JOURNAL OF VIROLOGY, June 1987, p. 1964–1971
0022-538X/87/061964-08$02.00/0

Vol. 61, No. 6

Retroviral DNA Integrated during Infection by an Integration-Deficient Mutant of Murine Leukemia Virus Is Oligomeric

KIMIKO HAGINO-YAMAGISHI,[1]† LAWRENCE A. DONEHOWER,[1]‡ AND HAROLD E. VARMUS[1,2]*

Departments of Microbiology and Immunology[1] and Biochemistry and Biophysics,[2] University of California, San Francisco, California 94143

Received 22 December 1986/Accepted 25 February 1978

Mutants with lesions engineered into the 3′ region of the *pol* gene of murine leukemia virus (MLV), the region that encodes p40^{pol}, have allowed us to define a function that is required for the normal integration of retroviral DNA. Two such mutants, MLV-SF1, which bears a single missense mutation (L. A. Donehower and H. E. Varmus, Proc. Natl. Acad. Sci. USA 81:6461–6465, 1984), and MLV-SF2, which has a frameshift mutation at the same site, have been used to explore the consequences of integration deficiencies. Pseudotypes that were formed with viral proteins supplied by these mutants and a replication-defective genome that carried a selectable genetic marker induced colony formation at least 100-fold less efficiently than did pseudotypes that were formed with proteins encoded by wild-type helper virus. Most of the proviruses from the mutant pseudotypes were manifestly aberrant when they were analyzed by mapping with restriction enzymes. Rare proviruses that appeared to be normal by restriction enzyme mapping were molecularly cloned and subjected to more detailed study. Two proviruses resulting from infection with the MLV-SF1 pseudotype were identical or nearly identical to wild-type MLV proviruses. Two base pairs were missing from the ends of the long terminal repeats at the host-viral junctions, and 4- or 5-base-pair duplications of host DNA flanked the proviruses. Thus, the missense mutant appears to retain a low level of normal integration activity. Four proviruses from cells that were infected with the MLV-SF2 pseudotype were flanked on both sides by viral DNA of either helper or vector origin, with a single long terminal repeat at each end of the intact vector DNA and with irregular junctions of viral and host DNAs. These results are most simply explained by the nonspecific integration of dimeric or trimeric forms of viral DNA, which may arise during infection by replicative or recombinational mechanisms.

During infection by retroviruses, the retrovirus RNA genomes are copied to form linear duplex DNA with long terminal repeats (LTRs). Conversion of linear DNA to covalently closed circles with one or two LTRs is followed by integration of viral DNA into host chromosomes, perhaps at random, producing a provirus that differs from linear DNA only by the loss of 2 base pairs (bp) from the termini of the LTRs (for a review, see reference 36). Results of genetic studies have indicated that the integration mechanism depends on at least two viral components: a function encoded in the 3′ portion of the *pol* reading frame (7, 24, 27) and a short sequence (called the circle junction) that includes a palindrome that is centered at the boundary between two LTRs in the larger species of monomeric circular DNA (5, 22, 23). Although the precise biochemical attributes of the integration function have not been determined, it is generally assumed to be responsible for accurate cutting of the proximal precursor to the provirus within the circle junction sequence and for staggered cleavage of chromosomal DNA to generate short duplications of 4, 5, or 6 bp that flank the provirus (36).

These assumptions have received partial support from results of enzymatic studies of the products of the avian and murine virus *pol* genes. The avian reverse transcriptase and a cleavage product called pp32^{pol} (15) nick single- and double-stranded circular DNAs preferentially near or at the correct sites within the circle junction sequence (8, 9, 14), and endonuclease activity has been attributed to a 46-kilodalton product of the 3′ end of the murine leukemia virus (MLV) *pol* gene (p46^{pol}) (17, 21).

We used site-directed mutagenesis of the *pol* gene of the Moloney strain of MLV (Mo-MLV) to define the integration function and examine the consequences of its deficiency (7). A missense mutation that changes a highly conserved arginine to cysteine at amino acid 114 of p46^{pol} incapacitates the replication competence of MLV. The mutated viral DNA can direct the synthesis of intact virus particles (MLV-SF1) that retain reverse transcriptase activity in vivo and in vitro, but little or no correct integration of newly synthesized viral DNA occurs. Because MLV-SF1 can synthesize *pol* proteins of normal size that may retain some integration functions, we also engineered a frameshift mutation at the same position in the *pol* gene. As expected, the new mutant DNA also generated particles (MLV-SF2) that behaved similarly to MLV-SF1.

Because it is difficult to detect and characterize rare integration events that may follow infection with MLV-SF1 or MLV-SF2, we used these mutants as helper viruses to generate pseudotypes with replication-defective MLV vectors carrying selectable markers, such as the herpes simplex virus (HSV) thymidine kinase gene (*tk*) or the bacterial neomycin phosphotransferase gene (*neo*). Under these conditions it is possible to isolate an occasional infected cell that contains proviral DNA that is synthesized and integrated

* Corresponding author.
† Present address: Department of Viral Oncology, Cancer Institute, Kami-Ikebukuro, Toshima-ku, Tokyo, Japan.
‡ Present address: Department of Virology, Baylor College of Medicine, Houston, TX 77030.

under the influence of the products of mutant *pol* genes (7); suitable proviruses can be molecularly cloned from such cells for further analysis.

By using this approach, we show here that MLV-SF1 is a slightly leaky mutant, because correct or nearly correct integration occurs at a low frequency. In contrast, no correct integrations were encountered with MLV-SF2 pseudotypes. However, we found several examples of vector DNA that appeared to have been integrated into chromosomal DNA as part of vector and helper DNA oligomers that were formed. The oligomeric precursors may be similar to the dimeric and trimeric closed circular forms of avian sarcoma virus DNA that have been reported previously to exist at low abundance in acutely infected cells (12, 18), and the existence of heterotrimers raises intriguing questions about the composition of retroviral RNA in infectious particles.

MATERIALS AND METHODS

Cells and viruses. Rat-2 cells (34) were grown in Dulbecco modified Eagle medium containing 10% fetal calf serum. COS-7 cells (11) were obtained from R. Tjian (University of California, Berkeley) via W. Lee (University of California, San Francisco) and grown in Dulbecco modified Eagle medium supplemented with 10% fetal calf serum and 3 g of glucose per liter. The cells that produced MLV were obtained by transfection of plasmid pZAP into rat-2 cells (7), and the medium was harvested from confluent virus-producing cells. Mutant viruses were sometimes concentrated by pelleting at 24,000 rpm in an SW27 rotor for 90 min. Infections with mutant MLVs were carried out with 5 to 10 ml of virus stock on plates (diameter, 100 mm) of semiconfluent cells in the presence of 8 μg of polybrene (Aldrich Chemical Co., Inc., Milwaukee, Wis.) per ml. Thymidine kinase-positive (TK^+) transformants were selected in medium containing hypoxanthine, aminopterin, and thymidine (25). G418-resistant colonies were selected in medium containing 250 to 300 μg of G418 (GIBCO Laboratories, Grand Island, N.Y.) per ml.

Cloned DNAs. pMLV-KT contains an integrated Mo-MLV provirus in which the *pol-env* region is replaced by the HSV *tk* gene in the opposite transcriptional orientation to the LTR (32). pZIPneoSVX (1) was obtained from C. Cepko. pFG5 contains the HSV *tk* *Bam*HI fragment (4). p8.2 is a clone of circular MLV DNA with one LTR opened at the *Hin*dIII site (29), and pZAP is a clone containing an infectious Mo-MLV provirus (32).

Generation of an *Sst*II frameshift mutant in pZAP. *Sst*II-digested pZAP was incubated with 7 U of T4 DNA polymerase (Bethesda Research Laboratories, Gaithersburg, Md.) for 2 min without deoxynucleotides; and then dATP, dCTP, dGTP, and TTP (75 μM each) were added for 30 min at 37°C. Reactions were carried out in 50 mM Tris acetate (pH 7.9), 50 mM potassium acetate, 5 mM magnesium acetate, and 2 mM dithiothreitol. The DNA was extracted with phenol followed by ethanol precipitation and was religated with 4×10^5 U of T4 ligase (New England BioLabs, Inc., Beverly, Mass.) for 3 h at 20°C. *Escherichia coli* HB101 was transformed by this DNA, and plasmids from transformants were screened for *Sst*II resistance.

DNA transfection. DNA transfections were carried out as described by Graham and Van der Eb (13). pMLV-KT or pZIPneoSVX (300 ng each) and 3 μg of pZAP or mutagenized derivatives of pZAP and 8 μg of salmon sperm DNA were cotransfected as calcium phosphate precipitates onto 10^6 rat-2 cells for 8 h.

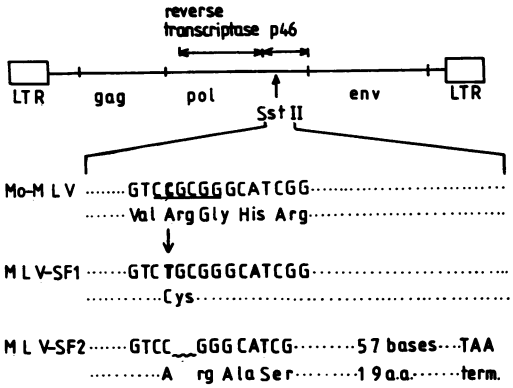

FIG. 1. Nature of the mutations in MLV-SF1 and MLV-SF2. The diagram shows an Mo-MLV provirus with LTRs (open boxes) encompassing the three retroviral coding domains; the position of the solitary *Sst*II site and the regions of *pol* encoding reverse transcriptase and p46 (the integration function) are indicated. Relevant nucleotide and deduced amino acid (a.a.) sequences are shown below for the wild-type and the two mutant viruses. The *Sst*II site (at positions 4949 to 4954 of the Mo-MLV sequence [28]) is underlined, and the altered nucleotides are highlighted; see text for further explanations.

COS-7 cell fusion rescue of provirus and flanking DNA sequences. A total of 10^6 G418-resistant cells and 10^6 COS-7 cells were mixed and plated onto a dish (diameter, 100 mm). After incubation for 2 days, the cells were fused by 50% (wt/vol) of polyethylene glycol 4000 (Merck & Co., Inc., Rahway, N.J.) in Dulbecco modified Eagle medium for 1 min at room temperature and washed 4 times with phosphate-buffered saline. The fused cells were incubated for an additional 3 days, unintegrated DNA from these cells was prepared as described by Hirt (16), the DNA was transfected onto *E. coli* HB101 (19), and kanamycin-resistant colonies were selected. Plasmid DNA from these colonies was analyzed by mapping with restriction enzymes, which were used according to specifications of the manufacturers.

DNA sequencing. DNA sequencing was done as described by Maxam and Gilbert (20) or Sanger et al. (26) with 20-base oligonucleotide primers for sequences near the ends of MLV LTRs (TGTATTTTTCCATGCCTTGC) and (TCCGACT TGTGGTCTCGCTG); the primers were purchased from the University of California, San Francisco, Biomolecular Resource Center.

RESULTS

MLV-SF1, a missense mutant in $p46^{pol}$, can occasionally mediate normal integration and hence appears leaky. We have reported previously (7) that an arginine to cysteine alteration at amino acid 114 of the 46-kilodalton protein encoded by the 3′ region of MLV *pol* (Fig. 1) impairs the ability of the resulting virus (MLV-SF1) to integrate viral DNA, although apparently normal unintegrated forms of viral DNA are synthesized at near normal efficiency. A pseudotype virus, MLV-KT (MLV-SF1), which is composed of the RNA genome of an MLV vector bearing the HSV *tk* gene in an opposing orientation and viral proteins encoded by the $p46^{pol}$ mutant, was more than 100-fold less effective than the pseudotypes that were formed with wild-type helper proteins in converting TK^- rat-2 cells to a TK^+ phenotype (Table 1). Moreover, in about 70% of the occasional TK^+ colonies, the vector DNA was structurally abnormal as gauged with restriction enzymes, implying aberrant integration (7).

TABLE 1. Efficiency of infection by the integrase-deficient pseudotype MLV-KT (MLV-SF2) is markedly impaired[a]

Virus stock	No. of TK^+ colonies/plate	No. of colonies containing potentially normal MLV-KT provirus/total
MLV-KT (Mo-MLV)	100–125	16/16
MLV-KT (MLV-SF1)	0–3	6/16
MLV-KT (MLV-SF2)	0–3	6/15

[a] pMLV-KT DNA was cotransfected into rat-2 cells with pZAP (wild type), pMLV-SF1, or pMLV-SF2 DNA. TK^+ cells were pooled and propagated to collect stocks of viral pseudotypes. Culture medium (5 ml) with similar levels of reverse transcriptase activity was used to infect 10^6 rat-2 cells per dish (diameter, 100 mm). TK^+ colonies were selected in medium containing hypoxanthine, aminopterin, and thymidine and were counted after 7 to 10 days. To determine the number of colonies that contained potentially normal MLV-KT proviruses, DNA from 15 or 16 colonies was analyzed (Fig. 2; see also Fig. 5 in reference 7).

To determine whether viral DNA in the remaining colonies was also integrated abnormally, with differences from wild-type proviruses that could not be detected by mapping with whole-cell DNA, we isolated molecular clones of DNA encompassing the virus-host junctions in two TK^+ colonies. DNA sequencing procedures were then applied to look for the signs that are characteristic of normal MLV integration events: the joining of viral DNA to host DNA at sites 2 bp from the ends of the LTRs and the duplication of 4 bp of host DNA (36). In both cases the correspondence between the determined sequences and those published previously for the LTR of Mo-MLV (6, 28, 35) disappeared exactly 2 bp from the distal ends of the LTRs, implying that the correct sites in viral DNA were used for integration of both proviruses (Table 2). In one of the cells, the same 4-bp sequence, presumably of host origin, was present on both sides of the provirus; in the other, identical 5-bp sequences were found flanking the provirus. We conclude that the MLV-SF1 mutant is slightly leaky, encoding a protein with a slightly altered sequence that can inefficiently mediate correct, or nearly correct, integration. (The significance of the potentially abnormal 5-bp duplication of host DNA is considered below.)

MLV-SF2, a frameshift mutant in $p46^{pol}$, has a phenotype similar to that of the missense mutant MLV-SF1. Because MLV-SF1 appeared to be a leaky mutant, we introduced a more radical change into the $p46^{pol}$ domain, thinking that a null mutation would allow us to examine the possibility that other cellular or viral functions might inefficiently complement an integration deficiency. We therefore engineered a frameshift mutation at the solitary *Sst*II site within the $p46^{pol}$ domain of pZAP, an infectious clone of Mo-MLV DNA (32), by using the exonuclease activity of T4 bacteriophage DNA polymerase to remove the two overhanging 3′ nucleotides at each end of the cleaved *Sst*II site, religating the DNA, and isolating plasmids that lack the *Sst*II site (see above). The expected mutation (Fig. 1) was confirmed by nucleotide sequencing (data not shown). By shifting the reading frame at this position, translation terminated 22 codons after the mutation, producing protein that contained only one-fourth the information that is encoded in $p46^{pol}$.

The mutated plasmid pMLV-SF2 was introduced into rat-2 (TK^-) cells by cotransfection with an HSV *tk* plasmid, and pooled colonies of TK^+ cells were shown to produce MLV-SF2 particles with reverse transcriptase activity; levels of activity were similar to those obtained from parallel cultures producing MLV-SF1 or wild-type Mo-MLV (data not shown). When these stocks of MLV-SF2 virus were used to infect rat-2 cells, full-sized linear and closed circular viral DNA was synthesized in amounts similar to those reported previously in cells infected by wild-type virus or MLV-SF1 (7); however, the proportion of circular DNA with a single LTR appeared to be less in cells that were infected by MLV-SF2. No virus production ensued in cells that were infected by either mutant (7) (data not shown). In concert, these results suggest that the defect in replication of MLV-SF2 affects a step between DNA synthesis and expression of viral genes, presumably proviral integration, as expected from results of similar studies with MLV-SF1 (7).

To examine the replication defect in MLV-SF2 more closely, we produced pseudotype viruses with MLV-KT genomes and proteins that were provided by the MLV-SF2 helper. The resulting pseudotypes, MLV-KT (MLV-SF2), were found to transform TK^- rat cells to a TK^+ phenotype about 100-fold less efficiently than did pseudotypes that were formed with the wild-type helper (Table 1). When DNA from several of the rare TK^+ clones of MLV-KT (MLV-SF2)-infected cells was examined with a restriction endonuclease (*Xba*I) that cleaves only in the LTR of MLV-KT, most samples yielded fragments that were different in size from the 8.3-kilobase (kb) product of digestion of correctly integrated MLV-KT DNA (Fig. 2), implying that the vector DNA was inserted aberrantly. In contrast, all of the examined MLV-KT proviruses established by wild-type pseudotypes (MLV-KT [Mo-MLV]) yielded the expected 8.3-kb *Xba*I fragment (Table 1). Thus, the helper activity provided by MLV-SF2 is inefficient and usually produces abnormal proviral DNA in the rare, successfully infected cells.

MLV-SF2 pseudotypes produce some proviruses that appear normal by restriction mapping. In some clones of MLV-KT (MLV-SF2)-infected TK^+ cells (ca. 40%; see Table 1 and Fig. 2, lanes c and f), the vector DNA appeared to be indistinguishable from normally integrated MLV-KT proviruses on digestion with *Xba*I, which is again similar to previously reported findings with pseudotypes formed with MLV-SF1 (7). Because the integration function of MLV-

TABLE 2. Host-viral junctions of MLV-KT proviruses from cells infected with MLV-KT (MLV-SF1) pseudotypes are normal or nearly normal[a]

Provirus	Cellular DNA	5′ LTR	3′ LTR	Cellular DNA
MLV linear DNA		AATGAAAGACCCCA---GGGGTCTTTCATT		
MLV-MSV proviral DNA	ATAAACG	TGAAAGACCCCA---GGGGTCTTTCA		AACGTA
MLV-KT (MLV-SF1) clone B	TTTATAA	TGAAAGACCCCA---GGGGTCTTTCA		ATAACTT
MLV-KT (MLV-SF1) clone E	GTGTTTG	TGAAAGACCCCA---GGGGTCTTTCA		GTTTGTA

[a] *Eco*RI fragments containing a potentially normal MLV-KT provirus from TK^+ colonies E and B infected with MLV-KT (MLV-SF1) were molecularly cloned and subjected to nucleotide sequencing (see text). Sequences at the viral-host junctions are shown in comparison with sequences of unintegrated linear DNA and the Mo-MSV provirus characterized by Dhar et al. (6). Nucleotides of viral origin are shown as capital letters; those of host origin are shown as small capital letters; flanking host sequences that are presumed to be duplicated during integration are underlined.

SF2, unlike that of MLV-SF1, seemed likely to be completely defective, we wished to determine whether the apparently normal proviruses were, in fact, incorrectly joined to cellular DNA, rather than correctly joined by an unexpected activity of viral or cellular origin.

To expedite the molecular cloning of proviral DNA that was required for this analysis, we generated new stocks of pseudotype virus with a genome that would permit cloning by direct transfection of bacteria with DNA from infected mammalian cells. Accordingly, the pZIPneoSVX plasmid, which harbors an MLV vector that is equipped with both simian virus 40 (SV40) and ColE1 origins and that has a gene (*neo*) that is selectable in both mammalian and bacterial cells, was used in place of pMLV-KT to cotransfect rat-2 cells with pMLV-SF2. Supernatants from pooled G418-resistant cells contained the viral pseudotypes ZIPneoSVX (MLV-SF2) and were used to infect rat-2 cells. After selection of the rare infected cells that were able to grow in the antibiotic G418, about 50% of the ZIPneoSVX proviruses were found to be grossly aberrant by restriction mapping with an enzyme (*Sac*I) that cleaves within the LTR of ZIPneoSVX DNA to produce a fragment of 4.2 kb (data not shown). The somewhat higher frequency of proviral DNA that appeared normal by this test in experiments with the ZIPneoSVX vector may be due to the smaller size of the vector provirus (4.8 versus 8.8 kb for MLV-KT) and to the requirement for a 5′ LTR to drive expression of the *neo* gene. (The HSV *tk* gene in MLV-KT is under the control of its native promoter.)

Preparatory to molecular cloning of the apparently normal ZIPneoSVX proviruses, we performed additional restriction

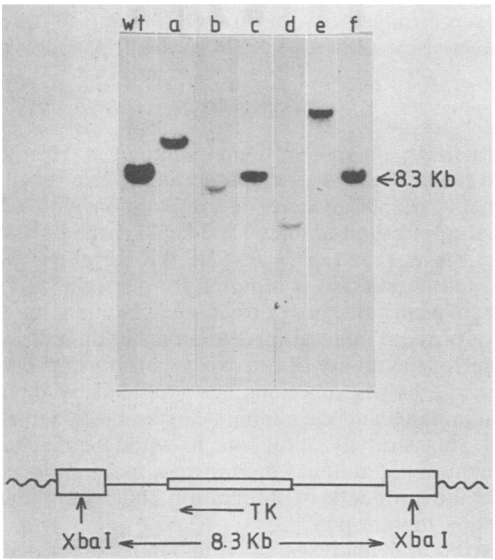

FIG. 2. Most MLV-KT proviruses in cells infected by MLV-KT (MLV-SF2) pseudotypes are aberrantly integrated. High-molecular-weight DNA was prepared from TK$^+$ clones of rat-2 cells that were infected with the viral pseudotypes that were formed with the integration-deficient helper MLV-KT (MLV-SF2) (lanes a to f) or with the wild-type helper MLV-KT (Mo-MLV) (lane wt) (see Table 1 and text). After digestion with *Xba*I, which cleaves MLV-KT DNA within each LTR as shown at the bottom of the figure, the DNA was electrophoresed in a 0.8% agarose gel, transferred to nitrocellulose, and analyzed by hybridization with a ^{32}P-labeled probe for HSV *tk*. The position of the 8.3-kb fragment, which is the expected product of digestion of a normally integrated provirus, is indicated.

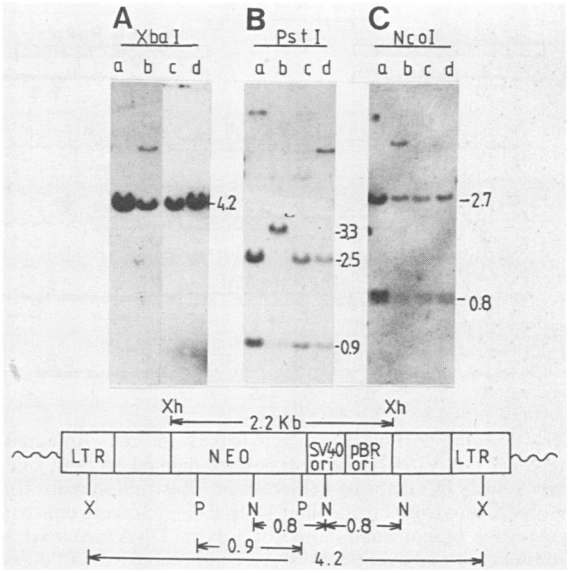

FIG. 3. Potentially normal ZIPneoSVX proviruses in cells infected by ZIPneoSVX (MLV-SF2) pseudotypes yield similar junction fragments on cleavage with restriction enzymes. High-molecular-weight DNA was prepared from four clones (a to d, as indicated in lanes a to d, respectively) of G418-resistant rat-2 cells after infection by the pseudotype ZIPneoSVX (MLV-SF2), which was digested with *Xba*I (X; panel A), *Pst*I (P; panel B), or *Nco*I (N; panel C), and analyzed as described in the legend to Fig. 2 with the 2.2-kb *Xho*I (Xh) fragment of pZIPneoSVX, which is shown above the diagram, used as the probe. The positions and origins of diagnostic internal fragments of ZIPneoSVX DNA are also indicated, as are the positions of important *Pst*I and *Nco*I fragments that extend beyond the LTR to the right (2.5- and 3.3-kb *Pst*I fragments) and to the left (2.7-kb *Nco*I fragment).

mapping of viral DNA in four independently infected, G418-resistant cell lines (clones a to d). *Xba*I, an enzyme that cleaves ZIPneoSVX DNA only in the LTR, produced the expected 4.2-kb internal fragment (Fig. 3A), confirming the results obtained with *Sac*I. The other two enzymes that were used (*Pst*I and *Nco*I) produced three types of fragment that were detectable with a probe prepared with the 2.2-kb internal *Xho*I fragment of ZIPneoSVX: the expected internal fragments from an intact ZIPneoSVX provirus (e.g., a 0.9-kb *Pst*I fragment and two 0.8-kb *Nco*I fragments); variably sized fragments that were novel to each clone, which were presumed to represent host-viral junction fragments that confirmed the independent origins of the clones; and unexpected, relatively small fragments, which were common to most or all clones, that appeared to extend just beyond an LTR of ZIPneoSVX either to the right (the 2.5-kb *Pst*I fragment in three of the four samples) or to the left (the 2.7-kb *Nco*I fragment in all four samples). These findings suggest that the ZIPneoSVX proviruses are immediately flanked on both the 3′ and 5′ sides by highly related sequences, although differences in flanking DNA are apparent at greater distances from the proviruses.

Mapping of molecularly cloned ZIPneoSVX proviruses reveals that they are flanked by viral DNA. To explain these perplexing observations, we recovered ZIPneoSVX DNA and flanking sequences from rat cell lines a to d by molecular cloning in *E. coli*. Each cell line was fused with COS-7 cells to amplify and excise ZIPneoSVX and adjacent DNA (1), and DNA that was prepared from the Hirt (16) supernatant

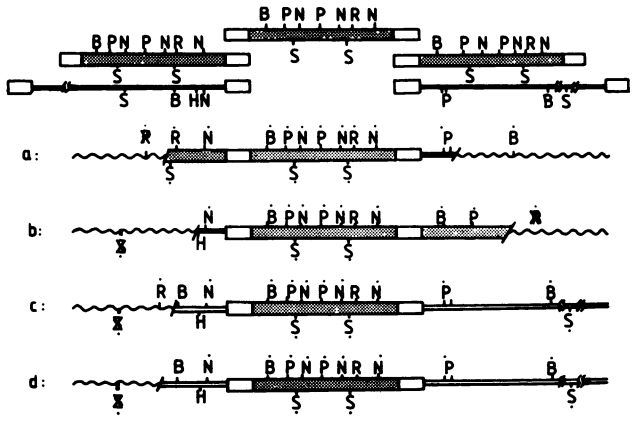

FIG. 4. Some ZIPneoSVX proviruses in cells infected by ZIPneoSVX (MLV-SF2) pseudotypes are flanked on both sides by viral sequences of vector or helper origin. Restriction maps for the ZIPneoSVX proviruses from clones a to d (Fig. 3) were constructed from digestion of molecular clones of proviral DNA (see text); most sites (letters with adjacent dots) were confirmed by direct digestion of unfractionated cellular DNA from each cell line. To facilitate interpretation of the maps, diagrams of vector (ZIPneoSVX) DNA (wide stippled internal domain) and helper (MLV-SF2) DNA (narrow open internal domain) are shown, with LTRs indicated as open boxes. Wavy lines denote flanking host cellular DNA, with slashes indicating the uncertain positions of host-viral junctions. Host DNA was identified by the presence of restriction sites that were not expected in either form of viral DNA or by the absence of sites that were predicted to be in viral DNA (indicated by an X overlying the symbol for a restriction enzyme). Restriction enzyme abbreviations: S, *Sph*I; N, *Nco*I; P, *Pst*I; H, *Hpa*I; R, *Eco*RI; B, *Bam*HI.

fractions was used to transfect *E. coli* and select for resistance to kanamycin. On examination by restriction enzyme mapping, most of the recovered plasmids (over 90%) were 4.2 kb in size; contained one LTR and no flanking DNA; and hence appeared to have been generated, as expected, by homologous recombination between the LTRs of ZIPneoSVX DNA. However, 2 to 5% of the recovered plasmids contained either one LTR and flanking sequences or both LTRs and sequences flanking both ends of ZIPneoSVX DNA. Several independent plasmids encompassing one or both ends of ZIPneoSVX DNA from each cell line were then analyzed with over 15 restriction enzymes. In addition, restriction sites for several enzymes (including *Hin*dIII, *Pst*I, *Nco*I, *Eco*RI, *Bam*HI, and *Sph*I) were confirmed by examination of unfractionated DNA from all four of the original cell lines.

The results of these extensive analyses are presented in Fig. 4, in conjunction with the relevant maps of ZIPneoSVX and MLV-SF2 DNAs. The cardinal feature in each case is the presence of viral DNA, originating from either vector or helper virus, on both sides of an apparently intact ZIPneoSVX provirus, with all of the viral DNA at each integration site being in the same transcriptional orientation. The viral DNAs adjacent to the intact ZIPneoSVX DNA are joined to heterogeneous cellular DNA at various positions within their internal regions. Inspection of the maps in Fig. 4 reveals that all of the commonly sized *Pst*I and *Nco*I fragments from different cell lines (Fig. 3) can be explained as fragments that contain LTRs flanked on both sides by viral DNA.

In clone a the 5′ LTR is flanked by a portion of the 3′ end of the internal region of ZIPneoSVX DNA, and the 3′ LTR is flanked by a short portion of the 5′ end of the internal region of MLV-SF2 DNA. In clone b the situation is reversed, with a short region from the 3′ end of MLV-SF2 flanking the 5′ LTR and a portion of the 5′ end of ZIPneoSVX flanking the 3′ LTR. In clones c and d helper DNA flanks both the 5′ and 3′ LTRs of ZIPneoSVX; although the configuration of viral DNA is similar in these two clones, there are differences in the flanking cellular DNA (e.g., the size difference between high-molecular-weight fragments of clones c and d generated by *Pst*I digestion [Fig. 3] and the *Eco*RI sites on the 5′ side [Fig. 4]) that validate the claim that the clones are independently derived. Because the viral DNAs that flank intact vector DNA in clones a and b have different origins, it is likely that the insertions are derived from heterotrimeric DNA; the insertions in clones c and d could have been derived from DNA with 1 or 2 units of helper DNA and 1 unit of vector DNA. (The mechanisms by which these oligomers might have been formed are considered below.)

The LTRs in an oligomeric insert do not show signs of retrovirus-mediated recombination. The tandem arrays of viral DNA illustrated in Fig. 4 could have arisen in part by the integration of molecules of viral DNA into each other via the specialized recombination system that normally mediates integration of retroviral DNA into chromosomal DNA at various sites. Although this seemed inherently unlikely in the presumed absence of the integration function, we sought the nucleotide sequence arrangements that are characteristic of normally integrated MLV DNA. Nucleotide sequencing of both ends of the 5′ and 3′ LTRs in the viral DNA cloned from line a revealed no loss of nucleotides from the LTR and no directly repeated sequences flanking the LTRs (data not shown). Apparently, mechanisms other than the normal form of retrovirus-mediated integration must be invoked to explain the arrangements of viral DNA that are observed in cells infected by ZIPneoSVX (MLV-SF2) pseudotypes.

DISCUSSION

Mutations in the p46^{pol} domain impair a viral function essential for normal retroviral integration. The results of the experiments described here, in conjunction with results of previously published studies (7, 24, 27), argue strongly for the hypothesis that the 3′ end of the retroviral *pol* gene encodes a function that is required for a specialized integration mechanism. Because direct measurement of the efficiency of proviral integration is technically difficult, we took advantage of the ability of retroviruses to form pseudotypes in which the helper functions are provided by integration-deficient mutants and the genome carries a selectable genetic marker. This strategy allows us to estimate the degree of impairment of the mutant function and, more importantly, to examine the products of integration that occur under the influence of mutant p46^{pol}.

The structural analyses of viral DNA integrated in the presence of our mutants, MLV-SF1 and MLV-SF2, confirm the importance of the p46^{pol} coding domain for correct integration. The missense mutant MLV-SF1, with an arginine residue that is highly conserved among retroviral *pol* products replaced by a cysteine residue, directs integration inefficiently but occasionally correctly. Pseudotypes that were formed with this mutant displayed a marked impairment of infectivity, and most infected cells acquired blatantly aberrant proviruses; but we recovered at least one provirus that appeared to be entirely normal, as judged by restriction mapping and sequencing of host-viral junctions. Another provirus exhibited a minor deviation of a 5-bp,

rather than a 4-bp, duplication of host DNA at the integration site. Because Colicelli and Goff (5) encountered a 5-bp duplication after integration of viral DNA with a mutant circle junction sequence, it is tempting to propose that the difference in length is a manifestation of abnormal integration machinery; however, 5-bp duplications of the target sequence have also been occasionally observed to be flanking products of presumably normal integration reactions (35). We conclude that MLV-SF1 has a markedly impaired, but slightly leaky, capacity to integrate retroviral DNA properly.

In contrast, we discovered no correct products of integration when helper functions were provided to the pseudotypes by a mutant, MLV-SF2, with a frameshift lesion near the 5′ end of the $p46^{pol}$ coding region. Again, the efficiency of infection was low, and most of the rare integration events produced frankly aberrant proviruses that presumably resulted from the use of alternative host recombination systems. The only proviruses that initially appeared normal by restriction mapping proved to be unusual hetero-oligomers of viral DNA of uncertain origins (see below). Thus, ablation of the ability to synthesize $p46^{pol}$ seems to prevent correct retroviral integration, implying that the protein is essential for the integration mechanism and is unlikely to be complemented by other viral or host functions.

The mechanism for generation of integrated hetero-oligomeric DNA. Several kinds of events can be envisioned to produce the sorts of integrated, oligomeric viral DNAs illustrated in Fig. 4.

(i) Homologous integration. Integration of free monomeric DNA into a related provirus by homologous recombination could explain at least some of our findings. For example, if a circular form of helper DNA containing a single LTR was first integrated aberrantly to form a permuted provirus, then homologous recombination between the internal LTR and the LTR in circular vector DNA would produce arrangements that were similar to those in clones c and d (Fig. 5, left). More baroque schemes are required to explain the arrangements of helper and vector DNA that were observed in clones a and b. In addition, the clones that we examined could be generated only by the sequential integration of circular DNAs with single LTRs, but there is no a priori reason to exclude the use of circles with two LTRs (and such circles are more abundant than those with one LTR in cells that are infected with SF2).

This mechanism, of course, presumes that homologous recombination can occur with reasonable efficiency between previously integrated and unintegrated retroviral DNA. Homology has not been observed to have a role during infection with wild-type virus, even in cells containing many copies of related endogenous proviruses, although its influence might be obscured in the face of an intact system for normal, nonhomologous, retroviral integration. Moreover, newly synthesized viral DNA is present in subviral nucleoprotein complexes (B. Bowerman and H. E. Varmus, unpublished data), in which it may be relatively inaccessible to host enzymes that mediate homologous recombination. On the other hand, there was strong selection in our experiments for rare integration events that generated an intact transcriptional unit for the *neo* gene. Homologous recombination between DNA that is introduced by transfection or by microinjection and DNA that is present in the host chromosome has been observed to occur with frequencies that might account for at least some of our findings (30, 33).

(ii) Formation and integration of unintegrated oligomeric forms of viral DNA. Less complicated schemes for generating oligomeric integrated species of viral DNA depend on the formation of oligomers prior to integration. Such schemes are favored by previously published reports (12, 18) of dimeric and trimeric circular forms of viral DNA that have been observed during infection of avian cells with wild-type Rous sarcoma virus. Although these species are too rare to be analyzed in great detail, the viral monomeric components are known to be arranged in a head-to-tail fashion (18). Integration of such oligomeric circles by random crossing-over with chromosomal DNA could explain the unusual integrated DNA that we observed (Fig. 5, center), provided that (i) only one LTR separates each internal region of viral DNA; (ii) recombination leaves an intact copy of ZIPneoSVX DNA; and (iii) heterodimeric and heterotrimeric molecules, containing both helper and vector DNA, can form. Similar schemes can be envisioned to involve unintegrated linear DNA (Fig. 5, right), but there is no experimental evidence for such forms.

How might hetero-oligomeric circular DNAs be generated? If they arise during synthesis of viral DNA, it would seem necessary to invoke both the existence of heterozygotic particles (which already have been predicted to occur as intermediates in high-frequency retroviral recombination [2]) and the packaging of at least three templates for reverse transcription in a single particle. Retroviral genomes are known to be diploid (3), but it has not been possible to establish whether some particles might incorporate three or four subunits of viral RNA, rather than just the two that are required for dimerization. In addition, subgenomic viral mRNA is sometimes associated with the 70S complex of retroviral RNA (31). Thus, a single particle could supply templates for the synthesis of heterotrimeric DNA, although the mechanics of synthesis are entirely speculative. We cannot completely exclude the possibility that the SV40 *ori*

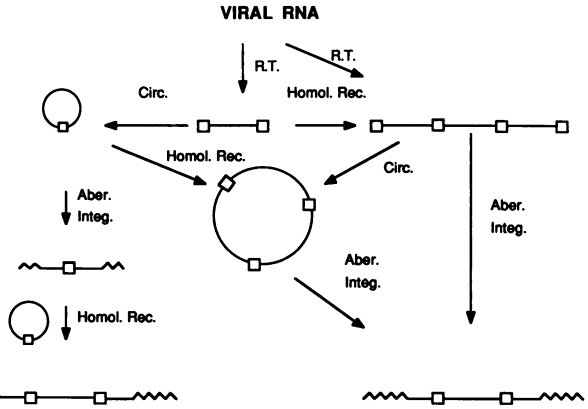

FIG. 5. Possible pathways to integrated oligomeric forms of viral DNA. On the left, aberrant integration (Aber. Integ.) of monomeric circular DNA produces a permuted species; homologous recombination (Homol. Rec.) with another monomeric circle (e.g., by crossing-over between the LTRs) could produce the form observed in Fig. 4c and d. Unintegrated oligomeric forms of viral DNA can be envisioned as arising either by synthesis of oligomeric linear DNA by reverse transcription of viral RNA (R.T.; top right), perhaps followed by circularization (Circ.) to form an oligomeric circle (center), or by homologous recombination between monomeric circular forms (top left). Aberrant integration of the oligomeric DNAs into the chromosome, with concurrent deletion of sequences at the viral integration sites, could also produce the forms observed in Fig. 4. Symbols: smooth lines, viral DNA; open boxes, LTRs; wavy lines, host cellular DNA.

in ZIPneoSVX had a role in generating the oligomers via a replication mechanism, but it is important to recall that the viral DNA flanking ZIPneoSVX proviruses was most commonly derived from helper DNA, lacking SV40 *ori*. Moreover, rat cells are nonpermissive for replication of SV40 DNA, and T antigen was not available.

Hetero-oligomeric circles could also arise by homologous recombination between newly synthesized monomeric circles. There are strong precedents for such recombination reactions that have been obtained from the study of naked DNA that is introduced into cells by transfection or microinjection procedures (10). Our recent findings that both linear and circular forms of viral DNA are largely confined to subviral particles in newly infected cells (Bowerman and Varmus, unpublished data) suggest that the products of reverse transcription might be shielded from such recombinational activities. The fate of retroviral DNA in the absence of a functional integration system, however, has not been determined; it is possible that complete uncoating ultimately renders the DNA susceptible to homologous recombination systems that could generate suitable oligomeric precursors to the integrated forms that we have observed.

ACKNOWLEDGMENTS

We thank William Lee, Cliff Tabin, and Connie Cepko for gifts of reagents; Pat Brown, Bruce Bowerman, and Shantanu Basu for advice; and Janine Marinos for assistance with the manuscript.

K.H.-Y. was the recipient of a Toyobo Biotechnology Foundation fellowship, and H.E.V. is an American Cancer Society research professor. This study was supported by Public Health Service grants from the National Institutes of Health.

LITERATURE CITED

1. **Cepko, C. L., B. E. Roberts, and R. C. Mulligan.** 1984. Construction and application of a highly transmissible murine retrovirus shuttle vector. Cell **37**:1053–1062.
2. **Coffin, J. M.** 1979. Structure, replication, and recombination of retrovirus genomes: some unifying hypotheses. J. Gen. Virol. **42**:1–26.
3. **Coffin, J. M.** 1982. Structure of the retroviral genome, p. 261–368. *In* R. Weiss, N. Teich, H. Varmus, and J. Coffin (ed.), RNA tumor viruses. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
4. **Colbere-Garapin, F., F. Horodniceau, P. Kourilsky, and A. C. Garapin.** 1981. A new dominant hybrid selective marker for higher eukaryotic cells. J. Mol. Biol. **150**:1–4.
5. **Collicelli, J., and S. P. Goff.** 1985. Mutants and pseudorevertants of Moloney murine leukemia virus with alterations at the integration site. Cell **42**:573–580.
6. **Dhar, R., W. L. McClements, L. W. Enquist, and G. F. Vande Woude.** Nucleotide sequences of integrated Moloney sarcoma provirus long terminal repeats and their host and viral junctions. Proc. Natl. Acad. Sci. USA **77**:3937–3941.
7. **Donehower, L. A., and H. E. Varmus.** 1984. A mutant murine leukemia virus with a single missense codon in *pol* is defective in a function affecting integration. Proc. Natl. Acad. Sci. USA **81**:6461–6465.
8. **Duyk, G., J. Leis, M. Longiaru, and A. M. Skalka.** 1983. Selective cleavage on the avian retroviral long terminal repeat sequence by the endonuclease associated with the form of avian reverse transcriptase. Proc. Natl. Acad. Sci. USA **80**:6745–6749.
9. **Duyk, G., M. Longiaru, D. Cobrinik, R. Kowal, P. DeHaseth, A. M. Skalka, and J. Leis.** 1985. Circles with two tandem long terminal repeats are specifically cleaved by *pol* gene-associated endonuclease from avian sarcoma and leukosis virus: nucleotide sequences required for site-specific cleavage. J. Virol. **56**:589–599.
10. **Folger, K. R., E. A. Wong, G. Wahl, and M. R. Capecchi.** 1982. Patterns of integration of DNA microinjected into cultured mammalian cells. Evidence for homologous recombination between injected plasmid DNA molecules. Mol. Cell. Biol. **2**:1372–1387.
11. **Gluzman, Y.** 1981. SV40-transformed simian cells support the replication of early SV40 mutants. Cell **23**:175–182.
12. **Goubin, G., and M. Hill.** 1979. Monomer and multimer covalently closed circular form of Rous sarcoma virus DNA. J. Virol. **29**:799–804.
13. **Graham, F. L., and A. J. Van der Eb.** 1973. A new technique for the assay of infectivity of human adenovirus 5 DNA. Virology **52**:456–467.
14. **Grandgenett, D. P., and A. C. Vora.** 1985. Site-specific nicking at the avian retrovirus LTR circle junction by the viral pp32 DNA endonuclease. Nucleic Acids Res. **13**:6205–6221.
15. **Grandgenett, D. P., A. C. Vora, and R. D. Schiff.** 1978. A 32,000-dalton nucleic acid-binding protein from avian retrovirus cone possesses DNA endonuclease activity. Virology **89**:119–132.
16. **Hirt, B.** 1967. Selective extraction of polyoma DNA from infected mouse cell culture. J. Mol. Biol. **26**:365–369.
17. **Kopchick, J. J., J. Harless, B. S. Geisser, R. Killam, R. R. Hewitt, and R. B. Arlinghaus.** 1981. Endodeoxyribonuclease activity associated with Rauscher murine leukemia virus. J. Virol. **37**:274–283.
18. **Kung, H.-J., P. R. Shank, J. M. Bishop, and H. E. Varmus.** 1980. Identification and characterization of dimeric and trimeric circular forms of avian sarcoma virus-specific DNA. Virology **103**:425–433.
19. **Lederberg, E. M., and S. N. Cohen.** 1974. Transformation of *Salmonella typhimurium* by plasmid deoxyribonucleic acid. J. Bacteriol. **119**:1072–1074.
20. **Maxam, A., and W. Gilbert.** 1980. Sequencing endolabeled DNA with base-specific chemical cleavage. Methods Enzymol. **65**:499–560.
21. **Nissen-Meyer, J., and I. F. Nes.** 1980. Purification and properties of DNA endonuclease associated with Friend leukemia virus. Nucleic Acids Res. **8**:5043–5055.
22. **Panganiban, A. T., and H. M. Temin.** 1983. The terminal nucleotides of retrovirus DNA are required for integration but not virus production. Nature (London) **306**:155–160.
23. **Panganiban, A. T., and H. M. Temin.** 1984. Circles with two tandem LTRs are precursors to integrated retrovirus DNA. Cell **36**:673–679.
24. **Panganiban, A. T., and H. M. Temin.** 1984. The retrovirus *pol* gene encodes a product required for DNA integration: identification of a retrovirus *int* locus. Proc. Natl. Acad. Sci. USA **81**:7885–7889.
25. **Pellicer, A., M. Wigler, R. Axel, and S. Silverstein.** 1978. The transfer and stable-integration of the HSV thymidine kinase gene into mouse cell. Cell **14**:133–141.
26. **Sanger, F., S. Nicklen, and A. R. Coulson.** 1977. DNA sequencing with chain terminating inhibitors. Proc. Natl. Acad. Sci. USA **74**:5463–5467.
27. **Schwartzberg, P., J. Collicelli, and S. P. Goff.** 1984. Construction and analysis of deletion mutations in the *pol* gene of Moloney murine leukemia virus: a new viral function required for productive infection. Cell **37**:1043–1052.
28. **Shinnick, T. M., R. A. Lerner, and J. Sutcliffe.** 1981. Nucleotide sequence of Moloney murine leukemia virus. Nature (London) **293**:543–548.
29. **Shoemaker, C. S., S. Goff, E. Gilboa, M. Paskind, S. W. Mitra, and D. Baltimore.** 1980. Structure of a cloned circular Moloney murine leukemia virus molecule containing an inverted segment: implications for retrovirus integration. Proc. Natl. Acad. Sci. USA **77**:3932–3936.
30. **Smithies, O., R. G. Gregg, S. S. Boggs, M. A. Koralewski, and R. S. Kucherlapati.** 1985. Insertion of DNA sequences into the human chromosomal β-globin locus by homologous recombination. Nature (London) **317**:230–234.
31. **Stacey, D. W.** 1979. Messenger activity of virion RNA for avian leukosis viral envelope glycoprotein. J. Virol. **29**:949–956.
32. **Tabin, C. J., J. W. Hoffman, S. Goff, and R. Weinberg.** 1982.

Adaptation of a retrovirus as a eukaryotic vector transmitting the herpes simplex virus thymidine kinase gene. Mol. Cell. Biol. **2**:426–436.

33. **Thomas, K. R., K. R. Folger, and M. R. Capecchi.** 1986. High frequency targeting of genes to specific sites in the mammalian genoime. Cell **44**:419–428.
34. **Topp, W. C.** 1981. Normal rat cell lines deficient in nuclear thymidine kinase. Virology **113**:408–411.
35. **Van Beveren, C., E. Rands, S. K. Chattopadhyay, D. R. Lowy, and I. M. Verma.** 1982. Long terminal repeat of murine retroviral DNAs: sequence analysis, host-proviral junctions, and preintegration site. J. Virol. **41**:542–556.
36. **Varmus, H. E., and R. Swanstrom.** 1985. Replication of retroviruses p. 369–512 (vol. 1); p. 75–134 (vol. 2). *In* R. Weiss, N. Teich, H. Varmus, and J. Coffin (ed.), RNA tumor viruses. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.