Modeling the Late Steps in HIV-1 Retroviral Integrase-catalyzed DNA Integration

Received for publication, August 1, 2000
Published, JBC Papers in Press, September 26, 2000, DOI 10.1074/jbc.M006959200

Elena Brin§, Jizu Yî§, Anna Marie Skalka§, and Jonathan Leis¶†

From the §Department of Microbiology and Immunology, Northwestern University School of Medicine, Chicago, Illinois 60611 and the ¶Fox Chase Cancer Center, Philadelphia, Pennsylvania 19111

Model oligodeoxyribonucleotide substrates representing viral DNA integration intermediates with a gap and a two-nucleotide 5’ overhang were used to examine late steps in human immunodeficiency virus, type 1 (HIV-1) retroviral integrase (IN)-catalyzed DNA integration in vitro. HIV-1 or avian myeloblastosis virus reverse transcriptase (RT) were capable of quantitatively filling in the gap to create a nicked substrate but did not remove the 5’ overhang. HIV-1 IN also failed to remove the 5’ overhang with the gapped substrate. However, with a nicked substrate formed by RT, HIV-1 IN removed the overhang and covalently closed the nick in a disintegration-like reaction. The efficiency of this closure reaction was very low. Such closure was not stimulated by the addition of HMG-(I/Y), suggesting that this protein only acts during the early processing and joining reactions. Addition of Flap endonuclease-1, a nuclease known to remove 5’ overhangs, abolished the closure reaction catalyzed by IN. A series of base pair inversions, introduced into the HIV-1 U5 long terminal repeat sequence adjacent to and/or including the conserved CA dinucleotide, produced no or only a small decrease in the HIV-1 IN-dependent strand closure reaction. These same mutations caused a significant decrease in the efficiency of concerted DNA integration by a modified donor DNA in vitro, suggesting that recognition of the ends of the long terminal repeat sequence is required only in the early steps of DNA integration. Finally, a combination of HIV-1 RT, Flap endonuclease-1, and DNA ligase is capable of quantitatively forming covalently closed DNA with these model substrates. These results support the hypothesis that cellular enzyme(s) may catalyze the late steps of retroviral DNA integration.

Insertion of viral DNA into the host genome is an obligatory step in the retroviral life cycle and requires a concerted mechanism that brings the two LTR1 ends of a single donor into a complex with IN and the host DNA. HIV IN is associated with DNA binding, specific DNA endonuclease activity, DNA joining, and disintegration (see Refs. 1 and 2). The mechanism for insertion of a donor DNA into an acceptor comprises IN processing the 3’ ends of viral LTRs, which are subsequently involved in a nucleophilic attack of phosphodiester bonds in the host DNA to covalently join the two DNAs. These reactions are catalyzed by IN alone (3, 4) but are stimulated by cellular HMG-(I/Y) (5, 6) and the viral NC protein (7). During the strand transfer reaction, the host DNA is cut in a staggered manner to create small gaps between the 5’ ends of the viral DNA and 3’ ends of host DNA. The repair of these gaps at both ends produces the direct repeats that flank the inserted viral DNA. In the case of HIV-1, the size of these duplications is 5 base pairs (8, 9). During the initial processing only the 3’ ends of the LTRs are shortened by two nucleotides. The complementary nucleotides on the opposite LTR strand form a 5’ two-nucleotide overhang after the strand transfer reaction. These overhangs have to be removed after the gaps are repaired so that the covalent joining of the cellular and viral DNA can be completed. The precise enzymes responsible for these reactions are not known, although it was reported that if presented with a nicked substrate, wild type IN or its central domain (10) is capable of removing the 5’ overhang and sealing the nick in a disintegration-like reaction (2, 10, 11).

Disintegration is a polynucleotidyl transfer reaction that is relatively independent of viral DNA sequence; DNA structure rather than sequence (12, 13) is the major determinant of the site of nucleophilic attack. Although the processing and joining steps require a full-length IN protein, disintegration can be performed by the isolated catalytic domain (14). This suggests that the catalytic domain includes the active site for polynucleotidyl transfer and is consistent with the crystal structure determined for this domain, which revealed a five-strand β-sheet and six α-helices similar to other polynucleotidyl transfer enzymes (15).

In this report we demonstrate with model substrates that the combination of RT and IN is sufficient to complete all of the late steps required for integration: filling in the gap, removing the 5’ overhang, and joining the second host DNA strand to the viral DNA. However, this reaction takes place with very low efficiency. We have also examined the ability of cellular proteins such as PCNA, HMG-(I/Y), or Flap endonuclease-1 (FEN-1) (16) to work in concert with RT and/or IN to increase the efficiency of the reaction. None of these enzymes stimulated IN-dependent closure of a nicked substrate with an overhang, whereas the addition of FEN-1 prevented this reaction. As expected from their known specificities, FEN-1 in combination with RT and DNA ligase quantitatively produced strand closure of gapped substrates with 5’ two-nucleotide overhangs.
**Experimental Procedures**

Reagents—[α-32P]dATP (3,000 Ci/mmol), [α-32P]d(TCTP3,000Ci/mmol), and [γ-32P]ATP (2,500 Ci/mmol) were purchased from Amer sham Pharmacia Biotech. Phosphodiesterase I type V (Bothrops atrox) was obtained from Sigma. Proteinase K (30 units/mg), glycogen, and T4 polynucleotide kinase (10 units/μl) were from Roche Molecular Biochemicals. Terminal deoxynucleotidyl transferase was from Amer sham Pharmacia-Biochrom. AMV RT was purchased by U. Biochemical Corp. HMG-(I/Y) was purified as described by Nissen and Reeves (17).

**Preparation of Model Substrates**—Model HIV-1 gapped substrates were prepared by annealing strands A, B, and C. Nicked HIV-1 or ASV substrates were formed by substituting the respective strand C for D. For the HIV-1 mutant LTR substrates 4, 5, and 6 (Fig. 1A), the respective mutated strands A and B were substituted for the wild type HIV-1 strands A and B. Oligos were labeled either at the 3' end (strands B or C) or at the 3' end (strand B). The 5' end was labeled with [γ-32P]ATP and T4 polynucleotide kinase (10 units/50 pmol of oligo). The 3' end was labeled with [α-32P]dATP and TdT as described by the manufacturer. Labeled oligos were purified by reverse phase HPLC (Type US, 0.025 μm) as described (20). 10 pmol of substrate, and HIV-1 RT and 10 μM deoxyribonucleotides were added. The reaction was terminated by the addition of 16 μl of 95% formamide, 10 mM EDTA, 1 μg/ml bromophenol blue, and 0.1 μg/ml xylene cyanole. The tubes were heated to 95 °C for 5 min and loaded onto a 20% denaturing polyacrylamide gel. The reaction products were visualized by autoradiography.

**Gap Repair**—The coupled gap repair by RT and overhang removal by FEN-1 was examined in a 15-μl reaction containing 50 pm HIV-1 DNA, pH 8.0, 10 mM MgCl2, 0.5 mM βME, 145 μg/ml BSA, and 1 pmol of substrate. Where indicated, 6 ng of FEN-1, 500 ng of PCNA, 5 pmol of IN, or 110 ng of HIV-1 RT and 10 μM deoxyribonucleotides were added. The reaction was terminated by the addition of 16 μl of 95% formamide, 10 mM EDTA, 1 μg/ml bromophenol blue, and 0.1 mg/ml xylene cyanole. The products were analyzed by electrophoresis using denaturing 20% polyacrylamide gels. Bands were detected by autoradiography.

**cDNA-Ligation Reactions**—cDNA-ligation activity of IN was examined in a 1-μl reaction containing 50 pm DNA, pH 8.0, 10 mM MgCl2, 0.5 mM βME, 145 μg/ml BSA, and 1 pmol of substrate. Where indicated, 5 ng of FEN-1, 400 ng of PCNA, 5 pmol of IN, or 110 ng of HIV-1 RT, and 10 μM deoxyribonucleotides were added. The reaction was terminated by the addition of 16 μl of 95% formamide, 10 mM EDTA, 1 μg/ml bromophenol blue, and 0.1 mg/ml xylene cyanole. The products were analyzed by electrophoresis using denaturing 20% polyacrylamide gels. Bands were detected by autoradiography.

**RT and IN Repair Assays**—The repair of model substrates by viral enzymes was examined in a 15-μl reaction containing 50 pm HIV-1 DNA, pH 8.0, 10 mM MgCl2, 0.5 mM βME, 145 μg/ml BSA, and 10 μM deoxyribonucleotides, and 1 pmol of substrate. Where indicated, 5 ng of FEN-1, 400 ng of PCNA, 5 pmol of IN, or 110 ng of HIV-1 RT and 10 μM deoxyribonucleotides were added. The reaction was terminated by the addition of 16 μl of 95% formamide, 10 mM EDTA, 1 mg/ml bromophenol blue, and 1 mg/ml xylene cyanole. The products were analyzed by electrophoresis using denaturing 20% polyacrylamide gels. Bands were detected by autoradiography.

**Flap Endonuclease Assay—**Overhang cleavage activity of FEN-1 was examined in a 15-μl reaction containing 50 pm HIV-1 DNA, pH 8.0, 10 mM MgCl2, 0.5 mM βME, 145 μg/ml BSA, and 1 pmol of substrate. Where indicated, 6 ng of FEN-1, 500 ng of PCNA, 5 pmol of IN, or 110 ng of HIV-1 RT and 10 μM deoxyribonucleotides were added. The reaction was terminated by the addition of 16 μl of 95% formamide, 10 mM EDTA, 0.1 mg/ml bromophenol blue, and 0.1 mg/ml xylene cyanole. The products were analyzed by electrophoresis using denaturing 20% polyacrylamide gels. Bands were detected by autoradiography.

**Plasmid Constructions and Preparations—**Plasmid pHIVII2, which was used in this study as a template to amplify donor DNA, is a variation of pBCKS* in which a wild type HIV-1 donor DNA polymerase chain reaction product was inserted into pBCKS* catalyzed by IN, resulting in the loss of 2 base pairs from the LTR ends. This plasmid was propagated in E. coli MC1061/P3 under the conditions described above. The integration acceptor was plasmid pBCSK* (Stratagene, La Jolla, CA), which was propagated in E. coli DH5α. Plasmids were purified with Qiaprep columns (Qiagen, Chatsworth, CA) according to the manufacturer's instructions. Growth of DH5α containing pBCKS* was selected for by addition of chloramphenicol (35 μg/ml).

**Preparation of Donor DNAs—**Integration donor DNA was purified by using thermostable Vent DNA polymerase and the primers listed above. 25 pmol of each primer and 50 ng of pHIVII2 DNA as the template were used for each polymerase chain reaction. Vent DNA polymerase was used according to manufacturer's instructions. A total of 20 rounds of amplification were performed in each reaction. The amplification conditions were 94 °C for 2 min, 50 °C for 1 min, and 72 °C for 1 min for three rounds. This was followed by amplification conditions that used 94 °C for 2 min, 57 °C for 1 min, and 72 °C for 45 s for 17 additional rounds. The resultant product donor DNA was isolated after electrophoresis through 2% agarose gels equilibrated with 0.5× Tris borate-EDTA (6). The purified DNA (600 ng) was recovered using Qiaex II resin (Qiagen) and then precipitated with ethanol. The recovered DNA was suspended in either TE buffer (10 mM Tris-HCl, pH 8.0, 1 mM EDTA) or denoized distilled water. The isolation donors, which were approximately 300 base pairs in length, were internally labeled during the polymerase chain reaction by the inclusion of [α-32P]dTTP. The concerted integration of deoxyribonucleotides or two deoxyribonucleotides uniting amplification reactions were performed after the labeled oligo. The strands A and B were substituted for the wild type HIV-1 strands A and B. Oligos were labeled either at the 3' end (strands B or C) or at the 3' end (strand B). The 5' end was labeled with [γ-32P]ATP and T4 polynucleotide kinase (10 units/50 pmol of oligo). The 3' end was labeled with [α-32P]dATP and TdT as described by the manufacturer. Labeled oligos were purified by reverse phase HPLC (Type US, 0.025 μm) as described (20). 10 μl of the reaction was recovered by incubation for 30 min at 37 °C followed by a 30-min drop dialysis against 10 mM Tris-HCl, pH 8.0, 10 mM MgCl2, 0.5 mM βME, 145 μg/ml BSA, and 1 pmol of acceptor DNA. Where indicated, 5 ng of FEN-1, 400 ng of PCNA, 5 pmol of IN, or 110 ng of HIV-1 RT and 10 μM deoxyribonucleotides were added. The reaction was terminated by the addition of 16 μl of 95% formamide, 10 mM EDTA, 0.1 mg/ml bromophenol blue, and 0.1 mg/ml xylene cyanole. The products were analyzed by electrophoresis using denaturing 20% polyacrylamide gels. Bands were detected by autoradiography.
FIG. 1. Model substrates representing integration intermediates. A, model oligodeoxyribonucleotide substrates representing an integration intermediate of an HIV-1 or ASV U5 LTR cell DNA junction. The HIV-1 or ASV LTR and the cell DNA sequences are as indicated, and the conserved CA dinucleotide is underlined. Substrate 1, HIV-1 junction with a 5-base gap and a 2-base 5' overhang. Substrate 2, HIV-1 junction with a nick and a 2-base 5' overhang. Substrate 3, ASV junction with a 6-base gap and a 2-base 5' overhang. Substrate 4, mutant HIV-1 junction, with base pair inversions at positions 5 and 6 (bold type), with a nick and a 2-base 5' overhang. The conserved CA dinucleotide is changed by the mutation to a GA dinucleotide. Substrate 6, mutant HIV-1 junction, with base pair inversions at positions 3 and 4 (bold type), with a nick and a 2-base 5' overhang. The conserved CA dinucleotide is changed by the mutation to a GG dinucleotide. B, stick figure representations of the gap and nicked substrates. The bottom strand A is a 36-mer. The size of the two upper strands (strands B–D) in the nicked and gapped substrate are indicated to either the right or left of the horizontal lines. The orientation is 5' to the left.

containing 30% glycerol, 0.5 M NaCl, 50 mM HEPES, pH 8.0, 1 mM dithiothreitol, and 0.1 mM EDTA. Where specified 100 ng of HMG-(I/Y) was added to the reaction mixtures. The preincubation reaction mixtures were placed on ice overnight. The volume of each preincubation mixture was then increased to 10 μl with the addition of MgCl₂ to a final concentration of 7.5 mM, and the incubation assay mixture was incubated at 37 °C for 2 h. The reactions were stopped by addition of the volume to 150 μl of 25 mM Tris, pH 8.0, 150 mM KCl, and 2.5 mM MgCl₂. After digestion for 60 min at 37 °C, the reaction mixtures were extracted with phenol followed by phenol-chloroform-isomyl alcohol (25:24:1 mixture). 15 μl of 3 M sodium acetate, pH 5.2, was added along with 1 μl of glycerogen (10 mg/ml stock solution). The reaction products were precipitated by the addition of 450 μl of 100% ethanol and washed twice with 70% ethanol prior to electrophoresis and autoradiography. The reaction products were separated on a 1% agarose gel run in 0.5× Tris borate, EDTA, and ethidium bromide at 10 V/cm for 2 h. Following electrophoresis, gels were submerged in 5% trichloroacetic acid for 20 min or until the bromophenol blue dye turned bright yellow. After being washed with water, the gels were dried on DE-81 paper (Whatman) in a Bio-Rad gel dryer at 80 °C for approximately 2 h under vacuum. The dried gels were exposed to autoradiographic film overnight at ~80 °C in a film cassette with GAFMED TA-3 or Kodak mid speed screens.

Cloning and Sequencing of Integrants—In all experiments, integration products pooled from several separate reactions were used directly for transformation of bacteria. The integration products were introduced into E. coli MC1061/P3 by electroporation, using a Bio-Rad electroporator with 0.1-cm electroporation cuvettes, 1.8-kV voltage, 25 μF microfarad capacitance, and 200-ohm resistance. The P3 episome is maintained at a low copy number. Therefore, only 40 μg/ml of ampicillin, 15 μg/ml of kanamycin, or 10 μg/ml of tetracycline were required for selection. Under these conditions, we detected no colonies after supF selection when the donor, acceptor, or donor and acceptor were electroporated into cells in the absence of IN. Plasmid DNAs were recovered from individual clones, and integration junctions were sequenced by using primers U3seq (for sequencing the U3 junction) and U5seq (for sequencing the U5 junction). Sequencing was performed using the Thermo-Sequenase kit (U.S. Biochemical Corp.).

RESULTS

Model Substrates—In the present study, we have employed a series of model HIV-1 and ASV oligodeoxyribonucleotide substrates to examine the action of viral and/or cellular enzymes in catalyzing the repair and strand closure reactions associated with integration of viral DNA. The substrates contain HIV-1 or ASV U5 LTR terminal sequences in which the conserved CA dinucleotide is covalently linked to an acceptor DNA sequence. Substrate 1 in Fig. 1A depicts a simplified HIV-1 integration intermediate with a gap in only one of two strands, representing a single viral-host DNA junction. For this substrate, the gap is five nucleotides in length. The U5 HIV-1 LTR upper strand also contains at its 5' end a two-nucleotide mismatch. To obtain fully repaired covalently closed DNA, the gap in substrate 1 has to be repaired, the 5' overhang has to be excised, and the resultant nick has to be sealed. Substrate 2 in Fig. 1A represents an HIV-1 DNA intermediate in which the gap is already repaired to form a nick but still contains the two-nucleotide 5' overhang. Substrate 3 is comparable with substrate 2 but contains the ASV instead of the HIV-1 U5 LTR sequence, and the gap is six rather than five nucleotides in length. We have also examined several HIV-1 substrate variants in which base pair inversions were introduced at LTR positions 3–7, adjacent to and/or including the conserved CA dinucleotide. The number system used here refers to the ends of the unintegrated LTRs where the CA dinucleotide is at positions 3 and 4. Substrate 4 (Fig. 1A) contains a 2-base pair substitution in the HIV-1 U5 LTR IN recognition sequences at positions 5 and 6. Substrate 5 and 6 contain base pair inversions at positions 4–7 and 3 and 4, respectively. For ease of discussion, we refer to the bottom covalently linked LTR cell DNA strand (36 mer) as strand A (Fig. 1B). Strand B refers to the upper LTR strand with the two-nucleotide 5' overhang (21-mer). Strand D refers to the upper cell DNA strand (12-mer) that serves as the primer to fill in the gap. Under these conditions, we detected no colonies after supF selection when the donor, acceptor, or donor and acceptor were electroporated into cells in the absence of IN. Plasmid DNAs were recovered from individual clones, and integration junctions were sequenced by using primers U3seq (for sequencing the U3 junction) and U5seq (for sequencing the U5 junction). Sequencing was performed using the Thermo-Sequenase kit (U.S. Biochemical Corp.).
The HIV-1 gapped substrate 1 (Fig. 1) is a 22-mer. Above each gel analysis is a stick figure representation of the substrate where the open star indicates the position of the radioactive label.

Removal of the Two-nucleotide 5' Overhang.—To look for activities that would remove the two-nucleotide 5' overhang, the HIV-1 gapped substrate 1 (Fig. 1A) was prepared with a 5'-32P end label on strand B. This substrate was incubated with different viral and/or cellular enzymes, and the gel electrophoresis analysis of the products is shown in Fig. 2. Incubation with purified HIV-1 RT in the presence or absence of deoxynucleotides (data not shown) or with IN (Fig. 2A, lane 7) failed to release the mismatched nucleotides. In contrast and as a control, the addition of FEN-1, which is known to remove 5' overhang strands from duplexes, resulted in the endonucleolytic release of the 5'-32P label as a series of bands representing dimers to hexamers and/or larger products (Fig. 2A, lane 2). The most abundant product was a tetranucleotide, even though the initial mismatch was only two nucleotides in length. The addition of PCNA to the reaction resulted in the release of larger 5'-labeled fragments (Fig. 2A, lane 3), whereas the addition of HIV-1 RT and/or IN did not influence the removal of the overhang strand (Fig. 2A, lanes 4 and 5). If unlabeled deoxynucleotides were added to the reaction with RT, a 22-mer 5'-end labeled product was observed, because of a nontemplated addition reaction to the 3' end of the strand B (Fig. 2A, lanes 5 and 6).

If a nicked substrate with a 5'-32P-labeled two-nucleotide overhang (Fig. 1A, Substrate 2) was substituted for substrate 1, FEN-1 released the 5' overhang, but the predominate product in this case was a dinucleotide. Some mononucleotides were detected (Fig. 2B, lanes 2 and 3). The presence of HIV-1 IN had no significant effect on FEN-1-mediated removal of the 5' mismatched sequences (Fig. 2B, lane 4). To confirm the size of FEN-1 released products as well as to follow the exonuclease activity of FEN-1, strand B was prepared with a 3'-32P end label (22-mer). In the presence of FEN-1, the 22-mer was converted to a series of smaller oligos ranging in size from 20 to 17 nucleotides or smaller (Fig. 2C). A 21-mer was not detected. Although FEN-1 has the ability to release products greater than two nucleotides in length, we asked whether there was a transient formation of a nick that could be trapped by DNA ligase. Substrate 2 with a 3' end label on strand B (22-mer) was incubated with FEN-1 in the presence or absence of T4 DNA ligase (Fig. 3, lanes 2 and 3, respectively). In the presence of both FEN-1 and ligase (lane 2), there was a quantitative conversion of the 22-mer to a ligated 37-mer product that would represent a covalent fusion of strands B and C. This indicates that a nick had formed in all of these substrates that could be sealed by ligase.

It has been reported that integrase alone is able to remove the overhang and seal the break with a substrate that contained a 5' overhang at a nick via a "disintegration-like" reaction (2, 10–11). We therefore incubated substrate 2 with a 3'-32P end label on strand B (22-mer) and looked for the appearance of a 37-mer product. As shown in Fig. 4A (lane 3) a small percentage (1–2%) of the labeled 22-mer was converted to the expected covalently linked 37-mer product by the action of HIV-1 IN. If FEN-1 was added along with IN to this reaction, the 37-mer product was no longer detected (Fig. 4A, lane 2). We further explored the specificity of the above trans-esterification reaction by presenting HIV IN with a nicked substrate mimicking the host DNA-U5 LTR junction of an ASV integration intermediate (substrate 3) again with the label introduced at the 3' end of strand B. The results are shown in Fig. 4B. Because strand B is a 16-mer in this substrate, the expected size of the covalently closed product would be a 32-mer. As shown in Fig. 4B (lane 2), a 32-mer was detected. The extent of the reaction was similar to that observed with the HIV-1 model substrate shown in panel A. We confirmed that HIV-1 IN catalyzed both overhang removal and ligation by using substrate 3, which contained a 5'-32P-end label introduced on strand C (18-mer). The expected product in this case would be a 31-mer, which was detected as shown in Fig. 4C (lane 2).

We further explored the specificity of the reaction by presenting HIV-1 IN with model nicked substrates mimicking the HIV-1 U5 integration intermediates but with base pair substitutions introduced at positions 3–7 in the LTR IN recognition...
sequence as shown in Fig. 1A (substrates 4–6). Incubation of the wild type substrate 2 with HIV-1 IN produces the expected 37-mer product (Fig. 5A, lane 2). Base pair inversions at positions 5 and 6 or positions 4–7 had little detectable effect on formation of the 37-mer covalently linked product (Fig. 5A, lanes 4 and 6). Base pair inversions at positions 3 and 4 resulted in a reduction of the amount of the 37-mer detected (Fig. 5A, lane 8) as previously noted (9, 10). As discussed below, all of these base pair substitutions adversely affect concerted DNA integration in vitro. Taken together, these results indicate that the formation of the covalently linked DNA product via a disintegration-like reaction is relatively independent of sequence.

Effect of HIV-1 U5 Base Pair Inversion on Integration in Vitro—Base pair substitutions at positions 5 and 6 and positions 4–7 in an ASV concerted DNA integration substrate change the efficiency and, in part, the mechanism of integration (18, 4). These mutations, as well as base pair inversions at positions 3 and 4, cause defects in concerted HIV-1 DNA integration in vitro. The HIV-1 reconstituted system uses purified HIV-1 IN, cellular HMG-(I/Y), a mini donor DNA containing a supF transcription unit flanked by 20 base pairs of the wild type HIV-1 U3 and U5 recognition sequences, and a large RFI acceptor DNA (6). The donor DNA is radiolabeled, and integration is followed by incorporation of the small donor into the larger acceptor. The integration products formed with a wild type HIV-1 donor DNA is shown in Fig. 5B (lane 1). When the base pair substitutions are introduced into the U5 IN recognition sequence at positions 5 and 6, 4–7, or 3 and 4, respectively, there is a progressive decrease in detectable integration products compared with wild type (Fig. 5B, lanes 2–4). This was similar to the effect of base pair inversions introduced into the ASV U3 LTR sequence (18).

The products from these reactions were introduced into bacteria containing a P3 plasmid with drug-resistant markers that included amber mutations in their coding sequence. Individual integrants were isolated and sequenced, and the results are summarized in Tables I and II. Using the wild type donor, integrants exhibit characteristics associated with integration in vivo. This includes the loss of the two base pairs from the ends of the LTRs and mostly five base pair duplications introduced into the acceptor DNA at the site of insertion (in Table I and Hindmarsh et al. (6)). For the mutated donors, no integrants were recovered from reactions that contained base pair substitutions at positions 4–7 or 3 and 4. However, the number of integrants recovered from reactions with the base pair substitutions at U5 positions 5 and 6 was reduced by 40% compared with wild type. Among these integrants, approximately 15% were derived from a nonconcerted DNA integration mechanism that produced large deletions in the acceptor DNA (Table II). This is similar to what is observed when base pair inversions are introduced into comparable U5 positions of an ASV donor DNA substrate (18).

Effect of HMG-(I/Y) on HIV-1 Overhang Removal and Ligation Reactions—We have previously reported that HMG-(I/Y) stimulates HIV-1 concerted integration using a reconstituted assay (6) and confirmed this observation as shown in Fig. 6A. Because this cellular protein stimulates the initial processing and joining reactions, we asked whether it would also stimulate the overhang removal/closure reactions catalyzed by HIV-1 IN using substrate 2. As shown in Fig. 6B, the formation of the 37-mer product was unaffected by the presence of HMG-(I/Y) (lanes 3 and 4).

Gap Repair—We next examined the ability of HIV-1 or AMV RT to fill the gap in our model substrate, forming a nicked duplex with the 5' overhang. The assay used to follow the repair reaction depended upon the formation of a nicked substrate that could be converted to a 37-mer product in the presence of FEN-1 and T4 DNA ligase. This assay was used because AMV and HIV-1 RT are known carry out strand displacement DNA synthesis (22, 23). When coupled with pausing of RT on template during reverse transcription, these activities would produce a heterogeneous population of different sized DNA products. This was, in fact, observed when substrate 1...
was incubated with RT and labeled deoxynucleotides (data not shown). Strand B of substrate 1 was 3’ end labeled with a [α-32P]ddATP to form the 22-mer and then incubated with HIV-1 RT (Fig. 7, lanes 2–4). Using the coupled assay with FEN-1 and ligase, substantial amounts of the 37-mer product were detected (Fig. 7, lane 4). A similar result was obtained if AMV RT was substituted for HIV-1 RT (Fig. 7, lane 7). These results indicate that both HIV-1 and AMV RT can fill in the gap to form a nicked substrate with a 5’ overhang equivalent to substrate 2. If FEN-1 was left out of the assay, no 37-mer product was detected (Fig. 7, lanes 3 and 6). In separate experiments, we observed that cellular DNA polymerase δ, PCNA, and RF-C would also fill in the gap to form the nicked substrate (data not shown).

**Coupled Gap Repair, Overhang Removal, and Ligation in the Presence of RT and HIV-1 IN**—The activity of IN on a nicked model substrate suggested that IN could remove the 5’ overhang if a DNA polymerase, such as RT, repaired the gap. The same substrate used in the experiment described in Fig. 7 was incubated with HIV-1 IN and RT in the presence of unlabeled deoxyribonucleotides to determine whether the combination of the two viral enzymes was sufficient to catalyze the final steps in integration without the need of cellular enzymes. As shown in Fig. 8 (lane 3), HIV-1 IN and RT were sufficient to form the 37-mer product. Some smaller sized DNA products were detected because of possible nuclease contamination of our HIV-1 RT preparation. The 37-mer, but not the smaller products, was detected if AMV RT was substituted for HIV-1 RT (Fig. 8, lane 4). DNA polymerase δ could also substitute for HIV-1 RT in the reaction (data not shown). Taken together, these results indicate that any DNA polymerase that can fill in the gap will be unable to recover integrants when the reaction products were introduced into bacteria. With the base pair inversions at U5 LTR positions 5 and 6, integrants were recovered, but at about 40% the efficiency of comparable reactions with the wild type donor DNA. Among these recovered integrants, approximately 18% were derived by a nonconcerted DNA integration mechanism that introduced deletions into the acceptor DNA (Table II). This is similar to what is observed when base pair inversions are introduced into comparable positions of an ASV donor DNA substrate.2 Of the concerted DNA integrants sequenced, another 18% contained deletions in the donor DNA, 80% of which were in U3 and the remainder in U5. Of the deletions found in U3, three resulted from a loss of 4 base pairs and utilized the same CT dinucleotide. The other lost 13 base pairs and utilized an internal GA dinucleotide that was previously found in ASV integrants containing base pair inversions at the comparable positions 5 and/or 6 in ASV donor DNA substrates. These results agree with other reports (18)2 that the highly conserved CA dinucleotide is not absolutely required for concerted DNA integration. Among the sequenced integrants, approximately 21% arose from multiple insertions of donor into donor DNA, which in turn were integrated into the acceptor DNA. These products were not observed with the ASV reconstituted system. Taken together, these results indicate that mutations introduced into positions 3–7 of the HIV-1 IN recognition sequence adversely affect the efficiency and mechanism of DNA integration catalyzed by HIV-1 IN.

To examine further the properties of the reconstituted integration reaction, we used a model DNA substrate with a five-nucleotide gap and a two-nucleotide 5’ overhang to examine the ability of HIV-1 RT and IN, respectively, to fill in the gap, remove the overhang, and covalently close the DNA. The sequences used in this substrate represent a U5 HIV-1 IN recognition sequence linked to an acceptor DNA at a site previously used for integration in vitro in the above reconstituted system (6). To detect the covalently closed product, the radioactive label had to be placed on the 3’ end of the overhang strand. The label was also introduced with a deoxyribonucleotide to prevent end addition to the overhang strand. Such labeling allowed discrimination of a covalently linked product from one in which RT displaced the overhang strand and extended the primer to the end of the template. In the presence of deoxyribonucleotides, RT derived from either HIV-1 or AMV readily

---

**DISCUSSION**

Integration *in vivo* and *in vitro* is dependent upon the sequences at the LTR ends of viral DNA (1, 6, 8, 18, 19, 25).2 In this report, we demonstrate that base pair substitutions at positions 3 and 4, 5 and 6, or 4–7 of the HIV-1 U5 donor DNA end significantly decrease the efficiency of concerted DNA integration *in vitro*, as detected by gel electrophoresis analysis of reaction products. In fact, the defect caused by base pair inversions at U5 LTR positions 3 and 4 or 4–7 was so severe that we

---

**TABLE I**

| Sequence of donor-acceptor integration junctions | Base pair duplications in acceptor DNA | Plasmid position of integration U5-U3 |
|-----------------------------------------------|--------------------------------------|-------------------------------------|
| U5                                           | U3                                   |                                     |
| cgcctACGAT TTCCAgcgaag                       | 5                                   | 2623–2627                           |
| gggttACGAT TTCCAcccagaa                     | 5                                   | 650–654                             |
| gggggACGAT TTCCAccccc                       | 5                                   | 1233–1229                           |
| cggtgACGAT TTCCAggag                        | 5                                   | 2900–2904                           |
| cggggACGAT TTCCAggagc                      | 5                                   | 1176–1180                           |
| cagccACGAT TTCCAgcctg                      | 5                                   | 533–528                             |
| ggcACGAT TTCCAgcattc                      | 5                                   | 808–812                             |
| acgacACGAT TTCCAgctgc                      | 5                                   | 551–547                             |
| ggaagACGAT TTCCAccccc                      | 5                                   | 350–354                             |
| ctaACGAT TTCCAgatag                        | 4                                   | 698–695                             |
| cacACGAT TTCCAgttg                         | 3                                   | 1568–1570                           |
| aagactACGAT TTCCAgctcagttt                 | 0a                                  | U5–737, U3–756                      |

*Deoxyribonucleotide sequence of the junction of the donor integration into the acceptor DNA. The sequence for only the 3’ overhang strand is presented. Lowercase letters denote duplication of the cell DNA; uppercase letters indicate the processed viral DNA sequences, which have lost 2 base pairs from each end unless otherwise indicated.

b Denotes no base pair duplication indicative of a nonconcerted DNA integration mechanism.

c Denotes deletion introduced into the acceptor DNA.

---
FIG. 6. Effect of HMG-(I/Y) on integration reactions. A, concerted integration reactions with (lane 1) or without (lane 2) HMG-(I/Y) (100 ng) using a wild type donor and purified HIV-1 IN (2.5 pmol) in a reconstituted system are as described in the legend to Fig. 5. B, HIV-1 IN (10 pmol), as indicated, was incubated with substrate 2 as described in the legend to Fig. 5 in the presence (lanes 2 and 4) or absence of (lanes 1 and 3) of HMG-(I/Y) (200 ng). All other notations are as in the legend to Fig. 5.
Gap Repair and 5' Overhang Removal

Fig. 7. Formation of a covalently linked product catalyzed by RT, FEN-1, and ligase. Gapped substrate 1 (1 pmol) with a 32P-end label on strand B (Fig. 1A) was incubated as indicated (lanes 1–7) with or without FEN-1 (125 fmol), T4 DNA ligase (400 units) and/or HIV-1 (110 μg) or AMV RT (2 units). All other notations are as described in the legend to Fig. 5.

Fig. 8. Formation of a covalently closed product catalyzed by RT and IN. Gapped substrate 1 (10 pmol) as described in the legend to Fig. 7 was incubated with equimolar amounts of HIV-1 (lanes 2 and 3) or AMV RT (lanes 4 and 5) and/or HIV-1 IN (lanes 3 and 4) as indicated.

repaired the gap. This repair was independent of sequence in the template and was catalyzed by other DNA polymerases such as cell DNA polymerase ε in the presence of PCNA and RF-C (data not shown). A covalently closed product was not detected in reactions containing only RT. Thus, neither HIV-1 nor AMV RT was capable of removing the two-nucleotide 5' overhang. However, if HIV-1 IN was added to the reaction, a small amount of a covalently closed product was observed. Thus RT and IN are sufficient to covalently close a gapped substrate. Substituting a nicked for a gapped substrate could eliminate the requirement for RT.

Covalent closure of the nicked substrate with a two-nucleotide 5' overhang is analogous to an IN-catalyzed disintegration reaction in which IN recognizes the overhang structure rather than sequence. The IN-catalyzed closure reaction was abolished when FEN-1 was added, most likely because this nuclease removed the 5' overhang. This interpretation was confirmed by the observation that a nicked DNA substrate lacking a 5' overhang was not sealed by IN (data not shown). Disintegration reactions are sequence-independent (12, 13). To determine whether the same property characterized the closure reaction, we prepared a series of mutant HIV-1 substrates with base pair inversions at positions 3–7 in the US HIV-1 LTR end sequence adjacent to and including the conserved CA dinucleotide. As noted above, these changes dramatically reduced the integration efficiency of an HIV-1 donor DNA in the reconstituted in vitro concerted integration reaction. However, although base pair substitutions at the conserved CA dinucleotide reduced the rate of covalent closure, base pair inversions at positions 5 and 6 and 4–7 had no effect on the reaction (Fig. 5A). These results are consistent with the findings of Kulkosky et al. (10) and indicate that the covalent closure of these substrates has the characteristics of a disintegration-like reaction.

Although the combination of RT and IN can repair the gap and remove the 5' overhang to form covalently closed DNA in vitro, this was accomplished with very low efficiency. The rate-limiting step is the disintegration-like closure reaction rather than gap repair. Disintegration is known to occur with low efficiency (2, 10) and, unlike the processing and joining reactions (5, 6), is not stimulated by the addition of the cellular protein HMGI(Y) (Fig. 6). This is consistent with a mechanism whereby HMGI(Y) acts on the LTR ends to stimulate the processing and joining reactions. During replication in its host cell, an infecting retrovirus catalyzes only a single DNA integration event. Thus, a high turnover rate may not be required for the final closure step. However, it is possible that other cellular proteins can affect the rate of this reaction in a natural infection. If, as estimated from our in vitro experiments, that only 1–2% of the DNA integration intermediate substrates are closed by IN, then within a given population of infected cells, most integration events would be incomplete unless cellular enzymes were recruited to complete the process. The requirement for cellular enzymes is consistent with the observation that in certain repair-deficient cells (e.g., DNA-PKcs, Ku86, or XRCC4-deficient), retroviral DNA integration is sensed as DNA damage and induces apoptosis (24).

Results from DNA transfection experiments indicate that cellular enzymes can repair the integration intermediates produced by IN in vitro. For example, the plasmid integrants we isolate after transfection of bacterial cells with the products of IN-catalyzed in vitro integration reactions are covalently closed plasmids with the expected host-viral DNA junction sequences. In this report, we demonstrate that FEN-1 nuclease together with T4 DNA ligase can close a nick with a two-nucleotide 5' overhang in our model substrates very efficiently. Nevertheless, it is likely that one of the eukaryotic cellular DNA ligases would serve this function during a natural infection. The importance of XRCC4 to retroviral DNA integration (24) suggests that its binding partner, ligase IV, is the likely candidate. FEN-1 is a structure-specific nuclease that recognizes and cleaves the Y structure in a DNA flap (16, 26–28). This enzyme is implicated in both DNA replication and mismatch repair (29, 30). In yeast, the FEN-1 equivalent RAD27 participates in the single strand-annealing pathway of repair that requires processing of 5' ends (31). Here, we demonstrate that FEN-1 can recognize the model retroviral integration intermediate and efficiently remove the two-nucleotide 5' overhang in either a gapped or nicked substrate. We also observed that this nuclease cleaved several deoxyribonucleotides into the duplex structure (Fig. 2). However, a transient nick was formed that was sealed quantitatively by DNA ligase. Neither RT nor IN affected the FEN-1 catalyzed 5' overhang cleavage with these model substrates. PCNA, a protein involved in DNA replication, was reported to stimulate FEN-1 activity (especially exonuclease) on flap substrates (21, 32). However, we did not observe any significant stimulation of FEN-1-mediated cleavage in our system. This may be explained by the small size of the overhang used in our integration intermediate. Although we have demonstrated that FEN-1 can remove the 5' overhang, it is possible that other cellular nucleases, such as the FEN-1 ho-
mologue, XPG (21) could participate in the reaction.

Although results from our in vitro reconstituted experiments do not rule out a direct role for retrovirus RT and IN in catalysis of the last steps of retroviral DNA integration, the low efficiency that we observed for this reaction does not support this simple mechanism. It seems more likely that host repair enzymes such as those described in this report are critical for completion of retroviral DNA integration in vivo.

Acknowledgments—We thank Dr. Min S. Park for generously providing purified preparations of FEN-1 and PCNA and Dr. Jerard Hurwitz (Sloan Kettering Institute) for DNA polymerase δ and RF-C. We thank Dr. Stuart LeGrice (NCI, NIH, Frederick, MD) for purified HIV-1 RT and Dr. Ray Reeves for HMG-(I/Y). We also thank Mark Andrake for critically reading this manuscript.

REFERENCES

1. Hindmarsh, P., and Leis, J. (1999) Microbiol. Mol. Biol. Rev. 63, 836–843
2. Chow, S. A., Vincent, K. A., Ellison, V., and Brown, P. O. (1992) Science 253, 836–843
3. Katz, R. A., Merkel, G., Kulkosky, J., Leis, J., and Skalka, A. M. (1990) Cell 63, 87–95
4. Craigie, R., Fujiwara, T., and Bushman, F. (1999) Cell 92, 829–837
5. Farnet, C. M., and Bushman, F. D. (1997) Cell 88, 483–492
6. Hindmarsh, P., Ridky, T., Reeves, R., Andrake, M., Skalka, A. M., and Leis, J. (1999) J. Virol. 73, 2994–3003
7. Carteau, S., Gorelick, R. J., and Bushman, F. D. (1999) J. Virol. 73, 6670–6679
8. Vincent, K. A., York-Higgins, D., Quiroga, M., and Brown, P. O. (1999) Nucleic Acids Res. 18, 6045–6047
9. Vink, C., Groenink, M., Elgersma, Y., Fouchier, R. A. M., Termette, M., and Plasterk, R. H. A. (1990) J. Virol. 64, 5626–5627
10. Kulkosky, J., Katz, R. A., Merkel, G., and Skalka, A. M. (1995) Virology 206, 448–456
11. Sherman, P. A., Dickson, M. L., and Fyfe, J. A. (1992) J. Virol. 66, 3593–3601
12. Balakrishnan, M., and Jonsson, C. B. (1997) J. Virol. 71, 1025–1035
13. Van Den Ent, F. L., Vink, C., and Plasterk, R. H. A. (1994) J. Virol. 68, 7825–7832
14. Bushman, F. D., Engelman, A., Palmer, I., Wingfield, P., and Craigie, R. (1993) Proc. Natl. Acad. Sci. U. S. A. 90, 3428–3432
15. Dyda, F., Hickman, A. B., Jenkins, T. M., Engelman, A., Craigie, R., and Davies, D. R. (1994) Science 266, 1981–1986
16. Harrington, J. J., and Lieber, M. R. (1994) EMBO J. 13, 1235–1246
17. Nissen, M. S., and Reeves, R. (1995) J. Biol. Chem. 270, 4355–4360
18. Aiyar, A., Hindmarsh, P., Skalka, A. M., and Leis, J. (1996) J. Virol. 70, 3571–3580
19. Katzman, M., Katz, R. A., Skalka, A. M., and Leis, J. (1989) J. Virol. 63, 5319–532720. Marayuk, R., and Sergeant, A. (1980) Anal. Biochem. 105, 403–404
21. Khungland, A., and Lindahl, T. (1997) EMBO J. 16, 3314–3348
22. Collett, M. S., Leis, J. P., Smith, M. S., and Faras, A. (1978) J. Virol. 26, 498–509
23. Hottiger, M., Produst, V., Thimmig, R. L., McHenry, C., and Hubscher, U. (1994) J. Biol. Chem. 269, 986–991
24. Daniel, R. Katz, R. A., and Shalka, A. M. (1999) Science 284, 644–647
25. Kukolj, G., and Skalka, A. M. (1995) Gene Dev. 9, 2556–2567
26. Murante, R., Huang, L., Turchi, J., and Bambara, R. (1994) J. Biol. Chem. 269, 1191–1196
27. Murante, R., Rust, L., and Bambara, R. (1994) J. Biol. Chem. 270, 30377–30383
28. Robins, P., Pappin, D., Wood, R. D., and Lindahl, T. (1994) J. Biol. Chem. 269, 26385–26388
29. Johnson, R., Kovovali, G., Prakash, L., and Prakash, S. (1995) Science 269, 238–240
30. Wu, X., Li, J., Li, X., Hsieh, C.-L., Burgers, P. M. J., and Lieber, M. R. (1996) Nucleic Acids Res. 24, 2036–2043
31. Wu, X., Wilson, T. E., and Lieber, M. R. (1999) Proc. Natl. Acad. Sci. U. S. A. 96, 1393–1398
32. Gary, R., Kim, K., Cornelius, H. L., Park, M. S., and Matsumoto, Y. (1999) J. Biol. Chem. 274, 4354–4363