JOURNAL OF VIROLOGY, Feb. 1982, p. 542–556
0022-538X/82/020542-15$02.00/0

Vol. 41, No. 2

Long Terminal Repeat of Murine Retroviral DNAs: Sequence Analysis, Host-Proviral Junctions, and Preintegration Site

CHARLES VAN BEVEREN,[1] ELAINE RANDS,[2] SISIR K. CHATTOPADHYAY,[2] DOUGLAS R. LOWY,[2] AND INDER M. VERMA[1]*

Tumor Virology Laboratory, The Salk Institute, San Diego, California 92138,[1] and Dermatology Branch, Laboratory of Tumor Virus Genetics, National Cancer Institute, Bethesda, Maryland 20205[2]

Received 27 July 1981/Accepted 25 September 1981

The nucleotide sequence of the long terminal repeat (LTR) of three murine retroviral DNAs has been determined. The data indicate that the U_5 region (sequences originating from the 5′ end of the genome) of various LTRs is more conserved than the U_3 region (sequences from the 3′ end of the genome). The location and sequence of the control elements such as the 5′ cap, "TATA-like" sequences, "CCAAT-box," and presumptive polyadenylic acid addition signal AATAAA in the various LTRs are nearly identical. Some murine retroviral DNAs contain a duplication of sequences within the LTR ranging in size from 58 to 100 base pairs. A variant of molecularly cloned Moloney murine sarcoma virus DNA in which one of the two LTRs integrated into the viral DNA was also analyzed. A 4-base-pair duplication was generated at the site of integration of LTR in the viral DNA. The host-viral junction of two molecularly cloned AKR-murine leukemia virus DNAs (clones 623 and 614) was determined. In the case of AKR-623 DNA, a 3- or 4-base-pair direct repeat of cellular sequences flanking the viral DNA was observed. However, AKR-614 DNA contained a 5-base-pair repeat of cellular sequences. The nucleotide sequence of the preintegration site of AKR-623 DNA revealed that the cellular sequences duplicated during integration are present only once. Finally, a striking homology between the sequences flanking the preintegration site and viral LTRs was observed.

During the life cycle of retroviruses, the viral genomic RNA is transcribed to DNA, some of which is integrated into the host chromosome (4). This crucial function carried out by viral encoded reverse transcriptase is obligatory for establishment of infection (42). Although the mechanism of reverse transcription is quite complex (16) and not fully understood, several forms of double-stranded viral DNA including linear, circular, and supercoiled DNAs have been reported in infected cells (41). Both the in vivo- and in vitro-synthesized viral DNAs have two types of genome-length molecules (2, 6, 15, 20, 32): (i) those that contain 5′-end genomic sequences (U_5) repeated at their 3′ end (U_3), forming a structure 5′ . . . $U_3U_5$3′ and (ii) those that, in addition to having 5′ genomic RNA sequences repeated at their 3′ end, also contain 3′ genomic RNA sequences repeated at their 5′ end, forming a structure 5′-U_3U_5 . . . U_3U_5-3′. However, analysis of the integrated viral DNA shows only structures containing U_3U_5 . . . U_3U_5 sequences (21, 31). The U_3U_5 unit is referred to as the long terminal repeat (LTR).

The biological function of the LTR is not clearly understood, but its structure warrants several speculations. The LTR contains control elements for both the promotion (14a, 44) and termination of viral RNA transcripts and may mediate or facilitate the integration of viral DNA into the host chromosomal DNA. Nucleotide sequence analysis of LTRs from several proviral DNAs (11, 19, 25, 34, 38) has shown: (i) direct duplication of sequences at the termini of viral DNA, i.e., LTRs; (ii) inverted repeats at the termini of each LTR; and (iii) direct repeat of adjacent cellular sequences at both termini of proviral DNA. These structural attributes of LTRs are reminiscent of structures associated with bacterial transposons (7, 8) TY1 elements of yeasts (14), and the *copia* element in *Drosophila* (13). In this manuscript, we have analyzed and compared the nucleotide sequences of several LTRs from murine retroviral DNA. The host-viral junctions of two molecularly cloned, integrated, infectious AKR-murine leukemia virus (MLV) DNAs (24) have also been determined. We have also identified the preintegration site of one of the integrated AKR-MLV DNAs.

MATERIALS AND METHODS

Recombinant DNA. The construction of recombinant DNA clones λ-AKR-MLV DNA (clones 623 and 614) (24), plasmid pMLV-1A containing the unintegrated form of Moloney MLV (Mo-MLV) DNA (3, 39), plasmid $pMLV_I$-101 containing the 5′ half of an integrated form of Mo-MLV (38), and clone λ·MSV-1 (43) containing unintegrated form of Moloney mouse sarcoma virus (Mo-MSV) with inverted LTRs have been described. The construction of the Mo-MLV cDNA clone pMLV-201 has also been previously described (38).

To molecularly clone from uninfected cells a DNA fragment in which the AKR-MLV DNA had integrated, we used the following approach. About 1 mg of cellular DNA from uninfected NIH/3T3 cells was digested with *Eco*RI and electrophoresed on a 0.7% agarose gel for 20 h at 70 V/cm. The DNA from a portion of the agarose gel was transferred onto nitrocellulose filters and hybridized to a probe containing the LTR and cellular flanking sequences of the AKR-MLV DNA clone 623 (pAKR-LC-623), prepared as follows. The λ-AKR-MLV DNA clone 623 (λ-AKR-623) contains a single *Hin*dIII site about 2.5 kilobase pairs (kbp) upstream from the 5′ LTR-cellular junction. The *Hin*dIII-*Eco*RI fragment of λ-AKR-623 containing the entire AKR-MLV DNA and some flanking cellular sequences was subcloned in pBR322. It was then cleaved with *Kpn*I, and the largest fragment, containing pBR322, 2.0 kbp of 5′ cellular sequences, 0.5 kbp of 5′ LTR, 0.1 kbp of 3′ LTR, and 1.0 kbp of 3′ cellular sequences, was religated and used to transform *Escherichia coli* C600 (see Fig. 4A). Two other plasmids were constructed to characterize the recombinant DNA containing the preintegration site. The plasmid pAKR-101 contained a *Kpn*I-digested DNA fragment encompassing about 500 nucleotides of 5′ LTR and 2.0 kbp of adjacent cellular sequences (see Fig. 4A). The plasmid pAKR-102 contained a *Kpn*I-digested DNA fragment containing 130 nucleotides of 3′ LTR and adjacent cellular sequences (see Fig. 4). In both cases, the DNA fragments were tailed with oligodeoxycytidine and annealed to oligodeoxyguanosine-tailed, *Pst*I-cleaved pBR322 DNA, followed by transformation as has been described (9, 22). The ^{32}P-labeled probes were made by nick translation, and specific activities of 3×10^7 to 5×10^7 cpm/μg of DNA were usually obtained.

After hybridization of the blot of NIH/3T3 DNA to nick-translated pAKR-LC-623 DNA, a single band in the size range of 7 to 8 kbp was detected. The 7- to 8-kbp region of the remainder of the agarose gel was cut, eluted, and retested for hybridization to pAKR-LC-623 DNA and to a fragment containing only cellular sequences. The size of the fragment was that predicted, since the λ-AKR-623 DNA is 16.4 kbp long and the AKR-MLV DNA is about 8.8 kbp. About 1 to 2 μg of the cellular DNA was ligated to 5 μg of separated arms of modified λ phage Charon 4a as described previously (5). The ligated material was packaged in vitro and assayed for viable phage. One positive clone was isolated after screening 10,000 to 20,000 plaques. The primary plaques were purified by two rounds of screening to obtain a pure plaque population.

Sequence determination procedures. The chemical modification method of DNA sequencing as described by Maxam and Gilbert (26) was used. Sequence ladders were displayed on 6, 12, or 18% polyacrylamide–8 M urea gels 0.04 cm thick.

RESULTS

Comparison of sequences of LTRs. Since the LTR sequences may be involved in the integration and transcription of viral DNA, we were interested to determine whether different murine retroviral LTRs share some common features. Size and sequence heterogeneity of the LTR of murine C-type viruses have been predicted by Rands et al. (28) on the basis of restriction endonuclease mapping of ecotropic MLV DNAs. We have refined this analysis further by determining the nucleotide sequence of several MLV LTRs. Figure 1 shows a diagrammatic sketch of the comparison of nucleotide sequences of LTRs from various murine retroviral DNAs (11, 29, 38). The complete nucleotide sequence of several of these LTRs is given below (see Fig. 6). Several generalizations can be made. (i) The size of each LTR is different, ranging from 519 nucleotides (Mo-MLV of Balb/*Mov*-1 locus) to 626 for AKR-614 or AKR-623 DNA. (ii) In comparing Mo-MLV and AKR LTRs, the U_5 region is conserved, whereas the U_3 region shows considerable variation. A comparison of the LTRs of various Mo-MSV isolates and that of MLV-1A suggests that clone m1 Mo-MSV is more closely related to Mo-MLV clone 1 than is Mo-MSV clone 124. (iii) The 5′-cap nucleotide is located at approximately the same position in all cases. (iv) In all cases, a transcriptional control signal like the RNA Pol II initiation site (TATA-like box) is present −25 to −31 nucleotides from the RNA 5′-cap nucleotide. A similar sequence at positions +46 to +52 may be involved in polyadenylation of RNA (27). Another control signal CCAAT, located around position −80 and implicated in transcription (10, 17), is present in all cases at a similar position. (v) Some LTRs, for instance those of AKR-614, pMLV-1A, and pMSV-12 reported here and that of m1-MSV (11), contain an internal duplication of sequences ranging from 58 to 100 base pairs (bp). However, other LTRs like $pMLV_1$-101 (38) or pMLV-201 (36), containing 3′ LTR, do not contain duplication of sequences. (vi) All proviral LTRs have lost two A residues at the 5′ and 3′ termini, whereas the LTRs from unintegrated viral DNAs retain the terminal two A residues. It should be noted that the loss of two A residues in the 3′ LTR sequence derived from cDNA clone pMLV-201 probably occurred during S1 nuclease treatment (36). (vii) The inverted repeats at the termini of all MLV- or MSV-related LTRs have the same sequence, namely TGAAAGACCCC . . . GGGGTCTTTCA.

To determine whether the LTRs encode a

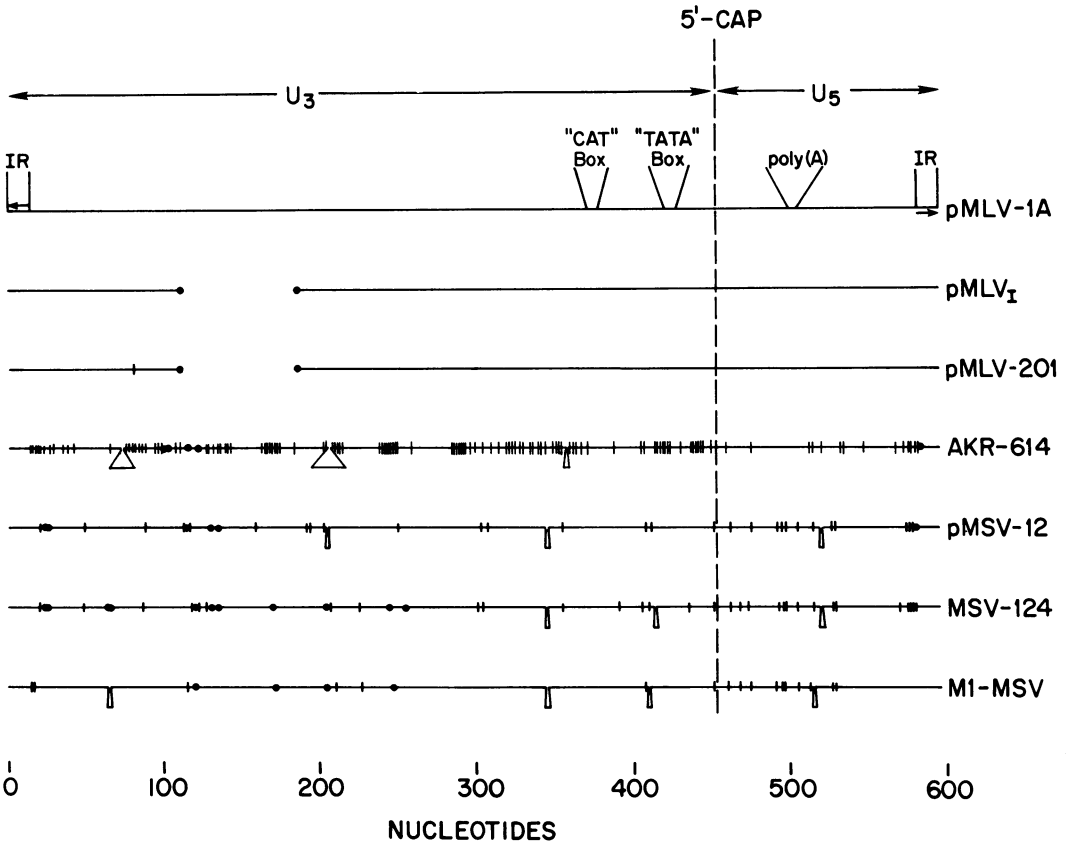

FIG. 1. Schematic comparison of various murine retroviral LTRs. The nucleotide sequence of the LTR of pMLV-1A (unintegrated Mo-MLV [39]) is compared to those of the integrated 5' LTR from the *Mov*-1 locus (pMLV$_I$-101 [38]), the Mo-MLV cDNA clone of the 3' LTR (pMLV-201; see text), AKR-614 (24), pMSV-12 (Fig. 2, structure II), an independent recombinant clone of Mo-MSV clone 124 (29), and a recombinant clone of Mo-MSV clone m1 (11). Comparisons were carried out using the ALIGN program (M. O. Dayhoff, W. C. Barker, and B. C. Orcutt, personal communication) with a unitary matrix and a gap penalty of 3. Symbols: ●, deletions versus pMLV-1A; △, insertions; and |, single base changes. Control signals: I.R., inverted repeat 5'-AATGAAAGACCCC-3'; CAT-box (10, 17), 5'-CCAAT-3'; TATA-box (10, 17), 5'-CAATAAA-3' for all except AKR-614 (5'-CTATAAA-3'); poly(A), possible polyadenylation signal (27) 5'-AATAAA-3'.

protein the three possible frames of translation of Mo-MLV, Mo-MSV, and AKR in the 5' → 3' direction are shown in the appendix. The largest possible open reading frame is for a protein of about 10 kilodaltons encoded by AKR-614 LTR. In the direction opposite to the genomic RNA transcription, the largest possible protein synthesized is less than 7 kilodaltons. Since no mRNA encoded by a murine type C retroviral LTR has been reported, the role of splicing in generation of an RNA molecule which can encode a protein cannot be ruled out.

Inversion of LTR in Mo-MSV. Although the LTR appears to be structurally analogous to movable genetic elements, no transposition of LTRs has been observed in the infected cell DNA. Shoemaker et al. (35) have, however, reported variants of molecularly cloned Mo-MLV DNA in which one of the LTRs integrated into the viral DNA, thus creating a molecule in which the two LTRs are not adjacent to each other. A similar situation has now been encountered in a molecularly cloned Mo-MSV DNA.

We have previously described the molecular cloning of the circular form of unintegrated Mo-MSV DNA containing two LTRs in bacteriophage λ and its subsequent subcloning in pBR322 (43). Since the circular DNA was cleaved at the *Hin*dIII site, the molecular clone was permuted with respect to the linear viral DNA. However, the two LTRs should be present adjacent to each other as shown in Fig. 2

(structure I). Out of a total of three clones analyzed, two clones appeared to have structure I, whereas in the third Mo-MSV DNA clone, λ-MSV-1 (43; subcloned in pBR322, pMSV-12), the LTRs are not adjacent but separated by a stretch of 321 nucleotides (structure II). Furthermore, the 5′ LTR is inverted from the expected orientation in Fig. 2 (compare structures I and II). Several conclusions can be made from the sequence of inverted LTR in pMSV-12. (i) The

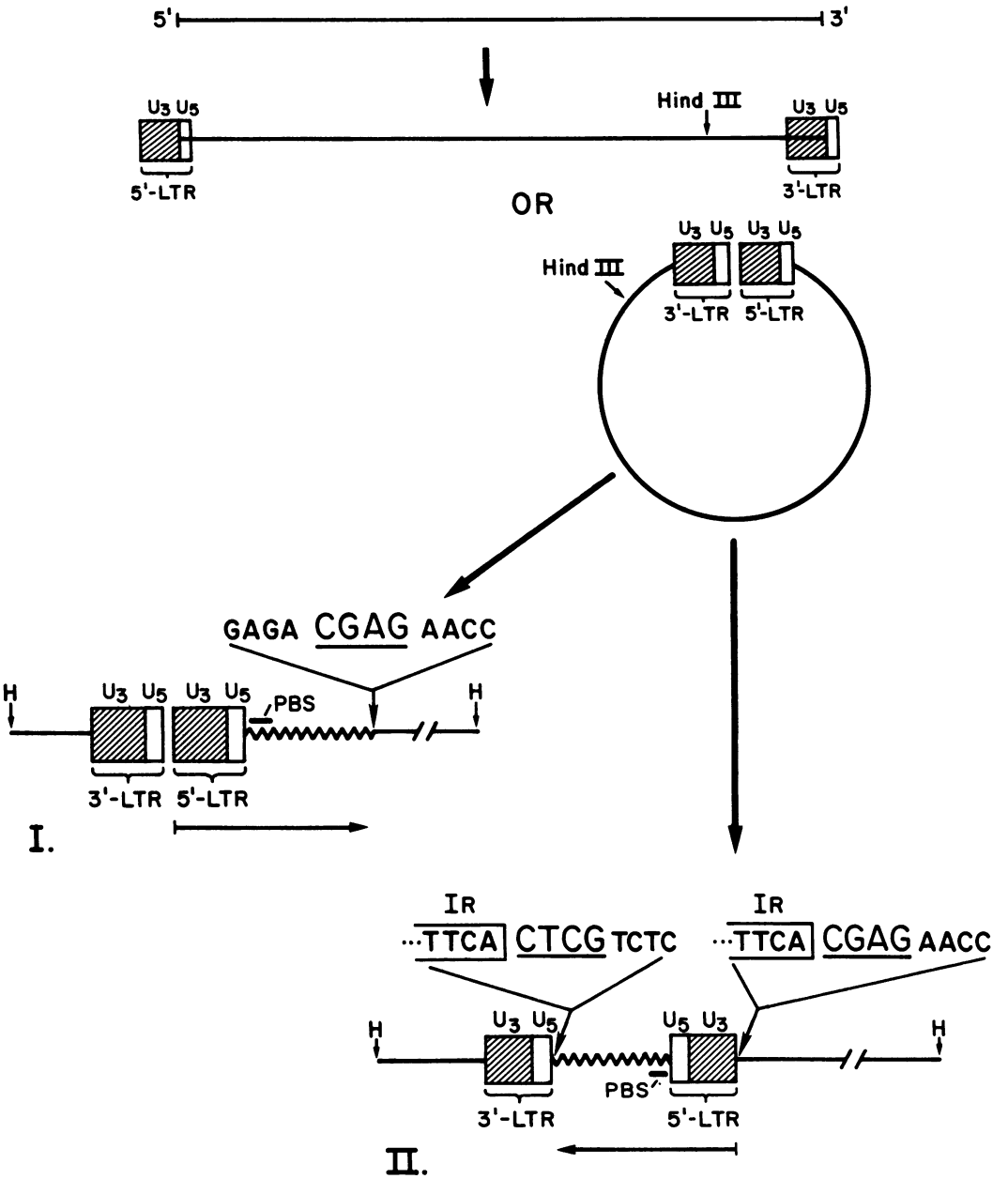

FIG. 2. Structure of recombinant clones of Mo-MSV clone 124. The genomic RNA gives rise to double-stranded DNA having one (not shown) or two LTRs in either linear or circular configuration. The arrangement of recombinant clones of supercoils, isolated from productively infected cells and cloned at the unique *Hin*dIII site, is generally that shown in structure I (e.g., clone pMSV-1L [39a]). In one recombinant clone, pMSV-12, the region indicated by the arrow, encompassing the 5′ LTR and the 321 nucleotides downstream from it (including the tRNA primer binding site [PBS]), has been inverted to give the arrangement shown in structure II. The target tetranucleotide CGAG, found once in structure I, is found repeated at the end of the inversion in structure II.

5′ LTR and 321 nucleotides downstream from it are inverted with respect to their orientation in pMSV-1L (structure I) (39a). (ii) The 321 nucleotides between the two LTRs correspond to sequences representing the tRNA binding site and downstream viral sequences. The inversion presumably occurred during viral DNA synthesis and is not an artifact of cloning (35). (iii) Two terminal A residues from the U_3 region of 5′ LTR and two terminal T residues from the U_5 region of 3′ LTR are lost. (iv) A 4-bp inverted repeat (Fig. 2) is observed at the U_3 terminus of the inverted 5′ LTR, and at the junction of the additional 321-bp sequences and U_5 terminus of the 3′ LTR. It may be noted that in contrast to the proviral DNA described below (Fig. 3 and 5) in which there is direct repeat of adjacent cellular sequences, the 5′ LTR and the additional 321-bp unit are inverted; thus, the flanking sequences display an inverted rather than direct orientation because the molecule has integrated into itself. (v) The 4-bp inverted repeat is present once in the parental DNA at the site of inversion.

Host-viral junctions of AKR-MLV 623 and 614. A hallmark of structures resembling bacterial transposons is the direct repeat of cellular sequences at both termini of the integrated molecule. We have recently reported the molecular cloning of two infectious integrated molecules of ecotropic AKR-MLV DNA (clones 623 and 614) from a chronically infected NIH/3T3 cell line (24). We wanted to determine the nucleotide sequence of host-viral junctions of these two proviral DNAs. In particular, we wanted to determine whether the viral DNA has any preference for host sites. Figure 3 shows the nucleotide sequence of the junctions of AKR-MLV DNAs and mouse cellular DNAs. The inverted repeats at the termini of an LTR define the boundary of the viral sequences. In the case of AKR-623 DNA, the 5′ LTR and host junction is ACAA, whereas the 3′ LTR and host junction is ACAT. In all the retroviral DNAs analyzed so far, the terminal two nucleotides are lost during the integration process. Assuming that similar rules apply to AKR-623 DNA, there appear to be either 3- or 4-nucleotide direct repeats of cellular sequences. If one out of the two terminal A residues is retained at the 5′ junction, then the repeat is 3 nucleotides. Alternatively, the repeat is 4 nucleotides, and 1 nucleotide at either the 5′ or 3′ junction has undergone transversion.

In the case of AKR-614, the direct cellular repeat is 5 nucleotides (Fig. 3). Again, the terminal two A residues are lost during integration. It is interesting to note that the two AKR MLV DNAs, depending on their site of integration,

FIG. 3. Nucleotide sequence at junctions of viral AKR and cellular DNA. The fragments sequenced for the junctions of AKR-623 are given in the legend to Fig. 5. To determine the junctions of AKR-614, the 12.9-kbp *Eco*RI insert of λ-AKR-614 was subcloned into the *Eco*RI site of pBR322. The left junction was determined by labeling the *Pst*I site in the 5′ LTR, cleaving the 2.5-kbp *Pst*I fragment with *Eco*RI, and sequencing the 1.75-kbp fragment. The right junction was sequenced by labeling the *Sma*I site in the 3′ LTR, cleaving the 9.1-kbp fragment with *Bam*HI, and sequencing the 2.9-kbp fragment. Duplicate nucleotides are enclosed in open boxes, and the 11-nucleotide inverted repeats at the termini of the LTRs are indicated by hatched boxes. A, Junctions in clone λ-AKR-623; B, junctions in clone λ-AKR-614.

generate direct repeats of cellular sequences of either 3 or 4 bp as in the case of AKR-623 DNA or a 5-bp repeat in AKR-614 DNA.

Nucleotide sequence of preintegration site. In the proposed models for integration of movable genetic elements, it is imperative that the direct cellular repeat sequences be present only once at the site of integration (33, 35). Since we were not sure whether the direct cellular repeat is 3 or 4 nucleotides in the case of AKR-623, we determined the nucleotide sequence of the preintegration site for AKR-623. Figure 4 shows the characterization of the recombinant DNA clone, λ-NIH-623, containing the preintegration site. The ethidium bromide staining pattern (Fig. 4B, lane a) shows that after cleavage with *Eco*RI, two inserts of an average size of 7.0 and 7.8 kbp, in addition to the λ Charon 4a arms, can be identified. However, when the DNA from the same gel is transferred to cellulose nitrate and hybridized to a nick-translated probe made from pAKR-LC-623 DNA, only the 7.8-kbp insert can be identified (Fig. 4B, lane b). The same 7.8-kbp band hybridizes when the filter is hybridized to nick-translated λ-AKR-623 DNA (Fig. 4B, lane c). In addition, however, the λ arms and some uncut DNA can also be identified. The λ-AKR-614 DNA does not hybridize to the λ-NIH-623 insert DNA (data not shown). Lanes d through i of Fig. 4B display the analysis of λ-NIH-623 DNA by hybridization to pAKR-101 and pAKR-102 DNA. The *Bam*HI (Quint and Berns, personal communication), *Kpn*I, and *Eco*RI restriction endonuclease map of AKR-623 DNA is shown in Fig. 4A, which also shows the sequences in plasmids pAKR-101 and pAKR-102. The restriction map of λ-NIH-623 should be identical to that of AKR-623 DNA minus the AKR-MLV DNA sequences. To test this prediction, we cleaved λ-NIH-623 DNA with *Bam*HI, *Eco*RI, and *Bam*HI plus *Eco*RI and hybridized to labeled pAKR-101 and pAKR-102 DNA as probes. In both cases, digestion with *Eco*RI should yield an insert of 7.8 kbp (Fig. 4B, lanes d and g). Digestion with *Bam*HI and hybridization to pAKR-101 DNA should show a major band at about 2.8 kbp and a 6.0-kbp band containing NIH-623 and λ right arm sequences (Fig. 4B, lane e). On the other hand, hybridization of the *Bam*HI digest to pAKR-102 DNA should show only the 6.0-kbp fragment (Fig. 4B, lane h). Digestion with *Bam*HI and *Eco*RI and hybridization to pAKR-101 DNA should yield the 2.8-kbp band and a smaller 1.4-kbp fragment because *Eco*RI cleaves at the junction of NIH-623 and λ right arm (Fig. 4B, lane f). However, hybridization of pAKR-102 DNA of the *Bam*HI-*Eco*RI-digested DNA should yield only one fragment of 1.4 kbp (Fig. 4B, lane i). Thus, it appears that the 7.8-kbp insert of λ-NIH-623 DNA has a restriction endonuclease pattern similar to that of the flanking cellular sequences of λ-AKR-623 DNA.

The nucleotide sequence of the preintegration site and flanking sequences is shown in Fig. 5. A sequence ACAT which constitutes the direct cellular repeat of λ-AKR-623 can be identified to occur once. This sequence is flanked at its 5′ and 3′ ends by sequences which overlap with flanking cellular sequences of λ-AKR-623. It is difficult to decide whether the direct cellular repeat in AKR-623 is 3 or 4 nucleotides.

A striking feature of the sequences flanking the preintegration site of AKR-623 DNA is the apparent homology with viral LTR sequences. A sequence TTCC at the 5′ end and TGAA at the 3′ end of the preintegration site have homologous sequences in the inverted repeats of 3′ and 5′ LTR sequences. No such homology can be observed in the case of AKR-614 DNA.

DISCUSSION

Establishment of infection by retroviruses requires the integration of viral DNA into host chromosomes. Usually more than one copy of the viral DNA is integrated, although not all proviral DNAs are transcribed. The mechanism of integration remains totally obscure. However, all proviral DNAs have a unique structure, namely, $U_3U_5 \ldots U_3U_5$. Detailed nucleotide sequence analyses of the structure of several retroviral DNAs provide a strong analogy with the structure of transposable genetic elements like bacterial transposons, the Ty*1* elements of yeasts, and *copia* in *Drosophila* (reviewed in reference 19). The retroviral DNA structure is characterized by the presence of LTRs. The size of the LTRs is variable, ranging from 273 bp for avian endogenous provirus (ev-1) (19) to 1,327 nucleotides for mouse mammary tumor virus (12). The murine retroviruses vary in size from 500 to 650 nucleotides, whereas the avian retroviruses have a wider size range, from 273 nucleotides for ev-1 virus to 330 nucleotides for RSV-SR-A (37), 350 nucleotides for RSV-SR-D (23), and 569 bp for spleen necrosis virus (34). Each direct repeat has inverted repeats at its termini. The number of nucleotides in the inverted repeats ranges from 3 for spleen necrosis virus to 11 for integrated Mo-MLV (38). As pointed out by Hishinuma et al. (19), the 5′ and 3′ termini of all proviral and many transposable elements are always TG . . . CA. The LTRs described here are no exception to this rule. In the case of unintegrated viral DNAs, however, the 5′ terminus is AATG, and the 3′ terminus is CATT. The last 2 nucleotides at each terminus are lost during integration. Presumably, the staggered cuts are made at AA ↓ TG and CA ↑ TT (33, 35).

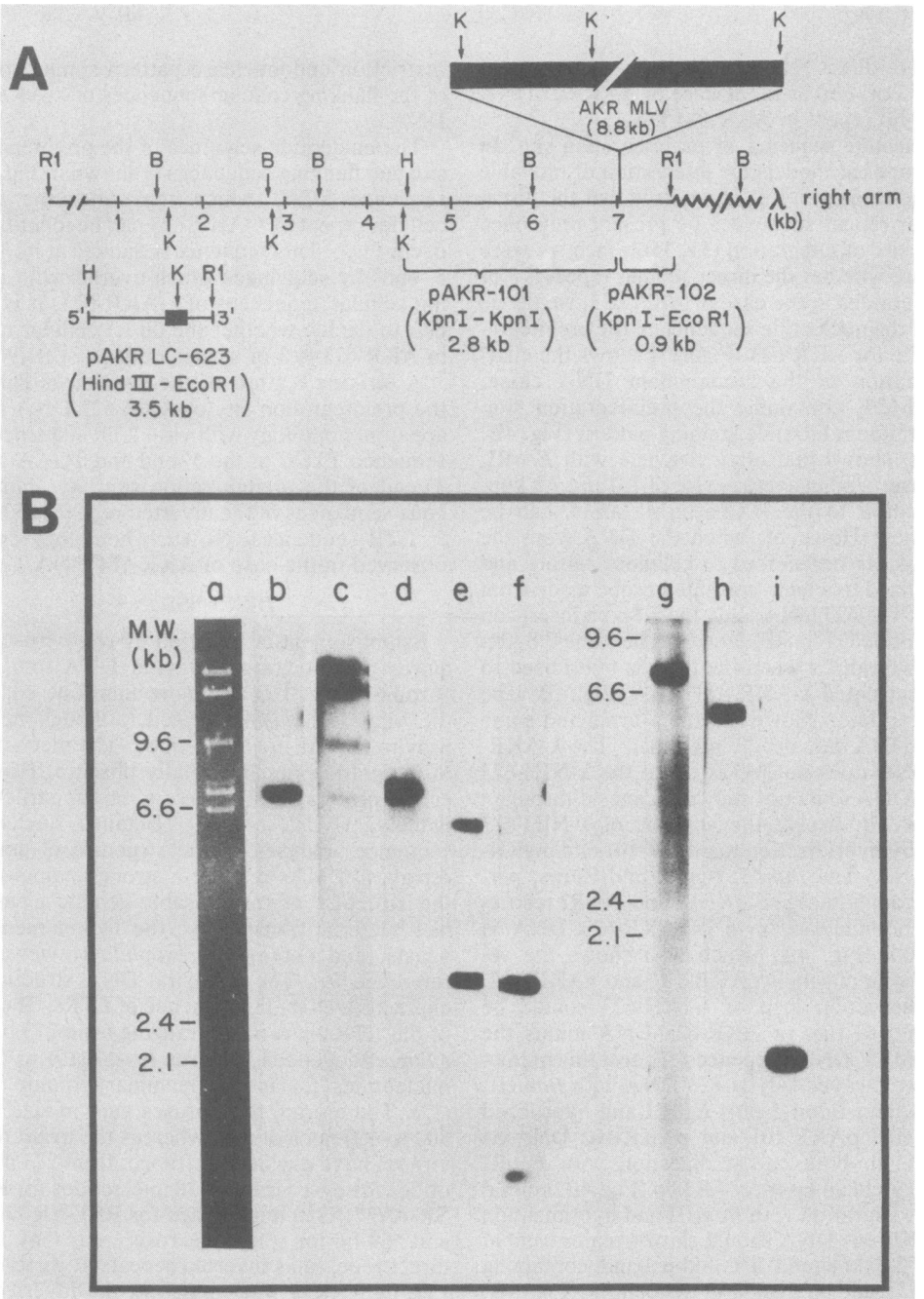

FIG. 4. Characterization of recombinant clone λ-NIH-623. A, Predicted restriction map of the 7.8-kbp *Eco*RI fragment containing the AKR-623 preintegration site. Lengths are shown in kbp. Restriction enzymes: RI, *Eco*RI; B, *Bam*HI; K, *Kpn*I; H, *Hin*dIII. (Data for *Bam*HI are from Quint and Berns, personal communication.) Regions of homology with AKR-623 recombinant subclones pAKR-101 and pAKR-102 are indicated with brackets. The location of the AKR-623 integration is shown by the closed box. The structure of the insert in the AKR-623 subclone pAKR-LC-623, from the *Hin*dIII site to the 5′ LTR *Kpn*I site, ligated to the 3′ LTR *Kpn*I site, and through to the *Eco*RI site is also shown. B, Restriction digest patterns of λ-NIH-623 DNA. λ-NIH-623 DNA was cleaved with *Eco*RI (lanes a, b, c, d, and g), *Bam*HI (lanes e and h), or *Eco*RI plus *Bam*HI (lanes f and i). In lane a, the DNA was stained with ethidium bromide. In all other cases, the DNA was transferred onto cellulose nitrate sheets and annealed to various nick-translated [^{32}P]DNAs, and the resultant bands were visualized by autoradiography (3). The [^{32}P]DNA probes used were: pAKR-LC-623 (lane b), λ-AKR-623 (lane c), pAKR-101 (lanes d, e, and f) and pAKR-102 (lanes g, h, and i). Wild-type phage λ DNA digested with *Hin*dIII was used for size markers.

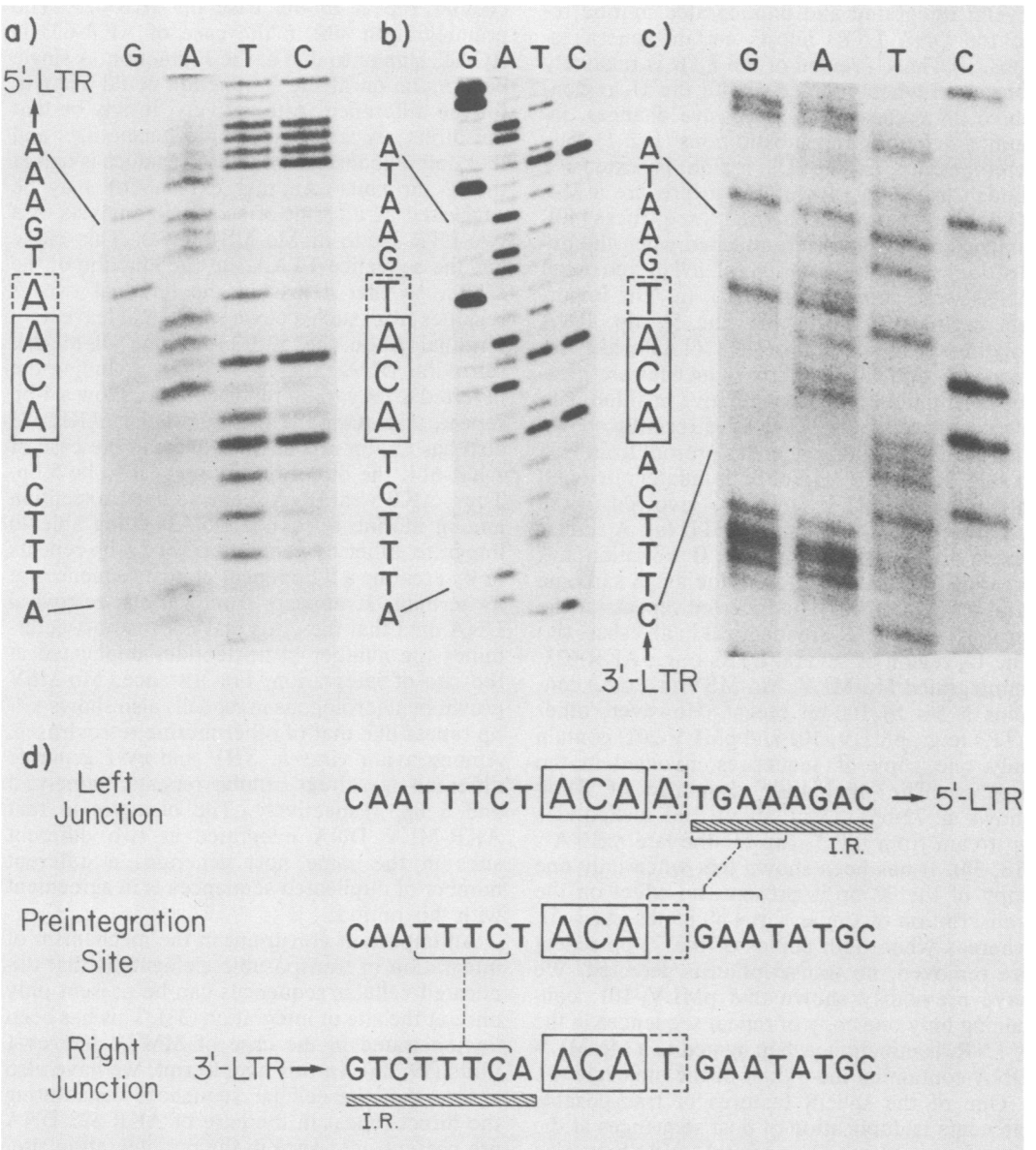

FIG. 5. Preintegration site and cellular-viral junctions of recombinant clone λ-AKR-623. All sequences are read from bottom to top in the 5′ to 3′ direction. The nucleotides enclosed in boxes are those shared by all three sequences, with the possible fourth nucleotide enclosed in a dotted box. a, Left junction. DNA sequence ladder of the faster migrating strand of a 5′ terminally labeled 410-bp *Sau*3A fragment of the 920-bp *Bam*HI-*Kpn*I fragment of subclone pAKR-101. Reactions (26): G, G alone; A, A > C; T, T + C; C, C alone. Partial digestion products were separated on a 12% polyacrylamide–8 M urea gel. b, Preintegration site. DNA sequence ladder of the slower migrating strand of a 5′ terminally labeled 255-bp *Sau*3A partial digestion fragment of a 925-bp *Bam*HI-*Eco*RI fragment of clone λ-NIH-623. Chemical degradation reactions and sequence gel were as described above. c, Right junction. DNA sequence ladder of the 800-bp *Kpn*I-*Eco*RI fragment of subclone pAKR-102 3′ terminally labeled at the *Kpn*I site. The lanes have been labeled to allow direct reading of the complementary strand. Partial chemical degradation products were separated on a 6% polyacrylamide–8 M urea gel. d, Alignment of 5′ and 3′ junction sequences with the preintegration site. Sequences shared between the preintegration site and the inverted repeats (I.R.) of the LTRs are shown by brackets.

A comparison of the nucleotide sequence of several integrated and unintegrated murine retroviral DNA LTRs allows certain generalizations. (i) The U_5 region of the LTR is relatively conserved when compared with the U_3 region, which undergoes rather extensive changes, including deletions and substitutions (Fig. 1). The heterogeneity in the U_3 region is extensive enough in that we were able to prepare a Mo-MLV specific probe from the U_3 sequences (40). Similar conclusions were reached when the nucleotide sequences of several avian retroviral DNAs were compared (23, 37). (ii) The important regulatory signals like the 5′ cap, RNA polymerase II initiation sites (TATA-like sequences), and CCAAT box sequences are present essentially at the same relative position with a minimum of change. We have recently shown that the TATA-like sequences present from −25 to −31 from the 5′ cap of both unintegrated and integrated Mo-MLV DNA are involved in the initiation of RNA polymerase II (14a). A similar region for RNA polymerase II initiation has previously been identified in the avian sarcoma viral LTR (44). (iii) The inverted repeats at the termini of the LTR are identical in all cases. (iv) The U_3 region of several LTRs (e.g., AKR-623, unintegrated Mo-MLV, Mo-MSV-m1, etc.) contains a 58- to 100-bp repeat. However, other LTRs (e.g., $pMLV_I$-101 and pMLV-201) contain only one copy of sequences involved in the repeats. The nucleotide sequence of SV40 shows a 72-bp repeat about 300 nucleotides upstream from the 5′ cap for the late mRNA's (18, 30). It has been shown that when only one copy of the 72 bp is present, no effect on the transcription of simian virus 40 is observed (1), whereas when both copies of the 72-bp repeat are removed, no transcription is detected. We have previously shown that $pMLV_I$-101, containing only one copy of repeat sequences in the 5′ LTR, is infectious when ligated to a Mo-MLV DNA containing the 3′ half of the molecule (3).

One of the salient features of transposable elements is duplication of host sequences at the site of integration. The proviral DNAs have also been shown to generate duplication of the host DNA as witnessed by the presence of direct repeat of cellular sequences at its termini (11, 19, 25, 34). The number of bases involved in the duplication ranges from 6 bp in MMTV and ev-1 proviral DNA to 5 bp in the case of spleen necrosis virus (SNV) and 4 bp in the case of several murine retroviral DNAs. Two integrated AKR-MLV DNAs have been analyzed in this report. One of them, AKR-623, generates a 3- or 4-bp repeat, whereas the other, AKR-614, creates a 5-bp direct repeat. Because the sequence at the 5′ junction is ACAA and that at the 3′ junction is ACAT, there is an uncertainty in the precise assignment of the size of the direct cellular repeat in the case of AKR-623. The preintegration site in the case of AKR-623 is ACAT, similar to that of the 3′ junction. A single base mutation at the 5′ junction could account for the difference. Alternatively, it may be that the direct repeat is only three nucleotides and that only one instead of two A residues is lost at the 5′ terminus. Another explanation may be suggested as a result of sequence analysis of a two-LTR clone of Mo-MSV (39a). This clone has the sequence TTAAA at the junction of the LTRs, so that a loss of the terminal two A residues from such a clone would still leave a 5′-terminal A on the 5′ LTR. Since all murine retroviral DNAs analyzed so far, including the inverted LTR variant reported here, show a 4-bp repeat, it is tempting to suggest that AKR-623 also has a 4-bp repeat. However, in the case of AKR-614, the direct repeat appears to be 5 bp. Thus, AKR viral DNA seems to be an exception among murine retroviral DNAs, being able to integrate either by generating 3- or 4-bp repeats or by creating a 5-bp repeat of host sequences at its termini. It appears from various retroviral DNA data that the virus and not the host determines the number of nucleotides duplicated at the site of integration. For instance, Mo-MSV grown in heterologous mink cells also shows a 4-bp repeat like that of other murine retroviruses. Among avian viruses, SNV and ev-1 generate different-size direct cellular repeats, namely, 5 and 6 bp, respectively. The observation that AKR-MLV DNA integrated at two different sites in the same host generates a different number of duplicated sequences is in agreement with this notion.

An important constraint in the mechanism of integration of transposable elements is that duplicated cellular sequences can be present only once at the site of integration (33). This has been demonstrated in the case of MMTV and ev-1 virus (19, 25). In this manuscript, we have also shown that the cellular sequences constituting the direct repeat in the case of AKR-623 DNA are present only once in the preintegration site. In the case of MMTV and ev-1, there seems to be no apparent homology between the flanking cellular sequences and viral LTRs. However, in the case of AKR-623, one nucleotide to the left of the preintegration site is a sequence TTTC homologous to TTTC at the 3′ terminus of 3′ LTR (Fig. 5). Similarly, a sequence TGAA immediately to the right of the preintegration site is homologous to TGAA present at the 5′ terminus of 5′ LTR (Fig. 5). Thus, had a circular form of AKR-MLV been the precursor to the AKR-623 clone, there would have been considerable sequence homology between cellular DNA and viral LTRs at the site of integration.

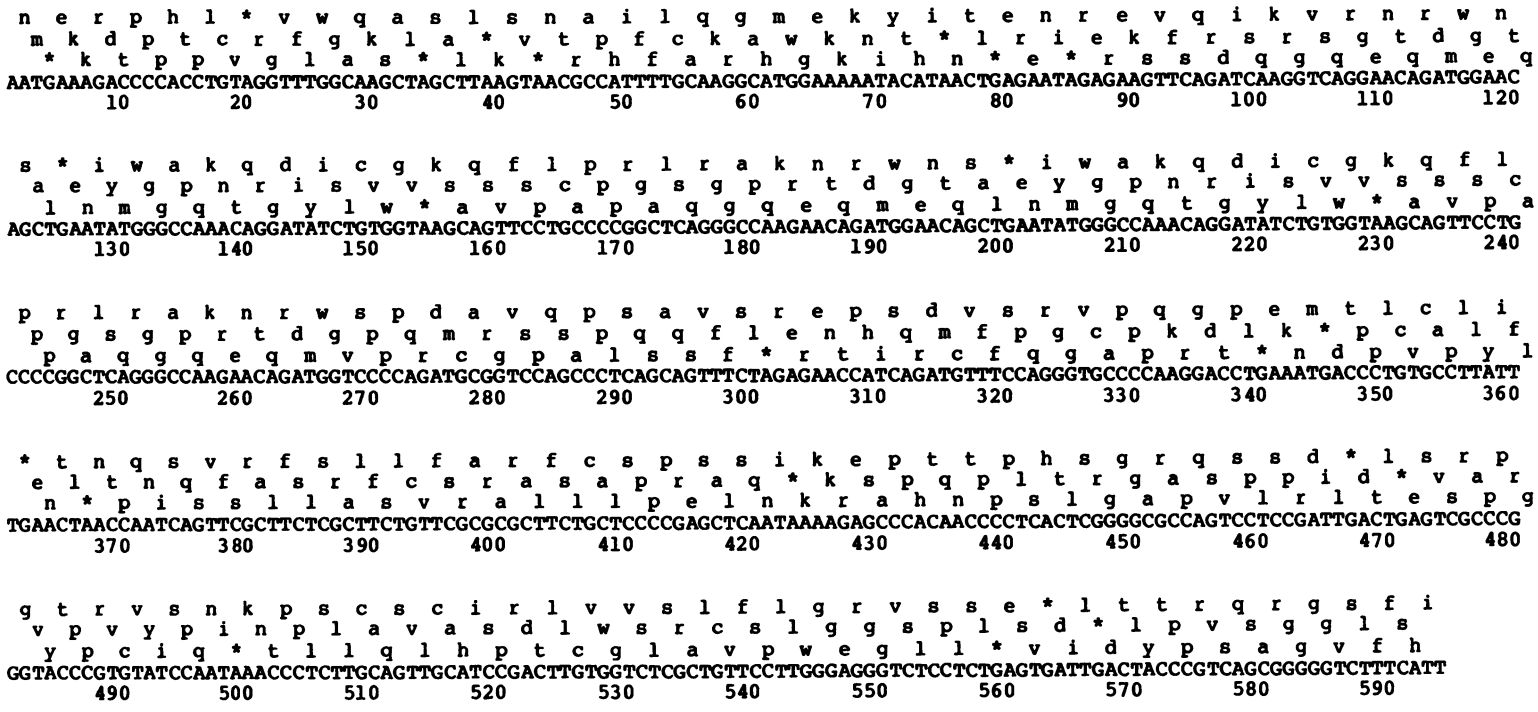

FIG. 6. DNA sequence and three-frame translation of LTRs. The single-stranded DNA sequence corresponding to the genomic (+) RNA strand of the LTRs of pMLV-1A (3), pMLV-201 (36), AKR-614 (24) and pMSV-12 (43) is presented in upper-case letters. Above the nucleotide sequence are the potential translation products, shown in lower-case letters, for the three possible frames. Translation terminators are depicted by (*).

2) pMLV-201

```
 n  e  r  p  h  l  *  v  w  q  a  s  l  s  n  a  i  l  q  g  m  e  k  y  i  t  e  n  r  k  v  q  i  k  v  r  n  r  w  n
  m  k  d  p  t  c  r  f  g  k  l  a  *  v  t  p  f  c  k  a  w  k  n  t  *  l  r  i  e  k  f  r  s  r  s  g  t  d  g  t
   *  k  t  p  p  v  g  l  a  s  *  l  k  *  r  h  f  a  r  h  g  k  i  h  n  *  e  *  k  s  s  d  q  g  q  e  q  m  e  q
AATGAAAGACCCCACCTGTAGGTTTGGCAAGCTAGCTTAAGTAACGCCATTTTGCAAGGCATGGAAAAATACATAACTGAGAATAGAAAAGTTCAGATCAAGGTCAGGAACAGATGGAAC
       10        20        30        40        50        60        70        80        90       100       110       120

 s  *  i  w  a  k  q  d  i  c  g  k  q  f  l  p  r  l  r  a  k  n  r  w  s  p  d  a  v  q  p  s  a  v  s  r  e  p  s  d
  a  e  y  g  p  n  r  i  s  v  v  s  s  s  c  p  g  s  g  p  r  t  d  g  p  q  m  r  s  s  p  q  q  f  l  e  n  h  q  m
   l  n  m  g  q  t  g  y  l  w  *  a  v  p  a  p  a  q  g  q  e  q  m  v  p  r  c  g  p  a  l  s  s  f  *  r  t  i  r  c
AGCTGAATATGGGCCAAACAGGATATCTGTGGTAAGCAGTTCCTGCCCCGGCTCAGGGCCAAGAACAGATGGTCCCCAGATGCGGTCCAGCCCTCAGCAGTTTCTAGAGAACCATCAGAT
      130       140       150       160       170       180       190       200       210       220       230       240

 v  s  r  v  p  q  g  p  e  m  t  l  c  l  i  *  t  n  q  s  v  r  f  s  l  l  f  a  r  f  c  s  p  s  s  i  k  e  p  t
  f  p  g  c  p  k  d  l  k  *  p  c  a  l  f  e  l  t  n  q  f  a  s  r  f  c  s  r  a  s  a  p  r  a  q  *  k  s  p  q
   f  q  g  a  p  r  t  *  n  d  p  v  p  y  l  n  *  p  i  s  s  l  l  a  s  v  r  a  l  l  l  p  e  l  n  k  r  a  h  n
GTTTCCAGGGTGCCCCAAGGACCTGAAATGACCCTGTGCCTTATTTGAACTAACCAATCAGTTCGCTTCTCGCTTCTGTTCGCGCGCTTCTGCTCCCCGAGCTCAATAAAAGAGCCCACA
      250       260       270       280       290       300       310       320       330       340       350       360

 t  p  h  s  g  r  q  s  s  d  *  l  s  r  p  g  t  r  v  s  n  k  p  s  c  s  c  i  r  l  v  v  s  l  f  l  g  r  v  s
  p  l  t  r  g  a  s  p  p  i  d  *  v  a  r  v  p  v  y  p  i  n  p  l  a  v  a  s  d  l  w  s  r  c  s  l  g  g  s  p
   p  s  l  g  a  p  v  l  r  l  t  e  s  p  g  y  p  c  i  q  *  t  l  l  q  l  h  p  t  c  g  l  a  v  p  w  e  g  l  l
ACCCCTCACTCGGGGCGCCAGTCCTCCGATTGACTGAGTCGCCCGGGTACCCGTGTATCCAATAAACCCTCTTGCAGTTGCATCCGACTTGTGGTCTCGCTGTTCCTTGGGAGGGTCTCC
      370       380       390       400       410       420       430       440       450       460       470       480

 s  e  *  l  t  t  r  q  r  g  s  f
  l  s  d  *  l  p  v  s  g  g  l  s
   *  v  i  d  y  p  s  a  g  v  f
TCTGAGTGATTGACTACCCGTCAGCGGGGGTCTTTCA
      490       500       510
```

3) AKR-614

```
 n e r p l h k a * p a n c s n a i l q g m g k y q s * c s q k n k n k e v q r g
  m k d p f i r l s q l t a v t p f c k a w e n t r a d v l r k t r t r k y r e a
   * k t p s * g l a s * l q * r h f a r h g k i p e l m f s e k q e q g s t e r l
AATGAAAGACCCCTTCATAAGGCTTAGCCAGCTAACTGCAGTAACGCCATTTTGCAAGGCATGGGAAAATACCAGAGCTGATGTTCTCAGAAAAACAAGAACAAGGAAGTACAGAGAGGC
        10        20        30        40        50        60        70        80        90       100       110       120

 w k v p g l g p n r i s v v k h * g p g p g p r t d g p q k q r g w k v p g l g
  g k y r d * g q t g y l w s s t r a p a q g q e q m v p r n r e a g k y r d * g
   e s t g t r a k q d i c g q a l g p r p r a k n r w s p e t e r l e s t g t r a
TGGAAAGTACCGGGACTAGGGCCAAACAGGATATCTGTGGTCAAGCACTAGGGCCCCGGCCCAGGGCCAAGAACAGATGGTCCCCAGAAACAGAGAGGCTGGAAAGTACCGGGACTAGGG
       130       140       150       160       170       180       190       200       210       220       230       240

 p n r i s v v k h * g p g p g p r t d g p q k * l k q q q f q e t q k l s q g s
  q t g y l w s s t r a p a q g q e q m v p r n s * n n n s f k r p r n c l k v p
   k q d i c g q a l g p r p r a k n r w s p e i a k t t t v s r d p e t v s r f p
CCAAACAGGATATCTGTGGTCAAGCACTAGGGCCCCGGCCCAGGGCCAAGAACAGATGGTCCCCAGAAATAGCTAAAACAACAACAGTTTCAAGAGACCCAGAAACTGTCTCAAGGTTCC
       250       260       270       280       290       300       310       320       330       340       350       360

 p d d r g s t p s l i * t n q s a r f s l l y p r l l l p s s i k r v r t p h s
  q m t g d q p q a s f k l t n q l a s r f c t r a y c c p a l * k g * e p h t r
   r * p g i n p k p h l n * p i s s l l a s v p a l i a a q l y k k g k n p t l g
CCAGATGACCGGGGATCAACCCCAAGCCTCATTTAAACTAACCAATCAGCTCGCTTCTCGCTTCTGTACCCGCGCTTATTGCTGCCCAGCTCTATAAAAAGGGTAAGAACCCCACACTCG
       370       380       390       400       410       420       430       440       450       460       470       480

 a r q s s d r l s r p g t r v s n k a f c c c i r i v v s l i l g r v s s e * l
  r a s p p i d * v a r v p v y p i k p f a v a s e s w s r * s l g g s p q s d *
   a p v l r * t e s p g y p c i q * s l l l l h p n r g l a d p w e g l l r v i d
GCGCGCCAGTCCTCCGATAGACTGAGTCGCCCGGGTACCCGTGTATCCAATAAAGCCTTTTGCTGTTGCATCCGAATCGTGGTCTCGCTGATCCTTGGGAGGGTCTCCTCAGAGTGATTG
       490       500       510       520       530       540       550       560       570       580       590       600

 t a q p g g l s
  l p s l g v f h
   c p a w g s f i
ACTGCCCAGCCTGGGGGTCTTTCATT
       610       620
```

4) pMSV-12

```
 n  e  r  p  h  p  *  v  a  s  *  l  k  *  r  h  f  a  r  h  g  k  i  h  n  *  e  *  k  s  s  d  q  g  q  e  q  r  n  s
  m  k  d  p  t  r  r  w  q  a  s  l  s  n  a  t  l  q  g  m  e  k  y  i  t  e  n  r  k  v  q  i  k  v  r  n  k  e  t  a
   *  k  t  p  p  v  g  g  k  l  a  *  v  t  p  l  c  k  a  w  k  n  t  *  l  r  i  e  k  f  r  s  r  s  g  t  k  k  q  l
AATGAAAGACCCCACCCGTAGGTGGCAAGCTAGCTTAAGTAACGCCACTTTGCAAGGCATGGAAAAATACATAACTGAGAATAGAAAAGTTCAGATCAAGGTCAGGAACAAAGAAACAGC
       10        20        30        40        50        60        70        80        90       100       110       120

 *  i  p  n  r  i  s  v  v  s  g  s  c  p  g  s  g  p  r  t  d  e  t  a  e  *  w  a  k  q  d  i  c  g  k  q  f  l  p  r
  e  y  q  t  g  y  l  w  *  a  v  p  a  p  a  q  g  q  e  q  m  r  q  l  s  d  g  p  n  r  i  s  v  v  s  s  s  c  p  g
   n  t  k  q  d  i  c  g  k  r  f  l  p  r  l  r  a  k  n  r  *  d  s  *  v  m  g  q  t  g  y  l  w  *  a  v  p  a  p  a
TGAATACCAAACAGGATATCTGTGGTAAGCGGTTCCTGCCCCGGCTCAGGGCCAAGAACAGATGAGACAGCTGAGTGATGGGCCAAACAGGATATCTGTGGTAAGCAGTTCCTGCCCCGG
       130       140       150       160       170       180       190       200       210       220       230       240

 l  g  a  k  n  r  w  s  p  d  a  v  q  p  s  a  v  s  s  e  s  s  d  v  s  r  v  p  q  g  p  e  n  d  p  v  p  y  l  n
  s  g  p  r  t  d  g  p  q  m  r  s  s  p  q  q  f  l  v  n  h  q  m  f  p  g  c  p  k  d  l  k  m  t  l  y  l  i  *  t
   r  g  q  e  q  m  v  p  r  c  g  p  a  l  s  s  f  *  *  i  i  r  c  f  q  g  a  p  r  t  *  k  *  p  c  t  l  f  e  l
CTCGGGGCCAAGAACAGATGGTCCCCAGATGCGGTCCAGCCCTCAGCAGTTTCTAGTGAATCATCAGATGTTTCCAGGGTGCCCCAAGGACCTGAAAATGACCCTGTACCTTATTTGAAC
       250       260       270       280       290       300       310       320       330       340       350       360

 *  p  i  s  s  l  l  a  s  v  r  a  l  p  l  s  e  l  n  k  r  a  h  n  p  s  l  g  a  p  v  f  r  *  t  a  s  p  g  y
  n  q  s  v  r  f  s  l  l  f  a  r  f  r  s  p  s  s  i  k  e  p  t  t  p  h  s  a  r  q  s  s  d  r  l  r  r  p  g  t
   t  n  q  f  a  s  r  f  c  s  r  a  s  a  l  r  a  q  *  k  s  p  q  p  l  t  r  r  a  s  l  p  i  d  c  v  a  r  v  p
TAACCAATCAGTTCGCTTCTCGCTTCTGTTCGCGCGCTTCCGCTCTCCGAGCTCAATAAAAGAGCCCACAACCCCTCACTCGGCGCGCCAGTCTTCCGATAGACTGCGTCGCCCGGGTAC
       370       380       390       400       410       420       430       440       450       460       470       480

 p  y  s  q  *  s  l  l  l  f  a  s  e  s  w  s  r  c  s  l  g  g  s  p  l  s  d  *  l  p  t  t  g  v  f  h
  r  i  p  n  k  a  s  c  c  l  h  p  n  r  g  l  a  v  p  w  e  g  l  l  *  v  i  d  y  p  r  r  g  s  f  i
   v  f  p  i  k  p  l  a  v  c  i  r  i  v  v  s  l  f  l  g  r  v  s  s  e  *  l  t  t  h  d  g  g  l  s
CCGTATTCCCAATAAAGCCTCTTGCTGTTTGCATCCGAATCGTGGTCTCGCTGTTCCTTGGGAGGGTCTCCTCTGAGTGATTGACTACCCACGACGGGGGTCTTTCATT
       490       500       510       520       530       540       550       560       570       580
```

The retroviruses have a strong structural homology with transposable elements. The mechanism of integration also suggests that they behave like movable genetic elements. Furthermore, like transposons, the yeast Ty*1* element or the *copia* element of *Drosophila*, retroviral DNAs integrate at multiple sites in the host chromosomal DNA. Unlike the bacterial transposons, however, the retroviral DNAs have not been shown to move from their integration site in the chromosome. The inverted LTR variant reported here and before (35) offers a direct proof that LTR can integrate like transposable elements by creating a duplication of the sequences at the site of integration.

APPENDIX

The complete nucleotide sequence of the LTRs and the possible open reading frames of the following murine retroviral DNAs have been determined (Fig. 6):

(i) pMLV-1A. The circular form of unintegrated Mo-MLV DNA containing one LTR was molecularly cloned as described previously (3).

(ii) pMLV-201. The construction and nucleotide sequence of the viral DNA contained in pMLV-201 has previously been described (36). There were reported to be a few nucleotide changes in the LTR sequence of pMLV-201 when compared to the LTR sequence of $pMLV_I$-101 (an integrated Mo-MLV DNA clone) (3, 38). In particular, at positions −25 to −31 from the 5′-cap nucleotide, the sequence in pMLV-201 was CAAAAAA as compared with CAATAAA in $pMLV_I$-101. We have independently determined the nucleotide sequence of the LTR in pMLV-201 and found only one of the original three changes (at position 88, i.e., −362 from the 5′-cap nucleotide).

(iii) pAKR-614. The 13-kbp insert from integrated λ-AKR-614 DNA (24) was cleaved with restriction endonuclease *Eco*RI and subcloned in the unique *Eco*RI site of plasmid pBR322. The nucleotide sequence of the LTR of AKR-623 DNA is identical to that of AKR-614 DNA.

(iv) pMSV-12. The circular form of unintegrated Mo-MSV DNA containing 2 LTRs was molecularly cloned as described previously (43). In pMSV-12, the 3′ LTR integrated into the viral DNA at 463 nucleotides from the 5′-cap nucleotide. Furthermore, the 3′ LTR is inverted in its orientation with respect to the 5′ LTR as shown in structure II of Fig. 2.

ACKNOWLEDGMENTS

We are indebted to Jim Ostlund for help in the computer analysis of the nucleotide sequences. We thank Janice Galleshaw and Doug Murdock for excellent technical assistance, and Maureen Brennan for typing the manuscript.

The work reported here was supported by grants from the National Cancer Institute and the American Cancer Society.

ADDENDUM IN PROOF

The nucleotide sequence of the LTR of an independently isolated clone of MSV-124 (29) was treated separately in Fig. 1 due to the occurrence of 26 differences with respect to the LTR of pMSV-12 (43). In a recent publication, however, the authors have made 26 base changes in their 582-nucleotide LTR sequence (E. P. Reddy, M. J. Smith, and S. A. Aaronson, Science **214**:445–450, 1981). The revised sequence now has only one difference from that of pMSV-12 (at position 85). The line in Fig. 1 for MSV-124 should, therefore, be almost identical to that for pMSV-12.

LITERATURE CITED

1. **Benoist, C., and P. Chambon.** 1981. *In vivo* sequence requirements of the SV40 early promoter region. Nature (London) **290**:304–310.
2. **Benz, E. W., Jr., and D. Dina.** 1979. Moloney murine sarcoma virions synthesize full genome length double-stranded DNA *in vitro*. Proc. Natl. Acad. Sci. U.S.A. **76**:3294–3298.
3. **Berns, A. J. M., M. H.-T. Lai, R. A. Bosselman, M. A. McKennett, L. T. Bacheler, H. Fan, E. C. Robanus Maandag, H.v.d. Putten, and I. M. Verma.** 1980. Molecular cloning of unintegrated and a portion of integrated Moloney murine leukemia viral DNA in bacteriophage lambda. J. Virol. **36**:254–263.
4. **Bishop, J. M.** 1978. Retroviruses. Annu. Rev. Biochem. **47**:35–88.
5. **Blattner, F. R., A. E. Blechl, K. Denniston-Thompson, H. E. Faber, J. E. Richards, J. L. Slightom, P. W. Tucker, and O. Smithies.** 1978. Cloning human fetal γ-globin and mouse α-type globin DNA: preparation and screening of shotgun collections. Science **202**:1279–1284.
6. **Bosselman, R. A., and I. M. Verma.** 1980. Genome organization of retroviruses. V. *In vitro* synthesized Moloney murine leukemia viral DNA has long terminal redundancy. J. Virol. **33**:487–493.
7. **Bukhari, A. I., J. A. Shapiro, and S. L. Adhya, ed.** 1977. DNA insertion elements, plasmids and episomes. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
8. **Calos, M. P., and J. H. Miller.** 1980. Transposable elements. Cell **20**:579–595.
9. **Chang, A. C. Y., J. H. Nunberg, R. J. Kaufman, H. A. Erlich, R. T. Schimke, and S. N. Cohen.** 1978. Phenotypic expression in *E. coli* of a DNA sequence coding for mouse dihydrofolate reductase. Nature (London) **275**:617–624.
10. **Corden, J., B. Wasylyk, A. Buchwalder, P. Sassone-Corsi, C. Kedinger, and P. Chambon.** 1980. Promoter sequences of eukaryotic protein-coding genes. Science **209**:1406–1414.
11. **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.
12. **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.
13. **Dunsmuir, P., W. J. Brorein, Jr., M. A. Simon, and G. M. Rubin.** 1980. Insertion of the *Drosophila* transposable element *copia* generates a 5 base pair duplication. Cell **21**:575–579.
14. **Farabaugh, P. J., and G. R. Fink.** 1980. Insertion of the eucaryotic transposable element Ty*1* creates a 5 base pair duplication. Nature (London) **286**:352–356.

14a. **Fuhrman, S. A., C. Van Beveren, and I. M. Verma.** 1981. Identification of an RNA polymerase II initiation site in the long terminal repeat of Moloney murine leukemia viral DNA. Proc. Natl. Acad. Sci. U.S.A. **78**:5411–5415.

15. **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.
16. **Gilboa, E., S. W. Mitra, S. Goff, and D. Baltimore.** 1979. A detailed model of reverse transcription and tests of crucial aspects. Cell **18**:93–100.
17. **Grosschedl, H., and M. L. Birnstiel.** 1980. Identification of

regulatory sequences in the prelude sequences of an H2A histone gene by the study of specific deletion mutants *in vivo*. Proc. Natl. Acad. Sci. U.S.A. **77**:1432–1436.

18. **Haegeman, G., and W. Fiers.** 1980. Characterization of the 5′-terminal cap structures of early Simian virus 40 mRNA. J. Virol. **35**:955–961.
19. **Hishinuma, F., P. J. DeBona, S. Astrin, and A. M. Skalka.** 1981. Nucleotide sequence of acceptor site and termini of integrated avian endogenous provirus ev-1: integration creates a 6 bp repeat of host DNA. Cell **23**:155–164.
20. **Hsu, T. W., J. L. Sabran, G. E. Mark, R. V. Guntaka, and J. M. Taylor.** 1978. Analysis of unintegrated avian RNA tumor virus double-stranded DNA intermediates. J. Virol. **28**:810–818.
21. **Hughes, S. H., P. R. Shank, D. H. Spector, H.-J. Kung, J. M. Bishop, H. E. Varmus, P. K. Vogt, and M. L. Breitman.** 1978. Proviruses of avian sarcoma viruses are terminally redundant, coextensive with unintegrated linear DNA and integrated at many sites. Cell **15**:1397–1410.
22. **Jones, M., R. A. Bosselman, F. A. v. d. Hoorn, A. Berns, H. Fan, and I. M. Verma.** 1980. Identification and molecular cloning of Moloney mouse sarcoma virus-specific sequences from uninfected mouse cells. Proc. Natl. Acad. Sci. U.S.A. **77**:2651–2655.
23. **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.
24. **Lowy, D. R., E. Rands, S. K. Chattopadhyay, C. F. Garon, and G. L. Hager.** 1980. Molecular cloning of infectious integrated murine leukemia virus DNA from infected mouse cells. Proc. Natl. Acad. Sci. U.S.A. **77**:614–618.
25. **Majors, J. E., and H. E. Varmus.** 1981. Nucleotide sequences at host-proviral junctions for mouse mammary tumor virus. Nature (London) **289**:253–258.
26. **Maxam, A. M., and W. Gilbert.** 1977. A new method for sequencing DNA. Proc. Natl. Acad. Sci. U.S.A. **74**:560–564.
27. **Proudfoot, N. J., and G. G. Brownlee.** 1976. 3′-non-coding region sequences in eukaryotic messenger RNA. Nature (London) **263**:211–214.
28. **Rands, E., D. R. Lowy, M. R. Lander, and S. K. Chattopadhyay.** 1981. Restriction endonuclease mapping of ecotropic murine leukemia viral DNAs: size and sequence heterogeneity of the long terminal repeat. Virology **108**:445–452.
29. **Reddy, E. P., M. J. Smith, E. Canaani, K. C. Robbins, S. R. Tronick, S. Zain, and S. A. Aaronson.** 1980. Nucleotide sequence analysis of the transforming region and large terminal redundancies of Moloney mouse sarcoma virus. Proc. Natl. Acad. Sci. U.S.A. **77**:5234–5238.
30. **Reddy, V. B., P. K. Ghosh, P. Lebowitz, M. Piatak, and S. M. Weissman.** 1979. Simian virus 40 early mRNAs. I. Genomic localization of 3′- and 5′-termini and two major splices in mRNA from transformed and lytically infected cells. J. Virol. **30**:279–296.
31. **Sabran, J. L., T. W. Hsu, C. Yeater, A. Kaji, W. S. Mason, and J. M. Taylor.** 1979. Analysis of integrated avian RNA tumor virus DNA in transformed chicken, duck and quail fibroblasts. J. Virol. **29**:170–178.
32. **Shank, P. R., S. H. Hughes, H.-J. Kung, J. E. Majors, N. Quintrell, R. V. Guntaka, J. M. Bishop, and H. E. Varmus.** 1978. Mapping unintegrated avian sarcoma virus DNA: termini of linear DNA bear 300 nucleotides present once or twice in two species of circular DNA. Cell **15**:1383–1395.
33. **Shapiro, J. A.** 1979. Molecular model for the transposition and replication of bacteriophage Mu and other transposable elements. Proc. Natl. Acad. Sci. U.S.A. **76**:1933–1937.
34. **Shimotohno, K., S. Mizutani, and H. M. Temin.** 1980. Sequence of retrovirus provirus resembles that of bacterial transposable elements. Nature (London) **285**:550–554.
35. **Shoemaker, C., S. Goff, E. Gilboa, M. Paskind, S. W. Mitra, and D. Baltimore.** 1980. Structure of a cloned circular Moloney murine virus DNA molecule containing an inverted segment: Implication for retrovirus integration. Proc. Natl. Acad. Sci. U.S.A. **77**:3932–3936.
36. **Sutcliffe, J. G., T. M. Shinnick, I. M. Verma, and R. A. Lerner.** 1980. Nucleotide sequence of Moloney leukemia virus: 3′-end reveals details of replication, analogy to bacterial transposons and an unexpected gene. Proc. Natl. Acad. Sci. U.S.A. **77**:3302–3306.
37. **Swanstrom, R., W. J. DeLorbe, J. M. Bishop, and H. E. Varmus.** 1981. Nucleotide sequence of cloned unintegrated avian sarcoma virus DNA: viral DNA contains direct and inverted repeats similar to those in transposable elements. Proc. Natl. Acad. Sci. U.S.A. **78**:124–128.
38. **Van Beveren, C., J. G. Goddard, A. Berns, and I. M. 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.
39. **Van Beveren, C., J. Galleshaw, V. Jonas, A. J. M. Berns, R. F. Doolittle, D. J. Donoghue, and I. M. Verma.** 1981. Nucleotide sequence and formation of the transforming gene of a mouse sarcoma virus. Nature (London) **289**:258–262.

39a. **Van Beveren, C., F. van Straaten, J. A. Galleshaw, and I. M. Verma.** 1981. Nucleotide sequence of the genome of a murine sarcoma virus. Cell **27**:97–108.

40. **Van der Putten, H., W. Quint, J. van Raaij, E. Robanus-Maandag, I. M. Verma, and A. Berns.** 1981. M-MuLV-induced leukemogenesis: integration and structure of recombinant proviruses in tumors. Cell **24**:729–739.
41. **Varmus, H. E., S. Heasley, H.-J. Kung, H. Oppermann, V. C. Smith, J. M. Bishop and P. R. Shank.** 1978. Kinetics of synthesis, structure and purification of avian sarcoma virus-specific DNA made in the cytoplasm of acutely infected cells. J. Mol. Biol. **120**:55–82.
42. **Verma, I. M.** 1977. The reverse transcriptase. Biochim. Biophys. Acta **473**:1–38.
43. **Verma, I. M., M. H.-T. Lai, R. A. Bosselman, M. A. McKennett, H. Fan, and A. Berns.** 1980. Molecular cloning of unintegrated Moloney mouse sarcoma virus DNA in bacteriophage lambda. Proc. Natl. Acad. Sci. U.S.A. **77**:1773–1777.
44. **Yamamoto, T., B. de Crombrugghe, and I. Pastan.** 1980. Identification of a functional promoter in the long terminal repeat of Rous sarcoma virus. Cell **22**:787–797.