2202–2210 *Nucleic Acids Research, 1999, Vol. 27, No. 10*

Activity of recombinant HIV-1 integrase on mini-HIV DNA

Peter Cherepanov, David Surratt[1], Jaan Toelen, Wim Pluymers, Jack Griffith[1], Erik De Clercq and Zeger Debyser*

Rega Institute for Medical Research, Minderbroedersstraat 10, B-3000 Leuven, Belgium and [1]The Lineberger Cancer Center, University of North Carolina, Chapel Hill, NC 27599-7295, USA

Received December 31, 1998; Revised and Accepted March 24, 1999

ABSTRACT

Integration of the human immunodeficiency virus type 1 (HIV-1) cDNA into the genome of a human cell is an essential step in the viral replication cycle. Understanding of the integration process has been facilitated by the development of *in vitro* assays using specific oligonucleotides and recombinant integrase. However, understanding of the biology of retroviral integration will require *in vitro* and *in vivo* model systems using long DNA substrates that mimic the HIV cDNA. We have now studied the activity of recombinant HIV-1 integrase on a linear 4.7 kb double-stranded DNA, containing flanking regions of ~200 bp that represent the intact ends of the HIV-1 long terminal repeat (LTR) sequences (mini-HIV). The strand transfer products of the integration reaction can be directly visualized after separation in agarose gels by ethidium bromide staining. The most prominent reaction product resulted from integration of one LTR end into another LTR end (U5 into U5 and U5 into U3). Sequence analysis of the reaction products showed them to be products of legitimate integration preceded by correct processing of the viral LTR ends. Hotspots for integration were detected. Electron microscopy revealed the presence of a range of reaction products resulting from single or multiple integration events. The binding of HIV-1 integrase to mini-HIV DNA was visualized. Oligomers of integrase seem to induce DNA looping whereby the enzyme often appears to be bound to the DNA substrate that adopts the structure of a three-site synapsis that is reminiscent of the Mu phage transposase complex.

INTRODUCTION

Integration of the retroviral genome into the host chromosome is a complex process that requires a number of viral and cellular proteins necessary to protect and transport the large viral cDNA molecule from the cytoplasm into the nucleus (1,2). It involves processing of the viral DNA termini followed by concerted insertion of both long terminal repeat (LTR) ends into the chromosome and a DNA repair step to fill in the remaining single-stranded DNA gaps. While numerous cellular and viral factors are probably required, the retroviral integrase nevertheless plays a pivotal role in this process. Many attempts have been undertaken to mimic the coordinated integration event using *in vitro* assay systems and to search for potential integrase inhibitors.

Retroviral integration was extensively studied in assays that made use of short duplex oligonucleotides (18–35 bp) mimicking the U5 or U3 LTRs. The HIV-1 integrase is known to possess three specific enzymatic activities *in vitro*: (i) the 3′ processing activity that removes the 3′ GT dinucleotide from the LTR end; (ii) the integration or DNA strand transfer activity which is accomplished through a nucleophilic attack of the free 3′ hydroxyl group of the processed LTR end on a phosphodiester bond in the target DNA; (iii) the disintegration reaction that is essentially a reversal of integration (reviewed in 2). DNA strand transfer *in vitro* occurs with low efficiency and can only be detected by radiolabeling the DNA substrates. With recombinant HIV-1 integrase the concerted integration of both LTR ends could not be demonstrated. Oligonucleotide-based assays have been instrumental in screening HIV-1 integrase inhibitors (3).

Another approach to study integration *in vitro* consists of isolating subviral particles, which contain integrase and viral cDNA, from acutely infected cells. When providing these pre-integration complexes (PICs) with double-stranded DNA (dsDNA), both ends of the viral cDNA will be integrated with high efficiency and in a concerted fashion (4). Integration is detected by Southern blotting with probes detecting specific viral DNA. Apart from integrase and reverse transcriptase, HIV-1 PIC particles also contain the viral matrix (MA) and Vpr proteins. Nuclear localization signals within MA, Vpr and integrase proteins are involved in targeting the PIC to the nucleus of the infected cell (5–7). Interestingly, it has been shown that a cellular host factor belonging to the high mobility group of proteins, termed HMG I(Y), is present in HIV-1 PICs and appears to be essential for *in vitro* integration (8). Various HIV-1 integrase inhibitors were shown to inhibit the enzymatic activity of PICs to a lesser extent than in oligonucleotide assays (9). The PIC assays seem to be more relevant for evaluating integration inhibitors.

Complete understanding of the enzymology of HIV-1 integration will require the full reconstruction of the integration process using purified components. An important feature of these integration assays is the design of the DNA substrate. Ideally, this substrate mimics HIV cDNA. Long DNA molecules flanked by LTR fragments have been used for this purpose. These mini-retroviral DNA substrates have been used to study retroviral integration *in vitro* with avian myeloblastosis virus (AMV) integrase (10) or

*To whom correspondence should be addressed. Tel: +32 16 33 21 60; Fax: +32 16 33 21 31; Email: zeger.debyser@uz.kuleuven.ac.be

HIV integrase (11,12). In the HIV integration system a genetic test was used to detect a minority of concerted integration products; the biochemical details of this reaction have not been reported so far.

We envisage the development of an intracellular integration system in which intracellularly expressed HIV-1 integrase interacts with a mini-HIV DNA substrate. As a first step, we studied the interaction of recombinant HIV-1 integrase with a 4.7 kb linear DNA containing the terminal fragments of the HIV-1 LTRs. We describe an efficient *in vitro* mini-HIV integration reaction that allows detection of the strand transfer products in non-denaturing agarose gels by ethidium bromide staining. We optimized the conditions for the assay and characterized some of the reaction products using restriction digestion, sequence analysis and electron microscopy (EM). The majority of reaction products resulted from integration of one LTR terminus into the LTR region of another mini-HIV molecule. Binding of integrase to mini-HIV DNA produced a typical DNA looping pattern that is reminiscent of the Mu phage transposase complex (13).

MATERIALS AND METHODS

Expression and purification of recombinant HIV-1 integrase

The enzyme used throughout this study was the His-tagged form of pNL4-3 HIV-1 integrase, prepared as described previously (16); but using the *Escherichia coli* B strain PC1 for protein production. This strain is lacking Endo I, a non-specific endonuclease for duplex DNA (14). It was obtained by P1 phage-mediated transduction of the Δ*endA*::Tc^R mutation from BT333 (15) into BL21(DE3)(pLysS) (Novagen, Madison, WI).

To remove the His-tag 0.8 mg of the purified protein was incubated overnight with 4 U of human restriction grade thrombin (Novagen) in 750 mM NaCl, 20 mM Tris–HCl pH 7.4, 7.5 mM CHAPS, 5 mM β-mercaptoethanol, 10% glycerol at 22°C. Cleavage was complete as determined by SDS–PAGE. Phenylmethylsulfonyl fluoride (PMSF) was added to 0.1 mM and integrase was further purified by chromatography on heparin–Sepharose (Pharmacia Biotech, Uppsala, Sweden).

Construction of the DNA substrates

The mini-HIV DNA was prepared by linearization of the plasmid pU3U5 with *Sca*I. This plasmid contains the intact HIV-1 LTR U3 and U5 terminal sequences joined to produce a unique *Sca*I site and cloned between *Eco*RI and *Kpn*I sites of the MCS of pBKRSV (Stratagene, La Jolla, CA). The DNA fragment carrying the terminal 175 bp from U3 was obtained using PCR with the primers U3A (5′-CGGAATTCTATCTTATCTGGCT) and U3B (5′-GGGAGTACTGGAAGGGCT AATTCACTC) on the template of the λHXB2 proviral clone (EMBL accession no. K03455). The terminal 191 bp from R-U5 were amplified in a PCR using the primers U5A (5′-GCGGTACCTGTAC TGGGTCTCTCTG-GTTAG) and U5B (5′-TTTAGTACTGCTAGAGATTTTCC ACACTGAC). The PCR fragments carrying U3 or U5 were digested with *Sca*I+*Eco*RI or *Sca*I+*Kpn*I, respectively, and inserted between the *Eco*RI and *Kpn*I sites of pBKRSV producing pU3U5 (4751 bp). Sequencing of the cloned regions in pU3U5 confirmed the absence of mutations.

The *Eco*RI–*Sca*I fragment of pU3U5 containing U3 was cloned into *Eco*RI and *Sca*I sites of pBKRSV, giving the plasmid pU3, which upon digestion with *Sca*I yielded a linear DNA molecule carrying only a single U3 sequence at one end. Cloning the *Kpn*I–*Sca*I fragment from pU3U5 into the *Kpn*I and *Sca*I sites of pBKRSV produced the plasmid pU5, which was a source of linear DNA having only a single U5 sequence at one terminus. These constructs, along with *Sca*I-digested pBKRSV, substituted for the mini-HIV substrate in integration reactions where indicated.

To label the ends of the mini-HIV DNA radioactively, *Sca*I-cleaved pU3U5 DNA was treated with ExoIII nuclease for exactly 1.5 min at 37°C. The resulting 5′ overhangs of approximately 350 bases were filled-in using T7 DNA polymerase (Sequenase v.2.0) in the presence of 10 μM dATP, dTTP and dGTP and 0.3 μM [α-^{32}P]dCTP. To remove possible 3′ A overhangs, the ends of the labeled mini-HIV substrate were then polished using *Pfu* DNA polymerase in the presense of 250 μM each dNTP at 72°C.

In vitro integration assay with the mini-HIV substrate

The typical reaction mixture (20 μl) contained 10 mM $MgCl_2$, 75 mM NaCl, 5 mM DTT, 10 mM HEPES, pH 7.6, 10% PEG-8000, 600 ng of the mini-HIV DNA (9.6 nM) and 150 ng integrase (210 nM). The quality of PEG used was important. Stock solutions of PEG had to be kept frozen; when stored at room temperature the stimulatory effect on integrase activity was lost. After addition of the enzyme, the samples were left for 5 min on ice, followed by incubation at 37°C for 60–90 min. Reactions were stopped by addition of SDS to 0.5%. The DNA was precipitated with ethanol, redissolved in the gel loading buffer (7% sucrose, 50 mM NaCl, 20 mM EDTA, 3 mM Tris, pH 7.8) and incubated at 72°C for 10 min to disrupt any possible secondary structures. After separation by electrophoresis in 0.8% agarose gels at 2 V/cm for 7 h in Tris/actetate/EDTA buffer the gels were stained with ethidium bromide. For publication, the images of the ethidium bromide stained gels were captured using the ImageMaster system (Pharmacia Biotech) connected to a MacIntosh computer and contrast adjusted with the NIH Image software (National Institutes of Health, Bethesda, MD).

To assess the 3′ processing activity, radioactively labeled reaction products were digested with *Hin*dIII or *Eco*RV restriction endonucleases and the resulting fragments were separated by electrophoresis in a denaturing polyacrylamide gel. The products of the ddNTP sequencing reactions performed on the pU3U5 template with the two primers U3-Eco (5′-ATCTTGTCTTCTTTGGG-AG) and U5-Hind (5′-AGCTTGCCTTGAGTG CTTC) were used as sequence-specific molecular weight markers. Gels were dried and autoradiographed. Quantification of the radioactive products was done using a PhosphoImager (Molecular Dynamics, Sunnyvale, CA).

Restriction analysis of the integration products and Southern blotting

From 0.5 to 7 ng of the gel-purified D product or mini-HIV DNA was digested with restriction enzymes in the presence of 1 μg of phage λ DNA. The restriction fragments were electrophoresed in 1% agarose gels in TAE buffer or in denaturing 6% polyacrylamide gels supplemented with 6 M urea in TBE. Southern blotting was performed using capillary transfer for agarose gels or electroblotting for polyacrylamide gels onto Hybond N^+ nylon membranes (Amersham, Little Chalfont, UK). Hybridization and chemiluminescent detection were done using the Gene Images system (Amersham) according to the manufacturer's instructions. To detect all restriction fragments blotted from agarose gels, a

full-length mini-HIV probe was used. The strand-specific U3 and U5 probes were synthesized using the PCR fluorescein labeling mixture (Boehringer Mannheim) in a linear amplification reaction. To produce the U3-specific probe, a reaction mixture containing 50 ng of the gel-purified U3 fragment, 10 μM primer 5′-ACTGGAAGGGCTAATTCACT and 1 U of AmpliTaq DNA polymerase was subjected to 12 cycles of 1 min at 95°C, 0.5 min at 55°C and 1 min at 72°C. The U5-specific probe was made in a similar way, using the U5 fragment and the primer 5′-ACTGCTAGAGATTTTCCACA. The probes were purified by precipitation with ethanol in the presence of 2 M ammonium acetate.

Amplification, cloning and sequencing of the products from the integration reaction

The PCR reactions contained ~0.5 ng of the template DNA (D-type products isolated from agarose gels), 200 nM each specific primer, 2 mM $MgCl_2$, 300 μM dNTPs and 1 U of AmpliTaq DNA polymerase. The optimized cycling conditions were: 95°C, 30 s; 60°C, 20 s; 72°C, 30 s (30–40 cycles). The PCR products appeared as a smear on non-denaturing gels and apparently consisted of dsDNA species from 400 bp to more than 3 kb; no amplification was detected when the mini-HIV DNA was used as a template. PCR products were digested with *Eco*RI+*Kpn*I, *Eco*RI+*Sca*I or *Kpn*I+*Sca*I depending on the experiment and were cloned into the vector pBluescript KS (+). The individual clones were sequenced on ABI Prism 310 using the Dye Terminator Cycle Sequencing Core kit (Perkin Elmer) with the primers 5′-TGTAAAACGACGGCCAGT (M13-universal) and 5′-CAGGAAACA GCTATGACC (M13-reversed).

Visualization of DNA and DNA–integrase complexes by electron microscopy

Integrase–DNA complexes were formed by incubating 600 ng mini-HIV DNA with 20 nM recombinant integrase in 20 μl of 20 mM HEPES (pH 7.5) buffer containing 10 mM $MgCl_2$, 2 mM DTT and 5–10% PEG. After 5 min on ice the samples were incubated for 5–10 min at 37°C. The complexes were fixed with 0.6% (v/v) glutaraldehyde for 5 min at room temperature, followed by gel filtration through 2 ml columns of Bio-Gel A5m (Bio-Rad, Hercules, CA). To mount the DNA–protein complexes for EM, the samples were mixed with a buffer containing 2 mM spermidine and applied for 30 s to thin carbon foils supported by 400 mesh copper grids. The grids were washed with a water/ethanol series, air dried and rotary shadow cast at 10^{-7} torr with tungsten (17). Samples were examined in a Philips EM400. For publication, images were digitized using a Nikon LS4500 film scanner camera attached to a MacIntosh computer and the contrast adjusted using Adobe Photoshop software.

For analysis of the integration products, integration reactions were performed as described. The products were separated by electrophoresis in low melting point agarose gels, bands of the reaction products were cut out from the gel under UV illumination and the agarose gel was digested using β-agarase enzyme and DNA was purified by precipitation with ethanol and gel filtration through 2 ml columns of Bio-Gel A5m. The purified integration products were prepared for electron microscopy by the basic protein film technique (18) using cytochrome C and formamide (19).

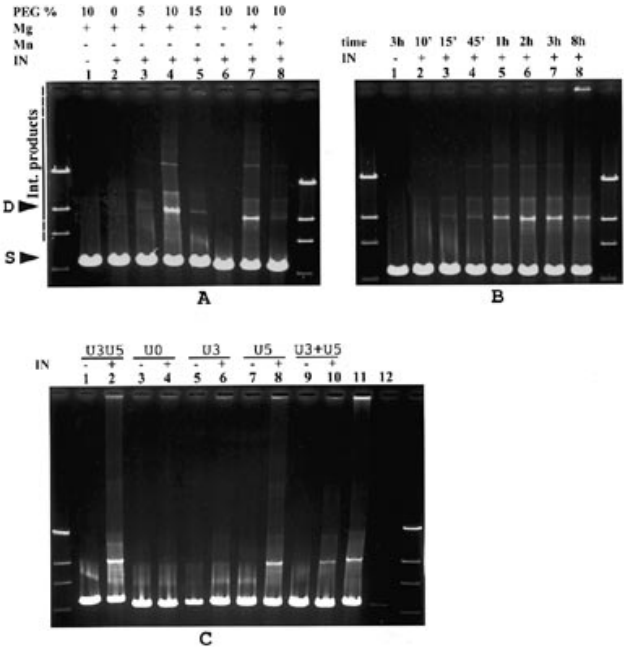

Figure 1. Activity of recombinant HIV-1 integrase on long DNA substrates. Reaction products were separated in agarose gels and stained with ethidium bromide. (**A**) The reactions were carried out with different concentrations of PEG in the presence or absence of 10 mM Mg^{2+} or Mn^{2+} as indicated. (**B**) Time course of the reaction under standard conditions. (**C**) Different substrates were used: U3U5 corresponds to the mini-HIV DNA substrate, containing both U3 and U5 terminal sequences; U0 is the vector pBKRSV, linearized with *Sca*I, containing no HIV LTR sequences; U3 and U5 are substrates carrying a U3 or a U5 sequence at one end; U3+U5 indicates reactions performed with the equimolar mixture of the U3 and U5 substrates. For each substrate a mock reaction without enzyme was performed. Lanes 11 and 12, as in 2, but aggregated (lane 11) and soluble (lane 12) DNA were separated after the reaction by centrifugation of the reaction mixture for 5 min at 20 000 *g*. The first and the last lane from each gel contain a molecular weight marker (λ phage DNA digested with *Hin*dIII; the four bands correspond to 23, 9.4, 6.5 and 4.4 kb). The substrate is referred to as S and the double size reaction product as D.

RESULTS

Optimization of *in vitro* integration reaction conditions using a mini-HIV DNA substrate

We have constructed a linear 4751 bp dsDNA substrate containing 175 and 191 bp of the intact U3 and R-U5 HIV-1 LTR terminal sequences, respectively (Fig. 4B). Under our conditions the enzyme efficiently converted this long mini-HIV DNA into distinct forms of higher molecular weight that could be easily visualized in agarose gels after staining with ethidium bromide (Fig. 1A–C). The optimal reaction conditions required the presence of Mg^{2+} and polyethylene glycol (PEG-8000) (Fig. 1A); PEG displayed a narrow concentration optimum of 10% (lanes 3–5). Mg^{2+} was preferred over Mn^{2+} as a divalent cation in all conditions tested, including different NaCl concentrations (lanes 6–8). Preference for Mg^{2+} has been observed in oligonucleotide integration reactions when longer substrates were used (35 bp) (29). Integrase was equally active over a range of NaCl concentrations between 35 and 145 mM. Higher salt concentrations inhibited the reaction, presumably by reducing the affinity of integrase for DNA (data not shown). Whereas addition of 10% DMSO contributed to the integrase activity in assays using

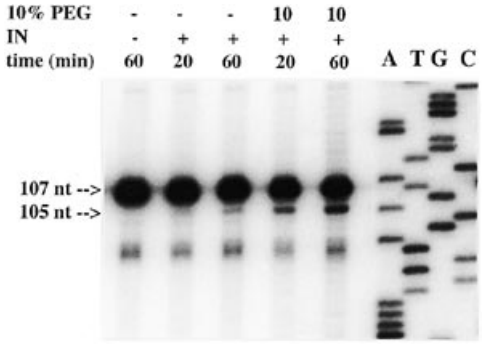

Figure 2. Specific 3′ processing of the U5 end of mini-HIV DNA by integrase. The integration reactions (in the presence or absence of PEG) were performed with the radioactively labeled substrate. The products were then digested with *Hin*dIII, which cuts 107 bp from the 3′-end of the U5 terminus of mini-HIV, and loaded on a 6% denaturing polyacrylamide gel. The sequencing reaction products generated from the primer U5-Hind on the pU3U5 template are used as sequence-specific molecular weight markers. The removal of the 3′ dinucleotide from the mini-HIV U5 end by integrase results in a 105 nt band.

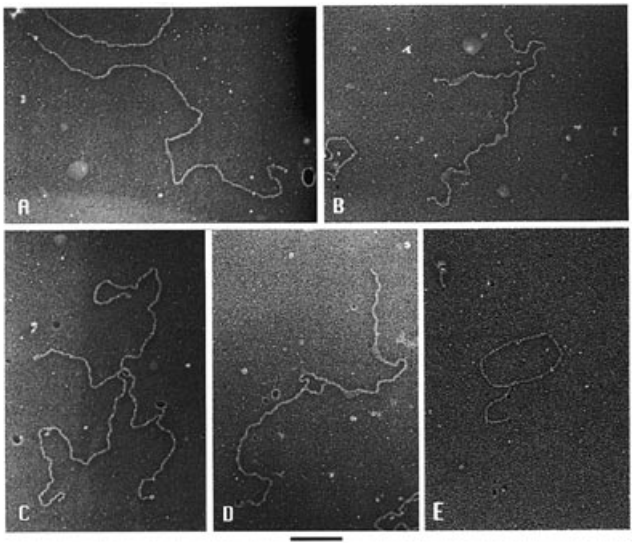

Figure 3. Analysis of the integration products by EM. (**A**) DNA extracted from agarose gel slices containing the D product appears to consist of linear molecules with a size twice that of the mini-HIV molecule. (B–E) Examples of DNA structures observed in the total pool of all integration products purified from the unreacted mini-HIV substrate DNA by agarose gel electrophoresis. (**B**) A Y-shaped molecule resulting from a single integration event. (**C**) Two or three integration events were necessary to produce this structure; it is possible that some branches of this structure were trimmed by non-specific nuclease activity. (**D**) This structure is made of two mini-HIV molecules integrated into each other, creating a bubble. (**E**) Apparently a product of integration of mini-HIV into itself. DNA was prepared for EM using the cytochrome C film method. Shown in reverse contrast. Bar equals a length of DNA equivalent to 1000 ± 50 bp except for (C) where it equals 900 ± 50 bp.

oligonucleotide substrates (21), integrase activity in the mini-HIV assay did not increase in the presence of DMSO. The optimal concentrations of integrase and mini-HIV DNA were found to be 210 nM (150 ng enzyme in a 20 µl reaction) and 9.6 nM, respectively. This may indicate that about 20 monomers of integrase per mini-HIV molecule are required, although the actual concentration of the active enzyme in our reactions was not determined.

Various reaction products could be distinguished (Fig. 1): a prominent band (D) representing DNA molecules with an apparent size twice the length of the mini-HIV DNA (we will refer to this type of integration product as D products), as well as distinct higher molecular weight products and a DNA smear. After prolonged incubation, DNA products too long to enter the gel appear in the wells of the agarose gel (Fig. 1B, lane 8). In general, 10–20% of the substrate DNA was converted into D product after 1.5 h as estimated by quantitative Southern blotting.

The mini-HIV integration reaction was clearly dependent on the presence of the HIV LTR terminal sequences in the substrate DNA. Thus, no reaction product was observed when the *Sca*I-linearized pBK-RSV vector was used instead of mini-HIV DNA (Fig. 1C, lane 4). Some reaction products, particulary the 'D band', were observed when a substrate containing only the U5 sequence on one end was used in the reaction (lane 8). The enzyme activity was negligible with the substrate containing only the U3 terminal sequence (lane 6). From the gel shown in Figure 1C, it seems that the reaction was more efficient when both U3 and U5 termini were supplied *in cis*, whereas mixing molecules containing only U3 and only U5 did not favor higher yields of reaction product (lane 10). At the end of the reaction, both substrate and product DNA could be recovered from the precipitate (lanes 11 and 12). This implies the presence of a high molecular weight reactive integration complex.

To obtain a preparation of integrase with low contaminating endonuclease activity we used an Endo I-deficient *E.coli* strain. Under the optimal conditions for the mini-HIV integration assay the purified protein retained some nicking activity as determined by relaxation of supercoiled plasmid DNA. The nicking activity in the presence of Mg^{2+} was significantly lower than with Mn^{2+} and was stimulated by addition of 10% PEG (data not shown). Integrase without a His-tag was obtained by overnight incubation of the purified His-tagged protein with thrombin at 22°C. This preparation displayed identical activity, producing the same pattern of reaction products in agarose gels (data not shown).

Analysis of the 3′ processing reaction

We investigated the 3′ processing activity of HIV-1 integrase in the mini-HIV DNA reaction. For this we utilized two restriction sites near the ends of the U5 and U3 sequences (see restriction map of mini-HIV DNA in Fig. 4B). Digestion of the DNA with *Eco*RV releases a 37 nt oligonucleotide from the U3 end and digestion with *Hin*dIII releases a 105 nt oligonucleotide from U5. After incubation of mini-HIV DNA, radiolabeled at the 3′-ends, with integrase, shortening of the oligonucleotides by integrase-mediated removal of the pdGpdT dinucleotide from the 3′-ends was detected after separation in a denaturing polyacrylamide gel. In the presence of 10% PEG the processing efficiency after 1 h incubation with integrase was ~5% for both U5 (Fig. 2) and U3 (not shown). In the absence of PEG the terminal nucleotides were removed from only 2% of mini-HIV DNA ends.

Characterization of the reaction products

Visualization of the reaction products by EM. The products of the HIV-1 integration reaction were examined using EM. In Figure 3A–E structures are shown that correspond to the different reaction products obtained in reactions with the mini-HIV

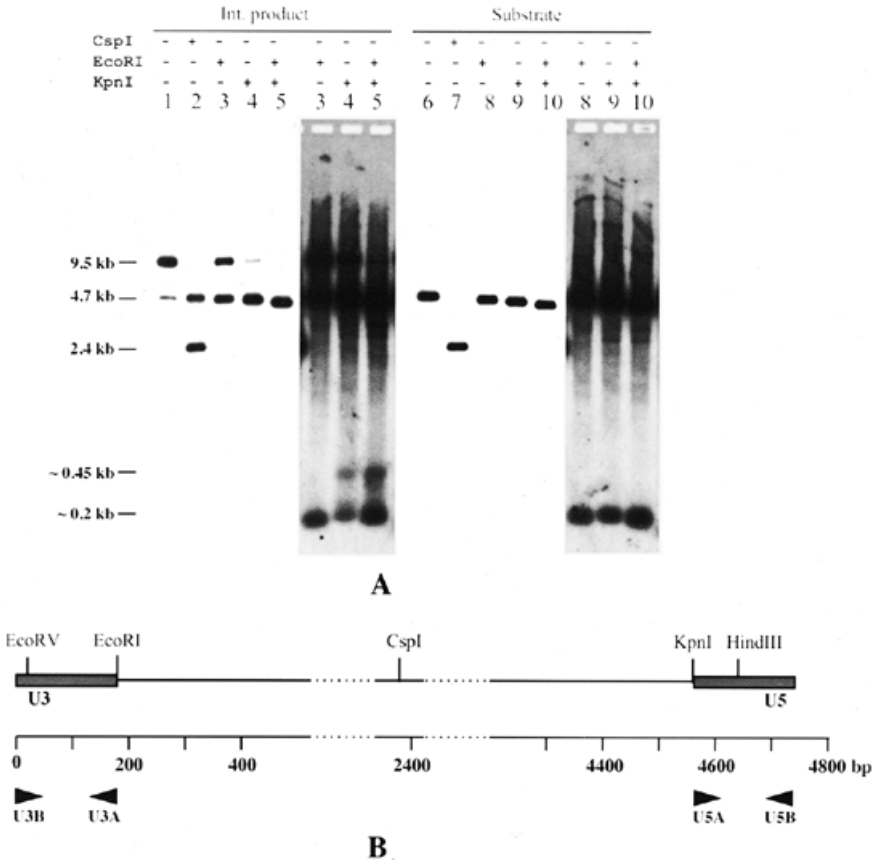

Figure 4. Restriction analysis of the major integration products. (**A**) The integration products with an apparent size twice that of the mini-HIV substrate were isolated from agarose gels and digested with different restriction enzymes. The resulting fragments were separated by electrophoresis in a native 1% agarose gel. The results of the Southern blotting are presented. The minor 4.7 kb band (lane 1) represents traces of mini-HIV substrate contaminating the product DNA. To visualize the shorter DNA fragments in lanes 3–5 and 8–10, longer exposure times were used. For comparison, analogous digests of the original mini-HIV DNA substrate are shown. (**B**) The restriction map of the mini-HIV DNA. The arrowheads indicate positions and directions of the annealing sites for the primers used for the PCR amplification and cloning of the U3 and U5 regions and integration products.

substrate. The prominent 'D band' consists of linear molecules with a length twice that of the mini-HIV DNA (Fig. 3A), as expected from the electrophoretic mobility. These observations do not explain how these molecules are formed. Indeed, a concerted integration event of both LTR ends of one mini-HIV molecule into another would result in a linear 9.5 kb DNA product. However, since the resolution of the cytochrome C film method used in these EM experiments is 100–150 bp, short branches or loops would not be detected. Therefore, these molecules could also result from one-sided integration of mini-HIV close to the end of another mini-HIV molecule (end-to-end integration). When the total pool of reaction products was studied by EM, a large variety of DNA structures was observed. The Y-shaped molecules (Fig. 3B) result from a single intermolecular integration event. The DNA smear seen in the agarose gel between 9.5 and 15 kb is probably due to molecules whose mobility is affected by their branched structure. A single intramolecular integration event results in the structure shown in Figure 3E. Other reaction products represent multiple integration events (for example Fig. 3C and D). We also observed very large DNA structures with multiple branches, with lengths up to 4-fold the length of mini-HIV (data not shown). Depending on the quality of the gel, other distinct bands could be detected (for example, Fig. 1A, lane 4). We did not do a specific study to match the various DNA structures with these minor bands.

Restriction analysis of the D-type reaction products using native gel electrophoresis. We examined the fine structure of the double-sized reaction products to determine whether they were formed through a concerted integration mechanism or by end-to-end integration of the mini-HIV molecules. The mini-HIV substrate contains unique restriction sites for the *Csp*I, *Eco*RI and *Kpn*I restriction enzymes; the former cuts the molecule exactly in the middle, while the other two release the U3 and R-U5 fragments, respectively (Fig. 4B). The gel-purified D product DNA was digested with these enzymes. The results are consistent with the end-to-end integration mechanism (Fig. 4A). Indeed, restriction with *Csp*I generates two distinct bands, one migrating as full-length (4.7 kb) and the other as half-size (2.4 kb) mini-HIV DNA in a native agarose gel (lane 2). A process of concerted integration whereby one mini-HIV molecule is inserted into another would generate a range of products differing in the location of the integration site that is supposedly selected more or less at random along the target molecule. Digestion of the pool of these products should not produce the clear band of the substrate

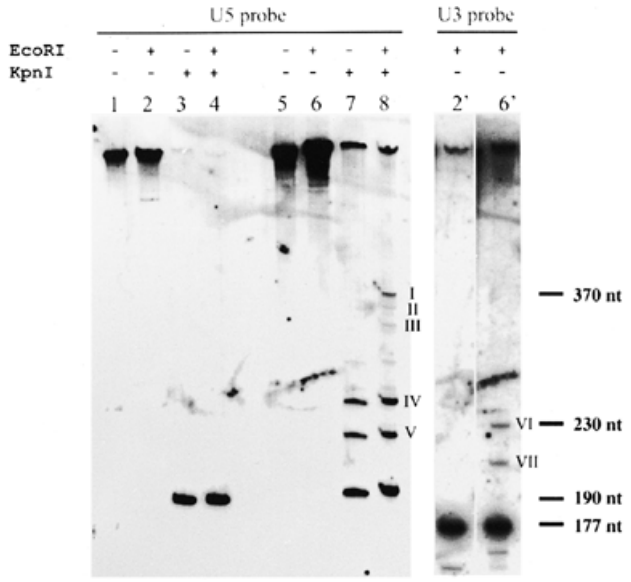

Figure 5. Extension of the U5 and U3 reacting strands. The mini-HIV and the D product DNA were digested with *Eco*RI and *Kpn*I enzymes and were separated by denaturing PAGE. The specific DNA fragments were detected using a U5 probe recognizing the reacting strand of U5. Lanes 1–4 contain the mini-HIV DNA and 5–8 the D product DNA. The same blot was reprobed with a U3 probe; only lanes 2' and 6' (as 2 and 6) are presented. Strand transfer resulted in the extension of the 3′-end of the reacting U5 or U3 terminus. The prominent bands corresponding to distinct integration hotspots are marked (I–VII).

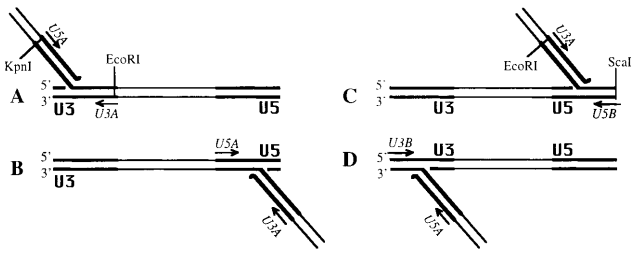

Figure 6. PCR amplification of integration products. Different types of integration products were PCR amplified with the following primer combinations: (**A** and **B**) U5A + U3A; (**C**) U5B + U3A; (**D**) U5A + U3B. Mini-HIV DNA (not to scale) is shown by parallel lines; bold lines represent U3 and R-U5 LTR fragments both in donor and target DNA; the primer annealing sites are indicated by arrows. Only a part of the donor mini-HIV molecule is shown.

sized DNA but rather a smear consisting of fragments with their lengths distributed between half- and full-size mini-HIV.

The restriction patterns generated by *Eco*RI and *Kpn*I digestion are also in accord with the end-to-end integration mechanism. Digestion of the D products with these enzymes produced distinct bands and not a smear as expected for a digest of concerted integration products (Fig. 4A, lanes 3 and 4). The DNA species migrating at ~450 bp and appearing after *Kpn*I or *Kpn*I+*Eco*RI digestion (Fig. 4A, lanes 4 and 5 after longer exposure) should represent branched molecules containing joined R-U5 and U3 fragments. Based on the appearance of this 450 bp band after digestion with *Kpn*I alone, we can deduce that a considerable portion of the D product must have been formed via insertion of the U5 terminus into U5 of another molecule. This is also suggested by the residual ~9.1 kb band after restriction with *Eco*RI (lane 3) (1000-fold excess of λ phage DNA present in every reaction showed complete digestion). Also, we can see that a fraction of the D product was probably formed through insertion of U3 into U3 (due to the residual ~9.1 kb DNA after *Kpn*I digestion in lane 4). The rest of the product can be the result of integration of U5 into U3 and vice versa. Note that in a fraction of D products integration may have occurred just outside the U3 or R-U5 regions.

Analysis of the D product using denaturing gel electrophoresis. To unambiguously prove that the observed products result from integration of the mini-HIV 3′-termini and are not due to secondary structure, the D product DNA was digested with *Eco*RI and *Kpn*I and separated by denaturing PAGE. The single-stranded DNA fragments containing the reacting strands of the U5 and U3 termini were detected by means of Southern blotting using strand-specific U5 and U3 probes (Fig. 5). Digestion of the mini-HIV DNA with *Kpn*I released a 190 nt fragment (lane 3), while the D product gave rise to two other major fragments of around 225 and 255 nt (labeled IV and V). Since they were released by a single digestion with *Kpn*I, these fragments apparently represent some integration hotspots of the U5 terminus at a distance of approximately 35 and 65 nt from the ends of U3 and/or U5. Alternatively, integration may preferentially take place at 35 and 65 nt from the *Kpn*I site. Double digestion with *Kpn*I and *Eco*RI produced a number of other fragments with estimated sizes ranging from 260 to 370 nt (I–III). The band of approximately 370 nt probably represents integration of U5 close to the end of U3.

The same blot was then reprobed with a strand-specific probe detecting the reacting strand of the U3 terminus. Digestion of the mini-HIV DNA with *Eco*RI yielded a 177 nt fragment (Fig. 5, lane 2'), while two additional fragments of 210 and 230 nt were detected upon digestion of the D product (Fig. 5, lane 6', fragments VI and VII). These fragments could result from specific integration of U3 either close to the ends of U3 and/or U5 or 35 and 55 nt from the *Eco*RI site.

The chemiluminiscent method employed for detection did not permit exact quantification of the bands. However, the data shown in Figure 5 are in a good agreement with the expectation of finding in total 25% of the LTR ends (U3 plus U5) in an integrated form (in this case all of the D product DNA would be the result of integration). It seems that a significant fraction of the D product is formed through integration of the U5 end into the two strongest hotspots.

Sequence analysis of the individual integration products. It was of interest to determine the sequence of the DNA junctions generated by integrase in the individual integration products. DNA from the D-type integration product was amplified using PCR and cloned. Three primer combinations were used: (i) U3A+U5A amplified both insertions of the U3 mini-HIV end into the U5 region and of the U5 end into U3 (Fig. 6A and B); (ii) U3A+U5B specifically amplified U3 end insertions (Fig. 6C); (iii) U5A+U3B was used to amplify the U5 end insertion sequences (Fig. 6D). Depending on the primer combination, there is specificity for products of integration into one of the strands of the mini-HIV DNA. However, using PCR for cloning implies some selective pressure against integration events that took place distantly from the primer binding site in the target molecule, i.e. presumably

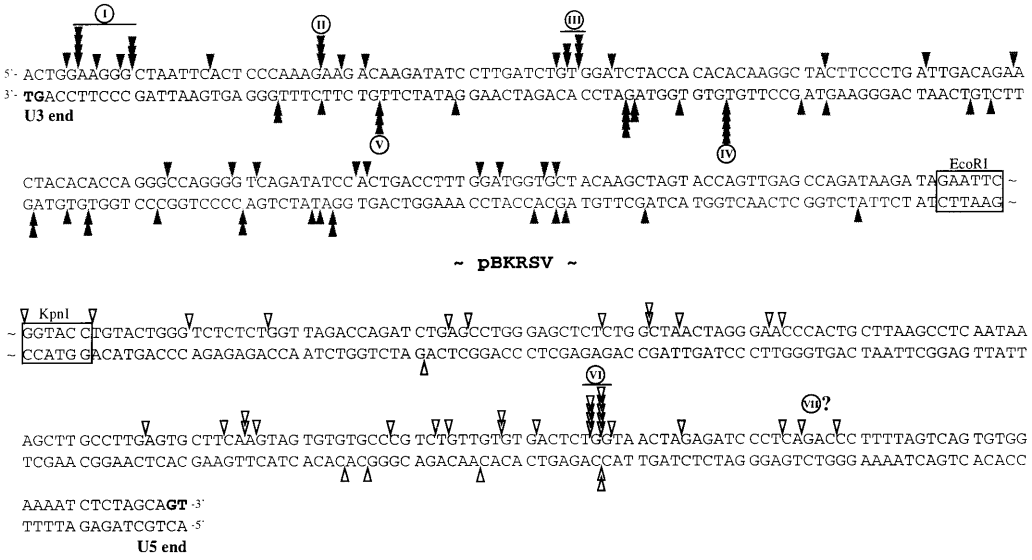

Figure 7. Sequence analysis of the integration products. Sequence of the LTR fragments present in the mini-HIV DNA is shown. The filled and open triangles point to the integration sites determined for the U5 and U3 ends, respectively. Five integration events for each U3 and U5 terminus, detected inside the vector sequence (6–90 bp from the *Kpn*I or *Eco*RI sites), are not indicated. The preferential integration sites are shown in bold and numbered I–VII. Position of hotspot VII is not clear. These hotspots apparently correspond to the bands observed in the Southern blot shown in Figure 5. The terminal GT dinucleotides that are removed in the processing step are in bold. The *Eco*RI and *Kpn*I restriction sites are boxed; in the mini-HIV molecule they are separated by 4.4 kb of vector DNA (pBKRSV).

more than 1000 bp from the ends of the mini-HIV DNA. With the primers U5B and U3B (Fig. 6C and D) integration events occurring less than 25 or 21 bp from the respective U5 and U3 ends could not be detected. The DNA amplified from the gel-purified D product was cloned into pBluescript KS(+) and 136 individual clones were sequenced. This experiment confirmed that strand transfer was preceded by correct 3′ processing, since all of the studied clones contained the U3 or U5 fragment with the two terminal base pairs removed and linked to some other part of the mini-HIV molecule.

Out of 49 clones obtained from PCR with the U3A+U5A primer combination, 35 were identified as the result of insertion of the U5 terminus close to the end of U3, whereas six contained insertions of U3 into U5 (Fig. 7). Strand transfer of U5 ends was thus 6-fold more efficient than that of U3 ends. The other eight clones had a structure in common that was unexpected: they contained both processed U5 and U3 ends with a short fragment in between. The fragment of 37 (five out of eight clones), 31, 59 or 60 bp is derived at random from the middle of the mini-HIV DNA (data not shown). We can only speculate as to the origin of these clones.

We also sequenced 87 clones obtained from PCRs with both other primer combinations (U3A+U5B and U5A+U3B). The U5 end insertions into the lower strand of U3 and the U3 insertions into the upper strand of U5 are also shown in Figure 7. Integration did not occur randomly throughout the DNA sequenced. Some hotspots for preferential integration were present. Some of the hotspots determined by sequence analysis apparently correspond to the data obtained by restriction digestion and denaturing Southern blot analysis (Fig. 5). The same labels (I–VII) are used to identify hotspots in both figures. The position of hotspot VII remains ambiguous. Due to technical complexity we could not clone and sequence a representative set of U5–U5 and U3–U3 junctions. Therefore, we cannot map other preferential integration sites. Possibly, hotspot VII in Figure 5 corresponds to integration of U3 into U3 about 35 bp from the end.

In total, we obtained 117 clones containing integration sites within the U3 and R-U5 regions of mini-HIV (shown in Fig. 7). We made a statistical analysis of the dinucleotide sequences targeted by integration in our assay. The relative contribution of each dinucleotide was calculated: 5′-Pu*Pu, 37%; 5′-Pu*Py, 26%; 5′-Py*Pu, 26%; 5′-Py*Py, 11%; asterisks indicate the phosphodiester bond cleaved during strand transfer. Based on this distribution a preference for purine bases at the site of DNA strand transfer can be deduced.

DNA–integrase complexes visualized by EM after crosslinking

To study the interaction between HIV-1 integrase and the mini-HIV DNA substrate a crosslinking reagent (0.6% glutaraldehyde) was added to the integration reaction and the resulting protein–DNA complexes were examined by EM. Unfortunately, polyethylene glycol, which was required for integrase activity, caused DNA to collapse. Therefore only 5–10% of the DNA molecules present could be analyzed. Most of the integrase complexes observed contained DNA loops and protein bound at the crossover points of the DNA loops. A characteristic pattern is shown in Figure 8A–C. Integrase was bound to the middle of the DNA molecule at a three site synapsis with two DNA loops. The enzyme seemed to interact with the DNA crossover point. The sites of the synapses and the sizes of the DNA loops were variable.

Complexes with enzyme bound to the ends of the substrate DNA or to sites of apparent DNA strand transfer were rare. The complexes shown in Figure 8D and E may represent the integration complex while the structure in Figure 8F could be an initial stable complex with integrase bound to the end of the mini-HIV DNA. Much larger protein oligomers were also observed. An example of such a complex is shown in Figure 8G.

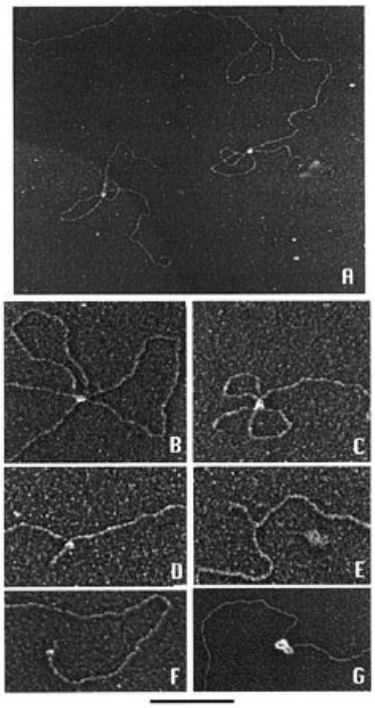

Figure 8. Crosslinking and visualization of the integrase–DNA complexes by EM. Integration reactions were allowed to proceed for 5–10 min, whereafter the nucleoprotein complexes were fixed using glutaraldehyde. The samples were visualized by adsorption onto charged thin carbon foils and dehydration, followed by rotary shadow casting with tungsten. Shown in reverse contrast. Bar equals a length of DNA equivalent to 1.0 (**A**) or 0.5 (**B**–**G**) kb.

This clump of protein bound to the DNA molecule could contain as many as 50–100 monomers of integrase and in these complexes DNA loops and DNA branches were often present. It was not clear whether the integration reaction *in vitro* is driven by some specific oligomer or by a population of differently sized complexes.

DISCUSSION

In vitro integration with mini-HIV DNA

Although the enzymatic activity of recombinant HIV-1 integrase could be studied in biochemical detail using short oligonucleotide substrates, authentic concerted integration has not been obtained *in vitro* so far. This work provides our first step toward reconstitution of the HIV integration process *in vivo* using mini-HIV DNA and intracellularly expressed integrase. We evaluated a linear 4.7 kb dsDNA flanked by intact U3 and R-U5 HIV-1 LTR fragments as a substrate for recombinant HIV-1 integrase. It is known that the initial binding preference of HIV-1 integrase for the viral DNA ends is not greater than that for non-specific DNA (24). DNA molecules (and polyanions in general) are potent inhibitors in oligonucleotide-based integration assays, acting as effective traps for the enzyme (16,25). In the 4.7 kb mini-HIV substrate non-specific DNA-binding sites are present in a large excess over the reactive terminal fragments. Hence, it was surprising to obtain efficient integration with long DNA molecules without addition of other viral or cellular factors. Previously it was shown that recombinant HIV-1 integrase can direct concerted integration using a mini-HIV DNA substrate but with such a low efficiency that it required a sensitive genetic approach for detection. The substrate contained 30 bp fragments of both HIV LTR ends with recessed 3′-termini (11). The mini-HIV DNA we designed contains blunt LTR fragments of approximately 200 bp that are substrates in both 3′ processing and strand transfer reactions. We found conditions in which integration products were readily detectable after separation by native agarose gel electrophoresis and ethidium bromide staining. In general, 20–30% of the substrate was converted into various integration products after 1.5 h incubation. We could directly demonstrate 3′-end processing. About 5% of the mini-HIV ends were cleaved, apart from the proportion that had undergone strand transfer. Notably, in the oligonucleotide-based assays, the extent of 3′-end processing clearly surpasses the efficiency of DNA strand transfer; this appears not to be the case in the mini-HIV assay. Apparently, the reactions of 3′ processing and strand transfer are coupled. Efficient coupling of both reactions was also found in the Mo-MLV PIC assay (26).

Both strand transfer and cleavage activities greatly depended on the presence of PEG. This macromolecular crowding agent was reported to promote strand transfer catalyzed by PICs of murine leukemia virus (MLV), HIV-1, Rous sarcoma virus, in integration reactions with purified AMV integrase and lysed HIV-1 virions as well as in oligonucleotide-based integration assays (4,20–23). PEG may be necessary to form a high molecular weight complex, possibly by increasing the effective protein concentration. Indeed, *in vivo* up to 100 molecules of integrase are believed to be present in the viral core, which corresponds to a protein concentration of ~20 mg/ml. In our mini-HIV integration assay Mg^{2+} was preferred over Mn^{2+} as the divalent cation. A similar observation has been made in experiments with longer oligonucleotide substrates (35 bp) (29).

In our mini-HIV reaction strand transfer was more efficient with U5 ends than with U3 ends (Figs 5 and 7), in accordance with other assays. It has been suggested that the sequence of U3 may not be as optimal for integration as the U5 end because of its coding function for Nef (27). It may also reflect different interactions of integrase with the two LTR ends that are formed asynchronously during the process of reverse transcription.

Because the reaction results predominantly in products with a size twice that of the mini-HIV DNA as seen by gel and EM analysis, concerted integration was expected. However, the presence of D products resulting from the reaction with a modified substrate that lacks the U3 fragment (Fig. 1C) was inconsistent with concerted integration. Analysis of the D-type reaction products revealed that a great majority of them were the result of one-sided integration of mini-HIV very close to the end of another mini-HIV molecule (end-to-end integration). We believe that this strong preference for end-to-end integration reflects the formation of an active complex where integrase interacts with two LTRs; however, in the absence of specific cofactors the reactive LTRs may not be aligned properly and consequently inserted into each other. Data obtained from both Southern blotting and sequencing were suggestive for strong hotspots inside the U3 and R-U5 regions of mini-HIV. As such, these hotspots may show a footprint of the enzyme bound to the LTR ends.

Evidence for a host factor involved in protection of retroviral DNA from auto-integration was suggested (30). This factor has been identified as a polypeptide of 88 amino acids, called BAF (barrier to auto-integration factor) (31). Obviously, the two LTR

ends in the pre-integration complex need to be correctly positioned for coordinated integration, perhaps with some rigid protein scaffold that holds both LTRs in place. It remains to be seen whether HMG I(Y), recently proposed as another cellular factor for HIV-1 integration, may serve this role (8). In the absence of such cofactors the complex may not be structured apropriately and illegitimate strand transfer between two LTRs could occur. Activity of purified AMV integrase on long DNA substrates has been studied before (10). It is interesting that reaction products approximately twice the length of the substrate DNA could also be detected, but these were not prominent. Probably specific reaction conditions, like concentration of PEG, account for this difference.

How do the specific integrase–DNA complexes form?

A large nucleoprotein complex is thought to assemble at the ends of viral DNA before mediating retroviral integration. This complex has been called the intasome (32). The problem remains unanswered how a specific intasome is formed at the viral DNA ends by a protein that displays no apparent sequence-specific binding. Integrase molecules bound to the mini-HIV DNA ends or in a strand transfer complex with target DNA were only infrequently seen by EM. The complexes may be too short-lived or may be too difficult to resolve after crosslinking with glutardehyde. Crosslinking of the active integrase complexes that may be aggregated might result in oligomers too large to analyze by EM.

Addition of a 300-fold molar excess of specific double-stranded oligonucleotides of 20 and 35 bp, with a sequence identical to the terminal bases of U5, did not effectively inhibit the integration process (data not shown). This lack of competition suggests a preferential binding of integrase to long DNA molecules and indicates that the primary target for integrase binding is not the LTR end, but that instead the enzyme after initially binding randomly to the DNA, scans the molecule in search of a specific end. The EM pictures of crosslinked HIV-1 integrase mostly show complexes of integrase bound to the middle of the DNA molecules with the DNA looping out. DNA loops in complexes with AMV integrase have been observed by Grandgenett and co-workers, albeit in the absence of Mg^{2+} or specific terminal LTR sequences (10). It is very tentative to propose that the loops are formed while integrase is scanning for the DNA end. That explanation suggests that the structures are active intermediates in the reaction. Alternatively, DNA loops may result from the oligomerization of different DNA-bound integrase monomers.

Interestingly, the observed DNA structures with two loops contain a three-way DNA crossover and resemble the specific three site synapsis reported to occur at the early stage of bacteriophage Mu transposase complex formation (13). It is also noteworthy that the three-way DNA crossover of the integrase–DNA complex consists of three dsDNA helices that seem to be positioned as in the putative coordinated integration complex. However, the latter should contain one continuous DNA molecule as target and the two viral DNA ends.

ACKNOWLEDGEMENTS

We appreciate the help of Dr A. Vandamme and K. Van Laethem with PCR and DNA sequencing. We thank Dr A. Vandamme for critical reading of the manuscript. Work in the laboratory of Dr J. Griffith was supported by a grant (GM42342) from the National Institutes of Health. Work at the Rega Institute was supported in part by the Biomedical Research Program of the European Commission and grants from the Belgian Fonds voor Wetenschappelijk Onderzoek and the Belgian Geconcerteerde Onderzoeksacties. Z.D. has a postdoctoral fellowship from the Flemish Fund for Scientific Research (FWO).

REFERENCES

1 Assante-Appiah,E. and Skalka,A.M. (1997) *Antiviral Res.*, **36**, 139–156.
2 Brown,P.O. (1997) In Coffin,J.M., Hughes,S.H. and Varmus,H.E. (eds), *Retroviruses*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, pp. 161–203.
3 Pommier,Y., Pilon,A.A., Bajaj,K., Mazumder,A. and Neamati,N. (1997) *Antiviral Chem. Chemother.*, **8**, 463–483.
4 Farnet,C.M. and Haseltine,W.A. (1990) *Proc. Natl Acad. Sci. USA*, **87**, 4164–4168.
5 Bukrinsky,M.I., Haggerty,S., Dempsey,M.P., Sharova,N., Adzhubei,A., Spitz,L., Lewis,P., Goldfarb,D., Emerman,M. and Stevenson,M. (1993) *Nature*, **365**, 666–669.
6 Heinzinger,N.K., Bukrinsky,M.I., Haggerty,S.A., Ragland,A.M., Kewalramani,S.O., Lee,M.A., Gendelman,H.E., Ratner,L., Stevenson,M. and Emerman,M. (1994) *Proc. Natl Acad. Sci. USA*, **91**, 7311–7315.
7 Gallay,P., Hope,T., Chin,D. and Trono,D. (1997) *Proc. Natl Acad. Sci. USA*, **94**, 9825–9830.
8 Farnet,C.M. and Bushman,F.D. (1997) *Cell*, **88**, 483–492.
9 Farnet,C.M. and Bushman,F.D. (1996) *Proc. Natl Acad. Sci. USA*, **93**, 9742–9747.
10 Grandgenett,D.P., Inman,R.B., Vora,A.C. and Fitzgerald,M.L. (1993) *Virology*, **67**, 2628–2636.
11 Bushman,F.D., Fujiwara,T. and Craigie, R. (1991) *Science*, **249**, 1555–1558.
12 Katz,R.A., Merkel,G., Kulkosky,J., Leis,J. and Skalka,A.M. (1990) *Cell*, **63**, 87–95.
13 Watson,M.A. and Chaconas,G. (1996) *Cell*, **85**, 435–445.
14 Wright,M. (1971) *J. Bacteriol.*, **107**, 87–94.
15 Cherepanov,P.P. and Wackernagel, W. (1995) *Gene*, **158**, 9–14.
16 Cherepanov,P., Esté,J.A., Rando,R.F., Ojwang,J.O., Reekmans,G., Steinfeld,R., David,G., De Clercq,E. and Debyser,Z. (1997) *Mol. Pharmacol.*, **52**, 771–780.
17 Griffith,J. and Christiansen,G. (1978) *Annu. Rev. Biophys. Bioeng.*, **7**, 19–35.
18 Kleinschmidt,A.K. (1968) *Methods Enzymol.*, **12**, 361–367.
19 Westmoreland,B.C., Szybalski,W. and Ris,H. (1969) *Science*, **163**, 1343–1348.
20 Brown,P.O., Bowerman,B., Varmus,H.E. and Bishop,J.M. (1987) *Cell*, **49**, 347–356.
21 Engelman,A. and Craigie,R. (1995) *J. Virol.*, **69**, 5908–5911.
22 Goodarzi,G., Im,G.-J., Brackmann,K. and Grandgenett,D. (1995) *J. Virol.*, **69**, 6090–6097.
23 Vora,A.C., McCord,M., Fitzgerald,M.L., Inman,R.B. and Grandgenett,D.P. (1994) *Nucleic Acids Res.*, **22**, 4454–4461.
24 van Gent,D.C., Elgersma,Y., Bolk,M.W.J., Vink,C. and Plasterk,R.H.A. (1991) *Nucleic Acids Res.*, **19**, 3821–3827.
25 Ellison,V. and Brown,P.O. (1994) *Proc. Natl Acad. Sci. USA*, **91**, 7316–7320.
26 Fujiwara,T. and Mizuuchi,K. (1988) *Cell*, **54**, 497–504.
27 Bushman,F.D. and Craigie,R. (1990) *Proc. Natl Acad. Sci. USA*, **88**, 1339–1343.
28 Hong,T., Murphy,E., Groarke,J. and Drlica,K. (1993) *J. Virol.*, **67**, 1127–1131.
29 Lee,S.P., Censullo,M.L., Kim,H.G. and Han,M.K. (1995) *Biochemistry*, **34**, 10215–10223.
30 Lee,M.S. and Craigie R. (1994) *Proc. Natl Acad. Sci. USA*, **91**, 9823–9827.
31 Lee,M.S. and Craigie,R. (1998) *Proc. Natl Acad. Sci. USA*, **95**, 1528–1533.
32 Wei,S.-Q., Mizuuchiand,K. and Craigie,R. (1997) *EMBO J.*, **16**, 7511–7520.