Mass Spectrometric Analysis of the HIV-1 Integrase-Pyridoxal 5′-Phosphate Complex Reveals a New Binding Site for a Nucleotide Inhibitor*

Kerry L. Williams‡, Yijun Zhang§, Nick Shkriabai‡, Rajeshri G. Karki‖, Marc C. Nicklaus‖, Nana Kotrikadze¶, Sonja Hess**, Stuart F. J. Le Grice§, Robert Craigie**, Vinay K. Pathak§, and Mamuka Kvaratskhelia‡ ‡‡

*From the ‡Ohio State University Health Sciences Center, College of Pharmacy, Center for Retrovirus Research and Comprehensive Cancer Center, Columbus, Ohio 43210, §HIV Drug Resistance Program and ‖Laboratory of Medicinal Chemistry, NCI-Frederick, Frederick, Maryland 21702, the ¶Department of Cellular and Molecular Biology, Tbilisi State University, Tbilisi 0128, Georgia, and **NIDDK, National Institutes of Health, Bethesda, Maryland 20892*

HIV-1 integrase (IN) is an important target for designing new antiviral therapies. Screening of potential inhibitors using recombinant IN-based assays has revealed a number of promising leads including nucleotide analogs such as pyridoxal 5′-phosphate (PLP). Certain PLP derivatives were shown to also exhibit antiviral activities in cell-based assays. To identify an inhibitory binding site of PLP to IN, we used the intrinsic chemical property of this compound to form a Schiff base with a primary amine in the protein at the nucleotide binding site. The amino acid affected was then revealed by mass spectrometric analysis of the proteolytic peptide fragments of IN. We found that an IC_{50} concentration (15 μM) of PLP modified a single IN residue, Lys^{244}, located in the C-terminal domain. In fact, we observed a correlation between interaction of PLP with Lys^{244} and the compound's ability to impair formation of the IN·DNA complex. Site-directed mutagenesis studies confirmed an essential role of Lys^{244} for catalytic activities of recombinant IN and viral replication. Molecular modeling revealed that Lys^{244} together with several other DNA binding residues provides a plausible pocket for a nucleotide inhibitor-binding site. To our knowledge, this is the first report indicating that a small molecule inhibitor can impair IN activity through its binding to the protein C terminus. At the same time, our findings highlight the importance of structural analysis of the full-length protein.

The continuing emergence of new HIV-1[1] variants resistant to current therapeutic treatments, which include small molecule inhibitors targeting HIV-1 reverse transcriptase and protease, makes the search for new anti-HIV-1 drugs imperative. In this regard, HIV-1 integrase (IN), which has no known human counterparts, is an attractive target. Furthermore, the fact that IN uses a common active site for 3′-processing and DNA strand transfer may constrain the range of mutations that can contribute to evolution of viable drug-resistant viruses.

HIV-1 IN catalyzes integration of the viral DNA, made by reverse transcription, into the host chromosome in a two-step reaction (reviewed in Ref. 1). In the first step, called 3′-processing, two nucleotides are removed at each 3′-end of the viral DNA. In the next step, called DNA strand transfer, concerted transesterification reactions integrate the viral DNA ends into the host genome.

HIV-1 IN is composed of three distinct structural and functional domains: the N-terminal domain (residues 1–50) that contains the HHCC zinc-binding motif, the core domain (residues 51–212) that contains the catalytic site, and the C-terminal domain (residues 213–270) that is thought to provide a platform for DNA binding. Crystallographic or NMR structural data are available for each of the individual domains (2–6). In addition, the structures of the combined core and C-terminal domains (7) and core and N-terminal domains (8) have been recently determined. However, efforts to obtain a structure of the full-length protein have been impeded by poor protein solubility.

Purified IN-based assays have been employed for screening of potential inhibitors. These studies have revealed several classes of compounds with anti-HIV-1 IN activity, including diketo acids, naphthyridines, pyranodipyrimidines, nucleotide analogs, hydroxylated aromatic compounds, DNA-interacting agents, peptides, and antibodies (9–22). Much effort has been devoted to dissecting the mechanism of inhibition and identifying the inhibitor binding sites. Crystallographic studies revealed two distinct binding sites in IN including the catalytic site for a diketo group-containing inhibitor (1-(5-chloroindol-3-yl)-3-hydroxy-3-(2H-tetrazol-5-yl)-propenone) and a site located near the IN dimer interface for 3,4-dihydroxyphenyltriphenylarsonium bromide (23, 24). However, these studies employed the isolated core domain of IN rather then the full-length protein.

Mass spectrometry is a powerful structural biology tool that allows the analysis of protein-inhibitor complexes under biologically relevant conditions and provides structural information complementary to NMR and crystallography. For example, we recently reported identification of a small molecule inhibitor binding site to full-length IN using the affinity acetylation and mass spectrometric analysis approach (25). These experiments

---

‡‡ To whom correspondence should be addressed: Ohio State University, 500 W. 12th Ave., Columbus, OH 43210. Tel.: 614-292-6091; Fax: 614-292-7766; E-mail: kvaratskhelia.1@osu.edu.

[1] The abbreviations used are: HIV-1, human immunodeficiency virus type 1; IN, integrase; PLP, pyridoxal 5′-phosphate; LTR, long terminal repeats; Q-ToF, quadrupole-time of flight; DTT, dithiothreitol; CHAPS, 3-[(3-cholamidopropyl)dimethylammonio]-1-propanesulfonic acid; IdU, 5-iododeoxyuracil; MOPS, 4-morpholinepropanesulfonic acid; MS, mass spectrometry; GFP, green fluorescent protein; PBGD, phorphobilinogen deminase gene.

were performed at low protein concentrations, where full-length IN as well as its complexes with inhibitor and DNA were fully soluble. In the present study, our efforts focused on interactions of a known nucleotide inhibitor, pyridoxal 5′-phosphate (PLP) with IN. Our interest in this compound was prompted for the following reasons. PLP has been previously shown to inhibit IN activity at low micromolar range (15, 16). More recently, certain PLP-derivatives were shown to exhibit potent antiviral activities in cell culture assays (26). In addition, PLP is known to form a Schiff base with the primary amine in the protein at the nucleotide-binding site (27). This intrinsic property coupled with mass spectrometric analysis could be exploited for determination of a nucleotide binding site in IN. Furthermore, protein interactions with the phosphate group of PLP could mimic that with the phosphate backbone of DNA. Finding the contact amino acids would benefit longstanding efforts in the field to define the DNA substrate binding cleft in HIV-1 IN. Indeed, our results reported here reveal a new plausible nucleotide-binding pocket in the protein. In particular, using mass spectrometry, we identified Lys[244] as a primary binding site of PLP. The importance of this amino acid for the IN function and for viral replication was confirmed by site-directed mutagenesis.

MATERIALS AND METHODS

Preparation of Recombinant Wild Type and K244E Mutant HIV-1 Integrase Proteins—pET-15b-IN1–288 (F185K/C280S) was expressed in *E. coli* strain BL21 as described previously (28). This plasmid was also used to introduce K244E mutation. For mutagenesis, we followed a Stratagene protocol using the QuikChange XL site-directed mutagenesis kit (Stratagene) and appropriate DNA oligonucleotides obtained from Integrated DNA Technologies (Coralville, IA). The presence of the desired mutation and the absence of undesired mutations were confirmed by DNA sequencing and further verified by mass spectrometric analysis of the recombinant protein. The proteins were purified by a nickel-nitrilotriacetic acid-Sepharose chromatography column as described previously (28). The fractions containing wild-type or K244E mutant protein were dialyzed against 50 mM HEPES, pH 7.5, 1 M NaCl, 100 μM $ZnCl_2$, 1 mM dithiothreitol (DTT), and 10% glycerol. Purified proteins were divided into separate aliquots and frozen in liquid nitrogen. The aliquots were thawed only once and used immediately.

IN Activity—3′-Processing and strand transfer activities were assayed in the reaction buffer containing 50 mM HEPES, pH 7.5, 50 mM NaCl, 10 mM $MnCl_2$, 2 mM β-mercaptoethanol, 1 mM CHAPS, 50 nM ^{32}P-labeled DNA substrate (21-mer blunt ended double-stranded DNA mimicking the U5 sequence), and 1 μM IN. The reaction products were separated by 20% denaturing PAGE and visualized by a Phosphor-Imager (Amersham Biosciences).

IN·DNA Photocross-linking—The protocol for specific photocross-linking of IN to cognate DNA substrate was adopted from Esposito and Craigie (29). HIV-1 U5 LTR sequence containing DNA oligonucleotide (5′-GTGTGGAAAATCTCTAGCAGT-3′) was annealed to the complementary DNA strand in which 5-iododeoxyuracil (IdU) was incorporated instead of the 5′-terminal adenine residue (5′-IdUCTGCTAGAGATTTTCCACAC-3′). The hybridized DNA substrates were radioactively labeled, purified through a Sephadex G25 desalting column, and used for cross-linking reactions. The reaction contained 50 mM HEPES, pH 7.8, 50 mM NaCl, 5 mM $MnCl_2$, 10 mM DTT, 20 nM DNA substrate, and 1 μM integrase. Samples were first incubated for 15 min at room temperature and then irradiated with 302-nm light from a handheld ultraviolet lamp for 30 min. The cross-linked products were analyzed by SDS-PAGE with NuPage 4–12% MOPS gels (Novex, Inc.). Quantitation of reaction products was carried out using a Phosphor-Imager and ImageQuant software (Amersham Biosciences).

Formation and Mass Spectrometric Analysis of the IN·Pyridoxal 5′-Phosphate Complex—The reaction mixture containing 50 mM HEPES, pH 7.8, 50 mM NaCl, 10 mM $MnCl_2$, 10 mM DTT, 1 μM IN was incubated with varying concentrations of PLP at 37 °C for 30 min. Sodium borohydride (5 mM final concentration) was added subsequently to reduce the Schiff base formed upon the PLP interaction with a primary amine. The protein was then unfolded by adding 40 mM DTT and incubating at 70 °C for 20 min. To modify cysteine residues, the reaction mixture was complemented with 100 mM iodoacetamide and incubated at 37 °C for 45 min. The reaction was quenched by the addition of 100 mM DTT. The modified protein samples were subjected to SDS-PAGE. The IN band was excised and subjected to in-gel proteolysis using 0.5 μg of one of the following three proteases: trypsin, LysC, or GluC (Roche Applied Science). Trypsin and LysC digestions were performed in 50 mM ammonium bicarbonate buffer, pH 8.0, at 37 °C for 16 h. GluC proteolysis was carried out in 50 mM ammonium acetate, pH 4.0, at 25 °C for 16 h. The proteolytic peptide fragments were desiccated in a SpeedVac (Thermo Savant, Holbrook, NY) and resuspended in an aqueous solution of 0.1% trifluoroacetic acid.

MS and MS/MS analyses of proteolytic peptide fragments were carried out using a Micromass Q-TOF-II instrument equipped with an electrospray source and a micromass cap-LC. Peptides were separated with a Waters Symmetry300 C18 precolumn and a Micro-Tech Scientific VC-10-C18–150 column using two sequential linear gradients of 5–40 and 40–90% acetonitrile for 35 and 10 min, respectively. MS/MS analysis data and the Mascot search engine (www.matrixscience.com) were used to identify IN peptide peaks from the NCBInr primary sequence data base.

Preparation and in Vivo Analysis of K244E Mutant—Together with the K244E mutation, we also analyzed HIV-1 NL4–3 viruses encoding wild type and D116N mutant IN as positive and negative controls, respectively. A 4.3-kb fragment of the NL4–3 vector spanning the SpeI and EcoRI sites was generated and subcloned into pBluescript SK(+) (Stratagene). Appropriate DNA oligonucleotides to introduce the K244E or D116N mutation were purchased from Integrated DNA Technologies (Coralville, IA). Preparation of mutant plasmids was carried out according to the Stratagene protocol using QuikChange XL site-directed mutagenesis kit (Stratagene). After confirming the mutation by DNA sequencing, the SpeI-EcoRI fragments were inserted back into the NL4–3 vector.

In vitro virus stock preparation and viral infectivity determination were carried out as follows. 293T cells (ATCC) were maintained in the presence of Dulbecco's modified Eagle's medium (Cellgro), 10% fetal calf serum (HyClone Laboratories), penicillin (50 units/ml; Invitrogen), and streptomycin (50 μg/ml; Invitrogen) and plated at a density of 5×10^6 cells/100-mm-diameter culture dish. After overnight incubation with 5% CO_2 at 37 °C, the cells were transfected using an MBS mammalian transfection kit (Stratagene) with 10 μg of HIV-1 NL4–3 plasmid DNA encoding wild-type and mutant (K244E or D116N) IN. Cell culture supernatant was replaced with fresh medium 6 h after transfection and further cultured for a total of 48 h. Thereafter, cell culture supernatants containing viruses were harvested, clarified through a 0.45-μm membrane filter (Nalgene), aliquoted, and stored at −80 °C. In a separate experiment, 293T cells were also co-transfected with 10 μg of plasmid DNAs together with 1 μg of plasmid DNA of pHCMV-G expressing vesiculostomatitis virus envelope glycoprotein G (30). The viral supernatant was cleared and concentrated 40-fold by centrifugation at 25,000 rpm for 90 min at 4 °C. The concentrated virus pellet was resuspended and digested in a buffer containing 50 mM Tris-HCl (pH 8.0), 10 mM $MgCl_2$, and 300 units/ml RNase-free DNase I (Roche Applied Science) for 60 min at room temperature. The viral p24 capsid protein amounts in the virus stocks were determined by using an enzyme-linked immunosorbent assay (p24 Core Profile Kit; DuPont). Human osteosarcoma cell line, GHOST (3) R3/X4/R5 (National Institutes of Health AIDS Research and Reference Reagent Program) transduced by HIV-2 LTR-GFP expressing both CD4 and CCR5 receptor were maintained in the same growth medium as 293T cells but with the additional selective reagents of G418 (500 μg/ml), hygromycin (1 μg/ml), and puromycin (1 μg/ml) (31). The cells were plated at a density of 1×10^5 cells in 0.5 ml of medium/well of a 24-well plate and cultured overnight under the conditions described above. Unconcentrated virus supernatants containing a range of 100–400 ng of p24 were added in duplicates and cultured for a total of 40 h. The cells were trypsinized off of the plate and inactivated by 1% paraformaldehyde in phosphate-buffered saline. A total of 10,000 events representing cells were acquired through flow cytometry to measure the percentage of GFP-positive cells. Thereafter, the percentage of GFP-positive cells was normalized to 100 ng of p24 equivalent viral inoculum.

Viral DNA Quantitation by Real Time PCR—CEMx174, the T and B lymphoid hybrid cell line (National Institutes of Health AIDS Research and Reference Reagent Program), was maintained in the presence of RPMI 1640 medium (Cellgro), 10% fetal calf serum (HyClone Laboratories), penicillin (50 units/ml; Invitrogen), and streptomycin (50 μg/ml; Invitrogen). 2×10^5 of CEMx174 cells in 0.5 ml of medium were infected by adding a 10-ng equivalent of p24 inoculum of vesiculostomatitis virus envelope glycoprotein G pseudotyped HIV-1 NL4-3 encoding wild-type and mutant (K244E or D116N) IN. Immediately and 12 h after infection, the cells were harvested, and total cellular DNA was ex-

FIG. 1. **Summary of the mass spectrometric data on proteolytic peptide fragments of the full-length IN.** The protein was subjected to proteolysis using trypsin, LysC, or GluC to generate small peptide fragments amenable to MS/MS analysis. The peptide ions were analyzed by the Q-TOF technique. *A*, a representative MS spectrum of IN tryptic peptides. *B*, peptide peak assignations with regard to the IN primary structure. The data show that application of all three proteases was essential to obtain complete coverage of the full-length IN amino acid sequence.
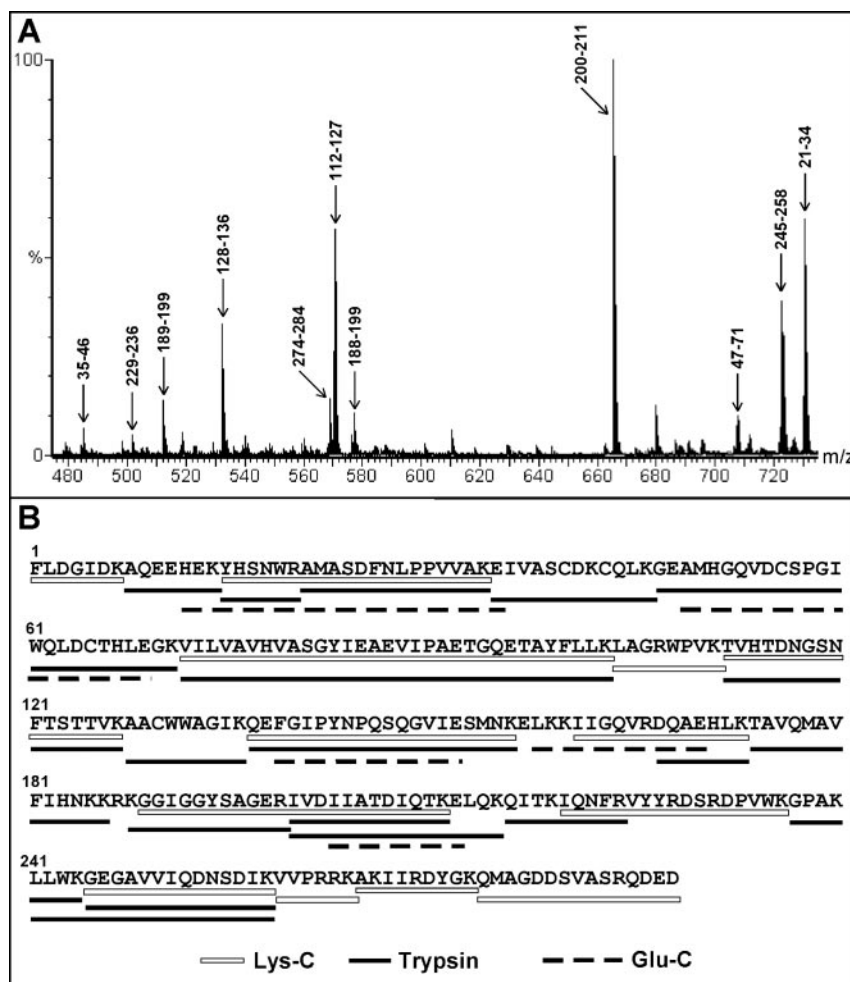
tracted (QIAmp DNA Blood Mini Kit; Qiagen) and dissolved in water. DNAs extracted from cells immediately after infection were used as a background control for the carryover plasmid DNA derived from the transfection procedure. The extracted DNAs were used in a real time PCR to quantitatively detect viral late reverse transcription product (U5-Ψ) and cellular DNA. The late reverse transcription product U5-Ψ DNA was quantified as described previously (32, 33). The following primer and probe sets were employed in the real time PCR assay to detect DNA products. For the late reverse transcription product U5-Ψ, the primer-probe sets were ES531 (5′-TGTGTGCCCGTCTGTTGTGT-3′), ES532 (5′-GAGTCCTGCGTCGAGAGATC-3′), and probe LRT-P (5′-6-carboxyfluorscein (FAM)-CAGTGGCGCCCGAACAGGGA-carboxytetramethylrhodamine (TAMRA)-3′) (32). The *PBGD* primer-probe sets consisted of the forward primer DF3 (5′-AGGGCAGGAACCAGG-GATTATG-3′), the reverse primer DR1 (5′-GGGCACCACACTCTCCT-ATCTTT-3′), and the probe *PBGD* (5′-ATGTCCACCACAGGGGACAA-GATTC-3′). All real time PCRs were performed using the TaqMan Universal PCR Master Mix Kit (Roche Applied Science), and the products were quantified using the ABI Prism sequence detector 7700. Primers were used at 600 nM, and probes were used at 75 nM concentration. The conditions used for real time PCRs were as follows: one cycle at 50 °C for 2 min, one cycle at 95 °C for 10 min, and 40 cycles of amplification at 95 °C for 0.15 min and 60 °C for 1 min.

Molecular Modeling—In order to better understand the binding mode and interactions of PLP with HIV-1 IN, a structural model of PLP complexed with full-length HIV-1 IN dimer was generated. Based on our experimental results and reported crystal structures of PLP complexed with different proteins in the protein data bank (34), the IN·PLP complex was modeled as a covalent interaction between Lys^{244} and PLP by adding a bond between terminal NH_2 of Lys^{244} and aldehyde carbonyl of PLP. The complex was then subjected to quenched molecular dynamics simulations using the Insight II 2000.0/Discover 97.0 modeling package (Molecular Simulations Inc., San Diego, CA), using the cff91 force field. Additional parameters for Mg^{2+} were determined using its ionic radius and small molecule crystallographic data. A dielectric

constant of 1.00 and the cell multipole method with fine accuracy were used for all nonbonding interactions except for the quenched molecular dynamics. The quenched molecular dynamics cycle consisted of 100 ps of molecular dynamics simulation at 800 K in an NVT ensemble, keeping all protein atoms beyond 15 Å of PLP fixed. The van der Waals interactions were scaled to 2%, and the coulombic interactions were scaled to 20%. The coordinates were saved every 200 fs and subsequently minimized by 300 steps of the Polak-Ribiere conjugate gradient algorithm. The lowest energy conformation of the 500 minimized structures is depicted in Fig. 7A. The electrostatic potential on the solvent-accessible surface of the full-length IN dimer and PLP was calculated using the Delphi module in the Insight II program and is depicted in Fig. 7B. The default Delphi formal charges were applied for the protein, whereas *ab initio* calculation with Gaussian 98 A11 was used to obtain charges for PLP.

RESULTS

To map the PLP binding site of HIV-1 IN, we exploited an intrinsic property of PLP to form a Schiff base with primary amines. The compound can be tethered covalently to the adjacent lysine by subsequent reduction of a Schiff base with sodium borohydride. Of note, lysine is the most abundant residue in HIV-1 IN, providing an excellent environment for detailed analysis of an inhibitor-binding site. Proteolytic hydrolysis coupled with mass spectrometric analysis yielded complete coverage of the full-length protein amino acid sequence, enabling us to monitor all of the lysine residues present in HIV-1 IN (Fig. 1).

To identify an inhibitory PLP binding site we examined IN interaction with a 15 μM concentration of the compound. This concentration was chosen because under our reaction conditions PLP inhibited 3′-processing and strand transfer activities
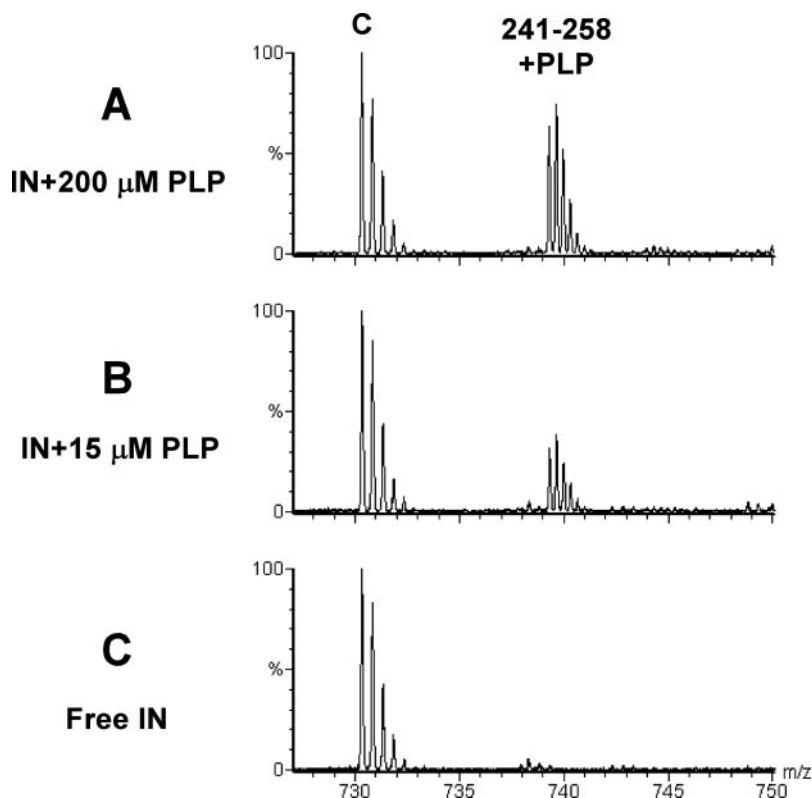
FIG. 2. **A segment of the Q-TOF spectra to show PLP binding to Lys[244].** Mass spectrometric analysis of PLP binding to HIV-1 IN. *A*, IN + 200 μM PLP; *B*, IN + 15 μM PLP; *C*, free IN. A single new peak (739.32^{3+}) was observed when tryptic fragments of the IN complex with 15 μM PLP (*B*) were compared with that of free IN (*C*). The molecular mass of this peptide (2,215 Da) corresponds to that of the peptide 241–258 (1,984 Da) plus one PLP modification (231 Da). *C*, unmodified tryptic peptide of HIV-1 IN.

with a comparable IC_{50} value of ∼15 μM. Interestingly, mass spectrometric analysis of tryptic peptide fragments of the IN·PLP complex revealed that a single new peak was formed when compared with free IN digests (Fig. 2). The molecular weight of this peak corresponded to the molecular weight of IN peptide amino acids 241–258 plus one molecule of PLP. This peptide contains two lysine residues, Lys[244] and Lys[258]. Lys[244], located within the peptide, was resistant to tryptic digestion. In contrast, Lys[258], located at the peptide C terminus, was readily hydrolyzed by trypsin. Normally, unmodified lysine residues in proteins are easily cleaved by trypsin, whereas modified lysines become resistant to proteolysis.

Conclusive evidence for Lys[244] modification by PLP emerged from MS/MS analysis (Fig. 3). Indeed, series of y ions (y2, y4–y9) indicated that Lys[258] did not contain any modified groups, whereas detection of ions b9, b10, and y17 enabled us to localize the modification site to Lys[244]. This lysine was the only modified residue observed in the presence of 15 μM PLP. Several additional surface-exposed lysines were modified when the PLP concentration was increased to 200 μM. However, detailed quantitative analysis of the modified peaks showed that even at such a high concentration (200 μM) of PLP, the primary site of modification remained Lys[244].

Lys[244] is located in the C-terminal domain of the enzyme. This domain is thought to be important for coordinating DNA and is significantly distanced from the catalytic site of IN. Therefore, we next examined whether modification of Lys[244] could interfere with formation of the IN·DNA complex. For this, we chose to employ the specific DNA substrate containing IdU at the 5′-end of the HIV-1 U5 LTR sequence. This particular DNA has been previously shown to cross-link with Gln[148] located in the core domain near the catalytic site of HIV IN (35). Thus, using this setup, we could monitor the specific DNA binding to IN without PLP directly interfering with the cross-linking site or chemistry. The results depicted in Fig. 4 show that PLP impairs DNA binding with an IC_{50} value of about 12 μM. In fact, there is a correlation between modification of

Lys[244] with PLP (Fig. 2), inhibition of the IN catalytic activities, and specific IN·DNA binding (Fig. 4).

To confirm an essential role of Lys[244] for the HIV-1 IN function, we performed *in vivo* and *in vitro* analyses of the K244E mutation. The results of *in vivo* analysis are depicted in Fig. 5. A reporter cell line Ghost-R3/X5/R5 was infected with the wild type and mutant (K244E or D116N) viruses. This cell line expresses HIV-1 receptors and was also transduced by HIV-2 LTR-GFP. Once the cells are infected by HIV-1, the integrated viral genome will express Tat protein that will subsequently transactivate LTR-driven GFP expression. Therefore, the percentage of GFP-positive cells provides a measure of the infectivity of a virus. When normalized at a 100-ng equivalent of p24 inoculum, HIV-1 NL4–3 wild type virus infection induced 74.4% of the cells to become GFP-positive, whereas only 0.8 and 1.0% of the cells were GFP-positive for the K244E and D116N mutants, respectively (Fig. 5*A*). This low percentage of GFP-positive cells is likely to be the background of the assay. It has been reported that up to 1.0% of GFP-positive cells might be induced by Tat protein expressed from the unintegrated LTR promoter (36).

The lack of infectivity of the K244E mutant could be attributed to defects at any steps of the viral life cycle including the steps preceding the integration of the viral genome. To address this question, a U5-Ψ viral DNA product derived from the late stage of viral reverse transcription was quantified by using a real time PCR assay. The data in Fig. 5*B* indicate that the wild-type as well as mutants D116N and K244E produced similar amounts of viral late reverse transcription products. These results indicated that reverse transcription was not affected by the K244E mutation.

Comparison of two-LTR circle copy numbers standardized over the late reverse transcription products revealed 2.29- and 5.11-fold increases for K244E and D116N mutants, respectively, over wild type NL4–3. A similar picture was observed previously (35) with other DNA binding mutations (K156E and K159E). Increases in two-LTR circles were relatively modest

FIG. 3. **MS/MS data of the peptide fragment (amino acids 241–258) to indicate modification of Lys[244] with PLP.** The b and y ion peaks are derived from internal fragmentation of peptide bonds and provide amino acid sequence information from the peptide N terminus to C terminus and from the C terminus to N terminus, respectively.
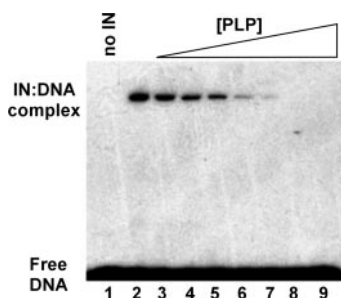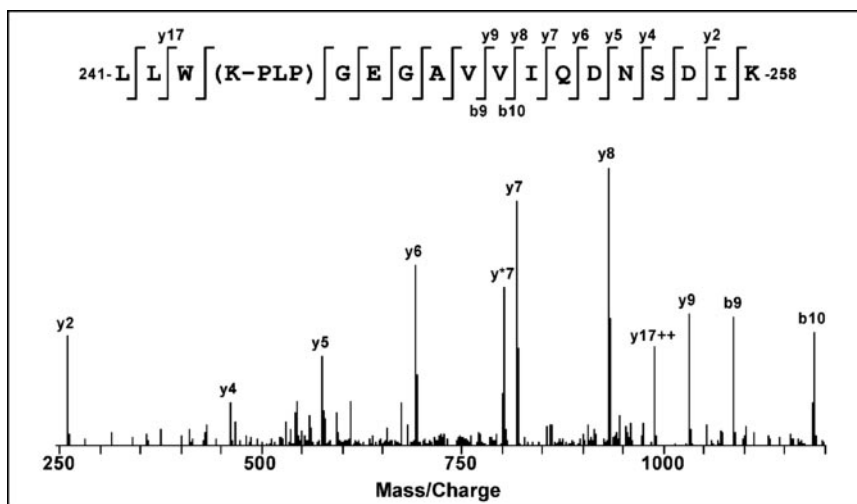

FIG. 4. **Effect of PLP on formation of the IN·DNA complex.** Radiolabeled specific 20-bp DNA containing 5-iododeoxyuracil at the 5′-end of HIV-1 U5 LTR sequence was cross-linked to HIV-1 IN using 302-nm light. The cross-linked products were analyzed by SDS-PAGE. *Lane 1* did not contain HIV-1 IN. *Lane 2* contained DNA and IN in the absence of PLP. Other *lanes* contained DNA, IN, and the following concentrations of PLP: 6.25 μM (*lane 3*), 12.5 μM (*lane 4*), 25 μM (*lane 5*), 50 μM (*lane 6*), 100 μM (*lane 7*), 200 μM (*lane 8*), and 400 μM (*lane 9*).

for these two mutants in comparison with the increase seen with D116N (35). This pattern could possibly be explained by the notion that formation of the IN·DNA complex that takes place in the cytoplasm of the infected cells may be essential for the efficient nuclear import. Unlike the DNA binding mutations, D116N exhibits wild type levels of DNA binding and is inactive due to its failure to coordinate divalent metal ions. Collectively, the data in Fig. 5 suggest that the K244E mutation is detrimental for the integration step and does not impact reverse transcription.

We next performed *in vitro* analysis of the K244E mutant. The data depicted in Fig. 6 indicate that the mutation fully impaired 3′-processing and strand transfer activities. Taken together, *in vivo* and *in vitro* data argue strongly for the importance of Lys[244] for IN function.

Our experimental results were employed to create a model of the IN·PLP complex. The modeling studies revealed a plausible nucleotide analog binding pocket in the HIV-1 IN (Fig. 7). In the lowest energy frame obtained from the molecular dynamics simulation, the phosphate group of PLP makes hydrogen bonding interactions with Arg[262], Lys[264], and Lys[266], whereas the hydroxy group at position 3 of PLP is hydrogen-bonded to D229 (Fig. 7A). Of note, the residues Arg[262], Lys[264], and Lys[266] were implicated in viral DNA binding (37, 38). A high degree of complementarity was observed in the electrostatic potentials of PLP and the IN residues surrounding it in this binding orientation (Fig. 7B).

Importantly, Lys[244] as well as the other PLP-interacting residues (Arg[262], Lys[264], and Lys[266]) are conserved in HIV-1,
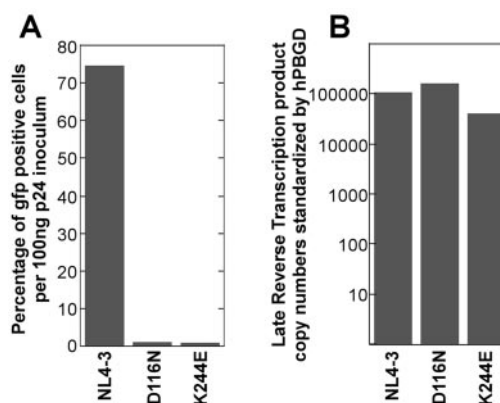

FIG. 5. *In vivo* **analysis of K244E mutation.** *A*, effects of the mutation on infectivity. Human osteosarcoma cell line, GHOST (3) R3/X4/R5 transduced by HIV-1 LRT-GFP were infected with HIV-1 NL4–3 encoding wild-type or mutant (K244E and D116N) IN, using 100-, 200-, and 400-ng equivalents of p24 inoculum. The percentage of GFP-positive cells induced by infection of a 100-ng equivalent of p24 inoculum was presented for all three viruses. One representative set of results is presented from two independent experiments, which generated similar results. *B*, comparison of HIV-1 late reverse transcription product (U5-Ψ) by quantitative real time PCR assay. A human T/B cell hybrid cell line CEMx174 was infected with HIV-1 NL4–3 encoding wild-type and mutant (K244E or D116N) IN using 10 ng of p24 equivalent inoculum, which has been digested by RNase-free DNase I to eliminate carryover DNA from transfection. Immediately and 12 h after infection, DNAs were purified and used to quantify viral late reverse transcription products (U5-Ψ) and cellular *PBGD* gene. The copy numbers of viral late reverse transcription product (U5-Ψ) detected at 12 h postinfection were normalized by subtracting carryover DNA copy numbers detected in samples harvested immediately after infection and also standardized by input cellular *PBGD* gene.

HIV-2, and SIV IN proteins. The structural fold of the HIV-1 IN C terminus is similar to that of other DNA-binding proteins exhibiting an Src homology 3-like fold (Sso7d and Sac7d) (7, 39–41). However, HIV-1 IN residues Lys[244], Arg[262], Lys[264], and Lys[266] are not conserved in Sso7d and Sac7d. Therefore, a nucleotide inhibitor binding site uncovered in the present work is of interest for designing small molecule antagonists selective to retroviral IN.

DISCUSSION

Our studies reveal a new plausible nucleotide inhibitor binding site in HIV-1 IN. We show a correlation between PLP modification of Lys[244] and inhibition of IN catalytic activities as well as formation of the IN·DNA complex. Our *in vitro* and *in vivo* site-directed mutagenesis analyses confirmed the importance of Lys[244] for IN function.
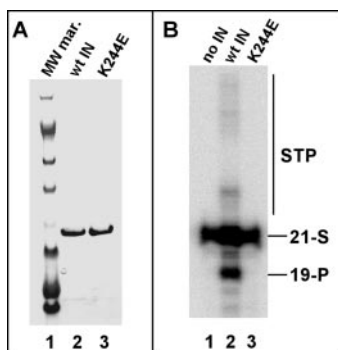
FIG. 6. *In vitro* analysis of the K244E protein. *A*, SDS-PAGE of purified proteins. *Lane 1*, MW markers; *lane 2*, wild type IN; *lane 3*, K244E. *B*, effects of the mutation on IN catalytic activities. Migration positions of 21-mer substrate (*21-S*), 19-mer product of the 3′-processing reaction (*19-P*), and products of the strand transfer reaction (*STP*) are indicated. *Lane 1* did not contain any protein. *Lanes 2* and *3* contained wild type and K244E mutant IN proteins, respectively.
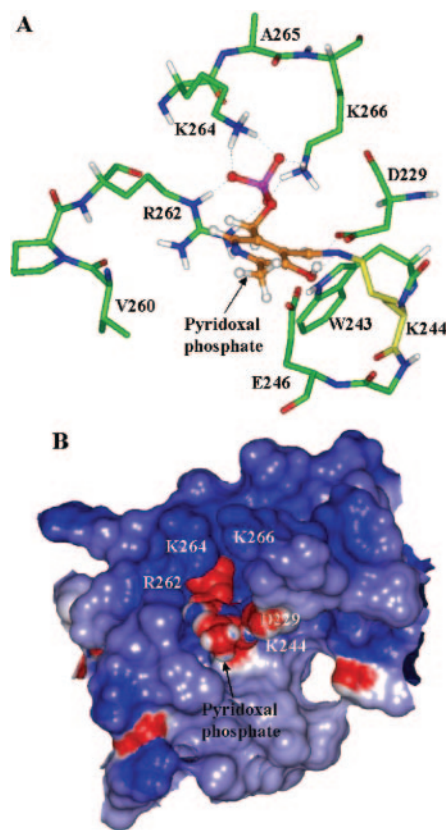

FIG. 7. **Hypothetical binding mode of PLP to HIV-1 IN.** *A*, the lowest energy frame from the molecular dynamics simulation. For clarity, only important residues and atoms are shown. Carbon atoms of Lys[244] are *colored* in *yellow*, and hydrogen bonds are depicted in *black* as *dashed lines*. PLP is covalently bonded with Lys[244]; the phosphate group engages in hydrogen bonding with Arg[262], Lys[264], and Lys[266]; the 3′-OH group hydrogen-bonds to Asp[229]. *B*, electrostatic potentials shown on the solvent-accessible surface of the IN·PLP complex. The surface charge distributions are *colored* as follows: *blue*, positive (6.0 kT); *red*, negative (-6.0 kT); *white*, neutral (where k represents the Boltzmann constant and T is the temperature).

To our knowledge, this is the first report indicating that a small molecule inhibitor can impair HIV-1 IN activity by binding to the C-terminal domain. A large molecule inhibitor, monoclonal antibody 33, was reported to bind to the C-terminal domain and impair IN coordination of the cognate DNA (17). Crystallographic and NMR efforts to determine small molecule inhibitor binding sites were restricted to analyses of the protein

core domain rather than full-length IN (23, 24). Recently, using full-length IN and mass spectrometric footprinting, we mapped an acetylated inhibitor (methyl-*N*,*O*-bis-(3,4-diacetoxycinnamoyl)serinate) binding site to Lys[173], which is located in the core domain (25). This inhibitor-binding site involves an architecturally critical protein dimer interface that is significantly distanced from the DNA binding cleft. Consistent with these structural data, the mechanistic studies indicated that the acetylated inhibitor does not interfere with formation of the IN·DNA complex.

Previous attempts to map a nucleotide analog binding site in IN included application of photoaffinity labeling with 3′-azido-3′-dideoxythymine analog 3′,5-diazido-2′,3′-dideoxyuridine 5′-monophosphate coupled with proteolytic digestion of the protein (42). Drake *et al.* (42) identified amino acid region 153–167 of IN as the site of photocross-linking. In this cross-linking experiment, the core domain of IN (amino acids 50–212) rather than the full length protein was used (42). In the present study with full-length IN, we found that PLP could also bind to the similar region by directly interacting with Lys[156]. However, the modification of Lys[156] could only be detected at an elevated concentration (200 μM) and not at 15 μM PLP (data not shown). Equally, it should be noted that the IC_{50} value for inhibition of IN with 3′-azido-3′-dideoxythymine was reported to be 360 μM (42). Our results indicate that Lys[244] binds PLP with a significantly higher affinity than Lys[156].

Lys[244] was accessible to PLP modification when the compound was introduced to the preassembled IN·DNA complex. Equally, we observed similar inhibition profiles whether PLP was added first to free IN and then DNA substrate was provided or the compound was exposed to the preassembled IN·DNA complex (data not shown). Our interpretation of this observation is as follows. The active IN protein is an oligomer where separate monomers appear to provide complementary rather than symmetrical contacts to the DNA substrate. Indeed, previous *in vitro* complementation assays have shown that two inactive IN proteins, one with the mutation in the core domain and another with the mutation in the C-terminal domain, can be combined to restore catalytic activity (43, 44). Thus, when an active site lysine in one monomer contacts the DNA substrate, the identical lysine in another monomer is exposed to solvent. Binding of PLP to the IN monomer in which Lys[244] is accessible to the surface could destabilize the IN·DNA structure and compromise the enzymatic activities.

Our results suggest that Lys[244] is an essential DNA binding residue. A study of Gao *et al.* (45) demonstrated that the E246C mutant preferentially interacted with position 7 of the U5 sequence. In particular, the disulfide cross-link between the cysteine residue and modified base in DNA was established through a 3-carbon linker arm. In our model, Glu[246] is located within 5 Å of PLP (Fig. 7*A*). Therefore, it is logical to propose that Lys[244] establishes direct contacts to DNA that could contribute to the observed cross-link of E246C with the respective DNA.

Our modeling studies revealed that Lys[244] together with other DNA binding residues (Arg[262], Lys[264], and Lys[266]) forms a plausible nucleotide inhibitor-binding pocket. The poor membrane permeability and reduced specificity restricts application of PLP as a potent antiviral inhibitor. However, certain PLP derivatives have shown promise as potent antiviral agents in cell culture assays (26). The structural information we report here facilitates a more rational approach toward improving selectivity and potency of this class of HIV-1 IN inhibitors.

REFERENCES

1. Brown, P. O. (1997) in *Retroviruses* (Coffin, J. M., Hughes, S. H., and Varmus, H. E., eds) pp. 161–204, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY

2. Cai, M., Zheng, R., Caffrey, M., Craigie, R., Clore, G. M., and Gronenborn, A. M. (1997) *Nat. Struct. Biol.* **4,** 567–577
3. Dyda, F., Hickman, A. B., Jenkins, T. M., Engelman, A., Craigie, R., and Davies, D. R. (1994) *Science* **266,** 1981–1986
4. 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
5. Eijkelenboom, A. P., Lutzke, R. A., Boelens, R., Plasterk, R. H., Kaptein, R., and Hard, K. (1995) *Nat. Struct. Biol.* **2,** 807–810
6. 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
7. 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
8. Wang, J. Y., Ling, H., Yang, W., and Craigie, R. (2001) *EMBO J.* **20,** 7333–7343
9. Nicklaus, M. C., Neamati, N., Hong, H., Mazumder, A., Sunder, S., Chen, J., Milne, G. W., and Pommier, Y. (1997) *J. Med. Chem.* **40,** 920–929
10. Maroun, R. G., Krebs, D., Roshani, M., Porumb, H., Auclair, C., Troalen, F., and Fermandjian, S. (1999) *Eur. J. Biochem.* **260,** 145–155
11. Zhang, X., Neamati, N., Lee, Y. K., Orr, A., Brown, R. D., Whitaker, N., Pommier, Y., and Burke, T. R., Jr. (2001) *Bioorg. Med. Chem.* **9,** 1649–1657
12. Zhao, L., O'Reilly, M. K., Shultz, M. D., and Chmielewski, J. (2003) *Bioorg. Med. Chem. Lett.* **13,** 1175–1177
13. Pommier, Y., Marchand, C., and Neamati, N. (2000) *Antiviral Res.* **47,** 139–148
14. De Clercq, E. (2001) *Curr. Med. Chem.* **8,** 1543–1572
15. Lipford, J. R., Worland, S. T., and Farnet, C. M. (1994) *J. Acquired Immune Defic. Syndr.* **7,** 1215–1223
16. Mazumder, A., Neamati, N., Sommadossi, J. P., Gosselin, G., Schinazi, R. F., Imbach, J. L., and Pommier, Y. (1996) *Mol. Pharmacol.* **49,** 621–628
17. Yi, J., Cheng, H., Andrake, M. D., Dunbrack, R. L., Jr., Roder, H., and Skalka, A. M. (2002) *J. Biol. Chem.* **277,** 12164–12174
18. Pannecouque, C., Pluymers, W., Van Maele, B., Tetz, V., Cherepanov, P., De Clercq, E., Witvrouw, M., and Debyser, Z. (2002) *Curr. Biol.* **12,** 1169–1177
19. Hazuda, D. J., Felock, P., Witmer, M., Wolfe, A., Stillmock, K., Grobler, J. A., Espeseth, A., Gabryelski, L., Schleif, W., Blau, C., and Miller, M. D. (2000) *Science* **287,** 646–650
20. Espeseth, A. S., Felock, P., Wolfe, A., Witmer, M., Grobler, J., Anthony, N., Egbertson, M., Melamed, J. Y., Young, S., Hamill, T., Cole, J. L., and Hazuda, D. J. (2000) *Proc. Natl. Acad. Sci. U. S. A.* **97,** 11244–11249
21. Grobler, J. A., Stillmock, K., Hu, B., Witmer, M., Felock, P., Espeseth, A. S., Wolfe, A., Egbertson, M., Bourgeois, M., Melamed, J., Wai, J. S., Young, S., Vacca, J., and Hazuda, D. J. (2002) *Proc. Natl. Acad. Sci. U. S. A.* **99,** 6661–6666
22. Hazuda, D. J., Anthony, N. J., Gomez, R. P., Jolly, S. M., Wai, J. S., Zhuang, L., Fisher, T. E., Embrey, M., Guare, J. P., Jr., Egbertson, M. S., Vacca, J. P., Huff, J. R., Felock, P. J., Witmer, M. V., Stillmock, K. A., Danovich, R., Grobler, J., Miller, M. D., Espeseth, A. S., Jin, L., Chen, I. W., Lin, J. H., Kassahun, K., Ellis, J. D., Wong, B. K., Xu, W., Pearson, P. G., Schleif, W. A., Cortese, R., Emini, E., Summa, V., Holloway, M. K., and Young, S. D. (2004) *Proc. Natl. Acad. Sci. U. S. A.* **101,** 11233–11238
23. Goldgur, Y., Craigie, R., Cohen, G. H., Fujiwara, T., Yoshinaga, T., Fujishita, T., Sugimoto, H., Endo, T., Murai, H., and Davies, D. R. (1999) *Proc. Natl. Acad. Sci. U. S. A.* **96,** 13040–13043
24. Molteni, V., Greenwald, J., Rhodes, D., Hwang, Y., Kwiatkowski, W., Bushman, F. D., Siegel, J. S., and Choe, S. (2001) *Acta Crystallogr. Sect. D. Biol. Crystallogr.* **57,** 536–544
25. Shkriabai, N., Patil, S. S., Hess, S., Budihas, 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
26. Sauve, G., and Stranix, B. R. (October 28, 2003) U. S. Patent 6,638,921
27. Lundblad, R. L. (1995) *Techniques in Protein Modification*, CRC Press, Inc., Boca Raton, FL
28. Jenkins, T. M., Engelman, A., Ghirlando, R., and Craigie, R. (1996) *J. Biol. Chem.* **271,** 7712–7718
29. Esposito, D., and Craigie, R. (1998) *EMBO J.* **17,** 5832–5843
30. Yee, J. K., Miyanohara, A., LaPorte, P., Bouic, K., Burns, J. C., and Friedmann, T. (1994) *Proc. Natl. Acad. Sci. U. S. A.* **91,** 9564–9568
31. Morner, A., Bjorndal, A., Albert, J., Kewalramani, V. N., Littman, D. R., Inoue, R., Thorstensson, R., Fenyo, E. M., and Bjorling, E. (1999) *J. Virol.* **73,** 2343–2349
32. Butler, S. L., Hansen, M. S., and Bushman, F. D. (2001) *Nat. Med.* **7,** 631–634
33. 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
34. Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H., Shindyalov, I. N., and Bourne, P. E. (2000) *Nucleic Acids Res.* **28,** 235–242
35. Jenkins, T. M., Esposito, D., Engelman, A., and Craigie, R. (1997) *EMBO J.* **16,** 6849–6859
36. Wu, Y., and Marsh, J. W. (2001) *Science* **293,** 1503–1506
37. Heuer, T. S., and Brown, P. O. (1998) *Biochemistry* **37,** 6667–6678
38. Heuer, T. S., and Brown, P. O. (1997) *Biochemistry* **36,** 10655–10665
39. Robinson, H., Gao, Y. G., McCrary, B. S., Edmondson, S. P., Shriver, J. W., and Wang, A. H. (1998) *Nature* **392,** 202–205
40. Krueger, J. K., McCrary, B. S., Wang, A. H., Shriver, J. W., Trewhella, J., and Edmondson, S. P. (1999) *Biochemistry* **38,** 10247–10255
41. Gao, Y. G., Su, S. Y., Robinson, H., Padmanabhan, S., Lim, L., McCrary, B. S., Edmondson, S. P., Shriver, J. W., and Wang, A. H. (1998) *Nat. Struct. Biol.* **5,** 782–786
42. Drake, R. R., Neamati, N., Hong, H., Pilon, A. A., Sunthankar, P., Hume, S. D., Milne, G. W., and Pommier, Y. (1998) *Proc. Natl. Acad. Sci. U. S. A.* **95,** 4170–4175
43. Engelman, A., Bushman, F. D., and Craigie, R. (1993) *EMBO J.* **12,** 3269–3275
44. van Gent, D. C., Vink, C., Groeneger, A. A., and Plasterk, R. H. (1993) *EMBO J.* **12,** 3261–3267
45. Gao, K., Butler, S. L., and Bushman, F. (2001) *EMBO J.* **20,** 3565–3576