JOURNAL OF VIROLOGY, July 1996, p. 4585–4597 Vol. 70, No. 7
0022-538X/96/$04.00+0

Functional Domains of Moloney Murine Leukemia Virus Integrase Defined by Mutation and Complementation Analysis

COLLEEN B. JONSSON,† GEORGE A. DONZELLA, ELENA GAUCAN, CHRISTINE M. SMITH, AND MONICA J. ROTH*

Department of Biochemistry, Robert W. Johnson Medical School, University of Medicine and Dentistry of New Jersey, Piscataway, New Jersey 08854

Received 16 January 1996/Accepted 15 April 1996

Retroviral integrases perform two catalytic steps, 3′ processing and strand transfer, that result in the stable insertion of the retroviral DNA into the host genome. Mutant M-MuLV integrases were constructed to define the functional domains important for 3′ processing, strand transfer, and disintegration by in vitro assays. N-terminal mutants had no detectable 3′ processing activity, and only one mutant which lacks the HHCC domain, NΔ105, had strand transfer activity. Strand transfer mediated by NΔ105 showed preference for one site in the target DNA. Disintegration activity of N-terminal mutants decreased only minimally. In contrast, all C-terminal mutants truncated by more than 28 amino acids had no integration or disintegration activity. Activity on a single-strand disintegration substrate did not require a functional HHCC domain but did require most of the C-terminal region. Complementation analysis found that the HHCC region alone was able to function in *trans* to a promoter containing only the DD(35)E and C-terminal regions and to enhance integration site selection. Increasing the reducing conditions or adding the HHCC domain to NΔ105 reaction mixtures restored the wild-type strand transfer activity and range of target sites. The reducing agent affected Cys-209 in the DD(35)E region. The presence of C-209 was required for complementation of NΔ105 by the HHCC region.

To persist in an eukaryotic cell, retroviruses require the stable insertion of a double-stranded copy of their genome into the host chromosome by the retroviral integrase (IN). This requires a complex series of replication and integration processes in the early stages of their life cycle (7, 23, 47). The virus-encoded enzymes responsible for these events are reverse transcriptase and IN, respectively. Replication of the single-stranded Moloney murine leukemia virus (M-MuLV) RNA genome by reverse transcriptase to a double-stranded DNA occurs within 2.5 h postinfection (39) within the host cytoplasm. During the replication process, viral sequences unique to the 3′-terminal (U3) and 5′-terminal (U5) regions are duplicated and transposed to opposing ends of the linear viral DNA to generate long terminal repeats (LTRs). Each LTR serves as a substrate for the retroviral IN. As few as 7 bp of the terminal M-MuLV LTR sequences are required for a productive integration event and, therefore, recognition by IN (33). The specific protein-DNA contacts responsible for LTR recognition by IN have not been clearly defined.

Prior to formation of the integrated M-MuLV provirus, IN processes both 3′ ends of the viral DNA by an endonucleolytic cleavage of the terminal dinucleotide (3′ processing) (39). Analysis of M-MuLV in vivo has suggested that the two viral termini must be coordinated by M-MuLV IN for 3′ processing to occur (34). Following 3′ processing, the viral DNA intermediate and IN form a stable complex that localizes to the host chromatin, where the second IN-dependent catalytic step required for retroviral integration occurs, strand transfer (1). In the murine retroviruses, strand transfer follows mitosis (31, 36). It is postulated that the IN and viral DNA complex enters into the nucleus after breakdown of the nuclear membrane (36). Strand transfer occurs precisely at the termini of the LTR following the ubiquitous 5′-CA-3′ sequence found in all retroviral and retrotransposon LTRs. The reaction is a one-step transesterification whereby each end of the viral DNA is integrated into a random site in the host's genome (16). The viral ends are positioned several base pairs apart (4–6), depending on the retrovirus. Repair of the gap created by the staggered integration event generates a duplication of the internal sequences that flank the provirus.

The 3′ processing and strand transfer reactions have been reconstituted in vitro with assays that use oligonucleotide duplexes containing the DNA sequences of the LTR termini (9, 24). These assays demonstrate that IN is the only protein required for integration. The retroviral IN is also active on an integration intermediate or disintegration substrate (6). The integration and disintegration activities have been characterized for the M-MuLV IN proteins produced in bacteria (5, 11, 12, 21). Although the activities of the retroviral INs have reaction similarities, we have observed apparent differences in the disintegration substrate requirements of M-MuLV IN and human immunodeficiency virus type 1 (HIV-1) IN (11, 22). This is not unexpected, since M-MuLV IN and HIV-1 IN are proteins with quite different compositions. These differences also extend to the macromolecular composition of their preintegration complexes and their abilities to infect nondividing and dividing cells (1, 17, 36).

IN proteins from different retroviral species vary in size and amino acid homology (20, 25, 29). Two regions of strong similarity are shared between retroviral IN proteins, a proposed zinc finger motif or HHCC region in the amino terminus, and a central core region containing the DD(35)E motif. Linker

* Corresponding author. Mailing address: Department of Biochemistry, University of Medicine and Dentistry of New Jersey, Robert W. Johnson Medical School, 675 Hoes Ln., Piscataway, NJ 08854. Phone: (908) 235-5048. Fax: (908) 235-4783. Electronic mail address: Roth@mbcl.rutgers.edu.

† Present address: Department of Chemistry and Biochemistry, New Mexico State University, Las Cruces, NM 88003.

insertion and point mutations in the HHCC domain are lethal in vivo because of their inability to mediate 3′ processing of the viral termini (10, 37). The central core domain is proposed to contain a single active site responsible for integration and disintegration activities (23). This region features three invariant amino acids essential for catalysis, two aspartic acids and a glutamic acid (15, 25, 26, 45). The C-terminal portion of retroviral INs is the least conserved and has been functionally characterized to be the site of nonspecific DNA binding (32, 35, 44, 48, 51, 52). The terminal 28 amino acids of the carboxyl terminus of IN are nonessential for viability of M-MuLV in Rat 2 cells (38). The M-MuLV IN is approximately 14 kDa larger than HIV-1 and avian sarcoma-leukosis virus INs. This larger size of M-MuLV IN is accounted for by differences in its N and C termini. Approximately 50 amino acids precede the HHCC domain of the M-MuLV IN, and the C terminus contains a unique 36-amino-acid insertion (20).

In this study, we sought to define the protein domains of M-MuLV IN required for catalysis by a mutation and complementation approach. Mutant INs were examined for enzymatic activity in vitro by assays for integration, disintegration, and single-strand disintegration. N- and C-terminal mutants were hindered more in their ability to catalyze integration than disintegration reactions. The minimal domain for enzymatic activity for M-MuLV IN included a complete DDE region and most of the C terminus. While not required for minimal enzymatic activity, the HHCC domain restored wild-type (WT) strand transfer activity and target site selection in complementation experiments. Mixing inactive IN mutant proteins more narrowly defined the overlap between complementing pairs of M-MuLV IN than that found for HIV-1 IN (14, 46). These results support a model for the M-MuLV IN in which HHCC forms an independent domain distinct from the DD(35)E and C-terminal regions. Further, the HHCC region can function in *trans* to directly influence target site selection.

MATERIALS AND METHODS

Materials. Crude [γ-^{32}P]adenosine 5′-triphosphate (7,000 Ci/mmol) was purchased from ICN. [α-^{32}P]TTP (800 Ci/mmol) was obtained from Amersham. T4 polynucleotide kinase was obtained from GIBCO BRL. Restriction enzymes were purchased from New England Biolabs.

Oligonucleotides. DNA oligonucleotides were purchased from the University of Medicine and Dentistry of New Jersey Biochemistry Department Synthesis Facility. Oligonucleotides used in this study are referred to by their synthesis numbers. Oligonucleotide substrates were ^{32}P labeled at the 5′ end with T4 polynucleotide kinase (GIBCO BRL) and hybridized with a fivefold excess of the complementary strand. Excess radioactivity was removed by using G-25 spin columns (Boehringer Mannheim) or by electrophoresis with 15% native polyacrylamide gels. For the Y substrates used in disintegration assays, the LTR DNA sequence is underlined. The standard Y substrates are as follows: oligonucleotide 3152, 5′-CAGCAACGCAAGCTTG; oligonucleotide 3192, 5′-AATGAAAGACCCCGCTGACTA; oligonucleotide 3154, 5′-TAGTCAGCGGGGTCTTTCAGGCTGCAGGTCGAC; and oligonucleotide 3527, 5′-GTCGACCTGCAGCCCAAGCTTGCGTTGCTG. The dumbbell Y substrate, Y_{DB}, is oligonucleotide 3808, 5′-AATGAAAGTTCTTTCAGGCTGTTCAGCCCAAGCTTGCTTG, which is self complementary. Oligonucleotide 3152 was 5′ end labeled, purified, and annealed to oligonucleotides 3192, 3527, and 3154 to produce the standard Y substrate or annealed to oligonucleotides 3527 and 3154 to produce the Y_{SS} substrate as previously described (11). Y_{DB} was 5′ end labeled, self-annealed, and purified as previously described (11). Y_{CD} was made from strands 3152 and 3527. For integration substrates, oligonucleotides 2784/2785 were prepared as described previously (22). The 3′-processing substrate sequence is 5′ AATGAAAGACCTTGGTCTTTCA-3′ and represents a new substrate that forms a hairpin LTR molecule.

Construction of M-MuLV IN mutants. Cloning of the M-MuLV IN into the expression vector pET11C (Novagen, Madison, Wis.) has been previously described and was used as the parent vector (pETINH1) for many of the subsequent manipulations (22). This construct has a Met introduced prior to the first amino acid of the M-MuLV IN; however, this residue is not counted in the numbering scheme of our mutant IN proteins. The first amino acid of the M-MuLV IN constructs presented here corresponds to nucleotide 4610 of the M-MuLV RNA (42). A second parent vector used in these constructions is pETINH3. pETINH3 contains additional nucleic acid sequence at the 5′ end of the gene, before the first M-MuLV amino acid and after the first methionine ATG, which encodes for a hexahistidine tag. The *Hin*dIII site in the vector portion of pETINH3 was removed to facilitate cloning of the C-terminal mutants. The site was destroyed by digestion of the vector with *Hin*dIII, filled in with Klenow fragment, and religated with T4 DNA ligase. Construction of C97Y (M-MuLV IN mutant with tyrosine at position 97 instead of cysteine) and *in*5247 has been previously reported (22, 37). NΔ105 was constructed by introducing a Met prior to amino acid 106 by PCR. The 5′-end PCR primer 3807, 5′-GATATACATATG**GCCGTTAAACAGGGA**-3′, adds an *Nde*I site (underlined) prior to nucleotide 316; the primer sequence for IN is shown in bold type. The 3′-end PCR primer, 2350, 5′-GGATCCAGTACTGACCCCTCTG-3′ adds a *Bam*HI site (underlined) outside the stop codon of the gene. The PCR product was digested with *Nde*I, and the resulting 475-bp fragment replaced with WT 791-bp *Nde*I fragment of pETINH1. NΔ176 was constructed in a similar manner by replacing the WT *Nde*I fragment with a 259-bp PCR-generated *Nde*I fragment. The 3′-end primer, 2350, was used in conjunction with a 5′-end primer, oligonucleotide 3904 (5′-CATATG**CCTCAGGTATTGGGA**) to create the deletion fragment. The Cys-to-Ala mutation, C209A, was created by oligonucleotide-directed mutagenesis using overlap PCR (19). The complementary primers, 3811, 5′-GGATTGATTGGAAATTGCATGCTGGATACAGACC-3′, and 3811C, 5′-GGTCTGTATCCAGCATGCAATTTCCAATCAATCC-3′ introduced a mutation at the underlined sequence. PCRs were made with primer combinations, primers 3811 and 2350 and primers 3811C and 2351 (5′-end primer [21]). Individual PCR products were separated on a 1.2% agarose gel and glass powder purified (50). The two products were then added to a PCR with primers 2350 and 2351. All PCRs were performed with Vent (New England Biolabs) or *Pfu* DNA polymerase (Stratagene). K227A and CΔ311/346 (deletion of the C-terminal 36-amino-acid insertion) were created in pHS by the method of Kunkel (18, 27). The DNA oligonucleotide corresponding to K227A, oligonucleotide 2076 (5′-GAACCATCGCGGAGACTTTAAC-3′), generates a *Bst*U1 site (underlined) that allowed screening for the mutation by the appearance of one additional *Bst*U1 site. The oligonucleotide for CΔ311/346, 5′-CGAAGTCTGGAGACCTCTGTGGAAAGGACCTTACACAGTCC-3′, resulted in the loss of a *Kpn*I site. Following construction of the mutation, a 750-bp fragment was digested with *Sac*II-*Pml*I and exchanged into pNCA-C (8, 18). The entire coding region was then moved into pETINH1. The construct encoding the mutant protein CΔ232 was made in pETINH3 (38). pETINH3 was first partially digested with *Sph*I. Linearized plasmid DNA was isolated, and the DNA termini were blunt ended with T4 DNA polymerase in the presence of deoxynucleoside triphosphates and ligated in the presence of 100-fold molar excess of *Nhe*I linkers (catalog no. 1060; New England Biolabs). The same protocol was followed for CΔ144; a partial *Nde*I digest of pETINH3 was made, the full-length linear DNA was isolated and filled in with T4 polymerase, and *Nhe* linkers were ligated into the linearized plasmid. Carboxyl-terminal deletions were subcloned from previously reported plasmid constructs by isolation of the *Hin*dIII-*Bam*HI fragment, and replacement of the *Hin*dIII-*Bam*HI fragment in the pETINH3 vector (38). The nucleotide sequence of each construct was verified by dideoxy sequencing with Sequenase version 2.0 (United States Biochemical) (40).

Protein purification. Mutant and WT M-MuLV IN proteins were expressed in *Escherichia coli* and purified by nickel nitriloacetate (NTA) chromatography (Qiagen), using methods described previously (21). One additional chromatography step was added to increase the purity of the preparation. Briefly, the material eluted from the nickel-NTA resin in buffer A (10 mM Tris base, 0.1 M Na_2HPO_4, 0.1% Nonidet P-40, 10 mM β-mercaptoethanol, 100 mM NaCl, 10% glycerol) at pH 4.5 was pooled, dialyzed for 2 h against buffer A at pH 8.0, and reapplied to the same 1 ml of reactivated nickel-NTA resin reequilibrated to pH 8.0. The column was then washed stepwise with 10 ml each of buffer A at pH 8.0, buffer A at pH 6.3, and buffer A at pH 5.9. M-MuLV IN was finally eluted from the resin with 10 ml of buffer A at pH 4.5 and refolded by a stepwise removal of denaturant (21). Chromatography steps were followed by sodium dodecyl sulfate-polyacrylamide gel electrophoresis and Western blot (immunoblot) analysis for each protein fraction. The protein concentrations in the fractions were quantitated by the Bradford method (2). To assess the enzymatic activity and stability of each mutant, proteins were prepared and isolated a minimum of three times. In vitro assays for 3′ processing, strand transfer, and disintegration activities were also measured for each protein for a minimum of three independent purifications. Two independent purifications of NΔ105/C209A were assayed and demonstrated consistent strand transfer and disintegration activities. Protein preparations that showed consistency in their activities are represented herein.

Integration and disintegration assays. Integration substrates were prepared from oligonucleotides, and assays were done as previously described at 30 or 37°C (21). Strand transfer activities were assayed in a reaction mixture containing 20 mM MES (morpholineethanesulfonic acid) (pH 6.2), 10 mM dithiothreitol (DTT), 10 mM KCl, 10 mM $MnCl_2$, 20% glycerol, 1 pmol of substrate, and 12 to 30 pmol of M-MuLV IN in a final volume of 15 μl. Reaction mixtures for 3′ processing were identical to strand transfer except for the inclusion of 10% ethylene glycol. The 3′ end of the substrate was filled in at room temperature for 30 min using 40 pmol of substrate and 80 pmol of [α-^{32}P]TTP (800 Ci/mmol) with 5 to 8 U of exonuclease-free Klenow fragment of *E. coli* DNA polymerase I. The unincorporated deoxynucleoside triphosphate label was removed by using a spin

column, and the labeled substrate was gel purified on 20% polyacrylamide gels, eluted, and resuspended in 10 μl of 0.1 M NaCl. The substrate was self annealed by slow cooling after heating to 95°C and subsequently was diluted to 0.5 pmol/μl and used as indicated in integration reactions. Preparation of Y substrates, enzymatic assays for disintegration activity, and product analysis were performed as previously described (11, 21). The standard Y substrate was Y3154, the single-stranded Y_{SS} substrate was Y3154 without the A strand, and the dumbbell substrate, Y_{DB}, was Y3808. All disintegration reactions were performed in a volume of 15 μl containing 20 mM piperazine-*N*,*N*′-bis(2-ethanesulfonic acid) (PIPES) (pH 6.4), 10 mM DTT, 25 mM $MnCl_2$, 10 mM 3-[(3-cholamidopropyl)-dimethyl-ammonio]-1-propanesulfonate (CHAPS), and 0.05% Nonidet P-40 at 30 or 37°C for 1 h, typically with 0.5 to 1.0 pmol of Y substrate, and 5 to 20 pmol of M-MuLV IN. All integration and disintegration reactions were terminated by the addition of 10 μl of stop buffer (95% formamide, 1 mM EDTA). Reaction mixtures were heated to 95°C for 3 to 5 min and then separated on 20% sequencing gels. Gels were dried, exposed to Kodak X-Omat X-ray film at −70°C or room temperature from 30 min to overnight, and analyzed by autoradiography.

Complementation assays. Equimolar amounts (7.5 pmol) of individual M-MuLV IN proteins were mixed and incubated for 15 min on ice or at 37°C in reaction buffer as indicated for integration assays in the absence of substrate. Following preincubation, substrate (1 pmol) for the 3′ processing or strand transfer reaction was added to each reaction mixture and placed at 30 or 37°C. After 1 h, reactions were terminated with 10 μl of stop buffer. Reaction mixtures were examined by the autoradiography method mentioned above.

Complementation (DTT) assay. DTT treatment of WT, C209A, CΔ232, NΔ105, and NΔ105/C209A M-MuLV IN proteins were performed as follows. Each protein (120 pmol) was diluted to a final concentration of 2 pmol/μl with ice-cold 20 mM HEPES (*N*-2-hydroxyethylpiperazine-*N*′-2-ethanesulfonic acid) (pH 7.4)–20% glycerol and equilibrated on ice for 5 min. For complementation assays, 60 pmol of each protein was mixed together and then diluted immediately to 2 pmol/μl (as above) with respect to total protein concentration. Untreated proteins were not subjected to this dilution step. The DTT concentration was then adjusted to 40 mM, and the mixture was allowed to incubate on ice for 20 min. Reactions were initiated by the addition of 10 or 20 pmol of the untreated or DTT batch-treated protein as indicated to 1 pmol of oligonucleotides 2784/2785 or Y standard substrate in their respective reaction buffer. The final DTT concentrations in the reaction mixtures were 10 mM for the untreated proteins and 40 mM for the DTT-treated proteins. Untreated complementation assays were conducted by adding 10 pmol of each protein pair to the sides of the microcentrifuge tube and mixing by a 1-s spin in the microcentrifuge. Reaction mixtures were incubated at 37°C for 1.5 h, and then the reactions were terminated. Both strand transfer and disintegration assays were performed concurrently with the same batch-treated M-MuLV IN preparations. Integration and disintegration assays were repeated three times with three separate DTT treatments of the M-MuLV IN proteins and yielded consistent results.

Product analysis. Data were quantitated directly from dried sequencing gels with a Molecular Dynamic PhosphorImager as previously described (22). Data were collected from gels with ImageQuant 3.1 software. Percent conversion to product was determined by dividing the amount of product by the sum of the substrate and product and multiplying by 100.

RESULTS

Construction and purification of M-MuLV IN mutants. WT and mutant M-MuLV IN proteins were produced in *E. coli* BL21(DE3). Protein expression was under the control of the T7 *lac* promoter (21). The open reading frames of the WT and mutant M-MuLV IN genes were constructed so as to contain a hexahistidine tag at the 5′ or 3′ end of the gene coding sequence. The hexahistidine tag facilitated the purification of IN proteins on Ni^{2+}-NTA affinity chromatography resin as previously described (21). No difference was noted in the specific activities of amino- and carboxyl-hexahistidine-tagged WT M-MuLV INs during in vitro integration or disintegration assays (data not shown). A schematic representation of each mutant M-MuLV IN protein and the position of the hexahistidine tag for each construct are presented in Fig. 1. In general, IN genes constructed with C-terminal truncations have the hexahistidine tag at the 5′ end, while genes with N-terminal truncations have the tag at the 3′ end. Mutants which alter the HHCC finger include *in*5247, C97Y, NΔ105, and NΔ176. Mutants with single amino acid substitutions within the catalytic core region include C209A and K227A. The C terminus was progressively deleted through the catalytic core region. Mutants with these deletions included CΔ22, CΔ28, CΔ34, CΔ40, CΔ45, CΔ86, CΔ144, and CΔ232. Mutants CΔ311/346 contains an internal deletion of the unique 36-amino-acid sequence found in the M-MuLV IN protein (20).

Figure 2 shows the protein products of whole-cell extracts expressing the WT M-MuLV IN and the 14 mutants used in this analysis. The protein products are expressed in large quantities, and they migrate to the positions predicted for the full-length product of each construct. A few of the IN proteins were unstable during subsequent solubilization, extraction, and purification, resulting in variable levels of enzymatic activity and an increased background of nonspecific nuclease activity (CΔ311/346 and CΔ34). Proteolytic degradation of these mutants was noted from Western blot analysis of crude protein extracts from *E. coli* cultures as well as in the final fractions eluted at pH 4.5 (data not shown). The poor yields and stabilities of CΔ34 and CΔ311/346 proteins precluded any further analysis. Activity of CΔ311/346 was noted in the standard disintegration reaction but was variable (data not shown).

Defining functional catalytic domains and amino acid residues for integration in M-MuLV IN by mutational analysis. Previous work with M-MuLV IN found that the HHCC region was necessary for 3′ processing and integration, but not for disintegration (22). In these experiments, we were interested in further refinement of the domains of M-MuLV IN required for enzymatic activity. Mutations targeted the three major domains of the IN protein: the HHCC finger, the catalytic core, and the C-terminal region. Proteins were assayed for strand transfer activity by using a 3′ recessed oligonucleotide substrate, serving as both the viral and target DNA (Fig. 3A). 3′ processing activity was also measured with a blunt-ended, self-annealing oligonucleotide substrate (Fig. 3C). The abilities of IN to remove the radiolabeled terminal dinucleotide and to mediate subsequent strand transfer reactions can be monitored by separation of the products on polyacrylamide gels. Under the assay conditions for 3′ processing, three forms of the released dinucleotide were identified previously (49).

The amino-terminal deletion of M-MuLV IN, NΔ105, deletes the HHCC region to a position homologous to the NΔ105 deletion of HIV-1 IN and Rous sarcoma virus IN (3, 4, 46). The increase in the size of this M-MuLV IN deletion is due to the additional 55 amino acids preceding the HHCC domain. NΔ105 is truncated through the 5 amino acids following C-97. NΔ176 is truncated through the HHCC region and the first D (D-125) of the DDE motif. The ability of NΔ105 to perform strand transfer varied slightly among protein purifications (Fig. 3A, lane 5, and Fig. 3B, lanes 3 to 5). NΔ105 had no detectable 3′ processing activity (FIg. 3C, lane 5). Upon further analysis of the strand transfer reactions, it was observed that strand transfer selectively occurred at one site approximately 4 to 5 nucleotides from the 3′ end of the labeled LTR strand (Fig. 3B). Several preparations of this protein were made to confirm the preference for one integration site. Three other preparations of NΔ105 are shown (Fig. 3B, lanes 3 to 5) in direct comparison with WT IN (Fig. 3B, lane 2). Two of these preparations (Fig. 3B, lanes 3 and 5) demonstrated marked specificity for a site-specific integration, while one preparation of NΔ105 showed no activity (Fig. 3B, lane 4). We show the activity of the inactive preparation because this protein was used in the complementation analysis below (see Fig. 5). NΔ176 had no detectable strand transfer (Fig. 3A and C, lanes 6) or 3′ processing activity (Fig. 3C, lane 6).

Additional mutants with mutations in the HHCC region, which were previously described (22) were further analyzed for

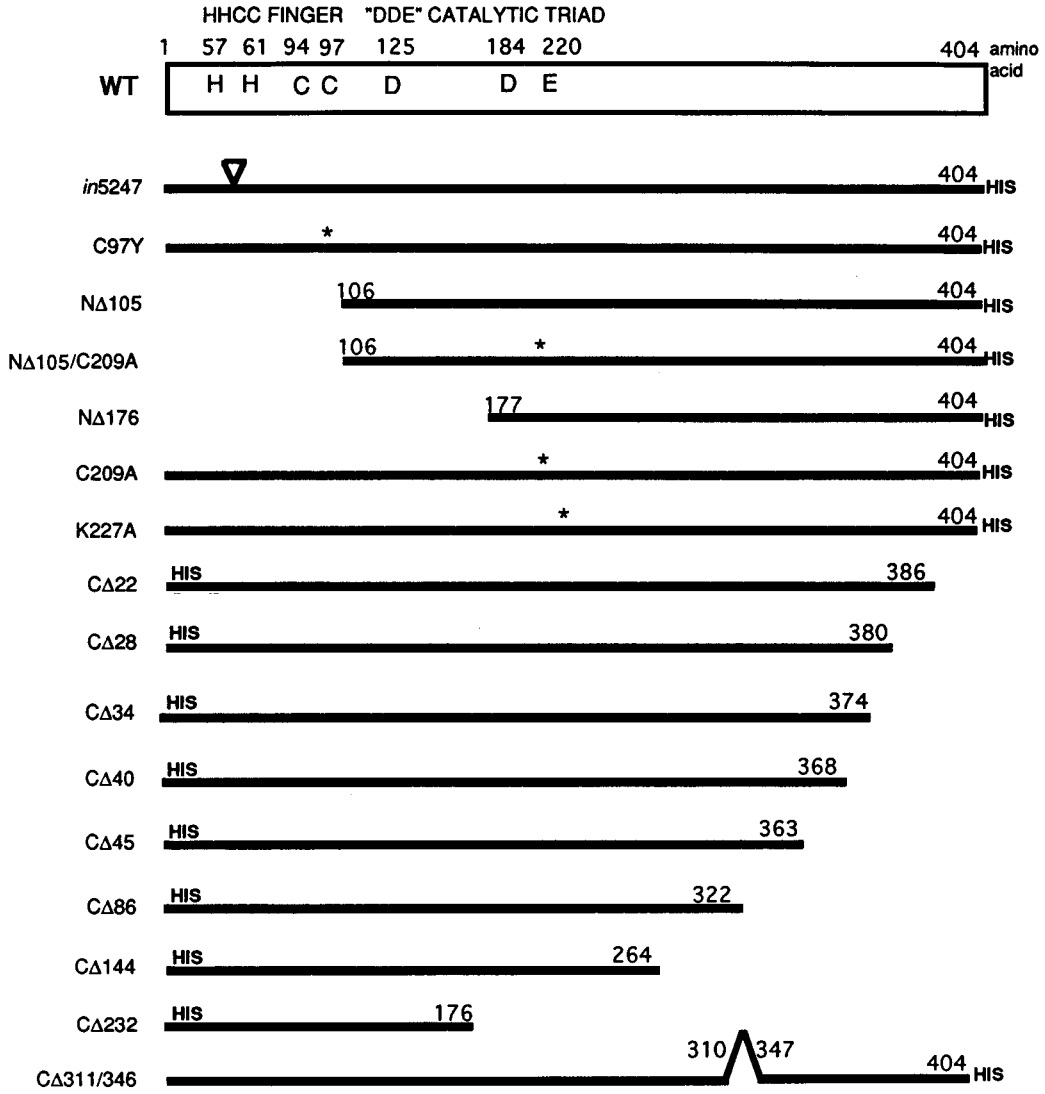

FIG. 1. Illustration of WT M-MuLV IN and mutant INs used in this study. The name of each mutant is indicated to the left of each protein schematic. The location of the hexahistadine tag (HIS) at the N- or C terminus of each mutant protein is indicated. Insertion of the hexahistadine sequence at the C terminus creates a deletion of the four terminal amino acids of the WT IN, resulting in the 404 amino acids (total) as described previously (20). Asterisks indicate the positions of single amino acid substitutions.

in vitro 3′ processing and strand transfer. These mutations include the linker insertion between the histidines, *in*5247 and the point mutation of the second cysteine for a tyrosine (C97Y; *bi*5247 [37]). 3′ processing and strand transfer activity were not observed with any preparation of *in*5247 or C97Y (Fig. 3A and C, lanes 3 and 4). The activities of these two mutants are also consistent with previous work in Rat 2 cells where these mutations render a nonviable virus with defective processing of the LTR termini (37).

Two amino acid substitutions were made in the catalytic core, C209A and K227A, which flank the conserved E-220 residue in the active site of M-MuLV IN. C209A is located 10 residues N terminal to the active-site E-220 residue, and K-227 is 6 residues C terminal to E-220 (20, 25). K-227 is conserved among retroviral IN proteins (14, 20). K227A did not show strand transfer or 3′ processing activity (Fig. 3A and C, lanes 8). The products produced with K227A correspond to nonspecific cleavage within the single-strand region of the substrate; whether this activity is intrinsic to this mutant or is from a

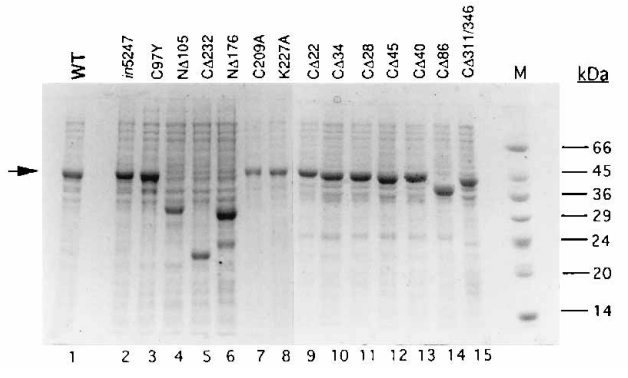

FIG. 2. Coomassie blue-stained sodium dodecyl sulfate-polyacrylamide gel electrophoresis of WT and mutant M-MuLV INs. Recombinant proteins from whole-cell extracts are shown. Lane numbers are shown below the gel, and the corresponding protein is indicated at the top of each lane. Lane M contains molecular mass markers. The arrow to the left of the gel points to the 46-kDa band, which contains WT M-MuLV IN.

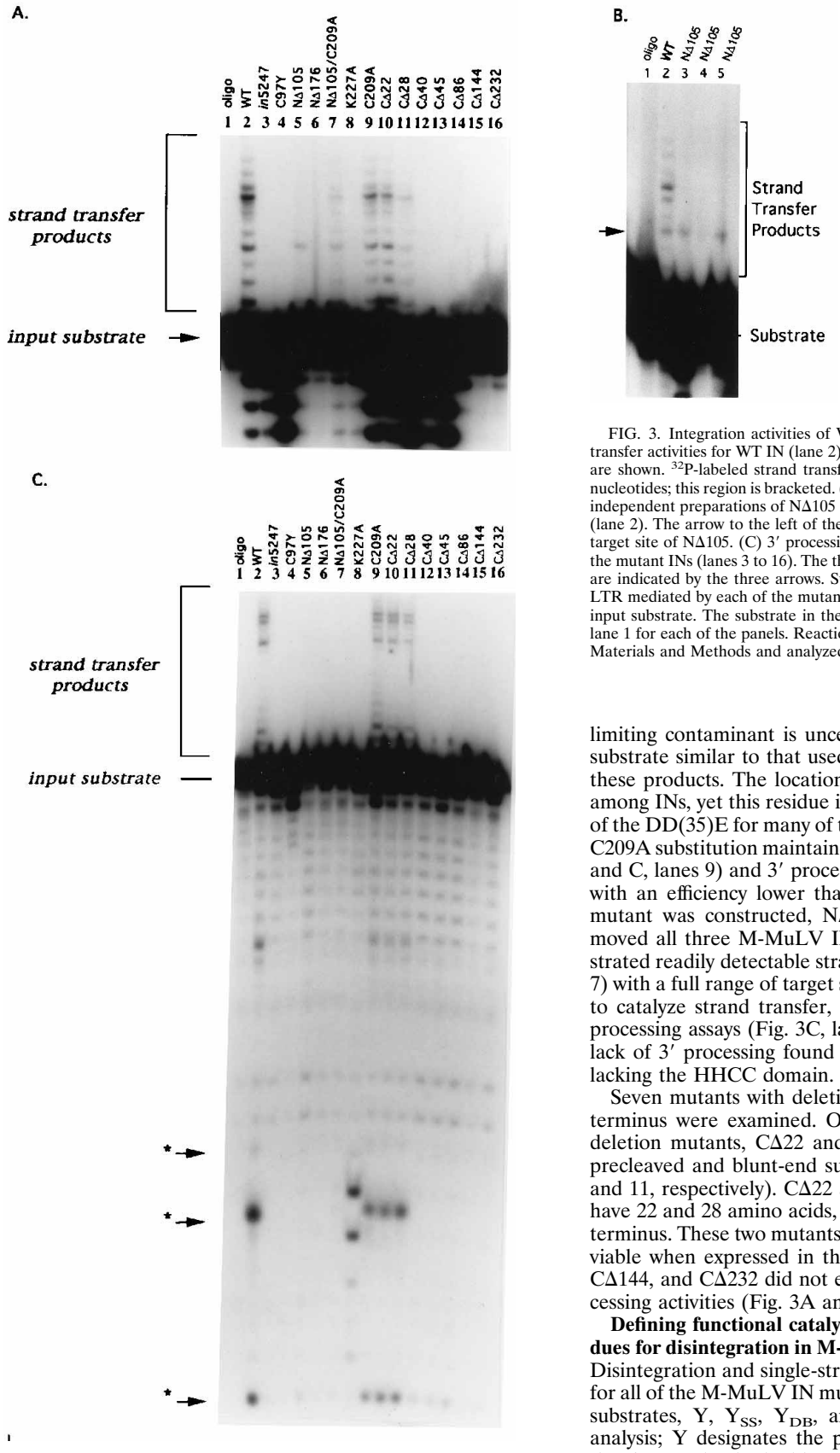

FIG. 3. Integration activities of WT and mutant M-MuLV INs. (A) Strand transfer activities for WT IN (lane 2) and each of the mutant INs (lanes 3 to 16) are shown. ^{32}P-labeled strand transfer product sizes may range from 19 to 36 nucleotides; this region is bracketed. (B) Strand transfer activity of NΔ105. Three independent preparations of NΔ105 are indicated in lanes 3 to 5 next to WT IN (lane 2). The arrow to the left of the gel indicates the preferred strand transfer target site of NΔ105. (C) 3′ processing activity for WT IN (lane 2) and each of the mutant INs (lanes 3 to 16). The three forms of the 3′ processing product (48) are indicated by the three arrows. Subsequent strand transfer of the processed LTR mediated by each of the mutants is indicated within the bracket above the input substrate. The substrate in the absence of protein (oligo) is presented in lane 1 for each of the panels. Reactions were performed at 37°C as described in Materials and Methods and analyzed on denaturing polyacrylamide gels.

limiting contaminant is uncertain. Blunt-ended 3′ processing substrate similar to that used for strand transfer did not yield these products. The location of the Cys-209 is not conserved among INs, yet this residue is found between the Asp and Glu of the DD(35)E for many of the retroviruses. IN containing the C209A substitution maintained strand transfer activity (Fig. 3A and C, lanes 9) and 3′ processing activity (Fig. 3C, lane 9) but with an efficiency lower than that of the WT IN. A double mutant was constructed, NΔ105/C209A, which in effect removed all three M-MuLV IN cysteines. This mutant demonstrated readily detectable strand transfer activity (Fig. 3A, lane 7) with a full range of target site usage. In contrast to its ability to catalyze strand transfer, NΔ105/C209A was inactive in 3′ processing assays (Fig. 3C, lane 7). This is consistent with the lack of 3′ processing found with the parental NΔ105 protein lacking the HHCC domain.

Seven mutants with deletions of the M-MuLV IN carboxyl terminus were examined. Only two of the seven C-terminal deletion mutants, CΔ22 and CΔ28, were capable of using a precleaved and blunt-end substrate (Fig. 3A and C, lanes 10 and 11, respectively). CΔ22 and CΔ28 represent mutants that have 22 and 28 amino acids, respectively, removed from the C terminus. These two mutants have been shown previously to be viable when expressed in the virus (38). CΔ40, CΔ45, CΔ86, CΔ144, and CΔ232 did not exhibit strand transfer and 3′ processing activities (Fig. 3A and C, lanes 12 to 16).

Defining functional catalytic domains and amino acid residues for disintegration in M-MuLV IN by mutational analysis. Disintegration and single-stranded splicing (6) were examined for all of the M-MuLV IN mutants (Fig. 4). Four disintegration substrates, Y, Y_{SS}, Y_{DB}, and Y_{CD}, were employed in this analysis; Y designates the prototype disintegration substrate that is completely double stranded. We have reported previ-

ously that the prototypic splicing substrate, Y_{SS}, for WT M-MuLV IN had a strict requirement for the presence of 17 or more nucleotides on the single-stranded LTR portion of the substrate (11). This activity is substantially different from that reported for HIV-1 IN, which is active even with a single-nucleotide overhang (6). Two alternative disintegration substrates, Y_{DB} and Y_{CD}, were also included in the analysis. The dumbbell substrate, Y_{DB}, is a self-complementary substrate that contains a 5-bp duplex of the LTR region. Variation in M-MuLV IN catalysis with the Y_{DB} substrate versus the standard disintegration Y substrate has been observed (11). Finally, a new substrate variant, Y_{CD}, which constitutes only the C and D strands of the disintegration substrate and lacks LTR sequences is introduced. No activity with the Y_{CD} substrate was expected. Surprisingly, a low level of activity was noted with this substrate.

Representative assays of the M-MuLV IN proteins for each of the Y substrates are shown in Fig. 4. The results of quantitation of multiple assays are summarized in Table 1. The WT M-MuLV IN activity was set to 100%, and the activities stated for the mutants are given as percentages of WT activity. The WT M-MuLV IN activity is shown for the Y, Y_{SS}, Y_{DB}, and Y_{CD} substrates (Fig. 4, lanes 2). Mutant proteins with mutations in the HHCC region, *in*5247 and C97Y (lanes 3 and 4), retained activity with Y, Y_{DB}, and Y_{SS}, although with efficiencies lower than that of the WT IN (Table 1). This was in contrast to the lack of activity noted for these enzymes with the integration substrates (Fig. 3A and B, lanes 3 and 4).

In contrast to mutations in the HHCC region, deletion of the N-terminal HHCC region, NΔ105, resulted in product accumulation using the Y and Y_{DB} substrates at levels similar to WT IN (Fig. 4A and C, lane 5). Activity on Y_{SS} remained low but detectable (Table 1 and Fig. 4B, lane 5). No disintegration activity was retained in an N-terminal M-MuLV IN mutant NΔ176 with any of the substrates examined (Fig. 4A to C, lanes 6). K227A had minimal activity with the Y, Y_{SS}, and Y_{DB} substrates (Fig. 4A to C, lanes 8; Table 1). C209A showed activity with all the substrates, Y, Y_{SS}, and Y_{DB} (Fig. 4A to C, lanes 9). The highest activity of C209A was noted with the Y substrate, which accumulated product to a level equivalent to that of WT IN (Fig. 4A, lane 9). C209A showed lower levels of activity with Y_{SS} converting 24% of the substrate to product compared with WT IN (Fig. 4B, lane 9). Similarly, a low level of activity was also observed with C209A for the Y_{DB} substrate, which had converted 27% of the substrate to product (Fig. 4C, lanes 9). The double mutant NΔ105/C209A retained activity with Y substrate (Fig. 4A, lane 7, and Fig. 7B, lanes 10 and 11) and Y_{SS} substrate (Fig. 4B, lane 7), both of which averaged less than 10% of WT IN (Fig. 4B, lane 7). NΔ105/C209A showed the least activity with the Y_{DB} at 3% of WT IN (Fig. 4C, lane 7). In general, the effects of the mutations and truncations at the N-terminal portions of M-MuLV IN were observed to produce a loss in 3′ processing and strand transfer activity, but not disintegration.

The disintegration activities of all seven of the carboxyl-terminus deletion mutants of M-MuLV IN were examined. The mutants displayed patterns of activity similar to that demonstrated for integration (Fig. 3A and C). Disintegration activity was readily detected for deletions CΔ22 and CΔ28 with the standard Y, Y_{SS}, and Y_{DB} substrates (Fig. 4A to C, lanes 10 and 11). No disintegration activity was noted for CΔ40, CΔ45, CΔ86, CΔ144, and CΔ232 (Fig. 4A, lanes 12 to 16). Enzymatic activity for CΔ40, CΔ45, CΔ86, CΔ144, and CΔ232 was also not detected with any of the other disintegration substrate analogs (Fig. 4B and C, lanes 12 to 16). Disintegration activities of mutants with C-terminal truncations were affected in manners similar to that observed for both integration activities examined.

The final enzymatic activity examined is one that has not previously been reported. The substrate contains no LTR sequences and consists of two strands, the 16-nucleotide C strand and a 30-nucleotide D strand that form a DNA molecule with double- and single-stranded regions. The 5′ end of the C strand is ^{32}P labeled. The reaction catalyzed by IN is a nucleophilic attack mediated by the 3′ OH of the C strand on the D strand, which results in the formation of a 32-nucleotide product. The ability to recognize the Y_{CD} substrate was limited to the WT IN, C209A (data not shown), and CΔ22 and CΔ28 (Fig. 4D, lanes 2, 9, and 10). The level of activity was lower for this substrate than observed with the other Y substrates analyzed and was not quantified.

Functional complementation of strand transfer activity using N- and C-terminal deletions of M-MuLV IN. The abilities of two of the seven C-terminal mutants to function in *trans* with NΔ105 and regain integration activity were examined (Fig. 5). Three concentrations of NΔ105 were examined, while the concentration of each mutant was held constant. NΔ105 used in this study did not have integration activity (Fig. 3B, lane 4). Each protein component was mixed in reaction buffer on ice and incubated for 15 min prior to the addition of strand transfer substrate.

In Fig. 5, the WT strand transfer activity of M-MuLV IN is shown in lane 8, and the activity of NΔ105 at the lowest concentration is presented in lane 1. The addition of NΔ105 to M-MuLV IN did not affect strand transfer activity, although a slight increase in product was noted (Fig. 5, lanes 8 to 11). A dramatic revival of strand transfer activity was noted in reaction mixtures with CΔ45 (Fig. 5, lanes 2 to 4) upon the addition of NΔ105. Mixing CΔ86 and NΔ105 did not reconstitute strand transfer activity as significantly (Fig. 5, lanes 5 to 7). These results indicate that a viral LTR one-end strand transfer reaction can be reconstituted by the assembly of one HHCC domain, two DD(35)E domains, and one carboxyl-terminal domain.

Functional complementation of NΔ105 and NΔ176 strand transfer activities by the HHCC domain. In the experiments shown in Fig. 5, complementation was performed with HHCC, DD(35)E and portions of the C-terminal region on one molecule and DD(35)E with the C terminus on the second molecule. In these experiments, we ask whether the HHCC region can function as an independent domain. The combinations of proteins used in this complementation analysis are schematically represented in Fig. 6A. The two C-terminal deletions were tested: CΔ144, which contains the HHCC and DD(35)E region, and CΔ232, which contains the HHCC domain and sequence up to and including the first D of the DD(35)E region. The C-terminal deletions were mixed with NΔ105 [contains the DD(35)E region and the C terminus] or NΔ176 [first D of the DD(35)E region and the C terminus]. The locations of the HHCC and DD(35)E invariant residues are depicted for the WT M-MuLV IN (Fig. 6A, #1). The terminal 72 amino acids of CΔ232 overlap with the first 72 amino acids of NΔ105 (Fig. 6A, #2). CΔ144 comprises the first 263 amino acids of M-MuLV IN and has an overlap of 159 amino acids with NΔ105 (Fig. 6A, #3). Again, these 159 amino acids comprise the entire DD(35)E region. CΔ232 protein contains the first 176 amino acids of M-MuLV IN. No overlap exists between CΔ232 and NΔ176 (Fig. 6A, #4). The NΔ176 protein starts where CΔ232 stops. CΔ144 has an overlap of 87 amino acids with NΔ176 (Fig. 6A, #5). These 87 amino acids comprise only the D(35)E portion.

The strand transfer activity was assayed for each of the four

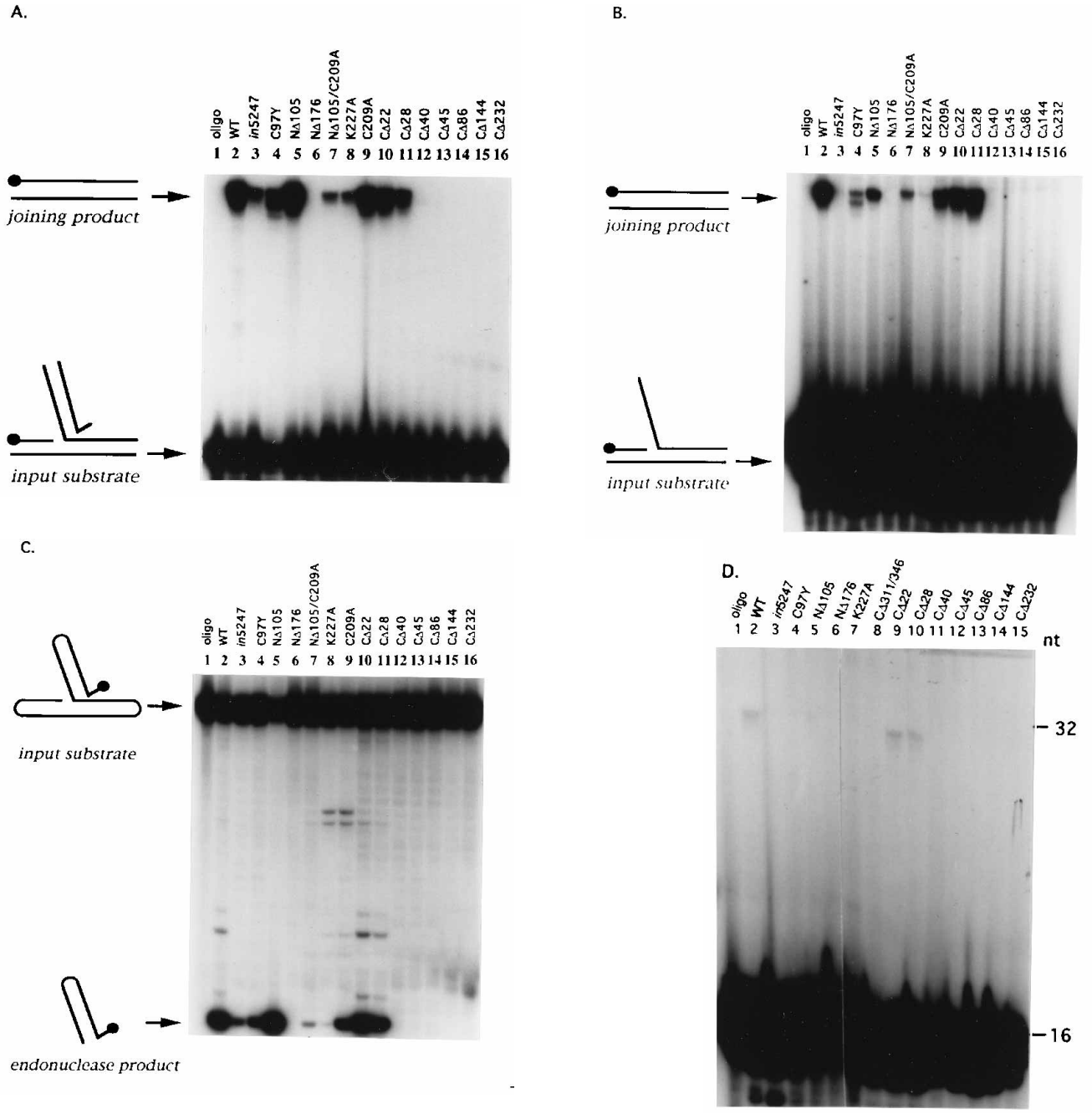

FIG. 4. Disintegration activities of WT and mutant M-MuLV INs. Enzymatic activities of WT and mutant M-MuLV INs with the following substrates: the standard disintegration substrate, Y (A); the single-strand disintegration substrate, Y_{SS} (B); the dumbbell substrate, Y_{DB} (C); and the non-LTR disintegration substrate, Y_{CD} (D). Reactions were performed at 37°C as described in Materials and Methods and analyzed on 20% denaturing polyacrylamide gels. In panels A and B, the substrate (input substrate) is 16 nucleotides long and the product (joining product) is 30 nucleotides long. In panel C, the Y_{DB} substrate is 40 nucleotides long and the product is 16 nucleotides long. In panel D, the Y_{CD} substrate is 16 nucleotides long and the product is 32 nucleotides long. The corresponding protein in each reaction is presented above each lane number. The substrate in the absence of protein (oligo) is presented in lane 1 in each panel. nt, nucleotide.

proteins individually and in combination following a short preincubation period at 37°C (Fig. 6B). Strand transfer activity was measured under identical conditions for individual proteins and for protein combinations used in mixing experiments. NΔ105 used in these experiments had a low level of the one-site strand transfer activity (Fig. 6B, lane 3). We expected that our assays would be more sensitive in detecting the minimal domains for activity with this preparation. WT strand transfer activity for M-MuLV IN is shown in Fig. 6B, lane 2. Strand transfer activity was not detected for NΔ176 (Fig. 6B, lane 4), CΔ232 (Fig. 6B, lane 5), or CΔ144 (Fig. 6B, lane 6). CΔ232 and CΔ144 proteins were able to complement strand transfer activity with NΔ105 (Fig. 6B, lanes 7 and 8). Interestingly, NΔ176 activity was not reconstituted by CΔ232 or CΔ144 (Fig. 6B, lanes 9 and 10). Similar results were also observed with a blunt-end substrate for 3′ processing (data not shown). This

TABLE 1. Summary of M-MuLV IN activities and complementation[a]

M-MuLV IN	Integration activity		Disintegration activity with substrate:				Complementation		
	3′ processing	Strand transfer	Y	Y_{SS}	Y_{DB}	Y_{CD}	NΔ105	CΔ232	NΔ176
WT	+++	+++	+++	+++	+++	+++	+++	+++	ND
*in*5247	−	−	+	+/−	+	−	−	ND	−
C97Y	−	−	+	+/−	+	−	−	ND	−
NΔ105	−	+/−	+++	+	+++	−	−	+++	−
NΔ176	−	−	−	−	−	−	−	ND	−
K227A	−	−	+	+/−	+/−	−	+/−	ND	−
CΔ311/346	−	−	+/−	−	−	−	ND	ND	−
C209A	+++	++	+++	+	+	+++	ND	+++	ND
CΔ22	+++	+++	++	++	+++	+++	ND	ND	ND
CΔ28	+++	+++	++	++	++	+++	ND	ND	ND
CΔ34	−	−	−	−	−	−	ND	ND	ND
CΔ40	−	−	−	−	−	−	ND	ND	ND
CΔ45	−	−	−	−	−	−	++	ND	−
CΔ86	−	−	−	−	−	−	++	ND	−
CΔ144	−	−	−	−	−	−	++	ND	−
CΔ232	−	−	−	−	−	−	++	ND	−
NΔ105/C209A	−	+	+	+	+/−	−	ND	−	ND

[a] Activity is based on WT IN activity. Symbols: +/−, 0 to 5%; +, 6 to 35%; ++, 36 to 75%, +++, 76 to 100%; −, no activity. ND, not determined.

analysis suggests that the central core functions along with the C terminus, rather than the N terminus, and does not need to be present on both molecules. The 71-amino-acid region, residues 106 to 177, was required in *cis* with the C-terminal portion of the IN, and not the N-terminal portion for reconstitution of integration activity.

Effects of protein concentration and high reductant levels on strand transfer and disintegration. IN mutants lacking the HHCC domain typically exhibit impaired strand transfer, although a low level of in vitro activity was noted with NΔ105 (Fig. 3C). Several biochemical observations indirectly correlated the redox state of C209A to the level of strand transfer activity. These observations include the following: (i) activities of C209A and WT M-MuLV IN (the activity of C209A was frequently higher than that of the WT M-MuLV IN), (ii) the site specificity of NΔ105 strand transfer, and (iii) increased yields of active NΔ105 and other M-MuLV IN proteins obtained when the protein concentration was limited to 2 to 5 μM during renaturation. As C-209 is the sole Cys present in NΔ105, these data suggested that the redox state of the M-MuLV IN proteins may affect activity. To address this possibility, the effects of elevated DTT and protein concentration on strand transfer (Fig. 7A) and disintegration reactions (Fig. 7B) were investigated.

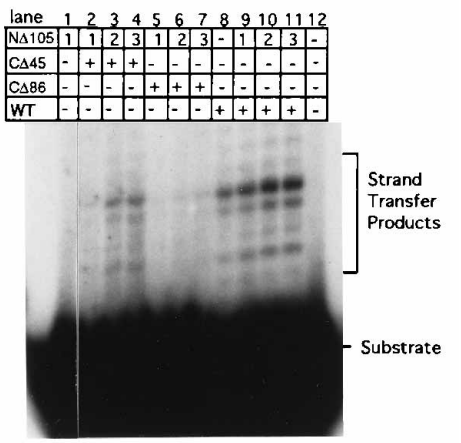

FIG. 5. Complementation of mutant M-MuLV INs for strand transfer activity. Strand transfer activity for NΔ105 (lane 1) and WT IN (lane 8) are presented. Strand transfer activity for pairs are shown for NΔ105 and CΔ45 (lanes 2 to 4), NΔ105 and CΔ86 (lanes 5 to 7), and NΔ105 and WT (lanes 9 to 11). Protein pairs were preincubated in reaction buffer on ice for 15 min. Reactions were performed at 30°C as described in Materials and Methods with 12 pmol of CΔ45, CΔ86, and WT IN and 0.5 pmol of oligonucleotide 2784/2785 strand transfer substrate. NΔ105 was added to each of the reaction mixtures as indicated: 1, 5 pmol; 2, 10 pmol; and 3, 15 pmol.

Standard assay conditions contain 10 mM DTT, and two levels of protein (10 or 20 pmol) which served as standards for protein concentrations used in complementation reactions. With WT M-MuLV IN, strand transfer was efficient (Fig. 7A, lanes 2 and 3). The CΔ232 finger-containing protein was inactive regardless of concentration (Fig. 7A, lanes 4 and 5). C209A, which contains the HHCC domain but lacks the sulfhydryl group of C-209, was capable of strand transfer at both protein concentrations (Fig. 7A, lanes 6 and 7) with an efficiency slightly lower than that of WT M-MuLV IN. NΔ105 demonstrated limited strand transfer activity, and integration appeared restricted to a single site (Fig. 7A, lanes 8 and 9). The NΔ105/C209A double mutant also displayed limited strand transfer activity at a lower protein concentration (Fig. 7A, lane 10); however, the sites of integration were more disperse than observed with NΔ105 (Fig. 7A, compare lane 10 with lane 8). This was more dramatic at the higher NΔ105/C209A protein concentration (Fig. 7A, compare lane 11 with lane 9), with the strand transfer pattern approximating that of WT M-MuLV IN. In complementation assays, CΔ232 had a negligible effect on the strand transfer activity of intermediate concentrations of WT and C209A M-MuLV IN proteins (Fig. 7A, lanes 12 and 13, respectively). The complementation of NΔ105 by CΔ232 restored strand transfer activity to near WT levels (Fig. 7A, lane 14). Significantly, the NΔ105/C209A double mutant could not be complemented by CΔ232 (Fig. 7A, lane 15).

To determine the effects of reducing agents on activity, a fixed concentration of each of the M-MuLV IN proteins was batch treated with 40 mM DTT. The DTT treatments are shown for both standard (10 pmol) and higher (20 pmol) levels

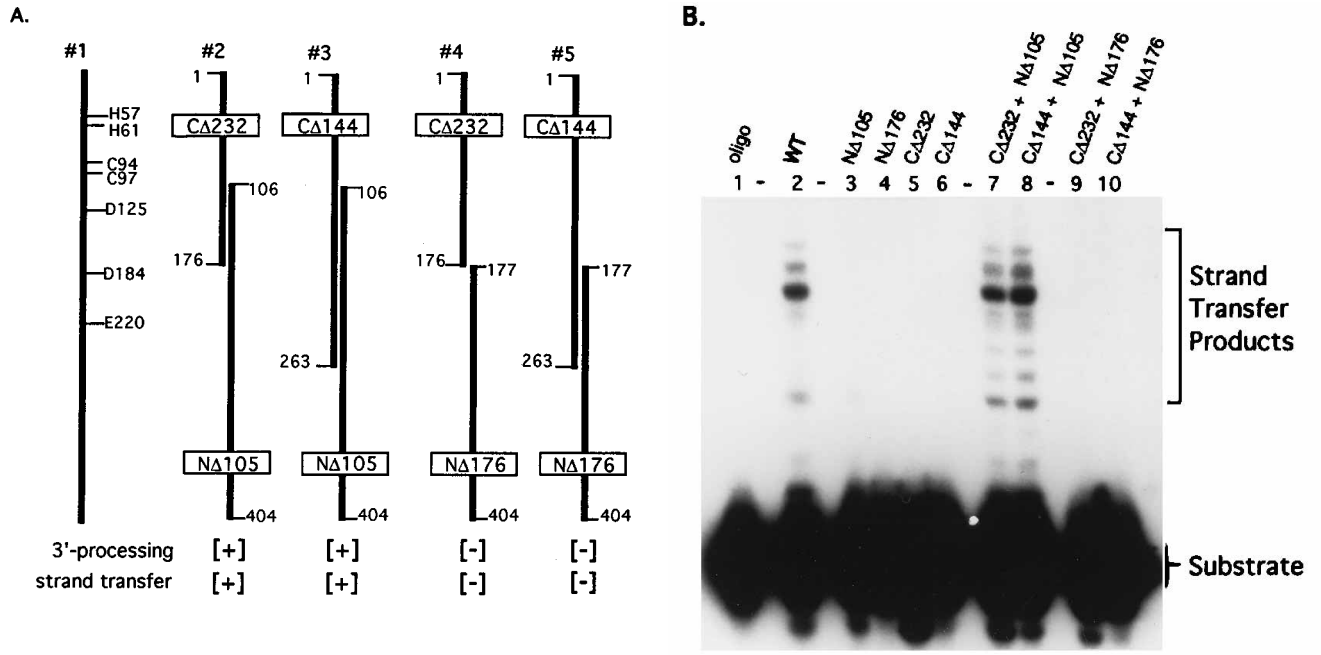

FIG. 6. Complementation of mutant M-MuLV INs for strand transfer activity. (A) Schematic illustration of protein pairs mixed for complementation analysis in panel B. Protein pairs, #1, WT M-MuLV IN shown with HHCC and DDE motifs; #2, CΔ232 and NΔ105; #3, CΔ144 and NΔ105; #4, CΔ232 and NΔ176; #5, CΔ144 and NΔ176. The ability of the protein pairs to complement is presented in brackets at the bottom of each set. +, complements; −, does not complement. Amino acids are numbered as described in Materials and Methods. (B) Strand transfer activities are presented for WT IN (lane 2), NΔ105 (lane 3), NΔ176 (lane 4), CΔ232 (lane 5), and CΔ144 (lane 6) individually and in pairs for CΔ232 and NΔ105 (lane 7), CΔ144 and NΔ105 (lane 8), CΔ232 and NΔ176 (lane 9), CΔ144 and NΔ176 (lane 10). Substrate without the addition of protein is shown in lane 1. Equimolar concentrations of protein pairs were preincubated at 37°C in reaction buffer without substrate. Reactions were performed at 37°C as described in Materials and Methods with the amounts of proteins indicated above and analyzed on 20% denaturing polyacrylamide gels.

of protein. The DTT-treated WT M-MuLV IN (Fig. 7A, lanes 16 and 17) and C209A proteins (Fig. 7A, lanes 20 and 21) demonstrated only slight enhancement of their strand transfer activity (compare these lanes with lanes 2 and 3 and lanes 6 and 7, respectively). The strand transfer levels mediated by 10 pmol of DTT-treated NΔ105 also showed only slight enhancement (Fig. 7A, compare lane 22 with lane 8). However, at a higher enzyme concentration, the DTT-treated NΔ105 protein displayed a dramatic enhancement in strand transfer activity and integration site usage (Fig. 7A, lane 23) versus untreated NΔ105 (Fig. 7A, compare lane 23 with lane 9). The basal strand transfer activity of DTT-treated NΔ105/C209A was essentially unaffected (Fig. 7A, lanes 24 and 25). The levels of strand transfer activity in complementation reaction mixtures containing DTT-treated CΔ232 plus WT M-MuLV IN (Fig. 7A, lane 26), CΔ232 plus C209A (Fig. 7A, lane 27), CΔ232 plus NΔ105 (Fig. 7A, lane 28), and CΔ232 plus NΔ105/C209A (Fig. 7A, lane 29) remained unchanged in comparison with those in the untreated complementation reaction mixtures.

Since all of the protein domains in M-MuLV IN are more stringently required for integration than for disintegration, it was possible that the effect of reducing conditions on M-MuLV IN activity could be specific to the strand transfer reactions. To more directly address this question, the DTT batch-treated M-MuLV IN proteins were simultaneously assayed in disintegration reactions (Fig. 7B). For reaction mixtures containing the WT M-MuLV IN (Fig. 7B, lanes 2 and 3) and C209A (Fig. 7B, lanes 6 and 7), NΔ105 (Fig. 7B, lanes 8 and 9), and NΔ105/C209A mutants, no significant differences were seen for disintegration reactions regardless of the reducing conditions. In the complementation reactions, mixtures of CΔ232 plus WT M-MuLV IN (Fig. 7B, lane 12), CΔ232 plus C209A (Fig. 7B, lane 13), CΔ232 plus NΔ105 (Fig. 7B, lane 14), and CΔ232 plus NΔ105/C209A (Fig. 7B, lane 15) did not show any stimulation of disintegration. Under stronger reducing conditions, complementation mixtures containing DTT-treated CΔ232 plus WT M-MuLV IN (Fig. 7B, lane 26), CΔ232 plus C209A (Fig. 7B, lane 27), CΔ232 plus NΔ105 (Fig. 7B, lane 28), and CΔ232 plus NΔ105/C209A (Fig. 7B, lane 29) also did not demonstrate stimulation of disintegration.

DISCUSSION

Sixteen mutants were generated from the M-MuLV IN to define the domains required for catalysis. Two of the C-terminal mutants exhibited proteolysis during expression and extraction and were not examined further. Overall, both N- and C-terminal deletions drastically reduced integration activity, while disintegration was affected more by the C-terminal truncations. A deletion into the N terminus through amino acid 105 was observed to retain partial strand transfer activity and near WT disintegration activity. The strand transfer activity showed a target preference for one site. HIV-1 IN with a comparable deletion does not exhibit integration activity (46), but Rous sarcoma virus IN with the same deletion exhibits 3′ processing and strand transfer activities (4). In the case of Rous sarcoma virus, the corresponding N-terminal deletion mutant retains greater activity when a short peptide is fused to its N terminus. The researchers suggest that the peptide may be providing stability to the folding of the protein or orientation of the active multimer. Work with the M-MuLV IN mutants suggested that the redox state of Cys-209 may play a role in the active state of NΔ105. To address the molecular basis for the

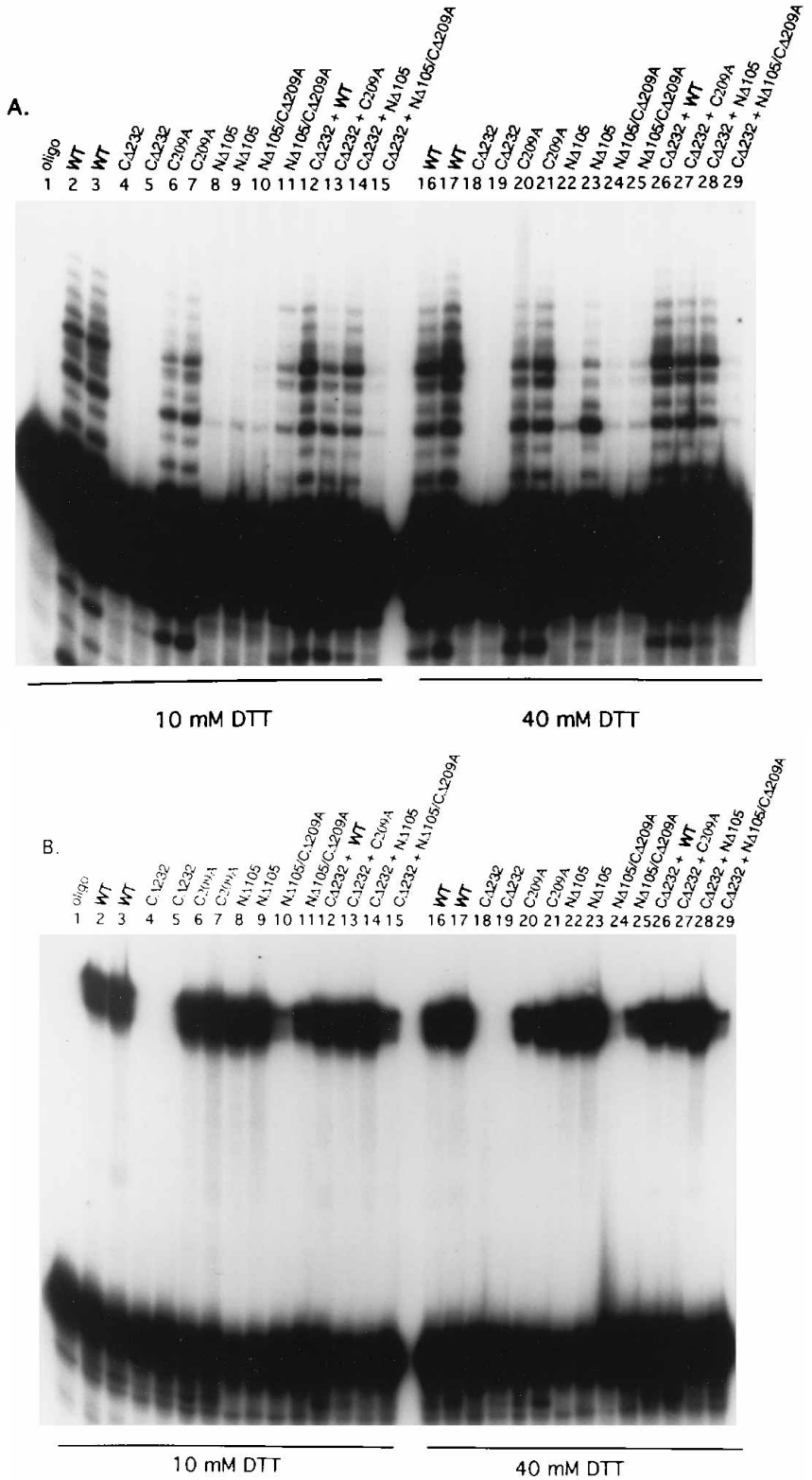

FIG. 7. Enzymatic activities in the presence of DTT (A) Strand transfer activities of untreated enzymes (10 mM DTT) (lanes 1 to 15) and enzymes treated with 40 mM DTT (lanes 16 to 29). Lane 1, substrate alone; lane 2, 10 pmol of WT IN; lane 3, 20 pmol of WT IN; lane 4, 10 pmol of CΔ232; lane 5, 20 pmol of CΔ232; lane 6, 10 pmol of C209A; lane 7, 20 pmol of C209A; lane 8, 10 pmol of NΔ105; lane 9, 20 pmol of NΔ105; lane 10, 10 pmol of NΔ105/C209A; lane 11, 20 pmol of NΔ105/C209A; lane 12, 10 pmol of CΔ232 and 10 pmol of WT IN; lane 13, 10 pmol of CΔ232 and 10 pmol of C209A; lane 14, 10 pmol of CΔ232 and 10 pmol of NΔ105; lane 15, 10 pmol of CΔ232 and 10 pmol of NΔ105/C209A; lane 16, substrate alone; lane 17, 10 pmol of WT IN; lane 18, 20 pmol of WT IN; lane 19, 10 pmol of CΔ232; lane 20, 20 pmol of CΔ232; lane 21, 10 pmol of C209A; lane 22, 20 pmol of C209A; lane 23, 10 pmol of NΔ105; lane 24, 20 pmol of NΔ105; lane 25, 10 pmol of NΔ105/C209A; lane 26, 20 pmol of NΔ105/C209A; lane 27, 10 pmol of CΔ232 and 10 pmol of WT IN; lane 28, 10 pmol of CΔ232 and 10 pmol of C209A and 10 pmol of C209A; lane 29, 10 pmol of CΔ232 and 10 pmol of NΔ105; lane 30, 10 pmol of CΔ232 and 10 pmol of NΔ105/C209A. (B) Disintegration activities of untreated enzymes (10 mM DTT) (lanes 1 to 15) and enzymes treated with 40 mM DTT (lanes 16 to 29). Lane 1, substrate alone; lane 2, 10 pmol of WT IN; lane 3, 20 pmol of WT IN; lane 4, 10 pmol of CΔ232; lane 5, 20 pmol of CΔ232; lane 6, 10 pmol of C209A; lane 7, 20 pmol of C209A; lane 8, 10 pmol of NΔ105; lane 9, 20 pmol of NΔ105; lane 10, 10 pmol of NΔ105/C209A; lane 11, 20 pmol of NΔ105/C209A; lane 12, 10 pmol of CΔ232 and 10 pmol of WT IN; lane 13, 10 pmol of CΔ232 and 10

observed activity of the M-MuLV N-terminal mutant NΔ105, reactions were titrated with NΔ105 protein pretreated with a reducing agent. At the higher enzyme concentrations examined, the strand transfer levels of the DTT-treated NΔ105 protein demonstrated enhanced activity and integration site usage. Further study of the aggregation state of the protein is warranted. One interpretation of this data is that the NΔ105 may form disulfide-linked multimers which are limited in access to the target DNA. High levels of DTT would break the cystine bonds and allow the assembly of active multimers. Similarly, evidence that a cysteine residue proximal to the HHCC domain of HIV-1 IN may assist in multimeric assembly of IN-DNA complexes has been suggested (13). Alternatively, these data may suggest two conformation states of NΔ105 brought about by either weak and strong reducing conditions. One can speculate that if the HHCC region were responsible for a more ordered active multimer, then the reduced conformation could promote either strand transfer randomly along the DNA or random association with DNA. NΔ176, which has a larger N-terminal deletion of 176 amino acids, was completely inactivated for integration and disintegration. The lack of activity for NΔ176 is not surprising, since the region deleted in this mutant includes the first D, D-125, of the invariant DD(35)E motif. This residue is absolutely required for mediation of integration and disintegration activity in all retroviral INs examined to date (15, 26, 28). Complementation studies showed that the M-MuLV D-125 residue was essential as part of the central core and C terminus and not the HHCC.

Two specific amino acid substitutions in the catalytic core region were examined, K227A and C209A. K-227 is conserved throughout all retroviral INs. Substitutions at this position in Rous sarcoma virus and HIV-1 IN retain some 3′ processing and strand transfer activity (26). It was thus unexpected that a mutation at this position was not tolerated in M-MuLV IN. This suggests that the K-227 residue plays a greater role in the integration processes for M-MuLV IN. A cysteine (C-209 for M-MuLV) is located within the D35E region of the majority of, but not all, INs. Given the reactive nature of Cys residues, studies addressed whether the residue afforded any specificity to the integration events. Surprisingly, C209A retained near WT levels of integration and disintegration activity. And as discussed in the context of NΔ105 above, the reduction of C-209 was required for a productive integration event in the absence of a HHCC domain. For disintegration, the effect of DTT treatment coupled to the mutation of C-209 to Ala is not as apparent as for integration. Disintegration is a unimolecular reaction, whereas strand transfer requires a bimolecular coordination of an LTR (donor) and target (acceptor) molecules. This suggests that the HHCC domain and redox state of C-209 are more important for the multimeric assembly and stability of bimolecular IN-DNA complexes. In unimolecular reactions such as disintegration, complex assembly may be assisted by the fixed structure of the Y substrate. This may diminish the reliance of M-MuLV IN on the more-specific interactions required for assembly of integration complexes. The decreased activity of NΔ105/C209A in disintegration might simply be due to protein-DNA complex instability.

The C terminus of the M-MuLV IN was sensitive to deletions and was observed to lose both integration and disintegration activities with deletions of 28 or more amino acids. A linker insertion at amino acid 368 results in a temperature-sensitive protein with defective M-MuLV IN function and replication (41). CΔ22 and CΔ28 functioned in ways similar to the WT IN and were active on all substrates including Y_{CD}. The lack of LTR sequence in Y_{CD} substrate indicates a unimolecular, nonspecific phosphoryl transfer reaction. This type of activity has been noted previously with M-MuLV IN disintegration substrate variants (22). Recently, HIV-1 IN has also been shown to catalyze a sequence-independent phosphoryl transfer (30, 43). Several of the M-MuLV IN mutants reported in this study, CΔ40, CΔ45, CΔ86, CΔ232, and CΔ144, had no detectable integration or disintegration activity. This is in contrast with HIV-1 IN which shows disintegration activity in C-terminal mutants with deletions to amino acid 212, albeit at very low levels (3). Several of the M-MuLV C-terminal mutants were able to function in *trans* with NΔ105. In the case of HIV-1 IN, efficient complementation does not occur if one molecule has more than 50 residues deleted from the N terminus and the other molecule has 70 residues deleted from the C terminus (46). In comparison to the results with HIV-1 IN, domains within the M-MuLV IN were able to act in *trans* in smaller functional units to promote integration activity. The results of the complementation analysis suggest that the HHCC region functions as one domain and that the catalytic core and C terminus function as a separate domain.

The HHCC region has been proposed to be involved in the recognition of the retroviral LTR termini, multimerization, positioning of the viral DNA into the catalytic core, and maintenance of the architecture of the protomeric complex (4, 22, 25). Studies of the strand transfer activity of NΔ105, which lacks the HHCC region, offered several new insights into a possible function for the HHCC region. Several observations were made. First, NΔ105 catalyzed strand transfer reactions at a low level; however, this activity was limited to one target site in the DNA substrate. This site is 4 nucleotides closer to the 3′ end than the site primarily targeted for strand transfer with the WT M-MuLV IN enzyme. Second, reactions with a mixture of NΔ105 with either of the two mutant M-MuLV IN proteins, CΔ232 or CΔ144, produced integration levels similar to that of WT IN. Thus, in these experiments, the HHCC domain functioned in *trans* to the core-C terminus to regain 3′ processing and strand transfer activities. Third, increasing the amount of DTT in the reaction environment of NΔ105 functioned in a manner similar to the addition of CΔ232, specifically, through the stimulation of integration activity. Last, while the integration activity of NΔ105 could be increased through the addition of DTT or CΔ232, the double mutant, NΔ105/C209A, could not be complemented by CΔ232. This suggests that the C-209 residue is in some manner required for IN-IN interactions between the NΔ105 protein and the HHCC region of the CΔ232 protein. Similarly, Ellison et al. have proposed a model for HIV-1 IN multimerization on the basis of interactions between the HHCC domain of one protomer and an *N*-ethylmaleimide-sensitive site on a second protomer (13). The lack of complementation of NΔ105/C209A could indicate improper folding of the active site. Although misfolding cannot be precluded, the demonstrated low-level strand transfer activity and

pmol of C209A; lane 14, 10 pmol of CΔ232 and 10 pmol of NΔ105; lane 15, 10 pmol of CΔ232 and 10 pmol of NΔ105/C209A; lane 16, 10 pmol of WT IN; lane 17, 20 pmol of WT IN; lane 18, 10 pmol of CΔ232; lane 19, 20 pmol of CΔ232; lane 20, 10 pmol of C209A; lane 21, 20 pmol of C209A; lane 22, 10 pmol of NΔ105; lane 23, 20 pmol of NΔ105; lane 24, 10 pmol of NΔ105/C209A; lane 25, 20 pmol of NΔ105/C209A; lane 26, 10 pmol of CΔ232 and 10 pmol of WT; lane 27, 10 pmol of CΔ232 and 10 pmol of C209A; lane 28, 10 pmol of CΔ232 and 10 pmol of NΔ105; lane 29, 10 pmol of CΔ232 and 10 pmol of NΔ105/C209A. Reactions were performed at 37°C as described in Materials and Methods and analyzed on 20% denaturing polyacrylamide gels.

multiple sites of integration demonstrated by NΔ105/C209A argue against this. These results suggest that the redox state of C-209 affects M-MuLV IN integration activity, particularly strand transfer, and that the HHCC domain assists in controlling the redox state of the C-209 residue. Whether this effect is due to a direct interaction of the HHCC and the core Cys or an indirect effect is an area of current investigation.

ACKNOWLEDGMENTS

We thank Kimberly C. Gray and Clarivick Macalalad for excellent technical assistance.

This work was supported by American Cancer Society grant VM-134 and by National Science Foundation grant DMB-9105091. M.J.R. was a Stohlmann Scholar of the Leukemia Society of America, Inc. C.B.J. was supported by NIH Postdoctoral Fellowship grant IF32CA09242. G.A.D. was supported by NIH training grant 5T32A107043-02.

REFERENCES

1. **Bowerman, B., P. O. Brown, J. M. Bishop, and H. E. Varmus.** 1989. A nucleoprotein complex mediates the integration of retroviral DNA. Genes Dev. **3:**469–478.
2. **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.
3. **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.
4. **Bushman, F. D., and B. Wang.** 1994. Rous sarcoma virus integrase protein: mapping functions for catalysis and substrate binding. J. Virol. **68:**2215–2223.
5. **Chow, S. A., and P. O. Brown.** 1995. Juxtaposition of two viral DNA ends in a bimolecular disintegration reaction mediated by multimers of human immunodeficiency virus type 1 or murine leukemia virus integrase. J. Virol. **68:**7869–7878.
6. **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.
7. **Coffin, J. M.** 1990. Retroviridae and their replication, p. 645–708. *In* B. N. Fields and D. M. Knipe (ed.), Fundamental virology. Raven Press, New York.
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. **Donehower, L. A.** 1988. Analysis of mutant Moloney murine leukemia viruses containing linker insertion mutations in the 3′ region of *pol*. J. Virol. **62:** 3958–3964.
11. **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.
12. **Dotan, I., B. P. Scottoline, T. S. Heuer, and P. O. Brown.** 1995. Characterization of recombinant murine leukemia virus integrase. J. Virol. **69:**456–468.
13. **Ellison, V., J. Gerton, K. A. Vincent, and P. O. Brown.** 1995. An essential interaction between distinct domains of HIV-1 integrase mediates assembly of the active multimer. J. Biol. Chem. **270:**3320–3326.
14. **Engelman, A., F. D. Bushman, and R. Craigie.** 1993. Identification of discrete functional domains of HIV-1 integrase and their organization within an active multimeric complex. EMBO J. **12:**3269–3275.
15. **Engelman, 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.
16. **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.
17. **Farnet, C. M., and W. A. Haseltine.** 1991. Determination of viral proteins present in the human immunodeficiency virus type 1 preintegration complex. J. Virol. **65:**1910–1915.
18. **Felkner, R. H., and M. J. Roth.** 1992. Mutational analysis of the N-linked glycosylation sites of the SU protein of Moloney murine leukemia virus. J. Virol. **66:**4258–4264.
19. **Ho, S. N., H. D. Huut, R. M. Horton, J. K. Pullen, and L. R. Pease.** 1989. Site-directed mutagenesis by overlap extension using the polymerase chain reaction. Gene **77:**51–59.
20. **Johnson, M. S., M. A. McClure, D. F. Feng, J. Gray, and R. F. Doolittle.** 1986. Computer analysis of retroviral pol genes: assignment of enzymatic functions to specific sequences and homologies with nonviral enzymes. Proc. Natl. Acad. Sci. USA **83:**7648–7652.
21. **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.
22. **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.
23. **Katz, R. A., and A. M. Skalka.** 1994. The retroviral enzymes. Annu. Rev. Biochem. **63:**133–173.
24. **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.
25. **Khan, E., J. P. G. Mack, R. A. Katz, J. Kulkosky, and A. M. Skalka.** 1990. Retroviral integrase domains: DNA binding and the recognition of LTR sequences. Nucleic Acids Res. **19:**851–860.
26. **Kulkosky, J., K. S. Jones, R. A. Katz, J. P. G. Mack, and A. M. Skalka.** 1992. Residues critical for retroviral integrative recombination in a region that is highly conserved among retroviral/retrotransposon integrases and bacterial insertion sequence transposases. Mol. Cell. Biol. **12:**2331–2338.
27. **Kunkel, T. A., J. D. Roberts, and R. A. Zakour.** 1987. Rapid and efficient site-specific mutagenesis without a phenotypic selection. Methods Enzymol. **154:**367.
28. **Leavitt, A. D., L. Shiue, and H. E. Varmus.** 1993. Site-directed mutagenesis of HIV-1 integrase demonstrates differential effects on integrase functions in vitro. J. Biol. Chem. **268:**2113–2119.
29. **Lin, T. H., T. P. Quinn, D. Grandgenett, and M. T. Walsh.** 1989. Secondary structural analysis of retrovirus integrase: characterization by circular dichroism and empirical prediction methods. Proteins Struct. Funct. Genet. **5:**156–165.
30. **Mazumder, A., A. Engelman, R. Craigie, M. Fesen, and Y. Pommier.** 1994. Intermolecular disintegration and intramolecular strand transfer activities of wild type and mutant HIV-1 integrase. Nucleic Acids Res. **22:**1037–1043.
31. **Miller, D. G., M. A. Adam, and A. D. Miller.** 1990. Gene transfer by retrovirus vectors occurs only in cells that are actively replicating at the time of infection. Mol. Cell. Biol. **10:**4239–4242.
32. **Mumm, S. R., and D. P. Grandgenett.** 1991. Defining nucleic acid-binding properties of avian retrovirus integrase by deletion analysis. J. Virol. **65:** 1160–1167.
33. **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.
34. **Murphy, J. E., and S. P. Goff.** 1992. A mutation at one end of Moloney murine leukemia virus DNA blocks cleavage of both ends by the viral integrase in vivo. J. Virol. **66:**5092–5095.
35. **Puras-Lutzke, R., C. Vink, and R. H. A. Plasterk.** 1994. Characterization of the minimal DNA-binding domain of the HIV integrase protein. Nucleic Acids Res. **22:**4125–4131.
36. **Roe, T., T. C. Reynolds, G. Yu, and P. O. Brown.** 1993. Integration of murine leukemia virus DNA depends on mitosis. EMBO J. **12:**2099–2108.
37. **Roth, M., P. Schwartzberg, N. Tanese, and S. P. Goff.** 1990. Analysis of mutations in the integration function of Moloney murine leukemia virus: effects on DNA binding and cutting. J. Virol. **64:**4709–4717.
38. **Roth, M. J.** 1991. Mutational analysis of the carboxyl terminus of the Moloney murine leukemia virus integration protein. J. Virol. **65:**2141–2145.
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. **Sanger, F., S. Nicklen, and A. R. Coulson.** 1978. DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. USA **74:**5463–5467.
41. **Schwartzberg, P. L., M. J. Roth, N. Tanese, and S. P. Goff.** 1993. Analysis of a temperature-sensitive mutation affecting the integration protein of Moloney murine leukemia virus. Virology **192:**673–678.
42. **Shinnick, T. M., R. A. Lerner, and J. G. Sutcliffe.** 1981. Nucleotide sequence of Moloney murine leukemia virus. Nature (London) **293:**543–548.
43. **van Den Ent, F. I., C. Vink, and R. H. A. Plasterk.** 1995. DNA substrate activities of the human immunodeficiency virus type 1 integrase protein. J. Virol. **68:**7825–7832.
44. **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.
45. **van Gent, D. C., A. A. M. O. Groenger, and R. H. A. Plasterk.** 1992. Mutational analysis of the integrase protein of human immunodeficiency virus type 2. Proc. Natl. Acad. Sci. USA **89:**9598–9602.
46. **van Gent, D. C., C. Vink, A. A. M. O. Groeneger, and R. H. A. Plasterk.** 1993. Complementation between HIV integrase proteins mutated in different domains. EMBO J. **12:**3261–3267.
47. **Varmus, H., and P. Brown.** 1989. Retroviruses. American Society for Microbiology, Washington, D.C.
48. **Vink, C., A. A. M. O. Groenger, 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. **Vink, C., E. Yeheskiely, G. A. van der Marel, J. H. van Boom, and R. H. A. Plasterk.** 1993. Site specific hydrolysis and alcoholysis of human immunodeficiency virus DNA termini mediated by the viral integrase protein. Nucleic Acids Res. **19:**6691–6698.
50. **Vogelstein, B., and D. Gillespie.** 1979. Preparative and analytical purification of DNA from agarose. Proc. Natl. Acad. Sci. USA **76:**615–619.
51. **Woerner, A. M., M. Klutch, J. G. Levin, and C. J. Marcus-Sekura.** 1992. Localization of DNA binding activity of HIV-1 integrase to the C-terminal half of the protein. AIDS Res. Hum. Retroviruses **8:**297–304.
52. **Woerner, A. M., and C. J. Marcus-Sekura.** 1993. Characterization of a DNA binding domain in the C-terminus of HIV-1 integrase by deletion mutagenesis. Nucleic Acids Res. **21:**3507–3511.