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

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 IC₅₀ concentration (15 μM) of PLP modified a single IN residue, Lys²⁴⁴, located in the C-terminal domain. In fact, we observed a correlation between interaction of PLP with Lys²⁴⁴ and the compound's ability to impair formation of the IN-DNA complex. Site-directed mutagenesis studies confirmed an essential role of Lys²⁴⁴ for catalytic activities of recombinant IN and viral replication. Molecular modeling revealed that Lys²⁴⁴ 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 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)-propene) and a site located near the IN dimer interface for 3,4-dihydroxyphenyltriptolylphenylarsonium bromide (23, 24). However, these studies employed the isolated core domain of IN rather than 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...
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 Lys243 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-IN–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 QuickChange XL site-directed mutagenesis kit (Stratagene) and appropriate DNA oligonucleotides obtained from Integrated DNA Technologies (Corvalle, IA). The presence of the desired mutation and the absence of undesired mutations were confirmed by DNA sequencing. A 20-μg sample of DNA elugene from the K244E or D116N mutation were purchased from Integrated DNA Technologies (Corvalle, IA). Preparation of mutant plasmids was carried out according to the Stratagene protocol using QuickChange 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 Dulbeco’s modified Eagle’s medium (Cellgro), 10% fetal calf serum (HyClone), and streptomycin (50 μg/ml; Invitrogen), and incubated at 37 °C, 5% CO2. The cells were transfected using a MBS mammalian transfection reagent (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 DNA harboring a K244E mutation for 24 h and infected with 100 TCID50 of HIV-1 NL4–3 virus stock 48 h after transfection. Viral titer was determined by an enzyme-linked immunosorbent assay (p24 Core Profile Kit; DuPont). Human osteosarcoma cell line, GHOST (3) R3X4/R5 (National Institutes of Health AIDS Research and Reference Reagent Program) was 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 (300 μg/ml), hygromycin (1 μg/ml), and puromycin (1 μg/ml) (31). The cells were plated at a density of 1 × 105 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. The percentage of GFP-positive cells was normalized to 100 ng of p24 equivalent viral in 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 RPMI 1640 medium (Cellgro), 10% fetal calf serum (HyClone Laboratories), penicillin (50 units/ml; Invitrogen), and streptomycin (50 μg/ml; Invitrogen). 2 × 105 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-
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 Lys244 and PLP. Of note, lysine is the most abundant residue among the lysine residues present 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 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).

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 mM concentration of the compound. This concentration was chosen because under our reaction conditions PLP inhibited 3'-processing and strand transfer activities.
with a comparable IC$_{50}$ value of $-15 \mu M$. 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 $\mu M$ PLP. Several additional surface-exposed lysines were modified when the PLP concentration was increased to 200 $\mu M$. However, detailed quantitative analysis of the modified peaks showed that even at such a high concentration (200 $\mu M$) 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 dIdU 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 $\mu M$. 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 analyses 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. 5A). 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-‘P 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. 5B 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.
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 Lys244 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 Arg262, Lys264, and Lys266, whereas the hydroxy group at position 3 of PLP is hydrogen-bonded to D229 (Fig. 7A). Of note, the residues Arg262, Lys264, and Lys266 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, Lys244 as well as the other PLP-interacting residues (Arg262, Lys264, and Lys266) are conserved in HIV-1, 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 Lys244, Arg262, Lys264, and Lys266 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 Lys244 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 Lys244 for IN function.
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\textsubscript{2}O-bis(3,4-diaceetoxyxynamoyl)serinate) binding site to Lys\textsuperscript{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'-dideoxythymidine analog 3',5-diazido-2',3'-dideoxyxuridine 5'-monophosphate coupled with proteolytic digestion of the protein (42). Drake \textit{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\textsuperscript{156}. However, the modification of Lys\textsuperscript{156} could only be detected at an elevated concentration (200 \textmu M) and not at 15 \textmu M PLP (data not shown).

Equally, it should be noted that the IC\textsubscript{50} value for inhibition of IN with 3'-azido-3'-dideoxythymidine was reported to be 360 \textmu M (42). Our results indicate that Lys\textsuperscript{244} binds PLP with a significantly higher affinity than Lys\textsuperscript{156}.

Lys\textsuperscript{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 \textit{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\textsuperscript{244} is accessible to the surface could destabilize the IN-DNA structure and compromise the enzymatic activities.

Our results suggest that Lys\textsuperscript{244} is an essential DNA binding residue. A study of Gao \textit{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\textsuperscript{246} is located within 5 Å of PLP (Fig. 7A). Therefore, it is logical to propose that Lys\textsuperscript{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\textsuperscript{244} together with other DNA binding residues (Arg\textsuperscript{262}, Lys\textsuperscript{264}, and Lys\textsuperscript{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 \textit{Retroviruses} (Coffin, J. M., Hughes, S. H., and Varmus, H. E., ed.) pp. 161–204, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY
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

J. Biol. Chem. 2005, 280:7949-7955.
doi: 10.1074/jbc.M413579200 originally published online December 22, 2004

Access the most updated version of this article at doi: 10.1074/jbc.M413579200

Alerts:
- When this article is cited
- When a correction for this article is posted

Click here to choose from all of JBC's e-mail alerts

This article cites 42 references, 21 of which can be accessed free at http://www.jbc.org/content/280/9/7949.full.html#ref-list-1