Nucleotide Sequence Analysis of the Long Terminal Repeat of Murine Virus-Like DNA (VL30) and Its Adjacent Sequences: Resemblance to Retrovirus Proviruses

AHUVA ITIN AND ELI KESHETr

Department of Virology, The Hebrew University—Hadassah Medical School, Jerusalem, Israel

Received 21 March 1983/Accepted 7 June 1983

VL30 DNA represents a retrovirus-like multigene family ubiquitous in all murine cells. For example, over 100 VL30 units are dispersed throughout the genomes of certain strains of Mus musculus (6). The origin of VL30 genetic information is not known since no homology has been detected between the nucleic acids of VL30 and those of any retrovirus tested, and no proteins encoded by VL30 DNA have as yet been identified. The list of retrovirus-like features displayed by VL30 elements includes the following. 30S RNA constitutively transcribed from certain VL30 copies is efficiently packaged in C-type virions. 30S RNA that is consequently rescued from the cell as a pseudovirion forms a dimer structure and can utilize an endogenous primer for reverse transcription (1, 5, 13). Moreover, VL30 genetic information can be transmitted to other cells via pseudovirion infection (12). We have previously shown that VL30 DNA shares a basic structure with retrovirus proviruses. Most notably retrovirus proviruses possess a long terminal direct repeat (LTR) of several hundred base pairs (7). LTRs of retroviruses bear a unique relationship with respect to their RNA templates, namely, the sequence organization U3-R-U5, where U3 and U5 designate unique sequences derived from the 3' and 5' ends of viral RNA, respectively, and R designates a short, terminally redundant sequence present at both termini of viral RNA. The proximity of the DNA synthesis priming site to the 5' end of viral RNA and the reverse transcription "jumps" are key features in generating this unique sequence arrangement. It is not known whether VL30 LTRs are also generated through a similar mechanism.

The integration of retrovirus proviruses is a highly ordered process. The specificity in provirus integration is best manifested by the characteristic sequence arrangement at the integration junctions. Specifically, the site of union with host DNA is precisely 2 base pairs from the ends of linear proviral DNA, and a short host sequence at the integration site is duplicated during integration (3, 14, 17, 18). The mechanism of VL30 integration is not known.

We reasoned, therefore, that determination of the primary nucleotide sequences and adjacent nucleotide sequences of VL30 LTRs might contribute to an understanding of the genetic relatedness of VL30 DNA to DNA of proviruses or other cellular elements. Specifically, we expected the nucleotide sequences at the inner boundaries of both 5' and 3' LTRs to be of particular importance since these are the putative primer binding sites for the minus and plus strands, respectively, of proviral DNA synthesis. In addition, we anticipated that the knowledge of the sequence arrangement at the junctions with mouse flanking sequences would provide insight into the mechanism of VL30 integration.
We have now determined the primary nucleotide sequences and the adjacent sequences of both LTR units of a randomly chosen VL30 copy. For this purpose we used a previously obtained phage clone derived from a BALB/c mouse embryonic library that contains a complete VL30 unit (8). This clone was designated clone 3. We have previously reported sequence heterogeneity within the VL30 family (6, 8). This heterogeneity also involves the LTRs (unpublished data). It should be noted, therefore, that the LTR whose sequence analysis is shown below was arbitrarily chosen from the collection of cloned VL30 units. Heteroduplex analysis was used to determine the approximate locations of the LTR units (7). The transcriptional orientation of VL30 clone 3 DNA was determined as follows. 30S RNA was partially fragmented, and polyadenylic acid-containing RNA fragments were selected by chromatography on an oligodeoxythymidylic acid-cellulose column. Fractions of polyadenylated RNA of increasing size were pooled from a preparative agarose gel and used as templates for randomly primed cDNA synthesis (16). Blots containing clone 3 DNA restriction fragments were challenged with these cDNA probes. Only DNA fragments residing at one side of clone 3 DNA (excluding the LTR-containing fragments) were detected by cDNA probes prepared from relatively short polyadenylated RNA (data not shown). This end was thus defined as corresponding to the 3’ end of 30S RNA (Fig. 1).

The sequencing procedure was initiated by digestion with a restriction enzyme that cut once within both LTR units and by determination of the sequences outwards in both directions beyond the points where the sequences of the 5’ and 3’ LTRs diverged. The points of sequence divergence defined the LTR boundaries (Fig. 1).

The primary nucleotide sequence of the 5’ LTR unit is shown in Fig. 2. The length of the LTR was 435 nucleotide base pairs. Also indicated in Fig. 2 are the minor sequence differences observed between the 5’ and 3’ LTR units—an additional CG at position 330 of the 3’ LTR and five single base pair differences clustered towards the 3’ end of the LTR. VL30 LTR possessed a perfect inverted repeat of five bases with the sequence 5’-TGAAA-3’ at its 5’ and 3’ termini (underlined sequence in Fig. 2).

The sequence arrangement at the LTR junctions is illustrated in Fig. 3. Analysis of these sequences revealed the following.

(i) Inner junction of the 5’ LTR. Immediately adjacent to the 3’ end of the 5’ LTR was a stretch of 18 bases (5’-TGAGCTGGTCCGGTATACG3’) which was complementary to the 3’ sequence of prolyl-tRNA (11). Retroviral genomes are known to contain at an identical position (i.e., at the 3’ boundary of the 5’ LTR) a tRNA binding site that serves as a primer for the synthesis of the minus DNA strand. This result suggested, therefore, that as for other mammalian retroviruses, tRNA may serve as a primer for reverse transcription of VL30 DNA. In fact, the VL30 sequence given above is identical to the 18 bases that follow the 5’ LTR of murine leukemia virus (18), spleen necrosis virus (14), simian sarcoma virus (2), and feline sarcoma virus (4).

(ii) Inner junction of the 3’ LTR. Synthesis of the proviral plus strand begins at a position that defined the 5’ boundary of U3. Although the primer for this event has not been identified, the priming site is always spanned by a purine-rich...

FIG. 1. Strategy for sequencing the LTR of VL30 DNA clone 3. The 5’→→→3’ orientation refers to the transcriptional orientation of VL30 DNA determined as explained in the text. To facilitate preparative resolution of end-labeled DNA fragments, we subcloned the fragments containing 5’ and 3’ LTRs in pBR322. Restriction enzyme sites used for sequencing are indicated by the vertical bars. Restriction sites with both protruding 5’ ends and blunt ends were labeled with the aid of the Klenow fragment of DNA polymerase I and the appropriate [α-32P]deoxynucleoside triphosphates. The sequencing procedure was that of Maxam and Gilbert (10). bp, Base pairs.
tract ending in AATG (19). Adjacent to the 5' boundary of the 3' LTR of VL30 clone 3 was the sequence 5'-AAGAAGAAGTGGGGATT-GATG-3'. This purine-rich tract (17 of 21 nucleotides) was followed by AATG. Thus, there exists a strong resemblance between this structure and the structures located at identical positions in retrovirus proviruses.

(iii) Junctions with mouse flanking sequences. The presumed locations of the joints between VL30 DNA and flanking DNA are indicated by the arrows in Fig. 3. At the site of union with mouse DNA was the VL30 sequence (5')TG . . . CA(3'). The same sequence was universally detected at the site of union of proviral DNA with host DNA (17). Integration of retrovirus proviruses is accompanied by the loss of two nucleotides from the predicted ends of the linear DNA. The structure shown in Fig. 3 can also be interpreted to be the result of integration from within, namely, the loss upon integration of the dinucleotide AA from the 5' end of the 5' LTR and the loss of the complementary dinucleotide TT from the 3' end of the 3' LTR. The LTRs shown in Fig. 3 are boxed in a fashion that illustrates this possibility. It should be pointed out, however, that establishing that the loss of a dinucleotide from each end of the LTR upon integration occurs awaits the sequencing of the unIntegrated precursor of VL30 DNA. VL30 DNA was flanked by a short tetranucleotide direct repeat (AAGA). During retroviral integration a short host sequence at the integration site is duplicated so that the duplicated sequence appears as a direct repeat flanking the provirus. It is reasonable to assume that the sequence AAGA flanking the VL30 DNA was also duplicated during the process of VL30 integration. Unfortunately, since VL30 is a natural resident of the cell, the nucleotide sequence spanning the point of insertion before VL30 DNA integration was not available to establish this point.

The question of whether VL30 LTRs are distinguished by the general structure U3-R-U5 is of particular interest. At the moment there is no evidence that VL30 RNA (30S RNA) is terminally redundant. The data presented above concerning the location of a tRNA^\text{pre} binding site suggested that reverse transcription was initiated close to the 5' end of the 30S RNA (and no further than 435 nucleotide base pairs from this end). This conclusion should be viewed in conjunction with a previous observation that a very short (<1 kilobase) polyadenylated VL30 RNA hybridized to the 5' LTR-containing DNA fragment (7). This suggests that the LTRs were assembled from VL30 sequences derived from both the 3' and 5' ends of 30S RNA.

FIG. 2. Nucleotide sequence of the LTR of VL30 DNA clone 3. The sequence shown is of the 5' LTR. Residues of the 3' LTR that were different are indicated underneath. The terminally inverted repeated sequence is underlined.

FIG. 3. Nucleotide sequences adjacent to the LTR inner and outer junctions. The presumed joints with mouse flanking DNA are indicated by arrows. The rationale for boxing the LTRs in the manner shown is explained in the text. Also indicated in the figure are a short direct repeat of flanking DNA (thin line), an 18-base-long sequence complementary to tRNA^\text{pre} (heavy line), and a purine-rich tract adjacent to the inner boundary of the 3' LTR (broken line).
Analysis of retroviral LTRs suggests that LTRs provide functions fundamental to the expression of eucaryotic genes, namely, promotion, initiation, and polyadenylation of transcripts (3, 10, 15, 17, 18). Some LTRs have been shown to have the additional capacity of transcriptional enhancement (9). There are indications that only a fraction of VL30 units are transcriptionally active, and it is not known whether the particular VL30 copy analyzed here expresses 30S RNA. Consequently, it is unclear whether sequences that resemble a TATA box or polyadenylation signals (e.g., TTAAAA at position 231 and ATTAAA at position 334) are indeed functionally significant. To select transcriptionally active VL30 LTRs we are currently cloning VL30 DNA through a protocol that involves the rescue of 30S RNA expressed in mouse cells as pseudovirion RNA, the subsequent infection of rat cells, and the isolation of preintegration VL30 DNA intermediates. Cloned VL30 LTRs obtained in this manner would be a more appropriate subject for the analysis of transcriptional control signals, including a possible transcription enhancer.

We have noticed that there is a great deal of sequence variation between the LTR of VL30 clone 3 and LTRs associated with some other VL30 units that we have cloned. The sequences residing at the 3' half of each LTR are particularly different. A comparative sequence analysis is currently being carried out to establish the exact relationships among several VL30-linked LTRs.

We have carried out a computer-aided comparison between the LTR sequence described here and the LTR of murine leukemia virus (18). The two LTR units are grossly nonhomologous. However, certain short DNA stretches show some resemblance. For example, sequences at the 3' ends of both LTRs are almost identical (the LTRs share 11 of the 12 3' terminal nucleotides).

In conclusion, the retrovirus-like nature of VL30 genetic information is strongly reinforced by the sequencing data presented here. Specifically, the detailed sequence arrangement at the junctions with mouse flanking sequences suggests that VL30 DNA uses the same strategy for integration as do proviruses and transposable elements. In addition, the nature of the sequences adjacent to the inner boundaries of 5' and 3' LTRs and their relationships to the primer binding site of the minus and plus strands, respectively, of retroviral DNA strongly suggest a similarity in mechanisms used for LTR synthesis.

This work was supported by a grant from the Israel-U.S.A. Binational Science Foundation.

LITERATURE CITED