

# Dynamic Modulation of HIV-1 Integrase Structure and Function by Cellular Lens Epithelium-derived Growth Factor (LEDGF) Protein<sup>\*[5]</sup>

Received for publication, July 30, 2008, and in revised form, September 15, 2008. Published, JBC Papers in Press, September 18, 2008, DOI 10.1074/jbc.M805843200

Christopher J. McKee<sup>†1</sup>, Jacques J. Kessl<sup>†1</sup>, Nikoloz Shkriabai<sup>‡</sup>, Mohd Jamal Dar<sup>§</sup>, Alan Engelman<sup>§</sup>, and Mamuka Kvaratskhelia<sup>‡2</sup>

From the <sup>‡</sup>Center for Retrovirus Research and Comprehensive Cancer Center, College of Pharmacy, The Ohio State University, Columbus, Ohio 43210 and the <sup>§</sup>Department of Cancer Immunology and AIDS, Dana-Farber Cancer Institute and Division of AIDS, Harvard Medical School, Boston, Massachusetts 02115

The mandatory integration of the reverse-transcribed HIV-1 genome into host chromatin is catalyzed by the viral protein integrase (IN), and IN activity can be regulated by numerous viral and cellular proteins. Among these, LEDGF has been identified as a cellular cofactor critical for effective HIV-1 integration. The x-ray crystal structure of the catalytic core domain (CCD) of IN in complex with the IN binding domain (IBD) of LEDGF has furthermore revealed essential protein-protein contacts. However, mutagenic studies indicated that interactions between the full-length proteins were more extensive than the contacts observed in the co-crystal structure of the isolated domains. Therefore, we have conducted detailed biochemical characterization of the interactions between full-length IN and LEDGF. Our results reveal a highly dynamic nature of IN subunit-subunit interactions. LEDGF strongly stabilized these interactions and promoted IN tetramerization. Mass spectrometric protein footprinting and molecular modeling experiments uncovered novel intra- and inter-protein-protein contacts in the full-length IN-LEDGF complex that lay outside of the observable IBD-CCD structure. In particular, our studies defined the IN tetramer interface important for enzymatic activities and high affinity LEDGF binding. These findings provide new insight into how LEDGF modulates HIV-1 IN structure and function, and highlight the potential for exploiting the highly dynamic structure of multimeric IN as a novel therapeutic target.

Integration of the reverse-transcribed RNA genome into a host chromosome is an obligatory step for HIV-1<sup>3</sup> replication

(reviewed in Ref. 1). This process is catalyzed by the retroviral enzyme integrase (IN) in two reaction steps. In the first step, which is called 3'-processing and takes place shortly after the cDNA is made, IN hydrolyzes a GT dinucleotide from each end of the viral DNA. In the second step, IN catalyzes concerted integration of the processed viral DNA ends into chromosomal DNA. The sites of attack on the two target DNA strands are separated by 5 bp, which leads to dissociation of the small double-stranded DNA fragment between the attachment sites. The subsequent repair of the intermediate species by cellular enzymes completes the integration reaction.

HIV-1 IN consists of three distinct structural and functional domains. The N-terminal domain (NTD) (residues 1–50) contains conserved pairs of histidine and cysteine residues that bind zinc (2, 3), which contributes to IN multimerization and its catalytic function (4, 5). The catalytic core domain (CCD) (residues 51–212) contains three acidic residues, Asp-64, Asp-116, and Glu-152, which play a key role in coordinating active site divalent metal ions (6, 7). The C-terminal domain (CTD) (residues 213–288) also contributes to functional IN multimerization (8, 9). Results of structural biology studies revealed each individual domain as a dimer (3, 6, 7, 10, 11) and more recent two-domain crystal structures comprised of the CCD and CTD (12) or NTD and CCD (13) likewise unveiled dimeric organizations. Functional studies suggested that a dimer of full-length IN could suffice to process each 3' end, whereas a tetramer is required to integrate both viral DNA ends into chromosomal DNA (14–16). Efforts to determine the complete IN structure have been impeded by limited protein solubility and/or the inherent flexibility of the three-domain enzyme. Full-length IN interestingly exists as a mixture of monomers, dimers, tetramers, and higher order species in the absence of DNA (9, 17–19).

While *in vitro* analysis with model DNA substrates demonstrated that IN alone could catalyze 3'-processing and DNA strand transfer reactions, the *in vivo* function of the enzyme is likely to be regulated by a number of viral and cellular proteins. Following the completion of reverse transcription, the newly synthesized cDNA remains associated with several viral pro-

<sup>\*</sup> This work was supported, in whole or in part, by National Institutes of Health Grants AI062520 (to M. K.) and AI039394 (to A. E.) from NIAID. 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.

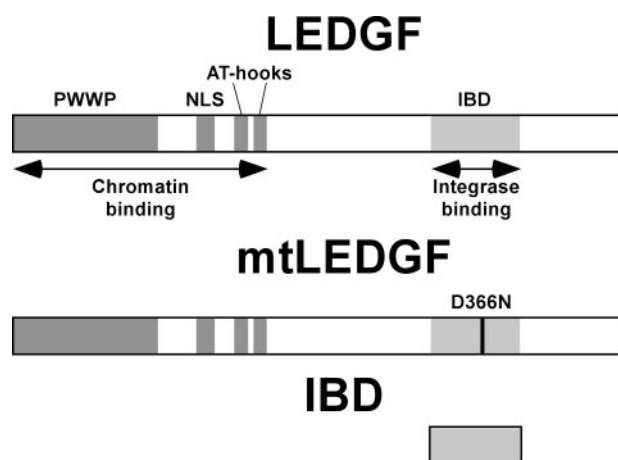
[5] The on-line version of this article (available at <http://www.jbc.org>) contains supplemental Figs. S1–S3 and Tables S1 and S2.

<sup>1</sup> Both authors contributed equally to this work.

<sup>2</sup> To whom correspondence should be addressed: The Ohio State University, 500 W. 12th Ave., 238 L.M. Parks Hall, Columbus, OH 43210. Tel.: 614-292-6091; Fax: 614-292-7766; E-mail: [kvaratskhelia.1@osu.edu](mailto:kvaratskhelia.1@osu.edu).

<sup>3</sup> The abbreviations used are: HIV-1, human immunodeficiency virus, type 1; D, donor; FS, full-site; HS, half-site; IBD, integrase-binding domain; IN, integrase; mt, mutant; NTA, nitrilotriacetic acid; NTD, N-terminal domain; CCD, catalytic core domain; CTD, C-terminal domain; LEDGF, lens epithelium-

derived growth factor; NHS, N-hydroxysuccinimide; HPG, p-hydroxyphenylglyoxal; S, substrate; P, product; PIC, preintegration complex; Di, dimer; Tet, tetramer; M, modifier; MOPS, 4-morpholinopropanesulfonic acid; CHAPS, 3-[(3-cholamidopropyl)dimethylammonio]-1-propanesulfonic acid.



**FIGURE 1. The LEDGF proteins used in our studies.** The upper wild-type protein exhibits two distinct chromatin and IN binding activities (green and blue, respectively). The middle D366N mutant (*mt*) protein is severely impaired for IN binding (47, 50). In contrast, the IBD interacts with IN but lacks the ability to bind DNA and chromatin (38, 40).

teins and recruits host factors to form the preintegration complex (PIC) (20–29). Of these, transcriptional co-activator p75, also known as lens epithelium-derived growth factor (LEDGF), is the principal cellular interactor of HIV-1 IN (27–29). A number of recent studies have indicated that LEDGF is critically important for effective HIV-1 integration and viral replication (30–33). RNA interference (RNAi)-mediated knock-down of endogenous LEDGF to below detectable levels resulted in reduction of infection to 3.5% of that observed in the presence of normal cells (33). Similarly significantly reduced levels of HIV-1 infection were detected in LEDGF knock-out mouse embryo fibroblasts (34, 35). Expression of recombinant HIV-1 IN in human cells revealed that LEDGF protects the viral protein from proteasomal degradation and tethers it to chromosomal DNA (25, 28, 36–38). Accordingly, LEDGF primarily functions during HIV-1 infection to tether PICs to active genes during integration (34). *In vitro* assays with purified recombinant proteins furthermore demonstrated that LEDGF binds directly to IN, which significantly stimulates its enzymatic activities (27, 39–44).

The N-terminal part of LEDGF contains a PWWP domain, nuclear localization signal, and dual copy of the AT-hook DNA binding motif (reviewed in Ref. 45 and Fig. 1). These conserved elements primarily mediate LEDGF association with chromatin (38, 40). An evolutionarily conserved domain in the C-terminal region (residues 347–429) mediates the interaction with IN and was thus termed the IN-binding domain (IBD) (39, 46). The solution structure of the LEDGF IBD and its co-crystallization with the IN CCD has been recently reported (47, 48). Interestingly, the IBD docks into a relatively small cavity at the CCD dimer interface, contacting both IN subunits (48). The importance of the interacting amino acids revealed from the crystal structure has been validated by site directed mutagenesis in the context of full-length recombinant proteins (29, 36, 47, 49, 50) and by the out-growth of resistant viral strains in the presence of a dominant-interfering LEDGF fragment (32). However, mutagenesis studies have also indicated that full-length IN-LEDGF interactions extend beyond the contacts observed in the co-crystal structure of the isolated domains (28).

We have undertaken a number of innovative biochemical approaches to characterize the structural and mechanistic foundations between the full-length interacting partners, which has revealed a highly dynamic nature for the interactions between free IN subunits. LEDGF moreover strongly stabilized the IN subunit-subunit contacts. Mass spectrometric surface topology studies furthermore uncovered novel protein-protein contacts, which lie outside of the central IBD-CCD co-crystal structure. Mutational analysis confirmed the importance of the identified residues and indicated a strong correlation between IN tetramer formation and high affinity LEDGF binding. These findings provide new insight into how LEDGF modulates HIV-1 IN structure/function, and highlight the potential to exploit the highly dynamic nature of IN subunit interactions as a novel therapeutic target.

## EXPERIMENTAL PROCEDURES

**Expression Plasmids and Recombinant Proteins**—HIV-1 IN proteins were expressed from pKBIN6Hthr, which was derived from pKB-IN6H (28) by replacing amino acids VDKLAAALE upstream from the C-terminal His<sub>6</sub> affinity tag with LVPRGSALE (thrombin cleavage site underlined) by PCR-directed mutagenesis. Mutations were also introduced into pKBIN6Hthr using PCR, and the coding regions of plasmids created via PCR were verified by DNA sequencing. Wild-type and mutant IN proteins were purified according to the previously described procedure (41). Purified recombinant LEDGF, mutant (*mt*) LEDGF, and IBD (Fig. 1) were obtained as described previously (28, 39, 47).

**IN 3'-Processing and DNA Strand Transfer Activities**—The <sup>32</sup>P-labeled 21-mer synthetic double-stranded DNA (50 nM) mimicking the U5 viral end sequence was used as substrate. The concentrations of wild type and mutant IN proteins as well as LEDGF and LEDGF IBD included in the reactions are indicated in the figure legends. The reactions were carried out at 37 °C for 1 h in buffer containing: 50 mM MOPS (pH 7.2), 2 mM β-mercaptoethanol, 10 mM MnCl<sub>2</sub>, 1 mM CHAPS, 50 mM NaCl, and stopped with 50 mM EDTA. Reaction products were subjected to denaturing polyacrylamide gel electrophoresis and visualized using a Storm 860 Phosphorimager (Amersham Biosciences).

**Concerted Integration Assay**—These assays were performed as described previously (43). Briefly, the 972 bp *Scal*-*Dra*III restriction fragment from pU3U5 (51) served as donor DNA and was 5'-end-labeled with [<sup>32</sup>P]ATP and T4 polynucleotide kinase. HIV-1 IN (400 nM) was assembled with the labeled donor substrate (18 nM) in the presence of 20 mM HEPES (pH 7.0), 5 mM dithiothreitol, 10 mM MgCl<sub>2</sub>, 25 μM ZnCl<sub>2</sub>, 100 mM NaCl, 5% DMSO, 10% PEG 6000. Ligands (IBD or LEDGF) were added before preincubation for 20 min at room temperature. Reactions (25-μl final volume) were initiated by adding 500 ng of circular target DNA (pGEM, Promega), and the mixtures were incubated for 1 h at 37 °C. Reactions were stopped by adding 10 mM EDTA, 0.2% SDS, and 1 mg/ml proteinase K. After ethanol precipitation, samples were subjected to 0.6% agarose gel electrophoresis for 6 h at 50 V. The gels were dried, and the labeled DNA products were detected using the Storm 860 Phosphorimager.

**Subunit Exchange Assay**—His-tagged IN (1  $\mu$ M) was preincubated with or without ligand (2  $\mu$ M LEDGF or 2  $\mu$ M mtLEDGF) in exchange buffer (25 mM Hepes, pH 7.1, 200 mM NaCl, 4% glycerol, 2 mM  $\beta$ -mercaptoethanol) for 30 min at room temperature. Tag-free IN (1  $\mu$ M) was then added and incubated for the indicated times. Aliquots were then briefly centrifuged 2 min at  $1,000 \times g$ , and supernatants were pulled-down by Ni-nitrilotriacetic acid (NTA) resin (GE Healthcare) for 10 min in the presence of bovine serum albumin (0.1 mg/ml). The IN-bound resin was then washed three times with buffer containing 50 mM HEPES (pH 7.1), 200 mM NaCl, 2 mM  $MgCl_2$ , 100 mM imidazole, and 0.1% (v/v) Nonidet P40. The bound proteins were subjected to SDS-PAGE separation and visualized by Coomassie Blue stain.

**Mass Spectrometric Footprinting**—In parallel reactions, free IN and IN+LEDGF were first incubated at room temperature for 30 min and then subjected to treatments at 37 °C with 1 mM *N*-hydroxysuccinimide (NHS)-biotin for 30 min or 20 mM *p*-hydroxyphenylglyoxal (HPG) for 60 min. These concentrations of modifying reagents were chosen because comparative pulldown experiments with untreated and modified IN-LEDGF complexes indicated that under these conditions the integrity of the preassembled protein-protein complex was fully preserved (data not shown). NHS-biotin treatment was carried out in buffer containing 50 mM HEPES (pH 8.0), 150 mM NaCl, 10 mM  $MgCl_2$ . The HPG modifications were performed in 50 mM HEPES (pH 8.0), 50 mM boric acid, 150 mM NaCl. The reactions were quenched by excess Lys and Arg using free amino acid forms. IN-LEDGF complexes were selectively pulled-down using Ni-NTA resin. The bound proteins were separated by denaturing SDS-PAGE and visualized by Microwave Blue stain (Protiga, Gaithersburg, MD). IN bands were excised, destained, and subjected to in-gel proteolysis with 0.5  $\mu$ g of trypsin. The tryptic peptides were analyzed with the Axima-CFR MALDI-ToF instrument (Shimadzu) using  $\alpha$ -cyano-4-hydroxy-cinnamic acid as a matrix.

**Size Exclusion Chromatography**—Experiments were performed with a Superdex 200 10/300 GL column (GE Healthcare) at 0.5 ml/min in buffer containing 50 mM HEPES (pH 7.4), 750 mM NaCl, and 10% glycerol. The column was calibrated with the following proteins: conalbumin (75,000 Da), carbonic anhydrase (29,000 Da), ribonuclease A (13,700 Da), and aprotinin (6,500 Da). Proteins were detected by absorbance at 280 nm.

**Molecular Modeling**—The model of the NTD-CCD tetramer bound to the LEDGF IBD was generated by overlaying the CCDs within PDB structures 2B4J (48) and 1K6Y (13) using the Insight II software package (Accelrys Inc., San Diego) on a Silicon Graphics O<sub>2</sub> work station. The constructed model was then energy-minimized by the same software package using the CFF91 force field and steepest descent method.

**LEDGF Binding Affinities to Wild-type and Mutant INs**—LEDGF (50–650 nm) was incubated with 100 nM His-tagged IN (WT or mutant) in binding buffer (50 mM Hepes (pH 7.1), 200 mM NaCl, 2 mM  $MgCl_2$ , 100 mM imidazole, 0.1% (v/v) Nonidet P40) for 60 min at room temperature. Samples were then briefly centrifuged for 2 min at  $1,000 \times g$ , and supernatants were pulled-down by Ni-NTA resin for 30 min in the presence of

bovine serum albumin (0.1 mg/ml). The resin was then washed three times with the same buffer, and the bound proteins were separated by SDS-PAGE. LEDGF was detected by Western blot analysis using a mouse monoclonal LEDGF antibody (BD Biosciences) and quantified using Image software (NIH). Plotting and curve fitting was performed with Origin 8 software (OriginLab). Nonspecific signal was not detected when LEDGF was incubated with Ni-NTA beads in the absence of IN (data not shown).

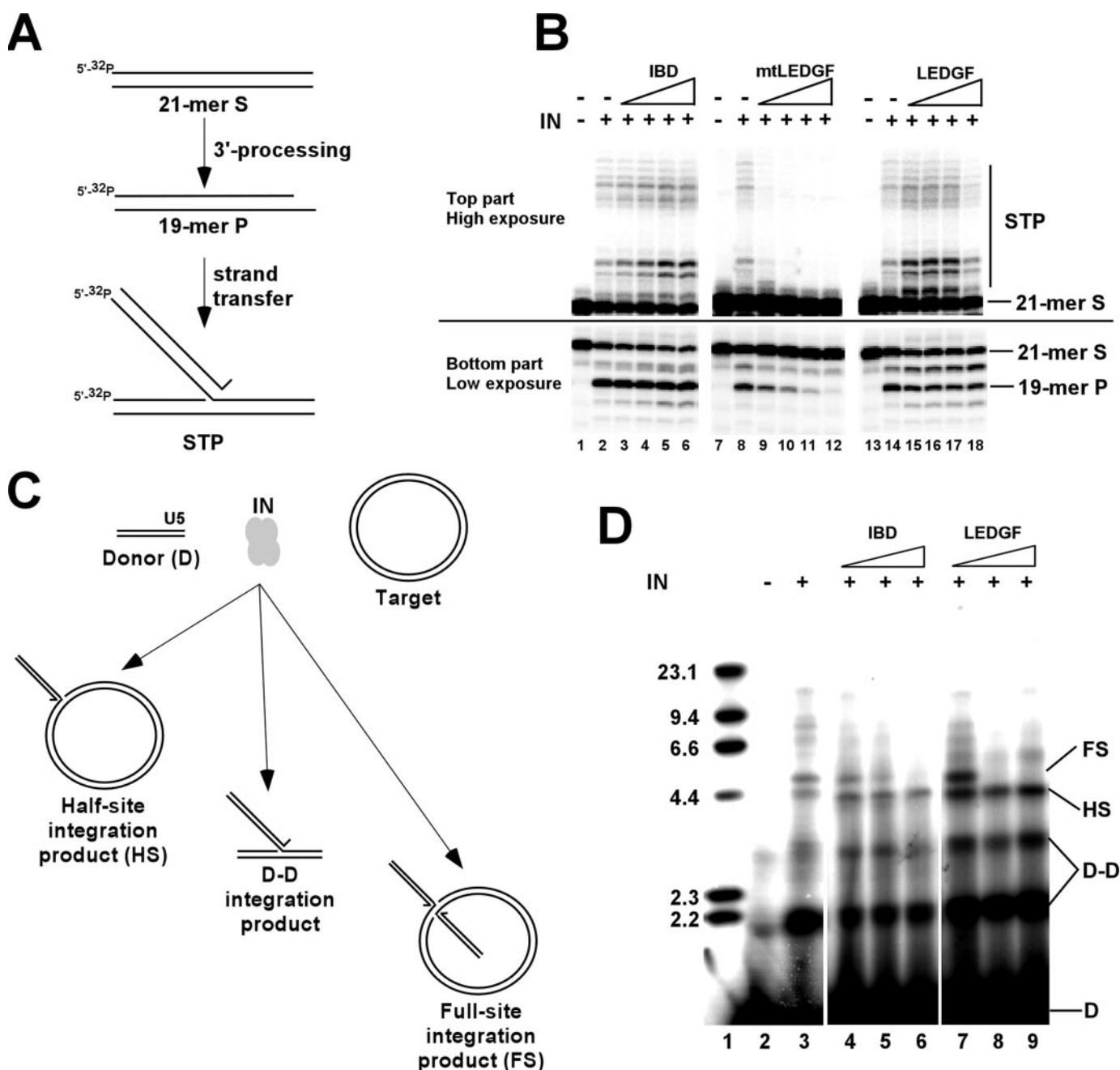
## RESULTS

We previously reported that LEDGF significantly stimulated the *in vitro* activities of HIV-1 IN whereas the isolated IBD failed to do so (39, 40). As these assays utilized relatively long blunt-ended viral DNA substrates and DNA strand transfer product formation as read-out, we reanalyzed the effects of these two proteins on IN function using an oligonucleotide-based assay that monitors the formation of 3'-processing and DNA strand transfer reaction products on denaturing sequencing gels (Fig. 2) (52). The results in *panel B (lanes 1–6)* revealed stimulation of IN DNA strand transfer activity by the LEDGF IBD under these assay conditions. It should be noted that in this setting the 19-mer 3'-processing reaction product is the substrate for the second catalytic step (Fig. 2A). To dissect if the IBD directly enhanced IN 3' processing activity, a selective IN strand transfer inhibitor (53, 54) was included in the experiment (supplemental Fig. S1). The 19-mer reaction product accumulated under these conditions, revealing significant stimulation of IN 3' processing activity by the LEDGF IBD (supplemental Fig. S1). For control experiments we analyzed the D366N point mtLEDGF, which is defective for IN binding *in vitro* (47) and in yeast cells (50), but retains LEDGF DNA binding activity (Fig. 1). In contrast with the LEDGF IBD, formation of 3'-processing and DNA strand transfer reaction products decreased with increasing mtLEDGF concentrations (Fig. 2B, *lanes 7–12*), probably due to competition between mtLEDGF and IN for binding to DNA. A bell-shape enhancement of IN activities was observed in the presence of wild-type LEDGF (Fig. 2B, *lanes 13–18*), likely due to the competition with IN for DNA binding at high LEDGF:IN ratios. The IBD, which is the weaker stimulant, could display this effect at relatively high stoichiometry (compare *lanes 6* and *18* in Fig. 2B) because it lacks the N-terminal LEDGF regions that mediate DNA binding (40) (Fig. 1).

We next analyzed the activity of the IBD using a longer donor DNA substrate and a second circular target DNA, a design that can distinguish the formation of single-end half-site (HS) integration products from those that form by the pairwise concerted integration of two viral DNA ends (labeled full-site (FS) in Fig. 2C) (42, 43). As expected (42, 43), wild-type LEDGF modestly stimulated FS and HS product formation at LEDGF:IN ratios of  $<1$ , while higher LEDGF concentrations selectively inhibited FS product formation (Fig. 2D, *lanes 7–9*). The selective reduction of FS product formation was also observed with increasing concentrations of IBD (Fig. 2D, *lanes 4–6*).

The above experiments (Fig. 2 and supplemental Fig. S1) demonstrated that direct binding of LEDGF or its IBD could

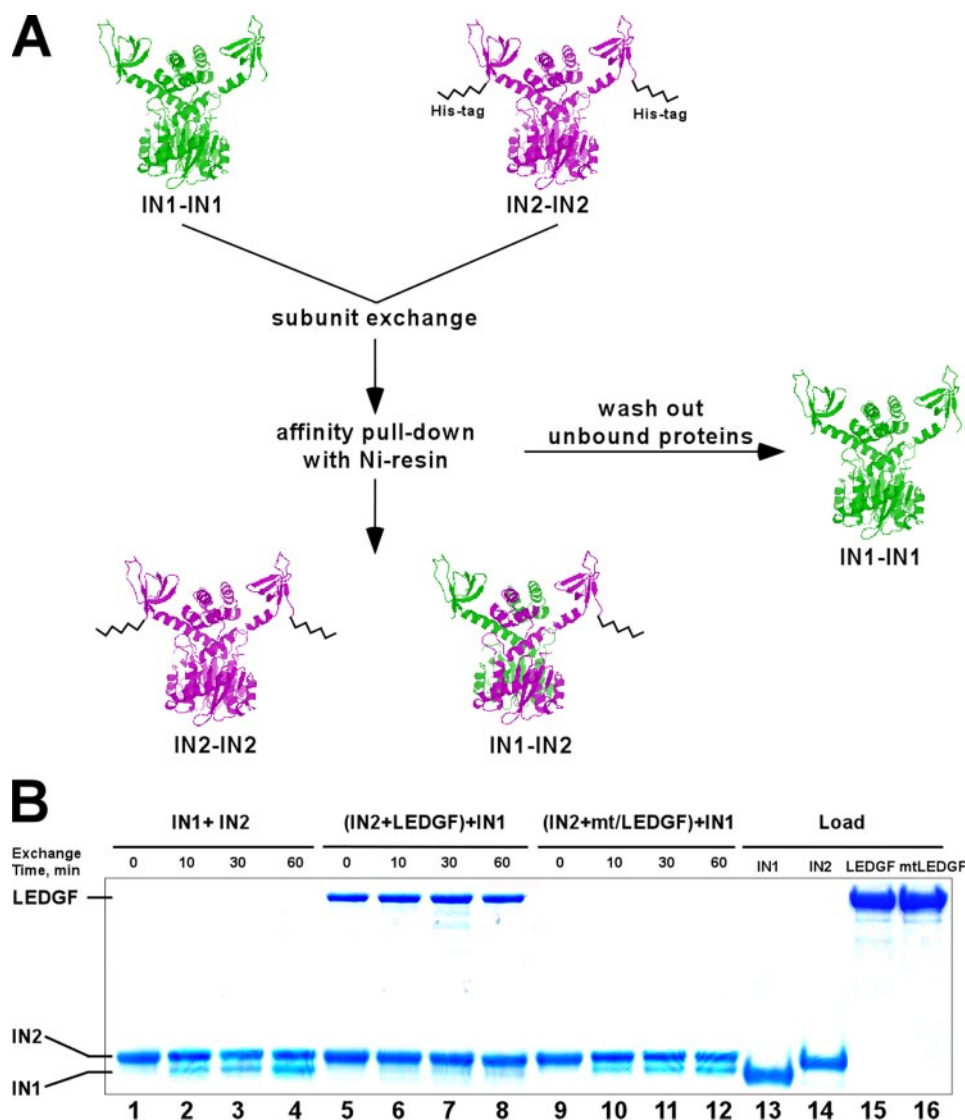




**FIGURE 2. Effects of LEDGF and the isolated IBD on IN 3'-processing and DNA strand transfer activities.** *A*, schematic of oligonucleotide-based integration assay. *B*, experimental results. The positions of 21-mer substrate (21-mer S) and the products of 3'-processing (19-mer P) and strand transfer (STP) reactions are indicated. Lanes 1, 7, and 13, 21-mer DNA substrate; lanes 2, 8, and 14, 50 nM DNA + 0.5  $\mu$ M IN. The following concentrations of the indicated LEDGF proteins were included in the reactions: lanes 3, 9, and 15, 0.25  $\mu$ M, lanes 4, 10, and 16, 0.5  $\mu$ M, lanes 5, 11, and 17, 1  $\mu$ M, and lanes 6, 12, and 18, 2  $\mu$ M. *C*, schematic diagram of concerted integration assay. Integration of one donor (D) DNA into the target plasmid yields half-site (HS) or single-end integration products. Integration of a pair of D ends into both strands of target DNA results in full-site (FS) or concerted integration. The D can otherwise integrate into separate D molecules to form D-D products. *D*, agarose gel image of reaction products. Lane 1,  $^{32}$ P-labeled lambda/HindIII DNA mass markers; lane 2, control reaction without IN; lane 3, reaction with 400 nM IN; lanes 4, 5, and 6, contained 0.25, 0.5, and 1.0  $\mu$ M IBD, respectively; lanes 7, 8, and 9 contained 0.25, 0.5, and 1.0  $\mu$ M LEDGF, respectively.

differentially influence IN activities. The co-crystal structure did not show any significant changes in the tertiary structure of the CCD upon IBD binding, but did reveal that the host factor engaged both monomers of the IN dimer at the CCD interface (48). Therefore, one possible mechanism could be that LEDGF binding influences the dynamics of IN subunit-subunit interactions. Indeed, prior analyses using IBD-based peptides hinted at this possibility (55, 56). We therefore devised the following experiment to analyze the dynamics of IN subunit exchange

(Fig. 3A). His-tagged IN (the IN2-IN2 multimer) was mixed with tag-free IN (IN1-IN1). Three IN populations could then form upon subunit exchange: IN1-IN1, IN1-IN2, and IN2-IN2. Of these, IN2-IN2 and IN1-IN2 could be pulled-down by NTA resin through affinity binding with the His-tag, while the tag-free IN1-IN1 would be washed out. Indeed, when incubated separately IN2 was affectively bound by NTA beads, while the interaction was not detected using IN1 (data not shown). However, the IN1 protein was quantitatively recovered following its



**FIGURE 3. Subunit exchange assay using His-tag IN and SDS-PAGE: design (A) and results (B).** A, subunit exchange between IN multimers was tested by mixing two different IN proteins: IN1, a tag-free form and IN2, containing the 6 $\times$  His tag at its C terminus. The full-length proteins are depicted as dimers (IN1-IN1) and (IN2-IN2). Both proteins contained wild-type IN sequences and displayed identical catalytic activities. After mixing, unbound proteins were washed from the resin and bound complexes were analyzed by SDS-PAGE. IN1 and IN2 were clearly separated based on their molecular weight differences. B, data in lanes 1–4 indicate that IN2 was able to quantitatively pull-down tag-free IN1. The recovered multimers contained a mixture of IN2-IN2 and IN1-IN2, while tag-free IN1-IN1 was washed out from the Ni-NTA resin. To assess the impact of LEDGF on IN subunit exchange, the IN2 multimer was preincubated with LEDGF and then exposed to IN1 (lanes 5–8). The results show that LEDGF interacted with IN2 and effectively prevented IN subunit exchange (lanes 5–8). In contrast, mtLEDGF did not bind IN2 or affect subunit exchange (lanes 9–12). Total amounts of input IN1, IN2, LEDGF, and mtLEDGF proteins are shown in lanes 13, 14, 15, and 16, respectively.

preincubation with IN2 due to the effective exchange of IN protein subunits. Kinetic analyses determined exchange within 10 min of mixing, reflecting the highly dynamic nature of protein-protein interactions between free IN subunits under these conditions (Fig. 3B, lanes 1–4).

We next asked how the IBD and full-length LEDGF would affect IN subunit exchange. For this, tag-free IN1 pre-incubated with LEDGF was then exposed to His-tagged IN2. LEDGF effectively prevented IN subunit exchange (Fig. 3B, lanes 5–8). Very similar results were obtained with the IBD (data not shown). In contrast, due to the inability to effectively bind IN, mtLEDGF failed to affect the dynamics of IN subunit exchange (Fig. 3B, lanes 9–12).

These results indicated that LEDGF or its IBD markedly affected IN subunit-subunit interactions but did not distinguish whether the cofactor stabilizes (“locks”) IN into a specific multimeric state or prevents the multimer from forming by interfering with subunit-subunit interactions. To delineate how the proteins modulated IN subunit exchange, we employed size exclusion chromatography (Fig. 4). The lowest detectable concentration of IN (2.5  $\mu$ M) was used to approach those employed above in activity-based assays. Under these conditions, wild type IN exhibited two main peaks with the predominant and minor species corresponding to tetramer and dimer, respectively (Fig. 4A and supplemental Table S1). The “shoulder” observed to the right of the dimer peak suggests that some monomeric protein was also present under these conditions. The addition of the IBD resulted in a single shifted peak with a retention time consistent with two LEDGF IBD molecules bound to the IN tetramer (Fig. 4A and supplemental Table S1). Because the CCD used in the crystallographic studies contained the F185K solubilizing mutation, we also assayed the F185K/C280S double mutant IN (dmIN), which is a soluble version of the full-length protein (9). By contrast to the wild type, dmIN contained greater quantities of the dimer than tetramer (Fig. 4B). Similar with the unmutated IN, the IBD shifted the equilibrium sharply in favor of the dmIN tetramer-IBD complex compared with the dimer dmIN-IBD complex. These experiments demonstrate that the IBD preferentially binds and stabilizes the IN tetramer.

To obtain more detailed information on the interaction between full-length IN and LEDGF, we turned to our mass spectrometric protein footprinting approach (Fig. 5). The method enables the comparison of surface topologies of free protein *versus* a protein-ligand complex using small chemical amino acid selective modifiers. The concentrations of the modifying reagents are optimized to ensure mild reaction conditions such that the integrity of the protein-ligand complex is preserved. Subsequent SDS-PAGE separation, proteolysis, and mass spectrometric analyses are carried out to reveal surface amino acids readily modified in free protein but shielded from

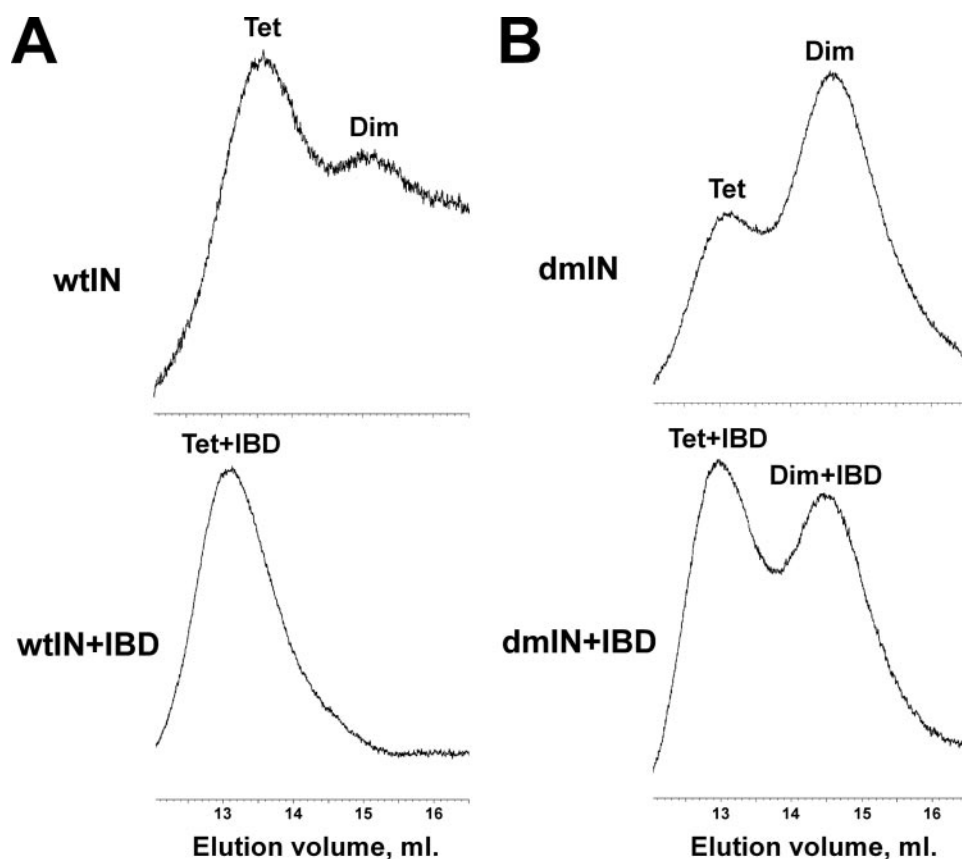


FIGURE 4. **Size exclusion chromatography of free IN and IN-IBD complexes.** 2.5  $\mu$ M IN and 5  $\mu$ M IBD were used in these experiments. *A*, chromatograms of wild-type IN protein in its free form (*top*) and complexed with IBD (*bottom*). *B*, elution profiles of the soluble double mutant (F185K/C280S) IN protein in its free form (*top*) and complexed with IBD (*bottom*). Peaks corresponding to tetramer (Tet) IN, dimer (Dim) IN, as well as IBD-bound IN tetramer (Tet+IBD) and dimer (Dim+IBD), are indicated.

modification by the interacting partner (Fig. 5, *top*). This approach has proved instrumental for studying a number of nucleic acid-protein interactions including the IN-viral DNA complex, and consistently revealed biologically relevant contact amino acids (57–64). Here, we used the method for the first time to examine protein-protein interactions. For this, in parallel reactions with free IN and pre-formed IN-LEDGF complexes were subjected to treatments with Lys- and Arg-modifying reagents. Despite tight binding, the solution mixture of IN and LEDGF was likely to contain unliganded proteins in addition to the protein complex. To ensure for selective analysis, we utilized His-tagged LEDGF and tag-free IN proteins. Following chemical modification, unliganded LEDGF and IN-LEDGF complexes recovered using Ni-NTA resin, and in parallel free IN, were separated by SDS-PAGE. IN amino acids protected from modification by LEDGF binding were then deciphered by comparing the modification patterns obtained with the different species of gel-isolated IN (Fig. 5, *top*). Control experiments utilized mtLEDGF that lacks the ability to bind IN.

Representative mass spectrometric fragments are shown in Fig. 5, and two distinct sets of modification patterns were readily observed. Three peaks containing modified Arg-107, Lys-186, and Lys-14 were detected with free IN and IN+mtLEDGF but were markedly reduced from the IN-LEDGF complex (Fig. 5, *A–C*). In contrast, two peaks containing modified Arg-228 and Lys-273, as well as an unmodified

residue 265–273 fragment, persisted in all three samples. A detailed summary of the modification patterns is presented in Table 1: of the 12 Lys and 9 Arg residues readily modified in free IN, Lys-14, Arg-107, Arg-166, Lys-186, Arg-187, and Lys-188 were selectively protected by LEDGF binding. These results are fully consistent with previous deletion analyses that revealed the CCD as the primary viral recognition determinant with the NTD donating secondary affinity-enhancing contacts. The CTD, by contrast, was dispensable for high affinity LEDGF-IN binding (28).

The footprinting results were next analyzed in the context of available protein structures to gain further insight into the details of the LEDGF-IN interaction. Protections of Arg-107 and Arg-166 are consistent with available data from unliganded (6, 7, 12, 13) as well as IBD-bound (48) CCD dimers. For example, Arg-166 is part of the so-called  $\alpha 4/5$  connector that forms the primary IBD recognition determinant from one of the IN monomers (48). The guanidino group of Arg-107 furthermore directly participates in IN dimerization through interactions with Glu-85

of the other CCD subunit (6, 7, 48). Intriguingly, our results indicated a relatively rapid exchange between IN subunits, consistent with Arg-107 accessibility in the free protein. In contrast, LEDGF binding stabilized the interacting IN subunits, rendering Arg-107 inaccessible to chemical modification (Figs. 3 and 5).

Our footprinting studies revealed new protein-protein contact residues Lys-14, Lys-186, Arg-187, and Lys-188 in the full-length IN-LEDGF complex. Of these, the latter three were unliganded and freely surface exposed in the IBD-CCD co-crystal structure (48). The co-crystal was comprised of a CCD dimer, while the results of size exclusion chromatography indicated affinity of the IBD for the full-length IN tetramer (Fig. 4). Interestingly, the two domain NTD/CCD crystal structure revealed interactions between two dimers, with Lys-14 from the NTD of one dimer and Lys-186/Arg-187/Lys-188 from the CCD of another dimer contributing to a tetrameric interface (13). In particular, supplemental Fig. S2 depicts the hydrogen bonding network between the two dimers involving the side chains of Lys-14, Lys-186, and Arg-187. The primary amine of Lys-188 is not directly involved in dimer-dimer interactions. However, two acidic residues (Glu-198 and Asp-25) located at  $\sim 4.5$  Å from Lys-188 could effectively restrict the access of NHS-biotin to this site (supplemental Fig. S2). Therefore, shielding of Lys-14, Lys-186, Arg-187, and Lys-188 in the context of the com-



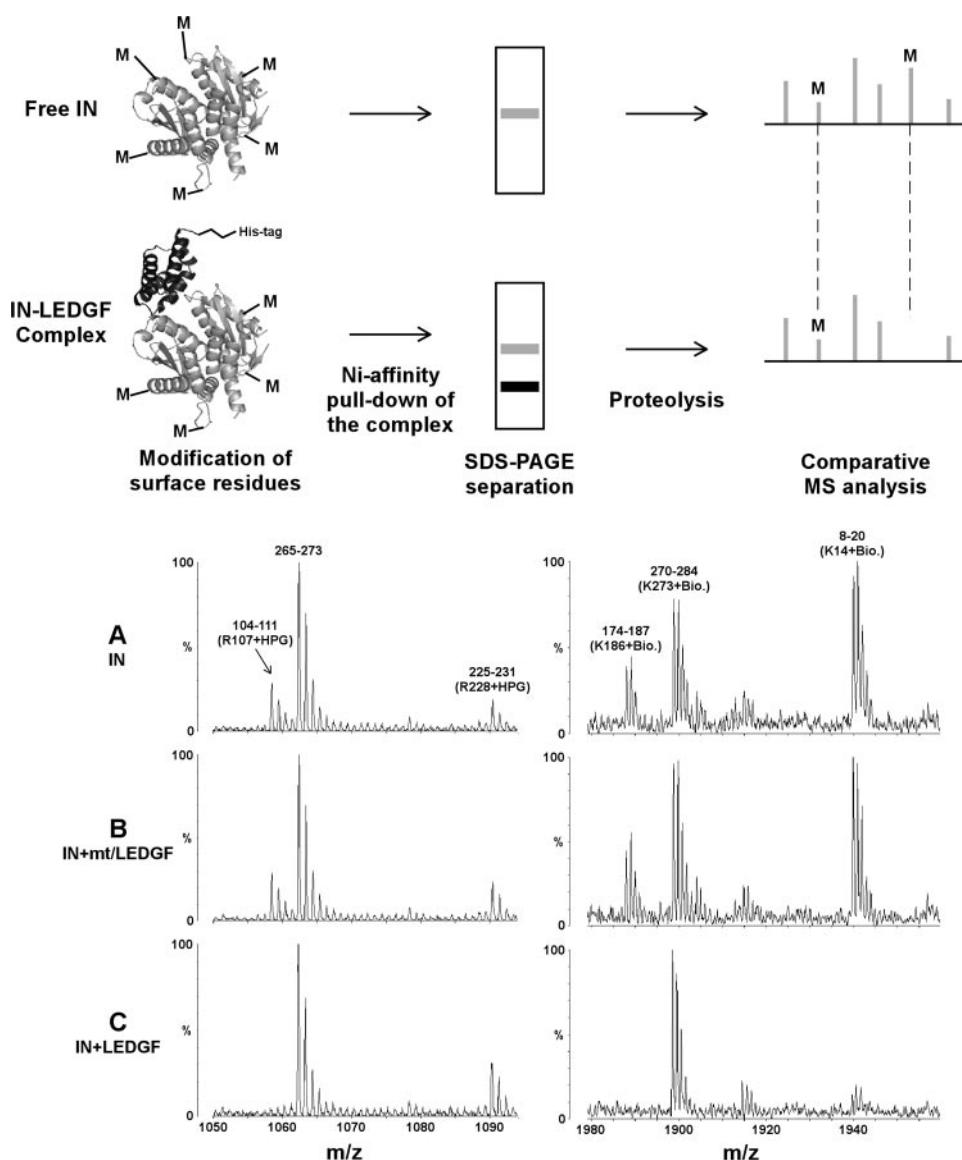


FIGURE 5. *Upper panel*, schematic presentation of the protein footprinting strategy. The structures of the CCD and the IBD-CCD complex are used for illustration while the experiments were performed with full-length IN and LEDGF. In parallel experiments free IN and the IN-LEDGF complex were subjected to treatment by small chemical modifiers (M). Surface residues in free IN and the complex were modified, but the interacting amino acids in the complex were shielded from modification. His-tag LEDGF and tag-free IN proteins were used. Following the modification reactions, the complex was pulled-down by NTA beads, which enabled recovery of only the LEDGF-bound form of IN from the reaction mixture. The interacting proteins were then separated by SDS-PAGE. The IN band was excised and subjected to in-gel proteolysis. Subsequent comparative mass spectrometry (MS) analyses revealed modification patterns in free protein and the complex. *Lower panel*, representative segments of the MALDI-ToF mass spectra. A, free IN was treated with HPG or NHS-biotin. B, IN was preincubated with mtLEDGF and then exposed to treatments with HPG or NHS-biotin. C, IN-LEDGF complexes were preformed and then exposed to HPG or NHS-biotin treatments. Start and end amino acid numbers of the detected peptide peaks are indicated. IN residues affected by modification are depicted in brackets.

plete IN-LEDGF complex could be due to interactions between two IN dimers. Also consistent with a role for the basic triad (Lys-186, Arg-187, Lys-188) in IN tetramerization is the observation that the mutation of the adjacent Phe-185 to Lys shifted the oligomeric state of IN in favor of a dimer while wtIN under the same conditions was predominantly tetrameric (compare dmIN and wtIN chromatograms in Fig. 4).

To test our hypothesis that protections of Lys-14, Lys-186, Arg-187, and Lys-188 resulted from LEDGF-mediated stabilization of tetrameric IN, we conducted site-directed mutagenesis

experiments (Fig. 6). Single point mutations of the target residues significantly compromised tetramer formation (Fig. 6A). For example, K14A, K186A, and R187A mutants were predominantly dimeric, even at relatively high protein concentration (10  $\mu$ M) (Fig. 6A and supplemental Table S2). The K188A substitution had relatively modest effect on tetramer formation (Fig. 6A) consistent with the lesser role of this residue in direct dimer-dimer interactions (supplemental Fig. S2).

We next examined the effects of these mutations on the LEDGF-IN interaction. For this, increasing concentrations of wild-type tag-free LEDGF were incubated with His-tagged IN proteins, and the fractions of LEDGF recovered by Ni-NTA pull-down were quantitated (Fig. 6, B and C). These experiments yielded an apparent  $K_d$  of  $\sim 200$  nM for the interaction between wild-type IN and LEDGF. The K14A, K186A, and R187A mutants exhibited significantly reduced affinity for LEDGF, while the K188A protein was relatively more effective at binding (Fig. 6, B and C). In fact, the comparison of size exclusion chromatography (Fig. 6A) and binding (Fig. 6, B and C) results suggested a strong correlation between IN tetramer formation and high affinity LEDGF binding.

The HIV-1 IN dimer can suffice to process viral DNA 3' ends whereas the tetramer has been implicated in DNA strand transfer activity (14–16); results of a separate study using chimera IN however indicated that the tetramer was required for efficient 3'-processing activity (65). Because the relevant protein-protein interfaces within the catalytic complexes

are for the most part unknown, we investigated whether the tetrameric interface important for high affinity LEDGF binding played a significant role in IN catalysis. Impressively, the K14A, K186A, and R187A substitutions adversely affected 3'-processing and DNA strand transfer activities (Fig. 6D). In contrast, the wild-type level of activities was observed with the K188A protein. Taken together, our results suggest that the tetramer interface involving basic residues Lys-14, Lys-186, and Arg-187 is important for HIV-1 IN 3'-processing and DNA strand transfer activities.

**TABLE 1**  
Summary of mass spectrometric footprinting results

Modified amino acids	Free IN	IN+LEDGF	IN+mtLEDGF
K7	+	+	+
K14	+	— <sup>b</sup>	+
R107	+	—	+
K111	+	+	+
K160	+	+	+
R166	+	—	+
K186	+	—	+
R187	+	—	+
K188	+	—	+
K215	+	+	+
K219	+	+	+
R224	+	+	+
R228	+	+	+
R231	+	+	+
K236	+	+	+
R262	+	+	+
R263	+	+	+
K264	+	+	+
K266	+	+	+
R269	+	+	+
K273	+	+	+

<sup>a</sup> +, surface residues susceptible to modification.

<sup>b</sup> —, residues protected from modification because of protein-protein contacts.

While the NTD-CCD tetramer was observed in a crystal lattice at high concentrations of the two-domain protein, it is less clear how LEDGF could promote tetramer formation in solution at significantly lower protein concentrations. To address this, we superimposed the structures of the CCD-IBD complex (two IBD molecules bound to the CCD dimer, see Ref. 48) and the two domain (NTD+CCD) tetramer (13) using molecular modeling (Fig. 7). The results indicated asymmetric interactions of the IBD molecules with the NTD+CCD tetramer. For example, two (colored magenta) of the four IBD molecules could effectively bridge between the two IN dimers by coordinating the CCD of one dimer and establishing additional electrostatic interactions with the NTD of another dimer (Fig. 7). These additional contacts could contribute to the high affinity IBD-IN interactions and stabilize the IN tetramer. The other two IBD molecules (colored gray) interacted with the CCD dimer interfaces but could not establish additional charge-charge contacts with the NTDs due to the spatial separation between the IN domains (Fig. 7). The lack of such interactions may reduce the binding affinity for these IBD molecules. In other words, we propose that the IN tetramer has two high affinity and two lower affinity binding sites for the LEDGF IBD.

## DISCUSSION

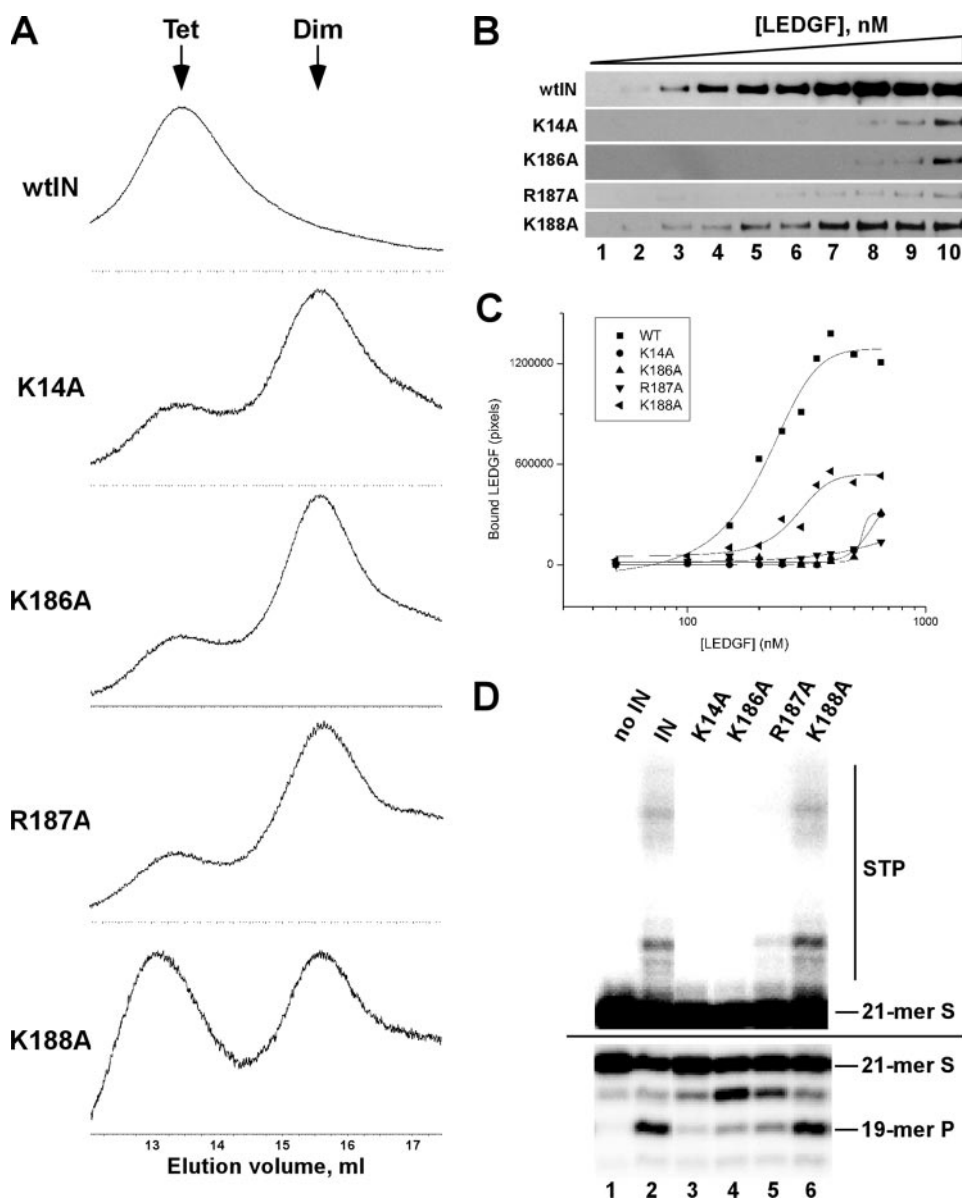
The present studies revealed a highly dynamic nature of interactions between IN subunits, which could be essential for its biological function as well as exploited as a novel therapeutic target (see below). The cellular cofactor LEDGF strongly modulated the dynamic structure of HIV-1 IN by stabilizing subunit-subunit interactions. Unlike the published co-crystal structure of the isolated domains that indicated binding of two IBD molecules to the CCD dimer, our experiments with the full-length proteins demonstrated the importance of the IN tetramer for high affinity LEDGF binding. These results are consistent with previous findings that endogenous LEDGF protein associated with tetrameric recombinant HIV-1 IN in human cells (27).

Our results also indicate that the stabilized IN tetramer is very effective at catalyzing IN 3'-processing and DNA strand transfer activities (Figs. 2, 6, and supplemental Fig. S1). Whereas free IN exists in solution as a mixture of monomer, dimer, tetramer, and high order species, the IBD strongly promoted formation of the IN tetramer (Fig. 4) and stimulated IN catalytic function (Fig. 2 and supplemental Fig. S1).

Our footprinting and molecular modeling studies uncovered novel intra- and inter-protein-protein contacts in the full-length IN complex with LEDGF (Figs. 5 and 7). Site-directed mutagenesis experiments confirmed the importance of Lys-14, Lys-186, and Lys-187 residues for tetramer formation and high affinity LEDGF binding (Fig. 6). Furthermore, the K14A, K186A, and R187A mutations also compromised IN 3'-processing and DNA strand transfer activities. These results suggest the importance of the common tetramer interface for IN interactions with DNA substrates and LEDGF. Consistently, our previous footprinting analysis of the IN-DNA complex revealed a role for K14 in DNA binding (62). The protections in the nucleoprotein complex could arise from direct protein-DNA or DNA-induced protein-protein interactions (62). The present studies clarify that Lys-14 could be a critical dimer-dimer contact essential for effective binding of IN with both DNA and LEDGF. Of note, our findings corroborate results of a number of virus-based assays that indicated essential roles for Lys-14, Lys-186, and Arg-187 (62, 66–70) in HIV-1 infection, whereas the K188A mutation resulted in reduced but reproducible levels of virus spread (69). Therefore, we propose that the IN tetramer is the biologically relevant form responsible for catalytic activities and high affinity binding to LEDGF.

The fact that LEDGF selectively impaired concerted integration of two HIV-1 DNA ends (Fig. 2, see also Refs. 42, 43) has been rather puzzling. How can this observation be reconciled *in vivo*? The following two scenarios can be considered. Free IN could first assemble onto the viral DNA ends before encountering LEDGF (43). The high degree of flexibility within individual IN subunits (62, 71, 72) as well as their dynamic interplay (Fig. 3) could be critical for effective assembly of the synaptic complex, where the two catalytic sites position themselves for concerted integration. The preassembled IN tetramer-viral DNA complex would engage LEDGF, which in turn would tether the nucleoprotein complex to active genes without significantly affecting structural arrangements of IN with its DNA substrates. Accordingly, *in vitro* experiments with model DNA substrates and purified proteins indicated the importance of the order of IN binding to DNA and LEDGF for effective concerted integration (42, 43). An alternative possibility is that the viral DNA-IN-LEDGF complex engages an as yet unidentified cellular chromatin factor(s). Binding this hypothetical cellular partner could trigger structural changes within the nucleoprotein complex to enable concerted integration. It has been reported that the IN tetramer catalyzes the sequential joining of the two viral DNA ends to target DNA *in vitro* (16), and that LEDGF strongly facilitates single end HIV-1 integration (41–43). Protein-protein interactions at chromatin could prompt LEDGF to partly or fully disengage the PIC to allow IN to regain its flexibility and complete the integration of the second viral DNA end.





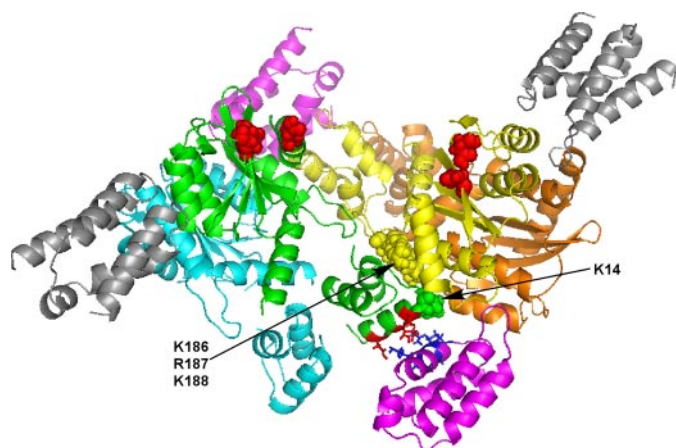
**FIGURE 6. Biochemical characterization of IN mutant proteins.** A, size exclusion chromatography of wild type and indicated mutants. All protein concentrations were 10  $\mu$ M. The elution time points corresponding to tetrameric and dimeric IN are indicated by arrows. B, Western blot results for LEDGF binding to wild type and mutant IN proteins. Increasing concentrations of LEDGF were incubated with 100 nM of the indicated IN. The concentrations of LEDGF were as follows: lane 1: 50 nM, lane 2: 100 nM, lane 3: 150 nM, lane 4: 200 nM, lane 5: 250 nM, lane 6: 300 nM, lane 7: 350 nM, lane 8: 400 nM, lane 9: 500 nM, lane 10: 650 nM. C, quantitative analysis of the Western blot results. D, 3'-processing and DNA strand transfer activities of the indicated IN proteins; other labeling is the same as in Fig. 2B.

Our structure-function studies have significant implications for exploiting the highly dynamic nature of HIV-1 IN as a novel therapeutic target. For example, inhibition of functional IN could be accomplished by the following mechanisms: 1) interfering with subunit-subunit assembly and 2) restricting IN flexibility by "locking" the protein into a functionally compromised multimeric state. Of note, such compounds would be complementary to the clinically approved IN active site inhibitor Raltegravir, as the resistant mutations developed to this drug are significantly distanced from the protein-protein interfaces (73). In general, protein-protein interactions are thought to be rather challenging drug intervention targets for the following reasons. The shape of

such interfaces is typically flat and comprises large surface areas ( $\sim 750$ – $1,500$  Å<sup>2</sup>) (reviewed in Ref. 74). However, the subset of interfaces that contribute to high affinity binding could be significantly smaller and affected by small molecule inhibitors. Indeed, significant progress has been made in discovering a number of protein-protein inhibitors in recent years (reviewed in Ref. 74).

Our results indicate that IN monomers could also be a plausible target for interfacial inhibitors. For example, the fact that IN subunits exchange rapidly (Fig. 3) provides biological targets for effective binding of small molecules. Consistently, our mass spectrometric footprinting revealed that Arg-107, which is located at the dimer interface, is readily modified by HPG (Fig. 5). While HPG cannot be viewed as a lead compound, detailed analysis of the CCD structure revealed the following intriguing role for this residue. Arg-107 together with Gly-106 form a bulge that effectively docks into a cavity within the interacting subunit (supplemental Fig. S3) to stabilize the CCD dimer. The structural dimensions depicted in supplemental Fig. S3 suggest that this site could be selectively targeted by small molecules. Of note, discovery of such structural pockets provide a strong rationale for exploiting IN-IN (supplemental Fig. S3) and IN-LEDGF (48) interactions as therapeutic targets.

An alternative intriguing mechanism could be to restrict the catalytically important flexibility of IN by stabilizing a multimeric state. For example, we showed that the LEDGF IBD forms a stable complex with the IN tetramer, which is active for 3'-processing and single-end DNA strand transfer but not concerted integration *in vitro* (Fig. 2). Previous reports showed that overexpression of LEDGF IBD proteins in target cells effectively impaired HIV-1 replication (30, 33). Furthermore, the IBD was significantly more effective at suppressing HIV-1 replication in LEDGF-deficient cells (555-fold) as compared with cells containing normal LEDGF levels ( $\sim 30$  fold) (33). These observations cannot be fully explained by a competition between the IBD and endogenous LEDGF for IN binding. Instead, our findings provide mechanistic clues that direct binding of the IBD to the IN tetramer restricts protein



**FIGURE 7. A model for the IBD bound NTD-CCD tetramer.** Two available crystal structures of the NTD-CCD tetramer and the IBD-CCD complex were superimposed. Individual subunits of IN are colored cyan, green, yellow, and orange. The acidic residues, which coordinate catalytic metals, in the green and yellow subunits are in spheres and colored red. For clarity only Lys-14 in the green subunit and the basic triad (Lys-186, Arg-187, Lys-188) in the yellow subunit are depicted as red sticks. These amino acids potentially establish charge-charge interactions with the side chains of Lys-401, Lys-402, Arg-404, and Arg-405 (blue sticks) of the IBD helix. The additional two lower affinity binding IBD molecules (gray), which coordinate the CCDs of either cyan-green or yellow-orange dimers, are significantly distanced from the tetramer interfaces.

flexibility required to form the fully functional DNA strand transfer complex.

While the studies with the LEDGF IBD provide proof-of-concept for a new mechanism of IN inhibition, the main interest is to discover small molecule inhibitors. Such compounds do not have to compete with IN subunit-subunit or IN-LEDGF interactions and overcome large energy barriers created by protein-protein interfaces. Instead, they could specifically exploit the structural “pockets” present in tetrameric or dimeric proteins to stabilize the interacting subunits and compromise the catalytically essential dynamic structure. Recent reports indicate that IBD-derived short peptides impaired IN catalytic activities *in vitro* (55, 56) and inhibited HIV-1 infection (55). In fact, earlier observations that certain small molecule inhibitors selectively bind at the IN dimer interface gain particular interest in the context of recent elucidation of structural and mechanistic details of LEDGF-IN interactions. For example, 3,4-dihydroxyphenyl-triphenylarsonium bromide has been reported to bind the IN dimer at a site that overlaps the IBD binding pocket (75). Our previous studies revealed highly selective binding of another small molecule, methyl *N,O*-bis(3,4-diacetoxycinnamoyl)-serinate, to the IN dimer at the site that is immediately adjacent to the IBD-binding pocket (60). A coumarin-based inhibitor was likewise determined to bind IN nearby the IBD interaction site (76). Further research in this direction may well lead to the discovery of novel types of clinically useful HIV-1 IN inhibitors.

**Acknowledgments**—We thank Jocelyn Norris for help with purifying LEDGF proteins. The integrase strand transfer inhibitor (118-D-24) was obtained through the National Institutes of Health AIDS Research and Reference Reagent Program, Division of AIDS, NIAID, National Institutes of Health.

## REFERENCES

- Brown, P. O. (1997) in *Retroviruses* (Coffin, J. M., Hughes, S. H., and Varmus, H. E., eds), pp. 161–204, Cold Spring Harbor Laboratory, Plainview, NY
- Bushman, F. D., Engelman, A., Palmer, I., Wingfield, P., and Craigie, R. (1993) *Proc. Natl. Acad. Sci. U. S. A.* **90**, 3428–3432
- Cai, M., Zheng, R., Caffrey, M., Craigie, R., Clore, G. M., and Gronenborn, A. M. (1997) *Nat. Struct. Biol.* **4**, 567–577
- Zheng, R., Jenkins, T. M., and Craigie, R. (1996) *Proc. Natl. Acad. Sci. U. S. A.* **93**, 13659–13664
- Lee, S. P., Xiao, J., Knutson, J. R., Lewis, M. S., and Han, M. K. (1997) *Biochemistry* **36**, 173–180
- Dyda, F., Hickman, A. B., Jenkins, T. M., Engelman, A., Craigie, R., and Davies, D. R. (1994) *Science* **266**, 1981–1986
- Goldgur, Y., Dyda, F., Hickman, A. B., Jenkins, T. M., Craigie, R., and Davies, D. R. (1998) *Proc. Natl. Acad. Sci. U. S. A.* **95**, 9150–9154
- Andrake, M. D., and Skalka, A. M. (1995) *J. Biol. Chem.* **270**, 29299–29306
- Jenkins, T. M., Engelman, A., Ghirlando, R., and Craigie, R. (1996) *J. Biol. Chem.* **271**, 7712–7718
- Eijkelenboom, A. P., Lutzke, R. A., Boelens, R., Plasterk, R. H., Kaptein, R., and Hard, K. (1995) *Nat. Struct. Biol.* **2**, 807–810
- Lodi, P. J., Ernst, J. A., Kuszewski, J., Hickman, A. B., Engelman, A., Craigie, R., Clore, G. M., and Gronenborn, A. M. (1995) *Biochemistry* **34**, 9826–9833
- Chen, J. C., Krucinski, J., Miercke, L. J., Finer-Moore, J. S., Tang, A. H., Leavitt, A. D., and Stroud, R. M. (2000) *Proc. Natl. Acad. Sci. U. S. A.* **97**, 8233–8238
- Wang, J. Y., Ling, H., Yang, W., and Craigie, R. (2001) *EMBO J.* **20**, 7333–7343
- Faure, A., Calmels, C., Desjobert, C., Castroviejo, M., Caumont-Sarcos, A., Tarrago-Litvak, L., Litvak, S., and Parissi, V. (2005) *Nucleic Acids Res.* **33**, 977–986
- Guiot, E., Carayon, K., Delelis, O., Simon, F., Tauc, P., Zubin, E., Gottikh, M., Mouscadet, J. F., Brochon, J. C., and Deprez, E. (2006) *J. Biol. Chem.* **281**, 22707–22719
- Li, M., Mizuuchi, M., Burke, T. R., Jr., and Craigie, R. (2006) *EMBO J.* **25**, 1295–1304
- van Gent, D. C., Elgersma, Y., Bolk, M. W., Vink, C., and Plasterk, R. H. (1991) *Nucleic Acids Res.* **19**, 3821–3827
- Vincent, K. A., Ellison, V., Chow, S. A., and Brown, P. O. (1993) *J. Virol.* **67**, 425–437
- Deprez, E., Tauc, P., Leh, H., Mouscadet, J. F., Auclair, C., and Brochon, J. C. (2000) *Biochemistry* **39**, 9275–9284
- Bukrinsky, M. I., Sharova, N., McDonald, T. L., Pushkarskaya, T., Tarpley, W. G., and Stevenson, M. (1993) *Proc. Natl. Acad. Sci. U. S. A.* **90**, 6125–6129
- Farnet, C. M., and Bushman, F. D. (1997) *Cell* **88**, 483–492
- Miller, M. D., Farnet, C. M., and Bushman, F. D. (1997) *J. Virol.* **71**, 5382–5390
- Li, L., Olvera, J. M., Yoder, K. E., Mitchell, R. S., Butler, S. L., Lieber, M., Martin, S. L., and Bushman, F. D. (2001) *EMBO J.* **20**, 3272–3281
- Lin, C. W., and Engelman, A. (2003) *J. Virol.* **77**, 5030–5036
- Llano, M., Delgado, S., Vanegas, M., and Poeschla, E. M. (2004) *J. Biol. Chem.* **279**, 55570–55577
- Iordanskiy, S., Berro, R., Altieri, M., Kashanchi, F., and Bukrinsky, M. (2006) *Retrovirology* **3**, 4
- Cherepanov, P., Maertens, G., Proost, P., Devreese, B., Van Beeumen, J., Engelborghs, Y., De Clercq, E., and Debyser, Z. (2003) *J. Biol. Chem.* **278**, 372–381
- Maertens, G., Cherepanov, P., Pluymers, W., Busschots, K., De Clercq, E.,

- Debyser, Z., and Engelborghs, Y. (2003) *J. Biol. Chem.* **278**, 33528–33539
29. Turlure, F., Devroe, E., Silver, P. A., and Engelman, A. (2004) *Front. Biosci.* **9**, 3187–3208
30. De Rijck, J., Vandekerckhove, L., Gijsbers, R., Hombrouck, A., Hendrix, J., Vercammen, J., Engelborghs, Y., Christ, F., and Debyser, Z. (2006) *J. Virol.* **80**, 11498–11509
31. Vandekerckhove, L., Christ, F., Van Maele, B., De Rijck, J., Gijsbers, R., Van den Haute, C., Witvrouw, M., and Debyser, Z. (2006) *J. Virol.* **80**, 1886–1896
32. Hombrouck, A., De Rijck, J., Hendrix, J., Vandekerckhove, L., Voet, A., De Maeyer, M., Witvrouw, M., Engelborghs, Y., Christ, F., Gijsbers, R., and Debyser, Z. (2007) *PLoS Pathog.* **3**, e47
33. Llano, M., Saenz, D. T., Meehan, A., Wongthida, P., Peretz, M., Walker, W. H., Teo, W., and Poeschla, E. M. (2006) *Science* **314**, 461–464
34. Shun, M. C., Raghavendra, N. K., Vandegraaff, N., Daigle, J. E., Hughes, S., Kellam, P., Cherepanov, P., and Engelman, A. (2007) *Genes Dev.* **21**, 1767–1778
35. Marshall, H. M., Ronen, K., Berry, C., Llano, M., Sutherland, H., Saenz, D., Bickmore, W., Poeschla, E., and Bushman, F. D. (2007) *PLoS ONE* **2**, e1340
36. Emiliani, S., Mousnier, A., Busschots, K., Maroun, M., Van Maele, B., Tempe, D., Vandekerckhove, L., Moisan, F., Ben-Slama, L., Witvrouw, M., Christ, F., Rain, J. C., Dargemont, C., Debyser, Z., and Benarous, R. (2005) *J. Biol. Chem.* **280**, 25517–25523
37. Llano, M., Vanegas, M., Fregoso, O., Saenz, D., Chung, S., Peretz, M., and Poeschla, E. M. (2004) *J. Virol.* **78**, 9524–9537
38. Llano, M., Vanegas, M., Hutchins, N., Thompson, D., Delgado, S., and Poeschla, E. M. (2006) *J. Mol. Biol.* **360**, 760–773
39. Cherepanov, P., Devroe, E., Silver, P. A., and Engelman, A. (2004) *J. Biol. Chem.* **279**, 48883–48892
40. Turlure, F., Maertens, G., Rahman, S., Cherepanov, P., and Engelman, A. (2006) *Nucleic Acids Res.* **34**, 1653–1675
41. Cherepanov, P. (2007) *Nucleic Acids Res.* **35**, 113–124
42. Pandey, K. K., Sinha, S., and Grandgenett, D. P. (2007) *J. Virol.* **81**, 3969–3979
43. Raghavendra, N. K., and Engelman, A. (2007) *Virology* **360**, 1–5
44. Yu, F., Jones, G. S., Hung, M., Wagner, A. H., MacArthur, H. L., Liu, X., Leavitt, S., McDermott, M. J., and Tsiang, M. (2007) *Biochemistry* **46**, 2899–2908
45. Engelman, A., and Cherepanov, P. (2008) *PLoS Pathog.* **4**, e1000046
46. Vanegas, M., Llano, M., Delgado, S., Thompson, D., Peretz, M., and Poeschla, E. (2005) *J. Cell Sci.* **118**, 1733–1743
47. Cherepanov, P., Sun, Z. Y., Rahman, S., Maertens, G., Wagner, G., and Engelman, A. (2005) *Nat. Struct. Mol. Biol.* **12**, 526–532
48. Cherepanov, P., Ambrosio, A. L., Rahman, S., Ellenberger, T., and Engelman, A. (2005) *Proc. Natl. Acad. Sci. U. S. A.* **102**, 17308–17313
49. Busschots, K., Voet, A., De Maeyer, M., Rain, J. C., Emiliani, S., Benarous, R., Desender, L., Debyser, Z., and Christ, F. (2007) *J. Mol. Biol.* **365**, 1480–1492
50. Rahman, S., Lu, R., Vandegraaff, N., Cherepanov, P., and Engelman, A. (2007) *Virology* **357**, 79–90
51. Cherepanov, P., Surratt, D., Toelen, J., Pluymers, W., Griffith, J., De Clercq, E., and Debyser, Z. (1999) *Nucleic Acids Res.* **27**, 2202–2210
52. Bushman, F. D., and Craigie, R. (1991) *Proc. Natl. Acad. Sci. U. S. A.* **88**, 1339–1343
53. Zhang, X., Pais, G. C., Svarovskaia, E. S., Marchand, C., Johnson, A. A., Karki, R. G., Nicklaus, M. C., Pathak, V. K., Pommier, Y., and Burke, T. R. (2003) *Bioorg. Med. Chem. Lett.* **13**, 1215–1219
54. Svarovskaia, E. S., Barr, R., Zhang, X., Pais, G. C., Marchand, C., Pommier, Y., Burke, T. R., Jr., and Pathak, V. K. (2004) *J. Virol.* **78**, 3210–3222
55. Hayouka, Z., Rosenbluh, J., Levin, A., Loya, S., Lebendiker, M., Veprintsev, D., Kotler, M., Hizi, A., Loyter, A., and Friedler, A. (2007) *Proc. Natl. Acad. Sci. U. S. A.* **104**, 8316–8321
56. Al-Mawsawi, L. Q., Christ, F., Dayam, R., Debyser, Z., and Neamati, N. (2008) *FEBS Lett.* **582**, 1425–1430
57. Liu, Y., Kvaratskhelia, M., Hess, S., Qu, Y., and Zou, Y. (2005) *J. Biol. Chem.* **280**, 32775–32783
58. Marchand, C., Krajewski, K., Lee, H. F., Antony, S., Johnson, A. A., Amin, R., Roller, P., Kvaratskhelia, M., and Pommier, Y. (2006) *Nucleic Acids Res.* **34**, 5157–5165
59. Shell, S. M., Hess, S., Kvaratskhelia, M., and Zou, Y. (2005) *Biochemistry* **44**, 971–978
60. Shkriabai, N., Patil, S. S., Hess, S., Budihis, S. R., Craigie, R., Burke, T. R., Jr., Le Grice, S. F., and Kvaratskhelia, M. (2004) *Proc. Natl. Acad. Sci. U. S. A.* **101**, 6894–6899
61. Shkriabai, N., Datta, S. A., Zhao, Z., Hess, S., Rein, A., and Kvaratskhelia, M. (2006) *Biochemistry* **45**, 4077–4083
62. Zhao, Z., McKee, C. J., Kessl, J. J., Santos, W. L., Daigle, J. E., Engelman, A., Verdine, G., and Kvaratskhelia, M. (2008) *J. Biol. Chem.* **283**, 5632–5641
63. Datta, S. A., Zhao, Z., Clark, P. K., Tarasov, S., Alexandratos, J. N., Campbell, S. J., Kvaratskhelia, M., Lebowitz, J., and Rein, A. (2007) *J. Mol. Biol.* **365**, 799–811
64. Kvaratskhelia, M., Miller, J. T., Budihis, S. R., Pannell, L. K., and Le Grice, S. F. (2002) *Proc. Natl. Acad. Sci. U. S. A.* **99**, 15988–15993
65. Bosserman, M. A., O'Quinn, D. F., and Wong, I. (2007) *Biochemistry* **46**, 11231–11239
66. Cannon, P. M., Wilson, W., Byles, E., Kingsman, S. M., and Kingsman, A. J. (1994) *J. Virol.* **68**, 4768–4775
67. Petit, C., Schwartz, O., and Mammano, F. (2000) *J. Virol.* **74**, 7119–7126
68. Tsurutani, N., Kubo, M., Maeda, Y., Ohashi, T., Yamamoto, N., Kannagi, M., and Masuda, T. (2000) *J. Virol.* **74**, 4795–4806
69. Lu, R., Limon, A., Devroe, E., Silver, P. A., Cherepanov, P., and Engelman, A. (2004) *J. Virol.* **78**, 12735–12746
70. Berthou, L., Sebastian, S., Muesing, M. A., and Luban, J. (2007) *Virology* **364**, 227–236
71. Engelman, A., Bushman, F. D., and Craigie, R. (1993) *EMBO J.* **12**, 3269–3275
72. van Gent, D. C., Vink, C., Groeneger, A. A., and Plasterk, R. H. (1993) *EMBO J.* **12**, 3261–3267
73. Malet, I., Delelis, O., Valantin, M. A., Montes, B., Soulie, C., Wirden, M., Tchertanov, L., Peytavin, G., Reynes, J., Mouscadet, J. F., Katlama, C., Calvez, V., and Marcelin, A. G. (2008) *Antimicrob. Agents Chemother.* **52**, 1351–1358
74. Arkin, M. R., and Wells, J. A. (2004) *Nat. Rev. Drug Discov.* **3**, 301–317
75. Molteni, V., Greenwald, J., Rhodes, D., Hwang, Y., Kwiatkowski, W., Bushman, F. D., Siegel, J. S., and Choe, S. (2001) *Acta Crystallogr. D. Biol. Crystallogr.* **57**, 536–544
76. Al-Mawsawi, L. Q., Fikkert, V., Dayam, R., Witvrouw, M., Burke, T. R., Jr., Borchers, C. H., and Neamati, N. (2006) *Proc. Natl. Acad. Sci. U. S. A.* **103**, 10080–10085