JOURNAL OF VIROLOGY, Feb. 1997, p. 1025–1035 Vol. 71, No. 2
0022-538X/97/$04.00+0

Functional Identification of Nucleotides Conferring Substrate Specificity to Retroviral Integrase Reactions

MINI BALAKRISHNAN[1] AND COLLEEN B. JONSSON[1,2]*

Graduate Program in Molecular Biology[1] and Department of Chemistry and Biochemistry,[2] New Mexico State University, Las Cruces, New Mexico 88003

Received 17 July 1996/Accepted 8 November 1996

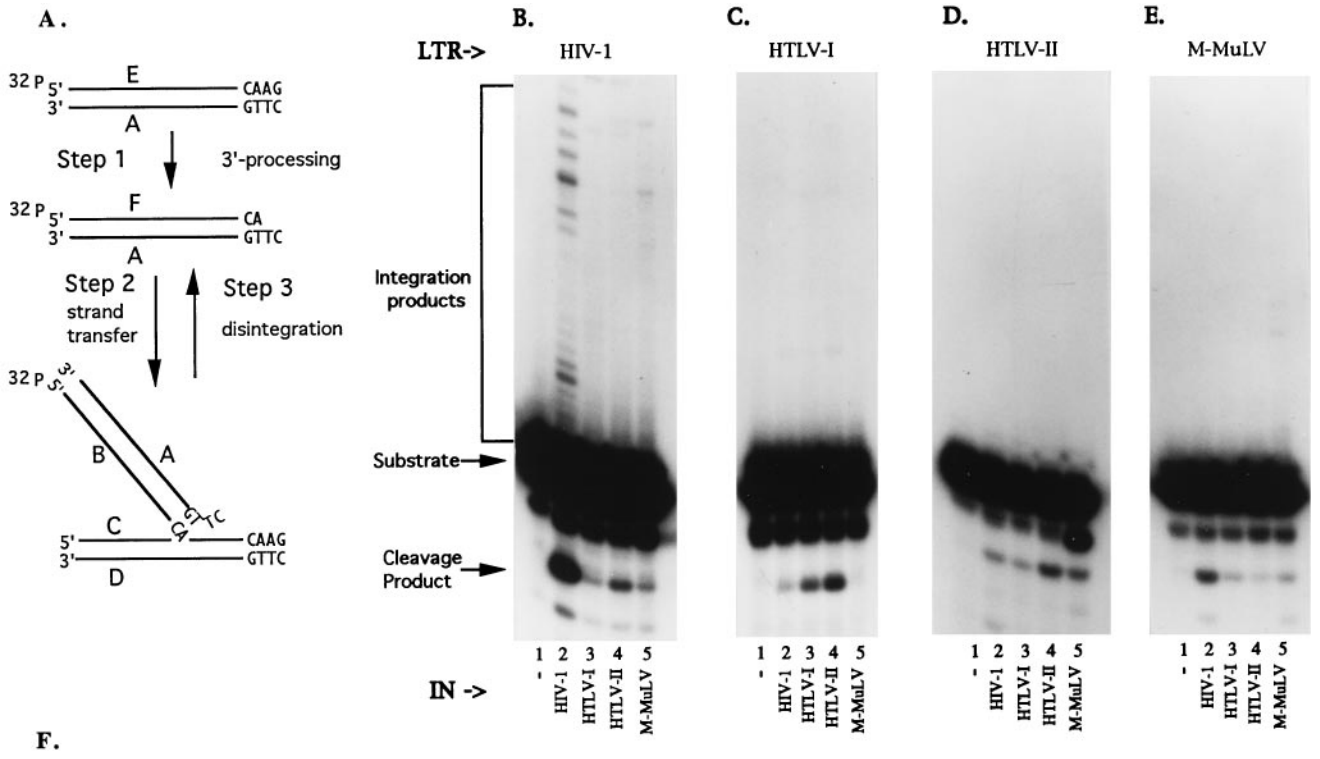
The long terminal repeats (LTRs) that flank the retroviral DNA genome play a distinct role in the integration process by acting as specific substrates for the integrase (IN). The role of LTR sequences in providing substrate recognition and specificity to integration reactions was investigated for INs from human immunodeficiency virus type 1 (HIV-1), Moloney murine leukemia virus (M-MuLV), human T-cell leukemia virus type 1 (HTLV-1), and human T-cell leukemia virus type 2 (HTLV-2). Overall, these INs required specific LTR sequences for optimal catalysis of 3′-processing reactions, as opposed to strand transfer and disintegration reactions. It is of particular note that in strand transfer reactions the sites of integration were similar among the four INs. In the 3′-processing reaction, sequence specificity for each IN was traced to the three nucleotides proximal to the conserved CA. Reactions catalyzed by M-MuLV IN were additionally influenced by upstream regions. The nucleotide requirements for optimal catalysis differed for each IN. HIV-1 IN showed a broad range of substrate specificities, while HTLV-1 IN and HTLV-2 IN had more defined sequence requirements. M-MuLV IN exhibited greater activity with the heterologous LTR substrates than with its own wild-type substrate. This finding was further substantiated by the high levels of activity catalyzed by the IN on modified M-MuLV LTRs. This work suggests that unlike the other INs examined, M-MuLV IN has evolved with an IN-LTR interaction that is suboptimal.

Retroviral replication requires the integration of a double-stranded DNA copy of the viral single-stranded RNA genome into a chromosome of the host cell (3, 17, 44). The viral DNA genome contains two long terminal repeat (LTR) sequences that flank the DNA. The virus-encoded protein integrase (IN) catalyzes the integration process (9, 16) for which the LTRs serve as specific substrates. Integration proceeds through two enzymatic steps (Fig. 1A). First, IN removes nucleotides at the 3′ ends of the LTR termini to generate recessed ends with the adenine of the conserved CA exposed (9, 16, 18, 34, 37, 49). This endonucleolytic cleavage, termed 3′ processing, occurs in the cytoplasm (34) and is catalyzed by IN through a single nucleophilic attack (11, 46). The 3′-processed DNA intermediate and IN subsequently move to the host chromatin in the nucleus. In the second step of integration, strand transfer, IN joins the two recessed viral DNA ends to the host DNA through a one-step transesterification reaction (11, 24). This concerted integration into the target DNA introduces a gap at the site of insertion on each target strand. Repair of these gaps results in duplication of the target sequences flanking the proviral DNA. The gap repair is thought to be directed by host enzymes, although the exact mechanism has not been resolved. Although integration is considered to be random, preferences for certain sites or regions in the target DNA have been observed (13, 20, 23, 29, 30, 33, 38, 45). In vitro, IN also catalyzes the reverse of the integration process, disintegration, on a substrate that mimics the integration intermediate (7).

Among retroviruses, the LTRs bear little homology to one another, except for the presence of the completely conserved dinucleotide CA near the 3′ end of the LTR termini (43). Mutagenesis of the CA sequence blocks cleavage and strand transfer; however, replacement of only the C or the A allows some level of activity (18, 22, 23, 36, 46). In addition to the conserved CA, the importance of the terminal LTR sequences has been shown for several INs. The requirement for specific LTR sequences for integration has been defined essentially by gross deletion and site-directed mutagenesis for avian sarcoma-leukosis virus (17), human immunodeficiency virus type 1 (HIV-1), (4, 22, 23, 36, 46, 50), and Moloney murine leukemia virus (M-MuLV) (5, 8, 25, 32, 34). The terminal 6 to 9 bp of the viral LTR are sufficient for integration activity, suggesting that the major determinants of LTR recognition and catalysis lie in this region. IN, through cooperative interaction, binds nonspecifically to the LTR as a high-order oligomer at a density of 10 monomers per 21-bp LTR (28). Sequence-specific recognition is not responsible for driving the binding of the LTR and IN (19, 22, 28, 35, 42, 48). Therefore, sequence specificity of the reaction may occur during catalysis, by alignment of the IN-substrate complex in a geometry favorable for precise cleavage and strand transfer. The terminal sequences thus play an important role in promoting catalysis of integration by presenting the LTR as the precise substrate. The nucleotide determinants in the LTR that confer substrate recognition and specificity on the reaction have not been well defined.

Based on the assumption that retroviral INs recognize and bind their LTRs by similar mechanisms that involve cognate recognition domains, a comparative approach was taken to identify nucleotides that may be involved in providing specificity to reactions catalyzed by HIV-1, human T-cell leukemia virus type 1 (HTLV-1), HTLV-2, and M-MuLV INs. A more pronounced requirement for specific nucleotide sequences was observed in 3′-processing reactions than in strand transfer and disintegration reactions. Therefore, further efforts were directed to defining nucleotides responsible for conferring substrate specificity during the 3′-processing reaction. Based on the levels of 3′ processing exhibited by the four INs on heterologous wild-type (WT) LTR sequences, substrates were de-

* Corresponding author. Mailing address: Department of Chemistry and Biochemistry, New Mexico State University, Dept. 3C, Box 30001, Las Cruces, NM 88003. Phone: (505) 646-3346. Fax: (505) 646-2649. E-mail: cjonsson@nmsu.edu.



F.

HIV-1 WT	5'ATGTGGAAAATCTCTAGCAGT
	3'TACTCCTTTTAGAGATCGTCA
HTLV-I WT	5'AGAGAAATTTAGTACACAGT
	3'TCTCTTTAAATCATGTGTCA
HTLV-II WT	5'TCCCGGGGAAGACAAACAAG
	3'AGGGCCCCTTCTGTTTGTTC
M-MuLV WT	5'GTCAGCGGGGGTCTTTCATT
	3'CAGTCGCCCCCAGAAAGTAA
Nucleotide No.	7 5 1

FIG. 1. (A) Schematic representation of the enzymatic activities catalyzed by the retroviral IN in vitro. Step 1, 3′ processing; 2, strand transfer; 3, disintegration. 3′-Processing substrates were prepared by hybridizing the ^{32}P-labeled E strand (20 nucleotides) with the complementary A strand (19) to generate a 20-bp blunt DNA fragment. 3′-Processing activity is readily observed by a reduction in size of the labeled E strand (20 nucleotides) by 2 nucleotides, thereby generating an 18-nucleotide recessed product (21-nucleotide substrate and 19-nucleotide product in the case of HIV-1). Strand transfer substrates are prepared by hybridizing the ^{32}P-labeled F strand (18 nucleotides) with the A strand. The substrate is identical to the 3′-processing substrate, except for the absence of two terminal nucleotides proximal to the CA dinucleotide. Strand transfer activity generates products both larger and smaller than the substrate, since integration occurs at random sites along the phosphate backbone of the target substrate DNA. The disintegration substrate represents a hypothetical strand transfer intermediate. The substrate is prepared by hybridizing the ^{32}P-labeled C strand (16-nucleotide) with the A (20-nucleotide), B (32-nucleotide), and D (30-nucleotide) strands. Disintegration reaction results in joining of the 3′-OH end of the C strand to the B strand, resulting in formation of the 30-nucleotide product. (B to E) 3′-Processing activities of the INs on WT native and heterologous LTR substrates. WT U5 LTR substrates from HIV-1 (B), HTLV-1 (C), HTLV-2 (D), and M-MuLV (E) were examined for 3′-processing activity with HIV-1 IN (lanes 2), HTLV-1 IN (lanes 3), HTLV-2 IN (lanes 4), M-MuLV IN and (lanes 5). Lane 1 in each panel represents the substrate with no protein. (F) Nucleotide sequences of the WT LTR substrates used in the assays.

signed to obtain a gain of function (increase in 3′-processing activity) for a particular IN. Results showed specific nucleotide requirements in the LTR for recognition leading to catalysis by HTLV-1 and HTLV-2 INs, while the HIV-1 IN had a comparatively relaxed LTR sequence requirement. For the M-MuLV IN, the M-MuLV WT LTR was a suboptimal substrate, as specific substitutions generated a much preferred LTR substrate for 3′ processing.

MATERIALS AND METHODS

Materials. Oligonucleotides were purchased from Ana-gen Technologies (Fabian Way, Calif.). Crude [γ-^{32}P]ATP was bought from Dupont-NEN (Boston, Mass.). Enzymes were purchased from Gibco Bethesda Research Laboratories (Gaithersburg, Md.) and New England Biolabs (Beverly, Mass.). DNA grade Sephadex G-50 was obtained from Pharmacia (Piscataway, N.J.). All chemicals were purchased from Sigma Chemical Co. (St. Louis, Mo.).

Cloning of HTLV-1 IN. The full-length HTLV-1$_{CH}$ genome was a gift from Lee Ratner (Washington University, St. Louis, Mo.) (31). The IN coding region from HTLV-1$_{CH}$ was PCR amplified by using Vent DNA polymerase and primers P17438-0194 (5′ AAGCTTCATATGGTCCTGCAGCTCTCTCCTGCA) and P20283-0294 (5′ CTCGAGCCCATGGTGTTGG), which introduced *Nde*I and *Xho*I restriction sites (underlined) at the 5′ and 3′ ends, respectively. The PCR-amplified HTLV-1 IN gene was cloned into the *Nde*I-*Xho*I sites in pET22B (Novagen, Madison, Wis.). Cloning into the *Xho*I site adds an additional 18 nucleotides, encoding six histidines, to the 3′ end of the open reading frame.

Protein purification. The INs from HIV-1, HTLV-1, HTLV-2, and M-MuLV used in this study were purified as hexahistidine-tagged fusion proteins. The presence of the histidine tag does not interfere with WT HIV-1 IN activity, as shown previously by Bushman et al. (6). Purified HIV-1 IN was a gift from Robert Craigie (Laboratory of Molecular Biology, National Institute of Diabetes and Digestive and Kidney Diseases, Bethesda, Md.). HTLV-1, HTLV-2, and M-MuLV INs were expressed in *Escherichia coli* BL21(DE3) cells and purified by using nickel nitrilotriacetate agarose (Qiagen, Chatsworth, Calif.) as described previously (2, 14). HTLV-1 IN was characterized prior to this study (1). Two separate purifications of HTLV-1 IN were performed which yielded proteins with similar specific activities and nearly 95% homogeneity. HTLV-1 IN exhibited solubility, stability, and reaction characteristics similar to those of the HTLV-2 IN (2). 3′-Processing and strand transfer activities catalyzed by the protein were optimal at 6.6 mM NaCl and inhibited at concentrations of above 100 mM. Disintegration activity was optimal at 6.6 mM NaCl and was inhibited at a concentration of 250 mM NaCl or greater. HTLV-1 IN catalyzed 3′-processing, strand transfer, and disintegration reactions optimally with 7.5 mM Mg^{2+} or Mn^{2+} as the divalent cation.

Oligonucleotide substrates. Oligonucleotide sequences corresponding to the WT U5 LTR regions of the HIV-1 (37), HTLV-1 (31), HTLV-2 (26), and M-MuLV (39) genomes were used as substrates in the assays (Fig. 1F). Oligonucleotides were purified on 20% denaturing polyacrylamide gels, labeled at the 5′ end with [γ-^{32}P]ATP and T4 polynucleotide kinase, and hybridized to complementary oligonucleotide strands as previously described (14). Unincorporated radioactivity was removed from substrate preparations by using G-50 spin columns.

Integration and disintegration assays. Results of titration experiments gave the protein concentration (0.43 μM HIV-1 IN and HTLV-1 IN, 0.33 μM HTLV-2 IN, and 0.93 μM M-MuLV IN) that yielded maximal activity for each IN with its WT blunt and recessed LTR substrate, as well as with disintegration substrates. This protein concentration was used throughout the study. All 3′-processing, strand transfer, and disintegration assays were performed with 15-μl reaction volumes containing the appropriate concentration of IN and 1 pmol of a labeled LTR substrate under the designated buffer conditions. Reactions were incubated at 37°C for 60 min. Integration and disintegration assays for HIV-1, HTLV-1, and HTLV-2 INs contained 25 mM morpholinepropanesulfonic acid (MOPS) (pH 7.2), 10 mM β-mercaptoethanol, 0.75 mM 3-[(3-cholamidopropyl)-dimethyl-ammonio]-1-propanesulfonate (CHAPS), 7.5 mM $MnCl_2$, and 10% glycerol. Strand transfer assays for M-MuLV IN contained 20 mM piperazine-1,4-bis(2-ethanesulfonic acid) (PIPES) (pH 6.2), 10 mM dithiothreitol, 10 mM KCl, 20 mM $MnCl_2$, and 20% glycerol. 3′-Processing reaction conditions for M-MuLV IN were identical to those for strand transfer, except for the inclusion of 10% ethylene glycol. M-MuLV IN disintegration reaction mixtures contained 20 mM PIPES (pH 6.4), 10 mM dithiothreitol, 10 mM CHAPS, 25 mM $MnCl_2$, and 0.05% Nonidet P-40. Reactions were stopped by a 30-min incubation with 25 μg of proteinase K in the presence of 25 mM EDTA and 0.01% sodium dodecyl sulfate at 37°C, followed by the addition of 10 μl of loading dye (95% formamide, 20 mM EDTA, 0.05% bromophenol blue, 0.5% xylene cyanol). Reaction products were separated on 20% polyacrylamide denaturing gels, subjected to autoradiography, and quantitated with a Hoefer GS-300 scanning densitometer and a GS-370 electrophoresis data system. Percent activity was calculated by dividing the sum of the products (3′ processed and strand transfer) by the sum of the substrate and products and multiplying by 100. Percent activity was calculated from a minimum of three trials for each experiment.

RESULTS

Substrate specificity during IN-catalyzed reactions. The terminal 6 to 9 bases of the viral LTRs are sufficient for substrate recognition for 3′ processing and strand transfer by retroviral INs (4, 5, 8, 22, 25, 36, 50). The specific nucleotide determinants within this LTR region that confer sequence specificity during catalysis by IN have not been well defined. It was of interest, therefore, to initially investigate the specificity requirements during catalysis of each of the enzymatic activities of IN. Highly purified HIV-1, HTLV-1, HTLV-2, and M-MuLV INs were thus examined for 3′-processing (Fig. 1), strand transfer (Fig. 2), and disintegration (Fig. 3) activities on their own LTRs and with heterologous LTR substrates.

Substrate specificity for the 3′-processing reaction was investigated by using WT blunt-end U5 LTR substrates. Each IN was examined for the ability to catalyze reactions with its own WT LTR and the WT LTRs of the other three INs (Fig. 1F). Figure 1B shows the 3′-processing activities of all four INs with the HIV-1 WT LTR. Among the four INs, HIV-1 IN was the most efficient in catalyzing 3′ processing on the HIV-1 WT LTR (Fig. 1B, lane 2). The HTLV-1 and M-MuLV INs catalyzed approximately one-fourth of the activity of HIV-1 IN on the HIV-1 WT LTR substrate (Fig. 1B, lanes 3 and 5). The HTLV-2 IN activity level was slightly higher than those of the HTLV-1 and M-MuLV INs (Fig. 1B, lane 4). With the HTLV-1 WT LTR (Fig. 1C), HTLV-2 IN exhibited the highest levels of 3′-processing activity (20%; Fig. 1C, lane 4), while the percentages of product conversion for HIV-1 IN (Fig. 1C, lane 2) and HTLV-1 IN (Fig. 1C, lane 3) were 5 and 10%, respectively. In contrast, the level of 3′ processing was less than 1% for M-MuLV IN (Fig. 1C, lane 5). HTLV-2 IN showed the greatest preference for the HTLV-2 WT LTR substrate (Fig. 1D, lane 4), and the other three INs supported comparatively lower activity levels (Fig. 1D). A comparison of the 3′-processing activities displayed by the four INs with the M-MuLV WT LTR (Fig. 1E) showed HIV-1 IN to be the most efficient (Fig. 1E, lane 2). Low but similar levels of activity (5%) were noted for M-MuLV IN (Fig. 1E lane 5), HTLV-1 IN (Fig. 1E, lane 3), and HTLV-2 IN (Fig. 1E, lane 4) with the M-MuLV WT LTR. Thus, for any LTR substrate, the different INs displayed a range of 3′-processing levels, indicating the influence of LTR sequences on substrate specificity. HIV-1 IN catalyzed the highest level of activity with the HIV-1 LTR and the least with the HTLV-1 LTR. In contrast to previous reports (22, 36, 46), our assays clearly indicated 3′-processing activity of HIV-1 IN on HTLV-1 and M-MuLV LTR substrates. This may reflect differences in assay conditions. HTLV-1 IN efficiently processed its own LTR but exhibited much lower levels with the other three LTRs. HTLV-2 IN recognized the WT LTRs from HIV-1, HTLV-1, and HTLV-2 but not the M-MuLV WT LTR. In contrast to the other INs, M-MuLV IN had the least activity in reactions with its own WT LTR, while efficiently processing the HTLV-2 and HIV-1 LTRs. This suggests that the M-MuLV WT LTR is not an optimal substrate for its IN.

To investigate LTR substrate specificity in the strand transfer reaction, the HIV-1, HTLV-1, HTLV-2, and M-MuLV INs were examined by using recessed WT LTR substrates (Fig. 2E). With the HIV-1 WT LTR, all four INs showed similar levels (26 to 35%) of strand transfer activity (Fig. 2A, compare lanes 2 to 5). This agrees with a previous study showing strand transfer of mini-HIV-1 and mini-HIV-2 substrates by cytoplasmic extracts from M-MuLV-infected cells (47). The HTLV-1 and HTLV-2 INs exhibited high activity levels (55 and 45%, respectively) with the HTLV-1 WT recessed substrate (Fig. 2B, lanes 3 and 4), while HIV-1 IN and M-MuLV IN activity levels were twofold lower (20 to 22%; Fig. 2B, lanes 2 and 5). Nearly equal levels of strand transfer activity were noted with the HIV-1, HTLV-1, and HTLV-2 INs on the recessed HTLV-2 WT LTR (Fig. 2C, lanes 2 to 4). The levels of strand transfer activity were minimal, however, for M-MuLV IN with this LTR (Fig. 2C, lane 5). Finally, analysis of the M-MuLV WT LTR (Fig. 3D) clearly showed that the HIV-1, HTLV-1, and HTLV-2 INs catalyzed higher levels of strand transfer (15 to 20%) than the M-MuLV IN (5%). In contrast to 3′ processing, all four INs showed less variability in the levels of product generated for a particular LTR substrate. Thus, strand transfer activity showed less dependence on specific sequences in the LTR. Interestingly, the panel of four INs produced similar patterns of strand transfer products for a given LTR sequence. These patterns suggest that recognition of the target DNA sequence was similar among these INs when the target DNA sequence was the LTR itself.

The final IN-mediated reaction examined for substrate specificity was disintegration using WT Y substrates (Fig. 3E). Three of the INs, HIV-1 IN, HTLV-1 IN, and M-MuLV IN, catalyzed identical levels of disintegration activity with each of the four disintegration substrates (Fig. 3A, B, and D). HTLV-2 IN generated slightly lower levels of disintegration products with HTLV-2 and M-MuLV Y substrates (Fig. 3C, lanes 6 and 8). In contrast to the marked changes in the levels of 3′-processing activity observed in Fig. 1, there was minimal variation in disintegration activity levels among the INs. These results further support previous findings that IN does not rely on the LTR sequence in the Y substrate for substrate recognition (10, 40, 41). The sequence flexibility of IN for recognition and catalysis of heterologous disintegration substrates indicates structure-based rather than sequence-based recognition of the Y substrates by IN (10). In addition to the flexibility in sequence requirements, reaction conditions also appear to be less stringent for the disintegration reaction (2, 14). Based on our observation that 3′ processing is more sequence depen-

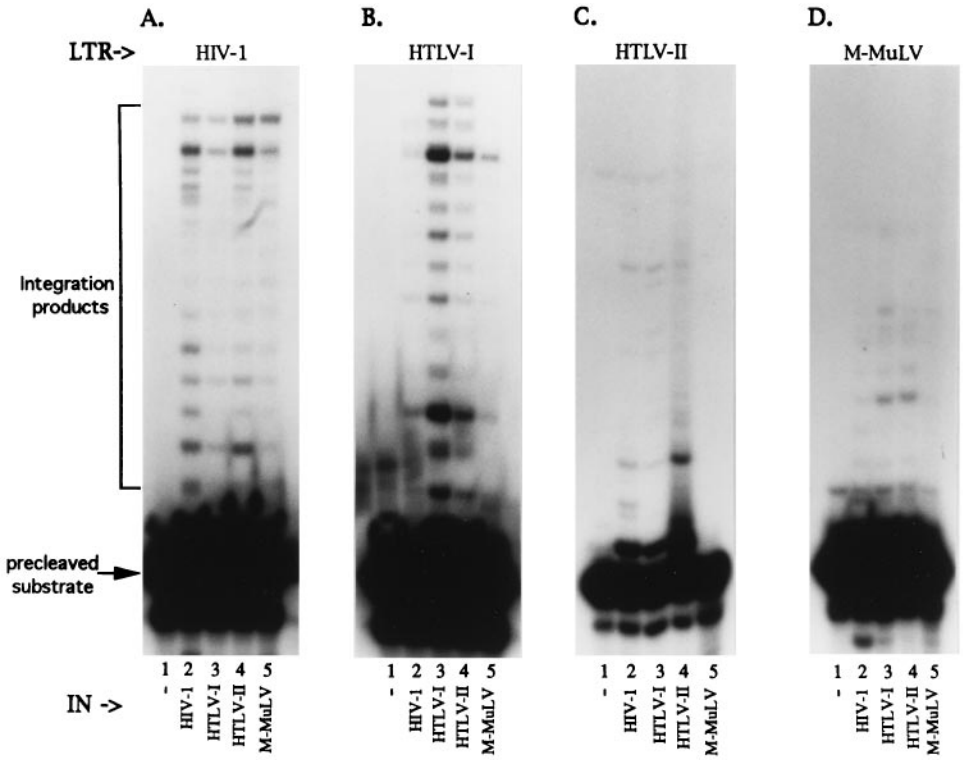


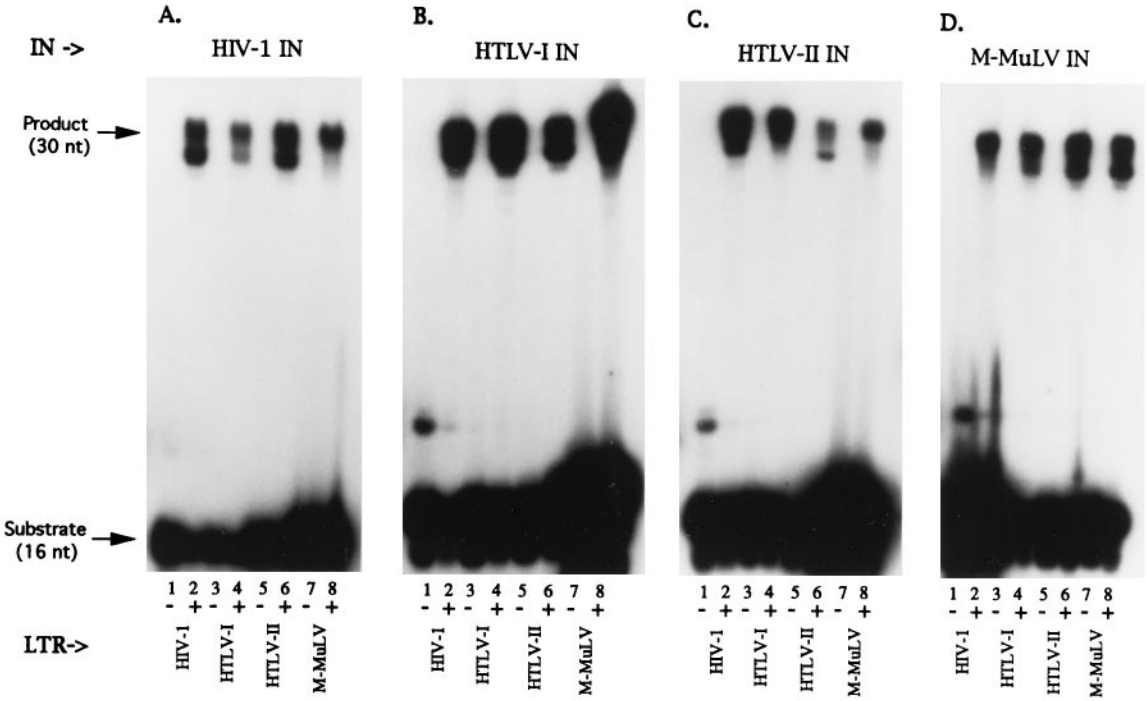
FIG. 2. Strand transfer activities of INs on native and heterologous recessed LTR substrates. HIV-1 (A), HTLV-1 (B), HTLV-2 (C), and M-MuLV (D) LTR substrates were examined for strand transfer activity with HIV-1 IN (lanes 2), HTLV-1 IN (lanes 3), HTLV-2 IN (lanes 4), and M-MuLV IN (lanes 5). Lane 1 in each panel represents the substrate with no protein. (C) Shorter exposure than that of the other panels. (F) Nucleotide sequences of the LTR substrates used in the assays.

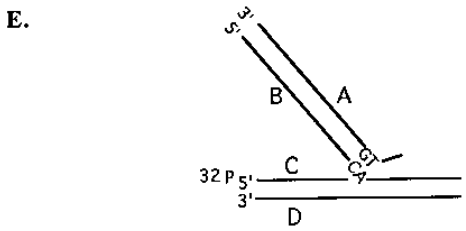
dent than strand transfer and disintegration, we predict that IN-LTR interactions in the preintegration complex involve specific interactions during the 3′-processing step and that these specific sequences may contribute only minimally toward directing strand transfer.

Nucleotide requirements for recognition of the M-MuLV U5 LTR by HIV-1, HTLV-1, HTLV-2, and M-MuLV INs. Based on the greatly reduced levels of 3′ processing of the M-MuLV WT LTR by HTLV-1 and HTLV-2 INs (Fig. 1E), we hypothesized that specific nucleotide substitutions in the M-MuLV LTR could create a substrate more efficiently cleaved by HTLV-1 and HTLV-2 INs. A comparison of HIV-1, HTLV-1, and HTLV-2 WT LTR sequences shows one to three adenines immediately 5′ to the CA, in positions 5, 6, and 7 (Fig. 1F). In the M-MuLV LTR, these positions are occupied by thymines. Since it had been established that the 6 to 9 bases preceding the CA are sufficient to serve as an LTR substrate for IN enzymatic activity (4, 5, 8, 22, 25, 36, 50), we speculated that changes within the first three bases next to the conserved CA in the M-MuLV substrate would generate a substrate that would be processed by the HTLV-1 and HTLV-2 INs but would produce a loss of activity for M-MuLV IN. Therefore, a substitution that conferred activity would be a determinant in substrate specificity. Modified M-MuLV LTRs were therefore designed with adenines substituted for thymines at positions 5 and 7, as well as individually at positions 5, 6, and 7 (Fig. 4F).

Replacement of thymine with adenine at positions 5 and 7 created a substrate (M-MuLV 5,7T→A) with substantially increased levels of 3′ processing by the HIV-1, HTLV-1, and HTLV-2 INs (Fig. 4B, lanes 2 to 4) compared with the M-MuLV WT LTR (Fig. 4A, lanes 2 to 4). Remarkably, an 11-fold increase in the M-MuLV IN activity level was noted with the M-MuLV 5,7T→A LTR (Fig. 4A, lane 5) compared with the M-MuLV WT LTR (Fig. 4B, lane 5). Efficient catalysis of the substrate by M-MuLV IN was also shown by the increased amounts of strand transfer products (Fig. 4B, lane 5). The M-MuLV 5,7T→A LTR substrate was therefore favored over the M-MuLV WT LTR by all four INs.

To examine the specific positional requirement for the adenines, modified M-MuLV LTRs were synthesized with individual adenine substitutions at positions 5 (M-MuLV 5T→A), 6 (M-MuLV 6T→A), and 7 (M-MuLV 7T→A) (Fig. 4F). The M-MuLV 5T→A substitution resulted in 3′-processing levels among the four enzymes similar to that observed for the M-MuLV 5,7T→A LTR substrate (compare Fig. 4B and C). In the case of the M-MuLV 6T→A LTR, the HIV-1, HTLV-1, and HTLV-2 INs maintained the same levels of activity as





HIV-1

A. 5' ACTGCTAGAGATTTTCCTCAT 3'
B. 5' ATGTGGAAAATCTCTAGCAGGCTGCAGGTCGAC 3'
C. 5' CAGCAACGCAAGCTTG 3'
D. 5' GTCGACCTGCAGCCCAAGCTTGCGTTGCTG 3'

HTLV-I

A. 5' ACTGTGTACTAAATTTCTCT 3'
B. 5' AGAGAAATTTAGTACACAGGCTGCAGGTCGAC 3'
C. 5' CAGCAACGCAAGCTTG 3'
D. 5' GTCGACCTGCAGCCCAAGCTTGCGTTGCTG 3'

HTLV-II

A. 5' CTTGTTTGTCTTCCCCGGGA 3'
B. 5' TCCCGGGGAAGACAAACAGGCTGCAGGTCGAC 3'
C. 5' CAGCAACGCAAGCTTG 3'
D. 5' GTCGACCTGCAGCCCAAGCTTGCGTTGCTG 3'

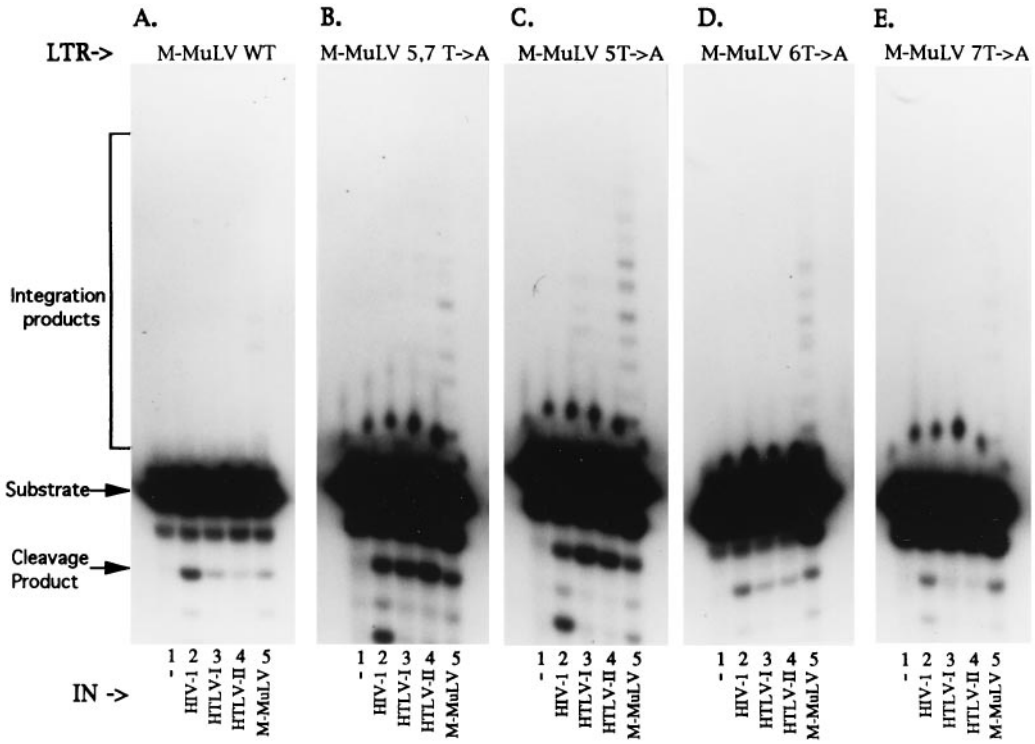
M-MuLV

A. 5' AATGAAAGACCCCGCTGAC 3'
B. 5' TAGTCAGCGGGGTCTTTCAGGCTGCAGGTCGAC3'
C. 5' CAGCAACGCAAGCTTG 3'
D. 5' GTCGACCTGCAGCCCAAGCTTGCGTTGCTG 3'

FIG. 3. Disintegration activities of INs on heterologous Y substrates. The HIV-1 (A), HTLV-1 (B), HTLV-2 (C), and M-MuLV (D) INs were examined for disintegration activity with HIV-1 (lanes 2), HTLV-1 (lanes 4), HTLV-2 (lanes 6), and M-MuLV (lanes 8) Y substrates. Lanes 1, 3, 5, and 7 (indicated by a minus sign) in each panel represent the substrate with no protein. (F) Schematic representation of the Y substrate and nucleotide sequences of the constitutive strands for the four Y substrates. Sequences in the A and B strands that correspond to the U5 LTR arms are highlighted. nt, nucleotides.

observed with the M-MuLV WT LTR (Fig. 4A and D, compare lanes 2 to 4). Strikingly, M-MuLV IN catalyzed eightfold higher activity on this substrate than on its WT substrate (Fig. 4A and D, compare lanes 5). The final thymine-to-adenine substitution at position 7 (M-MuLV 7T→A) resulted in low activity for HIV-1 IN and barely detectable levels of activity for the HTLV-1 and HTLV-2 INs (Fig. 4E, lanes 2 to 4). The M-MuLV IN showed about 20% activity on this substrate, indicating that the substitution was still favored over the WT sequence (Fig. 4E, lane 5). Overall, the single-base thymine-to-adenine substitution at position 5 in the M-MuLV WT LTR clearly generated a more competent substrate for the HTLV-1 and HTLV-2 INs, indicating its importance in 3′ processing for these enzymes. In M-MuLV IN, all of the substitutions, in particular, M-MuLV 5T→A, generated significantly higher levels of 3′ processing than observed for the M-MuLV WT substrate. This suggests that the M-MuLV WT LTR sequence has not evolved as an optimal substrate for M-MuLV IN. In contrast to the wide variation in activity observed for these INs, HIV-1 IN showed little variation in its 3′-processing activity levels.

Nucleotide requirements for recognition of the HTLV-2 U5 LTR. Based on the previous results, we further hypothesized that if the adenine at position 5 is crucial for substrate recognition by the HTLV INs, then replacement of this adenine with a thymine in the HTLV-2 WT LTR would result in loss of 3′-processing activity for the HTLV-1 and HTLV-2 INs, while the HIV-1 and M-MuLV INs would retain activity. Therefore, modified HTLV-2 LTRs were designed with substitutions converse to those introduced in the M-MuLV LTR. Specifically, the adenines at positions 5 and 7 in the WT HTLV-2 LTR were changed to thymines (Fig. 5E). Again, each of the modified substrates was screened for 3′-processing activity with each of the four INs (Fig. 5).



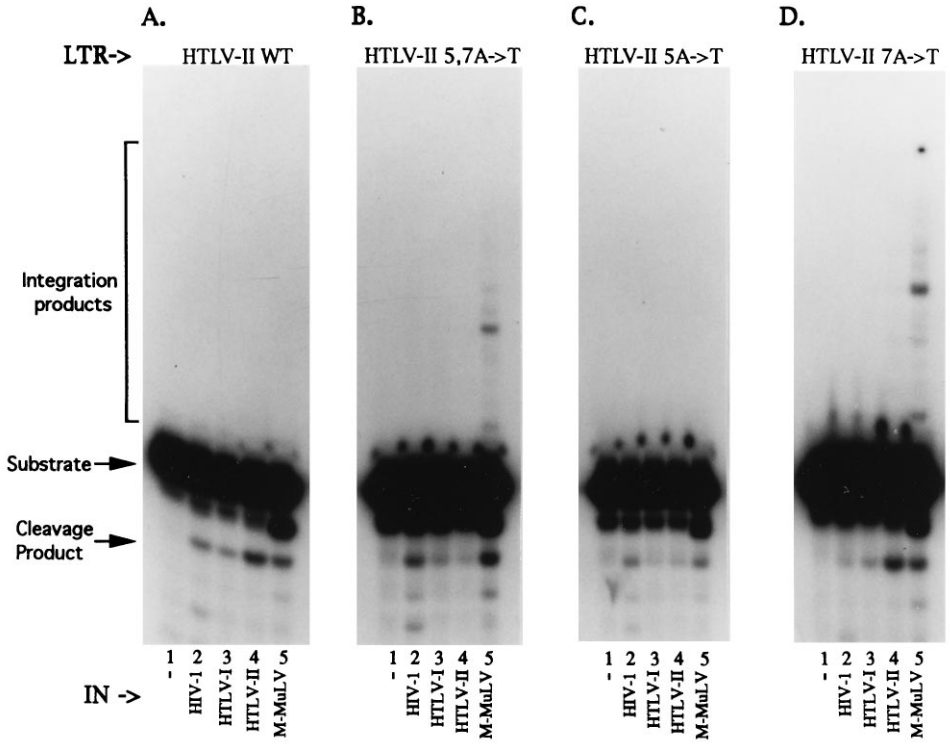
F.

```
M-MuLV WT          5'GTCAGCGGGGGTCTTTCATT
                   3'CAGTCGCCCCCAGAAAGTAA

M-MuLV 5,7T→A      5'GTCAGCGGGGGTCATACATT
                   3'CAGTCGCCCCCAGTATGTAA

M-MuLV 5T→A        5'GTCAGCGGGGGTCTTACATT
                   3'CAGTCGCCCCCAGAATGTAA

M-MuLV 6T→A        5'GTCAGCGGGGGTCTATCATT
                   3'CAGTCGCCCCCAGATAGTAA

M-MuLV 7T→A        5'GTCAGCGGGGGTCATTCATT
                   3'CAGTCGCCCCCAGTAAGTAA
                                 | |   |
Nucleotide No.                   7 5   1
```

FIG. 4. 3′-processing activities of INs on WT and modified M-MuLV LTR substrates. WT M-MuLV (A), M-MuLV 5,7T→A (B), M-MuLV 5T→A (C), M-MuLV 6T→A (D), and M-MuLV 7T→A (E) LTR substrates were used in assays with HIV-1 IN (lanes 2), HTLV-1 IN (lanes 3), HTLV-2 IN (lanes 4), and M-MuLV IN (lanes 5). Lane 1 in each panel represents the substrate with no protein. (F) Nucleotide sequences of the WT and modified M-MuLV U5 LTR substrates used in the assays. Base substitutions are in boldface.

In a comparison of IN activity levels on the HTLV-2 WT LTR (Fig. 5A) and HTLV-2 5,7A→T, HIV-1 IN retained similar amounts of activity with both substrates (Fig. 5A and B, lane 2), while the HTLV-1 and HTLV-2 INs showed a marked decrease in 3′-processing activity levels with HTLV-2 5,7A→T (Fig. 5B, lanes 3 and 4). Catalysis of the HTLV-2 5,7A→T substrate by M-MuLV IN produced higher levels of 3′-processing and strand transfer products (Fig. 5B, lane 5) than observed for the HTLV-2 WT LTR (Fig. 5A, lane 5). The individual substitution at position 5 (HTLV-2 5A→T LTR) dramatically decreased 3′-processing levels of the HTLV-1 and HTLV-2 INs (Fig. 5C, lanes 3 and 4) compared to those observed with the HTLV-2 WT LTR (Fig. 5A, lanes 3 and 4). The 3′-processing levels of the four INs noted with the HTLV-2 5A→T LTR (Fig. 5C) were comparable to those observed for the M-MuLV 7T→A LTR (Fig. 4E). Both of these substrates have an adenine at position 7 and a thymine at position 5 (Fig. 5E and 4F). However, a comparison of the M-MuLV IN strand transfer activities of these two modified LTRs showed more product with the M-MuLV 7T→A LTR (Fig. 4E, lane 5). The final substrate examined in this series, the HTLV-2 7A→T LTR (Fig. 5D), resulted in a loss of activity only for HIV-1 IN (Fig. 5D, lane 2). The HTLV-2 7A→T LTR supported activity levels similar to those of the HTLV-2 WT substrate for HTLV-1 IN and HTLV-2 IN (compare Fig. 5A and D, lanes 3 and 4), while higher activity was observed for the M-MuLV IN (Fig. 5D, lane 5). The HTLV-2 and M-MuLV INs showed high levels of activity with the HTLV-2 7A→T LTR (Fig. 5D, lanes 4 and 5) and the M-MuLV 5T→A LTR (Fig. 4C, lanes 4 and 5), both of which have an adenine at position 5 and a thymine at position 7. These combined results support the role of nucleotides at positions 5 and 7 in conferring sequence-specific recognition of LTR substrates for the HTLV-1, HTLV-2, and M-MuLV INs.

Nucleotide requirements for recognition of the HTLV-1 U5 LTR. Substitutions in the M-MuLV and HTLV-2 LTRs suggested a specific requirement for adenine at position 5 by HTLV-2 IN and, to some extent, by HTLV-1 IN. To further investigate this requirement, base pair substitutions in the HTLV-1 WT LTR were examined. Specifically, adenine-to-thymine substitutions at positions 5 and 7 in the HTLV-1 WT U5 LTR were tested (HTLV-1 5,7A→T, Fig. 6C). Based on earlier results, these modifications were predicted to decrease 3′-processing activity by the HTLV-1 and HTLV-2 INs, as well as confer substrate recognition by M-MuLV IN. Activity levels of the four INs on the HTLV-1 5,7A→T LTR (Fig. 6B) are



E.

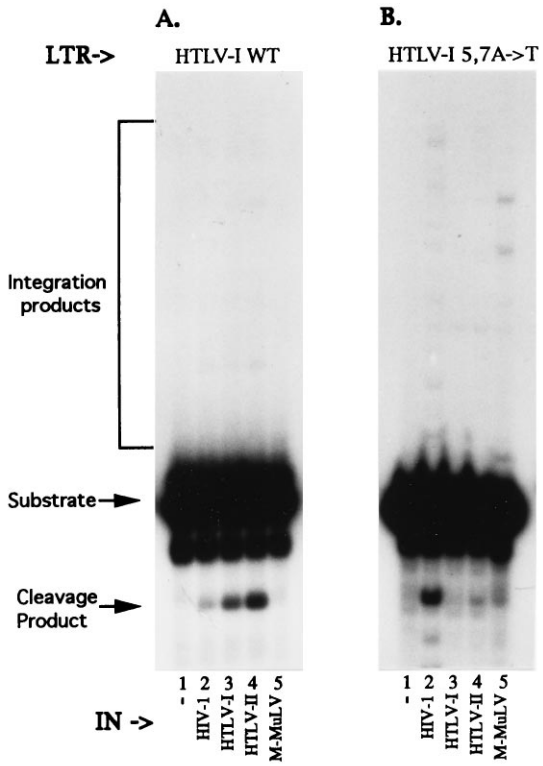
```
HTLV-II WT          5'TCCCGGGGAAGACAAACAAG
                    3'AGGGCCCCTTCTGTTTGTTC

HTLV-II 5,7A→T      5'TCCCGGGGAAGACTATCAAG
                    3'AGGGCCCCTTCTGATAGTTC

HTLV-II 5→T         5'TCCCGGGGAAGACAATCAAG
                    3'AGGGCCCCTTCTGTTAGTTC

HTLV-II 7→T         5'TCCCGGGGAAGACTAACAAG
                    3'AGGGCCCCTTCTGATTGTTC
                                  | |   |
Nucleotide No.                    7 5   1
```

FIG. 5. 3′-processing activities of INs on WT and modified HTLV-2 LTRs. WT HTLV-2 (A), HTLV-2 5,7A→T (B), HTLV-2 5A→T (C), and HTLV-2 7A→T (D) LTR substrates were used in assays with HIV-1 IN (lanes 2), HTLV-1 IN (lanes 3), HTLV-2 IN (lanes 4), and M-MuLV IN (lanes 5). Lane 1 in each panel represents the substrate with no protein. (E) Nucleotide sequences of the WT and modified HTLV-2 U5 LTR substrates used in the assays. Base substitutions are in boldface.

shown in comparison with those on the HTLV-1 WT LTR (Fig. 6A). HIV-1 IN displayed a significant gain in activity with the HTLV-1 5,7A→T LTR over the HTLV-1 WT LTR, with production of both 3′-processed and strand transfer products (Fig. 6B, lane 2). As expected, only basal 3′-processed product levels were observed for HTLV-1 IN and HTLV-2 IN with the modified LTR compared with the WT LTR (compare Fig. 6A and B, lanes 3 and 4). 3′ processing was confirmed only by the low, but nearly identical, levels of strand transfer products observed for both INs (Fig. 6B, lanes 3 and 4). Significantly, M-MuLV IN recognized the modified HTLV-1 substrate, as shown by the presence of strand transfer products (Fig. 6B, lane 5). Overall, the substitutions in the HTLV-1 LTR created a substrate recognized in a manner similar to those of the HIV-1 and M-MuLV WT LTRs.

Purine substitutions identify specific base requirements for the INs. Examination of purine-pyrimidine content at positions 5 to 7 revealed that among the HIV-1, HTLV-1, and HTLV-2 WT LTRs, two of three positions are purines; in the M-MuLV WT LTR, all three are pyrimidines. In the experiments described previously, changing the thymines in the M-MuLV LTR to adenines created substrates that could be recognized by the HTLV-1 and HTLV-2 INs. Significantly, these substitutions also increased the level of strand transfer activity for the M-MuLV IN. It was therefore of interest to determine whether the conferred recognition was sequence dependent or whether any purine at positions 5 and 7 could mediate catalysis. Thus, substitutions were made to examine whether guanine at these positions would also confer substrate recognition on the HTLV-1 and HTLV-2 INs for 3′ processing, as observed for the adenine substitution (M-MuLV 5T→A; Fig. 4C).

Two M-MuLV LTRs were designed with individual thymine-to-guanine substitutions at positions 5 (M-MuLV 5T→G) and 7 (M-MuLV 7T→G) (Fig. 7C). HIV-1 IN was active on both modified substrates, with a slight preference for the 5T→G modification (compare Fig. 7A and B, lanes 2). Interestingly, a previous report did not find activity with a similar substitution (46). Neither of the modified M-MuLV substrates was catalyzed by the HTLV-1 and HTLV-2 INs (Fig. 7A and B, lanes 3 and 4). M-MuLV IN had fourfold higher activity on the M-MuLV 5T→G LTR (Fig. 7A, lane 5) and twofold higher integration activity on the M-MuLV 7T→G LTR (Fig. 7B, lane 5) than on the M-MuLV WT LTR. Again, increased levels of strand transfer were observed for M-MuLV IN with these modified substrates. Similar purine substitutions (adenine to guanine) in the HTLV-2 WT LTR also created substrates recognized by HIV-1 and M-MuLV INs but not by either of



C.

```
HTLV-I WT          5'AGAGAAATTTAGTACACAGT
                   3'TCTCTTTAAATCATGTGTCA

HTLV-I 5,7A→T      5'AGAGAAATTTAGTTCTCAGT
                   3'TCTCTTTAAATCAAGAGTCA
                                 | |   |
Nucleotide No.                   7 5   1
```

FIG. 6. 3′-Processing activities of INs on WT and modified HTLV-1 LTRs. WT HTLV-1 (A) and HTLV-1 5,7A→T (B) LTR substrates were used in assays with HIV-1 IN (lanes 2), HTLV-1 IN (lanes 3), HTLV-2 IN (lanes 4), and M-MuLV IN (lanes 5). Lane 1 in each panel represents the substrate with no protein. (C) Nucleotide sequences of the WT and modified HTLV-1 U5 LTR substrates used in the assays. Base substitutions are in boldface.

the HTLV INs (data not shown). These results further confirm the specific requirement of the position 5 adenine for activity of the HTLV-1 and HTLV-2 INs. If either guanine or adenine substitutions had equally conferred activity on the substrate, it would have been inferred that the structure and size of this position(s) were more important than the chemistry inherent to the base. In effect, it would have provided a simple readout of the protein-DNA interactions based on a simple purine-pyrimidine sequence code. This, however, was not the case, indicating that the nucleotides contributed to the LTR sequence recognition via their specific functional group moieties.

Importance of upstream sequences for IN 3′-processing activity. Two hybrid substrates were synthesized to determine whether sequences upstream from LTR position 7 influenced substrate specificity: (i) 5′ HTLV-2–3′ M-MuLV, a substrate with the terminal 7 bp from M-MuLV and the remaining 13 from HTLV-2, and (ii) 5′ M-MuLV–3′ HTLV-2, a substrate with the terminal 7 bp from HTLV-2 and the remaining 13 from M-MuLV (Fig. 7F). With the 5′ HTLV-2–3′ M-MuLV hybrid LTR, activity levels displayed by the HIV-1, HTLV-1, and HTLV-2 INs (Fig. 7D, lanes 2 to 4) were comparable to their activity levels on the M-MuLV WT LTR (Fig. 4A, lanes 2 to 4). The M-MuLV IN catalyzed 12-fold higher activity with this hybrid LTR than that with its WT LTR (compare Fig. 4A and 7D, lanes 5). The 3′-processing activity levels of all four INs on the 5′ M-MuLV–3′ HTLV-2 hybrid were comparable to those on the HTLV-2 WT LTR (compare Fig. 1D and 7E). As observed with the HTLV-2 WT LTR, HTLV-2 IN exhibited the greatest activity with the 5′ M-MuLV–3′ HTLV-2 hybrid (Fig. 7E, lane 4). These results suggest that in the case of M-MuLV, the upstream HTLV-2 sequences contributed to substrate recognition. This is consistent with earlier reports signifying the importance of upstream sequences for catalysis by M-MuLV IN (15). For HIV-1, HTLV-1, and HTLV-2 INs, however, the sequences upstream from position 7 had little effect on LTR recognition.

DISCUSSION

In our comparisons of IN-catalyzed 3′ processing, strand transfer, and disintegration reactions, 3′ processing had the greatest requirement for specific LTR sequences. Therefore, the 3′-processing reaction served as our central focus to identify nucleotides within the viral LTR that determine substrate specificity for retroviral INs. Based on the levels of 3′-processing activity catalyzed by the HIV-1, HTLV-1, HTLV-2, and M-MuLV INs with their own and heterologous LTRs, the M-MuLV WT LTR was chosen as a suitable substrate for probing the substrate specificity determinants. This selection was based primarily on the low levels of activity catalyzed by the HTLV-1 and HTLV-2 INs on this LTR. In addition, sequence composition at positions 5 to 7 proximal to the conserved CA was distinctly different between the HTLV and M-MuLV LTRs. Based on these observations, we hypothesized that the substrate specificity determinants would be contained within the terminal 7 bp of the U5 LTR, since 6 to 9 bp of the LTR termini are sufficient for integration (5, 8, 22, 23, 25, 32, 50). The M-MuLV LTR was modified at positions 5 to 7 and assayed for a gain of function for the HTLV-1 and HTLV-2 INs with the hypothesis that we could engineer the M-MuLV substrate to be recognized as HTLV-1 and HTLV-2 substrates and vice versa. Mutational analyses, based on a gain-of-function approach, allowed us to identify specific nucleotides that converted nonfunctional heterologous LTRs into functional substrates for 3′ processing for a particular IN. These prediction-based analyses suggest that the role of the terminal nucleotides, specifically, those at positions 5, 6, and 7, was to confer substrate specificity during 3′ processing. Preferences for the nucleotide bases and their positions differed for each IN. Further, comparison of IN activities on the modified LTRs with adenine and guanine substitutions at these positions indicated that the specificity conferred by the nucleotide was associated with its individual chemical properties rather than the gross structure.

A summary of the activities of the four INs is presented in Table 1. The INs catalyzed higher levels of strand transfer activity on precleaved LTR substrates, compared with 3′ processing. With reference to 3′ processing, several observations were made concerning specific nucleotide requirements. HTLV-2 IN exhibited a more defined specificity, preferring a combination of adenine at position 5 and thymine or adenine at position 7 in its LTR. HTLV-1 maintained an overall low activity profile, although the adenine at LTR position 5 was important. M-MuLV IN exhibited a preference for A or G over T at position 5 in its LTR. Adenine and thymine substitutions in the M-MuLV WT and HTLV-2 WT LTRs were in

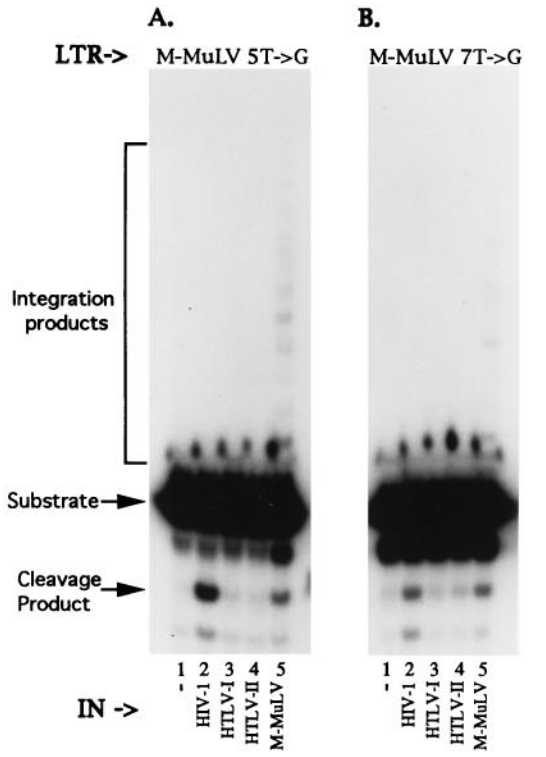


FIG. 7. 3′-Processing activities of INs on modified M-MuLV and HTLV-2–M-MuLV hybrid LTRs. M-MuLV 5T→G (A), M-MuLV 7T→G (B), 5′ HTLV-2–3′ M-MuLV (D), and 5′ M-MuLV–3′ HTLV-2 (E) LTR substrates were used in assays with HIV-1 IN (lanes 2), HTLV-1 IN (lanes 3), HTLV-2 IN (lanes 4), and M-MuLV IN (lanes 5). Lane 1 in each panel represents the substrate with no protein. (C and F) Nucleotide sequences of the modified M-MuLV LTRs and HTLV-2–M-MuLV hybrid LTRs, respectively. Base substitutions in the modified M-MuLV substrates and the M-MuLV sequences in the hybrid substrates are in boldface.

agreement with the requirement for adenine at position 5 and thymine at position 7 for the M-MuLV IN. In addition, the hybrid substrates indicated that certain upstream positions may also be involved in the LTR recognition by M-MuLV IN. Overall, the LTR substitution studies suggested that the M-MuLV WT LTR is not an optimal substrate for its IN. This observation is likely to have evolutionary significance (27). Therefore, contrary to its poor profile, as shown by the low activity on its own WT LTR, the M-MuLV IN efficiently processed alternate LTR substrates. It could be argued that the suboptimal WT LTR sequence is responsible for the protein's apparent low activity. Among the four INs, HIV-1 IN displayed the broadest range of substrate specificities. HIV-1 IN efficiently utilized the HIV-1, HTLV-2, and M-MuLV WT LTRs as substrates for 3′ processing and tolerated most substitutions in these substrates. Contrary to our observations, previous reports of HIV-1 IN have not shown activity with the M-MuLV and HTLV-1 LTRs (22, 36, 46). This may reflect differences in the IN assay conditions or purification schemes. Previous studies examining nucleotide requirements in the subterminal LTR regions have shown a decrease or loss of IN activity with nucleotide substitution, specifically, at positions 5 through 8 (22, 32, 46, 50). Mutations in the HIV-1 U3 LTR, at positions 6 to 8, significantly decrease 3′ processing in vitro, as well as lower reverse transcriptase activity and delay progeny formation in vivo (32). Similar deletion and substitution studies with cell cultures using M-MuLV highlight the requirement of nucleotides corresponding to positions 5 through 8 for reverse transcriptase activity, progeny formation, and viability of the virus (25). We have further explored these nucleotide requirements by using a gain-of-function approach and defined their role as providing substrate specificity to IN during 3′ processing.

Close examination of the strand transfer reaction showed that each of the four INs produced the same pattern of strand transfer products when a given LTR sequence was used as the target DNA. Further, single or double base pair substitutions in the WT LTRs significantly changed the preferred sites of integration. Although retroviral integration into a target sequence is not sequence specific, our results indicate that preferences for certain sites or regions exist, an observation consistent with earlier reports (4, 12, 13, 23, 29, 30). A preference for A-T-rich regions in the target DNA has been shown for integration catalyzed by avian myeloblastosis virus (12, 21) and M-MuLV (29) INs. Target site selection has been shown to be a function of the target sequence and is independent of the viral LTR that undergoes 3′ processing (23). Results from our comparative analysis are in agreement with this report. Over-

TABLE 1. Summary of activities of HIV-1, HTLV-1, HTLV-2, and M-MuLV INs on WT and modified LTR substrates

LTR	Activity[a] of:			
	HIV-1 IN	HTLV-1 IN	HTLV-2 IN	M-MuLV IN
WT blunt LTRs				
HIV-1	++++	+	++	+
HTLV-1	+	++	++	±
HTLV-2	++	++	++	++
M-MuLV	++	+	±	+
WT cleaved LTRs				
HIV-1	++++	+++	+++	+++
HTLV-1	+++	*	++++	+++
HTLV-2	+++	++	+++	ND
M-MuLV	++	++	++	±
M-MuLV 5T→A	+++	+++	+++	*
M-MuLV 6T→A	+	±	±	++++
M-MuLV 7T→A	+	±	±	++
M-MuLV 5,7T→A	+++	+++	+++	++++
HTLV-2 5A→T	+	±	±	+
HTLV-2 7A→T	±	+	++	++++
HTLV-2 5,7A→T	+	±	±	+++
HTLV-1 5,7A→T	+++	++	++	+++
M-MuLV 5T→G	++	±	±	+++
M-MuLV 7T→G	+	+	+	++++
5′ HTLV-2–3′ M-MuLV	+	±	+	*
5′ M-MuLV–3′ HTLV-2	+	++	+	+

[a] ±, <5%; +, 5 to 10%; ++, 11 to 20%; +++, 21 to 35%; ++++, 36 to 50%; *, >50%; ND, no detectable activity.

all, it may suggest that the domain for recognition of the target (host DNA) during strand transfer is conserved among retroviral INs. Further, the INs catalyzed higher levels of strand transfer activity with precleaved substrates than with 3′ processing with the blunt substrates (Table 1).

In summary, each of the retroviral INs exhibited unique substrate specificity traits. The comparative study presented here indicates that the nucleotides in the LTR termini present some structural complementarity, in terms of hydrogen bonding groups, to the LTR-binding sites in the IN. As with other sequence-specific DNA-binding proteins, recognition and binding by IN at specific LTR sequences may involve the interaction of a matrix of hydrogen bond donors and acceptors in the DNA grooves with the sterically complementary hydrogen bond acceptors and donors in the protein's binding site. These specific interactions may be further coupled with nonspecific interactions along the upstream regions of the LTR with IN. Since single base pair substitutions in the LTRs modulated IN activity at significant levels, it can be speculated that only a limited region in the LTR is involved in specific interactions with IN. Substitutions involving the critical nucleotides may therefore create disturbances in the hydrogen bond interactions reflected by low IN activity. However, nucleotide substitutions in the LTR that do not result in significant mispairing with the protein hydrogen bonding groups may maintain sufficient interactions with IN to mediate moderate levels of integration activity. Now that the critical nucleotide positions in the viral LTR that confer substrate specificity have been identified, the specific amino acids in the IN can be defined.

ACKNOWLEDGMENTS

We thank Robert Craigie, NIH, for providing the HIV-1 integrase. We gratefully acknowledge Rasika Harshey and Rita Varagona for critical reading of the manuscript and Devon Zastrow for technical assistance.

This work was supported by grants from the National Science Foundation (MCB-9509144) and the Leukemia Research Foundation Inc.

REFERENCES

1. **Balakrishnan, M., and C. B. Jonsson.** 1996. Unpublished data.
2. **Balakrishnan, M., D. Zastrow, and C. B. Jonsson.** 1996. Catalytic activities of the human T-cell leukemia virus integrase. Virology **219:**77–86.
3. **Brown, P. O.** 1990. Integration of retroviral DNA. Curr. Top. Microbiol. Immunol. **157:**19–48.
4. **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.
5. **Bushman, F. D., and R. Craigie.** 1990. Sequence requirements for integration of Moloney murine leukemia virus DNA in vitro. J. Virol. **64:**5645–5648.
6. **Bushman, F. D., A. Engelman, I. Palmer, P. Wingfield, and R. Craigie.** 1993. Domains of the integrase protein of human immunodeficiency virus type I responsible for polynucleotidyl transfer and zinc binding. Proc. Natl. Acad. Sci. USA **90:**3428–3432.
7. **Chow, S. A., K. A. Vincent, V. Ellison, and P. O. Brown.** 1992. Reversal of integration and DNA splicing mediated by integrase of human immunodeficiency virus. Science **255:**723–726.
8. **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.
9. **Craigie, R., T. Fujiwara, and F. Bushman.** 1990. The IN protein of Moloney murine leukemia virus processes the viral DNA ends and accomplishes their integration in vitro. Cell **62:**829–837.
10. **Donzella, G. A., C. B. Jonsson, and M. J. Roth.** 1993. Influence of substrate structure on disintegration by Moloney murine leukemia virus integrase. J. Virol. **67:**7077–7087.
11. **Engelman, A., K. Mizuuchi, and R. Craigie.** 1991. HIV-1 DNA integration: mechanism of viral DNA cleavage and DNA strand transfer. Cell **67:**1211–1221.
12. **Fitzgerald, M. L., A. C. Vora, W. G. Zeh, and D. P. Grandgenett.** 1992. Concerted integration of viral DNA termini by purified avian myeloblastosis virus integrase. J. Virol. **66:**6257–6263.
13. **Hong, T., E. Murphy, J. Groarke, and K. Drlica.** 1993. Human immunodeficiency virus type 1 DNA integration: fine structure target analysis using synthetic oligonucleotides. J. Virol. **67:**1127–1131.
14. **Jonsson, C. B., G. A. Donzella, and M. J. Roth.** 1993. Characterization of the forward and reverse integration reactions of the Moloney murine leukemia virus integrase protein purified from *Escherichia coli*. J. Biol. Chem. **268:**1462–1469.
15. **Jonsson, C. B., and M. J. Roth.** 1993. Role of the His-Cys finger of the Moloney murine leukemia virus integrase protein in integration and disintegration. J. Virol. **67:**5562–5571.
16. **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.
17. **Katz, R. A., and A. M. Skalka.** 1994. The retroviral enzymes. Annu. Rev. Biochem. **63:**133–173.
18. **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.
19. **Khan, E., J. P. G. Mack, R. A. Katz, J. Kulkosky, and A. M. Skalka.** 1991. Retroviral integrase domains: DNA binding and the recognition of LTR sequences. Nucleic Acids Res. **19:**851–860.
20. **Kitamura, Y., Y. M. H. Lee, and J. M. Coffin.** 1992. Nonrandom integration of retroviral DNA *in vitro*: effect of CpG methylation. Proc. Natl. Acad. Sci. USA **89:**5532–5536.
21. **Knaus, R. J. P., P. J. Hippenmeyer, T. K. Misra, D. P. Grandgenett, V. R. Muller, and W. M. Fitch.** 1984. Avian retrovirus pp32 DNA binding protein: preferential binding to the promoter region of long terminal repeat DNA. Biochemistry **23:**350–359.
22. **LaFemina, R. L., P. L. Callahan, and M. G. Cordingley.** 1991. Substrate specificity of recombinant human immunodeficiency virus integrase protein. J. Virol. **65:**5624–5630.
23. **Leavitt, A. D., R. B. Rose, and H. E. Varmus.** 1992. Both substrate and target oligonucleotide sequences affect in vitro integration mediated by human immunodeficiency virus type 1 integrase protein produced in *Saccharomyces cerevisiae*. J. Virol. **66:**2359–2368.
24. **Mizuuchi, K.** 1992. Polynucleotidyl transfer reactions in transpositional DNA recombination. J. Biol. Chem. **267:**21273–21276.
25. **Murphy, J. E., T. De Los Santos, and S. P. Goff.** 1993. Mutational analysis of the sequences at the termini of the Moloney murine leukemia virus DNA required for integration. Virology **195:**432–440.

26. **Pardi, D., W. M. Switzer, K. G. Hadlock, J. E. Kaplan, R. B. Lal, and T. M. Folks.** 1993. Complete nucleotide sequence of an Amerindian human T-cell lymphotropic virus type II (HTLV-II) isolate: identification of a variant HTLV-II subtype b from a Guaymi Indian. J. Virol. **67:**4659–4664.
27. **Partin, L., B. Milligan, and C. B. Jonsson.** 1996. Unpublished data.
28. **Pemberton, I. K., M. Buckle, and H. Buc.** 1996. The metal ion-induced cooperative binding of HIV-1 integrase to DNA exhibits a marked preference for Mn(II) rather than Mg(II). J. Biol. Chem. **271:**1498–1506.
29. **Pryciak, P. M., A. Sil, and H. Varmus.** 1992. Retroviral integration into minichromosomes in vitro. EMBO J. **11:**291–303.
30. **Pryciak, P. M., and H. E. Varmus.** 1992. Nucleosomes, DNA-binding proteins, and DNA sequence modulate retroviral integration target site selection. Cell **69:**769–780.
31. **Ratner, L., T. Philpott, and D. B. Trowbridge.** 1991. Nucleotide sequence analysis of isolates of human T-lymphotropic virus type I of diverse geographical origins. AIDS Res. Hum. Retroviruses **7:**923–941.
32. **Reicin, A. S., G. Kalpana, S. Paik, S. Marmon, and S. Goff.** 1995. Sequences in the human immunodeficiency virus type 1 U3 region required for in vivo and in vitro integration. J. Virol. **69:**5904–5907.
33. **Rohdewohld, H., H. Weiher, W. Reik, R. Jaenisch, and M. Breindl.** 1987. Retrovirus integration and chromatin structure: Moloney murine leukemia proviral integration sites map near DNase I-hypersensitive sites. J. Virol. **61:**336–343.
34. **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.
35. **Schauer, M., and A. Billich.** 1992. The N-terminal region of HIV-1 integrase is required for integration activity, but not for DNA-binding. Biochem. Biophys. Res. Commun. **185:**874–880.
36. **Sherman, P. A., M. L. Dickson, and J. A. Fyfe.** 1992. Human immunodeficiency virus type 1 integration protein: DNA sequence requirements for cleavage and joining reactions. J. Virol. **66:**3593–3601.
37. **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.
38. **Shih, C., J. P. Stoye, and J. M. Coffin.** 1988. Highly preferred targets for retrovirus integration. Cell **53:**531–537.
39. **Shinnick, T. M., R. A. Lerner, and J. G. Sutcliffe.** 1981. Nucleotide sequence of Moloney murine leukemia virus. Nature **293:**543–548.
40. **Stormann, K. D., M. C. Schlecht, and E. Pfaff.** 1995. Comparative studies of bacterially expressed integrase proteins of caprine arthritis-encephalitis virus, maedi-visna virus and human immunodeficiency virus type 1. J. Gen. Virol. **76:**1651–1663.
41. **van den Ent, F. M. I., C. Vink, and R. H. A. Plasterk.** 1994. DNA substrate requirements for different activities of the human immunodeficiency virus type 1 integrase protein. J. Virol. **68:**7825–7832.
42. **van Gent, D. C., Y. Elgersma, M. W. J. Bolk, C. Vink, and R. H. A. Plasterk.** 1991. DNA binding properties of the integrase proteins of human immunodeficiency viruses types 1 and 2. Nucleic Acids Res. **19:**3821–3827.
43. **Varmus, H. E., and P. Brown.** 1989. Retroviruses. American Society for Microbiology, Washington, D.C.
44. **Varmus, H. E., and R. Swanstrom.** 1984. Replication of retroviruses, p. 369–512. *In* R. Weiss, N. Teich, H. Varmus, and J. Coffin (ed.), RNA tumor viruses. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
45. **Vijaya, S., D. L. Steffen, and H. L. Robinson.** 1986. Acceptor sites for retroviral integrations map near DNase I-hypersensitive sites. J. Virol. **60:** 683–692.
46. **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.
47. **Vink, C., D. C. van Gent, and R. H. A. Plasterk.** 1990. Integration of human immunodeficiency virus types 1 and 2 DNA in vitro by cytoplasmic extracts of Moloney murine leukemia virus-infected mouse NIH 3T3 cells. J. Virol. **64:**5219–5222.
48. **Vink, C., A. A. M. O. Groeneger, and R. H. A. Plasterk.** 1993. Identification of the catalytic and DNA-binding region of the human immunodeficiency virus type 1 integrase protein. Nucleic Acids Res. **21:**1419–1425.
49. **Vora, A. C., M. L. Fitzgerald, and D. P. Grandgenett.** 1990. Removal of 3′-OH-terminal nucleotides from blunt-ended long terminal repeat termini by the avian retrovirus integration protein. J. Virol. **64:**5656–5659.
50. **Yoshinaga, T., and T. Fujiwara.** 1995. Different roles of bases within the integration signal sequence of human immunodeficiency virus type 1 in vitro. J. Virol. **69:**3233–3236.