JOURNAL OF VIROLOGY, Jan. 1995, p. 456–468 Vol. 69, No. 1
0022-538X/95/$04.00+0

Characterization of Recombinant Murine Leukemia Virus Integrase

IRIS DOTAN,[1]† BRIAN P. SCOTTOLINE,[2] TIMOTHY S. HEUER,[2] AND PATRICK O. BROWN[1,2]*

Howard Hughes Medical Institute[1] and Department of Biochemistry,[2] Stanford University Medical Center, Stanford, California 94305-5428

Received 8 August 1994/Accepted 17 October 1994

Retroviral integration involves two DNA substrates that play different roles. The viral DNA substrate is recognized by virtue of specific nucleotide sequences near the end of a double-stranded DNA molecule. The target DNA substrate is recognized at internal sites with little sequence preference; nucleosomal DNA appears to be preferred for this role. Despite this apparent asymmetry in the sequence, structure, and roles of the DNA substrates in the integration reaction, the existence of distinct binding sites for viral and target DNA substrates has been controversial. In this report, we describe the expression in *Escherichia coli* and purification of Moloney murine leukemia virus integrase as a fusion protein with glutathione *S*-transferase, characterization of its activity by using several model DNA substrates, and the initial kinetic characterization of its interactions with a model viral DNA substrate. We provide evidence for functionally and kinetically distinct binding sites for viral and target DNA substrates and describe a cross-linking assay for DNA binding at a site whose specificity is consistent with the target DNA binding site.

Integration of the retroviral genome into host cell chromosomal DNA is required for viral replication. Genetic studies have demonstrated that retroviral integration requires two viral functions: the viral IN protein, encoded by the 3′ end of the *pol* gene (14, 28, 38, 41), and *att* sites located at the termini of the viral long terminal repeats (11, 12, 37; for reviews, see references 2, 25, 26, 48, 53). Integration proceeds in three steps. After reverse transcription in the cytoplasm of the infected cell, the 3′ ends of each of the termini of the linear viral DNA are cleaved to remove the terminal two nucleotides. This step, termed 3′-end processing, generates a 2-nucleotide 5′-end overhang and 3′ ends that terminate with the phylogenetically conserved CA dinucleotide (4, 24, 39). In the nucleus, a concerted cleavage and ligation joins the processed 3′ viral DNA ends to 5′ staggered sites on opposite strands of the host chromosomal DNA (3, 4, 16, 20, 24). This strand transfer reaction results in a gapped intermediate in which the viral DNA 5′ ends and the host chromosomal DNA 3′ ends are unjoined. The integration process is completed by repair of the gapped intermediate. This step generates short direct repeats that flank the integrated provirus. The proteins involved in this final step remain to be identified.

The end-processing and strand transfer steps in integration require the activity of only a single protein, the virally encoded IN. Purified recombinant IN has been shown both to remove two nucleotides from the 3′ end of model viral DNA substrates and to join the recessed 3′ ends to target DNA in vitro (6, 8, 13, 30–32, 35, 36, 43, 46, 49, 50, 52). Both the 3′-end processing and strand transfer steps of integration proceed via one-step transesterification reactions (17) that require no exogenous energy source (3, 16, 20). The strand transfer reaction has been shown to be reversible in vitro by using a substrate that mimics a strand transfer intermediate (10). The products of the reverse reaction, termed disintegration, are substrates for a subsequent round of integration (49).

To facilitate studies of MoMLV IN, we have expressed this enzyme as a fusion protein with GST in *Escherichia coli* (45). We used the partially purified protein to analyze in detail the conditions that affect the catalytic activities of MoMLV IN. We have characterized the products of MoMLV IN-catalyzed 3′-end processing and compared them with those reported for HIV-1 (17, 51), HIV-2 (51), and FIV (50) INs. We report the results of kinetic measurements of 3′-end processing and present evidence that MoMLV IN is capable of catalytic turnover. Finally, we provide evidence that MoMLV IN contains two distinct DNA binding sites, one with specificity for a viral DNA end and one for target DNA.

MATERIALS AND METHODS

Abbreviations used. The following abbreviations are used in this paper: MoMLV, Moloney murine leukemia virus; HIV, human immunodeficiency virus; RSV, Rous sarcoma virus; FIV, feline immunodeficiency virus; ASLV, avian sarcoma leukosis virus; AMV, avian myeloblastosis virus; IN, integrase; GST, glutathione *S*-transferase; PNK, polynucleotide kinase; CIP, calf intestinal phosphatase; DTT, dithiothreitol; PMSF, phenylmethylsulfonyl fluoride; IPTG, isopropyl-1-thio-β-D-galactopyranoside; HEPES, 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid; MOPS, 4-morpholinopropanesulfonic acid; PIPES, 1,4-piperazinediethanesulfonic acid; MES, 4-morpholineethanesulfonic acid; SDS-PAGE, sodium dodecyl sulfate-polyacrylamide gel electrophoresis; NP-40, Nonidet P-40; BSA, bovine serum albumin; BCIP, 5-bromo-4-chloro-3-indolyl phosphate; BrdU, 5-bromo-2′-deoxyuridine; BrdUTP, 5-bromo-2′-deoxyuridine triphosphate.

Enzymes and chemicals. Restriction enzymes, T4 DNA ligase, and T4 PNK were from New England Biolabs (Beverly, Mass.). The large fragment of *E. coli* DNA polymerase I and CIP were from Boehringer-Mannheim (Indianapolis, Ind.). *Taq* DNA polymerase was from Perkin-Elmer Cetus (Norwalk, Conn.). Sequenase version 2.0 T7 DNA polymerase was from United States Biochemicals (Cleveland, Ohio). IPTG, PMSF, and lysozyme were from Sigma (St. Louis, Mo.). Aprotinin, pepstatin A, leupeptins, and antipain were from Boehringer-Mannheim. NP-40 was from Pierce (Rockford, Ill.).

Construction of pGEX-3X MLV IN. A 1.9-kb *Bam*HI-to-*Bam*HI molecular clone of MoMLV IN was created as follows. A 303-bp fragment spanning the first codon to the *Hin*dIII site in the IN coding region was amplified by PCR for 30 cycles from a molecular clone of MoMLV derived from pZAP (44) by using *Taq* DNA polymerase. The primer sequences were as follows: amino-terminal primer (*Bam*HI site underlined), 5′-GGGGGGATCCAATCATCACCCTA CACCTCAGAACATTTTCA-3′; *Hin*dIII primer (*Hin*dIII site underlined), 5′-GACTTGTGCACAAGCTTTGCAGGTCTCAGTG-3′ (Protein and Nucleic Acid facility, Stanford University). The amplified DNA was digested with *Bam*HI and *Hin*dIII and purified by Centricon-100 ultrafiltration (Amicon, Beverly, Mass.). The plasmid pSPendo46 (kindly provided by L. Donehower and H. E. Varmus, University of California, San Francisco), which contains a molecular clone of the MoMLV IN coding region, was digested with *Hin*dIII, treated with CIP, and ligated to the *Bam*HI-*Hin*dIII fragment of the MLV IN gene. The

* Corresponding author. Phone: (415) 725-7569. Fax: (415) 723-1399. Electronic mail address: pbrown@cmgm.stanford.edu.

† Present address: Department of Biochemistry, Faculty of Life Sciences, Tel Aviv University, Tel Aviv 69978, Israel.

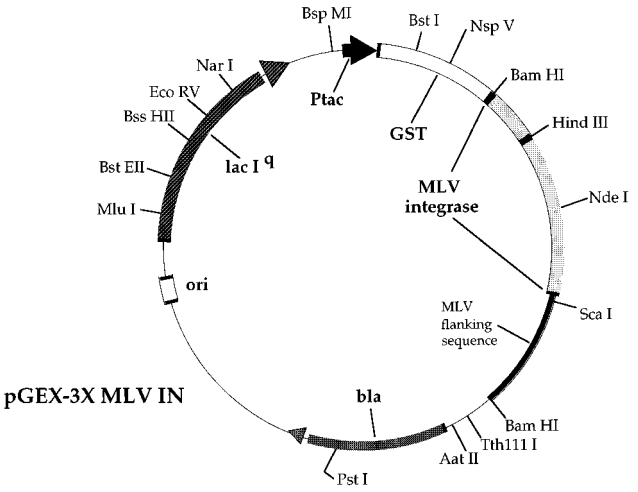


FIG. 1. Schematic diagram of the pGEX-3X-derived vector pGEX-3X MLV IN constructed to express MoMLV IN as the GST-MLV IN fusion protein.

resulting ligated DNA was then digested with *Bam*HI, and the *Bam*HI-to-*Bam*HI fragment containing the MoMLV IN coding region was purified and ligated into *Bam*HI-digested, CIP-treated pGEX-3X (Pharmacia, Piscataway, N.J.) to yield pGEX-3X MLV IN (Fig. 1). pGEX-3X MLV IN was then transformed into *E. coli* DH5α (27).

Purification of GST-MLV IN. DH5α(pGEX-3X MLV IN) was grown to an optical density at 600 nm of 0.8 in 10 liters of Luria broth at 30°C, and IPTG was added to a final concentration of 0.25 mM. The cells were then grown for an additional 3 h at 30°C and collected by centrifugation at 5,000 × *g*. The cell pellets were either frozen at −80°C and stored for later processing or resuspended in lysis buffer (50 mM Tris-Cl [pH 8.0], 10 mM EDTA, 140 mM NaCl, 10 mM DTT, 10% glycerol, 1 mM PMSF, 1 μg of pepstatin A per ml, 10 μg of leupeptins per ml, 1 μg of antipain per ml, 1 μg of aprotinin per ml), and lysozyme was added to 1 mg/ml. The cells were then incubated on ice for 1 h before being subjected to three cycles of freezing in liquid N_2 followed by thawing. The suspension was then sonicated on ice by five 30-s bursts at full power (Bronson Sonifier 450; Bronson Instruments, Danbury, Conn.). All subsequent steps were done at 4°C or on ice. The lysate was spun at 105,000 × *g* for 30 min. The supernatant was filtered through a 0.45-μm-pore-size filter, brought to 0.5% NP-40, and applied to a 10-ml glutathione-agarose column (Sigma) equilibrated in buffer A (50 mM Tris-Cl [pH 7.5], 100 mM NaCl, 10 mM DTT, 10% glycerol, 0.5% NP-40). The column was washed with 100 ml of buffer B (same as buffer A but with 0.05% NP-40) and then eluted with buffer C (50 mM Tris-Cl [pH 8.0], 100 mM KCl, 10 mM glutathione-HCl, 10% glycerol, 0.05% NP-40). The majority of GST-MLV IN eluted within 2 column volumes. Protein concentration was determined by the method of Bradford (1) (Bio-Rad, Emeryville, Calif.) by using BSA as the standard, and purity was estimated by laser scanning densitometry (Molecular Dynamics, Mountain View, Calif.) of 10% SDS-PAGE gels (33) stained with Coomassie brilliant blue R250 (Bio-Rad) by using ImageQuant 3.15 software. The fusion protein constituted approximately 10 to 20% of the total protein after this step. This protein was stored at −80°C in the elution buffer.

Further purification was accomplished by applying the partially purified protein to a 4-ml phosphocellulose column (Whatman Biosystems, Clifton, N.J.) equilibrated in buffer D (20 mM HEPES-OH [pH 7.4], 10 mM DTT, 10% glycerol, 0.05% NP-40) with 100 mM KCl. The column was washed with 40 ml of buffer D with 100 mM KCl and then developed with a linear gradient (60 ml) from 0.1 to 1.5 M KCl in buffer D. The peak of the fusion protein eluted at 750 mM KCl. Fractions eluting between 600 and 900 mM KCl were pooled, and GST-MLV IN constituted approximately 50% of the total protein in this pool as determined by SDS–10% PAGE and scanning laser densitometry.

SDS-PAGE and Western blotting. SDS–10% PAGE was performed as previously described (33) to monitor the expression and purification of the fusion protein. Western blotting (immunoblotting) was performed by standard procedures (40) with proteins blotted to polyvinylidene difluoride (Millipore, Bedford, Mass.). The blot was incubated in blocking solution (10 mM Tris-Cl [pH 7.4], 150 mM NaCl, 10% nonfat dry milk, 0.02% sodium azide) followed by incubation with the rabbit anti-MoMLV antiserum 76S-00124 (National Cancer Institute) diluted 1:1,000 in blocking solution, washed in TTBS (10 mM Tris-Cl [pH 7.5], 150 mM NaCl, 0.1% Tween 20), and incubated with alkaline phosphatase-conjugated swine anti-rabbit immunoglobulin G (Boehringer-Mannheim) diluted 1:1,000 in blocking solution. The blot was washed in TTBS and developed in a solution containing 150 mM Tris-Cl (pH 9.5), 100 μg of nitro blue tetrazolium per ml, 50 μg of BCIP per ml, and 5 mM $MgCl_2$.

3′-end processing and strand transfer assays. The 3′-end processing and strand transfer activities of IN were assayed in vitro by using a double-stranded synthetic oligonucleotide substrate corresponding to the terminal 23 nucleotides of the U5 end of viral DNA (Operon Technologies, Alameda, Calif.). The strand corresponding to the 3′ terminus of viral DNA (oligonucleotide A194, sequence 5′-ACCTACAGGTGGGGTCTTTCATT-3′) was labeled at the 5′ end with [γ-^{32}P]ATP (4,500 Ci/mmol; ICN, Irvine, Calif.) by using T4 PNK. The labeled strand was purified by gel filtration through two successive 1-ml-bed-volume Sephadex G-15 (Pharmacia) spin columns (40). The recovery was determined by scintillation counting. The labeled strand was then annealed to a 1.5-fold excess

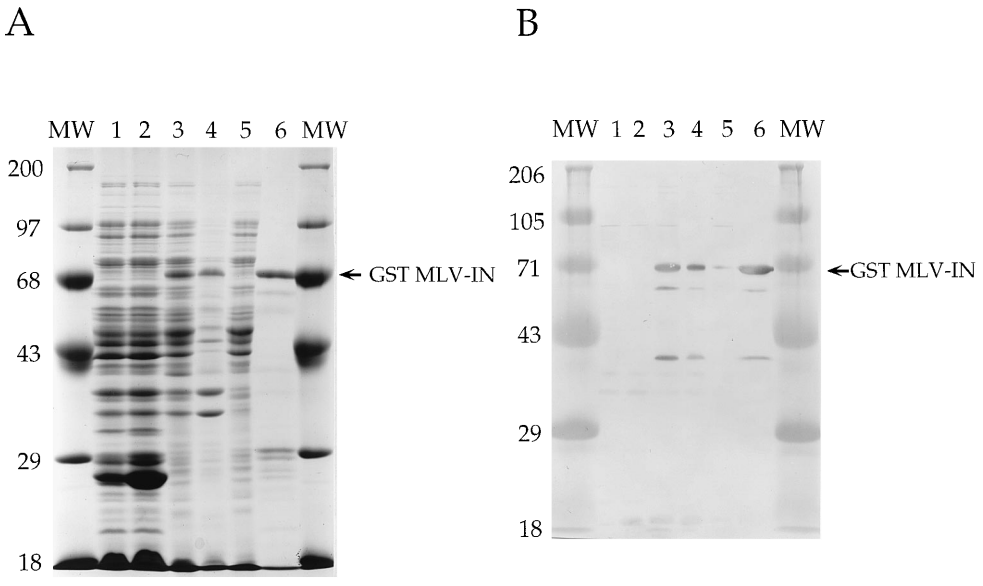


FIG. 2. SDS-PAGE and Western blot of partially purified GST-MLV IN fusion protein. (A) Coomassie blue-stained SDS-PAGE gel of proteins from uninduced DH5α harboring pGEX-3X (lane 1), DH5α(pGEX-3X) expressing GST (lane 2), and DH5α(pGEX-3X MLV IN) expressing GST-MLV IN (lane 3); insoluble protein from an extract of induced DH5α(pGEX-3X MLV IN) (lane 4); soluble fraction from induced DH5α(pGEX-3X MLV IN) (lane 5); and glutathione-agarose-purified GST-MLV IN (fraction I) (lane 6). MW, molecular weight standards. Numbers on the left are the masses of the molecular weight standard proteins (in kilodaltons). (B) Western blot analysis of the samples shown in panel A.

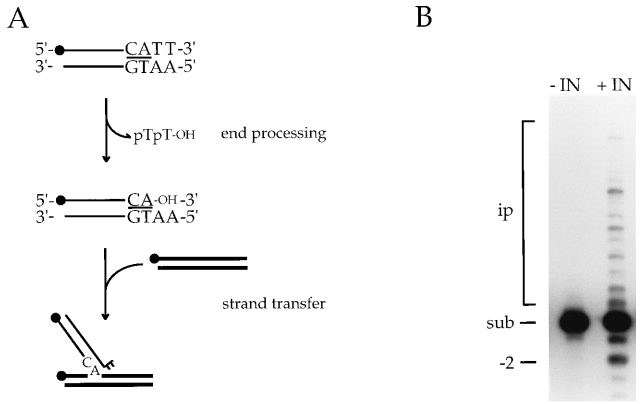


FIG. 3. Integration assay. (A) Diagram of the integration assay with MLV viral DNA end oligonucleotide substrates. The sequence of the double-stranded synthetic oligonucleotide matches that of the MLV U5 *att* site, located at one end of the viral DNA. Shown is the sequence of the terminal 4 bp of the *att* DNA, with the phylogenetically conserved CA dinucleotide underlined. The oligonucleotide is ^{32}P labeled at the 5′ end of the CA-containing strand, signified by the filled circle. During end processing, IN catalyzes the endonucleolytic cleavage of the phosphodiester bond immediately 3′ to the CA dinucleotide. The newly recessed 3′ hydroxyl serves as the nucleophile in the subsequent strand transfer reaction, which joins the viral DNA substrate to target DNA (thick lines). In an integration reaction without exogenous target, the target DNA is a second viral DNA end oligonucleotide. Assays for integration strand transfer activity alone employ a prerecessed double-stranded synthetic oligonucleotide substrate which has the exact structure of the end processed substrate. Integration shows little target sequence specificity (3, 13, 43a) and thus can occur at many positions in the target DNA, yielding multiple strand transfer products. (B) Integration reaction catalyzed by partially purified GST-MLV IN. The viral DNA end oligonucleotide substrate A194/A193 (sub) is labeled at the 5′ end and is converted into end processing products (−2) and integration products (ip). −IN indicates the reaction without IN; +IN indicates a reaction with IN.

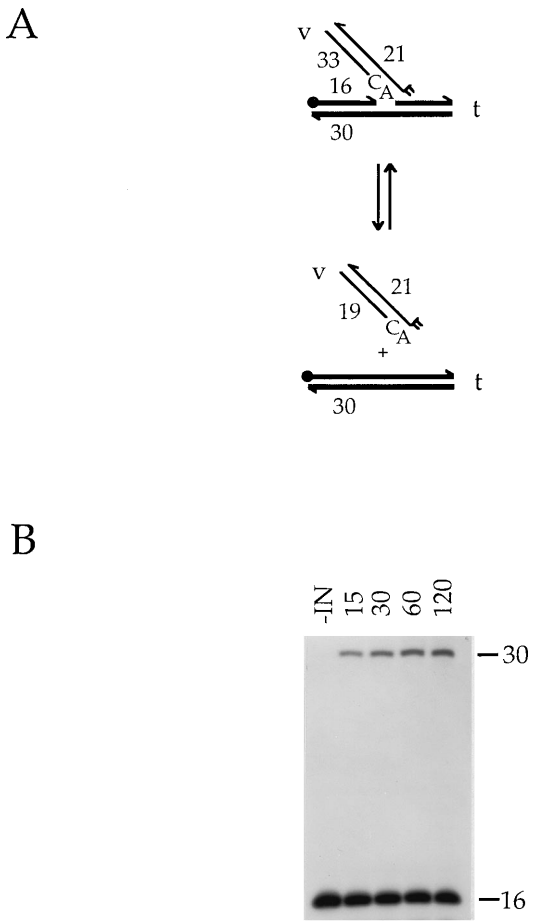


FIG. 4. Disintegration assay. (A) Diagram of the disintegration reaction with the Y-oligomer substrate. The Y oligomer mimics the structure of the product of an in vitro strand transfer reaction. The substrate is composed of both viral end sequences, including the conserved CA dinucleotide, represented by thin lines and the letter v, and a segment corresponding to target DNA, represented by thick lines and the letter t. The 3′ ends of the oligonucleotides in the Y oligomer are shown as half-arrowheads, while the 5′ ^{32}P-labeled end is shown as a filled circle. Numbers indicate the lengths (in nucleotides) of the oligonucleotides constituting the complete MLV and HIV disintegration Y-oligomer substrates. Incubation of the Y oligomer with IN results in a reaction in which the 16-mer 3′ hydroxyl serves as a nucleophile in a strand transfer reaction that joins the 16-mer to target DNA 3′ to the viral strand CA dinucleotide. This converts the labeled 16-mer into a 30-mer product and results in the release of the viral DNA sequences (thin lines) from the target DNA (thick lines). The products of the disintegration reaction can then undergo reintegration. (B) Reactions carried out with an MLV Y-oligomer. The label is on the 5′ end of the 16-mer in the substrate, and the reaction yields the 30-mer product. −IN indicates reaction in the absence of IN, and the numbering indicates reaction time (in minutes).

of its complement (oligonucleotide A193, sequence 5′-AATGAAAGACCCCA CCTGTAGGT-3′) in TEN (10 mM Tris-HCl [pH 7.4], 1 mM EDTA [pH 8.0], 100 mM NaCl) by heating to 95°C for 5 min followed by slow cooling to room temperature.

The in vitro assays of end processing and strand transfer activities (integration assay) were conducted as follows, except when indicated otherwise. One picomole of viral DNA end oligonucleotide substrate was incubated with 0.2 pmol of GST-MLV IN fusion protein in a reaction buffer consisting of 20 mM MOPS-OH (pH 7.2), 75 mM KCl, 5 mM $MnCl_2$, 10 mM DTT, 20% glycerol, and 50 μg of BSA per ml in a volume of 10 μl for 30 min at 30°C. The reactions were terminated by the addition of an equal volume of 95% deionized formamide–20 mM EDTA–0.05% xylene cyanol–0.05% bromophenol blue and heated to 95°C for 3 min prior to electrophoresis through a 20% polyacrylamide–7 M urea gel at 2,000 V for 6 h.

Assays for strand transfer alone were accomplished with a synthetic oligonucleotide substrate containing a 5′-end-labeled preprocessed strand (oligonucleotide A198, sequence 5′-ACCTACAGGTGGGGTCTTTCA-3′; Protein and Nucleic Acid facility) annealed to oligonucleotide A193. The reaction conditions were as described above. Products of end-processing and strand transfer reactions were visualized by autoradiography and were quantitated by using a Phosphorimager (Molecular Dynamics) equipped with ImageQuant 3.22 software. Total end-processing activity was defined as the sum of −2 end-processing and strand transfer activities, since strand transfer products are derived from end-processed substrates.

3′-end processing was measured directly in a manner similar to that of Engleman et al. (19) and Vink et al. (51) through the use of the 3′-end-labeled viral DNA end substrate A198tt/A193, which was prepared as follows. Oligonucleotides A198 and A195 (oligonucleotide A195 sequence, 5′-AATGAAAGAC CCCACCTGTAG-3′ [Protein and Nucleic Acid facility]) were annealed, and the recessed strand A198 was then filled in with [α-^{32}P]TTP (3,000 Ci/mmol; ICN) by using Sequenase version 2.0 for 15 min and then adding cold TTP to a concentration of 100 μM and incubating for 15 min. This results in a substrate which is almost exclusively labeled at the phosphates flanking the penultimate T. The filled-in material was then purified by two successive gel filtration steps through 1-ml Sephadex G-15 spin columns followed by electrophoresis through a 20% polyacrylamide–7 M urea gel. The products were visualized by autoradiography, and the desired oligonucleotide, termed A198tt, was excised from the wet gel, eluted, quantitated, and annealed to the complementary oligonucleotide, A193. This substrate was then used in the standard integration reaction for 30 min at 30°C, and the products of the reaction were separated by electrophoresis on 24% polyacrylamide–7 M urea gels run at 2,400 V for 3 h. The products were visualized and quantitated as described above. Total end-processing activity was defined as the sum of all radiolabeled products of end processing.

Analysis of 3′-end processing products. 3′-end processing reaction mixtures were diluted into 10-μl reaction volumes for treatment with CIP and T4 PNK in a manner similar to that described by Engleman et al. (19). For both CIP and T4 PNK treatments, reactions were done with 1.5 μl of 3′-end-processing reaction mixture. Incubations were at 37°C for 1 h. The products were analyzed as described above for 3′-end-processing assays.

Disintegration assays. Disintegration reactions were performed with several different disintegration substrates. Y-oligomer substrates were of two types, one with MLV U5 viral DNA end-specific sequence in the viral portion of the substrate and one with HIV U5 viral DNA end-specific sequence in the viral portion of the substrate. The HIV Y-oligomer is the same substrate as that used by Chow and coworkers (10). The sequences of the oligonucleotides used to construct the MLV disintegration substrates were as follows: MoDIS1, sequence

TABLE 1. Purification of MLV IN from *E. coli*

Purification step	Total protein (mg)	Total activity (units, 10^3)[a]	Sp act (units, 10^3/mg)	Yield (%)
Lysate	3,080 (estimated)	ND		
105,000 × *g* supernatant	1,500	ND		
Glutathione-agarose pool	19.2	27.2	1.4	100
Phosphocellulose pool	2	485	243	1,790

[a] Activity was quantitated in the dumbbell disintegration assay as described in Materials and Methods. One unit is defined as the amount of enzyme required to convert 1% of 1 pmol of substrate into product in 30 min at 30°C. ND, not done.

5′-CTACAGGTGGGGTCTTTCAACGCGAGGCTAATA-3′; MoDIS2, sequence 5′-AATGAAAGACCCCACCTGTAG-3′; MoDIS3, sequence 5′-CATGGATCGATTCGCG-3′; MoDIS4, sequence 5′-TATTAGCCTCGCGTCGCGAATCGATCCATG-3′ (Protein and Nucleic Acid facility). Both Y-oligomers were prepared, annealed, and purified as previously described (10). A dumbbell substrate similar to that described by Vincent et al. (49) incorporating both target and viral DNA end terminal sequences was also used to assay disintegration activity (oligonucleotide db-y-m, sequence 5′-TGAAAGTTCTTTCAGGCCCTTGGGCCGGCGCTTGCGCC-3′; Operon). The substrate was labeled at the 5′ end, purified, and quantitated. The dumbbell was then annealed in TEN by heating to 95°C followed by slow cooling. Disintegration reactions were carried out with 20 nM GST-MLV IN protein and 100 nM disintegration substrate in a reaction mixture which otherwise duplicates that for the integration assay, with the exception of a KCl concentration of 25 mM. The reactions were carried out at 30°C for 30 min. The products of the disintegration reaction were separated by electrophoresis through 20% polyacrylamide–7 M urea gels at 2,200 V for 1 h, visualized, and quantitated by using a Phosphorimager.

Competition of end processing. For competition experiments, reaction mixtures consisted of 10 nM GST-MLV IN fusion protein, 100 nM synthetic viral DNA end oligonucleotide A194/A193, and competitor molecules in various molar or mass ratios in the standard integration reaction buffer. For comparison of nucleic acid and non-nucleic acid competitors, a mass ratio of from 0- to 25-fold mass excess of competitor over substrate oligonucleotide A194/A193 was used. Competitors tested were unlabeled A194, unlabeled A194/A193, poly(dI-dC) (Boehringer-Mannheim), tRNA (Sigma), ϕX174RFI (New England Biolabs), and heparin (Sigma). In experiments with short oligonucleotide competitors, the range of molar excess of competitor over substrate oligonucleotide ranged from 0- to 60-fold. The double-stranded viral DNA end oligonucleotide A194/A193 and the nonviral competitor oligonucleotide A228/A227 were most frequently used (oligonucleotide A228, sequence 5′-ACCACTACCCGTCAGCGGGGT-3′; complementary oligonucleotide A227, sequence 5′-ACCCCGCTGACGGGTAGTGGT-3′; Operon). Competitor and viral DNA end oligonucleotides were mixed first in the reaction buffer, and then fusion protein was added. Reactions were carried out at 30°C for 30 min, and products were processed for analysis as described above.

For determination of the K_i for the competitor oligonucleotide A228/A227, 3′-end processing was measured in the absence and presence of A228/A227. Labeled viral DNA end substrate A194/A193 at concentrations ranging from 2 to 40 nM; unlabeled competitor A228/A227 at a constant concentration in each experiment, ranging from 0 to 80 nM; and 10 nM GST-MLV IN fusion protein were incubated in a standard integration assay reaction mixture. The reaction mixtures were incubated at 30°C for 20 min. Reactions were performed in triplicate. The amount of starting material that was converted into 3′-end-processed product was analyzed as described above. The K_i for the competitor A228/A227 was calculated from the apparent K_ms determined in the presence and absence of the competitor according to the equation $K_i = [i]/\Delta K_m$, where i = inhibitor (21).

Cross-linking reactions and competitions. Cross-linking was performed in integration reaction buffer in a total volume of 15 μl with 1 pmol of partially purified GST-MLV IN fusion protein and 2 pmol of viral DNA end oligonucleotide substrate labeled at the 3′ end with BrdU, termed A198BrdU/A193. The BrdU-labeled substrate was prepared as follows. The prerecessed viral DNA end oligonucleotide A198 was labeled at the 5′ end with [γ-^{32}P]ATP by using T4 PNK, purified, annealed to a 1.5-fold excess of oligonucleotide A193, and filled in with BrdUTP (Pharmacia) by using the large fragment of *E. coli* DNA polymerase I. The resulting double-stranded viral DNA end oligonucleotide A198BrdU/A193 was purified and quantitated. Reactions were prepared in capless microcentrifuge tubes which were inverted over a 300-nm UV transilluminator (Photodyne, New Berlin, Wis.) and illuminated for 3 to 15 min. SDS-PAGE loading buffer was added, and the products were boiled and loaded onto SDS–10% PAGE gels. Cross-linked products were visualized by autoradiography. Reaction mixtures including competitor were prepared and processed in the same way, with the addition of competitor oligonucleotide A194/A193 or A228/A227 in 0-, 1-, 2-, 5-, and 10-fold excess over A198BrdU/A193 prior to addition of IN.

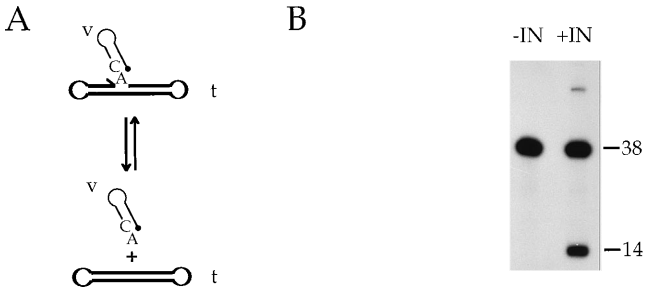


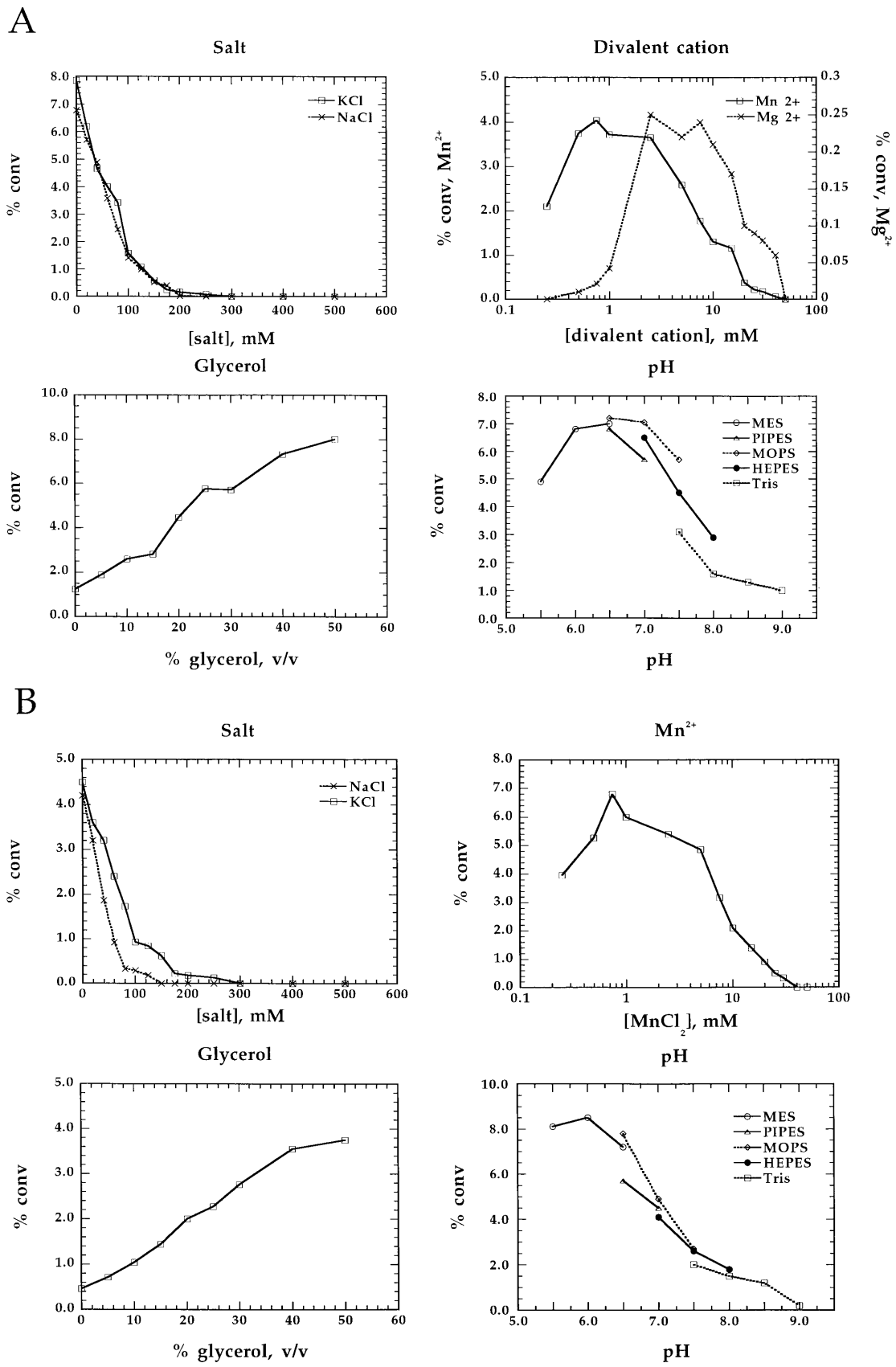
FIG. 5. Disintegration reactions with the MLV dumbbell substrate. (A) Diagram of the MLV dumbbell substrate db-y-m. The single oligonucleotide is labeled at the 5′ end with ^{32}P, symbolized by the filled circle. The 3′ end is symbolized by the half-arrowhead. Viral DNA end sequences including the conserved CA are denoted by thin lines and the letter v, and target sequences are represented by thick lines and the letter t. The dumbbell substrate lacks the two unpaired 5′ nucleotides of the viral DNA generated by end processing, which are not essential for disintegration (9). Incubation of the dumbbell substrate with IN results in the liberation of the viral DNA end hairpin from the target portion of the substrate. This viral DNA end may then undergo reintegration. (B) Disintegration reaction with the dumbbell substrate. −IN, disintegration reaction in the absence of GST-MLV IN fusion protein; +IN, disintegration reaction in the presence of fusion protein. The starting material is the 5′-end-^{32}P-labeled 38-mer, and the product of the reaction is the labeled 14-mer viral DNA end hairpin. The labeled band above the substrate 38-mer is believed to arise from reintegration.

RESULTS

Expression of MoMLV IN as a fusion protein in *E. coli*. The vector used for expression of MoMLV IN is diagrammed in Fig. 1. The pGEX-3X MLV IN vector encodes an amino-terminal 26-kDa fragment of the *Schistosoma japonicum* GST gene under the control of the hybrid *tac* promoter (45). The carboxyl terminus of the GST protein fragment includes a factor Xa cleavage site, and cloning of MoMLV IN into this vector results in a fusion of IN to GST immediately downstream of the factor Xa cleavage site. Fusing IN to GST in this manner results in the alteration of the first three amino acids of IN from Ile Glu Asn to Gly Ile Gln. Less than 5% of the total protein in *E. coli* after induction is found as the full-length GST-MLV IN protein, and an estimated 5% of that protein is found in the soluble fraction of an *E. coli* cell extract after high-speed centrifugation. Induction at 30°C for a period of 3 h resulted in the highest yield of active full-length fusion protein.

Partial purification of GST-MLV IN fusion protein. The soluble protein was absorbed to glutathione-agarose and, after extensive washing, was eluted in buffer C containing reduced glutathione. Figure 2A shows an SDS-PAGE gel of the partially purified protein. The presence of the 72-kDa fusion protein was confirmed by Western blotting (Fig. 2B). While the fusion protein was one of the most abundant proteins in the eluate, several proteins copurified with IN, including proteolytic fragments of the full-length fusion protein (Fig. 2). The protein in this fraction (fraction I) was approximately 10 to 20% pure. This purification step was not substantially improved by increasing the stringency of the glutathione-agarose wash step.

Integration and disintegration assays were performed with glutathione-agarose column-purified protein (Fig. 3B and 4B). In the integration assays, the substrate was a double-stranded


A
Salt
KCl
NaCl
% conv
[salt], mM
Divalent cation
Mn 2+
Mg 2+
% conv, Mn2+
% conv, Mg2+
[divalent cation], mM
Glycerol
% conv
% glycerol, v/v
pH
MES
PIPES
MOPS
HEPES
Tris
% conv
pH
B
Salt
NaCl
KCl
% conv
[salt], mM
Mn2+
% conv
[MnCl2], mM
Glycerol
% conv
% glycerol, v/v
pH
MES
PIPES
MOPS
HEPES
Tris
% conv
pH

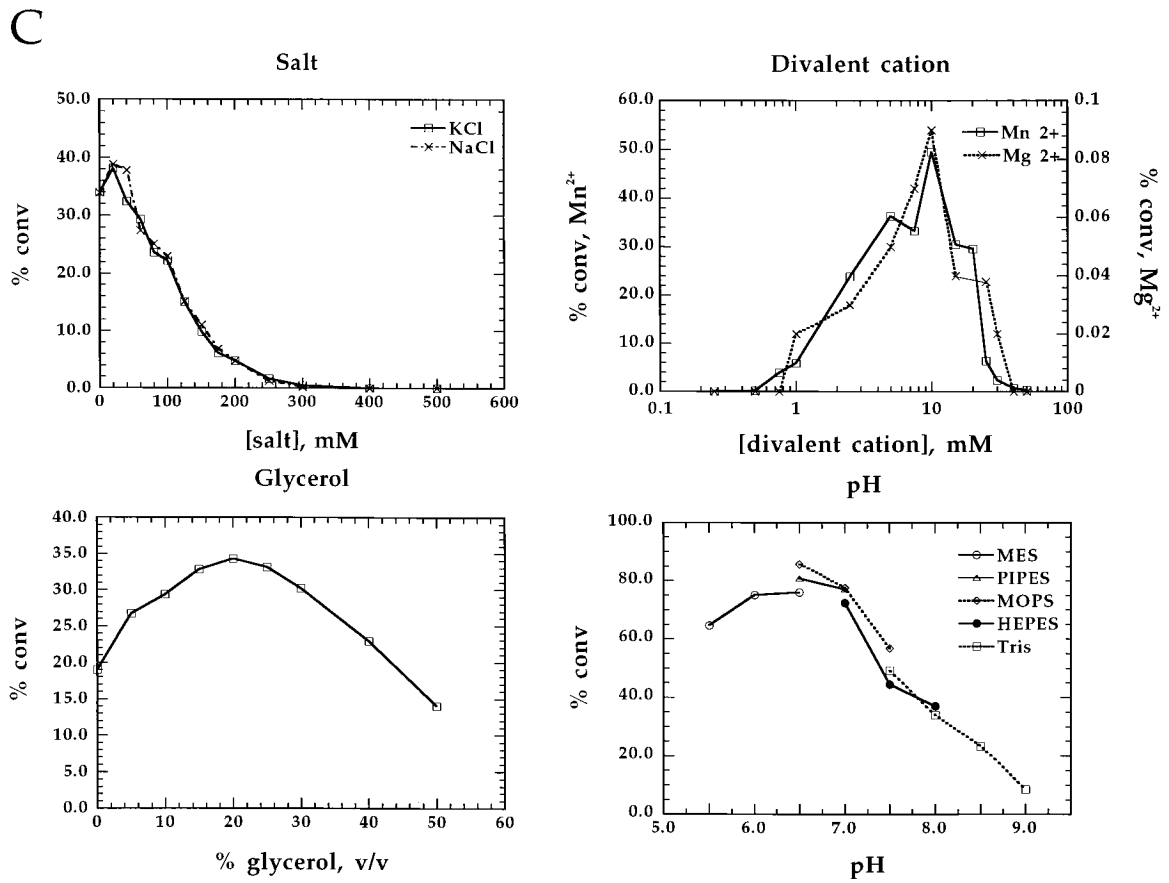


FIG. 6. Effect of reaction conditions on activities of the partially purified GST-MLV IN fusion protein in vitro. (A) 3′-end processing was assayed by using the double-stranded viral DNA end oligonucleotide substrate A194/A193 under standard integration reaction conditions, with the indicated variations. Activity is presented as percent conversion (conv) of substrate to 3′-end-processed product. (B) Integration activity was assayed by using the prerecessed viral end oligonucleotide substrate A198/A193 under standard integration reaction conditions, with the variations indicated. Because of difficulty in accurate quantitation, data are not presented for very weak integration activity in the presence of Mg^{2+} as the sole divalent cation. Activity is presented as percent conversion (conv) of substrate to integration strand transfer product. (C) Disintegration activity was assayed on the MLV dumbbell substrate db-y-m under standard disintegration reaction conditions with variations as indicated. Activity is presented as percent conversion (conv) of dumbbell substrate to disintegration product.

oligonucleotide labeled at one 5′ end. The sequence of the substrate duplicates that of the terminal 23 nucleotides of long terminal repeat DNA sequence from the U5 end of MLV DNA (Fig. 3A). Figure 3B shows the results of an integration reaction performed with a wild-type synthetic viral DNA end oligonucleotide substrate and glutathione-agarose column-purified fusion protein.

The disintegration reaction used the substrate depicted in Fig. 4A. The Y-oligomer substrate mimics the product of a strand transfer reaction with oligonucleotide substrates. Disintegration is the reverse of the strand transfer reaction, resulting in two products, a released viral DNA end and a target DNA (10). We used substrates with viral DNA end sequences corresponding to either MLV (Fig. 4B) or HIV (data not shown) ends. In both the integration and disintegration assays, IN activity coeluted with the GST-MLV IN fusion protein and the preparation was free of nuclease contamination that might interfere with analysis of IN activities.

The activity of the fusion protein after purification on glutathione-agarose was variable. In some cases, the protein was fully active; in others, only weak activity could be detected in integration or disintegration assays. In such cases, activity could be recovered through chromatography of the glutathione-agarose-purified protein (fraction I) on phosphocellulose to yield fraction II protein (Table 1), which resulted in the removal of an inhibitor of enzyme activity. The presence of this inhibitor was confirmed by mixing experiments (data not shown). Most experiments were performed with a fraction I preparation that was highly active. Purification of the fusion protein beyond 70% purity was limited by copurifying proteins, which were found in a high-molecular-weight fraction with IN by gel exclusion chromatography (data not shown). Further purification may require more stringent conditions to remove the contaminating proteins.

The fusion protein could also be solubilized from the insoluble fraction by extraction of a low-speed spin (12,000 × *g*) pellet with 8 M urea followed by slow dialysis to remove urea; however, the specific activity of the protein purified in this manner was significantly less than that of the protein purified from the soluble fraction (data not shown). Experiments with factor Xa-treated fusion protein showed no discernable difference in activity between the cleaved and uncleaved proteins (data not shown).

Effect of reaction conditions on the integration and disintegration reactions catalyzed by MLV IN. The effects of salt, divalent cation, and glycerol concentrations; pH; and reaction temperature on 3′-end processing, strand transfer, and disintegration activities of MoMLV IN were investigated

(see Fig. 6). Substrates used to study integration activity were those described and shown in Fig. 3. Disintegration assays used the substrate shown in Fig. 5A. The substrate is a single oligonucleotide of 38 residues with self-complementary sequences that allow the DNA to fold into a structure resembling the Y-oligomer disintegration substrate. This substrate is termed the dumbbell substrate (9, 49) and contains 6 bp of viral DNA end sequence and 10 bp of target DNA sequence. The products of a disintegration reaction of a 5′-end-labeled dumbbell are a labeled viral DNA end 14-mer and an unlabeled target 24-mer (Fig. 5B). Reactions were allowed to proceed for 30 min, during which the reactions were linear with time. Thus, the effects of the various conditions reflected the effects on initial rates and not on the extent of reaction.

(i) End processing and strand transfer. The maximum end processing and integration activity was observed in the absence of added KCl or NaCl. Activity decreased with increasing salt concentrations and was undetectable at salt concentrations higher than 250 mM. (Fig. 6A and B).

Both integration and 3′-end processing activities were completely dependent on a divalent cation in the reaction mixture, with Mn^{2+} supporting a much higher level of activity than Mg^{2+} (Fig. 6A and B). 3′-end processing was approximately 20-fold higher with Mn^{2+} than with Mg^{2+}. The difference in strand transfer activity supported by the two divalent cations was even greater, such that the small amounts of integration products obtained in the presence of Mg^{2+} could be visualized but were difficult to quantitate. The activity of the fusion protein in the presence of Mg^{2+} differs from that reported previously (13, 30). The optimal concentration of Mg^{2+} (2.5 to 10 mM) is slightly higher than that for Mn^{2+} (0.5 to 5 mM) for both activities. No activity was observed when other divalent cations (Zn^{2+}, Ca^{2+}, and Co^{2+}) were substituted (data not shown).

The optimal pHs for 3′-end processing and strand transfer were pH 6.5 and 6.0, respectively (Fig. 6A and B). Both activities were detectable at the lowest pH tested (pH 5.5) and at the highest pH tested (pH 9.0). The highest concentration of glycerol tested (50% [vol/vol]) resulted in the greatest activity for both end processing and strand transfer (Fig. 6A and B). The inclusion of BSA in the reaction mixture resulted in increased IN activity at a low enzyme concentration. End processing and strand transfer were observed at temperatures from 15 to 60°C, with maximal activity at 42°C.

(ii) Disintegration. The optimal conditions for the disintegration reaction with the dumbbell substrate were similar to those found for end processing and strand transfer (Fig. 6C). The disintegration activity of the enzyme was maximal at low NaCl and KCl concentrations, with the peak activity at a salt concentration of 25 mM. The disintegration reaction was completely dependent on the presence of a divalent cation, with a marked preference for Mn^{2+} over Mg^{2+}. The difference in activity observed when Mn^{2+} was used instead of Mg^{2+} was even greater than that seen for end processing, with the maximal activity supported by Mn^{2+} being 500-fold higher than that supported by Mg^{2+}. The optimal concentration of divalent cation was from 5 to 15 mM, with a peak at 10 mM, and very similar concentration dependence profiles were observed for the two divalent cations. The pH optimum for disintegration, pH 6.5, was similar to that for end processing and strand transfer, as was the pH range over which activity was observed. The range of temperatures over which the fusion protein supported dumbbell disintegration was from 10 to 50°C, with an optimum at 37 to 42°C. The preferred concentration of glycerol for the disintegration reaction differed from that observed for the end-processing and integration reactions, as activity peaked at 20% glycerol (vol/vol).

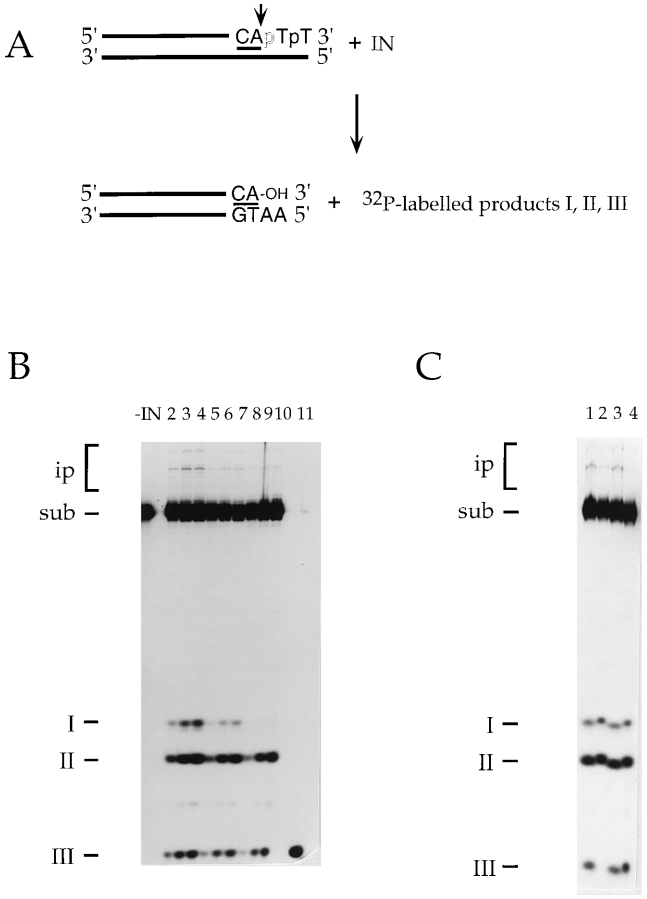


FIG. 7. 3′-end-processing activity of GST-MLV IN protein. (A) Schematic of the assay for the products of end processing. The viral DNA end oligonucleotide substrate A198tt/A193 was ^{32}P labeled at the 3′ end at the phosphate adjacent to the penultimate T nucleotide, shown as a letter p in outline. The arrowhead indicates the position at which endonucleolytic cleavage occurs to release labeled products of end processing. (B) Generation of products of end processing in the presence of decreasing amounts of glycerol. End-processing reactions were performed under standard integration reaction conditions, with the exception of variation in glycerol concentration, as indicated. −IN indicates the reaction in the absence of GST-MLV IN. Lanes are ordered in three groups of three. The first group of three, lanes 2, 3, and 4, contain end-processing reaction mixtures, sampled after 15, 30, and 60 min of reaction time, respectively, in the presence of 20% (vol/vol) glycerol. Lanes 5, 6, and 7 contain reaction mixtures sampled at the same intervals in the presence of 10% [vol/vol] glycerol. Lanes 8, 9, and 10 contain reaction mixtures sampled at the same intervals but with only trace amounts of glycerol from the enzyme storage buffer. Lane 11 contains the labeled dinucleotide standard $^{32}pTpT_{OH}$. Three labeled products of end processing, designated products I, II, and III, are released from the substrate (sub) in the presence of IN. Integration products are indicated as "ip". (C) Treatment of end-processing reaction mixtures with CIP and T4 PNK. End-processing reactions were performed, and aliquots were then treated with either CIP or T4 PNK. Lane 1, end-processing reaction products without CIP treatment; lane 2, end-processing reaction products treated with CIP (label is lost from product III with this treatment); lane 3, end-processing reaction without T4 PNK treatment; lane 4, end-processing reaction mixture treated with T4 PNK and ATP. Abbreviations are the same as those for panel B.

MLV IN generates multiple products of 3′-end processing. We analyzed the products of 3′-end processing generated by MLV IN using a model viral DNA end substrate labeled at the 3′ end (Fig. 7A). In the standard reaction buffer, which contains glycerol, three products of end processing were generated (Fig. 7B). One of the products, product III, ran

exactly coincident with the expected hydrolysis product, the dinucleotide 5′-pTpT$_{OH}$. Two products of slower mobility were also generated in this reaction, product I and product II. Analogy to the products generated by HIV and FIV IN end processing (19, 50, 51) suggests that these are a glycerol-linked pTpT$_{OH}$ dinucleotide and a 3′-5′ cyclic TT dinucleotide, respectively. Consistent with these structures, both products I and II were resistant to phosphorylation by T4 PNK, indicating the lack of a free 5′ hydroxyl in either product (Fig. 7C, lane 4). Both products were also resistant to treatment with CIP, indicating the lack of a free 5′ phosphate group. Only product III, the linear dinucleotide, labeled at the 5′ phosphate, was sensitive to CIP treatment (Fig. 7C, lane 2).

While the abundance of product I was linked to the amount of glycerol present in the reaction mixture, the other products of end processing were generated in the absence of glycerol. The slight reduction in their abundance with decreasing glycerol concentration is due to an overall reduction in 3′-end-processing activity. The glycerol dependence of product I is consistent with it being a glycerol-linked TT dinucleotide. The resistance of product II to treatment with PNK and CIP, its production in the absence of glycerol, and its mobility relative to that of products I and III parallel the characteristics of the 3′-5′ cyclic dinucleotide product of HIV-1 end processing (19). Additionally, to test if the terminal 3′ hydroxyl was necessary for formation of product II, reactions were carried out with a model viral DNA substrate lacking a 3′ hydroxyl on the terminal T nucleotide. Product II was not generated in these reactions, while product I and product III, the linear dinucleotide, were made in normal amounts (data not shown). Therefore, product II is presumed to be a 3′-5′ cyclic TT dinucleotide. The structures of products I and II were not further characterized.

The predominant product in these reactions was product II, the putative 3′-5′ cyclic TT dinucleotide, accounting for approximately one-half of the total products of end processing. The linear dinucleotide pTpT$_{OH}$ accounted for approximately 30% of the total product, with the remainder being the glycerol-linked product I. The three products were formed with indistinguishable kinetics (data not shown). Thus, with the wild-type synthetic viral DNA end substrate under the conditions of the in vitro reaction, the preferred nucleophile for end processing is the 3′ hydroxyl at the terminus of the viral DNA substrate.

Viral 3′-end processing kinetics. Even after prolonged incubation with a large molar excess of a 23-mer viral DNA end oligonucleotide substrate, GST-MLV IN fusion protein failed to turn over in either the 3′-end processing or the strand transfer reaction. During the linear portion of the reaction time course, in which IN was present at a concentration of 10 nM and substrate was present at a 10-fold-higher concentration, less than 1% of the substrate was converted into product. Thus, less than 0.1 pmol of product was formed for each pmol of IN monomer in 30 min (data not shown). Nevertheless, substrate saturation occurred as the concentration of viral DNA end substrate was increased (Fig. 8A). The apparent K_m for the blunt 23-mer viral DNA end oligonucleotide substrate was 6 ± 2 nM, determined with both 5′- and 3′-end-labeled substrates (Fig. 8B). A detailed interpretation of these results will require determination of the stoichiometry of the active IN enzyme and the kinetic parameters for individual steps in the reaction, such as the dissociation rate for products.

Jones and coworkers (29) reported that RSV IN is capable of multiple turnover in the 3′-end processing and strand transfer reactions using model viral DNA end oligonucleotide substrates. In reactions with GST-MLV IN fusion protein under conditions approximating those used by Jones and coworkers, turnover was not observed (data not shown).

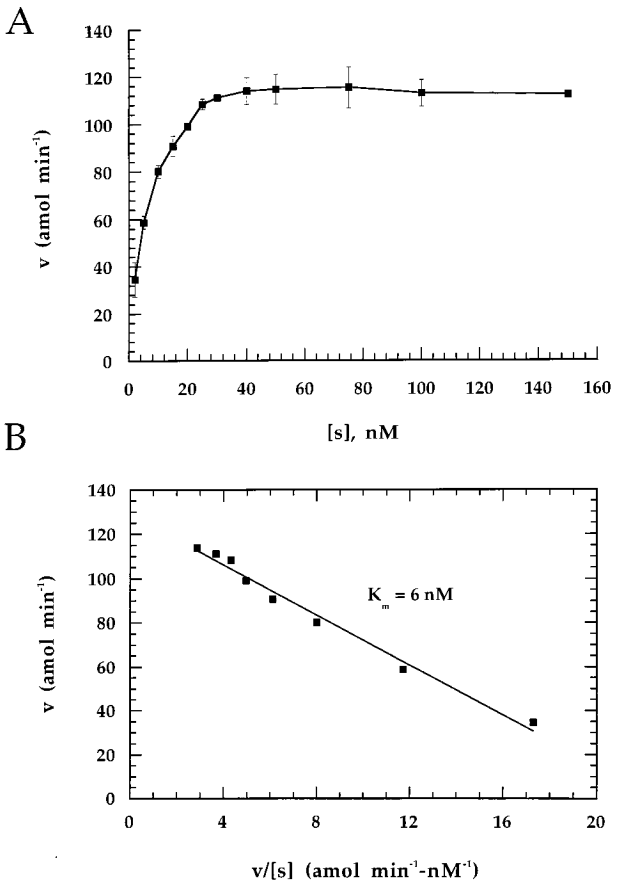


FIG. 8. 3′-end-processing activity substrate kinetics. (A) Reaction velocity (v) versus substrate concentration ([s]) plot for end processing. Integration reactions were performed with a range of concentrations from 2 to 100 nM either 5′ ^{32}P-labeled A194/A193 or 3′ ^{32}P-labeled A198tt/A193 viral DNA end oligonucleotide substrate and 10 nM GST-MLV IN and quantitated. Datum points are the averages for three determinations. Error bars indicate standard deviations. (B) Determination of apparent substrate K_m. The Eadie-Hofstee plot is from the data in panel A. Abbreviations are as for panel A.

MLV IN is an enzyme and is capable of turnover. We used disintegration of the MLV dumbbell substrate to investigate further the possibility that MLV IN is capable of turnover. We first determined an apparent K_m of 11 nM for the dumbbell substrate, to ensure that the dumbbell was saturating for IN in turnover experiments (Fig. 9A). MLV IN was incubated with a 20-fold molar excess of the dumbbell disintegration substrate, and the release of viral DNA end product was monitored over time. GST-MLV IN was capable of turnover, with a turnover value of approximately 0.035 min^{-1} per IN monomer during the first 3 h of the reaction (Fig. 9B). This turnover rate is similar to that reported for HIV IN with the dumbbell substrate (9). The ability of MLV IN to catalyze the reverse reaction of integration and to turn over establishes that MLV IN is an enzyme.

IN has distinct binding sites for viral and target DNAs. IN functionally discriminates between viral DNA ends and nonviral DNA, since only viral DNA ends, or oligonucleotides with sequences derived from viral ends, can serve as substrates for the 3′-end processing reaction or provide the attacking 3′ hydroxyl for strand transfer, while virtually any DNA can serve as a target for integration (5, 34, 35, 42). The ability of various

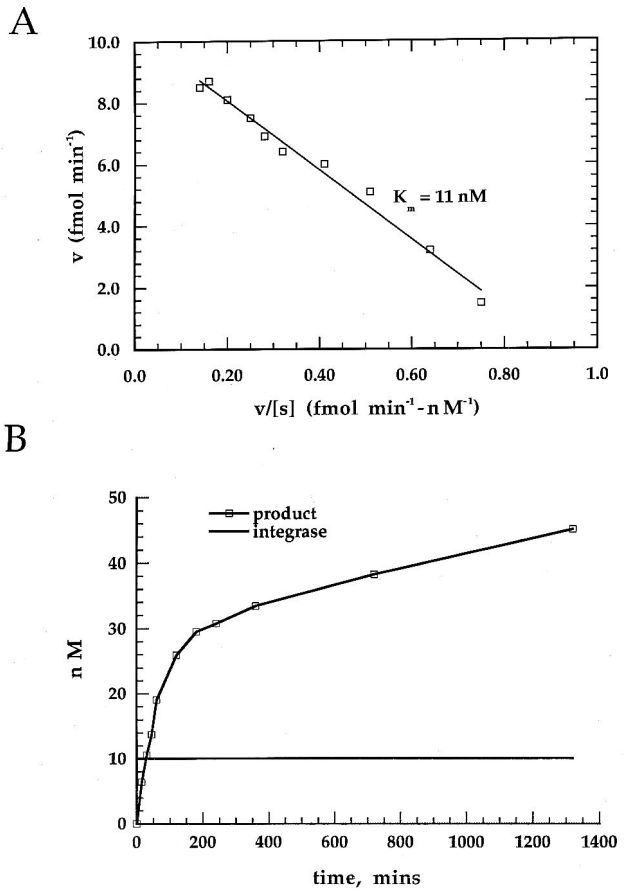


FIG. 9. Kinetic experiments with the MLV dumbbell substrate. (A) Determination of K_m for the dumbbell substrate db-y-m. Disintegration reactions were performed with a range of dumbbell substrate concentrations from 2 to 100 nM and 10 nM GST-MLV IN. Datum points are averages for three determinations. The data were then transformed into an Eadie-Hofstee plot to obtain the K_m. (B) Turnover of MLV IN. A 10 nM concentration of partially purified GST-MLV IN was incubated with 200 nM dumbbell substrate, and the extent of the reaction was quantified at each time point. v, reaction velocity; [s], substrate concentration.

DNAs and other molecules to compete for viral or target DNA binding was tested in a functional assay (Fig. 10). Double-stranded DNA and heparin were effective inhibitors of end processing and strand transfer, while single-stranded DNA and RNA were poor inhibitors. Some oligonucleotide competitors inhibited end processing only weakly yet were used efficiently as targets for strand transfer. Figure 11 compares inhibition of end processing and strand transfer by the nonspecific oligonucleotide A228/A227 and oligonucleotide A194/A193, which, except for being unlabeled, is identical to the viral DNA end substrate. Little inhibition of either 3′-end processing or strand transfer was observed in the presence of up to a 10-fold molar excess of the nonspecific oligonucleotide over labeled viral DNA end substrate. At a greater molar excess of the nonspecific oligonucleotide, inhibition was observed but residual activity was seen even in the presence of a 60-fold molar excess of this DNA. In contrast, significant inhibition by unlabeled viral DNA end oligonucleotide was observed even at low molar ratios to labeled viral DNA end substrate, with the amount of inhibition proportional to the relative amount of unlabeled to labeled viral DNA end. End processing and strand transfer of the labeled substrate were almost completely inhibited at a fivefold molar excess of unlabeled viral DNA end substrate. Thus, the nonspecific oligonucleotide A228/A227 competes poorly with viral DNA end substrates for the viral DNA binding site. A measure of the capacity of A228/A227 to compete for the viral DNA end binding site was obtained by measuring the K_m for the viral end oligonucleotide substrate in the presence and absence of a fixed concentration of the competitor oligonucleotide (Fig. 12). A228/A227 behaved as a simple competitive inhibitor of end processing, with a K_i of approximately 55 nM at fixed concentrations of 40 and 80 nM. This K_i is in contrast to the K_m for end processing with the wild-type viral DNA end oligonucleotide A194/A193, which is approximately 10 nM in the absence of competitor. In contrast to the specificity of the viral DNA binding site, nonspecific and viral DNA end oligonucleotides were similar in their ability to serve as targets for strand transfer. As the molar ratio of nonspecific oligonucleotide to viral DNA end substrate was increased, there was a clear shift toward integration products corresponding to strand transfer into the nonspecific competitor rather than the viral end DNA (Fig. 11).

The simplest explanation for these data is that IN has a viral DNA binding site that binds preferentially to DNAs that resemble the viral DNA end and a second binding site for target DNA that does not discriminate between viral and nonviral DNA sequences. If these two functionally distinct sites represent alternative states of a single site, there must be little interconversion between the two states on the time scale of these experiments.

Cross-linking of DNA to MLV IN. To further define the DNA binding characteristics of IN, we used an oligonucleotide corresponding in sequence to a viral DNA end, labeled at the 3′ end with BrdU, which allows photoreactive cross-linking to protein. Figure 13 shows that the GST-MLV IN fusion protein could be cross-linked to this oligonucleotide substrate by UV irradiation at 300 nm. Two major products appeared as bands with estimated molecular masses of 80 and 140 kDa. These molecular weights are consistent with monomeric and dimeric forms, respectively, of the fusion protein cross-linked to substrate. Cleavage of the cross-linked fusion protein-substrate complexes by factor Xa resulted in a single major labeled band at approximately 55 kDa, which is near the expected approximate molecular mass of unfused MLV IN (46 kDa) plus the oligonucleotide (15 kDa) (Fig. 13, lane 6). Thus, the cross-linked proteins were GST-MLV IN. While the basis for the apparent linkage of two GST-MLV IN polypeptides in the 140-kDa product is unknown, its sensitivity to factor Xa suggests that the linkage of the two polypeptides lies in the GST domain. Cross-linking occurred in the absence of Mn^{2+} or in the presence of 1 mM EDTA and was sensitive to salt concentration.

The binding specificity of this BrdU-labeled DNA substrate was analyzed by comparing the abilities of the viral DNA end A194/A193 oligonucleotide and the nonspecific A228/A227 oligonucleotide to inhibit cross-linking (Fig. 13, lanes 7 to 16). With the two competitors at equivalent molar excesses, the reduction in cross-linking was similar whether the competitor was a viral DNA end or a nonspecific oligonucleotide, as measured by densitometry. Thus, the inferred binding site(s) at which the majority of cross-linking occurred differs in its apparent binding specificity from the site that binds viral DNA for end processing and strand transfer. The lack of apparent specificity for viral DNA at the cross-linking site suggests that, although the photoreactive substrate was designed to mimic a viral DNA end, cross-linking may be occurring predominantly at the target DNA binding site.

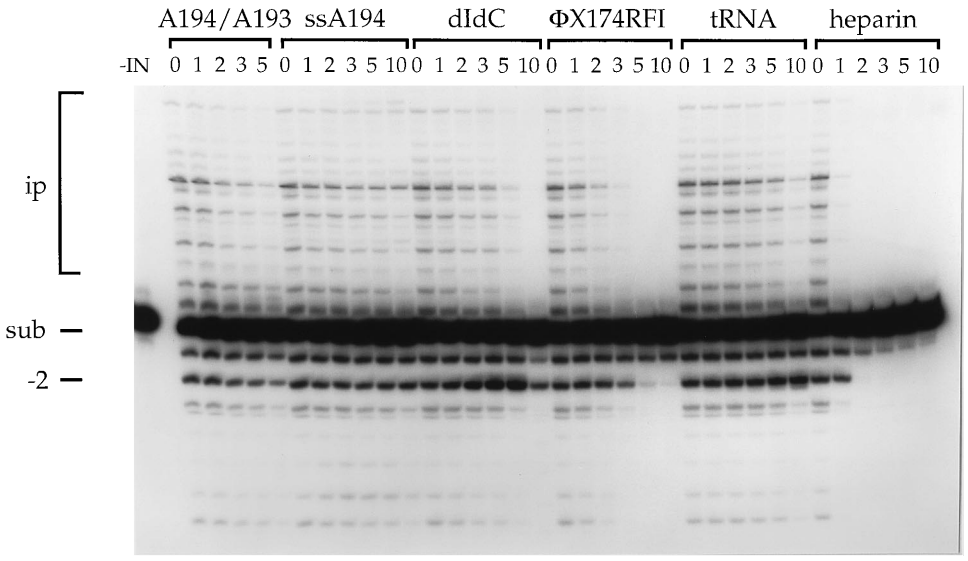


FIG. 10. 3′-end processing and strand transfer competition. Reactions were performed with the labeled viral DNA end oligonucleotide substrate A194/A193 in the presence of unlabeled competitors as indicated. The mass ratio of unlabeled competitor to 5′ ^{32}P-labeled viral end substrate is indicated above the lanes. Competitors were the double-stranded viral end oligonucleotide A194/A193, single-stranded (ss) viral end oligonucleotide A194, poly(dI-dC) (dIdC), double-stranded ϕX174 plasmid DNA, tRNA, and heparin. −IN indicates a reaction without GST-MLV IN. The positions of the labeled viral DNA end substrate (sub), end processing product (−2), and integration products (ip) are indicated.

DISCUSSION

We have produced soluble, active MoMLV IN as a fusion protein with GST in *E. coli*. Previously published studies of active recombinant MoMLV IN have employed extraction of the protein from an insoluble fraction under denaturing conditions followed by refolding (13, 30). Like these MLV INs, partially purified GST-MLV IN fusion protein possesses 3′-end processing, strand transfer, and disintegration activities. Active protein suitable for biochemical analysis is obtained after a maximum of two chromatography steps. While the fusion protein may be cleaved with factor Xa to release IN, we typically used the GST-MLV IN form of IN, as the fusion protein and the factor Xa-cleaved, unfused protein exhibit very similar activities.

All of the activities of MoMLV IN were found to be absolutely dependent on the presence of a divalent cation, either Mn^{2+} or Mg^{2+}. The data bear out a marked dependence of the fusion protein on Mn^{2+} for efficient strand transfer, both for integration and disintegration. In contrast, 3′-end processing activity was easily detected with either Mn^{2+} or Mg^{2+}. Previous reports of the activity of recombinant MoMLV IN indicated that catalysis by this protein was supported only by Mn^{2+} (13, 30). Similar differences in activity supported by the two divalent cations have been reported for purified recombinant HIV-1 IN (6, 34, 43, 49) and ASLV IN purified from virions (31). In contrast to the results for these purified INs, Mg^{2+} satisfies the divalent-cation requirement for IN in integration complexes from MLV-infected cells (3, 23, 24) and HIV-1-infected cells (16, 20), as well as for purified AMV IN from virions (22, 32, 52). The basis for the differences in divalent cation preference between purified IN and integration complexes is unknown. A second important difference between the fusion protein and the MoMLV integration complex is sensitivity to salt concentration. The GST-MLV IN is more sensitive to salt than the integration complex, which remains active at salt concentrations that the fusion protein and other recombinant MLV IN preparations do not tolerate (3). In general, the fusion protein and the MoMLV integration complex are active over similar pH ranges.

Analysis of the products of 3′-end processing catalyzed by MoMLV IN on model viral DNA end oligonucleotide substrates labeled at the 3′ end revealed three distinct products of the in vitro reaction. The linear dinucleotide $pTpT_{OH}$ is the expected hydrolysis product from an endonucleolytic cleavage reaction. The two additional products result from cleavages in which a glycerol hydroxyl or the terminal 3′ hydroxyl of the

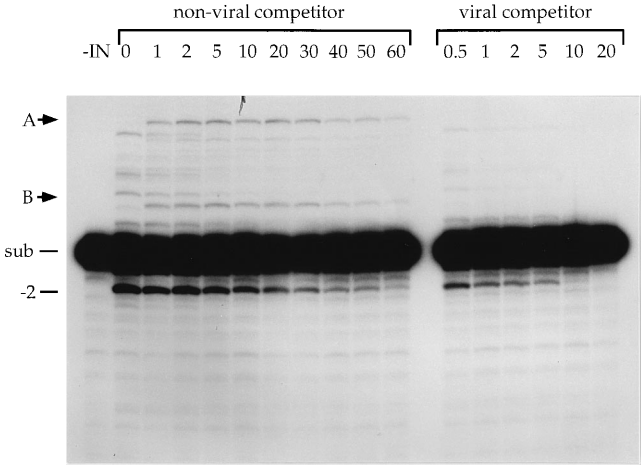


FIG. 11. 3′-end processing and strand transfer competition by viral DNA end oligonucleotide A194/A193 and nonviral oligonucleotide A228/A227. Standard integration reactions with the 5′ ^{32}P-labeled viral DNA end oligonucleotide substrate A194/A193 were carried out in the presence of competitor oligonucleotides at the molar ratios to viral DNA end substrate A194/A193 indicated above the lanes. −IN denotes a reaction without GST-MLV IN. The positions of the substrate (sub) and end-processing product (−2) are as shown. A indicates the position of a product of integration into the nonviral competitor oligonucleotide as target, while B indicates the position of a product of integration into a viral DNA end oligonucleotide as target.

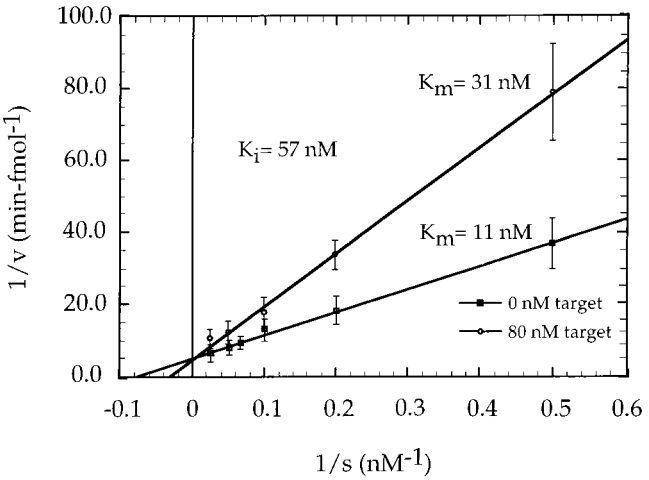


FIG. 12. Determination of the K_i for the nonviral oligonucleotide competitor A228/A227. A Lineweaver-Burk plot of end processing competition data is shown. Standard integration reactions with the viral DNA end oligonucleotide A194/A193 were carried out in the presence or absence of 80 nM A228/A227 as competitor. The K_m for end processing was determined in the presence and absence of competitor, and the K_i was determined from the respective K_ms. The datum points in the Lineweaver-Burk plot represent the averages from three experiments. Error bars indicate standard deviations. v, reaction velocity; s, substrate concentration.

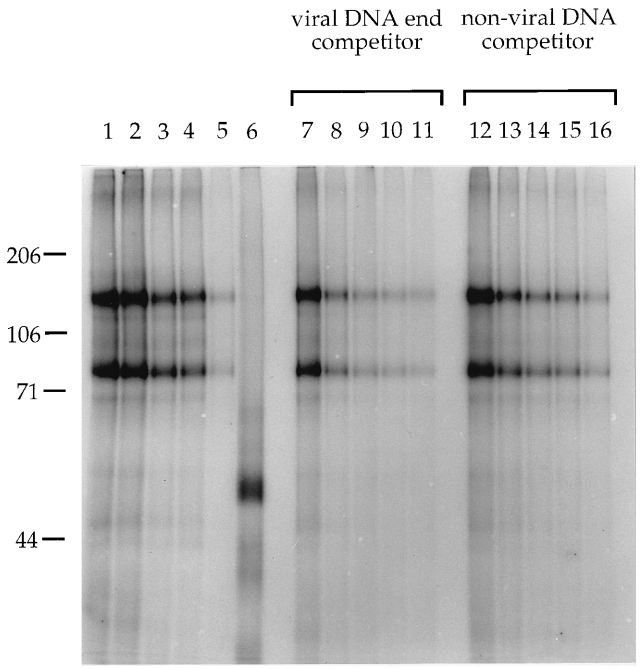


FIG. 13. Cross-linking competition. Cross-linking reactions with 5′ ^{32}P-labeled and 3′ BrdU-labeled viral DNA end oligonucleotide substrate A198 BrdU/A193 and partially purified GST-MLV IN were carried out in the absence and presence of competitor. The reaction mixtures were then electrophoresed on SDS–10% PAGE gels and visualized by autoradiography. Positions of migration of molecular mass standards are indicated in kilodaltons to the left. Lanes 1 through 6 show cross-linking in the absence of competitor with the variations from standard cross-linking conditions as indicated. Lanes: 1, standard integration reaction and cross-linking conditions; 2, 10 mM $MnCl_2$; 3, 0 mM $MnCl_2$; 4, 0 mM $MnCl_2$ and 1 mM EDTA; 5, 100 mM NaCl; 6, factor Xa cleavage of proteins subsequent to cross-linking. Lanes 7 through 11 show cross-linking in the presence of unlabeled viral DNA end oligonucleotide A194/A193 as competitor. Lanes: 7, standard cross-linking reaction without competitor; 8 through 11, 1×, 2×, 5×, and 10× excess, respectively, of unlabeled A194/A193 competitor. Lanes 12 through 16 show cross-linking in the presence of unlabeled nonviral oligonucleotide A228/A227 as competitor. Lanes: 12, standard cross-linking reaction without competitor; 13 through 16, 1×, 2×, 5×, and 10× excess, respectively, of unlabeled A228/A227. The band at 80 kDa is consistent with the expected size of a single GST-MLV IN protomer attached to the oligonucleotide, while the band at 140 kDa is consistent with the expected size of two GST-MLV IN proteins and the oligonucleotide.

substrate is used as the nucleophile instead of a water hydroxyl. Regardless of which nucleophile is used, the resulting processed viral DNA end has the same structure. This diversity in the structural context of the attacking hydroxyl group suggests that the catalytic site of IN has great flexibility in accommodating this substrate. Moreover, the similarity between the products of end processing by MoMLV, HIV (19, 51), and FIV (50) INs argues that this flexibility is a common theme of 3′-end processing. The flexibility is not limited to end processing; a 2′ hydroxyl can substitute for a 3′ hydroxyl as a nucleophile in the disintegration reaction (9, 15). Indeed, the use of the terminal 3′ hydroxyl as the nucleophile in end processing would necessitate distortion of the double-helical structure of the DNA at the terminus to allow access of the nucleophile to the penultimate phosphodiester bond. Consistent with this suggestion is the observation that IN is tolerant of distortion in the double-helical structure of substrates in end processing (41a) and in disintegration (9). The flexibility of IN for nucleophile and DNA substrate structure may be a consequence of a single catalytic domain catalyzing multiple distinct reactions. The substrates in 3′-end processing, strand transfer, and disintegration are likely to impose different requirements and constraints on the scissile phophodiester bond and the nucleophile in catalysis. Substrate distortion and promiscuity in accommodating the attacking hydroxyl group could potentially solve this catalytic problem by loosening the constraints that would otherwise be imposed by substrate structure.

Of the three donors of the hydroxyl nucleophile in MoMLV 3′-end processing, the terminal 3′ hydroxyl of the viral DNA end was preferred; thus, the most abundant product of end processing was the cyclic dinucleotide. This is in contrast with HIV IN 3′-end processing, where the 3′-5′ cyclic nucleotide is a minor product (19, 51). The relative usage of the 3′ hydroxyl can be increased by site-directed mutagenesis of the D,D(35)E region in the catalytic core domain of IN (7, 18, 47). Specific mutations in residues near D-116 and E-152 in the HIV INs result in a dramatic increase in the relative abundance of the cyclic dinucleotide. Comparison of the D,D(35)E regions of HIV and MoMLV INs reveals that, of the six sites in HIV IN which when mutated led to an increase in the relative abundance of the cyclic nucleotide, four share identity (amino acids 117, 143, 146, and 147) with MoMLV IN and two differ (residues 114 and 148 for HIV and residues 185 and 219 for MoMLV IN). Alteration of residue 148 in HIV-2 IN to the corresponding amino acid in MoMLV IN results in a mild reduction in end processing and converts the 3′ hydroxyl into the preferred nucleophile (47). It is possible that the differences in nucleophile selection between MoMLV IN and HIV IN may in part be explained by the differences in amino acid identity at these positions.

MLV IN failed to turn over in the end processing or strand transfer reaction, yet it catalyzed multiple reactions per IN protomer in the dumbbell disintegration reaction. To date, only RSV IN has been shown to be capable of turnover in the forward reaction (29), while HIV-1 IN has been shown to turn over in reactions using the dumbbell substrate (9). It has recently been shown that 3′-end processing and strand transfer are mediated by a stable complex between an IN multimer and the viral DNA end (17). This stable complex processively

catalyzes 3′-end processing and strand transfer. The complex is greatly stabilized by the presence of the two unpaired 5′ nucleotides at the viral DNA end after end processing. The stable association of IN with a viral DNA end could make product release rate limiting and thus preclude binding of additional substrates and catalytic turnover. The ability of IN to turn over in the disintegration of the dumbbell substrate may therefore be due to the small size of the product and the lack of the two unpaired nucleotides at the 5′ end of the processed viral DNA end. Both of these deficiencies would destabilize IN-DNA interactions, allowing relatively rapid product release and subsequent rounds of catalysis (9).

IN discriminates functionally between viral DNA end substrates and nonviral DNAs in the 3′-end-processing and strand transfer reactions. Competition experiments described here suggest that IN can bind with some specificity to viral DNA ends. We found that some double-stranded oligonucleotides competed weakly for viral DNA end binding in comparison to authentic viral DNA end substrates yet functioned equivalently to viral DNA ends as targets for strand transfer. These results suggest that IN has distinct sites for binding DNAs, one that discriminates between viral DNA ends and a nonviral DNA competitor and one that does not. We also found that cross-linking of a viral DNA end substrate to IN was equally inhibited by competition with viral DNA end and nonviral oligonucleotides, suggesting that cross-linking occurred at a site with characteristics of a nonspecific DNA binding site. Taken together, these data are evidence that IN has two distinct DNA binding sites, one that binds preferentially to viral DNA ends and a second that binds target DNA with little specificity. Further studies will be required to clarify the identity and specificity of these substrate binding sites.

ACKNOWLEDGMENTS

We thank Samson Chow, Viola Ellison, and other members of the Brown laboratory; Karl Reich; and members of the Varmus laboratory for helpful discussions and comments.

This work was supported by the Howard Hughes Medical Institute and NIH grant AI27205. P.O.B. is an assistant investigator of the Howard Hughes Medical Institute. I.D. was supported by the Howard Hughes Medical Institute. B.P.S. is a trainee in the Medical Scientist Training Program supported by the National Institute of General Medical Sciences. T.S.H. is supported by a predoctoral training grant from the National Institutes of Health.

REFERENCES

1. **Bradford, M. M.** 1976. A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. Anal. Biochem. **72:**248–254.
2. **Brown, P. O.** 1990. Integration of retroviral DNA. Curr. Top. Microbiol. Immunol. **157:**19–48.
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.** 1990. Sequence requirements for integration of Moloney murine leukemia virus DNA in vitro. J. Virol. **64:**5645–5648.
6. **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.
7. **Bushman, F. D., A. Engleman, I. Palmer, P. Wingfield, and R. Craigie.** 1993. Domains of the integrase protein of human immunodeficiency virus type 1 responsible for polynucleotidyl transfer and zinc binding. Proc. Natl. Acad. Sci. USA **90:**3428–3432.
8. **Bushman, F. D., T. Fujiwara, and R. Craigie.** 1990. Retroviral DNA integration directed by HIV-1 integration protein *in vitro*. Science **249:**1555–1558.
9. **Chow, S. A., and P. O. Brown.** 1994. Substrate features important for recognition and catalysis by human immunodeficiency virus type 1 integrase identified by using novel DNA substrates. J. Virol. **68:**3896–3907.
10. **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.
11. **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.
12. **Colicelli, J., and S. P. Goff.** 1988. Sequence and spacing requirements of a retrovirus integration site. J. Mol. Biol. **199:**47–59.
13. **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.
14. **Donehower, L., 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.
15. **Donzella, G., and M. J. Roth.** 1993. Influence of substrate structure on disintegration activity of Moloney murine leukemia virus integrase. J. Virol. **67:**7077–7087.
16. **Ellison, V., H. Abrams, T. Roe, J. Lifson, and P. O. Brown.** 1990. Human immunodeficiency virus integration in a cell-free system. J. Virol. **64:**2711–2715.
17. **Ellison, V., and P. O. Brown.** 1994. A stable complex between integrase and viral DNA ends mediates human immunodeficiency virus integration *in vitro*. Proc. Natl. Acad. Sci. USA **91:**7316–7320.
18. **Engleman, A., and R. Craigie.** 1992. Identification of conserved amino acid residues critical for human immunodeficiency virus type 1 integrase function in vitro. J. Virol. **66:**6361–6369.
19. **Engleman, A., K. Mizuuchi, and R. Craigie.** 1991. HIV-1 DNA integration: mechanism of viral DNA cleavage and DNA strand transfer. Cell **67:**1211–1221.
20. **Farnet, C. M., and W. A. Haseltine.** 1990. Integration of human immunodeficiency virus type 1 preintegration complex. Proc. Natl. Acad. Sci. USA **87:**4164–4168.
21. **Ferscht, A.** 1984. Enzyme structure and kinetics. W. H. Freeman & Co., New York.
22. **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.
23. **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.
24. **Fujiwara, T., and K. Mizuuchi.** 1988. Retroviral DNA integration: structure of an integration intermediate. Cell **54:**497–504.
25. **Goff, S. P.** 1992. Genetics of retroviral integration. Annu. Rev. Genet. **26:**527–544.
26. **Grandgenett, D. P., and S. R. Mumm.** 1990. Unravelling retroviral integration. Cell **60:**3–4.
27. **Hanahan, D.** 1983. Studies on transformation of *Escherichia coli* with plasmids. J. Mol. Biol. **166:**557–563.
28. **Hippenmeyer, P. J., and D. P. Grandgenett.** 1985. Mutants of the Rous sarcoma virus reverse transcriptase gene are defective in early replication events. J. Biol. Chem. **260:**8250–8256.
29. **Jones, K. S., J. Coleman, G. W. Merkel, T. M. Laue, and A. M. Skalka.** 1992. Retroviral integrase functions as a multimer and can turn over catalytically. J. Biol. Chem. **267:**16037–16040.
30. **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.
31. **Katz, R. A., G. Merkel, J. Kulkolsky, 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.
32. **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.
33. **Laemmli, U. K.** 1970. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. Nature (London) **227:**680–685.
34. **LaFemina, R. L., P. L. Callahan, and M. G. Cordingley.** 1991. Substrate specificity of recombinant human immunodeficiency virus integrase. J. Virol. **65:**5624–5630.
35. **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.
36. **Pahl, A., and R. M. Flugel.** 1993. Endonucleolytic cleavages and DNA joining activities of the integration protein of human foamy virus. J. Virol. **67:**5426–5434.
37. **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.
38. **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.
39. **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.
40. **Sambrook, J., E. F. Fritsch, and T. Maniatis.** 1989. Molecular cloning: a laboratory manual, 2nd ed. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
41. **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.
41a. **Scottoline, B. P., V. E. Ellison, and P. O. Brown.** Unpublished data.
42. **Sherman, P. A., M. L. Dickson, and J. A. Fyfe.** 1992. Human immunodeficiency virus type 1 integration protein: DNA cleaving activity. J. Virol. **66:**3593–3601.
43. **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.
43a. **Shih, C.-C., J. P. Stoye, and J. M. Coffin.** 1988. Highly preferred targets for retrovirus integration. Cell **53:**531–537.
44. **Shoemaker, C., J. Hoffman, S. P. Goff, and D. Baltimore.** 1981. Intramolecular integration within Moloney murine leukemia virus. J. Virol. **40:**164–172.
45. **Smith, D. B., and K. S. Johnson.** 1988. Single step purification of polypeptides expressed in *Escherichia coli* as fusions with glutathione S-transferase. Gene **67:**31–42.
46. **van Gent, D. C., Y. Elgersma, M. W. J. Bolk, C. Vink, and R. H. A. Plasterk.** 1991. DNA binding properties of the integrase of human immunodeficiency viruses types 1 and 2. Nucleic Acids Res. **19:**3821–3827.
47. **van Gent, D. C., A. A. M. Oude Groeneger, and R. H. A. Plasterk.** 1993. Identification of amino acids in HIV-2 integrase involved in site-specific hydrolysis and alcoholysis of viral DNA termini. Nucleic Acids Res. **21:**3373–3377.
48. **Varmus, H. E., and P. O. Brown.** 1989. Retroviruses, p. 53–108. *In* M. M. Howe and D. E. Berg (ed.), Mobile DNA. American Society for Microbiology, Washington, D.C.
49. **Vincent, K. A., V. Ellison, S. A. Chow, and P. O. Brown.** 1993. Characterization of human immunodeficiency virus type 1 integrase expressed in *Escherichia coli* and analysis of variants with amino-terminal mutations. J. Virol. **67:**425–437.
50. **Vink, C., K. H. van der Linden, and R. H. A. Plasterk.** 1994. Activities of the feline immunodeficiency virus integrase protein produced in *Escherichia coli*. J. Virol. **68:**1468–1474.
51. **Vink, C., E. Yeheskiely, G. A. van der Marel, J. H. van Bloom, and R. H. A. Plasterk.** 1991. Site-specific hydrolysis and alcoholysis of human immunodeficiency virus DNA termini mediated by the viral integrase protein. Nucleic Acids Res. **19:**6691–6698.
52. **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.
53. **Whitcomb, J. M., and S. H. Hughes.** 1992. Retroviral reverse transcription and integration: progress and problems. Annu. Rev. Cell Biol. **8:**275–306.