JOURNAL OF VIROLOGY, Apr. 1992, p. 2359–2368
0022-538X/92/042359-10$02.00/0

Vol. 66, No. 4

Both Substrate and Target Oligonucleotide Sequences Affect In Vitro Integration Mediated by Human Immunodeficiency Virus Type 1 Integrase Protein Produced in *Saccharomyces cerevisiae*

ANDREW D. LEAVITT,[1,2]* ROBERT B. ROSE,[3] AND HAROLD E. VARMUS[1,3]

Department of Microbiology and Immunology,[1] Cancer Research Institute,[2] and Department of Biochemistry and Biophysics,[3] The University of California, San Francisco, California 94143

Received 19 November 1991/Accepted 17 January 1992

Integration of retroviral DNA into the host cell genome requires the interaction of retroviral integrase (IN) protein with the outer ends of both viral long terminal repeats (LTRs) to remove two nucleotides from the 3′ ends (3′ processing) and to join the 3′ ends to newly created 5′ ends in target DNA (strand transfer). We have purified the IN protein of human immunodeficiency virus type 1 (HIV-1) after production in *Saccharomyces cerevisiae* and found it to have many of the properties described for retroviral IN proteins. The protein performs both 3′ processing and strand transfer reactions by using HIV-1 or HIV-2 attachment (*att*) site oligonucleotides. A highly conserved CA dinucleotide adjacent to the 3′ processing site of HIV-1 is important for both the 3′ processing and strand transfer reactions; however, it is not sufficient for full IN activity, since alteration of nucleotide sequences internal to the HIV-1 U5 CA also impairs IN function, and Moloney murine leukemia virus *att* site oligonucleotides are poor substrates for HIV-1 IN. When HIV-1 *att* sequences are positioned internally in an LTR-LTR circle junction substrate, HIV-1 IN fails to cleave the substrate preferentially at positions coinciding with correct 3′ processing, implying a requirement for positioning *att* sites near DNA ends. The 2 bp normally located beyond the 3′ CA in linear DNA are not essential for in vitro integration, since mutant oligonucleotides with single-stranded 3′ or 5′ extensions or with no residues beyond the CA dinucleotide are efficiently used. Selection of target sites is nonrandom when *att* site oligonucleotides are joined to each other in vitro. We modified an in vitro assay to distinguish oligonucleotides serving as the substrate for 3′ processing and as the target for strand transfer. The modified assay demonstrates that nonrandom usage of target sites is dependent on the target oligonucleotide sequence and independent of the oligonucleotide used as the substrate for 3′ processing.

The retroviral life cycle is characterized by the formation of a provirus, a DNA copy of the viral genome integrated into a host cell chromosome. Genetic studies have demonstrated that two viral components are essential for retroviral integration: (i) integrase (IN), the protein encoded by the 3′ end of *pol* (9, 10, 21, 22, 24), and (ii) attachment (*att*) sites, DNA sequences located at the ends of the viral long terminal repeats (LTRs) (6, 7, 13, 20, 23). IN first acts in the cytoplasm in a 3′ processing reaction which removes the terminal two nucleotides from the 3′ end of each strand of linear viral DNA (4, 14, 23) so that the 3′ ends of the viral DNA terminate with the dinucleotide CA. A nucleoprotein complex containing the processed viral DNA then enters the nucleus (2), where IN cleaves the chromosomal target DNA in a staggered fashion and ligates the processed 3′ ends of viral DNA to the 5′ ends of the newly cut target DNA in a strand transfer reaction (3, 11, 12). The product of the reaction is a gapped intermediate, which is thought to be repaired by host-derived DNA synthetic machinery, thereby generating the virus-specific short direct repeats that flank the provirus.

Recent advances in understanding IN activity have come from newly developed in vitro assays. The first in vitro assay for retroviral integration used nucleoprotein complexes from murine leukemia virus (MLV) infected cells to provide viral DNA and integration activity (3). The assay was then modified by substituting a heterologous DNA substrate with synthetic *att* sites to function in place of the viral DNA synthesized during infection (13). Subsequently, an in vitro integration assay which requires only purified retroviral IN protein and oligonucleotides representing the ends (*att* sites) of viral DNA was developed (8, 15, 16, 25). The *att* site oligonucleotides serve as both donor and target DNA in the reaction and can undergo both the 3′ processing and strand transfer reactions. As conventionally used, the assay does not require the coordinate strand transfer of two ends of viral DNA at a single target cleavage site, as occurs during integration in vivo; in this sense, it is not a complete integration reaction. Nonetheless, this simplified assay has helped to demonstrate that IN from MLV (8), avian leukosis virus (15), and human immunodeficiency virus type 1 (HIV-1) (5, 17, 29) can mediate all but the final step in the integration events that establish a provirus in an infected cell.

To better understand HIV-1 integration, we have produced HIV-1 IN in *Saccharomyces cerevisiae* and used the purified protein in the *att* site oligonucleotide-based assay to evaluate the DNA substrate requirements for HIV-1 IN. We have also modified the assay so that different oligonucleotides are monitored as the substrates for 3′ processing and as the targets for strand transfer. This has allowed us to demonstrate that in vitro integration occurs nonrandomly in a manner dependent upon the target rather than the donor oligonucleotide sequence.

* Corresponding author.

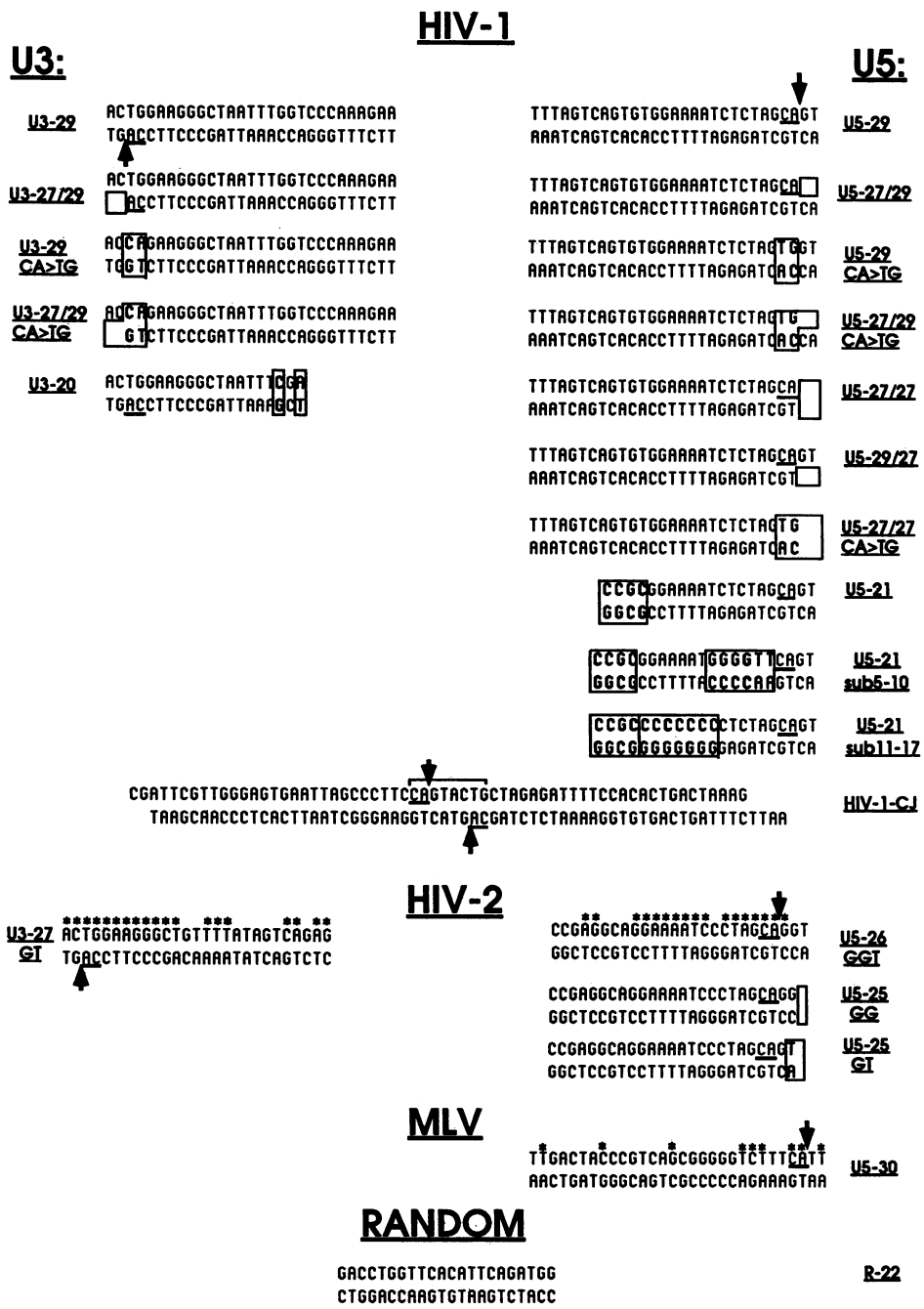

FIG. 1. Oligonucleotides used in assessment of HIV-1 IN activity. Synthetic DNA oligonucleotides were purified and annealed as described in Materials and Methods to produce the duplex oligonucleotides depicted. The conserved 3′ CA is underlined. Vertical arrows indicate the 3′ processing site. The numbers given after a hyphen refer to the length of each nucleotide strand. A single number after a hyphen implies both strands are a single length. Boxed, boldface base pairs indicate alterations from the wild-type sequence. HIV-1-CJ is the HIV-1 circle junction sequence. The bracketed sequence in HIV-1-CJ represents the inverted repeats. An asterisk over a base pair in the HIV-2 or MLV sequence indicates identity to the corresponding residue in the HIV-1 sequence. The R-22 sequence (RANDOM) comes from the N-II exon of mouse c-*src* and has no overt homology with HIV-1 *att* sites. HIV-1 sequences are from Myers et al. (19).

MATERIALS AND METHODS

Oligonucleotides. Oligonucleotides were obtained from the University of California, San Francisco, Biomolecular Resource Center, except for the HIV-2 oligonucleotides, which were a gift from Steve Hughes, National Cancer Institute, Frederick, Md. Oligonucleotide names and sequences are provided in Fig. 1. T4 polynucleotide kinase (Pharmacia) was used to label the 5′ end of oligonucleotides with [γ-^{32}P]ATP (3,000 Ci/mmol; Amersham). Klenow fragment (Pharmacia) was used to label the 3′ end of oligonucleotides with [α-^{32}P]dCTP (3,000 Ci/mmol; Amersham).

Yeast strain. *S. cerevisiae* AB116 (*MATα leu2 trp1 ura3-52 prB1-1122 pep4-3 prC1-407* [cir^0]) was propagated under conditions of leucine selection before liquid culture in enriched medium as described elsewhere (1).

Protein purification. The IN-coding sequence from HIV-1 (SF2) was cloned into the yeast expression vector pBS24Ub as described elsewhere (1). The construct creates a fusion gene consisting of the yeast ubiquitin-coding sequence followed in frame by the HIV-1 IN-coding sequence. The wild-type IN sequence, including the amino-terminal phenylalanine, is maintained, and the fusion gene product is accurately cleaved in vivo to yield HIV-1 IN (1). The construct places HIV-1 IN under the transcriptional control of the glucose-regulatable alcohol dehydrogenase 2/glyceraldehyde 3-phosphate dehydrogenase hybrid promoter (1).

After growth in enriched medium for 48 h, cells were pelleted and either lysed immediately or stored at −70°C. Cells were lysed (>90% lysis as determined by phase microscopy) in 50 mM HEPES (*N*-2-hydroxyethylpiperazine-*N'*-2-ethanesulfonic acid; pH 7.5)–5 mM EDTA–1 mM dithiothreitol (DTT) in a Dyno-Mill cell mill (Willy A. Bachofen, Basel, Switzerland), with a 1:2:1.5 ratio of pelleted cells (grams) to glass beads (grams) to buffer (milliliters). Glass beads were removed by filtration over a plastic screen. The lysate was then spun for 20 min at 12,000 × *g* at 4°C. After the pellet was washed once with lysis buffer, it was resuspended in 50 mM HEPES (pH 7.5)–1 mM EDTA–1 mM DTT–1 M NaCl–10 mM CHAPS {3-[(3-cholamidopropyl)-dimethylammonio]-1-propanesulfonate}. The resuspended pellet was stirred for 2 h at 4°C and then spun for 1 h at 100,000 × *g* at 4°C. The resulting supernatant was either frozen at −70°C in 20% glycerol or used for further purification. Whether used directly or after thawing from −70°C, the supernatant was diluted to give a final composition of 25 mM HEPES (pH 7.5), 1 mM EDTA, 1 mM DTT, and 7 mM CHAPS (buffer A) plus 150 mM NaCl and loaded onto an Aphigel-Heparin (Bio-Rad) column equilibrated with the same buffer. The column was washed with buffer A plus 200 mM NaCl, and adherent proteins were eluted with buffer A plus 350 mM NaCl. Peak elution fractions containing IN were pooled, and the buffer was changed to buffer A plus 100 mM NaCl and 1 M ammonium sulfate (buffer B). The pooled fractions were loaded onto a phenyl-Sepharose column (Pharmacia) equilibrated with buffer B; the column was then washed with buffer B and eluted with buffer A plus 100 mM NaCl. Peak fractions containing IN were pooled and dialyzed into buffer A plus 400 mM NaCl. When this sample was placed on a TSK (Pharmacia) high-performance liquid chromatography sizing column, IN eluted as a well-isolated protein of roughly 32 kDa.

Western blots (immunoblots). Western blots were performed with nitrocellulose filters for transfer and 1% yeast extract (Difco) in phosphate-buffered saline (PBS) as a blocking agent. The primary antiserum was a polyclonal anti-HIV-1 serum provided by Kathelyn S. Steimer (Chiron Corporation, Emeryville, Calif.). The antiserum was raised in rabbits against an HIV-1 IN–human superoxide dismutase fusion protein from *Escherichia coli* (26). The secondary antiserum was a goat anti-rabbit serum conjugated to alkaline phosphatase (Promega). Primary and secondary antisera were incubated at 4°C in PBS plus 1% yeast extract (Difco). The Western blots were developed with nitroblue tetrazolium and 5-bromo-4-chloro-3-indolyl phosphate (Sigma).

Assay for IN activity using radiolabeled oligonucleotides. Single-oligonucleotide reactions were performed in 50 mM MOPS (morpholinepropanesulfonic acid; pH 7.0)–1 mM DTT–15 mM $MnCl_2$–10% glycerol–100 μg of bovine serum albumin per ml–roughly 20 mM NaCl (from protein storage buffer) with roughly 5 pmol of IN and 0.5 pmol of radiolabeled *att* site oligonucleotide in a total volume of 10 μl. Two-oligonucleotide reactions differed only by the addition of roughly equal molar amounts of unlabeled *att* site oligonucleotide and 3′ radiolabeled target oligonucleotide. The reaction mixtures were incubated at 30°C for 1 h, and the reactions were stopped by adding 10 μl of loading buffer (95% formamide, 20 mM EDTA, 0.05% bromphenol blue, 0.05% xylene cyanol) and heating to 100°C for 5 min. The reaction products were then analyzed by electrophoresis on 20% polyacrylamide gels with 7 M urea in Tris-borate-EDTA (TBE). Autoradiography was performed with either Amersham Hyperfilm-MP or Kodak X-Omat film. Schematic representations of the single-oligonucleotide assay have been published elsewhere (8, 15). Figure 7 shows a schematic representation of our modification of the assay with separate donor and target oligonucleotides.

RESULTS

Isolation of HIV-1 IN expressed in *S. cerevisiae*. We expressed HIV-1 IN in *S. cerevisiae*, using the ubiquitin fusion construct described in Materials and Methods. The fusion protein is cleaved in vivo to yield IN with the wild-type amino acid sequence, including the amino-terminal phenylalanine, in an amount readily detected on a Coomassie blue-stained sodium dodecyl sulfate (SDS)-polyacrylamide gel of a total cell lysate (Fig. 2, lane 2). After centrifugation at 12,000 × *g*, most of the viral protein is located in the pellet of the IN-containing lysate (Fig. 2, lanes 3 and 4). After resuspension in lysis buffer plus 1 M NaCl and 10 mM CHAPS and centrifugation at 100,000 × *g*, IN remains predominantly in the supernatant (Fig. 2, lane 5). Sequential column chromatography on heparin sulfate and phenyl-Sepharose, as outlined in Materials and Methods, yields 1.5 to 2 mg of highly purified IN (>95% pure) from 1 liter of yeast culture (Fig. 2, lane 6). Western blot analysis of the total cell lysate (Fig. 2, lane 8) and of the material eluted from the phenyl-Sepharose column (Fig. 2, lane 9) demonstrated an immunoreactive protein at the expected molecular mass of roughly 32,000 Da. The protein was not detected in the total cell lysate of an isogenic yeast strain that lacks the IN-coding sequence (Fig. 2, lanes 1 and 7). All reactions described below were performed with IN eluted from the phenyl-Sepharose column.

HIV-1 IN from *S. cerevisiae* has specific 3′ nuclease and strand transfer activity in vitro that is impaired by alterations to the HIV-1 *att* site oligonucleotide sequence. We first tested our yeast-derived HIV-1 IN for its ability to mediate 3′ processing and strand transfer by using wild-type HIV-1 *att* site oligonucleotides 29 bp in length (U5-29 and U3-29). The predicted 27-nucleotide primary cleavage product generates the predominant signal after the addition of IN to the 29-bp U3 or U5 *att* site oligonucleotide (Fig. 3A, lanes 1 to 4). The lengths of the primary cleavage products were confirmed by DNA markers (data not shown). Like Bushman and Craigie (5) and LaFemina et al. (17), we found that HIV-1 IN acts more efficiently on U5 than U3 *att* site oligonucleotides. In contrast, Vink et al. (29) observed equal activities with U5 and U3 *att* site oligonucleotides. We also found minimal 3′ processing of an oligonucleotide representing the U5 end of MLV (Fig. 3A, lanes 9 and 10). As previously noted by others with HIV-1 IN produced in insect cells (5) or *E. coli* (17, 29), HIV-1 IN-mediated reactions showed a preference

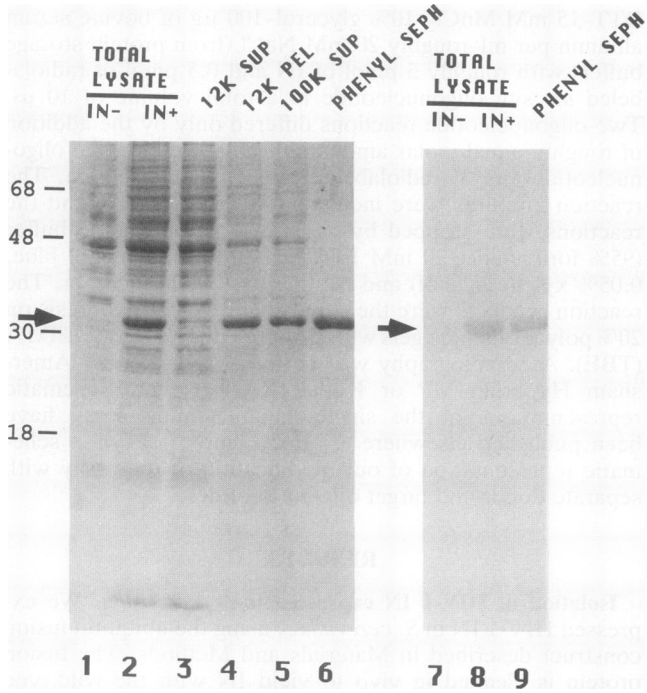

FIG. 2. Production and purification of HIV-1 IN expressed in *S. cerevisiae*. Yeast cells (strain AB116) containing an inducible expression plasmid without (pBS24Ub) or with (pBS24Ub-IN) the HIV-1 IN-coding sequence were grown as described in Materials and Methods. Samples from the purification steps outlined in Materials and Methods were analyzed by SDS–12.5% polyacrylamide gel electrophoresis. Lanes 1 to 6 were stained with Coomassie blue. Lanes 7 to 9 underwent Western blot analysis. Lanes 1 and 7, AB116(pBS24Ub) total cell lysate; lanes 2 and 8, AB116(pBS24Ub-IN) total cell lysate; lane 3, AB116(pBS24Ub-IN) supernatant from centrifugation at 12,000 × *g*; lane 4, AB116(pBS24Ub-IN) pellet from centrifugation at 12,000 × *g*, resuspended in 1 M NaCl–10 mM CHAPS; lane 5, AB116(pBS24Ub-IN) supernatant from centrifugation at 100,000 × *g*; lanes 6 and 9, AB116(pBS24Ub-IN) pooled phenyl-Sepharose eluate. The arrow indicates the location of HIV-1 IN. Molecular mass markers (in thousands of daltons) are indicated on the left.

for $MnCl_2$ over $MgCl_2$ and an absolute requirement for a divalent cation (data not shown). The strand transfer step of integration is detected in vitro by the appearance of a series of slowly migrating bands that represent oligonucleotides longer than the starting probe (Fig. 3) (8, 16). To confirm that the more slowly migrating products we observed were the result of IN-mediated strand transfer, they were isolated from gels after electrophoresis and subjected to Maxam-Gilbert sequencing. From assays which employed either U3-29 or U5-29, the sequences were as predicted for IN-mediated strand transfer (data not shown), in agreement with published reports from other groups (5, 8, 16). In an infected cell, 3′ processing occurs in the cytoplasm and strand transfer occurs in the nucleus, implying that the two steps are not coupled. Successful strand transfer with preprocessed oligonucleotides, U5-27/29 and U3-27/29, demonstrates that these two steps are also separable in vitro (Fig. 3A, lanes 5 to 8).

A 3′ CA dinucleotide constitutes the terminal two nucleotides at each 3′ end of all known proviruses. U5 *att* site oligonucleotides with the CA changed to TG significantly impaired both 3′ processing and strand transfer (Fig. 3B, lanes 1 to 4). Preprocessed oligonucleotides with the same CA-to-TG mutation also showed negligible strand transfer (Fig. 3B, lanes 5 and 6). The 3′ CA dinucleotide is therefore independently required for both 3′ processing and strand transfer. Identical requirements for the CA dinucleotide were seen with U3-derived *att* site oligonucleotides (data not shown).

We initially used oligonucleotides containing 29 bp of terminal HIV-1 sequence (Fig. 3A). However, our HIV-1 IN mediates both 3′ processing and strand transfer with oligonucleotides containing only the distal 17 bp from the U5 or U3 end of HIV-1 (Fig. 3C, lanes 1 to 4). To determine the importance of the 13 bp located adjacent to the U5 CA, we changed, as a unit, residues 1 to 6 or 7 to 13 internal to the CA dinucleotide (U5-21 sub5-10 and U5-21 sub11-17 [Fig. 1]). When incubated with HIV-1 IN, both mutants showed impaired 3′ processing and strand transfer (Fig. 3C, lanes 5 to 8), suggesting that IN has important interactions throughout at least 7 and as many as 13 bp adjacent to the U5 CA. The data do not address the importance of individual residues.

The terminal nucleotides removed during integration are not required for in vitro integration. To determine the integration-related function of the two terminal base pairs removed during retroviral integration, a series of mutant U5 *att* site oligonucleotides with altered 3′ and 5′ termini were assayed for 3′ processing and strand transfer (Fig. 4). As also demonstrated above, strand transfer was only slightly greater with wild-type preprocessed oligonucleotides, which contained two unpaired bases at the 5′ terminus, than for wild-type unprocessed oligonucleotides, which contained 2 bp beyond the conserved CA (Fig. 4, lanes 7 and 8 and lanes 1 and 2, respectively). When the starting oligonucleotide terminated with two unpaired bases 3′ to the CA, wild-type levels of 3′ processing and strand transfer were observed (Fig. 4, lanes 3 and 4). When the *att* site oligonucleotide terminated bluntly at the CA, wild-type levels of strand transfer were still observed (Fig. 4, lanes 5 and 6). To confirm that the strand transfers observed in Fig. 4, lanes 4 and 6, occurred by a mechanism similar to that seen with wild-type *att* site oligonucleotides, we tested another mutant oligonucleotide that contains TG in place of the 3′ CA dinucleotide and terminates bluntly after the TG. This mutant failed to undergo strand transfer (Fig. 4, lanes 9 and 10), as would be expected for an IN-mediated reaction. The results suggest that neither the 3′-terminal nor the 5′-terminal nucleotides removed from each end of the viral DNA during the integration reaction are essential in vitro. Furthermore, 3′ processing and strand transfer can occur in the absence of base pairing of the two nucleotides beyond the CA. Possible conflicts between these findings and results obtained with MLV mutants in vivo (6) are addressed in the Discussion.

An LTR-LTR circle junction oligonucleotide is not preferentially cleaved 3′ to the conserved CA sequences. Recent studies have established that MLV uses linear and not circular DNA forms as the precursor to provirus formation (3, 4, 14, 18). However, the LTR-LTR circle junction form contains the same sequences that are found at each end of the linear DNA species. To determine whether events in the in vitro assay with HIV-1 IN conform to in vivo observations with MLV, we radiolabeled the 5′ end of each strand of an oligonucleotide representing the HIV-1 circle junction and annealed each to unlabeled cDNA. DNA fragments of the lengths predicted for the products of nicking 3′ to the conserved CA residues (30 bp for CJ-3 and 32 bp for CJ-4

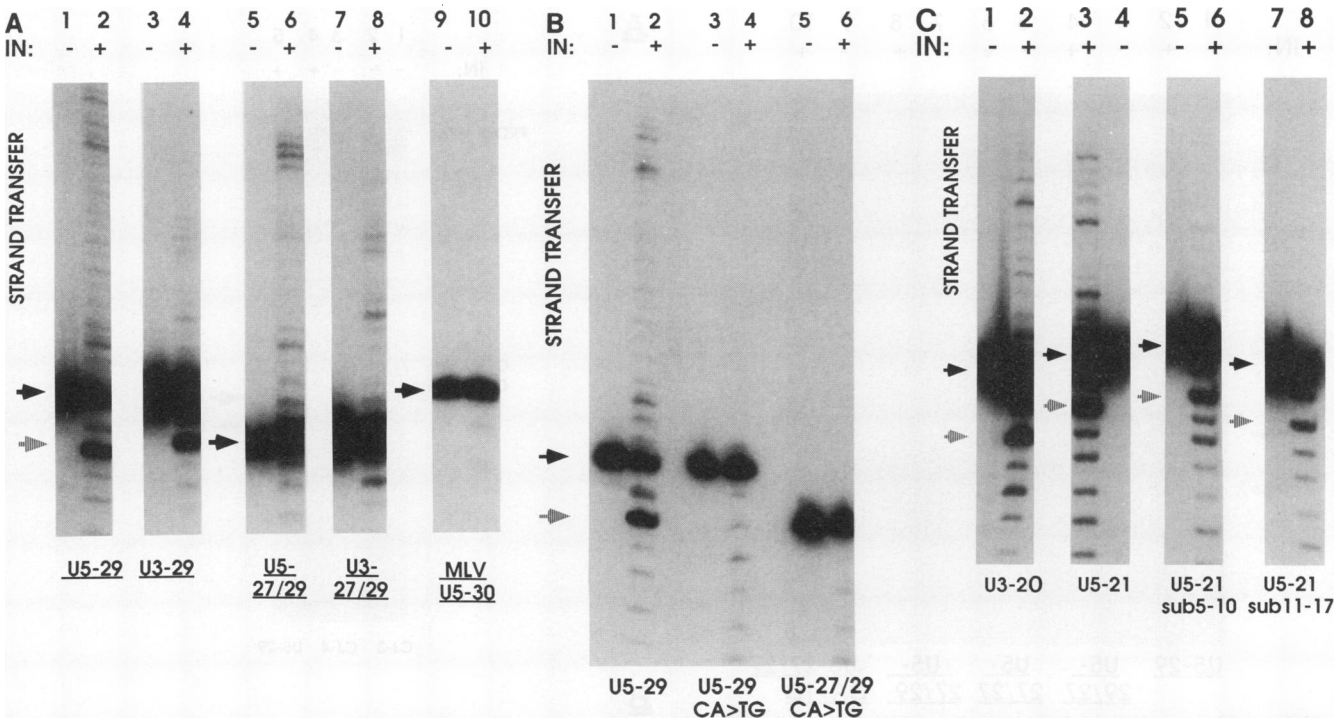

FIG. 3. Yeast-derived HIV-1 IN processes and recombines wild-type HIV-1 *att* site oligonucleotides and is sensitive to *att* site sequence alterations. Assays were performed as described in Materials and Methods. Oligonucleotides are identified below each pair of lanes (Fig. 1). Each oligonucleotide probe was incubated either without (−) or with (+) purified HIV-1 IN. (A) Lanes 1 to 4 contain the wild-type HIV-1 oligonucleotides. Lanes 5 to 8 contain preprocessed HIV-1 oligonucleotides. Lanes 9 and 10 contain the wild-type U5 end of MLV. The solid arrows represent the starting oligonucleotide probe, and the hatched arrow represents the product of 3′ processing. Strand transfer products are indicated. (B) Lanes 1 to 6 contain a series of mutants with CA mutations in the HIV-1 U5 *att* site. The solid arrow represents the starting oligonucleotide probe in lanes 1 to 4; the hatched arrow represents the product of 3′ processing in lane 2 and the labeled, preprocessed probe in lanes 5 and 6. (C) Lanes 1 to 4 contain wild-type HIV-1 oligonucleotides which contain 17 bp from the HIV-1 U3 or U5 *att* site. Lanes 5 to 8 assess the effect of altering groups of base pairs within 13 residues of the conserved CA in the U5 *att* site. The solid arrows represent starting oligonucleotide probes, and the hatched arrows represent the products of 3′ processing.

[Fig. 5B]) were not generated when oligonucleotide HIV-1-CJ was incubated with HIV-1 IN (Fig. 5A). Since 17 bp from either the U3 or U5 end is sufficient for 3′ processing (Fig. 3C, lanes 1 to 4), the lack of 3′ processing of CJ-3 or CJ-4 is not a consequence of an inadequate *att* site sequence but instead suggests a requirement that functional *att* sequences be near an end of the oligonucleotide. The several bands corresponding to oligonucleotides shorter than the starting probe (Fig. 5A, lanes 2 and 4) may be the result of contaminating nucleases, or they may represent the products of *att* site-independent endonuclease activity of IN. The locations of scissions that generated the prominent signals are indicated with solid arrows in Fig. 5B.

HIV-2 ends are functional substrates for HIV-1 IN in vitro. The HIV-1 and HIV-2 genomes have highly similar nucleotide sequences (19). For instance, at the left (U3) end of viral DNA, 15 of 17 nucleotides are identical, and at the right (U5) end, 18 of 23 are identical. However, unique among known retroviruses, unintegrated HIV-2 has 3 bp beyond the conserved 3′ CA at the U5 end, while the conventional 2 bp are present beyond the conserved 3′ CA at the U3 end (30). In contrast, HIV-1 has 2 bp distal to the 3′ CA at both the U5 and the U3 ends (28, 31).

To determine whether HIV-1 IN can mediate 3′ processing on the termini of HIV-2, a series of wild-type and mutant HIV-2 *att* site oligonucleotides were analyzed for 3′ processing and strand transfer by HIV-1 IN. Figure 6 demonstrates that HIV-1 IN preferentially processed HIV-2 *att* site oligonucleotides immediately 3′ of the conserved CA at both the U3 and U5 ends. At the HIV-2 U5 end, this was the case with the wild-type oligonucleotides (with 3 bp beyond the 3′ CA [Fig. 6, lanes 1 and 2]) and with either of two mutant oligonucleotides, each with 2 bp beyond the 3′ CA (Fig. 6, lanes 3 to 6). However, 3′ processing was consistently more efficient with 2 bp 3′ of the CA than with 3 bp 3′ of the CA. These findings agree with the observations of others that the location of the 3′ cleavage is dependent on the position of the CA dinucleotide and is not defined by a fixed distance from the end of the DNA molecule (5, 29). As we observed with HIV-1 *att* site oligonucleotides, the patterns of strand transfer products with the HIV-2 U5 and U3 oligonucleotides differ from each other, and IN-mediated reactions appear to be more efficient with U5 than with U3 *att* site oligonucleotides (Fig. 6, lanes 1 and 2 and lanes 7 and 8, respectively).

The target oligonucleotide influences the strand transfer pattern. There is no known sequence specificity for the target site of retroviral integration. Additionally, there are no known structural requirements for target DNA that are likely to be maintained in the oligonucleotide-based integration assay. Consequently, it is reasonable to assume that all sites within an oligonucleotide may serve equally as target sites for strand transfer. However, the strand transfer patterns show that all target sites are not used equally (Fig. 3A), as also observed by others studying IN from MLV (8) and

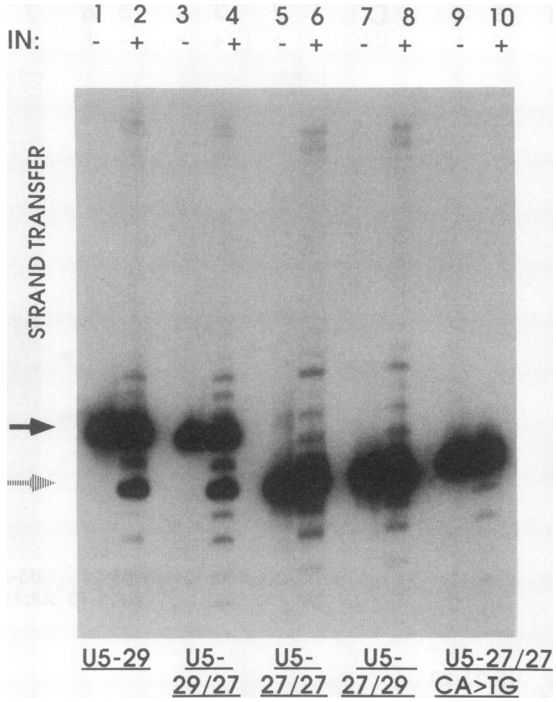

FIG. 4. The terminal nucleotides removed during provirus formation are not required for HIV-1 IN-mediated 3′ processing or strand transfer. Assays were performed as described in Materials and Methods. Oligonucleotides are identified below each pair of lanes (Fig. 1). Each oligonucleotide probe was incubated without (−) or with (+) purified HIV-1 IN. Strand transfer products migrated more slowly than the starting probe. The solid arrow represents the starting oligonucleotide probe in lanes 1 to 4; the hatched arrow represents the products of 3′ processing in lanes 2 and 4 and the starting probe in lanes 5 to 10.

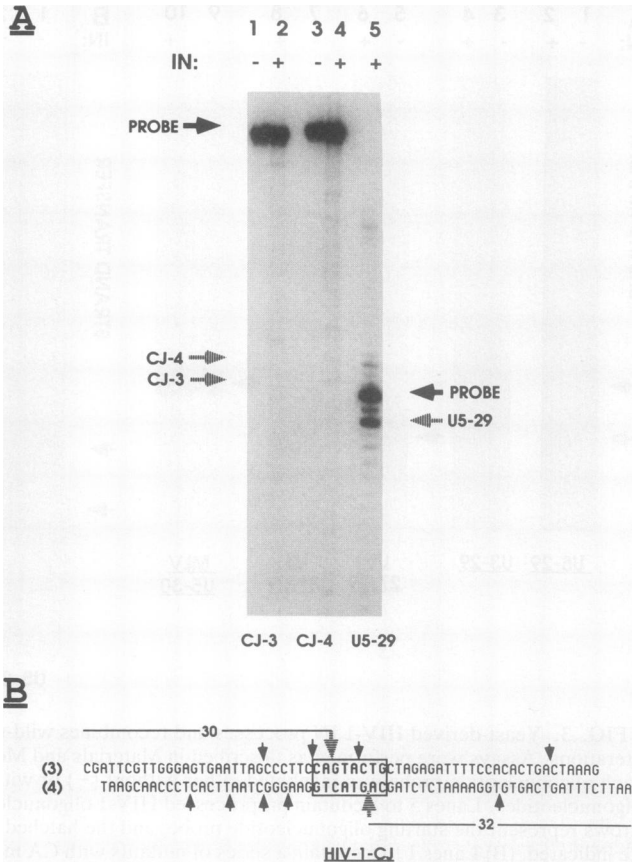

FIG. 5. HIV-1 IN does not cut LTR-LTR circle junction oligonucleotides at the 3′ processing site. Assays were performed as described in Materials and Methods. (A) Oligonucleotides are identified below each pair of lanes. CJ-3 and CJ-4 refer to HIV-1-CJ with either the upper or lower strand (as shown in panel B) 5′ end labeled. Paired lanes with a given oligonucleotide probe are either without (−) or with (+) purified HIV-1 IN. The solid arrows represent the starting probe. The hatched arrows indicate the expected location of the labeled product after 3′ processing. (B) The HIV-1-CJ oligonucleotide sequence is shown. The boldface, boxed sequence depicts the perfect, inverted repeat of the LTR-LTR circle junction. The hatched arrows are located 3′ to the conserved CA in each strand, the usual site of 3′ processing. The solid arrows indicate the location of cleavage corresponding to the prominent bands in lanes 2 and 4 in panel A. The lengths of the labeled products expected for 3′ processing of CJ-3 and CJ-4 are indicated as 30 and 32 nucleotides, respectively.

HIV-1 (5). Not only did we find nonrandom use of target sites, but different patterns were produced with HIV-1 U3 and U5 oligonucleotides. Such nonrandom use of potential target sites could be a function of the substrate or the target.

To examine the separate roles of the substrate and target oligonucleotides, we altered the standard reaction to include two different oligonucleotides, only one of which has the *att* site sequence that permits it to serve as the substrate for 3′ processing (Fig. 7). Although both oligonucleotides can serve as targets of integration, the HIV-1 *att* site oligonucleotide is not radiolabeled and therefore cannot be detected when it acts as a target in a strand transfer reaction. The second oligonucleotide, which does not contain HIV-1 *att* site sequences and therefore cannot serve as a substrate for 3′ processing, is 3′ radiolabeled and must be joined to the unlabeled oligonucleotide to generate the slowly migrating products characteristic of strand transfer. The modified oligonucleotide-based integration assay thus detects strand transfer products but not the products of 3′ processing.

We first used 3′ radiolabeled MLV U5-30 as the target oligonucleotide, because we had shown that it undergoes negligible 3′ processing when mixed with HIV-1 IN (Fig. 3A, lanes 9 and 10). Neither the 3′-radiolabeled target alone nor the target plus unlabeled MLV U5-30 yielded significant amounts of strand transfer products in the presence of IN (Fig. 8A, lanes 1 to 3). However, when unlabeled HIV U5-29 or unlabeled HIV U3-29 was mixed with 3′-radiolabeled MLV U5-30 DNA and IN, strand transfer products were readily seen (Fig. 8A, lanes 4 and 5). Identical strand transfer patterns were observed with the HIV-1 U3 and U5 *att* site oligonucleotides; recall that, in contrast, different patterns were seen with the two HIV-1 *att* sites in the standard assay (Fig. 3, lanes 1 to 4). Therefore, target site selection appears to be independent of the oligonucleotide that undergoes 3′ processing.

Although the use of MLV U5-30 as a radiolabeled target produced identical strand transfer patterns for unlabeled HIV-1 U5 and U3 *att* site oligonucleotide substrates, it did not produce strand transfer patterns consistent with random use of potential target sites. We wondered if this lack of randomness could be due to unsuspected interactions between HIV-1 IN and MLV U5-30 that might not be expected with a target unrelated to known retroviral sequences. We

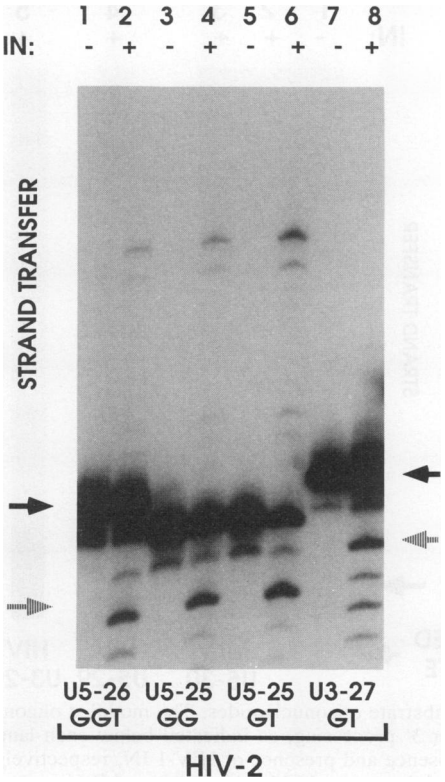

FIG. 6. HIV-2 *att* site oligonucleotides are correctly processed and recombined by HIV-1 IN. Assays were performed as described in Materials and Methods. Oligonucleotides are identified below each pair of lanes (Fig. 1). Each oligonucleotide probe was incubated either without (−) or with (+) purified HIV-1 IN. Solid arrows represent the starting oligonucleotide probe, and hatched arrows represent the products of 3′ processing (removal of three nucleotides in lane 2; removal of two nucleotides in lanes 4, 6, and 8). Strand transfer products are indicated.

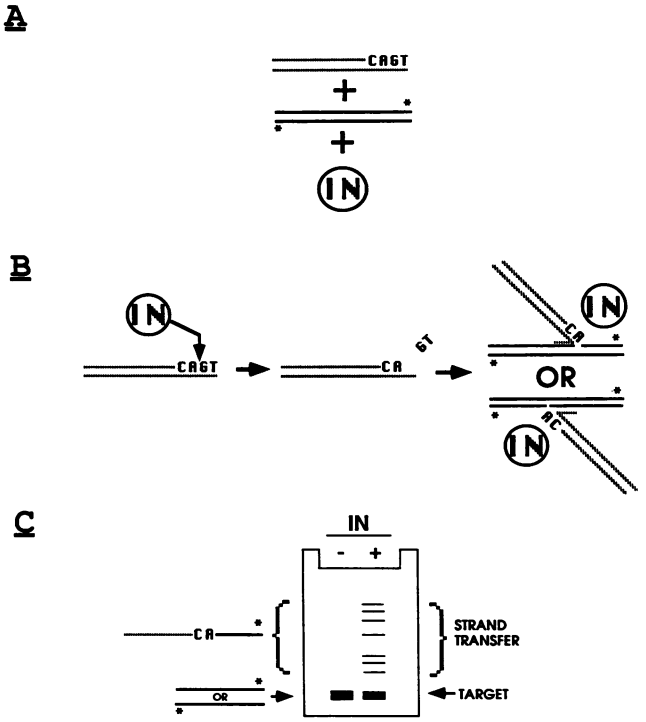

FIG. 7. Outline of an in vitro integration assay which distinguishes the effects of target and substrate oligonucleotides. (A) HIV-1 IN is incubated with two different oligonucleotides; one is a substrate for 3′ processing, and the other is a target for detectable strand transfer events. The substrate for 3′ processing is not radiolabeled and is depicted with broken lines; the target for insertion of 3′ processed DNA is radiolabeled at its 3′ ends and depicted with solid lines. *, the 3′ radiolabel on the target oligonucleotides. (B) The product of 3′ processing is unlabeled and therefore not detectable by autoradiography. Some of the 3′ processed unlabeled *att* site oligonucleotides will be joined to either unlabeled *att* site oligonucleotides (dashed lines) or oligonucleotides radiolabeled at their 3′ ends (solid lines). OR indicates that strand transfer of the *att* site oligonucleotide can occur into either of the target strands. (C) After heating and electrophoresis on a denaturing 20% acrylamide gel, the products of strand transfer into the 3′ radiolabeled target oligonucleotide are detected by autoradiography.

therefore used the R-22 oligonucleotide, which is based on the sequence from a c-*src* intron, as a radiolabeled target. When HIV U5-29 or HIV U3-29 was mixed with the 3′-radiolabeled R-22 in the presence of IN, strand transfer products were readily formed (Fig. 8B). Again, in contrast to the standard assay results (Fig. 3A), the strand transfer patterns were identical when unlabeled HIV U5-29 or HIV U3-29 was used. However, R-22 is also used nonrandomly as a target, suggesting that HIV-1 IN prefers certain target sequences for integration in vitro.

DISCUSSION

Retroviral integration, an essential step in the virus life cycle, involves a series of DNA cutting and joining events. To better understand the role of HIV-1 IN in these events, we expressed the protein in *S. cerevisiae* and developed a procedure to obtain highly purified IN for in vitro analysis. Using a now standard oligonucleotide-based in vitro assay for retroviral integration (8, 15), we have characterized the functional activities of the purified IN and assessed the sequence and structural requirements for oligonucleotide substrates. In addition, by modifying the assay (Fig. 7), we have shown that nonrandom use of target sites by yeast-derived HIV-1 IN in vitro is a function of target and not *att* site DNA sequence.

The HIV-1 IN isolated from *S. cerevisiae* has the native amino acid sequence of HIV-1 IN, and we can recover 1.5 to 2 mg of highly purified (>95%) protein per liter of cell culture (Fig. 2). The yeast-derived IN performs the 3′ processing and strand transfer reactions involved in retroviral integration, showing a strong preference for cognate *att* site sequences (Fig. 3). Our results corroborate the findings of others, using MLV (8), avian leukosis virus (15), and HIV-1 (5, 17, 29) IN proteins, that IN alone is sufficient for these two important steps in the integration of retroviral DNA.

To address the requirement for a functional *att* site for HIV-1 IN, a series of *att* site oligonucleotides were assayed for their effects on 3′ processing and strand transfer. Using blunt-ended and preprocessed U5 and U3 *att* site oligonucleotides, we showed that alteration of the conserved 3′ CA to TG has an independent and marked impairment of both 3′ processing and strand transfer (Fig. 3B). These findings are in agreement with those of LaFemina et al. (17), obtained with HIV-1 U5 *att* site oligonucleotides. Others have also

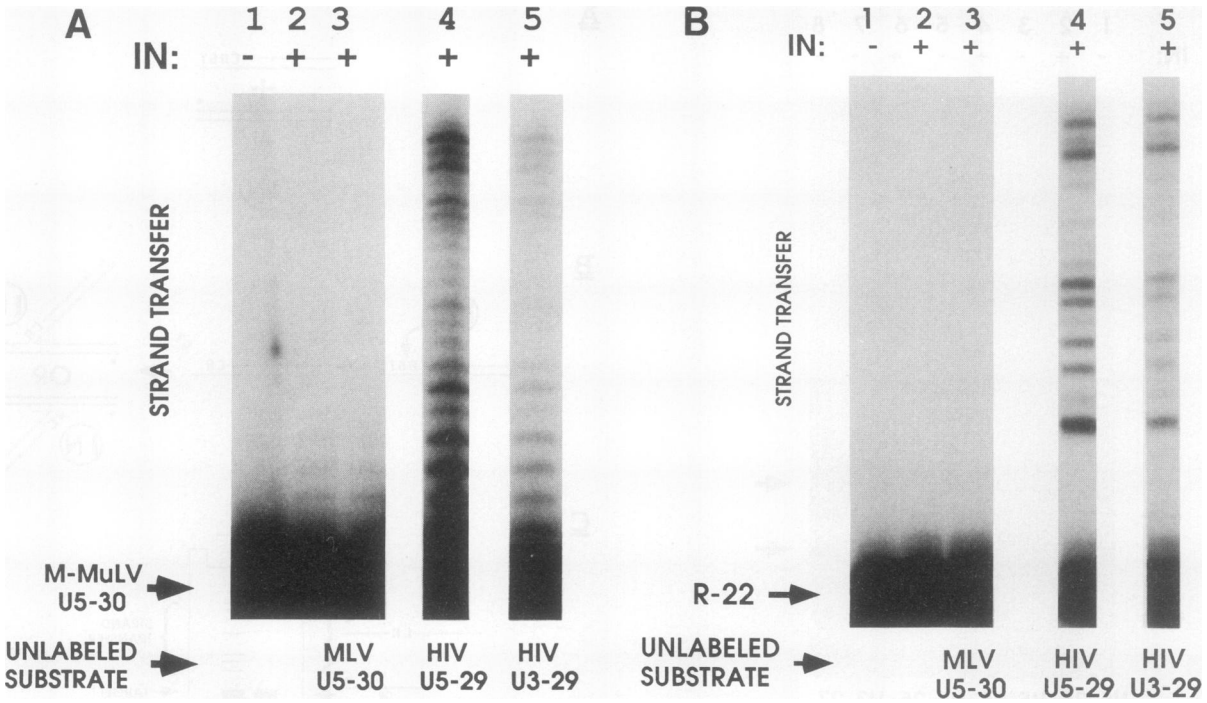

FIG. 8. Target site selection is dependent on the target and independent of the substrate oligonucleotides. The modified oligonucleotide assay shown in Fig. 7 was performed with unlabeled substrate oligonucleotides for 3′ processing, as indicated below each lane. The 3′ radiolabeled target DNA was MLV U5-30 (A) or R-22 (B). − and + indicate the absence and presence of HIV-1 IN, respectively. Strand transfer products are indicated. See Fig. 7 and its legend for a detailed description of the assay. M-MuLV, Moloney MLV.

demonstrated the importance of the U3 (5) and U5 (29) CA on 3′ processing, but preprocessed mutants were not tested, precluding conclusions about independent effects on strand transfer. In addition, two groups have shown that both the C residue and the A residue must be altered to achieve nearly complete ablation of 3′ processing (17, 29).

While the 3′ CA is necessary for both efficient 3′ processing and strand transfer, it is not sufficient. For example, the MLV U5 *att* site oligonucleotide is a poor *att* site substrate for HIV-1 IN (Fig. 3A), yet it possesses an appropriately placed CA dinucleotide. Furthermore, nucleotides internal to the conserved CA are important for IN function: alteration of the 6 bp immediately adjacent to the conserved CA (U5-21 sub5-10) or those located 7 to 13 bp from the CA (U5-21 sub11-17) impaired 3′ processing and associated strand transfer (Fig. 3C, lanes 5 to 8). Notably, changes located 7 to 13 bp from the conserved CA resulted in a greater reduction of IN activity than did changes located 1 to 6 bp from the conserved CA. Similar findings have been recently reported by Vink et al. (29); however, Bushman and Craigie (5) did not see significant impairment of IN activity with a grouped alteration of base pairs located 6 to 11 residues from the U3 CA. These differences may reflect the choice of new sequences as well as the loss of wild-type sequences. In general, mutations that change only one to three residues are unlikely to impair IN-mediated functions significantly unless they include the 3′ CA (17, 29). Thus, while no single base pair outside the HIV-1 3′ CA has been shown to have a strong inhibitory effect on 3′ processing or strand transfer, nucleotides up to 13 positions from the CA may influence IN activity in vitro if enough residues are altered. Definition of the terminal sequence requirements for integration of HIV-1 DNA in vivo has not yet been achieved.

We also investigated the role of the two terminal base pairs which are removed during provirus formation (Fig. 4). These 2 bp appear dispensable for either IN-mediated reaction. The terminal 2 bp can be removed completely (Fig. 4, lanes 5 and 6), or unpaired nucleotides can be present as 3′ or 5′ extensions (Fig. 4, lanes 2 and 3 and lanes 7 and 8, respectively), yet 3′ processing (U5-29/27) and strand transfer (U5-29/27, U5-27/29, and U5-27/27) occur as with wild-type (U5-29/29) *att* site oligonucleotides. We have confirmed that the strand transfer reaction seen with oligonucleotides that end bluntly at the conserved CA is dependent on the CA dinucleotide, as would be expected of an IN-mediated reaction (Fig. 4, lanes 9 and 10). In contrast, an MLV mutant which lacks the 2 bp beyond the U5 CA is unable to establish a provirus in vivo, despite synthesis of normal amounts of viral DNA (6). This apparent conflict between our in vitro and published in vivo results could represent either an unexpected difference between HIV-1 and MLV IN proteins or a difference between the demands of the in vivo and in vitro integration reactions. This question could be resolved by appropriate comparisons of MLV and HIV *att* site mutants in vivo and in vitro, tests that might challenge the verisimilitude of the in vitro assay.

After reverse transcription of the incoming retroviral RNA, linear viral DNA enters the nucleus, where some DNA molecules remain linear and others become one- or two-LTR circles (for a review, see reference 27). Recent studies with MLV strongly argue that linear DNA is the precursor to proviral DNA (3, 4, 14, 18). In agreement with these findings, linear (Fig. 3A) but not two-LTR circle junction (Fig. 5) forms of HIV-1 *att* site oligonucleotides serve as substrates for 3′ processing and strand transfer for HIV-1 in vitro. Others have shown a progressive decline of

3′ processing by HIV-1 IN as the number of base pairs 3′ to the conserved CA is increased from 2 to 5 (29). Taken together, these results suggest that correct 3′ processing by HIV-1 IN occurs only when the *att* sequence is very near the end of a DNA molecule. However, the location of the 3′ processing site for HIV-1 IN is not a fixed distance from the end, but is determined by the position of the CA dinucleotide (Fig. 6) (5, 29). In contrast, avian sarcoma virus IN preferentially nicks avian sarcoma virus DNA 3′ to the conserved CA dinucleotide in circle junction oligonucleotides (16) or in larger fragments containing a circle junction (14a). However, the activity observed with circle junction oligonucleotides was less than that seen with terminal *att* sites (16). These findings may be due to intrinsic differences between HIV-1 and avian sarcoma virus IN proteins or differences in reaction conditions.

The strand transfer patterns seen with our yeast-derived HIV-1 IN in the now standard oligonucleotide assay demonstrate nonrandom use of target sites for integration (Fig. 3A), an observation also made by others (5, 8, 15). In addition, we noticed that the usage patterns of U3 and U5 *att* site oligonucleotides differ (Fig. 3A). To distinguish the influence of substrate and target oligonucleotides on target site usage, we altered the assay so that target and substrate oligonucleotides can be independently assessed. In the modified assay, a 3′-radiolabeled oligonucleotide serves as a target for another unlabeled oligonucleotide. Only the unlabeled oligonucleotide has an HIV-1 *att* site and can serve as a donor in a strand transfer reaction, while both oligonucleotides can serve as targets for integration. However, only integration into the radiolabeled target can be detected. When a common oligonucleotide was used as a target, HIV-1 U5 and U3 *att* site oligonucleotide substrates produced identical strand transfer patterns (Fig. 8). This contrasts with the different strand transfer patterns seen when HIV-1 U5 or U3 *att* site oligonucleotides served as both substrate and target in the standard in vitro integration reaction (Fig. 3A, lanes 1 to 4). However, regardless of the target used, target site usage is nonrandom, and the patterns are different with *src*-derived or MLV-derived oligonucleotide targets. These results suggest that the strand transfer pattern is a function of target sequence and not a function of the donor DNA sequence undergoing 3′ processing.

The influence of the nucleotide sequence on the selection of the integration site appears not to be confined to the oligonucleotide-based integration assay described here. We have also observed that target nucleotide sequences influence the location of integration sites when minichromosomes or naked circular plasmids are used as targets in in vitro integration reactions, with purified IN or viral nucleoprotein complexes as a source of integration activity targets (21a). The significance of such site selectivity during integration in vivo is under investigation.

ACKNOWLEDGMENTS

We thank Phil Barr for assistance with protein expression, Paul Feucht and Ian Bathurst for assistance with the Dyno-Mill, Kathelyn Steimer for HIV-1 IN antisera, and Steve Hughes for donating HIV-2 *att* site oligonucleotides. We also thank Paul Bates, Pat Brown, Peter Pryciak, Lily Shiue, and John Young for critical reading of the manuscript.

A.D.L. is supported by a Burroughs Wellcome Research Foundation Award. H.E.V. is an American Cancer Society Research Professor. This work was supported by grants from the National Institutes of Health.

REFERENCES

1. **Barr, P. J., E. A. Sabin, C. T. L. Ng, K. E. Landsberg, K. S. Steimer, I. C. Bathurst, and J. R. Shuster.** 1991. Ubiquitin fusion approach to heterologous gene expression in yeast: high level production of amino-terminal authentic proteins. *In* R. T. Hatch, C. Goochee, T. Moreira, and Y. Alroy (ed.), Expression systems and processes for recombinant DNA products. American Chemical Society, Washington, D.C.
2. **Bowerman, B., P. O. Brown, J. M. Bishop, and H. E. Varmus.** 1989. A nucleoprotein complex mediates the integration of retroviral DNA. Genes Dev. **3**:469–478.
3. **Brown, P. O., B. Bowerman, H. E. Varmus, and J. M. Bishop.** 1987. Correct integration of retroviral DNA *in vitro*. Cell **49**:347–356.
4. **Brown, P. O., B. Bowerman, H. E. Varmus, and J. M. Bishop.** 1989. Retroviral integration: structure of the initial covalent product and its precursor, and a role for the viral IN protein. Proc. Natl. Acad. Sci. USA **86**:2525–2529.
5. **Bushman, F. D., and R. Craigie.** 1991. Activities of human immunodeficiency virus (HIV) integration protein *in vitro*: specific cleavage and integration of HIV DNA. Proc. Natl. Acad. Sci. USA **88**:1339–1343.
6. **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.
7. **Colicelli, J., and S. P. Goff.** 1988. Sequence and spacing requirements of a retrovirus integration site. J. Mol. Biol. **199**:47–59.
8. **Craigie, R., T. Fujiwara, and R. 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.
9. **Donehower, L. A.** 1988. Analysis of mutant murine leukemia viruses containing linker insertion mutations in the 3′ region of *pol*. J. Virol. **62**:3958–3964.
10. **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.
11. **Ellison, V., H. Abrams, T. Roe, J. Lifson, and P. Brown.** 1990. Human immunodeficiency virus integration in a cell-free system. J. Virol. **64**:2711–2715.
12. **Farnet, C. M., and W. A. Haseltine.** 1990. Integration of human immunodeficiency virus type 1 DNA *in vitro*. Proc. Natl. Acad. Sci. USA **87**:4164–4168.
13. **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.
14. **Fujiwara, T., and K. Mizuuchi.** 1988. Retroviral DNA integration: structure of an integration intermediate. Cell **54**:497–504.

14a. **Grandgenett, D. P., and A. C. Vora.** 1985. Site specific nicking at the avian retrovirus LTR circle junction by the viral pp32 DNA endonuclease. Nucleic Acids Res. **13**:6205–6221.

15. **Katz, R. A., G. Merkel, J. Kulkosky, J. Leis, and A. M. Skalka.** 1990. The avian retroviral IN protein is both necessary and sufficient for integrative recombination *in vitro*. Cell **63**:87–95.
16. **Katzman, M., R. A. Katz, A. M. Skalka, and J. Leis.** 1989. The avian retroviral integration protein cleaves the terminal sequences of linear viral DNA at the in vivo sites of integration. J. Virol. **63**:5319–5327.
17. **LaFemina, R. L., P. A. Callahan, and M. G. Cordingley.** 1991. Substrate specificity of recombinant human immunodeficiency virus integrase protein. J. Virol. **65**:5624–5630.
18. **Lobel, L. I., J. E. Murphy, and S. P. Goff.** 1989. The palindromic LTR-LTR junction of Moloney murine leukemia virus is not an efficient substrate for proviral integration. J. Virol. **63**:2629–2637.
19. **Myers, G., S. F. Josephs, and J. A. Berzofsky (ed.).** 1989. Human retroviruses and AIDS 1989: a compilation and analysis of nucleic acid and amino acid sequences. Los Alamos National Laboratory, Los Alamos, N.Mex.
20. **Panganiban, A. T., and H. M. Temin.** 1983. The terminal nucleotides of retrovirus DNA are required for integration but not virus production. Nature (London) **306**:155–160.

21. **Panganiban, A. T., and H. M. Temin.** 1984. The retrovirus *pol* gene encodes a product required for DNA integration: identification of a retrovirus *int* locus. Proc. Natl. Acad. Sci. USA **81**:7885–7889.

21a. **Pryciak, P. M., and A. D. Leavitt.** Unpublished data.

22. **Quinn, T. P., and D. P. Grandgenett.** 1988. Genetic evidence that the avian retrovirus DNA endonuclease domain of *pol* is necessary for viral integration. J. Virol. **62**:2307–2312.

23. **Roth, M. J., 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.

24. **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.

25. **Sherman, P. A., and J. A. Fyfe.** 1990. Human immunodeficiency virus integration protein expressed in *Escherichia coli* possesses selective DNA cleaving activity. Proc. Natl. Acad. Sci. USA **87**:5119–5123.

26. **Steimer, K. S., K. W. Higgins, M. A. Powers, J. C. Stephans, A. Gyenes, C. George-Nascimento, P. A. Luciw, P. J. Barr, R. A. Hallewell, and R. Sanchez-Pescador.** 1986. Recombinant polypeptide from the endonuclease region of the acquired immune deficiency syndrome retrovirus polymerase (*pol*) gene detects serum antibodies in most infected individuals. J. Virol. **58**:9–16.

27. **Varmus, H. E., and P. O. Brown.** 1989. Retroviruses, p. 53–108. *In* D. E. Berg and M. M. Howe (ed.), Mobile DNA. American Society for Microbiology, Washington, D.C.

28. **Vink, C., M. Groenink, Y. Elgersma, R. A. M. Fouchier, M. Tersmette, and R. H. A. Plasterk.** 1990. Analysis of the junctions between human immunodeficiency virus type 1 proviral DNA and human DNA. J. Virol. **64**:5626–5627.

29. **Vink, C., D. C. van Gent, Y. Elgersma, and R. H. A. 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.

30. **Whitcomb, J. M., and S. H. Hughes.** 1991. The sequence of human immunodeficiency virus type 2 circle junction suggests that integration protein cleaves the ends of linear DNA asymmetrically. J. Virol. **65**:3906–3910.

31. **Whitcomb, J. M., R. Kumar, and S. H. Hughes.** 1990. Sequence of the circle junction of human immunodeficiency virus type 1: implications for reverse transcription and integration. J. Virol. **64**:4903–4906.