA with oligo$(dT)_{12-18}$ and oligo$(dG)_{12-18}$ [14]. Recovery of DNA fragments from agarose gels was performed by electroelution into DE-81 paper [17].

Sequence Determination Procedures

BLV cDNA was subcloned using specialized vectors M13 mp7, M13 mp8 and M13 mp9 [18], and the nucleotide sequence was determined by the dideoxy chain termination method [19] with some modifications. Sequence ladders were displayed on 8% polyacryl-

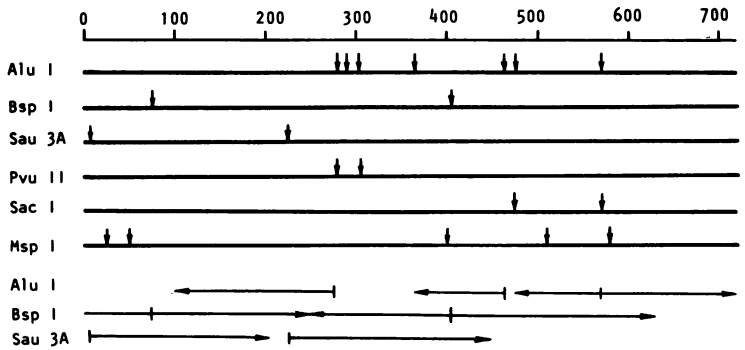

Fig.1. Restriction map of the plasmid pLV 12 insert. Horizontal arrows show the strategy of DNA sequencing.

amide - 8 M urea 30 x 40 x 0.04 cm gels (acrylamide : bisacrylamide - 20 : 1) during 2-5 hr at 40 mA. Radioautography was performed at -70° during 10-48 hr.

RESULTS

Poly(A)-RNA isolated from purified BLV virions served as template for the synthesis of cDNA in a reverse transcription system. Utilizing the 3'-terminal part of single-stranded cDNA as a primer for DNA polymerase (Klenow fragment), double-stranded cDNA was synthesized. Following S1 nuclease treatment, the double-stranded cDNA was inserted into the Pst I site of pBR322 by using oligo(dG : dC)-connector technique.

After transformation of 500 Tc^rAp^s colonies were obtained and screened for the presence of BLV sequences using 5'-^{32}P-labelled partially alkali-fragmented RNA as a hybridization probe, prepared from BLV RNA purified in a sucrose gradient.

Twenty-five colonies found positive by the RNA hybridization test were further screened for poly(dA) tract by hybridization with 5'-^{32}P-oligo$(dT)_{12-18}$. Among the poly(dA) containing clones, plasmid pLV 12 (insert of 718 bp) was selected for structural studies. Another plasmid, pLV 10 (insert length - 1070 bp), containing a cDNA fragment overlapping with plasmid pLV 12 insert over 200 bp, was also found.

The restriction map of plasmid pLV 12 for Alu I, Sau 3A, Bsp I , Pvu II, Sac I is illustrated in Fig.1. The same figure

```
  1   TCCCCGCGAT CGACCTATTC CTAACCGGTC CCCCTTCCCC ATGCGACCGG
 51   TTAGACGTAT GGTCCAGTCC TCAGGCCTTA CAGCGCTTCC TTCATGACCC
101   TACGCTAACC TGGTCCGAAT TAGTTGCTAG CAGAAAAATA AGACTTGATT
151   CCCCCTTAAA ATTACAACTG CTAGAAAATG AATGGCTCTC CCGCCTTTTT
201   TGAGGGGGAA TCATTTGTAT GAAAGATCAT GCCGACCTAG GCGCCGCCAC
251   CGCCCCGTAA ACCAGACAGA GACGTCAGCT GCCAGAAAAG CTGGTGACGG
301   CAGCTGGTGG CTAGAATCCC CGTACCTCCC CAACTTCCCC TTTCCCGAAA
351   AATCCACACC CTGAGCTGCT GACCTCACCT GCTGATAAAT TAATAAAATG
401   CCGGCCCTGT CGAGTTAGCG GCACCAGAAG CGTTCTTCTC CTGAGACCCT
451   CGTGCTCAGC TCTCGGTCCT GAGCTCTCTT GCTCCCGAGA CCTTCTGGTC
501   GGCTATCCGG CAGCGGTCAG GTAAGGCAAA CCACGGTTTG GAGGGTGGTT
551   CTCGGCTGAG ACCACGCGAG CTCTATCTCC GGTCCTCTGA CCGTCTCCAC
601   GTGGACTCTC TCCTTTGCCT CCTGACCCCG CGCTCCAAGG GCGTCTGGCT
651   TGCACCCGCA(A)n
```

Fig.2. Nucleotide sequence of BLV cDNA U3-R region. "Hogness-Goldberg box" and "CAAT-box" are framed; underlined "CA" sequence is the polyadenylation site; horizontal arrows represent direct and inverted repeats. Putative mRNA cap sites are marked with asterisks.

shows the sequencing strategy for pLV 12 plasmid insert based on the mapping data. The nucleotide sequence of the insert in all subregions was determined for DNA fragments obtained with two restriction enzymes, 80 per cent of the nucleotide composition having been determined by sequencing both DNA strands. The results of sequencing are consistent with the physical map.

The nucleotide sequence determined for BLV cDNA insert in pLV 12, consisting of 660 bp and a poly(dA) tract of about 30 dA bases at the 3'-end, corresponds to a major part of BLV LTR: U3 and R regions (Fig.2).

DISCUSSION

The sequence in Fig.2. consisting of 660 bp corresponds to the coding strand of BLV RNA and belongs to the 3'-terminal region of virion RNA. This is confirmed by the presence of a poly(dA) tract at one end of the insert within recombinant pLV 12 plasmid. Two estimates of BLV LTR length are available

in the literature - 750 bp [4] and around 600 bp [20]. If we take the length of BLV LTR equal to 750 bp and that of R+U5 (strong-stop viral cDNA) equal to 320 bp, as indicated in [4], the length of the U3 region would be 430 bp. Assuming G_{430} or G_{427} as RNA transcript initiation site (cap site) from LTR promoter (see below), the beginning of the sequence presented in Fig.2 should coincide with the left-hand border of U3. However, the given structure lacks, close to its beginning, the universal TG sequence serving as U3 border in all known viral LTR. Consequently, the structure in Fig.2 either fails to include the complete U3 start or our estimations are insufficiently accurate and U3 begins at one of the intrinsic TG (not earlier than at nucleotide 42). In the latter case, it appears more probable that the beginning of U3 coincides with TG_{147} or TG_{184}, since they are preceded by a purine-rich tract characteristic of the border region between U3 and the intrinsic viral sequence in many LTR [3]. If this assumption is correct the length of BLV LTR must be approximately 601 (604) bp or 564 (567) bp,which is closer to the value of 600 bp from [20].

The above nucleotide sequence contains no extensive open reading frames coding for possible virus-specific polypeptides. The same is true for the LTR of other retroviruses, except for MMTV [21].

According to our calculations, BLV has an unexpectedly long R (230 or 233 bp) reminding in this respect ATLV (R - 228 bp [4]) which is in marked contrast with other retroviruses containing a much shorter R of 16-79 bp [1,3]. This is not the only structural similarity feature shared by the U3-R region in the two retroviruses (BLV and ATLV). The nucleotide sequence itself is also somewhat homologous, although the homology is not so manifest. As in the case of ATLV, there is no $A^{A}_{G}TAAA$ sequence before the polyadenylation site at the end of R. The latter sequence is typical of all retroviruses and of a great many eukaryotic genes [1] and serves as a signal for mRNA polyadenylation. In fact, the structure of the polyadenylation site and the flanking nucleotides in BLV is similar to consensus $^{C}_{T}TTTGCN^{G}_{C}TTG\underline{CA}$ [22]. In this respect, BLV is unlike

ATLV containing no polyadenylation site CA characteristic of many LTR [4].

The LTR of retroviruses in the U3 region carries a promoter that structurally resembles eukaryotic promoters recognized by RNA polymerase II. An AT-rich tract ATAAATTAATAAAA can be found in the BLV U3 sequence at positions 385-398, part of which can be assigned to the Hogness-Goldberg box, consensus $\begin{matrix}GT\\CA\end{matrix}AT\begin{matrix}TT\\AA\end{matrix}AAG$ [1] or $TATA\begin{matrix}A\\T\end{matrix}A\begin{matrix}T\\A\end{matrix}$ [23]. Transcription initiation site, or cap site, is located 25-30 nucleotides downstream of the Hogness-Goldberg box, which in BLV may be G_{430} or G_{427}. Another structural determinant of promoter is the CAAT box (consensus - GGPyCAATCT) , generally 70-90 nucleotides upstream of the cap site [24]. Although a similar sequence is not so readily identifiable at the appropriate location of the BLV U3 region, the CCAACTT located some 90 nucleotides upstream from the cap site may qualify, to a certain extent, as a candidate for the CAAT box.

Still another regulatory element dramatically affecting transcription efficiency has been found in the LTR of some retroviruses and in the case of several other eukaryotic promoters. This is the so-called enhancer (or activator), a region located usually 100-200 or more nucleotides upstream of the cap site, whose structure is still poorly understood [25-27]. Enhancer usually resides in two closely located direct repeats (DR). The consensus $(G)TGG\begin{matrix}AAA\\TTT\end{matrix}(G)$ structure postulated for them [25] is not always detectable in the U3 of retroviruses. However, the presence of direct repeats over the wide sequence range prior to promoter is believed to indicate the occurrence of this regulatory element. One DR comprising 10 nucleotides having some similarity features with the postulated consensus is located in the BLV U3 region over 130-140 bp upstream of the cap site. Actually, tandem DR in BLV U3 occur at positions prior to the promoter, as is obvious from Fig.2. In ATLV U3, one tandem DR is located at a much greater distance from the promoter [4].

Two inverted repeats (IR) of medium size are detectable in BLV U3. Such IR have been discovered in the U3 sequences of several retroviruses 80 to 150 bp upstream from the cap site,

as well as prior to the eukaryotic promoters [28]. The significance of IR is not clear. IR as well as enhancers possibly participate in the nucleosome phasing by locating promoter sequences in positions accessible for efficient transcription.

Consequently, the U3-R regions of BLV LTR contain both the structural features common for retroviruses and certain distinctive characteristics. The latter permit to place BLV and ATLV (possibly HTLV, too) in a special class of retroviruses, at least as far as their U3-R structure is concerned. Some avian and mammalian retroviruses in an integrated provirus state activate the transcription of the flanking cellular sequences from the LTR promoter. This is not observable in the case of BLV [29]. It is not clear as yet whether this difference is due to a somewhat different structure of their LTR. At least, the functional activity of BLV LTR promoter has not been directly confirmed.

REFERENCES

1 Temin, H.M. (1981) Cell 27, 1-3.
2 Temin, H.M. (1982) Cell 28, 3-5.
3 Varmus, H.E. (1982) Science 216, 812-820.
4 Seiki, M., Hattori, S. and Yoshida, M. (1982) Proc. Natl. Acad. Sci. USA 79, 6899-6902.
5 Oroszlan, S., Sarngadharan, M.G., Copeland, T.D., Kalyanaraman, V.S., Gilden, R.V. and Gallo, R.C. (1982) Proc. Natl. Acad. Sci. USA 79, 1291-1294.
6 Van der Maaten, M.J. and Miller, J.M. (1976) Bibliogr. Haematol. 43, 360-362.
7 Nagata, S., Taira, H., Hall, A., Johnsrud, L., Streuli, H., Escödi, J., Boll, W., Cantell, K. and Weissmann, C. (1980) Nature 284, 316-320.
8 Aviv, H. and Leder, P. (1972) Proc. Natl. Acad. Sci. USA 69, 1408-1412.
9 Kurtz, D.T. and Nicodemus, C.F. (1981) Gene 13, 145-152.
10 Roychoudhury, R. and Wu, R. (1980) in Methods in Enzymology, Grossman, L. and Moldave, K. Eds., Vol.65, pp.43-62, Academic Press, New York.
11 Villa-Komaroff, L., Efstratiadis, A., Broome, S., Lomedico, R., Tizard, R., Naber, S.P., Chick, W.L. and Gilbert, W. (1978) Proc. Natl.Acad. Sci. USA 75, 3727-3731.
12 Dagert, M. and Ehrlich, S.D. (1979) Gene 6, 23-28.
13 Gergen, J.P., Stern, R.H. and Wensink, P.C. (1979) Nucl. Acids Res. 7, 2115-2136.
14 Wallace, R.B., Schaffer, J., Murphy, R.F., Bonner, J., Hirose, T. and Itakura, K. (1981) Nucl. Acids Res. 9, 879-894.

15 Birnboim, H.C. and Doly, J. (1979) Nucl. Acids Res. 7, 1513-1523.
16 Helling, R.B., Goodman, H.M. and Boyer, H.W. (1974) J.Virol. 14, 1235-1244.
17 Dretzen, G., Bellard, M., Sassone-Corsi, P. and Chambon, P. (1981) Anal. Biochem. 112, 295-298.
18 Messing, J. and Vieira, J. (1982) Gene 19, 269-276.
19 Sanger, F., Nicklen, S. and Coulson, A.R. (1977) Proc. Natl. Acad. Sci. USA 74, 5463-5467.
20 Kashmiri, S.V.S., Medhi, R. and Ferrer, J.F. (1983) J. Virol. 45, 1172-1176.
21 Kennedy, N., Knedlitschek, G., Groner, B., Hynes, N.E., Herrlich, P., Michalides, R. and van Ooyen, A.J.J. (1982) Nature 295, 622-624.
22 Bonner, T.I., O'Connell, C. and Cohen, M. (1982) Proc. Natl. Acad. Sci. USA 79, 4709-4713.
23 Corden, J. Wasylik, B., Buchwalder, A., Sassone-Corsi, P., Kedinger, C. and Chambon, P. (1980) Science 209, 1406-1414.
24 Chambon, P. and Breathnach, R. (1981) in Ann. Rev. Biochem., Snell, E.S., Boyer, P.D., Meister, A. and Richardson, C.C. Eds., Vol.50, pp.349-383.
25 Weiher, H., König, M. and Gruss, P. (1983) Science 219, 626-631.
26 McKnight, S.L. and Kingsbury, R. (1982) Science 217, 316-324.
27 Levinson, B., Khoury, G., Woude, G.V. and Gruss, P. (1982) Nature 295, 568-572.
28 Clark, S.P., and Mak, T.W. (1982) Nucl. Acids Res. 10, 3315-3330.
29 Kettmann, R., Deschamps, J., Cleuter, Y., Couez, D., Burny, A. and Marbaix, G. (1982) Proc. Natl. Acad. Sci. USA 79, 2465-2469.