Molecular Pharmacology Fast Forward. Published on July 28, 2009 as doi: 10.1124/mol.109.058883
This article has not been copyedited and formatted. The final version may differ from this version.

MOL#58883

An allosteric mechanism for inhibiting HIV-1 integrase with a small molecule

Jacques J. Kessl, Jocelyn O. Eidahl, Nikolozi Shkriabai, Zhuojun Zhao[1], Christopher J. McKee, Sonja Hess, Terrence R. Burke, Jr., Mamuka Kvaratskhelia

Center for Retrovirus Research and Comprehensive Cancer Center, College of Pharmacy, The Ohio State University, Columbus, OH (J.J.K, J.O.E., N.S., Z.Z., C.J.M., M.K.); Proteome Exploration Laboratory, Beckman Institute, California Institute of Technology, CA (S.H.); and Laboratory of Medicinal Chemistry, Center for Cancer Research, National Cancer Institute – Frederick, Frederick, MD (T.R.B.).

Molecular Pharmacology Fast Forward. Published on July 28, 2009 as DOI: 10.1124/mol.109.058883
This article has not been copyedited and formatted. The final version may differ from this version.

Running title: **HIV-1 integrase inhibitors**

Address correspondence to: Mamuka Kvaratskhelia, The Ohio State University, 500 West 12th Avenue, Room 238, L.M. Parks Hall, Columbus, OH 43210; Phone: 614-292-6091; Fax: 614-292-7766; E-mail: Kvaratskhelia.1@osu.edu

Number of text pages: 23

Number of tables: 1

Figures: 7

References: 39

Number of words in Abstract: 239

Number of words in Introduction: 705

Number of words in Discussion: 1,294

List of abbreviations: IN, integrase; CCD, catalytic core domain; NTD, N-terminal domain; PIC, preintegration complex; LEDGF, lens epithelium-derived growth factor; IBD, integrase binding domain; BSA, bovine serum albumin; MS, mass spectrometry; MALDI-ToF, matrix-assisted laser desorption ionization - time of flight; Q-ToF, quadrupole - time of flight; CA, chicoric acid.

Molecular Pharmacology Fast Forward. Published on July 28, 2009 as DOI: 10.1124/mol.109.058883
This article has not been copyedited and formatted. The final version may differ from this version.

Abstract

HIV-1 integrase (IN) is a validated target for developing anti-retroviral inhibitors. Using affinity acetylation and mass spectrometric (MS) analysis we previously identified a tetraacetylated inhibitor (**1**) that selectively modified Lys173 at the IN dimer interface. Here we extend our efforts to dissect the mechanism of inhibition and structural features that are important for the selective binding of **1**. Using a subunit exchange assay we have found that the inhibitor strongly modulates dynamic interactions between IN subunits. Restricting such interactions does not directly interfere with IN binding to DNA substrates or cellular cofactor LEDGF, but it compromises formation of the fully functional nucleoprotein complex. Studies comparing **1** with a structurally related IN inhibitor, the tetraacetylated-chicoric acid derivative (**2**), indicated striking mechanistic differences between these agents. The structures of the two inhibitors differ only in their central linker regions with **1** and **2** containing a single methyl ester group and two carboxylic acids, respectively. MS experiments highlighted the importance of these structural differences for selective binding of **1** to the IN dimer interface. Moreover, molecular modeling of **1** complexed to IN identified a potential inhibitor binding cavity and provided structural clues regarding a possible role of the central methyl ester group in establishing an extensive hydrogen bonding network with both interacting subunits. The proposed mechanism of action and binding site for the small molecule inhibitor identified in the present study provide an attractive venue for developing allosteric inhibitors of HIV-1 IN.

Molecular Pharmacology Fast Forward. Published on July 28, 2009 as DOI: 10.1124/mol.109.058883
This article has not been copyedited and formatted. The final version may differ from this version.

Recent development of the first clinically useful strand-transfer inhibitor (STI) Raltegravir (MK-0518) has validated HIV-1 integrase (IN) as a important anti-retroviral target (Summa et al., 2008). The enzyme functions as a multimer to insert the reverse transcribed RNA genome into the host chromosome through two reaction steps. In the first step, called 3'-processing, IN cleaves a GT dinucleotide from each end of the viral DNA. In the second step, termed strand transfer, IN catalyzes concerted integration of the processed viral DNA ends into chromosomal DNA. Studies with purified recombinant protein and model DNA substrates indicated that the individual protein monomers establish complementary contacts with DNA substrates with the subunit-subunit contacts playing a crucial role in the formation of the functional nucleoprotein complexes (Engelman et al., 1993; van den Ent et al., 1999; van Gent et al., 1993; Zhao et al., 2008). While a dimeric protein is sufficient to process each 3'-end, a tetramer is needed to carry out the concerted integration of both viral ends (Faure et al., 2005; Guiot et al., 2006; Li et al., 2006).

Following the discovery of the transcriptional co-activator lens epithelium-derived growth factor (LEDGF) as a key cellular cofactor for HIV-1 integration, LEDGF-IN interactions have become a new venue for antiviral drug design (Al-Mawsawi et al., 2008; Busschots et al., 2009; De Rijck et al., 2006; Hou et al., 2008). LEDGF directly interacts with IN through a C-terminal region termed the integrase binding domain (IBD) (Cherepanov et al., 2004; Vanegas et al., 2005) and tethers the preintegration complex to chromatin (Emiliani et al., 2005; Llano et al., 2006b; Maertens et al., 2003; Shun et al., 2007). Overexpression of the IBD, which tightly binds IN but lacks an N-terminal nuclear localization signal or the chromatin binding domain of LEDGF, severely inhibits HIV-1 replication in cell culture assays (De Rijck et al., 2006). Furthermore, HIV-1 strains resistant to the STI were fully susceptible to inhibition by the IBD

Molecular Pharmacology Fast Forward. Published on July 28, 2009 as DOI: 10.1124/mol.109.058883
This article has not been copyedited and formatted. The final version may differ from this version.

(De Rijck et al., 2006). Intriguingly, the IBD is significantly more effective at suppressing HIV-1 replication in LEDGF-deficient cells (555-fold) when compared to cells containing LEDGF at normal levels (~30-fold) (Llano et al., 2006a). These observations suggest that in the absence of competing LEDGF levels, the IBD could effectively engage IN and adversely affect its function.

We recently observed that the high exchange rate among IN subunits in multimers can be reduced by IBD binding (McKee et al., 2008). The preformed IN-IBD complex is capable of binding donor DNA but fails to catalyze concerted integration (McKee et al., 2008). These findings suggested that the dynamic interplay between IN subunits is essential for the assembly of the fully functional nucleoprotein complex and that restricting the molecular movement of individual subunits within a multimer could compromise catalytic processes. Consistent with this, IBD derived peptides have been shown to stabilize a multimeric form of IN and impair HIV-1 infection (Al-Mawsawi et al., 2008; Hayouka et al., 2007).

While studies with the IBD have provided proof-of-concept for targeting the dynamic structure of free IN, a focus of developmental efforts is to discover small molecule inhibitors that can modulate the subunit-subunit interactions. In this regard, small molecule inhibitors that have been shown previously to selectively bind at the IN dimer interface are worth revisiting. For example, 3,4-dihydroxyphenyltriphenylarsonium bromide and a coumarin-based inhibitor have been reported to bind IN at sites that partly overlap the IBD binding pocket (Al-Mawsawi et al., 2006; Molteni et al., 2001). Using affinity acetylation and MS analysis we previously mapped one contact between the small molecule inhibitor methyl N,O-bis(3,4,-diacetyloxycinnamoyl)-

Molecular Pharmacology Fast Forward. Published on July 28, 2009 as DOI: 10.1124/mol.109.058883
This article has not been copyedited and formatted. The final version may differ from this version.

serinate (**1**, see Figure 1) and Lys173 located at the IN dimer interface (Shkriabai et al., 2004). However, the mechanisms of actions of this and similar compounds have not been elucidated.

In the present study, we extend our previous research with **1**. Using a subunit exchange assay we have observed that the inhibitor modulates dynamic interactions between IN subunits in a dose-dependent fashion. Furthermore, we have found that the central linker region of **1** is essential for selective binding of the inhibitor to the IN dimer interface. Moreover, molecular modeling experiments have enabled us to propose contact amino acids and define a potential small molecule binding cavity within HIV-1 IN, which may represent a new therapeutic target.

MATERIALS AND METHODS

Preparation of wild type and mutant recombinant HIV-1 IN.

Full-length wild type and soluble (F185K/C280S) integrase proteins were expressed in *E. coli*. The point mutations were introduced in the wild type IN sequence using the Qiagen PCR mutagenesis kit (Qiagen). Wild-type and mutant IN proteins were purified according to previously described procedures (McKee et al., 2008).

Assay for HIV-1 IN inhibitors:

1 and **2** were prepared as reported previously (Lin et al., 1999). MALDI-ToF analysis confirmed successful synthesis and a high degree of purity of the full length products (see supplemental Figure 1). The preparations of **1** and **2** were dissolved in DMSO to make 10 mM stock solutions and stored at -20 ^{o}C until use. The compounds have been highly stable throughout the course of our studies.

Molecular Pharmacology Fast Forward. Published on July 28, 2009 as DOI: 10.1124/mol.109.058883
This article has not been copyedited and formatted. The final version may differ from this version.

Measurement of *in vitro* IC_{50} values was carried out as described previously (Hazuda et al., 1994). Briefly, a short double stranded DNA (a hybrid of the 21-mer and 19-mer synthetic oligonucleotides) representing the preprocessed HIV-1 U5 LTR end was immobilized on a Covalink microtiter plate (NUNC) via carbodiimide-mediated condensation. After incubation with IN, a biotinylated target oligonucleotide was added and the strand transfer reaction was allowed to proceed. The plates were washed thoroughly and blocked before incubation with alkaline phosphatase-conjugated streptavidin (Pierce). After additional washing, para-nitrophenylphosphate (pNPP) was added and the development of p-nitrophenolate was determined spectrophotometrically at 405nm.

Mass spectrometric analyses:

The acetylation reactions for short control peptides and recombinant IN have been described previously (Shkriabai et al., 2004). Briefly 5 μM peptides were incubated with 500 μM **1** or **2** for 30 min at 37 ºC and the reaction products were analyzed with MALDI-ToF using an Axima CFR instrument (Shimadzu Biotech, Manchester, UK) and the α-cyano-4-hydroxy-cinnamic acid matrix. To identify the sites in HIV-1 IN modified by **1** and **2**, the protein was first incubated with increasing concentrations of inhibitors, then digested by trypsin and the proteolytic fragments were analyzed with Shimadzu MALDI-ToF and Waters Q-ToF-II instruments (Manchester, U.K.) as described previously (Shkriabai et al., 2004).

DNA-IN interactions.

Disulfide-mediated IN-DNA crosslinking has been described previously (Zhao et al., 2008). The assays employed IN(E152C) protein and a 21-mer specific DNA with a crosslinkable

Molecular Pharmacology Fast Forward. Published on July 28, 2009 as DOI: 10.1124/mol.109.058883
This article has not been copyedited and formatted. The final version may differ from this version.

analog at the G2 position (see (Zhao et al., 2008) for detailed preparation of the protein and DNA). The cross-linking reactions were performed by incubating 1 μM IN with equimolar DNA duplex in 50 mM HEPES, pH 7.0, 100 mM NaCl, 5 mM $MgCl_2$, and 10% glycerol at 37 ^oC for 20 min. The reactions were quenched by addition of 20 mM methyl methanethiosulfonate and subjected to SDS-PAGE analysis. Free-protein and protein-DNA complexes were visualized by Western blots using IN anti-serum.

Assay for IN subunit exchange

A subunit exchange assay was performed as described previously (McKee et al., 2008). Briefly, His-tagged IN (50 nM) was pre-incubated in pull-down (PD) buffer (50 mM Hepes, pH 7.1, 300 mM NaCl, 100 mM Imidazole, 2 mM $MgCl_2$, 2 mM β-mercaptoethanol, 0.1% (v/v) Nonidet P40) for 30 min at room temperature with and without increasing amounts of ligands (**1** or **2**). An equimolar amount of tag-free IN (50 nM) was then added and subunit exchange was allowed for 60 min at room temperature. Samples were briefly centrifuged for 2 min at 1,000 g to remove non-specific aggregates. Supernatants were incubated with nickel-agarose beads (GE) for 30 min in the presence of BSA (0.1 mg/ml). Following the affinity pull-down of protein-protein complexes, the resin was washed three times with the PD buffer and boiled in SDS-PAGE loading buffer. His-tagged and tag-free INs were then separated by SDS-PAGE and visualized by Western blots using a mouse monoclonal IN antibody (the AIDS Research and Reference Reagent Program, NIH).

IN Crosslinking Assay

The IN crosslinking assay was performed as described previously (Li et al., 2006). Briefly, IN (100 nM) was pre-incubated in CL buffer (50 mM Hepes, pH 7.1, 300 mM NaCl, 2 mM MgCl2,

Molecular Pharmacology Fast Forward. Published on July 28, 2009 as DOI: 10.1124/mol.109.058883
This article has not been copyedited and formatted. The final version may differ from this version.

1 mM DTT, 0.1% (v/v) Nonidet P40) for 30 min. at room temperature with or without **1**. The bifunctional reagent BS_3 (Pierce) was then added at 100 µM final concentration and the mixture was incubated for 15 min. at room temperature. Reactions were quenched by addition of Tris/NaOAc and boiled in the SDS-PAGE loading buffer. INs were separated by SDS-PAGE and visualized by Western blot using a mouse monoclonal IN antibody (HIV-1 IN Monoclonal Antibody (2C11) from Dr. Dag E. Helland obtained through the AIDS Research and Reference Reagent Program, Division of AIDS, NIAID, NIH).

Protein solubility test

IN (100 nM) was pre-incubated in PD buffer (50 mM Hepes, pH 7.1, 300 mM NaCl, 100 mM imidazole, 2 mM $MgCl_2$, 2 mM β-mercaptoethanol, 0.1% (v/v) Nonidet P40) for 60 min at room temperature with or without 250 µM ligand (**1** or **2**). Samples were then centrifuged for 10 min at 5,000 g. Pellets and supernatants were analyzed by SDS-PAGE and detected by Western blot as described above.

Monitoring affects of compounds on IN-LEDGF interactions

LEDGF-IN binding assays were performed as described previously (McKee et al., 2008). Briefly, His-tagged IN (100 nM) was pre-incubated in PD buffer (50 mM Hepes, pH 7.1, 300 mM NaCl, 100 mM imidazole, 2 mM $MgCl_2$, 2 mM β-mercaptoethanol, 0.1% (v/v) Nonidet P40) for 30 min at RT with and without increasing amounts of ligands (**1** or **2**). Tag-free LEDGF (300 nM) was then added and incubated for 60 min at room temperature. Samples were centrifuged and pulled-down by a nickel-agarose resin as described above. The bound proteins were subjected to SDS-PAGE separation and LEDGF was visualized by Western blot using a mouse monoclonal LEDGF antibody (BD Biosciences).

Molecular Pharmacology Fast Forward. Published on July 28, 2009 as DOI: 10.1124/mol.109.058883
This article has not been copyedited and formatted. The final version may differ from this version.

LEDGF binding to wild-type and mutant IN

LEDGF (1 µM) was incubated with 600 nM His-tagged IN (WT or mutant) in binding buffer (50 mM Tris-HCl, pH 7.5, 150 mM NaCl, 2 mM $MgCl_2$, 35 mM imidazole, 2 mM β-mercaptoethanol, 0.1% (v/v) Nonidet P40) for 60 min at room temperature. Samples were pulled-down by a nickel-affinity resin for 30 min in the presence of BSA (0.1 mg/ml). The resin was washed three times with the same buffer, and the bound proteins were subjected to SDS-PAGE separation and visualized by Coomassie blue staining.

Size Exclusion Chromatography:

10 µm protein solutions in storage buffer (50 mM HEPES, pH 7.4, 1 M NaCl, 7.5 mM CHAPS, 10 mM DTT, 10 % glycerol) were subjected to size exclusion chromatography using a Superdex 200 10/300 GL column (GE Healthcare) and running buffer (50 mM HEPES, pH 7.4, 1 M 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 docking studies

All simulations were performed on a Silicon Graphics O2 work station. The coordinates of the catalytic core domain (CCD) of IN were extracted from the crystal structure 2B4J (Cherepanov et al., 2005) obtained from the RCSB Protein Data Bank. The initial 3D model of **1** was generated using the biopolymer module within the Insight II software package (Accelrys Inc., San Diego). Molecular docking studies were performed using the Autodock program (version 3.0) which allows virtual docking of a fully flexible ligand within a rigid defined binding pocket

Molecular Pharmacology Fast Forward. Published on July 28, 2009 as DOI: 10.1124/mol.109.058883
This article has not been copyedited and formatted. The final version may differ from this version.

(Goodsell and Olson, 1990). A 'target grid' covering both putative binding cavities (including K173, H171, D167 W131 and T125 residues) was generated. Polar hydrogens were added, Kollman charges were assigned and 3-D affinity grid maps were calculated for each atom type (C, A (aromatic C), N, O, S, and H) using the Autogrid module. For the ligand, hydrogens were added, Gasteiger charges were assigned and the rotatable bonds were determined by the AutoTors module. The docking parameters were as follows: genetic algorithm, trials of 50 dockings, random starting position and conformation, rotation step ranges of 5° and 1 million energy evaluations. The Autodock pose showing the lowest binding energy was extensively energy-minimized using the Discover module of Insight II software using the CFF91 force field and the steepest descent method. Both protein and ligand were left flexible and no constraints were applied.

RESULTS

1 and **2** (Figure 1A) have been synthesized previously (Lin et al., 1999) as structural analogs of chicoric acid (CA) (see supplemental figure 2). **2** represents tetraacetylated CA. **1** also contains di-*O*-tetraacetyloxycinnamoyl groups but it differs significantly from **2** and CA in the central linker region. For example, **1** possesses a non-charged methyl ester group, while **2** and CA contain two polar carboxylic acids (Figure 1). Despite these structural differences the compounds inhibited IN with similar IC_{50} values ((Lin et al., 1999), also see Figure 1B) and they were assumed to exhibit similar mechanisms of action. However, more recent studies have suggested that **1** may differ from authentic CA compounds in its interactions with IN. For example, using affinity acetylation coupled with MS analysis we have demonstrated that **1** selectively acetylates Lys173 located at the protein dimer interface ((Shkriabai et al., 2004),

Molecular Pharmacology Fast Forward. Published on July 28, 2009 as DOI: 10.1124/mol.109.058883
This article has not been copyedited and formatted. The final version may differ from this version.

see also Figure 2). In contrast, CA has been reported to interact at the IN active site (Healy et al., 2009), which is significantly distanced from the dimer interface.

To clarify these differences we compared **1** and **2** for their interactions with HIV-1 IN. We chose **2** as a representative CA for the following reasons. The detailed *in vitro* IN activity and cell culture infectivity assays indicated that **2** and hydroxylated CA exhibit the same mechanism of action (Lin et al., 1999; Pluymers et al., 2000). Furthermore, in common with **1**, **2** contains di-*O*-acetyl groups that would enable us to compare interactions of the two compounds with IN using the affinity acetylation and MS analysis approach (Shkriabai et al., 2004).

Using model peptides we have previously demonstrated that aryl di-*O*-acetyl compounds could effectively acetylate Lys, Cys and Tyr side chains (Shkriabai et al., 2004). Therefore, to compare acetylation reactions for **1** and **2** we used a short control peptide (HDMNKVLDL) containing a single Lys residue. As expected no acetylation was detected at low compound concentrations (5 μM) due to the lack of specific interactions. In contrast, 500 μM **1** and **2** effectively acetylated the target Lys residue (data not shown). Comparison of the modified peak intensities indicated that aryl di-*O*-acetyl groups in the tested compounds exhibit very similar chemical reactivities.

In contrast, a striking difference was observed between **1** and **2** in their interactions with HIV-1 IN (Figure 2). **1** at concentrations similar to its IC_{50} value effectively acetylated Lys173, while **2** failed to modify this residue even at very high concentrations (200 μM) (Figure 2). No specific acetylation of IN was detected with low micromolar concentrations of **2** suggesting that in the

Molecular Pharmacology Fast Forward. Published on July 28, 2009 as DOI: 10.1124/mol.109.058883
This article has not been copyedited and formatted. The final version may differ from this version.

protein-inhibitor complex di-*O*-acetyl groups were not positioned close enough to react with the side chains of Lys, Cys or Tyr residues. At elevated concentrations (200 μM) of **2** multiple Lys residues were acetylated. This could be due to non-specific modes of modifying surface residues. To test this hypothesis, we performed control experiments using 500 μM of primary amine modifying reagent sulfo-NHS-biotin. The modification profiles obtained with **2** and sulfo-NHS-biotin were very similar indicating that the affected residues were indeed targeted for their surface exposure rather than specific interactions. Thus, our experiments did not reveal the specific binding site for **2** but indicated marked differences in **1** and **2** binding to IN.

Further evidence for different mechanisms of actions for **1** and **2** emerged from IN-DNA binding studies. The results in Figure 3 show that **2** interfered with IN-DNA complex formation in a dose dependent manner. These results are consistent with a recently proposed molecular model (Healy et al., 2009), where the CA binding site partly overlaps with the presumed donor DNA binding cleft. In contrast, **1** did not directly interfere with IN-DNA interactions suggesting an alternative mechanism of action for this compound.

Further experiments were performed to elucidate a mechanism of action for **1** (Figures 4 and 5) and define the inhibitor binding site in more detail (Figures 6 and 7). Given that **1** selectively interacts with Lys173 located at the protein dimer interface (Figure 2), we examined whether **1** could affect the dynamic interactions between IN subunits. For this purpose we employed the subunit exchange assay devised by our group ((McKee et al., 2008), also see Figure 4A). This method monitors kinetic interactions between the two wild type proteins: His-tagged IN (indicated as IN2 in the monomeric state and IN2-IN2 in the dimeric state) and tag-free IN (indicated as IN1 in the monomeric state and IN1-IN1 in the dimeric state). Due to the effective

Molecular Pharmacology Fast Forward. Published on July 28, 2009 as DOI: 10.1124/mol.109.058883
This article has not been copyedited and formatted. The final version may differ from this version.

exchange between individual protein subunits, the tag-free protein within the IN1-IN2 complex can be quantitatively pulled-down by a nickel-affinity resin. To reveal whether the compounds can modulate these interactions, IN2 was first pre-incubated with increasing concentrations of **1** or **2** and then exposed to IN1. The data in Figure 4B (lanes 5-10) show that **1** impaired dynamic exchange between IN subunits in dose dependent manner. In contrast, **2** tested at high concentrations had no detectable effect on IN subunit-subunit exchange (Figure 4B, lanes 11 and 12).

Dynamic modulation of subunit-subunit exchange can be achieved either through **1** stabilizing the interacting IN monomers into a multimeric state or by preventing multimer formation through interference with subunit-subunit assembly. To differentiate between these possibilities order-of-addition experiments were performed (Figure 4C). Pre-incubation of **1** with IN2 and subsequent addition of IN1 to the reaction mixture effectively interfered with the subunit-subunit exchange (Figure 4C, lane 2). However, **1** did not significantly affect the pre-assembled IN1-IN2 complex (Figure 4C, lane 4). These findings suggest that **1** does not dissociate IN multimers. Instead, **1** binds to multimeric IN and restricts the ability of interacting subunits to exchange freely among multimers.

Effects of **1** on IN multimerization were also tested using bifunctional crosslinkers and SDS-PAGE separation (Figure 5A). This approach has been instrumental in probing IN structures in the synaptic complex and infected cells (Cherepanov et al., 2003; Li et al., 2006). As expected (Li et al., 2006) different multimeric species were observed with the free protein. Upon addition of **1,** the intensity of the monomeric band was significantly reduced. Similar quantities of dimeric IN bands were observed in free IN and the IN-**1** complex samples. Whereas, a band

Molecular Pharmacology Fast Forward. Published on July 28, 2009 as DOI: 10.1124/mol.109.058883
This article has not been copyedited and formatted. The final version may differ from this version.

corresponding to a tetramer was enhanced at least by 50 % in the protein-inhibitor complex compared with free IN. The observed crosslinking patterns (Figure 5A) were highly reproducible. Since in the **1**-IN complex a monomer band was reduced and multiple oligomeric species were detected, it is logical to suggest that the inhibitor promotes protein multimerization in general rather than stabilizes one particular multimeric form of IN. For example, **1** could affect conformations of interacting subunits so that the **1**-IN complex is less flexible (Figure 4) and more susceptible for multimerization (Figure 5A) than free IN.

Next, we examined whether **1** induced protein multimerization could affect IN solubility. Figure 5B shows that under our assay conditions free IN remained in solution and that **1** did not alter IN solubility. Given that LEDGF binds at IN dimer interface adjacent to K173, we checked whether **1** could affect the IN-LEDGF complex formation. The results in Figure 5C indicate that **1** did not significantly affect protein-protein interactions indicating that the primary binding sites of **1** and LEDGF differ.

The following efforts were devoted to a more detailed characterization of the sites of **1** binding to IN. Specific acetylation of Lys173 provided an important clue regarding one of the contacts between the terminal di-*O*-acetyl group of the inhibitor and the protein. However, comparative SAR studies of **1** and **2** revealed an essential role of the central linker region for determining binding and mechanistic properties of these inhibitors. Indeed, **1** containing the methyl ester group within the linker interacted with the IN dimer and modulated dynamic interactions between the subunits (Figures 2 and 4) but it did not directly affect IN-DNA interactions (Figure 3). In contrast, **2** possessing two central carboxylic acid groups failed to bind to the IN dimer

Molecular Pharmacology Fast Forward. Published on July 28, 2009 as DOI: 10.1124/mol.109.058883
This article has not been copyedited and formatted. The final version may differ from this version.

interface (Figure 2) or modulate subunit-subunit interactions (Figure 4). Instead, **2** interfered with the IN-DNA complex formation (Figure 3).

To identify additional contact amino acids contributing to specific binding of **1** to the protein dimer interface, we used molecular modeling. While full length IN has not been amenable to high resolution structural studies, atomic structures of individual protein domains are available. Lys173 is located in the catalytic core domain (CCD). Crystallographic studies indicated a dimeric organization of this domain (Dyda et al., 1994). The CCD contains the DDE motif coordinating the catalytic divalent metals and an IBD binding site (Figures 6). Analysis of the protein surface surrounding Lys173 revealed two adjacent pockets depicted in orange and magenta (Figure 6, A and B). The pocket in orange has been shown to tightly interact with the LEDGF (Cherepanov et al., 2005; Hare et al., 2009). Furthermore, the two small molecule inhibitors (3,4-dihydroxyphenyltriphenylarsonium bromide and a coumarin-based compound) have also been demonstrated to target this site (Al-Mawsawi et al., 2006; Molteni et al., 2001). In contrast, the similarly sized cavity depicted in magenta has not previously been implicated in small molecule binding. Interestingly, Lys173 is located at the boundary separating these distinct pockets (Fig. 6 A and B). Therefore, **1** could potentially target either of the two similarly sized cavities.

To delineate between interactions at the two possible binding sites (magenta and orange), docking simulations were performed using the Autodock software and a ‘target grid’ covering both putative binding cavities on the dimeric CCD structure (see virtual docking details in the Materials and Methods section). Table 1 shows results for the 10 lowest binding energy, which indicate a striking preference for **1** binding to the site represented by “magenta” cavity (Figure

Molecular Pharmacology Fast Forward. Published on July 28, 2009 as DOI: 10.1124/mol.109.058883
This article has not been copyedited and formatted. The final version may differ from this version.

6). The lowest docking pose (-5.30 kcal/mol) was further minimized using Insight II software (see Materials and Methods).

The resulting molecular model for the CCD dimer-**1** complex is shown in Figure 6, C and D. The inhibitor interacts with both IN monomers. One di-*O*-acetyl group of **1** is located at 3.8 Å within K173 (subunit yellow), a distance compatible with the specific acetylation of this position observed in the MS experiments (Figure 2). The central methyl ester group of the inhibitor is buried deeply within the cavity where it binds to IN through tight electrostatic and hydrophobic interactions (Figure 6, C and D). IN amino acids engaging the central linker region include side chains of E87, E96, Y99 and K103. These interactions could explain the importance of the linker region for the selective binding of **1** to the IN dimer.

To examine significance of these residues for IN function and **1** binding we prepared purified recombinant proteins containing the following point substitutions: E87A, E96A, Y99A and K103A in the wild type IN sequence. Size exclusion chromatography results in Figure 7A show that these mutations affected IN multimerization. For example, a peak corresponding to a tetrameric protein was reduced in all mutant proteins compared with wild type IN. We also examined acetylation of the mutant proteins with **1**. In sharp contrast with wild type IN, K173 was not acetylated in any of the four mutant proteins even in the presence of elevated concentrations (200 μM) of **1** (data not shown). These experiments, however, did not fully delineate whether E87, E96, Y99 and K103 directly or indirectly contribute to **1** binding to IN as their substitutions affected IN multimerization. Nevertheless, our findings highlight the importance of a correctly assembled IN multimer for selective binding of **1**. Figure 7B shows that all of the mutant proteins were catalytically inactive. Furthermore, E87A failed to bind

Molecular Pharmacology Fast Forward. Published on July 28, 2009 as DOI: 10.1124/mol.109.058883
This article has not been copyedited and formatted. The final version may differ from this version.

LEDGF, and the E96A, Y99A and K103A substitutions reduced IN affinity for the cellular cofactor (Figure 7C). Taken together, the mutagenesis studies highlighted the architectural importance of this newly identified small molecule binding cavity for IN function and provide proof-of-concept for targeting this site with allosteric inhibitors.

DISCUSSION

Here we show for the first time that a small molecule compound inhibits HIV-1 IN catalytic function through allosteric modulation of dynamic interactions between the individual protein subunits (Figure 4). The highly dynamic nature of free IN subunits is essential for the productive assembly of the fully functional IN-vDNA complex (McKee et al., 2008). Binding of **1** to free IN stabilizes the interacting subunits (Figure 4) and promotes protein multimerization (Figure 5A). The preformed protein-protein complex is capable of binding DNA substrates (Figure 3) as **1** does not appear to directly mask the substrate binding channel in the retroviral enzyme (Figures 2 and 6). Instead, **1** could compromise IN catalytic activities by the following allosteric mechanisms. The restricted flexibility of the preformed **1**-IN complex may impede correct positioning of the catalytic site on the cognate DNA substrate. Alternatively, **1** could limit a conformational flexibility of IN, which is needed for formation of the fully functional nucleoprotein complex (Zhao et al., 2008).

Proposed mechanisms of action for **1** are reminiscent of modulation of IN structure and function by LEDGF and its IBD. For example, *in vitro* experiments have indicated that the cellular cofactor effectively impairs IN subunit-subunit exchange and promotes IN tetramerization (McKee et al., 2008). The pre-assembled protein-protein complexes are capable of binding donor DNA but are defective for the concerted integration (McKee et al.,

Molecular Pharmacology Fast Forward. Published on July 28, 2009 as DOI: 10.1124/mol.109.058883
This article has not been copyedited and formatted. The final version may differ from this version.

2008; Pandey et al., 2007; Raghavendra and Engelman, 2007). These findings have suggested that IN tetramers formed in IN-LEDGF and IN-vDNA complexes are not identical. In contrast, the full length cofactor stimulates the catalytic reactions when IN-vDNA complexes are pre-formed and then LEDGF is provided (Hare et al., 2009; Pandey et al., 2007; Raghavendra and Engelman, 2007). These *in vitro* observations, highlighting the importance of the order of vDNA and LEDGF binding to IN, corroborate with the sequence of events during early stages of HIV-1 replication. IN first encounters vDNA in the cytoplasm, where the exchange of subunits in multimers may be essential for effective assembly of the fully functional nucleoprotein complex termed the preintegration complex (PIC). LEDGF then engages the preassembled IN-vDNA complex in the nucleus and tethers PICs to the chromatin without disturbing the structural arrangements of IN with its DNA substrates.

This chronology of events would suggest that IN prior to its binding to vDNA rather than PICs could be a plausible target for allosteric inhibitors. Indeed, PICs have a very stable structure and are likely to be resistant to conformational challenges with small allosteric compounds. In contrast, unliganded IN at earlier stages of viral replication, before synthesis of vDNA by reverse transcriptase is complete, could be vulnerable to the attack by allosteric inhibitors. This notion is supported by the published data (De Rijck et al., 2006; Llano et al., 2006a) indicating that the overexpression of the eGFP-IBD, which lacks nuclear localization signal and interacts with unliganded IN in the cytoplasm, effectively inhibits HIV-1 integration. Consistently, the IBD derived short peptides have also been shown to stabilize IN multimers *in vitro* and impair HIV-1 integration in the infected cells (Al-Mawsawi et al., 2008; Hayouka et al., 2007).

Molecular Pharmacology Fast Forward. Published on July 28, 2009 as DOI: 10.1124/mol.109.058883
This article has not been copyedited and formatted. The final version may differ from this version.

Despite certain mechanistic similarities between the IBD and **1** dependent inactivation of IN activities, binding modes of and oligomerization pathways induced by these ligands could differ. For example, the IBD establishes extensive contacts with the CCD of one dimer and the N-terminal domain (NTD) of another dimer to stabilize a tetramer of IN (McKee et al., 2008). In contrast, the binding site of **1** is likely to be restricted to a much smaller area. Our results (Figures 2 and 6) suggest that **1** binds at the IN dimer interface at an allosteric pocket, which differs from the IBD binding sites. The resulting inhibitor-IN complex may lack sufficient flexibility to carry out catalytic reactions. Future structural studies of the inhibitor-protein complex are necessary to gain important details for how **1** affects the multimeric structure of IN.

Our MS (Figure 2) and molecular modeling experiments (Figure 6) have provided initial clues about the amino acids interacting with **1**. Unfortunately, the mutagenesis studies did not allow us to elucidate whether the proposed residues directly engage the inhibitor since the single amino acid substitutions adversely affected both the inhibitor binding and protein multimerization (Figure 7). Nevertheless, mutagenesis experiments have highlighted the architectural importance of residues within the proposed inhibitor binding site, and suggest that a correctly assembled IN multimer is a selective target for **1**. There are the following key advantages for targeting the proposed allosteric pocket with small molecule compounds. The pocket is significantly distanced from presumed STI binding sites. For example, IN mutations conferring resistance to Raltegravir are localized in the vicinity of the catalytic DDE motif (Cooper et al., 2008). Therefore, agents selectively binding at the protein dimer interface are likely to be active against IN phenotypes resistant to the STI. Furthermore, the inhibitors targeting the allosteric cavity (Figure 6) would not have to compete with IN-IN or IN-LEDGF

Molecular Pharmacology Fast Forward. Published on July 28, 2009 as DOI: 10.1124/mol.109.058883
This article has not been copyedited and formatted. The final version may differ from this version.

interactions. Structural studies indicate extensive monomer-monomer interfaces within dimeric IN (Chen et al., 2000; Wang et al., 2001). The LEDGF contacts in IN include two separate areas in the CCD and the NTD (Cherepanov et al., 2005; Hare et al., 2009; McKee et al., 2008). Therefore, finding small compounds that interfere with intra- or inter- protein-protein interactions could be challenging. In contrast, **1** is not required to overcome large energy barriers created by numerous interacting amino acids. Instead, the inhibitor exploits the preformed dimer interface, which is fully distinct from the LEDGF binding site.

The other two small molecule inhibitors (3,4-dihydroxyphenyltriphenylarsonium bromide and a coumarin-based compound) of IN have also been shown to interact with the CCD dimer interface (Al-Mawsawi et al., 2006; Molteni et al., 2001). However, their mechanisms of inhibition have not been elucidated. It is possible that these agents also function by dynamic modulation of IN multimeric structure. Our recently reported subunit-subunit exchange assay (see Figure 4 and (McKee et al., 2008)) provides means to examine SAR for these and related compounds.

Our assays enabled us to delineate mechanisms of action for **1** and authentic CA class inhibitors (see Figures 2, 3, 4 as well as supplemental Figure 2). Unlike **1**, **2** failed to interact with the IN dimer interface, and instead, impaired IN-DNA complex formation. The hydroxylated and tetraacetylated CAs have been previously examined in the cell culture assays (King and Robinson, 1998; Pluymers et al., 2000). Robinson and co-workers isolated a CA resistant strain containing a mutation in the IN gene corresponding to the G140S IN variant (King and Robinson, 1998). However, subsequent studies with hydroxylated and

Molecular Pharmacology Fast Forward. Published on July 28, 2009 as DOI: 10.1124/mol.109.058883
This article has not been copyedited and formatted. The final version may differ from this version.

tetraacetylated CAs mapped resistant mutations to the gp120 gene, suggesting that these compounds may target multiple sites in infected cells (Pluymers et al., 2000).

1 has not been tested in the cell culture assays. However, the presence of di-*O*-acetyl groups in this inhibitor is likely to adversely affect its specificity. For example, these groups are substrates for cellular acetyltransferases and the resulting catechols could deter **1** from its retroviral target. Comparative SAR and molecular modeling studies (Figures 2, 4 and 6) suggest that the central linker establishes extensive hydrogen bonding network with IN subunits, while contributions of di-*O*-acetyl groups to the **1**-IN complex formation are relatively limited. Therefore, the tetraacetate groups, in theory, could be replaced by alternative structures without compromising the binding specificity. In fact, continued optimization of the **1** structure could lead to the identification of more potent inhibitors with desired specificity in the infected cells.

Taken together, the mechanism of action and binding site for the small molecule inhibitor described herein provides an attractive venue for targeting IN. Furthermore, our findings facilitate wider ongoing efforts in the field to develop new types of clinically useful allosteric inhibitors of IN that could effectively complement current antiretroviral therapies.

Molecular Pharmacology Fast Forward. Published on July 28, 2009 as DOI: 10.1124/mol.109.058883
This article has not been copyedited and formatted. The final version may differ from this version.

ACKNOWLEDGEMENTS:

HIV-1 IN Monoclonal Antibody (2C11) was obtained from Dr. Dag E. Helland through the AIDS Research and Reference Reagent Program, Division of AIDS, the National Institute of Diabetes and Digestive and Kidney Diseases (NIDDK), National Institutes of Health (NIH). Experiments by S.H. were conducted at the Proteomics and Mass Spectrometry Facility of NIDDK.

Molecular Pharmacology Fast Forward. Published on July 28, 2009 as DOI: 10.1124/mol.109.058883
This article has not been copyedited and formatted. The final version may differ from this version.

REFERENCES:

Al-Mawsawi, L.Q., Christ, F., Dayam, R., Debyser, Z. and Neamati, N. (2008) Inhibitory profile of a LEDGF/p75 peptide against HIV-1 integrase: insight into integrase-DNA complex formation and catalysis. *FEBS Lett*, **582**, 1425-1430.

Al-Mawsawi, L.Q., Fikkert, V., Dayam, R., Witvrouw, M., Burke, T.R., Jr., Borchers, C.H. and Neamati, N. (2006) Discovery of a small-molecule HIV-1 integrase inhibitor-binding site. *Proc Natl Acad Sci U S A*, **103**, 10080-10085.

Busschots, K., De Rijck, J., Christ, F. and Debyser, Z. (2009) In search of small molecules blocking interactions between HIV proteins and intracellular cofactors. *Mol Biosyst*, **5**, 21-31.

Chen, J.C., Krucinski, J., Miercke, L.J., Finer-Moore, J.S., Tang, A.H., Leavitt, A.D. and Stroud, R.M. (2000) Crystal structure of the HIV-1 integrase catalytic core and C-terminal domains: a model for viral DNA binding. *Proc Natl Acad Sci U S A*, **97**, 8233-8238.

Cherepanov, P., Ambrosio, A.L., Rahman, S., Ellenberger, T. and Engelman, A. (2005) Structural basis for the recognition between HIV-1 integrase and transcriptional coactivator p75. *Proc Natl Acad Sci U S A*, **102**, 17308-17313.

Cherepanov, P., Devroe, E., Silver, P.A. and Engelman, A. (2004) Identification of an evolutionarily conserved domain in human lens epithelium-derived growth factor/transcriptional co-activator p75 (LEDGF/p75) that binds HIV-1 integrase. *J Biol Chem*, **279**, 48883-48892.

Cherepanov, P., Maertens, G., Proost, P., Devreese, B., Van Beeumen, J., Engelborghs, Y., De Clercq, E. and Debyser, Z. (2003) HIV-1 integrase forms stable tetramers and associates with LEDGF/p75 protein in human cells. *J Biol Chem*, **278**, 372-381.

Cooper, D.A., Steigbigel, R.T., Gatell, J.M., Rockstroh, J.K., Katlama, C., Yeni, P., Lazzarin, A., Clotet, B., Kumar, P.N., Eron, J.E., Schechter, M., Markowitz, M., Loutfy, M.R., Lennox, J.L., Zhao, J., Chen, J., Ryan, D.M., Rhodes, R.R., Killar, J.A., Gilde, L.R., Strohmaier, K.M., Meibohm, A.R., Miller, M.D., Hazuda, D.J., Nessly, M.L., DiNubile, M.J., Isaacs, R.D., Teppler, H. and Nguyen, B.Y. (2008) Subgroup and resistance analyses of raltegravir for resistant HIV-1 infection. *N Engl J Med*, **359**, 355-365.

De Rijck, J., Vandekerckhove, L., Gijsbers, R., Hombrouck, A., Hendrix, J., Vercammen, J., Engelborghs, Y., Christ, F. and Debyser, Z. (2006) Overexpression of the lens epithelium-derived growth factor/p75 integrase binding domain inhibits human immunodeficiency virus replication. *J Virol*, **80**, 11498-11509.

Dyda, F., Hickman, A.B., Jenkins, T.M., Engelman, A., Craigie, R. and Davies, D.R. (1994) Crystal structure of the catalytic domain of HIV-1 integrase: similarity to other polynucleotidyl transferases. *Science*, **266**, 1981-1986.

Emiliani, S., Mousnier, A., Busschots, K., Maroun, M., Van Maele, B., Tempe, D., Vandekerckhove, L., Moisant, F., Ben-Slama, L., Witvrouw, M., Christ, F., Rain, J.C., Dargemont, C., Debyser, Z. and Benarous, R. (2005) Integrase Mutants Defective for Interaction with LEDGF/p75 Are Impaired in Chromosome Tethering and HIV-1 Replication. *J Biol Chem*, **280**, 25517-25523.

Engelman, A., Bushman, F.D. and Craigie, R. (1993) Identification of discrete functional domains of HIV-1 integrase and their organization within an active multimeric complex. *Embo J*, **12**, 3269-3275.

Faure, A., Calmels, C., Desjobert, C., Castroviejo, M., Caumont-Sarcos, A., Tarrago-Litvak, L., Litvak, S. and Parissi, V. (2005) HIV-1 integrase crosslinked oligomers are active in vitro. *Nucleic Acids Res*, **33**, 977-986.

Goodsell, D.S. and Olson, A.J. (1990) Automated docking of substrates to proteins by simulated annealing. *Proteins*, **8**, 195-202.

Guiot, E., Carayon, K., Delelis, O., Simon, F., Tauc, P., Zubin, E., Gottikh, M., Mouscadet, J.F., Brochon, J.C. and Deprez, E. (2006) Relationship between the oligomeric status of HIV-1 integrase on DNA and enzymatic activity. *J Biol Chem*, **281**, 22707-22719.

Hare, S., Shun, M.C., Gupta, S.S., Valkov, E., Engelman, A. and Cherepanov, P. (2009) A novel co-crystal structure affords the design of gain-of-function lentiviral integrase mutants in the presence of modified PSIP1/LEDGF/p75. *PLoS Pathog*, **5**, e1000259.

Hayouka, Z., Rosenbluh, J., Levin, A., Loya, S., Lebendiker, M., Veprintsev, D., Kotler, M., Hizi, A., Loyter, A. and Friedler, A. (2007) Inhibiting HIV-1 integrase by shifting its oligomerization equilibrium. *Proc Natl Acad Sci U S A*, **104**, 8316-8321.

Hazuda, D.J., Hastings, J.C., Wolfe, A.L. and Emini, E.A. (1994) A novel assay for the DNA strand-transfer reaction of HIV-1 integrase. *Nucleic Acids Res*, **22**, 1121-1122.

Healy, E.F., Sanders, J., King, P.J. and Robinson, W.E., Jr. (2009) A docking study of L-chicoric acid with HIV-1 integrase. *J Mol Graph Model*, **27**, 584-589.

Hou, Y., McGuinness, D.E., Prongay, A.J., Feld, B., Ingravallo, P., Ogert, R.A., Lunn, C.A. and Howe, J.A. (2008) Screening for antiviral inhibitors of the HIV integrase-LEDGF/p75 interaction using the AlphaScreen luminescent proximity assay. *J Biomol Screen*, **13**, 406-414.

King, P.J. and Robinson, W.E., Jr. (1998) Resistance to the anti-human immunodeficiency virus type 1 compound L-chicoric acid results from a single mutation at amino acid 140 of integrase. *J Virol*, **72**, 8420-8424.

Li, M., Mizuuchi, M., Burke, T.R., Jr. and Craigie, R. (2006) Retroviral DNA integration: reaction pathway and critical intermediates. *Embo J*, **25**, 1295-1304.

Lin, Z., Neamati, N., Zhao, H., Kiryu, Y., Turpin, J.A., Aberham, C., Strebel, K., Kohn, K., Witvrouw, M., Pannecouque, C., Debyser, Z., De Clercq, E., Rice, W.G., Pommier, Y. and Burke, T.R., Jr. (1999) Chicoric acid analogues as HIV-1 integrase inhibitors. *J Med Chem*, **42**, 1401-1414.

Llano, M., Saenz, D.T., Meehan, A., Wongthida, P., Peretz, M., Walker, W.H., Teo, W. and Poeschla, E.M. (2006a) An essential role for LEDGF/p75 in HIV integration. *Science*, **314**, 461-464.

Llano, M., Vanegas, M., Hutchins, N., Thompson, D., Delgado, S. and Poeschla, E.M. (2006b) Identification and characterization of the chromatin-binding domains of the HIV-1 integrase interactor LEDGF/p75. *J Mol Biol*, **360**, 760-773.

Maertens, G., Cherepanov, P., Pluymers, W., Busschots, K., De Clercq, E., Debyser, Z. and Engelborghs, Y. (2003) LEDGF/p75 is essential for nuclear and chromosomal targeting of HIV-1 integrase in human cells. *J Biol Chem*, **278**, 33528-33539.

McKee, C.J., Kessl, J.J., Shkriabai, N., Dar, M.J., Engelman, A. and Kvaratskhelia, M. (2008) Dynamic modulation of HIV-1 integrase structure and function by cellular lens epithelium-derived growth factor (LEDGF) protein. *J Biol Chem*, **283**, 31802-31812.

Molteni, V., Greenwald, J., Rhodes, D., Hwang, Y., Kwiatkowski, W., Bushman, F.D., Siegel, J.S. and Choe, S. (2001) Identification of a small-molecule binding site at the dimer interface of the HIV integrase catalytic domain. *Acta Crystallogr D Biol Crystallogr*, **57**, 536-544.

Pandey, K.K., Sinha, S. and Grandgenett, D.P. (2007) Transcriptional coactivator LEDGF/p75 modulates human immunodeficiency virus type 1 integrase-mediated concerted integration. *J Virol*, **81**, 3969-3979.

Pluymers, W., Neamati, N., Pannecouque, C., Fikkert, V., Marchand, C., Burke, T.R., Jr., Pommier, Y., Schols, D., De Clercq, E., Debyser, Z. and Witvrouw, M. (2000) Viral entry as the primary target for the anti-HIV activity of chicoric acid and its tetra-acetyl esters. *Mol Pharmacol*, **58**, 641-648.

Raghavendra, N.K. and Engelman, A. (2007) LEDGF/p75 interferes with the formation of synaptic nucleoprotein complexes that catalyze full-site HIV-1 DNA integration in vitro: implications for the mechanism of viral cDNA integration. *Virology*, **360**, 1-5.

Shkriabai, N., Patil, S.S., Hess, S., Budihas, S.R., Craigie, R., Burke, T.R., Jr., Le Grice, S.F. and Kvaratskhelia, M. (2004) Identification of an inhibitor-binding site to HIV-1 integrase with affinity acetylation and mass spectrometry. *Proc Natl Acad Sci U S A*, **101**, 6894-6899.

Shun, M.C., Raghavendra, N.K., Vandegraaff, N., Daigle, J.E., Hughes, S., Kellam, P., Cherepanov, P. and Engelman, A. (2007) LEDGF/p75 functions downstream from preintegration complex formation to effect gene-specific HIV-1 integration. *Genes Dev*, **21**, 1767-1778.

Summa, V., Petrocchi, A., Bonelli, F., Crescenzi, B., Donghi, M., Ferrara, M., Fiore, F., Gardelli, C., Gonzalez Paz, O., Hazuda, D.J., Jones, P., Kinzel, O., Laufer, R., Monteagudo, E., Muraglia, E., Nizi, E., Orvieto, F., Pace, P., Pescatore, G., Scarpelli, R., Stillmock, K., Witmer, M.V. and Rowley, M. (2008) Discovery of raltegravir, a potent, selective orally bioavailable HIV-integrase inhibitor for the treatment of HIV-AIDS infection. *J Med Chem*, **51**, 5843-5855.

van den Ent, F.M., Vos, A. and Plasterk, R.H. (1999) Dissecting the role of the N-terminal domain of human immunodeficiency virus integrase by trans-complementation analysis. *J Virol*, **73**, 3176-3183.

van Gent, D.C., Vink, C., Groeneger, A.A. and Plasterk, R.H. (1993) Complementation between HIV integrase proteins mutated in different domains. *Embo J*, **12**, 3261-3267.

Vanegas, M., Llano, M., Delgado, S., Thompson, D., Peretz, M. and Poeschla, E. (2005) Identification of the LEDGF/p75 HIV-1 integrase-interaction domain and NLS reveals NLS-independent chromatin tethering. *J Cell Sci*, **118**, 1733-1743.

Wang, J.Y., Ling, H., Yang, W. and Craigie, R. (2001) Structure of a two-domain fragment of HIV-1 integrase: implications for domain organization in the intact protein. *Embo J*, **20**, 7333-7343.

Zhao, Z., McKee, C.J., Kessl, J.J., Santos, W.L., Daigle, J.E., Engelman, A., Verdine, G. and Kvaratskhelia, M. (2008) Subunit-specific protein footprinting reveals significant structural rearrangements and a role for N-terminal Lys-14 of HIV-1 Integrase during viral DNA binding. *J Biol Chem*, **283**, 5632-5641.

Molecular Pharmacology Fast Forward. Published on July 28, 2009 as DOI: 10.1124/mol.109.058883
This article has not been copyedited and formatted. The final version may differ from this version.

FOOTNOTES:

This work has been supported by National Institutes of Health grants [AI062520 and AI077341] as well as by the Intramural Research Programs of National Institutes of Health for the Center for Cancer Research, the National Cancer Institute, and National Institute of Diabetes and Digestive and Kidney Diseases.

[1] Current affiliation: Center for Drug Evaluation and Research, U.S. Food and Drug Administration, Rockville, MD.

Molecular Pharmacology Fast Forward. Published on July 28, 2009 as DOI: 10.1124/mol.109.058883
This article has not been copyedited and formatted. The final version may differ from this version.

LEGENDS FOR FIGURES:

Figure 1. A. Structures of **1** and **2**. B. The inhibition profiles of recombinant wild type IN with compounds **1** (triangles) and **2** (circles). The mean values for at least three experiments are presented. The deviation for each measurement was within ± 10%.

Figure 2. Segments of Q-ToF mass spectra demonstrating selective acetylation of Lys173 with **1**. A. IN+ 6.25 μM **1**; B. IN + 25 μM **1**; C. IN + 10 μM **2**; D. IN + 200 μM **2**; E. free IN. The tryptic peptide of IN containing acetylated Lys173 is indicated. C1 is unmodified tryptic peptide (AMASDFNLPPVVAK) of IN, which serves as an internal control.

Figure 3. Effects of **1** and **2** on IN-DNA cross-linking. IN(E152C) was first incubated with increasing concentrations of **1** (A) and **2** (B), and then cross-linked to DNA(G2). The reaction products were resolved by SDS-PAGE. The bands corresponding to free IN and IN-DNA complex are indicated. A. Lane 1 molecular weight markers; lane 2, IN-DNA cross-linking in the absence of **1**. Lanes 3 to 10 contained increasing concentrations of **1** (1, 2, 4, 8, 16, 32, 64, 128 μM). B. Lane 1, molecular weight markers; Lane 2, IN-DNA cross-linking in the absence of **2**; lanes 3 to 9 contained increasing concentrations of **2** (2, 4, 8, 16, 32, 64, 128 μM)

Figure 4. Effects of **1** and **2** on IN subunit-subunit interactions. Experimental design (A) and results (B). A. The subunit exchange between IN multimers was tested by mixing the two wild type IN proteins: IN1, a tag-free form and IN2, containing the His-tag at its C terminus. The full-length proteins are depicted as dimers (IN1-IN1) and (IN2-IN2). Upon subunit-subunit exchange three IN populations can be formed: IN1-IN1, IN1-IN2 and IN2-IN2. Of these, IN2-IN2 and IN1-IN2 can be pulled-down by nickel-affinity resin through binding with the His-tag,

Molecular Pharmacology Fast Forward. Published on July 28, 2009 as DOI: 10.1124/mol.109.058883
This article has not been copyedited and formatted. The final version may differ from this version.

while the tag free IN1-IN1 is washed out. B. The IN1 and IN2 proteins from the bound complexes were subjected to SDS-PAGE and detected by Western blot. Lane 1: IN2 load; lane 2: IN1 load; lane 3: no ligand or IN1 (non specific pull-down control); lane 4: the subunit-subunit exchange in the absence of compounds; lanes 5-10: the subunit exchange reactions in the presence of increasing **1** concentrations (8, 16, 32, 64, 128, 256 µM); lane 11 and 12: reactions in the presence of 128 and 256 µM 2. C. Order-of-addition experiments. 1 and 3 subunit exchange reactions in the absence of compounds; lane 2, IN2 was first pre-incubated with **1** and then exposed to IN1; lane 4, IN2 and IN1 were first mixed to carry out subunit exchange and then **1** was added to the mixture.

Figure 5. Effects of **1** on IN multimerization (A), solubility (B) and LEDGF-IN binding (C). A. In parallel reactions free IN (lane 1) and the IN-**1** complex (lane 2) were subjected to crosslinking with BS_3 and the reaction products were separated by SDS-PAGE. Migrations of molecular weight markers and IN bands are indicated. B. Lane 1: molecular weigh markers; lane 2: (T) total sample before centrifugation; lanes 3 and 4: supernatant (S) and precipitate (P) fractions formed after centrifugation of IN in the presence of 10 % DMSO; lanes 5 and 6: IN solubility in the presence of 256 µM **1**. C. Lane 1: LEDGF input; lane 2: no ligand or IN (non-specific pull down control); lane 3: LEDGF and IN interactions in the absence of **1**; lanes 4-8: assays with increasing **1** concentrations (8, 16, 32, 64, 128 µM).

Figure 6. Molecular docking studies for **1**. A and B depict two potential binding pockets (colored in magenta and orange) at the CCD dimer interface. These distinct sites are located immediately adjacent to one another. A and B provide two different views of the CCD dimer to better visualize individual binding pockets. Individual subunits of IN are colored green and

Molecular Pharmacology Fast Forward. Published on July 28, 2009 as DOI: 10.1124/mol.109.058883
This article has not been copyedited and formatted. The final version may differ from this version.

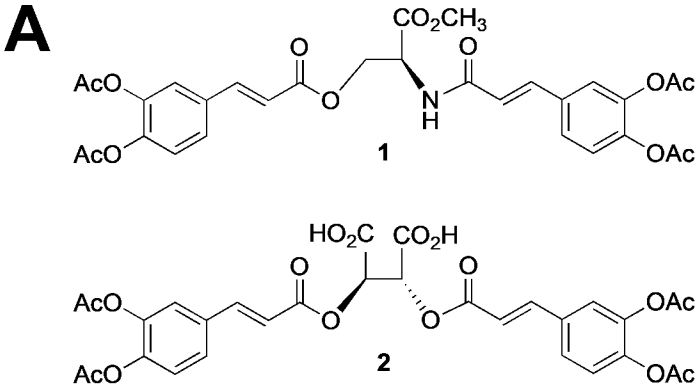
yellow. The location of Lys173 and height (H), depth (D) and width (W) for each cavity is indicated. C. The space-filling model for the CCD dimer-**1** complex showing that the inhibitor simultaneously interacts with both IN monomers (Green and Yellow). The central methyl ester group is deeply buried in the cavity and cannot be seen in this picture. Integrase active site residues (D64, D116 and E152) on each monomer are shown in red. D. Key interactions with IN residues established by the central methyl ester group and the linker region. The amino acids from green or yellow subunits are colored accordingly.

Figure 7. Biochemical analysis of IN mutants. A. Size exclusion chromatography of wild type and mutant proteins. Peaks corresponding to tetramer (Tet) IN with estimated molecular weighs of ~111 kDa and a dimeric (Dim) protein with estimate molecular weights of ~54 kDa, are indicated. B. Effects of amino acid substitutions on recombinant IN activities. Upper image depicts strand transfer activities. Positions of 21-mer substrate (21-S) and reaction products (STP) are indicated. Lower image displays 3'-processing activities. The positions of 21-S and specific 19-mer products (19-P) are shown. B. LEDGF binding to wild-type and mutant INs: lane 1: molecular weight markers, lane 2: LEDGF input, lane 3: assay without IN (non specific pull down control), lane 4: assay with wild type IN, lane 5: assay with E87A mutant, lane 6: assay with E96A mutant, lane 7: assay with Y99A mutant, lane 8: assay with K103A mutant.

Molecular Pharmacology Fast Forward. Published on July 28, 2009 as DOI: 10.1124/mol.109.058883
This article has not been copyedited and formatted. The final version may differ from this version.

Table 1. Binding energy results for docking **1** in the CCD dimer.

Conformation #	Binding Energy (kcal/mol)	Cavity selected
1	-5.30	magenta
2	-4.34	magenta
3	-3.81	magenta
4	-3.25	magenta
5	-2.69	magenta
6	-2.69	magenta/orange
7	-2.64	magenta
8	-2.09	magenta/orange
9	-1.49	magenta
10	-1.42	magenta


A
CO2CH3
O
O
AcO
AcO
O
N
H
OAc
OAc
1
HO2C
CO2H
O
O
AcO
AcO
O
O
OAc
OAc
2

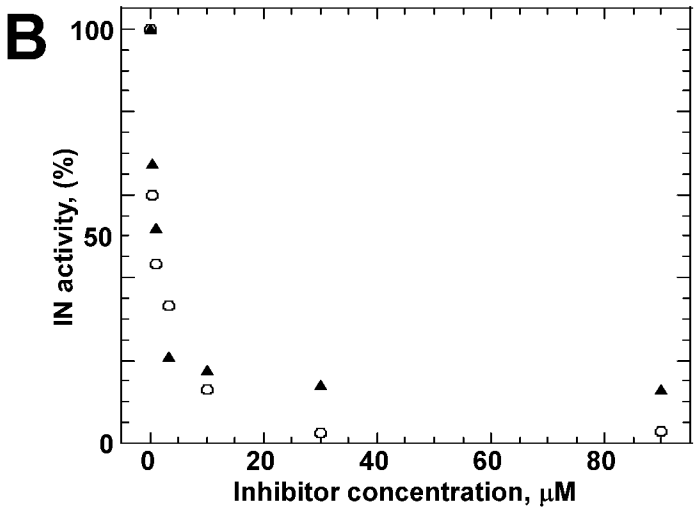

B
100
50
0
IN activity, (%)
0
20
40
60
80
Inhibitor concentration, μM

Figure 1

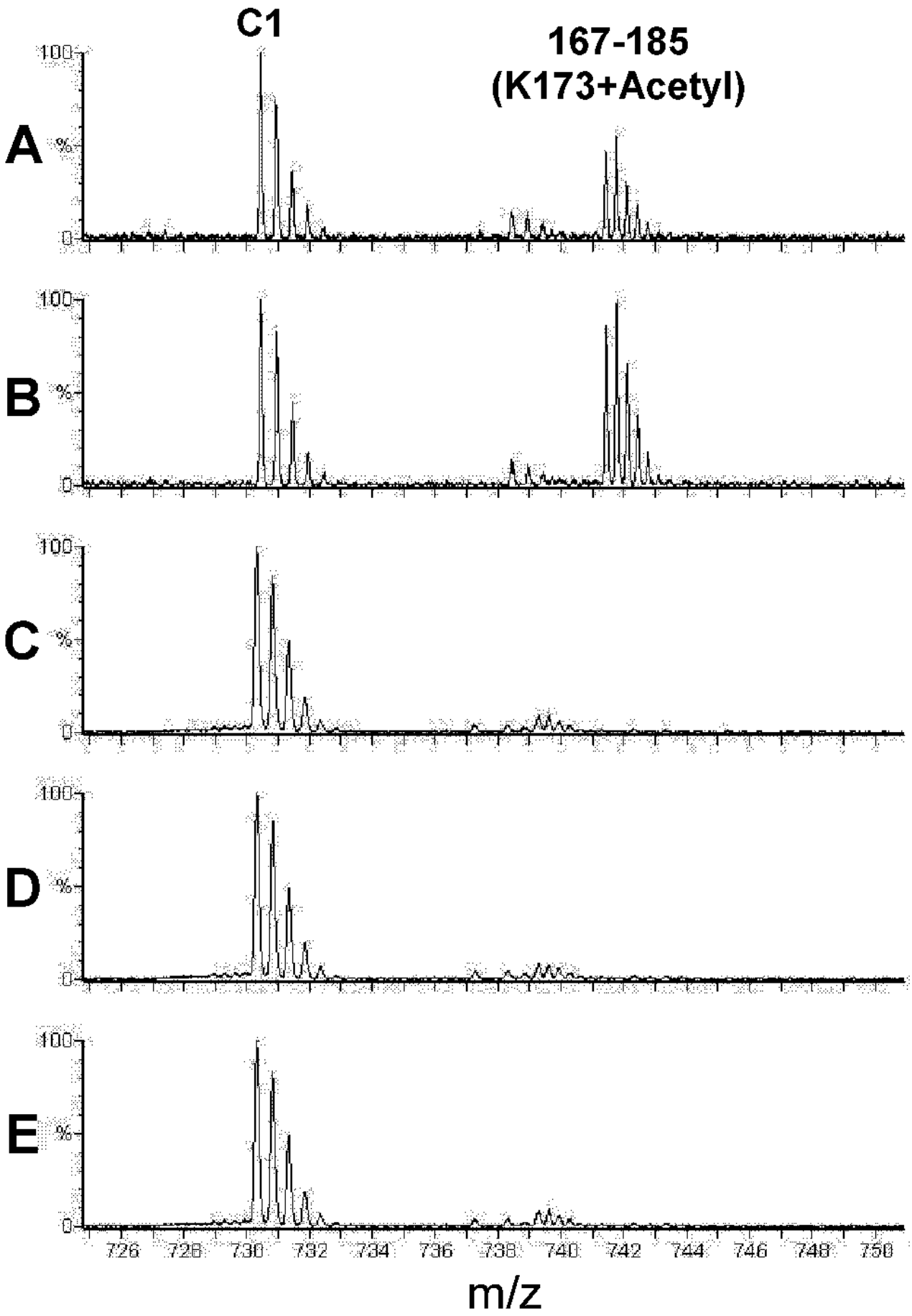

C1
167-185
(K173+Acetyl)
A
B
C
D
E
100
%
0
726
728
730
732
734
736
738
740
742
744
746
748
750
m/z

Figure 2

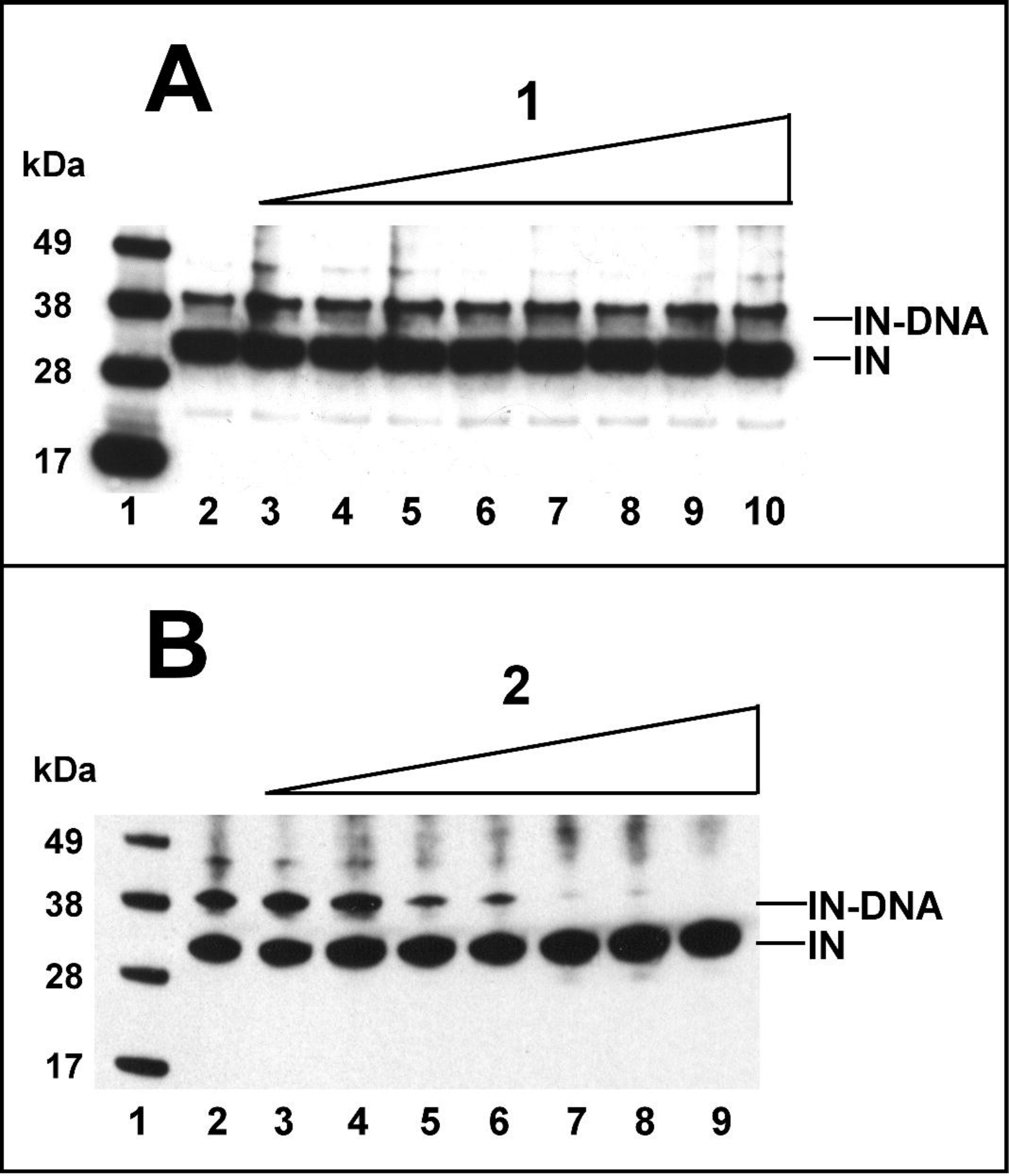

A
1
kDa
49
38
28
17
IN-DNA
IN
1 2 3 4 5 6 7 8 9 10
B
2
kDa
49
38
28
17
IN-DNA
IN
1 2 3 4 5 6 7 8 9

Figure 3

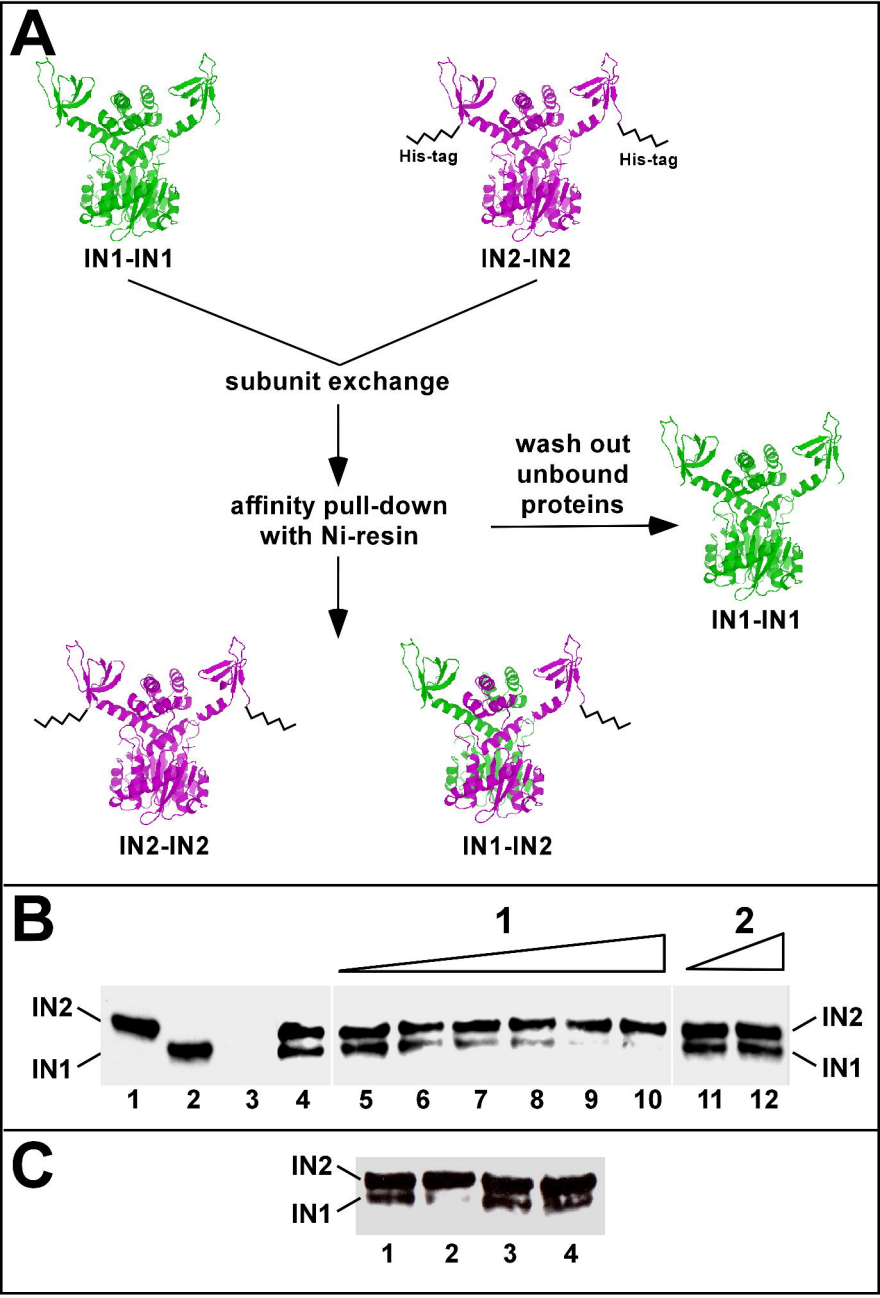

A
IN1-IN1
His-tag
His-tag
IN2-IN2
subunit exchange
affinity pull-down
with Ni-resin
wash out
unbound
proteins
IN1-IN1
IN2-IN2
IN1-IN2
B
1
2
IN2
IN1
IN2
IN1
1 2 3 4 5 6 7 8 9 10 11 12
C
IN2
IN1
1 2 3 4

Figure 4

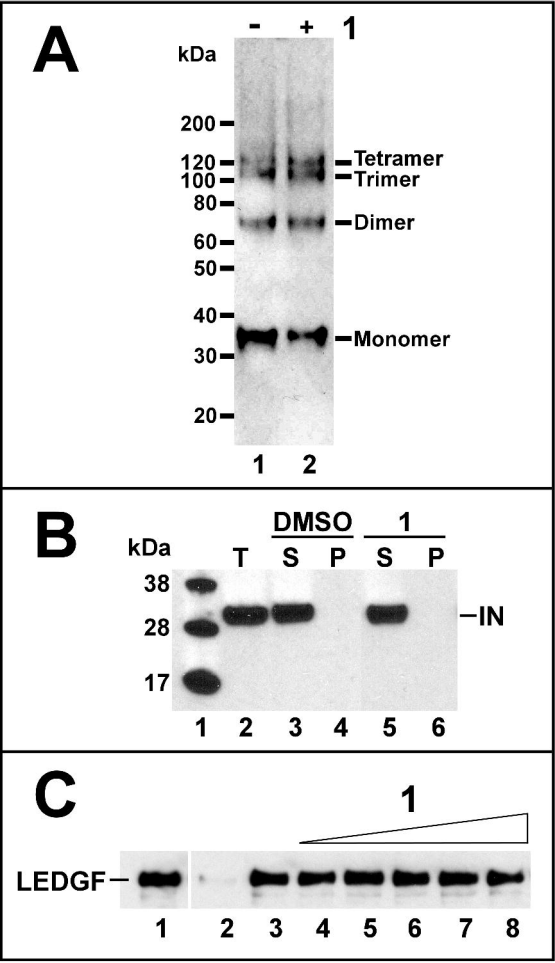

A
- + 1
kDa
200
120
100
80
60
50
40
30
20
Tetramer
Trimer
Dimer
Monomer
1 2
B
DMSO
1
kDa
T S P S P
38
28
17
IN
1 2 3 4 5 6
C
1
LEDGF
1 2 3 4 5 6 7 8

Figure 5

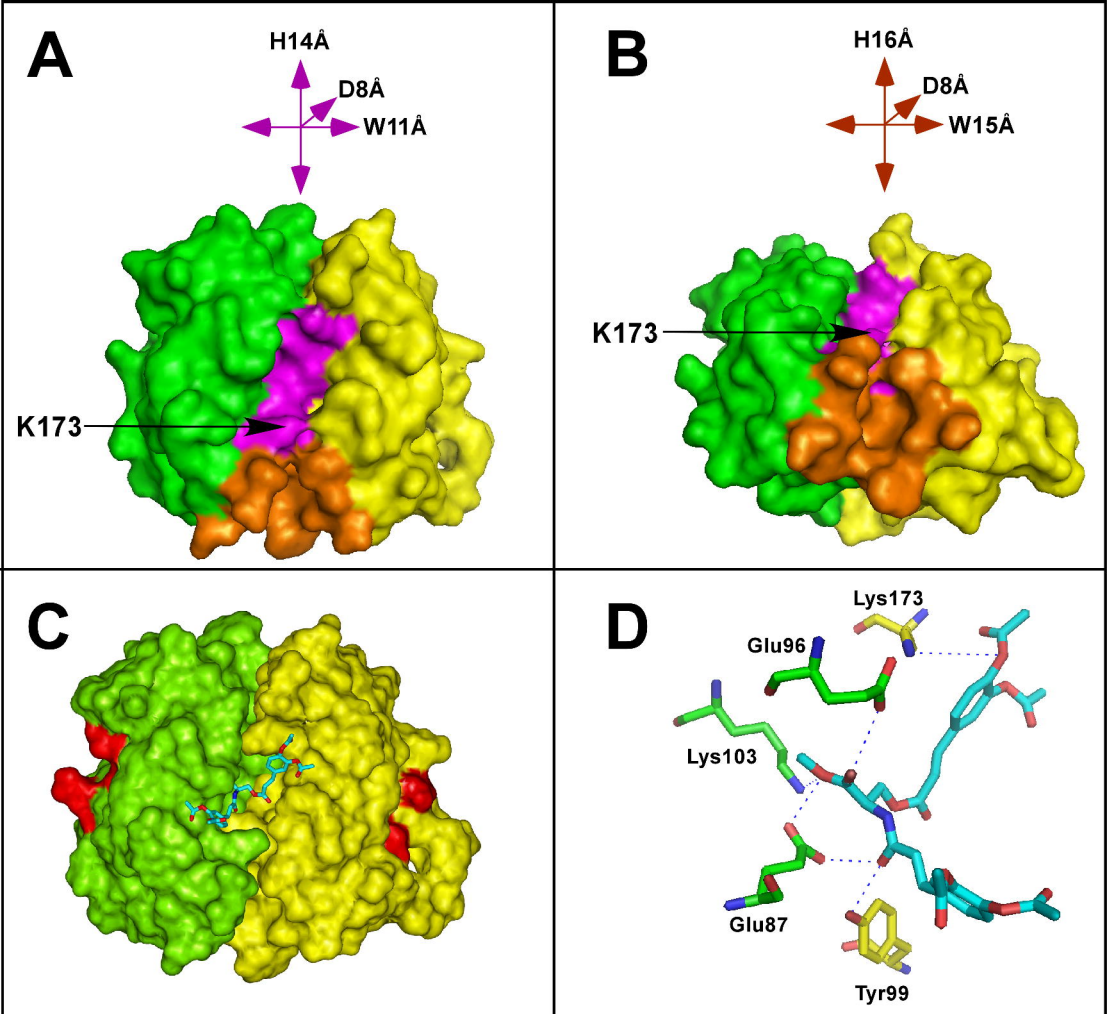

A
H14Å
D8Å
W11Å
K173
B
H16Å
D8Å
W15Å
K173
C
D
Lys173
Glu96
Lys103
Glu87
Tyr99

Figure 6

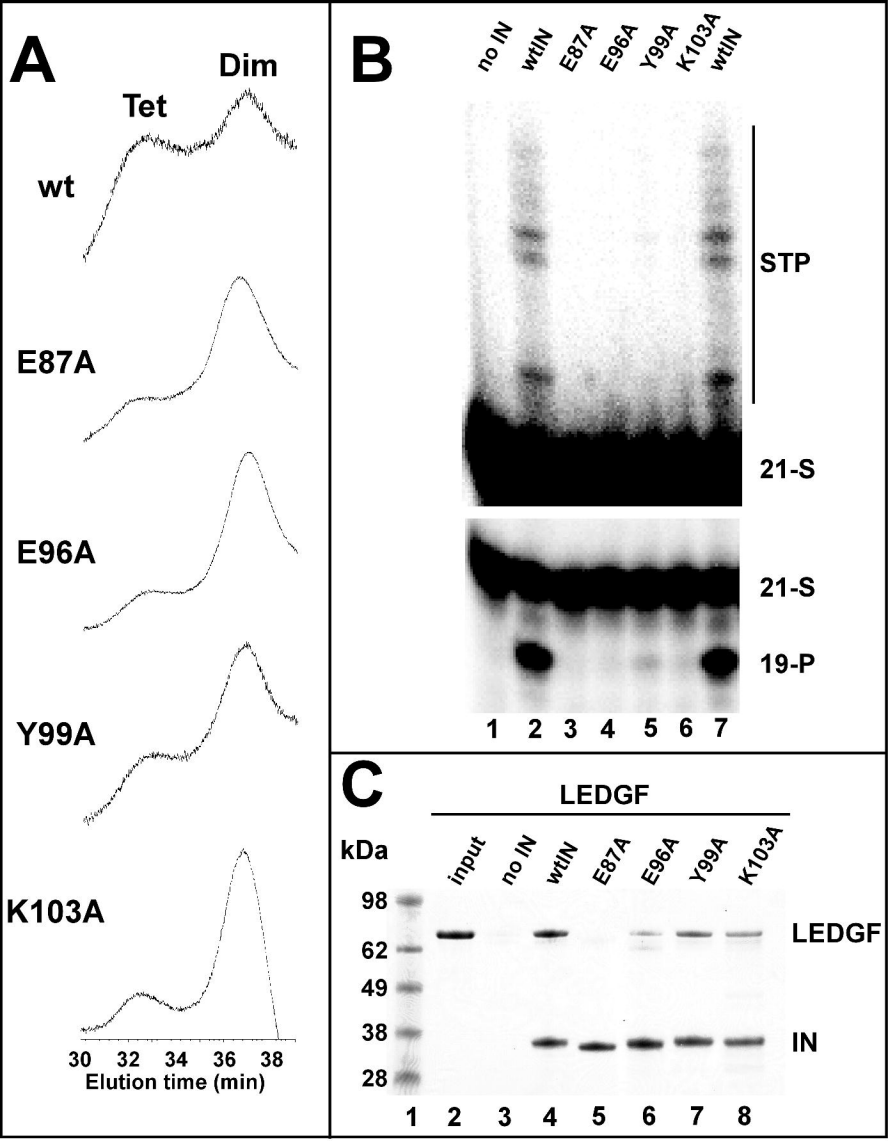

A
Dim
Tet
wt
E87A
E96A
Y99A
K103A
30 32 34 36 38
Elution time (min)
B
no IN
wtIN
E87A
E96A
Y99A
K103A
wtIN
STP
21-S
21-S
19-P
1 2 3 4 5 6 7
C
LEDGF
kDa
input
no IN
wtIN
E87A
E96A
Y99A
K103A
98
62
49
38
28
LEDGF
IN
1 2 3 4 5 6 7 8

Figure 7