JOURNAL OF VIROLOGY, June 1994, p. 3896–3907
0022-538X/94/$04.00+0

Vol. 68, No. 6

Substrate Features Important for Recognition and Catalysis by Human Immunodeficiency Virus Type 1 Integrase Identified by Using Novel DNA Substrates

SAMSON A. CHOW[1]† AND PATRICK O. BROWN[1,2,3]*

Department of Pediatrics,[1] Department of Biochemistry,[2] and Howard Hughes Medical Institute,[3] Stanford University Medical Center, Stanford, California 94305-5428

Received 17 December 1993/Accepted 17 March 1994

The integrase encoded by human immunodeficiency virus type 1 (HIV-1) is required for integration of viral DNA into the host cell chromosome. In vitro, integrase mediates a concerted cleavage-ligation reaction (strand transfer) that results in covalent attachment of viral DNA to target DNA. With a substrate that mimics the strand transfer product, integrase carries out disintegration, the reverse of the strand transfer reaction, resolving this integration intermediate into its viral and target DNA parts. We used a set of disintegration substrates to study the catalytic mechanism of HIV-1 integrase and the interaction between the protein and the viral and target DNA sequence. One substrate termed dumbbell consists of a single oligonucleotide that can fold to form a structure that mimics the integration intermediate. Kinetic analysis using the dumbbell substrate showed that integrase turned over, establishing that HIV-1 integrase is an enzyme. Analysis of the disintegration activity on the dumbbell substrate and its derivatives showed that both the viral and target DNA parts of the molecule were required for integrase recognition. Integrase recognized target DNA asymmetrically: the target DNA upstream of the viral DNA joining site played a much more important role than the downstream target DNA in protein-DNA interaction. The site of transesterification was determined by both the DNA sequence of the viral DNA end and the structure of the branched substrate. Using a series of disintegration substrates with various base modifications, we found that integrase had relaxed structural specificity for the hydroxyl group used in transesterification and could tolerate distortion of the double-helical structure of these DNA substrates.

Integration of a double-stranded DNA copy of the retrovirus genome into a chromosome of the host cell is essential for efficient viral replication (9, 24, 27, 30). Genetic studies have shown that retrovirus integration requires two viral components: integrase (encoded by the 3′ end of the *pol* gene) and DNA sequences at the ends of the viral long terminal repeats (for reviews on retrovirus DNA integration, see references 3, 17, and 35). Analysis of the integration reaction in vivo and in vitro has demonstrated that it proceeds in the following three steps: (i) 3′ processing, in which integrase removes 2 nucleotides from the 3′ end of each strand of linear viral DNA so that the viral 3′ ends terminate with the CA dinucleotide (5, 14, 22, 28); (ii) strand transfer, a concerted cleavage-ligation reaction during which integrase makes a staggered cut in the target DNA and ligates the recessed 3′ ends of the viral DNA to the 5′ ends of the target DNA at the cleavage site (5, 10, 13, 21), producing a gapped intermediate; and (iii) gap repair, in which the integration process is completed by removal of the 2 unpaired nucleotides at the 5′ ends of the viral DNA and repair of the gaps between the viral and target DNA sequences, thereby generating the short direct repeats that flank the provirus. The protein factors involved in the last step of the reaction remain to be characterized.

The first two steps of the reaction, which are integrase dependent, do not require the presence of a nucleotide cofactor as an energy source (4, 10), and each step proceeds by a one-step transesterification mechanism (12). Using a substrate that mimics the initial integration intermediate, we showed previously that human immunodeficiency virus type 1 (HIV-1) integrase can catalyze the strand transfer reaction in reverse and resolve the intermediate into its viral and target DNA parts. The reverse reaction is called disintegration (8). Integrase can also mediate DNA cleavage-ligation on an altered disintegration substrate in which the entire viral portion is single stranded, an activity we termed DNA splicing (7, 8, 31, 36, 37). The discovery of this DNA splicing activity led us to propose that integrase may have other biological roles during the retrovirus life cycle, such as viral 5′-end joining and viral recombination. As exemplified by the DNA splicing activity, the sequence and structure requirements for disintegration are less stringent than those for 3′ processing and strand transfer (8). This characteristic allows many genetic variants of integrase that lack detectable activity in 3′ processing and strand transfer to retain disintegration activity (7, 11, 26, 34, 36, 37). Thus, the disintegration assay has played an essential role in locating the catalytic domain of integrase and will be useful in mapping other functional domains of the protein.

Compared with the substrates used in other in vitro assays for integrase activity, the disintegration substrate has the unique property that the site of integration of viral DNA into target DNA is predetermined and can be manipulated. In this study, this property was exploited by using novel DNA substrates to identify features of viral and target DNA structure important for integrase recognition, the structural specificity of the nucleophile, and the sequence and structure involved in determining the site of transesterification during disintegra-

* Corresponding author. Mailing address: Howard Hughes Medical Institute, B255 Beckman Center, Stanford University Medical Center, Stanford, CA 94305-5428. Phone: (415) 725-7569. Fax: (415) 723-1399. Electronic mail address: pbrown@CMGM.stanford.edu.

† Present address: Department of Molecular and Medical Pharmacology, UCLA School of Medicine, Los Angeles, CA 90024.

tion. Our results show that integrase recognizes target DNA asymmetrically and that several features of the substrates do not require double-helical structure for recognition by the protein.

The reversibility of the strand transfer reaction suggested that integrase may function as an enzyme. The present study provides kinetic evidence proving that HIV-1 integrase is an enzyme.

MATERIALS AND METHODS

Enzymes. HIV-1 integrase was expressed in *Escherichia coli* using the T7 polymerase-promoter system (32) and purified as described previously (36). T4 polynucleotide kinase, T4 DNA ligase, and restriction endonucleases were purchased from New England Biolabs. Exonuclease-free Klenow fragment of *E. coli* DNA polymerase I was from United States Biochemical Corp. *E. coli* exonuclease III was from Stratagene and New England Biolabs. Terminal deoxyribonucleotidyl transferase was obtained from Boehringer Mannheim.

Nucleotides and oligonucleotides. Deoxyribonucleotides were purchased from Pharmacia LKB. [α-^{32}P]dATP, [α-^{32}P]dCTP, and [γ-^{32}P]ATP with specific activities of 3,000 to 4,500 Ci/mmol were obtained from ICN. [α-^{32}P]3′-dATP (cordycepin 5′-triphosphate) with a specific activity of 5,000 Ci/mmol was from New England Nuclear. [α-^{32}P]ddATP with a specific activity of 3,000 Ci/mmol was from Amersham. A controlled-pore glass column containing 2′,3′-dideoxyadenosine for oligonucleotide synthesis was obtained from Midland Certified Reagent Company. Oligonucleotides were purchased from Operon Technologies, Inc., and were purified by electrophoresis through a 15% denaturing polyacrylamide gel before use.

Disintegration assay for integrase activity. Oligonucleotides were labeled either at the 5′ end with T4 polynucleotide kinase and [γ-^{32}P]ATP or at the 3′ end with exonuclease-free Klenow fragment of *E. coli* DNA polymerase I and [α-^{32}P]dCTP (28a). Unincorporated radioactive nucleotides were removed from the labeled oligonucleotide by centrifugation through small columns (0.9-ml bed volume) of Sephadex G-15 (Sigma). Preparation of the Y oligomer (Fig. 1B) and its structurally related substrates for the disintegration assay was done as described previously (8). The novel DNA substrate, the dumbbell (Fig. 1C), and its derivatives were prepared by heating the oligonucleotide in a solution containing 10 mM Tris-HCl, (pH 7.5), 1 mM EDTA, and 0.1 M NaCl for 3 min at 80°C, then slowly cooling to room temperature, and finally chilling to 4°C. The reaction conditions for the disintegration assay were essentially the same as previously described (8). Typically in a 20-μl volume, the DNA substrate was incubated with HIV-1 integrase for 60 min at 37°C in a reaction buffer containing final concentrations of 20 mM *N*-2-hydroxyethylpiperazine-*N*′-2-ethanesulfonic acid (HEPES) (pH 7.5), 10 mM $MnCl_2$, 10 mM dithiothreitol, 0.1 mM 3-[(3-cholamidopropyl)-dimethylammonio]-1-propanesulfonate (CHAPS), and 0.05% Nonidet P-40. The concentrations of integrase used throughout this report refer to protomers and not necessarily active enzyme molecules. The reaction was stopped by the addition of 18 mM EDTA and 0.1% sodium dodecyl sulfate. The reaction products were heated at 90°C for 3 min before analysis by electrophoresis on 15% polyacrylamide gels with 7 M urea in Tris-borate-EDTA buffer. Quantitation of the products was carried out with a Molecular Dynamics Densitometer or a Molecular Dynamics PhosphorImager.

Ethylation interference. The conditions of ethylation were essentially the same as those of Siebenlist and Gilbert (31a). Oligonucleotides, labeled at the 5′ termini, were suspended in 50 mM sodium cacodylate (pH 8.0). An equal volume of ethanol, saturated with *N*-nitroso-*N*-ethylurea (Sigma) was added, and the solution was incubated at 50°C for 30 min. After incubation, the solution was spun through a Sephadex G-15 column (0.9-ml bed volume). The DNA was further purified by electrophoresis on a 15% denaturing polyacrylamide gel. The labeled, ethylated DNA was then annealed with other oligonucleotides to form the Y oligomer. Disintegration of the ethylated substrate was carried out as described above, and the labeled products were isolated on a 15% native or denaturing polyacrylamide gel. The isolated disintegration products were resuspended in 0.15 M NaOH and cleaved at ethyl phosphates by heating at 90°C for 30 min. An equal volume of formamide buffer was added, and the cleaved products were analyzed on a 15% denaturing polyacrylamide gel.

RESULTS

The dumbbell substrate undergoes disintegration mediated by HIV-1 integrase. The dumbbell substrate is made up of a single oligonucleotide 38 nucleotides long. The folded structure of the dumbbell substrate depicted in Fig. 1C was based on previous results obtained from studies examining the formation of hairpins using similar DNA sequences (1, 2). We have not determined the structure of the dumbbell substrate experimentally or the influence of other factors, such as helical stacking, on the final structure of the substrate. However, we believe that most, if not all, of the structural features shown are correct because of the following results. (i) The substrate was recognized by integrase and formed expected products (see below). (ii) The substrate migrated at a faster rate on a denaturing (7 M urea) polyacrylamide gel than oligonucleotides of an identical length but composed of random sequences (data not shown). The rapid mobility is a characteristic shared by other oligonucleotides that form stable hairpins (18, 19). (iii) More than 95% of the substrate migrated as a single band with a size corresponding to that of a monomolecule on a native polyacrylamide gel, indicating that species resulting from annealing of two or more molecules were not formed (data not shown). The folded substrate has 5 bp of viral sequence and 10 bp of nonviral sequence. The nonviral sequences, which are arbitrary, were used as target DNA and are referred to as such in this report. Disintegration, the reverse of the strand transfer reaction, is a concerted strand cleavage-ligation reaction (8). The strand cleavage occurs precisely at the junction between the viral and target DNA sequences and is coupled to the rejoining of the interrupted target DNA sequences. Disintegration of the dumbbell substrate was therefore expected to form two products: a 14-nucleotide hairpin and a 24-nucleotide circular molecule. The reaction was monitored by labeling either the 5′ or 3′ end of the oligonucleotide. In both 5′- and 3′-labeled substrates, we detected the appearance of products with the expected sizes (Fig. 2). Similar to the disintegration of the Y oligomer, the reaction required integrase and Mn^{2+} (8). The 24-nucleotide molecule, after its isolation from the gel, showed the expected cleavage patterns after digestion with *Hpa*II and *Hae*III (data not shown). We verified that the isolated product was a closed loop by its inability to undergo chain elongation by terminal deoxyribonucleotidyl transferase and its resistance to digestion by *E. coli* exonuclease III under conditions where control substrates corresponding to the unligated target DNA part of the dumbbell molecule were acted on by these enzymes (data not shown). Demonstration of disintegration activity with the

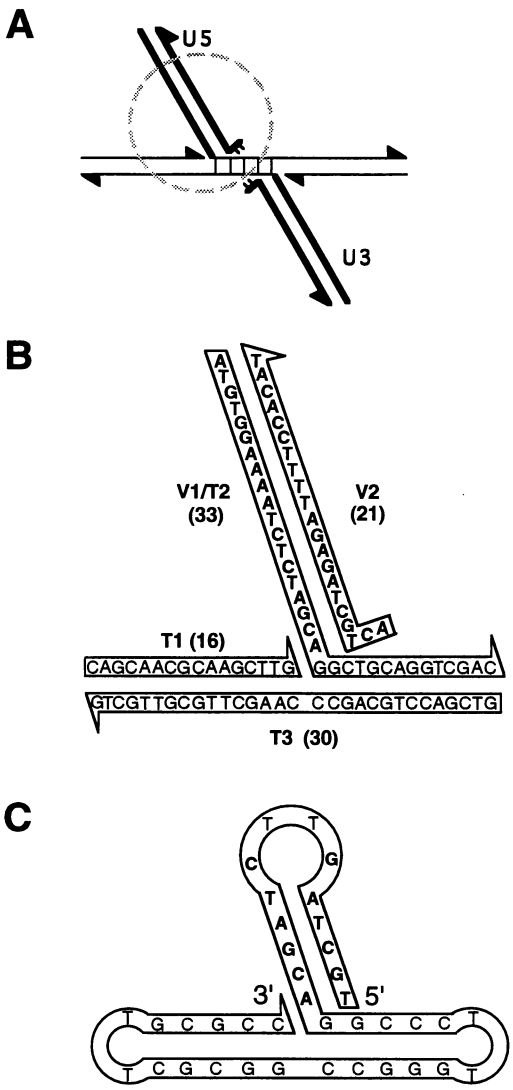

FIG. 1. Novel synthetic DNA substrates that mimic the recombination intermediate of retrovirus integration. (A) Diagram of the integration intermediate. Thick lines represent the U5 and U3 ends of the long terminal repeats of provirus DNA. Thin horizontal lines represent target DNA. Vertical lines represent the 5 bp of target DNA between the staggered breaks. Half-arrows denote the locations of the 3′-OH ends of DNA strands. The region of the intermediate mimicked by the synthetic DNA substrates (see panels B and C) is highlighted by the dashed circle. (B) Y oligomer. The substrate is formed by annealing four separate oligonucleotides as described previously (8). The DNA sequences of the U5 end of the HIV-1 (HXB2) long terminal repeat (in boldface type) and the target DNA sequences derived from plasmid πAN13 (not in boldface type) are shown. The names and lengths (in nucleotides) of the oligonucleotides are indicated. (C) Dumbbell substrate. The substrate consists of a single oligonucleotide 38 nucleotides long. The two unpaired 5′ nucleotides of the viral DNA, which are not essential for disintegration (8), are not included in the dumbbell substrate. The predicted secondary structure shown and the numbers of nucleotides in the hairpin loops are based on the results of Blommers et al. (2) and Antao et al. (1). The DNA sequences of the HIV-1 U5 end (in boldface type) in the hairpin stem and the arbitrary DNA sequences representing target DNA (not in boldface type) are shown.

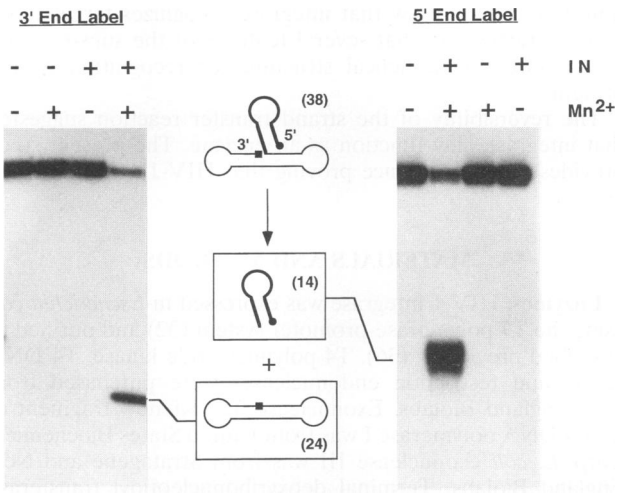

FIG. 2. Disintegration of the dumbbell substrate by HIV-1 integrase. Correct strand cleavage and ligation of the dumbbell substrate (38-mer) is expected to form two products: a 14-nucleotide hairpin and a 24-nucleotide closed circular DNA. The dumbbell substrate was labeled at the 5′ end (closed circle) with T4 polynucleotide kinase and [γ-^{32}P]ATP. The DNA sequence of the starting oligonucleotide (37-mer) for 3′-end labeling is identical to the first 37 nucleotides of the 38-mer dumbbell substrate (Fig. 1C), and the terminal nucleotide was filled in with exonuclease-free Klenow fragment of *E. coli* DNA polymerase I and [α-^{32}P]dCTP (closed square). The labeled oligonucleotide 38 nucleotides long was purified by electrophoresis through a 15% denaturing polyacrylamide gel. The oligonucleotide, labeled either at the 5′ or 3′ end, was incubated with (+) or without (−) HIV-1 integrase (IN) and Mn^{2+} as indicated above the gels. The reaction was carried out as described in Materials and Methods. The reaction products were analyzed on a 15% denaturing polyacrylamide gel. As size markers, oligonucleotides with sequences identical to those of the two expected disintegration products were synthesized. The ends of the synthetic 24-nucleotide molecule were then ligated with T4 DNA ligase and used as a marker for the circular product (markers are not shown in the figure).

dumbbell substrate indicated that small fragments of viral DNA and target DNA were sufficient for disintegration.

As previously observed with the Y oligomer (8), the viral DNA sequence influenced the disintegration reaction catalyzed by integrase. Substitution of the conserved CA dinucleotide at the viral 3′ end with TC reduced disintegration fivefold (data not shown).

HIV-1 integrase is capable of turning over. We used disintegration of the dumbbell substrate as an assay for kinetic analysis of HIV-1 integrase. By determining the initial velocity of the reaction at different substrate concentrations, we calculated the K_m for this substrate to be 140 nM (Fig. 3A). To determine whether integrase is capable of turning over, we incubated the protein with a 30-fold molar excess of the dumbbell substrate and monitored the formation of product over time (Fig. 3B). The concentration of the dumbbell substrate used in the turnover experiment was sevenfold higher than the K_m to ensure that the substrate was saturating the integrase. After 12 h of incubation, a reaction mixture containing 33 nM integrase generated more than 600 nM product. This result indicates that integrase was capable of turning over, with an apparent k_{cat} (catalytic constant) of 1.8 h^{-1} (Fig. 3B). The demonstration that integrase could turn over catalytically proves that HIV-1 integrase is an enzyme. Recently, Jones and coworkers (20) showed that integrase of Rous sarcoma virus is

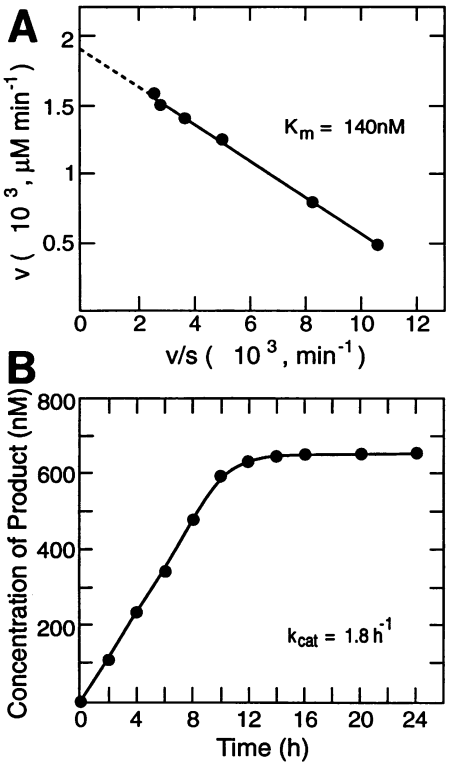

FIG. 3. Kinetic experiments using the dumbbell substrate. (A) Determination of K_m. Various concentrations (50 to 1,500 nM) of the dumbbell substrate labeled at the 5′ end were incubated with 50 nM HIV-1 integrase under standard reaction conditions. Aliquots of the reaction mixture were removed at 2, 5, and 10 min after the start of the reaction, and the formation of the labeled 14-mer hairpin was determined. The rate of the reaction at each substrate concentration was calculated, and the K_m value of the dumbbell substrate was determined with an Eadie-Hofstee plot. v, reaction rate; s, substrate concentration. (B) Turnover of HIV-1 integrase. The dumbbell substrate (1 mM) labeled at the 5′ end was incubated with 33 nM HIV-1 integrase. The formation of the labeled 14-mer hairpin product was determined at 0, 2, 4, 8, 10, 12, 14, 16, 20, and 24 h after the start of the reaction. The apparent k_{cat} was determined by calculating the slope of the curve during the first 10 h of the reaction. The apparent k_{cat} obtained is likely an underestimation, since the calculation method assumes that all integrase in the reaction is enzymatically active and that the active unit of integrase is a monomer.

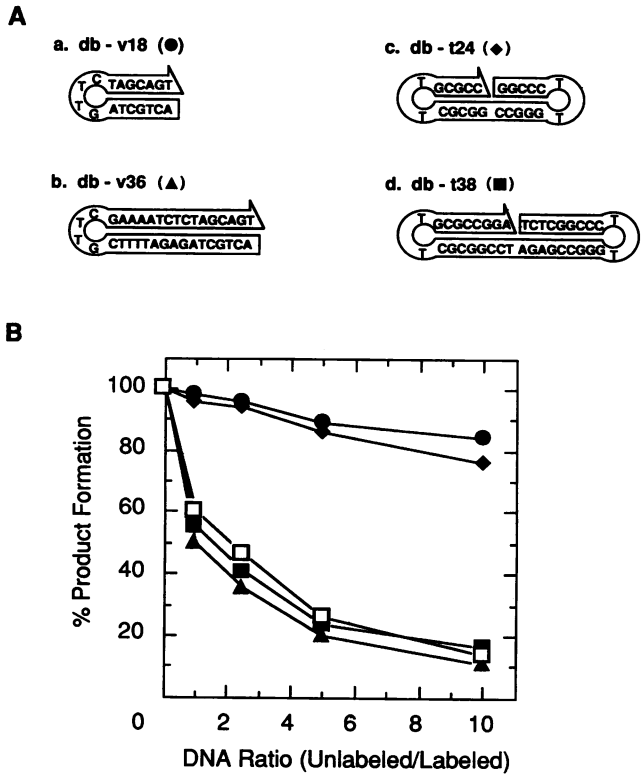

FIG. 4. Neither the viral nor target DNA part of the dumbbell substrate competes with the dumbbell substrate for HIV-1 integrase. (A) Sequences and predicted secondary structures of DNA substrates used in the competition experiment. The predicted secondary structures and the numbers of nucleotides in the hairpin loops are based on the results of Blommers et al. (2) and Antao et al. (1). Analysis of these DNA substrates on a 15% native polyacrylamide gel showed that they did not form bimolecular or higher-molecular-weight species under the described reaction conditions (data not shown). In substrates a and b, the DNA sequences correspond to the U5 end of HIV-1. In substrates a and c, the DNA sequences correspond to the viral and target parts of the dumbbell substrate, respectively. The DNA sequence in substrate d is arbitrary and represents target DNA. The 3′-OH ends of DNA strands are denoted by half-arrows. The lengths of substrates a, b, c, and d are 18, 36, 24, and 38 nucleotides, respectively. (B) Competition experiment. The dumbbell substrate (50 nM) labeled at the 5′ end was mixed with 0, 50, 125, 250, or 500 nM of unlabeled dumbbell substrate (□) or one of the four substrates shown in panel A. The mixture was incubated with 50 nM HIV-1 integrase under the standard reaction conditions for 25 min at 37°C. The reaction products were separated by electrophoresis on a 15% denaturing polyacrylamide gel. The formation of the labeled 14-mer hairpin product was quantitated with a Molecular Dynamics PhosphorImager.

capable of turnover in the 3′ processing and strand transfer reactions.

Both the viral and target DNA parts of the dumbbell substrate are required for integrase recognition. Since the dumbbell substrate contains both viral and target DNA sequences, it was not known whether its recognition by integrase was mediated mainly via the viral or target DNA part or both. This question was addressed by carrying out a competition experiment in which integrase was incubated with an equimolar concentration of labeled dumbbell substrate (Fig. 4). In the positive control, a 1- to 10-fold molar excess of unlabeled dumbbell substrate was added simultaneously with the labeled molecule and the formation of product was measured 25 min after the start of the reaction. As shown in Fig. 4B, when unlabeled dumbbell substrate was added so that the overall substrate concentration ranged from 0.7 to 4 times the K_m, the unlabeled substrate competed with the labeled molecule for the fixed concentration of integrase and resulted in a proportional decrease in product formation. When we divided the dumbbell into its viral (db-v18 [Fig. 4A]) and target (db-t24 [Fig. 4A]) DNA parts, neither fragment was efficient in competing with the labeled dumbbell substrate (Fig. 4B). We attributed this to the small size of the dumbbell oligonucleotide; loss of stabilizing interactions provided by larger viral and target DNA substrates leads to an absolute requirement for interactions with both viral and target DNA sequences of the dumbbell substrate. When the dumbbell was separated into individual parts, each part was too small to provide stable interaction. Indeed, when we increased the length of the target part of the dumbbell from 24 to 38 nucleotides (db-t38 [Fig. 4A]) or the viral part from 18 to 36 nucleotides (db-v36 [Fig.

4A]), we observed effective competition against the labeled dumbbell molecule (Fig. 4B). The observed competition was not an artifact caused by the hairpin structure of the competitors: similar results were obtained with double-stranded, blunt-ended oligonucleotides containing viral or target DNA sequences (data not shown). The inefficient binding of db-v18 to integrase was consistent with its inability to undergo 3′ processing and strand transfer reactions, whereas db-v36, which was an efficient competitor, showed normal levels of 3′ processing and strand transfer under identical reaction conditions (data not shown). The effectiveness of nonspecific target DNA as a competitor is consistent with previous reports showing that oligonucleotides containing nonspecific sequences can compete effectively against oligonucleotides containing viral end sequences for binding to HIV-1 integrase (29, 33, 39).

Asymmetric recognition of target DNA by HIV-1 integrase. Previous results from our laboratory suggested that integrase may recognize the target DNA asymmetrically, i.e., the 5′ and 3′ sides of the target DNA, in reference to the site of integration of viral DNA, are recognized unequally by integrase. When the lengths of the products formed from the strand transfer reaction were analyzed, products with sizes corresponding to integration near the 3′ end were abundant, but products of integration events near the 5′ end were not detected (Fig. 5B). To test the possibility that the asymmetric pattern of integration was due to a sequence effect imposed by the HIV-1 U5 DNA (C220/120 [Fig. 5A]), we examined the sites of integration by using a modified strand transfer assay similar to that described by Leavitt et al. (25). The unlabeled U5 DNA was used as a donor molecule. Oligonucleotides, which were made up of arbitrary sequences and have no overt homology among themselves or to the U5 or U3 region of HIV long terminal repeat, were singly labeled at the 3′ end and were used as target DNA (Fig. 5A). Integration of the processed U5 DNA (19 nucleotides) at the furthest 3′ position of any 3′-labeled target DNA generates a 20-nucleotide product, whereas integration at the furthest 5′ position generates a 40-nucleotide product for S21/22, T11/12, and T227/228, and a 39-nucleotide product for T45C/45. Similar to the results obtained with U5 DNA, using the described random oligonucleotides as target DNA, we observed integration near the 3′ end and the internal region of the DNA molecules, but no integration occurred near the 5′ end (Fig. 5C). The absence of an integration product longer than 36 nucleotides indicates that the terminal 4 nucleotides at the 5′ end were not used as target sites. This led us to hypothesize that recognition of target DNA by integrase is asymmetric.

We studied further the asymmetric requirement of target DNA with three modified dumbbell substrates (Fig. 6). In the standard dumbbell substrate (Fig. 1C), there are 5 bp of target sequences on each side of the viral DNA-target DNA junction. In the modified substrates, the 5′ deleted (Fig. 6, substrate A) and 3′-deleted (Fig. 6, substrates B and C) substrates, the right (corresponding to the 5′ end) and left (corresponding to the 3′ end) sides, respectively, of the disrupted target DNA strand were unchanged, while the remaining sides were reduced to 2 bp. The 5′-deleted substrate showed an activity that was almost indistinguishable from the standard dumbbell, whereas both of the 3′-deleted molecules had only 8 to 18% of the standard activity. It is possible that the poor activities of the 3′-deleted substrates were not due to the deletion of the target DNA but to instability of the short hairpin stem. We did not favor this possibility because substrate C (Fig. 6) contains a sequence (5′-GCGAAAGC) that forms an unusually stable minihairpin with a melting temperature of 76.5°C (19). The result obtained

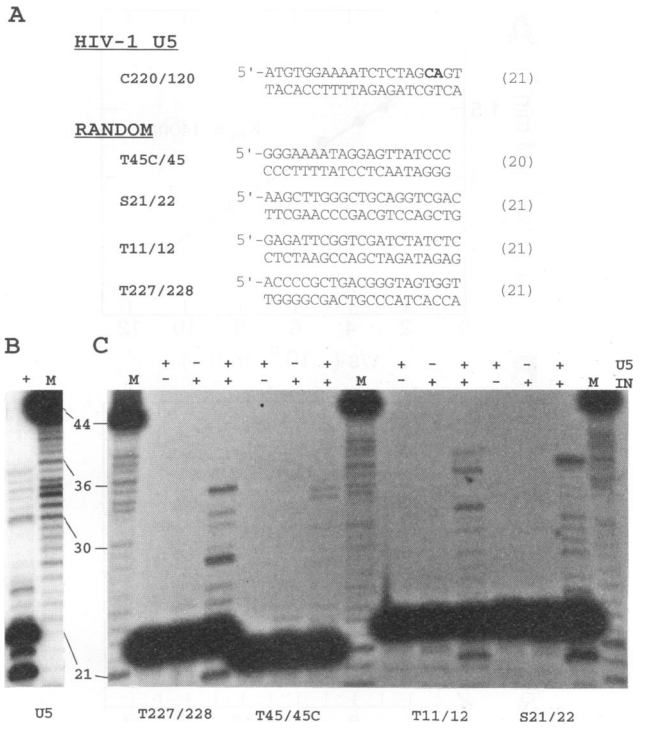

FIG. 5. Absence of integration near the 5′ end of target DNA. (A) Oligonucleotides used in determining sites of integration. C220/120 DNA contains DNA sequences of the U5 end of the HIV-1 (HXB2) long terminal repeat, and the conserved CA dinucleotide is indicated in boldface type. The DNA sequences of random oligonucleotides are arbitrary and have no overt homology among themselves or to the U5 or U3 region of HIV-1 long terminal repeat. Synthetic oligonucleotides were purified as described in Materials and Methods. The purified oligonucleotides were labeled at the 5′ or 3′ end, and unincorporated radioactive nucleotides were removed by centrifugation through spin columns or electrophoresis in 15% denaturing polyacrylamide gels. The labeled DNAs were annealed with their complementary strands by heating to 80°C, followed by slow cooling to room temperature. The numbers in parentheses indicate the lengths (in nucleotides) of the oligonucleotides. (B) Strand transfer assay using HIV-1 U5 DNA. C220/120 DNA was labeled at the 5′ end of the C220 strand using T4 polynucleotide kinase and [γ-^{32}P]ATP. The labeled substrate (0.1 pmol) was incubated with 1.5 pmol of HIV-1 integrase, and the reaction conditions were as described in Materials and Methods. (C) Strand transfer assay using random oligonucleotides as target DNA. Oligonucleotides S22, T12, T228, and T45 were labeled at the 3′ end using terminal deoxyribonucleotidyl transferase and [α-^{32}P]ddATP before annealing with their corresponding complementary strands (panel 5A). Unless otherwise indicated, unlabeled U5 DNA (0.1 pmol) was incubated with 1.5 pmol of HIV-1 integrase in a total volume of 20 μl for 1 min at 37°C before the addition of 0.1 pmol of labeled target DNA. The reaction mixture was then further incubated for 60 min at 37°C. In panels B and C, the size markers were prepared by digesting a 5′ ^{32}P-labeled 44-nucleotide DNA (38 μM) with 2 μg of P1 nuclease per ml for 5 min at 37°C. The absence (−) or presence (+) of C220/120 DNA (U5) and HIV-1 integrase (IN) is indicated above the gels. The numbers between the two panels denote the lengths (in nucleotides) of the size markers (M).

with the 5′-deleted substrate indicated that for full activity the 5′ side of the target DNA was dispensable. The large difference in activity between the 5′- and 3′-deleted substrates is consistent with the hypothesis that the recognition of target DNA is asymmetric.

Mismatches in the target DNA part of a disintegration

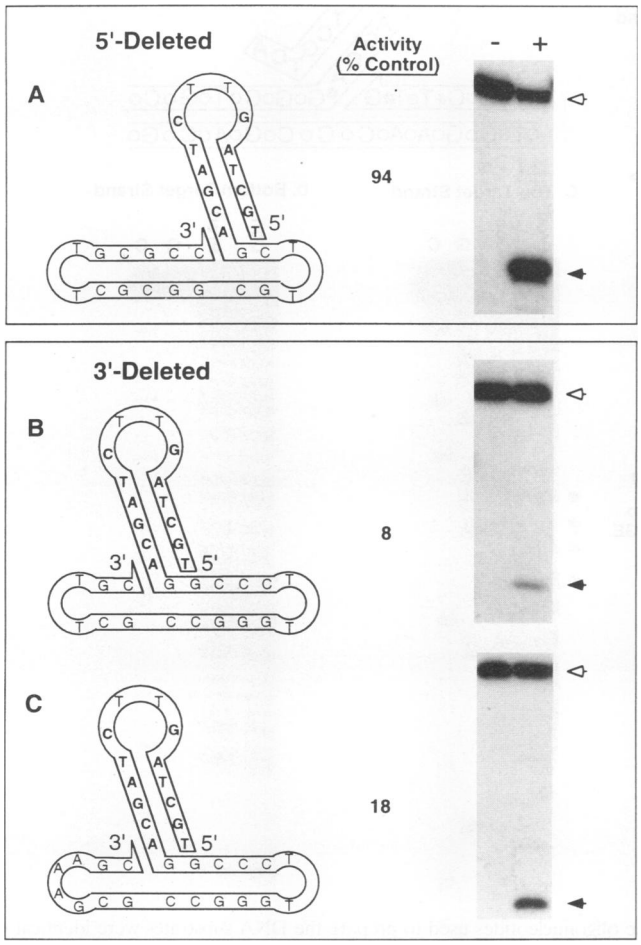

FIG. 6. Differential effect on disintegration of deleting the 5′ or 3′ side of the target DNA of the dumbbell substrate. Substrates A, B, and C were modified from the standard dumbbell molecule (Fig. 1C). The viral DNA sequences (in boldface type) and target DNA sequences (not in boldface type) are shown. The 5′- or 3′-deleted molecule (5 nM) was labeled at the 5′ end and incubated with (+) or without (−) 75 nM HIV-1 integrase under the standard reaction conditions. The formation of the labeled 14-mer hairpin was determined, and the activities of the modified substrates were expressed as percentages of that obtained using standard dumbbell as a control. Under identical conditions, 85% of the labeled standard dumbbell was converted to the labeled 14-mer hairpin molecule. For each set of reactions, the positions of the substrate (open arrowhead) and disintegration product (solid arrowhead) are shown.

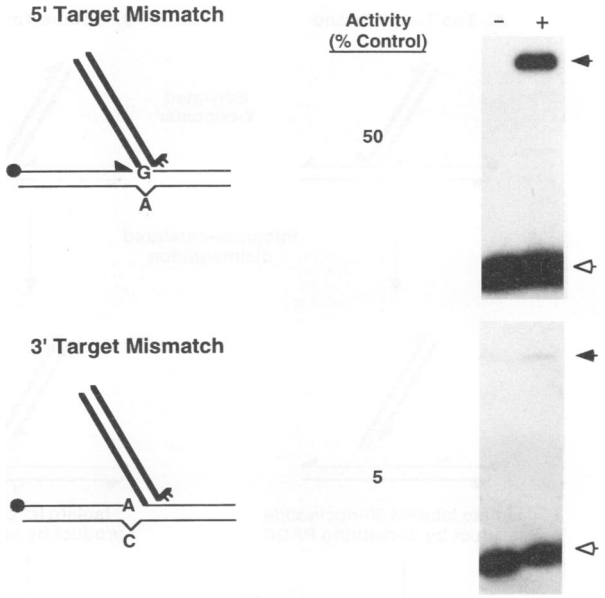

FIG. 7. Differential effect on disintegration of a mismatch located upstream or downstream of the viral DNA-target DNA junction site. The substrates were modified from the Y oligomer (Fig. 1B) and contained a mismatch in the target DNA downstream (5′ target mismatch) or upstream (3′ target mismatch) of the viral DNA joining site. Except for the mismatched nucleotides shown, the remaining DNA sequences and the lengths of the strands are identical to those of the Y oligomer. The labeled, modified substrates were prepared in the same manner as that of the Y oligomer (8). Thick lines represent viral DNA, and thin lines represent target DNA. A solid circle denotes the location of the 5′ ^{32}P label, and a half-arrow denotes the 3′-OH end of the labeled strand. The formation of the disintegration products was determined, and the activities of the modified substrates were expressed as percentages of that obtained using the standard Y oligomer (Fig. 1B) as a control. Under identical conditions, 40% of the labeled standard Y oligomer was converted to the labeled 30-mer product. The absence (−) and presence (+) of integrase are indicated. For each set of reactions, the positions of the substrate (open arrowhead) and disintegration product (solid arrowhead) are shown.

substrate showed analogous asymmetric effects on disintegration. The mismatched substrates were modified from the Y oligomer, rather than the dumbbell substrate, to minimize the effect of a mismatch on the overall helical structure of the DNA molecule. We found drastic differences in disintegration activities, depending on whether the mismatch was located upstream or downstream of the viral DNA joining site (Fig. 7). The activity obtained with the substrate containing a mismatch downstream of the joining site was 10-fold higher than that obtained with the substrate containing a mismatch upstream of the joining site.

Asymmetric recognition of target DNA at the integration site was further corroborated by an ethylation interference experiment (Fig. 8). Ethylation interference has been used recently to probe the positions on viral DNA that are required for integrase-mediated strand transfer in vitro (6). Treatment of DNA with *N*-nitroso-*N*-ethylurea attaches ethyl groups to the sugar-phosphate backbone (31a). In this study, we ethylated the target part of the Y oligomer and mapped the sites at which ethylation affected the ability of integrase to recognize the modified DNA as a substrate (Fig. 8A and B). DNA ethylation at two positions inhibited disintegration, while ethylation at two other positions enhanced the reaction (Fig. 8C). These positions were clustered at and immediately upstream of the junction between viral and target DNA. We did not determine, however, whether DNA ethylation affected disintegration directly by interfering with integrase binding or indirectly by altering the structure of the substrate. The earlier experiment carried out by Bushman and Craigie (6) was designed to determine functional contacts on viral DNA. The substrate they used did not permit detection of possible contacts between integrase and the upstream target DNA. However, they found that ethylation of the target DNA at the position immediately downstream of the viral joining site inhibited strand transfer. We do not believe that the downstream site represents an essential contact with integrase because we did not detect such an interference for disintegration and because strand transfer events can occur one base

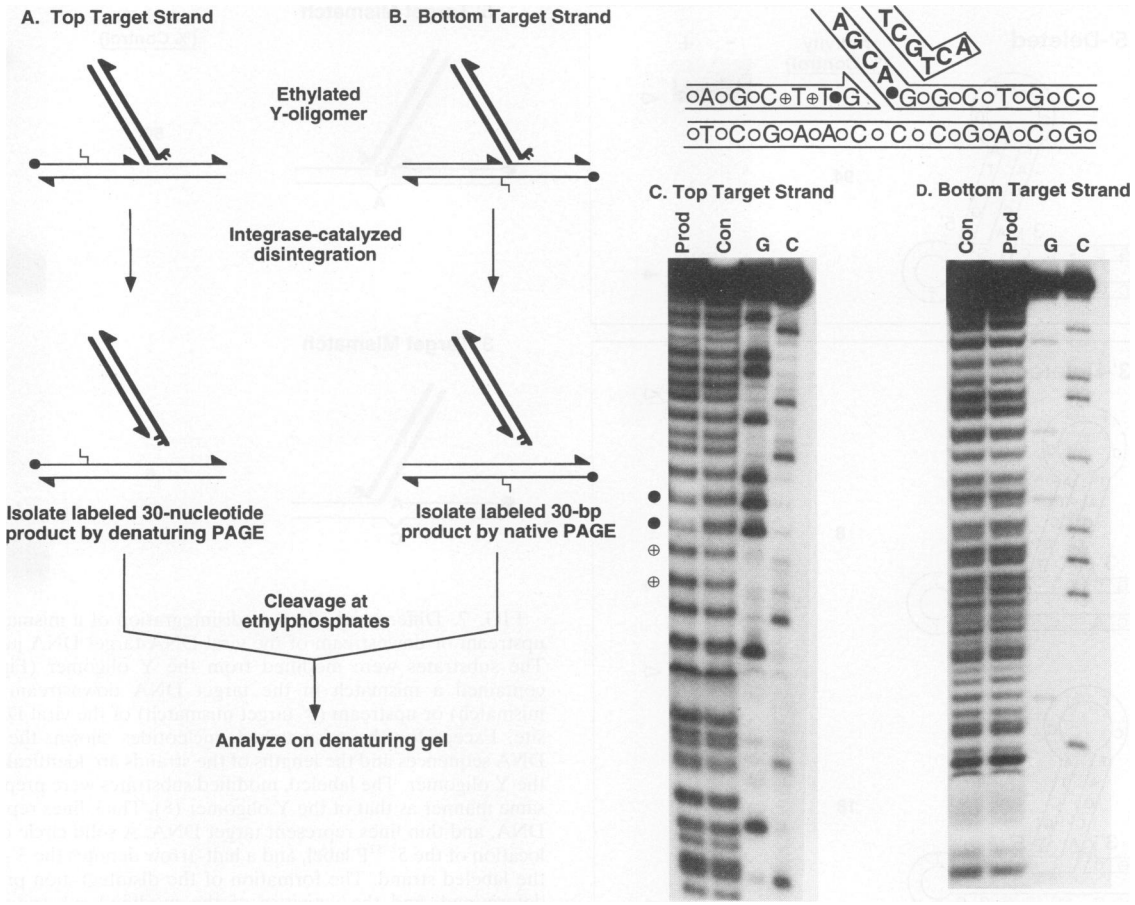

FIG. 8. Probing integrase-DNA interactions by ethylation interference. The oligonucleotides used to prepare the DNA substrates were identical to those of the Y oligomer (Fig. 1B). The zigzag lines represent ethyl groups. For the other symbols, see the legend to Fig. 6. The top target strand refers to the target strand that is joined to the viral DNA; the bottom target strand refers to the complement of the top strand. (A) Ethylated substrate for detection of interference on the top target strand. The T1 oligonucleotide was labeled at the 5′ end. The labeled T1 and unlabeled V1/T2 DNA were ethylated by using *N*-nitroso-*N*-ethylurea as described in Materials and Methods. The ethylated strands were annealed with untreated V2 and T3 DNA to form the Y oligomer as previously described (8). Each Y oligomer had, on average, less than one ethyl group randomly placed on the sugar-phosphate backbone of the T1 or V1/T2 strand. Disintegration was then carried out under the standard reaction conditions. A labeled 30-nucleotide DNA, formed after integrase-catalyzed disintegration, was isolated by electrophoresis on a 15% denaturing polyacrylamide gel. (B) Ethylated substrate for detection of interference on the bottom target strand. The T3 oligonucleotide was labeled at the 5′ end and ethylated as described in Materials and Methods. The labeled, ethylated strand was then annealed with oligonucleotides T1, V2, and V1/T2 to form the Y oligomer. The disintegration product, a 30-bp DNA fragment, was isolated by electrophoresis on a 15% native polyacrylamide gel. In both panels A and B, the isolated disintegration products were cleaved at the positions of ethyl phosphates by heating in 0.15 M NaOH. The NaOH cleavage products were analyzed on a 15% denaturing polyacrylamide gel. PAGE, polyacrylamide gel electrophoresis. (C) Integrase contacts on the top target strand. Lane Prod contains products of NaOH cleavage of the disintegration product made with the ethylated Y oligomer shown in Fig. 8A. Lane Con contains products of NaOH cleavage of an unreacted, ethylated 30-nucleotide DNA (T1-2) that is identical in DNA sequence to that of the top target strand formed after disintegration. Therefore, the DNA sequence of T1-2 strand is complementary to that of T3. Lanes G and C contain products of the Maxam-Gilbert sequencing reactions performed on the unethylated T1-2 DNA (26a). (D) Integrase contacts on the bottom target strand. Lane Prod contains products of the NaOH cleavage of the disintegration product made with the ethylated Y oligomer shown in Fig. 8B. Lane Con contains products of NaOH cleavage of an unreacted, ethylated T3 DNA. Lanes G and C contain products of the Maxam-Gilbert sequencing reactions performed on the unethylated T3 DNA. NaOH treatment of a population of the unreacted, randomly ethylated, and end-labeled DNA yields a ladder of bands after separation on a denaturing gel. A position at which ethylation interferes with substrate recognition by integrase is recognized as a change in band intensity by comparing the NaOH cleavage patterns obtained with the disintegration product and the unreacted substrate DNA. Positions where ethylation inhibited product formation (●), where ethylation enhanced product formation (⊕), and where ethylation did not appreciably alter product formation (○) are shown. Only the relevant part of the Y oligomer is shown.

from the 3′ end of the joining strand of target DNA (Fig. 5), indicating that integrase does not absolutely require contacts with the phosphodiester bond immediately downstream from the site of integration.

Relaxed structural specificity for the nucleophilic group during disintegration. The mechanism of 3′ processing and strand transfer mediated by integrase is a one-step in-line transesterification (12). The attacking nucleophile in strand transfer is the 3′-OH of DNA; in 3′ processing, a number of nucleophiles, such as water, glycerol, and certain aliphatic alcohols, can be used (12, 38). Disintegration normally involves a nucleophilic attack by the 3′-OH of target DNA on the phosphodiester bond between the viral and target DNA sequences (8). A 2′-OH group, however, could replace the

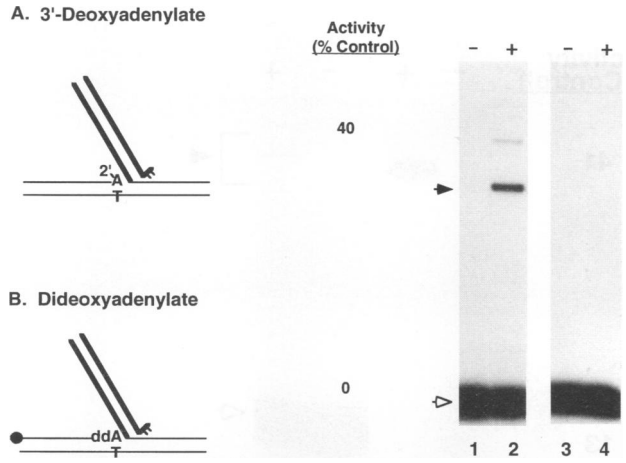

FIG. 9. Nucleophile usage during disintegration. The DNA substrates were modified from the Y oligomer. For explanation of the symbols, see the legend to Fig. 7. In panel A and lanes 1 and 2, the 3′ end of the T1 oligonucleotide was labeled with terminal deoxyribonucleotidyl transferase and [α-^{32}P]3′-dATP (cordycepin 5′-triphosphate [A-2′]). The labeled oligonucleotide therefore had a 2′-OH end instead of a 3′-OH end of normal DNA. The bands appearing above the product and those between the product and substrate represent products from reintegration of a viral DNA end liberated by disintegration. In panel B and lanes 3 and 4, the terminal nucleotide of the 5′ ^{32}P-labeled target strand was 2′,3′-dideoxyadenylate (ddA).

3′-OH group, but with a 2.5-fold decrease in activity (Fig. 9A). The target DNA product after disintegration of such a substrate presumably contained an unusual 2′-5′ phosphodiester linkage. The essential role of the -OH group of DNA was confirmed in Fig. 9B in which substitution of the terminal nucleotide with a 2′,3′-dideoxy analog completely abolished the disintegration activity that rejoined the interrupted target strand. Integrase, however, could mediate an inefficient cleavage reaction at the viral DNA joining site of the dideoxy-containing substrate using water or various alcohols as nucleophiles (data not shown).

Reaction site determinants and effects of distortion of the double-helical structure on disintegration. We showed previously that integrase can mediate disintegration using substrates in which the viral sequences are single stranded (8). Correct disintegration also occurs, albeit with a sixfold reduction in activity, using a substrate in which the highly conserved CA dinucleotide at the viral 3′ end is replaced with TC (8). The results indicate that integrase can recognize DNA that does not have a B-form structure and that substrate features other than the CA dinucleotide play a role in determining the site of transesterification. We introduced base deletions or insertions near the viral DNA-target DNA junction site in a Y-oligomer DNA substrate and examined their effects on disintegration. Substrates containing a 1- or 2-nucleotide deletion at the 3′ end of the interrupted target strand could still support disintegration (Fig. 10A). The disintegration product of the 1-base deletion substrate was isolated from the gel, and its identity was confirmed by DNA sequencing. Sequence analysis showed that the cleavage still occurred precisely 3′ to the CA dinucleotide and that cleavage was coupled to the joining of the interrupted target strand (data not shown). Since the cleavage-ligation reaction is mediated by a nucleophilic attack of the 3′-OH on the phosphodiester bond 3′ to the CA dinucleotide (12), the activity observed with these deletion substrates implies that there must be a considerable degree of flexibility at the catalytic site or that the DNA in this region can be distorted to allow correct transesterification to occur.

In the other group of modifications, we found that insertion of a base pair beyond the CA dinucleotide at the viral end decreased the activity only twofold; insertion of 2 bp decreased the activity to around 10% of that seen with the standard Y oligomer (Fig. 10B). Analysis of the disintegration products from these insertion substrates provided some interesting observations (Fig. 10B). When the insertion was 1 bp, a majority of the cleavage-ligation events took place at the phosphodiester bond immediately 3′ to the CA dinucleotide, i.e., between the terminal A nucleotide of the viral sequence and the inserted nucleotide, while a small fraction occurred at the branch point, i.e., the phosphodiester bond between the inserted nucleotide and its 3′ neighbor in the target DNA (Fig. 10B). When the insertion was 2 bp, the amount of activity at the branch point remained the same as that of 1-bp insertion, but the frequency of cleavage-ligation events occurring at the bond immediately 3′ to the CA dinucleotide decreased, so that the activities at these two sites became comparable. However, no product was detected with a length that corresponded to disintegration at the phosphodiester bond between the 2 inserted bp. The ability of the modified substrates to undergo correct disintegration again demonstrated the tolerance of the catalytic site to structural distortion. This result shows that during disintegration the viral DNA sequence is the major factor in determining the site of transesterification. When the position of the viral sequence is shifted by base insertions, there is residual activity that is dependent on the structure, namely, the branch point.

DISCUSSION

During the retrovirus life cycle, integrase is required for integration of retrovirus DNA into the host cell genome. In vitro, integrase alone is sufficient to carry out the major steps of retrovirus integration, namely, 3′ processing and strand transfer. The strand transfer reaction that results in the formation of an integration intermediate is reversible. Using a model substrate termed a Y oligomer, prepared by annealing four separate oligonucleotides to form a structure that mimics the integration intermediate, we showed previously that HIV-1 integrase efficiently and precisely resolves the intermediate into its viral and target parts, a process we termed disintegration (8). In the present study, we devised a novel DNA substrate called a dumbbell. The substrate consists of a single oligonucleotide containing sequences that can form a secondary structure resembling the integration intermediate. Although the structure of the dumbbell substrate has not been determined experimentally, for reasons described earlier (see Results), we believe that the structure shown in Fig. 1C is likely correct. We showed that with such a substrate, integrase can mediate correct and efficient disintegration. Experiments using the dumbbell substrate, the Y oligomer, and their structural variants provided information on the functional and enzymatic properties of HIV-1 integrase.

Compared with the Y oligomer, the dumbbell substrate has the advantages of being small, simple, and easy to prepare. The dumbbell substrate also offers a means of studying aspects of integrase-catalyzed disintegration that cannot be easily addressed by using the Y oligomer. As shown by the competition experiment, because of its small size, recognition of the dumbbell substrate required interactions of integrase with both the viral and target DNA parts of the substrate. This is in contrast with the larger Y oligomer wherein either the viral or target DNA part of the substrate contributes substantially to

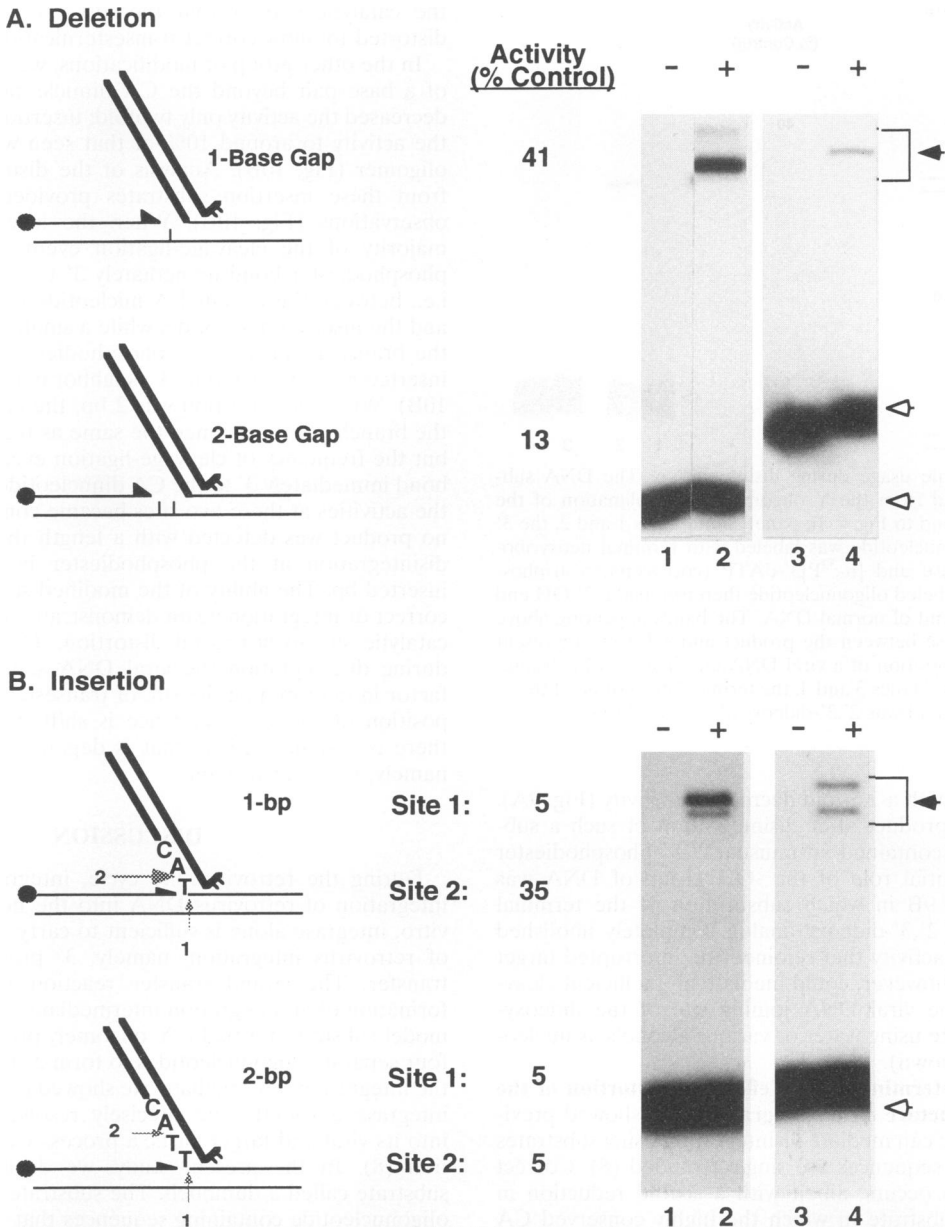

FIG. 10. Effects of nucleotide deletion or insertion on disintegration. All of the DNA substrates are modified from the Y oligomer. For explanation of the symbols, see the legend to Fig. 7. (A) Deletion. The modified substrates were prepared using a 5′-end-labeled target strand that is 1 (lanes 1 and 2) or 2 nucleotides (lanes 3 and 4) shorter than that of the standard Y oligomer (Fig. 1B). The resulting unpaired nucleotides are denoted by vertical lines. (B) Insertion. The location of the conserved CA dinucleotide in the viral DNA strand and the 1- (lanes 1 and 2) or 2-bp insertion (lanes 3 and 4) after the CA dinucleotide are indicated. The two sites of transesterification are denoted by shaded arrows. For each set of reactions, the top and bottom disintegration products correspond to transesterification occurring at sites 2 and 1, respectively.

integrase recognition (36). Several recent studies on mutant integrases indicate that certain integrase variants are unable to catalyze 3′ processing and strand transfer but are capable of carrying out disintegration (7, 11, 26, 34, 36, 37). We interpret this observation as an indication that interaction of integrase with either the viral or target DNA part of the Y oligomer is sufficient for substrate binding. Therefore, the differential abilities of mutant integrases to use Y oligomer and dumbbell as disintegration substrates can serve as an indicator for variant proteins that are defective in recognizing either the viral or target DNA (36).

Another distinguishing feature between the dumbbell and Y oligomer was that HIV-1 integrase showed turnover using the dumbbell substrate (see below) but not with the Y oligomer. To explain the difference, we hypothesize that in disintegration, release of reaction products is ordinarily the rate-limiting step. Stable association of disintegration products with integrase occludes a DNA binding site and prevents turnover. There is evidence that HIV-1 integrase forms stable protein-DNA complexes during 3′ processing, strand transfer, and disintegration reactions; the formation of stable complexes is dependent on the presence of the two unpaired 5′ nucleotides

of the viral DNA (10a). However, the disintegration products of the dumbbell substrate, perhaps because of their small sizes and the absence of the two terminal unpaired 5′ nucleotides, do not form stable complexes with the enzyme and thus allow the protein to turn over.

We tested the hypothesis that integrase recognizes target DNA asymmetrically. Depending on the position of target DNA relative to the viral sequences, dumbbell substrates of identical lengths had a 5- to 10-fold difference in activity. A similar difference in activity was observed when we compared the effect of locations of mismatches using Y oligomers. The cleavage pattern obtained from the ethylation interference experiment also pointed to essential interactions between integrase and target DNA upstream, but not downstream, of the viral DNA-target DNA junction site. The results are consistent with the hypothesis that the target DNA upstream of the viral joining site plays a more important role than the downstream target DNA in interacting with integrase. In corroboration with this observation is the result obtained from strand transfer reaction showing that the 5′-terminal region of target DNA was rarely used as an integration site. Experimental evidence is consistent with the notion that the nature of target recognition is similar in the forward and reverse directions of the integration reaction. Our result therefore suggests that in mediating integration, integrase interacts strongly with about 4 bp of target DNA upstream of the site of integration, but contacts with target DNA downstream are of less importance. In the in vivo reaction where two viral ends are joined at staggered sites on opposite strands, these key contacts with target DNA would flank the two joining sites, such that contacts with target DNA would span at least 13 bp (Fig. 11). Data obtained from sequencing integration targets used by murine leukemia virus (27a) and avian myeloblastosis virus (16) provide consistent evidence of sequence bias in target DNA immediately flanking the two viral DNA-target DNA junction sites, while no sequence bias is observed in the staggered region between the junction sites. Therefore, the positions of the nucleotides that show sequence bias are in good agreement with the protein-DNA contacts revealed by the ethylation interference experiment.

The findings that integrase could catalyze disintegration on substrates that contained base substitutions, deletions, insertions, or mismatches suggested that the catalytic site of integrase was promiscuous and could tolerate helical distortion of DNA substrates. Integrase normally uses the 3′-OH group of DNA as a nucleophile for disintegration, but a 2′-OH group could substitute for the 3′-OH group of DNA as the nucleophile. Moreover, the position of the attacking -OH group relative to the junctional phosphodiester bond was not highly restricted. We also found that, in addition to the DNA sequence at the viral end, the structure of the branched substrate played a role in determining the site of transesterification. Site-directed mutagenesis studies of HIV-1 integrase showed that substitutions of conserved amino acids in the central region, the D,D(35)E domain, render the protein inactive in carrying out 3′ processing, strand transfer, and disintegration (11, 23, 26, 37). This result suggests that the D,D(35)E motif defines a single catalytic domain that is responsible for all the catalytic activities of integrase. If this is indeed the case, the same catalytic domain that mediates a hydrolytic cleavage during 3′ processing is also involved in mediating transesterification during strand transfer and disintegration. Since the substrates in these different reactions are likely to pose different structural constraints on the nucleophile and the phosphodiester bond involved in breakage and joining, we propose that accommodation of these differences

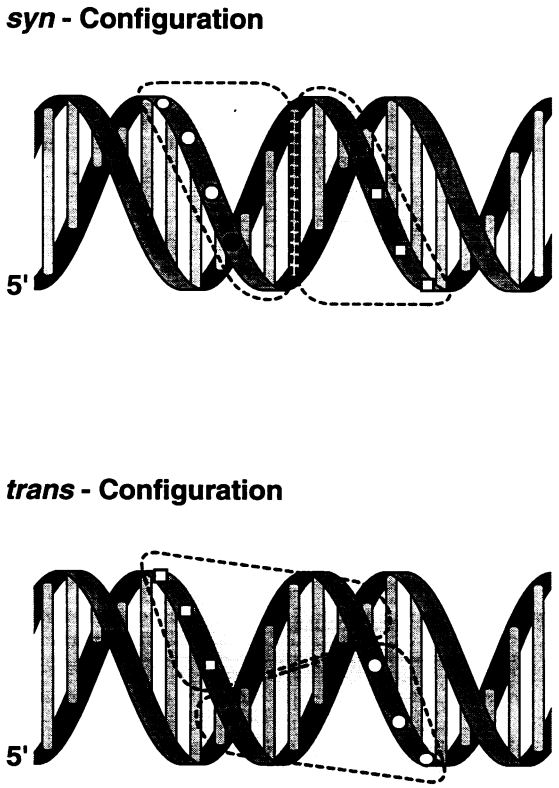

FIG. 11. A model showing key sites of HIV-1 integrase contact with target DNA. Dashed lines represent the positions of integrase protomers. Open and filled symbols denote the positions at which ethylation influences disintegration. Filled symbols denote the sites of viral DNA joining, which are staggered by 5 bp. If we assume that joint integration of two viral ends in vivo involves two symmetrically placed integrase protomers, the interaction between an integrase protomer and target DNA can have two possible configurations. In the *syn* configuration, the integrase protomer that catalyzes the cleavage-ligation reaction on a target strand makes contacts with target DNA immediately upstream on the same strand. In the *trans* configuration, the integrase protomer that catalyzes cleavage-ligation on a target strand makes contact with the complementary strand across the major groove. Filled and open circles represent the joining and contact sites, respectively, of one protomer; whereas filled and open rectangles represent the joining and contact sites, respectively, of the other protomer. Even though the contact sites are located on both target strands across the major groove, we believe that for each viral DNA end, in both *syn* and *trans* configurations, the contact sites on target DNA that are most critical for integrase interaction are those positioned immediately upstream of the viral DNA-target DNA junction. This is based on the observations that the strand transfer reaction in vitro can occur one base from the 3′ end of the target DNA and that disintegration of a Y oligomer was not affected by ethylation of target DNA at positions downstream of the viral DNA-target DNA junction site.

may be best accomplished by locally distorting the double-helical structure of both viral and target DNA sequences, so that the structure of the substrate in the active site need not be determined by the secondary structure of the free substrate. Thus, the ability of integrase to tolerate various modifications on the DNA substrates may be a reflection of a single catalytic site being required to catalyze two or more distinct reactions.

It has been estimated that there are about 100 copies of integrase per virion (35), and presumably in vivo integrase

needs to act only once at each viral DNA end for joining retrovirus DNA to cellular DNA. In vitro, the strand transfer reaction mediated by purified recombinant HIV-1 integrase is generally inefficient and requires molar excess of integrase. This led to the speculation that HIV-1 integrase may act stoichiometrically during integration (8a). Our earlier result showing that the cleavage-ligation reaction is reversible cast doubts on the hypothesis (8). In the present report, kinetic experiments using the dumbbell substrate showed that HIV-1 integrase could turn over catalytically. The turnover number was comparable to those obtained with the integrase of Rous sarcoma virus (20), and the FLP recombinase of the 2μm plasmid of yeasts (15). The observations that integrase turns over and that the integrase-mediated cleavage-ligation is reversible clearly demonstrate that integrase is not a reactant but indeed functions as an enzyme.

Although the results described in this report were obtained with HIV-1 integrase, similar properties were also observed with murine leukemia virus integrase. Like HIV-1 integrase, murine leukemia virus integrase could mediate disintegration using the dumbbell substrate, showed asymmetric recognition of target DNA, and tolerated distortions in the double-helical structure of the DNA substrate (data not shown). MLV integrase was also capable of turning over (9a).

Many small, simple DNA substrates are currently available for studying specific properties of integrase. Miniature substrate analogs, such as a dumbbell substrate containing a modified 3′ end, may provide a starting point for development of inhibitors of integrase. Finally, the simple, efficient assays described here can be easily adapted as screens for potential antiretroviral agents.

ACKNOWLEDGMENTS

We thank the other members of our laboratory and Harold Varmus and Andrew Leavitt for helpful discussions. We thank Janice Chow, TaiYun Roe, Brian Scottoline, George Yu, and Nancy Valente for comments on the manuscript.

This work was supported by the Howard Hughes Medical Institute and by NIH grant AI27205. P.O.B. is an assistant investigator of the Howard Hughes Medical Institute.

REFERENCES

1. **Antao, V. P., S. Y. Lai, and I. Tinoco, Jr.** 1991. A thermodynamic study of unusually stable RNA and DNA hairpins. Nucleic Acids Res. **19:**5901–5905.
2. **Blommers, M. J. J., J. A. L. I. Walters, C. A. G. Haasnoot, J. M. A. Aelen, G. A. van der Marel, J. H. van Boom, and C. W. Hilbers.** 1989. Effects of base sequence on the loop folding in DNA hairpins. Biochemistry **28:**7491–7498.
3. **Brown, P. O.** 1990. Integration of retroviral DNA. Curr. Top. Microbiol. Immunol. **157:**19–48.
4. **Brown, P. O., B. Bowerman, H. E. Varmus, and J. M. Bishop.** 1987. Correct integration of retroviral DNA in vitro. Cell **49:**347–356.
5. **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.
6. **Bushman, F. D., and R. Craigie.** 1992. Integration of human immunodeficiency virus DNA: adduct interference analysis of required DNA sites. Proc. Natl. Acad. Sci. USA **89:**3458–3462.
7. **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.
8. **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.

8a. **Craigie, R., K. Mizuuchi, F. Bushman, and A. Engelman.** 1991. A rapid *in vitro* assay for HIV DNA integration. Nucleic Acids Res. **19:**2729–2734.

9. **Donehower, L. A., and H. E. Varmus.** 1984. A mutant murine leukemia virus with a single missense codon in *pol* is defective in a function affecting integration. Proc. Natl. Acad. Sci. USA **81:**6461–6465.

9a. **Dotan, I., B. P. Scottoline, T. S. Heuer, and P. O. Brown.** Unpublished data.

10. **Ellison, V., H. Abrams, T. Roe, J. Lifson, and P. O. Brown.** 1990. Human immunodeficiency virus integration in a cell-free system. J. Virol. **64:**2711–2715.

10a. **Ellison, V., and P. O. Brown.** A stable complex between integrase and viral DNA ends mediates HIV integration in vitro. Proc. Natl. Acad. Sci. USA, in press.

11. **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:**6363–6369.
12. **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.
13. **Fujiwara, T., and R. Craigie.** 1989. Integration of mini-retroviral DNA: a cell-free reaction for biochemical analysis of retroviral integration. Proc. Natl. Acad. Sci. USA **86:**3065–3069.
14. **Fujiwara, T., and K. Mizuuchi.** 1988. Retroviral DNA integration: structure of an integration intermediate. Cell **54:**497–504.
15. **Gates, C. A., and M. M. Cox.** 1988. FLP recombinase is an enzyme. Proc. Natl. Acad. Sci. USA **85:**4628–4632.
16. **Grandgenett, D. P., R. B. Inman, A. C. Vora, and M. L. Fitzgerald.** 1993. Comparison of DNA binding and integration half-site selection by avian myeloblastosis virus integrase. J. Virol. **67:**2628–2636.
17. **Grandgenett, D. P., and S. R. Mumm.** 1990. Unraveling retrovirus integration. Cell **60:**3–4.
18. **Hirao, I., Y. Nishimura, T. Naraoka, K. Watanabe, Y. Arata, and K. Miura.** 1989. Extraordinary stable structure of short single-stranded DNA fragments containing a specific base sequence: d(GCGAAAGC). Nucleic Acids Res. **17:**2223–2231.
19. **Hirao, I., Y. Nishimura, Y. Tagawa, K. Watanabe, and K. Miura.** 1992. Extraordinarily stable mini-hairpins: electrophoretical and thermal properties of the various sequence variants of d(GCGAAAGC) and their effect on DNA sequencing. Nucleic Acids Res. **20:**3891–3896.
20. **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.
21. **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.
22. **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.
23. **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.
24. **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.
25. **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.
26. **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.

26a. **Maxam, A. M., and W. Gilbert.** 1980. Sequencing end-labeled DNA with base-specific chemical cleavages. Methods Enzymol. **65:**499–560.

27. **Panganiban, A. T., and H. M. Temin.** 1984. The retrovirus pol gene

encodes a product required for DNA integration: identification of a retrovirus int locus. Proc. Natl. Acad. Sci. USA **81**:7885–7889.

27a. **Pryciak, P. M., A. Sil, and H. E. Varmus.** 1992. Retroviral integration into minichromosomes *in vitro*. EMBO J. **11**:291–303.

28. **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.

28a. **Sambrook, J., E. F. Fritsch, and T. Maniatis.** 1989. Molecular cloning: a laboratory manual, 2nd ed. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.

29. **Schauer, M., and A. Billich.** 1992. The N-terminal region of HIV-1 integrase is required for integration activity, but not for DNA binding. Biochim. Biophys. Acta **186**:874–888.

30. **Schwartzberg, P., J. Colicelli, and S. P. Goff.** 1984. Construction and analysis of deletion mutations in the pol gene of Moloney murine leukemia virus: a new viral function required for productive infection. Cell **37**:1043–1052.

31. **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.

31a. **Siebenlist, U., and W. Gilbert.** 1980. Contacts between *Escherichia coli* RNA polymerase and an early promoter of phage T7. Proc. Natl. Acad. Sci. USA **77**:122–126.

32. **Tabor, S., and C. C. Richardson.** 1985. A bacteriophage T7 RNA polymerase/promoter system for controlled exclusive expression of specific genes. Proc. Natl. Acad. Sci. USA **82**:1074–1078.

33. **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 virus types 1 and 2. Nucleic Acids Res. **19**:3821–3827.

34. **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.

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

36. **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.

37. **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.

38. **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. J. Virol. **19**:6691–6698.

39. **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**:2433–2437.