Structure-based modeling of the functional HIV-1 intasome and its inhibition

Lavanya Krishnan, Xiang Li, Hema L. Naraharisetty, Stephen Hare, Peter Cherepanov, and Alan Engelman

The intasome is the basic recombination unit of retroviral integration, comprising the integrase protein and the ends of the viral DNA made by reverse transcription. Clinical inhibitors preferentially target the DNA-bound form of integrase as compared with the free protein, highlighting the critical requirement for detailed understanding of HIV-1 intasome structure and function. Although previous biochemical studies identified integrase residues that contact the DNA, structural details of protein–protein and protein–DNA interactions within the functional intasome were lacking. The recent crystal structure of the prototype foamy virus (PFV) integrase–viral DNA complex revealed numerous details of this related integration machine. Structures of drug-bound PFV intasomes moreover elucidated the mechanism of inhibitor action. Herein we present a model for the HIV-1 intasome assembled using the PFV structure as template. Our results pinpoint previously identified protein–DNA contacts within the quaternary structure and reveal hitherto unknown roles for Arg20 and Lys266 in DNA binding and integrase function. Models for clinical inhibitors bound at the HIV-1 intasome active site were also constructed and compared with previous studies. Our findings highlight the structural basis for HIV-1 integration and define the mechanism of its inhibition, which should help in formulating new drugs to inhibit viruses resistant to first-class compounds.

HIV/AIDS | integrase | integration | drug resistance | raltegravir

The integration of the linear DNA molecule made during reverse transcription into a cell chromosome is an essential step in the retroviral lifecycle. The key players in retroviral DNA integration are the integrase (IN) protein and the LTR end regions of the reverse transcript. Together with other viral and cellular factors, they comprise a high-molecular-weight preintegration complex (PIC) wherein IN catalyzes two successive chemical reactions. During 3′ processing, IN hydrolyzes two or three nucleotides from one or both DNA ends, exposing invariant CA dinucleotides. In the nucleus, IN catalyzes DNA strand transfer using the reactive 3′-OHs to cut phosphodiester bonds on opposing strands of target DNA, at the same time joining the LTR ends to the chromosome. The resulting strand transfer reaction intermediate is repaired by cellular enzymes to yield the integrated provirus (see ref. 1 for a comprehensive review).

Retroviral INs and other nucleic acid metabolizing enzymes, including RNase H, RuvC, and bacterial and eukaryotic transposases such as Tn5 and Mos1, respectively, belong to the polynucleotidyl transferase superfamily (2–4). Common features among these proteins include an RNase H-like fold adopted by their catalytic domains and divalent metal ion-dependent bimolecular nucleophilic substitution (S_N2) reactions at nucleic acid phosphodiester bonds mediated by carboxylate side chains of active-site residues. The retroviral INs minimally harbor two functional domains in addition to their catalytic core domain (CCD), the N-terminal domain (NTD) that folds into a helical bundle coordinating a Zn atom via invariant His and Cys residues and the C-terminal domain (CTD), which adopts an SH3 fold and binds nonspecifically to DNA (reviewed in ref. 1).

The integration activity of PICs extracted from infected cells persists when the complexes are treated with high salt (5, 6), and the term “intasome” was coined to describe the PIC substructure that catalyzes integration in vitro (6, 7). Integration can furthermore be reconstituted from purified components, minimally requiring IN protein and DNA mimics of the viral LTR ends (1). Integration in vitro proceeds through a series of functional IN–DNA complexes. An initial stable synaptic complex (SSC) consists of an IN tetramer and two DNA end substrates (8). 3′ Processing converts the SSC to the cleaved donor complex (CDC), which is functionally analogous to cell-extracted PICs because these invariably harbor processed termini (9, 10). Subsequent target DNA capture and integration generates the strand transfer complex (STC) (8, 11), which shares key attributes of viral PICs, including resistance to high-ionic-strength conditions. Herein, “intasome” refers to CDCs derived from virus infection or assembled from purified components.

Because of its essential nature, HIV-1 IN has been extensively studied as an antiviral drug target, and raltegravir (RAL), which preferentially inhibits DNA strand transfer activity, was licensed for the clinic in 2007 (12). Other compounds in clinical trials, typified by elvitegravir (EVR), likewise inhibit DNA strand transfer activity, defining the IN strand transfer inhibitor (INSTI) class of antiviral drugs (13). Although the 3D structure of each IN domain as well as NTD-CCD (14) and CCD-CTD (15) two-domain fragments have been solved, to date there is no structure for the active HIV-1 intasome. Because INSTIs preferentially bind to and inhibit the IN–DNA complex as compared with free IN (16), the elusive 3D structure would predictably play an important role in developing second-generation INSTIs to inhibit viruses resistant to front-line compounds. Available IN fragment structures have revealed highly variable interdomain (NTD-CCD and CCD-CTD) linker conformations (14, 15, 17–19), and DNA binding is moreover required for the active site to assume its functional state. Consequently, intasome modeling based on partial HIV-1 IN structures has been less than straightforward and resulted in numerous, nonconverging results (20–25). We recently reported the X-ray crystal structure of the prototype foamy virus (PFV) intasome together with chemically inhibited structures in the presence of RAL or EVG (26). Taking advantage of the relatively close relationship among retroviruses, we construct here a set of realistic models for the HIV-1 intasome in its active and INSTI-inhibited forms.

Results and Discussion
HIV-1 Intasome Construction. The previous two-domain NTD-CCD (14) and CCD-CTD (15) crystal structures have been widely used...
to construct HIV-1 IN–DNA models. Some approaches relied on results of protein–DNA cross-linking (20, 27) or protein modification (24) experiments as guides for DNA placement. Others used the Tn5 transpososome (3), which until recently was the most closely related nucleoprotein complex crystal structure, as a modeling template (21, 23, 28). Because PFV IN is much more closely related to HIV-1 IN than is Tn5 transposase, we have modeled the HIV-1 intasome on the basis of the recent PFV X-ray crystal structure (26).

The crystallographic asymmetric unit contained a PFV IN dimer bound to one molecule of substrate DNA, and a pair of symmetry-related dimers formed an oblong tetramer with two synapsed DNA ends (26). Within the asymmetric unit, a near fully resolved IN monomer (residues 10–374) contacted the DNA, whereas the N-terminal extension domain (NED; see Fig. S1A for a PFV/HIV-1 IN structure-based alignment), NTD, and CTD of the second monomer were unseen in electron density maps (26). Structure determination of MnCl₂-soaked crystals furthermore revealed two metal ions coordinated at the DNA-bound active site. The HIV-1 model was constructed in a step-wise fashion starting from the Mn²⁺-bound PFV asymmetric unit [Protein Data Bank code 3L2S] and two-domain HIV-1 IN crystal structures (14, 15). The resulting DNA-bound IN tetramer was optimized for stereochemistry and energy minimized as described in Materials and Methods.

**Protein Interactions Within the HIV-1 Intasome.** Previous HIV-1 CCD crystal structures revealed globular dimers with the two sets of active-site residues separated by ≈35 Å on opposing surfaces. Because 5 bp, roughly equivalent to 18 Å in ideal B-form DNA, separates the points of attack on target DNA phosphodiester bonds during strand transfer, it was unlikely that this pair of active sites catalyzed the integration of both viral DNA ends (2). Inspection of the intasome structure (26) and HIV-1 model (Fig. 1) reveals the answer to this long-standing riddle. The common CCD-CCD dimer interface is maintained but occurs between the outer and inner monomer within each half of the IN tetramer (blue and green protomers in Fig. 1). It is the inner subunits of the tetramer (green and cyan in Fig. 1D) that make all visible DNA contacts and thus harbor the two functional active sites—the outer CCDs (blue and yellow) provide an apparent supportive role(s) to the overall structure. The DNA-contacting subunits moreover adopt an extended conformation, with the CTD positioning between the CCD and NTD (Fig. 1B). Because of lack of electron density for the analogous regions of PFV (26) and two-domain HIV-1 CCD-CTD (15) INs, we note that structural information for the C-terminal 18 amino acids (HIV-1 IN residues 271–288) is not available in our model.

Relatively long α helices extending from the common CCD dimer to each CTD were observed in the HIV-1 IN CCD-CTD crystal structure (15). Other CCD-CTD structures, however, did not possess extended α helical linker regions. In the analogous Rous sarcoma virus construct, for example, the linkers adopted extended variable conformations (18). These results suggested considerable flexibility of CCD-CTD linker regions within IN deletion constructs. A dramatic increase in the sensitivity of Arg199 to small modifying reagents upon viral DNA binding moreover confirmed flexibility in this region of full-length, active HIV-1 IN (24). Consistent with these observations, the CCD-CTD as well as NTD-CCD linker regions adopt extended conformations within the PFV structure (26) and HIV-1 model (Fig. 1).

The CCD of each inner monomer engaged with the reactive DNA terminus at its active site interface with the NTD of the other inner monomer in trans (green CCD and cyan NTD in upper portion of Fig. 1A). The X-ray crystal structure of the two-domain NTD-CCD fragment from maedi-visna virus IN revealed critical intermolecular interactions between NTD residues Glu11 and Glu25 and CCD residues Lys188 and Lys190, respectively (19). Salt bridges between corresponding HIV-1 positions Glu11–Lys186 and Asp25–Lys188 were notably seen in the initial NTD-CCD structure, although at that time it was unclear these formed intermolecularly (14). E11K and K186E mutations each similarly disrupted HIV-1 IN tetramerization and intasome function, with a mixture of E11K and K186E proteins supporting partial restoration of multimerization and IN activities (19). This functionally validated NTD-CCD interface is notably conserved in the PFV structure (26) and HIV-1 model (Fig. S2). Previously observed hydrogen (H)-bond contacts between the side chains of CCD residues Gln164 and Arg187 and the backbones of NTD residues Lys14 and Tyr15, respectively (14, 19), also persist in the model (Fig. S2).

**IN–DNA Interactions.** Residues within each domain of the inner HIV-1 IN monomers as well as their linker regions contact the DNA in our model (Fig. S1A). Table S1 summarizes these and compares them with the contacts observed in the PFV crystal structure. Also tabulated are results from previous biochemical experiments that implicated HIV-1 residues in DNA binding. We note that whereas some experiments identified specific amino acid–base interactions (20, 29–31), others did not discern precise DNA (22, 24) or IN (25, 30, 32) contact points.

For the majority of HIV-1 IN amino acids that contact DNA, the analogous PFV IN residue interacts with its substrate, although notably some PFV interactions are absent owing to different interdomain linker lengths or absence of the NED in lentiviral IN (Fig. S1A and Table S1). As observed in the PFV structure (26), the majority of HIV-1 IN residues contact the nontransferred DNA strand (Table S1). Likewise, the majority of amino acids interact with the DNA backbone, although some base contacts are observed, implicating these in defining specificity. Sequence-specific interactions include the main chain carbonyl of Gly149, which H-bonds with G4 of the nontransferred strand (Fig. S3A). This interaction is maintained in both structures because Gly218 in PFV IN, equivalent to Gly149 in HIV-1 (Fig. S1A), likewise contacts G4 of the nontransferred PFV strand (Table S1) (26). Because of the invariance of G4 among retroviral and retrotransposon LTRs and the conservation of Gly or similar small nonpolar...
Ala at this IN position (33, 34), we speculate this contact does not tolerate the presence of a bulky side chain and moreover may be conserved among a diverse set of intasome complexes.

The side chain of PFV IN residue Arg222 interacts with TS and C6 of the nontransferred strand (26), whereas its HIV-1 counterpart, Ser153, also contacts bases in the nontransferred strand (G4 and C5 in this case). Results based on chimera avian sarcoma virus–HIV-1 IN activities (22) and NMR chemical shift spectra (35) previously implicated Ser153 in viral DNA binding. PFV IN residue Asn348 contacts T2 of the nontransferred strand, whereas the analogous HIV-1 residue, Glu246, interacts with nontransferred strand position T3 (Fig. S3 and Table S1). Glu246 was previously shown to interact with viral DNA, initially to A7 of the nontransferred strand (20) although subsequently in a less specific manner (24).

Because of the relatively large number of previous studies, many of the other contact residues in our model were previously implicated, either directly or indirectly, in DNA binding (15, 22, 24, 27, 29, 30, 32, 35, 36) (Table S1). Some unique contacts were nevertheless noted, and potential roles for Asn18, Arg20, and Lys211 in IN function were probed by correlating 3′ processing and DNA binding activities of site-directed mutant proteins. Arg228 (32) and Lys266 (25, 30, 32), which reside within peptides that cross-linked to DNA, were additionally targeted because specific roles in DNA binding and IN function were untested. Arg262 (36), Lys219, Arg263, and Arg269 (24), which are known to contact DNA, were included because systematic comparison of mutants in DNA binding and activity assays were lacking. Because the final energy-minimized model placed Lys219 approximately 7 Å from the DNA, this contact was excluded from Table S1. However, consistent with previous results (24), Lys219 approached within 4 Å of the DNA backbone during Molecular Dynamics simulations.

DNA binding was assessed via covalent IN–DNA complex formation after UV irradiation and polyacrylamide gel electrophoresis (Fig. 2A). Under these conditions, IN bound a U5 DNA substrate approximately 8-fold more efficiently than a base composition-matched, scrambled control sequence. IN controls included active-site mutant D64A and DNA binding defective K156E/K159E (29). As expected, D64A effectively bound DNA under conditions that precluded K156E/K159E binding (Fig. 2A and B).

The previously undescribed IN mutant proteins displayed wide spectra of 3′ processing and DNA binding activities (Fig. 2B). As anticipated from previous studies (24, 36), K219E, R262E, R263E, and R269E harbored 3′ processing and DNA binding defects (Fig. 2B). Protein–protein cross-linking in the absence of DNA was conducted to assess potential mutational affects on inherent IN dimerization (19). Because the efficiencies of K219E, R262E, R263E, and R269E tetramerization were similar to the WT (Fig. S4), impaired catalysis correlated with defective DNA binding in these cases. K211E harbored approximate 2- to 2.5-fold DNA binding and 3′ processing defects (Fig. 2B). Because this mutant failed to effectively tetramerize (Fig. S4), it is unclear whether defective multimerization or DNA binding predominantly caused the K211E catalytic defect. Next to K156E/K159E, R228E and K266E were the most impaired for catalysis and DNA binding (Fig. 2B). R228E failed to effectively tetramerize, whereas K266E multimerized similar to the WT (Fig. S4). We therefore conclude that Lys266 contributes significantly to viral DNA binding and IN catalysis. Finding that the substitution of Glu perturbed function more so than Ala at positions 262, 263, 266, and 269 is consistent with the modeled interactions of these residues with the DNA backbone (Table S1).

Arg20 harbored an approximate 3-fold processing defect yet bound DNA as efficiently as the WT. R20A by contrast was highly defective for both activities (Fig. 2B). Because R20A formed tetramers as efficiently as the WT (Fig. S4), we conclude its DNA binding and catalytic defects specifically correlate. Arg20 is predicted to H-bond with nontransferred strand residues G8 and G10 (Fig. S5 and Table S1). Ala at this position would effectively negate potential base contacts, whereas the Glu side chain could potentially H-bond with C10 of the transferred strand (Fig. S5). Accordingly, and in contrast to the results obtained with residues predicted to interact with the DNA backbone, R20A was significantly more defective than R20E for 3′ processing and DNA binding (Fig. 2B). The same situation moreover applied to Asn18, predicted to H-bond with nontransferred strand residue G8 (Table S1). N18D displayed an approximate 5-fold processing defect yet, like R20E, it bound DNA as the WT (Fig. 2B). Considering that N18G displayed similar 2- to 2.5-fold 3′ processing and DNA binding defects, the lack of a side-chain at this position disrupted DNA binding more so than the electrongative Asp substitution.

**HIV-1 Intasome Active Site.** A theme common to transpososome (3, 4) and retroviral intasome (26) structures is the unpairing of the donor DNA duplex at the enzyme active site, which is probably important to unveil the scissile phosphodiester bond for 3′ processing and resulting 3′-OH nucleophile for subsequent DNA strand transfer. Accordingly, the terminal adenine (A17) of the reactive HIV-1 DNA strand in the model is completely separated from its T3 complement on the nontransferred strand, whereas the first three bases (A1, C2, and T3) of the nontransferred strand loop around the back of CCD α4 helix and active site, emerging to make extensive contacts with the CTD and CCD–CTD linker (Fig. S34). As predicted from previous biochemical analyses (22, 27, 29–31, 35), many of the CCD–DNA contacts involve the active site loop (residues 140–148) and α4 helix (Figs. S14 and S34).

Consistent with a two metal ion reaction mechanism (26, 37), two Mn2+ ions (Fig. S34, labeled A and B) are coordinated by the side-chains carboxylates of the catalytic triad in the modeled active site. During DNA strand transfer, metal ion B engaged by Asp64 and Glu152 would activate the 3′-OH of the reactive DNA strand, whereas metal ion A, coordinated by Asp64 and Asp116, would destabilize the scissile phosphodiester group in target DNA. By superimposing the Cα atoms of the Ts5 transpososome and PFV intasome Asp–Asp–Glu (DDE) active site residues, we previously

**Fig. 2.** IN DNA binding and 3′ processing activities. (A) DNA binding assay. Left: Representative gel loaded with reactions conducted in the absence of IN or containing WT, D64A, or K156E/K159E (EE) IN and U5 or sequence non-specific DNA; Right: quantified results of n = 4 experiments. (B) 3′ Processing (black) and DNA-binding (gray) activities of indicated mutants expressed as percentage WT function for two (3′ processing) or three (DNA binding) experimental replicates.
noted marked conservation in the positioning of donor DNA 3′-OH ends and divalent metal ions (26). Incorporating the Mos1 transpososome (4) into the structural overlay reinforces the universal mechanism at the heart of these DNA recombination machines (Fig. S3B).

**INSTIT Binding and Mechanism of Inhibition.** A common feature of clinical INSTIs is coplanar oxygen atoms reminiscent of an original diketo acid pharmacophore (12, 13), which has been proposed to chelate the essential divalent metal ions at the IN active site and effectively shield their availability for DNA strand transfer (38). Predecessor diketo acids moreover displayed affinity for binding to the intasome as compared with free IN, highlighting a role for the viral DNA end in forming the complete drug interaction site (16). Because of the importance of discerning the underlying mechanism of drug binding, numerous studies have modeled INSTIs at the HIV-1 IN active site, the most recent of these focusing on RAL and EVG (28, 39, 40).

PFV intasome activity, as well as cell infection by PFV, is inhibited by RAL and EVG (41). PFV intasome crystals soaked in the presence of RAL or EVG, each with MgCl₂ or MnCl₂, led to the determination of drug-bound structures (26). These findings afforded an optimal platform from which to model RAL and EVG binding at the HIV-1 IN active site. The resulting interactions were mediated via two conserved features of each partner: coplanar RAL and EVG oxygen atoms engage the divalent metal ions, whereas their halobenzyl moieties stack against the penultimate cytosine (C16) of the reactive DNA strand, forcing the 3′-OH of the terminal A17 nucleotide away from the active-site carboxylates and divalent metal ions (compare the drug-bound models in Fig. 3 B and C with the drug-free state in Fig. 3A). These results are consistent with the demonstrated importance of the terminal reactive adenosine for the effective binding of a related INSTI (42). We conclude that the dramatic displacement of the 3′-OH nucleophile from the active site upon drug binding explains the mechanism of inhibition.

We note that the induced fit mode of RAL and EVG binding to divalent metal ions and donor DNA bases was achieved only through modeling these salient components from the corresponding PFV crystal structures. Concordantly, automated docking of RAL or EVG onto the Mg-bound intasome (Fig. 3A) using Autodock 4.2 (43) under conditions that assumed flexibility in receptor and binding ligands failed to achieve the drug binding mode observed through manual docking. We can only assume that the rather dramatic rearrangement in adenosine positioning upon INSTI engagement lies outside the myriad of possibilities sampled by automated docking programs. Accordingly, comparing the positioning of RAL and EVG in our models with results of recent INSTI docking studies (28, 39, 40) reveals strikingly different binding modes.

**Mechanism of HIV-1 Resistance to RAL and EVG.** A select number of amino acids in addition to the DDE active-site residues contact the drugs in our models: Gln146 and Arg231 interact with each compound, RAL additionally contacts Asn117, Tyr143, Asn144, and Pro145, whereas EVG is engaged by Cys65 (Fig. 3). Each drug additionally interacts with three of the four bases that constitute the invariant LTR DNA end (Fig. S1B): C16 and A17 of the reactive DNA strand and G4 of the nontransferred strand (Fig. 3A–C).

Numerous mutations recovered from ex vivo virus evolution experiments and/or patients confer resistance to RAL and/or EVG (see refs. 13 and 44 for recent overviews). Individual amino acid changes are defined as primary determinants if they confer ≥10-fold increases in drug resistance, and secondary mutations have been identified that work with primary changes to increase overall resistance. Substitutions of Gln148, Asn155, or Thr16 confer primary resistance to both drugs (45, 46). Mutations at Tyr143 confer resistance to RAL (47), whereas changes at Glu92, Phe121, Pro145, Gln146, Ser147, and Val151 are observed in EVG-resistant strains (45, 48) (Fig. S6 A and B, residues painted cyan). Sites of secondary resistance changes are shown as salmon color in Fig. S6. Clearly, only a limited number of resistance mutations arise at residues that directly contact the drugs in our models: Tyr143 in the case of RAL, and Gln146 for EVG (compare Fig. S6A and B with Fig. 3B and C, respectively). Our analysis therefore highlights numerous instances in which a mutant side chain that predictably will not contact the drug nevertheless confers significant resistance to it. At least in some of these cases, the mutation is likely to affect the positions of the key drug contacts—the viral DNA end and/or coordinated metal ions. Conceivably, subtle detuning of these crucial ligands will significantly reduce the affinity of the intasome–drug interaction while at the same time retain sufficient IN activity to enable integration during HIV-1 infection. Occurrence of secondary mutations might predominantly compensate for loss of enzymatic function instilled by primary resistance changes to increase the replication fitness of drug-resistant strains (49).
Conclusions

HIV-1 Intasome Structure and Function. Several features of our model, which shed light on molecular details important for HIV-1 integration, are supported by experimental data. Critical intermolecular NTD-CCD interactions (19) were maintained in the final energy-minimized structure (Fig. S2), and results reported here moreover validate important roles for Arg20 and Lys266 in DNA binding and IN function. Because mutations at Ser230, Asp232, or Arg263 can confer INSTI resistance (13), the CTD was expected to impinge on the CCD in the functional nucleoprotein complex, and our results indeed highlight the juxtaposition of these residues in the vicinities of the RAL and EVG binding sites (Fig. S6 and Table S1).

Some previously observed specificity interactions, for examples between Gln148 and Tyr143 with A1 (30) or Gln148 with C2 (31) of the nontransferred DNA strand, were not observed here. These differences may very well be attributable to experimental design. Some cross-linking substrates (30, 31) harbored the pGTpOH dinucleotide that is removed during 3′ processing and absent from our model, and its presence could conceivably position the 5′ end of the nontransferred strand closer to Tyr143 and/or Gln148 (Fig. 3 and Figs. S3A and S6). Although a precleaved DNA modified at C2 also cross-linked with Gln148 (31), in this case a used alkanethiol tether may have provided sufficient flexibility. Consistent with this interpretation, the adjacent residue in the CCD flexible loop, Ser147, interacted with the neighboring T3 position of the nontransferred strand in our model (Fig. S3A and Table S1). Lys159 was reported to interact with the adenine base of the terminal reactive nucleotide (29), and although close by, we note that this side chain in the HIV-1 model and analogous PFV IN residue Lys228 contacts the adjoining DNA backbone (Fig. S3A and Table S1).

The functional organization of IN protomers and DNA ends in the intasome model is fully consistent with results of complementation experiments that indicated that the NTD functioned in trans to the active site, whereas the CTD could be supplied in cis or in trans (50, 51). Notably, these early experiments used conditions under which only one DNA end instead of two predominantly became integrated, implying that the intasomal tetramer may be the protagonist of single-end in addition to concerted HIV-1 integration. The IN–DNA complex that mediates single-end integration is far less stable than the intasome (8), suggesting that DNA end synapsis is crucial for proper assembly and/or stability. The results of some experiments have indicated that an HIV-1 IN dimer can suffice to process a DNA end (52, 53), whereas we and others have concluded that processing is likely to occur within the context of the tetramer (8, 19, 54). If 3′ processing occurred before synapsis, the missing NTD of the top blue protomer in Fig. 1 would have to assume the position filled by the cyan NTD in the intact intasome. The cyan NTD would then have to displace this blue NTD upon tetramerization and DNA end synthesis.

INSTI Binding Mechanism and Inhibition. The mechanism of INSTI–IN binding has been difficult to ascertain in the absence of intasome structural information. The underlying features of the drugs, coplanar oxygen atoms and aromatic rings that tend to be halogenated (12, 13, 42), had suggested that coordination of critical divalent metal ions via pharmacophore oxygens underscores the mechanism of inhibition (38), and our results indeed highlight these interactions between RAL and EVG with the intact PFV (26) and HIV-1 (Fig. 3) IN active sites. Until recently, the aromatic elements of these compounds were proposed to interact with a protein component(s) of the intasome (23). Because of the new structural information, we however speculate that an additional common and critical theme among INSTIs is an interaction between the aromatic features of the drugs and nucleotides that constitute the viral DNA end, specifically A17 and C16 of the reactive DNA strand and G4 of the nontransferred strand. Because of a strict requirement for these bases during retroviral integration (1) and as-of-yet lack of evidence for LTR mutations that confer INSTI resistance, next-generation compounds should take advantage of these invariant design features of the HIV-1 integration machine.

Materials and Methods

Model Building and Validation. The secondary structure-based alignment of PFV and HIV-1 INs (26) was used to guide the construction of the HIV-1 intasome (see also Fig. S1A). RAL and EVG were modeled at the resulting IN active site using corresponding drug-bound PFV structures (26). In vitro integration and cross-linking assays were used to validate IN-DNA contacts; see SI Materials and Methods for salient details.

ACKNOWLEDGMENTS. This work was supported by UK Medical Research Council Grant G0000116 (to P.C.), US National Institutes of Health Grant AI070042 (to A.E.), and the Harvard University Center for AIDS Research, a National Institutes of Health-funded program (P30AI60635) that is supported by multiple National Institutes of Health Institutes and Centers (National Institute of Allergy and Infectious Diseases, National Cancer Institute, National Institute of Mental Health, National Institute on Drug Abuse, National Institute of Child Health and Human Development, National Heart, Lung, and Blood Institute, and National Center for Complementary and Alternative Medicine).

Supporting Information

Krishnan et al. 10.1073/pnas.1002346107

SI Materials and Methods

Intasome Model Construction and Validation. Despite a lack of recognizable amino acid sequence homology between TN5 transposase and prototype foamy virus (PFV) integrase (IN), aligning the Ca atoms of their active-site residues yielded striking spatial conservation in the positioning of the reactive donor DNA 3’-OH ends and coordinated divalent metal ions (1). On the basis of these key points of structural conservation, the HIV-1 model was initiated by overlaying the Ca positions of active-site residues Asp64, Asp116, and Glu152 from one catalytic core domain (CCD) monomer of the HIV-1 IN N-terminal domain (NTD)-CCD dimer (2) [Protein Data Bank (PDB) code 1K6Y] onto those of the inner CCD monomer of the Mn-bound PFV crystallographic asymmetric unit (PDB code 3L2S). The NTD and C-terminal domain (CTD) coordinates from the two-domain HIV-1 IN structures (CCD-CTD dimer PDB code 1EX4; ref. 3) were next superimposed using the Ca atoms of the conserved secondary structural elements within each domain as guides. The NTD within the two-domain HIV-1 IN structure (2) notably harbors three α helices, whereas a fourth C-terminal helix was observed in the PFV structure as well in the solution structures of the isolated HIV-1 (4) and HIV-2 (5) domains. The analogous region of the NTD (residues 42–45) within the HIV-1 model was therefore built as α-helix. The coordinates of the HIV-1 IN active-site loop (residues 140–148), absent from the NTD-CCD structure (2), were generated using the Ca trace of the equivalent PFV IN residues (Fig. S1A).

The structures of the NTD-CCD (residues 47–58) and CCD-CTD (residues 203–222) interdomain linkers were guided by the equivalent PFV IN linkers and optimized for their shorter lengths and altered sequences (Fig. S1A) by defining them as loops in the loop refinement protocol of Discovery Studio 2.5 (Accelrys Software). Ten models were built and refined for each linker, with the best models selected on the basis of lowest total energy. Alternatively, possible linker conformations searched using the Scan Loop Database protocol of DeepView (Swiss-PdbViewer 4.0.1) (6) were evaluated according to number of clashes, hydrogen bond formation potential, and GROMOS96 energy (7). Optimal models were selected for least number of clashes and lowest threading formation potential, and GROMOS96 energy minimized (100 steps of Smart Minimizer method) and analyzed in terms of total energy and RMS deviation of Ca atoms, neither of which significantly deviated from the starting structure.

IN Strand Transfer Inhibitor Modeling at the HIV-1 IN Active Site. The compounds from the Mg-bound PFV structures [PDB codes 3L2T for raltegravir (RAL) and 3L2U for elvitegravir (EVR)] were modeled into the HIV-1 intasome using the Ca atoms of the respective catalytic triads as guide. The conformations of the terminal A17 nucleotides from the inhibited PFV structures were then introduced using these key Ca atom guide positions. The Mg ions from the PFV structures were next modeled, again guided by catalytic triad Ca atoms, after which the drug-bound HIV-1 structures were energy minimized as described above for the tetrameric IN–viral DNA complex.

Figs. 1 and 3 of the main text as well as Figs. S2, S3, S5, and S6 were drawn using PyMOL version 0.99 (http://www.pymol.org).

Protein Expression and Purification. HIV-1 IN proteins expressed from pCPH6P-HIV1-IN were purified essentially as previously described (11). Briefly, transformed Escherichia coli strain PC2 grown at 30 °C overnight and subcultured at 1:20 in 750 mL LB–100 µg/mL ampicillin at 30 °C until A600 of 0.5 were induced by the addition of 0.4 mM isopropyl-β-D-thiogalactopyranoside for 4.5 h. The bacterial pellet was resuspended in ice-cold buffer A [50 mM Tris-HCl (pH 7.4), 1 mM NaCl, 1 mM DTT, 1 mM phenylmethylsulfonylfluoride, 7.5 mM CHAPS] containing 25 mM imidazole and sonicated. After centrifugation for 30 min at 40,000 × g, the supernatant was incubated with 1.5 mL Ni2+-nitrilotriacetic acid agarose beads (Qiagen) equilibrated in buffer A–25 mM imidazole at 4 °C for 1 h. The beads were washed with 35 volumes of buffer A–25 mM imidazole, and His6-IN was eluted with buffer A–200 mM imidazole. The His6 tag was removed by digestion with 40 U HRV14 3 C protease (GE Healthcare) overnight at 4 °C, and the mixture was dialyzed overnight against buffer A–10% glycerol. Purified tag-free IN was concentrated, flash frozen in liquid N2, and stored at −80 °C.

3' Processing and Cross-Linking Assays. 3' End filled-in 30 bp U5 HIV-1 DNA was used to monitor IN 3' processing activity (12). DNA binding assays used 32-pb oligonucleotides that modeled the U5 end [AE3651 (5'-CCCTTTATGATCTTGGAAATCTTCATTAGAGT), AE3652 (5'-ACTGCTAGAATTTTCACCTGACTGACTAAAG)], or a base-composition matched, scrambled control sequence [AE4167 (5'-CAAGAAGCATCTACATGTTGCCATG), AE4168 (5'-TTAACCGCAATTGTGGAACGTGATGCTTCTG)]. AE3652 and AE4168 5'-end labeled with [γ-32P]ATP (3000 Ci/mmol; Perkin-Elmer) using OptiKinase (USB) were annealed to
AE3651 and AE4167, respectively. Unincorporated radionuclide was removed using Bio-Spin 6 columns (Bio-Rad) equilibrated with 10 mM Tris-HCl (pH 8.0), 20 mM NaCl, and 0.1 mM EDTA. DNA binding and 3′ processing reactions contained 17.5 mM Mops (pH 7.2), 7 mM DTT, 7 mM MgCl₂, 3.5% polyethylene glycol 8000, 3.5% DMSO, 125 mM NaCl, 20 mM DNA, and 1 μM IN. After 15 min on ice, binding reactions (10 μL) transferred to Terasaki plate wells (Greiner Bio-one) were UV (254 nm) cross-linked at 4 °C for 15 min with a mean flux of 2.1 mW/cm². Samples boiled in SDS/PAGE sample buffer were fractionated through 4–20% Tris-Glycine gels (Invitrogen). Dried gels were visualized by autoradiography, and DNA binding activity was quantified as percentage substrate converted to cross-linked product using Image Quant version 1.2.

3′ Processing reactions (10 μL) incubated at 37 °C for 2 h were stopped by adding an equal volume of sequencing gel buffer [95% (vol/vol) formamide, 125 mM NaCl, 20 nM DNA, and 1 μM IN] cross-linking was performed essentially as previously described (13). In brief, 6 μL of WT or mutant IN (0.53 mg/mL protein in 1 M NaCl, 2.5 mM DTT, 7.5 mM CHAPS, 25 mM Tris-HCl, pH 7.4) was diluted with 20 μL reaction buffer [0.75 M NaCl, 2 mM MgSO₄, 25 μM ZnCl₂, 25 mM Hepes-NaOH (pH 7.5), and 2.5 mM DTT]. Cross-linking initiated by addition of 4 μL BS³ (Sigma-Aldrich; fresh 5 mM or 1.5 mM stock in water) proceeded for 30 min at 18 °C. Reactions stopped by addition of SDS/PAGE sample buffer were separated through Novex 10% Bis-Tris gels (Invitrogen); products were detected by staining with Sypro Orange (Invitrogen). Percentage input WT and mutant IN (monomer) converted to tetramer at 0.7 mM BS³ was calculated using Alpha Innotech FluorChem FC2.

Fig. S1. PFV/HIV-1 sequence alignments. (A) Structure-based amino acid sequence alignment with secondary elements (α, β, and η) represent α helix, β strand, and 3_10 helix, respectively shown atop the IN sequences numbered for individual domains (NED, N-terminal extension domain; NTD, CCD, and CTD as defined in SI Materials and Methods). PFV assignments (Upper) are from PDB code 3L2S, whereas the HIV-1 elements (Lower) were deciphered herein. Residues that interact with DNAs are labeled with filled circles and open ovals (blue, PFV; black, HIV-1) to denote hydrogen bond and van der Waals contacts, respectively. Conserved residues are bold, with identical residues highlighted in gray; underlined residues belonging to the NTD HHCC, and CCD DDE motifs are colored green and red, respectively. (B) Viral U5 DNA end sequences. The nontransferred and reactive strands are colored orange and magenta, respectively, with bases conserved between viruses highlighted in gray background. The CA/GT bp conserved among retroviruses and LTR retrotransposons is boxed; A17 within this motif harbors the reactive 3'-OH nucleophile for DNA strand transfer.

Fig. S2. Intermolecular NTD-CCD interactions in the HIV-1 intasome. Highlighted are Glu11–Lys186 and Asp25–Lys188 salt bridges (black dashed lines) between the NTD of one inner monomer (cyan) and the CCD of the second inner monomer (green). Also shown are hydrogen bonds (gray dashes) between the side chains of CCD residues Gln164 and Arg187 with the backbones of NTD residues Lys14 and Tyr15. Secondary structural elements are labeled as in Fig. S1A.
Fig. S3. Details of the HIV-1 intasome active site. (A) Protein–DNA interactions. The majority of CCD, CTD, and CCD-CTD linker residues (Table S1) from the green IN protomer (Fig. 1) that interact with the nontransferred (orange) and reactive (magenta) DNA strand termini are shown as sticks, with red, blue, and yellow denoting O, N, and S atoms, respectively. The side chains of the catalytic triad, coordinated Mn ions (labeled A and B), and 3′-OH of the reactive DNA strand are also colored red. Secondary structural elements are labeled as in Fig. S1A. (B) Superimposition of retroviral intasome and DNA transpososome active sites. Side chains of catalytic triad residues, coordinated divalent metal ions, and 3′-OHs of reactive terminal nucleotides are shown for Mos1 (green), Tn5 (gray), PFV (pink), and HIV-1 (salmon). Tn5, PFV, and HIV-1 each use DDE active-site residues, whereas Mos1 uses DDD. The Mos1 structure harbored only one Mn ion, at position A (1).

**Fig. S4.** IN cross-linking. (A) The mobilities of the indicated mutant proteins incubated in the absence or presence of 0.7 or 0.2 mM BS3 were compared with the WT IN loaded on each gel. The positions of mass standards are on the left, whereas those of cross-linked IN dimer, trimer, and tetramer as well as unmodified monomer are indicated on the right. (B) Quantitation of results after two (most mutants) to four (K219A and K219E) experimental replicates.

**Fig. S5.** Potential base contacts mediated by Arg20 and mutant side chains. In the model, Arg20 is within hydrogen bonding distance to A9 and G10 of the nontransferred strand. The presence of Ala at this position destroys this hydrogen bond network, whereas the carboxylate group of R20E is in bonding distance to C10 of the reactive DNA strand.
Fig. S6. Sites of resistance mutations mapped onto drug-bound HIV-1 intasome models. (A and B) Side chains that undergo mutation to confer resistance to RAL and EVG, respectively, are shown as cyan (primary resistance mutations) or salmon (secondary changes). Ser153 in B is on the backside of α-helix 4 and thus not visible in this projection; remaining labeling is the same as in Fig. 3 of the main text.
Table S1. Protein–DNA contact details

<table>
<thead>
<tr>
<th></th>
<th>PFV</th>
<th>HIV-1</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Nt*</td>
<td>aa²</td>
</tr>
<tr>
<td>A1</td>
<td>T363</td>
<td>v, B</td>
</tr>
<tr>
<td>T2</td>
<td>E207</td>
<td>v, bb</td>
</tr>
<tr>
<td>T2</td>
<td>N348</td>
<td>H, B</td>
</tr>
<tr>
<td>T3</td>
<td>S209</td>
<td>v, bb</td>
</tr>
<tr>
<td>T3</td>
<td>T210</td>
<td>H, bb</td>
</tr>
<tr>
<td>T3</td>
<td>S216</td>
<td>H, bb</td>
</tr>
<tr>
<td>T3</td>
<td>H183</td>
<td>H, bb</td>
</tr>
<tr>
<td>T3</td>
<td>E207</td>
<td>H, B</td>
</tr>
<tr>
<td>T3</td>
<td>K124</td>
<td>H, B</td>
</tr>
<tr>
<td>T3</td>
<td>T351</td>
<td>v, bb</td>
</tr>
<tr>
<td>T3</td>
<td>G215</td>
<td>v, bb</td>
</tr>
<tr>
<td>G4</td>
<td>L113</td>
<td>v, bb</td>
</tr>
<tr>
<td>G4</td>
<td>R350</td>
<td>H, bb</td>
</tr>
<tr>
<td>G4</td>
<td>H213</td>
<td>H, bb</td>
</tr>
<tr>
<td>G4</td>
<td>G218</td>
<td>H, B</td>
</tr>
<tr>
<td>T5</td>
<td>R114</td>
<td>H, bb</td>
</tr>
<tr>
<td>T5</td>
<td>P115</td>
<td>v, bb</td>
</tr>
<tr>
<td>T5</td>
<td>R222</td>
<td>H, B</td>
</tr>
<tr>
<td>C6</td>
<td>R222</td>
<td>H, B</td>
</tr>
<tr>
<td>C6</td>
<td>R117</td>
<td>H, bb</td>
</tr>
<tr>
<td>C6</td>
<td>K219</td>
<td>H, bb</td>
</tr>
<tr>
<td>A7</td>
<td>S258</td>
<td>H, bb</td>
</tr>
<tr>
<td>T8</td>
<td>R229</td>
<td>H, bb</td>
</tr>
<tr>
<td>T8</td>
<td>G68</td>
<td>H, bb</td>
</tr>
<tr>
<td>T8</td>
<td>N106</td>
<td>H, B</td>
</tr>
<tr>
<td>G9</td>
<td>H374</td>
<td>H, bb</td>
</tr>
<tr>
<td>G10</td>
<td>K107</td>
<td>H, bb</td>
</tr>
<tr>
<td>G10</td>
<td>S373</td>
<td>v, bb</td>
</tr>
<tr>
<td>G10</td>
<td>H374</td>
<td>H, bb</td>
</tr>
<tr>
<td>G10</td>
<td>R69</td>
<td>H, B</td>
</tr>
<tr>
<td>G10</td>
<td>R41¹</td>
<td>H, bb</td>
</tr>
<tr>
<td>T18</td>
<td>K26¹</td>
<td>H, bb</td>
</tr>
<tr>
<td>A7</td>
<td>N84</td>
<td>H, bb</td>
</tr>
<tr>
<td>T8</td>
<td>R66</td>
<td>H, bb</td>
</tr>
<tr>
<td>T9</td>
<td>K262</td>
<td>H, bb</td>
</tr>
<tr>
<td>C11</td>
<td>R313³</td>
<td>H, B</td>
</tr>
<tr>
<td>T13</td>
<td>S311</td>
<td>H, bb</td>
</tr>
<tr>
<td>G14</td>
<td>S105</td>
<td>v, bb</td>
</tr>
<tr>
<td>A15</td>
<td>R326</td>
<td>H, bb</td>
</tr>
<tr>
<td>C16</td>
<td>R336</td>
<td>H, bb</td>
</tr>
<tr>
<td>A17</td>
<td>K228</td>
<td>H, bb</td>
</tr>
<tr>
<td>C16</td>
<td>R336</td>
<td>H, bb</td>
</tr>
</tbody>
</table>

*Nt, nucleotide: orange, nontransferred DNA strand; magenta, reactive strand.

¹aa, amino acid: green, residue from DNA-bound green IN protomer in Fig. 1A in main text; cyan, residue from transacting cyan monomer.

²Con, contact; representing the following <4-Å IN–DNA interactions: H, hydrogen bond; v, van der Waals; B, base; bb, backbone; ND, not determined; NA, not applicable.

³Residue implicated in DNA binding but not necessarily to this Nt position.

⁴Analogous amino acid position absent from HIV-1 sequence (see Fig. S1A alignment).

