Catalytically-active complex of HIV-1 integrase with a viral DNA substrate binds anti-integrase drugs

Akram Alian^a, Sarah L. Griner^a, Vicki Chiang^b, Manuel Tsiang^c, Gregg Jones^c, Gabriel Birkus^c, Romas Geleziunas^c, Andrew D. Leavitt^b, and Robert M. Stroud^{a,d,1}

Departments of ^aBiochemistry and Biophysics, ^bLaboratory Medicine, and ^cPharmaceutical Chemistry, University of California, San Francisco, CA 94158; and ^dGilead Sciences, Foster City, CA 94404

Edited by Patrick O. Brown, Stanford University School of Medicine, Stanford, CA, and approved March 31, 2009 (received for review November 22, 2008)

HIV-1 integration into the host cell genome is a multistep process catalyzed by the virally-encoded integrase (IN) protein. In view of the difficulty of obtaining a stable DNA-bound IN at high concentration as required for structure determination, we selected IN–DNA complexes that form disulfide linkages between 5′-thiolated DNA and several single mutations to cysteine around the catalytic site of IN. Mild reducing conditions allowed for selection of the most thermodynamically-stable disulfide-linked species. The most stable complexes induce tetramer formation of IN, as happens during the physiological integration reaction, and are able to catalyze the strand transfer step of retroviral integration. One of these complexes also binds strand-transfer inhibitors of HIV anti-viral drugs, making it uniquely valuable among the mutants of this set for understanding portions of the integration reaction. This novel complex may help define substrate interactions and delineate the mechanism of action of known integration inhibitors.

covalent | cross-linking | disulfide | DNA binding | strand transfer

Integration of the viral genome into host cell DNA is an obligate step in the HIV-1 life cycle. Integrase (IN), encoded by the *pol* gene, mediates this 2-step process. In the first step, termed 3′ processing, IN cleaves a distal dinucleotide adjacent to a conserved CA located at each 3′ end of the DNA copy of the viral genome. In the second step, termed strand transfer, IN covalently attaches the 3′ processed viral DNA to the host genome (1).

IN consists of 3 functional domains: the N-terminal domain (NTD; residues 1–51) that contains a conserved “HH-CC” zinc-binding motif, the catalytic core domain (CCD; residues 52–210) with the catalytic residues (D64, D116, and E152), and the C-terminal domain (CTD; residues 210–288) that contributes to DNA binding (2). In solution, recombinant IN exists in a dynamic equilibrium between monomers, dimers, tetramers, and higher-order oligomers (3, 4). Monomers are reportedly inactive *in vitro*, whereas dimers are able to catalyze 3′ processing and integration of 1 viral end (4–9). Tetramers, which have also been isolated from human cells expressing HIV-1 IN (10), can catalyze integration of 2 viral DNA ends into target DNA (7, 11), but the exact nature of the IN complex mediating 3′ processing and strand transfer reactions remains to be determined. The integration step is an attractive drug target given its essential role in the viral life cycle and the lack of a cellular IN homologue. Strand transfer inhibitors appear to bind significantly better to IN when it is assembled on its DNA substrate than to IN alone (12). To date there is only 1 structure of an inhibitor bound to IN (13), and that is in the absence of DNA. The compound binds at the active site; however, it dimerizes across a crystallographic 2-fold axis and therefore might not be in its bioactive configuration.

Structure-based understanding of the mechanisms of the action of IN inhibitors and optimization of compounds as potential drugs targeting HIV-1 IN have been hampered by the inability to capture and crystallize IN–DNA complexes. Two key factors have contributed to this problem: first, the high salt concentration (≈ 1 M NaCl) required to maintain full-length IN in solution interferes with DNA binding; second, IN has intrinsically low affinity for DNA. To

overcome these 2 obstacles, we used disulfide cross-linking to generate soluble, catalytically-active, covalent IN–DNA complexes. A similar strategy, covalent disulfide cross-linking between HIV-1 reverse transcriptase (RT) and DNA, mediated crystallization of the RT–DNA complex (14).

Previous cross-linking from cysteineal mutations in the CTD (6) and CCD (15) of IN with thiolated DNA substrates suggested that the CTD of 1 protomer of dimeric IN binds 1 end of viral DNA *in trans* with the CCD of the other protomer. However, while complexes were selected on the basis of IN–DNA cross-linking (6, 15), enzymatic activities of the covalent IN–DNA complexes were not reported.

Here, we describe an IN cysteine mutant, IN^{Y143C}, which is able to form IN–DNA complexes efficiently. The IN^{Y143C}–DNA complexes form stable tetramers in solution, retain single-end strand transfer activity, show increased resistance to protease and nuclease digestion, and bind a strand transfer inhibitor. This IN–DNA complex can serve as an *in vitro* platform to identify and evolve strand transfer inhibitors of HIV integration and as a means of understanding the basis for a key part of the integration reaction.

Results

Selection of Most Stable Disulfide Cross-Linked IN–DNA Complexes.

To trap IN–DNA complexes with a viral DNA substrate bound in a biologically-relevant manner, we used available IN structures (16–18) to guide the selection of sites for the introduction of cysteine residues near the active site. We started with IN^{C56S/W131D/F185D/C280S/C65S}, termed IN_P. This protein incorporates 4 previously-described mutations designed to diminish surface hydrophobicity for improved solubility (termed IN_O) (17) plus the introduction of C65S to avoid potential reactivity with the thiolated DNA. Hence, IN_P retains 3 cysteines: C130 and C40 and C43 of the zinc finger. Two clusters of mutant sites were chosen (Fig. 1A). Cluster 1 (D167C, Q164C, K160C, and L68C) is near K159, a residue previously shown to interact with the conserved 3′-A of the penultimate CA dinucleotide (underlined in Fig. 1B) in the viral DNA (19), and also near the catalytic triad (D64, D116, E152) (20). Cluster 2 (N117C, G118C, G140C, Y143C, and Q148C) is associated with a flexible loop that contributes to the IN active site (21). None of the 9 single-cysteine mutations significantly affected protein expression or the solubility of full-length or truncated IN_P constructs. All of those tested, including IN_O, the starting quadruple mutant, retained 64–90% WT activity for single-end strand transfer (Fig. S1).

To identify and isolate cross-linked complexes, each of the mutant proteins was incubated with 18/20 3′-processed DNA

Author contributions: A.A. designed research; A.A., S.L.G., V.C., G.J., and G.B. performed research; M.T. and R.G. contributed new reagents/analytic tools; A.A., A.D.L., and R.M.S. analyzed data; and A.A., M.T., R.G., A.D.L., and R.M.S. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

¹To whom correspondence should be addressed. E-mail: stroud@msg.ucsf.edu.

This article contains supporting information online at www.pnas.org/cgi/content/full/0811919106/DCSupplemental.

Fig. 3. Impact of DNA substrate on IN-DNA complex formation. (A) IN-DNA complex formation using 3'-processed (18/20) and unprocessed DNA (20/20) for full-length IN_p constructs, each containing the indicated cysteine mutation. Thiol groups are indicated as TNB. (B) Nonreducing SDS/PAGE (4–12%), stained with Coomassie blue, of cross-linking analysis of full-length IN_p^{Y143C} and IN_p^{Y143C}-DNA complex without homo-bifunctional cross-linkers (lanes 2 and 6), and in the presence of homo-bifunctional cross-linkers of 2 different spacer arm lengths, 6.4 Å (lanes 3 and 7) and 11.4 Å (lanes 4 and 8). Lanes 1 and 5 are molecular standards. M: monomer; C: IN-DNA complex; D: dimer. Molecular masses (kDa) were determined by electrophoretic mobility. #: masses are the mean of 3 independent experiments. (C) Calculated molecular masses (kDa) of full length IN_p and IN_p-DNA. Asterisks indicate calculated masses of IN_p-DNA as expected to show on SDS/PAGE (≈7 kDa for each DNA as observed on gels). Masses in parentheses are as observed on the gel in B. Mono: monomer (IN); Di: dimer (IN₂ or [IN₂-DNA]); Tet: tetramer (IN₄ or [IN₂-DNA]₂); Oct: octamer (IN₈ or [IN₂-DNA]₄).

determined the ability of IN_p^{K160C}, IN_p^{Q164C}, and IN_p^{Y143C} to bind unprocessed, blunt-ended, thiolated DNA substrates that mimic the true end-product of reverse transcription. Only IN_p^{Y143C} cross-linked with blunt-end DNA (Fig. 3A), suggesting that the flexible loop can maintain contact with the viral DNA during the proposed IN conformational change between the 3'-processing and strand transfer steps (6, 7, 12, 21), and residues K160 and Q164 come into close proximity with the viral end only during the strand transfer step.

To test the hypothesis that DNA binding triggers rearrangement of IN monomers within the DNA-bound IN oligomer, lysines in full-length IN_p^{Y143C} and IN_p^{Y143C}-DNA complexes were cross-linked by using homo-bifunctional cross-linking reagents with 2 different lengths of spacer arm, disuccinimidyl tartarate (DST) (6.4 Å) and bis(sulfosuccinimidyl)suberate (BS³) (11.4 Å). Whereas the IN_p^{Y143C}-DNA complex yielded tetramers and higher-order oli-

Fig. 4. DNA tethering overcomes the IN NTD requirement for strand transfer. Single-end strand transfer activity of CCD + CTD untethered IN_p^{K160C/52-288} (◇), K160C and IN_p^{Y143C/52-288} (◆, Y143C), and tethered complexes of IN_p^{K160C/52-288}-DNA (○, K160C/DNA) and IN_p^{Y143C/52-288}-DNA (●, Y143C/DNA). Error bars represent SEM of 3 independent experiments. RLU: relative luminescence units.

gomers when cross-linked using both short (DST) and long (BS³) spacer arms (Fig. 3B, lanes 7 and 8), apo IN_p formed tetramers and higher-order oligomers only when using the longer spacer arm (Fig. 3B, lane 3 vs. lane 4). The cross-linked apo IN_p and IN_p-DNA species migrated predominantly with molecular masses of 111 ± 7.7 and 126 ± 2.3 kDa, respectively (Fig. 3B). The difference is consistent with the addition of 2 DNA molecules (≈7 kDa each for 18/20 DNA as observed on SDS/PAGE gels) to each tetramer. Higher-order oligomers of ≈197 kDa for IN_p, and ≈239 kDa for IN_p-DNA, possibly octamers (Fig. 3C), were also visible. The electrophoretic mobility of Lys-cross-linked IN and Lys-cross-linked IN-DNA was slightly faster than expected relative to the noncross-linked species and the calculated molecular masses (Fig. 3C), a phenomenon that has been attributed to conformational constraints imposed by Lys cross-linking (11).

To further probe the conformational changes upon IN-DNA complex formation, we tested protease and nuclease protection (Fig. S2). Trypsin digestion of IN_p^{Y143C/52-288} (CCD plus CTD) yielded a small amount of a 16-kDa protease-resistant fragment (molecular mass determined by MALDI-TOF MS and SDS/PAGE). In contrast, trypsin digestion of the IN_p^{Y143C/52-288}-DNA complex yielded a much larger amount of a ≈20-kDa protease-resistant fragment. DTT treatment of the ≈20-kDa IN_p^{Y143C/52-288}-DNA fragment yielded a 16-kDa protein fragment that comigrates with the fragment from apo-IN. N-terminal sequencing of the 16-kDa protease-resistant fragment showed its amino terminus to be the same as that of the N terminus of IN_p^{Y143C/52-288}. Based on fragment size, the cleavage site is predicted to be within the flexible loop near residue R187 (Fig. 1A).

We also probed the nuclease resistance of the bound 18/20 substrate DNA. DNaseI digestion of the IN_p^{Y143C/52-288}-DNA protease-resistant ≈20-kDa fragment yielded a complex whose electrophoretic mobility suggests that ≈10 bp of IN-bound DNA is highly protected (Fig. S2). Proteolytic and nuclease protection was maintained under a variety of different salt, inhibitor, or purification conditions.

DNA Tethering Overcomes IN Requirement of the NTD for Strand Transfer. Lack of either the NTD or CTD of IN severely impairs strand transfer activity (5, 22). As expected, the truncated variants IN_p^{K160C/52-288} and IN_p^{Y143C/52-288} demonstrate very low strand transfer activity (Fig. 4). When donor DNA was cross-linked to these truncated mutants, strand transfer activity was increased by ≈5- and 10-fold, respectively (Fig. 4). The ≈40% lower strand transfer activity of cross-linked IN_p^{K160C/52-288}-DNA relative to

Fig. 5. Inhibitor binding to IN_p-DNA complexes. Binding of GS-9160 to cross-linked IN_p^{K160C/52-288}-DNA (diamonds, K160C/DNA) and IN_p^{Y143C/52-288}-DNA complexes (circles, Y143C/DNA) at 2 different immobilization concentrations of IN_p-DNA complexes: 500 nM (open symbols) and 1,500 nM (filled symbols). Error bars represent SEM of triplicates.

cross-linked IN_p^{Y143C/52-288}-DNA was correlated with a $\approx 25\%$ lower stoichiometry of DNA binding by IN_p^{K160C/52-288} (Fig. S3A). Whereas IN_p^{Y143C/52-288}-DNA forms dimers mixed with some tetramers and higher-order oligomers, with IN_p^{K160C/52-288}-DNA only dimers were detectable (Fig. S3B).

The Tethered IN_p^{Y143C/52-288}-DNA Complex Binds Strand Transfer Inhibitors. Having identified an IN-DNA complex that is capable of performing single-end strand transfer reaction (Fig. 4), we sought to ask whether it would bind currently used drugs that are reportedly inhibitors of strand transfer. We therefore validated a scintillation proximity assay (SPA) for measuring the affinity of a known strand transfer inhibitor, [benzene-2-³H]-GS-9160 (23), for noncross-linked IN-DNA complexes (Fig. S4). We next determined the affinity of the anti-IN drug GS-9160, a representative strand transfer inhibitor, for 2 of our cross-linked IN-DNA complexes, IN_p^{K160C/52-288}-DNA and IN_p^{Y143C/52-288}-DNA (Fig. 5). Although IN_p^{K160C/52-288}-DNA failed to bind the strand transfer inhibitor, IN_p^{Y143C/52-288}-DNA bound GS-9160 in a saturable manner with a K_d of ≈ 67 nM (the full-length WT IN K_d is 5.4 nM; Fig. S4), indicating that IN_p^{Y143C/52-288}-DNA contains a functional binding pocket for strand transfer inhibitor HIV antiviral drugs.

Discussion

We describe a functional complex of HIV-1 IN (IN_p^{Y143C}) covalently linked to a short dsDNA that matches the end of the viral DNA. The complex is able to catalyze single-end strand transfer reactions and binds to a strand transfer inhibitor only after complex formation. This complex can help define the basis of substrate binding and may therefore assist in the development of therapeutic inhibitors of IN-mediated integration.

Dynamic Nature of the IN-DNA Complex. Three of our observations support the previously-reported conformational change upon DNA binding and 3' processing (6, 7, 12, 21). First, of the 3 mutants assayed for their ability to bind both blunt-ended and 3'-processed DNA, Q164C and K160C effectively bound only the 3'-processed DNA, suggesting that the orientation between IN and the DNA substrate differs for the 2 substrates. Second, IN and IN-DNA complexes in solution showed distinct differences in the ability of homo-bifunctional Lys cross-linkers to cross-link IN molecules within soluble oligomeric complexes. Specifically, with DNA

bound, monomers of IN can be cross-linked by using shorter (6.4 Å) linker arms, demonstrating different interactions among IN molecules depending on whether they are engaged with DNA and suggesting that DNA binding induces some reorganization of the oligomer that brings lysine sites closer to each other as detected by cross-linking of IN protomers. Third, 3'-processed DNA binding induces changes that significantly protect a subdomain (residues 52–187) of the IN CCD from proteolysis.

Upon formation of the IN-DNA complex, the CCD of IN_p also protects the bound DNA (≈ 10 bp) from nuclease action. This result is consistent with the finding of others that IN binds to the 10 terminal base pairs of the viral DNA (24) that influence the efficiency of integration (11). Ten base pairs is approximately half of the DNA length protected by the full-length IN oligomer in previously-reported stable IN-DNA complexes (11, 25). From a biological perspective, having IN protect the viral DNA ends from nucleases strategically ensures the integrity of the viral ends, which are essential for viral integration and, thereby, for viral propagation.

Stoichiometry of the IN-DNA Complex. Under denaturing conditions (SDS/PAGE) our IN-DNA complexes consistently display approximately equal amounts of unbound IN_p and DNA-bound IN_p, consistent with a 2:1 ratio of IN to DNA (Fig. 1C, lanes 3 and 4). This 2:1 ratio pertains even when DNA was present in molar excess, suggesting that only 1 of the 2 DNA binding pockets in an IN dimer readily binds a viral DNA end. The 2:1 ratio is consistent with the findings of numerous studies including cross-linking (4, 6), complementation (5), fluorescence anisotropy (8, 9), and small-angle X-ray scattering (7), each of which indicate that a dimer of IN binds a single viral end and is sufficient for 3' processing and half-site integration. The (CCD + CTD) IN_p^{K160C/52-288}-DNA complex, which only formed dimers and was unable to form tetramers (Fig. S3B), was capable of catalyzing half-site integration (Fig. 4).

The 2 IN active sites lie on opposite sides of the IN dimer (17, 18, 26) such that DNA binding to 1 active site would not occlude the other site from binding a second DNA. However, there is evidence that binding of viral DNA to only 1 face of the dimer induces asymmetry in the dimer, detected by a decrease in exposure of Arg-199 (located in the linking helix between CCD and CTD) of the DNA-bound monomer, and an increase in exposure of Arg-199 on the other, relative to unbound IN dimers (15). The asymmetry could therefore be the basis for inactivating the other monomer in the dimer and formation of a so-called tetramer by dimerization of the heterodimer.

Mechanism of Assembly of Active IN-DNA Tetramers. DNA-induced conformational changes have been suggested to play a role in promoting IN-DNA tetramerization required for concerted integration (8). Different lines of evidence support the sequential assembly of the tetrameric DNA complexes from DNA-bound dimers. First, cross-linked IN tetramer does not bind DNA directly; rather, IN-DNA tetramers form only by the interaction between 2 DNA-bound IN dimers (4). Second, small peptides have been described that inhibit DNA binding to IN by shifting the IN oligomerization equilibrium from an active dimer toward an inactive tetramer (9). Third, DNA binding induces dissociation of the multimeric IN (27). Taken together, these observations suggest that a tetrameric form of the apo IN must dissociate to bind DNA and then reorganize into a tetrameric IN-DNA complex. Consistent with this hypothesis, we find that upon DNA binding to IN_p, a higher-order oligomer of ≈ 197 kDa (Fig. 3B, lane 4) is replaced with a lesser amount of distinctly different oligomers, one corresponding to a DNA-bound oligomer of ≈ 239 kDa (Fig. 3B, lane 8) and the other to a DNA-induced dimeric IN_p-DNA of ≈ 64 kDa (Fig. 3B, lanes 4 vs. lane 8). Interestingly, our data show that tetramer formation does not require the NTD as indicated by the ability of the IN_p^{Y143C/52-288}-DNA to form tetramers.

We propose that only after the viral DNA has undergone 3'

Fig. 6. DNA binding by IN dimer. A model of IN dimer (1 protomer in blue and 1 in yellow) bound to DNA (orange ribbon). The 6C-thiol linker (≈ 9 -Å length) is attached to the indicated 5' end of the DNA. Dashed lines represent distances between the 5' end and K160 (C α in red sphere, 15Å) or Y143 (C α in red sphere, 23Å). CCD flexible loop spanning Y143 (residues 139–150) is colored red. The inhibitor (L-870810, yellow sticks) and the definition of inhibitor binding pocket (magenta surface) have been described (30). Model of viral DNA bound to IN dimer has been described (15), and coordinates were generously obtained from M. Kvaratskhelia (Ohio State University, Columbus, OH). Loop coordinates were obtained from the CCD structure (1BI4) (16). The figure was made with Pymol (DeLano Scientific).

processing by an IN dimer does the dimer complex assume the conformation required for interaction with a second IN–DNA dimer of similar conformation, which is required to form the tetramer complex essential for concerted integration. Such a specific sequence of events required to generate IN–DNA tetramers may have biological implications for the retroviral life cycle by preventing unwanted single-ended integration events that could be a dead end for the replicating virus. Sequential reorganization of the IN–DNA complex has been suggested to channel the integration reaction along the correct pathway toward concerted integration (11).

The IN–DNA Dimer. There are 2 possible modes of viral DNA binding within active IN tetramers consistent with current models. In 1 mode, the strand with the 3' end that undergoes 3' processing would bind to an active site from 1 dimer, while the other strand is somehow unwound from the helix such that the complementary 5' end binds to another active site in a second dimer *in trans* (28–31). In the second mode, the complementary 5' and 3' ends remain double-stranded and bind to a single active site of a single dimer (15). Our observed half-site integration activity of the cross-linked $\text{IN}_p^{\text{K160C/52-288}}$ -DNA, cross-linked to DNA through its 5' end, which is essentially all dimer as assayed by size exclusion chromatography and MALLS, requires that the 3' end also be in the active site of the same monomer. With the slight caveat that possible transient tetramers of $\text{IN}_p^{\text{K160C/52-288}}$, although undetected, might carry out the observed half-site integration, our results support the case that each end of the viral DNA duplex (with its 3' and complementary 5' end) bind to 1 active site of a dimer.

A plausible model for the IN dimer bound to 1 DNA duplex, in which the viral DNA is stabilized by interactions to the CTD and NTD and its 3'-CA bound to the CCD of the same dimer, has been described (15). The model illustrates how the 5'-thiol group of the modified DNA is in close proximity to and may cross-link to either K160C or Y143C while at the same time the 3' end of the DNA remains at the active site (Fig. 6). The length of the 6-carbon thiol linker (≈ 9 Å), the flexibility of the DNA 5' overhang, the flexibility of the active site loop (Y143C) that may swing nearly 14 Å (21), and the conformational changes in the DNA and IN upon association

are expected to allow the thiol group access to either of the cysteines.

The Drug Binding Pocket. Viral DNA 3' end processing by IN is proposed to create a hydrophobic drug binding pocket in the space vacated by the removal of 3'-GT, bounded by the flexible active site loop (30) and the 5'-CA overhang of the viral DNA (32). Inhibitor binding is attenuated when either the flexibility of the active site loop is impaired (33) or the 5'-CA overhang of the viral DNA is deleted (32). Although the thiol group on the 5'-CA overhang cross-links efficiently to either K160C or Y143C, only cross-linked Y143C binds IN inhibitor. We suggest that 5' cross-linking to Y143C retains enough flexibility to allow formation of the natural binding pocket that is bounded by the 5'-CA overhang and active site loop, whereas 5'-CA cross-linking to K160C constrains the 5'-CA overhang from moving toward the inhibitor to contribute to the binding pocket. Although 5'-CA cross-linking to K160C attenuates drug binding, it does not seem to affect the position of the 3' end in the catalytic site as indicated by strand transfer activity of $\text{IN}_p^{\text{K160C/52-288}}$ -DNA. The cross-linking data presented add to the repertoire of IN–DNA modeling constraints, which have so far been insufficient to generate a unified model.

In conclusion, we harnessed a series of IN cysteine mutations and selective reducing conditions to identify a stable IN–DNA complex, $\text{IN}_p^{\text{Y143C}}$ -DNA, that retains its ability to catalyze half-site integration and contains a binding site for a validated strand transfer inhibitor. This complex can serve as a platform for structural analysis and optimization of drug candidates that target integration of HIV.

Experimental Procedures

Mutagenesis. Mutations C56S, W131D, F185D, and C280S were introduced into a synthetic full-length HIV-1 IN (SF1), termed IN_0 (Quadra mutant), and cloned into pT7-7 (17). We removed an additional native cysteine, C65, to generate $\text{IN}_p^{\text{C56S/W131D/F185D/C280S/C65S}}$, termed IN_p (Penta mutant). An N-terminal 6-histidine (6-His) tag followed by a thrombin cleavage site (Stratagene) was added for purification. $\text{IN}_p^{\text{52-210}}$, $\text{IN}_p^{\text{1-210}}$, and $\text{IN}_p^{\text{52-288}}$ were generated from $\text{IN}_p^{\text{1-288}}$. Additional cysteine mutations were introduced by site-directed mutagenesis using the QuikChange Kit (Stratagene). All constructs were confirmed by DNA sequencing.

Expression and Purification. IN_p constructs were expressed and purified as described (17) except for 1 M NaCl for the full-length protein. Proteins were dialyzed into a final solution consisting of 2 mM DTT, 20 mM Tris (pH 8.0), 0.1 mM ZnCl₂, and 1 M NaCl for the full-length IN_p or 0.5 M NaCl for the truncated variants.

Oligodeoxynucleotides. The thiol group was located at the end of a 6-carbon linker attached to a phosphate at the 5' end of the oligonucleotide (5'-Xactgctagagattttccaca-3' [X: 5'-C6 Thiol linker] (TriLink). Complementary sequences of either 18 bp (5'-tgtggaaaattcttagca-3') or 20 bp (5'-tgtggaaaattcttagcagt-3') were annealed by slow cooling after 3 min at 95 °C. After deprotection and activation (40 mM DTT, 170 mM phosphate buffer, pH 8.0, at 37 °C for 16 h), the thiolated dsDNAs were desalted by using G-25 Microspin columns (GE Healthcare Bio-Sciences) and treated with 1 mM 5,5'-dithio-bis(2-nitrobenzoic acid) (DTNB) (Sigma) at room temperature for 1 h (6), followed by a second desalting.

Reaction Conditions to Form IN–DNA Complexes. Reaction mixtures contained IN_p and thiolated-DNA (in molar excess) in 50 mM Tris (pH 8.0), ≈ 2 mM DTT (supplemented by the protein solution), 10 mM MnCl₂, and 0.5M NaCl for full-length IN_p or 0.15 M NaCl for 1- and 2-domain IN_p . To allow the assembly of IN_p on the donor DNA under mild reducing conditions, MnCl₂ was added at the final step to help oxidize the reducing agent and initiate the cross-linking reaction (≈ 1 h at room temperature).

Evaluation of IN_p -DNA Complex Formation. Complex formation was determined by nonreducing SDS/PAGE (4–12% Bis-Tris Gel; Invitrogen). Gels were scanned and analyzed with Scion Image. The percentage of complex formation = density of IN_p -DNA band / (density of IN_p monomer band + density of IN_p -DNA band).

IN_p -DNA Complex Purification. IN_p -DNA complexes were purified by size exclusion chromatography (SEC) using G3000SW TSK-Gel columns (Tosoh Bioscience).

Elution buffers consisted of 20 mM Hepes, pH 7.0, and 1 M NaCl for IN_p^{1–288} and IN_p^{1–288}-DNA complex or 0.5 M NaCl for the IN_p^{52–288} and IN_p^{52–288}-DNA complex. The SEC chromatograph, from which the IN_p-DNA complex peak was taken for further light scattering analysis, is shown in Fig. S5.

Static Light Scattering. Static light scattering was performed by using Shodex Protein KW-803 (Thomson Instrument), attached to 3 detectors in series: a Thermo System UV1000 absorbance detector (at 280 nm), a Wyatt Technologies DAWN HELEOS MALLS detector, and a Wyatt Technologies Optilab rEX refractive index detector. The refractive index increment (dn/dc) of 0.185 was used for all analysis of IN_p and IN_p-DNA. MALLS results were reproduced and confirmed by an independent analysis performed by Alliance Protein Laboratories. Elution buffers for light scattering are identical to those used for SEC. MALLS machine error is 0.3–5%. Standard deviations were from the indicated different sample concentrations.

IN Single-End Strand Transfer by SPA. The 2 strands of the 3'-processed donor DNA used in the strand transfer assay were: LTR1, 5'-Biotin-ACCTTTTGTAGT-CAGT-GTGGAAATCTCTAGCA-3' and LTR2, 3'-GAAATCATGTCACACCTTTT-GAGATCGTCA-5'. The 2 strands of the DIG-tagged target DNA were: TargetDIG1, 5'-TCGACCAAGGGCTAATCACT-DIG-3' and TargetDIG2, 3'-DIG-ACTGGTTC-CGATTAAGTGA-5'. The standard strand transfer assay has been described (34). Briefly, biotinylated 3'-processed donor DNA was used to coat Reacti-Bind High Binding Capacity Streptavidin (SA)-coated white plates (0.14 μ M) and was then incubated with IN_p (0.1–1 μ M). The luminescence (DIG)-tagged target DNA (0.25 μ M) was added to start the strand transfer reaction (30 min at 37 °C). Reactions contained 20 mM Hepes (pH 7.3), 10 mM DTT, 75 mM NaCl, 10 mM MgCl₂, 1% glycerol, 0.1–1 μ M IN, and 0.25 μ M target DNA. After washing, integrated target

DNA was detected by ELISA. To assay cross-linked complexes, IN_p^{K160C/52–288} and IN_p^{Y143C/52–288} were cross-linked to biotinylated 3'-processed donor DNA (18/20 with Biotin on the 5' end of the strand containing the conserved CA). After SEC purification, the biotinylated complexes were used to coat SA plates. Strand transfer conditions (without DTT) and detection of joined target DNA are identical to the standard assay.

Inhibitor Binding by SPA. Biotinylated IN_p^{K160C/52–288}-DNA and IN_p^{Y143C/52–288}-DNA complexes, purified by SEC (Fig. S3B, peak c), were directly bound to SA SPA beads (Amersham) at either 500 nM or 1500 nM in binding buffer [27.8 mM Hepes (pH 7.8), 27.8 mM MnCl₂, 111.1 μ g/mL BSA, and 67 mM NaCl] for 2 h at 22 °C with rocking. Beads were treated as in the standard SPA assay (12) (detailed in SI Text). The IN_p-DNA-beads were resuspended to 5 mg/mL in binding buffer and incubated in 96-well plates (12 h at 22 °C with shaking) with inhibitor. The inhibitor was [benzene-2-³H]-GS-9160 or compound 9 (23) that was synthesized by Moravsek Biochemicals (20 Ci/mmol specific activity). Plates were read on a Top Count scintillation counter (PerkinElmer). Binding data were analyzed by curve fitting using the exact binding equation to extract K_d values (SI Text).

Lysine Cross-Linking. Short spacer-arm (6.4 Å) DST and long spacer-arm (11.4 Å) BS³ (Pierce) were added to SEC-purified IN_p^{Y143C/1–288} and IN_p^{Y143C/1–288}-DNA to a final concentration of 0.5–1 mM ($\approx$ 1 h at room temperature).

ACKNOWLEDGMENTS. We thank Dr. Demetri Moustakas for assistance in early-stage modeling of the IN-DNA complexes for mutant design, Dr. Daniel Southworth for assistance in MALLS analysis, and Patricia Greene and Janet Finer-Moore for critical reading of the manuscript. This work was supported by the National Institutes of Health Grant P50 GM082250 via the HARC Center (to R.M.S.).

- Kukolj G, Skalka AM (1995) Enhanced and coordinated processing of synapsed viral DNA ends by retroviral integrases in vitro. *Genes Dev* 9:2556–2567.
- Brown PO (1997) in *Retroviruses*, eds Coffin JM, Hughes SH, Varmus HE (Cold Spring Harbor Laboratory, Cold Spring Harbor, NY), pp 161–203.
- Deprez E, et al. (2000) Oligomeric states of the HIV-1 integrase as measured by time-resolved fluorescence anisotropy. *Biochemistry* 39:9275–9284.
- Faure A, et al. (2005) HIV-1 integrase cross-linked oligomers are active in vitro. *Nucleic Acids Res* 33:977–986.
- Engelman A, Bushman FD, 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.
- Gao K, Butler SL, Bushman F (2001) Human immunodeficiency virus type 1 integrase: Arrangement of protein domains in active cDNA complexes. *EMBO J* 20:3565–3576.
- Baranova S, et al. (2007) Small-angle X-ray characterization of the nucleoprotein complexes resulting from DNA-induced oligomerization of HIV-1 integrase. *Nucleic Acids Res* 35:975–987.
- Guiot E, et al. (2006) Relationship between the oligomeric status of HIV-1 integrase on DNA and enzymatic activity. *J Biol Chem* 281:22707–22719.
- Hayouka Z, et al. (2007) Inhibiting HIV-1 integrase by shifting its oligomerization equilibrium. *Proc Natl Acad Sci USA* 104:8316–8321.
- Cherepanov P, et al. (2003) HIV-1 integrase forms stable tetramers and associates with LEDGF/p75 protein in human cells. *J Biol Chem* 278:372–381.
- Li M, Mizuuchi M, Burke TR, Jr, Craigie R (2006) Retroviral DNA integration: Reaction pathway and critical intermediates. *EMBO J* 25:1295–1304.
- Grobler JA, et al. (2002) Diketo acid inhibitor mechanism and HIV-1 integrase: Implications for metal binding in the active site of phosphotransferase enzymes. *Proc Natl Acad Sci USA* 99:6661–6666.
- Goldgur Y, et al. (1999) Structure of the HIV-1 integrase catalytic domain complexed with an inhibitor: A platform for antiviral drug design. *Proc Natl Acad Sci USA* 96:13040–13043.
- Huang H, Chopra R, Verdine GL, Harrison SC (1998) Structure of a covalently trapped catalytic complex of HIV-1 reverse transcriptase: Implications for drug resistance. *Science* 282:1669–1675.
- Zhao Z, et al. (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.
- Maignan S, Guilloteau JP, Zhou-Liu Q, Clement-Mella C, Mikol V (1998) Crystal structures of the catalytic domain of HIV-1 integrase free and complexed with its metal cofactor: High level of similarity of the active site with other viral integrases. *J Mol Biol* 282:359–368.
- Chen JC, et al. (2000) Crystal structure of the HIV-1 integrase catalytic core and C-terminal domains: A model for viral DNA binding. *Proc Natl Acad Sci USA* 97:8233–8238.
- Wang JY, Ling H, Yang W, 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.
- Jenkins TM, Esposito D, Engelman A, Craigie R (1997) Critical contacts between HIV-1 integrase and viral DNA identified by structure-based analysis and photo-cross-linking. *EMBO J* 16:6849–6859.
- Engelman A, Craigie R (1992) Identification of conserved amino acid residues critical for human immunodeficiency virus type 1 integrase function in vitro. *J Virol* 66:6361–6369.
- Lee MC, Deng J, Briggs JM, Duan Y (2005) Large-scale conformational dynamics of the HIV-1 integrase core domain and its catalytic loop mutants. *Biophys J* 88:3133–3146.
- Schauer M, Billich A (1992) The N-terminal region of HIV-1 integrase is required for integration activity, but not for DNA binding. *Biochem Biophys Res Commun* 185:874–880.
- Jin H, et al. (2008) Tricyclic HIV integrase inhibitors: Potent and orally bioavailable C5-aza analogs. *Bioorg Med Chem Lett* 18:1388–1391.
- Vora A, Grandgenett DP (2001) DNase protection analysis of retrovirus integrase at the viral DNA ends for full-site integration in vitro. *J Virol* 75:3556–3567.
- Vora A, Bera S, Grandgenett D (2004) Structural organization of avian retrovirus integrase in assembled intasomes mediating full-site integration. *J Biol Chem* 279:18670–18678.
- Dyda F, et al. (1994) Crystal structure of the catalytic domain of HIV-1 integrase: Similarity to other polynucleotidyl transferases. *Science* 266:1981–1986.
- Deprez E, et al. (2001) DNA binding induces dissociation of the multimeric form of HIV-1 integrase: A time-resolved fluorescence anisotropy study. *Proc Natl Acad Sci USA* 98:10090–10095.
- Wielens J, Crosby IT, Chalmers DK (2005) A three-dimensional model of the human immunodeficiency virus type 1 integration complex. *J Comput Aided Mol Des* 19:301–317.
- Chen A, Weber IT, Harrison RW, Leis J (2006) Identification of amino acids in HIV-1 and avian sarcoma virus integrase subsites required for specific recognition of the long terminal repeat ends. *J Biol Chem* 281:4173–4182.
- Chen X, et al. (2008) Modeling, analysis, and validation of a novel HIV integrase structure provide insights into the binding modes of potent integrase inhibitors. *J Mol Biol* 380:504–519.
- Dolan J, Chen A, Weber IT, Harrison RW, Leis J (2009) Defining the DNA substrate binding sites on HIV-1 integrase. *J Mol Biol* 385:568–579.
- Dicker IB, et al. (2007) Changes to the HIV long terminal repeat and to HIV integrase differentially impact HIV integrase assembly, activity, and the binding of strand transfer inhibitors. *J Biol Chem* 282:31186–31196.
- Greenwald J, Le V, Butler SL, Bushman FD, Choe S (1999) The mobility of an HIV-1 integrase active site loop is correlated with catalytic activity. *Biochemistry* 38:8892–8898.
- Yu F, et al. (2007) HIV-1 integrase preassembled on donor DNA is refractory to activity stimulation by LEDGF/p75. *Biochemistry* 46:2899–2908.