HHS Public Access

Author manuscript

Nat Rev Microbiol. Author manuscript; available in PMC 2023 January 01.

Published in final edited form as:

Nat Rev Microbiol. 2022 January ; 20(1): 20–34. doi:10.1038/s41579-021-00586-9.

Structure and function of retroviral integrase

Goedele N. Maertens[1], **Alan N. Engelman**[2,3], **Peter Cherepanov**[1,4]

[1]Department of Infectious Disease, St-Mary's Campus, Imperial College London, London, W2 1PG, UK.

[2]Department of Cancer Immunology & Virology, Dana-Farber Cancer Institute, Boston, MA 02215, USA.

[3]Department of Medicine, Harvard Medical School, Boston, MA 02115, USA.

[4]Chromatin Structure and Mobile DNA Laboratory, The Francis Crick Institute, London, NW1 1AT, UK.

Abstract

A hallmark of retroviral replication is establishment of the proviral state, wherein a DNA copy of the viral genome is stably incorporated into a host cell chromosome. Integrase is the viral enzyme responsible for the catalytic steps involved in this process, and integrase strand transfer inhibitors are widely used to treat people living with HIV. In the past decade, a series of X-ray crystallography and cryogenic electron microscopy studies revealed the structural basis of retroviral DNA integration. A variable number of integrase molecules congregate on viral DNA ends to assemble a conserved intasome core machine that facilitates integration. The structures additionally informed on the modes of integrase inhibitor action and the means by which HIV acquires drug resistance. Recent years have witnessed the development of allosteric integrase inhibitors, a highly promising class of small molecules that antagonise viral morphogenesis. This review focusses on the recent insights into the organisation and mechanism of the retroviral integration machinery and highlights open questions and new directions in the field.

Retroviruses are enveloped diploid (+) strand RNA viruses that require integration of a DNA copy of the viral genome into a host cell chromosome to establish infection. The phylogeny of this well-studied viral family includes two subfamilies, Orthoretrovirinae and Spumaretrovirinae, each of which contains several genera: α-, β-, γ-, δ-, ε-retroviruses, and lentiviruses comprise Orthoretrovirinae, while Spumaretrovirinae includes simii-, prosimii-, bovi-, feli-, and equispumaviruses. The first human retrovirus, human T-cell lymphotropic virus (HTLV, a δ-retrovirus), was isolated in 1979 (REF.[1,2]). Human immunodeficiency virus (HIV, a lentivirus) was discovered four years later[3,4]. Both viral species comprise two main types (*e.g.*, HIV-1 and HIV-2) that largely share biological properties although they differ in severity of clinical manifestations. While members of the remaining retroviral

Correspondence: G.Maertens@imperial.ac.uk (GNM), Alan_Engelman@dfci.harvard.edu (ANE), Peter.Cherepanov@crick.ac.uk (PC).

Conflicts of Interest

A.N.E. consults for ViiV Healthcare, Co. The other authors have no conflicts to declare.

genera are not known to cause pathology in humans, research on these species has been invaluable to elucidate cellular transformation and tumorigenesis mechanisms, develop anti-HIV/AIDS therapies, and biomedical science in general. Retroviral vectors based on murine leukaemia virus (MLV, a γ-retrovirus) are widely used in biological research. Thanks to the many additional years of basic studies of avian sarcoma-leukosis virus (ASLV, an α-retrovirus) and the discovery of reverse transcriptase (RT)[5,6], the development of anti-retroviral drugs had an early start, resulting in the rollout of the first HIV inhibitor, zidovudine, in 1987.

The obligatory integration step distinguishes retroviruses from all other viral families. This process is catalysed by integrase (IN), which together with protease and RT comprise the canonical triad of enzymes carried by replication-competent retroviruses. Produced as part of the Gag-Pol polypeptide, IN is released during proteolytic maturation within HIV particles[7]. It is estimated that a mature HIV-1 virion contains about 120 copies of IN protein[8,9]. Upon entry of the viral core into a host cell, RT synthetises a linear double-stranded DNA copy of the viral genome (vDNA), flanked by direct long terminal repeats (LTRs). The exact cellular location and timing at which vDNA is liberated from the HIV-1 core is unclear, though recent results indicate this transpires in the cell nucleus close to the site of integration[10–13]. In addition to vDNA and IN, retroviral pre-integration complexes (PICs) were reported to contain cell-derived components including barrier-to-autointegration factor (BAF)[14,15], high-mobility group protein A1 (HMGA1)[16], lens epithelium-derived growth factor (LEDGF/p75)[17], and histones[18,19]. Acting upon the ends of the vDNA molecule, IN accomplishes two enzymatic reactions (Fig. 1). During the 3’-processing reaction, IN hydrolyses the vDNA to liberate 3’-hydroxyl groups attached to the invariant dCdA dinucleotides, which, prior to processing, are located within three bases of the vDNA 3’ termini. During the strand transfer reaction, IN uses the 3’-OH groups to cut both strands of chromosomal DNA, simultaneously joining the 3’ ends of the vDNA molecule to the target. Integration initially results in a hemi-integrated form of vDNA flanked by short single-stranded gaps and short 5’ overhangs. This integration intermediate subsequently relies on thus far poorly characterized host cell enzymes to join the 5’ vDNA ends to the chromosome[20,21]. Because the 3’ ends of vDNA are inserted in a staggered fashion, across the major groove of target DNA, the hemi-integrated provirus is initially flanked by 4–6-nucleotide gaps (the gap length varies between viral species). Upon repair, these gaps give rise to short direct repeats of the target DNA sequence flanking the integrated virus.

Highlighting this field of research, inhibitors of HIV IN are common components of frontline antiretroviral therapies. Herein we overview results of IN structural biology that over the past decade have elucidated the structural basis for retroviral integration and the mechanism of clinical inhibitor action. We also describe approaches used by different retroviruses to target host chromatin for integration, which informed the development of a preclinical class of HIV-1 IN inhibitors. Unexpectedly, these compounds block the morphological transformation of the immature virus particle, a step in virus replication that is far removed from the catalytic steps of vDNA integration. Our goal is to describe the current status of the different types of IN inhibitors and what could be done to improve their efficacies in coming years. At the same time, we aim to highlight exciting new directions in the basic science of retroviral integration.

Architectures of retroviral intasomes

INs are encoded within the 3’ portions of retroviral *pol* genes[7] and are composed of ~300 amino acid residues. All INs harbour three structural domains connected by flexible linkers: the N-terminal domain (NTD), the catalytic core domain (CCD), and the C-terminal domain (CTD). Structures of the individual HIV-1 IN domains were determined in the 1990s. The NTD folds into a compact three-helical bundle, stabilised by coordination of a Zn^{2+} ion by invariant His and Cys residues of the HHCC motif[22]. The CCD harbours the active site and shares its characteristic α/β-fold with a family of diverse polynucleotidyl transferases and nucleases including RNase H, Holliday junction resolvases, and DNA transposases[23]. The C-terminal domain features an SH3-like β-barrel fold[24]. The INs from the γ-, ε-retroviruses, and spumaviruses additionally contain a small N-terminal extension domain that interacts with vDNA[25,26].

Retroviral integration is highly similar to prokaryotic transposition, and work with bacteriophage Mu greatly informed the early years of viral research[27]. To accomplish its function, a multimer of IN assembles at the ends of vDNA to form the highly stable intasome nucleoprotein complex. The intasome exists in four sequential functional states: the initial stable synaptic complex, which converts into the cleaved synaptic complex upon 3’-processing, the target capture complex, and, finally, the post-catalytic strand transfer complex (Fig. 2A, bottom panel). The transitions between these states do not seem to be accompanied by profound remodeling of intasome structure[28].

Akin to functionally equivalent transposase-DNA complexes, retroviral intasomes are resistant to challenge with high-salt conditions, and this property was instrumental in their biochemical isolation. The intasomes from several retroviral species were assembled from purified components and visualised by X-ray crystallography or single-particle cryogenic electron microscopy (cryo-EM)[25,29–39] (Figs. 2A, 3). Consistent with their function, all intasomes feature 2-fold symmetry and harbour a pair of juxtaposed active sites engaged with 3’ vDNA ends. The structures revealed unexpected architectural diversity among intasomes from different retroviral genera. While the intasomes from the prototype foamy virus (PFV, a simiispumavirus) as well as the δ-retroviruses HTLV-1 and simian T-cell lymphotropic virus (STLV) contain four IN subunits (arranged as a dimer-of-dimers)[25,29,37,38], those from ASLV and mouse mammary tumour virus (MMTV, β-retrovirus) harbour eight IN chains (a tetramer-of-dimers)[30,31,39]. The intasome from maedi-visna virus (MVV, a lentivirus), containing sixteen IN subunits (a tetramer-of-tetramers), is the largest characterized to date[32]. The stoichiometric state of primate lentiviral intasomes from HIV-1 and the simian immunodeficiency virus (SIV) remain a subject of debate, with complexes containing four and eight to sixteen IN molecules observed by cryo-EM and negative stain electron microscopy[33–36]. Despite these differences, all intasomes incorporate the conserved intasome core (CIC) assembly, which is the minimal functional unit and is represented by the PFV structure (Figs. 2A, 3). The CIC contains two IN CCD dimers (painted cyan and green in Figs. 2–4), each of which contributes one active site for catalysis. Within the CIC, the IN dimers are bridged via the exchange of associated NTDs and are joined by a pair of CTDs (referred to as the synaptic CTDs), which act as rigid spacers.

Inserted into this assembly, the vDNA ends make extensive interactions with all three types of IN domains within the CIC.

The flanking IN dimers within α-, β-retroviral and lentiviral intasomes visualised by cryo-EM display considerable flexibility[31,33,34,39]. The expansion of the intasome architecture in these viruses is dictated by three factors. While in δ-retroviral and spumaviral INs comparatively long and extended CCD-CTD linkers allow the CIC IN subunits to provide the synaptic CTDs, these are donated by additional IN molecules in the cases where the CCD-CTD linker is too short (α- and β-retroviruses) or too compact due to formation of an α-helix (lentiviruses). Secondly, the multimeric state of the IN protein before assembly partly determines the building unit of the resulting intasome. Thus, α- and β- retroviral INs, which tend to form dimers in solution[40,41], assemble into octameric intasomes: two dimers donate one active site each, while the additional two dimers provide the synaptic CTDs. Conversely, the propensities of MVV and HIV-1 INs to form tetramers in solution[32,42,43] lead to formation of the larger intasome assemblies (Fig. 3). The lentiviral IN tetramers formed in the absence of DNA are structurally dissimilar from the intasomal spumaviral and δ-retroviral IN tetramers that only assemble on vDNA ends. Finally, the architecture of the intasome is controlled by the properties of the CTD, which shows the lowest extent of amino acid sequence conservation among the three canonical IN domains. While all intasomes harbour a pair of CTDs in the synaptic positions, the remaining CTDs form diverse protein-protein and protein-DNA contacts. In particular, the lentiviral IN CTDs form stable dimers, which in the case of MVV multimerise further into tetrads providing intra- and inter-tetramer contacts within the hexadecameric intasome[32,33,44]. Yet, these biochemical and structural insights shed little light on the functional significance of intasome expansion in the α-, β-retroviruses and lentiviruses. Furthermore, while the intasome structure covers only ~20 bp of vDNA, footprinting studies suggested that as much as 200–250 bp at each vDNA end are protected by a stable protein complex within retroviral PICs[45,46]. Thus, the field would greatly benefit from visualisation of the intasome as part of native retroviral PICs.

The mechanism of integration at the level of the IN active site

The active site of retroviral IN is fully structured only upon engagement of the vDNA end in the context of the intasome[25]. The placement of the 3’ vDNA end in the active site is afforded by fraying of the vDNA termini by insertion of a CCD 3_{10} helix between vDNA strands. Within the CIC, the vDNA duplex is split at the invariant dCdA dinucleotide: while dC remains paired to its dG mate from the opposing vDNA strand, 3’ dA and the scissile dinucleotide are unpaired. The IN active site contains three acidic residues comprising the essential D,D-35-E motif, and even conservative substitutions of these residues ablate the enzymatic functions. The side chains of these residues serve to coordinate the essential pair of Mg^{2+} ions, which facilitate 3’-processing and strand transfer reactions. Due to similarities in coordination geometry and size, Mn^{2+} can substitute for Mg^{2+} *in vitro*. The mechanism of two-metal ion catalysis at DNA phosphodiester bonds was elucidated in the context of the RNase H active site and can be extended to the reactions executed by retroviral IN[28,47,48]. Both reactions catalysed by IN follow the general S_N2 nucleophilic substitution pathway at a phosphorus atom in vDNA (during 3’-processing) or target DNA (during

strand transfer)[49]. During 3'-processing, a water molecule is used as a nucleophile, and the reaction results in cleavage of a phosphodiester bond to reveal a 3' dA hydroxyl group. The latter hydroxyl takes the role of the nucleophile during the strand transfer reaction to attack the phosphodiester backbone of chromosomal DNA, leading to transesterification and covalent joining of vDNA to the chromosome. Both of these reactions were visualised in PFV intasome crystals[28].

3'-Processing reaction.

The PFV intasome crystallised with blunt-ended vDNA underwent catalysis upon incubation of the crystals in the presence of Mg^{2+} or Mn^{2+} salts. Briefly soaking the crystals in Mn^{2+} allowed visualisation of the active site in the Michaelis complex form just prior to 3'-processing[28]. The metal ions coordinate the scissile phosphodiester and a water molecule for in-line nucleophilic attack. In this pre-reaction state, both active site metal ions (designated A and B, Fig. 2B) interact with the scissile phosphodiester bond. While Mn^{2+} ion A, which interacts with the water nucleophile, exists in a near perfect coordination geometry, the coordination of metal B deviates significantly from the ideal, potentially promoting destabilization of the scissile phosphodiester[28,47,48]. Following catalysis and dissociation of the 3' dinucleotide, both metals exist in low energy states. Prior to 3'-processing, the retroviral IN active site is fully occupied, presenting no available pocket, likely explaining why effective catalytic inhibitors of this step have yet to be found.

Target DNA capture and strand transfer.

The PFV intasome assembled using pre-processed vDNA was co-crystallized with a double-stranded DNA oligonucleotide in the form of the target capture complex[28]. The target DNA bound a groove between the active sites of the intasome, nestling on the synaptic CTDs. Sharp bending of the target DNA duplex substantially widened the major groove at the site of integration, permitting the insertion of opposing strand phosphodiester bonds into the intasome active sites (Fig. 2A). Brief incubation of the crystals in the presence of Mn^{2+} ions prior to snap-freezing in liquid nitrogen trapped the active site prior to strand transfer. A pair of metal ions interact with the scissile phosphodiester, and it is metal B that coordinates the vDNA 3'-hydroxyl nucleophile for in-line nucleophilic substitution (Fig. 2B). Thus, the metal ions in the IN active site swap their roles during 3'-processing versus strand transfer. Because transesterification does not change the balance of high-energy bonds, the strand transfer reaction should be reversible. Indeed, retroviral INs can cleave branched DNA constructs that mimic strand transfer products, in a reaction termed disintegration[50]. However, in the context of the intasome, strand transfer results in reconfiguration of the active site followed by dissociation of metal ion B[28]. Relocation of the new phosphodiester bond that joins vDNA to the target protects it from a nucleophilic attack by the 3' hydroxyl of the target DNA[28,29]. The driving force for this conformational change is thought to originate from the tension within sharply bent target DNA. Breaking a pair of phosphodiester bonds in the target duplex during transesterification could allow local relaxation of the structure. Intasomes from several retroviral species as well as prokaryotic and eukaryotic transpososomes have been visualised as post-catalytic strand transfer complexes, all of which display sharp deformations of target DNA[29,30,32,33,37,51–54]. Thus,

the mechanism of channelling transesterification towards the product is likely conserved throughout this highly diverse family of enzymes.

PFV and MLV INs cut the target DNA major groove with 4-bp spacings and strongly disfavour rigid purine/pyrimidine (RY) dinucleotides at the centres of their integration sites[29,55]. Considering the extent of base stacking among the four possible dinucleotide pairs, RY and YR dinucleotide steps are the most and the least physically constrained, respectively. The observed integration target site preference accordingly supports the large target DNA distortion required to accommodate scissile phosphodiester bonds at opposing IN active sites, as observed in the PFV structure[29]. HIV-1 IN cuts target DNA with a 5-bp spacing, and subsequent analyses revealed the RYNRY consensus at the centre of these sites, which would secure a malleable YR dinucleotide close to the middle of the integration site regardless of the central nucleotide's identity[56]. IN residues that interact with target DNA which, when altered, significantly changed base preferences at integration sites[29,55–57]. While the PFV structures provided a plausible empirical explanation of how the IN active site works, quantum mechanical simulation of the reactions may be required to explain the process at a fundamental level.

Integration in the context of chromatin

The bulk of chromosomal DNA exists in the form of nucleosomes, and the nucleosomal structure imposes severe constraints on DNA conformation, greatly restricting accessibility of the major groove[58]. Moreover, DNA wrapping around a histone octamer results in a relatively smooth bend in contrast to the sharply distorted target DNA conformations visualised in retroviral intasomes. Yet, chromatinized DNA was shown to be a suitable, or even preferred target for retroviral integration *in vitro* and *in vivo*[59–64]. Cryo-EM structures of the PFV intasome engaged with a mononucleosome revealed that the latter adapts to present the DNA in a conformation optimal for strand transfer[64,65]. The intasome forms a multivalent interface with the nucleosome, contacting both nucleosomal DNA gyres and the core histones (Fig. 4). These interactions are thought to offset the energy required to shift nucleosomal DNA along the histone octamer core to create the slack for lifting the target DNA at the site of integration (Fig. 4)[65]. Consequently, the PFV intasome displays highly preferred nucleosomal integration sites at superhelix ±3.5 locations that overlay the H2A-H2B heterodimers[64]. The same nucleosomal positions are favoured by yeast retrotransposon Ty1 *in vivo*[63]. However, *in vitro* studies with intasomes of several retroviruses suggested genus-specific modes of chromosomal integration[66], which is consistent with predicted variations in target DNA conformations required to integrate across 4–6 bp spacings[67]. The expansion of the intasome structure may allow lentiviruses to form more extensive interactions with chromatin, possibly spanning several nucleosomes. Thus, it would be of considerable interest to visualize expanded orthoretroviral intasome structures engaged to chromatinized targets.

Selection of a suitable integration site

Retroviral integration into host genomes is non-random and is influenced by several factors including the mode of viral nuclear entry and PIC-host interactions (see REF.[68] for a recent

in-depth review). Within the retroviral family, only lentiviruses have the ability to infect cycling and growth-arrested cells with similar efficiency. In particular, while HIV-1 PICs are actively transported through nuclear pore complexes[69], their MLV and PFV cousins lack this function and require disassembly of the nuclear envelope during the M phase of the cell cycle[70,71]. The well-understood integration preferences of HIV-1 for active transcription units[72] and MLV for promoter/enhancer regions[73–75] are in large part dictated via specific interactions of the respective IN proteins with chromatin-bound host factors.

Nuclear entry and access to host chromatin.

HIV-1 nuclear import is mediated via the viral capsid, which interacts directly with several components of the cellular nuclear import machinery, including the nuclear pore complex components Nup153 and Nup358 (REF.[76–78]). Cleavage and polyadenylation specificity factor 6 (CPSF6), which is a component of the cellular pre-mRNA cleavage and polyadenylation machinery[79], binds the same region of capsid as Nup153 (REF.[78,80]). During nuclear import, CPSF6 is thought to compete with Nup153 for capsid binding to release the viral core from the nuclear pore complex[81]. In addition to its role in nuclear import, CPSF6 plays a critical role in regulating nuclear PIC incursion. Lamina-associated domains (LADs), which tend to be gene-sparse heterochromatin, segregate to the nuclear periphery due to associations with lamina proteins[82]. Conversely, transcriptionally active speckle-associated domains (SPADs) are located towards the interior of the nucleus[83]. Under normal infection conditions, HIV-1 PICs colocalized with nuclear speckles and strongly favoured SPADs for integration while avoiding LADs[84]. In *CPSF6* knockout cells, PICs congregated at the nuclear periphery and uncharacteristically targeted LADs for integration while losing the preference for SPADs[84,85]. The mechanistic basis for the CPSF6-mediated HIV-1 capsid interaction with chromatin is currently unknown. Possibly, the Arg-Ser-like domain of CPSF6, which has been implicated in nuclear speckle condensation, is sufficient to direct PICs to speckles for SPAD-proximal integration[86]. Interestingly, non-primate lentiviruses displayed minimal preferences for integrating into SPADs, and the respective capsid proteins did not bind CPSF6 (REF.[87]). Thus, additional work is required to ascertain whether the involvement of capsid in integration targeting is shared among the non-primate lentiviruses. Depletion of cellular Nup153 or Nup358 also shifted HIV-1 integration away from gene-dense regions of chromatin[88,89], highlighting a functional connection between PIC nuclear entry and integration site selection.

C-terminal proteolytic processing of spumaviral Gag proteins, which is required for infection, occurs adjacent to the chromatin binding sequence (CBS)[90]. The PFV Gag CBS binds nucleosomes via an acidic patch predominantly composed of core histone H2A/H2B heterodimers[91,92], which is similarly leveraged by other nucleosome binding proteins such as herpesvirus latency-associated nuclear antigen[92]. The R540Q substitution in PFV Gag, which altered an Arg residue conserved across this group of H2A/H2B binding proteins, decoupled CBS-nucleosome binding *in vitro* and redirected a large portion of PFV integration events into centromeres[92]. In MLV, the chromatin tethering function is accomplished by p12, a small phosphoprotein that is encoded within the *gag* gene, which acts in a manner seemingly indistinguishable from the PFV Gag CBS[93–95].

IN-host factor interactions.

Given the key role of IN in mediating integration, it is not surprising that IN-host interactions play fundamental roles in retroviral integration targeting. Best understood for the lentiviruses and γ-retroviruses, these interactions influence integration site distribution at the scale of individual genomic features such as transcription units, promoters, and enhancers. The lentiviral IN binding protein LEDGF/p75 significantly stimulated IN strand transfer activity *in vitro*[17,43,96,97]. Knockout of the *PSIP1* gene, which encodes for LEDGF/p75, similarly suppressed levels of vDNA integration in cells several fold[98,99]. The genic proviruses that did form under these conditions uncharacteristically congregated around transcription start sites[87,99–101]. LEDGF/p75 has been shown to bind several pre-mRNA splicing factors and to assist transcriptional elongation on chromatinized DNA *in vitro*[100,102]. Whether the role of LEDGF/p75 in lentiviral integration site targeting is dependent on specific interactions with mRNA splicing and/or transcriptional elongation machineries warrants further investigation.

LEDGF/p75 appears to function as a bimodal tether in lentiviral integration targeting. Elements including a PWWP chromatin reader domain and two copies of an AT-hook DNA binding motif in the N-terminal half of the protein mediate its association with chromatin, while the C-terminal IN-binding domain (IBD) confers binding to IN[103–105]. The CCD is the main IN binding determinant, with the NTD contributing secondary interactions for full binding affinity[106,107]. The LEDGF/p75 IBD is an α-helical domain topologically related to HEAT repeat proteins[108]. Residues Ile356 and Asp366 located on the loop between α-helices 1 and 2 mediate interactions with both IN monomers of the CCD dimer while electropositive residues on the outward face of α-helix 4 interface with electronegative residues of IN NTD α-helix 1 (REF.[107,109]). The LEDGF/p75 PWWP domain can bind nucleosomes *in vitro* that are modified to carry a chemical mimic of trimethylated lysine at position 36 of the histone H3 tail (H3K36me3)[110,111], and spatial distribution of LEDGF/p75 on chromatin coincided with H3K36me2 and H3K36me3 epigenetic marks[102]. Alterations that disrupted LEDGF/p75 binding to chromatin or IN ablated its cofactor function during HIV-1 infection[99,103,112]. LEDGF/p75 similarly tethers several cellular proteins to chromatin, including MLL1 and MML2 methyltransferases, via IBD interactions with a linear IBM (for IBD-binding motif) amino acid sequence[113]. Thus, lentiviruses have leveraged the juxtaposition of independent IN NTDs and CCDs to evolve a functional mimic of the cellular IBM.

MLV and other γ-retroviral INs interact directly with bromo- and extra-terminal domain (BET) proteins, including bromodomain proteins 2, 3 and 4 (REF.[114–116]). Several similarities can be drawn between the γ-retroviral IN-BET and lentiviral IN-LEDGF/p75 interactions. BET proteins significantly stimulate γ-retroviral IN strand transfer activities *in vitro*[67,115,116] and appear to act as bimodal tethers to direct PICs to preferred genomic locations[74,75,114,117,118]. N-terminal bromodomains confer binding to acetylated histone H3 and H4 tails, while the C-terminal extraterminal domain interacts with chromatin modifying enzymes such as histone methyltransferase NSD3 (REF.[119,120]). While LEDGF/p75 interacts primarily with the lentiviral IN CCD dimer interface, BET proteins more simply engage a conserved linear sequence motif located at the C-terminus of γ-retroviral IN

proteins[114–116,121]. The C-terminal peptide of MLV IN folds onto the extraterminal domain, contributing two β-strands to complete a 3-stranded β-sheet within a compact quaternary structure[122,123].

Other retroviral IN binding host factors have also been shown to significantly stimulate the strand transfer activities of purified IN proteins *in vitro*. B56 (also called B' or PR61), a regulatory subunit of protein phosphatase 2A, significantly stimulated δ-retroviral IN[124], while purified histone chaperone FACT complex facilitated ASLV IN activity *in vitro*[125]. Although PP2A-B56 does not bind chromatin directly, it has a range of chromatin-associated substrates[126]. It is currently unclear if these host factors play roles in targeting δ- or α-retroviral integration, which display less dramatic site preferences than their lentiviral and γ-retroviral cousins[68,127,128]. Plausibly, the primary reason for IN to evolve an interaction with a chromatin-associated host factor is to stimulate strand transfer activity at an opportune place and time. Encagement of PICs within capsids during trafficking towards a suitable nuclear location may prevent premature integration[10,11,13]. Be this as it may, B56 proteins additionally served as key structural components of the recent δ-retroviral intasome structures[37,38]. Similarly, LEDGF/p75 was instrumental to reconstitute lentiviral intasomes for structural characterisation[32–36]. While the biochemistry of integration is now well understood, the organisation of the native PIC, the mechanisms of its assembly and disassembly, and the host factors involved in the post-integration DNA repair[20,21] remain to be characterized. Recent reconstruction of reverse transcription and integration in permeabilised HIV-1 virions may provide inroads into solving some of these mysteries[129].

Beyond integration

Mutations in HIV-1 IN are often associated with baffling phenotypes that do not lend themselves to straightforward interpretations. Thus, ablation of HIV-1 IN via a premature stop within the *pol* gene disrupted viral assembly and egress[130], while partial deletions within the IN coding region or point mutations, such as H12N, additionally impaired HIV-1 core formation[131]. The literature is replete with examples of HIV-1 IN point mutants that retain enzymatic activity yet fail to support viral replication (REF.[132] and references therein). These observations led to suggestions that IN plays essential functions outside of the integration process. The pleiotropic phenotypes of HIV-1 IN mutants were generally characterized by a reduction in the levels of reverse transcription in infected cells. On that basis, HIV-1 IN mutants were divided into two groups: class I, which harbours variants specifically defective at the step of integration, and the much larger class II that incorporates variants with a variable degree of reverse transcription defects[133]. Accumulation of 2-LTR circles, the products of end-to-end ligation of unintegrated vDNA ends, in cells infected with class I variants was consistent with a specific defect at the stage of integration[131,134]. In particular, the majority of amino acid substitutions within the HIV-1 IN D,D-35-E catalytic triad fell into this category. At least in some cases, the class II phenotype was accompanied by premature proteolytic maturation, misfolding or perhaps mislocalisation of the mutant Gag-Pol polyprotein[130,135,136]. However, there are many examples of class II HIV-1 IN mutants that incorporated normal levels of mature viral proteins[132]. HIV-1 IN can bind RT and stimulate its activity *in vitro*[137]. Studies using nuclear magnetic resonance identified amino acid residues within the IN CTD responsible for the interaction, and viruses carrying

mutations at these IN positions displayed pronounced defects in reverse transcription[138]. Moreover, in avian α-retroviruses, the large RT subunit retains the entire IN polypeptide due to alternative proteolytic processing[7]. These observations suggest that IN may play a role in reverse transcription, accounting for some of the class II phenotypes.

A mature HIV-1 particle harbours a conical capsid shell pregnant with the ribonucleoprotein complex composed of viral nucleocapsid protein and genomic RNA, which appears as a dark body on electron micrographs of stained virions[139]. Class II IN mutations additionally affected viral particle morphology, with the ribonucleoprotein material found outside of the capsid shell, oftentimes in association with the viral membrane[131,132,139–142]. The mislocalisation of the viral genomic material provides an attractive explanation as to why class II mutants fail to synthetize normal levels of vDNA upon infection[132]. Strikingly, recent cross-linking immunoprecipitation experiments demonstrated that HIV-1 IN directly interacts with viral genomic RNA[132,141]. Examined class II HIV-1 IN mutant proteins failed to engage genomic RNA, though for varying reasons including defects in virion incorporation, tetramerization, and/or RNA binding[132]. Because HIV-1 IN tetramerization was important for RNA binding, the lack of IN-RNA complex within virions seems to be the unifying property of the three subclasses. These data suggest that IN may have a role in organising RNA within infectious particles, which is not inconsistent with the proposed co-factor function during reverse transcription[138]. If the IN-RT interaction is physiologically relevant, it is important to distinguish whether it is required for viral particle morphogenesis or the ensuing step of reverse transcription. Another unanswered question in the field is whether IN-RNA binding plays important roles in particle morphogenesis and/or reverse transcription of other retroviruses. Similar to HIV-1, mutations in MLV IN can strongly reduce reverse transcription[143]. Pertinently, class I and II HIV-1 IN mutant phenotypes are phenocopied by strand transfer and allosteric IN inhibitors, respectively (see next section).

It is becoming increasingly clear that the classification of IN mutants introduced more than 2 decades ago is outdated. Two recent studies described HIV-1 IN mutants with normal reverse transcription and integration, but delayed proviral expression[19,21]. The handover of the hemi-integrated vDNA from IN to cellular machinery for 5’-end joining and subsequent proviral gene expression are unlikely to be disordered processes left to chance and represent an exciting frontier in the field. Accordingly, a recent report indicated that HIV-1 IN might persist on proviral DNA after integration[19].

Small molecule antagonists of HIV-1 integrase

Strand transfer inhibitors

Because human cells lack a close IN homolog, the viral enzyme would appear an ideal target for drug development. However, by contrast to RT and protease, HIV IN inhibitors had to be created without the benefit of prior mechanistic insights or known natural inhibitors[144,145]. Moreover, unlike RT and protease, both of which must catalyse thousands of individual reactions, the IN active site carries out a single strand transfer event during infection. Thus, a successful inhibitor would have an extremely slow dissociation rate, remaining bound for the lifetime of the PIC. Therefore, it is not surprising that 24 years transpired from the discovery of HIV-1 before the first IN strand transfer inhibitor (INSTI), raltegravir, became

available in clinical practice[146]. In the last 14 years, four more INSTIs were approved for treatment of HIV infection: elvitegravir, dolutegravir, bictegravir and cabotegravir[147–150]. The latter three, classified as second-generation compounds, are characterized by reduced susceptibility to viral mutations that render the first-generation drugs, raltegravir and elvitegravir, ineffective[151–153]. Currently, dolutegravir is recommended by the Word Health Organisation for use in first-line combination antiretroviral therapy[154].

The prototypical INSTIs, diketo acids, were discovered over 20 years ago by Merck scientists in a screen for small molecule inhibitors of pre-assembled HIV-1 intasomes[155]. The INSTI pharmacophore consists of a metal chelating core - generally a co-planar triad of oxygen atoms - and a halobenzyl side chain (Fig. 5A, shown in red and blue, respectively)[156]. These compounds bind to the IN active site only when it is engaged with vDNA and compete with target DNA[157]. Because retroviral IN active sites are highly conserved, INSTIs are broadly active across the retroviral family[32,158,159] save the occasional example when an IN residue, such as Ser150 in ASLV IN, confers baseline resistance, in this case to elvitegravir[39,160]. The mode of action of these compounds was visualized in crystals of the PFV intasome[25,161,162] and later refined in HIV-1 and SIV cryo-EM structures[35,36]. The INSTI heteroatoms interact with the essential metal pair in the IN active site, while the halobenzyl sidechain burrows into the vDNA-IN interface, replacing the base of the 3’ adenosine of the processed vDNA end (Fig. 5B). Thus, INSTIs take advantage of the enhanced mobility of the 3’ vDNA nucleotide that is separated from its Watson-Crick pair in the intasome. Because the 3’ adenosine must vacate its normal position to allow inhibitor binding, the INSTIs display slow binding kinetics[163]. The displaced vDNA nucleotide forms a stacking interaction with the metal chelating core of the inhibitor, greatly enhancing the strength of binding[35,36,163]. The second-generation compounds feature extended scaffolds, allowing them to fill the active site more completely, and make additional interactions with the IN backbone[35,36,162]. Due to their interaction with the metal ions, stacking with vDNA bases and extensive van der Waals interactions with IN, INSTIs display unusually long residence in the active site. Thus, the dissociation half-times of dolutegravir and bictegravir from the wild-type HIV-1 intasome were reported to be ~100 and 160 h, respectively[164,165].

Although the majority of INSTI interactions with the intasome occur via immutable features - the catalytic Mg^{2+} ion pair, the invariant dCdA vDNA dinucleotide, and the IN protein backbone - the virus can develop resistance to these compounds. Three main genetic pathways lead to high-level resistance to raltegravir, involving changes of HIV IN residues Tyr143, Gln148, or Asn155 (REF.[166]). The oxadiazole group of raltegravir, unique to this inhibitor, stacks with the side chain of Tyr143, explaining why amino acid substitutions such as Y143R/C cause resistance to this drug[161]. Yet, the most nefarious mutations that contribute to complete or partial loss of INSTI susceptibility involve changes of Gln148 and Asn155, which do not directly contact the drugs. The primate lentiviral intasome structures revealed that these residues participate in the secondary coordination spheres of the metal cations (Fig. 5B). Accordingly, substitutions at these positions exert their effects by destabilizing the coordination spheres of the Mg^{2+} ion pair, indirectly affecting drug binding[35]. Thus, the sensitivity of metal ions for the geometry and electronic properties of the ligand chelating cluster underpins the major mechanism of INSTI resistance. Despite

the remarkable success of INSTIs, it is becoming increasingly clear that HIV can acquire high-level resistance even to the second-generation compounds, highlighting the need to continue developing this family as well as exploring new targets[167–169]. The urgency is underscored by the rising prevalence of HIV resistance to RT inhibitors, which comprise the backbone of combination antiretroviral therapy[170].

Allosteric inhibitors

Allosteric IN inhibitors (ALLINIs) are a distinct class of small molecule HIV-1 IN inhibitors (Fig. 6A). ALLINIs engage the LEDGF/p75 binding cleft at the IN CCD dimer interface, which is distal from the IN active site[109,171]. Lead compounds in this class were discovered using two very different high-throughput approaches: a screen for inhibitors of HIV-1 IN 3'-processing activity[172], and structure-based design of small molecule inhibitors of the HIV-1 IN-LEDGF/p75 interaction[171]. Many, though not all, of the ensuing compounds effectively inhibit IN-LEDGF/p75 binding *in vitro*[171,173–175]. The common biochemical consequence of ALLINI binding is aberrant HIV-1 IN hyper-multimerization[140,142,174–181]. In addition to ALLINI, other utilized acronyms for this drug class include LEDGIN for LEDGF-IN site[171], NCINI for non-catalytic IN inhibitor[142], IN-LAI for IN-LEDGF allosteric inhibitor[174], and MINI for multimerization IN inhibitor[177]. Compounds of the latter class do not effectively antagonise HIV-1 IN-LEDGF/p75 binding. Although ALLINIs have yet to advance in clinical trials, several compounds display exquisite low-nM potencies to inhibit HIV-1 replication *in vitro*[172,180,182,183].

Counterintuitively, the underlying mode of ALLINI action is not via inhibition of integration[140,142,174–177,179–181]. Critical clues for dissecting the mechanism of ALLINI antiviral activity came from staging the ingress versus egress phases of HIV-1 replication. Egress can be modelled by plasmid DNA transfection, which supports infectious virus particle formation while bypassing the early steps of virus entry, reverse transcription and integration. The early infection steps are readily staged via so-called single-round virus constructs that harbour a reporter gene for rapid readout of virus infection and a deletion in a late-acting gene such as *env*. Functional envelope glycoprotein is provided *in trans* during transfection, and the viruses are thus restricted to the single round of integration and infection. Under these conditions, it became clear that HIV-1 was more susceptible to ALLINIs when exposed to the compounds during egress as compared to ingress[140,142,174–176]. MINI compounds moreover were ineffective at inhibiting the early stage of HIV-1 infection[177]. Detailed analysis of virus particles produced in the presence of ALLINIs revealed that although viral RNA and proteins were incorporated properly, IN-RNA binding was disrupted, and the ribonucleoprotein complex was located eccentrically (Fig. 6B)[140–142,174–176]. Accordingly, ALLINIs are inhibitors of HIV-1 maturation/morphogenesis[139]. Consistent with the known role of LEDGF/p75 in HIV-1 integration targeting, genic integration targeting during HIV-1 ingress was suppressed by ALLINIs[177,184,185]. Resultant proviruses were marginally recalcitrant to transcriptional activation via latency reversal agents[183,185], indicating that ALLINIs could potentially be employed as preexposure prophylactic compounds to reduce both the size and reactivation potential of the latent viral reservoir.

ALLINIs display limited structural diversity, sharing a warhead comprising three chemical functions: a carboxylate, a compact aliphatic (typically tert-butoxy) sidechain and a relatively bulky hydrophobic (usually aromatic) group (Fig. 6A, shown in red, green and blue, respectively). Many ALLINIs were co-crystallised with the isolated HIV-1 IN CCD, and the structures revealed the molecules grip onto the LEDGF/p75 IBD binding pocket at the IN CCD dimerization interface[171,172,174,175,180,186,187]. The carboxylate establishes a bidentate hydrogen bond with main chain amides of IN Glu170 and His171, perfectly recuperating the interactions made by Asp366 of LEDGF/p75 (Fig. 6C)[109]. The bulky hydrophobic side chain contacts several hydrophobic residues from both subunits of the IN dimer, thus mimicking Ile365 of LEDGF/p75. The compact aliphatic side chain of the inhibitor explores the base of the pocket. Resistance to ALLINIs maps to the IN CCD dimer interface in the vicinity of the LEDGF/p75 binding site[171,173,179,182,188]. Some of the residues that line the ALLINI binding pocket are highly polymorphic among circulating HIV-1 strains, inspiring efforts to design molecules that remain active in the face of such amino acid sequence variability[186,187].

While the primary ALLINI binding site on HIV-1 IN CCD is well characterized, the mechanism of IN aggregation by these compounds is far from clear. A co-crystal structure of full-length HIV-1 IN with GSK1264 revealed that binding of this ALLINI at the CCD dimer interface contributes to a novel interface for the CTD, leading to formation of linear IN polymers (Fig. 6D)[179]. Solution measurements using mass spectrometry footprinting techniques, mutagenesis and computational chemistry strongly support this mode of IN aggregation by ALLINIs[178,189]. A more recent study detected branching of the GSK1264-induced IN polymers via homomeric (CTD-CTD) interactions[181]. Unfortunately, the limited resolution of the available X-ray diffraction data precluded detailed description of the interactions involving the small molecule with the CTD[179]. The absence of structural similarity between ALLINIs outside of the common warhead (Fig. 6A) likely contributes to differential IN aggregation mechanisms reported for some of these compounds[180], calling for more research and development of this fascinating class of antiviral agents.

Conclusions and Perspectives

The PFV intasome crystal structures reported in the early 2010s supercharged the field of retroviral integration structural biology[25,29,161]. Not only did they provide the first glimpses of the active DNA recombination machine, but they also helped to elucidate the mechanism of INSTI drug action. By leveraging the resolution revolution, subsequent single-particle cryo-EM structures of α-, β-, δ-retroviral and lentiviral intasomes revealed unexpected diversity among the number of IN molecules required to construct the active CIC[30–33,37,38]. Although the α-retroviral strand transfer intasome complex structure was solved by X-ray crystallography[30], it seems highly doubtful, given their propensities to artificially stack upon one another in solution[35,36], that primate lentiviral intasomes would ever crystallize.

In the coming years, it will be good to see research uniting integration with the anterior (reverse transcription and PIC assembly) and posterior (repair of the chromosomal lesions initially flanking the provirus, chromatinization and the onset of transcription) events of the retroviral life cycle. We also hope that the exciting developments in *in situ* cryo-electron

tomography[190] and super-resolution correlative microscopy[191] will allow to visualise the viral nucleoprotein complexes in their natural environment, provided many remaining technical challenges can be solved. Such a fit may help to clarify how the PIC avoids suicide by autointegration and how it interacts with the cellular trafficking machinery and chromatin.

The key translational aspect of retroviral IN research is the ability to aid in the development of retroviral/lentiviral vectors and anti-HIV/AIDS drugs. Recent work uncovered surprising differences between mechanisms involved in targeting retroviral integration, even within a single genus[87]. These results will inform the design of vectors for future gene therapy applications. Although highly potent compounds, ALLINIs to date have proven too toxic for in-depth clinical trials[183], warranting attempts to develop safer compounds that can hopefully advance to the clinic. The recent primate lentiviral intasome structures importantly resolved INSTI occupancies in the most relevant retroviral IN active sites and helped to elucidate the structural basis for INSTI drug resistance[35,36]. Visualising the HIV-1 intasome bound to INSTIs at even higher resolution will be necessary to refine the precise positions of the key water molecules that complete the metal chelating cluster. These basic research outcomes will accordingly be invaluable to design improved inhibitors as long as INSTI-containing regimens remain topical to the ongoing fight against the HIV pandemic.

Acknowledgements

We apologize to colleagues whose work we were unable to cite due to space constraints. This work was supported by US National Institutes of Health grant P50 AI150481 (P.C. and A.N.E.), R01 AI070042 (A.N.E.), the Francis Crick Institute (P.C), which receives its core funding from Cancer Research UK (FC001061), the UK Medical Research Council (FC001061), and the Wellcome Trust (FC001061), and Wellcome Trust Investigator award 107005/Z/15Z (G.N.M.).

References

1. Poiesz BJ et al. Detection and isolation of type C retrovirus particles from fresh and cultured lymphocytes of a patient with cutaneous T-cell lymphoma. Proc Natl Acad Sci U S A 77, 7415–7419, doi:10.1073/pnas.77.12.7415 (1980). [PubMed: 6261256]
2. Coffin JM The discovery of HTLV-1, the first pathogenic human retrovirus. Proc Natl Acad Sci U S A 112, 15525–15529, doi:10.1073/pnas.1521629112 (2015). [PubMed: 26696625]
3. Barre-Sinoussi F et al. Isolation of a T-lymphotropic retrovirus from a patient at risk for acquired immune deficiency syndrome (AIDS). Science 220, 868–871, doi:10.1126/science.6189183 (1983). [PubMed: 6189183]
4. Gallo RC et al. Isolation of human T-cell leukemia virus in acquired immune deficiency syndrome (AIDS). Science 220, 865–867, doi:10.1126/science.6601823 (1983). [PubMed: 6601823]
5. Baltimore D RNA-dependent DNA polymerase in virions of RNA tumour viruses. Nature 226, 1209–1211, doi:10.1038/2261209a0 (1970). [PubMed: 4316300]
6. Temin HM & Mizutani S RNA-dependent DNA polymerase in virions of Rous sarcoma virus. Nature 226, 1211–1213, doi:10.1038/2261211a0 (1970). [PubMed: 4316301]
7. Schiff RD & Grandgenett DP Virus-coded origin of a 32,000-dalton protein from avian retrovirus cores: structural relatedness of p32 and the beta polypeptide of the avian retrovirus DNA polymerase. J Virol 28, 279–291, doi:10.1128/JVI.28.1.279-291.1978 (1978). [PubMed: 81316]
8. Jacks T et al. Characterization of ribosomal frameshifting in HIV-1 gag-pol expression. Nature 331, 280–283, doi:10.1038/331280a0 (1988). [PubMed: 2447506]

9. Carlson LA et al. Three-dimensional analysis of budding sites and released virus suggests a revised model for HIV-1 morphogenesis. Cell Host Microbe 4, 592–599, doi:10.1016/j.chom.2008.10.013 (2008). [PubMed: 19064259]
10. Chen NY et al. HIV-1 capsid is involved in post-nuclear entry steps. Retrovirology 13, 28, doi:10.1186/s12977-016-0262-0 (2016). [PubMed: 27107820]
11. Burdick RC et al. HIV-1 uncoats in the nucleus near sites of integration. Proc Natl Acad Sci U S A 117, 5486–5493, doi:10.1073/pnas.1920631117 (2020). [PubMed: 32094182]
12. Zila V et al. Cone-shaped HIV-1 capsids are transported through intact nuclear pores. Cell, doi:10.1016/j.cell.2021.01.025 (2021).
13. Li C, Burdick RC, Nagashima K, Hu WS & Pathak VK HIV-1 cores retain their integrity until minutes before uncoating in the nucleus. Proc Natl Acad Sci U S A 118, doi:10.1073/pnas.2019467118 (2021).
14. Lee MS & Craigie R A previously unidentified host protein protects retroviral DNA from autointegration. Proc Natl Acad Sci U S A 95, 1528–1533, doi:10.1073/pnas.95.4.1528 (1998). [PubMed: 9465049]
15. Lin CW & Engelman A The barrier-to-autointegration factor is a component of functional human immunodeficiency virus type 1 preintegration complexes. J Virol 77, 5030–5036, doi:10.1128/jvi.77.8.5030-5036.2003 (2003). [PubMed: 12663813]
16. Farnet CM & Bushman FD HIV-1 cDNA integration: requirement of HMG I(Y) protein for function of preintegration complexes in vitro. Cell 88, 483–492, doi:10.1016/s0092-8674(00)81888-7 (1997). [PubMed: 9038339]
17. Llano M et al. LEDGF/p75 determines cellular trafficking of diverse lentiviral but not murine oncoretroviral integrase proteins and is a component of functional lentiviral preintegration complexes. J Virol 78, 9524–9537, doi:10.1128/JVI.78.17.9524-9537.2004 (2004). [PubMed: 15308744]
18. Machida S et al. Exploring histone loading on HIV DNA reveals a dynamic nucleosome positioning between unintegrated and integrated viral genome. Proc Natl Acad Sci U S A 117, 6822–6830, doi:10.1073/pnas.1913754117 (2020). [PubMed: 32161134]
19. Winans S & Goff SP Mutations altering acetylated residues in the CTD of HIV-1 integrase cause defects in proviral transcription at early times after integration of viral DNA. PLoS Pathog 16, e1009147, doi:10.1371/journal.ppat.1009147 (2020). [PubMed: 33351861]
20. Yoder KE & Bushman FD Repair of gaps in retroviral DNA integration intermediates. J Virol 74, 11191–11200, doi:10.1128/jvi.74.23.11191-11200.2000 (2000). [PubMed: 11070016]
21. Knyazhanskaya E et al. NHEJ pathway is involved in post-integrational DNA repair due to Ku70 binding to HIV-1 integrase. Retrovirology 16, 30, doi:10.1186/s12977-019-0492-z (2019). [PubMed: 31690330]
22. Cai M et al. Solution structure of the N-terminal zinc binding domain of HIV-1 integrase. Nat Struct Biol 4, 567–577, doi:10.1038/nsb0797-567 (1997). [PubMed: 9228950]
23. Dyda F et al. Crystal structure of the catalytic domain of HIV-1 integrase: similarity to other polynucleotidyl transferases. Science 266, 1981–1986, doi:10.1126/science.7801124 (1994). [PubMed: 7801124]
24. Eijkelenboom AP et al. The DNA-binding domain of HIV-1 integrase has an SH3-like fold. Nat Struct Biol 2, 807–810, doi:10.1038/nsb0995-807 (1995). [PubMed: 7552753]
25. Hare S, Gupta SS, Valkov E, Engelman A & Cherepanov P Retroviral intasome assembly and inhibition of DNA strand transfer. Nature 464, 232–236, doi:10.1038/nature08784 (2010). [PubMed: 20118915]
26. Guan R et al. X-ray crystal structure of the N-terminal region of Moloney murine leukemia virus integrase and its implications for viral DNA recognition. Proteins 85, 647–656, doi:10.1002/prot.25245 (2017). [PubMed: 28066922]
27. Harshey RM The Mu story: how a maverick phage moved the field forward. Mob DNA 3, 21, doi:10.1186/1759-8753-3-21 (2012). [PubMed: 23217166]
28. Hare S, Maertens GN & Cherepanov P 3'-processing and strand transfer catalysed by retroviral integrase in crystallo. EMBO J 31, 3020–3028, doi:10.1038/emboj.2012.118 (2012). [PubMed: 22580823]

29. Maertens GN, Hare S & Cherepanov P The mechanism of retroviral integration from X-ray structures of its key intermediates. Nature 468, 326–329, doi:10.1038/nature09517 (2010). [PubMed: 21068843]
30. Yin Z et al. Crystal structure of the Rous sarcoma virus intasome. Nature 530, 362–366, doi:10.1038/nature16950 (2016). [PubMed: 26887497]
31. Ballandras-Colas A et al. Cryo-EM reveals a novel octameric integrase structure for betaretroviral intasome function. Nature 530, 358–361, doi:10.1038/nature16955 (2016). [PubMed: 26887496]
32. Ballandras-Colas A et al. A supramolecular assembly mediates lentiviral DNA integration. Science 355, 93–95, doi:10.1126/science.aah7002 (2017). [PubMed: 28059770]
33. Passos DO et al. Cryo-EM structures and atomic model of the HIV-1 strand transfer complex intasome. Science 355, 89–92, doi:10.1126/science.aah5163 (2017). [PubMed: 28059769]
34. Li M et al. A Peptide Derived from Lens Epithelium-Derived Growth Factor Stimulates HIV-1 DNA Integration and Facilitates Intasome Structural Studies. J Mol Biol 432, 2055–2066, doi:10.1016/j.jmb.2020.01.040 (2020). [PubMed: 32061936]
35. Cook NJ et al. Structural basis of second-generation HIV integrase inhibitor action and viral resistance. Science 367, 806, doi:10.1126/science.aay4919 (2020). [PubMed: 32001525]
36. Passos DO et al. Structural basis for strand-transfer inhibitor binding to HIV intasomes. Science 367, 810, doi:10.1126/science.aay8015 (2020). [PubMed: 32001521]
37. Bhatt V et al. Structural basis of host protein hijacking in human T-cell leukemia virus integration. Nature Communications 11, 3121, doi:10.1038/s41467-020-16963-6 (2020).
38. Barski MS et al. Cryo-EM structure of the deltaretroviral intasome in complex with the PP2A regulatory subunit B56γ. Nature Communications 11, 5043, doi:10.1038/s41467-020-18874-y (2020).
39. Pandey KK et al. Cryo-EM structure of the Rous sarcoma virus octameric cleaved synaptic complex intasome. Commun. Biol, in press.
40. McCord M et al. Purification of recombinant Rous sarcoma virus integrase possessing physical and catalytic properties similar to virion-derived integrase. Protein Expr Purif 14, 167–177, doi:10.1006/prep.1998.0954 (1998). [PubMed: 9790878]
41. Ballandras-Colas A, Naraharisetty H, Li X, Serrao E & Engelman A Biochemical characterization of novel retroviral integrase proteins. PLoS One 8, e76638, doi:10.1371/journal.pone.0076638 (2013). [PubMed: 24124581]
42. Hare S et al. Structural basis for functional tetramerization of lentiviral integrase. PLoS Pathog 5, e1000515, doi:10.1371/journal.ppat.1000515 (2009). [PubMed: 19609359]
43. Cherepanov P et al. HIV-1 integrase forms stable tetramers and associates with LEDGF/p75 protein in human cells. J Biol Chem 278, 372–381, doi:10.1074/jbc.M209278200 (2003). [PubMed: 12407101]
44. Lodi PJ et al. Solution structure of the DNA binding domain of HIV-1 integrase. Biochemistry 34, 9826–9833, doi:10.1021/bi00031a002 (1995). [PubMed: 7632683]
45. Wei SQ, Mizuuchi K & Craigie R A large nucleoprotein assembly at the ends of the viral DNA mediates retroviral DNA integration. EMBO J 16, 7511–7520, doi:10.1093/emboj/16.24.7511 (1997). [PubMed: 9405379]
46. Chen H, Wei SQ & Engelman A Multiple integrase functions are required to form the native structure of the human immunodeficiency virus type I intasome. J Biol Chem 274, 17358–17364, doi:10.1074/jbc.274.24.17358 (1999). [PubMed: 10358097]
47. Yang W, Lee JY & Nowotny M Making and breaking nucleic acids: two-Mg2+-ion catalysis and substrate specificity. Mol Cell 22, 5–13, doi:10.1016/j.molcel.2006.03.013 (2006). [PubMed: 16600865]
48. Nowotny M & Yang W Stepwise analyses of metal ions in RNase H catalysis from substrate destabilization to product release. EMBO J 25, 1924–1933, doi:10.1038/sj.emboj.7601076 (2006). [PubMed: 16601679]
49. Engelman A, Mizuuchi K & Craigie R HIV-1 DNA integration: mechanism of viral DNA cleavage and DNA strand transfer. Cell 67, 1211–1221, doi:10.1016/0092-8674(91)90297-c (1991). [PubMed: 1760846]

50. Chow SA & Brown PO Juxtaposition of two viral DNA ends in a bimolecular disintegration reaction mediated by multimers of human immunodeficiency virus type 1 or murine leukemia virus integrase. J Virol 68, 7869–7878, doi:10.1128/JVI.68.12.7869-7878.1994 (1994). [PubMed: 7966577]
51. Richardson JM, Colloms SD, Finnegan DJ & Walkinshaw MD Molecular architecture of the Mos1 paired-end complex: the structural basis of DNA transposition in a eukaryote. Cell 138, 1096–1108, doi:10.1016/j.cell.2009.07.012 (2009). [PubMed: 19766564]
52. Montano SP, Pigli YZ & Rice PA The mu transpososome structure sheds light on DDE recombinase evolution. Nature 491, 413–417, doi:10.1038/nature11602 (2012). [PubMed: 23135398]
53. Morris ER, Grey H, McKenzie G, Jones AC & Richardson JM A bend, flip and trap mechanism for transposon integration. Elife 5, doi:10.7554/eLife.15537 (2016).
54. Ghanim GE, Kellogg EH, Nogales E & Rio DC Structure of a P element transposase-DNA complex reveals unusual DNA structures and GTP-DNA contacts. Nat Struct Mol Biol 26, 1013–1022, doi:10.1038/s41594-019-0319-6 (2019). [PubMed: 31659330]
55. Aiyer S et al. Structural and sequencing analysis of local target DNA recognition by MLV integrase. Nucleic Acids Res 43, 5647–5663, doi:10.1093/nar/gkv410 (2015). [PubMed: 25969444]
56. Serrao E et al. Integrase residues that determine nucleotide preferences at sites of HIV-1 integration: implications for the mechanism of target DNA binding. Nucleic Acids Res 42, 5164–5176, doi:10.1093/nar/gku136 (2014). [PubMed: 24520116]
57. Demeulemeester J et al. HIV-1 integrase variants retarget viral integration and are associated with disease progression in a chronic infection cohort. Cell Host Microbe 16, 651–662, doi:10.1016/j.chom.2014.09.016 (2014). [PubMed: 25525795]
58. Luger K, Mader AW, Richmond RK, Sargent DF & Richmond TJ Crystal structure of the nucleosome core particle at 2.8 A resolution. Nature 389, 251–260, doi:10.1038/38444 (1997). [PubMed: 9305837]
59. Pryciak PM & Varmus HE Nucleosomes, DNA-binding proteins, and DNA sequence modulate retroviral integration target site selection. Cell 69, 769–780, doi:10.1016/0092-8674(92)90289-o (1992). [PubMed: 1317268]
60. Pruss D, Bushman FD & Wolffe AP Human immunodeficiency virus integrase directs integration to sites of severe DNA distortion within the nucleosome core. Proc Natl Acad Sci U S A 91, 5913–5917, doi:10.1073/pnas.91.13.5913 (1994). [PubMed: 8016088]
61. Wang GP, Ciuffi A, Leipzig J, Berry CC & Bushman FD HIV integration site selection: analysis by massively parallel pyrosequencing reveals association with epigenetic modifications. Genome Res 17, 1186–1194, doi:10.1101/gr.6286907 (2007). [PubMed: 17545577]
62. Roth SL, Malani N & Bushman FD Gammaretroviral integration into nucleosomal target DNA in vivo. J Virol 85, 7393–7401, doi:10.1128/JVI.00635-11 (2011). [PubMed: 21561906]
63. Baller JA, Gao J, Stamenova R, Curcio MJ & Voytas DF A nucleosomal surface defines an integration hotspot for the Saccharomyces cerevisiae Ty1 retrotransposon. Genome Res 22, 704–713, doi:10.1101/gr.129585.111 (2012). [PubMed: 22219511]
64. Maskell DP et al. Structural basis for retroviral integration into nucleosomes. Nature 523, 366–369, doi:10.1038/nature14495 (2015). [PubMed: 26061770]
65. Wilson MD et al. Retroviral integration into nucleosomes through DNA looping and sliding along the histone octamer. Nat Commun 10, 4189, doi:10.1038/s41467-019-12007-w (2019). [PubMed: 31519882]
66. Benleulmi MS et al. Intasome architecture and chromatin density modulate retroviral integration into nucleosome. Retrovirology 12, 13, doi:10.1186/s12977-015-0145-9 (2015). [PubMed: 25807893]
67. Serrao E, Ballandras-Colas A, Cherepanov P, Maertens GN & Engelman AN Key determinants of target DNA recognition by retroviral intasomes. Retrovirology 12, 39, doi:10.1186/s12977-015-0167-3 (2015). [PubMed: 25924943]
68. Engelman AN, Maertens GN in Retrovirus-Cell Interactions (ed Parent LJ) Ch. 4, 163–199 (Elsevier Inc., 2018).

69. Bukrinsky MI et al. Active nuclear import of human immunodeficiency virus type 1 preintegration complexes. Proc Natl Acad Sci U S A 89, 6580–6584, doi:10.1073/pnas.89.14.6580 (1992). [PubMed: 1631159]
70. Roe T, Reynolds TC, Yu G & Brown PO Integration of murine leukemia virus DNA depends on mitosis. Embo j 12, 2099–2108 (1993). [PubMed: 8491198]
71. Bieniasz PD, Weiss RA & McClure MO Cell cycle dependence of foamy retrovirus infection. J Virol 69, 7295–7299, doi:10.1128/JVI.69.11.7295-7299.1995 (1995). [PubMed: 7474157]
72. Schroder AR et al. HIV-1 integration in the human genome favors active genes and local hotspots. Cell 110, 521–529, doi:10.1016/s0092-8674(02)00864-4 (2002). [PubMed: 12202041]
73. Wu X, Li Y, Crise B & Burgess SM Transcription start regions in the human genome are favored targets for MLV integration. Science 300, 1749–1751, doi:10.1126/science.1083413 (2003). [PubMed: 12805549]
74. De Ravin SS et al. Enhancers are major targets for murine leukemia virus vector integration. J Virol 88, 4504–4513, doi:10.1128/jvi.00011-14 (2014). [PubMed: 24501411]
75. LaFave MC et al. MLV integration site selection is driven by strong enhancers and active promoters. Nucleic Acids Research 42, 4257–4269, doi:10.1093/nar/gkt1399 (2014). [PubMed: 24464997]
76. Yamashita M & Emerman M Capsid is a dominant determinant of retrovirus infectivity in nondividing cells. J Virol 78, 5670–5678, doi:10.1128/jvi.78.11.5670-5678.2004 (2004). [PubMed: 15140964]
77. Schaller T et al. HIV-1 capsid-cyclophilin interactions determine nuclear import pathway, integration targeting and replication efficiency. PLoS Pathog 7, e1002439, doi:10.1371/journal.ppat.1002439 (2011). [PubMed: 22174692]
78. Matreyek KA, Yucel SS, Li X & Engelman A Nucleoporin NUP153 phenylalanine-glycine motifs engage a common binding pocket within the HIV-1 capsid protein to mediate lentiviral infectivity. PLoS Pathog 9, e1003693, doi:10.1371/journal.ppat.1003693 (2013). [PubMed: 24130490]
79. Rüegsegger U, Beyer K & Keller W Purification and characterization of human cleavage factor Im involved in the 3' end processing of messenger RNA precursors. J. Biol. Chem 271, 6107–6113, doi:10.1074/jbc.271.11.6107 (1996). [PubMed: 8626397]
80. Price AJ et al. Host cofactors and pharmacologic ligands share an essential interface in HIV-1 capsid that is lost upon disassembly. PLoS Pathog. 10, e1004459, doi:10.1371/journal.ppat.1004459 (2014). [PubMed: 25356722]
81. Bejarano DA et al. HIV-1 nuclear import in macrophages is regulated by CPSF6-capsid interactions at the nuclear pore complex. eLife 8, e41800, doi:10.7554/eLife.41800 (2019). [PubMed: 30672737]
82. Bizhanova A & Kaufman PD Close to the edge: Heterochromatin at the nucleolar and nuclear peripheries. Biochim Biophys Acta Gene Regul Mech 1864, 194666, doi:10.1016/j.bbagrm.2020.194666 (2021). [PubMed: 33307247]
83. Chen Y et al. Mapping 3D genome organization relative to nuclear compartments using TSA-Seq as a cytological ruler. J. Cell Biol 217, 4025–4048, doi:10.1083/jcb.201807108 (2018). [PubMed: 30154186]
84. Francis AC et al. HIV-1 replication complexes accumulate in nuclear speckles and integrate into speckle-associated genomic domains. Nat. Commun 11, 3505, doi:10.1038/s41467-020-17256-8 (2020). [PubMed: 32665593]
85. Achuthan V et al. Capsid-CPSF6 interaction licenses nuclear HIV-1 trafficking to sites of viral DNA integration. Cell Host Microbe 24, 392–404.e398, doi:10.1016/j.chom.2018.08.002 (2018). [PubMed: 30173955]
86. Greig JA et al. Arginine-Enriched Mixed-Charge Domains Provide Cohesion for Nuclear Speckle Condensation. Mol Cell 77, 1237–1250.e1234, doi:10.1016/j.molcel.2020.01.025 (2020). [PubMed: 32048997]
87. Li W et al. CPSF6-Dependent Targeting of Speckle-Associated Domains Distinguishes Primate from Nonprimate Lentiviral Integration. mBio 11, doi:10.1128/mBio.02254-20 (2020).

88. Koh Y et al. Differential effects of human immunodeficiency virus type 1 capsid and cellular factors nucleoporin 153 and LEDGF/p75 on the efficiency and specificity of viral DNA integration. J Virol 87, 648–658, doi:10.1128/JVI.01148-12 (2013). [PubMed: 23097450]

89. Ocwieja KE et al. HIV integration targeting: a pathway involving Transportin-3 and the nuclear pore protein RanBP2. PLoS Pathog 7, e1001313, doi:10.1371/journal.ppat.1001313 (2011). [PubMed: 21423673]

90. Müllers E The foamy virus Gag proteins: what makes them different? Viruses 5, 1023–1041, doi:10.3390/v5041023 (2013). [PubMed: 23531622]

91. Tobaly-Tapiero J et al. Chromatin tethering of incoming foamy virus by the structural Gag protein. Traffic 9, 1717–1727, doi:10.1111/j.1600-0854.2008.00792.x (2008). [PubMed: 18627573]

92. Lesbats P et al. Structural basis for spumavirus GAG tethering to chromatin. Proc Natl Acad Sci U S A 114, 5509–5514, doi:10.1073/pnas.1621159114 (2017). [PubMed: 28490494]

93. Elis E, Ehrlich M, Prizan-Ravid A, Laham-Karam N & Bacharach E p12 tethers the murine leukemia virus pre-integration complex to mitotic chromosomes. PLoS Pathog 8, e1003103, doi:10.1371/journal.ppat.1003103 (2012). [PubMed: 23300449]

94. Schneider WM et al. Viral DNA tethering domains complement replication-defective mutations in the p12 protein of MuLV Gag. Proc Natl Acad Sci U S A 110, 9487–9492, doi:10.1073/pnas.1221736110 (2013). [PubMed: 23661057]

95. Wanaguru M, Barry DJ, Benton DJ, O'Reilly NJ & Bishop KN Murine leukemia virus p12 tethers the capsid-containing pre-integration complex to chromatin by binding directly to host nucleosomes in mitosis. PLoS Pathog 14, e1007117, doi:10.1371/journal.ppat.1007117 (2018). [PubMed: 29906285]

96. Busschots K et al. The interaction of LEDGF/p75 with integrase is lentivirus-specific and promotes DNA binding. J Biol Chem 280, 17841–17847, doi:10.1074/jbc.M411681200 (2005). [PubMed: 15749713]

97. Cherepanov P LEDGF/p75 interacts with divergent lentiviral integrases and modulates their enzymatic activity in vitro. Nucleic Acids Res 35, 113–124, doi:10.1093/nar/gkl885 (2007). [PubMed: 17158150]

98. Marshall HM et al. Role of PSIP1/LEDGF/p75 in lentiviral infectivity and integration targeting. PLoS One 2, e1340, doi:10.1371/journal.pone.0001340 (2007). [PubMed: 18092005]

99. Shun MC et al. LEDGF/p75 functions downstream from preintegration complex formation to effect gene-specific HIV-1 integration. Genes Dev 21, 1767–1778, doi:10.1101/gad.1565107 (2007). [PubMed: 17639082]

100. Singh PK et al. LEDGF/p75 interacts with mRNA splicing factors and targets HIV-1 integration to highly spliced genes. Genes Dev 29, 2287–2297, doi:10.1101/gad.267609.115 (2015). [PubMed: 26545813]

101. Sowd GA et al. A critical role for alternative polyadenylation factor CPSF6 in targeting HIV-1 integration to transcriptionally active chromatin. Proc. Natl. Acad. Sci. USA 113, E1054–1063, doi:10.1073/pnas.1524213113 10.1073/pnas.1524213113. Epub 2016 Feb 8. (2016). [PubMed: 26858452]

102. LeRoy G et al. LEDGF and HDGF2 relieve the nucleosome-induced barrier to transcription in differentiated cells. Sci. Adv 5, eaay3068, doi:10.1126/sciadv.aay3068 (2019).

103. Llano M et al. Identification and characterization of the chromatin-binding domains of the HIV-1 integrase interactor LEDGF/p75. J Mol Biol 360, 760–773, doi:10.1016/j.jmb.2006.04.073 (2006). [PubMed: 16793062]

104. Turlure F, Maertens G, Rahman S, Cherepanov P & Engelman A A tripartite DNA-binding element, comprised of the nuclear localization signal and two AT-hook motifs, mediates the association of LEDGF/p75 with chromatin in vivo. Nucleic Acids Res 34, 1653–1665, doi:10.1093/nar/gkl052 (2006).

105. Cherepanov P, Devroe E, Silver PA & Engelman A 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, doi:10.1074/jbc.M406307200 (2004). [PubMed: 15371438]

106. Maertens G et al. LEDGF/p75 is essential for nuclear and chromosomal targeting of HIV-1 integrase in human cells. J Biol Chem 278, 33528–33539, doi:10.1074/jbc.M303594200 (2003). [PubMed: 12796494]
107. Hare S et al. 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, doi:10.1371/journal.ppat.1000259 (2009). [PubMed: 19132083]
108. Cherepanov P et al. Solution structure of the HIV-1 integrase-binding domain in LEDGF/p75. Nat Struct Mol Biol 12, 526–532, doi:10.1038/nsmb937 (2005). [PubMed: 15895093]
109. Cherepanov P, Ambrosio AL, Rahman S, Ellenberger T & Engelman A Structural basis for the recognition between HIV-1 integrase and transcriptional coactivator p75. Proc Natl Acad Sci U S A 102, 17308–17313, doi:10.1073/pnas.0506924102 (2005). [PubMed: 16260736]
110. Eidahl JO et al. Structural basis for high-affinity binding of LEDGF PWWP to mononucleosomes. Nucleic Acids Res 41, 3924–3936, doi:10.1093/nar/gkt074 (2013). [PubMed: 23396443]
111. Wang H, Farnung L, Dienemann C & Cramer P Structure of H3K36-methylated nucleosome-PWWP complex reveals multivalent cross-gyre binding. Nat Struct Mol Biol 27, 8–13, doi:10.1038/s41594-019-0345-4 (2020). [PubMed: 31819277]
112. Shun MC et al. Identification and characterization of PWWP domain residues critical for LEDGF/p75 chromatin binding and human immunodeficiency virus type 1 infectivity. J Virol 82, 11555–11567, doi:10.1128/JVI.01561-08 (2008). [PubMed: 18799576]
113. Tesina P et al. Multiple cellular proteins interact with LEDGF/p75 through a conserved unstructured consensus motif. Nat Commun 6, 7968, doi:10.1038/ncomms8968 (2015). [PubMed: 26245978]
114. De Rijck J et al. The BET family of proteins targets moloney murine leukemia virus integration near transcription start sites. Cell Rep 5, 886–894, doi:10.1016/j.celrep.2013.09.040 (2013). [PubMed: 24183673]
115. Gupta SS et al. Bromo- and extraterminal domain chromatin regulators serve as cofactors for murine leukemia virus integration. J Virol 87, 12721–12736, doi:10.1128/JVI.01942-13 (2013). [PubMed: 24049186]
116. Sharma A et al. BET proteins promote efficient murine leukemia virus integration at transcription start sites. Proc Natl Acad Sci U S A 110, 12036–12041, doi:10.1073/pnas.1307157110 (2013). [PubMed: 23818621]
117. Larue RC et al. Bimodal high-affinity association of Brd4 with murine leukemia virus integrase and mononucleosomes. Nucleic Acids Res 42, 4868–4881, doi:10.1093/nar/gku135 (2014). [PubMed: 24520112]
118. LeRoy G et al. Proteogenomic characterization and mapping of nucleosomes decoded by Brd and HP1 proteins. Genome Biol 13, R68, doi:10.1186/gb-2012-13-8-r68 (2012). [PubMed: 22897906]
119. Morinière J et al. Cooperative binding of two acetylation marks on a histone tail by a single bromodomain. Nature 461, 664–668, doi:10.1038/nature08397 (2009). [PubMed: 19794495]
120. Rahman S et al. The Brd4 extraterminal domain confers transcription activation independent of pTEFb by recruiting multiple proteins, including NSD3. Mol Cell Biol 31, 2641–2652, doi:10.1128/mcb.01341-10 (2011). [PubMed: 21555454]
121. Aiyer S et al. Altering murine leukemia virus integration through disruption of the integrase and BET protein family interaction. Nucleic Acids Res 42, 5917–5928, doi:10.1093/nar/gku175 (2014). [PubMed: 24623816]
122. Crowe BL et al. Structure of the Brd4 ET domain bound to a C-terminal motif from γ-retroviral integrases reveals a conserved mechanism of interaction. Proc Natl Acad Sci U S A 113, 2086–2091, doi:10.1073/pnas.1516813113 (2016). [PubMed: 26858406]
123. Aiyer S et al. A common binding motif in the ET domain of BRD3 forms polymorphic structural interfaces with host and viral proteins. Structure, doi:10.1016/j.str.2021.01.010 (2021).
124. Maertens GN B'-protein phosphatase 2A is a functional binding partner of delta-retroviral integrase. Nucleic Acids Res 44, 364–376, doi:10.1093/nar/gkv1347 (2016). [PubMed: 26657642]

125. Winans S et al. The FACT Complex Promotes Avian Leukosis Virus DNA Integration. J Virol 91, doi:10.1128/jvi.00082-17 (2017).

126. Hertz EPT et al. A Conserved Motif Provides Binding Specificity to the PP2A-B56 Phosphatase. Mol Cell 63, 686–695, doi:10.1016/j.molcel.2016.06.024 (2016). [PubMed: 27453045]

127. Mitchell RS et al. Retroviral DNA integration: ASLV, HIV, and MLV show distinct target site preferences. PLoS Biol 2, E234, doi:10.1371/journal.pbio.0020234 (2004). [PubMed: 15314653]

128. Gillet NA et al. The host genomic environment of the provirus determines the abundance of HTLV-1-infected T-cell clones. Blood 117, 3113–3122, doi:10.1182/blood-2010-10-312926 (2011). [PubMed: 21228324]

129. Christensen DE, Ganser-Pornillos BK, Johnson JS, Pornillos O & Sundquist WI Reconstitution and visualization of HIV-1 capsid-dependent replication and integration in vitro. Science 370, doi:10.1126/science.abc8420 (2020).

130. Bukovsky A & Gottlinger H Lack of integrase can markedly affect human immunodeficiency virus type 1 particle production in the presence of an active viral protease. J Virol 70, 6820–6825, doi:10.1128/JVI.70.10.6820-6825.1996 (1996). [PubMed: 8794322]

131. Engelman A, Englund G, Orenstein JM, Martin MA & Craigie R Multiple effects of mutations in human immunodeficiency virus type 1 integrase on viral replication. J Virol 69, 2729–2736, doi:10.1128/JVI.69.5.2729-2736.1995 (1995). [PubMed: 7535863]

132. Elliott JL et al. Integrase-RNA interactions underscore the critical role of integrase in HIV-1 virion morphogenesis. Elife 9, doi:10.7554/eLife.54311 (2020).

133. Engelman A In vivo analysis of retroviral integrase structure and function. Adv Virus Res 52, 411–426, doi:10.1016/s0065-3527(08)60309-7 (1999). [PubMed: 10384245]

134. Leavitt AD, Robles G, Alesandro N & Varmus HE Human immunodeficiency virus type 1 integrase mutants retain in vitro integrase activity yet fail to integrate viral DNA efficiently during infection. J Virol 70, 721–728, doi:10.1128/jvi.70.2.721-728.1996 (1996). [PubMed: 8551608]

135. Mohammed KD, Topper MB & Muesing MA Sequential deletion of the integrase (Gag-Pol) carboxyl terminus reveals distinct phenotypic classes of defective HIV-1. J Virol 85, 4654–4666, doi:10.1128/JVI.02374-10 (2011). [PubMed: 21367893]

136. Hoyte AC et al. Resistance to pyridine-based inhibitor KF116 reveals an unexpected role of integrase in HIV-1 Gag-Pol polyprotein proteolytic processing. J Biol Chem 292, 19814–19825, doi:10.1074/jbc.M117.816645 (2017). [PubMed: 28972144]

137. Dobard CW, Briones MS & Chow SA Molecular mechanisms by which human immunodeficiency virus type 1 integrase stimulates the early steps of reverse transcription. J Virol 81, 10037–10046, doi:10.1128/JVI.00519-07 (2007). [PubMed: 17626089]

138. Tekeste SS et al. Interaction between Reverse Transcriptase and Integrase Is Required for Reverse Transcription during HIV-1 Replication. J Virol 89, 12058–12069, doi:10.1128/JVI.01471-15 (2015). [PubMed: 26401032]

139. Fontana J et al. Distribution and Redistribution of HIV-1 Nucleocapsid Protein in Immature, Mature, and Integrase-Inhibited Virions: a Role for Integrase in Maturation. J Virol 89, 9765–9780, doi:10.1128/jvi.01522-15 (2015). [PubMed: 26178982]

140. Jurado KA et al. Allosteric integrase inhibitor potency is determined through the inhibition of HIV-1 particle maturation. Proc Natl Acad Sci U S A 110, 8690–8695, doi:10.1073/pnas.1300703110 (2013). [PubMed: 23610442]

141. Kessl JJ et al. HIV-1 Integrase Binds the Viral RNA Genome and Is Essential during Virion Morphogenesis. Cell 166, 1257–1268 e1212, doi:10.1016/j.cell.2016.07.044 (2016). [PubMed: 27565348]

142. Balakrishnan M et al. Non-catalytic site HIV-1 integrase inhibitors disrupt core maturation and induce a reverse transcription block in target cells. PLoS One 8, e74163, doi:10.1371/journal.pone.0074163 (2013). [PubMed: 24040198]

143. Steinrigl A et al. Mutations in the catalytic core or the C-terminus of murine leukemia virus (MLV) integrase disrupt virion infectivity and exert diverse effects on reverse transcription. Virology 362, 50–59, doi:10.1016/j.virol.2006.11.037 (2007). [PubMed: 17258786]

144. Wlodawer A & Vondrasek J Inhibitors of HIV-1 protease: a major success of structure-assisted drug design. Annu Rev Biophys Biomol Struct 27, 249–284, doi:10.1146/annurev.biophys.27.1.249 (1998). [PubMed: 9646869]

145. Pattishall KH in The Search for Antiviral Drugs: Case Histories from Concept to Clinic (eds Adams Julian & Merluzzi Vincent J.) 23–43 (Birkhäuser Boston, 1993).

146. Murray JM et al. Antiretroviral therapy with the integrase inhibitor raltegravir alters decay kinetics of HIV, significantly reducing the second phase. AIDS 21, 2315–2321, doi:10.1097/QAD.0b013e3282f12377 (2007). [PubMed: 18090280]

147. Sato M et al. Novel HIV-1 integrase inhibitors derived from quinolone antibiotics. J Med Chem 49, 1506–1508, doi:10.1021/jm0600139 (2006). [PubMed: 16509568]

148. DeJesus E et al. Antiviral activity, pharmacokinetics, and dose response of the HIV-1 integrase inhibitor GS-9137 (JTK-303) in treatment-naive and treatment-experienced patients. J Acquir Immune Defic Syndr 43, 1–5, doi:10.1097/01.qai.0000233308.82860.2f (2006). [PubMed: 16936557]

149. Tsiang M et al. Antiviral Activity of Bictegravir (GS-9883), a Novel Potent HIV-1 Integrase Strand Transfer Inhibitor with an Improved Resistance Profile. Antimicrob Agents Chemother 60, 7086–7097, doi:10.1128/AAC.01474-16 (2016). [PubMed: 27645238]

150. Markham A Cabotegravir Plus Rilpivirine: First Approval. Drugs 80, 915–922, doi:10.1007/s40265-020-01326-8 (2020). [PubMed: 32495274]

151. Hassounah SA et al. Antiviral Activity of Bictegravir and Cabotegravir against Integrase Inhibitor-Resistant SIVmac239 and HIV-1. Antimicrob Agents Chemother 61, doi:10.1128/aac.01695-17 (2017).

152. Smith SJ, Zhao XZ, Burke TR Jr. & Hughes SH Efficacies of Cabotegravir and Bictegravir against drug-resistant HIV-1 integrase mutants. Retrovirology 15, 37, doi:10.1186/s12977-018-0420-7 (2018). [PubMed: 29769116]

153. Zhang WW et al. Accumulation of Multiple Mutations In Vivo Confers Cross-Resistance to New and Existing Integrase Inhibitors. J Infect Dis 218, 1773–1776, doi:10.1093/infdis/jiy428 (2018). [PubMed: 30010985]

154. WHO. Update of recommendations on first- and second-line antiretroviral regimens. WHO Policy Brief, 1–16 (2019). [PubMed: 30840413]

155. Hazuda DJ et al. Inhibitors of strand transfer that prevent integration and inhibit HIV-1 replication in cells. Science 287, 646–650, doi:10.1126/science.287.5453.646 (2000). [PubMed: 10649997]

156. Grobler JA et al. Diketo acid inhibitor mechanism and HIV-1 integrase: implications for metal binding in the active site of phosphotransferase enzymes. Proc Natl Acad Sci U S A 99, 6661–6666, doi:10.1073/pnas.092056199 (2002). [PubMed: 11997448]

157. Espeseth AS et al. HIV-1 integrase inhibitors that compete with the target DNA substrate define a unique strand transfer conformation for integrase. Proc Natl Acad Sci U S A 97, 11244–11249, doi:10.1073/pnas.200139397 (2000). [PubMed: 11016953]

158. Valkov E et al. Functional and structural characterization of the integrase from the prototype foamy virus. Nucleic Acids Res 37, 243–255, doi:10.1093/nar/gkn938 (2009). [PubMed: 19036793]

159. Barski MS, Minnell JJ & Maertens GN Inhibition of HTLV-1 Infection by HIV-1 First- and Second-Generation Integrase Strand Transfer Inhibitors. Front Microbiol 10, 1877, doi:10.3389/fmicb.2019.01877 (2019). [PubMed: 31474960]

160. Koh Y, Matreyek KA & Engelman A Differential sensitivities of retroviruses to integrase strand transfer inhibitors. J Virol 85, 3677–3682, doi:10.1128/jvi.0254110 (2011). [PubMed: 21270168]

161. Hare S et al. Molecular mechanisms of retroviral integrase inhibition and the evolution of viral resistance. Proc Natl Acad Sci U S A 107, 20057–20062, doi:10.1073/pnas.1010246107 (2010). [PubMed: 21030679]

162. Hare S et al. Structural and functional analyses of the second-generation integrase strand transfer inhibitor dolutegravir (S/GSK1349572). Mol Pharmacol 80, 565–572, doi:10.1124/mol.111.073189 (2011). [PubMed: 21719464]

163. Langley DR et al. The terminal (catalytic) adenosine of the HIV LTR controls the kinetics of binding and dissociation of HIV integrase strand transfer inhibitors. Biochemistry 47, 13481–13488, doi:10.1021/bi801372d (2008). [PubMed: 18991395]

164. Hightower KE et al. Dolutegravir (S/GSK1349572) exhibits significantly slower dissociation than raltegravir and elvitegravir from wild-type and integrase inhibitorresistant HIV-1 integrase-DNA complexes. Antimicrob Agents Chemother 55, 4552–4559, doi:10.1128/AAC.00157-11 (2011). [PubMed: 21807982]

165. White KL et al. Long Dissociation of Bictegravir from HIV-1 Integrase-DNA Complexes. Antimicrob Agents Chemother, doi:10.1128/AAC.02406-20 (2021).

166. Grobler JA & Hazuda DJ Resistance to HIV integrase strand transfer inhibitors: in vitro findings and clinical consequences. Curr Opin Virol 8, 98–103, doi:10.1016/j.coviro.2014.07.006 (2014). [PubMed: 25128610]

167. Santoro MM et al. Susceptibility to HIV-1 integrase strand transfer inhibitors (INSTIs) in highly treatment-experienced patients who failed an INSTI-based regimen. Int J Antimicrob Agents 56, 106027, doi:10.1016/j.ijantimicag.2020.106027 (2020). [PubMed: 32450199]

168. Ndashimye E et al. Accumulation of integrase strand transfer inhibitor resistance mutations confers high-level resistance to dolutegravir in non-B subtype HIV-1 strains from patients failing raltegravir in Uganda. J Antimicrob Chemother 75, 3525–3533, doi:10.1093/jac/dkaa355 (2020). [PubMed: 32853364]

169. Rhee SY et al. A systematic review of the genetic mechanisms of dolutegravir resistance. J Antimicrob Chemother 74, 3135–3149, doi:10.1093/jac/dkz256 (2019). [PubMed: 31280314]

170. Engone-Ondo JD et al. High rate of virological failure and HIV drug resistance in semi-rural Gabon and implications for dolutegravir-based regimen efficacy. J Antimicrob Chemother, doi:10.1093/jac/dkaa537 (2020).

171. Christ F et al. Rational design of small-molecule inhibitors of the LEDGF/p75-integrase interaction and HIV replication. Nat Chem Biol 6, 442–448, doi:10.1038/nchembio.370 (2010). [PubMed: 20473303]

172. Fader LD et al. Discovery of BI 224436, a Noncatalytic Site Integrase Inhibitor (NCINI) of HIV-1. ACS Med Chem Lett 5, 422–427, doi:10.1021/ml500002n (2014). [PubMed: 24900852]

173. Tsiang M et al. New class of HIV-1 integrase (IN) inhibitors with a dual mode of action. J Biol Chem 287, 21189–21203, doi:10.1074/jbc.M112.347534 (2012). [PubMed: 22535962]

174. Le Rouzic E et al. Dual inhibition of HIV-1 replication by integrase-LEDGF allosteric inhibitors is predominant at the post-integration stage. Retrovirology 10, 144, doi:10.1186/1742-4690-10-144 (2013). [PubMed: 24261564]

175. Gupta K et al. Allosteric inhibition of human immunodeficiency virus integrase: late block during viral replication and abnormal multimerization involving specific protein domains. J Biol Chem 289, 20477–20488, doi:10.1074/jbc.M114.551119 (2014). [PubMed: 24904063]

176. Desimmie BA et al. LEDGINs inhibit late stage HIV-1 replication by modulating integrase multimerization in the virions. Retrovirology 10, 57, doi:10.1186/1742-4690-10-57 (2013). [PubMed: 23721378]

177. Sharma A et al. A new class of multimerization selective inhibitors of HIV-1 integrase. PLoS Pathog 10, e1004171, doi:10.1371/journal.ppat.1004171 (2014). [PubMed: 24874515]

178. Deng N et al. Allosteric HIV-1 integrase inhibitors promote aberrant protein multimerization by directly mediating inter-subunit interactions: Structural and thermodynamic modeling studies. Protein Sci 25, 1911–1917, doi:10.1002/pro.2997 (2016). [PubMed: 27503276]

179. Gupta K et al. Structural Basis for Inhibitor-Induced Aggregation of HIV Integrase. PLoS Biol 14, e1002584, doi:10.1371/journal.pbio.1002584 (2016). [PubMed: 27935939]

180. Koneru PC et al. HIV-1 integrase tetramers are the antiviral target of pyridine-based allosteric integrase inhibitors. Elife 8, doi:10.7554/eLife.46344 (2019).

181. Gupta K et al. Allosteric HIV integrase inhibitors promote formation of inactive branched polymers via homomeric carboxy-terminal domain interactions. Structure, doi:10.1016/j.str.2020.12.001 (2020).

182. Amadori C et al. The HIV-1 integrase-LEDGF allosteric inhibitor MUT-A: resistance profile, impairment of virus maturation and infectivity but without influence on RNA packaging or

virus immunoreactivity. Retrovirology 14, 50, doi:10.1186/s12977-017-0373-2 (2017). [PubMed: 29121950]

183. Bruggemans A et al. GS-9822, a preclinical LEDGIN candidate, displays a block- and-lock phenotype in cell culture. Antimicrobial Agents and Chemotherapy, AAC.02328–02320, doi:10.1128/aac.02328-20 (2021).
184. Feng L et al. The Competitive Interplay between Allosteric HIV-1 Integrase Inhibitor BI/D and LEDGF/p75 during the Early Stage of HIV-1 Replication Adversely Affects Inhibitor Potency. ACS Chem Biol 11, 1313–1321, doi:10.1021/acschembio.6b00167 (2016). [PubMed: 26910179]
185. Vranckx LS et al. LEDGIN-mediated Inhibition of Integrase-LEDGF/p75 Interaction Reduces Reactivation of Residual Latent HIV. EBioMedicine 8, 248–264, doi:10.1016/j.ebiom.2016.04.039 (2016). [PubMed: 27428435]
186. Bonnard D et al. Structure-function analyses unravel distinct effects of allosteric inhibitors of HIV-1 integrase on viral maturation and integration. J Biol Chem 293, 6172–6186, doi:10.1074/jbc.M117.816793 (2018). [PubMed: 29507092]
187. Peese KM et al. 5,6,7,8-Tetrahydro-1,6-naphthyridine Derivatives as Potent HIV-1-Integrase-Allosteric-Site Inhibitors. J Med Chem 62, 1348–1361, doi:10.1021/acs.jmedchem.8b01473 (2019). [PubMed: 30609350]
188. Fenwick C et al. Preclinical profile of BI 224436, a novel HIV-1 non-catalytic-site integrase inhibitor. Antimicrob Agents Chemother 58, 3233–3244, doi:10.1128/aac.02719-13 (2014). [PubMed: 24663024]
189. Shkriabai N et al. A critical role of the C-terminal segment for allosteric inhibitor-induced aberrant multimerization of HIV-1 integrase. J Biol Chem 289, 26430–26440, doi:10.1074/jbc.M114.589572 (2014). [PubMed: 25118283]
190. Sutton G et al. Assembly intermediates of orthoreovirus captured in the cell. Nat Commun 11, 4445, doi:10.1038/s41467-020-18243-9 (2020). [PubMed: 32895380]
191. Hoffman DP et al. Correlative three-dimensional super-resolution and block-face electron microscopy of whole vitreously frozen cells. Science 367, 1–12, doi:10.1126/science.aaz5357 (2020).

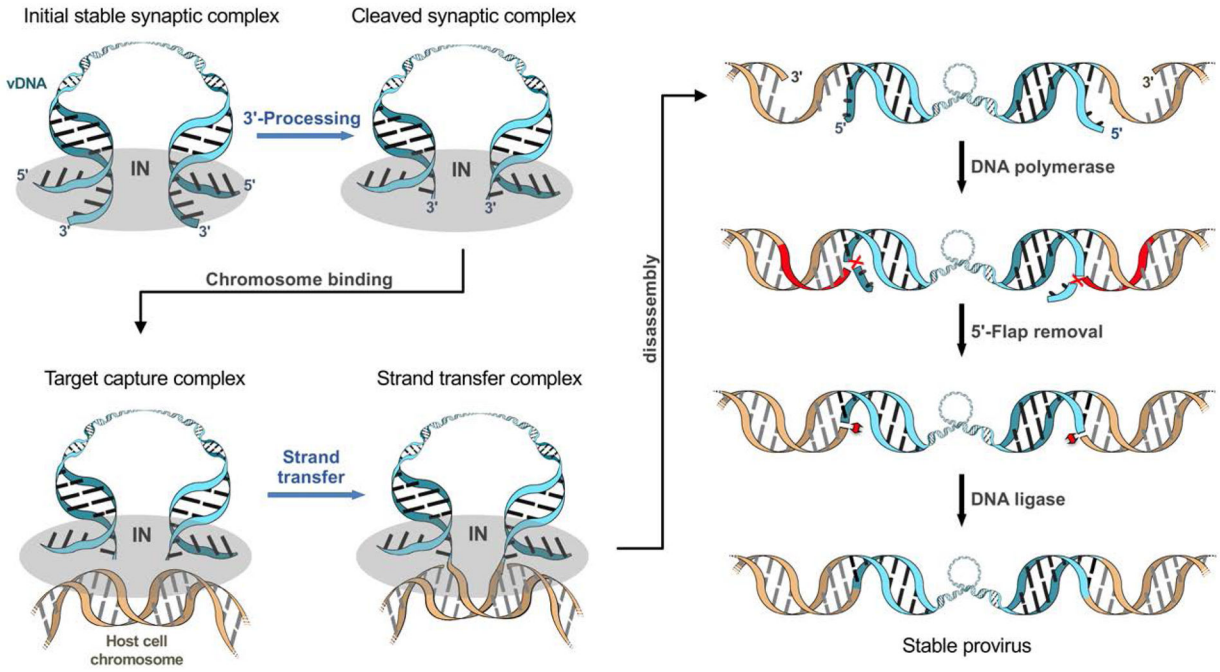

Figure 1. Schematic of the retroviral integration process.

The intasome undergoes two enzymatic steps (blue arrows) *en route* from the initial stable synaptic complex to the post-catalytic strand transfer complex, wherein 3' vDNA ends are joined to chromosomal DNA. Following intasome disassembly, ligation of the 5' vDNA ends to the chromosome requires at least three additional enzymatic functions (thick black arrows) provided by host cell factors.

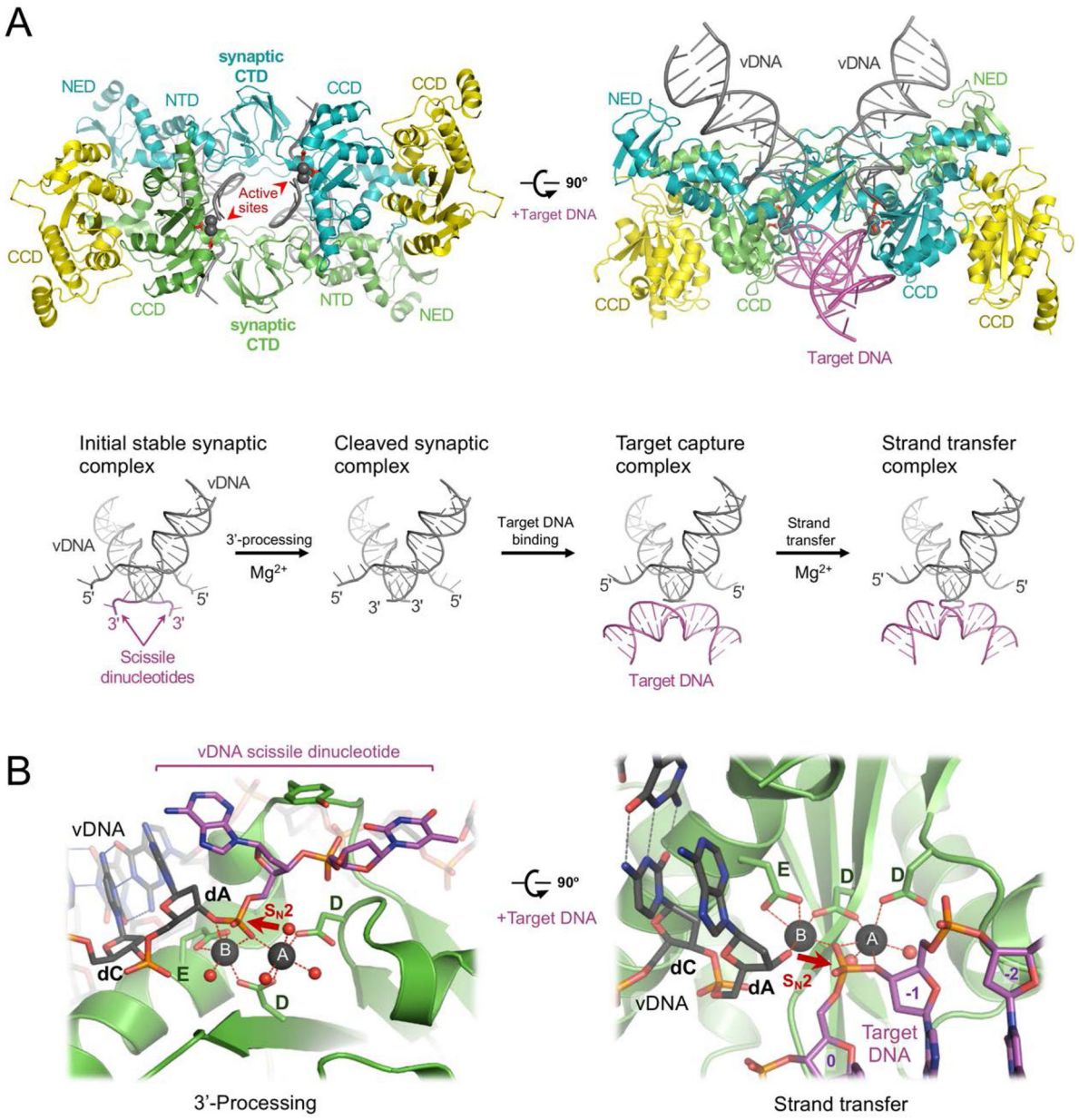

Figure 2. PFV intasome structures and IN active site mechanics.

(A) Crystal structures of the PFV cleaved synaptic complex (left) and the target capture complex (right)[28,161]. IN and DNA chains are shown as cartoons and color-coded: target DNA and the scissile vDNA dinucleotides (below) are magenta, the rest of the vDNA is dark grey; IN chains providing the active sites are green and cyan, and the outer IN chains are yellow. Grey spheres are catalytic metal ions. Active sites carboxylates comprising the IN D,D-35-E motif are shown as red sticks. Locations of the individual IN domains (NTD, CCD, and CTD) and the intasome active sites (red arrowheads) are indicated. Bottom panel shows target DNA and/or vDNA within four successive functional states of the intasome as it transitions from the initial stable synaptic complex (left) to the post-catalytic strand transfer complex (right). (B) IN active sites primed for 3'-processing (left) and strand

transfer (right). DNA chains and IN active site residues (indicated as D, D, and E) are shown as sticks. DNA chains are coloured dark grey, except for target DNA and scissile vDNA dinucleotide (magenta). Selected water molecules are shown as small red spheres. Direction of the S_N2 nucleophilic substitution at each step is indicated with red arrows. Grey and red dashes are hydrogen and metal coordination bonds, respectively.

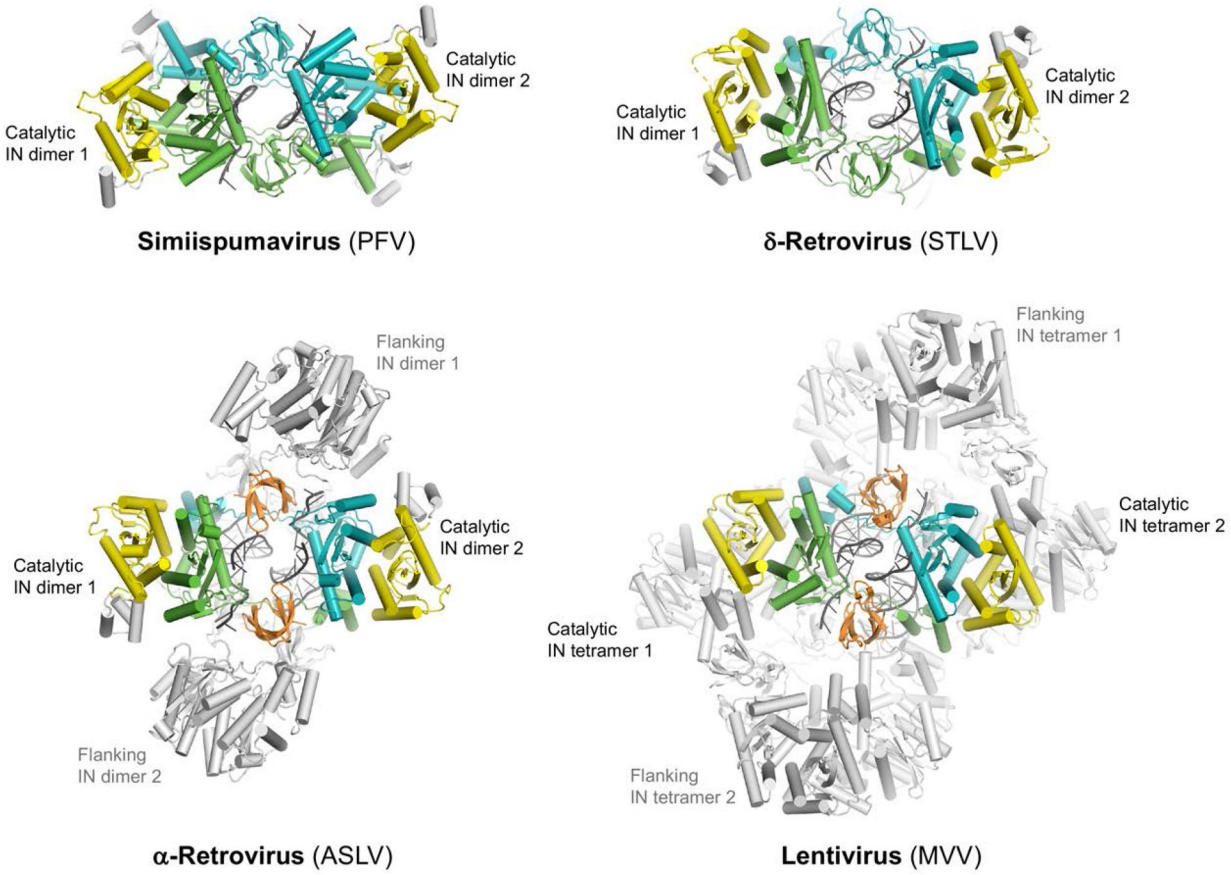

Figure 3. Diversity of retroviral intasome architectures.

Examples of intasome structures from simiispumavirus (PFV, containing a tetramer of IN), δ-retrovirus (STLV, a tetramer), α-retrovirus (ASLV, an octamer), and lentivirus (MVV, a hexadecamer) are shown as cartoons with cylinders representing α-helices[25,30,32,38]. In each case, IN domains contributing to the CIC are shown in colour with the remainder of the structure in grey. Synaptic CTDs in ASLV and MVV intasomes are coloured orange to indicate their origins from flanking IN protomers.

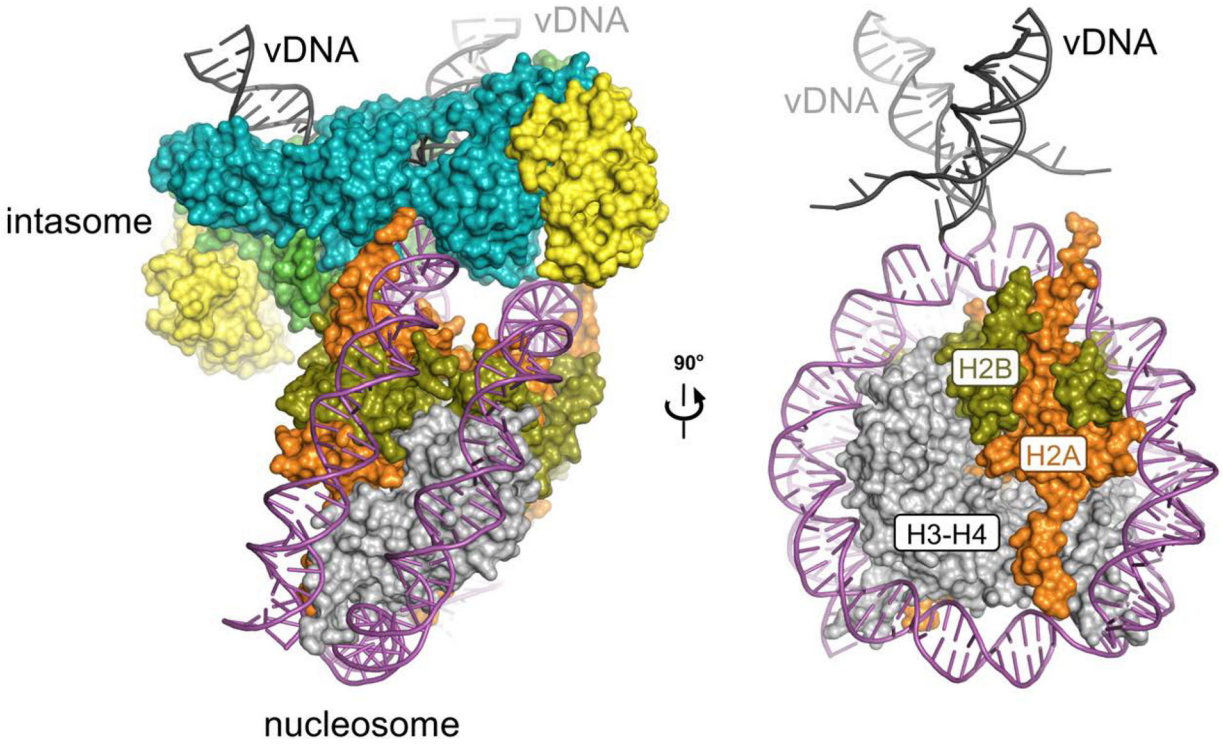

Figure 4. Integration into nucleosomal DNA.

Cryo-EM structure of the PFV intasome in a strand transfer complex with a mononucleosome[65]. The structure is shown in two orthogonal orientations. DNA (vDNA and nucleosomal DNA) and protein (IN and histones) are depicted as cartoons or in spacefill, respectively (H2A, orange; H2B, olive; H3 and H4 grey). For clarity, IN chains are hidden in the right panel. Note that the nucleosomal DNA is lifted from the surface of the histone octamer at the site of integration[64].

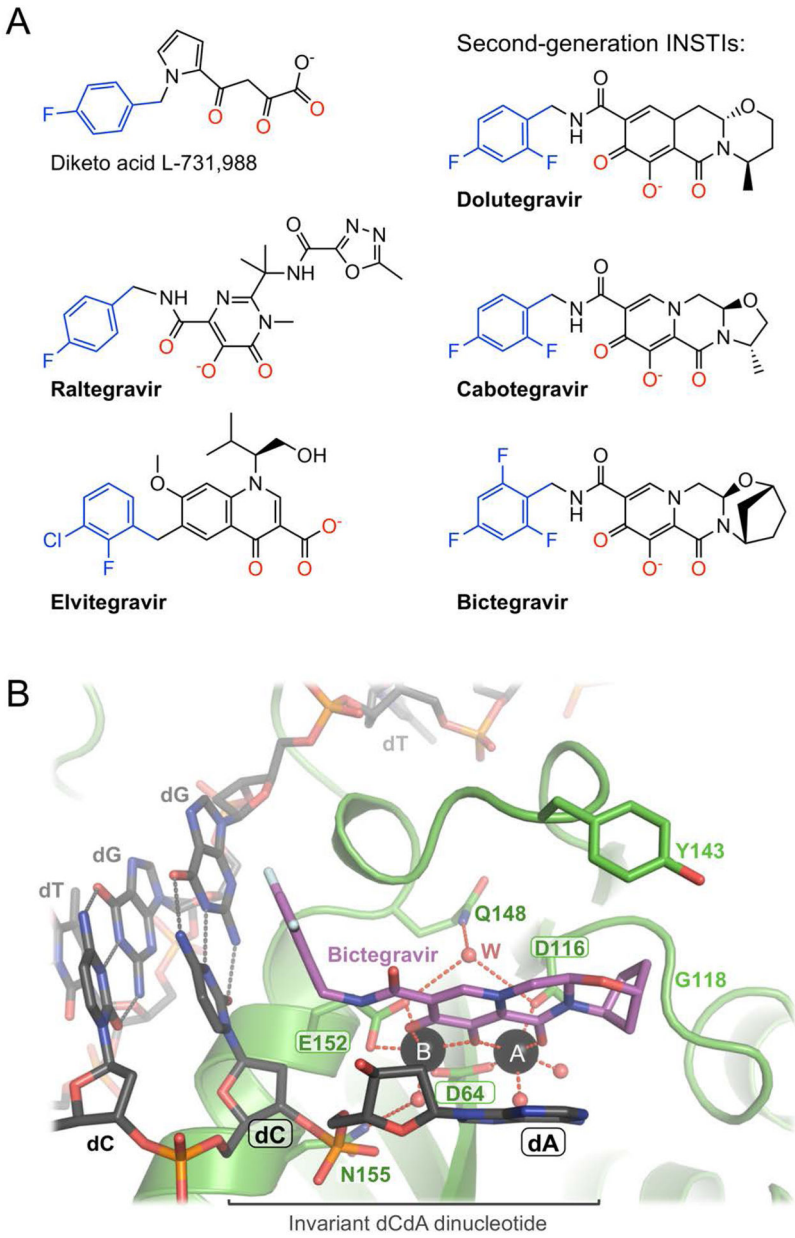

Figure 5. IN strand transfer inhibitors.

(A) Chemical structures of a lead diketo-acid compound (L-731,988) and of the five INSTIs approved for clinical use. Co-planar oxygen atoms involved in metal chelation are highlighted in red, and halo-benzyl groups in blue. (B) Bictegravir bound to the active site of the red-capped mangabey SIV intasome visualised by cryo-EM[35]. The drug molecule is shown as sticks with carbon atoms in magenta. Note that Asn155 and Gln148 participate in the secondary coordination spheres of the metal ions. In particular, Q148H leads to displacement of a key water molecule (W) bonded to carboxylates of Asp116 and Glu152 (REF.[35]).

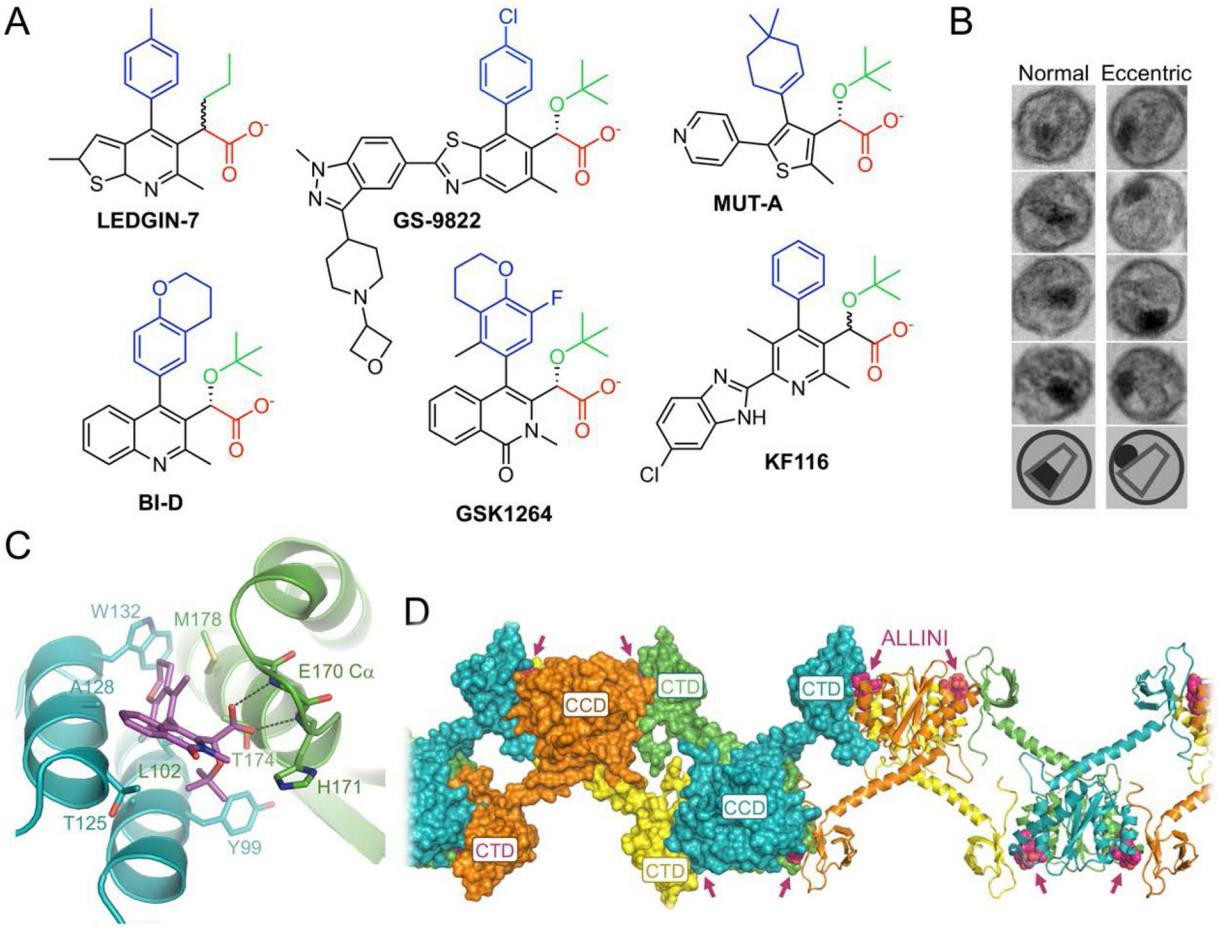

Figure 6. Allosteric HIV-1 IN inhibitors.

(A) Examples of small molecules of this class. Functional groups comprising the shared ALLINI warhead structure are shown in colour: red, a carboxyl group; green, an aliphatic side chain; and blue, a bulky hydrophobic group. (B) Thin section electron microscopy of stained HIV-1 virions produced under normal conditions (left) or in the presence of an ALLINI (right)[139]. Idealised schematics of the respective morphologies are shown on the bottom. (C) Co-crystal structure of HIV-1 IN CCD dimer with GSK1264 (REF.[179]). IN chains are depicted as cartoons, with the amino acid residues lining the ALLINI binding pocket as sticks. Carbon atoms of the inhibitor are shown in magenta. (D) A chain of IN dimers observed in the co-crystal structure of full-length HIV-1 IN with GSK1264 (REF.[179]). Positions of IN CCDs and CTDs are indicated; NTDs were not resolved in this structure. The ALLINI molecule (positions indicated with arrowheads), found at each of the CTD-CCD interfaces, is shown in spacefill, with carbon atoms in magenta.