

Rous Sarcoma Virus (RSV) Integration In Vivo: a CA Dinucleotide Is Not Required in U3, and RSV Linear DNA Does Not Autointegrate[∇]

Jangsuk Oh,¹ Kevin W. Chang,¹ Rafal Wierzchoslawski,¹
W. Gregory Alvord,² and Stephen H. Hughes^{1*}

HIV Drug Resistance Program¹ and Data Management Services,² National Cancer Institute at Frederick,
Frederick, Maryland 21702-1201

Received 2 July 2007/Accepted 15 October 2007

The sequences required for integration of retroviral DNA have been analyzed in vitro. However, the in vitro experiments do not agree on which sequences are required for integration: for example, whether or not the conserved CA dinucleotide in the 3' end of the viral DNA is required for normal integration. At least a portion of the problem is due to differences in the experimental conditions used in the in vitro assays. To avoid the issue of what experimental conditions to use, we took an in vivo approach. We made mutations in the 5' end of the U3 sequence of the Rous sarcoma virus (RSV)-derived vector RSVP(A)Z. We present evidence that, in RSV, the CA dinucleotide in the 5' end of U3 is not essential for appropriate integration. This result differs from the results seen with mutations in the U5 end, where the CA appears to be essential for proper integration in vivo. In addition, based on the structure of circular viral DNAs smaller than the full-length viral genome, our results suggest that there is little, if any, integrase-mediated autointegration of RSV linear DNA in vivo.

The retroviral life cycle is characterized by the conversion of the single-stranded RNA genome found in virions into a double-stranded linear DNA that is subsequently integrated into the host cell genome. Viral DNA synthesis takes place in the cytoplasm of the infected cell. The RNA genome of the virus is copied into DNA by a virally encoded enzyme, reverse transcriptase (RT). RT contains two enzymatic activities, a DNA polymerase that can copy either an RNA or a DNA template, and an RNase H that cleaves RNA only if the RNA is part of an RNA/DNA hybrid. Like many other DNA polymerases, RT requires both a template and a primer. First (minus)-strand DNA synthesis is initiated from a cellular tRNA primer that is base paired at the primer binding site (PBS), which is near the 5' end of the viral genomic RNA. During the reverse transcription process, this tRNA primer is removed from the end of the minus-strand DNA by RNase H; this defines the right (U5) end of the linear viral DNA. Second (plus)-strand DNA synthesis is initiated from a polypurine tract (PPT) primer adjacent to U3. The RNase H cleavages that generate and remove PPT primer define the left (U3) end of the linear viral DNA (31).

After reverse transcription, the linear viral DNA is transported to the nucleus, where the viral DNA is inserted into the host genome in a reaction mediated by the viral enzyme integrase (IN). Retroviral integration is an essential step in the viral life cycle. The integrated viral DNA is called the provirus. The provirus is copied into RNA by the host DNA-dependent RNA polymerase. This viral RNA then is used both as the viral genome and as viral mRNA (2, 8, 9, 17, 27, 30). In the integration step, IN recognizes sequences at the ends of linear viral

DNA, termed the attachment site (*att*), and usually removes two nucleotides adjacent to the conserved CA dinucleotide on each of the 3' ends of the linear viral DNA (6, 11, 15, 28). In the next step, the processed ends of the viral DNAs are joined to the host chromosomal DNA. The reaction catalyzed by IN leaves short gaps between the viral DNA and host DNA which are repaired by host enzymes, creating a short direct repeat (4 to 6 bp) of the host sequence that flanks the integrated provirus (7, 13, 14, 28).

The sequences required for integration have been analyzed in vitro for both avian leukosis-sarcoma virus (ASLV) and human immunodeficiency virus type 1 (HIV-1) IN. In integration reactions carried out in vitro, the canonical CA in the viral sequence is thought to be important; when the CA of either the U5- or the U3-derived *att* site is changed, several groups have reported that 3' processing and strand transfer are significantly impaired (10, 16, 18, 32). However, different researchers have reported different sequence requirements for in vitro integration by HIV-1 IN: for example, that the CA is required (10) or that it is not (33). Zhou et al. (33) showed that sequences other than the CA dinucleotide can be successfully integrated by ASLV IN in an in vitro reaction. It is possible that some of the differences seen in the in vitro assays come from the fact that different laboratories have used different preparations of IN and have used somewhat different conditions for the integration reactions. One solution to this problem is to determine the sequence requirements in vivo. However, at least for ASLV, a linear viral DNA with one defective end can be integrated in the host genome with good efficiency even though the aberrant end is not inserted by IN (24). For this reason, titer is not sufficient to determine whether IN can properly insert an aberrant end. Colicelli and Goff (4, 5) reported that murine leukemia virus (MLV) could, with reduced efficiency, replicate if there were one or four nucleotides between the CA and the PBS but not with an end with no nucleotides between the CA and the PBS. Analysis of autointegrants suggested that MLV

* Corresponding author. Mailing address: HIV Drug Resistance Program, NCI at Frederick, P.O. Box B, Bldg. 539, Rm. 130A, Frederick, MD 21702-1201. Phone: (301) 846-1619. Fax: (301) 846-6966. E-mail: Hughes@ncifcrf.gov.

[∇] Published ahead of print on 24 October 2007.

IN could remove one or four nucleotides beyond the CA; however, it was not clear in these experiments whether a CA was required for IN-mediated integration of MLV DNA (4, 5). In HIV-1, changing the CA dinucleotides at both ends of the linear HIV-1 DNA to TG caused a replication defect and the resulting virus had an integration efficiency approximately 1% of that of the wild-type virus (1, 21). However, mutating the canonical CA sequence at only one end of HIV-1 linear DNA gave rise to viruses with an integration efficiency about 10 to 30% of that of the wild type (1, 22). In these experiments, the structure of the proviruses produced by inserting HIV-1 viral DNAs with one defective end was not determined. For this reason, it is not clear whether mutating either of the conserved CAs found at the ends of the HIV-1 linear DNA leads to the generation of normal or aberrant proviruses. In the experiments with RSV(A)Z, we showed that the CA in U5 was essential for efficient integration *in vivo*. However, the sequences of the U5 and U3 ends of the RSV(A)Z vector are different, and it was not clear that mutating the CA in U3 would have the same effect on the integration of viral DNA.

Here, we made mutations that changed the sequence of the 5' end of the U3 sequence of RSV(A)Z and analyzed the effects of the mutations on the viral titer, two-long terminal repeat (two-LTR) circle junctions, and integration. RSV(U3CATG) and RSV(U3CATC) did not measurably affect the relative virus titer. The relative titer of RSV(U3TCTT) was about 75% of that of the wild type; the titer of RSV(U3CAT) was about half of that of the wild type. The two-LTR circles are derived by the ligation of the ends of the linear viral DNA and can be used as surrogates that can be used to infer the sequences at the ends of the linear viral DNA, which are difficult to analyze directly. The two-LTR circle junction sequences that we obtained suggest that RSV RNase H removes the entire PPT primer regardless of the sequence changes in U3. In the case of RSV(U3CAT), in which there is only a single nucleotide beyond the CA dinucleotide at the U3 end, 30% of the proviruses that we isolated were from aberrant integrations in which the insertion of the U3 end into host DNA apparently did not involve viral IN. These aberrant integrations were similar to the aberrant integrations that we reported in experiments involving mutant viruses with aberrant U5 ends, in which there is one nucleotide between the CA and the PBS (24). However, in the case of RSV(U3TCTT), in which the CA of the U3 was mutated to TC, about 60% of the proviruses were properly integrated using the TC and the resulting proviruses were flanked by a normal 5- or 6-bp duplication of host sequence. This result was different from our recent finding with RSV(CATC), in which the CA sequence of U5 was changed to TC (24). All of the proviruses derived from infections with RSV(CATC) were aberrant. This result suggests that the conserved CA dinucleotide is not required for appropriate integration of the U3 end *in vivo*, although, in U5, the conserved CA dinucleotide was essential.

In the nucleus of an infected cell, some of the unintegrated linear viral DNA is converted into several different kinds of circular DNAs; one-LTR circles are formed by homologous recombination between the LTRs of the linear viral DNA, two-LTR circles are formed by joining the ends of the linear viral DNA, and other circular forms are generated both by IN-mediated autointegration and by other, less well defined

processes. If there is IN-mediated autointegration, products can be formed by the integration of one or both of the viral DNA ends into viral DNA. In the present study, we isolated 130 smaller circular viral DNAs, in which aberrant U5 ends were joined to various positions of the viral DNA, deleting a portion of the U3 end of the viral DNA. The sequences of the viral/viral DNA junctions show that the conserved CA sequence used in the integration of the linear viral DNA into host DNA was not used to generate any of these smaller viral DNAs. Normal viral DNA integration does not involve microhomology; however, most of the viral/viral DNA junctions in the smaller circular viral DNAs involved microhomologies that ranged in size from 1 to 5 nucleotides. These results suggest that, in contrast to the linear viral DNAs of other retroviruses, RSV linear DNA does little, if any, IN-mediated autointegration *in vivo*.

MATERIALS AND METHODS

Plasmid construction. Mutations were introduced into the 5' end of the U3 sequence next to the PPT using *in vitro* site-directed mutagenesis with KS/RSV(A)Z as a template and appropriate sets of primers using the QuikChange site-directed mutagenesis kit (Stratagene, La Jolla, CA), following the manufacturer's recommendations. KS/RSV(A)Z has a 0.9-kb MluI-to-SacI fragment containing the PPT, both LTRs, and the PBS region of RSV(A)Z (24). The resulting plasmids (KS/U3CATG, KS/U3CATC, KS/U3CAT, and KS/U3TCTT) were sequenced to ensure that only the desired mutations were introduced. The MluI-SacI fragments from KS/U3CATG, KS/U3CATC, KS/U3CAT, and KS/U3TCTT were used to replace the corresponding MluI-SacI fragment in RSV(A)Z, which generated RSV(U3CATG), RSV(U3CATC), RSV(U3CAT), and RSV(U3TCTT), respectively.

To measure virus titer, the MluI-PvuI fragment of RSV(A)Z1-LTRgfp (3) was replaced with the corresponding MluI-PvuI fragments in RSV(U3CATG), RSV(U3CATC), RSV(U3CAT), and RSV(U3TCTT), generating RSV/GFP(U3CATG), RSV/GFP(U3CATC), RSV/GFP(U3CAT), and RSV/GFP(U3TCTT), respectively.

In order to express recombinant Lac repressor, the nucleotide sequence corresponding to nucleotides 366734 to 365652 of the *Escherichia coli* K-12 strain (GenBank accession no. NC_000913) was PCR amplified from plasmid pLS1 (a generous gift from Kathleen S. Matthews, Rice University) with oligonucleotides Lac5 (5'-CGCGGATCCAAACCAGTAACGTTATACGATGTCGC-3') and Lac3 (5'-CCGGAATTCCTGCCCGCTTCCAGTCGGGAAACCTGTC-3') carrying BamHI and EcoRI restriction sites, respectively (both sites are underlined). The PCR product was digested with EcoRI and BamHI and ligated between the corresponding sites of the pTrcHisA expression plasmid (Invitrogen, Carlsbad, CA), in frame with the polyhistidine (six-His) region, creating the N-terminal fusion six-His-Lac (plasmid pTrc-His-Lac).

The C terminus of the Lac repressor was modified by inserting a recombinant tobacco etch virus (rTEV) protease cleavage site, followed by a 23-residue peptide (BP; amino acid sequence, MASSLRQILDSQKMEWRNAGGS) previously identified as the minimal substrate for BirA-catalyzed biotinylation *in vivo*. Proteins carrying this sequence are biotinylated by the endogenous *E. coli* biotin synthetase BirA, which catalyzes the transfer of biotin to the epsilon amino group of a lysine residue in the BP sequence (underlined in the amino acid sequence). The rTEV-BP coding sequence was PCR amplified using a pair of self-annealing oligonucleotides, RTEVB-5 (5'-CCGGAATTCGAAAACCTGTATTTTCAGGGCATGGCTTCTTCCTCAGGCAATCCTGGACTCTC-3') and RTEVB-3 (5'-CAGCCAAGCTTCGTCATCGGGATCCGCCAGCGTTAGATCTCCACTCCATCTTCTGAGAGTCCAGGATTGCCTCAG-3') carrying EcoRI and HindIII restriction sites, respectively (both sites are underlined, the nucleotide sequence corresponding to the BP peptide is shown in bold, and rTEV is italicized). The PCR product was cleaved with EcoRI and HindIII and ligated into the corresponding sites of the pTrc-His-Lac expression plasmid.

Cells, transfection, and infection. DF-1, a continuous line of chicken fibroblasts, was derived from EV-O embryos (12, 29). The cells were maintained in Dulbecco's modified Eagle's medium (Gibco, Carlsbad, CA) supplemented with 5% fetal bovine serum, 5% newborn calf serum, 100 U of penicillin per ml, and 100 µg of streptomycin (Quality Biological, Inc., Gaithersburg, MD) per ml. DF-1 cells were incubated at 39°C with 5% CO₂, and 293-tva cells were incubated

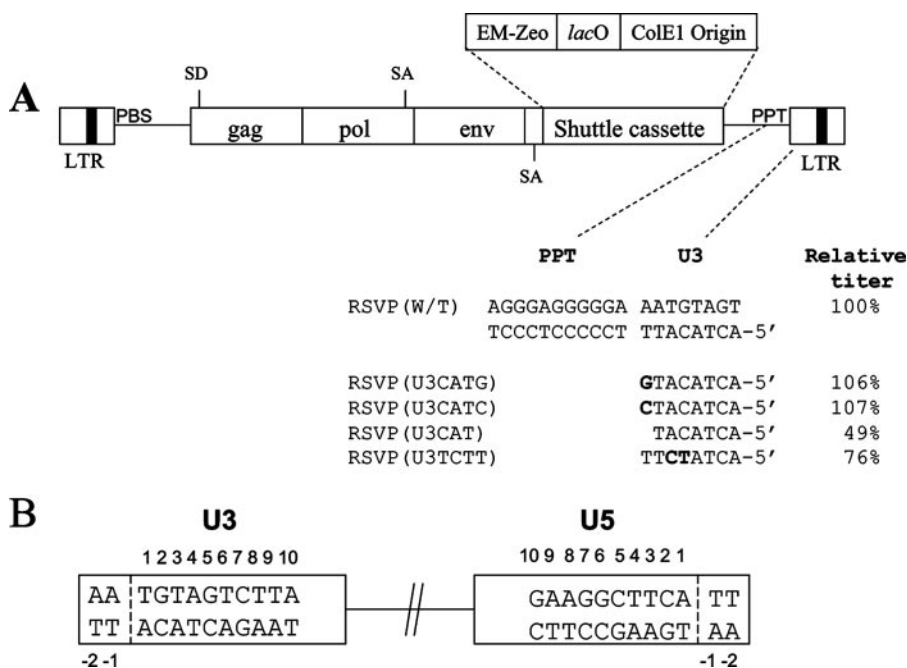


FIG. 1. Structures of the RSVP vectors. (A) Schematic representation (not to scale) of the recovery cassette and the mutants. Mutations are indicated in bold. The virus titers, normalized to the amount of p27 (capsid), are shown relative to that of the wild-type virus. These results are the averages from three independent experiments. EM-Zeo is the bacterial promoter linked to the zeocin resistance gene; lacO is the lac operator. (B) Schematic numbering of the nucleotide positions in the ends of the provirus. The last nucleotide on each end of a normal provirus is number 1.

at 37°C with 5% CO₂. Cells were passaged 1:5 at confluence with trypsin DeLarco (pH 6.8). Ten micrograms of plasmid DNA encoding RSVP(U3CATG), RSVP(U3CATC), RSVP(U3CAT), and RSVP(U3TCTT) was introduced into DF-1 cells by using the calcium phosphate transfection kit (Invitrogen), following the manufacturer's recommendations. DF-1 cells were incubated with medium containing 15% glycerol for 5 min at 39°C 16 h after transfection. The cells were washed twice with phosphate-buffered saline and incubated in fresh medium for 48 h. The 48-h supernatants were harvested and subjected to low-speed centrifugation to remove cellular debris. A portion of the infectious virions was used to infect fresh DF-1 cells. Selection for zeocin resistance was initiated 48 h postinfection with 300 µg/ml of zeocin (Invitrogen).

Measurement of virus titer. Titers of viral stocks generated by transfection with the RSVP(A)Z1-LTRgfp vectors carrying the mutations were determined on 293-tva cells, and the percentage of green fluorescent protein (GFP)-positive cells was determined by flow cytometry 48 h after infection. The 293 cells expressing the tva receptor (293-tva) were kindly provided by John Young. The

titers were normalized based on the amount of p27 antigen present in the viral stocks, as measured by p27 antigen capture enzyme-linked immunosorbent assay, giving a relative titer by normalizing the data to the values for wild-type RSVP/GFP.

Recovery of two-LTR circle junctions. Genomic DNA was isolated from infected DF-1 cells ca. 48 h after infection by using a DNeasy tissue kit (Qiagen). A portion of the recovered DNA was used to transform ElectroMax DH10B or DH5α (Invitrogen) by electroporation. Electroporation was performed as described previously (23, 24).

Lac repressor protein expression. To express the recombinant Lac repressor protein, the pTrc-His-Lac-biotin plasmid was heat shock transformed into the competent *E. coli* Stbl2 strain (Invitrogen). A single colony was used to inoculate

TABLE 1. Recovery of full-length integrated viral DNA

Provirus	U3 ^a	U5 ^a	Size of duplication (bp)	No. of cases
RSVP(U3CATG)	1	1	15	3
	1	1	6	17
	-62	1	6	1
RSVP(U3CATC)	1	1	5	3
	1	1	6	9
	1	26	191	1
RSVP(U3CAT)	1	1	5	2
	1	1	6	9
	Variable	1	Variable	5
RSVP(U3TCTT)	1	1	5	2
	1	1	6	8
	Variable	1	Variable	6

^a The last viral nucleotide at each end of the proviruses is indicated by a number, using the numbering system in Fig. 1B.

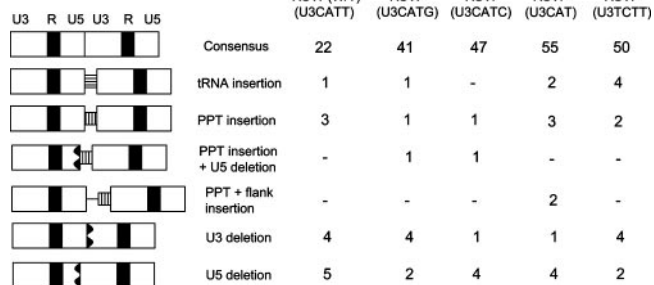


FIG. 2. Two-LTR circle junctions isolated from infected cells. A consensus circle junction is shown at the top. Different types of aberrant two-LTR circle junctions are shown. The tRNA insertion is indicated by a white box with black horizontal bars; the PPT is a white box with black vertical bars; a deletion is indicated by a black jagged end.

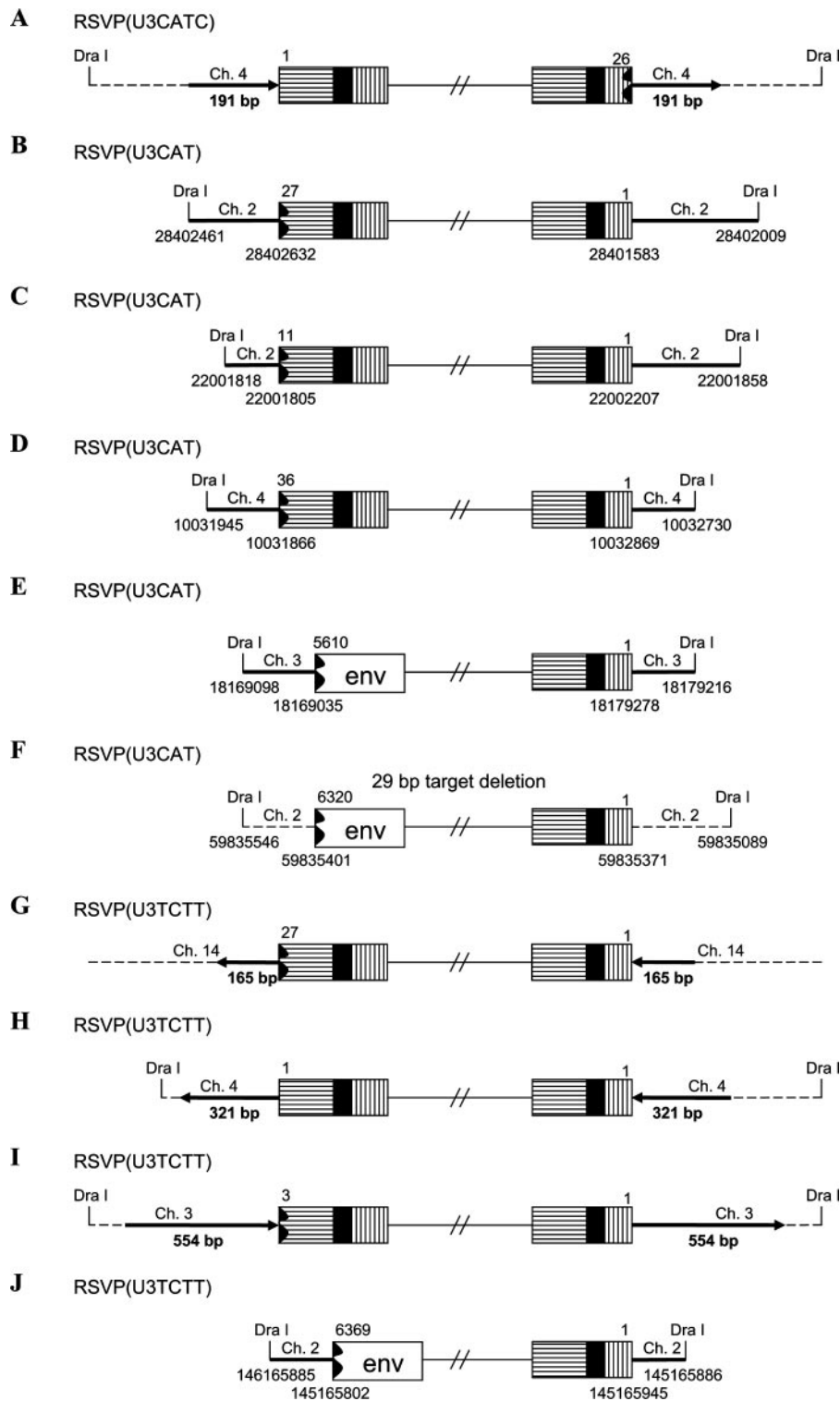


FIG. 3. Structures of the aberrant proviruses. U3 is shown as a box with horizontal bars, R is shown as a black box, and U5 is shown as a box with vertical bars. A deletion in U5 and/or U3 is indicated by a black jagged end. Insertion of PPT and flanking viral sequence is shown by a box with black dots. Host DNA is shown as a line. Large duplications of host sequences are indicated by thick black lines/arrows (not to scale). The direction of the arrow corresponds to the numbering of the host genomic DNA. In some cases in which the nature of the origin of the viral DNA is ambiguous (see the text), the positions of *Dra*I sites and the viral/host DNA junctions are indicated by nucleotide position numbers from a BLAT search.

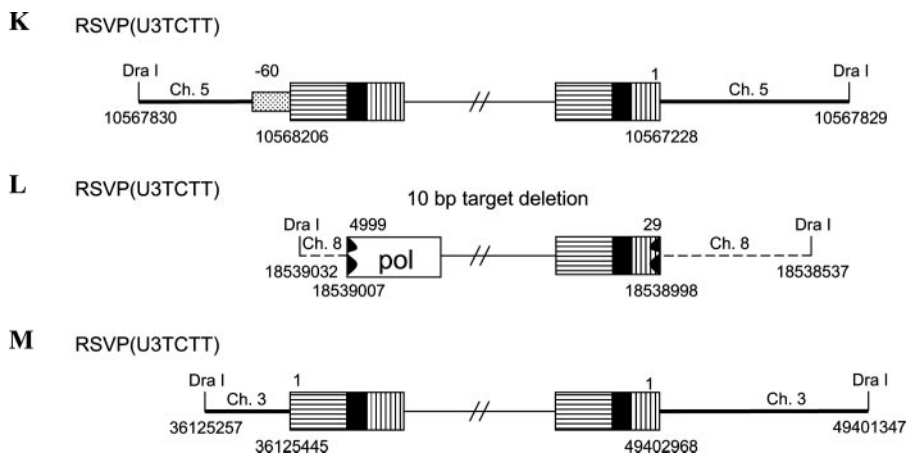


FIG. 3—Continued.

2 ml of LB medium (supplemented with 50 μ g/ml ampicillin for selection of transformed colonies; 2 mg/ml biotin was provided as a substrate for BP biotinylation), and the cells were grown overnight at 37°C with shaking. The next day, 1 liter of LB containing ampicillin and biotin was inoculated with 1 ml of the overnight culture and grown at 37°C with shaking to an optical density at 600 nm of 0.6 (mid-log phase of the cells). Lac expression was induced with 1 mM isopropyl- β -D-thiogalactopyranoside (IPTG) and, after growth of the culture for another 5 h and removal of a 1-ml aliquot for protein analysis, the suspension was centrifuged and the resulting pellet was stored at -20°C for further use.

To confirm the expression of the recombinant Lac repressor protein, the pellet from the 1-ml aliquot was resuspended in 100 μ l of 20 mM phosphate buffer at neutral pH, frozen in liquid nitrogen, and thawed at 42°C (the freeze-thaw was repeated three additional times) followed by pelleting of the insoluble proteins for 10 min in a conventional tabletop microcentrifuge at maximum speed. The supernatant was fractionated on a 4 to 12% sodium dodecyl sulfate-polyacrylamide gel which was stained with Coomassie blue. This procedure confirmed the presence of a 43-kDa band corresponding to the recombinant Lac repressor monomer.

To confirm the presence of introduced C- and N-terminal modifications, the sample was subjected to standard Western blot analysis using two primary antibodies which detected the His tag (anti-His antibody from Sigma) and the biotinylated BP peptide (streptavidin-alkaline phosphatase conjugate), respectively (both antibodies were from Sigma). Both antibodies detected the 43-kDa protein, identifying it as the modified recombinant Lac repressor.

Lac repressor protein purification. To extract the Lac repressor protein, the *E. coli* cell pellet was resuspended in 10 ml of lysis buffer (50 mM NaPO₄, pH 8.0, 50 mM NaCl, 0.75 mg/ml lysozyme, 1 mM phenylmethylsulfonyl fluoride) and incubated on ice for 30 min, after which the concentration of NaCl in the lysate was adjusted to 300 mM. The lysate was sonicated three times for 30 s (60% power plus 70% pulse) and centrifuged at 85,000 \times g for 1 h, and the supernatant was diluted 1:1 with 50 mM NaPO₄, pH 6.8, 300 mM NaCl and passed over a Poly-Prep chromatography column (Bio-Rad) prepacked with 1 ml of BD Talon resin (BD Biosciences) that had been preequilibrated with 50 mM NaPO₄, pH 7.0, 300 mM NaCl, 6 mM imidazole. The column was washed four times with 20 ml of the equilibrating buffer, and the bound protein was eluted with 20 ml of the elution buffer (50 mM NaPO₄, pH 6.0, 300 mM NaCl, 600 mM imidazole, 10% glycerol).

Lac repressor-mediated recovery of integrated retroviral DNA. Genomic DNA was isolated from infected DF-1 (24, 25) cells that survived zeocin selection using a QIAamp DNA blood maxikit (Qiagen). To recover full-length proviruses, 100 to 200 μ g of genomic DNA was digested with DraI overnight at 37°C. One milligram of Dynabeads M-280 streptavidin (DynaL Biotech) was washed three times with a wash buffer (150 mM NaCl, 10 mM EDTA) and incubated with 3 to 6 μ g of purified biotinylated Lac repressor protein for 1 to 2 h at room temperature with gentle rotation of the tube. The Dynabeads-biotinylated Lac repressor mixture was washed three times with a wash buffer. The digested DNA was adjusted to 150 mM NaCl, 10 mM EDTA, 50 μ g of bovine serum albumin per ml, and 10% (vol/vol) glycerol in a final volume of 600 μ l and then incubated with the Dynabeads-biotinylated Lac repressor mixture for 2 h at room temperature with gentle rotation of the tube. The DNA-Dynabeads-biotinylated Lac repressor

mixture was washed three times with 1 ml of wash buffer and eluted twice with 1 ml of elution buffer (10 mM Tris-HCl [pH 7.5], 10 mM EDTA, 10 mM IPTG) for 30 min at 37°C. The enriched DNA was extracted with phenol-chloroform and precipitated with ethanol. The precipitated DNA was ligated with T4 DNA ligase (30 U/200 μ l; Roche, Indianapolis, IN) for 18 h at 16°C. The ligated DNA was then used to transform *E. coli* as described above. After a 3-h recovery period in SOC medium at 37°C, the transformed bacteria were plated onto low-salt Luria-Bertani plates containing 50 μ g of zeocin per ml.

Sequencing of two-LTR circle junctions and integration sites. Recovered plasmids were directly sequenced using the following primers: the PBS primer (5'-ACTATCAGTCGGGGTCACCA) for the two-LTR circle junction and the U3 integration sites and the Ori-2 primer (5'-GCAAGCAGCAGATTACGC GCA) for the U5 integration sites. The Ori-2 primer was also used to confirm that the mutations were present in the recovered plasmids. In those cases in which the provirus was flanked by substantial duplications, the ends of the duplications were sequenced to demonstrate that the duplicated regions were present on both ends of the provirus. Chicken genomic sequences were analyzed by BLAT searches (<http://genome.ucsc.edu/cgi-bin/hgBlat>).

RESULTS AND DISCUSSION

The U3-end mutations affected virus titer. Starting with RSV(A)Z, we changed the sequence that is normally found at the 5' end of U3 adjacent to the PPT (TT) and the CA dinucleotide that is normally found at the U3 end of the provirus. The behavior of the mutant viruses should help us understand how the sequence adjacent to the PPT affects the generation of the linear viral DNA and how IN inserts the linear DNAs produced by the mutants into the host genome. We also changed the CA dinucleotide in the 5' end of U3 to TC to ask how this substitution affects integration. The changes made in the viral mutants are shown in Fig. 1.

To measure virus titer, viruses were generated by transfecting DF-1 cells with plasmids containing the U3-end mutations and the resulting viral stock was used to infect 293-tva cells. GFP expression was measured using fluorescence-activated cell sorting; the titers were normalized relative to the amount of p27 (capsid) in the infecting viral stock. RSV(A)Z(U3CATG) and RSV(A)Z(U3CATC) had relative virus titers that were indistinguishable from those of the wild type. However, RSV(A)Z(U3CAT) decreased virus titer to about 49% of that of the wild type. A similar result was obtained with the corresponding U5 mutant, RSV(A)Z(HIV1), in which the two nucleotides (TT) between the PBS and the CA dinucleotide of RSV(A)Z were

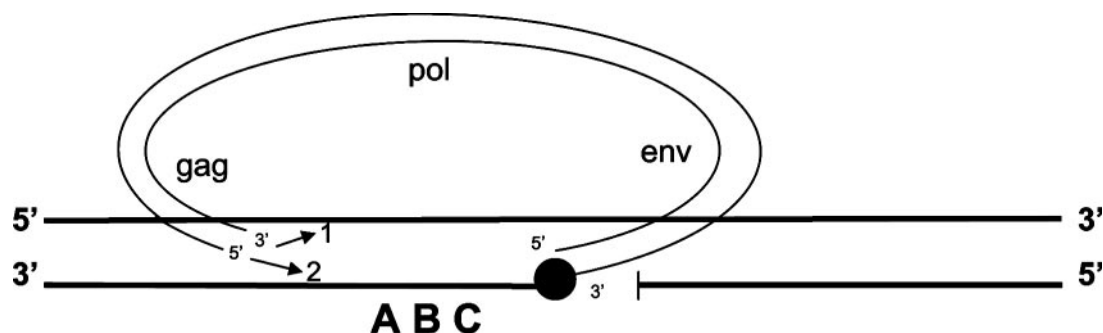


FIG. 4. A model for the generation of a duplication of host DNA during integration or the capture of host DNA into an unintegrated circular viral DNA. The U5 end is inserted normally by IN, which links the 3' end of the viral DNA to the host DNA (black ball on the right side of the figure). On the U3 end (left side of the figure) an aberrant joining event can involve either the 3' end or the 5' end. Arrow 1 shows the joining of the 3' end of the U3 end, which causes a duplication of host sequence (ABC) that lies between the two viral/host junctions, giving rise to an aberrant provirus flanked on both sides by the host sequence ABC. Arrow 2 shows, when the 5' end of the U3 end is joined to the host DNA, that the host sequence ABC is captured between the two ends of a circular viral DNA.

changed to the one nucleotide that matches the HIV-1 sequence (G). The RSVP(U3CAT) mutation should put the canonical CA in the U3 end one nucleotide from the end of the linear DNA, just as the RSVP(HIV1) mutation puts the canonical CA on the U5 end one nucleotide from the other end of the linear DNA. The RSVP(HIV1) mutation decreased virus titer to about 46% of the wild-type level (24). Surprisingly, the RSVP(U3TCTT) mutation had a more modest effect on viral titer, reducing the titer to 76% of the wild-type level, even though the canonical CA dinucleotide in the U3 end was changed to TC. This result differed from our previous result with the corresponding RSVP(CATC) mutant, in which the canonical CA dinucleotide in the U5 end was changed to TC; RSVP(CATC) decreased virus titer to about 36% of the wild-type level (24). Because the titer of the mutants was within twofold of that of the wild type, we did not attempt to measure the synthesis of the viral DNA using real-time PCR. In our hands, the accuracy of real-time PCR is approximately twofold, so we were unable to determine whether any of the mutations had an effect on the efficiency of viral DNA synthesis; if there is an effect, it would be small (twofold or less).

Analysis of two-LTR circle junctions. Analysis of the two-LTR circle junctions from the mutants is shown in Fig. 2. Although the two-LTR circle junctions obtained from the mutants included both insertions and deletions, the proportion of consensus sequence junctions that were derived from the joining of complete, normal copies of both ends of the linear viral DNA was not significantly different between the mutants and the wild type. Although in Table 2 the percentage of two-LTR circles in the wild type is slightly lower than the percentage of two-LTR circles seen with the mutants (62% versus approximately 80%), in previous experiments we found that the percentage of two-LTR circles produced in a wild-type infection was approximately 80% (24). We previously reported that replacing the terminal 3' A of the RSV PPT with a G caused a preferential miscleavage one nucleotide into U3, following the newly created GA dinucleotide sequence. This miscleavage resulted in the deletion of the terminal adenine normally present at the 5' end of the U3, making the resulting linear viral DNA 1 bp shorter on the U3 end (3). Here, we replaced the T at the 3' end of U3 with a C (U3CATC) or with a G

TABLE 2. Microhomology at the host and viral DNA junctions in aberrant proviruses^a

Provirus	U3	U5
A	Acat-5'	5'-tacGAG
B	AGAaca-5'	5'-ttca
C	ACgtta-5'	5'-ttca
D	taccat-5'	5'-ttca
E	aggctc-5'	5'-ttca
F	Cgatgg-5'	5'-ttca
G	AGAcag-5'	5'-ttca
H	CTatca-5'	5'-ttca
I	ATcaga-5'	5'-ttca
J	Taaggg-5'	5'-ttca
K	TGgtag-5'	5'-ttca
L	Cccct-5'	5'-gacTAC
M	ctatca-5'	5'-gcttca

^a Four nucleotides from normal host/virus DNA junctions and six nucleotides from aberrant junctions are shown. Capital letters indicate sequences showing microhomology with the host DNA. The letters A to M correspond to the proviruses described in Fig. 3.

TABLE 3. Recovery of circles derived from abortive integrations

Provirus	U3 ^a	Host DNA		U3-end junction
		No.	Size (bp)	
RSVP(U3CATC)	Env	3	220	GGcgtt-5'
	Env	1	284	ataccg-5'
RSVP(U3CAT)	Env	23	358	AACgga-5'
	Env	7	189	AAGTct-5'
RSVP(U3TCTT)	U3	17	374	acgtta-5'
	Env	19	300	TTTATG-5'
	Pol	1	260	CCATgt-5'
	Env	4	241	Tgacga-5'
	Pol	4	240	aacctt-5'
	Env	9	60	Cccta-5'
	Env	2	76	GTCCgt-5'

^a At the U5 junctions, the last nucleotide was a 1 in all cases, using the numbering system in Fig. 1B. Six nucleotides from U3-end junctions are shown, and capital letters indicate sequences showing microhomology with the host DNA.

(U3CATG), which resulted in a linear DNA in which there was either a 5'-GA or a 5'-CA at the U3 end of the linear viral DNA. Neither of these mutations affected the specificity of the cleavage of the PPT by Rous sarcoma virus (RSV) RNase H. These results suggest that, as expected, the specific cleavage that removes the PPT primer *in vivo* is much more dependent on the sequence of the PPT than it is on the adjacent sequence in U3.

Recovery of full-length integrated viral DNA. Although RSV RNase H properly cleaves the PPT primer in RSVP(U3CAT) and RSVP(U3TCTT), the mutations lead to the generation of linear viral DNAs with aberrant U3 ends, which raises the question of whether the aberrant ends are substrates that can be properly integrated by RSV IN. Full-length integrated viral DNA was recovered as described in Materials and Methods (see also references 24 and 25).

Based on sequence analysis, most of the proviruses derived from infections with RSVP(U3CATG) and RSVP(U3CATC) arose from normal integrations in which two nucleotides were removed from each end of the linear viral DNA and the viral DNA was inserted into the host genome by IN, creating a 5- or 6-bp duplication of the target site sequence flanking the provirus (Table 1). However, in one case, an RSVP(U3CATG) provirus was isolated in which the viral DNA was integrated using an internal CA sequence in U3, giving rise to a provirus flanked by a normal host DNA repeat of 6 bp (Table 1). In another case, a provirus derived from infection with the RSVP(U3CATC) mutant virus had a deletion in U5 but had a normal U3 junction (Table 1; Fig. 3A). Infections with this mutant gave rise to two-LTR circle junctions with U5 deletions; suggesting that some of the linear viral DNAs synthesized by this mutant had aberrant U5 ends. If we compare the aberrant U3 junctions obtained in infections with RSVP(U3CATC) (in which the mutations are in U3) with the aberrant U5 junctions of proviruses that arise from infections with a virus with mutations in the U5 end, the aberrant U3 and U5 junctions are similar (24).

In the case of RSVP(U3CAT), about 70% of the proviruses were integrated normally, indicating that RSV IN can remove a single nucleotide from the U3 LTR terminus of the linear viral DNA during integration (Table 1). However, 30% of the proviruses were aberrant (Fig. 3B to F). In each of the aberrant proviruses, the terminus of the U3 LTR was deleted and none of the aberrant proviruses were flanked by a 5- or 6-bp duplication at the target site. In one case, there was a 29-bp deletion of host sequence at the target site. The proviruses were obtained by digesting host genomic DNA with *Dra*I, which does not cleave in viral DNA. If an aberrant provirus is flanked by a large direct repeat of host sequence and the *Dra*I sites lie outside the repeat, *Dra*I cleavage allows us to recover host DNA that contains the complete direct repeats. However, if there is a *Dra*I site in the repeated portion of the flanking host DNA, portions of both repeats are lost, and the recovered viral DNA is indistinguishable from an unintegrated circular DNA with host sequences between the two LTRs that could arise from an abortive integration event like the one shown in Fig. 4 (described below; see also reference 24). We suggest that most of the ambiguous DNAs that we have recovered are derived from proviruses for two reasons. (i) We selected cells that were zeocin resistant. The fact that we were able to select the cells

suggests that a significant portion of the viral DNA is integrated because zeocin is expressed at a high-enough level to allow the cells to survive the selection. (ii) Many of the ambiguous DNAs have a *Dra*I site in the host sequences. We did recover some circular DNAs in which the host DNA lacks a *Dra*I site. These DNAs almost certainly derive from abortive integrations; these abortive integrations are discussed in a later section.

In the case of an infection with RSVP(U3TCTT), about 60% of the proviruses that we recovered were integrated normally, indicating that RSV IN can successfully remove two nucleotides beyond the mutated TC sequence, which, in this mutant, replaces the canonical CA that is normally found two nucleotides from the U3 LTR terminus of the linear viral DNA. In about 40% of the RSVP(U3TCTT) proviruses, the U3 LTR terminus was deleted and, more rarely, there was an insertion of the PPT and flanking sequences. In the aberrant integrations there were either large duplications or a deletion of the flanking host sequence, instead of the normal 5- or 6-bp duplication. For some of the aberrant proviruses (Fig. 3G to I), we sequenced through the duplicated regions of host DNA at both ends of the proviruses into the nonduplicated sequences. One apparently normal provirus had what appeared to be the mutated TC sequence in the U3 LTR junction linked directly to host DNA. This would suggest that this was an IN-mediated insertion; however, there was a 321-bp duplication of the host sequence at the target site, which suggests that this provirus was not derived from a normal integration event (Fig. 3H). This interpretation was supported by the fact that the last two nucleotides of the U3 LTR terminus were identical with the first two nucleotides of the host DNA. Microhomology at the host/virus DNA junction is not a characteristic of normal IN-mediated integration events; however, microhomologies are often found at aberrant host/virus DNA junctions that are generated by host enzymes (24, 25) (Table 2). The fact that there is microhomology at the host/virus DNA junction also means we cannot define the exact U3 junction for this provirus (Table 2, H). We isolated another apparently normal provirus; however, it was flanked by an inversion of host sequences (Fig. 3M). There was no microhomology between the viral and host sequences at either of the host/virus junctions. We previously proposed a model to explain how an aberrant proviral insertion could generate an inversion of the flanking host sequence (25).

Others previously showed that mutating the CA in either U3 or U5 caused a moderate reduction in the titer of HIV-1 (1, 22). However, at least for RSV, a linear viral DNA with one defective end can be integrated into the host genome with good efficiency, even though the aberrant end is not inserted by IN. For this reason, titer is not sufficient to determine whether IN can properly insert an aberrant end. In RSV, mutating the CA in the viral DNA that is joined to host sequences has different consequences in U3 and U5. We previously reported that changing the CA dinucleotide in the U5 end to TC, the mutation in RSVP(CATC), prevented the normal IN-mediated insertion of the mutated U5 end in all the proviruses that we isolated (24). However, in the experiments that we present here, making the corresponding mutation in U3, changing the CA sequence in the U3 end to TC, allowed IN-mediated insertion of the mutant end in 60% of the RSVP(U3TCTT) proviruses; these proviruses were flanked by the normal 5- or

TABLE 4. Microhomology at the virus-virus DNA junctions^a

Virus	U3	U5	Microhomology	Virus	U3	U5	Microhomology	
RSVP(W/T)	Gag	-20	5'-cgtgat	Gag	-9	5'-ggTGAC		
	Env	-2	5'-ttcaTT	Gag	-7	5'-ttggtG		
	Gag	-2	5'-ttcatt	Gag	-5	5'-atttGG		
	Pol	-2	5'-ttcaTT	Gag	1	5'-gcttcA		
	Env	4	5'-aagGCT	Env	5	5'-gaaggC		
	Gag	6	5'-agaaGG	U3	6	5'-agaagg		
	Env	7	5'-cagAAG	Pol	7	5'-aagaaG		
	Env	7	5'-cagaaG	Env	8	5'-gcagAA		
	U3	14	5'-catgAA	U3	14	5'-catgAA		
	Pol	16	5'-tgcacG	Env	17	5'-ctgcAT		
	U3	16	5'-tgcATG	Env	18	5'-cctGCA		
	U3	16	5'-tgcacG	U3	19	5'-acctgc		
	PBS	20	5'-caccTG	Env	19	5'-acctgC		
	U3	26	5'-tacgaG	Env	20	5'-cacctG		
	Gag	31	5'-acgacT	U3	22	5'-agcacC		
	Pol	37	5'-tcccTG	Env	25	5'-acgAGC		
	Gag	70	5'-gtgtgc	Gag	35	5'-cctgAC		
	RSVP(U3CATC)	Env	-20	5'-ccTGAT	Env	36	5'-ccCTGA	
Env		-8	5'-tggTGA	U3	37	5'-tcccTG		
Env		-5	5'-atttgG	Gag	58	5'-gttgAT		
Gag		1	5'-gcttca	Gag	69	5'-tgtgCA		
U3		3	5'-aggctt	Env	130	5'-atattG		
Pol		5	5'-gaaGGC	D64K	Pol	-18	5'-gacgtG	
Env		18	5'-cctGCA		Gag	-2	5'-ttcatt	
Gag		24	5'-cgagCA		Gag	1	5'-gcttCA	
Env		27	5'-ctacgA		Env	2	5'-ggcttC	
Env		29	5'-gacTAC		Gag	4	5'-aaggct	
Pol		38	5'-ttccCT		Gag	5	5'-gaaggc	
Env		134	5'-agaGAT		Gag	5	5'-gaaGGC	
Env		176	5'-gggtcT		Pol	6	5'-aGAAGG	
Gag		205	5'-gateGT		Pol	7	5'-cagaAG	
RSVP(U3TCTT)		Pol	-15		5'-cccgaC	Pol	10	5'-aacgAG
	Gag	-13	5'-accCCG		Gag	15	5'-gcatGA	
	Pol-Env	-12	5'-gaccCC		Env	16	5'-tGCATG	
	Env	-8	5'-tggTGA		Gag	18	5'-cctgCA	
	Env	-2	5'-ttcatt		Env	20	5'-cacCTG	
	Env	-2	5'-ttcatT		Env	22	5'-agcacc	
	Env	1	5'-gcttcA	Pol	23	5'-gagcaC		
	Pol	1	5'-gcttCA	Gag	28	5'-actacG		
	Gag	1	5'-gcttCA	Env	37	5'-tccctG		
	Gag	5	5'-gaaggc	Gag	58	5'-gttgaT		
	U3	5	5'-gaaggc	RSVP(U3CAT)	Env	-16	5'-ccgacG	
	Env	5	5'-gaagGC		Env	-9	5'-ggtgAC	
	Pol	6	5'-agaaGG		Gag	-2	5'-ttcatT	
	Gag	6	5'-agaAGG		Env	-2	5'-ttcatt	
	Env	8	5'-gcaGAA		Env	3	5'-aggCTT	
	Gag	9	5'-agcaGA		Gag	6	5'-agaagG	
	Gag	9	5'-agcagA		Gag	7	5'-cagaAG	
	Pol	9	5'-agcaGA		Env	9	5'-agcAGA	
	Env	13	5'-atgaAG		Gag	13	5'-atgaAG	
	U3	17	5'-ctgcat		Env	19	5'-acctGC	
	Env	20	5'-cacctg		U3	20	5'-cacctg	
	Gag	20	5'-cacctg		U3	20	5'-caccTG	
	Gag	21	5'-gcacCT		U3	21	5'-gcaccT	
	Pol-Env	21	5'-gcaccT	Env	23	5'-gagcaC		
	Gag	26	5'-tacgAG	Pol	25	5'-acgaGC		
	Pol	30	5'-cgaCTA	Gag	25	5'-acGAGC		
	Gag	33	5'-tgacgA	U3	27	5'-ctacga		
	Env	47	5'-gaccgT	Gag	27	5'-ctaCGA		
Gag	54	5'-aTGGCC	Gag	28	5'-actacG			
Gag	73	5'-ttggtG	Gag	28	5'-actaCG			
Pol	86	5'-accatt	Gag	30	5'-cgacta			
RSVP(U3CATG)	Env	-30	5'-agggAA	Pol	69	5'-tgtgca		
	Gag	-13	5'-accccG	Env	299	5'-tagtCT		
	Env	-9	5'-ggtgAC	Gag	423	5'-gggacT		

^a Six nucleotides from U5-end junctions are shown, and capital letters indicate sequences showing microhomology between the viral DNAs. The name of the viral mutant from which the circular DNAs were derived is given in the left column for each group. D64K has a mutation in the active site of IN. W/T, wild type. For Pol-Env, the position of the junction is between the *pol* and *env* genes.

6-bp duplication of host sequence. Not surprisingly the titer of the U3 mutant RSV(U3TCTT) was higher than the titer of the U5 mutant RSV(CATC), 76% and 36% of the wild-type level, respectively. These results suggest that, in the context of the U3 end, TC, instead of the canonical CA, can be used reasonably efficiently by RSV IN, which means that the CA dinucleotide is not required at the U3 end for appropriate integration *in vivo*.

Recovery of circular viral DNAs derived from abortive integrations. When we recovered full-length proviruses, we also recovered 11 circular viral DNAs that contained host DNA sequences ranging in size from tens to hundreds of nucleotides, in which there was no DraI site (Table 3). All of these circles contained a normal U5 junction; however, portions of the U3 end of the viral DNA were deleted and the new U3 end was joined to the host DNA in a way that caused the host sequence to be captured between the ends of a circular viral DNA (Fig. 3). There were, in 8 out of 11 cases, microhomologies involving one to five nucleotides between the ends of viral sequences and host sequences at the aberrant junction (Table 3); as has already been discussed, this is a characteristic of aberrant host/virus junctions that are generated by host enzymes (24).

Recovery of smaller circular viral DNAs. When we recovered the two-LTR circle junctions, we also recovered 130 circular viral DNAs in which a portion of the viral DNA was deleted from the U3 end (Table 4). In most of these smaller circles, the U5 end of the viral DNAs was abnormal: either the U5 end was deleted or a portion of the PBS/flanking sequence was retained. The U5 end of the viral DNA was then joined to various positions of the viral DNA, creating smaller circular viral DNAs. In six cases, a canonical CA sequence was present at the U5 end of the viral DNA, which suggests that these circular forms could have come from IN-mediated autointegration events. However, in five out of these six cases, we cannot define the exact U5 junction, because there are microhomologies involving one or two nucleotides between the two ends of the viral DNAs.

Olsen et al. (26) previously reported the structure of circular viral DNAs produced in RSV infections. Approximately 30% of the circular viral DNAs had aberrant U5 ends, and there were deletions in the 5' portion of the viral genome, similar to the circular viral DNAs reported here. Although they pointed out the common nucleotides present at each boundary of the deletion, the statistical significance of the microhomologies in their circular DNA isolates was not determined, due to limited numbers of circular viral DNAs recovered. Olsen et al. discuss the possibility that IN has a role in generating the small circles that arise during RSV infections without reaching a firm conclusion; however, their data are consistent with our interpretation that IN-mediated autointegration occurs rarely, if at all, in RSV infections. We have recently done experiments using an HIV-1-based shuttle vector that is similar to the RSV-based vectors used in the experiments reported here and recovered numerous circular viral DNAs that were obviously derived from one-ended and two-ended IN-mediated autointegrations, which shows that the technology would have allowed us to recover IN-mediated autointegrated viral DNAs if they had been present (unpublished observations).

In the present study, when the sequences used to form the viral/viral DNA junctions were compared, there were micro-

homologies involving from one to five nucleotides between the ends (Table 4). Although the junctions with one homologous nucleotide were found at a frequency that was not higher than what could have occurred by chance, the frequencies of junctions in which there were two or more homologous nucleotides were high enough to be statistically significant. The probabilities (*P*) of obtaining the number of two, three, four, and five homologous nucleotides by chance were <0.0003, <0.0001, 0.0018, and 0.0022, respectively. In addition, we did not recover any circles arising from a concerted autointegration event in which both ends are processed correctly and inserted into the viral DNA in an orientation opposite from their own polarity, which creates a circular viral DNA that contains two unlinked LTRs with a 5- or 6-bp duplication at the target site. This type of autointegrated circular DNA is commonly found in cells infected by MLV and HIV-1. To exclude the involvement of IN in generating smaller circles, we recovered a set of small circular DNAs from cells infected with a virus mutated in the IN active site (D64K). When the viral/viral DNA junctions were examined, the smaller viral DNAs recovered in the absence of IN enzymatic activity were similar in their structure to those that we recovered from an infection with a virus that has a normal wild-type IN (Table 4), suggesting that the generation of smaller circles is independent of integrase activity. Taken together, these results suggest that RSV linear DNA, in contrast to MLV and HIV-1 linear DNA, does little, if any, autointegration. This should provide RSV with an important advantage. From the point of view of the virus, autointegration is a dead end that wastes a linear viral DNA that might otherwise be used to establish a provirus. Despite the fact that linear RSV DNA does not appear to autointegrate *in vivo*, it autointegrates efficiently when the preintegration complexes (PICs) are isolated from infected cells (19, 20). The PICs of HIV and MLV appear to interact with a host factor, BAF (barrier to autointegration factor), whose presence reduces, but does not eliminate, autointegration. The striking contrast between the *in vivo* and *in vitro* results with RSV autointegration suggests the possibility that the RSV PIC interacts with some as-yet-unidentified host factor that efficiently blocks autointegration of linear RSV DNA. Presumably, the procedures used by Lee and Coffin to isolate RSV PICs from infected cells removed this protective factor (19, 20).

ACKNOWLEDGMENTS

We are grateful to Terri Burdette for help in preparing the manuscript.

This research was supported by the Intramural Research Program of the NIH, National Cancer Institute, Center for Cancer Research.

REFERENCES

1. Brown, H. E. V., H. Chen, and A. Engelman. 1999. Structure-based mutagenesis of the human immunodeficiency virus type 1 DNA attachment site: effects on integration and cDNA synthesis. *J. Virol.* **73**:9011–9020.
2. Cannon, P. M., W. Wilson, E. Byles, S. M. Kingsman, and A. J. Kingsman. 1994. Human immunodeficiency virus type 1 integrase: effect on viral replication of mutations at highly conserved residues. *J. Virol.* **68**:4768–4775.
3. Chang, K. W., J. G. Julias, W. G. Alvord, J. Oh, and S. H. Hughes. 2005. Alternate polypurine tracts (PPTs) affect the Rous sarcoma virus RNase H cleavage specificity and reveal a preferential cleavage following a GA dinucleotide sequence at the PPT-U3 junction. *J. Virol.* **79**:13694–13704.
4. Colicelli, J., and S. P. Goff. 1985. Mutants and pseudorevertants of Moloney murine leukemia virus with alterations at the integration site. *Cell* **42**:573–580.
5. Colicelli, J., and S. P. Goff. 1986. Isolation of a recombinant murine leukemia virus utilizing a new primer tRNA. *J. Virol.* **57**:37–45.

6. Craigie, R., T. Fujiwara, and F. Bushman. 1990. The IN protein of Moloney murine leukemia virus processes the viral DNA ends and accomplishes their integration *in vitro*. *Cell* **62**:829–837.
7. 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. USA* **77**:3937–3941.
8. 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.
9. Englund, G., T. S. Theodore, E. O. Freed, A. Engelman, and M. A. Martin. 1995. Integration is required for productive infection of monocyte-derived macrophages by human immunodeficiency virus type 1. *J. Virol.* **69**:3216–3219.
10. Esposito, D., and R. Craigie. 1998. Sequence specificity of viral end DNA binding by HIV-1 integrase reveals critical regions for protein-DNA interaction. *EMBO J.* **17**:5832–5843.
11. Fujiwara, T., and R. Craigie. 1989. Integration of mini-retroviral DNA: a cell-free reaction for biochemical analysis of retroviral integration. *Proc. Natl. Acad. Sci. USA* **86**:3065–3069.
12. Himly, M., D. N. Foster, I. Botoli, J. S. Iacovoni, and P. K. Vogt. 1998. The DF-1 chicken fibroblast cell line: transformation induced by diverse oncogenes and cell death resulting from infection by avian leucosis viruses. *Virology* **248**:295–304.
13. Hughes, S. H., A. Mutschler, J. M. Bishop, and H. E. Varmus. 1981. A Rous sarcoma virus provirus is flanked by short direct repeats of a cellular DNA sequence present in only one copy prior to integration. *Proc. Natl. Acad. Sci. USA* **78**:4299–4303.
14. 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.
15. Julias, J. G., M. J. McWilliams, S. G. Sarafianos, W. G. Alvord, E. Arnold, and S. H. Hughes. 2004. Effects of mutations in the G tract of the human immunodeficiency virus type 1 polypurine tract on virus replication and RNase H cleavage. *J. Virol.* **78**:13315–13324.
16. LaFemina, R. L., P. L. Callahan, and M. G. Cordingley. 1991. Substrate specificity of recombinant human immunodeficiency virus integrase protein. *J. Virol.* **65**:5624–5630.
17. LaFemina, R. L., C. L. Schneider, H. L. Robbins, P. L. Callahan, K. LeGrow, E. Roth, W. A. Schleif, and E. A. Emimi. 1992. Requirement of active human immunodeficiency virus type 1 integrase enzyme for productive infection of human T-lymphoid cells. *J. Virol.* **66**:7414–7419.
18. Leavitt, A. D., R. B. Ross, and H. E. Varmus. 1992. Both substrate and target oligonucleotide sequences affect *in vitro* integration mediated by human immunodeficiency virus type 1 integrase protein produced in *Saccharomyces cerevisiae*. *J. Virol.* **66**:2359–2368.
19. Lee, Y. M., and J. M. Coffin. 1990. Efficient autointegration of avian retrovirus DNA *in vitro*. *J. Virol.* **64**:5958–5965.
20. Lee, Y. M., and J. M. Coffin. 1991. Relationship of avian retrovirus DNA synthesis to integration *in vitro*. *Mol. Cell. Biol.* **11**:1419–1430.
21. Masuda, T., V. Planelles, P. Krogstad, and I. S. Y. Chen. 1995. Genetic analysis of human immunodeficiency virus type 1 integrase and the U3 *att* site: unusual phenotype of mutants in the zinc finger-like domain. *J. Virol.* **69**:6687–6696.
22. Masuda, T., M. J. Kuroda, and S. Harada. 1998. Specific and independent recognition of U3 and U5 *att* sites by immunodeficiency virus type 1 integrase *in vivo*. *J. Virol.* **72**:8396–8402.
23. Oh, J., J. G. Julias, A. L. Ferris, and S. H. Hughes. 2002. Construction and characterization of a replication-competent retroviral shuttle vector plasmid. *J. Virol.* **76**:1762–1768.
24. Oh, J., K. W. Chang, and S. H. Hughes. 2006. Mutations in the U5 sequences adjacent to the primer binding site do not affect tRNA cleavage by Rous sarcoma virus RNase H but do cause aberrant integrations *in vivo*. *J. Virol.* **80**:451–459.
25. Oh, J., K. W. Chang, W. G. Alvord, and S. H. Hughes. 2006. Alternate polypurine tracts affect Rous sarcoma virus integration *in vivo*. *J. Virol.* **80**:10281–10284.
26. Olsen, J. C., C. Bova-Hill, D. P. Grandgenett, T. P. Quinn, J. P. Manfredi, and R. Swanstrom. 1990. Rearrangements in unintegrated retroviral DNA are complex and are the results of multiple genetic determinants. *J. Virol.* **64**:5475–5484.
27. Panganiban, A. T., and H. M. Temin. 1983. The terminal nucleotides of retrovirus DNA are required for integration but not virus production. *Nature* **306**:155–160.
28. Roth, M. C., P. L. Schwartzberg, and S. P. Goff. 1989. Structure of the termini of DNA intermediates in the integration of retroviral DNA: dependence on IN function and terminal DNA sequence. *Cell* **58**:47–54.
29. Schaefer-Klein, J., I. Givol, E. V. Barsov, J. M. Whitcomb, M. VanBrocklin, D. N. Foster, M. J. Federspiel, and S. H. Hughes. 1998. The EV-O derived cell line DF-1 supports the efficient replication of avian leucosis-sarcoma viruses and vectors. *Virology* **248**:305–311.
30. Schwartzberg, P., J. Colicelli, 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.
31. Telenitsky, A., and S. P. Goff. 1997. Reverse transcriptase and the generation of retroviral DNA, p. 121–160. *In* J. M. Coffin, S. H. Hughes, and H. E. Varmus (ed.), *Retroviruses*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
32. Vink, C., D. C. van Gent, Y. Elgersma, and R. H. Plasterk. 1991. Human immunodeficiency virus integrase protein requires a subterminal position of its viral DNA recognition sequence for efficient cleavage. *J. Virol.* **65**:4636–4644.
33. Zhou, H., J. Rainey, S. K. Wong, and J. M. Coffin. 2001. Substrate sequence selection by retroviral integrase. *J. Virol.* **75**:1359–1370.