JOURNAL OF VIROLOGY, Sept. 1995, p. 5687–5696 Vol. 69, No. 9
0022-538X/95/$04.00+0

Mapping Domains of Retroviral Integrase Responsible for Viral DNA Specificity and Target Site Selection by Analysis of Chimeras between Human Immunodeficiency Virus Type 1 and Visna Virus Integrases

MICHAEL KATZMAN[1,2]* AND MALGORZATA SUDOL[1]

Department of Medicine[1] and Department of Microbiology and Immunology,[2] The Milton S. Hershey Medical Center, Pennsylvania State University College of Medicine, Hershey, Pennsylvania 17033

Received 21 March 1995/Accepted 13 June 1995

Human immunodeficiency virus type 1 (HIV-1) and visna virus integrases were purified from a bacterial expression system and assayed on oligonucleotide substrates derived from each terminus of human immunodeficiency virus type 1 and visna virus linear DNA. Three differences between the proteins were identified, including levels of specific 3′-end processing, patterns of strand transfer, and target site preferences. To map domains of integrase (IN) responsible for viral DNA specificity and target site selection, we constructed and purified chimeric proteins in which the N-terminal, central, and C-terminal regions of these lentiviral integrases were exchanged. All six chimeric proteins were active for disintegration, demonstrating that the active site in the central region of each chimera maintained a functional conformation. Analysis of endonucleolytic processing activity indicated that the N terminus of IN does not contribute to viral DNA specificity; this function must reside in the central region or C terminus of IN. In the viral DNA integration assay, chimeric proteins gave novel patterns of strand transfer products which did not match that of either wild-type IN. Thus, target site selection with a viral DNA terminus as nucleophile could not be mapped to regions of IN defined by these boundaries and may involve interactions between regions. In contrast, when target site preferences were monitored with a new assay in which glycerol stimulates IN-mediated cleavage of nonviral DNA, chimeras clearly segregated between the two wild-type patterns. Target site selection for this nonspecific alcoholysis activity mapped to the central region of IN. This report represents the first detailed description of functional chimeras between any two retroviral integrases.

Oligonucleotide-based assays have demonstrated three enzymatic activities for purified retroviral integrases. Processing refers to the site-specific endonuclease (29) or alcoholysis (60) activity that places nicks following the invariant CA typically found two nucleotides from the 3′ ends of viral DNA. Strand transfer denotes the joining of processed DNA ends to various sites on other oligonucleotides in reactions that model integration (10, 27). Disintegration indicates the ability of integrase (IN) to reverse the strand transfer reaction by resolving a complex that mimics one end of viral DNA integrated into host DNA (9).

Three regions of human immunodeficiency virus (HIV) IN have been defined on the basis of sequence alignments (30, 32), sensitivity to proteolysis (16, 57), refolding studies (6), mutagenesis, and functional complementation experiments (15, 55). The N terminus (amino acid residues 1 to 49) contains an HHCC zinc finger motif conserved among retroviral and retrotransposon integrases (4, 6, 32). Single-amino-acid substitutions of the conserved His and Cys residues generally have adverse effects on processing and strand transfer (16, 32, 35, 37, 53, 56), and proteins with deletions of this region are greatly impaired or inactive for these two activities (6, 13, 15, 45, 55–57). In contrast, this region is not required for disintegration (6, 16, 37, 53, 55–57). The central region (residues 50 to 186) has three acidic amino acids forming a D,D(35)E motif that is highly conserved among retroviral and retrotransposon integrases and the transposases of some bacterial insertion sequence elements (32). Single-amino-acid substitutions at these sites can knock out all three actions of IN (13, 16, 33, 35, 37, 53), and this domain is sufficient to catalyze disintegration (6, 57). The C terminus (residues 187 to 288), which is conserved only among related retroviruses, binds DNA nonspecifically (17, 38, 45, 57, 61, 62). Proteins with deletions of this end of IN lose processing and strand transfer activities but retain disintegration function (13, 55, 57). On the basis of these results, the active site that mediates all three phosphoryl or polynucleotidyl transfer reactions described above has been assigned to the central region of IN. However, the locations of the viral and host DNA binding sites have not been determined.

Analysis of chimeric proteins can be a useful strategy for mapping functional domains of enzymes if the wild-type proteins meet certain criteria. Unless the domains to be swapped are entirely independent, chimeras are more likely to be soluble and functional when the two proteins are related by sequence or structure. Optimal reaction conditions for the proteins should be similar so that activities of chimeras aren't masked by assay parameters. Finally, reliable assays should yield distinguishable and trackable results for the wild-type proteins. By these criteria, visna virus IN is particularly well suited for forming functional chimeras with HIV-1 IN.

A previous alignment and phylogenetic analysis showed that the integrases of HIV-1 and visna virus have separated from those of avian, murine, bovine, and human leukemia viruses (49). These two proteins share 89 of 288 (30.9%) amino acid

* Corresponding author. Mailing address: Department of Medicine, The Milton S. Hershey Medical Center, Pennsylvania State University College of Medicine, P.O. Box 850, Hershey, PA 17033. Phone: (717) 531-8881. Fax: (717) 531-4633.

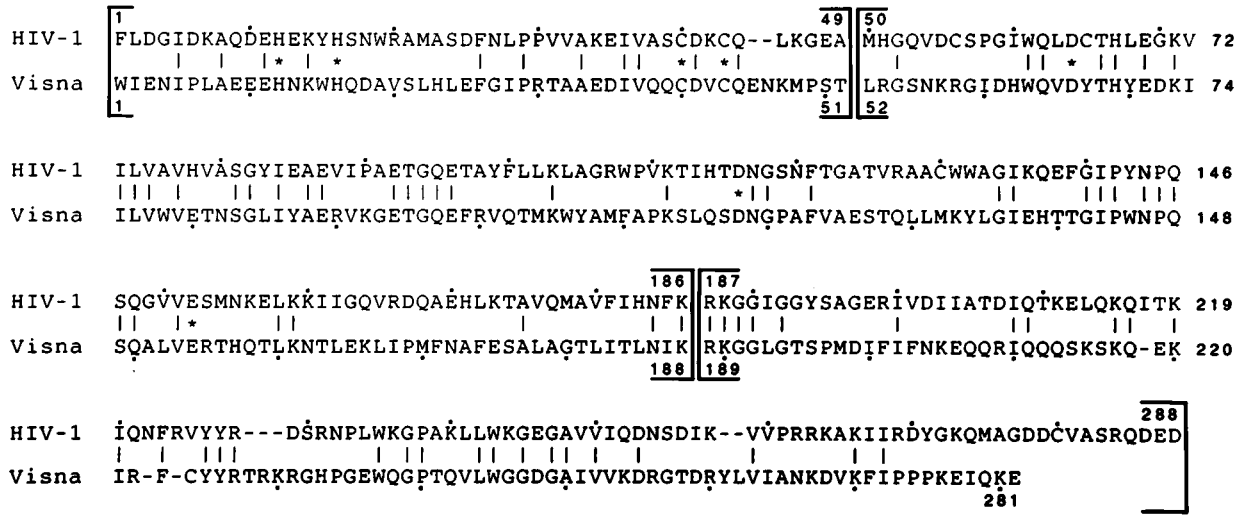

FIG. 1. Sequence comparison between HIV-1 and visna virus integrases. The single-letter amino acid code is shown; dashes represent gaps introduced by the alignment. The conserved zinc finger and D,D(35)E motifs are highlighted by asterisks; other identical amino acids are noted by a vertical line. Every 10th amino acid is marked by a dot, and the number of the last residue on each line is indicated at the right. Previously described IN domains that formed the basis for the chimeric proteins are indicated by brackets and numbering of the boundary positions.

positions, including 15 of 49 (30.6%) N-terminus amino acids, 45 of 137 (32.8%) central region residues, and 29 of 102 (28.4%) C-terminus residues (Fig. 1 [data expressed with reference to the HIV-1 sequence]). The integrases of HIV-1 and HIV-2 (and the related simian immunodeficiency virus) are even more closely related. However, HIV-1 IN and HIV-2 IN were reported to exhibit comparable levels of processing activity on terminal DNA sequences derived from either virus (52), precluding the chance to map viral DNA specificity by creating chimeras between these proteins.

We recently reported successful expression, purification, and analysis of visna virus IN (31). Optimal reaction conditions defined in that report match those for HIV-1 IN purified in our laboratory (28). We now report that purified HIV-1 IN and visna virus IN exhibit various degrees of specific 3′-end processing activity on oligonucleotide substrates derived from either end of HIV-1 or visna virus linear DNA and demonstrate distinctly different patterns of strand transfer and target site preferences on the same substrates. To map regions of IN that are responsible for viral DNA specificity (the putative viral DNA binding site) and target site selection (the putative target DNA binding site), we constructed chimeras in which the three regions of the two enzymes were reciprocally exchanged and utilized various in vitro integrase assays to analyze the purified proteins.

MATERIALS AND METHODS

Sources of IN coding sequences. Plasmid pLJS10 contains the IN coding sequence from HIV-1_{HXB2} clone p22K56, which was derived from clone pHRT25; the latter contains the *pol* gene sequence of proviral clone pHXBc2 (44, 50). The 288-amino-acid IN protein encoded by pLJS10 thus derives from an infectious clone. Molecular clone 8-5 of Icelandic strain 1514 of visna virus was obtained from Janice Clements of Johns Hopkins University (2) and encodes a 281-amino-acid IN protein identical to that of infectious clones (31).

Cloning of IN sequences. The HIV-1 IN coding region of pLJS10 was amplified by PCR. The 5′ primer for PCR, 5′GCGGATCC(ATCGAAGGTAGA)**TTT**TTGGATGGAATAGATAAGGCCCAAG3′, includes a *Bam*HI restriction site (underlined), followed by sequences coding for a factor Xa cleavage site (Ile-Glu-Gly-Arg in parentheses), and the first 28 nucleotides coding for IN (the codon for the first residue, Phe, is in boldface). The 3′ primer for PCR, 5′AG CAGCGAGCTC**CTA**ATCCTCATCCTGTCTACTTGC3′, includes an *Sst*I restriction site (underlined) and hybridizes at the 3′ end of the IN coding region (the triplet complementary to the termination codon is in boldface). Analogous cloning of visna virus IN sequences, but with a *Bgl*II site on the 5′ primer because of the occurrence of a *Bam*HI site within the coding region, was described previously (31). PCR products were digested with *Bam*HI (or *Bgl*II) and *Sst*I and ligated into plasmid pQE-30 (QIAGEN, Inc., Chatsworth, Calif.) that had been digested with *Bam*HI and *Sst*I. Bacterial strain M15[pREP4] (QIAGEN) was transformed with the ligation reaction, and colonies resistant to ampicillin and kanamycin were screened for the presence of insert DNA by restriction endonuclease digestion of plasmid DNA and for production of IN after induction with isopropyl-β-D-thiogalactopyranoside (IPTG), as described previously (31).

For chimeric integrases, the amino acid sequences of HIV-1 IN and visna virus IN were aligned and boundaries of previously defined domains of HIV-1 IN (15, 16), which are areas of low hydrophobicity, were chosen as points of crossover. Chimeric DNA was produced by the overlap extension method (22), taking care that all overlap primers started 3′ to a T so as to minimize errors introduced by any nontemplated addition of a 3′ A by *Taq* polymerase (48). Target sequences for amplification were the wild-type HIV-1 IN and visna virus IN cloned into PQE-30. Cassettes encoding HIV-1 IN amino acids 1 to 49, 1 to 186, 50 to 186, 50 to 288, and 187 to 288 (Fig. 1) were amplified by the following pairs of primers, respectively: H1 and V2H1, H1 and V3H2, V1H2 and V3H2, V1H2 and H3, and V2H3 and H3. Analogous cassettes encoding visna virus amino acids 1 to 51, 1 to 188, 52 to 188, 52 to 281, and 189 to 281 (Fig. 1) were amplified by primer pairs V1 and H2V1, V1 and H3V2, H1V2 and H3V2, H1V2 and V3, and H2V3 and V3, respectively. Note that amino acids 186 to 190 of HIV-1 IN are identical to residues 188 to 192 of visna virus IN (Fig. 1). The outermost primers for HIV-1 IN or visna virus IN sequences, respectively, were as follows: H1, 5′GGAGATCT(ATCGAAGGTAGA)**TTT**TTAGATGG3′; H3, 5′AGCAGCGAGCTC**TTA**ATCCTCATCCTGTCTACTTGC3′; V1, 5′GGAGATCT(ATCGAAGGTAGA)**TGG**ATAGAAAATATTCCCCTAGCAGA3′; and V3, 5′CTAAGTGGAGCTCCTTTCCATGCCCATAGTGGCA3′, where *Bgl*II (H1 and V1) and *Sst*I (H3 and V3) sites are underlined, factor Xa coding sequences (H1 and V1) are in parentheses, and codons for the first amino acid (H1 and V1) or complementary to a termination triplet (H3) are in boldface; V3 hybridizes approximately 25 nucleotides downstream of the visna virus IN coding region and thus does not have a triplet in boldface. The sequences of the overlap primers were as follows (HIV-1 sequences are in uppercase, visna virus sequences in lowercase, and common sequences are underlined): V2H1, 5′gcctct taaGGCTTCTCCTTTAGCTGAC3′; V3H2, 5′agcccacccttttctTTTAAAATTG TGGATGAATACTG3′; V1H2, 5′gctagtacaATGCATGGACAAGTAGACTG TAG3′; V2H3, 5′ctaaatataaaaAGAAAAGGGGGGATTG3′; H2V1, 5′GTCC ATGCATtgtactaggcattttatttcttgac3′; H3V2, 5′CCCCCCTTTTCTtttatatttagagt aatgagggtc3′; H1V2, 5′AAAAGGAGAAGCCttaagaggcagtaataaaaggggca3′; and H2V3, 5′CCACAATTTTAAAagaaagggtgggctagg3′. Products of the expected size were purified and appropriately mixed in a second round of PCR with the necessary outermost primers (H1 or V1 and H3 or V3) to yield full-length chimeras. These products were digested with *Bgl*II and *Sst*I and ligated into the *Bam*HI and *Sst*I sites of pQE-30. Transformation of bacteria and selection for production of IN were as described above. As a result of the cloning scheme, all proteins carried 16 additional N-terminal amino acids: Met-Arg-Gly-Ser-His-His-His-His-His-His-Gly-Ser-Ile-Glu-Gly-Arg. The entire IN coding sequences of plasmids containing wild-type and chimeric IN DNA were determined in

order to ensure that the amino acid sequences of expressed proteins were correct. The sequences of a total of 15 clones were determined in order to identify a complete set of eight intact proteins.

Expression and purification of integrases. All proteins were expressed with the QIAexpress system type IV construct (QIAGEN), which adds six histidine residues to the N terminus of proteins expressed from the DNA insert. Culture and induction by IPTG were performed as described previously (31) with minor modifications. The pellet from a 250- to 500-ml culture was suspended in 2.5 ml of sonication buffer (50 mM sodium phosphate [pH 7.6], 10 mM 2-mercaptoethanol, 10% glycerol, 1 mM phenylmethylsulfonyl fluoride) and lysed by addition of lysozyme to 1 mg/ml and NaCl to 1 M and by sonication. The lysate was passed through an 18-gauge needle five times to decrease viscosity and centrifuged at 40,000 × *g* and 4°C for 45 min, the supernatant was harvested, and a second extraction was performed. The combined extracts were adjusted to 30 to 40 mM imidazole and stirred for 1 h at 4°C with 0.2 to 0.3 ml of washed Ni^{2+}-nitrilotriacetic acid agarose (Ni-NTA resin; QIAGEN). The slurry was poured into a small polypropylene column and allowed to settle. Unbound material was collected, the resin was washed with SBNT (sonication buffer with 1 M NaCl and 0.1% Triton X-100) containing 30 to 40 mM imidazole, and recombinant proteins were eluted with SBNT containing 300 mM imidazole. Purification was monitored by sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) with 12.8% separation gels and 3% stacking gels (acrylamide/methylene-bisacrylamide ratio, 37.5:1). IN-containing fractions were dialyzed individually two times against 500 volumes of dialysis buffer (50 mM Tris-HCl [pH 7.6], 1 M NaCl, 1 mM dithioerythritol, 0.1 mM EDTA, 0.1% Triton X-100, 10% glycerol). The dialyzed product was adjusted to 40% glycerol by addition of 1/2 volume of 100% glycerol, yielding a protein storage buffer containing 33 mM Tris-HCl (pH 7.6), 0.67 M NaCl, 0.7 mM dithioerythritol, 0.07 mM EDTA, 0.07% Triton X-100, and 40% glycerol, and aliquots were frozen. In some cases, detergent and glycerol were removed from purified proteins by dialysis to permit concentration in a centrifugal microconcentrator (Amicon, Beverly, Mass.), followed by adjustment to the same storage buffer. Protein concentrations were measured by comparison with Coomassie blue-stained standards with quantitation by a laser densitometer (Molecular Dynamics, Sunnyvale, Calif.) with software from Protein Database, Inc. (Huntington, N.Y.).

Oligonucleotides. All oligodeoxynucleotides used as assay substrates were gel purified after synthesis. The sequences of terminal 18-mer oligonucleotide substrates were as follows (the invariant CA dinucleotides are in boldface, and complementary strands are not shown): HIV-1 U5 plus strand, 5′TG GAAAATCTCTAG**CA**GT3′; HIV-1 U3 minus strand, 5′TGAATTAGCCCT TC**CA**GT3′; visna virus U5 plus strand, 5′CGGAGCGGATCTCG**CA**GC3′; and visna virus U3 minus strand, 5′GTTCTCTGTCCTGA**CA**GT3′. The 5′-end-labeled sequences were prepared with [γ-^{32}P]ATP (3,000 Ci/mmol; New England Nuclear, Boston, Mass.) and T4 polynucleotide kinase (Promega, Madison, Wis.). Labeled oligomers were purified on 20% polyacrylamide–7 M urea denaturing gels, visualized by autoradiography, eluted from gel slices in 0.5 M ammonium acetate-1 mM EDTA at 37°C, concentrated by binding to C_{18} resin minicolumns (Waters Chromatography, Milford, Mass.), eluted with 25% acetonitrile, and evaporated to dryness. Specific activities of radiolabeled oligonucleotides were 10^6 to 10^7 cpm/pmol. Sequence-specific markers for gel analysis were produced by the 3′- to 5′-exonuclease activity of snake venom phosphodiesterase (Sigma, St. Louis, Mo.) on 5′-radiolabeled oligonucleotides, as described previously (29).

Endonuclease (processing) and strand transfer (integration) assays. Double-stranded DNA substrates were prepared by addition of fourfold excess of unlabeled complementary oligonucleotide to the labeled strand and heating at 95°C for 5 min, followed by incubation at 37°C for 30 min and 4°C for 10 min. Standard 10-μl reaction mixtures contained 0.2 to 0.5 pmol of double-stranded DNA, 25 mM Tris-HCl (pH 8.0), 10 mM dithiothreitol, 10 mM $MnCl_2$, and 0.5 to 2.0 μl of IN or protein storage buffer. Some reaction mixtures utilized preprocessed duplex oligonucleotides in which the 5′-labeled strand was missing two nucleotides from the 3′ end. Reaction mixtures were incubated for 60 to 120 min at 37°C and then stopped by addition of 10 μl of loading buffer (95% formamide, 20 mM EDTA, 0.05% bromophenol blue, 0.05% xylene cyanol) and heating at 95°C for 5 min. For gel analysis, aliquots were loaded onto 20% polyacrylamide (acrylamide/methylene-bisacrylamide ratio, 19:1)–7 M urea denaturing gels, followed by electrophoresis at 55 W until the bromophenol blue dye had migrated 23 cm. Wet gels were autoradiographed at −80°C. Specific cleavage was demonstrated by the appearance of radiolabeled oligomers two nucleotides shorter than the starting substrate, and strand transfer was demonstrated by the appearance of radiolabeled products longer than the substrate (Fig. 2). The radioactivity of bands in wet gels was quantified with a Betascope (Betagen, Waltham, Mass.).

Disintegration assays. The reversal of strand transfer assays were done under conditions identical to those above, except that 0.03 pmol (10^5 cpm) of a Y-shaped complex of four oligonucleotides was used as substrate. This complex was designed to mimic the immediate product of integration of the HIV-1 U5 DNA end into a nonviral DNA sequence (Fig. 2). Complexes were prepared and then gel purified on native 15% polyacrylamide gels, as described previously (31).

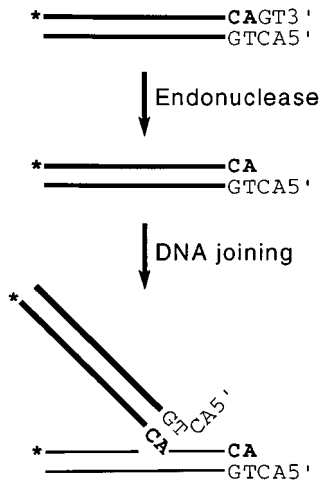

FIG. 2. In vitro integrase assays. Assays are shown schematically for the HIV-1 U5 substrate. With blunt-ended viral DNA, the endonuclease (processing) activity of IN produces a site-specific nick after the conserved CA, yielding a radioactive oligomer (asterisk) two nucleotides shorter than the substrate. The DNA joining (integration, strand transfer) activity of IN inserts the 3′-OH of the recessed CA terminus into various sites along other oligonucleotides (shown as thin lines), yielding labeled integration products longer than the substrate; for clarity of presentation, the donor DNA representation was rotated. The target strand may or may not be processed or labeled; some shorter labeled products can be accounted for when a radioactive oligonucleotide acts as the target. Both DNA nicking and joining can occur sequentially in one assay. IN also can catalyze the reverse of the DNA joining reaction, termed disintegration.

RESULTS

Viral DNA specificities and target site selectivities of HIV-1 and visna virus integrases. We recently reported that purified visna virus IN exhibits the full repertoire of in vitro activities characteristic of retroviral integrases. We also noted that visna virus IN had specific cleavage and strand transfer activities on oligonucleotide substrates derived from either end of HIV-1 DNA (31). HIV-1 IN purified in our laboratory in a similar manner has optimal reaction conditions identical to those for visna virus IN, including a temperature of 37°C, Tris buffer at pH 8.0, and 10 mM $MnCl_2$ as divalent cation (28). When assayed on four sets of duplex oligonucleotides that were designed to mimic both termini of blunt-ended linear DNA from either virus (29), three distinct differences between the two enzymes could be discerned.

Each IN exhibited specific cleavage after the invariant CA near the 3′ ends of all four substrates as demonstrated by the appearance of prominent bands two nucleotides shorter than the substrates (Fig. 3A, region B, and the shorter radiographic exposure of this region shown in Fig. 3B). Cleavage at these sites was specific on each substrate, but the efficiency of cleavage varied reproducibly between different combinations of DNA and IN. Although both enzymes cleaved the HIV-1 U5 terminus efficiently (Fig. 3B, lanes 3 and 4), only HIV-1 IN cleaved the HIV-1 U3 end well (compare lanes 7 and 8). In contrast, the visna virus U5 and U3 ends were both processed to a greater extent by visna virus IN (lanes 12 and 16) than by HIV-1 IN (lanes 11 and 15). No cleavage was evident if IN was omitted from reaction mixtures (lanes 2, 6, 10, and 14) or if parallel fractions obtained from a mock purification were utilized (data not shown). Comparison of specific cleavage of the two U3 substrates best discriminated between the sources of IN.

The oligonucleotide cleavage assay also reveals products longer than the substrate (Fig. 2). These bands have been

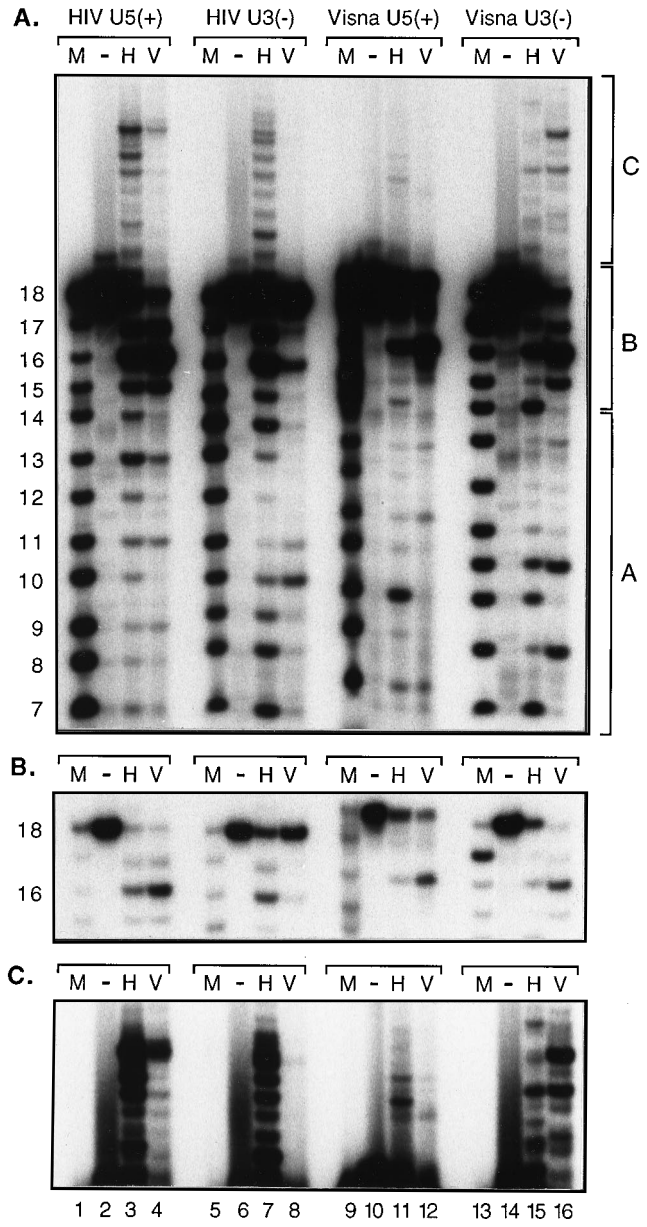

FIG. 3. Endonuclease and strand transfer activities of HIV-1 and visna virus integrases. Duplex oligonucleotide 18-mer substrates derived from the U5 or U3 termini of HIV-1 or visna virus DNA were 5′ labeled on the plus (+) or minus (−) strand, as indicated. For each DNA substrate, the four lanes represent a sequence-specific oligonucleotide ladder as markers (M), and incubation with protein buffer (−), HIV-1 IN (H), or visna virus IN (V) under standard conditions for 60 min. Analysis was performed as described in Materials and Methods. Site-specific cleavage is indicated by the appearance of a prominent band at the position of 16-mers (region B, a shorter radiographic exposure is shown in panel B). Distinct bands longer than the substrate represent strand transfer products (region C, a longer radiographic exposure is shown in panel C). Each IN also created a unique pattern of shorter products (region A). No major bands were detected in reaction lanes below the position of 4-mers. The sizes (nucleotides) of markers for the HIV-1 U5 plus strand are indicated at the left.

shown to represent insertion of processed 3′ ends of radiolabeled oligonucleotides, produced by the specific nicking activity of IN, into various sites on other oligonucleotides that act as surrogates for host DNA (10, 27). Both HIV-1 IN and visna virus IN created longer strand transfer products with all four DNA substrates (Fig. 3A, region C, and the longer exposure of this region shown in Fig. 3C). However, the amounts of these products varied for different combinations of DNA and IN. For substrates derived from either end of HIV-1 DNA, only HIV-1 IN was efficient at strand transfer (compare lanes 3 and 4 or 7 and 8). In contrast, neither enzyme was efficient for strand transfer on visna virus U5 DNA (lanes 11 and 12), but both yielded easily detected products with visna virus U3 DNA (lanes 15 and 16). Curiously, HIV-1 IN was efficient at strand transfer with visna virus U3 DNA, even though it was not efficient at cleaving this substrate (lane 15), suggesting that cleavage may have been rate limiting. However, visna virus IN processed the HIV-1 U5 terminus well but integrated it poorly (lane 4). Moreover, only moderate increases in strand transfer products were obtained by using preprocessed substrates, and the relative yields of products by the two enzymes were not altered (data not shown). Figure 3 also reveals that the pattern of strand transfer products on each substrate differed for the two enzymes, in both the location of some bands and the prominence of other bands. It has been shown previously that the pattern of strand transfer products is a function of the source of IN (7, 43, 52, 58) as well as the target DNA sequence (21, 36, 52). Product yield and ease of distinguishing strand transfer patterns (lanes 15 and 16) allowed visna virus U3 DNA to discriminate best between the sources of IN.

This assay revealed a third distinguishing characteristic between the two integrases. Each protein created a variety of shorter products in addition to the specific cleavage products two nucleotides shorter than the substrate (Fig. 3A, region A). Although the portion of the gel below the position of the 7-mers is not included in the figure, shorter products were seen to the level of 4-mers; only positions 1, 2, and 3 were not sites of cleavage. On a given DNA substrate, each IN created a distinctive pattern of these products and with an efficiency that made these patterns easily detectable. Site preferences on any of the DNA substrates for this endonuclease activity of IN, to be discussed in more detail later, clearly discriminated between the sources of IN.

Construction and purification of chimeras between HIV-1 and visna virus integrases. Having ascertained that these two related integrases had similar reaction conditions and three activities that could distinguish between them, we decided to create chimeric proteins in which the N-terminal, central, and C-terminal regions of HIV-1 IN and visna virus IN were exchanged. By convention, we refer to the wild-type proteins as HHH and VVV and to the six chimeras as VHH, HVV, HVH, VHV, HHV, and VVH, where the three letters represent the N terminus, central region, and C terminus, respectively, and H (HIV-1) or V (visna virus) indicates the source of that region. Chimeric DNA was constructed by the method of overlap extension PCR and cloned into a bacterial expression system. The wild-type and chimeric proteins were expressed with a short N-terminal extension including six histidine residues, facilitating rapid and efficient protein purification by metal affinity chromatography with a nickel chelating resin. Each native purification yielded a single prominent band by SDS-PAGE that migrated at positions appropriate for proteins with sizes of approximately 300 amino acids (Fig. 4). Densitometric scanning of a gel that included a series of protein standards for calibration indicated that the final concentrations of the purified proteins ranged from 33 ng/μl for VHV IN (lane 7) to 375 ng/μl for HVV IN (lane 5). Thus, each of the eight proteins was purified at a final concentration of ≥1 pmol/μl, comparable to or exceeding that of functional HIV-1 IN (52) or the avian retroviral IN utilized in the initial report that introduced the oligonucleotide cleavage assay (29).

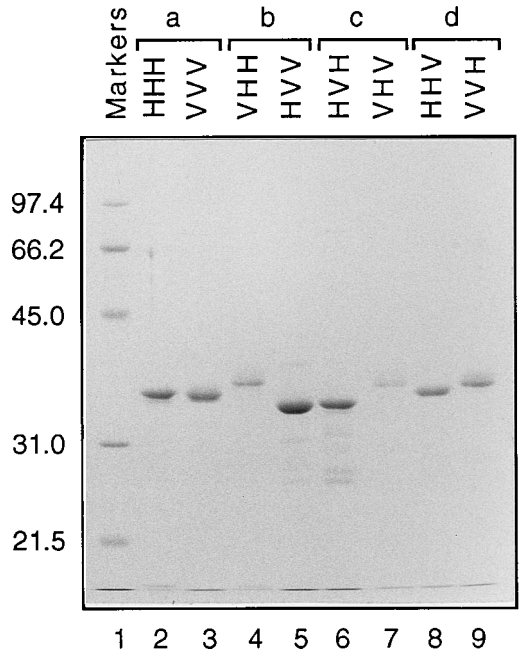

FIG. 4. SDS-PAGE demonstrating purification of chimeric integrases. Three microliters of each purified IN (indicated above the lanes, nomenclature as in Results) was heated in sample buffer and separated by SDS-PAGE, and the gel was stained with Coomassie blue. The proteins are paired in this and subsequent figures as follows: a, wild-type proteins; b, proteins with N termini exchanged; c, proteins with middle regions exchanged; d, proteins with C termini exchanged. Molecular mass markers are in lane 1 (sizes are shown in kilodaltons at the left) and represent 200 ng per band.

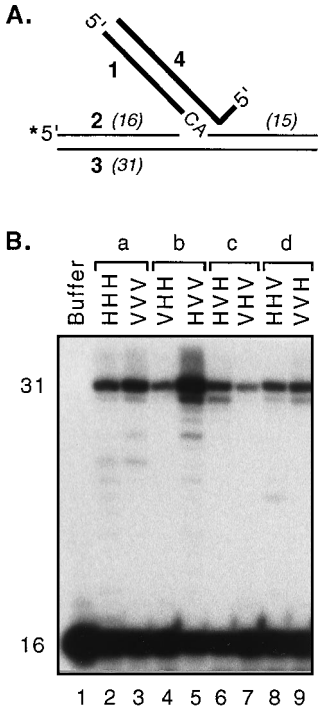

FIG. 5. Disintegration activity of chimeric integrases. (A) Schematic of the four-oligonucleotide Y-shaped substrate representing the predicted immediate product of the HIV-1 U5 DNA end (thick lines) integrated into host DNA (thin lines). Numbers in boldface designate the oligomers, and numbers in parentheses are lengths in nucleotides. Oligomer 1 is the HIV-1 U5 plus strand without the final 2 nucleotides but extended by 15 irrelevant nucleotides (…AGCTCGAG GTCGACG3′). Oligomer 2 (5′GAGCTACGGATCCTCG3′) and oligomer 3 (5′CGTCGACCTCGAGCTCGAGGATCCGTAGCTC3′) are irrelevant 16- and 31-mers, respectively. Oligomer 4 is identical to the HIV-1 U5 minus strand 18-mer, and the two-nucleotide 5′ overhang is indicated. Reversal of the integration reaction (disintegration) occurs when the viral DNA end is released as a result of cleavage after the CA, with concomitant joining of 5′-radiolabeled oligomer 2 (asterisk) to the 15 nucleotides at the 3′ end of oligomer 1, yielding a new radioactive 31-mer. (B) The complex shown in panel A was incubated with protein buffer or purified integrases under standard conditions for 90 min and analyzed as described in Materials and Methods. The two wild-type integrases and all six chimeric integrases were active in this assay. Similar relative activities were observed when visna virus U5 sequences were substituted for HIV-1 sequences. For definitions of protein pairs (a, b, c, and d), see the legend to Fig. 4.

Demonstration of disintegration activity by wild-type and chimeric integrases. Of the processing, strand transfer, and disintegration activities of IN, the last is the least constrained by reaction conditions (24) and substrate DNA sequence requirements (9, 46) and is detected even when the N terminus, C terminus, or both are deleted from IN. This activity thus is an excellent way to test the integrity of the active site in the catalytic central region of IN. Using a four-oligonucleotide Y-shaped substrate that represented the predicted immediate product of the HIV-1 U5 terminus integrated into host DNA (Fig. 5A), we found that the two wild-type and all six chimeric integrases were active in this assay. Each protein converted the radiolabeled 16-mer to a 31-mer (Fig. 5B, lanes 2 to 9). The faint bands migrating between these positions may represent breakdown products or reintegration events (56, 57). The relative yields of 31-mer products were observed consistently with different preparations or amounts of the chimeric proteins and by substitution of visna virus U5 DNA sequences in place of HIV-1 U5 DNA sequences (oligomers 1 and 4 in Fig. 5A). Indeed, the comparable activities of the two wild-type integrases on these substrates indicate that the in vitro disintegration assay cannot discriminate between the sources of IN. More importantly, we conclude from this experiment that the catalytic core that resides in the central region of each chimeric protein was maintained in a functional conformation.

Specific endonuclease activity of purified integrases on terminal viral DNA sequences. Each purified protein was tested for processing activity on the four duplex oligonucleotide substrates described earlier. As before, preference for one or the other of the two U3 substrates most clearly distinguished between the wild-type integrases (compare cleavage by HHH IN [lane 2] and VVV IN [lane 3] in Fig. 6B and D). Of the six chimeric proteins, only HVV IN demonstrated specific processing activity, selectively cleaving two nucleotides from three of the four DNA substrates (lane 5 in Fig. 6A, C, and D). This protein appeared to be most active on visna virus U3 DNA. To provide confidence for this observation, we compared the results of replicate reactions ($n = 4$) of the HHH, VVV, and HVV proteins with the two U3 DNA substrates. The viral DNA sequence preference of HVV IN closely matched that of the wild-type VVV IN and was very different from that of the wild-type HHH IN, as indicated by the ratio of specific cleavage on visna U3 DNA to that on HIV-1 U3 DNA (Table 1). These results indicate that the N terminus of IN does not contribute to specificity on viral DNA.

Longer radiographic exposures of the gel shown in Fig. 6 revealed that each of the other chimeric proteins also nicked at sites corresponding to shortening of substrate DNA by two nucleotides. However, none of these proteins demonstrated specific nicking at that site. Since HVH IN appeared to have a relatively high level of cleaving activity (Fig. 6, lane 6), we also performed replicate reactions with this protein and found that the ratio of cleavage at the relevant site on the two U3 DNA

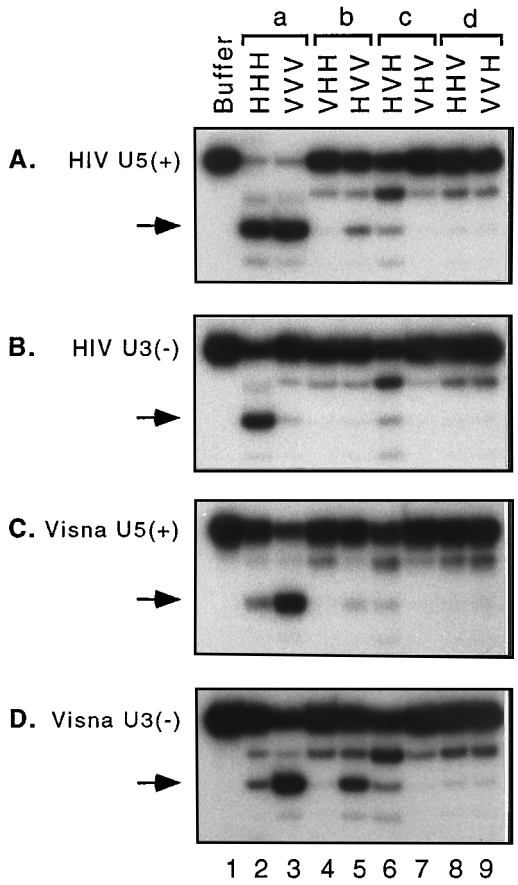

FIG. 6. Specific endonuclease activity of chimeric integrases. Duplex oligonucleotide 18-mer substrates derived from the U5 or U3 termini of HIV-1 or visna virus DNA and 5′ labeled on the plus (+) or minus (−) strand, as indicated, were incubated with protein buffer or purified integrases under standard conditions for 90 min and analyzed as described in Materials and Methods. Autoradiographs from a region of the gel comparable to that in Fig. 3B are shown. Biologically relevant specific cleavage products two nucleotides shorter than the substrate are indicated by arrows. For definitions of protein pairs (a, b, c, and d), see the legend to Fig. 4.

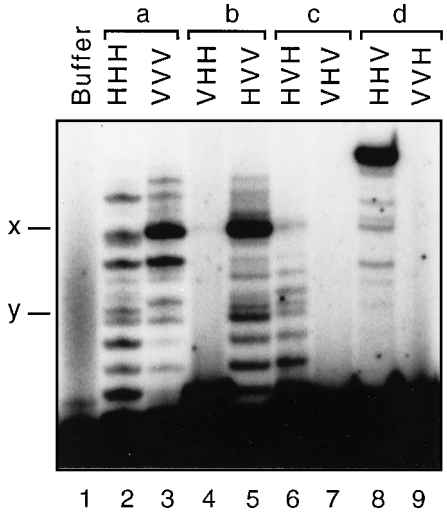

FIG. 7. Strand transfer activity of chimeric integrases. A duplex oligonucleotide derived from the visna virus U3 end but preprocessed by omission of the final two nucleotides from the minus strand was used as the substrate for the purified integrases. An autoradiograph from a region of the gel comparable to that in Fig. 3C is shown. The distinctive patterns produced by HHH IN and VVV IN are evident in lanes 2 and 3. Of the six chimeras, HVV IN (lane 5) was the most active in this assay; to display the patterns produced by the other chimeras, ~20× more cpm was loaded in lanes 4, 6, 7, 8, and 9, and a 50-fold concentrated preparation of HHV IN was assayed (lane 8). x denotes a heavy band common to lanes 3 and 5, and y indicates a light band common to lanes 2 and 5. For definitions of protein pairs (a, b, c, and d), see the legend to Fig. 4.

substrates was equal to 1. Use of concentrated protein preparations or mixing the chimeric proteins in all possible pairs before assaying for activity provided no new information.

Strand transfer patterns of wild-type and chimeric integrases. Each purified protein was tested for strand transfer activity on a preprocessed duplex oligonucleotide substrate derived from the visna virus U3 end. The distinctive patterns produced by the wild-type integrases are again evident (Fig. 7, lanes 2 and 3). HVV IN was the most active chimera in this assay (Fig. 7, lane 5). Somewhat surprisingly, none of the strand transfer patterns displayed by the chimeric proteins matched that of either wild-type IN. For example, HVV IN (Fig. 7, lane 5) created a prominent band (designated x) that aligns with a prominent band created by the wild-type VVV IN (lane 3) and a less-prominent yet distinct band (designated y) that is common to the wild-type HHH IN pattern (lane 2). Similarly, the locations and intensities of bands created by the action of HVH IN (Fig. 7, lane 6) did not match the pattern of either wild-type IN. The uniqueness of strand transfer patterns by chimeric integrases is best illustrated by HHV IN (Fig. 7, lane 8), which created a novel pattern of longer products. Although the reaction shown for HHV IN utilized a concentrated protein preparation, the slowly migrating prominent band near the top of the lane was detected with the unconcentrated HHV IN preparation (data not shown).

Strand transfer activities of the VHH IN (Fig. 7, lane 4), VHV IN (lane 7), and VVH IN (lane 9) were not enhanced by use of concentrated preparations of these proteins, addition of excess target DNA to reaction mixtures, or mixing of chimeric proteins. Testing purified proteins on processed substrates derived from the other three viral DNA termini provided no additional information. Similar results were obtained with a longer viral DNA-processed substrate, as well as with 3′-labeled nonviral DNA sequences as targets for insertion of unlabeled processed viral DNA ends (21, 36). In particular, HVV IN created a pattern of bands which by position and intensity matched that of neither wild-type IN, and HHV IN created a novel and prominent longer band (data not shown). We conclude from this series of experiments that target site selection with a viral DNA terminus as the nucleophile does not map clearly to one of these three regions of IN.

TABLE 1. Specific processing activity of HHH, VVV, and HVV integrases

IN	Specific cleavage of[a]:		Ratio of specific cleavages[b]
	HIV-1 U3 DNA	Visna virus U3 DNA	
HHH	33.4 ± 3.7	7.9 ± 4.1	0.2
VVV	3.5 ± 0.9	42.7 ± 9.9	12
HVV	1.8 ± 0.5	17.3 ± 3.4	10

[a] Duplex substrates were incubated with IN under standard conditions for 90 min and analyzed by denaturing gel electrophoresis, and counts per minute were quantified by Betascope counting. Specific cleavage was calculated as (cpm of 16-mer product + cpm of integration products)/(total cpm in lane). Values are means ± standard deviations of four reactions.

[b] (Specific cleavage of visna virus U3 DNA)/(specific cleavage of HIV-1 U3 DNA).

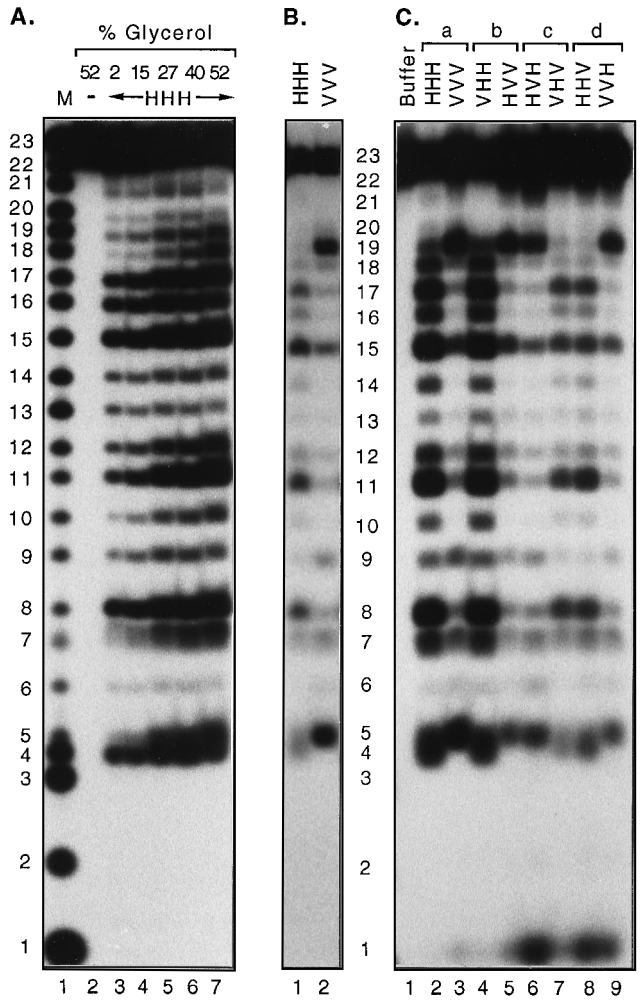

FIG. 8. Glycerol-stimulated alcoholysis assay on nonviral DNA. (A) A 5′-labeled 23-mer of nonviral sequence (5′GAGACTACGTTCGAGGATCCGA G3′) was annealed to a complementary oligonucleotide and incubated for 60 min with wild-type HIV-1 IN in the presence of increasing concentrations of glycerol (2% glycerol was contributed by IN and 0, 12.5, 25, 37.5, or 50% glycerol was added for reactions shown in lanes 3 to 7, respectively). Increasing amounts of cleavage products, representing IN-mediated cleavage of DNA by a nucleophile other than viral DNA, most likely glycerol itself (28), are noted at all positions except those very close to the end. A negative control incubated with 52% glycerol and protein buffer is in lane 2. A sequence-specific oligonucleotide ladder as markers (M) is in lane 1, and the sizes (nucleotides) are indicated at the left and between panels B and C. (B) The two wild-type integrases were tested on the substrate used in panel A in the presence of 44% glycerol during 90-min reactions. This exposure demonstrates the distinctive patterns of products created by the two enzymes. (C) Each IN was tested for cleavage activity on the substrate used in panel A during 90-min incubations in the presence of 44% glycerol. Equal volumes of reaction mixtures were loaded so that relative intensities reflect efficiencies of the different proteins. The six chimeras created patterns that segregated clearly to the HHH pattern (chimeras VHH, VHV, and HHV in lanes 4, 7, and 8, respectively) or to the VVV pattern (chimeras HVV, HVH, and VVH in lanes 5, 6, and 9, respectively). For definitions of protein pairs (a, b, c, and d), see the legend to Fig. 4.

Mapping target site selection in the absence of viral DNA. When extra glycerol was added to reaction mixtures to try to increase strand transfer (11), some shorter products were enhanced instead (28). This observation suggested that a nucleophile other than the processed viral DNA terminus was attacking target DNA and may have been responsible for some of the bands observed in region A of Fig. 3A. We therefore performed the following experiment. A 5′-labeled 23-mer of nonviral sequence was annealed to its unlabeled complementary sequence (in the form of a 24-mer that created a one-nucleotide 5′ overhang). This duplex DNA was incubated with HIV-1 IN in the absence of viral DNA ends but with 0 to 50% glycerol (in addition to the 2% provided by the protein buffer). Increasing amounts of glycerol clearly resulted in increasing amounts of cleavage products (Fig. 8A, lanes 3 to 7). Nicking was not evident in the presence of high concentrations of glycerol if IN was omitted from the reaction mixture (Fig. 8A, lane 2). These products obviously represent glycerol-stimulated, IN-mediated cleavage of DNA by a nucleophile other than viral DNA. Additional experiments indicated that glycerol itself acts as the nucleophile in these phosphoryl transfer reactions, which thus reveal a nonspecific alcoholysis activity for HIV-1 IN (28). These results were not dependent on the 5′ overhang in the substrate DNA, because extending the labeled 23-mer to a 24-mer to create flush ends yielded each of the bands seen in Fig. 8A, as did annealing to the complementary 23-mer (data not shown). A contaminating bacterial activity was unlikely given the results described below and was excluded as a possibility when parallel fractions from a mock purification did not exhibit this activity (data not shown).

Although cleavages were noted in Fig. 8A at all positions except those very close to the end, each site was not selected equally. Target site selection with the nucleophile utilized in this assay provided another way of discriminating between HIV-1 IN and visna virus IN. When the wild-type integrases were tested on the substrate used in Fig. 8A in the presence of 44% glycerol, easily distinguishable patterns of cleavage products were created (Fig. 8B). When the chimeric integrases were tested in this assay, all six proteins were active (Fig. 8C). In addition, each chimeric protein created a pattern of bands that closely matched the pattern of one or the other wild-type IN. Chimeras VHH, VHV, and HHV (lanes 4, 7, and 8, respectively), as did wild-type HHH IN (lane 2), created prominent bands at the 17, 15, 11, and 8 positions. Chimeras HVV, HVH, and VVH (lanes 5, 6, and 9, respectively), similar to wild-type VVV IN (lane 3), produced prominent bands at the 19, 15, and 5 positions. Thus, target site selection for the nonspecific alcoholysis activity of IN maps to the central region of IN.

DISCUSSION

Retroviral integrases must carry out two endonuclease events that are distinct in specificity, time, and location. After reverse transcription yields blunt-ended linear viral DNA, IN produces site-specific nicks following the invariant CA near the 3′ end of each DNA strand; this action can occur in the cytoplasm of the infected cell (3, 19). Within the cell nucleus, IN is responsible for insertion of the processed viral DNA ends into each strand of host DNA at cleavage sites that are separated by a small number of base pairs and that exhibit no sequence specificity. These actions are modeled with simple oligonucleotide substrates designed to mimic the viral DNA termini (29). For example, the integrases of avian myeloblastosis virus (29), HIV-1 (47), Moloney murine leukemia virus (10), Rous sarcoma virus (RSV) (27), HIV-2 (52), human foamy virus (42), feline immunodeficiency virus (58), and visna virus (31) specifically nick their cognate DNA ends and exhibit various levels of activity on sequences from noncognate DNA (7, 29, 34, 36, 46, 52, 58, 59). The specificity of integrases exhibited in vitro may even exceed that required in vivo, since the only selectivity necessary within the viral nucleoprotein complex (1) inside an infected cell may be recognition of the CA dinucleotides near viral DNA ends. In contrast, there is markedly diminished specificity for host DNA cleavage sites, both in vitro and in

vivo, when the processed viral DNA termini act as nucleophiles during strand transfer.

How IN distinguishes whether viral or host DNA is the target for nucleophilic attack remains to be determined, but one possibility is that IN recognizes viral DNA and host DNA in different binding sites (56, 60). Experimental support for the existence of two separate DNA binding sites recently was provided by kinetic data for Moloney murine leukemia virus IN (11) and HIV-1 IN (14). The viral DNA binding site has been proposed to reside in the zinc finger-containing N terminus, as suggested by some experimental data (6, 20, 25, 32, 35, 53, 56). The C terminus of IN, which has nonspecific DNA binding activity, is an attractive location for the host DNA binding site. However, convincing evidence in support of these assignments is lacking and the possibility of a single DNA binding site for HIV-1 IN exists (18, 38, 57). Our analysis of functional chimeras between two lentiviral integrases provides information directly relevant to these issues.

Viral DNA specificity. We found that the viral DNA preference of HVV IN matched that of wild-type VVV IN in the standard processing reaction (Fig. 6 and Table 1), proving that visna virus IN doesn't require its N-terminal amino acid sequences for viral DNA specificity. Furthermore, the results obtained with this chimera demonstrated that the N terminus of HIV-1 IN did not confer the specificity of HIV-1 IN on this active protein. A reasonable conclusion is that the N terminus and its zinc finger motif do not contribute to viral DNA specificity for either of these closely related lentiviral integrases. This conclusion is consistent with results from the avian system, in which RSV IN proteins with substitution or deletion of part of the zinc finger region sometimes exhibited residual processing and strand transfer activity (32). In addition, wild-type levels of strand transfer and modest levels of specific viral DNA processing were noted when the entire N-terminal region of RSV IN was replaced by short peptides (7). It has been difficult to evaluate specificity of HIV-1 IN proteins with N-terminal deletions, since these proteins were not active for processing or strand transfer (13, 45, 55–57). However, a small amount of residual processing and strand transfer activity by HIV-1 IN with N-terminal deletions replaced by short peptides has been noted (6, 7, 15). Further support for assignment of recognition of viral DNA ends to a region outside the N terminus of HIV-1 IN comes from a *trans* disintegration assay utilizing a crossbones substrate (8). HIV-1 IN proteins with an N-terminus substitution or deletion retained residual activity in this assay, which was shown to be dependent upon viral DNA ends (8). Retention of any specific activity in various assays in the absence of the N-terminal sequences of HIV-1 IN implies either that primary responsibility for viral DNA specificity lies outside this region or that there is redundancy of function.

The evidence by which others reasoned that the N terminus of IN is responsible for binding viral DNA must be reexamined. Bushman et al. (6) tentatively assigned a role for viral DNA binding to the N terminus, because a protein lacking this region was less active than wild-type IN on the standard disintegration substrate but had comparably low activity on a substrate that did not contain viral DNA sequences. The apparent sensitivity to the absence of viral DNA (i.e., greater loss of activity) when the N terminus of IN was present suggested that these parts interact. However, the levels of activity on the altered substrate were so low (only 0.1% conversion to products) that firm conclusions were difficult. Vincent et al. (56) found that proteins with amino acid substitutions in the conserved HHCC region had diminished activity on disintegration substrates in which the target DNA portion was altered. Dependence on interactions with target sequences suggested that the viral DNA binding sites of the mutated proteins had been perturbed and could not compensate for interference with target DNA binding. However, these proteins also were sensitive to alterations of the viral component of the disintegration substrates, especially the invariant CA dinucleotide, indicating that the HHCC motif is not required for recognizing the most conserved aspect of viral DNA termini. The authors noted that protein-protein interactions also could explain their findings. Analogous experiments have led to similar inferences for viral DNA specificity of Moloney murine leukemia virus IN (25). Others have noted that point substitutions of the HHCC residues sometimes impaired processing more than strand transfer (35, 53). It was suggested that the N terminus of IN must contain the viral DNA binding site, because processing depends upon viral DNA binding alone, whereas strand transfer involves more interactions. Whether these differences were statistically significant is not clear. Finally, Hazuda et al. (20) found that binding and UV cross-linking of HIV-1 IN to viral DNA terminal sequences were diminished when Ser was substituted for Cys at position 40. Function for the UV cross-linked complexes was not described. Moreover, the correlation of binding with in vitro activity was not precise, since the altered protein had diminished processing activity but wild-type levels of strand transfer on preprocessed substrates (35).

Experiments supporting a viral DNA binding site in the N terminus of IN were indirect and often required interpretation of subtle differences. In contrast, the retention of processing and strand transfer activity by RSV IN with various replacements of the N terminus (7), the residual activity of similar HIV-1 proteins (6, 15), and our functional analysis of chimeras of two lentiviral integrases strongly suggest that the viral DNA binding site of retroviral integrases resides in the central or C-terminal regions. Whereas the central region and C terminus of IN do bind many forms of DNA, the isolated N terminus does not (32, 41, 45, 57, 61, 62). An indirect role for the N terminus in positioning or recognizing viral DNA ends is still possible (12, 25, 57).

Target site selection. In the viral DNA integration assay, we found that strand transfer patterns of the chimeric integrases did not match that of either wild-type IN (Fig. 7). Thus, we could not map the host DNA binding site for viral DNA insertion to any of the three regions that formed the basis for these chimeric constructs. Chimeras utilizing regions defined by other boundaries might have resulted in the ability to map the target DNA binding site for strand transfer. It also is possible that regions of IN that are not contiguous along the primary amino acid sequence form the host DNA binding site for viral DNA insertion. Such interactions between different parts of IN might occur within one IN monomer and define a noncontiguous domain or between different IN monomers in the higher-order protein multimer (15, 23, 26, 55) and reveal a complex degree of communication between regions of IN.

The standard oligonucleotide IN assay also revealed distinct patterns of products that indicated DNA cleavages at sites other than the biologically relevant ones two nucleotides from the 3′-OH ends (Fig. 3A). These products cannot all be attributed to insertion of processed viral DNA ends into the 50% of DNA strands that are 5′ labeled (5, 10, 27, 31), because the total intensities of these bands far exceed that predicted by such a mechanism (i.e., 50% of the intensities of the longer strand transfer products [Fig. 2]). As suggested by Fig. 8, many of these products may stem from a newly identified activity of IN in which glycerol acts as a nucleophile in a nonspecific alcoholysis reaction (28). This alcoholysis activity of IN occurred at many sites in nonviral DNA and thus can be distinguished from the site-specific alcoholysis activity previously

described for IN on viral DNA termini (18, 31, 60). Whatever the significance of this activity, the patterns of target site selection in this assay provided a clearly discernible difference between the wild-type integrases. In addition, each of the chimeric proteins was active in this assay, as they were in the disintegration reaction. Target site selectivity for VHH, VHV, and HHV matched that of HHH IN, whereas the preferred sites for HVV, HVH, and VVH matched that of VVV IN. Thus, the site that binds target DNA for the nucleophile utilized in this assay appears to reside in the central region of IN. This result is consistent with the ability of the isolated central region to catalyze strand transfer between two gapped duplex DNA molecules in a reaction termed intermolecular disintegration (39); this reaction has been suggested to represent a sequence-independent phosphoryl transfer rather than a true reversal of integration (51).

The choice of nucleophile for viral DNA processing, which actually is a site-specific alcoholysis reaction (60), also maps to the central region of IN (54). Point mutations near active site residues Asp-116 and Glu-152 of HIV-2 IN interfered with its ability to use water or other soluble nucleophiles in the processing reaction (54) and increased utilization of the nearby 3′-OH end of viral DNA to produce cyclic dinucleotide products (16, 54). The central region of IN thus plays a role in the choice of a nucleophile and in positioning target DNA for that nucleophile. Yet, the central region of IN does not seem to be capable in a simple way of correctly positioning target DNA when the nucleophile is the processed viral DNA end, as suggested by the inability to map this activity to one region of IN. This paradox points out an important unresolved question regarding integration: how does viral DNA act sequentially as the target for processing and then as the nucleophile for integration? For bacteriophage Mu, higher-order protein-DNA complexes partially explain how a single active site can catalyze successive reactions, and similar mechanisms may be involved for retroviral integration (40).

This report represents the first detailed description of functional chimeras between any two retroviral integrases. These studies have mapped viral DNA specificity to the central or C-terminal region of IN. They also reveal that target site preferences with a viral DNA terminus as the nucleophile may involve interactions between regions of IN, but target site selection for the nonspecific alcoholysis activity of IN maps to the central region of IN. Future studies may localize these activities with greater precision and contribute to both our understanding of retroviral integration and efforts to interfere with this key step in HIV replication.

ACKNOWLEDGMENTS

This work was supported by Public Health Service grant R29 AI30759 from the National Institute of Allergy and Infectious Diseases. The following reagent was obtained through the AIDS Research and Reference Reagent Program, Division of AIDS, NIAID, NIH: pLJS10 (catalog number 1820) from J. M. Groarke, J. V. Hughes, and F. J. Dutko.

We thank James Groarke of Sterling Winthrop, Inc., Collegeville, Pa., for providing supplemental information about the derivation and sequence of pLJS10. We thank John W. Wills and Rebecca C. Craven for advice and encouragement.

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. **Braun, M. J., J. E. Clements, and M. A. Gonda.** 1987. The visna virus genome: evidence for a hypervariable site in the *env* gene and sequence homology among lentivirus envelope proteins. J. Virol. **61:**4046–4054.
3. **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.
4. **Burke, C. J., G. Sanyal, M. W. Bruner, J. A. Ryan, R. L. LaFemina, H. L. Robbins, A. S. Zeft, C. R. Middaugh, and M. G. Cordingley.** 1992. Structural implications of spectroscopic characterization of a putative zinc finger peptide from HIV-1 integrase. J. Biol. Chem. **267:**9639–9644.
5. **Bushman, F. D. and R. Craigie.** 1991. Activities of human immunodeficiency virus (HIV) integration protein *in vitro*: specific cleavage and integration of HIV DNA. Proc. Natl. Acad. Sci. USA **88:**1339–1343.
6. **Bushman, F. D., A. Engelman, 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.
7. **Bushman, F. D., and B. Wang.** 1994. Rous sarcoma virus integrase protein: mapping functions for catalysis and substrate binding. J. Virol. **68:**2215–2223.
8. **Chow, S. A., and P. O. Brown.** 1994. 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.
9. **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.
10. **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.
11. **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.
12. **Drelich, M., M. Haenggi, and J. Mous.** 1993. Conserved residues Pro-109 and Asp-116 are required for interaction of the human immunodeficiency virus type 1 integrase protein with its viral DNA substrate. J. Virol. **67:**5041–5044.
13. **Drelich, M., R. Wilhelm, and J. Mous.** 1992. Identification of amino acid residues critical for endonuclease and integration activities of HIV-1 IN protein *in vitro*. Virology **188:**459–468.
14. **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.
15. **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.
16. **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.
17. **Engelman, A., A. B. Hickman, and R. Craigie.** 1994. The core and carboxyl-terminal domains of the integrase protein of human immunodeficiency virus type 1 each contribute to nonspecific DNA binding. J. Virol. **68:**5911–5917.
18. **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.
19. **Fujiwara, T., and K. Mizuuchi.** 1988. Retroviral DNA integration: structure of an integration intermediate. Cell **54:**497–504.
20. **Hazuda, D. J., A. L. Wolfe, J. C. Hastings, H. L. Robbins, P. L. Graham, R. L. LaFemina, and E. A. Emini.** 1994. Viral long terminal repeat substrate binding characteristics of the human immunodeficiency virus type 1 integrase. J. Biol. Chem. **269:**3999–4004.
21. **Hong, T., E. Murphy, J. Groarke, and K. Drlica.** 1993. Human immunodeficiency virus type 1 DNA integration: fine structure target analysis using synthetic oligonucleotides. J. Virol. **67:**1127–1131.
22. **Horton, R. M., S. N. Ho, J. K. Pullen, H. D. Hunt, Z. Cai, and L. R. Pease.** 1993. Gene splicing by overlap extension. Methods Enzymol. **217:**270–279.
23. **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.
24. **Jonsson, C. B., G. A. Donzella, and M. J. Roth.** 1993. Characterization of the forward and reverse integration reactions of the Moloney murine virus integrase protein purified from *Escherichia coli*. J. Biol. Chem. **268:**1462–1469.
25. **Jonsson, C. B., and M. J. Roth.** 1993. Role of the His-Cys finger of Moloney murine leukemia virus integrase protein in integration and disintegration. J. Virol. **67:**5562–5571.
26. **Katz, R. A., J. P. G. Mack, G. Merkel, J. Kulkosky, Z. Ge, J. Leis, and A. M. Skalka.** 1992. Requirement for a conserved serine in both processing and joining activities of retroviral integrase. Proc. Natl. Acad. Sci. USA **89:**6741–6745.
27. **Katz, R. A., G. Merkel, J. Kulkosky, J. Leis, and A. M. Skalka.** 1990. The avian retroviral IN protein is both necessary and sufficient for integrative recombination in vitro. Cell **63:**87–95.
28. **Katzman, M.** Unpublished data.
29. **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.
30. **Katzman, M., J. P. G. Mack, A. M. Skalka, and J. Leis.** 1991. A covalent

complex between retroviral integrase and nicked substrate DNA. Proc. Natl. Acad. Sci. USA **88:**4695–4699.
31. **Katzman, M., and M. Sudol.** 1994. In vitro activities of purified visna virus integrase. J. Virol. **68:**3558–3569.
32. **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.
33. **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.
34. **LaFemina, R. L., P. L. Callahan, and M. G. Cordingley.** 1991. Substrate specificity of recombinant human immunodeficiency virus integrase protein. J. Virol. **65:**5624–5630.
35. **LaFemina, R. L., C. L. Schneider, H. L. Robbins, P. L. Callahan, K. LeGrow, E. Roth, W. A. Schleif, and E. A. Emini.** 1992. Requirement of active human immunodeficiency virus type 1 integrase enzyme for productive infection of human T-lymphoid cells. J. Virol. **66:**7414–7419.
36. **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.
37. **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.
38. **Lutzke, R. A. P., 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.
39. **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.
40. **Mizuuchi, K.** 1992. Transpositional recombination: mechanistic insights from studies of Mu and other elements. Annu. Rev. Biochem. **61:**1011–1051.
41. **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.
42. **Pahl, A., and R. M. Flügel.** 1993. Endonucleolytic cleavages and DNA-joining activities of the integration protein of human foamy virus. J. Virol. **67:**5426–5434.
43. **Pryciak, P. M., and H. E. Varmus.** 1992. Nucleosomes, DNA-binding proteins, and DNA sequence modulate retroviral integration target site selection. Cell **69:**769–780.
44. **Ratner, L., W. Haseltine, R. Patarca, K. J. Livak, B. Starcich, S. F. Josephs, E. R. Doran, J. A. Rafalski, E. A. Whitehorn, K. Baumeister, L. Ivanoff, S. R. Petteway, Jr., M. L. Pearson, J. A. Lautenberger, T. S. Papas, J. Ghrayeb, N. T. Chang, R. C. Gallo, and F. Wong-Staal.** 1985. Complete nucleotide sequence of the AIDS virus, HTLV-III. Nature (London) **313:**277–283.
45. **Schauer, M., and A. Billich.** 1992. The N-terminal region of HIV-1 integrase is required for integration activity, but not for DNA-binding. Biochem. Biophys. Res. Commun. **185:**874–880.
46. **Sherman, P. A., M. L. Dickson, and J. A. Fyfe.** 1992. Human immunodeficiency virus type 1 integration protein: DNA sequence requirements for cleaving and joining reactions. J. Virol. **66:**3593–3601.
47. **Sherman, P. A., and J. A. Fyfe.** 1990. Human immunodeficiency virus integration protein expression in *Escherichia coli* possesses selective DNA cleaving activity. Proc. Natl. Acad. Sci. USA **87:**5119–5123.
48. **Smith, K. D., A. Valenzuela, J. L. Vigna, K. Aalbers, and C. T. Lutz.** 1993. Unwanted mutations in PCR mutagenesis: avoiding the predictable. PCR Methods Appl. **2:**253–257.
49. **Sonigo, P., M. Alizon, K. Staskus, D. Klatzmann, S. Cole, O. Danos, E. Retzel, P. Tiollais, A. Haase, and S. Wain-Hobson.** 1985. Nucleotide sequence of the visna lentivirus: relationship to the AIDS virus. Cell **42:**369–382.
50. **Tanese, N., J. Sodroski, W. A. Haseltine, and S. P. Goff.** 1986. Expression of reverse transcriptase activity of human T-lymphotropic virus type III (HTLV-III/LAV) in *Escherichia coli*. J. Virol. **59:**743–745.
51. **van den Ent, F. M. I., C. Vink, and R. H. A. Plasterk.** 1994. DNA substrate requirements for different activities of the human immunodeficiency virus type 1 integrase protein. J. Virol. **68:**7825–7832.
52. **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.
53. **van Gent, D. C., A. A. M. Oude Groeneger, 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.
54. **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.
55. **van Gent, D. C., C. Vink, A. A. M. Oude Groeneger, and R. H. A. Plasterk.** 1993. Complementation between HIV integrase proteins mutated in different domains. EMBO J. **12:**3261–3267.
56. **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.
57. **Vink, C., A. A. M. Oude Groeneger, and R. H. A. Plasterk.** 1993. Identification of the catalytic and DNA-binding region of the human immunodeficiency virus type 1 integrase protein. Nucleic Acids Res. **21:**1419–1425.
58. **Vink, C., K. H. van der Linden, and R. H. Plasterk.** 1994. Activities of the feline immunodeficiency virus integrase protein produced in *Escherichia coli*. J. Virol. **68:**1468–1474.
59. **Vink, C., D. C. van Gent, Y. Elgersma, and R. H. A. Plasterk.** 1991. Human immunodeficiency virus integrase protein requires a subterminal position of its viral DNA recognition sequence for efficient cleavage. J. Virol. **65:**4636–4644.
60. **Vink, C., E. Yeheskiely, G. A. van der Marel, J. H. van Boom, 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.
61. **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. Human Retroviruses **8:**297–304.
62. **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.