Nucleic Acids Research, Vol. 19, No. 24 6691–6698

Site-specific hydrolysis and alcoholysis of human immunodeficiency virus DNA termini mediated by the viral integrase protein

Cornelis Vink, Esther Yeheskiely[1], Gijs A.van der Marel[1], Jacques H.van Boom[1] and Ronald H.A.Plasterk*

Division of Molecular Biology, The Netherlands Cancer Institute, Plesmanlaan 121, 1066 CX Amsterdam and [1]Gorlaeus Laboratories, PO Box 9502, 2300 RA Leiden, The Netherlands

Received November 13, 1991; Accepted November 27, 1991

ABSTRACT

Before integration of the human immunodeficiency virus (HIV) DNA, two nucleotides are removed from the 3′ ends of the viral DNA by the integrase (IN) protein. We studied the chemistry of this reaction, and found that IN mediates site-specific hydrolysis of a phosphodiester bond, resulting in release of a dinucleotide. A class of alcohols (including glycerol, 1,2-propanediol, but not 1,3-propanediol) can also act as nucleophile in this reaction, and likewise the alcoholic amino acids L-serine and L-threonine can be covalently linked to the dinucleotide. No evidence was found for a covalent linkage between the IN protein and this dinucleotide, suggesting that IN directs a single nucleophilic attack of water at the specific phosphodiester bond.

INTRODUCTION

Integration of a double-stranded DNA copy of the viral RNA genome into a chromosome of the infected cell is essential for retroviral replication. The integrated provirus serves as the template for the production of RNAs that function either as genomes of progeny virions or as messenger RNAs for the production of viral proteins. Assembled virions bud from the cell membrane and the retroviral replication cycle is completed by the infection of other cells.

One protein is known to be necessary for retroviral integration, the integrase (IN) protein (1–4). The IN protein is encoded by the 3′ part of the viral *pol* gene and is a proteolytic cleavage product of a *gag-pol* fusion protein precursor. The following steps are involved in the integration reaction. After reverse transcription of the RNA and synthesis of the second DNA strand, a linear, double-stranded DNA molecule with blunt ends is formed. This DNA contains at its termini direct repeat sequences, so-called long terminal repeats (LTRs). At the ends of the LTRs, short, imperfect inverted repeats are present. These sequences are required for retroviral integration (5, 6). The IN protein cleaves off the two terminal nucleotides at the 3′ ends of the flush-ended DNA. The resulting DNA molecule is the immediate precursor for integration (7–9). In the following step the viral 3′ hydroxyl ends are coupled to a site in the target DNA that is cleaved in a staggered fashion. The length of the staggered cut is virus specific (10), and probably measures 5 bp in the case of human immunodeficiency virus (HIV) integration. This was derived from the 5-bp direct repeats that flank integrated HIV proviruses (11, 12). Finally, the single-stranded DNA gaps that flank the provirus are repaired, presumably by host cell proteins. It is probably in this same step that the terminal nucleotides at the proviral 5′ ends are lost. The integrated provirus is now flanked by a short direct duplication of the target DNA and terminates with the dinucleotides 5′-TG and CA-3′.

Information about the properties of IN has come from the development of various *in vitro* DNA cleavage and integration assays. It was shown that (partially) purified IN proteins from HIV, Moloney murine leukemia virus (MoMLV), and avian sarcoma-leukosis virus (ASLV), can specifically cleave and integrate synthetic duplex oligonucleotides that resemble the termini of the retroviral DNA (2–4, 13–15). The sequence requirements for cleavage and integration by IN are limited (14–17). The presence of the conserved CA-3′ dinucleotide at the viral DNA ends is necessary, but not sufficient for cleavage (13, 15). Cleavage of the viral DNA termini by IN is site-specific; cutting can still be observed at the phosphodiester bond that is present 3′ of the CA-3′ dinucleotide when this bond is located one, three, four or five nucleotides away from the DNA terminus (15).

Although it is clear that most IN proteins cleave off two nucleotides from the 3′ end of the retroviral DNA, it has not been established whether this cleavage is carried out in a processive fashion, by removing one nucleotide at a time from the terminus, or by nicking the viral ends only once (immediately 3′ of the conserved dinucleotide CA-3′). Neither has it been investigated what the chemistry of the cleavage is. It has been found that, after the cleavage, the viral DNA terminates with a 3′ hydroxyl group (7–9). What happens to the 5′ ends of the

* To whom correspondence should be addressed

removed (oligo)nucleotides? One recent report strongly suggested that these 5′ ends are covalently linked to the IN protein as a necessary intermediate in the reaction (18). Here we show that the IN proteins of HIV-1 and HIV-2 cleave the HIV DNA ends by hydrolysis of the phosphodiester bond 3′ of the CA-3′ dinucleotide. Besides a specific oligonucleotide product that resulted from this hydrolysis, other products were found when glycerol or other 1,2-diols were included in the reaction mixture. We find that IN can use these alcohols as substrates in a site-specific alcoholysis of the HIV DNA ends. Two other alcohols, the amino acids L-serine and L-threonine, can also be used in the IN-directed alcoholysis, and thus become covalently linked to DNA. This may explain why a low level of covalent attachment of DNA to these amino acid residues in the ASLV IN protein was observed (18).

MATERIALS AND METHODS

Expression and purification of IN from HIV-1 and HIV-2

The expression and purification of the IN proteins of HIV-1 and HIV-2 have previously been described (15, 19). Reversible immobilization of the IN proteins to Thiopropyl Sepharose 6B (Pharmacia) will be described elsewhere.

Oligonucleotide substrates

The synthetic oligonucleotides that were tested for specific cleavage by IN are listed in Figure 1. To test whether these substrates could indeed be cleaved specifically by IN, we first radiolabeled the 5′ end of the strand that is supposed to be cleaved by IN. Twenty picomoles of oligonucleotide 3 in (Figure 1) were incubated with 80 μCi of [γ-^{32}P]ATP (3,000 Ci/mmol; Amersham corp.) and 0.2 U of T4 polynucleotide kinase (Boehringer Mannheim) in a reaction volume of 20 μl. After incubation for 1 hr at 37°C, the reaction mixture was incubated at 90°C for 3 min, and 6 μl of the mixture was added to either 2 μl (20 pmol) of oligonucleotide 7, 9 or 11 (Figure 1). The oligonucleotides were heated to 90°C, and annealed by slow cooling to room temperature. The recessed 3′ ends of the substrates were then filled-in by the Sequenase Version 2.0 system (United States Biochemical Corp.) in a reaction volume of 15 μl. After incubation for 45 min at room temperature, 10 μl of formamide loading dye (95% formamide, 20 Mm EDTA, 0.05% bromophenol blue, 0.05% xylene cyanol) was added. The sample was incubated at 80°C for 3 min and loaded onto a 12% denaturing polyacrylamide (38:2 acrylamide/bisacrylamide) gel. After electrophoresis an X-Ray film (X-OMAT AR; Eastman Kodak Co.) was exposed to the gel, to localize the labeled oligonucleotides. The oligonucleotides were then cut from the gel and eluted for 16 hr at 37°C in elution buffer (21). After precipitation with ethanol, the oligonucleotides were resuspended in H_2O to an end concentration of 0.2 pmol/μl. Then 1 μl (10 pmol) of the unlabeled complementary strands (Figure 1) was added to 10 μl of these oligonucleotide samples, and the strands were annealed as described above. 3′ End-labeling of the substrates was performed by first annealing oligonucleotide 3 with either oligonucleotide 7, 9 or 11, and thereafter filling-in the 3′ recessed ends as described above in the presence of [α-^{32}P] dCTP (3000 Ci/mmol; Amersham Corp.). The phosphate group immediately 3′ of the conserved CA-3′ dinucleotide was radiolabeled by annealing oligonucleotide 1 with oligonucleotide 4 or 9, and then filling-in the 3′ recessed ends in the presence of [α-^{32}P]dGTP (800 Ci/mmol; Amersham Corp.). The 3′ end-labeled substrates were purified and annealed in a similar way as the 5′ end-labeled substrates. Size markers were obtained by using a strand of 15 nucleotides that is identical to the 3′ terminal nucleotides of oligonucleotide 3 (oligonucleotide 2 in Figure 1). This oligonucleotide was annealed to either strand 7, 9 or 11, and thereafter 3′ end-labeled similarly as described above. In this way radiolabeled strands of respectively 16 (M1), 17 (M2) and 18 residues (M3) were obtained.

Cleavage reactions

Standard reactions (10 μl) contained 20 Mm MOPS Ph 7.2, 25 mM NaCl, 3 mM $MnCl_2$ or 5 mM $MgCl_2$, 3 mM DTT, 0.2 pmol of duplex oligonucleotide and approximately 50 ng of

```
                    25   20   15   10    5    1
                    |    |    |    |    |    |
HIV-1 U5    5′-TTAGTCAGTGTGGAAAATCTCTAGCAGT-3′

            1  -------------------------CA       (26)
            2                ----------CAGT     (15)
            3  -------------------------CAGT     (28)
            4  -------------------------GTCA     (28)
            3  -------------------------CAGT     (28)
            5  -------------------------GT       (26)
            6  -------------------------CAGTC    (29)
            7  -------------------------GTCAG    (29)
            8  -------------------------CAGTTC   (30)
            9  -------------------------GTCAAG   (30)
           10  -------------------------CAGTTTC  (31)
           11  -------------------------GTCAAAG  (31)
```

Figure 1. The oligonucleotide substrates that were used in this study. The sequence of the HIV-1 U5 terminus has previously been described (20). The 3′ end of the strand at the top of the figure represents the terminus of the viral DNA. The conserved CA-3′ dinucleotide is indicated in bold print. The numbers of the oligonucleotide strands are shown on the left, and the length of the strands is indicated by the numbers in parentheses. Residues at positions 3–26 in the oligonucleotides 1–3, 6, 8 and 10 are similar to the wild type HIV-1 U5 strand that is shown at the top. Oligonucleotides 4, 5, 7, 9 and 11 are complementary to this strand.

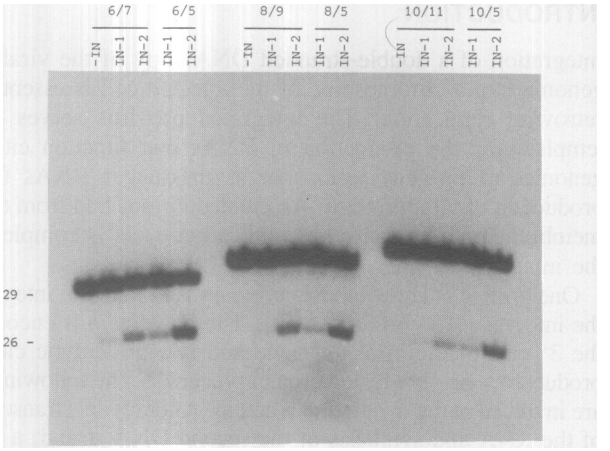

Figure 2. Site-specific cleavage of 5′ end-labeled substrates by HIV IN. Double-stranded oligonucleotides that were radiolabeled at the 5′ end of the strand that is cleaved by IN, were incubated in the absence (-IN) or presence of soluble HIV-1 (IN-1) or HIV-2 IN (IN-2) (see Materials and methods). The numbers of the substrates are indicated above the lanes, and are similar as in Figure 1. The lengths of the oligonucleotide strands are shown on the left. Specific cleavage of the oligonucleotides was enhanced when the sequence 3′ of the CA-3′ dinucleotide was present as single-stranded DNA. The same observation has previously been made with substrates that contain either the sequence GTT-3′, GTTT-3′ or GTTTT-3′ at the 3′ side of the CA-3′ sequence (15).

immobilized HIV-1 or 200 ng of HIV-2 IN protein. In addition, when soluble IN was used (approximately 50 ng for HIV-1 IN and 100 ng of HIV-2 IN), glycerol was present in the reaction mixture at a concentration of 21% (vol/vol) and NaCl at a concentration of 75 mM. The reactions were at 30°C for 1 h, and were stopped by addition of 10 μl of formamide loading dye. DNase I degradation patterns were obtained by incubation of the 3′ end-labeled substrates (0.2 pmol) for 15 min at 30°C in 50 mM Tris-HCl (pH 7.5), 10 mM $MgSO_4$, 0.1 mM DTT and 10 ng of DNase I, in an end volume of 10 μl. After the addition of 10 μl of formamide loading dye, the samples were incubated at 80°C for 3 min and 5 μl of the samples was loaded onto a 22% polyacrylamide, 8 M urea, 1×TBE (89 mM Tris, 89 mM boric acid, 2 mM EDTA) gel, and electrophoresed. Reaction products were visualized by autoradiography.

Synthesis and purification of the 1,2-ethanediol-oligonucleotide product

The solid-phase synthesis (22) of $hoch_2ch_2opGpTpTpTpC$-oh 3′ (Figure 7A) was carried out on an automated Gene assembler (Pharmacia) using a standard elongation protocol (23). Functionalized monobeads (controlled pore glass) were purchased from Pharmacia. N-acyl-5′-O-(4,4′-dimethoxytrityl)-2′-deoxynucleoside 3′-O-[(2-cyanoethyl)-phosphoramidites] were obtained from Synorchem. 1-O-(2-cyanoethoxy-N,N-diisopropyl-phosphoramidite)-2-O-(4,4′-dimethoxytrityl) ethane was prepared by reaction of mono-1-(4,4′-dimethoxytrityloxy)-ethan-2-ol with 2-cyanoethoxy-(N,N-diisopropylamino)-chlorophosphine (23). The fully protected oligonucleotide was deblocked and released from the solid-support by treatment with aqueous ammonia (25%). The thus obtained 4,4′-dimethoxytrityl protected compound was purified over a Pro RPC HR 5/10 column. The 4,4′-dimethoxytrityl group was removed with 80% aqueous acetic acid. The fully-deblocked $hoch_2ch_2opGpTpTpTpC$-oh 3′ was further purified and desalted on a Hiload superdex 75 HR 16/120 column.

Reactions with terminal transferase

First, cleavage reactions were performed with substrate 10/5 and immobilized HIV-2 IN, either in the presence or absence of 1,2-ethanediol. After the reactions, the mixtures were incubated at 60°C for 10 min. To 5 μl of the reaction mixtures was added:

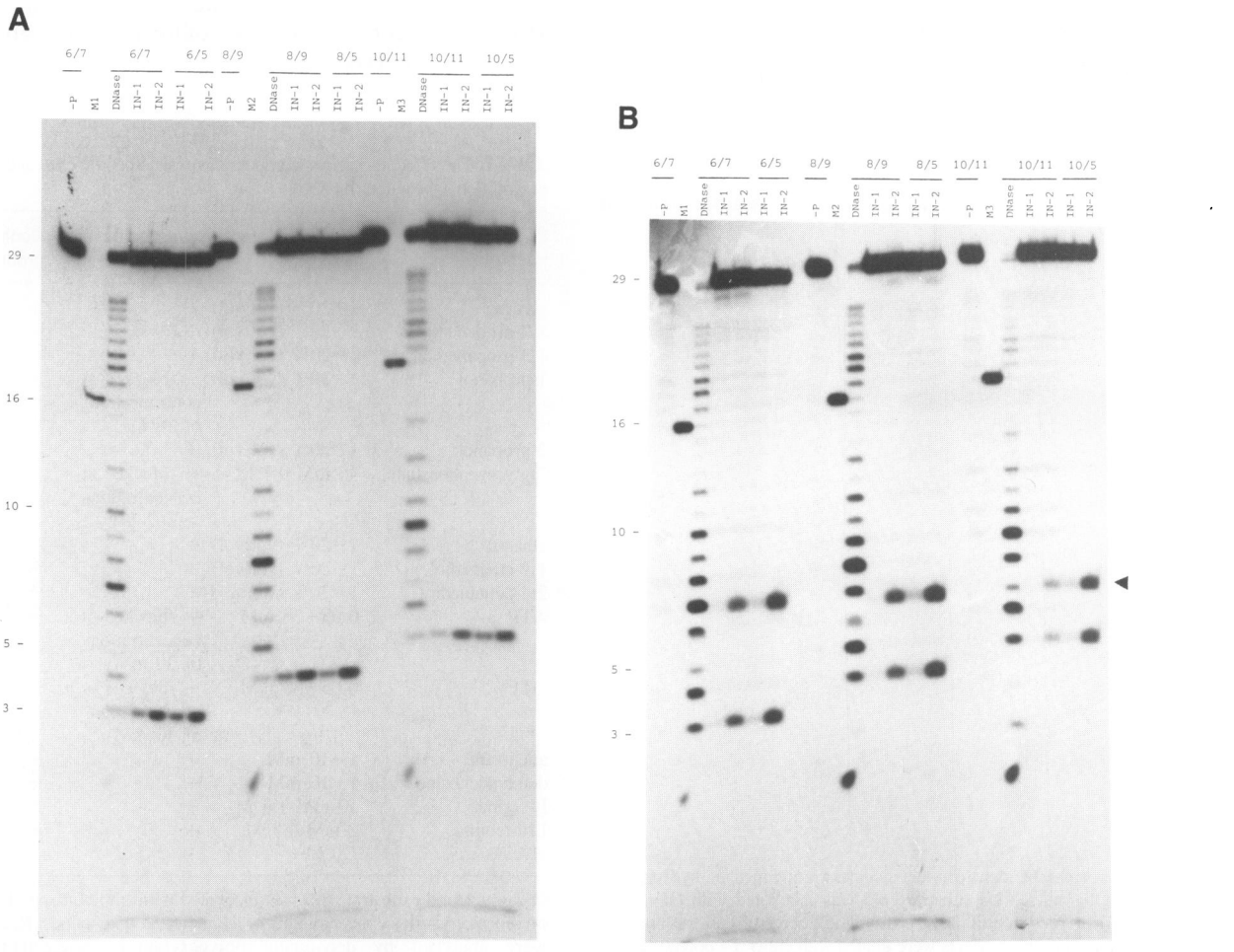

Figure 3. Site-specific cleavage of 3′ end-labeled substrates by HIV IN. The double-stranded oligonucleotides were radiolabeled at the 3′ ends of the strands that are cleaved by IN (see Materials and methods). **(A)** Site-specific cleavage of the oligonucleotides by immobilized HIV-1 (IN-1) or HIV-2 IN (IN-2), in the absence of glycerol. The numbers of the oligonucleotide substrates are depicted above the lanes (see Figure 1). The lengths of the oligonucleotides are shown on the left. The size markers (M1, M2 and M3), and the DNase I (DNase) degradation ladder were obtained as described in Materials and methods. −P, incubation of the substrates in the absence of protein. **(B)** Site-specific cleavage of the oligonucleotides by HIV-1 (IN-1) or HIV-2 IN (IN-2) in the presence of 20% (vol/vol) glycerol in the reaction mixture. The position of the glycerol-dependent products is indicated with a closed arrowhead.

10 μl of 1 M potassium cacodylate/1.25 mg/ml bovine serum albumin/125 mM Tris-HCl pH 6.6, 5 μl of 25 mM $CoCl_2$, 5 μl of 5 μM ddATP, 24 μl of H_2O and 1 μl of 25 U/μl terminal transferase (Boehringer Mannheim). After incubation for 1 h at 37°C, 35 μl of formamide-loading dye was added. Two μl of this mixture was diluted with 3 μl of formamide-loading dye, incubated for 3 min at 80°C, and loaded onto a denaturing polyacrylamide gel (see above).

The synthetic 1,2-ethanediol-oligonucleotide (Figure 7A) was 3′ end-labeled as follows. One pmol of the product was incubated with 0.2 M potassium cacodylate, 25 mg/ml bovine serum albumin, 2.5 mM $CoCl_2$, 10 μCi [α-^{32}P]ddATP (3000 Ci/mmol; Amersham Corp.), 25 mM Tris-HCl pH 6.6 and 25 U of terminal transferase in an end volume of 50 μl. After incubation for 1 h at 37°C, 35 μl of formamide-loading dye was added. One μl of a 35 fold dilution of this mixture was added to 4 μl of formamide-loading dye, incubated at 80°C for 3 min, and loaded onto a denaturing polyacrylamide gel.

RESULTS

Cleavage of 5′ and 3′ end-labeled substrates

The IN proteins of HIV-1 and HIV-2 can specifically remove two nucleotides from the 3′ termini of the retroviral flush-ended DNA (see Introduction). Removal of these two nucleotides could be carried out either in one step, by making one nick at the phosphodiester bond directly 3′ of the conserved CA-3′ dinucleotide, or in a processive fashion. To address the mechanism of cleavage by IN, we constructed double-stranded oligonucleotides that mimic the U5 end of HIV-1 DNA, and contain either 3, 4 or 5 nucleotides at the 3′ side of the CA-3′ dinucleotide (Figure 1). First, we determined whether these oligonucleotides were indeed substrates for specific cleavage by the HIV-1 and HIV-2 IN proteins. This was tested by introducing a radiolabeled phosphate group at the 5′ end of the oligonucleotide strands that are cleaved by IN. As shown in Figure 2, the 3 substrates were indeed specifically cleaved at the phosphodiester bond 3′ of the conserved CA-3′. We then tested the same oligonucleotide substrates, but now containing a radiolabeled phosphate group between the 2 terminal nucleotides 3′ of the C-A-3′ sequence, so that we could follow the nucleotides that are removed by IN (Figure 3A). The lengths of the reaction products were determined by comparison with products of a DNase I degradation ladder of the oligonucleotides. A specific product with a length of 3 nucleotides was generated when substrates 6/7 and 6/5 (Figure 1) were used. A specific product of 4 nucleotides was found with substrates 8/9 and 8/5, and of 5 nucleotides with substrates 10/11 and 10/5. The presence of a 5′ phosphate group on the oligonucleotide products was confirmed by their sensitivity to calf intestinal phosphatase and the presence of a free 3′ hydroxyl group was determined by the addition of nucleotides to the 3′ ends of the oligonucleotides by terminal transferase (see below). Our results show that site-specific cleavage of the viral DNA ends occurs by hydrolysis of the phosphate group directly 3′ of the conserved CA-3′.

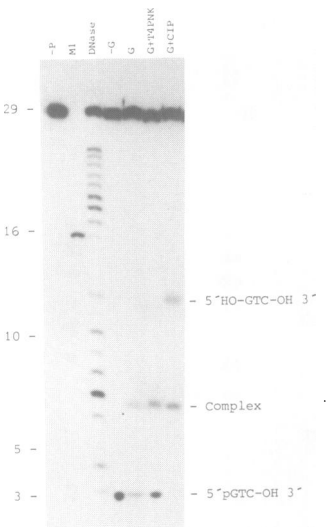

Figure 4. The glycerol-dependent product does not contain a 5′ hydroxyl or phosphate group. Oligonucleotide substrate 6/5 was incubated with HIV-2 IN in the absence (-G) or presence of 20% (vol/vol) glycerol (G) in the reaction mixture. After the reaction, the products were treated with either T4 polynucleotide kinase (+T4PNK), or calf intestinal phosphatase (+CIP) as previously described (21). The lengths of the oligonucleotides are shown on the left. −P, incubation of oligonucleotide 6/5 in the absence of protein; M1, size marker; DNase, DNase I degradation ladder of oligonucleotide 6/7; 5′ pGTC-OH 3′, the specific hydrolysis product; 5′ HO-GTC-OH, the specific hydrolysis product from which the 5′ phosphate group has been removed by calf intestinal phosphatase; Complex, glycerol-dependent product.

Table I. The effect of various compounds on site-specific cleavage and formation of additional products by IN.

compound	concentrations tested	site-specific cleavage[a]	alcohol-dependent product[b]
glycerol	1−20% (vol/vol)	+	+
1,2-propanediol	1−20% (vol/vol)	+	+
1,3-propanediol	1−20% (vol/vol)	+	−
1-propanol	1−20% (vol/vol)	+ (blocked at a concentration of 20%)	−
2-propanol	1−20% (vol/vol)	+	−
2-glycerophosphate	46 mM-0.7 M	+ (blocked at concentrations ≥ 0.2 M)	−
ethanol	1−20% (vol/vol)	+	−
1,2-ethanediol	1−20% (vol/vol)	+	+
2,4-pentanediol	1−20% (vol/vol)	+	−
ATP	0.05−30 mM	+ (blocked at a concentration of 30 mM)	−
GTP	0.05−30 mM	+ (blocked at a concentration of 30 mM)	−
adenosine	1−10 mM	+	−
deoxyadenosine	1−10 mM	+	−
L-serine	10 mM-1.3 M	+	+
L-threonine	2 mM-0.3 M	+	+
glycine	1.0 M	+	−

Oligonucleotide substrate 8/5 was incubated with immobilized HIV-2 IN in the presence of various concentrations of the indicated compounds. Reaction products were analyzed by denaturing polyacrylamide gelelectrophoresis and autoradiography, as described in Materials and methods.

[a]Site-specific cleavage was scored (+) when the product of site-specific hydrolysis by IN could clearly be observed on autoradiographs.

[b]Formation of an alcohol-dependent product was scored (+) when an additional band (besides the band of the product of site-specific hydrolysis) could be seen on autoradiographs, of which the formation depended on the concentration of the compound tested.

Site-specific cleavage in the presence of glycerol

The reactions described above were carried out in the absence of glycerol. Many of the cleavage reactions described in the literature, however, were performed in the presence of glycerol (2, 13, 15, 19). To investigate its effect on site-specific cleavage by IN, we performed the same reactions as described above, but now in the presence of glycerol (Figure 3B). All substrates tested were cleaved in a site-specific manner. Besides the specific band, another prominent band was seen with each of the 3 substrates. The additional product that was generated when substrates 8/9 and 8/5 were used, did not comigrate with a band from the DNase I degradation ladder of the oligonucleotides, and therefore does probably not represent an oligonucleotide that contains a phosphate group at its 5′ end and a hydroxyl group at its 3′ end. It is not possible that the additional products are generated by nonspecific cleavage of the substrates, because additional bands are not observed when 5′ end-labeled substrates are used (Figure 2). We found that the generation of the extra products was dependent on the presence of glycerol in the reaction mixture (compare Figure 3A and 3B). This suggests either that glycerol stimulates a side reaction, or that glycerol is one of the substrates in the reaction and is covalently coupled to a radiolabeled part of the oligonucleotide.

Properties of the glycerol-dependent products

To investigate the properties of the glycerol-dependent products, we performed a series of enzymatic reactions on the cleavage reaction products. We found that treatment with Proteinase K and subsequent phenol extraction had no effect on the mobility of the glycerol-dependent product (data not shown). This result suggests that the radiolabeled product does not contain (part of) a protein. Treatment of the reaction mixture with T4 polynucleotide kinase and calf intestinal phosphatase had no effect on the mobility of the glycerol-dependent product, whereas the hydrolysis product shifted to a higher position on the gel upon phosphatase treatment (Figure 4). These findings show that the oligonucleotide part of the product does not contain a free 5′ hydroxyl or phosphate group. The product does, however, contain a free 3′ hydroxyl group. This was shown by the addition of nucleotides to its 3′ end by terminal transferase (see below). We conclude that the product is non-proteinaceous, does not have a free 5′ phosphate nor a 5′ hydroxyl group, but does have a free 3′ hydroxyl group.

Formation of products with compounds other than glycerol

As stated above, the glycerol-dependent products could have been generated in two ways. First, the products are formed by a covalent linkage of oligonucleotides to glycerol. Second, they are generated by a glycerol-stimulated side reaction. To address these two possibilities, we tested whether the presence of glycerol-related alcohols in the reaction mixture would also lead to the formation of additional products. If these products are indeed formed and have the same mobility on gel as the glycerol-dependent product, a side reaction is probably stimulated by glycerol and glycerol-related compounds. If, however, products

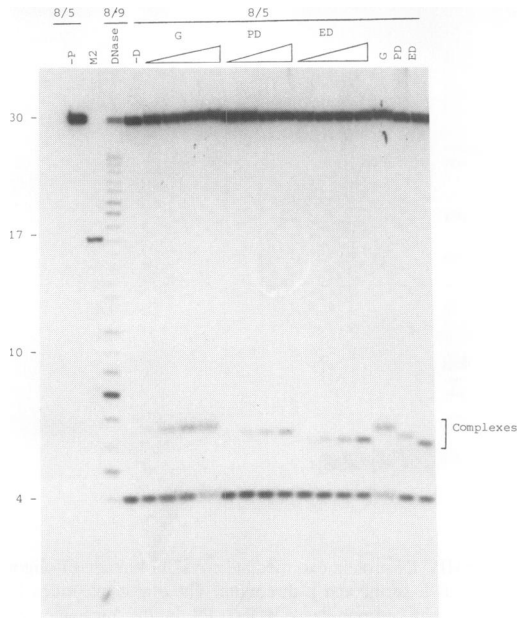

Figure 5. Site-specific cleavage by IN in the presence of 1,2-diols. Oligonucleotide substrate 8/5 was incubated with immobilized HIV-2 IN in the absence (-D) or presence of increasing concentrations of the alcohols glycerol (G), 1,2-propanediol (PD) or 1,2-ethanediol (ED) in the reaction mixture. The alcohols were present at a concentration of 1, 5, 10 and 20% (vol/vol). The reactions mixtures that contained 20% of the alcohols were also loaded on the outermost right of the gel. −P, incubation of oligonucleotide 8/5 in the absence of protein; M2, size marker; DNase, Dnase I degradation pattern of oligonucleotide 8/9. The lengths of the oligonucleotides are shown on the left. The position of the alcohol-dependent products is depicted on the right (Complexes).

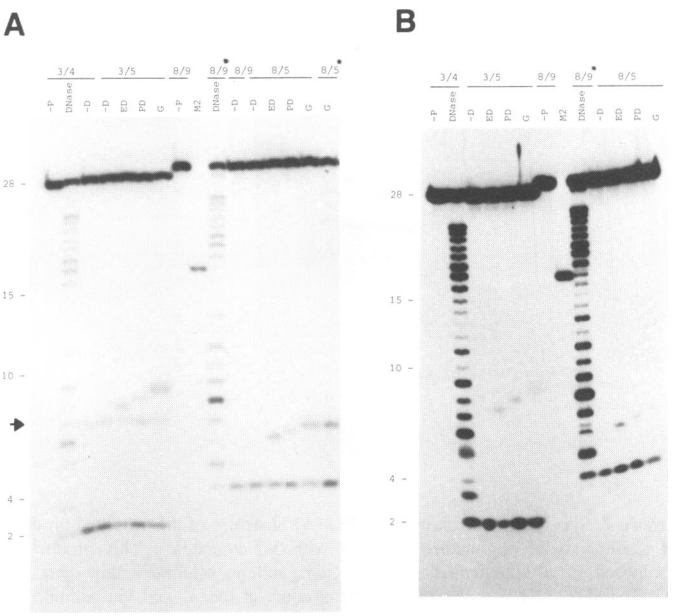

Figure 6. Site-specific cleavage of oligonucleotides that have a radiolabeled phosphorous atom in the phosphodiester bond that is cleaved by IN. We radiolabeled the phosphorous atom in the phosphodiester bond that is specifically cleaved by IN in oligonucleotide substrates 3/4, 3/5, 8/9 and 8/5. These substrates were incubated with immobilized HIV-2 IN in the presence of Mn^{2+} **(A)** or Mg^{2+} **(B)**. The numbers of the substrates are shown above the lanes. The substrates that are depicted with an asterisk have a radiolabeled phosphorous atom in the 3′ outermost phosphodiester bond, and are the same as the substrates that were tested in Figure 4. Lengths of oligonucleotides are shown on the left. The arrow in **(A)** points to a product of which the formation was dependent on the presence of Mn^{2+} in the reaction mixture (see text). −P, incubation in the absence of protein; M2, size marker; −D, incubation in the absence of 1,2-diols; ED, PD and G, incubation in the presence of 20% (vol/vol) 1,2-ethanediol, 1,2-propanediol and glycerol, respectively. DNase, DNase I degradation ladder.

with a different mobility are generated, it is likely that they have been generated by a covalent coupling of oligonucleotides to the glycerol-related compounds. The effect of several glycerol-related alcohols is shown in Table I. Glycerol, 1,2-propanediol and 1,2-ethanediol (ethylene glycol) gave rise to the formation of an additional product. Increasing concentrations of these substrates in the reaction mixture resulted in an increase in the amount of product that was formed (Figure 5). Additional products were not generated when the 1,3-diols 1,3-propanediol or 2,4-pentanediol were used, suggesting that two hydroxyl groups on adjacent carbon atoms are required for product formation. The products that were generated in the presence of glycerol, 1,2-propanediol and 1,2-ethanediol have different mobilities (Figure 5). This suggests that these alcohols do not stimulate a side reaction, but instead are involved in a IN-dependent reaction in which the specifically removed oligonucleotide is directly transferred to the 1,2-diols. It is likely that the glycerol-dependent product is composed of an oligonucleotide that is covalently coupled by its 5′ phosphate group to one of the hydroxyl groups of glycerol. The complex can be generated by an IN-dependent nucleophilic attack of one of the oxygen atoms of glycerol at the phosphate group 3′ of the CA-3′ sequence in the oligonucleotide substrate. In this way the 1,2-diols would replace water as the nucleophile, and besides site-specific hydrolysis, HIV IN would also be able to mediate site-specific alcoholysis of the viral DNA termini.

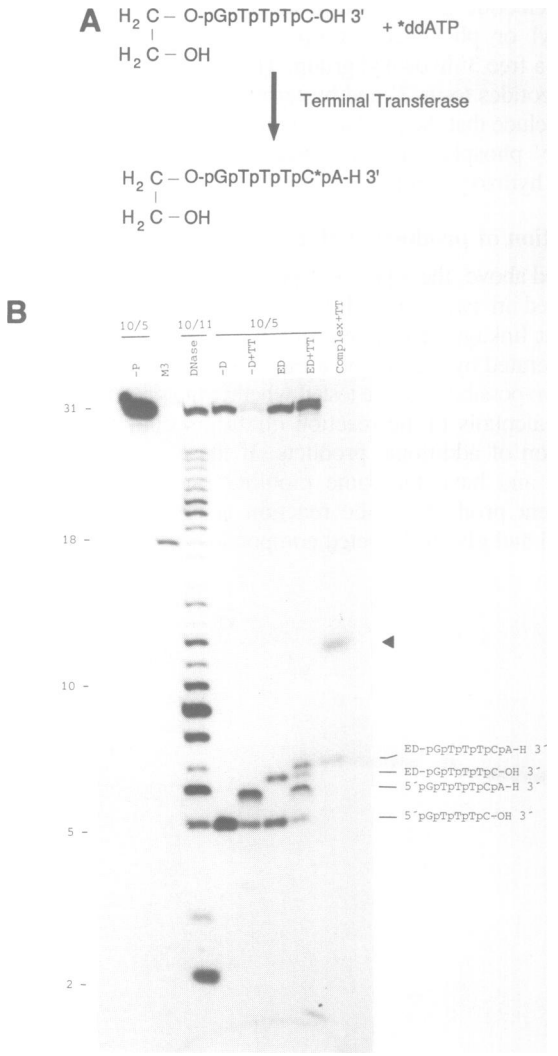

Figure 7. Site-specific alcoholysis by IN. **(A)** Structure of the proposed product of alcoholysis of oligonucleotide 10/5 with 1,2-ethanediol. This product is composed of a 1,2-ethanediol-derived part and an oligonucleotide part (5′ pGpTpTpTpTpC-OH 3′). We labeled the product at the 3′ end via addition of [α-^{32}P]ddATP by terminal transferase (see Materials and methods). **(B)** Comparison of the synthetic product from **(A)** with the product formed by incubation of oligonucleotide 10/5 with immobilized HIV-2 IN in the presence of 20% (vol/vol) of 1,2-ethanediol. After cleavage of oligonucleotide 10/5 in the absence (-D) or presence of 1,2-ethanediol (ED), one half of each reaction was treated with terminal transferase and ddATP (-D+TT and ED+TT, respectively), and subsequently loaded on gel next to the half of each reaction that was not treated with terminal transferase. The synthetic 1,2-ethanediol-product was labeled as shown in **(A)**, and run along on gel with the different reaction mixtures (Complex+TT). The structure of the various products is shown on the right. The product ED-pGpTpTpTpC-OH 3′ is probably identical to the product that is drawn at the top in **(A)** The lengths of the oligonucleotide strands are depicted on the left. −P, incubation of substrate 10/5 in the absence of protein; M3, size marker; DNase, DNase I degradation ladder. The closed arrowhead points to a fuzzy band high up in the gel, which is probably a side product of the chemical synthesis, and was not investigated further.

Site-specific alcoholysis by HIV IN

If site-specific alcoholysis of the viral DNA can indeed occur, then the phosphorous atom in the phosphodiester bond that is

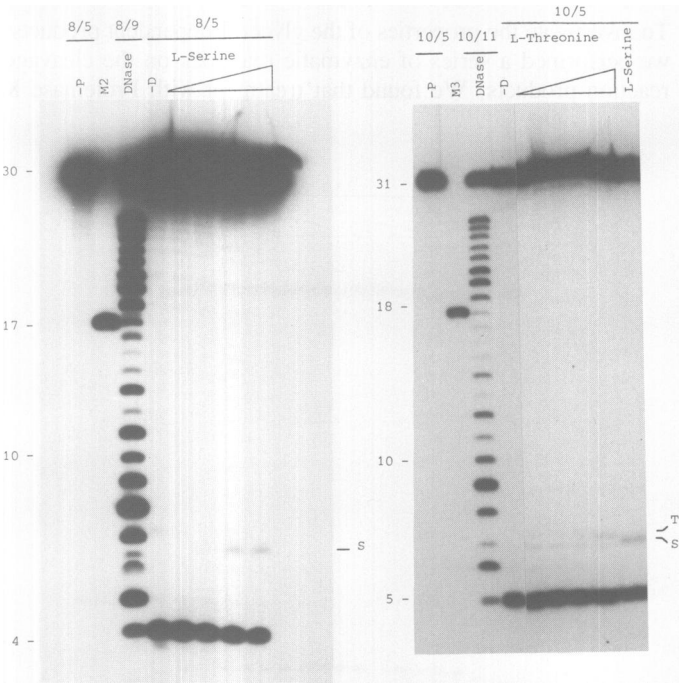

Figure 8. The HIV IN protein can mediate covalent linkage of oligonucleotides to the amino acids L-serine and L-threonine. The oligonucleotides 10/5 and 8/5 were incubated with immobilized HIV-2 IN in the presence of increasing concentrations of the amino acids L-serine or L-threonine. The concentrations used were 40, 80, 160, and 240 mM for L-serine (the panel on the left) and 3, 15, 67 and 335 mM for L-threonine (the panel on the right). When oligonucleotide 10/5 was cleaved in the presence of L-serine (the panel on the right), the amino acid was present at a concentration of 1,3 M. The numbers of the substrates that were tested are shown above the lanes. The alcohol-dependent products that are formed are indicated on the right of each panel with 'S' (for serine-product) or 'T' (for threonine-product). Lengths of oligonucleotides are shown on the left. −P, incubation in the absence of protein; M2 and M3, size markers; DNase, DNase I degradation pattern; −D, incubation of the oligonucleotides with immobilized HIV-2 IN in a standard reaction, in the absence of additional compounds.

cleaved should also be present in the alcohol-dependent products. To investigate this, we radiolabeled this phosphorous atom in four oligonucleotide substrates: oligonucleotides 3/4, 3/5, 8/9 and 8/5 (Figure 1). Besides the product of site-specific hydrolysis by IN, an additional product was found when either of the 3 1,2-diols were included in the reaction mixture (Figure 6). The products that were formed have the same mobility as the alcohol-dependent products that were generated in previous experiments (compare Figure 5 and 6). Apparently, the phosphorous atom in the phosphodiester bond that is broken, becomes a covalent part of the product.

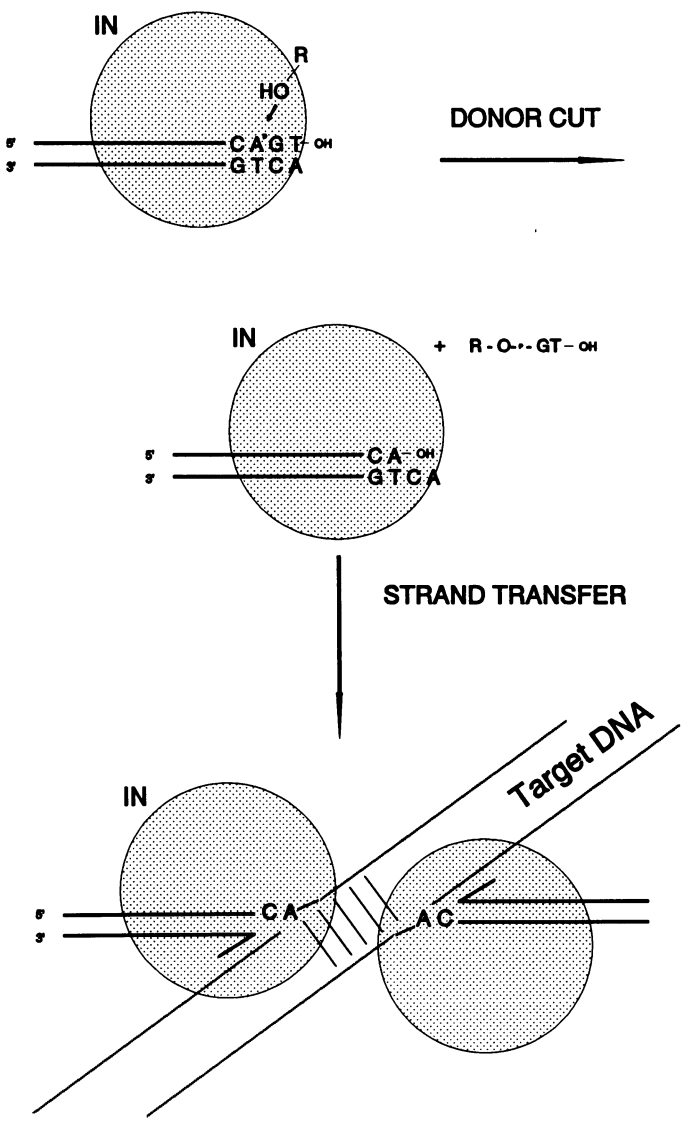

Figure 9. Model for the donor cut and the strand transfer reaction in retroviral DNA integration. First, two nucleotides at the 3′ ends of the viral DNA (indicated by black lines) are removed (the donor cut), either via site-specific hydrolysis (in which case 'R-OH' is a water molecule) or site-specific alcoholysis (in which case 'R-OH' is a vicinal diol) mediated by IN. The integrase protein ('IN') is depicted as a stippled circle. In the strand transfer step the recessed 3′ ends of the viral DNA are coupled via a direct nucleophilic attack to phosphodiester bonds in both strands of the target DNA (indicated by interrupted lines). These phosphodiester bonds are located five nucleotides away from each other. Pairing between the bases that are present in between the two new phosphodiester bonds is broken, and the remaining gaps are repaired, probably by cellular enzymes. Note that the two reactions are here shown to be carried out by the same active site on the same IN molecule (see text).

One other prominent product was formed when oligonucleotides 3/5 or 3/4 were tested in the presence of Mn^{2+} in the reaction mixture (see arrow in Figure 6A). When we substituted Mg^{2+} for Mn^{2+}, the product was not formed (compare Figure 6A and 6B). Additional Mn^{2+}-dependent products were not generated with any of the other oligonucleotide substrates (Figure 6 and data not shown). The Mn^{2+}-dependent product has not been investigated further.

The proposed structure of the product that is formed between 1,2-ethanediol and oligonucleotide 10/5 is drawn in Figure 7A. To have a reference, we chemically synthesized this oligonucleotide derivative, and radiolabeled it at its 3′ hydroxyl terminus via addition of [α-^{32}P]ddATP by terminal transferase (see Materials and methods). The radiolabeled product was run on a gel along with the cleavage products that were formed after incubation of oligonucleotide 10/5 with IN, in the presence of 1,2-ethanediol. These products were treated with terminal transferase in the presence of unlabeled ddATP, to obtain a product that has the same structure as the synthetic product. As shown in Figure 7B, the synthetic product has the same mobility on gel as the 1,2-ethanediol-specific cleavage product. When electrophoresis of the products was carried out in an 18% polyacrylamide gel instead of a 22% gel, both the IN product and the synthetic reference oligonucleotide-derivative again had the same mobility on gel, but shifted to a different position with respect to the DNase I ladder (data not shown). These results support that the structure drawn in Figure 7A is the product of IN-mediated site-specific alcoholysis of the HIV DNA termini.

Formation of complexes with L-serine and L-threonine

It has previously been shown that some enzymes that catalyze hydrolytic reactions, can also use alcohols as substrates (24). The enzyme alkaline phosphatase from E. coli for example, can transfer phosphate groups to acceptors that contain two hydroxyl groups or one hydroxyl and one amino group (24). To investigate whether HIV IN can also use hydroxy compounds that contain an amino group instead of a second hydroxyl group on an adjacent carbon atom, we tested the amino acids L-serine and L-threonine as substrates. As shown in Table I and Figure 8 these amino acids are indeed substrates for site-specific alcoholysis by IN.

It has been reported that a low level of covalent attachment of viral DNA sequences to serine and threonine residues in the ASLV IN protein occurs (18). We tested this for the IN proteins from HIV-1 and HIV-2. The products of a cleavage reaction were analyzed by SDS-polyacrylamide gelelectrophoresis and (prolonged) autoradiography, but we were not able to detect a radiolabeled band at or around the position of the IN proteins in the gel. The presence or absence of β-mercaptoethanol in the loading-buffer (25), or omission of the heating step prior to loading on gel, did not have an effect (data not shown).

DISCUSSION

The IN proteins of HIV-1 and HIV-2 can specifically cleave off 2 nucleotides from the 3′ termini of the viral DNA. We show here, that this cleavage reaction involves hydrolysis of a single phosphodiester bond 3′ of the conserved CA-3′ dinucleotide. Besides hydrolysis, we found another activity of the IN proteins: site-specific alcoholysis of the viral DNA termini. In this reaction, a number of hydroxy compounds that contain either a second hydroxyl group or an amino group on adjacent carbon atoms can be used as substrates. The 1,2-diols 1,2-ethanediol,

1,2-propanediol and glycerol, and the amino acids L-serine and L-threonine were substrates for IN. We found that 1,3-diols were not used efficiently as substrates (Table I). An explanation for this might be that 1,2-diols are better nucleophiles than 1,3-diols, due to the closer proximity of both hydroxyl groups in 1,2-diols than in 1,3-diols. Considering the concentration of alcohols that is required to get IN-mediated alcoholysis, we do not think the process plays a role *in vivo*. Alcoholysis has previously been described for several enzymes that catalyze hydrolytic reactions (24). For instance, the enzyme alkaline phosphatase from E. coli catalyzes the transfer of phoshate groups from substrates like p-nitrophenyl phosphate to hydroxy compounds containing a second hydroxyl group or amino group at a distance not greater than 1,3 (24).

We here show that IN has in fact two substrates in the donor cut reaction: viral DNA and water or some alcohols. It is possible that IN acts by making the phosphodiester bond 3′ of the CA-3′ dinucleotide accessible to any nucleophile in the solution. It is however also possible that there is some specificity in the recognition of the alcohol substrate by IN, and in that case a detailed study of this recognition may serve to define inhibitors of IN. Elucidation of the three-dimensional structure of the IN protein plus its substrate will help to clarify this.

Several other recombination, nuclease and transposition mechanisms have been studied *in vitro*, and in several cases glycerol has been used as a standard component in the reaction (for example see references 2, 13 and 26). It might be interesting to check whether site-specific alcoholysis of DNA can also be mediated by other proteins.

A one-step transesterification mechanism has recently been proposed for the joining of phage Mu DNA to the target DNA (the strand transfer reaction) (26). This strongly suggests that there is no covalent transposase-DNA intermediate involved in transposition of phage Mu. A similar mechanism of transesterification has been suggested for the strand transfer reaction that is mediated by HIV-1 IN. This was based on an analysis of the stereochemical course of this reaction (27). This raises the question how a single transesterification step can be reconciled with the finding that oligonucleotides of various lengths that are removed from the 3′ end of the viral DNA are transferred to serine, and to a lesser extent to threonine residues of the ASLV IN protein (18). Based on this observation, Katzman *et al.* proposed a model in which a covalent IN-DNA intermediate is involved in the integration reaction (18). A mechanism that involves such an intermediate, however, would require at least two transesterification steps. We show here, that the transfer of oligonucleotides to the IN protein is not in contradiction with a one-step transesterification mechanism, since the free amino acids L-serine and L-threonine can also be used as acceptors for IN-mediated transfer of DNA. Therefore, the finding of a covalent linkage of oligonucleotides to the ASLV IN protein cannot be taken as evidence for the involvement of a covalent intermediate in the integration reaction. Some amino acid residues in the ASLV IN protein might be used as acceptors of DNA in the alcoholysis of the viral DNA, but these residues do not need to be specific residues that are important for the activity of ASLV IN. We have been unable to detect any transfer of oligonucleotides to the HIV-1 and −2 IN proteins.

We propose that the IN protein cuts the viral (donor) DNA, by making one specific phosphodiester bond available for direct attack by nucleophiles, usually water. The simplest model for the subsequent integration reaction is, that the IN molecule remains bound to the cleaved HIV DNA end, and directs a second transesterification event where the 3′ OH of the viral DNA terminus is ligated to a 5′ phosphate in the target DNA (Figure 9). This model implies that in chemical terms the strand transfer of precut HIV DNA to the target DNA can be the reversal of the donor DNA cleavage reaction: a substitution of the hydrogen atom on the 3′ OH of the viral DNA by a phosphodiester group. It is therefore conceivable that both IN mediated reactions (donor cutting and integration) are directed by a single active site on the protein. The presence of a single active site is further strongly suggested by the identical substrate DNA requirements of HIV IN for both the donor cut and the strand transfer reaction (13, 28).

ACKNOWLEDGEMENTS

We would like to thank Hidde Ploegh, Chris Vos and Dik van Gent for critical reading of the manuscript. We thank Hans van den Elst for technical assistence. This research was supported by grant 900-502-119 from the Dutch Organization for Scientific Research, by grant 88033 from the Dutch Ministry of Health, and by a PIONIER grant from the Dutch Organization for Scientific Reseach to R.H.A.P.

REFERENCES

1. Bushman,F.D., Fujiwara,T. and Craigie,R. (1990) *Science*, **249**, 1555–1558.
2. Craigie,R., Fujiwara,T. and Bushman,F. (1990) *Cell*, **62**, 829–837.
3. Katz,R.A., Merkel,G., Kulkosky,J., Leis,J. and Skalka,A.-M. (1990) *Cell*, **63**, 87–95.
4. Sherman,P.A. and Fyfe,J.A. (1990) *Proc. Natl. Acad. Sci. U.S.A.*, **87**, 5119–5123.
5. Colicelli,J. and Goff,S.P. (1985) *Cell*, **42**, 573–580.
6. Colicelli,J. and Goff,S.P. (1988) *J. Mol. Biol.*, **199**, 47–59.
7. Brown,P.O., Bowerman,B., Varmus,H.E. and Bishop,J.M. (1989) *Proc. Natl. Acad. Sci. U.S.A.*, **86**, 2525–2529.
8. Fujiwara,T. and Mizuuchi,K. (1988) *Cell*, **54**, 497–504.
9. Roth,M.J., Schwartzberg,P.L. and Goff,S.P. (1989) *Cell*, **58**, 47–54.
10. Varmus,H.E. and Brown,P.O. (1989) In Berg,D.E. and Howe,M.M. (eds), *Mobile DNA*. American Society for Microbiology, Washington, DC.
11. Vincent,K.A, York-Higgins,D., Quiroga,M. and Brown,P.O. (1990) *Nucleic Acids Res.*, **18**, 6045–6047.
12. Vink,C., Groenink,M., Elgersma,Y., Fouchier,R.A.M., Tersmette,M. and Plasterk,R.H.A. (1990) *J. Virol.*, **64**, 5626–5627.
13. Bushman,F.D. and Craigie,R (1991) *Proc. Natl. Acad. Sci. U.S.A.*, **88**, 1339–1343.
14. Katzman,M., Katz,R.A., Skalka,A.-M. and Leis,J. (1989) *J. Virol.*, **63**, 5319–5327.
15. Vink,C., van Gent,D.C., Elgersma,Y. and Plasterk,R.H.A. (1991) *J. Virol.*, **65**, 4636–4644.
16. Bushman,F.D. and Craigie,R. (1990) *J. Virol.*, **64**, 5645–5648.
17. Vink,C., van Gent,D.C. and Plasterk,R.H.A. (1990) *J. Virol.*, **64**, 5219–5222.
18. Katzman,M., Mack,J.P.G., Skalka,A.-M. and Leis,J. (1991) *Proc. Natl. Acad. Sci. U.S.A.*, **88**, 4695–4699.
19. van Gent,D.C., Elgersma,Y., Bolk,M.W.J., Vink,C. and Plasterk,R.H.A. (1991) *Nucleic Acids Res.*, **19**, 3821–3827.
20. Wain-Hobson,S., Sonigo,P., Danos,O., Cole,S. and Alizon,M. (1985) *Cell*, **40**, 9–17.
21. Sambrook,J., Fritsch,E.F. and Maniatis,T. (1989) Molecular Cloning: A Laboratory Manual. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
22. Roelen,H.C.P.F., Kamer,P.C.J., van den Elst,H., van der Marel,G.A. and van Boom,J.H. (1991) *Recl. Trav. Chim. Pays-Bas*, **110**, 325–331.
23. Sinha,N.D., Biernat,J., McManus,J. and Koster,H. (1984) *Nucleic Acids Res.*, **12**, 4539–4557.
24. Wilson,I.B., Dayan,J. and Cyr,K. (1964) *J. Biol. Chem.*, **239**, 4182–4185.
25. Laemmli,U.K. (1970) *Nature*, **227**, 680–685.
26. Mizuuchi,K. and Adzuma,K. (1991) *Cell*, **66**, 129–140.
27. Engelman,A., Craigie,R. and Mizuuchi,K. (unpublished).
28. LaFemina,R.L., Callahan,P.L. and Cordingley,M.G. (1991) *J. Virol.*, **65**, 5624–5630.