THE JOURNAL OF BIOLOGICAL CHEMISTRY
© 2003 by The American Society for Biochemistry and Molecular Biology, Inc.

Vol. 278, No. 1, Issue of January 3, pp. 372–381, 2003
Printed in U.S.A.

HIV-1 Integrase Forms Stable Tetramers and Associates with LEDGF/p75 Protein in Human Cells*

Received for publication, September 10, 2002, and in revised form, October 23, 2002
Published, JBC Papers in Press, October 28, 2002, DOI 10.1074/jbc.M209278200

Peter Cherepanov‡§, Goedele Maertens¶‖, Paul Proost‡, Bart Devreese, Jozef Van Beeumen**, Yves Engelborghs¶, Erik De Clercq‡, and Zeger Debyser‡ ‡‡**

*From the ‡Rega Institute for Medical Research, Katholieke Universiteit Leuven, Minderbroedersstraat 10, B-3000, Leuven, Belgium, the ¶Laboratory of Biomolecular Dynamics, Katholieke Universiteit Leuven, B-3001, Heverlee, Belgium, and the **Laboratory for Protein Biochemistry & Protein Engineering, Ghent University, K. L. Ledeganckstraat 35, B-9000 Ghent, Belgium*

We studied human immunodeficiency virus, type 1 (HIV-1) integrase (IN) complexes derived from nuclei of human cells stably expressing the viral protein from a synthetic gene. We show that in the nuclear extracts IN exists as part of a large distinct complex with an apparent Stokes radius of 61 Å, which dissociates upon dilution yielding a core molecule of 41 Å. We isolated the IN complexes from cells expressing FLAG-tagged IN and demonstrated that the 41 Å core is a tetramer of IN, whereas 61 Å molecules are composed of IN tetramers associated with a cellular protein with an apparent molecular mass of 76 kDa. This novel integrase interacting protein was found to be identical to lens epithelium-derived growth factor (LEDGF/p75), a protein implicated in regulation of gene expression and cellular stress response. HIV-1 IN and LEDGF co-localized in the nuclei of human cells stably expressing IN. Furthermore, recombinant LEDGF robustly enhanced strand transfer activity of HIV-1 IN *in vitro*. Our findings indicate that the minimal IN molecule in human cells is a homotetramer, suggesting that at least an octamer of IN is required to accomplish coordinated integration of both retroviral long terminal repeats and that LEDGF is a cellular factor involved in this process.

Establishment of the provirus, a DNA copy of the viral genome integrated into the host cell chromosome, is an obligatory step in retroviral replication. Moreover, stable integration into the human genome is the primary reason for the persistence of the human immunodeficiency virus (HIV),[1] which leads to AIDS. Therefore, HIV IN, the enzyme orchestrating the insertion of the DNA replica of the viral genome into the cellular chromosomal DNA, is an important target for antiretroviral therapy (1–3).

Mechanistically and structurally, retroviral integrases are similar to the well studied prokaryotic Mu phage and Tn5 transposases and belong to a family of DNA strand transferases that catalyze DNA cutting and joining via direct transesterification (reviewed in Refs. 4–7). In the course of retroviral infection, HIV IN performs two enzymatic reactions using the viral DNA as substrate. The first reaction is the removal of the 3′-GT dinucleotides from both LTRs (the 3′-end processing reaction). The second reaction is the insertion of the recessed viral DNA ends into the opposite strands of the target DNA, whereby the 3′ hydroxyls of the processed LTR ends attack two phosphodiester bonds in the target DNA molecule (the strand transfer or integration reaction). *In vivo*, insertion of the two viral LTRs takes place in a coordinated fashion across the major groove of the target DNA (concerted or full site integration). As a result, the integrated provirus is flanked by two 5-nucleotide gaps as well as two unmatched 5′-AC dinucleotides, which are then repaired by cellular enzymes.

The stoichiometry of the native retroviral IN complex has not been established. Based on the available crystal structure information, it appears that at least a tetramer or even an octamer of IN would be necessary to accomplish concerted integration of both LTRs (8–10). The distance between the target DNA phosphates (~18 Å) presents one important constraint for the modeling of the active IN multimer. Monomers, dimers, and tetramers were observed in preparations of recombinant HIV and avian sarcoma virus integrases (11–14). The presence of octamers and larger complexes has been suggested in some reports (15, 16). Virion-associated HIV IN was also shown to be in a multimeric form, whereby dimers and higher order complexes appeared to be stabilized by disulfides, although the complexes were not studied under native conditions (17). Although recombinant HIV IN forms enzymatically active multimers (18, 19), reconstitution of the integration reaction *in vitro* using recombinant enzyme preparations results in predominantly uncoupled (half-site) integration of LTR DNA substrates.

In vivo, retroviral DNA integration is preceded by the assembly of a stable and compact preintegration complex (PIC) that contains a DNA copy of the viral genome associated with viral and cellular proteins. Several cellular proteins have been suggested to play auxiliary roles during retroviral integration. Thus, barrier-to-autointegration factor (BAF) has been reported to protect Moloney murine leukemia virus PICs against suicidal self-integration (20). Another cellular protein, HMG-I(Y) was

* This work was supported in part by the Concerted Research Action Fund of the Flemish Government. The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

§ To whom correspondence may be addressed. Tel.: 32-16-332176; Fax: 32-16-337340; E-mail: Peter.Cherepanov@uz.kuleuven.ac.be.
‖ Aspirant of the Belgian National Fund for Scientific Research.
‡‡ Postdoctoral Fellow of the Belgian National Fund for Scientific Research (F.W.O. Flanders). To whom correspondence may be addressed. Tel.: 32-16-332176; Fax: 32-16-337340; E-mail: Zeger.Debyser@uz.kuleuven.ac.be.

[1] The abbreviations used are: HIV, human immunodeficiency virus; CHAPS, 3-[(3-cholamidopropyl)dimethyl-ammonio]propanesulfonate; DTSSP, 3,3′-dithiobis[sulfosuccinimidyl propionate]; HDGF, hepatoma-derived growth factor; IN, integrase; IN_f, FLAG-tagged IN; IN^s, synthetic integrase gene; LEDGF, lens epithelium-derived growth factor; LTR, long terminal repeat; MS, mass spectrometry; PEG, polyethylene glycol; PIC, preintegration complex; Pipes, 1,4-piperazinediethanesulfonic acid; BAF, barrier-to-autointegration factor; Ini1, integrase interactor 1; LC, liquid chromatography; DNA PKcs, catalytic subunit of DNA-dependent protein kinase.

This paper is available on line at http://www.jbc.org

This is an Open Access article under the CC BY license.

found in HIV PICs and appeared to be essential for their integration activity *in vitro* (21, 22). Conversely, BAF could substitute for HMG-I(Y) at least *in vitro*, partially restoring integration activity of salt-denatured HIV-1 PICs (23). Yet, it remains to be shown that BAF co-fractionates with retroviral PICs. Both BAF and HMG-I(Y) are small DNA-binding proteins able to bridge and deform DNA molecules and are thought to play structural roles within retroviral PICs, possibly by juxtaposing both LTRs. Similarly, Mu phage transposase and λ phage integrase require the DNA-bending host proteins IHF and/or HU to form committed synaptic complexes (7, 24). Another potential co-factor for HIV integration, the integrase interactor 1 (Ini1), was originally discovered in a yeast two-hybrid screen for human proteins interacting with HIV-1 IN (25). Cellular Ini1 is a subunit of the 2-MDa SWI/SNF chromatin-remodeling complex (26). It has been proposed that Ini1 plays a role during retroviral replication by directing the PICs to open chromatin regions or by modulating expression of the integrated provirus. Recent studies demonstrated that green fluorescent protein-tagged Ini1 was exported from the nuclei of infected cells and co-localized with incoming subviral particles (27). Ini1 has also been reported to enhance the release of infectious HIV particles (28).

Using a synthetic gene, we have been able to achieve efficient expression of HIV-1 IN in human cells (29). We have now characterized HIV-1 IN protein complexes present in nuclear extracts from cells stably expressing this viral protein. We now report the first HIV integrase-interacting protein that forms a *distinct* complex with IN in human cells. Our results also provide an insight into the oligomeric state of intracellular HIV IN, indicating that the minimal cellular IN complex is a homotetramer.

EXPERIMENTAL PROCEDURES

Recombinant DNA—The HIV-1 integrase expression constructs were based on the episomal pCEP4 vector (Invitrogen). The plasmid pCEP-IN^sala is almost identical to the published pCEP-IN^s plasmid (29), with the sole difference that the Gly codon in the second position of the synthetic open reading frame was mutated to Ala. As a result, the construct expressed native HIV-1 IN with an addition of Met-Ala dipeptide at the N terminus. To create the FLAG epitope-tagged IN expression construct pCEP-IN^salaFLAG, the IN^s gene from pCEP-IN^sala was amplified in two consecutive steps with the sense primer 5′-GGCTAGATATCACTAGCAACCTCAAACAG plus the two antisense primers 5′-GTCGTCCTTGTAATCGCCGTCCTCATCTTGACGAGAG and 5′-GGCGCTCGAGTTACTTGTCATCGTCGTCCTTGTAATCGC; the resulting PCR fragment was digested with *Xho*I and cloned between the *Pvu*II and *Xho*I sites of pCEP4. This plasmid expressed HIV-1 IN carrying the C-terminal FLAG epitope (DYKDDDDK). The plasmid pRP1012, for bacterial expression of His_6-tagged HIV-1 IN, was a gift of Dr. R. Plasterk (Netherlands Cancer Institute, Amsterdam, The Netherlands). To obtain pCP6H75, the plasmid used for bacterial expression of His_6-tagged LEDGF, a DNA fragment coding for LEDGF/p75 was amplified from a sample of total HeLa RNA by reverse transcription-PCR using the primers 5′-GGCCGGATCCGACTCGCGATTTCAAACCTGGAGAC and 5′-CCGCGAATTCTAGTT ATCTAGTGTAGAATCCTTC. The PCR fragment was digested with *Bam*HI and *Eco*RI and subcloned into pRSETB (Invitrogen). To prepare the mini-HIV DNA substrate for IN, the plasmid pU3U5 was digested with *Sca*I (30).

Cells—The human embryonic kidney cells expressing SV40 large T antigen, 293T were obtained from Dr. O. Danos (Evry, France). The cells were grown in Dulbecco's modified Eagle's medium (Invitrogen) supplemented with 10% fetal calf serum, 2 mM glutamine, and 20 μg/ml gentamicin at 37 °C in 5% CO_2 humidified atmosphere. To establish stable cell lines, 293T cells were transfected by electroporation with the integrase expression constructs and selected with 200 μg/ml of hygromycin B (Invitrogen). For radioactive immunoprecipitation experiments, the cells were labeled in methionine/cysteine-free Dulbecco's modified Eagle's medium (Invitrogen) supplemented with 10% dialyzed fetal calf serum plus 0.1 mCi/ml of TRAN^{35}S-LABEL (ICN Biomedicals, Asse-Relegem, Belgium) for 24 h.

Preparation of Nuclear Extracts—293T-IN^sala or 293T-IN^salaFLAG cells grown to a confluency of 80–90% were harvested by trypsinization, washed with phosphate-buffered saline, and resuspended in modified CSK buffer (10 mM Pipes pH 6.8, 10% (w/v) sucrose, 1 mM dithiothreitol, 1 mM $MgCl_2$ plus the EDTA-free protease inhibitor mixture (Roche Molecular Biochemicals)) (31) containing 100 mM NaCl (referred to as 100mCSK buffer). The cells were lysed for 10 min on ice with 0.5% Nonidet P-40, and the nuclei were pelleted and washed with 100mCSK. To extract IN, the nuclei were resuspended in 400mCSK buffer (same as 100mCSK, but containing 400 mM NaCl) and left on ice for 5 min; the chromatin was removed by centrifugation at 7,500 rpm for 2 min. The total protein content of the nuclear extracts was measured using the BCA protein assay (Pierce), with bovine serum albumin as the standard.

Chemical Cross-linking and Gel Filtration Chromatography—The nuclear extracts were diluted using 400mCSK buffer to adjust the total protein concentration. DTSSP (Pierce) was dissolved in water immediately prior to the experiment. The cross-linking reactions were allowed to proceed for 15 min at room temperature and were terminated by the addition of ¼ volume of 4× SDS sample buffer (200 mM Tris, pH 6.8, 4% SDS, and 40% (v/v) glycerol) and further incubation at room temperature for 20 min.

Nuclear extracts and affinity-purified FLAG-tagged IN (IN_f) were fractionated on a Superdex 200 HR 10/30 gel filtration column (Amersham Biosciences); 400mCSK buffer was used in all of the chromatography experiments. The column was operated at 0.6 ml/min, 4 °C and calibrated using low and high molecular weight gel filtration standards from Amersham Biosciences (blue dextran; thyroglobulin, R_S/molecular mass 85 Å/669 kDa; ferritin, 61 Å/440 kDa; catalase, 52.2 Å/232 kDa; aldolase, 48.1 Å/158 kDa; bovine serum albumin, 35.5 Å/67 kDa; chymotrypsinogen A, 20.9 Å/25 kDa). The sample volume was kept at 200 μl; fractions of 300 μl were collected and analyzed by Western blotting using polyclonal anti-IN antibodies. When necessary, gel filtration fractions were concentrated by precipitation with trichloroacetic acid. Stokes radii (R_S) and approximate molecular masses of the IN complexes were determined from their experimental partition coefficients (K_{av}) as described (32).

Western Blotting and Immunoprecipitation—The gradient 4–12 and 4–20% Novex Tris-glycine gels were purchased from Invitrogen. The proteins were transferred onto polyvinylidene difluoride membranes (Bio-Rad); detection was done with ECL+ (Amersham Biosciences). The rabbit polyclonal anti-HIV-1 IN antibody has been described previously (29). The anti-FLAG M2 monoclonal antibody was from Sigma-Aldrich, the monoclonal anti-DNA-PKcs Ab-4 mixture was from NeoMarkers (Fremont, CA), and the monoclonal anti-LEDGF p75/p52 was from BD Biosciences (Erembodegem, Belgium). The affinity purified anti-hMCM3 polyclonal antibody (33) was a kind gift from Dr. R. Knippers (University of Konstanz, Konstanz, Germany). A combination of prestained molecular weight markers (New England Biolabs, Hitchin, Hertfordshire, UK) and Mark12 (Invitrogen) was used to estimate molecular weights of the cross-linking products and p76. DNA PKcs detected in a 293T nuclear lysate sample using the anti-DNA PKcs Ab4 antibody served as the 470-kDa marker in some Western blots.

In the initial immunoprecipitation experiments, 30 μl of protein G-agarose (Roche Molecular Biochemicals) and 1–3 μg of the anti-FLAG M2 antibody was added to the nuclear extracts prepared in 400mCSK and diluted to obtain total protein concentration of 200 μg/ml. The suspension was stirred at 4 °C overnight (12–18 h). The agarose beads were washed once with 400mCSK and four times with 100mCSK plus 0.1% Nonidet P-40. The protein was eluted in 400mCSK buffer by the addition of 200 μg/ml FLAG peptide (Sigma-Aldrich) or in SDS-PAGE sample buffer. To purify IN_f-p76 complexes, immunoprecipitation was carried out using undiluted nuclear extracts (600–1000 μg/ml total protein) for 3–5 h. To identify the p76 protein by N-terminal sequencing and mass spectrometry, the procedure was upscaled. 293T-IN^salaFLAG cells grown to confluency on five 500-cm^2 dishes (VWR International, Leuven, Belgium) were harvested and lysed with 0.5% Nonidet P-40. IN complexes were extracted from the nuclear pellets into 13 ml of 400mCSK buffer and incubated with 300 μl of protein G-agarose beads and 40 μg of the anti-FLAG M2 antibody for 4.5 h. The IN_f complexes were eluted in 700 μl of 400mCSK buffer with 200 μg/ml FLAG peptide.

N-terminal Sequencing and Mass Spectrometry—Immunopurified IN_f-p76 complexes were precipitated with trichloroacetic acid and redissolved in SDS-PAGE sample buffer. Approximately 3 μg of the p76 protein, electroblotted onto a Sequi-Blot polyvinylidene difluoride membrane (Bio-Rad) from an SDS-PAGE gel, was subjected to Edman degradation on a pulsed liquid phase Procise 491cLC protein sequencer (Applied Biosystems, Lennik, Belgium). For mass spectrometry analy-

sis the Coomassie Blue-stained band of p76 was cut from an SDS-PAGE gel, destained in a 200 mM ammonium bicarbonate, 50% acetonitrile, air-dried, and soaked in 8 μl of trypsin solution (16 ng of trypsin (Promega) in 50 mM ammonium bicarbonate) on ice for 20 min. Following overnight digestion at 37 °C, the supernatant was recovered, and the gel slice was extracted twice using 60% acetonitrile, 0.1% formic acid. The extracts and the supernatant were pooled and dried in a Speedvac concentrator. The peptides were redissolved in 0.1% formic acid and analyzed by on-line nanoflow high performance liquid chromatography tandem mass spectrometry (LC/MS/MS) on an UltiMate capillary LC system (LC-Packings, Amsterdam, The Netherlands) coupled to a Q-Tof mass spectrometer (Micromass, Manchester, UK) equipped with an electrospray ionization source. All of the spectra were processed using the MassLynx and MaxEnt software delivered with the mass spectrometer.

Indirect Immunofluorescence Microscopy—293T-IN^{s}alaFLAG cells grown in Lab-Tek II glass chamber slides (VWR International) were fixed with 4% formaldehyde in phosphate-buffered saline for 10 min and permeabilized in ice-cold methanol. The cells were further blocked in phosphate-buffered saline supplemented with 20 mM ammonium chloride and 10% fetal calf serum and incubated with rabbit polyclonal anti-FLAG antibodies (diluted 1:10.000 in phosphate-buffered saline, 10% fetal calf serum) (Sigma-Aldrich) and monoclonal anti-LEDGF (1:300) or anti-DNA PKcs (1:300) followed by Alexa-555 anti-rabbit and Alexa-488-conjugated anti-mouse IgG antibodies (Molecular Probes, Leiden, The Netherlands). The nuclear DNA was labeled with 5 μM To-Pro3 iodide (Molecular Probes). Confocal laser scanning fluorescent microscopy and imaging was carried with an LSM510 system (Carl Zeiss, Zaventem, Belgium) using a 488-nm argon ion laser with a 505–530-nm band pass filter for Alexa-488, a 543-nm HeNe laser with a 565–615-nm filter for Alexa-555, and a 633-nm HeNe laser with a low pass 650-nm filter for ToPro-3. All of the acquisitions were done in the multi-track mode.

Recombinant Proteins—The His_6-tagged HIV-1 IN was produced from the plasmid pRP1012 in the Endo I-free host *Escherichia coli* strain PC1 (BL21(DE3), Δ*endA*::Tc^{R}, pLysS) (30). The protein was purified from the soluble fraction by chromatography on nickel-nitrilotriacetic acid-agarose (Qiagen) and Heparin-Sepharose (Amersham Biosciences) in the presence of 7.5 mM CHAPS (Sigma-Aldrich). The His_6 tag was removed by incubation of the purified protein with thrombin (Novagen). The His_6-tagged LEDGF was expressed from the plasmid pCP6H75 in PC1 cells by induction with 1 mM isopropylthiogalactopyranoside in LB medium at 29 °C. The cells harvested 3 h post induction were disrupted using a French press in 1 M NaCl, 50 mM Tris pH 7.4. The soluble His_6-tagged LEDGF protein was enriched by chromatography on nickel-nitrilotriacetic acid-agarose and further purified on a 1-ml HiTrap Heparin-Sepharose column (Amersham Biosciences). The protein was eluted from the Heparin-Sepharose column using a linear NaCl gradient in 30 mM Tris, pH 7.0. Peak fractions collected at ~800 mM NaCl were pooled and concentrated using Centricon-30 (Millipore, Brussels, Belgium). The purified protein supplemented with 5 mM dithiothreitol and 10% glycerol was kept frozen at −80 °C.

RESULTS

HIV-1 IN Is Present in the Insoluble Nuclear Fraction—The 293T-IN^{s}ala cell line used in this work was similar to the previously reported 293T-IN^{s} (29), except that it expressed HIV-1 IN with the Met-Ala dipeptide (instead of Met-Gly) at its N terminus. This change was introduced to prevent potential myristoylation of the protein and did not affect either cell line stability or IN expression levels. The integrase protein was nuclear in both cell lines as determined by indirect immunofluorescence microscopy (data not shown) (29). After lysis of 293T-IN^{s}ala cells with digitonin or Nonidet P-40 and centrifugation, most of the IN protein was retained in the nuclear pellet (Fig. 1*A*). Nonidet P-40 permeabilizes both the plasma membrane and the nuclear envelope. Hence, the bulk of IN present in the cell is stably associated with insoluble nuclear structures. IN could be readily extracted from the Nonidet P-40-permeabilized nuclei in high salt conditions (Fig. 1*B*).

We observed no elution of IN when nuclei prepared from 293T-IN^{s}ala cells were treated with DNase I (Fig. 1*C*) that completely digested nuclear DNA to fragments of less then 200 bp (data not shown). In accordance with Fujita *et al.* (31), MCM3, a chromosomal replication factor, was removed from the nuclei by gentle DNase I treatment (Fig. 1*C*). The same salt concentration was required to extract IN from the nuclease-digested nuclei (data not shown). However, exposure of the DNase-digested nuclei to low ionic strength conditions led to efficient elution of the protein (Fig. 1*D*). Nondigested nuclei did not release any detectable amount of IN in the hypotonic medium (Fig. 1*D*). These results indicate that IN is associated both with chromosomal DNA and with some other nuclear components, which are destabilized in low ionic strength conditions. Similar results were obtained when 293T-IN^{s}ala nuclei were exposed to micrococcal nuclease (data not shown).

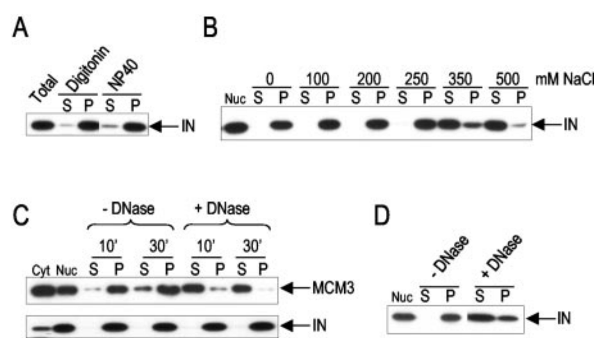

FIG. 1. **Extraction of IN from nuclei of 293T-IN^{s}ala cells.** *A*, 293T-IN^{s}ala cells were lysed in 100mCSK buffer in the presence of 4 μg/ml digitonin or 0.5% Nonidet P-40 on ice for 10 min. After centrifugation, the supernatant (*S*) and nuclear pellet (*P*) fractions were recovered and analyzed by Western blotting with an anti-IN antibody. The *first lane* contained the total cell extract. *B*, 293T-IN^{s}ala cells were lysed in 100mCSK buffer supplemented with 0.5% Nonidet P-40 on ice for 10 min, and the extracted nuclei were resuspended in CSK buffer containing 0–500 mM NaCl. Following centrifugation, supernatants (*S*) and nuclear pellets (*P*) were analyzed by Western blotting with an anti-IN antibody. The total nuclear protein was loaded in the *first lane* (*Nuc*). *C*, Nonidet P-40-permeabilized nuclei from the 293T-IN^{s}ala cells, prepared as above, were incubated in 100mCSK buffer with (+ *DNase I*) or without (− *DNase I*) DNase I (250 units/ml) at 25 °C for 10 min (*10′*) or 30 min (*30′*). The supernatants (*S*) and pellets (*P*) were separated in an 11% SDS-PAGE gel, of which the upper part of which was used for the immunoblot to detect MCM3 (91 kDa) and the lower part was used to detect IN (32 kDa). The total cytoplasmic and nuclear protein fractions were loaded in the *first* (*Cyt*) and the *second* (*Nuc*) *lanes*, respectively. *D*, Nonidet P-40-permeabilized nuclei from 293T-IN^{s}ala cells were incubated in 100mCSK buffer with or without DNase I (250 units/ml) for 10 min, pelleted by centrifugation, and resuspended in ice-cold hypotonic buffer (2 mM EDTA, 2 mM Hepes, pH 7.5). After centrifugation, supernatants (*S*) and pellets (*P*) were analyzed by Western blotting with anti-IN antibodies. The *first lane* (*Nuc*) contained total nuclear protein.

Chemical Cross-linking of IN Complexes Present in Nuclear Extracts—We used DTSSP, an amine-specific *N*-hydroxysuccinimid ester, to cross-link protein complexes present in the nuclear extracts of 293T-IN^{s}ala cells. The nuclear proteins were extracted from Nonidet P-40-permeabilized 293T-IN^{s}ala nuclei using cytoskeleton (CSK) buffer supplemented with 400 mM NaCl. The total protein concentration was adjusted to 100, 20, and 4 μg/ml; the samples were incubated with DTSSP and separated in a nonreducing 4–12% SDS-PAGE gel. IN-containing cross-linking adducts were detected by Western blotting using polyclonal anti-IN antibodies. A typical result is shown in Fig. 2*A*. In the non-cross-linked samples (*lanes 2*, *6*, and *10*) as well as in the samples cross-linked in the presence of SDS (*lane 1*), only IN monomer and a band corresponding to IN dimer (~60 kDa) were apparent. Addition of 0.1–2 mM DTSSP yielded cross-linked complexes of 60 kDa (p60^{cl}), 150 kDa (the p150^{cl} band, clearly visible in *lanes 4*, *8*, *9*, *12*, and *13*), 250–300 kDa (p300^{cl}; *lanes 4*, *5*, and *9*), and less resolved higher molecular mass species (*lane 5*). Strikingly, detection of the cross-linked IN complexes with our polyclonal anti-IN an-

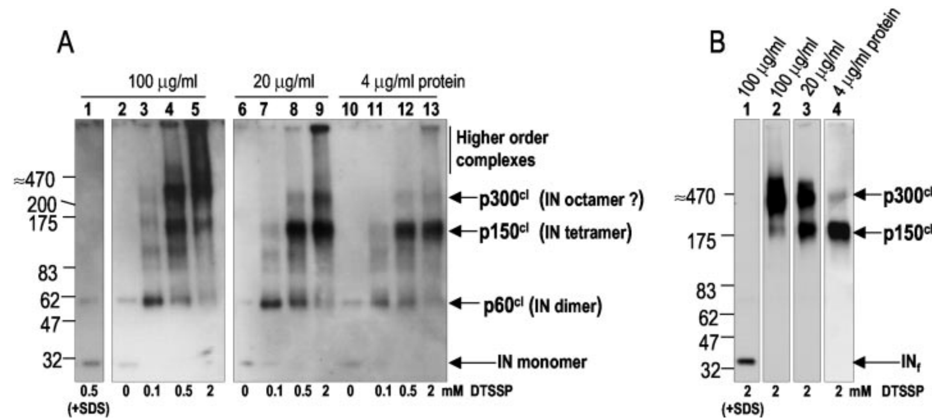

FIG. 2. **Cross-linking of the IN and FLAG-tagged IN complexes with DTSSP.** *A*, the nuclear extract from 293T-IN^{s}ala cells was prepared in 400mCSK buffer, incubated with DTSSP, and separated in a nonreducing 4–12% SDS-PAGE gel. The IN-containing cross-linking adducts were detected by Western blotting with a polyclonal anti-IN antibody. Prior to cross-linking, the extract was adjusted to 100 μg/ml (*lanes 1–5*), 20 μg/ml (*lanes 6–9*), or 4 μg/ml (*lanes 10–13*) of total protein. The concentration of DTSSP was 0.1 mM (*lanes 3*, *7*, and *11*), 0.5 mM (*lanes 1*, *4*, *8*, and *12*), or 2.0 mM (*lanes 5*, *9*, and *13*). No cross-linker was added to the samples in *lanes 2*, *6*, and *10*. The sample in *lane 1* was cross-linked in the presence of 0.2% SDS. p300^{cl}, p150^{cl}, p60^{cl}, the IN monomer bands, and the positions of the molecular weight markers are indicated. The ≈470-kDa mark corresponds to the band of the catalytic subunit of DNA PK (469 kDa), which was detected in a separate lane with a monoclonal anti-DNA PKcs antibody. *B*, cross-linking of the FLAG-tagged IN complexes was done under conditions similar to those in *A*. Prior to cross-linking, the extract was adjusted to 100 μg/ml (*lanes 1* and *2*), 20 μg/ml (*lane 3*), or 4 μg/ml (*lane 4*) of total protein. Only lanes containing samples cross-linked with 2 mM DTSSP are shown; the sample in *lane 1* was cross-linked in the presence of 0.2% SDS. Positions of p300^{cl}, p150^{cl}, and the band of monomeric IN$_{f}$ are indicated.

tibodies was far more sensitive than detection of the non-cross-linked IN. Probably some strong conformational epitopes were better preserved within cross-linked IN during SDS-PAGE and Western blotting. Importantly, no unspecific bands were revealed in nuclear extracts from 293T cells before and after cross-linking with DTSSP, confirming that all of the bands detected in the Western blot correspond to IN-containing complexes (data not shown). The cross-linking of the 293T-IN^{s}ala nuclear extract was clearly dependent on the concentration of both DTSSP and protein. Cross-linking of the diluted nuclear extract (4 μg/ml protein) with 2 mM DTSSP yielded the p150^{cl} band (*lane 13*), whereas in more concentrated extracts, p300^{cl} was the most prominent (*lane 5*). Hence, there exist at least two different IN complexes: a large complex at higher protein concentrations and a smaller complex in the diluted extract. The p150^{cl} product seems to be the result of complete cross-linking because no significant change in cross-linking occurs when the concentration of DTSSP was increased from 0.5 to 2 mM (compare *lanes 12* and *13*) and higher (data not shown). Thus, p150^{cl} probably represents the IN complex present in the diluted nuclear extract. Moreover, this complex is a dissociation product and importantly, a *component* of the larger complex, because it appeared as a partial cross-linking adduct in the reactions with more concentrated protein extracts at 0.5 mM DTSSP, and it decreased at 2 mM DTSSP (compare *lanes 4* and *5*). However, the p300^{cl} band is probably not the result of complete cross-linking of the larger complex, because a strong smear and some less resolved bands are present above p300^{cl} on the Western blot; aggregated material not able to enter 4% polyacrylamide gel is also evident (*lanes 5* and *9*). Some of the high molecular weight adducts in the reactions with 100 μg/ml extracts may result from nonspecific intermolecular cross-linking of proteins. Cross-linking of IN complexes in the nuclear extracts using oxidizing Cu^{2+}-[1,10-phenanthroline]$_3$ complex (Cys-Cys cross-linker) (34) were also suggestive for the presence of a large protein complex that dissociated upon dilution, releasing a molecule with an apparent molecular mass of ~120 kDa after cross-linking (data not shown).

Apparent Stokes Radii of the Two Nuclear Integrase Complexes—To confirm the presence of both IN complexes and deduce their size, we used gel filtration. Nuclear salt extracts from 293T-IN^{s}ala cells were run on a calibrated Superdex 200 column, and the IN elution was followed by immunoblotting the collected fractions (Fig. 3*A*). We observed two distinct elution volumes corresponding to two different IN complexes. Thus, after chromatography of the undiluted extract (600 μg/ml total protein), IN eluted symmetrically with a peak maximum in fractions 8 and 9 corresponding to the elution volume (V_e) of a molecule with a Stokes radius (R_S) of 61 Å (Fig. 3, *A* and *B*). However, IN behaved as a 41 Å molecule, when the sample was diluted 20-fold prior to gel filtration (Fig. 3, *A* and *B*). Assuming that both complexes are globular, their molecular masses can be calculated to be 380 and 115 kDa, respectively (Fig. 3*C*). The smaller dilution-resistant molecule (R_S = 41 Å) most likely corresponds to the p150^{cl} cross-linked complex observed in the previous experiment, whereas partial cross-linking of the 61 Å IN complex probably resulted in p300^{cl}. When the gel filtration fractions containing the 61 Å IN complex were incubated with DTSSP immediately after chromatography, a mixture of p150^{cl} and p300^{cl} products was obtained (data not shown).

Purification and Characterization of FLAG-tagged IN Complexes—To facilitate isolation of native IN complexes from cell extracts, we modified the IN expression construct adding the FLAG epitope tag at the C terminus of IN. The 293T-IN^{s}ala-FLAG cell line, obtained by stable transfection of 293T cells with the tagged expression construct, was very similar to 293T-IN^{s}ala in stability and levels of IN expression (data not shown). FLAG-tagged IN (IN$_f$) localized predominantly in the nuclei in a diffuse pattern and was associated with chromosomes during mitosis (see below), as has been previously reported for non-tagged HIV-1 IN (29). IN$_f$ could be extracted from the nuclei of 293T-IN^{s}alaFLAG cells in the same conditions as for non-tagged IN. The cross-linking pattern of IN$_f$ with DTSSP was very similar to that of nontagged IN (Fig. 2*B*). The two major cross-linking products of IN$_f$ showed slightly slower migration in SDS-PAGE gels than the original p150^{cl} and p300^{cl}, which can be attributed to the negative charge of the FLAG tag and the increased molecular mass of the tagged protein. For convenience, however, we refer to the IN$_f$ cross-linking adducts as p150^{cl} and p300^{cl}. The gel filtration profiles were as observed for the nontagged IN extracted from 293T-IN^{s}ala cells (data not shown).

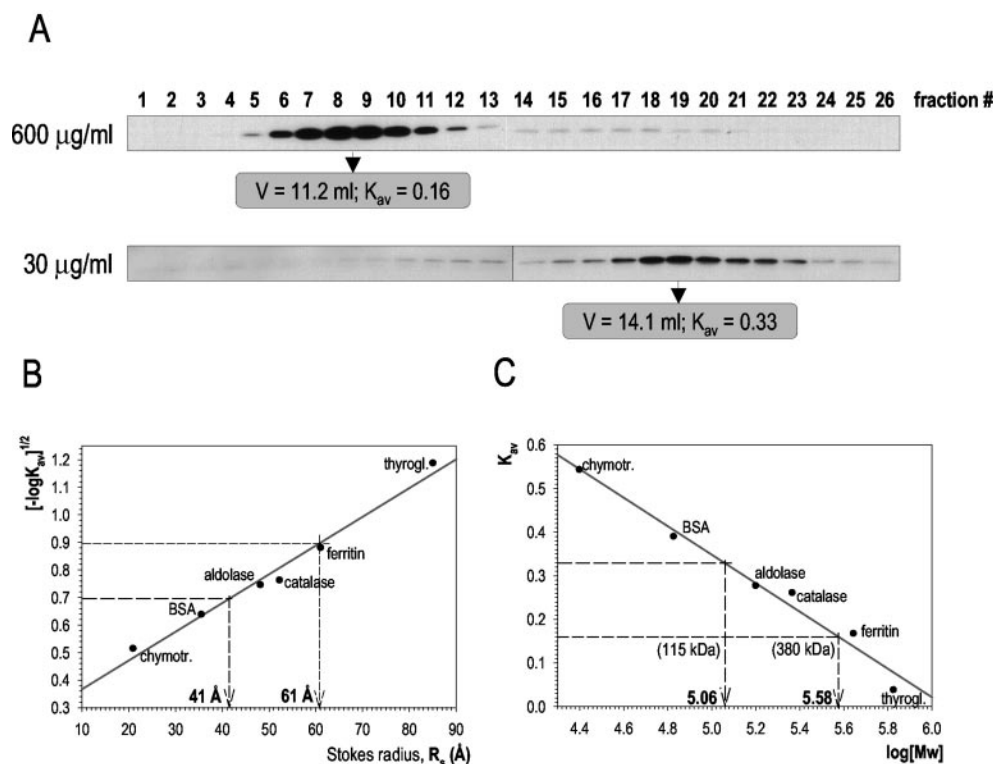

FIG. 3. **Determination of sizes and molecular masses of the IN complexes.** *A*, chromatography of nuclear extracts from 293T-IN^{s}ala cells was carried out on a calibrated Superdex 200 column. Prior to chromatography, the extract was adjusted to 600 or 30 μg/ml of total protein. The collected fractions (*lanes 1–26*) were tested for the presence of IN by Western blotting. The elution volumes (V_e) and the respective partition coefficients (K_{av}) for the observed IN peaks are indicated. *B* and *C*, determination of the Stokes radii and approximate molecular masses of the IN complexes from the experimental K_{av} values. The partition coefficients for the standard proteins were determined in the same conditions (thyroglobulin (*thyrogl.*), K_{av} = 0.039; ferritin, K_{av} = 0.17; catalase, K_{av} = 0.26; aldolase, K_{av} = 0.28; bovine serum albumin (*BSA*), K_{av} = 0.39; chymotrypsinogen A (*chymotr.*), K_{av} = 0.59).

In initial immunoprecipitation experiments, we incubated diluted nuclear extracts from metabolically labeled 293T-IN^{s}alaFLAG cells with the anti-FLAG M2 antibody and protein G-agarose overnight. The protein isolated in this way displayed a single specific band in SDS-PAGE gels migrating at the expected position for the FLAG-tagged IN (33.5 kDa) (Fig. 4*A*). Isoelectrofocusing of immunoprecipitated IN$_f$ in denaturing pH gradients showed a major band close to the predicted pI, which reacted with anti-IN serum in immunoblot (data not shown). When the IN$_f$ immunoprecipitated from a nuclear extract of 293T-IN^{s}alaFLAG cells was eluted from the anti-FLAG M2 antibody with synthetic FLAG peptide and incubated with DTSSP, the p150^{cl} cross-linking product was readily obtained (Fig. 4*B*). When higher IN$_f$ concentrations were used in cross-linking, the immunoreactive reaction products accumulated at the top of the gel, suggesting aggregation of the protein (data not shown). We were not able to find reaction conditions to reproduce the p300^{cl} cross-linking product with IN$_f$ preparations purified this way. Fractionation of purified IN$_f$ on a Superdex column showed a peak with a K_{av} value very close to that of the 41 Å complex (Fig. 4*C*). The presence of the 41 Å complex in the purified IN$_f$ preparation and the apparent molecular mass of 115–150 kDa, based on gel filtration and cross-linking experiments, suggest that the 41 Å molecule is a homotetramer of IN.

Apparently, the native 61 Å IN complex was not stable enough to withstand immunoprecipitation under the original conditions. When we tried shorter incubation times (3–5 h) starting from more concentrated nuclear extracts, the overall yield of IN$_f$ was decreased, but the immunoprecipitated samples were found to contain an additional protein. It had an

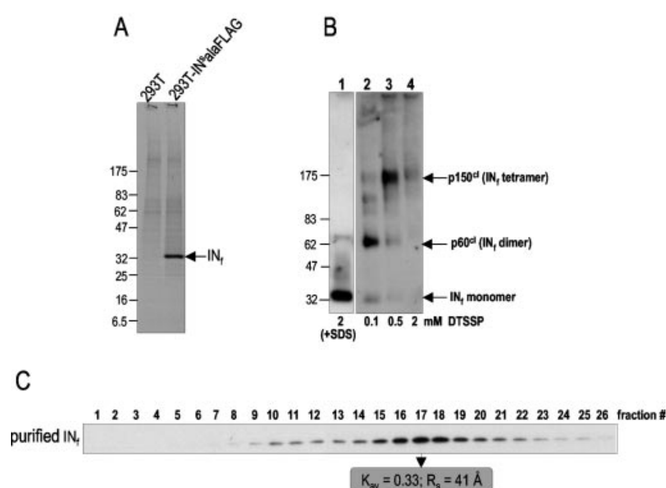

FIG. 4. **Immunoprecipitation of the FLAG-tagged IN from diluted nuclear extracts.** *A*, nuclear extracts prepared from metabolically labeled 293T-IN^{s}ala and 293T-IN^{s}alaFLAG cells were diluted to 200 μg/ml of total protein and immunoprecipitated with the anti-FLAG M2 antibody and protein G-agarose beads for 16 h at 4 °C. The protein was eluted by boiling in SDS-PAGE sample buffer and separated in a 4–20% SDS-PAGE gel. An autoradiograph of the gel is shown. *B*, FLAG-tagged IN was immunoprecipitated from a diluted nuclear extract of nonlabeled 293T-IN^{s}alaFLAG cells overnight. The protein was eluted with FLAG peptide in 400mCSK buffer and cross-linked with DTSSP. The reaction conditions are similar to those in Fig. 2. *C*, the protein immunoprecipitated and eluted as in *B* was subjected to gel filtration on a Superdex 200 column. The fractions collected were analyzed by Western blotting with anti-IN antibodies. The R_S value corresponding to the observed peak was determined as for Fig. 3*B*.

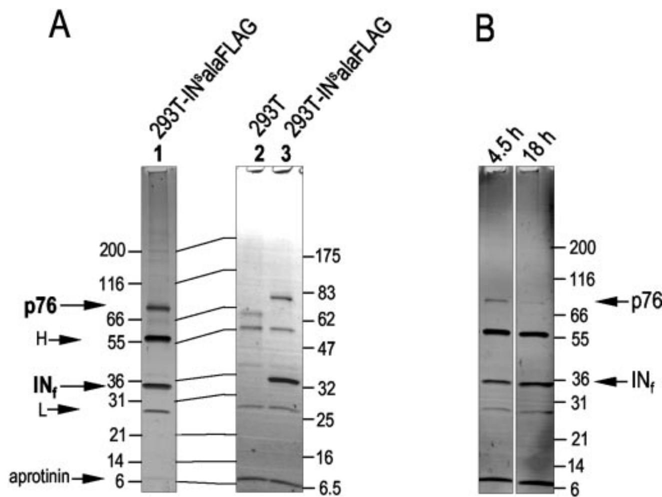

FIG. 5. **Co-immunoprecipitation of FLAG-tagged IN with p76 from nondiluted nuclear extracts.** *A*, nuclear extracts from 293T-IN^{s}alaFLAG and 293T cells (700 μg/ml of total protein) were incubated with anti-FLAG M2 antibody and protein G-agarose beads for 4 h. The beads were washed as described under "Experimental Procedures," and the bound proteins were eluted with FLAG peptide in 400mCSK buffer. The eluted proteins were concentrated by precipitation with trichloroacetic acid, redissolved in SDS sample buffer, separated in 4–20% denaturing PAGE gels, and visualized by silver staining. *Lane 1*, immunoprecipitate of a nuclear extract from 60×10^6 293T-IN^{s}alaFLAG cells. *Lanes 2* and *3*, immunoprecipitation was done in parallel with nuclear extracts from 293T and 293T-IN^{s}alaFLAG cells. The bands of IN$_f$, p76, the heavy and the light chains of the anti-FLAG M2 IgG1 antibody, and aprotinin (protease inhibitor present in 400mCSK buffer) are indicated. Two sets of molecular mass markers were used in both gels to determine the apparent molecular mass of p76. The positions of the molecular mass markers are shown. *B*, nuclear extract from 20×10^6 293T-IN^{s}alaFLAG cells ($\approx$700 μg/ml total protein) was incubated with anti-FLAG antibody and protein G-agarose beads for either 4.5 h (*left lane*) or 18 h (*right lane*). The immunoprecipitated protein was then eluted and analyzed as in *A*.

apparent molecular mass of ~76 kDa, as determined by SDS-PAGE (Fig. 5*A*) and was present at variable ratios to IN$_f$ in different preps. This protein, here referred to as p76, was specifically associated with IN$_f$, because it could not be immunoprecipitated from the parental 293T cells with the anti-FLAG antibody (Fig. 5*A*, compare *lanes 2* and *3*). When undiluted nuclear extract from 293T-IN^{s}alaFLAG cells was immunoprecipitated with anti-FLAG antibody for 4.5 h, both IN$_f$ and p76 bands were readily detected (Fig. 5*B*). Although extending immunoprecipitation to 18 h improved IN$_f$ recovery, the yields of p76 were greatly reduced (Fig. 5*B*). Intriguingly, the p300^{cl} band, detected after DTSSP cross-linking of the nuclear salt extracts, was also observed when the p76-containing IN$_f$ preparations were cross-linked with DTSSP (see Fig. 7), suggesting that p76 is part of the large IN complex present in the nuclear extracts.

Identification of the p76 Protein as LEDGF/DFS70/p75—By upscaling immunoprecipitation, we were able to isolate sufficient amounts of p76 for characterization by Edman degradation and mass spectrometry (Fig. 6*A*). The N-terminal sequence obtained from p76 was *XX*RDFKPGD (the first two residues were not resolved because of background noise). Searching the TrEMBL protein data base for human proteins carrying this sequence tag using TagIdent (us.expasy.org/tools/tagident.html) (35) resulted in four hits with accession numbers O95368, Q9UER6, Q9NZI3, and O75475, all corresponding to the two alternative products of one gene: LEDGF/DFS70/p75 (referred to as LEDGF) and the p52 protein (36, 37). Although the actual molecular mass of LEDGF is ~60 kDa, it is known to migrate as a 75-kDa band in SDS-PAGE gels (36). On-line LC/MS/MS analysis of tryptic peptides obtained by in-gel digestion of p76 provided further evidence that p76 is indeed identical to LEDGF. Half of the predicted LEDGF tryptic peptides within the mass range of 1000–2500 Da could be identified in the sample, and their MS/MS spectra readily matched LEDGF covering ~18% of its sequence (Fig. 6*B* and Table I). Moreover, p76 strongly reacted with a commercially available monoclonal anti-LEDGF antibody (data not shown). Most of IN$_f$ present in nuclear extracts could be immunoprecipitated with the anti-LEDGF antibody, whereas only about 10% of LEDGF could be recovered with the anti-FLAG antibody (Fig. 6*C*), suggesting that LEDGF is present in an excess over IN$_f$ in the extract. Nontagged IN could also be efficiently precipitated with the anti-LEDGF antibody, and a fraction of the LEDGF could be precipitated with polyclonal anti-IN antibody in similar conditions from nuclear salt extracts of 293T-IN^{s}ala cells (data not shown).

LEDGF Is Part of the 61 Å HIV IN Complex—To determine whether the 61 Å complex contains LEDGF, we preincubated the nuclear salt extract from 293T-IN^{s}alaFLAG cells with a monoclonal anti-LEDGF antibody prior to chromatography on a Superdex column. The IN$_f$ elution profile changed dramatically; the peak eluted now near the void volume of the column (Fig. 6*D*). Elution of the 61 Å complex was not altered by preincubation of the extract with an unrelated mouse IgG1 (Fig. 6*D*). Predictably, elution of the 41 Å IN complex (the presumed IN tetramer) did not change after preincubation of the diluted nuclear extracts with the anti-LEDGF antibody (data not shown).

When the IN$_f$-LEDGF complex was purified by immunoprecipitation and cross-linked with DTSSP, the p300^{cl} band could be readily detected in an immunoblot with anti-IN antibody (Fig. 7). However, p300^{cl} did not react with a monoclonal anti-LEDGF antibody; instead, a Western blot with the anti-LEDGF antibody revealed two bands migrating at higher positions in the gel (pHMW$_1^{cl}$ and pHMW$_2^{cl}$, *lane 2′* in Fig. 7) (the molecular masses of these molecules are too high to be determined with SDS-PAGE). In addition, both pHMW$_1^{cl}$ and pHMW$_2^{cl}$ products were detected with the anti-LEDGF antibody in the cross-linked nuclear extracts of 293T-IN^{s}alaFLAG cells but not of parental 293T cells (data not shown). These results suggest that p300^{cl} is a product of incomplete cross-linking of the 61 Å IN-LEDGF complex. We speculate that p300^{cl} probably represents an octamer of IN (*i.e.* dimer of tetramers). Contacts between IN and LEDGF within the 61 Å complex may be less prone to cross-linking with DTSSP than those between IN protomers. We cannot exclude the possibility, however, that the target epitope for the monoclonal anti-LEDGF antibody used is masked or destroyed within p300^{cl}. In addition to the major p300^{cl} product, a band at a position close to pHMW$_1^{cl}$ is present on the anti-IN immunoblot of the purified and cross-linked IN$_f$-LEDGF complex (Fig. 7, *lane 2*). Thus, the pHMW$_1^{cl}$ adduct is probably the smallest cross-linked IN complex containing LEDGF.

LEDGF Co-localizes with IN within Nuclei of 293T-IN^{s}alaFLAG Cells—Immunofluorescent detection of both IN$_f$ and LEDGF in fixed 293T-IN^{s}ala cells revealed strikingly similar intranuclear distribution patterns for both proteins (Fig. 8*A*). In accordance with previous reports, both proteins were bound to condensed chromosomes in mitotic cells (Fig. 8*B*) (29, 38). The distribution of another nuclear protein, the catalytic subunit of DNA-dependent protein kinase (DNA PKcs) clearly differed from that of IN$_f$ (Fig. 8*C*). In addition, DNA PKcs was excluded from condensed chromosomes in mitotic cells (data not shown). Intriguingly, the nuclear localization of IN$_f$ and LEDGF did not precisely correspond to the overall DNA staining pattern, arguing against the possibility that the apparent

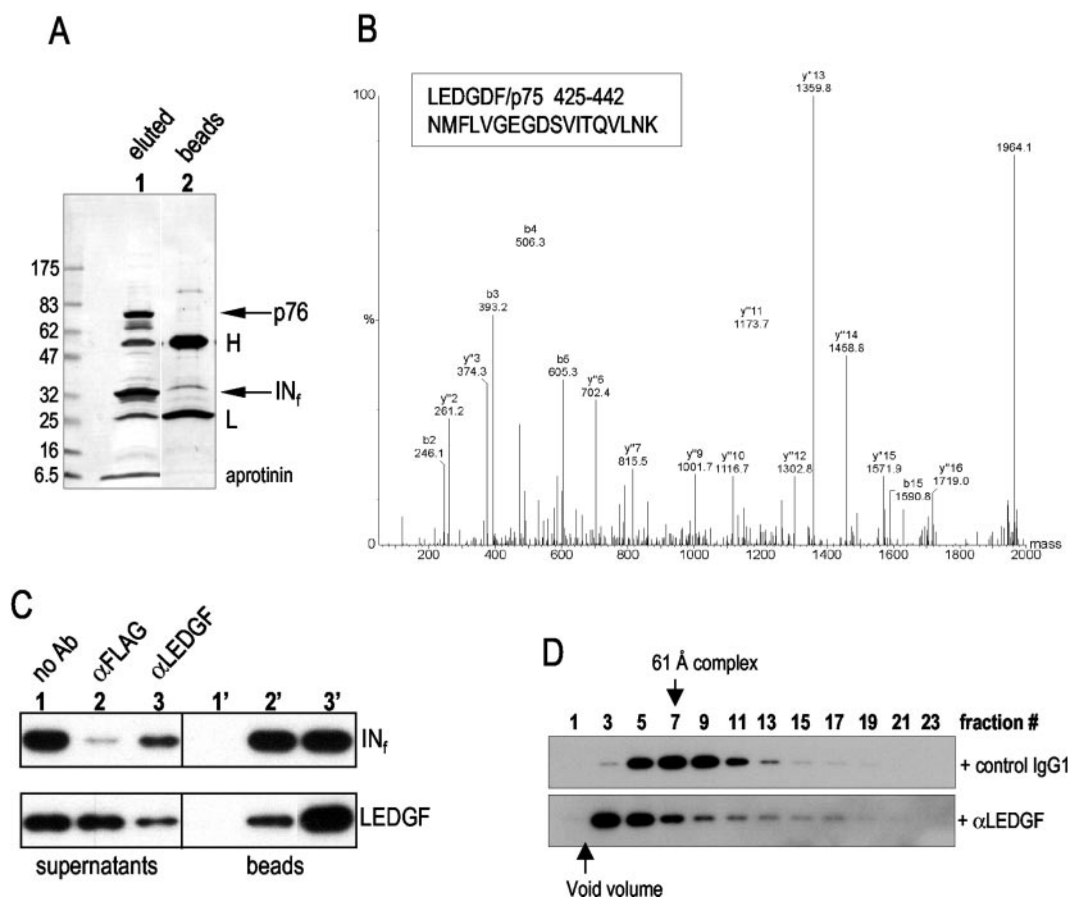

FIG. 6. **P76 is identical to LEDGF/DFS70/p75 and is part of the 61 Å complex.** *A*, the Coomassie Blue-stained polyvinylidene difluoride membrane used for N-terminal microsequencing of p76. The IN_f-p76 complex was eluted from protein G-agarose-immobilized anti-FLAG M2 antibody with FLAG peptide, separated in reducing 4–20% SDS-PAGE gel, and transferred onto the polyvinylidene difluoride membrane (*lane 1*). The proteins left on the beads after incubation with FLAG peptide were eluted with SDS sample buffer (*lane 2*). The bands corresponding to p76, IN_f, the heavy (*H*) and light (*L*) chains of the anti-FLAG M2 antibody, aprotinin, and the molecular weight markers are indicated. *B*, the MS/MS spectrum of a doubly charged peptide ion (*m/z* 982.57) obtained from the in-gel tryptic digest of p76 corresponding to the LEDGF peptide N425-K442. The observed b- and y″-dominant fragment ions are indicated (for nomenclature see Ref. 56). *C*, co-immunoprecipitation of LEDGF and FLAG-tagged IN from a nuclear extract of 293T-IN^salaFLAG cells. Immunoprecipitation was carried with anti-FLAG (*lanes 2* and *2′*), anti-LEDGF (*lanes 3* and *3′*), or no antibody (*lanes 1* and *1′*). After 4 h of incubation, the protein G-agarose beads containing the precipitated protein complexes were washed three times with 400mCSK buffer and were resuspended in reducing SDS-PAGE sample buffer. Western blotting was done to detect IN_f and LEDGF (*lanes 1′–3′*). *Lanes 1–3* contained the immunoprecipitation supernatants. *D*, elution of the 61 Å IN complex from a gel filtration column is shifted after preincubation with anti-LEDGF antibody. A nuclear extract of 293T-IN^salaFLAG cells was preincubated with 3 μg/ml anti-HA (control mouse IgG1) or anti-LEDGF antibody and separated by chromatography on a Superdex 200 column. IN_f was detected in the fractions by Western blotting. Only odd numbered fractions are shown. The void volume of the column was 8.3 ml, approximately corresponding to fraction 2.

TABLE I
LEDGF/p75 tryptic peptides identified from the in-gel digest of p76 and confirmed by MS/MS sequencing

Residues[a]	Sequence	Mass[b]	Error[c]
			Da
4–14	DFKPGDLIFAK	1249.67	+0.05
40–56	LPIFFFGTHETAFLGPK	1921.01	+0.19
57–67	DIFPYSENKEK	1368.66	+0.14
76–89	GFNEGLWEIDNNPK	1631.76	−0.10
136–143	AVDITTPK	843.47	+0.02
156–175	QVETEEAGVVTTATASVLNK	2046.05	+0.15
425–442	NMFLVGEGDSVITQVLNK	1963.00	−0.16

[a] Tryptic digestion fragments of LEDGF/p75.
[b] Mass values are the calculated monoisotopic masses of the sequences in the table.
[c] Errors are the differences between the calculated monoisotopic masses and those observed in the sample.

co-localization of the two proteins might merely reflect their independent association with chromosomal DNA. In a control experiment, we visualized IN_f using a mixture of polyclonal and monoclonal anti-FLAG antibodies; the obtained two-color IN_f staining was similar to that of IN_f and LEDGF (data not shown).

LEDGF/p75 Is an Activator of HIV-1 IN in Vitro—We have previously described the activities of recombinant HIV-1 IN on the mini-HIV substrate, a linear 4.7-kb double-stranded DNA molecule, carrying the U3 and U5 terminal fragments of the viral LTR sequences (30). Recombinant HIV-1 IN on itself was proficient in carrying-out 3′-end processing and strand transfer using this long DNA substrate. Mini-HIV served as both donor and target DNA in this assay. Presence of 5–12% polyethylene glycol (PEG) in the reaction was required for the enzymatic activity. To ascribe a possible function to the observed IN-LEDGF interaction, we examined whether LEDGF could modulate enzymatic activity of HIV-1 IN *in vitro*. The mini-HIV DNA substrate was incubated with recombinant IN and His_6-tagged LEDGF, and the reaction products were analyzed by native agarose gel electrophoresis (Fig. 9). Although in the absence of PEG and LEDGF, strand transfer products were almost undetectable (Fig. 9*A*, *lane 3*), the addition of LEDGF alone resulted in a robust stimulation of the reaction (*lanes 4–8*). In some conditions, approximately half of the substrate

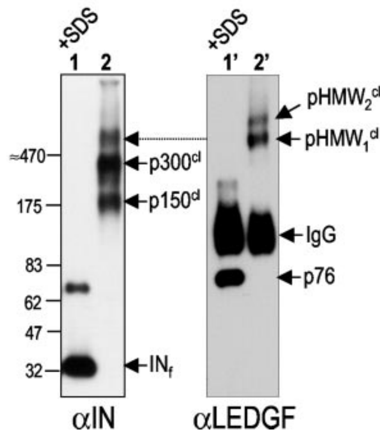

FIG. 7. **Cross-linking of the IN_f-LEDGF complex with DTSSP.** The IN_f-LEDGF complex was immunoprecipitated from a nuclear extract prepared from 293T-IN^{s}alaFLAG cells with anti-FLAG M2 antibody and protein G-agarose for 4.5 h. The protein was eluted with FLAG peptide and incubated with 2 mM DTSSP in the presence (*lanes 1* and *1′*) or absence (*lanes 2* and *2′*) of 0.2% SDS. The cross-linked samples were then separated in a nonreducing 4–12% SDS-PAGE gel and immunoblotted with polyclonal anti-IN (*left blot*) or monoclonal anti-LEDGF (*right blot*) antibodies. The positions of IN_f and LEDGF as well as of the cross-linking adducts $p150^{cl}$, $p300^{cl}$, $pHMW_1^{cl}$ and $pHMW_2^{cl}$ are indicated. Anti-FLAG M2 IgG1 present in the sample is detected on the anti-LEDGF Western blot.

DNA was converted into various strand transfer products, including those that were too large to enter the gel (*lanes 7* and *8*). Remarkably, both the overall efficiency of the reaction and the range of the strand transfer products depended on the concentration of LEDGF. No significant variation in the yield of the strand transfer products was detected when the order of addition of LEDGF and IN to the mini-HIV reaction was reversed (data not shown). In agreement with previous results, the addition of PEG stimulated IN activity. In the presence of 10% PEG, various strand transfer products could be detected, including the major 9.4-kb product, which results from end-to-end integration of mini-HIV molecules (Fig. 9*B*, *lane 4*) (30). However, at least in the conditions tested, PEG did not have a significant effect on the LEDGF-dependent reaction (Fig. 9*B*, *lanes 5* and *6*).

DISCUSSION

A serious obstacle in working with recombinant retroviral integrases is their poor solubility and propensity for aggregation. All of the crystal structure and some of the *in vitro* multimerization studies have been carried out with the soluble mutants. Furthermore, recent reports raised concern that the stoichiometry and enzymatic activities of recombinant IN can be affected by the enzyme preparation (16, 39). Our goal was to study the protein complexes that HIV-1 IN forms within the nuclei of human cells. The bulk of HIV-1 IN present in 293T cells, which stably produce this viral protein, is associated with the insoluble nuclear fraction. Although IN seems to be directly or indirectly bound to chromosomal DNA, this may not be the only factor in nuclear retention of IN, because digestion of the detergent-permeabilized nuclei with nucleases was not sufficient to elute IN. In this work, we concentrated on the study of IN complexes extracted from the detergent-permeabilized nuclei in hypertonic conditions. We found that salt-eluted IN exists as part of a distinct 61 Å complex, which is not stable in diluted nuclear extracts and dissociates, releasing a 41 Å core molecule. Our cross-linking and gel filtration data suggest that the latter molecule is a homotetramer of IN. We estimated that the concentrations of IN and IN_f in the nuclear extracts did not exceed 10 nM (≤0.3 μg/ml IN at 100 μg/ml total protein). Therefore, the IN tetramer was stable even at subnanomolar concentrations (*i.e.* in the extracts diluted to 4 μg/ml of total protein), implying that the minimal nuclear IN complex is a homotetramer.

All of the HIV-1 IN present in nuclear extracts appears to be in complex with LEDGF. However, at its concentration in nuclear extracts, the 61 Å complex was not stable enough to allow measurement of its sedimentation coefficient, which is required to determine its precise molecular mass (32). Assuming that the 61 Å complex is globular, we estimated its molecular mass to be around 400 kDa (Fig. 3*C*). The simplest model compatible with this molecular mass suggests a symmetrical complex containing a pair of IN tetramers and two subunits of LEDGF, corresponding to a macromolecule of 370 kDa. Reconstitution of the IN-LEDGF complex from the recombinant proteins will help to confirm the proposed stoichiometry. At this time, we cannot rule out the possibility that the native 61 Å complex contains another cellular protein lost during immunoprecipitation. Purified LEDGF-containing IN samples displayed complex gel filtration profiles, probably because of partial dissociation of the native complex (data not shown).

Retroviral IN within PIC: a Dimer of Dimers or a Dimer of Tetramers?—During reverse transcription, the two retroviral cDNA termini are not completed simultaneously, and both seem to be substrates for the 3′-end processing activity of IN as soon as they appear (22). Moreover, at least in the case of HIV, 3′-processing of one LTR end was observed in conditions where the second end was nonfunctional and not supportive of normal intasome assembly (40). On the other hand, two functional LTRs were found to be required for strand transfer activity of isolated HIV PICs. Therefore, although LTRs can be processed asymmetrically, a synaptic complex involving both LTRs must be formed to allow strand transfer, ensuring that only legitimate integration of both retroviral cDNA ends occurs. Based on a comparison with the Mu phage transposase and available crystal structure data, it has been suggested that the active form of retroviral IN is a tetramer (a dimer of dimers) (9, 10). Our results suggest that HIV IN expressed in human cells is indeed present as a stable tetramer. Intriguingly, both Mu phage and Tn5 transposases form functional multimers (tetramers and dimers, respectively) only within their synaptic complexes (41–43). Thus, independent transposase protomers must first bind to the ends of the transposon genome, before being brought together to form the synaptic complex. Extrapolating this scheme to retroviral PIC assembly, the stable IN tetramers can be looked at as such independent protomers, which first have to bind to one LTR end each before interacting with each other. Accordingly, each individual tetramer would be capable of carrying out 3′-end processing, whereas a *dimer of tetramers* would be necessary to accomplish the strand transfer. Our model is thus in agreement with Heuer and Brown (8), who have argued that an octamer of IN is minimally required to mediate concerted integration. It has been postulated to be a universal feature shared by transposases that only one pair of active sites within the functional multimers are involved in catalysis (43). In the case of Mu transposase, which forms a tetramer within the synaptic complex, only two subunits are catalytically active, performing two reactions each, whereas the remaining two subunits are thought to play a structural role (44). In this respect, a model involving eight IN subunits in the synaptic complex is plausible. Of interest, an octamer of retroviral IN (250 kDa) would roughly match in size the functional multimers of Tn5 and Mu transposases (a dimer of 220 kDa and a tetramer of 300 kDa, respectively). This model, however, remains speculative and needs further experimental verification.

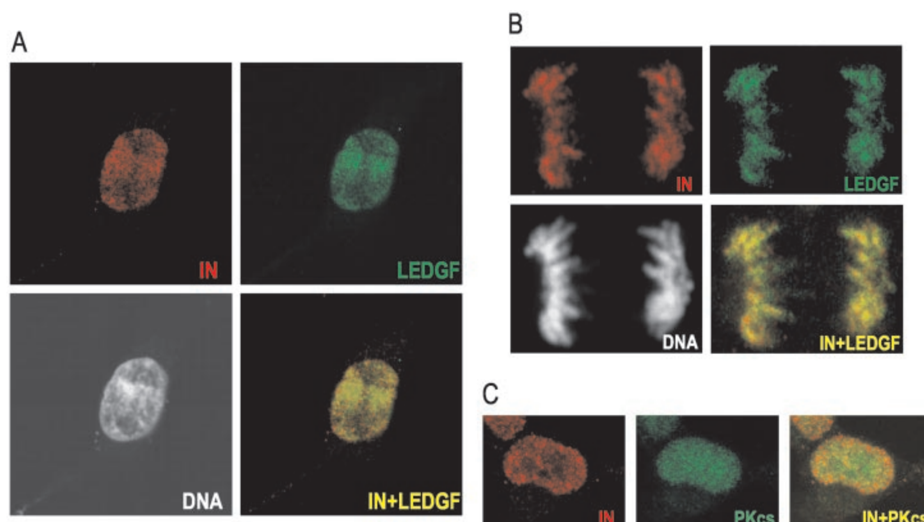

FIG. 8. **Co-localization of FLAG-tagged IN and LEDGF in 293T-IN^{s}alaFLAG cells.** ***A,*** Confocal laser scanning micrographs of a fixed and permeabilized cell fluorescently stained with a combination of monoclonal anti-LEDGF plus Alexa-488 conjugated anti-mouse antibodies to detect LEDGF (*green*, *LEDGF*) and rabbit polyclonal anti-FLAG plus Alexa-555 conjugated anti-rabbit antibodies (*red*, *IN*) to localize FLAG-tagged IN. DNA was stained with To-Pro3 iodide (shown as *white*). The two-color merged image (*IN+LEDGF*) was produced by overlaying the IN and LEDGF images. *B*, both IN_f and LEDGF are associated with condensed chromosomes during mitosis. Immunofluorescent staining was performed as described for *A*. *C*, DNA PKcs and IN_f display no significant co-localization. IN_f (*red*, *IN*) was detected as in *A*; DNA PKcs (*red*, *PKcs*) was localized with monoclonal anti-DNA PKcs antibody plus Alexa-555 conjugated anti-mouse antibody. The two color *IN+PKcs* image is an overlay of the IN and PKcs images.

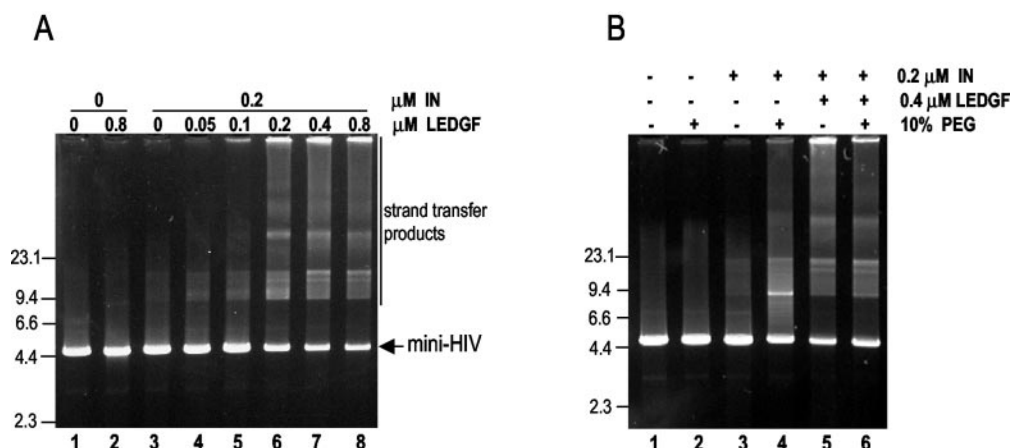

FIG. 9. **Recombinant LEDGF enhances HIV-1 IN strand transfer activity *in vitro*.** *A*, mini-HIV DNA was preincubated with HIV-1 IN for 7 min at room temperature. Next, 0–0.8 μM His_6-tagged LEDGF was added to the reactions that were further incubated at 37 °C for 90 min. The concentrations of IN and LEDGF used in the reactions are indicated. The reactions contained 150 ng of mini-HIV DNA, 110 mM NaCl, 20 mM Hepes, pH 7.5, 5 mM dithiothreitol, and 5 μM $ZnCl_2$ in a final volume of 20 μl. The reactions were stopped by addition of 0.5% SDS and 25 mM EDTA, and the samples were digested with 0.25 mg/ml proteinase K at 37 °C for 30 min to completely disrupt protein-DNA complexes. DNA was then precipitated with ethanol, redissolved in Tris-EDTA, and analyzed by electrophoresis in an 0.8% agarose gel. *B*, mini-HIV DNA was incubated with 0.2 μM HIV-1 IN (*lanes 3–6*) in the presence of 10% PEG-8000 (*lanes 2*, *4*, and *6*) and/or 0.4 μM His_6-tagged LEDGF (*lanes 5* and *6*). The other reaction conditions and the sample preparation were as in *A*. The positions of the DNA molecular mass markers (23.1, 9.4, 6.6, 4.4, and 2.3 kb) are indicated. The gels were stained with SybrGold (Molecular Probes).

What Is the Role of LEDGF in Retroviral Replication?— Based on sequence similarity, LEDGF/p75 is a member of the hepatoma-derived growth factor (HDGF) family that includes HDGF and several other HDGF-related proteins (reviewed in Ref. 45). A high degree of homology exists between the N-terminal regions of these proteins. The PWWP motif (70 residues containing the Pro-Trp-Trp-Pro core sequence) is located within the N-terminal homology region of HDGF-related proteins and relates them to a larger and functionally diverse nuclear protein family that includes DNA-binding transcription factors and enzymes involved in DNA repair and DNA methylation. PWWP domains are thought to be implicated in protein-protein interactions (46).

The p75 protein has first been described as the positive transcription co-factor PC4-interacting protein (36). It has also been shown to interact with components of the general transcription machinery and with the transcription activation domain of VP16. Independently, a cDNA clone coding for a protein identical to p75 has been isolated from a lens epithelium cell library (47). Overexpression of the protein-stimulated survival of diverse primary cells and cell lines and enhanced their resistance to oxidative and hyperthermic stress (hence, lens epithelium-derived growth factor). The same protein has also been identified as the DFS70 autoantigen, antibodies to which were found in some cases of atopic dermatitis, asthma, and interstitial cystitis (48). LEDGF has been shown to be a DNA-binding protein with affinity for heat shock and stress-related DNA elements (49). Searching its sequence for known protein motifs found in the Blocks+ data base (www.blocks.fhcrc.org/) (50) revealed fragments with similarity to the HMG-I(Y) DNA

AT hook sequence (data not shown). However, it remains to be determined whether these sequence elements are involved in DNA binding. Recent reports suggested that LEDGF plays an important role in regulating expression of the stress response genes (51, 52).

Alternative splicing of LEDGF pre-mRNA allows expression of the second protein, p52, from the same gene (36, 37). The transcripts coding for p75 and p52 were detected in different cell types and tissues, with p52 being most abundant in testis and p75 being most abundant in thymus. A growing body of evidence suggests that p75 and p52 may have different functions. Although they both can interact with PC4, VP16, and general transcription factors, at least *in vitro*, p52 displays higher transcription activation activity (36). In addition, p52 and not p75 has been shown to functionally interact with the ASF/SF2 splicing factor *in vitro* (53). The proteins also differ in their nuclear distribution patterns (38). Intracellular levels of p52 appear to be much lower than those of p75, at least in the cell lines we evaluated (HEK-293, 293T, HeLa, and CEM) (data not shown). We have not detected co-immunoprecipitation of p52 with IN from nuclear extracts of IN-expressing 293T cells. However, it remains to be determined whether p52 is able to interact with HIV IN.

So far, LEDGF/p75 has not been linked to retroviral replication. The fact that the recombinant protein was able to dramatically stimulate HIV-1 IN activity *in vitro* suggests a direct involvement of LEDGF in the integration process. We are currently investigating whether LEDGF can specifically promote coupling and concerted integration of both mini-HIV DNA termini. During viral infection, LEDGF, being a chromosome-associated IN-binding protein, may play the role of a docking factor or a receptor for PICs. LEDGF might thus be functionally similar to the Mu phage transposition co-factor MuB, which, by associating with the acceptor DNA, makes it a preferred target for transposition (7). The specific association of LEDGF with the PC4 transcription co-activator and the general transcription machinery could explain the recent data that HIV favors transcription units for integration (54). LEDGF was also one of the genes up-regulated in SupT1 cells following HIV infection (54). In addition, as a proposed stress response-related transcription factor, LEDGF may be an important element in the mechanism of HIV activation by stress stimuli (55). Preferential HIV integration into the LEDGF-associated chromosomal loci, for example, could contribute to more efficient expression of the provirus under chronic oxidative stress conditions observed in AIDS.

Experiments are currently underway to establish the exact role of LEDGF in HIV replication. If proved essential for HIV DNA integration, LEDGF may constitute a novel target for anti-retroviral therapy. Alternatively, a therapeutic strategy based on a modified LEDGF protein, designed to capture the viral IN in a catalytically quiescent complex, may also be pursued.

Acknowledgments—We thank Johan Van Griensven and the other members of the Molecular Virology and Gene Therapy Laboratory of the Rega Institute for helpful discussions. We are grateful to Dr. M. Fujita (Aichi Cancer Center, Nagoya, Japan) for advice on DNase I elution of hMCM3 and to Dr. R. Knippers (Universty of Konstanz, Konstanz, Germany) for providing us with anti-hMCM3 antibodies. We also thank Frank Vanrobaeys (Ghent University, Ghent, Belgium) for help with MS analysis of p76/LEDGF.

REFERENCES

1. Debyser, Z., Cherepanov, P., Van Maele, B., De Clercq, E., and Witvrouw, M. (2002) *Antivir. Chem. Chemother.* **13,** 1–15
2. Miller, M. D., and Hazuda, D. J. (2001) *Curr. Opin. Microbiol.* **4,** 535–539
3. Pommier, Y., Marchand, C., and Neamati, N. (2000) *Antiviral Res.* **47,** 139–148
4. Asante-Appiah, E., and Skalka, A. M. (1997) *Antiviral Res.* **36,** 139–156
5. Brown, P. O. (1997) in *Retroviruses* (Coffin, J. M., Huges, S. H., and Varmus, H. E., eds) pp. 161–203, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY
6. Craigie, R. (2001) *J. Biol. Chem.* **276,** 23213–23216
7. Mizuuchi, K. (1992) *Annu. Rev. Biochem.* **61,** 1011–1051
8. Heuer, T. S., and Brown, P. O. (1998) *Biochemistry* **37,** 6667–6678
9. Wang, J. Y., Ling, H., Yang, W., and Craigie, R. (2001) *EMBO J.* **20,** 7333–7343
10. Yang, Z. N., Mueser, T. C., Bushman, F. D., and Hyde, C. C. (2000) *J. Mol. Biol.* **296,** 535–548
11. Coleman, J., Eaton, S., Merkel, G., Skalka, A. M., and Laue, T. (1999) *J. Biol. Chem.* **274,** 32842–32846
12. Deprez, E., Tauc, P., Leh, H., Mouscadet, J. F., Auclair, C., and Brochon, J. C. (2000) *Biochemistry* **39,** 9275–9284
13. Jenkins, T. M., Engelman, A., Ghirlando, R., and Craigie, R. (1996) *J. Biol. Chem.* **271,** 7712–7718
14. Jones, K. S., Coleman, J., Merkel, G. W., Laue, T. M., and Skalka, A. M. (1992) *J. Biol. Chem.* **267,** 16037–16040
15. Lee, S. P., Xiao, J., Knutson, J. R., Lewis, M. S., and Han, M. K. (1997) *Biochemistry* **36,** 173–180
16. Leh, H., Brodin, P., Bischerour, J., Deprez, E., Tauc, P., Brochon, J. C., LeCam, E., Coulaud, D., Auclair, C., and Mouscadet, J. F. (2000) *Biochemistry* **39,** 9285–9294
17. Petit, C., Schwartz, O., and Mammano, F. (1999) *J. Virol.* **73,** 5079–5088
18. Ellison, V., Gerton, J., Vincent, K. A., and Brown, P. O. (1995) *J. Biol. Chem.* **270,** 3320–3326
19. Engelman, A., Bushman, F. D., and Craigie, R. (1993) *EMBO J.* **12,** 3269–3275
20. Lee, M. S., and Craigie, R. (1998) *Proc. Natl. Acad. Sci. U. S. A.* **95,** 1528–1533
21. Farnet, C. M., and Bushman, F. D. (1997) *Cell* **88,** 483–492
22. Miller, M. D., Farnet, C. M., and Bushman, F. D. (1997) *J. Virol.* **71,** 5382–5390
23. Chen, H., and Engelman, A. (1998) *Proc. Natl. Acad. Sci. U. S. A.* **95,** 15270–15274
24. Friedman, D. I. (1992) *Curr. Opin. Genet. Dev.* **2,** 727–738
25. Kalpana, G. V., Marmon, S., Wang, W., Crabtree, G. R., and Goff, S. P. (1994) *Science* **266,** 2002–2006
26. Wang, W., Cote, J., Xue, Y., Zhou, S., Khavari, P. A., Biggar, S. R., Muchardt, C., Kalpana, G. V., Goff, S. P., Yaniv, M., Workman, J. L., and Crabtree, G. R. (1996) *EMBO J.* **15,** 5370–5382
27. Turelli, P., Doucas, V., Craig, E., Mangeat, B., Klages, N., Evans, R., Kalpana, G., and Trono, D. (2001) *Mol. Cell* **7,** 1245–1254
28. Yung, E., Sorin, M., Pal, A., Craig, E., Morozov, A., Delattre, O., Kappes, J., Ott, D., and Kalpana, G. V. (2001) *Nat. Med.* **7,** 920–926
29. Cherepanov, P., Pluymers, W., Claeys, A., Proost, P., De Clercq, E., and Debyser, Z. (2000) *FASEB J.* **14,** 1389–1399
30. Cherepanov, P., Surratt, D., Toelen, J., Pluymers, W., Griffith, J., De Clercq, E., and Debyser, Z. (1999) *Nucleic Acids Res.* **27,** 2202–2210
31. Fujita, M., Kiyono, T., Hayashi, Y., and Ishibashi, M. (1997) *J. Biol. Chem.* **272,** 10928–10935
32. Siegel, L. M., and Monty, K. J. (1965) *Biochim. Biophys. Acta* **112,** 346–362
33. Schulte, D., Burkhart, R., Musahl, C., Hu, B., Schlatterer, C., Hameister, H., and Knippers, R. (1995) *J. Cell Sci.* **108,** 1381–1389
34. Ji, T. H. (1983) *Methods Enzymol.* **91,** 580–609
35. Wilkins, M. R., Gasteiger, E., Sanchez, J. C., Appel, R. D., and Hochstrasser, D. F. (1996) *Curr. Biol.* **6,** 1543–1544
36. Ge, H., Si, Y., and Roeder, R. G. (1998) *EMBO J.* **17,** 6723–6729
37. Singh, D. P., Kimura, A., Chylack, L. T., Jr., and Shinohara, T. (2000) *Gene (Amst.)* **242,** 265–273
38. Nishizawa, Y., Usukura, J., Singh, D. P., Chylack, L. T., Jr., and Shinohara, T. (2001) *Cell Tissue Res.* **305,** 107–114
39. Sinha, S., Pursley, M. H., and Grandgenett, D. P. (2002) *J. Virol.* **76,** 3105–3113
40. Chen, H., and Engelman, A. (2001) *Mol. Cell. Biol.* **21,** 6758–6767
41. Baker, T. A., and Mizuuchi, K. (1992) *Genes Dev.* **6,** 2221–2232
42. Bhasin, A., Goryshin, I. Y., Steiniger-White, M., York, D., and Reznikoff, W. S. (2000) *J. Mol. Biol.* **302,** 49–63
43. Rice, P. A., and Baker, T. A. (2001) *Nat. Struct. Biol.* **8,** 302–307
44. Williams, T. L., Jackson, E. L., Carritte, A., and Baker, T. A. (1999) *Genes Dev.* **13,** 2725–2737
45. Dietz, F., Franken, S., Yoshida, K., Nakamura, H., Kappler, J., and Gieselmann, V. (2002) *Biochem. J.* **366,** 491–500
46. Stec, I., Nagl, S. B., van Ommen, G. J., and den Dunnen, J. T. (2000) *FEBS Lett.* **473,** 1–5
47. Singh, D. P., Ohguro, N., Chylack, L. T., Jr., and Shinohara, T. (1999) *Invest. Ophthalmol. Visual Sci.* **40,** 1444–1451
48. Ochs, R. L., Muro, Y., Si, Y., Ge, H., Chan, E. K., and Tan, E. M. (2000) *J. Allergy Clin. Immunol.* **105,** 1211–1220
49. Singh, D. P., Fatma, N., Kimura, A., Chylack, L. T., Jr., and Shinohara, T. (2001) *Biochem. Biophys. Res. Commun.* **283,** 943–955
50. Henikoff, S., Henikoff, J. G., and Pietrokovski, S. (1999) *Bioinformatics* **15,** 471–479
51. Fatma, N., Singh, D. P., Shinohara, T., and Chylack, L. T., Jr. (2001) *J. Biol. Chem.* **276,** 48899–48907
52. Shinohara, T., Singh, D. P., and Fatma, N. (2002) *Prog. Retin. Eye Res.* **21,** 341–358
53. Ge, H., Si, Y., and Wolffe, A. P. (1998) *Mol. Cell* **2,** 751–759
54. Schroder, A. R., Shinn, P., Chen, H., Berry, C., Ecker, J. R., and Bushman, F. (2002) *Cell* **110,** 521–529
55. Israel, N., and Gougerot-Pocidalo, M. A. (1997) *Cell. Mol. Life Sci.* **53,** 864–870
56. Roepstorff, P., and Fohlman, J. (1984) *Biomed. Mass Spectrom.* **11,** 601