JOURNAL OF VIROLOGY, July 1997, p. 5382–5390 Vol. 71, No. 7
0022-538X/97/$04.00+0

Human Immunodeficiency Virus Type 1 Preintegration Complexes: Studies of Organization and Composition

MICHAEL D. MILLER,† CHRIS M. FARNET,‡ AND FREDERIC D. BUSHMAN*

Infectious Disease Laboratory, The Salk Institute for Biological Studies, La Jolla, California 92037

Received 9 December 1996/Accepted 20 March 1997

We have investigated the organization and function of human immunodeficiency virus type 1 (HIV-1) preintegration complexes (PICs), the large nucleoprotein particles that carry out cDNA integration in vivo. PICs can be isolated from HIV-1-infected cells, and such particles are capable of carrying out integration reactions in vitro. We find that although the PICs are large, the cDNA must be condensed to fit into the measured volume. The ends of the cDNA are probably linked by a protein bridge, since coordinated joining of the two ends is not disrupted by cleaving the cDNA internally with a restriction enzyme. cDNA ends in PICs were protected from digestion by added exonucleases, probably due to binding of proteins. The intervening cDNA, in contrast, was susceptible to attack by endonucleases. Previous work has established that the virus-encoded integrase protein is present in PICs, and we have reported recently that the host protein HMG I(Y) is also present. Here we report that the viral matrix and reverse transcriptase (RT) proteins also cofractionated with PICs through several steps whereas capsid and nucleocapsid proteins dissociated. These data support a model of PIC organization in which the cDNA is condensed in a partially disassembled remnant of the viral core, with proteins tightly associated at the apposed cDNA ends but loosely associated with the intervening cDNA. In characterizing the structure of the cDNA ends, we found that the U5 DNA ends created by RT were ragged, probably due to the terminal transferase activity of RT. Only molecules correctly cleaved by integrase protein at the 3′ ends were competent to integrate, suggesting that one role for terminal cleavage by integrase may be to create a defined end at otherwise heterogeneous cDNA termini.

Early during infection, the human immunodeficiency virus (HIV) RNA genome is reverse transcribed to form a cDNA copy, which is then inserted into host cell DNA (20, 43). The substrate for HIV-1 DNA integration is a linear, mostly double-stranded viral cDNA molecule with terminal direct repeats and blunt ends (6, 10, 16, 21, 23). To carry out integration, the virus-encoded integrase protein (IN) first cleaves two nucleotides from each 3′ viral DNA end in a step called terminal cleavage (6, 23, 36) and then joins the newly exposed 3′ hydroxyl groups to phosphodiester bonds in the target (host) DNA (6, 23). The cDNA is inserted by coordinately joining both viral cDNA ends to nearby phosphodiesters on opposite strands of the target DNA. Resulting breaks in the DNA are then repaired, probably by host DNA repair enzymes, to complete the integration reaction. Because insertion of only one viral cDNA end would generate a branched DNA structure that would probably be eliminated by host DNA repair systems, the coupled integration of both cDNA ends is important for the survival of the virus.

Retrovirus cDNA integration can be partially reproduced by using purified components in vitro. Purified HIV-1 IN performs end-processing and joining reactions on model DNA substrates in vitro, but such reactions differ from authentic integration because coordinate joining of two viral DNA ends is inefficient (9, 10, 12, 30, 31, 40). Although higher levels of coupled joining can be achieved with integrase extracted from virions (45), the efficiency is still far below that expected for integration in vivo.

Subviral preintegration complexes (PICs) can be obtained from cells freshly infected with HIV-1 and are competent to integrate the endogenous cDNA into a target DNA supplied in vitro (5, 16, 21). In marked contrast to the products of reactions involving purified IN and model DNA substrates, all detected products of PIC integration reactions have structures indicative of coupled joining (16, 27, 34a). The basis of this difference with reaction mixtures containing purified IN protein is unknown. Possible explanations include the presence of unknown modifications on IN, the participation of unknown protein or DNA components, a unique mode of IN DNA assembly, or all of these. Here we use assays of active HIV-1 PICs to characterize the architecture and composition in detail.

MATERIALS AND METHODS

Preparation of PICs and use in integration reactions. HIV-1 infection of SupT1 cells by coculture with Molt-IIIB cells and preparation of cytoplasmic extracts were performed as described previously (21, 33), except that cells were permeabilized in buffer K (20 mM HEPES, 150 mM KCl, 5 mM $MgCl_2$, 1 mM dithiothreitol, 50 U of aprotinin per ml [pH 7.3]) containing 0.025% digitonin instead of 0.025% Brij 96. Extracts of uninfected SupT1 cells were made in the same way, but coculture with Molt-IIIB cells was omitted. Integration reactions were carried out by incubating cytoplasmic extracts with φX174 RFI DNA at 37°C; the amount of extract, concentration of target DNA, form of target DNA (linear or circular), and time of incubation are described in the figure legends. DNAs were purified by sodium dodecyl sulfate (SDS)-proteinase K treatment, phenol-chloroform extraction, and ethanol precipitation as described previously (33). In some cases, DNAs were digested with restriction enzymes by standard techniques and then repurified by digestion with RNase A, phenol-chloroform extraction, and ethanol precipitation.

Electrophoresis and Southern blotting. Native agarose gel electrophoresis and Southern blotting were done essentially as described previously (33). DNA samples for denaturing gel electrophoresis were prepared by resuspending RNase-treated DNA pellets in 5 μl of Tris-EDTA (TE), adding 5 μl of 95% formamide–20 mM EDTA–0.05% bromophenol blue–0.05% xylene cyanol, and boiling for 5 min. The samples were resolved by electrophoresis through 6%

* Corresponding author. Mailing address: Infectious Disease Laboratory, The Salk Institute for Biological Studies, 10010 N. Torrey Pines Rd., La Jolla, CA 92037.

† Present address: Department of Antiviral Research, Merck Research Laboratories, West Point, PA 19486.

‡ Present address: Bio-Mega Research Division (Canada), Boehringer Ingelheim, Montreal, Quebec, Canada.

polyacrylamide–8 M urea sequencing-type gels in 1× Tris-borate-EDTA (TBE) (Sequagel) and then electroblotted onto Duralon-UV (Stratagene) for 1 h at 200 mA in 0.2× TBE. The filters were rinsed in 10× SSC (1× SSC is 0.15 M NaCl plus 0.015 M sodium citrate), air dried, baked for 2 h at 80°C, and cross-linked with 10^5 μJ of UV (Stratalinker). DNAs were detected by hybridization to ^{32}P-labeled RNA probes (5×10^6 cpm/ml in QuickHyb [Stratagene]) for 2 h at 68°C. The filters were washed twice for 15 min in 2× SSC–0.1% SDS at room temperature and then twice for 10 min in 0.1× SSC–0.1% SDS at 60°C, air dried, and exposed to Kodak XAR-5 film. Sequencing reactions for markers were carried out with the control template (M13) and primer (−40) included in the Sequenase kit (U.S. Biochemical).

Gel filtration. Sephacryl S-1000 chromatography was carried out with a Pharmacia XK 16/100 column attached to a fast protein liquid chromatography apparatus. The total volume (elution volume of ATP) was 176.5 ml, and the void volume (elution volume of 491 ± 8-nm-diameter latex spheres [all latex spheres from Polysciences]) was 67.5 ml. The elution volumes of additional standards were 71 ml (215 ± 9-nm-diameter latex spheres), 101.5 ml (107 ± 13-nm-diameter latex spheres), 120 ml (64 ± 10-nm-diameter latex spheres), and 165 ml (bovine serum albumin; Stokes radius, 4.5 nm). For chromatography of PICs, the column was equilibrated in buffer K containing 10% dimethyl sulfoxide (DMSO) (in the absence of DMSO, PICs were found in the excluded volume). The flow rate was 1 ml/min. Crude cytoplasmic extract (10 ml) was loaded and run in the same buffer at a flow rate of 1 ml/min. Fractions were collected and assayed for viral DNA and integration activity as described above. The peak elution volume of PICs was 132.5 ml. The chromatographic mobility of latex spheres could not be analyzed in buffer containing 10% DMSO because the spheres dissolve in this solvent.

Sucrose density gradient sedimentation. Gradients (12 ml) of 10 to 50% (wt/wt) sucrose in buffer K were poured into polyallomer SW40 tubes with an Auto Densi-flow III apparatus (Buchler). Cytoplasmic extracts were treated with 100 U of *Sal*I per ml for 10 min at 37°C and then placed on ice. Aliquots were removed for analysis, and then a 0.75-ml sample was layered on top of the gradient and centrifuged in an SW40 rotor for 2 h at 35,000 rpm and 4°C. Fractions (1 ml) were collected with the Auto DensiflowIII and assayed for *Sal*I and integration activities.

Nuclease digestion of preintegration complexes. Naked DNA from PICs was purified as described above and resuspended in TE at 1/10 the original volume of extract. For exonuclease digestions, naked DNA was added to a 10-fold excess volume of uninfected cytoplasmic extracts, while for endonuclease digestions, naked DNA was added to buffer K. In all cases, comparisons of nuclease sensitivities were done with intact PICs and naked DNA prepared from the same batch of complexes. To improve the conditions for various nucleases, cytoplasmic extracts and DNA solutions were supplemented as follows: for exonuclease III, Triton X-100 was added to 0.5%; for λ exonuclease, Triton X-100 was added to 0.5% and Tris-HCl (pH 9.0) was added to 100 mM; and for micrococcal nuclease, Triton X-100 was added to 0.5% and $CaCl_2$ was added to 10 mM. PICs showed wild-type activity in all of these buffers, typically with >50% of the viral DNA molecules becoming integrated into target DNA. Nucleases were added to the concentrations indicated in the figure legends, incubated at 37°C for the time indicated in the figure legends, and stopped by SDS-proteinase K treatment as described above.

Probes. The specific sequences used to generate RNA probes were as follows: pMM104 (U3 region), 111-bp fragment corresponding to bp 1 to 111 of LAV; pMM105 and 106 (U5 region), 101-bp fragment corresponding to bp 9605 to 9706; pBW13 (internal fragment), 1,820-bp fragment corresponding to bp 2010 to 3830 of pNL4-3. All were cloned into pBluescriptII SK (Stratagene). Probes to detect the plus strand were made with pMM104 and pMM105 linearized with *Bam*HI and *Hin*dIII, respectively, and T7 RNA polymerase. Probes to detect the minus strand were made with pMM104, pMM106, and pBW13 cut with *Eco*RV, *Bst*YI, and *Bgl*II, respectively, and T3 RNA polymerase. RNA probes were synthesized by incubating 1 μg of linearized plasmid in 1× RNA polymerase buffer (NEB)–0.5 mM each GTP, ATP, and CTP–5.5 U of placental RNase inhibitor (USB) per μl–50 μCi of [α-^{32}P]UTP (400 Ci/mmol; Amersham)–5 U of T3 RNA polymerase (Stratagene) or T7 RNA polymerase (NEB) per μl in a volume of 10 μl. After incubation at 37°C for 1 h, 10 U of RNase-free DNase I (Pharmacia) was added and the reaction mixtures were incubated at 37°C for 10 min. Unincorporated nucleotides were removed by passing the reaction products through a G-50 spin column.

Purification of PICs and analysis by Western blotting. PICs (2 ml of crude extract prepared as above) were purified by low-salt precipitation as described previously (19). PICs were purified by using gel filtration spin columns and Microcon-100 ultrafiltration as described previously (18). Western blotting was carried out as described previously (8, 18, 24). Antibodies against MA (a gift of D. Trono), reverse transcriptase (RT) (Advanced Biotechnologies Inc.), and nucleocapsid (NC) (a gift of L. Henderson, National Cancer Institute) were rabbit polyclonal antibodies, and antibody against capsid (CA) (Advanced Biotechnologies Inc.) was a mouse monoclonal antibody.

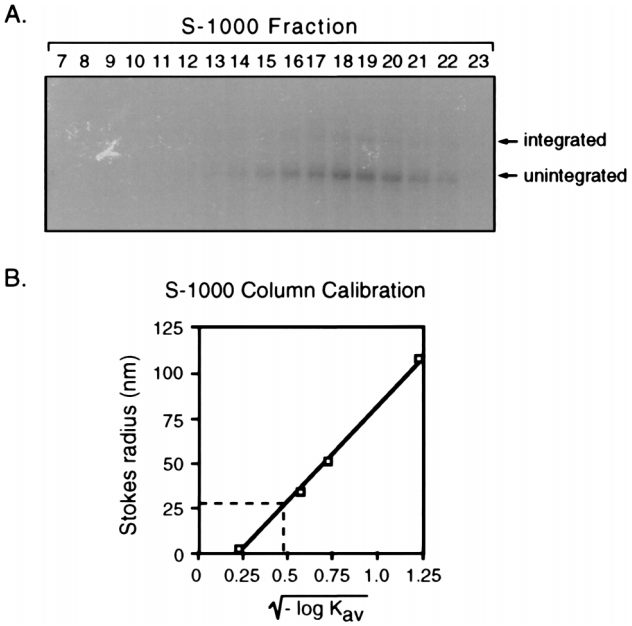

FIG. 1. Biophysical characterization of HIV-1 PICs. (A) Southern blot analysis of PICs subjected to Sephacryl S-1000 chromatography. Aliquots of S-1000 fractions were incubated with 3 μg of linear φX174 RFI DNA per ml for 30 min at 37°C, deproteinized, and analyzed by Southern blotting with an LTR probe. (B) Calibration curve of the S-1000 column. The standards used were latex beads (Polysciences) with a radius of 245 ± 4 (void volume, not shown), 107 ± 5, 53 ± 7, or 32 ± 5 nm; bovine serum albumin (Stokes radius, 4.5 nm), and ATP (total volume, not shown).

RESULTS

Biophysical characterization of HIV-1 preintegration complexes. We used size exclusion chromatography to obtain an estimate for the Stokes radius of HIV-1 PICs. Crude extract containing PICs was fractionated on an S-1000 column, and the fractions were assayed for the presence of viral cDNA and integration activity by native agarose gel electrophoresis, Southern blotting, and probing with labeled long terminal repeat (LTR) sequences (Fig. 1A). Complexes containing linear viral cDNA eluted in a single broad, included peak. Complexes were competent to integrate their cDNA into a target DNA in vitro, and the peak of integration activity coincided with the peak of viral DNA. The S-1000 column was calibrated (41) with latex spheres of known diameter, allowing the Stokes radius to be estimated as 28 nm (Fig. 1B). It is clear from these measurements that the viral cDNA, with an estimated contour length of 3.3 μm, must be compacted to fit in a structure of this size.

Cleavage of PIC DNA by endonucleases. To begin to investigate protein-DNA interactions in HIV-1 PICs, we asked whether the cDNA in PICs was less exposed to endonuclease digestion than was naked cDNA. We treated intact PICs or naked cDNA purified from PICs with various amounts of micrococcal nuclease (Fig. 2A) and analyzed the reaction products by Southern blotting with LTR probes as described above.

Micrococcal nuclease was able to completely digest both PIC DNA and naked DNA with roughly equal efficiency (Fig. 2A), indicating that the bulk of the viral DNA is not packaged so tightly that nucleases cannot attack it. Micrococcal nuclease cleaved both PIC and naked DNAs with some specificity (0.15 and 1.5 U/ml), generating prominent fragments of 4 to 5 and 0.4 to 0.7 kb; the smaller fragments are notable because they are roughly the size of an LTR. The activity of complexes was

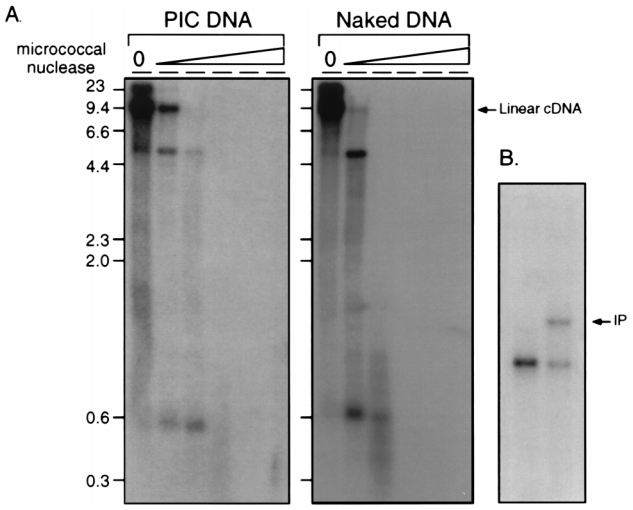

FIG. 2. DNA in PICs and naked viral cDNA are similarly susceptible to digestion by micrococcal nuclease. (A) HIV-1 PICs are active in micrococcal nuclease buffer. PICs in cytoplasmic extracts were supplemented with Triton X-100 to 0.5% and with $CaCl_2$ to 10 mM and then incubated in the absence or presence of 3 μg of linear φX174 DNA per ml for 30 min at 37°C. Reaction products were deproteinized, resolved by electrophoresis through a 1.5% agarose gel, transferred to a nylon membrane, and detected by hybridization with a probe complementary to the viral LTR. (B) Activity of PICs in the micrococcal nuclease digestion buffer.

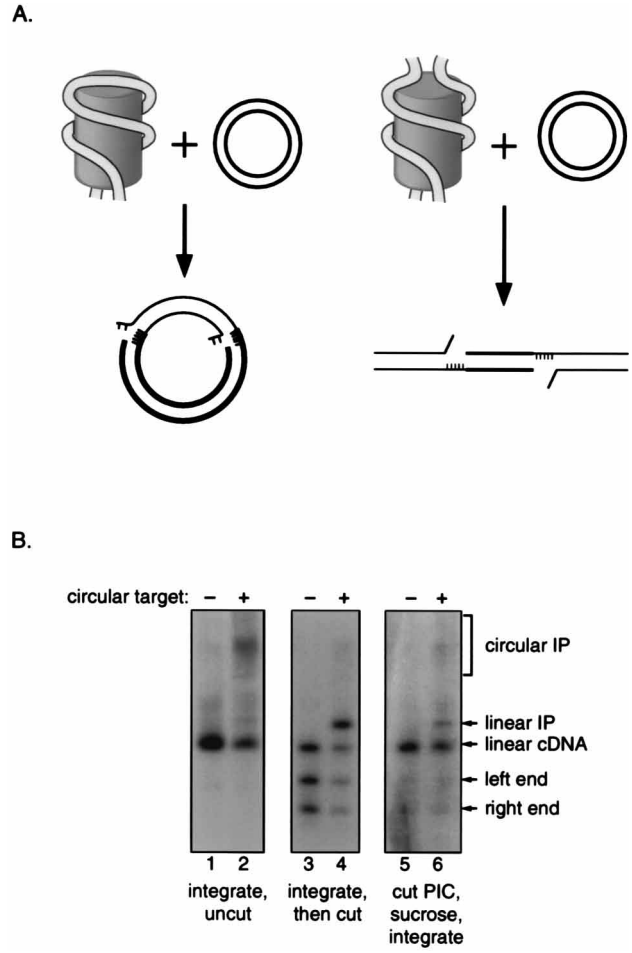

FIG. 3. The ends of the HIV-1 cDNA are held together in the PIC. (A) Diagram of the experimental strategy. The PIC is depicted as a cylinder with the viral DNA (thin parallel lines) wrapped around the outside; the drawing is not meant to indicate any specific arrangement of the viral DNA. Target DNA is shown as thick parallel lines. Unpaired bases are shown as short thin lines. The integration reaction yields an intermediate in which the viral DNA 3′ ends are attached to the target, the viral cDNA 5′ ends are free and contain a 2-nucleotide overhang with respect to the 3′ ends, and the target DNA has 5 unpaired bases flanking the integrated viral cDNA. The left side depicts the integration of untreated viral cDNA into a circular target. The right side depicts the integration of viral cDNA cleaved with a restriction endonuclease into a circular target. (B) Southern blot analysis of integration reactions with untreated or *Sal*I-treated PICs. Lanes: 1, PICs incubated without target; 2, PICs incubated with a circular target; 3 and 4, same reaction products as in lanes 1 and 2 but cleaved with *Sal*I; 5 and 6, PICs treated with *Sal*I and then fractionated over a sucrose gradient. *Sal*I activity was detectable in fractions 1 to 4, and viral DNA was detectable in fractions 7 to 12 (not shown). Fraction 11 was incubated in the absence (lane 5) or presence (lane 6) of a circular target DNA. Purified DNA reaction products were resolved by electrophoresis through 0.75% agarose gels in 1× TAE, transferred to a nylon membrane, and detected by hybridization with a ^{32}P-labeled HIV-1 LTR probe.

unaffected by incubation in the micrococcal nuclease buffer (Fig. 2B). Nuclease *Bal* 31 generated a similar cleavage pattern (data not shown). For both micrococcal nuclease and *Bal* 31, the cleavage patterns of PIC DNA and naked DNA were essentially identical and suggest the presence of nuclease-hypersensitive sites near the LTRs. The 4- to 5-kb fragments generated by partial digestion with micrococcal nuclease or *Bal* 31 probably resulted from cleavage of single-stranded viral DNA present near the central polypurine tract (11, 33).

These findings parallel those of Bowerman et al. (4), who found that micrococcal nuclease digestion of murine leukemia virus (MLV) PICs yielded a prominent LTR-sized DNA fragment. We extend these results by showing that in HIV-1, the production of the LTR-sized fragments is associated with the DNA itself rather than with proteins bound to the DNA, since the LTR-sized bands are seen in digests of naked cDNA.

The two viral DNA ends are probably linked by a protein bridge. Since the points of cDNA integration on each target DNA strand are always separated by a defined distance, it seems likely that the cDNA ends are held together by proteins. To verify that the ends are paired, we asked whether two viral cDNA ends could still be coordinately joined to target DNA after being severed from one another by a restriction endonuclease. The experimental strategy is based on the idea that coordinated integration of linear viral cDNA into a circular target yields a circular product whereas coordinate integration of two unlinked cDNA should yield a linear product (Fig. 3A).

PICs were first incubated in the presence or absence of a circular DNA target, and then an aliquot of each reaction mixture was cleaved with *Sal*I, which cuts once in the HIV-1 cDNA but does not cut the target DNA. Cleaved and uncleaved reaction products were then analyzed by Southern blotting as above. In the reaction mixture lacking target DNA and *Sal*I, only the linear viral cDNA was detected (Fig. 3B, lane 1); addition of target resulted in the production of a diffuse band corresponding to the circular integration product (lane 2). The second panel shows the products of the same integration reactions after cleavage with *Sal*I. In the reaction mixture lacking target DNA (lane 3), the products included the two *Sal*I fragments generated by cleavage of the linear cDNA, as well as a species the size of the uncleaved linear cDNA (presumably representing either incomplete digestion of the linear DNA or linearization of circular forms of the cDNA). The reaction mixture containing target DNA included these products, as well as a sharp band of the size predicted for the linear integration product (lane 4). These experiments dem-

onstrate that the circular and linear integration products diagrammed in Fig. 3A can be distinguished by electrophoresis.

To ask whether the viral cDNA ends were held together, intact PICs were treated with *Sal*I before the integration reactions were performed. To rule out the possibility that linearization of the integration product occurred after the integration itself, PICs were separated from the *Sal*I enzyme before reaction with target. *Sal*I-treated PICs were centrifuged through a sucrose gradient, and aliquots of a fraction containing PICs but no detectable *Sal*I (data not shown) were incubated in the absence (Fig. 3B, lane 5) or presence (lane 6) of target DNA and analyzed as above. Analysis of products of the reaction lacking target DNA (lane 5) revealed that only a fraction of the viral cDNA was cleaved by the restriction enzyme. Incubation of this material with target DNA yielded a diffuse band corresponding to the circular integration product. In addition, a distinct band corresponding in size to the linear integration product was observed (lane 6). This product must have resulted from the coupled joining of two viral cDNA ends that remained paired despite cleavage of the intervening cDNA.

Active preintegration complexes contain processed 3′ ends. To begin to examine protein-DNA interactions at the cDNA ends, we first characterized the fine structure of viral cDNA termini in active PICs. Earlier studies presented evidence of 3′-end processing in viral cDNA isolated from active Moloney MLV (MoMLV) PICs (6, 23) or from cells infected with MoMLV (38) or HIV-1 (36). We asked whether the viral cDNA 3′ ends in our active HIV-1 PICs had been processed to remove the expected 2 nucleotides from each 3′ end (Fig. 4). Complexes were incubated on ice (Fig. 4B, lanes 1) or at 37°C for 30 min in the absence (lanes 2) or presence (lanes 3) of a target DNA. Following the reactions, viral cDNAs were deproteinized and then digested with the restriction endonucleases *Hae*III and *Hin*dIII, which cleave near the left and right ends, respectively. DNA fragments were resolved by electrophoresis through denaturing polyacrylamide gels, transferred to nylon membranes, and detected by hybridization with ^{32}P-labeled RNA probes specific for each strand. Figure 4A shows the experimental strategy and the lengths of each cDNA fragment expected to appear on the autoradiogram.

Probes for viral DNA 5′ left and right ends each detected a single band of the predicted size (Fig. 4B, lanes 1, left top strand and bottom right strand). Probes for viral cDNA 3′ left and right ends each hybridized with two bands of the sizes expected for processed and unprocessed forms (lanes 1, top right strand and bottom left strand). Most of the viral cDNA 3′ ends in freshly isolated PICs were processed (about 70% processed by PhosphorImager quantification [data not shown]). Incubation of complexes at 37°C resulted in little additional 3′-end processing (compare lanes 1 and 2).

Inclusion of an integration target DNA in the reaction mixture resulted in the selective depletion of the processed 3′ ends with little change in the amount of unprocessed 3′ ends present (Fig. 4B, top right and bottom left, compare lanes 2 and 3). In contrast, probes for the 5′-end fragments detected similar amounts of DNA in reactions carried out in the absence or presence of target DNA (top left and bottom right; compare lanes 2 and 3), indicating that, as expected, the 5′ ends are not joined to target DNA in vitro. These findings indicate that processed cDNA ends are integrated selectively.

Time course of the generation and cleavage of the cDNA ends in vivo. A time course of viral cDNA synthesis during HIV-1 infection shows that apparently full-length linear cDNA is first detectable 3 h after infection and that maximal synthesis is apparent 4 h after infection (Fig. 5A). Figure 5B shows an analysis of the fine structure of cDNA 3′ ends at various times after infection.

Interestingly, the right (U5) 3′ end was markedly heterogeneous early in the time course (lanes 2 and 3), probably as a result of the previously described terminal transferase activity of HIV-1 RT (35, 37). The terminal transferase activity of purified RT was also more pronounced on a U5 (right)-end substrate than on a U3 (left)-end substrate (35), consistent with the pattern observed here for cDNA from infected cells. By 3 h, there was a prominent fragment of the size expected for a full-length, unprocessed U5 3′ end and a small amount of the processed form (lane 4). The abundance of processed U5 ends increased after 3 h, with about 50% of the U5 ends processed at 4 h and about 80% processed at 6 h.

A simpler pattern was seen at the left (U3) viral DNA 3′ end (Fig. 5B). The completed left (U3) end of the minus strand first appeared 3 h after initiation of coculture, and most of this end was unprocessed. About 50% of the U3 ends were processed at 4 h, and additional processing was seen at 5 and 6 h. Thus, most of the processing of the viral cDNA ends occurs between 3 and 5 h after initiation of reverse transcription, soon after synthesis of the end is completed.

The plus-strand strong stop DNA, an intermediate in reverse transcription, is also recognized by the U5 plus-strand probe (Fig. 5B). Plus-strand strong stop first appeared 1 h after infection, and then its abundance increased and later declined. This is as expected if the plus-strand strong stop DNA completes the econd template switch and is subsequently extended.

These data also specify the order of late steps in HIV-1 reverse transcription. This analysis demonstrated that the right (U5) viral cDNA end is completed before the left end (1 and 3 h, respectively). This result is as expected if plus-strand cDNA synthesis is initiated in the middle of the minus strand at internal polypurine tracts, as predicted from earlier studies (11, 33). Thus, the plus strand can extend to the right end before the plus-strand strong stop transfers to the left end of the minus strand to allow completion of the left end (see reference 33 for diagrams and further references).

Proteins bind the HIV-1 cDNA ends in the PIC. To examine protein binding at the HIV-1 cDNA ends, we compared the sensitivity to exonucleases of cDNA in intact PICs and deproteinized cDNA. Intact PICs were tested as part of a crude extract. To provide a similar reaction milieu in the controls, purified cDNA was tested in a mock extract from uninfected cells. Exonuclease III (exoIII), a 3′-to-5′ exonuclease, was used to analyze the sensitivity of viral cDNA 3′ ends, and λ exonuclease (λexo), a 5′-to-3′ exonuclease, was used to analyze the sensitivity of viral cDNA 5′ ends. Nuclease reaction mixtures were incubated for 10 min at 37°C, and then the reactions were stopped. DNAs were purified, digested with *Hae*III and *Hin*dIII, and analyzed by electrophoresis and Southern blotting as in Fig. 4.

Figure 6A shows that HIV-1 cDNA was protected from digestion by exoIII when it was part of the PIC but not when it had been deproteinized (compare lanes 1 and 8). A control probe hybridizing to an internal restriction fragment, in contrast, was not digested (bottom panel). The 3′-end probes (top two panels) detected both processed and unprocessed forms as described above. Interestingly, for cDNAs in PICs, the processed and unprocessed forms differed in their nuclease sensitivity. The processed ends were completely resistant to exoIII, but the unprocessed ends were digested. The selective sensitivity of the unprocessed viral cDNA ends suggests that they may be present in disrupted PICs. In naked cDNA, both forms were digested by exoIII, with nearly complete digestion at 333 U/ml. For unknown reasons, the minus strand (left-end probe)

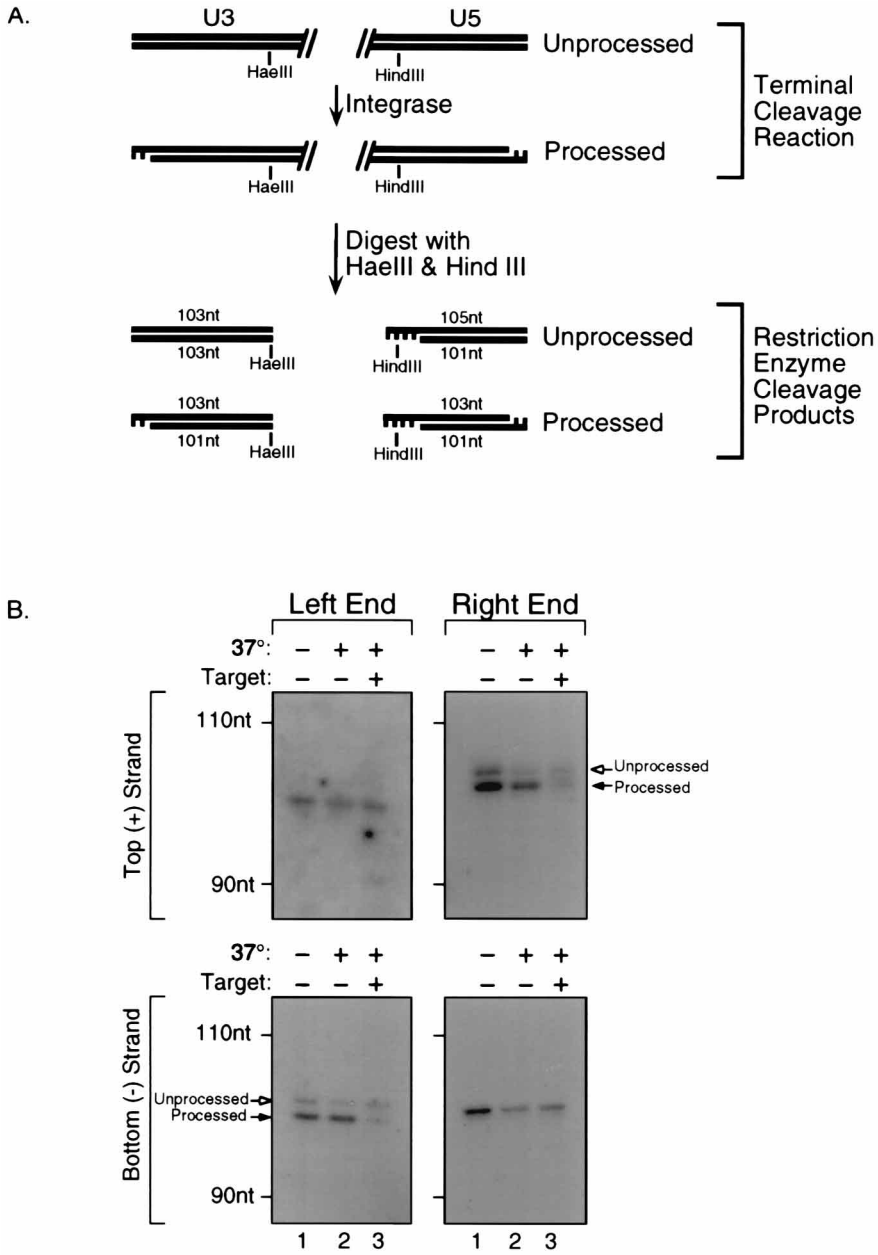

FIG. 4. The 3′ ends of PIC DNAs are mostly processed. (A) Diagram of the experimental strategy. The two viral cDNA ends are depicted at the top of the figure. Terminal cleavage by IN removes 2 nucleotides (nt) from each 3′ end, leaving a 2-nucleotide 5′ overhang at each end. *Hae*III cleavage yields blunt DNA ends, and *Hin*dIII cleavage yields DNA ends with 4-nucleotide 5′ overhangs. The expected length (in nucleotides) of each DNA fragment is shown. The drawing is not to scale. (B) Electrophoretic analysis of HIV-1 cDNA 3′ ends. Lanes: 1, PICs were incubated on ice; 2, PICs were incubated at 37°C in the absence of a target DNA; 3, PICs were incubated at 37°C in the presence of a target DNA. Purified DNA reaction products were cleaved with *Hin*dIII and *Hae*III, repurified, and resolved on a 6% acrylamide–8 M urea sequencing gel in 1× TBE. DNAs were electroblotted to a nylon membrane and detected by hybridization with ^{32}P-labeled RNA probes to detect the top (+) or bottom (−) strand of the U3 (left end) or U5 (right end) region of the viral cDNA. In other experiments, the lengths of the detected fragments were found to be of the expected sizes by comparison to the products of dideoxynucleotide sequencing reactions run in adjacent lanes.

was more resistant to exoIII than was the plus strand (right-end probe). Overall, the increase in nuclease sensitivity following protease treatment indicates that proteins in the PIC protected the processed cDNA ends.

One possible trivial explanation for the apparent protection of the cDNA ends is that these ends are contained in aggregated, nonfunctional PICs. To rule out this possibility, we asked whether PICs would still be integration competent after treatment with exoIII. To assay integration activity, aliquots of nuclease-treated or control PICs were incubated for an additional 30 min in the absence or presence of target DNA. The products of integration reactions were deproteinized, resolved by native agarose gel electrophoresis, and analyzed by Southern blotting (Fig. 6B). We found that exoIII treatment of PICs had no effect on the yield of integration product (Fig. 6B, right panel). Thus, the protected ends were nevertheless functional, supporting the view that biologically relevant protein complexes were responsible for the observed protection.

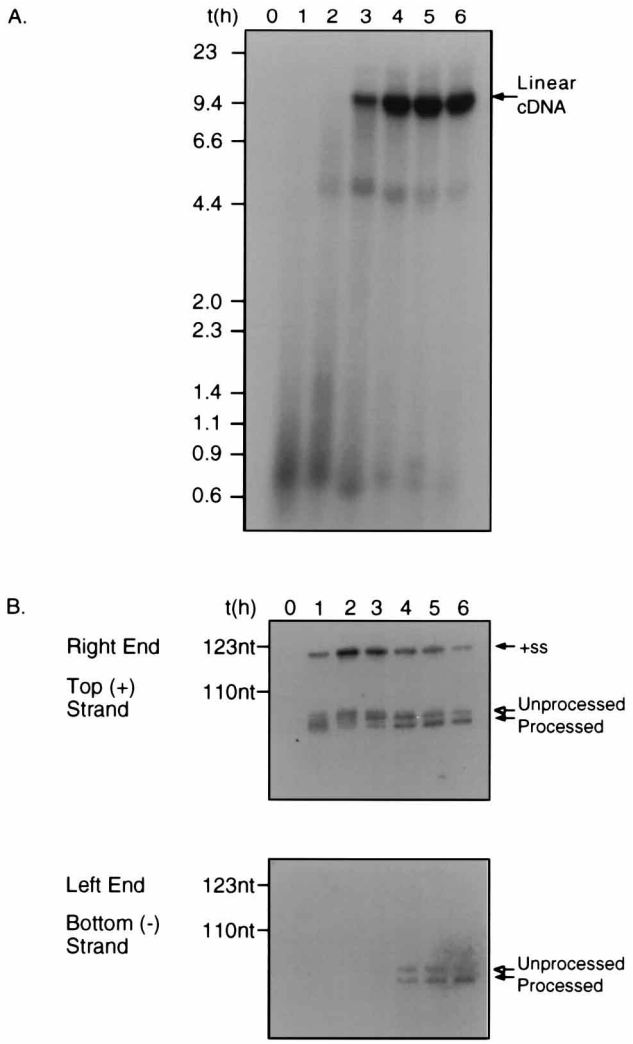

FIG. 5. Kinetics of viral cDNA synthesis and 3′-end processing. Cytoplasmic extracts were prepared at various times after acute HIV-1 infection of SupT1 cells. cDNA was purified from the extracts and analyzed by native agarose gel electrophoresis and denaturing gel electrophoresis. (A) Southern blot of a native agarose gel. Viral cDNA was detected with an LTR probe. (B) Southern blot of a denaturing gel to analyze the viral cDNA end structures. Purified cDNAs were cleaved with *Hae*III and *Hin*dIII and analyzed as in Fig. 4. Extracts prepared 3 h after infection contained detectable PIC integration activity; those prepared 4 to 6 h after infection showed abundant integration activity (data not shown). +ss, plus-strand strong stop DNA; nt, nucleotides.

The 5′ ends of viral cDNA in PICs were also protected from nuclease attack (Fig. 6C). When present in functional PICs, both viral cDNA 5′ ends were resistant to digestion with λexo, but deproteinized DNA was digested. For unknown reasons, λexo always failed to fully digest naked viral DNA even at high concentrations of enzyme, but there was consistently more digestion of naked DNA than of PIC DNA (e.g., compare lanes 1 and 7).

Protein components of PICs. The protein composition of PICs was investigated following partial purification. PICs were partially purified by precipitation in buffers of low ionic strength and then subjected to gel filtration (18, 19). These steps remove >95 and >90% of total proteins when tested separately; when they are used in combination, the total protein concentration drops below the level detectable by the bicinchoninic acid assay (Pierce) for bulk protein. Previous studies of the composition of PICs have been carried out with much less pure PIC preparations (7, 24) or with PICs purified in the presence of detergents which apparently removed some associated proteins (22). For the experiment in Figure 7, gel filtration was carried out in either 150 or 600 mM KCl. PICs purified in 150 mM KCl retained full integration activity, whereas PICs purified in 600 mM KCl lost activity but could be restored to activity by addition of human HMG I(Y) protein (reference 18 and data not shown). Partially purified PICs were then concentrated by Microcon 100 ultrafiltration, and the viral proteins present were analyzed by Western blotting.

Sequential probing of fractions revealed that matrix protein (MA) and RT were present but that NC protein was absent (Fig. 7). Small amounts of CA could be detected after prolonged exposure in samples gel filtered in 150 mM KCl, but CA was absent after gel filtration in 600 mM KCl. Since PICs can be restored to full activity after the high-salt treatment, these data indicate that CA is dispensable for integration.

Previous studies have identified two further proteins in PICs. Several groups have detected integrase in HIV-1 PICs (7, 18, 22, 24), and our previous work has indicated that a host factor, HMG I(Y) protein, is also present and important for function in vitro (18).

DISCUSSION

Here we present data on the organization and function of PICs from HIV-1-infected cells. We find that PICs are large but that the cDNA must nevertheless be compacted to fit into particles of the size measured. The ends of the viral cDNA are paired and bound by protein, but the bulk of the cDNA is accessible to nucleases, implying that the major specific protein-DNA interactions in PICs occur at the cDNA ends. Proteins cofractionating through several purification steps include IN, MA, RT, and human HMG I(Y). NC was not detected, and only traces of CA were detected, indicating that PICs as isolated represent partially disassembled derivatives of the viral core.

Size of complexes. Biophysical measurements provide a tentative picture of PIC size. Size exclusion chromatography yielded an estimate for the Stokes radius of PICs of about 28 nm. Studies of the sedimentation coefficient of PICs from various retroviruses including HIV have ranged from 160S to 640S (4, 21, 29; unpublished data). Several technical issues could confound these measurements. Interaction of PICs with the column matrix during gel filtration could lead to aberrantly slow migration, leading to underestimates of the Stokes radius. Aggregation of PICs could increase the apparent Stokes radius and alter the apparent sedimentation coefficient. Such effects may explain the wide range of sedimentation values. However, these measurements yield a reasonable picture of PICs as large particles and provide a point of departure for further refinements of these values. The Stokes radius determined for PICs is about half that measured for virions by electron microscopy (25).

DNA packaging in the PIC. A 3.3-μm-long HIV-1 cDNA molecule must be significantly condensed to fit into a 56-nm-diameter PIC, but little is known about the mechanism of this condensation. By contrast with nucleic acids in viral capsids, we find that DNA in HIV-1 PICs is vulnerable to attack by many endonucleases, including micrococcal nuclease and *Bal* 31, as well as DNase I and many restriction endonucleases (data not shown). Similarly, DNA in MoMLV PICs can also be digested by restriction enzymes and micrococcal nuclease (4), and it was suggested that the viral cDNA may either be packaged inside a permeable, capsid-like protein shell or be wrapped around

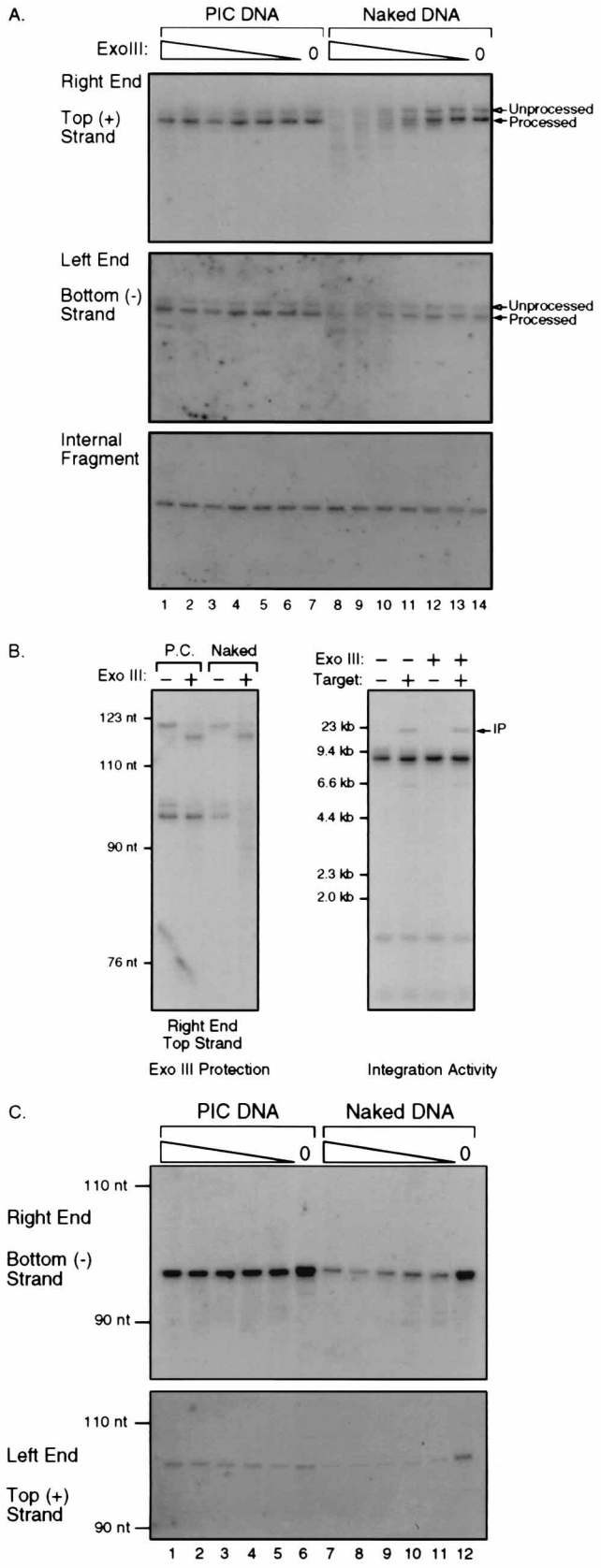

FIG. 6. PIC proteins protect the viral cDNA from exonucleases. (A) exoIII protection assay. Cytoplasmic extracts containing PICs or purified naked viral cDNA were treated with exoIII (threefold serial dilutions of 333 to 1.4 U/ml, or no exoIII), and then the DNAs were repurified, cleaved with *Hae*III and *Hin*dIII, and analyzed by Southern blotting of denaturing gels as in Fig. 4. Probes detected the U3 minus strand, the U5 plus strand, or an internal *Hae*III fragment of 95 nucleotides. (B) exoIII-treated PICs are competent to integrate into a target DNA. PICs or naked cDNA were treated with exoIII (1,000 U/ml) for 10 min at 37°C or left untreated. Before the cDNA end structures were analyzed as in panel A (left panel), aliquots of PICs were removed and incubated with or without 3 μg of circular ϕX174 DNA per ml for 30 min at 37°C to allow integration. Integration activity was assessed by digesting the reaction products with *Bam*HI, which cleaves the viral cDNA at two nearby sites, and analyzing the products by Southern blotting of a native agarose gel (right panel). IP, integration product; nt, nucleotides. (C) λ exonuclease protection. Cytoplasmic extracts containing PICs or purified naked viral cDNA were supplemented with Tris-HCl to 100 mM (pH 9.0) and with Triton X-100 to 0.5% and then digested with λexo (serial threefold dilutions of 830 to 10 U/ml, or no λexo) for 10 min at 37°C. The DNAs were analyzed as in panel A but with probes detecting the U3 plus strand or the U5 minus strand. Similar amounts of DNA were loaded in all lanes (data not shown). nt, nucleotides.

the outside of a protein core (4). Our data do not distinguish between these models, but they are consistent with the idea that the bulk of the viral DNA is loosely associated with a protein scaffold.

Regardless of how the intervening viral cDNA is packaged, it must form a loop because the DNA ends are held together by a protein bridge. Our data do not demonstrate directly that the protein bridge is bound at the ends of the cDNA. In principle, the bridge could be between any two sites in the two fragments generated by restriction enzyme cleavage. Nevertheless, the two ends must be held together in such a way that coupled joining is preserved after the cDNA is severed, and a close interaction of the two ends is the simplest model to explain this finding.

IN may mediate this DNA looping, because it probably binds near each of the DNA ends and is capable of multimerizing (15, 17, 28, 42). Electron microscopic studies of avian myeloblastosis virus IN bound to DNA have identified looped molecules (26). Other scenarios involving additional viral or host proteins are possible, but IN-mediated looping is attractive because such an arrangement would be analogous to that found in the well-studied phage Mu transpososome. For phage Mu integration, the viral DNA ends are held together by a tetramer of transposase (2, 32, 34) and the DNA cleavage and joining reactions require contributions from transposase monomers bound to different phage Mu DNA ends (1, 3, 39, 46). This requirement ensures that only correctly assembled transpososomes are competent to integrate and thereby prevents harmful side reactions such as single-ended integration events. The rarity of single-ended joining by PICs indicates that a related mechanism to control IN activity may also exist in HIV.

Protein binding at the cDNA ends. Here we present the first evidence that in HIV-1 PICs, a protein binds sufficiently close to the ends of HIV-1 cDNA to protect the ends from attack by exonucleases. Two modes of protein-DNA binding can be imagined to explain such protection. A protein might bind directly at or near the ends and thereby prevent exonuclease attack by steric hindrance. Alternatively, since exoIII requires a double-stranded substrate, a protein that destabilizes the DNA helix would also confer protection. However, the cDNA 3′ ends in the PIC are not digested by the single-strand-specific mung bean nuclease (data not shown), and so it is probable that steric hindrance mediates protection.

IN is a strong candidate for the protein bound at the DNA ends. IN is present in PICs (7, 18, 22, 24), and most of the cDNA ends in PICs were processed, indicating that they had encountered IN. Previous studies have shown that after per-

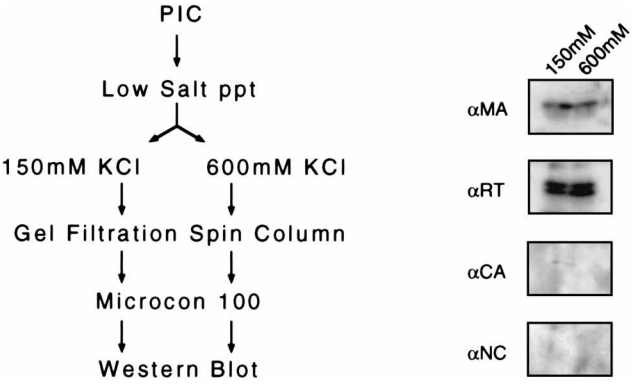

FIG. 7. Some proteins cofractionating with PICs. PICs were partially purified by precipitation in low-ionic-strength buffers and then passed over gel filtration spin columns in the presence of either 150 or 600 mM KCl. Fractions were further purified and concentrated by ultrafiltration in C100 units (Amicon), and aliquots were analyzed by Western blotting with the indicated antibodies. Bound antibody-protein complexes were revealed by probing with ^{125}I-labeled protein A and autoradiography, except for the case of the CA antibody, which was revealed by reaction with donkey anti-rabbit immunoglobulin conjugated to horseradish peroxidase and by enhanced chemiluminescence (Amersham). Fifteen nanograms of NC was readily detectable in control lanes on the blot (data not shown).

forming the terminal cleavage reaction on model DNA substrates in vitro, purified HIV-1 IN remains tightly bound to DNA ends (14, 44). Thus, the simplest interpretation of our data is that in functional PICs, IN binds to and processes the newly completed viral DNA ends and then remains stably associated until a target DNA molecule is encountered.

Composition of PICs. Several further proteins, in addition to IN, cofractionate with PICs. Human HMG I(Y) protein is not only present but required for function in vitro (18), representing another candidate for a DNA binding protein contributing to protection of the ends. RT and MA also cofractionate with PICs. Whether their apparent association contributes to obstructing nuclease digestion is unclear. At present, there is no evidence that RT and MA are directly involved in integration. Previous studies of less pure fractions have led to the proposal that NC is present in PICs (7, 24), but we did not detect NC after more extensive fractionation. It is unclear whether NC cofractionated adventitiously in these earlier studies or whether it was stripped off PICs by the methods used here. CA is present in trace amounts in fractions gel filtered in 150 mM KCl and absent in fractions gel filtered in 600 mM KCl, indicating that it is at most loosely associated with PICs and unimportant for integration. Taken together, these data indicate that active PICs contain a specific subset of the proteins present in the viral core upon entry.

Role for the IN terminal cleavage reaction: removal of non-template-encoded nucleotides added by RT. The kinetic analysis of events at the cDNA ends (Fig. 5B) allowed us to suggest an additional purpose for the IN-catalyzed terminal cleavage reaction. Reverse transcription yields a blunt cDNA end, but prior to integration, 2 nucleotides are removed from each 3′ end to expose the recessed 5′-CA-3′ sequence that becomes joined to target. Previous work has suggested that IN may bind tightly to the resulting processed end (14), supporting the view that forming this stable intermediate is a use of this cleavage reaction. However, work in a related retrotransposon system suggests that active complexes may form without this terminal cleavage step (13).

We propose that an additional purpose of the terminal cleavage reaction may be to form a precise end at otherwise heterogeneous DNA termini. We found that 2 h after initiating infection, DNA 3′ ends 1 nucleotide longer than the expected length are the most prominent form for the U5 plus strand (Fig. 5B). Purified HIV-1 RT is known to add non-template-encoded bases to DNA ends in vitro, probably accounting for the added sequences, and the U5 end is a favored substrate (35, 37). The terminal cleavage reaction removes such extra bases, and the cleaved form predominates 4 to 6 h after infection. An important role of the terminal cleavage reaction may be the removal of these non-template-encoded bases, which, if present as extensions on the cDNA 3′ end, might block integration.

ACKNOWLEDGMENTS

We thank Amy Espeseth for comments on the manuscript, members of the Bushman laboratory for suggestions and comments on the manuscript, and Leslie Barden for artwork.

This work was supported by National Institutes of Health grant RO1 AI34786 and AI37489 (F.D.B.) and NRSA fellowship 1F32 AI09068-01 (M.D.M.). F.D.B. is a Scholar of the Leukemia Society of America.

REFERENCES

1. **Aldaz, H., E. Schuster, and T. A. Baker.** 1996. The interwoven architecture of the Mu transposase couples DNA synapsis to catalysis. Cell **85:**257–269.
2. **Baker, T. A., E. Kremenstova, and L. Luo.** 1994. Complete transposition requires four active monomers in the Mu transposase tetramer. Genes Dev. **8:**2416–2428.
3. **Baker, T. A., M. Mizuuchi, H. Savilahti, and K. Mizuuchi.** 1993. Division of labor among monomers within the Mu transposase tetramer. Cell **74:**723–733.
4. **Bowerman, B., P. O. Brown, J. M. Bishop, and H. E. Varmus.** 1989. A nucleoprotein complex mediates the integration of retroviral DNA. Genes Dev. **3:**469–478.
5. **Brown, P. O., B. Bowerman, H. E. Varmus, and J. M. Bishop.** 1987. Correct integration of retroviral DNA in vitro. Cell **49:**347–356.
6. **Brown, P. O., B. Bowerman, H. E. Varmus, and J. M. Bishop.** 1989. Retroviral integration: structure of the initial covalent complex and its precursor, and a role for the viral IN protein. Proc. Natl. Acad. Sci. USA **86:**2525–2529.
7. **Bukrinsky, M. I., N. Sharova, T. L. McDonald, T. Pushkarskaya, G. W. Tarpley, and M. Stevenson.** 1993. Association of integrase, matrix, and reverse transcriptase antigens of human immunodeficiency virus type 1 with viral nucleic acids following acute infection. Proc. Natl. Acad. Sci. USA **90:**6125–6129.
8. **Bushman, F., and M. D. Miller.** 1997. Tethering human immunodeficiency virus type 1 preintegration complexes to target DNA promotes integration at nearby sites. J. Virol. **71:**458–464.
9. **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.
10. **Bushman, F. D., T. Fujiwara, and R. Craigie.** 1990. Retroviral DNA integration directed by HIV integration protein in vitro. Science **249:**1555–1558.
11. **Charneau, P., and F. Clavel.** 1991. A single-stranded gap in human immunodeficiency virus unintegrated linear DNA defined by a central copy of the polypurine tract. J. Virol. **65:**2415–2421.
12. **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.
13. **Eichinger, D. J., and J. D. Boeke.** 1990. A specific terminal structure is required for Ty1 transposition. Genes Dev. **4:**324–330.
14. **Ellison, V., and P. O. Brown.** 1994. A stable complex between integrase and viral DNA ends mediates human immunodeficiency virus integration *in vitro*. Proc. Natl. Acad. Sci. USA **91:**7316–7320.
15. **Ellison, V., J. Gerton, K. A. Vincent, and P. O. Brown.** 1995. An essential interaction between distinct domains of HIV-1 integrase mediates assembly of the active multimer. J. Biol. Chem. **270:**3320–3326.
16. **Ellison, V. H., H. Abrams, T. Roe, J. Lifson, and P. O. Brown.** 1990. Human immunodeficiency virus integration in a cell-free system. J. Virol. **64:**2711–2715.
17. **Engelman, A., F. D. Bushman, and R. Craigie.** 1993. Identification of discrete functional domains of HIV-1 integrase and their organization within an active multimeric complex. EMBO J. **12:**3269–3275.
18. **Farnet, C., and F. D. Bushman.** 1997. HIV-1 cDNA integration: requirement of HMG I(Y) protein for function of preintegration complexes in vitro. Cell **88:**483–492.
19. **Farnet, C., R. Lipford, B. Wang, and F. D. Bushman.** 1996. Differential inhibition of HIV-1 preintegration complexes and purified integrase protein by small molecules. Proc. Natl. Acad. Sci. USA **93:**9742–9747.

20. **Farnet, C. M., and F. D. Bushman.** 1996. HIV cDNA integration: molecular biology and inhibitor development. AIDS **10**(Suppl. A)**:**3–11.
21. **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.
22. **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.
23. **Fujiwara, T., and K. Mizuuchi.** 1988. Retroviral DNA integration: structure of an integration intermediate. Cell **54:**497–504.
24. **Gallay, P., S. Swingler, J. Song, F. Bushman, and D. Trono.** 1995. HIV nuclear import is governed by the phosphotyrosine-mediated binding of matrix to the core domain of integrase. Cell **17:**569–576.
25. **Gelderblom, H. R., E. H. S. Hausmann, M. Ozel, G. Pauli, and M. A. Koch.** 1987. Fine structure of human immunodeficiency virus (HIV) and immunolocalization of structural proteins. Virology **156:**171–176.
26. **Grandgenett, D. P., R. B. Inman, A. C. Vora, and M. L. Fitzgerald.** 1993. Comparison of DNA binding and integration half-site selection by avian myeloblastosis virus integrase. J. Virol. **67:**2628–2636.
27. **Hansen, M., and F. D. Bushman.** 1997. Human immunodeficiency virus type 2 preintegration complexes: activities in vitro and response to inhibitors. J. Virol. **71:**3351–3356.
28. **Jones, K. S., J. Coleman, G. W. Merkel, T. M. Laue, and A. M. Skalka.** 1992. Retroviral integrase functions as a multimer and can turn over catalytically. J. Biol. Chem. **267:**16037–16040.
29. **Karageorgos, L., P. Li, and C. Burrell.** 1993. Characterization of HIV replication complexes early after cell-to-cell infection. AIDS Res. Hum. Retroviruses **9:**817–823.
30. **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.
31. **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.
32. **Lavoie, B. D., B. S. Chan, R. G. Allison, and G. Chaconas.** 1991. Structural aspects of a higher order nucleoprotein complex: induction of an altered DNA structure at the Mu-host junction of the Mu type 1 transpososome. EMBO J. **10:**3051–3059.
33. **Miller, M. D., B. Wang, and F. D. Bushman.** 1995. HIV-1 preintegration complexes containing discontinuous plus strands are competent to integrate in vitro. J. Virol. **69:**3938–3944.
34. **Mizuuchi, M., T. A. Baker, and K. Mizuuchi.** 1992. Assembly of the active form of the transposase-Mu DNA complex: a critical control point in Mu transposition. Cell **70:**303–311.
34a. **Myrick, K., and C. Farnet.** Unpublished data.
35. **Patel, P. H., and B. D. Preston.** 1994. Marked infidelity of human immunodeficiency virus type 1 reverse transcriptase at RNA and DNA template ends. Proc. Natl. Acad. Sci. USA **91:**549–553.
36. **Pauza, C. D.** 1990. Two bases are deleted from the termini of HIV-1 linear DNA during integrative recombination. Virology **179:**886–889.
37. **Peliska, J. A., and S. J. Benkovic.** 1992. Mechanism of DNA strand transfer reactions catalyzed by HIV-1 reverse transcriptase. Science **258:**1112–1118.
38. **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 of IN function and terminal DNA sequence. Cell **58:**47–54.
39. **Savilahti, H., and K. Mizuuchi.** 1996. Mu transpositional recombination: donor DNA cleavage and strand transfer in *trans* by the Mu transposase. Cell **85:**271–280.
40. **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.
41. **Siegel, L. M., and K. J. Monty.** 1966. Determination of molecular weights and frictional ratios of proteins in impure systems by use of gel filtration and density gradient centrifugation. Application to crude preparations of sulfite and hydroxylamine reductases. Biochim. Biophys. Acta **112:**346–362.
42. **van Gent, D. C., C. Vink, A. A. M. Oude Groeneger, and R. H. A. Plasterk.** 1993. Complementation between HIV integrase proteins mutated in different domains. EMBO J. **12:**3261–3267.
43. **Vink, C., and R. H. A. Plasterk.** 1993. The human immunodeficiency virus integrase protein. Trends Genet. **9:**433–437.
44. **Vink, C., R. A. Puras Lutzke, and R. H. A. Plasterk.** 1994. Formation of a stable complex between the human immunodeficiency virus integrase protein and viral DNA. Nucleic Acids Res. **22:**4103–4110.
45. **Vora, A. C., M. McCord, M. L. Fitzgerald, R. B. Inman, and D. P. Grandgenett.** 1994. Efficient concerted integration of retrovirus-like DNA in vitro by avian myeblastosis virus integrase. Nucleic Acids Res. **22:**4454–4461.
46. **Yang, J.-Y., M. Jayaram, and R. M. Harshey.** 1996. Positional information within the Mu transposase tetramer: catalytic contributions of individual monomers. Cell **85:**447–455.