

Human Immunodeficiency Virus Integrase Protein Requires a Subterminal Position of Its Viral DNA Recognition Sequence for Efficient Cleavage

CORNELIS VINK, DIK C. VAN GENT, YPE ELGERSMA, AND RONALD H. A. PLASTERK*

*Division of Molecular Biology, The Netherlands Cancer Institute,
Plesmanlaan 121, 1066 CX Amsterdam, The Netherlands*

Received 28 November 1990/Accepted 24 May 1991

Retroviral integration requires *cis*-acting sequences at the termini of linear double-stranded viral DNA and a product of the retroviral *pol* gene, the integrase protein (IN). IN is required and sufficient for generation of recessed 3' termini of the viral DNA (the first step in proviral integration) and for integration of the recessed DNA species *in vitro*. Human immunodeficiency virus type 1 (HIV-1) IN, expressed in *Escherichia coli*, was purified to near homogeneity. The substrate sequence requirements for specific cleavage and integration of retroviral DNA were studied in a physical assay, using purified IN and short duplex oligonucleotides that correspond to the termini of HIV DNA. A few point mutations around the IN cleavage site substantially reduced cleavage; most other mutations did not have a drastic effect, suggesting that the sequence requirements are limited. The terminal 15 bp of the retroviral DNA were demonstrated to be sufficient for recognition by IN. Efficient specific cutting of the retroviral DNA by IN required that the cleavage site, the phosphodiester bond at the 3' side of a conserved CA-3' dinucleotide, be located two nucleotides away from the end of the viral DNA; however, low-efficiency cutting was observed when the cleavage site was located one, three, four, or five nucleotides away from the terminus of the double-stranded viral DNA. Increased cleavage by IN was detected when the nucleotides 3' of the CA-3' dinucleotide were present as single-stranded DNA. IN was found to have a strong preference for promoting integration into double-stranded rather than single-stranded DNA.

Efficient retroviral replication requires integration of a double-stranded DNA copy of the viral RNA genome into a chromosome of the infected cell. Transcription of the integrated provirus produces viral RNAs that function as genomes of progeny virions or as the template for translation of viral proteins. Assembled virions bud from the cell membrane, and infection of other cells completes the retroviral replication cycle.

Studies of integration in the Moloney murine leukemia virus (MoMLV), human immunodeficiency virus (HIV), and avian sarcoma-leukosis virus (ASLV) systems have indicated that the following steps are involved in the integration process. After reverse transcription and second-strand DNA synthesis, a linear, double-stranded DNA molecule with flush ends is formed. This DNA copy of the viral RNA contains at its termini direct repeat sequences, termed long terminal repeats (LTRs). At the outer edges of the LTRs, short, imperfect inverted repeats are present. These sequences are necessary and sufficient for correct proviral integration (6, 7). The two terminal nucleotides at the 3' ends of the blunt-ended DNA are cleaved off, resulting in the putative immediate precursor for integration (2, 15, 26). This cleavage reaction is executed by the viral integrase protein (IN), which is the only protein known to be required for integration (3, 5, 8, 18, 19, 30). In the next step, the target DNA is cut in a staggered fashion and joined to the viral 3' OH ends (3, 8, 15, 18). The length of the staggered cut is virus specific (35). In the case of HIV integration, presumably a 5-bp staggered cut is generated, which is inferred from the characteristic 5-bp direct repeats in the target DNA that flank integrated HIV proviruses (10, 21, 36). The last step in

proviral integration is repair of the single-stranded DNA gaps present in the integration intermediate. It is possibly in this step that the terminal nucleotides at the proviral 5' ends are lost. The resulting provirus is now flanked by a short direct duplication of the target DNA and terminates with the dinucleotides 5'-TG and CA-3'. These same dinucleotides are also present at the termini of some eukaryotic and prokaryotic transposable elements and in all retrotransposons.

The HIV-1 IN protein is encoded near the 3' end of the viral *pol* gene and is a 32-kDa proteolytic cleavage product of a *gag-pol* fusion protein precursor. Mutations in the IN-coding region have no influence on reverse transcription, but integration of the viral DNA can be blocked (9, 28, 34). An understanding of the role of the IN protein in the integration process has come from the development of different *in vitro* integration assays. First it was shown that extracts from retrovirus-infected cells were able to integrate endogenous DNA as well as exogenously added retroviral DNA (2, 10, 11, 14, 37). Thereafter it was demonstrated by a direct physical assay that the IN proteins of HIV, MoMLV, and ASLV are the only viral proteins that are required both to generate the recessed 3' ends of the viral DNA and to accomplish their integration *in vitro* (4, 8, 18, 30). In this assay, oligonucleotides that represent the termini of the retroviral DNA were incubated with IN, and it was shown that besides specific cutting of the oligonucleotides, IN is also capable of integrating one oligonucleotide into another.

The IN protein of HIV-1 has been overexpressed in insect cells (5). This protein was partially purified, and it was shown that it is able to promote integration of mini-HIV DNA in a genetic assay. HIV-1 IN was also overproduced in and partially purified from *Escherichia coli* (4, 30). For this protein, it was demonstrated in a physical assay that it is

* Corresponding author.

capable of generating the 3' recessed ends of the retroviral DNA. In this case, double-stranded, blunt-ended oligonucleotides that represent the termini of the LTRs of HIV-1 DNA were used as substrate viral DNAs.

We overproduced HIV-1 IN in *E. coli* and purified it to near homogeneity. We show here that IN has specific integration activity as well as specific cleavage activity. The sequence requirements for efficient specific cleavage and integration of HIV DNA were investigated by using a series of oligonucleotides that contain mutations relative to an oligonucleotide that mimics the ends of HIV DNA. We demonstrate that for specific cutting by IN, the cleavage site (the phosphodiester bond at the 3' side of the conserved CA-3' dinucleotide) must be located at a subterminal position in the viral DNA. We further investigated whether IN could carry out integration into single-stranded as well as double-stranded DNA.

MATERIALS AND METHODS

Construction of a plasmid expressing HIV-1 IN. The HIV-1 IN expression plasmid that we constructed is essentially the same as the plasmid described by Sherman and Fyfe (30). The HIV-1 IN-coding sequence was derived from plasmid pHXB-2D (13). The IN gene was subcloned into plasmid pIAN7 (29) to generate plasmid pRP177. The sequence 5'-ATAGAT-3' at positions 3822 to 3827 in the IN-coding sequence (24) was changed by oligonucleotide-directed mutagenesis to the sequence 5'-ATCGAT-3', which is a *Cla*I restriction site. In plasmid pRP177, the *Stu*I site that is located at position 4985 (24) was lost by fusion to the *Hinc*II site of vector pIAN7. The adjacent *Bam*HI site in the polylinker of pIAN7 and the *Cla*I site in the IN-coding region were used to clone the IN gene into the *Nde*I and *Bam*HI sites of the expression vector pET-3c (25), together with the synthetic oligonucleotide 5'-TATGTTTCTAGATGGAAT-3' and its complement, having *Nde*I and *Cla*I cohesive ends. The ATG in the oligonucleotide (in bold print) is the translation start codon and precedes the TTT codon, which encodes the actual IN N-terminal amino acid, phenylalanine. In the resulting HIV-1 IN expression vector, pRP274, the IN gene is under the control of a bacteriophage T7 RNA polymerase promoter. *E. coli* BL21(DE3) was transformed with the IN expression plasmid. In this strain, a copy of the T7 RNA polymerase gene is present in the chromosome and is under the control of the *lacUV5* promoter.

Purification of HIV-1 IN. IN was purified as described by Sherman and Fyfe (30), with some modifications in the method of growing and lysis of the bacteria. Bacteria harboring plasmid pRP274 were grown for 16 h at 37°C in LB broth supplemented with 50 µg of ampicillin per ml. The culture was diluted 50 times in 1 liter of TB medium (27) containing 10 mM MgSO₄ and 50 µg of ampicillin per ml. After 4 h of growth at 37°C, isopropyl β-D-thiogalactopyranoside was added to a final concentration of 0.3 mM. Three hours after induction, cells were harvested by centrifugation and resuspended in cold 50 mM Tris-HCl (pH 7.5)–5 mM dithiothreitol–1 mM EDTA–1 mg of lysozyme per ml. After incubation on ice for 15 min, the suspension was sonicated. The sonic extract was centrifuged for 1 h at 12,000 × *g*. We found that with these modifications, we obtained virtually pure, soluble IN (see Results).

SDS-polyacrylamide gel electrophoresis. Proteins were separated on 10% sodium dodecyl sulfate (SDS)-polyacrylamide gels (20). Gels were silver stained by using a silver stain kit (Sigma Chemical Co.).

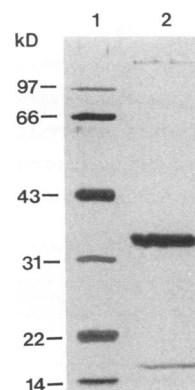


FIG. 1. HIV-1 IN expressed in and purified from *E. coli*. Purified IN was subjected to SDS-polyacrylamide gel electrophoresis and silver stained (lane 2). Lane 1, molecular size markers.

Oligonucleotide substrates. The synthetic oligonucleotides that were tested for specific cleavage and integration by IN are depicted in Fig. 2 to 4. Before labeling and annealing, the oligonucleotides were purified by electrophoresis in a denaturing 12% polyacrylamide (20:1 acrylamide/bisacrylamide) gel. Ten picomoles of the oligonucleotide strand to be labeled with ³²P at its 5' end was incubated with 40 µCi of [γ -³²P]ATP (3,000 Ci/mmol; Amersham Corp.) and 0.1 U of T4 polynucleotide kinase (Boehringer Mannheim) in a reaction volume of 10 µl. After incubation for 1 h at 37°C, 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 gel. After electrophoresis, an X-ray film (X-Omat AR; Eastman Kodak Co.) was exposed to the gel to localize the labeled oligonucleotide. The oligonucleotide was then cut from the gel and eluted for 16 h at 37°C in elution buffer (27). After precipitation with ethanol, the oligonucleotide was resuspended in H₂O to an end concentration of 0.2 pmol/µl. Then 1 µl (10 pmol) of the unlabeled complementary strand was added to 10 µl of this oligonucleotide sample, and the mixture was heated to 90°C. The DNA was annealed by slow cooling to room temperature.

Cleavage and integration reactions. Reaction conditions for specific cleavage and integration of the oligonucleotide substrates were generally as described by Craigie et al. for the MoMLV IN protein (8). Reaction mixtures (10 µl) contained 75 mM NaCl, 20 mM morpholinepropanesulfonic acid (MOPS; pH 7.2), 3 mM MnCl₂, 10 mM dithiothreitol, 20% (vol/vol) glycerol, 100 µg of bovine serum albumin per ml, 0.2 pmol of duplex oligonucleotide, and about 50 ng of HIV IN protein. The reaction mixtures were at 30°C for 1 h, and the reactions were stopped by addition of 10 µl of formamide loading dye. After heating to 80°C for 3 min, 5 µl of the samples was loaded onto a 12% polyacrylamide–8 M urea–1 × TBE (89 mM Tris, 89 mM boric acid, 2 mM EDTA) gel and electrophoresed. The gel was fixed by incubation in 10% methanol–10% acetic acid for 30 min and dried on Whatman 3MM paper. Reaction products were visualized by autoradiography.

RESULTS

Purification of HIV-1 IN. The IN protein of HIV-1 was purified to near homogeneity essentially as described by

A		25	20	15	10	5	1	sp.cl.	int.	
HIV-1 U5	5'-	TTAGT	CAGT	GTGG	AAAAT	CTCT	TAGCAGT	-3'	++	++
HIV-1 U3		TCGTTGGG	GAGTGAAT	TAGCC	CTTCC				++	++
HIV-2 U5		AAACCGAG	GCAGG	AAAAT	CCCTAG				++	++
HIV-2 U3		TCTTTACT	GTAAAAC	ATCC	CTTCC				++	++
MoMLV U5		GACTACCC	GTGAG	CGGGGG	TCTTTC				-	-
HIV-1 U5	5'-	TTAGT	CAGT	GTGG	AAAAT	CTCT	TAGCAGT	-3'		
mutants									sp.cl.	int.
G(-1)>T									++	+
CA(2-1)>TG									-	-
A(1)>T									+	-
A(1)>G									+	-
C(2)>T									+	+/-
C(2)>G									+	+/-
AG(4-3)>GA									+	+
T(5)>A									++	++
C(6)>T									++	++
C(8)>G									++	++
A(10)>G									+	+
AAATC(12-8)>GGGGG									+	+/-
A(13)>C									++	+
T<->G, A<->C(26-14)									++	++
MoMLV U5	5'-	GACTACCC	GTGAG	CGGGGG	TCTTTC					
mutants										
T(3)>G									-	-
TT(4-3)>AG									+/-	-
GGGGG(12-8)>AAATC									-	-
C(13)>A									+/-	-

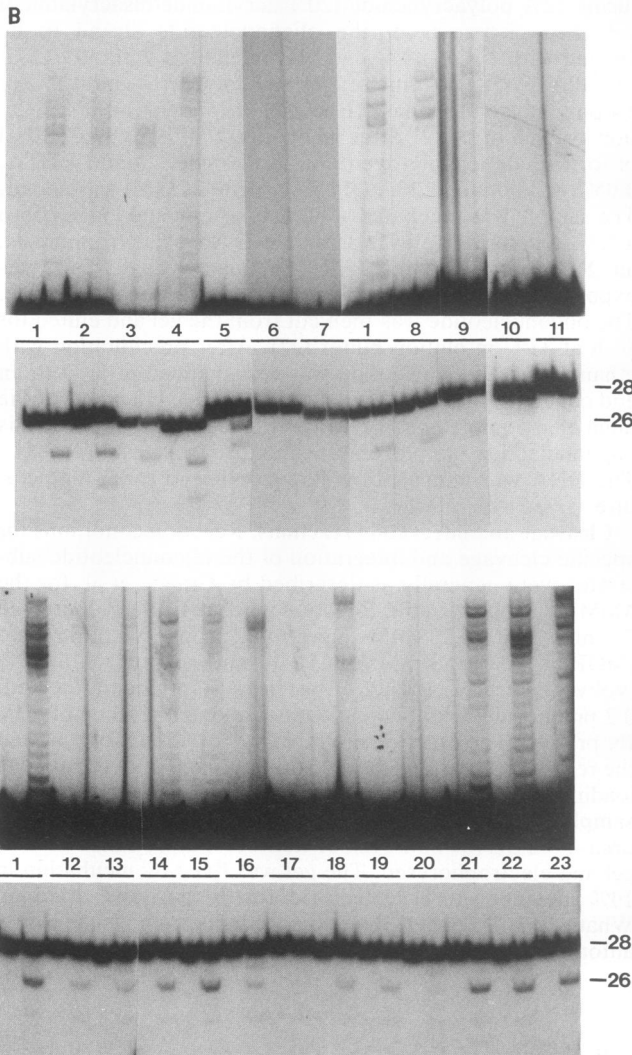


FIG. 2. Specific cleavage and integration of oligonucleotide substrates by HIV-1 IN. (A) Sequences of the 28-bp oligonucleotides used and summary of results from the cleavage and integration assay. The HIV-1 sequences are from reference 38, the HIV-2 sequences are from reference 16, and the MoMLV sequences are from reference 31. Numbers in parentheses refer to positions of the nucleotides shown at the top. Only one strand of the double-stranded substrates is shown. The 3' end of each strand represents the terminus of the viral DNA. The conserved dinucleotide CA-3' is indicated in bold print. sp.cl., specific cleavage at the phosphodiester bond 3' of the CA-3' dinucleotide. Specific cleavage was scored when the band at the position on the gel of the specific product was the predominant product. Int., integration. ++, high level of specific cleavage or integration; +, moderate level of specific cleavage or integration; +/-, low level of specific cleavage or integration; -, no specific cleavage or integration. (B) Specific cleavage and integration of the various substrates. The oligonucleotides were incubated in the presence (right lane) or absence (left lane) of HIV-1 IN. The top panels show a longer exposure of the gel containing products longer than 28 nucleotides, to identify integration products. Lengths of the oligonucleotide strands are shown on the right. Lanes contain the following substrates: 1, HIV-1 U5; 2, HIV-1 U3; 3, HIV-2 U5; 4, HIV-2 U3; 5, MoMLV U5; 6, AAATC(12-8)>GGGGG; 7, MoMLV T(3)>G; 8, C(8)>G; 9, A(13)>C; 10, MoMLV GGGGG(12-8)>AAATC; 11, MoMLV TT(4-3)>AG; 12, A(1)>G; 13, C(2)>T; 14, A(10)>G; 15, G(-1)>T; 16, AG(4-3)>GA; 17, CA(2-1)>TG; 18, C(2)>G; 19, A(1)>T; 20, MoMLV C(13)>A; 21, C(6)>T; 22, T(5)>A; 23, T<->G, A<->C(26-14).

Sherman and Fyfe (30). Purified IN is shown on a silver-stained SDS-polyacrylamide gel in Fig. 1. Only a few other minor bands can be detected in the IN preparation. The identity of the 32-kDa IN band was determined by immunoblotting, using polyclonal antisera that were raised against an oligopeptide that is identical to the C-terminal 11 amino acids of HIV-1 IN (data not shown). We found that by using some modifications in the method of induction and lysis of the bacteria that express IN (see Materials and Methods), we obtained protein preparations that had a higher purity than those previously described (30).

Sequence requirements for specific cleavage and integration. The sequence requirements for specific cleavage and integration of retroviral DNA by HIV-1 IN were determined in a physical assay. Initially, the activities of HIV-1 IN were tested on synthetic oligonucleotides that mimic the terminal 28 bp of HIV-1, HIV-2, and MoMLV DNA. The sequences of these oligonucleotide substrates are shown in Fig. 2A. Purified IN was incubated with the blunt-ended oligonucleotides under the appropriate reaction conditions (see Materials and Methods), and the products were electrophoresed in a denaturing polyacrylamide gel. Reaction products were visualized by autoradiography (Fig. 2B). As was shown previously (4, 30), HIV-1 IN is capable of specifically removing two nucleotides from the 3' ends of the HIV-1 U3 and U5 termini. After incubation with IN, a major product of 26 nucleotides could be seen, in addition to the band corresponding to the substrate. A minor but clear product of 24 nucleotides could also be detected, indicating that HIV-1 IN is also able to cleave, with a low efficiency, the phosphodiester bond between the G nucleotide at position 3 and the C nucleotide at position 2 of the U5 sequence and between the two C nucleotides at positions 2 and 3 in the U3 sequence of HIV-1 (Fig. 2A). The relevance of cleavage at these positions is unknown. The higher reactivity of HIV-1 U5 ends than of HIV-1 U3 ends, which has previously been described (4), was not seen; the HIV-1 U3 and U5 oligonu-

cleotides were cleaved at approximately equal levels (Fig. 2B, lanes 1 and 2). Oligonucleotides that represent the U3 and U5 ends of HIV-2 were also good substrates for specific cleavage by HIV-1 IN; a major product with a length of 26 nucleotides could be detected in both instances. Reaction products that are longer than 28 nucleotides could be detected after incubation of IN with both the HIV-1 and HIV-2 substrates (Fig. 2B, upper panel). It was shown for MoMLV IN (8), ASLV IN (18), and HIV-1 IN (4) that these bands represent products of integration of one oligonucleotide into another. Products that are shorter than 26 nucleotides can also be explained by integration, because the 3' ends of the cut made in the (5') labeled strand of the target DNA remain unjoined in the integration product. Specific cleavage and integration were not observed when the oligonucleotides were incubated with an extract from *E. coli* BL21(DE3) transformed with plasmid pET-3c (data not shown). This extract was subjected to the same purification procedure as was the HIV-1 IN preparation.

HIV-1 IN is not able to cleave the MoMLV U5 oligonucleotide with high specificity (Fig. 2). A nonspecific degradation ladder could be observed when the MoMLV substrate was incubated with IN.

We set out to determine the nucleotides present at the LTR ends of HIV-1 that are required for specific cleavage and integration by IN. For this purpose, we tested oligonucleotides that contain mutations relative to the oligonucleotide that represents the terminal 28 bp of the HIV-1 U5 sequence. The sequences of the different mutant substrates are shown in Fig. 2A. The results obtained from the cleavage and integration assay are depicted in Fig. 2B and summarized in Fig. 2A. Mutations that are located in the vicinity and at the 5' side of the IN cleavage site (the phosphodiester bond at the 3' side of the CA-3' dinucleotide at positions 1 and 2 in Fig. 2A) all had a reducing effect on specific cleavage and integration. The most drastic effect was seen with the oligonucleotide that contains the CA(2-1)>TG mutation. This substrate was not cleaved at all by IN. When the A nucleotide at position 1 and the C nucleotide at position 2 were mutated separately, a reduced but still visible level of specific cutting was observed. Apparently, both positions are important, and the added effect of the double mutant [CA(2-1)>TG] is a total block of cleavage. Although integrations of mutants C(2)>G and C(2)>T were observed, no integrations were seen of the oligonucleotides that are mutated at position 1 [A(1)>T and A(1)>G] (see Discussion). It should be noted that the changes that are introduced into the mutant substrates could influence the function of the oligonucleotides as targets for integration.

The A nucleotides at positions 10 and 13 were also shown to be important for cleavage and integration by IN; the substrates A(10)>G and A(13)>C were cleaved and integrated with a lower efficiency than the wild-type substrate. Other point mutations did not reduce the degree of site-specific cleavage and integration. When all of the nucleotides at positions 14 to 26 were changed [T↔G, A↔C(26-14)] in Fig. 2A], no effect on cleavage and integration could be detected. This suggests that the terminal 15 bp of the retroviral DNA are sufficient for recognition by IN (see below).

To investigate whether we could change the MoMLV U5 end sequence in such a way that it would become a target for specific cutting by HIV-1 IN, we altered some of the nucleotides in the MoMLV oligonucleotide to bases that are present at similar positions in the HIV-1 U5 substrate and were shown above to be important for cleavage and integra-

tion. For example, we showed that the A(13)>C mutation in HIV-1 U5 resulted in a lower level of cutting by HIV-1 IN. The MoMLV U5 sequence contains a C nucleotide at position 13 (Fig. 2A). We changed this C nucleotide in the MoMLV U5 sequence to an A nucleotide that is present at the same position in the HIV-1 U5 oligonucleotide. The resulting oligonucleotide, C(13)>A, was not a good substrate for cleavage and integration by HIV-1 IN. The other MoMLV mutants were also not cleaved efficiently by IN.

The results presented here indicate that the conserved CA-3' dinucleotide is required for specific cleavage and integration. Most other nucleotides at the viral DNA ends, however, are not absolutely required for specific IN activity. The substrate specificity is therefore probably determined by several nucleotides at the viral DNA termini that all contribute to specific recognition. This explains why no single point mutation in the HIV-1 U5 terminus completely blocked IN activity and why changing single nucleotides of the MoMLV U5 end into nucleotides that are present at similar positions in the HIV-1 U5 end did not result in efficient site-specific cleavage by HIV-1 IN.

Spacing requirements for specific cleavage. HIV-1 IN normally cleaves the 3' end of the blunt-ended retroviral DNA at the phosphodiester bond that is located two nucleotides away from the 3' terminus. Since we showed that the sequence requirements for cutting and integration are not very strong, we wanted to determine whether additional specificity is provided by the position of the recognition site with regard to the viral DNA termini. We tested whether IN would tolerate additional nucleotides at the 3' end of the HIV-1 U5 sequence for specific cleavage and integration. We studied a series of mutant double-stranded oligonucleotides that contain either no, one, two (the wild-type situation), three, four, five, or six nucleotides 3' of the conserved CA-3' dinucleotide. The sequences of these substrates are depicted in Fig. 3A; the results of the cleavage and integration assay are shown in Fig. 3B. It is clear that all substrates were cleaved at the phosphodiester bond that is located at the 3' side of the CA-3' dinucleotide. The blunt-ended substrate that does not contain any residues 3' of the CA-3' (substrate 1/2) cannot be specifically cleaved by IN. However, integrations of this substrate could be detected (Fig. 3B). The wild-type oligonucleotide, having two nucleotides (GT-3') 3' of the CA-3' dinucleotide, was the best substrate for site-specific cleavage and integration by IN. Specific cleavage gradually decreased when more nucleotides were added to the 3' end. The oligonucleotide having six nucleotides 3' of the CA-3' sequence (substrate 13/14 in Fig. 3B) was not a good substrate for cleavage. A specific band at the position of 26 nucleotides can be seen, but this band is less intense than the nonspecific bands at the positions of 31, 30, 29, and 24 nucleotides. The substrate containing only one nucleotide (a G) 3' of the CA-3' was cleaved less efficiently than the wild-type substrate. Integration events, as seen in Fig. 3B from the products on the gel that are larger than 28 nucleotides, were most frequent when the wild-type substrate was used. The other oligonucleotides that have less than six nucleotides 3' of the CA-3' were integrated at approximately equally low frequencies. Integrations of the largest substrate (substrate 13/14) were not detected.

The inhibitory effect of additional base pairs at the viral ends may result from steric hindrance of the added sequences in the recognition by IN. To investigate this further, we also tested substrates in which only one DNA strand was extended. We first investigated whether HIV-1 IN would cleave duplex oligonucleotides in which the sequence 3' of

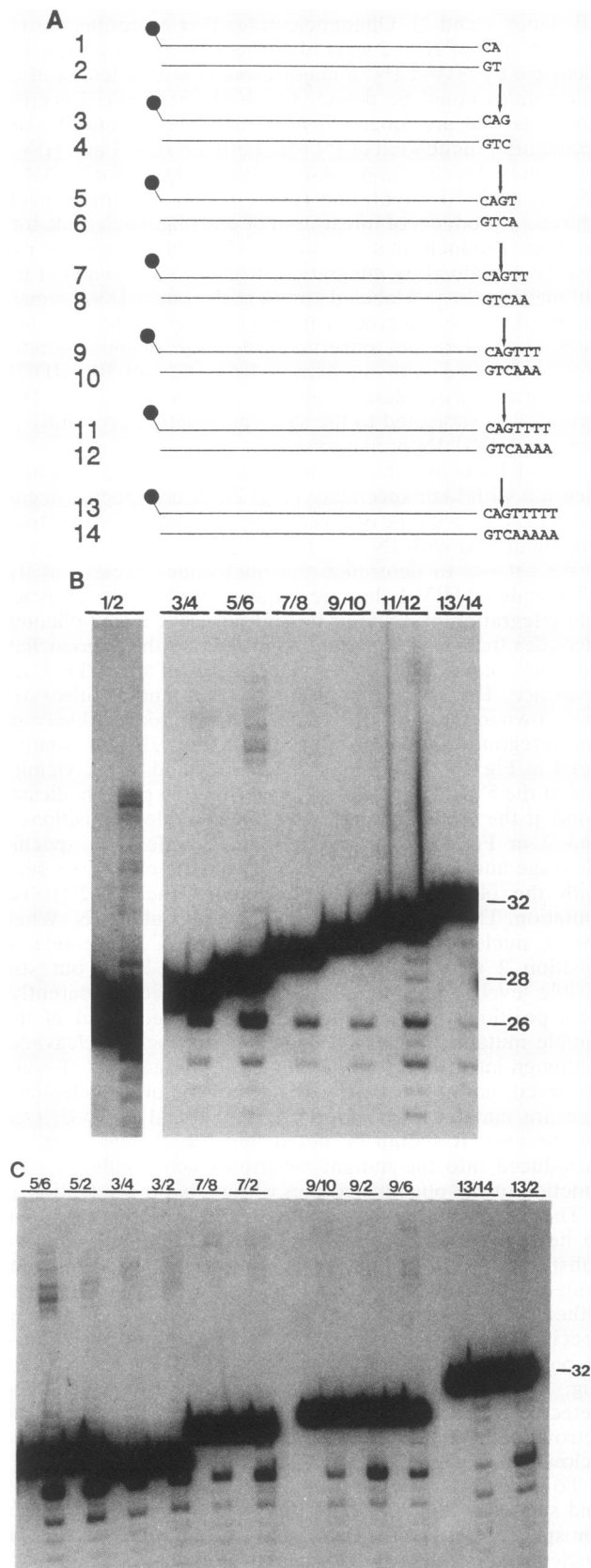
the conserved CA-3' dinucleotide was present as single-stranded DNA. The substrate containing a 3' overhang of two nucleotides (substrate 5/2) was cleaved more efficiently than its flush-ended counterpart, the wild-type oligonucleotide (substrate 5/6) (Fig. 3C). However, integration was slightly reduced. The other substrates tested all showed similar patterns of cleavage; when the sequence at the 3' side of the CA-3' dinucleotide was present as single-stranded DNA, cleavage was greater than with the substrates that contain double-stranded extensions. Even the substrate that contains six bases 3' of the CA-3' sequence (substrate 13/2) was cleaved and integrated. We tested whether the enhanced cleavage of substrates that contain 3' protrusions was the result of specific IN activity and not of nonspecific, 3' single-stranded exonuclease activity caused by contaminating proteins in the IN preparation. For this purpose, the labeled single-stranded oligonucleotide of 30 nucleotides that contains four bases at the 3' side of the CA-3' dinucleotide (oligonucleotide 9) was annealed with the oligonucleotide that has the wild-type sequence (oligonucleotide 6). IN-specific cleavage of this substrate would result in a product of 26 nucleotides, whereas nonspecific 3' single-stranded exonuclease activity would result in a product of 28 nucleotides. In Fig. 3C, a major band of 26 nucleotides can be observed after incubation of substrate 9/6 with IN. This finding demonstrates that the enhanced cleavage is indeed the result of IN activity.

We also tested whether 5' single-stranded extensions in the DNA strand that is not cleaved by IN would influence specific cleavage. The labeled wild-type strand of 28 nucleotides (oligonucleotide 5) was annealed with oligonucleotides that are longer by one (oligonucleotide 8), two (oligonucleotide 10), three (oligonucleotide 12), or four (oligonucleotide 14) nucleotides, and the resulting substrates were incubated with IN (Fig. 3D). The level of site-specific cleavage and integration (which could be observed after longer exposure of the gel; data not shown) was comparable to the wild-type level in all cases. Consequently, IN does not seem to be hindered by these 5' single-stranded extensions.

The cleavage and integration assays with the oligonucleotides shown in Fig. 3A were also performed with purified HIV-2 IN. The purification of HIV-2 IN will be described elsewhere (34a). HIV-2 IN was indeed able to specifically cleave the HIV-1 U5-derived substrates and showed a cleavage pattern similar to that of HIV-1 IN (Fig. 3E). Also, in this case the wild-type substrate was cleaved more efficiently than the other oligonucleotides, and cleavage decreased gradually when more nucleotides were added to the 3' side of the conserved CA-3' dinucleotide. Enhanced cleavage was seen when the bases 3' of the CA-3' were present as single-stranded DNA, as was observed for the HIV-1 IN protein.

Specific cleavage and integration of 15-bp oligonucleotides.

We showed that the terminal 15 bp of the LTRs of HIV-1 are sufficient for specific cleavage and integration by IN. These terminal 15 bp, however, were present on an oligonucleotide of 28 bp (Fig. 2). Therefore, we wanted to determine whether an oligonucleotide of only the terminal 15 bp of the HIV-1 U5 sequence would be sufficient for specific cleavage and integration by HIV-1 IN. Cleavage of the 15-bp substrate was compared with cleavage of the 28-bp wild-type substrate (oligonucleotide 5/6) and of the 28-bp substrate that contains mutations at positions 26 to 14 (Fig. 2). We also tested duplex oligonucleotides that have 15 nucleotides in one strand and 28 in the other strand. The structures of the oligonucleotides are depicted in Fig. 4A; the results are



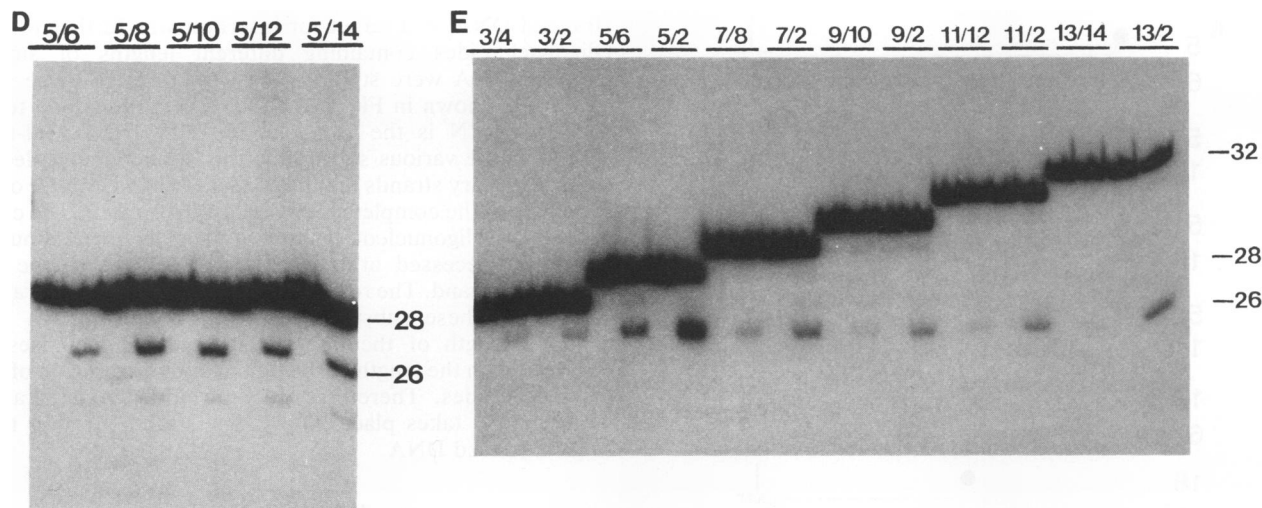


FIG. 3. Cleavage and integration of mutant HIV-1 U5 oligonucleotides that contain deletions or extensions at the 3' side of the conserved CA-3' dinucleotide. (A) Structures of the oligonucleotide substrates. In each case, the top line represents the HIV-1 U5 sequence as depicted in Fig. 2A, and the bottom line represents the complementary strand. The lines correspond to a length of 24 nucleotides. Each dot indicates the radiolabeled phosphate group at the 5' end of the viral strand. Each IN cleavage site is depicted by an arrow. (B) Specific cleavage and integration of the substrates shown in panel A. The substrates were incubated in the absence (left lane) or presence (right lane) of HIV-1 IN. The autoradiograph shown in lane 1 (substrate 1/2) was exposed twice as long as those in the other lanes. Substrate numbers (from panel A) are indicated above the lanes. Lengths of the oligonucleotide strands are shown on the right. (C) Specific cleavage and integration of substrates that contain single-stranded extensions in the oligonucleotide strand that is cleaved by IN. Substrate numbers (from panel A) are indicated above the lanes. (D) Specific cleavage and integration of substrates that have single-stranded extensions in the strand that is not radiolabeled. Substrate numbers are indicated above the lanes. (E) Specific cleavage and integration of different substrates by HIV-2 IN. Substrate numbers are indicated above the lanes.

shown in Fig. 4B. The 15-bp substrate could indeed be cleaved by IN, but there were about equal amounts of cleavage product of 14 nucleotides (a nonspecific product) and of 13 nucleotides (the specific product). Integration products could also be observed when the 15-bp substrate was used. We conclude that the 15-bp oligonucleotide is a substrate for specific cleavage and integration but not as good a substrate as the 28-bp oligonucleotide. Specific cleavage was more efficient with the substrate in which the strand that is cleaved by IN had a length of 15 nucleotides and the complementary strand had a length of 28 nucleotides (substrate 18/6; Fig. 4B). In addition, integration events were more frequent with this substrate than with the 15-bp substrate. It might be suggested that more nucleotides than the terminal 15 bp of the HIV-1 U5 sequence are required for recognition by IN. We have shown above, however, that all residues located more than 15 nucleotides away from the viral termini could be altered without a detectable effect on specific cutting. In combination with the observation that a single-stranded extension beyond position 15 from the terminus is sufficient to rescue high-efficiency cleavage, we suggest that IN is helped by the presence of any (nonspecific) extension to hold on to its viral recognition site.

Target specificity for integration. The largest products of integration of the 15-bp oligonucleotide were 24 nucleotides in length (Fig. 4B and data not shown). This finding demonstrates that integration can take place at phosphodiester bonds that are located 4 nucleotides or more removed from the 5' end of the target DNA (13 nucleotides are obtained from the 3' recessed strand that integrates, and apparently the maximum length of the target DNA strand is 11 nucleotides). The largest product that resulted from integration of the wild-type substrate had a length of 50 nucleotides (as observed after longer electrophoresis of reaction products

and overexposure of the gel; data not shown), which implies that the 26 nucleotides of the recessed donor oligonucleotide integrated 4 nucleotides or more removed from the 5' termini of a 28-bp target DNA. When substrate 18/6 was used, the largest integration products that could be detected were 38 nucleotides in length (Fig. 4B), which means that in this case the integration site was 3 nucleotides or more removed from the termini of the target DNA. These experiments show that the integration site must be located at an internal position in the target DNA and at least three nucleotides away from the 5' end of the target DNA.

An interesting observation can be made when the maximal length of the integration product of substrate 5/17 is considered. The radiolabeled strand of the donor DNA that can integrate is 26 nucleotides long (after site-specific removal of two bases from the initial 28-nucleotide strand by IN); the complementary strand is 15 nucleotides long. If the target phosphodiester bond could be any bond in either oligonucleotide strand that is located 3 or more nucleotides away from a 5' end (see above), then the maximum length of the integration products is expected to be 51 nucleotides (26 plus 25). As shown in Fig. 4B, the ladder of integration products extends to a position corresponding to 40 nucleotides. This shows that the target site for efficient integration cannot be found in the 5' single-stranded region of the 28-nucleotide strand of substrate 5/17. Apparently, integration preferentially takes place into double-stranded DNA. Close inspection of Fig. 4B reveals faint bands that correspond to reaction products longer than 40 nucleotides. We can exclude that these bands represent the products of rare double integration events, since this explanation is not consistent with the observation that products larger than 40 nucleotides were not generated when substrate 18/6 was used. An explanation might be that integration can take place with a

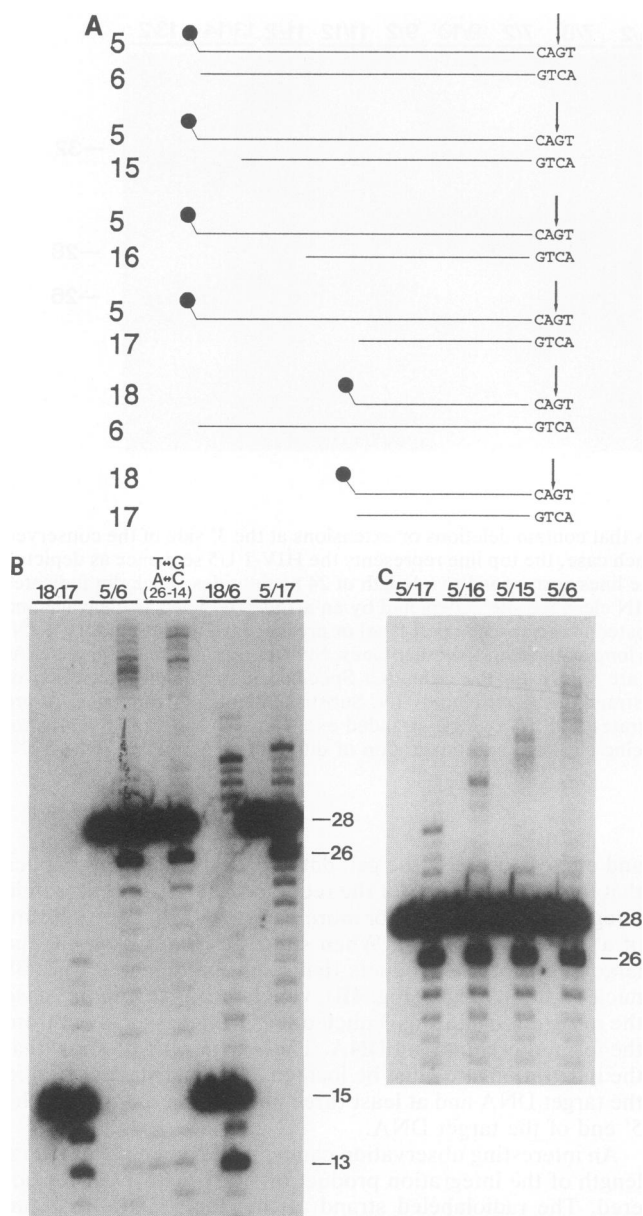


FIG. 4. Specific cutting and integration of 15-bp oligonucleotides and of oligonucleotides that contain stretches of single-stranded DNA of different lengths. (A) Structures of the substrates. Substrate 5/6 is the same as in Fig. 3. Oligonucleotide strands 15, 16, and 17 represent the terminal 23, 19, and 15 nucleotides of the HIV-1 U5 strand that is not cleaved by IN. Each IN cleavage site is indicated by an arrow. (B) Specific cleavage and integration of 15- and 28-bp substrates and of substrates that contain both a 15- and a 28-nucleotide strand. The substrates were incubated in the absence (left lane) or presence (right lane) of HIV-1 IN. Substrate numbers (from panel A) are indicated above the lanes. The structure of substrate T↔G,A↔C(26-14) is shown in Fig. 2A. Lengths of the oligonucleotide strands are shown on the right. (C) Specific cutting and integration of substrates having different stretches of single-stranded DNA. Substrate numbers are indicated above the lanes.

low frequency into single-stranded DNA. However, we cannot rule out that some base pairing of the single-stranded DNA occurs under the reaction conditions that were used.

The conclusion that IN preferentially uses double-

stranded DNA as a target for insertion was further tested. Oligonucleotides containing different lengths of single-stranded DNA were studied. The structures of these substrates are shown in Fig. 4A. In all cases, the strand to be cleaved by IN is the 28-nucleotide HIV-1 U5 wild-type strand. In the various substrates, this strand is annealed to complementary strands that have a length of 15, 19, 23, or 28 nucleotides (the complementary wild-type strand). The complementary oligonucleotides that are smaller than 28 nucleotides are recessed at their 3' ends relative to the 28-nucleotide strand. The results of the cleavage and integration assay with these substrates are shown in Fig. 4C. The maximal length of the integration products increases as expected with the length of the double-stranded region of the oligonucleotides. Therefore, we conclude that integration preferentially takes place into double-stranded rather than single-stranded DNA.

DISCUSSION

The IN protein of HIV-1 was expressed in *E. coli* and purified to near homogeneity. The sequence and spacing requirements for efficient specific cleavage and integration of HIV DNA by IN were studied in a direct physical assay. HIV-1 IN was shown to be capable of specifically cutting and integrating 28-bp oligonucleotides that mimic the U3 and U5 termini of HIV-1 as well as HIV-2 DNA. Despite the differences in sequence between the U3 and U5 end of HIV-1 or HIV-2, these substrates were cleaved and integrated with about equal efficiencies.

Oligonucleotide substrates that contain mutations relative to the oligonucleotide that represents the terminal 28 bp of the HIV-1 U5 sequence were tested. The results presented here demonstrate that there is only a limited sequence requirement for site-specific cleavage and integration by HIV-1 IN. Only point mutations in the direct vicinity and at the 5' side of the IN cleavage site could drastically reduce cleavage and integration. Mutation of the conserved CA-3' dinucleotide to the sequence TG-3' totally blocked cleavage and, as a consequence integration. When the A at position 1 (Fig. 2A) was changed to a T, low-efficiency cleavage was detected. However, integration events could not be detected when this substrate was used. This result could be explained by the fact that the A at position 1 of the HIV 3' ends is the nucleotide that is coupled to the target DNA. The importance of this A nucleotide for integration has also been observed for MoMLV (3). In this case, the integration of plasmids, linearized by digestion with *EcoRI*, was studied in a genetic assay. It was shown that integration was favored when the termini of the plasmids were modified in such a way that an A nucleotide was present at the 3' ends.

A limited sequence requirement for integration has been described previously for the MoMLV IN protein, which was provided either as cytoplasmic extract from MoMLV-infected cells (37) or as a partially purified protein (3). MoMLV IN was shown to be capable of integrating HIV DNA, despite the considerable difference in sequence between the termini of MoMLV and HIV DNA (3, 37); 7 of 15 nucleotides at the U3 end and 9 of 15 residues at the U5 end are different between HIV-1 and MoMLV. The integration of HIV substrates by MoMLV IN was less efficient than the integration of MoMLV substrates (3, 37). We show here that despite the limited sequence requirements for integration, HIV-1 IN is not able to specifically cleave MoMLV DNA. Even when nucleotides 1 to 4 or 8 to 12 of the MoMLV U5 sequence in Fig. 2A were changed to the residues that

are present at corresponding positions in the HIV-1 U5 sequence [substrates TT(4-3)>AG and GGGGG(12-8)>AAATC, respectively], no efficient specific cleavage could be seen. It should be emphasized, however, that the criterion for specific cutting is rather stringent; low levels of cleavage cannot be detected in the assay that we used.

The limited sequence requirement for specific cleavage and integration of retroviral DNA termini raises the question of how IN can specifically recognize its viral target site. Besides the sequence context of the viral ends, the spacing between the IN cleavage site and the DNA terminus is important for cleavage and integration (see below). *In vivo*, in the cytoplasm of the infected cell, the viral DNA is part of a nucleoprotein complex that contains the IN protein as the only detectable protein component (12). It is possible that the IN proteins are helped, by their interaction with each other, to recognize the retroviral DNA ends. Thus, the interaction between the IN proteins or the continued presence of IN molecules in the right position in the viral core particle might circumvent the need for high-specificity recognition of the viral ends by IN.

The distance between the IN cleavage site (the phosphodiester bond 3' of the CA-3' sequence) and the retroviral DNA terminus is important for specific cleavage by IN. Efficient specific cleavage occurred when the IN cleavage site is spaced two nucleotides away from the DNA end. In the wild-type situation, the HIV-1 IN cleavage site is probably also located two nucleotides away from the DNA terminus. This was inferred from the sequences of junction fragments (so-called circle junctions) that are generated *in vivo* after circularization of linear double-stranded viral DNA molecules (17, 22, 23, 33, 39). We found that when only one nucleotide was present at the 3' side of the CA-3' sequence, the level of cleavage was less than with the wild-type substrate. The addition of more than two nucleotides to the 3' side of the CA-3' sequence resulted in a gradual decrease of site-specific cleavage. The finding that all oligonucleotides with more (or less) than two residues 3' of the conserved CA-3' were indeed cleaved at the phosphodiester bond 3' adjacent to the CA-3' implies that the cleavage site is determined by sequences internal to the LTR termini. This view is consistent with the observation that correct 3' recessed termini are formed *in vivo* and *in vitro* by MoMLV mutants that contain additional nucleotides 3' of the CA-3' dinucleotide (8, 26). It was demonstrated, however, that cleavage at the IN cleavage site did not occur *in vivo* when this site was located 12 bases internal to the terminus of the MoMLV DNA (26). This result is in agreement with our finding that HIV-1 IN is not able to specifically cleave retroviral DNA substrates when the IN cleavage site is located more than five nucleotides internal to the DNA ends. Our results are further consistent with the observation that ASLV circle junction sequences, in which the ASLV IN recognition sites are present at internal positions, are not efficient substrates for cleavage and, as a consequence, for integration (18, 19). Our observations could provide an explanation for why integrated proviruses are not substrates for IN and therefore cannot be excised from the host DNA.

Site-specific cleavage by IN was enhanced when the residues 3' adjacent to the conserved CA-3' dinucleotide were present as single-stranded DNA. A stretch of single-stranded DNA might be more accessible for specific cleavage by IN than is double-stranded DNA, because a single-stranded extension may provide less steric hindrance than does a double-stranded extension. Single-stranded exten-

sions at the 5' end near the IN cleavage site did not have any effect on cutting and integration.

The terminal 15 bp of HIV-1 DNA are sufficient for specific cleavage and integration by IN. Although a 28-bp oligonucleotide that mimics the ends of HIV-1 DNA was a better substrate for cutting and integration than was a 15-bp substrate, it was demonstrated that this difference was not the result of additional specific recognition of sequences in the region that is located more than 15 bp away from the terminus. The additional stretch of single- or double-stranded DNA that was present in the substrates that are longer than 15 nucleotides possibly plays a role in helping the IN protein to hold on to its viral DNA substrate. Similar observations have been made for many restriction endonucleases that require that their recognition site be a few base pairs removed from the ends of DNA.

It should be emphasized that the assay in which cleavage and integration were tested represents an artificial system. For example, the requirement for Mn^{2+} rather than Mg^{2+} as the divalent cation for efficient specific cleavage and integration (8, 30) is not consistent with the observations that *in vitro* integration of endogenous viral DNA, which is present in nucleoprotein complexes, takes place efficiently with Mg^{2+} as the divalent cation (1, 10, 15). The results concerning the substrate specificity of purified IN are, however, in agreement with what was observed in *in vivo* systems. The limited requirement for specific sequences at the retroviral DNA termini that we demonstrated for purified HIV-1 IN has previously been shown *in vivo* for MoMLV IN (26). The inability of HIV-1 IN to efficiently cleave its viral target site when this target site is located at an internal position has also been demonstrated in the *in vivo* MoMLV system (26). Therefore, it is probable that the results that we obtained do indeed reflect true properties of the IN protein. Further analysis of the properties of the IN protein might therefore contribute considerably to the knowledge of retroviral integration.

ACKNOWLEDGMENTS

We thank E. Wientjens and H. van der Gulden for synthesis of the oligonucleotides, and we thank M. Bolk for the construction of plasmid pRP177.

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 Research to R.H.A.P.

REFERENCES

1. Brown, P. O., B. Bowerman, H. E. Varmus, and J. M. Bishop. 1987. Correct integration of retroviral DNA *in vitro*. *Cell* 49:347-356.
2. Brown, P. O., B. Bowerman, H. E. Varmus, and J. M. Bishop. 1989. Retroviral integration: structure of the initial covalent product and its precursor, and a role for the viral IN protein. *Proc. Natl. Acad. Sci. USA* 86:2525-2529.
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. 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.

7. Colicelli, J., and S. P. Goff. 1988. Sequence and spacing requirements of a retrovirus integration site. *J. Mol. Biol.* **199**:47–59.
8. 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.
9. Donehower, L. A., and H. E. Varmus. 1984. A mutant murine leukemia virus with a single missense codon in *pol* is defective in a function affecting integration. *Proc. Natl. Acad. Sci. USA* **81**:6461–6465.
10. Ellison, V., H. Abrams, T. Roe, J. Lifson, and P. Brown. 1990. Human immunodeficiency virus integration in a cell-free system. *J. Virol.* **64**:2711–2715.
11. 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.
12. 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.
13. Fisher, A. G., E. Collalti, L. Ratner, R. C. Gallo, and F. Wong-Staal. 1985. A molecular clone of HTLV-III with biological activity. *Science* **316**:262–265.
14. 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.
15. Fujiwara, T., and K. Mizuuchi. 1988. Retroviral DNA integration: structure of an integration intermediate. *Cell* **54**:497–504.
16. Guyader, M., M. Emerman, P. Sonigo, F. Clavel, L. Montagnier, and M. Alizon. 1987. Genome organization and transactivation of the human immunodeficiency virus type 2. *Nature (London)* **326**:662–669.
17. Hong, T., K. Drlica, A. Pinter, and E. Murphy. 1991. Circular DNA of human immunodeficiency virus: analysis of circle junction nucleotide sequences. *J. Virol.* **65**:551–555.
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. Laemmli, U. K. 1970. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature (London)* **227**:680–685.
21. Muesing, M. A., D. H. Smith, C. D. Cabradilla, C. V. Benton, L. A. Lasky, and D. J. Capon. 1985. Nucleic acid structure and expression of the human AIDS/lymphadenopathy retrovirus. *Nature (London)* **313**:450–458.
22. Pauza, C. D. 1990. Two bases are deleted from the termini of HIV-1 linear DNA during integrative recombination. *Virology* **179**:886–889.
23. Pullen, K. A., and J. J. Champoux. 1990. Plus-strand origin for human immunodeficiency virus type 1: implications for integration. *J. Virol.* **64**:6274–6277.
24. Ratner, L., W. Haseltine, R. Patarca, K. J. Livak, B. Starcich, S. F. Josephs, E. R. Doran, J. A. Rafalski, E. A. Whitehorn, K. Baumeister, L. Ivanoff, S. R. Petteway, Jr., M. L. Pearson, J. A. Lautenberger, T. S. Papas, J. Ghayeb, N. T. Chang, R. C. Gallo, and F. Wong-Staal. 1985. Complete nucleotide sequence of the AIDS virus, HTLV-III. *Nature (London)* **313**:277–284.
25. Rosenberg, A. H., B. N. Lade, D. Chui, S. Lin, J. J. Dunn, and F. W. Studier. 1987. Vectors for selective expression of cloned DNAs by T7 RNA polymerase. *Gene* **56**:125–135.
26. 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.
27. Sambrook, J., E. F. Fritsch, and T. Maniatis. 1989. Molecular cloning: a laboratory manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
28. Schwartzberg, P., J. Colicelli, and S. P. Goff. 1984. Construction and analysis of deletion mutations in the *pol* gene of Moloney murine leukemia virus: a new viral function required for productive infection. *Cell* **37**:1043–1052.
29. Seed, B. 1983. Purification of genomic sequences from bacteriophage libraries by recombination and selection in vivo. *Nucleic Acids Res.* **11**:2427–2447.
30. 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.
31. Shinnick, T. M., R. A. Lerner, and J. G. Sutcliffe. 1981. Nucleotide sequence of Moloney murine leukaemia virus. *Nature (London)* **293**:543–548.
32. Shoemaker, C., S. Goff, E. Gilboa, M. Paskind, S. W. Mitra, and D. Baltimore. 1980. Structure of a cloned circular Moloney murine leukemia virus DNA molecule containing an inverted segment: implications for retrovirus integration. *Proc. Natl. Acad. Sci. USA* **77**:3932–3936.
33. 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.
34. 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.
- 34a. van Gent, D., Y. Elgersma, C. Vink, and R. H. A. Plasterk. Submitted for publication.
35. Varmus, H. E. 1983. Retroviruses, p. 411–503. *In* J. Shapiro (ed.), *Mobile genetic elements*. Academic Press, Inc., New York.
36. Vink, C., M. Groenink, Y. Elgersma, R. A. M. Fouchier, M. Tersmette, and R. H. A. Plasterk. 1990. Analysis of the junctions between human immunodeficiency virus type 1 proviral DNA and human DNA. *J. Virol.* **64**:5626–5627.
37. Vink, C., D. C. van Gent, and R. H. A. Plasterk. 1990. Integration of human immunodeficiency virus types 1 and 2 DNA in vitro by cytoplasmic extracts of Moloney murine leukemia virus-infected mouse NIH 3T3 cells. *J. Virol.* **64**:5219–5222.
38. Wain-Hobson, S., P. Sonigo, O. Danos, S. Cole, and M. Alizon. 1985. Nucleotide sequence of the AIDS virus, LAV. *Cell* **40**:9–17.
39. 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.