Proc. Natl. Acad. Sci. USA
Vol. 88, pp. 1339–1343, February 1991
Biochemistry

Activities of human immunodeficiency virus (HIV) integration protein *in vitro*: Specific cleavage and integration of HIV DNA

(retroviral DNA integration/integrase protein/long terminal repeat/integrative recombination)

FREDERIC D. BUSHMAN AND ROBERT CRAIGIE

Laboratory of Molecular Biology, National Institutes of Diabetes, Digestive and Kidney Diseases, Bethesda, MD 20892

Communicated by Howard Nash, November 16, 1990 (received for review October 18, 1990)

ABSTRACT **Growth of human immunodeficiency virus (HIV) after infection requires the integration of a DNA copy of the viral RNA genome into a chromosome of the host. Here we present a simple *in vitro* system that carries out the integration reaction and the use of this system to probe the mechanism of integration. The only HIV protein necessary is the integration (IN) protein, which has been overexpressed in insect cells and then partially purified. DNA substrates are supplied as oligonucleotides that match the termini of the linear DNA product of reverse transcription. In the presence of HIV IN protein, oligonucleotide substrates are cleaved to generate the recessed 3′ ends that are the precursor for integration, and the cleaved molecules are efficiently inserted into a DNA target. Analysis of reaction products reveals that HIV IN protein joins 3′ ends of the viral DNA to 5′ ends of cuts made by IN protein in the DNA target. We have also used this assay to characterize the sequences at the ends of the viral DNA involved in integration. The assay provides a simple screen for testing candidate inhibitors of HIV IN protein; some such inhibitors might have useful antiviral activity.**

Growth of a retrovirus after infection of a permissive cell requires establishment of the proviral state. Early after infection, reverse transcriptase synthesizes a DNA copy of the viral RNA genome, which is subsequently integrated into a chromosome of the host cell. The viral integration (IN) protein is required for this integration reaction (refs. 1–5; for recent reviews, see refs. 6 and 7). Circular and linear forms of the viral DNA are found in cells after infection, but only the linear form appears to be integrated (8–10).

Studies of Moloney murine leukemia virus (MoMLV) and avian sarcoma-leukosis virus (ASLV) have delineated the DNA cleavage and strand transfer reactions that mediate integration. The blunt ends of the linear product of reverse transcription are first cleaved to remove two bases from each 3′ end. The IN protein of MoMLV is required for this cleavage *in vivo* (11), and the IN proteins of MoMLV and ASLV carry out this reaction *in vitro* (12–14). After this first cleavage reaction, a DNA strand transfer step joins each 3′ each of the viral DNA to the protruding 5′ ends of a staggered break in the DNA target (8, 9). The IN proteins of MoMLV and ASLV carry out the strand transfer reaction *in vitro* (12, 15). The single-stranded connections between the ends of the viral and host DNAs are then converted to double-strand joints, presumably by cellular DNA repair enzymes, to yield the integrated provirus.

As a first step toward developing a fully defined system that carries out integration of human immunodeficiency virus (HIV) DNA, we partially purified HIV IN protein from insect cells expressing a cloned copy of the IN coding region. A low level of correct DNA integration was detected *in vitro* in the presence of this IN protein fraction by a genetic assay that monitored integration of a model substrate DNA into λ DNA (16). The model substrate, termed miniHIV, was a linear DNA molecule with ends that resemble those of unintegrated HIV DNA following the cleavage step; thus only the subsequent DNA strand transfer step of integration was assayed.

Here we report a simple and more efficient assay for both the long terminal repeat (LTR) cleavage and DNA strand transfer activities of HIV IN protein and the use of this assay in dissecting the mechanism of HIV DNA integration. Sherman and Fyfe have also reported a specific LTR cleavage activity of HIV IN protein (17).

MATERIALS AND METHODS

Partial Purification of Soluble HIV IN Protein. Overexpression and partial purification of insoluble HIV IN protein have been described (16). The HIV IN-containing insoluble fraction from two 150-cm^2 tissue culture flasks, prepared as described (16), was solubilized in a buffer containing 4 M urea, fractionated on a Superose 12 column, and dialyzed to remove the urea exactly as described for MoMLV IN protein (12). HIV IN protein was the predominant species in the preparation as determined by Coomassie blue staining of SDS/PAGE gels. The concentration of HIV IN protein was estimated by comparison to bovine serum albumin standards.

***loxP-cre* Recombination.** The 174-base-pair (bp) circular DNA target was prepared by recombination of plasmid pAB253 (18) in the presence of the Cre protein of bacteriophage P1. Cre protein was kindly provided by Andrew Bates (University of Leicester). Reactions were carried out as described (19). One-hundred seventy-four-base-pair circular DNA products were isolated on a native polyacrylamide gel, eluted from the gel slice, and purified by chromatography on a G-25 Sephadex column.

LTR Cleavage and Insertion Assays. Reaction mixtures (16 μl) contained 25 mM KCl, 50 mM potassium glutamate, 10 mM 2-mercaptoethanol, 30 mM Mes (pH 6.2), 15 mM $MnCl_2$, 10% glycerol, 100 μg of bovine serum albumin per ml, 0.3 pmol of labeled duplex LTR DNA, and ≈10 pmol of HIV IN protein. Reaction mixtures were incubated for 1 hr at 30°C. End labeling and hybridization of complementary strands of oligonucleotide substrates were carried out as described (12). Analysis of reaction products on DNA sequencing gels was conducted as described (12). Preparative reaction mixtures for characterization of LTR insertion products were scaled up 4-fold, and the products of four such reactions were gel isolated and eluted from the gel slice by the crush-soak method (20). Aliquots were then cleaved by the Maxam–Gilbert chemical sequencing method (21). Assays of insertion of LTR substrates into a heterologous target were as above except 40 fmol of the 174-bp circular DNA was added to each reaction mixture, the incubation temperature was increased

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

Abbreviations: HIV, human immunodeficiency virus; LTR, long terminal repeat; MoMLV, Moloney murine leukemia virus.

to 37°C, and the incubation time was increased to 2 hr. Reaction products were separated on 5% native polyacrylamide gels in TBE buffer (20) and visualized by autoradiography.

RESULTS

Specific Cleavage of the LTR DNA Termini by HIV IN Protein. The production of HIV IN protein in insect cells infected with a baculovirus expressing the HIV IN coding region, and partial purification of insoluble IN protein from this source, have been described (16). The IN protein used for the experiments described here was solubilized and further purified, yielding a preparation in which soluble IN protein comprises at least 50% of the total protein.

To determine whether HIV IN protein can specifically cleave the termini of the unintegrated linear viral DNA, oligonucleotides matching the LTR sequences at the predicted left (U3) and right (U5) ends of the unintegrated HIV DNA (22, 23) were prepared and assayed *in vitro* (Fig. 1; LTR A and LTR B). Oligonucleotide LTR substrates were labeled on the 5′ end of the DNA strand expected to be cleaved by IN protein (the bottom strand as shown in Fig. 1). Reaction products were separated by electrophoresis on a denaturing polyacrylamide gel and visualized by autoradiography (Fig. 2*A*). A labeled DNA two bases shorter than the LTR A and LTR B substrates was produced in the reaction mixtures containing HIV IN protein (lane 4 of each panel). No product is detected in reaction mixtures in which a control fraction from insect cells infected with a wild-type baculovirus was substituted for the fraction containing HIV IN protein (LTR A and LTR B, lane 3). Substituting EDTA for $MnCl_2$ eliminates cleavage (LTR A, lane 5). The product of cleavage by HIV IN protein migrates slightly more slowly than the

SUBSTRATE	OLIGONUCLEOTIDE SEQUENCE	LTR CLEAVAGE	STRAND TRANSFER
	1 21		
LTR A (U3 end)	5'ACTGGAAGGGCTAATTCACTC 3' 3'TGACCTTCCCGATTAAGTGAG 5'	+	+
LTR B (U5 end)	5'ACTGCTAGAGATTTTCCACAC 3' 3'TGACGATCTCTAAAAGGTGTG 5'	+	+
LTR C (precut U3 end)	5'ACTGGAAGGGCTAATTCACTC 3' 3' ACCTTCCCGATTAAGTGAG 5'	NA	+
LTR D (U3 end +1 bp)	5'GACTGGAAGGGCTAATTCACTC 3' 3'CTGACCTTCCCGATTAAGTGAG 5'	+	+
LTR E (U3 end -1 bp)	5'CTGGAAGGGCTAATTCACTC 3' 3'GACCTTCCCGATTAAGTGAG 5'	+	+
LTR F (U3 T3->C)	5'AC[C]GGAAGGGCTAATTCACTC 3' 3'TG[G]CCTTCCCGATTAAGTGAG 5'	-	-
LTR G (SnaB 1 4-9)	5'ACT[TACGTA]GCTAATTCACTC 3' 3'TGA[ATGCAT]CGATTAAGTGAG 5'	-	-
LTR H (SnaB 1 10-15)	5'ACTGGAAGG[TACGTA]TCACTC 3' 3'TGACCTTCC[ATGCAT]AGTGAG 5'	+	+
LTR I (SnaB 1 16-21)	5'ACTGGAAGGGCTAAT[TACGTA] 3' 3'TGACCTTCCCGATTA[ATGCAT] 5'	+	+

FIG. 1. Sequence and activities of LTR substrate DNAs. The U3 and U5 LTR substrates match the predicted LTR ends of the lymphadenopathy-associated virus/human T-lymphotropic virus type III isolate of HIV (24, 25). For LTR A and LTR B, base pair 1 marks the inferred flush end of the unintegrated linear precursor of the proviral DNA. A minus sign in the "LTR Cleavage" column indicates a lack of detectable cleavage above the slight background seen in control reaction mixtures containing the corresponding protein fraction from insect cells infected with a wild-type baculovirus. A minus sign in the "Strand Transfer" column indicates that no strand transfer product was detected. Plus signs denote sequences that are cleaved or inserted at a frequency comparable to LTR A and LTR B. Whether LTR F and LTR G would function as a substrate for strand transfer after removal of the two 3′ bases (positions 1 and 2) has not been tested. NA, not applicable.

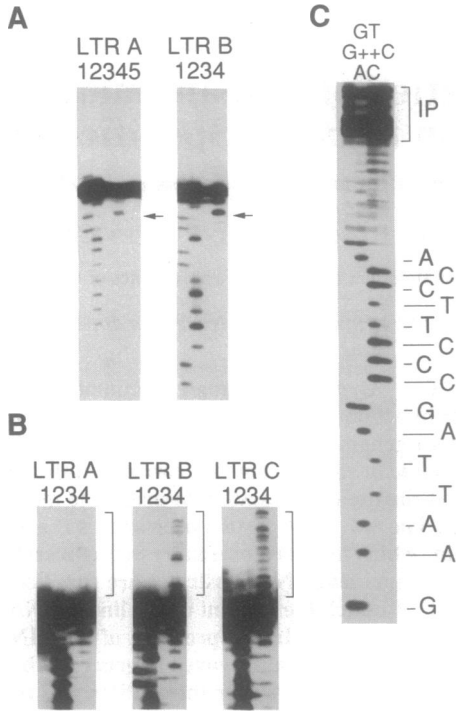

FIG. 2. LTR cleavage and DNA strand transfer activities of HIV IN protein. (*A*) Cleavage of LTR A and LTR B (see Fig. 1 for sequences) in the presence of HIV IN protein. In each panel, lanes 1 and 2 contain purine and pyrimidine chemical cleavage products as size markers, lane 3 contains the products of a complete reaction except that a control protein fraction from insect cells infected with a wild-type baculovirus was substituted for IN protein, lane 4 contains a complete reaction mixture with HIV IN protein, and lane 5 contains a reaction mixture identical to lane 4 except that 15 mM EDTA is substituted for 15 mM $MnCl_2$. The arrow marks the expected mobility of the product of cleavage by IN. (*B*) Transfer of a LTR DNA strand to an oligonucleotide target. Reactions are as in *A*, except longer exposures to the autoradiograms are shown. LTR substrates are as indicated; lanes are labeled as in *A*. Brackets above the unreacted substrate mark the position of the signal from the strand transfer products. (*C*) DNA sequence of the labeled strand in a population of strand transfer products. Products made with LTR A were purified from a denaturing polyacrylamide gel and subjected to base-specific chemical cleavage reactions. The products were then separated by electrophoresis in a second sequencing gel. Chemical cleavage reactions are as indicated. Strand transfer products that remained uncleaved after the chemical cleavage reactions are labeled IP.

products of chemical cleavage of the LTR substrate (lanes 1 and 2 of each panel), as would be expected if the molecules cleaved by IN protein lack a 3′ phosphate.

Insertion of Cleaved LTR Substrates into Target DNA. In a previous study, we found that oligonucleotides matching the MoMLV LTR termini could be cleaved and subsequently inserted into other LTR substrate molecules in the presence of MoMLV IN protein (12). A schematic diagram of this DNA strand transfer reaction (Fig. 3*A*) shows that two new labeled species are produced. A product longer than the initial LTR substrate is formed by joining of the 3′ end of the cleaved LTR substrate to the 5′ end of a cut made in another LTR substrate molecule that acts as the target DNA. A second product smaller than the initial substrate corresponds to the remaining unjoined fragment of the target DNA strand. These products are expected to be heterogenous in size because insertion can occur at many positions in the target DNA.

Such products are also obtained in reaction mixtures containing HIV IN protein. Longer exposures of autoradio-

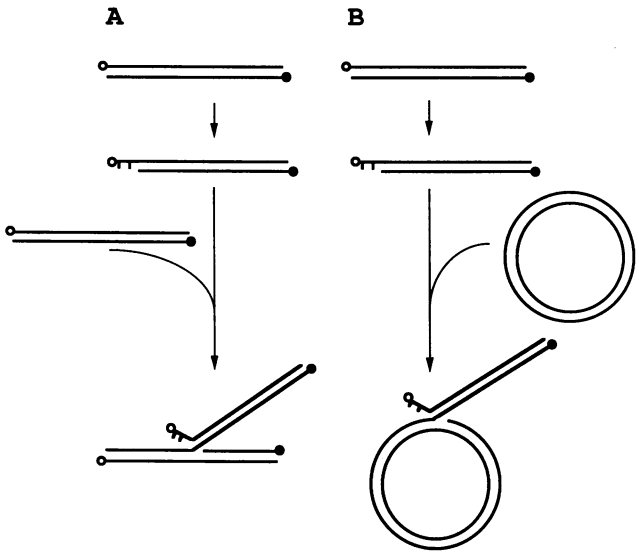

FIG. 3. (*A*) Schematic diagram of the LTR cleavage and strand transfer reactions. LTR oligonucleotides can be cleaved in the presence of IN. Cleaved oligonucleotides can then be inserted into other oligonucleotides present in the reaction mixture such that the cleaved 3′ ends become joined to 5′ ends in the cleaved target DNA. The 5′ ends are marked with circles; the label at the 5′ end of the strand that is expected to be cleaved by IN protein at its 3′ end is distinguished by a filled circle. (*B*) Schematic diagram of insertion of a LTR substrate into the 174-bp circular DNA target to produce the "tagged circle" product.

grams such as those presented in Fig. 2*A* reveal the presence of products resembling those expected for proper strand transfer. A "ladder" of labeled bands of higher molecular weight than the unreacted substrate is seen with LTR A and LTR B that terminates at a length approximately twice that of two substrate molecules (Fig. 2*B*, lane 4 of each panel). No such products are detectable when the HIV IN fraction is replaced by a control fraction from insect cells infected with a wild-type baculovirus (lane 3 of each panel). The lower molecular weight DNA products expected from the strand transfer reaction are also detected. Some low molecular weight products are also seen with control reactions and are probably products of cleavage by contaminating cellular nucleases.

Characterization of Insertion Products. The ladder of longer DNA products made in reaction mixtures containing HIV IN protein was analyzed to determine whether these DNAs have the structure expected of authentic strand transfer products. DNA from the upper two-thirds of a product ladder was purified from a gel, and aliquots were cleaved by the Maxam–Gilbert chemical sequencing method. The resulting chemical cleavage products were visualized by autoradiography after electrophoresis on a denaturing gel (Fig. 2*C*). The scheme shown in Fig. 3*A* predicts that the sequence should be that of the LTR substrate up to the 3′ A residue at the site of cleavage by IN protein. Beyond this position, a labeled band should be seen in all lanes, since the cleaved LTR DNA is inserted adjacent to different target bases in different product molecules. The sequence of reaction products shown in Fig. 2*C* exactly matches that expected.

Effects of Modifications of the LTR Substrate. To ask whether LTR cleavage by IN protein itself is required for the subsequent strand transfer reaction, a substrate with a recessed 3′ end resembling the product of cleavage by IN was assayed (LTR C; Fig. 1). Fig. 2*B* shows that the higher molecular weight product ladder was also obtained with this substrate. In fact, LTR C is reproducibly more active than LTR A, the matching blunt-ended substrate, which requires cleavage before strand transfer.

The importance of the distance between the A residue at the site of cleavage (position 3 in the bottom strand) and the proximal LTR end was examined by analyzing LTR D and LTR E (Fig. 1). These substrates contain 3 bp (LTR D) or 1 bp (LTR E) 3′ of the A, instead of the usual 2 bp (LTR A). LTR D and LTR E were cleaved adjacent to the correct A residue and then inserted into target DNA, indicating that the internal sequence, rather than the distance from the proximal DNA end, primarily dictates the position of cleavage (data not shown, summarized in Fig. 1).

The internal LTR sequences required for LTR cleavage and strand transfer were mapped by analyzing the activity of LTR substrates containing base substitutions. The A residue at position 3 on the bottom strand is conserved in all retroviruses sequenced and is important for integration in the MoMLV system *in vitro* (12, 26). The importance of this A residue in the HIV system was tested with LTR F, a LTR substrate in which the conserved A residue is changed to G. No specific cleavage or DNA insertion was detected (Fig. 1). Substitution of bases 4–9 with a recognition site for the enzyme *Sna*BI (LTR G) abolished detectable cleavage and strand transfer, but replacing bases 10–15 (LTR H) or bases 16–21 (LTR I) with the *Sna*BI site had no obvious effect on the efficiency of cleavage and strand transfer (Fig. 1).

Insertion into a Heterologous DNA Target. To determine whether the strand transfer products detected in these experiments result primarily from insertion of single LTR substrate molecules or concerted insertion of pairs of LTR substrates, a 174-bp circular DNA was tested as a target in the *in vitro* assay. A new labeled product was detected in reaction mixtures containing a 5′ labeled LTR substrate, the circular target DNA, and partially purified HIV IN protein after

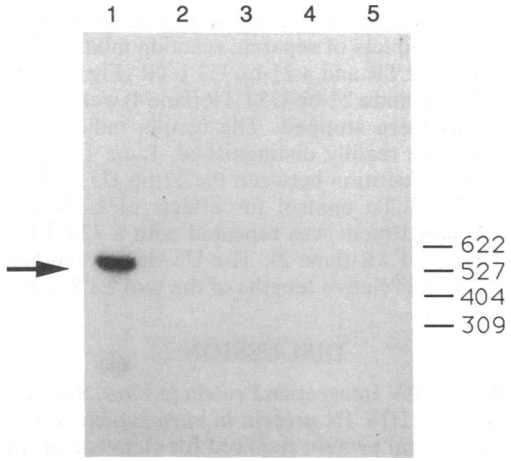

FIG. 4. Insertion of model LTR substrates into a circular DNA target. Standard strand transfer reactions, modified as described, were carried out in the presence of 5 ng of a 174-bp circular DNA. Lane 1, complete reaction mixture. The predominant product is marked with the arrow. Lane 2, 15 mM $MgCl_2$ substituted for 15 mM $MnCl_2$; lane 3, 15 mM EDTA substituted for 15 mM $MnCl_2$; lane 4, target DNA omitted; lane 5, control protein fraction from insect cells infected with a wild-type baculovirus substituted for the IN protein-containing fraction. The dashes beside the gel mark the mobility of size standards (pBR322 digested with *Msp* I) of the indicated length (given in bp). The labeled LTR substrate used in this experiment, LTR J, is identical to LTR A except for an addition of 11 bp to the right side as drawn in Fig. 1. The sequence of this addition was 5′-GGATCCTATCG-3′ and its complement. The use of the 174-bp circular target, synthesized using the *loxP-cre* recombination system of phage P1, and this lengthened LTR substrate facilitated subsequent characterization of reaction products after denaturation (see text).

electrophoresis on a native polyacrylamide gel and autoradiography (Fig. 4, lane 1). Production of the labeled species was undetectable in reaction mixtures in which $MgCl_2$ (lane 2) or EDTA (lane 3) was substituted for $MnCl_2$, in which the target DNA was omitted (lane 4), or in which a control fraction from insect cells infected with a wild-type baculovirus was substituted for the HIV IN preparation (lane 5). The labeled product migrates slightly more slowly than the most relaxed topoisomer of the 174-bp circle (data not shown), as is expected if the product is formed by insertion of one labeled LTR duplex into the circular target to yield a tagged circle (Fig. 3*B*). In support of this view, it was found that reaction mixtures containing 404-bp or 2686-bp circular target DNAs yield products comigrating with the relaxed circular forms of these target DNAs (data not shown). The labeled linear product resulting from concerted integration of pairs of LTR substrates was not detected in these experiments (see *Discussion*).

To further analyze the structure of the putative tagged circle DNA, this product was excised from a gel, denatured in formamide, and analyzed by electrophoresis on a denaturing gel (data not shown). A labeled species migrating slightly more slowly than a 201-base marker was observed, the size expected for the labeled strand of the tagged circle.

The U3 and U5 Ends of the HIV LTR Are Functionally Distinct. The U5 end of the HIV LTR (LTR B) was reproducibly more active in assays of LTR cleavage and strand transfer than was the U3 end (LTR A). To characterize this difference further, the activities of the U3 and U5 ends were compared in a competition experiment. Equal amounts of each LTR DNA were added to reaction mixtures containing the 174-bp circular DNA target, and the products were analyzed by autoradiography after prolonged electrophoresis on a 38-cm native polyacrylamide gel. The signals corresponding to insertion of the U3 or U5 ends were distinguished by varying the lengths of the U3 and U5 substrates. To verify that the two signals could be adequately resolved in a single gel lane, the products of separate reaction mixtures containing a 32-bp U5 LTR and a 21-bp U3 LTR (Fig. 5, lane 3) or a 32-bp U3 LTR and a 21-bp U5 LTR (lane 4) were mixed after the reactions were stopped. The results indicate that the products can be readily distinguished. Lane 1 displays the results of a competition between the 32-bp U5 LTR and the 21-bp U3 LTR. To control for effects of LTR substrate length, the experiment was repeated with a 32-bp U3 LTR and a 21-bp U5 LTR (lane 2). The U5 signal predominates, regardless of the relative lengths of the two LTR substrates.

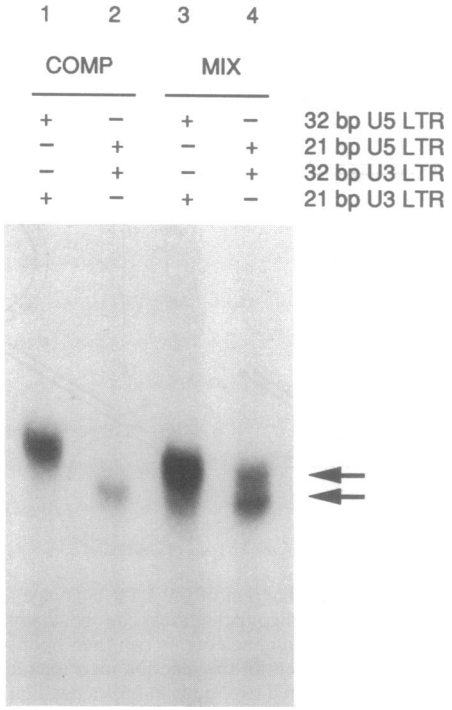

FIG. 5. Competition analysis of the relative activities of the U3 and U5 LTR ends. The activities of the U3 and U5 ends were compared by adding equal masses of a U3 and a U5 LTR substrate to an *in vitro* reaction containing the 174-bp circular DNA target. Lane 1, competition between a 32-bp U5 LTR (LTR K) and a 21-bp U3 LTR (LTR A); lane 2, competition between a 32-bp U3 LTR (LTR J) and a 21-bp U5 LTR (LTR B). COMP, competition. To establish that the tagged circular products containing the 32-bp and 21-bp LTR substrates could be distinguished, products of reactions containing different length LTRs were prepared separately and then pooled after the reactions were stopped. Lane 3, LTR K plus LTR A; lane 4, LTR J plus LTR B. MIX, mixed. Arrows mark the mobilities of the two species. The results of a competition experiment between LTR A and LTR J indicate that the longer U3 substrate is utilized more efficiently by IN protein (unpublished data), thus explaining the relative weakness of the signal from LTR A in lane 3. LTR K is identical to LTR B except for the addition of sequence 5′-GGATCCTAGGC-3′ and its complement to the right end as shown Fig. 1.

DISCUSSION

Activities of HIV Integration Protein *in Vitro*. Studies of the activities of the HIV IN protein *in vitro* establish that IN is the only retroviral protein required for cleavage of the HIV LTR and insertion of the cleaved LTR DNA into a target DNA. Because the HIV IN protein used here is only partially purified, it is not ruled out that contaminating cellular or baculoviral proteins play a stimulatory role in the reaction. However, HIV IN protein expressed in *Escherichia coli* and purified to near homogeneity by a different procedure shows similar specific cleavage and DNA strand transfer activities (A. Engelman, F.D.B., and R.C., unpublished data; for cleavage, see also ref. 17), reinforcing the view that the activities detected are intrinsic properties of the HIV IN protein.

Integration reaction mixtures containing HIV IN protein produced in insect cells and a circular DNA target yielded primarily products in which one LTR substrate molecule was inserted into the target rather than products in which pairs of LTR substrates were inserted in a concerted fashion as is expected of the reaction *in vivo*. In contrast, HIV core particles isolated from infected cells do integrate the two ends of the endogenously synthesized viral DNA in a concerted manner; single-ended events have not been reported (27, 28). The lack of concerted integration in the experiments presented here, however, need not be an intrinsic feature of the assay. In more efficient integration reaction mixtures that contain HIV IN protein produced in *E. coli*, the one-ended integration products still predominate, but products consistent with concerted integration of pairs of substrate oligonucleotides are also detectable (F.D.B., A. Engelman, and R.C., unpublished data). We have also reported that such concerted reaction products are produced in a similar reaction with MoMLV IN protein (12). The factors that promote pairwise integration remain to be investigated.

Because the integration assay presented here is so simple, requiring only HIV IN protein, labeled oligonucleotide substrates, and $MnCl_2$ in a simple buffer, it may be particularly suitable for rapid screening of candidate inhibitors of HIV integration. To assess inhibition, compounds to be tested could be added to integration reaction mixtures and the relative integration activity could be scored. Some such inhibitors of the integration step may block HIV growth and so serve as potential antiviral agents.

Mechanistic Studies of HIV Integration. The experiments presented here establish the chemical polarity of the insertion

of HIV DNA into target DNA. The sequencing of strand transfer products generated *in vitro* indicates that a 3′ end of the LTR is linked to a 5′ end in the DNA target. The 5′ end of the LTR remains unlinked in this intermediate (unpublished data). The 3′ LTR end that becomes linked to target DNA is inferred to have a 3′ OH and not a 3′ phosphate by (*i*) the electrophoretic mobility of the product of cleavage by IN protein and (*ii*) the observation that a "precleaved" substrate synthesized with a 3′ OH (LTR C) is efficiently inserted into target DNA. HIV thus appears to be a member of the large family of DNA insertion elements, including MoMLV, Ty*1*, phage Mu, and Tn*10*, that join terminal 3′ OH ends of the element to the 5′ ends of a cut made in the target DNA (8, 9, 29–31).

The finding that a precleaved substrate, LTR C, can be properly integrated establishes that the cleavage of the LTR ends is not energetically coupled to insertion of this DNA into target DNA. The energy of the phosphodiester bond broken in the LTR cleavage reaction cannot be stored and used for the subsequent strand transfer reaction. Because there is also no requirement for a high-energy cofactor, it seems probable that the energy for strand transfer derives from phosphodiester bonds broken in the target DNA during the reaction.

We find that sequences near the end of the LTR DNA are required for function of HIV IN protein, but the relation of these sequences to the proximal DNA end is not strictly fixed. The critical bases required for function of HIV IN protein apparently lie between positions 2 and 9, since LTR substrates containing base substitutions in this region are drastically impaired in LTR cleavage and DNA strand transfer. The importance of this region is also highlighted by the observation that sequences of the 9 terminal bp are more conserved between HIV isolates than the sequence of the LTR as a whole (32). The exact spacing between the proximal 3′ end of the LTR and base pairs 2–9 is not critical: addition or subtraction of a base pair at the proximal 3′ end does not eliminate cleavage at the proper A residue (position 3 in Fig. 1) or subsequent strand transfer. The MoMLV IN protein also requires sequences of the MoMLV LTR in the corresponding region for function. In addition to the internal sequences, function of MoMLV IN protein seems to require the presence of a DNA end roughly 2–5 bases from the A residue at the site of cleavage (11, 26, 33). A similar flexible requirement for a nearby DNA end may also apply for the HIV case.

The ends of the unintegrated HIV DNA are functionally distinct: the U5 end is more active in competition assays than the U3 end. The difference in function reflects the difference in sequence at the two ends, which match at only 8 of the 13 terminal bp. The U3 and U5 ends of MoMLV DNA, in contrast, match at all 13 terminal bp. The sequence of the U3 region of HIV may represent a compromise between competing requirements not present in MoMLV: in addition to its role in integration, the U3 region of HIV encodes part of the Nef protein (24, 25), whereas the U3 region of MoMLV does not encode any known proteins. The sequence of U3 of HIV may be constrained by both the integration and coding requirements and so may not be optimized for integration alone.

We thank Tania Baker, Janet Leatherwood Bushman, Alan Engelman, Martin Gellert, and Kiyoshi Mizuuchi for comments on the manuscript and Andrew Bates for Cre protein, pAB253, and advice on *loxP-cre* recombination. This work was supported by the National Institutes of Health Intramural AIDS Targeted Antiviral Program. F.D.B. is a Fellow of the Leukemia Society of America.

1. Donehower, L. A. & Varmus, H. E. (1984) *Proc. Natl. Acad. Sci. USA* **81**, 6461–6465.
2. Schwartzberg, P., Colicelli, J. & Goff, S. P. (1984) *Cell* **37**, 1043–1052.
3. Panganiban, A. T. & Temin, H. M. (1984) *Proc. Natl. Acad. Sci. USA* **81**, 7885–7889.
4. Hippenmeyer, P. J. & Grandgenett, D. P. (1984) *Virology* **137**, 358–370.
5. Clavel, F., Hoggan, M. D., Willey, R. L., Strebel, K., Martin, M. & Repaske, R. (1989) *J. Virol.* **63**, 1455–1459.
6. Grandgenett, D. P. & Mumm, S. R. (1990) *Cell* **60**, 3–4.
7. Varmus, H. E. & Brown, P. O. (1989) in *Mobile DNA*, eds. Berg, D. E. & Howe, M. M., pp. 53–108.
8. Fujiwara, T. & Mizuuchi, K. (1988) *Cell* **54**, 497–504.
9. Brown, P. O., Bowerman, B., Varmus, H. E. & Bishop, J. M. (1989) *Proc. Natl. Acad. Sci. USA* **86**, 2525–2529.
10. Lobel, L. I., Murphy, J. E. & Goff, S. P. (1989) *J. Virol.* **63**, 2629–2637.
11. Roth, M. J., Schwartzberg, P. L. & Goff, S. P. (1989) *Cell* **58**, 47–54.
12. Craigie, R., Fujiwara, T. & Bushman, F. D. (1990) *Cell* **62**, 829–837.
13. Katzman, M., Katz, R. A., Skalka, A. M. & Leis, J. (1989) *J. Virol.* **63**, 5319–5327.
14. Vora, A. C., Fitzgerald, M. L. & Grandgenett, D. P. (1990) *J. Virol.* **64**, 5656–5659.
15. Katz, R. A., Merkel, G., Kulkosky, J., Leis, J. & Skalka, A. M. (1990) *Cell* **63**, 87–95.
16. Bushman, F. D., Fujiwara, T. & Craigie, R. (1990) *Science* **249**, 1555–1558.
17. Sherman, P. A. & Fyfe, J. A. (1990) *Proc. Natl. Acad. Sci. USA* **87**, 5119–5123.
18. Bates, A. D. & Maxwell, A. (1989) *EMBO J.* **8**, 1861–1866.
19. Abremski, K. & Hoess, R. (1984) *J. Biol. Chem.* **259**, 1509–1514.
20. Maniatis, T., Fritsch, E. & Sambrook, J. (1982) *Molecular Cloning: A Laboratory Manual* (Cold Spring Harbor Lab., Cold Spring Harbor, NY).
21. Maxam, A. M. & Gilbert, W. (1977) *Proc. Natl. Acad. Sci. USA* **74**, 560–564.
22. Whitcomb, J. M., Kumar, R. & Hughes, S. H. (1990) *J. Virol.* **64**, 4903–4906.
23. Smith, J. S., Kim, S. & Roth, M. J. (1990) *J. Virol.* **64**, 6286–6290.
24. Ratner, L., Haseltine, W., Patarca, R., Livak, K. J., Starchich, B., Josephs, S. R., Doran, E. R., Rafalski, J. A., Whitwhoen, E. A., Baumeister, K., Ivanoff, L., Petteway, S. R., Pearson, M. L., Lautenberger, J. A., Papas, T. S., Ghrayeb, J., Chang, N. T., Gallo, R. C. & Wong-Staal, F. (1985) *Nature (London)* **313**, 277–284.
25. Wain-Hobson, S. P., Somigo, O., Danos, S., Cole, S. & Alizon, M. (1985) *Cell* **40**, 9–17.
26. Bushman, F. D. & Craigie, R. (1990) *J. Virol.* **64**, 5645–5648.
27. Ellison, V., Abrams, H., Roe, T. Y., Lifson, J. & Brown, P. (1990) *J. Virol.* **64**, 2711–2715.
28. Farnet, C. M. & Haseltine, W. A. (1990) *Proc. Natl. Acad. Sci. USA* **87**, 4164–4168.
29. Eichinger, D. J. & Boeke, J. D. (1990) *Genes Dev.* **4**, 324–330.
30. Craigie, R. & Mizuuchi, K. (1985) *Cell* **41**, 867–876.
31. Benjamin, H. W. & Kleckner, N. (1989) *Cell* **59**, 373–383.
32. Myers, G., Rabson, A. B., Josephs, S. F., Smith, T. F. & Wong-Staal, F. (1990) *Human Retroviruses and AIDS 1990* (Los Alamos Natl. Lab., Los Alamos, NM).
33. Colicelli, J. & Goff, S. P. (1988) *J. Mol. Biol.* **199**, 47–59.