

Human Immunodeficiency Virus Type 1 Integration Protein: DNA Sequence Requirements for Cleaving and Joining Reactions

PAULA A. SHERMAN,* MATTHEW L. DICKSON, AND JAMES A. FYFE

Experimental Therapy, Wellcome Research Laboratories, Research Triangle Park, North Carolina 27709

Received 18 November 1991/Accepted 11 March 1992

Using purified integration protein (IN) from human immunodeficiency virus (HIV) type 1 and oligonucleotide mimics of viral and target DNA, we have investigated the DNA sequence specificity of the cleaving and joining reactions that take place during retroviral integration. The first reaction in this process is selective endonucleolytic cleaving of the viral DNA terminus that generates a recessed 3' OH group. This 3' OH group is then joined to a 5' phosphoryl group located at a break in the target DNA. We found that the conserved CA located close to the 3' end of the plus strand of the U5 viral terminus (also present on the minus strand of the U3 terminus) was required for both cleaving and joining reactions. Six bases of HIV U5 or U3 DNA at the ends of model substrates were sufficient for nearly maximal levels of selective endonucleolytic cleaving and joining. However, viral sequence elements upstream of the terminal 6 bases could also affect the efficiencies of the cleaving and joining reactions. The penultimate base (C) on the minus strand of HIV U5 was required for optimal joining activity. A synthetic oligonucleotide mimic of the putative in vivo viral "DNA" substrate for HIV IN, a molecule that contained a terminal adenosine 5'-phosphate (rA) on the minus strand, was indistinguishable in the cleaving and joining reactions from the DNA substrate containing deoxyadenosine instead of adenosine 5'-phosphate at the terminal position. Single-stranded DNA served as an in vitro integration target for HIV IN. The DNA sequence specificity of the joining reaction catalyzed in the reverse direction was also investigated.

Integration, the insertion of a double-stranded DNA copy of the viral RNA genome into the host cellular DNA, is required for the establishment of a productive retroviral infection. This is true for human immunodeficiency virus (HIV) as well as for other retroviruses (2, 7, 28). Because there is no known equivalent to this specialized DNA recombination event in human cells, integration is an attractive target for selective antiretroviral therapy.

The retroviral integration protein (IN) is the only viral protein required to carry out a series of reactions that occur during the integration process (10, 18). These reactions include sequence-selective cleaving of nucleotides from the 3' termini of linear double-stranded viral DNA and joining of the resulting recessed 3' OH groups at the viral DNA termini to the 5' phosphoryl groups at a double-stranded break in the target chromosomal DNA. Repair of the single-stranded gaps in the integration intermediate, presumably by cellular enzymes, results in a mature provirus.

Synthetic double-stranded oligonucleotides that mimic the U3 and U5 termini of retroviral DNA can function as both viral and target DNA for purified retroviral integration proteins (4, 10, 18). We have used such oligonucleotide substrates and HIV type 1 (HIV-1) IN purified to near homogeneity from an *Escherichia coli* expression system (25) to investigate in detail the DNA sequence requirements for the selective cleaving reaction. Oligonucleotide mimics of the products of the selective cleaving reaction, which contained preformed recessed 3' termini, were used to examine the sequence specificity of the joining reaction independently of the cleaving reaction. The present study confirms and extends the results of two similar studies of the DNA substrate specificity of HIV IN that have recently been published (22, 29).

MATERIALS AND METHODS

Purification of HIV-1 IN protein from *E. coli*. HIV-1 IN was expressed in *E. coli* and purified by butyl- and heparin-Sepharose chromatography as described previously (25).

DNA cleaving and joining assays. DNA oligonucleotide mimics of viral termini were prepared with a Biosearch model 8600 DNA synthesizer (W. Dallas, Wellcome) and were purified by electrophoresis on 20% denaturing polyacrylamide gels. Molar absorption coefficients at 260 nm for the purified oligonucleotides were estimated by summing the contributions of the individual bases (T, 8,820; A, 15,400; G, 11,700; and C, 7,400), and concentrations were calculated by using these values. An oligonucleotide mimic of the minus strand of HIV U5 DNA with a 5'-terminal adenosine ribonucleotide and an oligonucleotide with a 3' biotin (attached to the 3' phosphate via an *N*-caproyl linkage to an 11-atom spacer arm) were synthesized by Midland Certified Reagent Company. The appropriate DNA strands were labeled at the 5' termini with T4 polynucleotide kinase (New England Biolabs) and [γ -³²P]ATP (3,000 Ci/mmol; New England Nuclear). The labeling reactions were terminated by boiling for 5 min, after which an equimolar amount of the complementary DNA strand was added and annealed to the labeled strand by slow cooling. The oligonucleotide models of the products of the joining reaction were prepared in a similar fashion, except that 2 or 3 oligonucleotides were added at the annealing step. Reaction mixtures (10 μ l) for cleaving and joining contained 1 pmol of DNA, 20 mM Tris-HCl (pH 8), 5 mM 2-mercaptoethanol, 1 mM MgCl₂ (contributed from the labeling reaction mixture), 1 mM MnCl₂, and 0.5 μ g of protein. A different buffer system that was found to enhance the joining activity (and also enhanced a nonselective cleaving activity with double-stranded substrates) was used for some of the reactions. In this case, reaction mixtures contained the same amounts of DNA and protein as were used for the Tris buffer system, 20 mM MOPS (morpholinepro-

* Corresponding author.

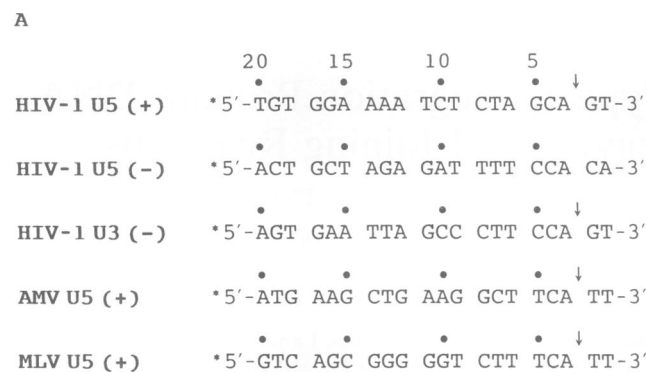
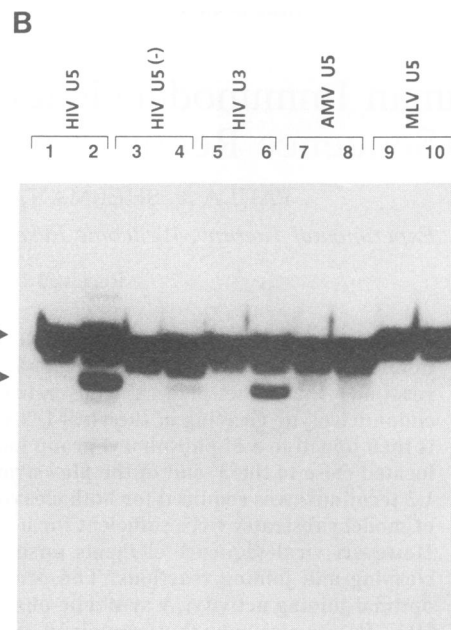


FIG. 1. HIV IN selectively cleaves oligonucleotide mimics of HIV DNA. (A) Sequences of synthetic oligonucleotide substrates. The plus strands for HIV, AMV, and MLV U5 substrates are shown, as are the minus strands for HIV U5 and HIV U3 substrates. Asterisks indicate the location of the radiolabel. Arrows indicate the expected cleavage sites (adjacent to the conserved CA). (B) Analysis of reaction products on a denaturing polyacrylamide gel. DNA substrates are indicated above the lanes of the autoradiogram. The first lane in each set of two was a no-protein control. Numbers to the left indicate molecular size (in bases).



panesulfonic acid) (pH 7.2), 10 mM 2-mercaptoethanol, 1 mM MgCl₂, 10 mM MnCl₂, and 10% glycerol. Unless otherwise specified (see the legends to the figures), the Tris buffer system was used. Incubations were for 1 h at 37°C, and reactions were stopped by adding 15 μl of 90% formamide–0.025% bromophenol blue–0.025% xylene cyanol–89 mM Tris–89 mM boric acid–2 mM EDTA and then heating for 3 min at 90°C. The samples were subjected to electrophoresis on 20% denaturing polyacrylamide gels. Reaction products were visualized by autoradiography of the wet gels.

RESULTS

DNA sequence requirements for selective cleaving of oligonucleotide mimics of viral termini. The integrated proviral form of a retrovirus is identical to the precursor viral DNA, except for the loss of, in general, 2 bp at each end, at the points of attachment to the cellular DNA. Furthermore, all proviruses are punctuated at both termini by a CA dinucleotide (on the minus strand of U3 and the plus strand of U5) so that the provirus plus strand takes the form TG-----CA. Therefore, one function for the retroviral IN is selective cleaving on the 3' side of the A residue in the conserved CA located at the ends of the viral DNA. This conclusion is substantiated by observations that the immediate precursor for integration in virus-infected cells is a linear DNA species that is full length at its 5' ends but recessed by 2 bases at its 3' ends (1, 16) and that a functional viral IN is required for the formation of this 3' recessed precursor (1, 24). More recently, we and others have demonstrated that the selective processing of viral DNA termini by retroviral INs can be simulated in vitro with synthetic double-stranded oligonucleotide mimics of the termini of viral DNA (4, 10, 18, 19, 25, 26).

We have extended our earlier examination (25, 26) of the ability of purified HIV IN to selectively cleave oligonucleotide mimics of HIV viral termini compared with its ability to cleave the termini from other retroviruses (sequences of oligonucleotide substrates are shown in Fig. 1A). As shown in Fig. 1B, only the HIV U3 (minus-strand) and U5 (plus-

strand) substrates were efficiently and selectively cleaved; in both cases, cleaving was adjacent to the conserved CA dinucleotide. The products of this cleaving reaction were a viral DNA terminus with a recessed 3' OH and a dinucleotide containing a 5' phosphate (26). The minus strand of the HIV U5 substrate was not cleaved efficiently, despite the fortuitous presence of a CA dinucleotide located 2 bases from the 3' end (Fig. 1B, lane 4). In agreement with results recently reported by two other groups (4, 22), the HIV U5 substrate was reproducibly cleaved more efficiently than the HIV U3 substrate. Very little, if any, selective cleaving was observed with mimics of the U5 termini of avian myeloblastosis virus (AMV) and Moloney murine leukemia virus (MLV). The AMV U5 substrate was efficiently cleaved, however, by purified AMV IN (data not shown) (protein provided by D. Grandgenett, St. Louis University Medical Center).

Next, we examined the effect of changing the conserved CA (with its paired GT) on the ability of HIV IN to cleave oligonucleotide mimics of the U5 terminus of HIV DNA. We also looked at the effect of changing the identity and the number of nucleotides following the conserved CA. As shown in Fig. 2, cleaving of the HIV U5 substrate was greatly reduced when the C and the A of the conserved CA were changed to T and C, respectively (Fig. 2, lanes 4 and 6); changing both the C and the A abolished selective cleaving (Fig. 2, lane 8). Changing the terminal GT of HIV U5 DNA to a TA caused a smaller reduction in cleaving (Fig. 2, lane 10), as did changing (decreasing or increasing by one) the number of nucleotides following the CA (Fig. 2, lanes 12 and 14).

We next investigated the effect of changing the DNA sequence upstream of the conserved CA in the HIV U5 substrate on the selective cleaving reaction. Effort was concentrated on 10 bases at the 3' terminus of HIV U5, since our earlier studies had shown that this length of wild-type sequence was sufficient for selective cleaving activity (26). We replaced increasing amounts of viral sequence (starting at the 5' end of the plus strand of U5) with an unrelated DNA

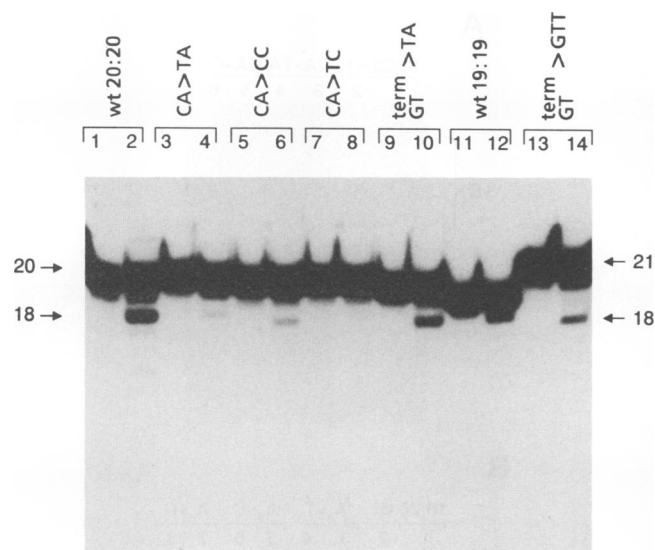


FIG. 2. Selective cleaving activity of HIV IN with HIV U5 substrates containing changes in the conserved CA and changes in the identity and the number of nucleotides following the conserved CA. Reaction products were analyzed on a denaturing polyacrylamide gel. Changes in the HIV U5 DNA sequence are shown above the lanes of the autoradiogram. For simplicity, changes in only the plus strand are indicated; the complementary strand was also altered. In all cases, the plus strand was labeled. The first lane in each pair was a no-protein control. Numbers to the left and right indicate molecular size (in bases).

sequence (modified substrates are shown in Fig. 3A). As shown in Fig. 3B, cleaving was greatly diminished when the length of HIV sequence was reduced to 5 bases (Fig. 3B, mut 15, lane 8) and was barely detectable with 4 bases of wild-type sequence (Fig. 3B, mut 16, lane 10). The 6 terminal bases of HIV U5 DNA appeared to be sufficient for a nearly maximal level of selective cleaving (Fig. 3B, mut 14, lane 6); the same result was observed with an oligonucleotide mimic of the HIV U3 terminus (data not shown). However, evi-

dence for a role of viral sequence elements upstream of the 6 terminal bases was obtained with the substrate designated rev. mut 15. Cleaving activity lost by decreasing the length of wild-type viral sequence from 6 to 5 bases was largely restored when selected residues upstream of position 6 were reverted to wild type (Fig. 3B, lane 12); these residues were chosen for reversion because they are identical in HIV U3 and U5 DNA.

We further examined the role of the A at position 6 and the G at position 5 in a series of single-base substitutions. As shown in Fig. 4A, changing the A at position 6 to T had little effect on the cleaving reaction (Fig. 4A, lane 4); changing this A to C (Fig. 4A, lane 6) or G (Fig. 4A, lane 8) caused a small (approximately 10%), reproducible decrease in cleaving. That the A-to-C change at position 6 was well tolerated as a single-base substitution further supports a role for viral DNA sequence elements upstream of the terminal six bases. As described above, the A-to-C change was also tolerated in the substrate rev. mut 15, which contained a wild-type viral DNA backbone but was not as well tolerated in mut 15, which contained a nonspecific DNA backbone (Fig. 3B). As shown in Fig. 4B, changing the G at position 5 to C, A, or T had little effect on the cleaving reaction.

DNA sequence requirements for the *in vitro* joining reaction. In addition to selectively cleaving an oligonucleotide mimic of a viral DNA terminus, purified retroviral IN can join the resulting recessed 3' end to the 5' end of a cut in a second identical oligonucleotide (10, 18). The second oligonucleotide thus serves as the integration target or mimic of host cellular DNA. This *in vitro* joining reaction is evidenced by products that are both longer (viral plus target DNA) and shorter (the 5' portion of the target DNA strand released by joining) than the unreacted substrate DNA. These products are easily observed upon extended exposure of autoradiography films to gels after electrophoresis of cleaving reaction mixtures.

Joining follows the selective 3' processing of 20:20 oligonucleotide mimics of HIV U3 and U5; it is also observed when the starting DNA substrate is preprocessed, e.g., an 18:20 mimic of HIV U5 (4, 10). Such a preprocessed substrate may be used to distinguish the effects of DNA se-

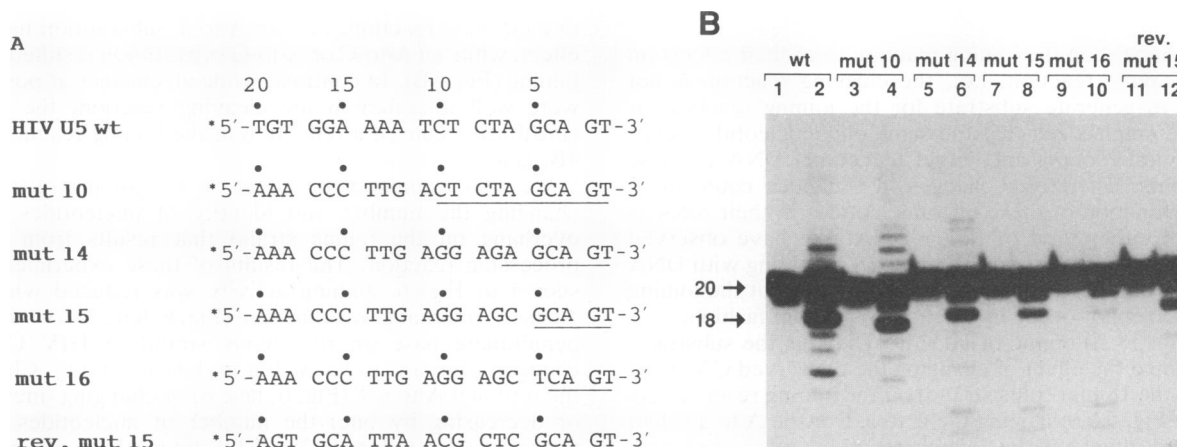


FIG. 3. Length of HIV U5 sequence required for selective cleaving by HIV IN. (A) Sequences of oligonucleotide substrates. All sequences are of the plus strands. Asterisks indicate the location of the radiolabel. Wild-type viral sequence is underlined. The upstream nucleotides returned to wild type in rev. mut 15 are those nucleotides that are identical in HIV U3 (minus strand) and HIV U5 (plus strand). (B) Analysis of reaction products on a denaturing polyacrylamide gel. DNA substrates are shown above the lanes of the autoradiogram. The first lane in each pair was a no-protein control. Numbers to the left indicate molecular size (in bases).

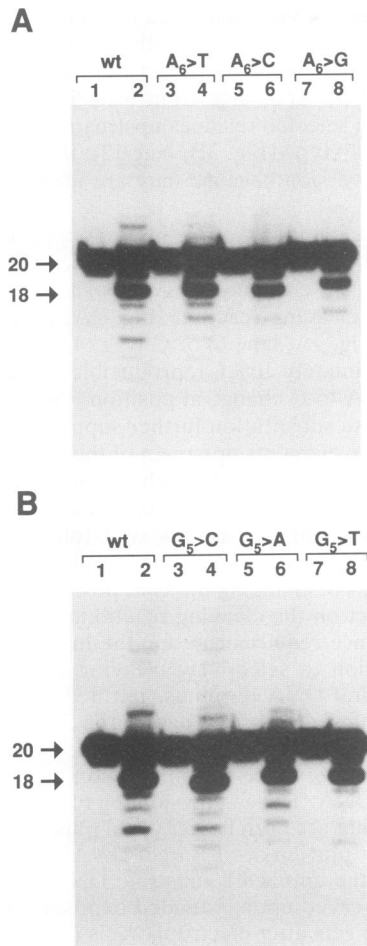


FIG. 4. Effect of changing the A at position 6 and the G at position 5 on selective cleaving of a mimic of HIV U5 DNA. Changes in the DNA sequence are indicated above the lanes of the autoradiogram. The first lane in each pair was a no-protein control. (A) The A at position 6 was changed to T, C, and G. (B) The G at position 5 was changed to C, A, and T. Numbers to the left indicate molecular size (in bases).

quence changes on the joining reaction from their effects on the 3' cleaving reaction, i.e., the cleaving reaction is not required to generate substrate for the joining reaction. It should be emphasized that the same oligonucleotides serve as both viral (donor) and target (acceptor) DNA in these experiments. Therefore, changes in sequence could influence the function of these oligonucleotides in their roles as mimics of either viral or target DNA. We have observed differences in the ladder of products from joining with DNA substrates of different sequences, indicating that the joining reaction does not occur in an entirely random fashion.

Using an 18:20 mimic of HIV U5 DNA as the substrate, we examined the effect of changing the conserved CA at the 3' end of the 18-mer (plus strand) on the joining reaction. As shown in Fig. 5A, changing the C to a T or the A to a C had a dramatic effect on the joining reaction; even after prolonged autoradiography, larger-than-substrate products of joining were barely detectable. The effect of single-base substitutions at positions 6 and 5 on the joining reaction was also examined. Changes at position 6 appeared to affect the joining reaction in the same way as they affected the selec-

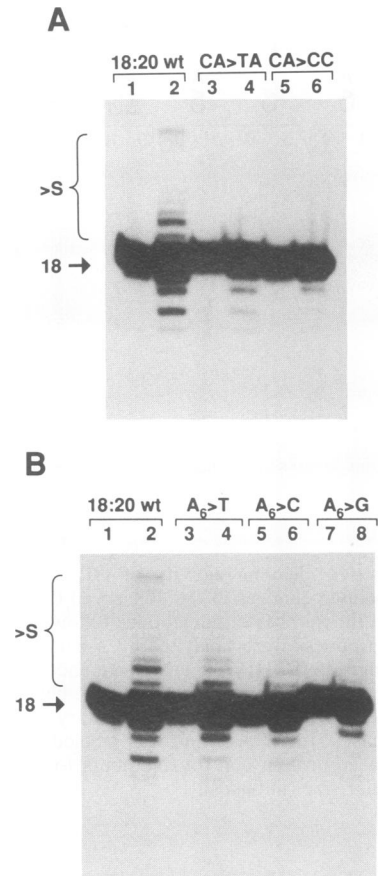


FIG. 5. Effect of changing the conserved CA and the A at position 6 on the joining reaction. An 18:20 mimic of HIV U5 DNA was the substrate. Changes in the DNA sequence are indicated above the lanes of the autoradiogram. The first lane in each pair was a no-protein control. Greater-than-substrate-length products of joining are indicated with brackets. (A) The conserved CA at the 3' end of the 18-mer (plus strand) was changed to TA and CC. (B) The A at position 6 was changed to T, C, and G. Numbers to the left indicate molecular size (in bases).

tive cleaving reaction, i.e., an A-to-T substitution had little effect, while an A-to-C or A-to-G substitution resulted in less joining (Fig. 5B). In contrast, while all changes at position 5 were well tolerated in the cleaving reaction, the G-to-T substitution caused a reduction in the joining reaction (Fig. 4B, lane 8).

We also examined the effect on the joining reaction of changing the number and identity of nucleotides in the overhang on the minus strand that results from the 3' processing reaction. The results of these experiments are shown in Fig. 6. Joining activity was reduced when the 2-base 5' overhang was removed (Fig. 6, lane 16) or when the penultimate base on the minus strand of HIV U5 was changed from a C to an A (Fig. 6, lanes 4 and 8). Changing the terminal A to a T (Fig. 6, lane 6) or changing (increasing or decreasing by one) the number of nucleotides in the overhang (Fig. 6, lanes 12 and 14) did not have a significant effect on the joining reaction. Taken together, these results suggest that an overhang containing at least 1 base (preferably not an A) is required for efficient joining and that the joining reaction is less sensitive than the cleaving reaction to the number of nucleotides following the conserved CA

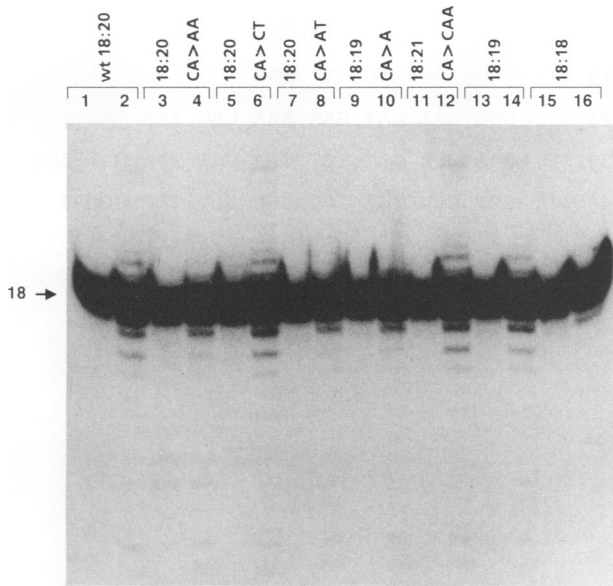


FIG. 6. Effect of changing the identity and the number of nucleotides in the 5' overhang (minus strand) of HIV U5 on the joining reaction. An 18:20 mimic of HIV U5 DNA was the substrate. Changes in the DNA sequence are shown above the lanes of the autoradiogram. The first lane in each pair was a no-protein control. The number to the left indicates molecular size (in bases).

(compare Fig. 6, lanes 12 and 14, with Fig. 2, lanes 12 and 14).

As was observed for the cleaving reaction, viral DNA sequence upstream of the terminal 6 bases influenced the joining reaction. Evidence for this influence was obtained with 20:20 mimics of HIV U3 and U5 DNA. The ends of HIV U3 and U5 DNA that are cleaved by HIV IN are different (the minus strand of U3 is TCCAGT and the plus strand of U5 is AGCAGT; Fig. 1A), yet products of selective cleaving and joining reactions were observed with both substrates when the appropriate strands were radiolabeled (Fig. 7, lanes 2 and 4). No joining was observed when the minus strand of HIV U5 was radiolabeled (Fig. 7, lane 6). The terminal sequence of the minus strand of HIV U5 is TC-CACA, which is similar to the terminal sequence of the minus strand of HIV U3 (TCCAGT). Joining was detected when the U3 minus strand sequence was changed to TC CACA (Fig. 7, lane 8), indicating that, in the context of U3 sequence elements upstream of the terminal 6 bases, the terminal sequence of the U5 minus strand (also TCCACA) could function as a substrate for joining. This observation was confirmed with a substrate in which selected upstream bases in the U5 minus strand were changed to a consensus (same in U3 minus and U5 plus strands) sequence, resulting in the restoration of joining activity (Fig. 7, lane 10).

The joining reaction was uncoupled from the selective cleaving reaction when the DNA substrate was a preprocessed 18:20 mimic of HIV U5. Another way to uncouple the two reactions was to remove a portion of the minus strand at the viral U5 DNA terminus. As shown in Fig. 8, substrates lacking 2 or 3 nucleotides at the 5' end of the minus strand (20:18 and 20:17) were cleaved more efficiently than the 20:20 HIV U5 substrate. However, no joining was observed with these substrates, even after prolonged autoradiography (data not shown), which is consistent with the earlier observation that an overhang (on the minus strand) of at least 1

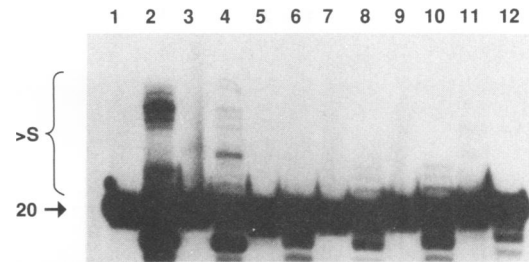


FIG. 7. Influence of viral sequence upstream of the terminal 6 bases on the joining reaction. The MOPS buffer system (see Materials and Methods) was used for these reactions. DNA substrates were 20:20 mimics of HIV U3 and U5 DNA. The first lane in each pair was a no-protein control. Lanes: 1 and 2, wild-type HIV U5, plus strand radiolabeled; 3 and 4, wild-type HIV U3, minus strand radiolabeled; 5 and 6, wild-type HIV U5, minus strand radiolabeled; 7 and 8, HIV U3 with terminal GT (minus strand) changed to CA, minus strand radiolabeled; 9 and 10, HIV U5 with the following changes in the minus strand: T8→C, T15→A, and C19→G, minus strand radiolabeled; 11 and 12, wild-type HIV U3, plus strand radiolabeled. The number to the left indicates molecular size (in bases).

base was required for efficient joining. The selective cleaving reaction, as well as the joining reaction, was greatly reduced when 5 nucleotides were removed from the 5' end of the minus strand (20:15; Fig. 8, lane 8). Lowering the temperature of the reaction mixture to 20°C (to ensure that the DNA was double stranded in the region of the first 15 bases) did not increase cleaving of this substrate.

The "authentic" HIV U5 terminus as a substrate for in vitro cleaving and joining. In virus-infected cells, the U5 terminus of the double-stranded nucleic acid product of HIV-1 reverse transcriptase may contain an adenosine 5'-phosphate (rA) on the minus strand resulting from the incomplete removal of the tRNA primer by HIV-1 RNase H (17, 30). To test this putative authentic HIV U5 terminus as a substrate for HIV IN, a 20-base oligonucleotide mimic of the minus strand of HIV U5 with a terminal rA was synthesized. This rA-terminated minus strand was then annealed to either a 20- or an 18-base oligonucleotide mimic of the HIV U5 plus strand (plus strands radiolabeled at the 5' termini). The resulting

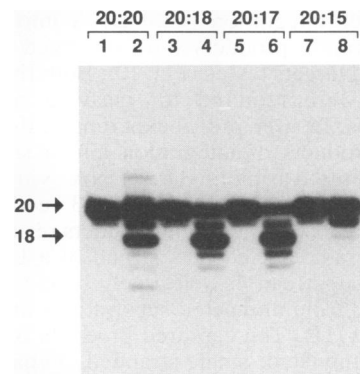


FIG. 8. Effect of removing nucleotides at the 5' end of the HIV U5 minus strand on cleaving and joining reactions. Oligonucleotide mimics of HIV U5 lacking 2 (20:18), 3 (20:17), or 5 (20:15) nucleotides on the 5' end of the minus strand were tested as substrates with HIV IN. DNA substrates are indicated above the lanes of the autoradiogram. The first lane in each pair was a no-protein control. Numbers to the left indicate molecular size (in bases).

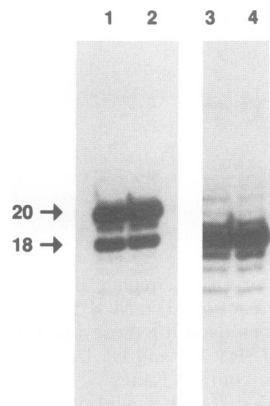


FIG. 9. The "authentic" HIV U5 terminus as a substrate for *in vitro* cleaving and joining by HIV IN. A 20-base oligonucleotide mimic of the minus strand of HIV U5 with a terminal rA was synthesized. This rA-terminated minus strand was annealed to a radiolabeled 20- or 18-base oligonucleotide mimic of the HIV U5 plus strand and tested as a substrate in the cleaving and joining reactions. Lanes: 1 (20:20) and 3 (18:20), DNA substrates; 2 (20:20) and 4 (18:20), rA-terminated substrates. Numbers to the left indicate molecular size (in bases).

20:20 and 18:20 duplex oligonucleotides were used as substrates in the cleaving and joining reactions. As shown in Fig. 9, the rA-terminated substrates were indistinguishable from the DNA substrates in both reactions.

Single-stranded DNA is an *in vitro* integration target for HIV IN. To further explore the DNA requirements for the joining step of the integration reaction, we tested the ability of single-stranded oligonucleotides to serve as integration targets. Various single-stranded oligonucleotides with sequences unrelated to the sequence of the HIV U5 terminus were radiolabeled at the 5' termini and incubated with HIV IN in the presence and absence of the duplex HIV U5 mimic shown in Fig. 1A. Under these conditions, integration of the U5 viral DNA mimic into the single-stranded targets would be expected to result in the release of labeled products smaller than the input DNA. The radiolabel was placed on the target rather than on the incoming viral DNA strand (U5 plus strand) in order to distinguish the products of integration of viral DNA into single-stranded target DNA from the products of self-integration (viral DNA into viral DNA). Putative integration products were observed for all of the single-stranded targets tested (Fig. 10). Both IN and the viral DNA mimic were required for cleaving of the single-stranded DNAs. In a reciprocal experiment, the larger-than-input DNA products of integration into a single-stranded target were captured as biotinylated DNA-avidin complexes. If a biotin moiety is attached to the 3' end of a single-stranded target DNA, integration products resulting from the joining of the radiolabeled plus strand of a U5 viral DNA mimic to the target can be selectively bound to avidin and hence resolved from unreacted substrate or the products of self-integration (11). The captured products of the reaction between a biotinylated, single-stranded, 20-base target and HIV U5 20:20 were the sizes predicted by the target cleavage experiment (data not shown).

The ability of a single-stranded oligonucleotide mimic of a preprocessed HIV U5 plus strand (18-mer ending in CA) to integrate into either a single-stranded or a double-stranded target DNA of sequence unrelated to HIV was also examined. The single-stranded viral DNA mimic was radiola-

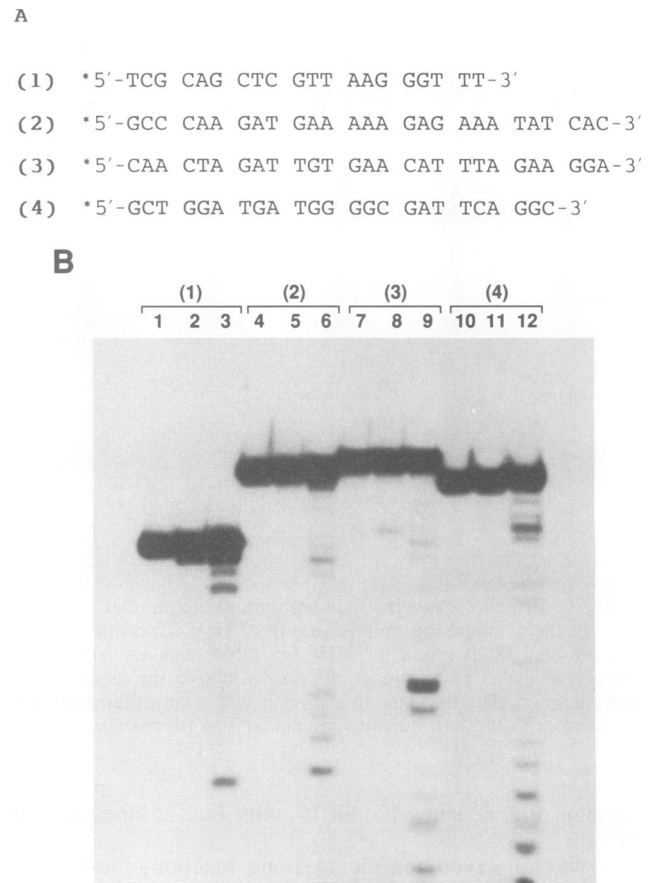


FIG. 10. Single-stranded DNA functions as an *in vitro* integration target for HIV IN. (A) Sequences of single-stranded oligonucleotides that were tested as targets for *in vitro* integration. Asterisks indicate the position of the radiolabel. (B) Analysis of reaction products on a denaturing polyacrylamide gel. Reactions contained 1 pmol of radiolabeled target DNA and 1 pmol of unlabeled HIV U5 20:20 "viral" DNA. The MOPS buffer system (see Materials and Methods) was used for these reactions. Target DNAs are indicated above the lanes of the autoradiogram. The first lane in each set of three was a no-protein control; the second lane was a reaction with no viral DNA; and the third lane was a reaction with viral DNA and target DNA.

beled. No integration products (larger-than-input DNA) were observed (data not shown).

Substrate specificity of the joining reaction catalyzed in the reverse direction. Chow et al. recently discovered that the joining reaction catalyzed by purified HIV IN with synthetic oligonucleotide substrates is reversible (6). To explore the substrate specificity of this reverse reaction, we prepared several oligonucleotide models of products of the joining reaction (Fig. 11A) and tested them as substrates with HIV IN. As shown in Fig. 11B, HIV IN converted several models (composed of single-stranded mimics of viral DNA joined to double-stranded targets) to unjoined single-stranded viral DNA termini and resealed targets. This was surprising, since we had earlier observed that a single-stranded mimic of viral DNA cannot be joined to either a single- or a double-stranded target DNA. Also surprising was the result that a model containing a viral DNA mimic that was only 2 nucleotides long (Fig. 11B, substrate B, lane 4) was converted almost as well as a product containing a longer

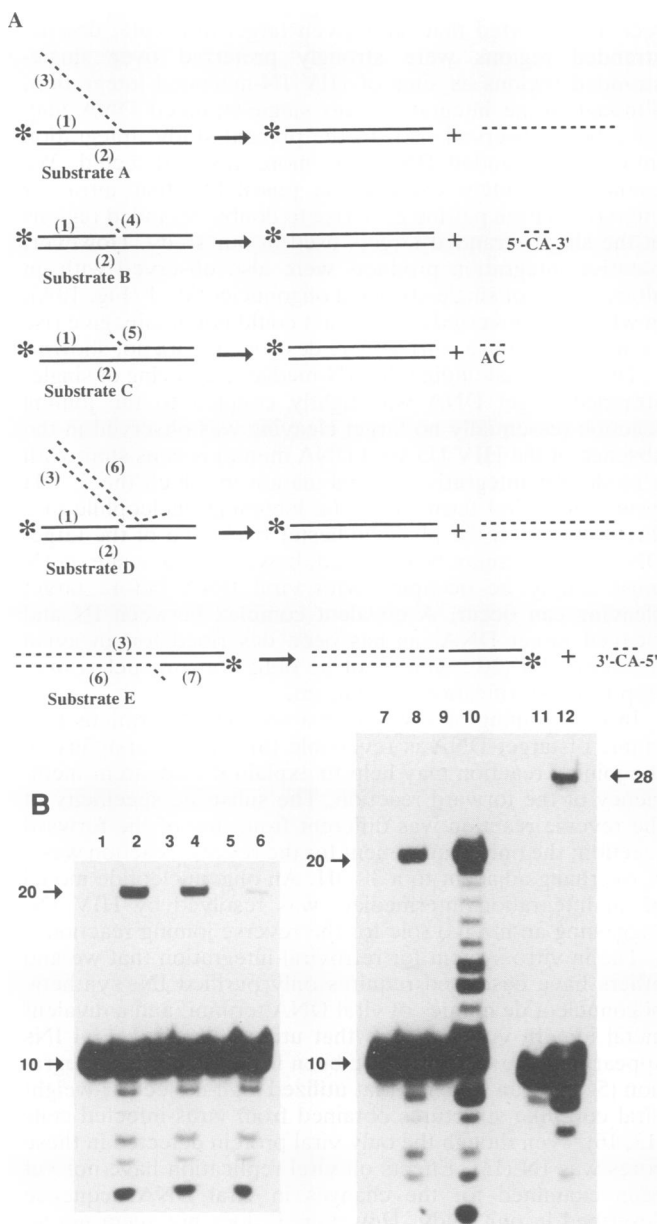


FIG. 11. Substrate specificity of the joining reaction catalyzed in the reverse direction. (A) Schematic of reverse reactions with different model substrates. Asterisks indicate the positions of the radiolabel. HIV DNA sequences are shown as broken lines, and sequences unrelated to HIV DNA are shown as solid lines. The oligonucleotides used to prepare the model substrates are indicated as follows: (1), 5'-TCGCAGCTCG-3'; (2), 5'-AAACCCTTAACGAGCTGCGA-3'; (3), 5'-TGTGGAAAATCTCTAGCATTAAGGGTTT-3', 18 bases of HIV U5 plus strand are underlined; (4), 5'-CATTAAGGGTTT-3'; (5), 5'-ACTTAAGGGTTT-3'; (6), 5'-ACTGCTAGAGATTTTCCACA-3', 20 bases of HIV U5 minus strand; and (7), 5'-AAACCCTTAA-3'. (B) Analysis of reaction products on denaturing polyacrylamide gels. The MOPS buffer system was used for these reactions. The first lane in each pair was a no-protein control. Lanes: 1 and 2, substrate A; 3 and 4, substrate B; 5 and 6, substrate C; 7 and 8, substrate A; 9 and 10, substrate D; 11 and 12, substrate E. Numbers to the left indicate molecular size (in bases).

(18-nucleotide) viral mimic (Fig. 11B, substrate A, lane 2). The efficiency of the reaction was greatly reduced when the 2-base viral mimic was AC instead of CA (compare substrate B, lane 4, with substrate C, lane 6, in Fig. 11B). A model product of the joining of an 18:20 double-stranded mimic of viral DNA to a double-stranded target (Fig. 11B, substrate D) was converted much more efficiently than the model containing an 18-base, single-stranded viral DNA mimic (Fig. 11B, lanes 8 and 10). Finally, Chow et al. have suggested that the final ligation step in the resolution of the integration intermediate, which has been attributed to host-cell enzymes, is similar to the reverse of the joining reaction catalyzed by purified HIV IN. We tested an oligonucleotide model of such an integration intermediate as a substrate for IN and found that conversion to a model of a mature provirus did occur (Fig. 11B, substrate E, lane 12).

DISCUSSION

Retroviral integration proteins are the only viral proteins required to carry out integrative recombination *in vitro*. This process involves selective endonucleolytic cleaving of the viral DNA termini followed by insertion of the cleaved viral DNA into a target DNA. The simplest *in vitro* system for the study of retroviral integration utilizes purified IN and synthetic, double-stranded oligonucleotide mimics of viral DNA termini, which can function as both viral and target DNA. This simple system offers the opportunity to investigate in detail the DNA sequence requirements for the IN-mediated cleaving and joining reactions.

Although the sequences at the extreme termini of HIV-1 DNA have been a matter of speculation, it has now been well established that 2 bases (GT) are deleted from both ends of HIV-1 linear DNA during integrative recombination (21, 23, 27, 30). It is, therefore, not surprising that in an HIV IN-mediated selective cleaving reaction, an oligonucleotide mimic of HIV U5 DNA with 2 bases following the conserved CA was preferred over substrates containing 1 or 3 bases following the CA. A similar result was recently reported by Vink et al. (29). We also observed some preference for the wild-type GT in the terminal position over other 2-base combinations. These findings contrast with the results obtained *in vivo* for MLV in which the MLV IN appeared to tolerate substantial changes in the number and identity of nucleotides at positions following the conserved CA (9). Our results do suggest that the low efficiencies observed in other *in vitro* integration systems, especially with blunt-ended mimics of viral DNA, may have been due in part to nonauthentic viral termini (15, 18).

We also observed, as have others (10, 22, 29), that both the C and the A in the CA dinucleotide found near the termini of all retroviral DNAs were required for efficient endonucleolytic processing. However, a C-to-T change in the conserved CA of MLV DNA resulted in only a slight reduction in the level of the 3'-recessed viral DNA in infected cells and in the rate of replication compared with wild type (24). The apparent discrepancies between our *in vitro* results with the HIV IN and the *in vivo* results with the MLV IN may reflect differences in the assay systems or in the viral integration proteins. It would be interesting to determine the effect of a double mutation in the conserved CA, which resulted in no detectable selective cleaving in our *in vitro* assay, on MLV replication.

We previously reported that a double-stranded oligonucleotide mimic of HIV U5 DNA must be 12 to 15 bases long in order to be selectively cleaved by HIV IN (26). However,

adding 10 bases of unrelated sequence to just 10 bases of wild-type HIV U5 sequence resulted in a substrate that was cleaved almost as well as the substrate with 20 bases of wild-type sequence. In the present study, we found that as few as 6 bases of HIV U3 or U5 DNA at the ends of model substrates were sufficient for nearly maximal levels of selective cleaving. In related studies, the terminal 9 bases of HIV U3 DNA (4) or the terminal 15 bases of HIV U5 DNA (29) were shown to be sufficient for *in vitro* cleaving and joining by HIV IN, the terminal 9 bases of MLV U3 and U5 DNA were sufficient for nearly maximal levels of *in vitro* integration by MLV IN (3), and the minimal length of a duplex oligonucleotide substrate for selective cleaving by AMV IN was 15 bases of AMV U5 and 13 to 15 bases of AMV U3 DNA (19). Although very few bases of viral DNA sequence appeared to be required for *in vitro* cleaving and joining reactions, we observed that additional upstream viral sequence elements influenced the reaction. Five terminal bases of HIV U5 DNA in the context of a U5 DNA backbone (some, but not all, of the upstream nucleotides were wild type) supported the cleaving reaction.

Alterations in the sequence of the viral DNA termini had an effect on the joining of the viral DNA to target DNA as well as an effect on the selective cleaving reaction. A preprocessed oligonucleotide mimic of HIV U5 DNA made it possible to separate the effects of DNA sequence changes on the joining reaction from effects on the selective cleaving reaction. We found that the conserved CA and the G at position 5 of the HIV U5 substrate were required for optimal joining activity. In addition, an overhang of at least 1 nucleotide on the minus strand of the preprocessed HIV U5 substrate was required for joining, and joining was more efficient when the base in this overhang was a C rather than an A. The former requirement is in agreement with a report that an MLV mutant lacking the terminal 2 bp of linear viral DNA was unable to establish a productive infection (8). The latter requirement might help to explain the low efficiency of *in vitro* integration observed in an HIV system in which the C in the overhang was replaced by an A (5).

The cleaving reaction was uncoupled from the joining reaction by "precleaving" the HIV U5 DNA mimic at the 3' end of the plus strand. Removing 2 or 3 nucleotides from the 5' end of the minus strand of a mimic of HIV U5 DNA also appeared to uncouple the two reactions, resulting in enhanced cleaving and no detectable joining. The decrease in joining is in agreement with the earlier observation that a 5' overhang on the minus strand was required for optimal joining. Vink et al. also reported an increase in cleaving activity with a 20:18 U5 substrate (29). The 18:18 and 18:17 products may be released more rapidly from the protein than the normal 18:20 product, resulting in a faster turnover in the cleaving reaction and less conversion of products of cleaving to products of joining.

Integration of retroviral DNA is expected to occur as a pairwise insertion of the two processed ends of the viral DNA into the site on the host cellular DNA where a double-stranded, staggered break is made. Reaction products, consistent with the concerted integration of pairs of oligonucleotide mimics of viral DNA termini, have been observed with purified MLV IN (10). However, in this same system with HIV IN, a predilection for integrating oligonucleotide mimics of HIV DNA in a singular rather than a pairwise fashion was observed (4). Cooperative interaction between IN subunits could facilitate pairwise integration, which probably requires a double-stranded DNA target. Perhaps because of a lack of cooperativity between IN

subunits, we have found that single-stranded oligonucleotides can serve as integration targets for HIV IN. Vink et al. recently reported that on a given target molecule, double-stranded regions were strongly preferred over single-stranded regions as sites of HIV IN-mediated integration, although some integration into single-stranded DNA may have been observed (29). In the present study, integration into single-stranded DNA was more easily detected. We cannot completely exclude the possibility that intra- or interstrand base pairing gave rise to double-stranded regions in the single-stranded targets used in our study. However, putative integration products were also observed with an altered form of single-stranded oligonucleotide 1 (Fig. 10A), in which two inverted repeats that could potentially give rise to interstrand base pairing were destroyed (data not shown).

The additional finding that IN-mediated cleaving of single-stranded target DNA was tightly coupled to the joining reaction (essentially no target cleaving was observed in the absence of the HIV U5 viral DNA mimic) is consistent with a model for integrative recombination in which the 3' OH group on a viral terminus is the incoming nucleophile in a direct attack on the phosphodiester backbone of the target DNA (12). It cannot be excluded, however, that a site on IN must simply be occupied with viral DNA before target cleaving can occur. A covalent complex between IN and cleaved target DNA, as has been described for an avian retrovirus IN (20), would not be obligatory in such a one-step transesterification mechanism.

In vitro joining of a mimic of a viral DNA terminus to a mimic of target DNA is reversible (6). The reversibility of the joining reaction may help to explain the apparent inefficiency of the forward reaction. The substrate specificity of the reverse reaction was different from that of the forward reaction; the only requirement for the reverse reaction was a 5' overhang adjacent to a 3' OH. An oligonucleotide model of an integration intermediate was resolved by HIV IN, suggesting an *in vivo* role for the reverse joining reaction.

The *in vitro* system for retroviral integration that we and others have described requires only purified IN, synthetic oligonucleotide mimics of viral DNA termini, and a divalent metal ion. *In vitro* systems that utilized purified viral INs appear in some cases to have been less efficient for integration (5, 18) than systems that utilized high-molecular-weight viral core-like structures obtained from virus-infected cells (13, 16), even though the only viral protein detected in these cores was IN (14). Effects on viral replication have not yet been examined for the changes in viral DNA sequence described in our study. However, it does not seem unreasonable to predict that the simplified *in vitro* system may be a fairly accurate representation of integration in a virus-infected cell, particularly with respect to DNA substrate specificity. As such, this system should be useful for future studies aimed at understanding the mechanism of retroviral integration and inhibiting this essential step in the viral life cycle.

ACKNOWLEDGMENTS

Thanks to W. Dallas (Wellcome Research Laboratories) for synthesis of oligonucleotides and T. Krenitsky (Wellcome Research Laboratories) for continued support of this work.

REFERENCES

1. 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 integrase protein. *Proc. Natl. Acad. Sci. USA* **86**:2525-2529.

2. **Buchow, H. D., E. Tschachler, R. C. Gallo, and M. Reitz, Jr.** 1989. HIV-1 replication requires an intact integrase reading frame. *Hematol. Bluttransfus.* **32**:402-405.
3. **Bushman, F. D., and R. Craigie.** 1990. Sequence requirements for integration of Moloney murine leukemia virus DNA *in vitro*. *J. Virol.* **64**:5645-5648.
4. **Bushman, F. D., and R. Craigie.** 1991. Activities of human immunodeficiency virus (HIV) integration protein *in vitro*: specific cleavage and integration of HIV DNA. *Proc. Natl. Acad. Sci. USA* **88**:1339-1343.
5. **Bushman, F. D., T. Fujiwara, and R. Craigie.** 1990. Retroviral DNA integration directed by HIV integration protein *in vitro*. *Science* **249**:1555-1558.
6. **Chow, S., K. Vincent, V. Ellison, and P. Brown.** Science, in press.
7. **Clavel, F., M. D. Hoggan, R. W. Willey, K. Strebel, M. A. Martin, and R. Repaske.** 1989. Genetic recombination of human immunodeficiency virus. *J. Virol.* **63**:1455-1459.
8. **Colicelli, J., and S. P. Goff.** 1985. Mutants and pseudorevertants of Moloney murine leukemia virus with alterations at the integration site. *Cell* **42**:573-580.
9. **Colicelli, J., and S. P. Goff.** 1988. Sequence and spacing requirements of a retrovirus integration site. *J. Mol. Biol.* **199**:47-59.
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. **Craigie, R., K. Mizuuchi, F. D. Bushman, and A. Engelman.** 1991. A rapid *in vitro* assay for HIV DNA integration. *Nucleic Acids Res.* **19**:2729-2734.
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. **Farnet, C. M., and W. A. Haseltine.** 1990. Integration of human immunodeficiency virus type 1 DNA *in vitro*. *Proc. Natl. Acad. Sci. USA* **87**:4164-4168.
14. **Farnet, C. M., and W. A. Haseltine.** 1991. Determination of viral proteins present in the human immunodeficiency virus type 1 preintegration complex. *J. Virol.* **65**:1910-1915.
15. **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.
16. **Fujiwara, T., and K. Mizuuchi.** 1988. Retroviral DNA integration: structure of an integration intermediate. *Cell* **55**:497-504.
17. **Furfine, E. S., and J. E. Reardon.** 1991. Human immunodeficiency virus reverse transcriptase ribonuclease H: specificity of tRNA^{Lys3}-primer excision. *Biochemistry* **30**:7041-7046.
18. **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.
19. **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.
20. **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.
21. **Kulkosky, J., R. A. Katz, and A. M. Skalka.** 1990. Terminal nucleotides of the preintegrative linear form of HIV-1 DNA deduced from the sequence of circular DNA junctions. *J. Acquired Immune Defic. Syndr.* **3**:852-858.
22. **LaFemina, R. L., P. L. Callahan, and M. G. Cordingley.** 1991. Substrate specificity of recombinant human immunodeficiency virus integrase protein. *J. Virol.* **65**:5624-5630.
23. **Pauza, C. D.** 1990. Two bases are deleted from the termini of HIV-1 linear DNA during integrative recombination. *Virology* **179**:886-889.
24. **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.
25. **Sherman, P. A., and J. A. Fyfe.** 1990. Human immunodeficiency virus integration protein expressed in *Escherichia coli* possesses selective DNA cleaving activity. *Proc. Natl. Acad. Sci. USA* **87**:5119-5123.
26. **Sherman, P. A., and J. A. Fyfe.** 1991. DNA cleaving activity of purified human immunodeficiency virus integration protein, p. 27-33. *In* A. Kumar (ed.), *Advances in molecular biology and targeted treatment for AIDS*. Plenum Press, New York.
27. **Smith, J. S., S. Kim, and M. J. Roth.** 1990. Analysis of long terminal repeat circle junctions of human immunodeficiency virus type 1. *J. Virol.* **64**:6286-6290.
28. **Stevenson, M., T. L. Stanwick, M. P. Dempsey, and C. A. Lamonica.** 1990. HIV-1 replication is controlled at the level of T cell activation and proviral integration. *EMBO J.* **9**:1551-1560.
29. **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.
30. **Whitcomb, J. M., R. Kumar, and S. H. Hughes.** 1990. Sequence of the circle junction of human immunodeficiency virus type 1: implications for reverse transcription and integration. *J. Virol.* **64**:4903-4906.