Nucleic Acids Research, Vol. 19, No. 14 3821–3827

DNA binding properties of the integrase proteins of human immunodeficiency viruses types 1 and 2

Dik C.van Gent, Ype Elgersma, Marian W.J.Bolk, Cornelis Vink and Ronald H.A.Plasterk*
Netherlands Cancer Institute, Division of Molecular Biology, Plesmanlaan 121, 1066 CX Amsterdam, The Netherlands

Received May 23, 1991; Accepted June 21, 1991

ABSTRACT

Integration of retroviral DNA into the host chromosome requires the integrase protein (IN). We overexpressed the IN proteins of human immunodeficiency viruses types 1 and 2 (HIV-1 and HIV-2) in *E. coli* and purified them. Both proteins were found to specifically cut two nucleotides off the ends of linear viral DNA, and to integrate viral DNA into target DNA. This demonstrates that HIV IN is the only protein required for integration of HIV DNA. Although the two types of IN proteins have only 53% amino acid sequence similarity, they act with equal efficiency on both type 1 and type 2 viral DNA. Binding of IN to DNA was tested: purified IN does not bind very specifically to viral DNA ends. Nevertheless, only viral DNA ends are cleaved and integrated. We interpret this as follows: *in vitro* quick aspecific binding to DNA is followed by slow specific cutting and integration. IN can not find viral DNA ends in the presence of an excess of aspecific DNA; *in vivo* this is not required since the IN protein is in constant proximity of viral DNA in the viral core particle.

INTRODUCTION

Acquired Immunodeficiency Syndrome (AIDS) is caused by HIV-1 or HIV-2. HIV-1 is widespread, whereas HIV-2 is primarily found in West-Africa. Although both viruses have only limited sequence similarity (42% on the nucleotide level), they cause a disease with similar symptoms. HIV-2 is more related to simian immunodeficiency virus (SIV) than to HIV-1 (1).

An essential step in the retroviral lifecycle is the integration of a double stranded DNA copy of the viral genomic RNA into the host genome. From studies of retroviral integration *in vitro* the following picture of the reaction has emerged (reviewed in 2). The 3′ terminal two nucleotides are removed from both ends of the double-stranded linear viral DNA (3–6). Subsequently, target DNA is cut in a staggered fashion, which leaves 5′ protruding ends, and the 3′ ends of the viral DNA are ligated to the 5′ protrusions of the target DNA (7, 8). This reaction does not require ATP. It is not yet clear whether there is a covalent intermediate between integrase (IN) and the target DNA or whether cleavage of the target DNA and ligation to the viral DNA are accomplished in a concerted reaction. *In vivo* the remaining single-stranded gaps are then filled in, probably by cellular repair enzymes, which results in a target DNA duplication.

One retroviral protein is required for the integration reaction: IN. Mutations in the IN coding region of the *pol* gene result in retroviruses that are defective in integration (9, 10). Various assays have been developed to test for activities of (partially) purified IN protein. IN proteins of several retroviruses have been shown to be able to cut off two nucleotides from the 3′ end of double-stranded oligonucleotides that mimic the viral DNA ends (6, 11, 12). In the same assay, integration of one oligonucleotide into another was observed (3, 4, 13–16). IN has only limited substrate specificity: most point mutations in the viral DNA do not severely affect specific cleavage and integration (14, 17), and the IN protein of Moloney murine leukemia virus (MoMLV) can mediate integration of DNA substrates with HIV DNA ends (13, 18). Probably IN function does not require much specificity, since IN remains in the viral core particle when the double-stranded DNA copy is synthesized, and it does not need to search a whole cellular genome for its specific viral sites (19, 20).

Although it has been established that partially purified HIV IN can carry out the integration reaction, it had not been excluded that contaminating host proteins are essential for the reaction. The IN protein of avian sarcoma leukosis virus, judged to be pure on a coomassie stained gel, is sufficient for integration *in vitro* (4), and by analogy one would expect that pure IN would also be sufficient for integration of mammalian retroviruses such as HIV. In this study we addressed this by determining the activity of purified HIV-1 and HIV-2 IN.

We studied the IN proteins of both HIV types for the following reasons:

1. The similarity between the IN proteins of HIV-1 and HIV-2 is 57%; the sequences of their viral DNA substrates differ in 3 out of 15 positions at the U3 end and in 2 out of 15 positions at the U5 end. In search for sequence specific DNA recognition domains, we tested whether the IN proteins of HIV-1 and HIV-2 prefer to bind and cleave their cognate viral DNA ends over those of the other.

2. *In vitro* assays for IN activity can be used to screen, test and improve drugs that may interfere with this reaction and have therapeutic potential. It will be important to check that these drugs act on the IN proteins of both HIV types.

* To whom correspondence should be addressed

In order to mediate viral integration, IN has to recognize viral DNA ends specifically and target DNA aspecifically. General DNA binding has been detected for the IN proteins of several retroviruses in South-Western assays and filter binding assays (6, 21–25). Also specific binding of MoMLV IN to the viral DNA ends in a South-Western assay was reported (26). The IN proteins of MoMLV and HIV-1 were reported to bind specifically to viral DNA termini in a mobility shift assay (27–29). Using purified and active HIV-1 and HIV-2 IN we found quite different results: IN has only an approximately three-fold preference for specific viral sequences over aspecific DNA. Following quick aspecific binding IN cuts and mediates integration only when it encounters viral substrate sequences.

MATERIALS AND METHODS

Materials

Restriction enzymes, T4 DNA ligase and T4 polynucleotide kinase were purchased from Boehringer Mannheim. Taq polymerase was purchased from Perkin Elmer Cetus, radiochemicals from Amersham.

DNA techniques

Standard DNA techniques were carried out as described (30). The HIV-1 IN expression vector (pRP 274) is essentially the same as the one described by Sherman and Fyfe (6), and is described in detail by Vink *et al.* (17). The HIV-2 IN expression plasmid was constructed as follows. The IN coding region of HIV-2 (clone pROD4.8, kindly provided by M. Emerman; ref. 1) was amplified by PCR, using oligonucleotides 5′-TTTCATATGTT-CCTGGAAAAAATAGAGCCC-3′ and 5′-TTTAGATCTAT-GCCATTTCTCCATCC-3′ as primers. The resulting PCR product had a NdeI site at the 5′ end, and a BglII site at the 3′ end of the IN coding region. By digesting the PCR product with NdeI and BglII and cloning the resulting fragment into the NdeI and BamHI digested pET-3c expression vector (31), the IN gene was placed under transcriptional control of the bacteriophage T7 promoter. In the resulting construct (pRP 279) an ATG translation initiation codon precedes the TTC triplet, encoding the N-terminal phenylalanine of HIV-2 IN. *E. coli* BL21(DE3), which carries the T7 RNA polymerase gene under control of the *lac*UV5 promoter, was used as host bacterium.

Protein production and purification

Bacteria expressing either the HIV-1 or the HIV-2 IN gene were grown overnight in LB medium supplemented with 50 μg/ml ampicillin. The o/n culture was diluted 1:50 in 1 liter TB medium (30) supplemented with 10 mM $MgSO_4$, 50 μg/ml ampicillin and grown for 4 hours at 37°C. IN expression was started by addition of isopropyl β-D-thiogalactopyranoside (IPTG) to a final concentration of 0.3 mM. Three hours after induction cells were collected by centrifugation and resuspended in 40 ml ice-cold lysis buffer (50 mM Tris.HCl, pH 7.5, 5 mM dithiothreitol (DTT), 1 mM EDTA, 1 mg/ml lysozyme) and incubated at 0°C for 15 minutes. After sonication of the cells, we followed the same procedure as described by Sherman and Fyfe for purification of HIV-1 IN (6). For the purification of HIV-2 IN we modified the procedure as follows. The 1 M NaCl extract was dialyzed for 2 hours against buffer 1 (50 mM Tris.HCl, pH 7.5, 0.5 M ammonium sulfate, 0.4 M NaCl, 5 mM DTT, 0.1 mM EDTA). After centrifugation (20 minutes, 10.000 × *g*) the supernatant was loaded onto a Butyl-Sepharose 4B (Pharmacia) column of 10 ml, equilibrated with buffer 1. After washing the column with 75 ml buffer 1, elution was started with a linear gradient from buffer 1 to buffer 2 (50 mM Tris.HCl, pH 7.5, 100 mM NaCl, 5 mM DTT, 10% (vol/vol) glycerol). The gradient (100 ml) was followed by 50 ml buffer 2. The integrase containing fractions (detected by SDS-PAGE analysis) were pooled and diluted with 1 volume buffer 3 (50 mM Tris.HCl, pH 7.5, 5 mM DTT, 10% (vol/vol) glycerol) before loading onto the Heparin-Sepharose CL-6B (Pharmacia) column. The protein was eluted from this column as described by Sherman and Fyfe (6). The HIV-1 IN preparation contained 30 ng of protein per μl and the HIV-2 IN preparation 100 ng/μl in buffer 3 containing 750 mM NaCl. Batches of integrase-containing fractions were frozen in liquid nitrogen and stored at −70°C.

Gel exclusion chromatography

A 180 ml Sephacryl S-300 HR (Pharmacia) column was equilibrated with 20 mM TRIS.HCl, pH 7.5, 1 M NaCl, 1 mM DTT and calibrated with Blue Dextran 2000, Bovine serum albumin, Ovalbumin, and Chymotrypsinogen A (sizemarkers purchased from Pharmacia). Proteins were detected by continuously monitoring the OD_{280} of the flow through. The position of IN was visualized by SDS-PAGE.

SDS-PAGE analysis

Protein samples were subjected to electrophoresis in 10% SDS-polyacrylamide minigels (Biorad) and visualized by either coomassie brilliant blue staining or silver staining (silver stain kit, purchased from Sigma). To obtain increased sensitivity, the silver staining procedure was done twice (as described by the supplier).

Cleavage and integration reactions

For the cleavage and integration reactions the following double-stranded oligonucleotide substrates were used: 28-mers representing the HIV-1 U5 (5′ TTAGTCAGTGTGGAAAAT-CTCTAGCAGT-3′ and the complementary strand) and U3 ends (5′-TCGTTGGGAGTGAATTAGCCCTTCCAGT-3′ and the complementary strand, sequence from ref. 32), the HIV-2 U5 (5′-AAACCGAGGCAGGAAAATCCCTAGCAGG-3′ and the complementary strand) and U3 ends (5′-TCTTTCA-CTGTAAAACATCCCTTCCAGT-3′ and the complementary strand, sequence from ref. 1) and the MoMLV U5 end (5′-GACTACCCGTCAGCGGGGGTCTTTCATT-3′ and the complementary strand, sequence from ref. 33). Also aspecific double-stranded oligonucleotides were used (5′-GCGTGTATG-AATCCGTCGAGAGCTACTA-3′ and the complementary strand). The strands of which the sequence is given were labeled using [γ-^{32}P]-ATP and T4 polynucleotide kinase. Full length oligonucleotides were isolated from a 12% denaturing polyacrylamide gel, and resuspended in TE (10 mM Tris.HCl, pH8, 1 mM EDTA) to a final concentration of 0.2 pmoles/μl, mixed with a 4-fold excess of the unlabeled complementary strand, heated to 80°C and allowed to anneal by slow cooling to room temperature.

The cleavage and integration reactions were carried out essentially as described for MoMLV IN (7). The standard reaction mixture contained 20 mM MOPS pH 7.2, 75 mM NaCl, 10 mM DTT, 20% glycerol (v/v), 100 μg/ml bovine serum albumin, 3 mM $MnCl_2$ (HIV-1) or 1 mM $MnCl_2$ (HIV-2), 0.2 pmoles of labeled double-stranded LTR substrate and 1 pmol

HIV-1 IN or 3 pmoles HIV-2 IN in a total volume of 10 μl. For the competition studies equal amounts of both strands of the competitor DNA were mixed, heated to 80°C and allowed to anneal by slow cooling to room temperature. The reaction mixtures were incubated for one hour at 30°C.

Reactions were stopped by addition of 10 μl 95% formamide, 20 mM EDTA, 0.05% bromophenol blue and 0.05% xylene cyanol. Samples were heated to 80°C for 2 minutes and 5 μl samples were analyzed on 12% denaturing polyacrylamide gels in 1× TBE (30). Gels were dried and reaction products were visualized by autoradiography. To quantitate the reaction, cleaved and uncleaved products were excised from the gel. Gelslices were incubated overnight in scintillation cocktail (Ultima Gold, Packard) and their radioactivity was determined in a scintillation counter.

Mobility shift assay

Binding reactions were performed under the same conditions as the cleavage reactions or under conditions that were essentially the same as described for HIV-1 IN binding (29): 25 mM HEPES, pH 7.5, 75 mM NaCl, 5 mM $MgCl_2$, 1 mM DTT, 5% glycerol (v/v), 0.4 pmoles of double-stranded oligonucleotide and 2 pmoles of HIV-1 IN or 6 pmoles of HIV-2 IN in a total volume of 20 μl. Reaction mixtures were incubated for 15 minutes at 30°C (which are the conditions of the cleavage reaction) or for 10 minutes at room temperature. The glycerol concentration was adjusted to 20% and the samples were separated in 6% polyacrylamide gels in 1/2×TBE at room temperature (5V/cm). Gels were dried and products were visualized by autoradiography.

RESULTS

Overexpression and purification of HIV-1 and HIV-2 IN

The IN coding regions of HIV-1 and HIV-2 were cloned behind the T7 promoter and an ATG initiation codon in the expression vector pET-3c. Both proteins were overproduced and purified essentially as described (6), with some modifications (see materials and methods). Figure 1A shows an SDS-PAGE analysis of a total cell lysate of HIV-2 IN expressing *E. coli* cells (lane 2), proteins solubilized in 1M NaCl (lane 3) and purified HIV-2 IN (lane 4). On a silver stained gel (figure 1B) the HIV-2 IN preparation appears to be pure; HIV-1 IN contains some minor contaminants. The yield of protein from one liter of *E. coli* culture was approximately 0.3 mg for HIV-1 IN and 1 mg for HIV-2 IN.

The subunit stoichiometry of HIV-1 IN and HIV-2 IN in solution (containing 1M NaCl) was determined by gel exclusion chromatography. Both proteins eluted between ovalbumin and chymotrypsinogen A, corresponding to a MW of approximately 32 kD, as expected for IN monomers (figure 1C). At lower salt concentration the proteins precipitate out of solution (see below).

Activities of HIV-2 IN

HIV-1 IN can specifically cleave the ends of viral DNA and integrate viral DNA substrates into target DNA (6, 14, 17). HIV-2 IN was tested for the same activities. As shown in figure 2 (lanes 9 and 12), purified HIV-2 IN was able to cut off two nucleotides from the 3′ end of oligonucleotide substrates that represent the HIV-2 DNA ends. We investigated whether the IN proteins of HIV-1 and HIV-2 recognize each others substrates and whether they have a preference for their own DNA ends. As shown in figure 2, HIV-1 and HIV-2 IN have similar substrate preferences: they both act on HIV-1 and HIV-2 substrates with equal efficiency. For both IN proteins U5 is a somewhat better substrate than U3. However, the U5 end of MoMLV, a more distantly related retrovirus with DNA ends that differ significantly from HIV DNA ends (8 out of 15 nucleotides are similar at the U3 end and 6 out of 15 at the U5 end), was cleaved with much lower efficiency than HIV DNA ends, and no integration products were detected (figure 2 lanes 14 and 15), not even after long exposure times (not shown). HIV-2 IN cleaved the MoMLV substrate primarily at the expected position, 3′ of the conserved CA sequence. Since there is some aspecific nuclease activity in the HIV-1 IN preparation, no specific cutting of MoMLV DNA ends could be detected above the background.

In this assay also products were formed, that were longer than the substrate (figure 2 lanes 6 and 12). It has previously been shown, that those are the result of integration of one

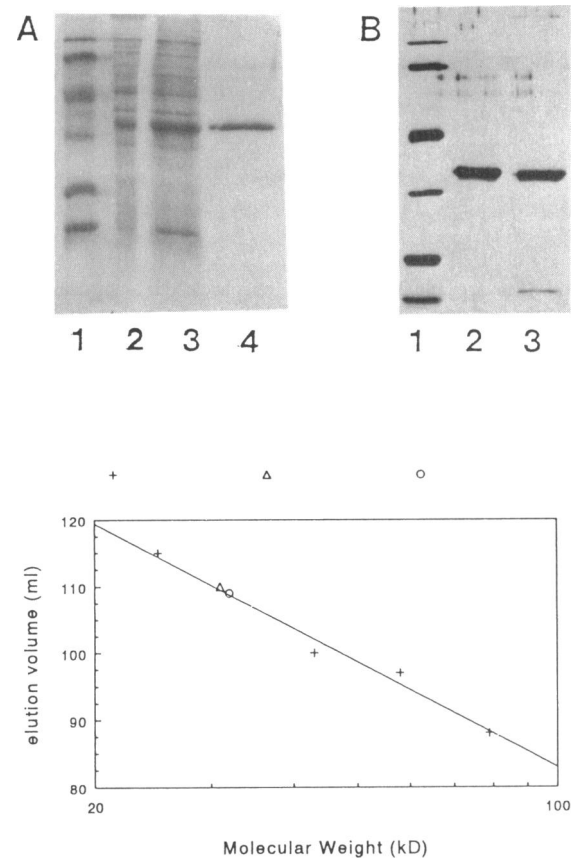

Figure 1. HIV integrase proteins. (A) Purification of HIV-2 IN. Total *E. coli* cell lysate (lane 2), 1M NaCl supernatant (lane 3) and purified HIV-2 IN (lane 4). Lane 1 contains marker proteins: rabbit muscle phosphorylase b (97.4 kD), bovine serum albumin (66.2 kD), hen egg white ovalbumin (42.7 kD), bovine carbonic anhydrase (31 kD), soybean trypsin inhibitor (21.5 kD) and hen egg white lysozyme (14.4 kD). The gel was stained with coomassie brilliant blue. (B) Silver stained gel with purified HIV-2 IN (lane 2) and HIV-1 IN (lane 3). Lane 1 contains the same marker proteins as in (A). The faint bands around the position of 60 kD are not proteins in the sample, but contaminants that are present in the sample buffer. (C) Gel exclusion chromatography of IN. Elution volume is plotted against log(MW). The positions of the main peaks of HIV-1 and HIV-2 IN are marked with a triangle and a circle respectively. For HIV-1 IN very faint peaks were detected at the predicted dimer position (elution volume of 90 ml) and in the void volume (60 ml). The pluses indicate proteins with known MW.

oligonucleotide into another (3, 4, 14). Note that integrations are more frequent near the ends of an oligonucleotide than in the central region (figure 2, lanes 6 and 12). Moreover, integration patterns are somewhat different for the various substrates (compare lanes 5, 8 and 11) and they are somewhat different for HIV-1 and HIV-2 IN (compare lanes 5 and 6), indicating that under these conditions integration is non-random. The significance for target preference *in vivo* (34) remains to be determined. Since this is the first reported case of activity of pure HIV IN we can now conclude that IN is the only protein required for HIV integration *in vitro*. We can however not rule out the possibility that host proteins enhance the efficiency of the reaction.

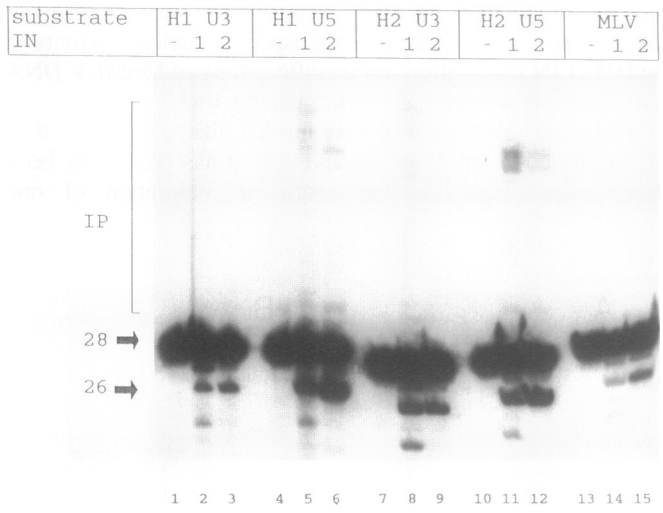

Figure 2. Cleavage and integration of oligonucleotide substrates. Cleavage reactions were performed on double-stranded 28-mer oligonucleotides that represent HIV-1 U3 (lanes 1–3), HIV-1 U5 (lanes 4–6), HIV-2 U3 (lanes 7–9), HIV-2 U5 (lanes 10–12) and MoMLV U5 ends (lanes 13–15). Reactions were carried out without IN (lanes 1, 4, 7, 10 and 13), with HIV-1 IN (lanes 2, 5, 8, 11 and 14) or with HIV-2 IN (lanes 3, 6, 9, 12 and 15). The samples were separated on a 12% denaturing polyacrylamide gel. A protein preparation was also purified from cells that harbor only the expression vector pET-3c, in the same way as HIV-1 IN. The same reactions were carried out with this protein preparation. It did not contain nuclease activity (results not shown). The arrows indicate the positions of the uncleaved (28) and cleaved (26) oligonucleotides, IP = integration products.

In the standard cleavage reaction with HIV-2 IN the protein to DNA ratio is 15 : 1 on a molar basis. Under these conditions only 30–50% of the substrate was cleaved. Possibly the reaction is very slow. We therefore investigated the kinetics of the cleavage reaction. As shown in figure 3A the reaction rate was constant for approximately 20 minutes and thereafter decreased. Between one and two hours after start of the reaction some additional cleavage occurred, but the amount of cleaved product did not increase significantly at longer reaction times (results not shown). With a four-fold lower substrate concentration a comparable curve was obtained for the time course (figure 3A), showing that the decrease in reaction rate is independent of the concentration of substrate or product. The level of cleavage was linear with IN concentration (figure 3B) and not with substrate concentration (figure 3C), supporting the conclusion that the amount of IN is rate limiting in this assay. The incomplete conversion of substrate to cut product can be explained by slow inactivation, probably due to aggregation and precipitation of IN (see below).

Competition of the cleavage reaction

The cutting and integration reaction is sequence specific: only linear DNA with a specific viral sequence at its end can be cut and integrated. This implies that IN specifically recognizes these sequences. We investigated whether this specific recognition could be outcompeted by an excess of DNA of aspecific as well as specific sequence. As shown in figure 4A (lanes 3–6) the cleavage reaction was strongly competed by oligonucleotides that mimic the HIV-1 U5 as well as U3 DNA ends. Surprisingly, MoMLV DNA ends, and even a double-stranded 28-mer oligonucleotide of a sequence unrelated to viral DNA ends, also competed in the cleavage reactions (figure 4A lanes 7–10). The aspecific oligonucleotides competed approximately three-fold less strongly than specific competitors (as determined by counting radioactivity in excised gelslices).

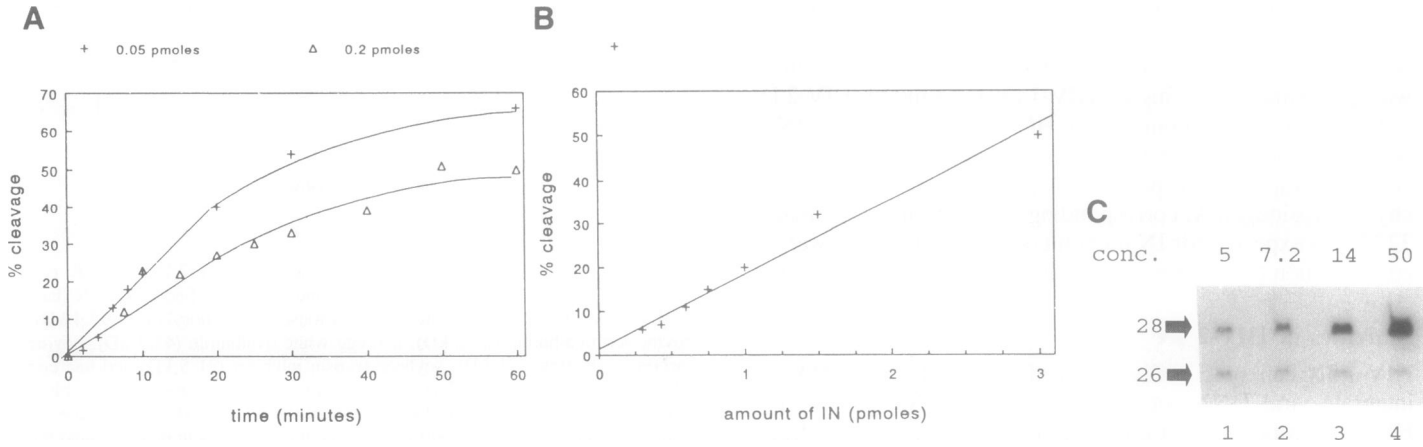

Figure 3. The cleavage reaction. (A) Timecourse of the cleavage reaction. The amount of cleaved product (as percentage of the total amount of substrate) is plotted against incubation time for the reaction with 0.05 pmoles (pluses) and 0.2 pmoles of HIV-1 U5 substrate per 10 μl (triangles). (B) 0.02 pmoles of U5 substrate per μl was incubated for one hour with HIV-2 IN. The amount of cleaved product (as percentage of the total amount of substrate) is plotted against the IN concentration. (C) 3 pmoles of HIV-2 IN were incubated with 50 (lane 1), 72 (lane 2), 140 (lane 3), and 500 (lane 4) fmoles of HIV-1 U5 substrate. The samples were separated on a 12% denaturing polyacrylamide gel. Above the lanes the substrate concentration in fmoles per μl is indicated.

Since efficient cleavage requires that the specific viral DNA sequence is present near the end of linear DNA (17), we tested whether DNA ends are required for competition in the cutting reaction. As shown in figure 4A (lanes 11) this is not the case: circular pBR322 competes at least as strongly as oligonucleotides (when the same mass of DNA is added). This shows that large circular as well as small linear double-stranded DNA can inhibit the cleavage reaction. We tested whether DNA needs to be double-stranded. As shown in figure 4B single-stranded DNA did not compete as efficiently as double-stranded DNA. To obtain the same level of competition approximately 15 times more single-stranded than double-stranded DNA is needed (data not shown). tRNA could also inhibit the cleavage reaction. It competed approximately as efficiently as double-stranded DNA (figure 4A lanes 13 and 14). We conclude, that cleavage by HIV IN can be competed by aspecific double stranded DNA and by tRNA, and (much less efficiently) by single-stranded DNA. How can aspecific nucleic acids prevent sequence specific cutting by IN? An explanation could be that IN quickly binds to nucleic acids without much sequence specificity, and then (slowly) finds and cuts specific viral sequences. To investigate this further we studied binding of HIV IN to DNA.

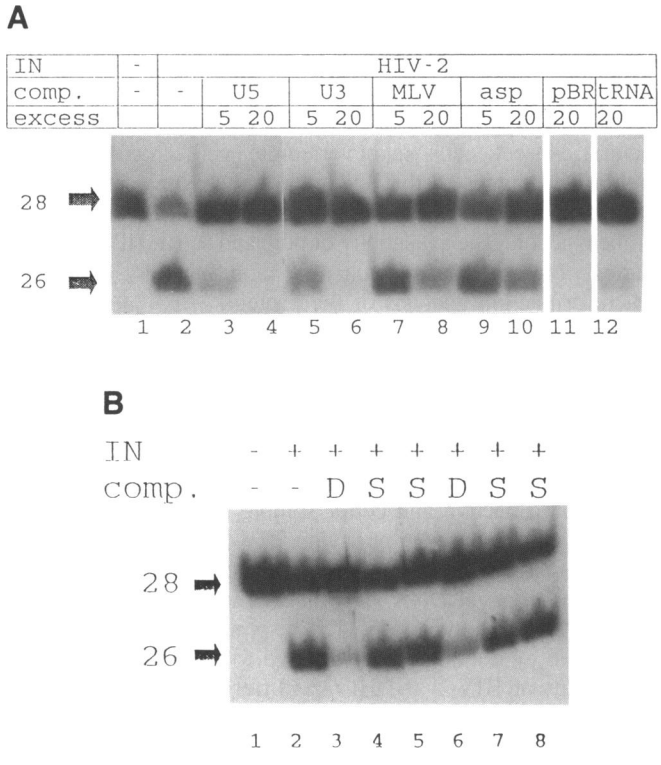

Figure 4. Competition of the cleavage reaction. (A) 0.05 pmoles of HIV-1 U5 substrate was incubated without IN (lane 1), with 3 pmoles of HIV-2 IN (lane 2) or with 3 pmoles of HIV-2 IN and as competitor DNA oligonucleotides that mimic the HIV-1 U5 ends (lanes 3 and 4), the HIV-1 U3 ends (lanes 5 and 6), or the MoMLV U5 ends (lanes 7 and 8), aspecific double-stranded 28-mer oligonucleotides (lanes 9 and 10), pBR322 (lane 11), or yeast tRNA (lane 12). Competition experiments were done with a 5-fold (lanes 3, 5, 7 and 9) or a 20-fold (lanes 4, 6, 8 and 10–12) mass excess of competitor DNA or RNA. (B) Reactions as in (A) without IN (lane 1), with HIV-2 IN (lane 2), or with HIV-2 IN and a 20-fold mass excess of double-stranded U5 substrate (lane 3), single strands of the U5 substrate (lanes 4 and 5), double-stranded aspecific 28-mer oligonucleotide (lane 6), or single strands of the aspecific 28-mer (lanes 7 and 8). All competitor preparations were heated to 80°C and allowed to anneal by slow cooling.

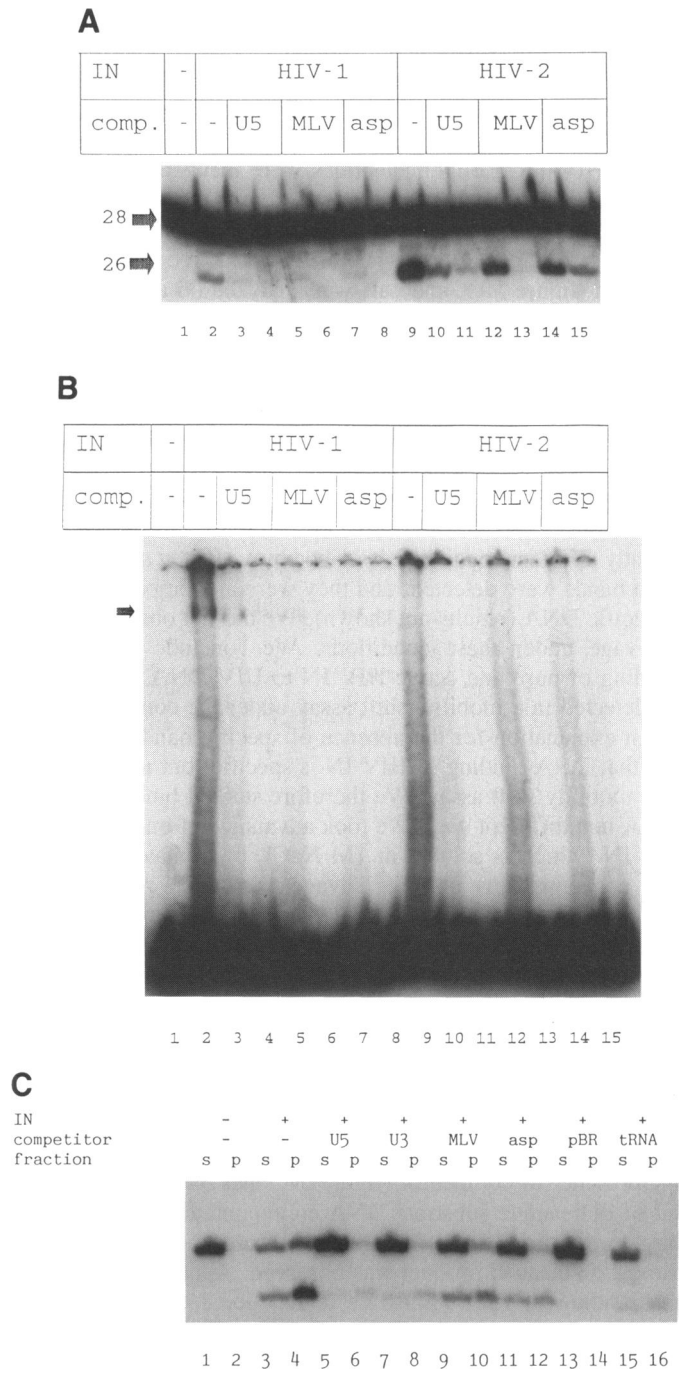

Figure 5. DNA binding of HIV-1 and HIV-2 IN. (A) Analysis of cleavage of oligonucleotides representing the HIV-1 U5 end. 0.2 pmoles of the HIV-1 U5 substrate was incubated for 15 minutes under cleavage conditions without IN (lane 1), with 2 pmol HIV-1 IN (lanes 2–8), or with 6 pmoles HIV-2 IN (lanes 9–15) in 20 μl. The reaction mixtures were incubated for 15 minutes at 30°C without competitor (lanes 1, 2 and 9), or with a 20-fold (lanes 3, 5, 7, 10, 12 and 14) or a 100-fold (lanes 4, 6, 8, 11, 13, 15) excess of HIV-1 U5 substrate (lanes 3, 4, 10 and 11), MoMLV U5 substrate (lanes 5, 6, 12 and 13), or an aspecific double-stranded 28-mer oligonucleotide (lanes 7, 8, 14 and 15). Samples were separated on a 12% denaturing polyacrylamide gel. (B) Mobility shift gel with the same reaction mixtures. Lanes 1–15 correspond to lanes 1–15 in (A). (C) 0.05 pmoles of HIV-1 U5 substrate were incubated without IN (lanes 1 and 2), with HIV-2 IN (lanes 3 and 4), with HIV-2 IN and a 20-fold mass excess of HIV-1 U5 ends (lanes 5 and 6), HIV-1 U3 ends (lanes 7 and 8), MoMLV U5 ends (lanes 9 and 10), aspecific double-stranded 28-mer oligonucleotides (lanes 11 and 12), pBR322 (lanes 13 and 14), or yeast tRNA (lanes 15 and 16). After one hour incubation, the reactions were centrifuged for 30 minutes at 10,000×*g*. Lanes labeled 'S' contain supernatants and lanes labeled 'P' contain pellets.

Lack of specific binding of HIV-1 and HIV-2 IN to HIV DNA ends

For the IN proteins of MoMLV and HIV-1 specific binding to viral DNA ends was reported (26–29). We investigated binding of HIV-1 and HIV-2 IN to the HIV-1 U5 end in a mobility shift assay. HIV-1 U5 oligonucleotides were incubated with HIV-1 or HIV-2 IN under the conditions used in the cleavage reaction; part of the reaction mixture was analyzed on a denaturing polyacrylamide gel to test whether IN was active in the cleavage reaction (figure 5A), and part was analyzed on a mobility shift gel (figure 5B). Although we observed efficient cleavage, only minor shifted bands were detected. Those bands were competed equally well by both viral and aspecific DNA (figure 5B, lanes 3–8). As shown in figure 5B (lane 9) HIV-2 IN does not produce a strong band shift even in the absence of competitor; HIV-1 IN, which contains some impurities, shows a stronger aspecific band shift, and we conclude this is probably caused by contaminants. We also performed mobility shift assays under the slightly different conditions described by Leavitt *et al.* (29); again faint bands were detected, and they were all competed away by aspecific DNA (results not shown). We did not observe specific cleavage under these conditions. We conclude that specific binding of pure and active HIV IN to HIV DNA ends can not be detected in a mobility shift assay under the conditions used.

An explanation for the absence of specific band shifts might be, that DNA binding of HIV IN is specific, but not detectable in a mobility shift assay. We therefore studied binding of IN to DNA in a different way. We took advantage of our observation that IN, which is soluble in 1M NaCl, precipitates at the low salt concentrations in the cleavage/integration reaction. The reaction mixture was centrifuged after incubation and the fraction of radioactive DNA that coprecipitated with IN was determined. Without competition more than 50% of the DNA coprecipitated, whereas in the reactions with a 20-fold excess of specific as well as aspecific competitor hardly any radioactive DNA was found in the pellet (figure 5C). The ratio of cleaved product to uncleaved substrate was not similar in pellet and supernatant: the pellet contained a larger proportion of cleaved product than the supernatant (figure 5C lanes 3 and 4). This effect is stronger in the presence of competitor (figure 5C lanes 5–16): only a little bit of radioactive substrate DNA coprecipitated with IN, but a large proportion of that was cleaved, whereas in the supernatant almost no cleaved product was observed. Apparently, although the binding of IN to substrate is functional, nevertheless the binding is outcompeted by aspecific DNA. This provides a third confirmation that IN does not bind with strong preference to viral DNA termini.

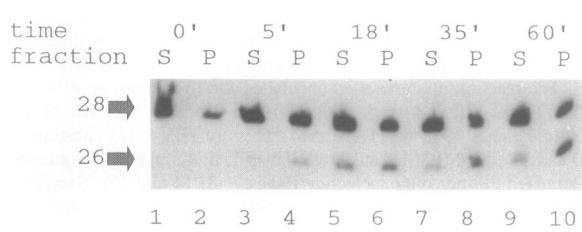

Figure 6. Kinetics of binding. 0.05 pmoles of HIV-1 U5 substrate were incubated with 3 pmoles of HIV-2 IN for 0 (lanes 1 and 2), 5 (lanes 3 and 4), 18 (lanes 5 and 6), 35 (lanes 7 and 8) or 60 minutes (lanes 9 and 10). Reaction mixtures were centrifuged for 5 minutes at 10,000×*g*. Lanes labeled 'S' contain supernatants and lanes labeled 'P' contain pellets. Samples were separated on a 12% denaturing polyacrylamide gel.

We determined how fast IN binds to DNA. In a time course of the precipitation reaction we found, that already after 5 minutes a large fraction of the DNA coprecipitated. The amount of coprecipitated DNA did not increase much in the next 55 minutes (figure 6). However, the amount of cleaved product increased in time, especially in the precipitate. We interpret these results as follows: as shown above most IN protein binds to DNA with little sequence specificity within a few minutes. This is probably an irreversible reaction under these reaction conditions. IN quickly forms aggregates that can be pelleted by centrifugation, but nevertheless it remains active for over an hour in cutting and integration of viral DNA sequences that are bound to it. This hypothesis would imply that competitor DNA only competes in the cleavage reaction if it is added before all active IN is bound to DNA. The cleavage reaction is indeed hardly affected by addition of specific as well as aspecific competitor 5 minutes after the addition of substrate (data not shown), whereas there is a strong effect when competitor is added simultaneously with substrate and an even stronger effect when the competitor is added before the substrate. We conclude, that IN rapidly binds to DNA without much specificity and that subsequently the viral DNA ends are slowly and specifically cleaved.

DISCUSSION

We overexpressed the IN proteins of HIV-1 and HIV-2 in *E. coli* and purified them. The IN proteins can mediate both specific cleavage of HIV DNA ends and integration of viral DNA. The fact that IN, partially purified from various sources such as bacteria, insect cells and vertebrate cells, could mediate integration, suggested that IN is the only protein required for integration. We show here that indeed pure HIV-2 IN is sufficient for the reaction. The availability of pure and active HIV IN allows studies of the reaction mechanism and kinetics, as well as structure/function analysis of the protein.

We have not yet been able to determine whether the donor cleavage and integration reactions are enzymatic, since the reaction reaches a plateau long before one substrate molecule per IN molecule has been converted into product, probably as a result of aggregation of IN. This problem might be solved by cross-linking IN to a matrix to stabilize it or by making a fusion protein that is active and better soluble than IN.

One of our purposes in purifying HIV-2 as well as HIV-1 IN was to investigate a possible preference of each protein for the corresponding viral DNA substrates. This could have allowed mapping of the domain of IN responsible for viral DNA recognition by domain swapping. We found however, that the IN proteins of HIV-1 and HIV-2 (differing 43% in sequence) act on both HIV-1 and HIV-2 DNA with comparable efficiency. We even observed a low level of specific cutting of a MoMLV substrate. The corresponding opposite observation has also been made: MoMLV IN integrates DNA substrates with HIV DNA ends, albeit with low efficiency (13, 18). From the results presented here we conclude that IN can recognize a wide range of substrates with similarity to viral DNA ends. An extensive analysis of HIV-1 IN substrate requirements by mutagenesis of viral DNA sequences resulted in a similar conclusion (17).

We found no specific DNA binding of IN in mobility shift assays, nor did we see any specificity if we isolated aggregated IN by centrifugation and determined whether viral DNA was bound preferentially. In the latter case we know the interaction between IN and substrate was productive, because a large

proportion of the substrate DNA that was bound was cut within an hour; nevertheless this binding of substrate to IN could be competed out by aspecific DNA. Together with the observation that addition of competitor DNA after a short preincubation with the labeled DNA hardly affects the extent of cleavage, this leads to the following picture of IN action *in vitro*: in reaction buffer (under near physiological salt concentrations) IN quickly binds to nucleic acids with little sequence specificity (and also aggregates). Once bound to DNA, IN can slowly cut a specific viral sequence. It probably remains bound to the cleaved DNA and can mediate integration into target DNA. If IN does not find a specific site it also remains bound. This picture of IN action *in vitro* probably does not reflect the mechanism of recognition of viral DNA *in vivo*: in the infected cell IN is in constant proximity of HIV RNA, or, after reverse transcription, of HIV DNA, in the viral core particle. This suggests that IN is probably never free in solution; it may be bound first to viral RNA and later to viral DNA, which can explain why IN does not need to be very soluble under physiological salt concentrations. The presence of IN in core particles can also explain its loose substrate specificity. IN is probably continuously present near the viral DNA, and it only needs to recognize the specific sequence for precise positioning. This requires a specificity of another order of magnitude than e.g. the specific recognition of enhancers in a cellular genome by transcription factors.

In light of these results it is not easy to understand the observations of others who reported specific binding by recombinant MoMLV or HIV IN (27, 29). The HIV IN preparations that showed specific DNA binding were not pure, and no specific cutting or integration by these protein preparations was reported. Using pure and active HIV IN we find competition of any binding by aspecific DNA, and this is consistent with our observations that also cutting and integration of specific substrates is competed out by aspecific DNA. Whatever the explanation is for the band shifts observed by others, specific DNA binding does not seem to be an intrinsic property of active HIV IN.

ACKNOWLEDGEMENTS

We want to thank Hanneke van der Gulden and Ellen Wientjens for synthesizing the oligonucleotide substrates. We thank Piet Borst, Chris Vos, Hans Westerhoff and Joost Zomerdijk for critical comments on the manuscript. We thank Michael Emerman (Institut Pasteur, Paris) for the kind gift of plasmid pROD 4-8. This work was supported by grant 88033 from the Dutch Ministry of Health (RGO) and grant 900-502-119 of the Netherlands Organization for Scientific Research (NWO), and a NWO PIONIER grant to RHAP.

REFERENCES

1. Guyader, M., Emerman, M., Sonigo, P., Clavel, F., Montagnier, L. and Alizon, M. (1987) *Nature*, **326**, 662–669.
2. Varmus, H.E. and Brown, P. (1989) In D.E. Berg and M.M. Howe (eds.), Mobile DNA. American Society for Microbiology, Washington, DC.
3. Craigie, R., Fujiwara, T. and Bushman, F D. (1990) *Cell*, **62**, 829–837.
4. Katz, R.A., Merkel, G., Kulkosky, J., Leis, J. and Skalka. A.M. (1990) *Cell*, **63**, 87–95.
5. Roth, M.J., Schwartzberg, P.L. and Goff, S.P. (1989) *Cell* **58**, 47–54.
6. Sherman, P.A. and Fyfe, J.A. (1990) *Proc. Natl. Acad. Sci. USA*, **87**, 5119–5123.
7. Brown, P.O., Bowerman, B., Varmus, H.E. and Bishop, J.M. (1987) *Cell*, **49**, 347–356.
8. Fujiwara, T. and Mizuuchi, K. (1988) *Cell*, **54**, 497–504.
9. Donehower, L.A. (1988) *J. Virol.*, **62**, 3958–3964.
10. Schwartzberg, P.J., Collicelli, J. and Goff, S.P. (1984) *Cell*, **37**, 1043–1052.
11. Katzman, M., Katz, R.A., Skalka, A.M. and Leis, J. (1989) *J. Virol.*, **63**,5319–5327.
12. Vora, A.C., Fitzgerald, M.L. and Grandgenett, D.P. (1990) *J. Virol.*, **64**, 5656–5659.
13. Bushman, F.D. and Craigie, R. (1990) *J. Virol.*, **64**, 5645–5648.
14. Bushman, F.D. and Craigie, R. (1991) *Proc. Natl. Acad. Sci. USA*, **88**, 1339–1343.
15. Bushman, F.D., Fujiwara, T. and Craigie, R. (1990) *Science*, **249**, 1555–1558.
16. Craigie, R., Mizuuchi, K., Bushman, F.D. and Engelman, A. (1991) *Nucleic Acids Res.*, in press.
17. Vink, C., van Gent, D.C., Elgersma, Y. and Plasterk, R.H.A. (1991) *J. Virol.*, in press.
18. Vink, C., van Gent, D.C. and Plasterk, R.H.A. (1990) *J. Virol.*, **64**, 5219–5222.
19. Bowerman, B., Brown, P.O., Bishop, J.M. and Varmus, H.E. (1989) *Genes Dev.*, **3**, 469–478.
20. Farnet, C.M. and Haseltine, W.A. (1991) *J. Virol.*, **65**, 1910–1915.
21. Hizi, A. and Hughes, S.H. (1988) *Virology*, **167**, 634–638.
22. Luk, K.C., Gilmore, T.D. and Panganiban, A.T. (1987) *Virology*, **157**, 127–136.
23. Misra, T.K., Grandgenett, D.P. and Parsons, J.T. (1982) *J. Virol.*, **44**, 330–343.
24. Roth, M.J., Tanese, N. and Goff, S.P. (1988) *J. Mol. Biol.*, **203**, 131–139.
25. Terry, R., Soltis, D.A., Katzman, M., Cobrinik, D., Leis, J. and Skalka, A.M. (1988) *J. Virol.*, **62**, 2358–2365.
26. Krogstad, P.A. and Champoux, J.J. (1990) *J. Virol.*, **64**, 2796–2801.
27. Basu, S., and Varmus, H.E. (1990) *J. Virol.*, **64**, 5617–5625.
28. Ishimoto, L.K., Halperin, M. and Champoux, J.J. (1991) *Virology*, **180**, 527–534.
29. Leavitt, A.D., Barr, P.J., Feucht, P.H., Bathurst, I.C. and Varmus, H.E. (1991) *Virology*, in press.
30. Sambrook, J., Fritsch, E.F. and Maniatis, T. (1989) Molecular Cloning: A Laboratory Manual. Cold Spring Harbor University Press, Cold Spring Harbor, NY.
31. Rosenberg, A.H., Lade, B.N., Chui, D., Lin, S., Dunn, J.J. and Studier, F.W.(1987) *Gene*, **56**, 125–135.
32. Ratner, L., Haseltine, R., Patarca, R., Livac, K.J., Starcich, B., Josephs, S.F., Doran, E.R., Rafalski, J.A., Whitehorn, E.A., Baumeister, K., Ivanoff, L., Petteway Jr., S.R., Pearson, M.L., Lautenberger, J.A., Papas, T.S., Ghrayeb, J., Chang, N.T., Gallo, R.C. and Wong-Staal,F. (1985) *Nature*, **313**, 277–284.
33. Shinnick, T.M., Lerner, R.A. and Sutcliffe, J.G. (1981) *Nature*, **293**, 543–548.
34. Shih, C.C., Stoye, J.P. and Coffin, J.M. (1988) *Cell*, **53**, 531–537.