



# Structural Implications of Genotypic Variations in HIV-1 Integrase From Diverse Subtypes

Leonard Rogers<sup>1</sup>, Adetayo E. Obasa<sup>2,3</sup>, Graeme B. Jacobs<sup>2</sup>, Stefan G. Sarafianos<sup>4</sup>, Anders Sönnnerborg<sup>3,5</sup>, Ujjwal Neogi<sup>3\*</sup> and Kamalendra Singh<sup>1,3,6</sup>

<sup>1</sup> Christopher S. Bond Life Sciences Center, University of Missouri, Columbia, MO, United States, <sup>2</sup> Division of Medical Virology, Department of Pathology, Faculty of Medicine and Health Sciences, Stellenbosch University, Stellenbosch, South Africa, <sup>3</sup> Division of Clinical Microbiology, Department of Laboratory Medicine, Karolinska Institute, Stockholm, Sweden, <sup>4</sup> Laboratory of Biochemical Pharmacology, Department of Pediatrics, Emory University School of Medicine, Atlanta, GA, United States, <sup>5</sup> Division of Infectious Diseases, Department of Medicine Huddinge, Karolinska Institute, Stockholm, Sweden, <sup>6</sup> Department of Molecular Microbiology and Immunology, University of Missouri, Columbia, MO, United States

## OPEN ACCESS

### Edited by:

Akio Adachi,  
Kansai Medical University, Japan

### Reviewed by:

Takao Masuda,  
Tokyo Medical and Dental University,  
Japan

Tino Sanchez,  
Loma Linda University, United States  
Nanjie Deng,  
Pace University, United States

### \*Correspondence:

Ujjwal Neogi  
ujjwal.neogi@ki.se  
Kamalendra Singh  
singhka@missouri.edu

### Specialty section:

This article was submitted to  
Virology,  
a section of the journal  
Frontiers in Microbiology

**Received:** 28 May 2018

**Accepted:** 13 July 2018

**Published:** 02 August 2018

### Citation:

Rogers L, Obasa AE, Jacobs GB, Sarafianos SG, Sönnnerborg A, Neogi U and Singh K (2018) Structural Implications of Genotypic Variations in HIV-1 Integrase From Diverse Subtypes. *Front. Microbiol.* 9:1754. doi: 10.3389/fmicb.2018.01754

Human immunodeficiency virus type 1 (HIV-1) integrase (IN) integrates viral DNA into the host genome using its 3'-end processing and strand-transfer activities. Due to the importance of HIV-1 IN, it is targeted by the newest class of approved drugs known as integrase strand transfer inhibitors (INSTIs). INSTIs are efficient in maintaining low viral load; however, as with other approved antivirals, resistance mutations emerge in patients receiving INSTI-containing therapy. As INSTIs are becoming increasingly accessible worldwide, it is important to understand the mechanism(s) of INSTI susceptibility. There is strong evidence suggesting differences in the patterns and mechanisms of drug resistance between HIV-1 subtype B, which dominates in United States, Western Europe and Australia, and non-B infections that are most prevalent in countries of Africa and Asia. IN polymorphisms and other genetic differences among diverse subtypes are likely responsible for these different patterns, but lack of a full-length high-resolution structure of HIV-1 IN has been a roadblock in understanding the molecular mechanisms of INSTI resistance and the impact of polymorphisms on therapy outcome. A recently reported full-length medium-resolution cryoEM structure of HIV-1 IN provides insights into understanding the mechanism of integrase function and the impact of genetic variation on the effectiveness of INSTIs. Here we use molecular modeling to explore the structural impact of IN polymorphisms on the IN reaction mechanism and INSTI susceptibility.

**Keywords:** HIV-1 integrase, strand transfer inhibitor, polymorphism, drug-resistance, HIV-1 subtypes

## INTRODUCTION

Combination antiretroviral therapy (cART) targets several steps of viral replication. In many cases, cART can suppress viral load below the detection level and make HIV infection a chronic yet manageable disease with a near-normal life expectancy (Antiretroviral Therapy Cohort Collaboration, 2017). However, use of cART has been constantly challenged by the emergence of both acquired and transmitted drug resistance mutations (DRMs). In addition, anti-HIV drugs have associated toxicity and bioavailability issues, although to different extents. These challenges

have spurred the development of new antiretrovirals that have a high genetic barrier to resistance and low toxicity and that are effective against resistant viruses. Integrase strand transfer inhibitors (INSTIs) are the newest class of approved anti-HIV drugs. As the name implies, INSTIs inhibit HIV-1 integrase (IN), which is one of the three enzymes encoded by the *pol* gene.

Retrovirus INs have two major catalytic activities: a 3'-end processing (3'EP) that excises a dinucleotide at the 3'-end, and a strand transfer (ST) activity that integrates HIV DNA into the host chromosome. Both activities are conducted by the same active site (Craigie, 2001, 2012; Chiu and Davies, 2004). In addition, two more IN associated activities have been reported. These are (i) specific endonucleolytic cleavage at the terminal sequences of each LTR, and (ii) disintegration, which can be considered as the reverse of the ST reaction (Thierry et al., 2016).

HIV-1 IN, a 32 kDa protein, functions as a tetramer and/or higher-order oligomer (Craigie, 2012; Craigie and Bushman, 2012). HIV-1 IN has three distinct domains: the N-terminal domain (NTD) (residues 1–46), the catalytic core domain (CCD) (residues 56–186), and the C-terminal domain (CTD) (195–288). The active site (D64, D116, and E152) resides in the CCD. The NTD contains a conserved Zn<sup>2+</sup>-binding motif (His-His-Cys-Cys). The CTD is basic in nature and adopts an SH3-fold. The CTD has been implicated in DNA binding and oligomerization of HIV-1 IN (Kessl et al., 2009). The high resolution structure of full-length HIV-1 IN is not known (Craigie, 2012), although structures of IN domains (NTD/CCD and CCD/CTD) have been determined (Cai et al., 1997; Chen et al., 2000; Wang et al., 2001).

Currently used first-line cART includes an INSTI in the backbone of two nucleoside reverse transcriptase inhibitors (NRTIs). At present, four INSTIs have been approved. These are raltegravir (RAL), elvitegravir (EVG), dolutegravir (DTG), and bictegravir (BIC). Of these, RAL and EVG are referred to as first-generation, whereas DTG and BIC are second-generation INSTIs (Anstett et al., 2017). Another INSTI, cabotegravir (CAB), is in the late phases of clinical trials, and it has shown great potential, especially as a long-acting (LA) antiviral (Whitfield et al., 2016). Resistance pathways to first-generation INSTIs and DTG are well established. However, the same for CAB and BIC are not well understood. Molecular details of resistance to INSTIs are also not well known, mostly due to the lack of a high-resolution full-length structure of HIV-1 IN.

Recently, great strides have been made in the structural biology of INs. CryoEM structures of the HIV-1 intasome were reported in early 2017 (Passos et al., 2017). These structures, together with a 3.8 Å resolution X-ray crystal structure of the Rous sarcoma virus (RSV) intasome (Yin et al., 2016), the cryoEM structures of the prototype foamy virus (PFV) intasome (Ballandras-Colas et al., 2017), and that of the mouse mammary tumour virus (MMTV) intasome (Ballandras-Colas et al., 2016), provide significant insights into the assembly of retroviral IN nucleoprotein complexes at different stages of the ST reaction (Engelman and Cherepanov, 2017). These structures reveal many common features of the IN mechanism of action across retroviruses (Engelman and Cherepanov, 2017). In addition, the crystal structures of PFV INs in complex with DNA and DNA/INSTI have been solved (Maertens et al., 2010;

Hare et al., 2010a, 2012). The crystal structures of PFV IN in complex with DNA and INSTIs have been serving as the model systems for understanding INSTI resistance mechanisms.

Limited solubility of HIV-1 IN in buffers containing isotonic salt concentrations has hindered structure determination of full-length HIV-1 IN. To overcome this limitation, Li et al. (2014) generated a chimera containing full-length HIV-1 IN fused with *Sulfolobus solfataricus* protein Sso7d that greatly enhanced the solubility and activity of the HIV-1 IN. This construct facilitated the solution of a cryoEM structure of the HIV-1 IN strand transfer complex (STC) intasome (Passos et al., 2017). Despite some limitations (discussed later), this structure provided the first glimpse of nucleoprotein organization that could be used to deduce the effect of a polymorphism (PM, defined as the change in genomic sequence variation that is common in more than 1% of the sequences) in the IN of different HIV-1 subtypes at the atomic level. Here we present the identification of IN PMs from diverse HIV-1 subtypes and assess the impact of these PMs in the structure-function mapping of HIV-1 IN.

## MATERIALS AND METHODS

### Identification of PMs From Diverse HIV-1 Subtypes

HIV-1 integrase sequences ( $n = 8114$ ) of viruses isolated from individual patients were downloaded from the HIV-1 Stanford Database (Rhee et al., 2003). After excluding the low quality and shorter sequences, we included HIV-1 A1/A2 ( $n = 483$ ), HIV-1B ( $n = 4379$ ), HIV-1C ( $n = 1155$ ), CRF01\_AE ( $n = 1581$ ), and CRF02\_AG ( $n = 522$ ) sequences. These five subtypes and Circulating Recombinant Forms represent >90% of global infections (Hemelaar, 2012). Multiple sequence alignment was performed using ClustalX (Larkin et al., 2007) against the HIV-1 HXB2 sequence (the reference sequence). Variant calling of each residue was performed using an in-house R script (Krzywinski et al., 2009). We also used the IN sequences ( $n = 91$ ) from the Tygerberg Virology (TV) cohort (Jacobs et al., 2009). This cohort contains treatment-naïve patients from a variety of ethnic groups and sexual orientations. The patient samples of the TV cohort were collected between 2000 and 2001, before the initiation of South Africa's national HIV treatment program. Recently, we amplified and sequenced IN genes from the TV cohort ( $n = 91$ ) for identification of INSTI DRMs (Brado et al., 2018). Subtype specific consensus sequences were generated using the Consensus Maker<sup>1</sup> tool. Naturally occurring polymorphisms (PMs) were defined as any mutations that were present in >50% of sequences.

### Structures of IN Intasomes From Different Subtypes

Homology-derived molecular models of HIV-1 IN tetramers from different subtypes were generated using the cryoEM structure of the HIV-1B IN intasome (PDB file 5U1C) (Passos

<sup>1</sup><https://www.hiv.lanl.gov/content/sequence/CONSENSUS/consensus.html>

et al., 2017) as a template, using Prime (version 4.2) through Maestro (Schrodinger, New York, NY, United States) as previously described (Hagglblom et al., 2016). The homology models were minimized using Prime and MacroModel to eliminate steric overlaps and to optimize sidechain conformations, respectively (Singh et al., 2012). The cryoEM intasome structure (PDB file 5U1C) is missing residues A205 to N222 (Passos et al., 2017). In addition, residues 1–55 (comprising the entire NTD and the linker region between the NTD and CCD), residues 135–150, and residues 186–195 could not be resolved in the outer IN molecules (definition of outer and inner INs presented below). These loops were constructed with Prime's loop modeling utility (Schrödinger Suite). All modeled structures were submitted to the Structure Analysis and Verification Server (SAVES)<sup>2</sup> as well as the Protein Structure Preparation tool of SYBYL-X (version 2.1). No bad contacts were noted in the final models of INs. The backbone torsion angles were checked by Ramachandran plot for allowed conformations of  $\phi$  and  $\psi$  angles.

## RESULTS AND DISCUSSION

### IN PMs

Among all non-B subtype sequences, 17 naturally occurring IN PM positions with 18 PMs were observed (**Figure 1A**). Distribution of these PMs in different domain contexts is depicted in **Figure 1B**. Fifteen of these PMs were successfully mapped structurally. Three of these (K14R, D25E, and V31I) belong to the NTD, whereas M50I belongs to the loop region connecting the NTD and CTD. Eight PMs (I72V, L74M/I, F100Y, L101I, T124A, K136Q, and D167E) are part of the CCD, and the remaining six (V201I, T218I, L234I, A265V, R269K, and S283G) belong to the CTD. The other two were in regions for which structural data are unavailable. IN PM T218I lies within the missing region (residues 205–222). While we have modeled this region as a loop, an ambiguity remains in the secondary structure assignment, as this region in the crystal structure of CCD + CTD assumes a helical conformation. A polymorphism at position 283 could not be structurally mapped, since the structure of the C-terminal end of IN (amino acids 270–288) is unavailable.

### Structure of HIV-1 Intasome

The tetrameric cryoEM structure represents a post-catalysis synaptic STC of the HIV-1 intasome formed by four IN subunits arranged in two-fold symmetry (Passos et al., 2017). A pair of IN dimers encapsulates the viral/host DNA chimera, in which the two inner molecules directly interact with DNA, while the outer molecules have protein–protein interactions with the inner molecules (**Figure 2**). This arrangement of IN oligomers bound to the viral/host DNA chimera has been referred to as the intasome core structure (Engelman and Cherepanov, 2017). The structure of the NTD in the outer IN molecules is not available in PDB coordinate file 5U1C (Passos et al., 2017). The structure of residues 205–222 is missing in all IN structures, whereas the

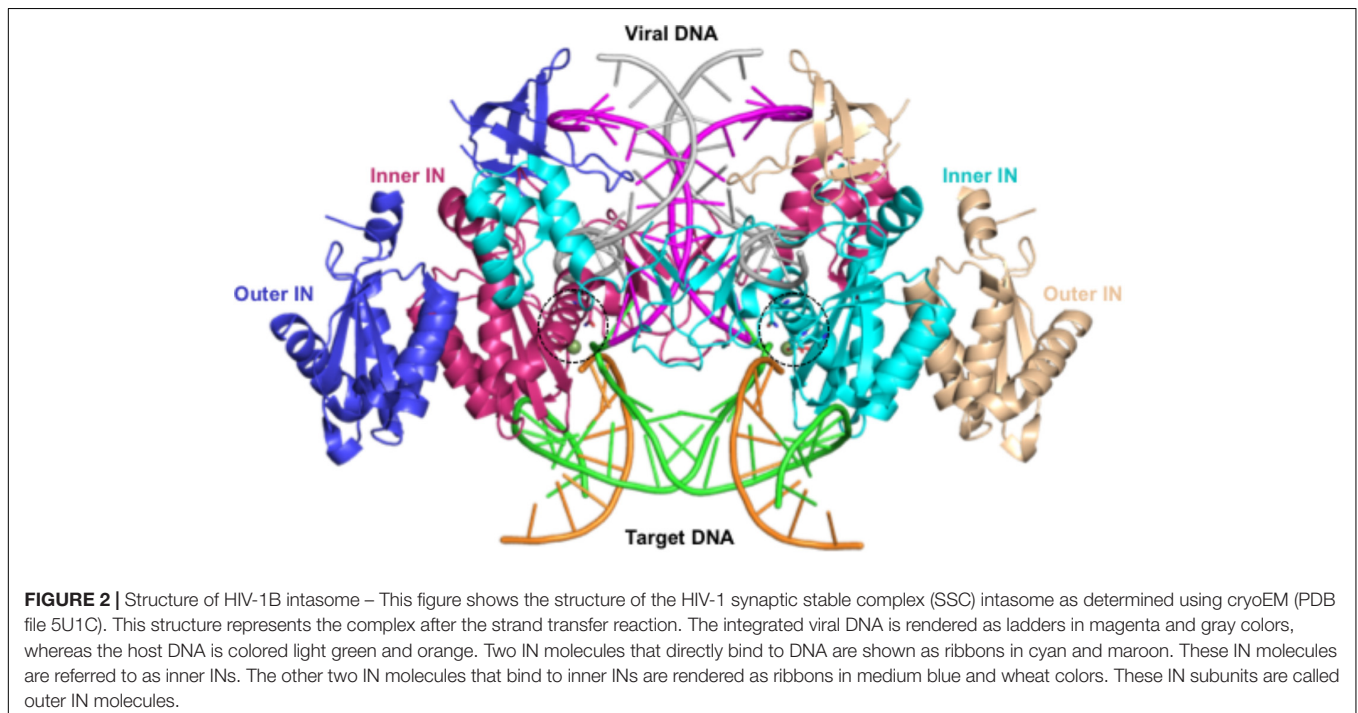
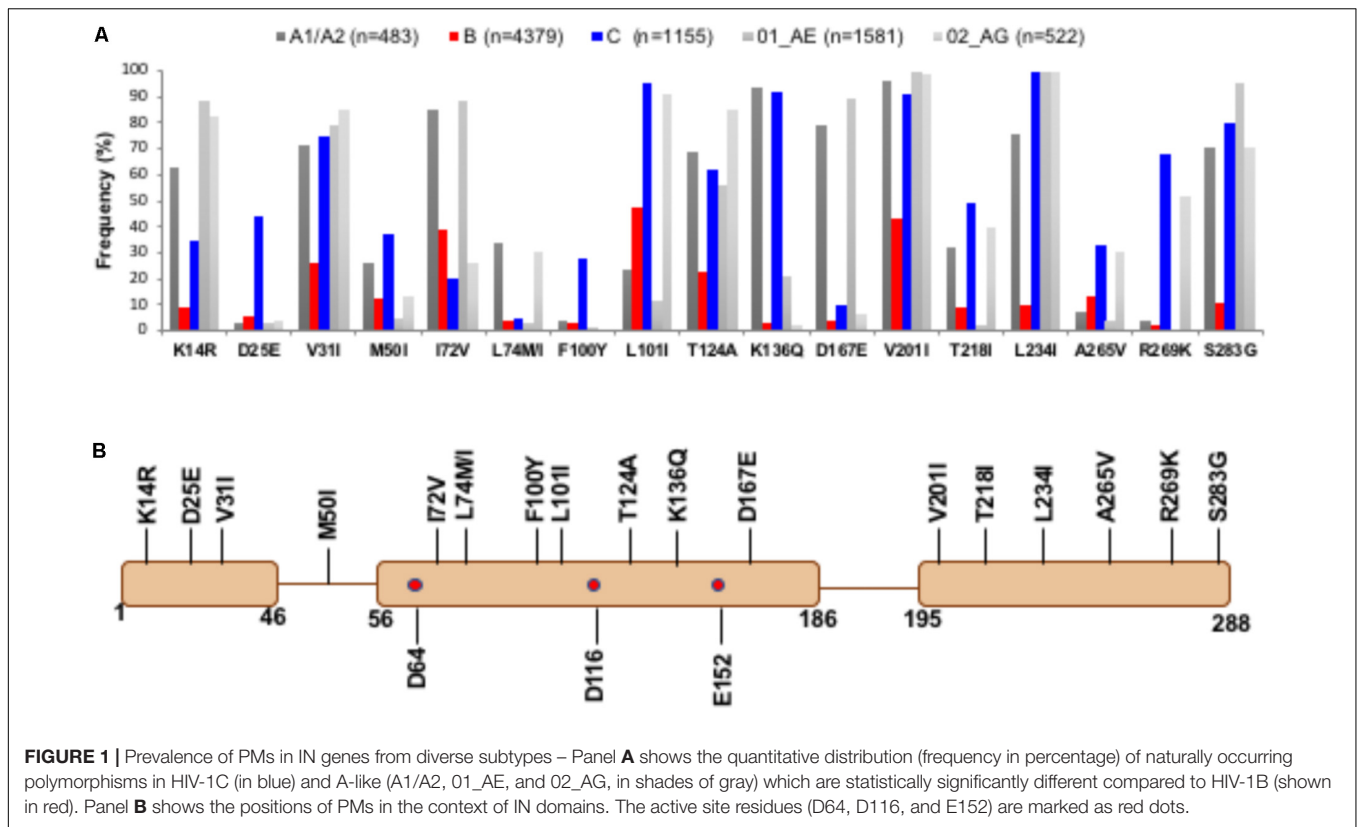
structures corresponding to residues 135–150 and 186–195 of the outer IN molecules are also unresolved. Residues 135–139 assume a  $\beta$ -strand conformation in the structure of NTD+CCD [PDB file 1K6Y, (Wang et al., 2001)], whereas residues 186–195 exist as a loop structure in the crystal structure of CCD+CTD [PDB file 1EX4, (Chen et al., 2000)]. In an earlier structure of IN CCD and CTD domains, residues 205–222 are part of the helix (PDB file 1EX4). However, due to the topological position of the CCD and CTD in the STC, a loop conformation of this region appears more feasible, which is the conformation in our modeled structures. We set the conformation of residues 135–139 as a  $\beta$ -strand, whereas residues 186–195 were kept in loop conformation in all IN molecules.

### PMs in the NTD

Our analyses showed a total of three NTD PM positions in different subtypes. These are K14R, D25E, and V31I. PM M50I is located in the linker region connecting the NTD and CCD. Of these, D25E, V31I, and M50I were also noted in the cohort of South African HIV-1C patients (Brado et al., 2018). PM K14R was most prominent in CRFs 01\_AE and 02\_AG, followed by subtypes A1/A2. Approximately 33% of HIV-1C sequences had 14R. Amino acid K14 is located on  $\alpha$ A-helix. The sidechain of K14 in HIV-1B forms a salt-bridge with the backbone C = O of W131 of the outer IN subunits, suggesting that K14 is involved in tetramer formation (**Figure 3A**). Substitution of K to R results in an additional salt-bridge formation (with C = O of W132). Typical salt-bridge energy ranges between 1 and 4 kcal/mol (Honig and Hubbell, 1984; Wimley et al., 1996), suggesting that R14 may induce a change in interaction energy due to a pair of salt-bridges and reduced compared to one in case of K14. In addition, both K14 and R14 form H-bonds with Y15. Previously, Y15 has been shown to be crucial for the assembly of IN and HIV-1 RT on viral RNA through the RT-IN precursor form (Takahata et al., 2017). In addition to the above mentioned interactions of K/R14, the K14 side chain forms a cation- $\pi$  interaction with the indole ring of the W131 side chain, which may further stabilize the tetramer. Residue K14R is between zinc-binding residues H12 and H16. Due to the location of these residues on the  $\alpha$ -helix, the sidechain of K14/R14 is extended away from the zinc-binding residues, and does not appear to affect the geometry of the zinc-binding motif.

PM D25E was noted exclusively in HIV-1C in the sequences from the HIV database as well as in our TV cohort (Brado et al., 2018). D25 is located at the beginning of  $\alpha$ B-helix. D25 forms symmetrical interactions between the inner INs. D25 from one inner IN forms a salt-bridge with K188 of the other inner IN molecule (**Figure 3B**). K188 is one of the three (K186, R187, and K188) positively charged residues in the vicinity of D25. Of these, R187 directly interacts with the phosphate backbone of viral DNA (**Figure 3B**). In the modeled structure of HIV-1C IN, E25 interacts with both K186 and K188 (**Figure 3B**). Additional interactions of E25 may influence DNA binding by IN, which in turn can affect INSTI binding, since the INSTI binding pocket is formed by both DNA and protein molecules. The third PM of the NTD, V31I, exists in ~75–85% of sequences among non-B subtypes (**Figure 1**). In the STC, V31 is close to viral DNA

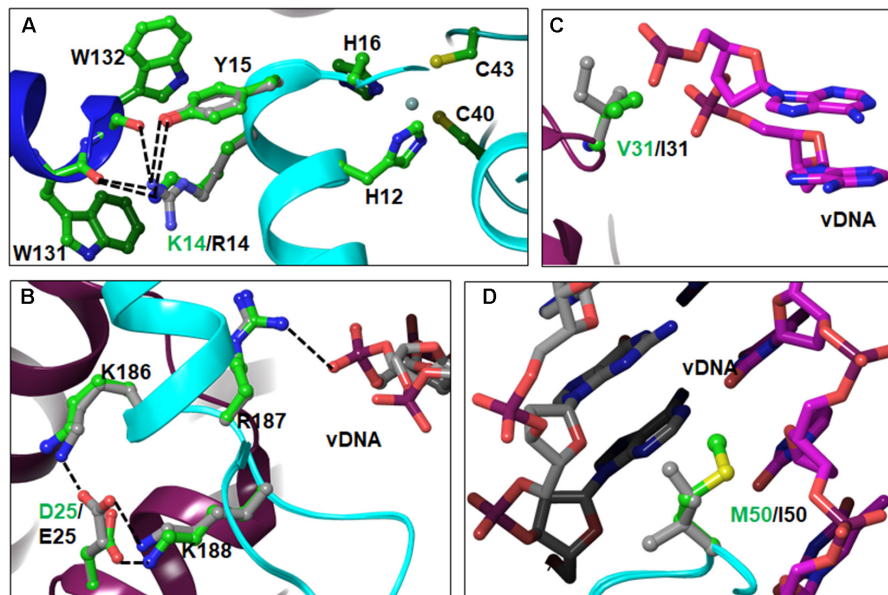
<sup>2</sup><https://services.mbi.ucla.edu/SAVES/>



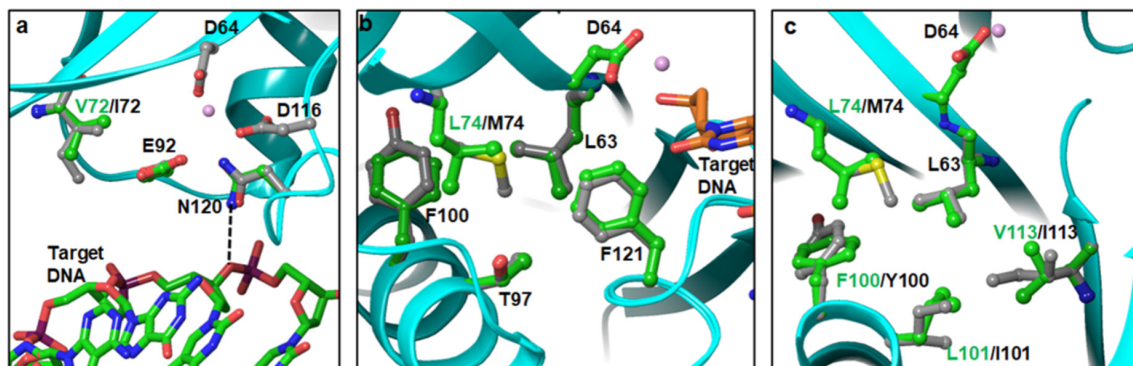
(**Figure 3C**), although not within interacting distance. The longer sidechain of I31 brings DNA and protein within interacting distance. Hence, the V31I PM can influence binding of DNA with IN in non-B subtypes.

Amino acid M50 is located on the linker region between the NTD and CCD. In the STC, M50 from both inner IN molecules interact with viral DNA through hydrophobic interactions with base moieties (**Figure 3D**). Substitution I50 will enhance the





**FIGURE 3** | Locations and interactions of NTD PMs – Panels **A–C** show PMs and their interactions in the NTD. The M50I PM belongs to the linker region and is shown in panel **D**. The interactions (hydrogen bonding and/or ion-pairs) are shown as dotted lines. In this and following figures, the amino acid residues are shown in ball-and-stick representation and colored as green carbons for HIV-1B and gray carbons in non-B HIV-1. The other atoms are colored as red (oxygen), blue (nitrogen) and yellow (sulfur). The DNA is shown as sticks. The viral DNA is colored with magenta and gray carbons, and target DNA is colored with green and orange carbons. For reference purposes, we have included the zinc-binding motif in panel **A**.



**FIGURE 4** | Topological positions of IN PMs near the active site of CCD – Panels **a**, **b**, and **c** show the positions of V72I, L74M and F100Y/L101I, respectively. As a reference point, D64 and metal ion at the active site are also shown.

hydrophobic interaction due to greater hydrophobicity of I compared to M. In breakthrough selections, M50I mutations have been seen to emerge after R263K (Tsiang et al., 2016) and to provide a replication advantage to R263K-containing viruses. While M50I alone did not change the EC<sub>50</sub> in seven of 24 recombinant viruses in our previous report (Neogi et al., 2018), it is possible that the M50I PM in non-B subtypes may be advantageous for R263K-containing viruses.

### PMs in the CCD

A total of eight PMs (I72V, L74M/I, F100Y, L101I, T124A, K136Q, and D167E) in the CCD were identified in our sequence analyses. In the cryoEM intasome structure, position 72 is occupied by

valine. It is not surprising, since I72 is highly polymorphic in HIV-1B (Figure 1). To evaluate the contribution of isoleucine, we modeled I72 in place of V72 and conducted energy minimization of the structure. Our results showed some rearrangement of neighboring sidechains including E92. Mutations E92Q/A are known DRMs against first-generation INSTIs RAL and EVG (Wensing et al., 2017). E92 is positioned between I/V72 and N120, and does not interact directly with DNA (Figure 4a). However, N120 has a direct interaction with the bridging phosphate oxygen of target DNA (shown as dotted line in Figure 4a). In the model of I72 containing IN, we noticed that the positions of the sidechain carbonyl and NH<sub>2</sub> groups of N120 are flipped relative to the conformation of N120 in the V72

structure. This flipped position of the C = O requires a water-mediated contact, indicating that the I72 PM may influence emergence of the DRM at E92. To assess the effect of the E92Q/A mutation, we superposed the crystal structure of PFV IN bound to RAL and DNA (PDB file 3OYA) (Hare et al., 2010b) onto the structure of the HIV-1 intasome (PDB file 5U1C) (Passos et al., 2017). This superposition showed that at least two nucleotides at the 3' end of the 8-nt long target DNA must be displaced to accommodate the oxadiazole moiety of RAL. A second observation that we made was that there are several water molecules in the vicinity of RAL, and three of these water molecules appear to complete the coordination geometry of Mg<sup>2+</sup> ions. In the superposed structure, these water molecules are within interacting distance of E92. The mutation E92Q/A is expected to disrupt the spatial arrangement of these waters which, in turn, may affect the coordination of Mg<sup>2+</sup> ions and thereby result in reduced binding of RAL. L74 is part of a hydrophobic cluster including L63, T97, F100/Y100, L101, L113/I113, and F121 near the active site of IN (Figures 4b,c). F100Y and L101I are also highly polymorphic in our sequence and in the HIV-1 sequence database. Both T97 and F121 are known INSTI DRM positions (Wensing et al., 2017). In addition, a recent report showed that the L74F mutation increased resistance to second-generation INSTIs (Hachiya et al., 2017). Hence, the impact of PMs in this hydrophobic cluster appears rather complex. In our modeled HIV-1C IN structure, M74 is closer to T97 and F121 than L74 (Figure 4b). Previous *in vitro* selection studies have shown that Q148H/R and G140S in combination with mutations L74I/M, E92Q, T97A, E138A/K, G140A, or N155H are associated with 5- to 20-fold reduced DTG susceptibility (Kobayashi et al., 2011). It is possible that polymorphism L74M is related to T97A mutation evolutionarily. F100Y and L101I PMs can impact the core structure and thereby affect the local geometry of the active site. The sequence containing I101 also had I113 in our model of HIV-1C IN. It is possible that two mutations arise simultaneously for preferred replication of the virus.

T124 interacts with target DNA (Figure 5a). This is a highly polymorphic position, noted here as well as reported previously (Kobayashi et al., 2011). In a previous report, the T124A mutation alone or in combination with L101I (T124A/L101I) was identified more frequently in RAL failing patients than in INSTI-naïve HIV-1B patients (Malet et al., 2011). In another study, the T124A mutation was highly prevalent in INSTI-naïve and RAL-failed patients, and it was significantly associated with HIV-nonB (Saladini et al., 2012). While the T124A mutation was selected under DTG pressure, it does not significantly affect the efficacy of DTG (Kobayashi et al., 2011; Vavro et al., 2013). These results suggest that the T124A mutation affects the binding of RAL more than the binding of DTG. T124 from the inner subunit is within interacting distance with the phosphate group bridging two nucleotides that base pair with the fourth and fifth nucleotides from the 3'-end of the 8-nt long target DNA (Figure 5a). This interaction will be lost with T124A. It is possible that this mutation does not affect DTG binding significantly compared to RAL, as RAL is a larger molecule due to the presence of an oxadiazole ring. In the outer subunits, T124 is exposed at the surface. Hence, the implication of T124A on other viral functions

such as viral fitness cannot be deduced from available structures. From the modeled structures, the interactions of K136Q are difficult to determine for two reasons: (i) the position of K136 in the outer molecules of 5U1C is not known, and (ii) K136 is near the segment that is missing in the cryoEM structure (205–222). Nonetheless, one can speculate that K136 in the inner molecule interacts with residues in the region 205–222. It should also be pointed out here that residues in the 205–222 region in the crystal structure of CCD+CTD (PDB file 1EX4) (Chen et al., 2000) assume a helical conformation, whereas in the HIV-1 intasome structure (PDB file 5U1C) (Passos et al., 2017), they form an unordered structure that appears close to K126/Q136. Amino acid D167 from one inner IN molecule interacts with K42 of the other inner IN molecule (Figure 5b). This is a symmetric interaction similar to that seen for the D25E polymorphism. Due to the longer sidechain of E167 compared to that of D167, there are two interactions with E167, which contributes to additional stability of the intasome tetramer in non-B.

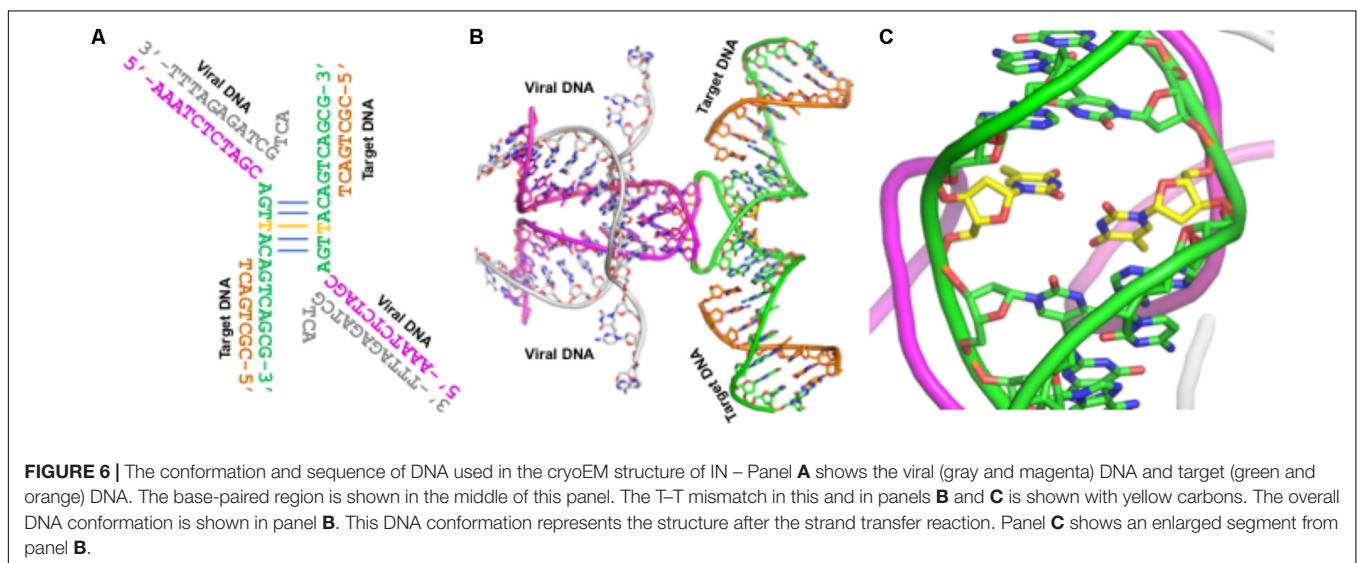
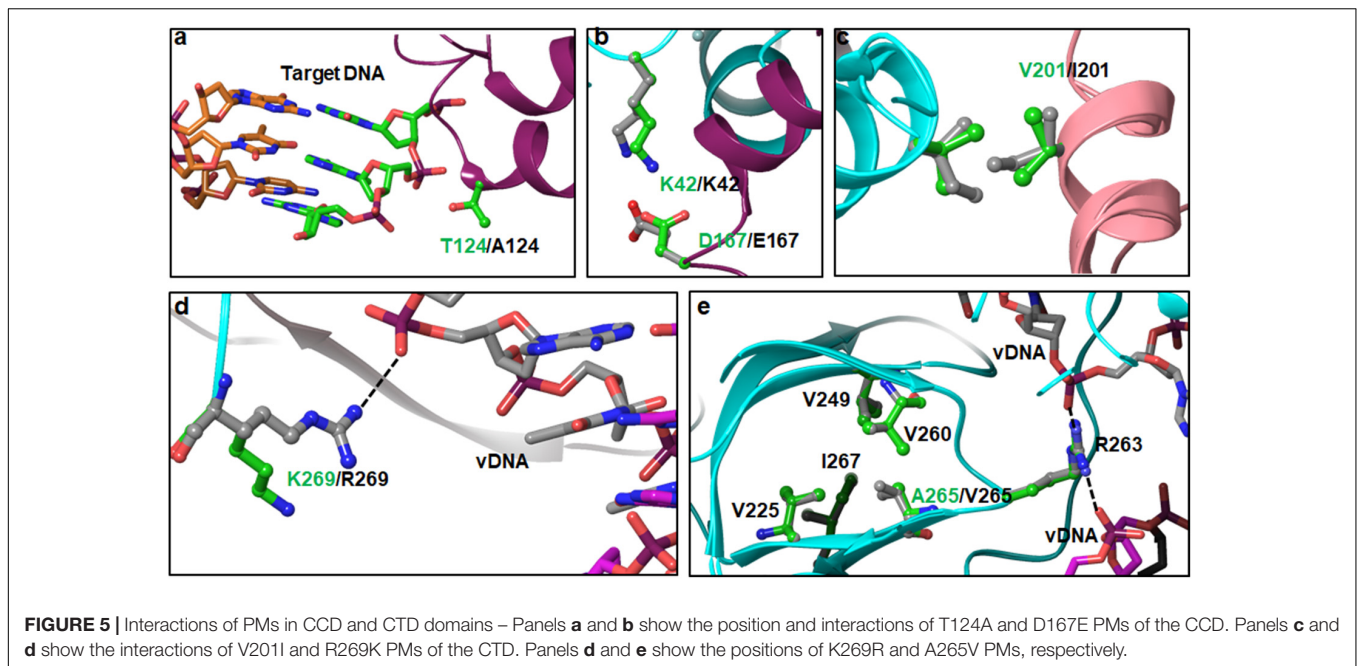
## PMs in the CTD

There are six PMs (V201I, T218I, L234I, A265V, R269K, and S283G) in the CTD. Of these six positions, L234I and S283G are highly polymorphic in HIV-nonB. V201I is highly polymorphic in both HIV-1B and non-B (Figure 1). Amino acid positions 218 and 283 are missing in structures of IN. Hence, the structural impact of PMs at these positions cannot be deduced unambiguously. V201 is at the interface of inner and outer IN molecules (Figure 5c), and it participates in a symmetric interaction. Mutation V201I may increase the buried surface area between two IN molecules by a factor of 265 Å<sup>2</sup>, which empirically can account for ~15 kcal/mol greater binding energy between the two molecules.

In the structures of IN tetramers bound to DNA, amino acid L234 from the outer IN molecules is adjacent to viral DNA. However, it is not within interacting distance. With mutation L234I, although the sidechain is closer to the DNA, it is still far from any sort of interaction with DNA. In the inner IN molecules, L234 is exposed to the solution and does not participate in either protein–protein interactions or protein–DNA interactions.

Amino acid position 265 resides in the SH3-like fold of CTD. We noticed PM A263V in our sequence analyses. In HIV-1B, A/V265 is part of a hydrophobic cluster constructed by V225, V249, V260, A265, and I267 (Figure 5e). A critical amino acid residue with respect to INSTI resistance (R263) is part of the loop connecting V260 and A265 (Figure 5e). The sidechain of R263 interacts with the phosphate backbone of viral DNA. Mutation A265V is expected to change the geometry of the hydrophobic cluster, which may affect the interaction of R263 with viral DNA. In fact, our modeled structure of HIV-1C IN shows a slight change in the conformation of R263K, resulting in altered distances of NH1 and NH2 atoms from the phosphate groups of viral DNA.

In the cryoEM structure, K269 is the last residue that could be resolved. Although our sequence analyses show very low R269K PM in HIV-1B, the cryoEM structure contains K269, and it does not interact with the viral DNA. Hence, we modeled



R269 in place of K269. The results shown in **Figure 5d** indicate that R269 interacts with the phosphate backbone of the viral DNA molecule (shown as dotted line). While this interaction is not seen with K269, there are two confounding factors associated with this phenomenon. First, since it is the last residue in the solved structure, the conformation of the sidechain cannot be unambiguously deduced. The second factor is that the lysine sidechain is quite flexible, leaving the possibility open that K269 will adopt conformations which can interact with DNA.

The cryoEM structure of the HIV-1 STC is a milestone in the structural biology of integrase. By modeling the IN nucleoprotein complexes from different subtypes in combination with sequence analyses, we have elucidated structural aspects

and potential functional impacts of IN PMs. However, our analyses and the conclusions drawn from these analyses should be considered with caution, since there are some unusual features associated with the cryoEM structure of the IN intasome. First, the structure was solved as a fusion protein with Sso7. Whether Sso7 fusion has affected any domain rearrangement of IN remains unknown. Second, the IN construct that was used in structure determination contains an active site mutation (E152Q), although it is highly unlikely that this mutation may have affected the overall structure of the IN/DNA complex. The third and most important feature is that the DNA used in this structure determination contains a T-T mismatch in the double-stranded region (**Figure 6**). This mismatch (shown in yellow carbons



in **Figure 6**) may have affected DNA bending, which in turn would have affected spatial oligomerization of IN molecules in the intasome.

In summary, we have presented extensive analyses of IN PMs in the structural context. Our analyses suggest that there are several naturally occurring polymorphisms that may affect the structural stabilities of the IN and vDNA binding, and drug binding propensity. Future biochemical and virological experiments will provide deeper insights into the functional impacts of sequence variations among IN genes from different subtypes. Importantly, these studies will also provide guidance for investigating how naturally occurring polymorphisms can affect treatment response in large real-life cohorts.

## AUTHOR CONTRIBUTIONS

KS and UN conceived and designed the study. UN conducted sequence analyses. LR and KS conducted structural analyses. GJ and AO provided sequence information on TV cohort. KS, and UN wrote the first draft of the manuscript reviewed by GJ, AO,

SS, AS, and LR. All the authors approved the final version of the manuscript.

## FUNDING

The study is partially funded by the Swedish Research Council Establishment grant (2017-01330) and Jeansson's Stiftelser (JS2016-0185) to UN, the Swedish Research Council grant 2016-01675 to AS and ALF-grant from Stockholm County Council (ALF 20160074) to UN and AS. Part of this research was supported by National Institute of Health grant GM118012 (SS). KS acknowledges support from NIH CTSA grant UL1 TR002345.

## ACKNOWLEDGMENTS

The authors acknowledge the support received from Anoop T. Ambikan to perform statistical analysis and SNP calling. They also acknowledge comments on manuscript by Dr. Donald Burke.

## REFERENCES

- Anstett, K., Brenner, B., Mesplede, T., and Wainberg, M. A. (2017). HIV drug resistance against strand transfer integrase inhibitors. *Retrovirology* 14:36. doi: 10.1186/s12977-017-0360-7
- Antiretroviral Therapy Cohort Collaboration (2017). Survival of HIV-positive patients starting antiretroviral therapy between 1996 and 2013: a collaborative analysis of cohort studies. *Lancet HIV* 4, e349–e356. doi: 10.1016/S2352-3018(17)30066-8
- Ballandras-Colas, A., Brown, M., Cook, N. J., Dewdney, T. G., Demeler, B., Cherepanov, P., et al. (2016). Cryo-EM reveals a novel octameric integrase structure for betaretroviral intasome function. *Nature* 530, 358–361. doi: 10.1038/nature16955
- Ballandras-Colas, A., Maskell, D. P., Serrao, E., Locke, J., Swuec, P., Jonsson, S. R., et al. (2017). A supramolecular assembly mediates lentiviral DNA integration. *Science* 355, 93–95. doi: 10.1126/science.aah7002
- Brado, D., Obasa, A. E., Ikomey, G. M., Cloete, R., Singh, K., Engelbrecht, S., et al. (2018). Analyses of HIV-1 integrase sequences prior to South African national HIV-treatment program and available of integrase inhibitors in Cape Town, South Africa. *Sci. Rep.* 8, 4709–4718. doi: 10.1038/s41598-018-22914-5
- Cai, M., Zheng, R., Caffrey, M., Craigie, R., Clore, G. M., and Gronenborn, A. M. (1997). Solution structure of the N-terminal zinc binding domain of HIV-1 integrase. *Nat. Struct. Biol.* 4, 567–577. doi: 10.1038/nsb0797-567
- Chen, J. C., Krucinski, J., Miercke, L. J., Finer-Moore, J. S., Tang, A. H., Leavitt, A. D., et al. (2000). Crystal structure of the HIV-1 integrase catalytic core and C-terminal domains: a model for viral DNA binding. *Proc. Natl. Acad. Sci. U.S.A.* 97, 8233–8238. doi: 10.1073/pnas.150220297
- Chiu, T. K., and Davies, D. R. (2004). Structure and function of HIV-1 integrase. *Curr. Top. Med. Chem.* 4, 965–977. doi: 10.2174/1568026043388547
- Craigie, R. (2001). HIV integrase, a brief overview from chemistry to therapeutics. *J. Biol. Chem.* 276, 23213–23216. doi: 10.1074/jbc.R100027200
- Craigie, R. (2012). The molecular biology of HIV integrase. *Future Virol.* 7, 679–686. doi: 10.2217/fvl.12.56
- Craigie, R., and Bushman, F. D. (2012). HIV DNA integration. *Cold Spring Harb. Perspect. Med.* 2:a006890. doi: 10.1101/cshperspect.a006890
- Engelman, A. N., and Cherepanov, P. (2017). Retroviral intasomes arising. *Curr. Opin. Struct. Biol.* 47, 23–29. doi: 10.1016/j.sbi.2017.04.005
- Hachiya, A., Kirby, K. A., Ido, Y., Shigemitsu, U., Matsuda, M., Okazaki, R., et al. (2017). Impact of HIV-1 integrase L74F and V75I mutations in a clinical isolate on resistance to second-generation integrase strand transfer inhibitors. *Antimicrob. Agents Chemother.* 61, e315–e317. doi: 10.1128/AAC.00315-17
- Haggblom, A., Svedhem, V., Singh, K., Sonnerborg, A., and Neogi, U. (2016). Virological failure in patients with HIV-1 subtype C receiving antiretroviral therapy: an analysis of a prospective national cohort in Sweden. *Lancet HIV* 3, e166–e174. doi: 10.1016/S2352-3018(16)00023-0
- Hare, S., Gupta, S. S., Valkov, E., Engelman, A., and Cherepanov, P. (2010a). Retroviral intasome assembly and inhibition of DNA strand transfer. *Nature* 464, 232–236. doi: 10.1038/nature08784
- Hare, S., Vos, A. M., Clayton, R. F., Thuring, J. W., Cummings, M. D., and Cherepanov, P. (2010b). 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
- Hare, S., Maertens, G. N., and Cherepanov, P. (2012). 3'-processing and strand transfer catalysed by retroviral integrase in crystallo. *EMBO J.* 31, 3020–3028. doi: 10.1038/emboj.2012.118
- Hemelaar, J. (2012). The origin and diversity of the HIV-1 pandemic. *Trends Mol. Med.* 18, 182–192. doi: 10.1016/j.molmed.2011.12.001
- Honig, B. H., and Hubbell, W. L. (1984). Stability of "salt bridges" in membrane proteins. *Proc. Natl. Acad. Sci. U.S.A.* 81, 5412–5416. doi: 10.1073/pnas.81.17.5412
- Jacobs, G. B., Loxton, A. G., Laten, A., Robson, B., Van Rensburg, E. J., and Engelbrecht, S. (2009). Emergence and diversity of different HIV-1 subtypes in South Africa, 2000–2001. *J. Med. Virol.* 81, 1852–1859. doi: 10.1002/jmv.21609
- Kessler, J. J., Mckee, C. J., Eidahl, J. O., Shkriabai, N., Katz, A., and Kvaratskhelia, M. (2009). HIV-1 Integrase-DNA Recognition Mechanisms. *Viruses* 1, 713–736. doi: 10.3390/v1030713
- Kobayashi, M., Yoshinaga, T., Seki, T., Wakasa-Morimoto, C., Brown, K. W., Ferris, R., et al. (2011). In vitro antiretroviral properties of S/GSK1349572, a next-generation HIV integrase inhibitor. *Antimicrob. Agents Chemother.* 55, 813–821. doi: 10.1128/AAC.01209-10
- Krzywinski, M., Schein, J., Birol, I., Connors, J., Gascoyne, R., Horsman, D., et al. (2009). Circos: an information aesthetic for comparative genomics. *Genome Res.* 19, 1639–1645. doi: 10.1101/gr.092759.109
- Larkin, M. A., Blackshields, G., Brown, N. P., Chenna, R., McGettigan, P. A., McWilliam, H., et al. (2007). Clustal W and clustal X version 2.0. *Bioinformatics* 23, 2947–2948. doi: 10.1093/bioinformatics/btm404
- Li, M., Jurado, K. A., Lin, S., Engelman, A., and Craigie, R. (2014). Engineered hyperactive integrase for concerted HIV-1 DNA integration. *PLoS One* 9:e105078. doi: 10.1371/journal.pone.0105078



- Maertens, G. N., Hare, S., and Cherepanov, P. (2010). The mechanism of retroviral integration from X-ray structures of its key intermediates. *Nature* 468, 326–329. doi: 10.1038/nature09517
- Malet, I., Wirlden, M., Fourati, S., Armenia, D., Masquelier, B., Fabeni, L., et al. (2011). Prevalence of resistance mutations related to integrase inhibitor S/GSK1349572 in HIV-1 subtype B raltegravir-naive and -treated patients. *J. Antimicrob. Chemother.* 66, 1481–1483. doi: 10.1093/jac/dkr152
- Neogi, U., Singh, K., Aralaguppe, S. G., Rogers, L. C., Njenda, D. T., Sarafianos, S. G., et al. (2018). Ex-vivo antiretroviral potency of newer integrase strand transfer inhibitors cabotegravir and bictegravir in HIV type 1 non-B subtypes. *AIDS* 32, 469–476. doi: 10.1097/QAD.0000000000001726
- Passos, D. O., Li, M., Yang, R., Rebensburg, S. V., Ghirlando, R., Jeon, Y., et al. (2017). Cryo-EM structures and atomic model of the HIV-1 strand transfer complex intasome. *Science* 355, 89–92. doi: 10.1126/science.aah5163
- Rhee, S. Y., Gonzales, M. J., Kantor, R., Betts, B. J., Ravela, J., and Shafer, R. W. (2003). Human immunodeficiency virus reverse transcriptase and protease sequence database. *Nucleic Acids Res.* 31, 298–303. doi: 10.1093/nar/gkg100
- Saladini, F., Meini, G., Bianco, C., Monno, L., Punzi, G., Pecorari, M., et al. (2012). Prevalence of HIV-1 integrase mutations related to resistance to dolutegravir in raltegravir naive and pretreated patients. *Clin. Microbiol. Infect.* 18, E428–E430. doi: 10.1111/j.1469-0691.2012.03917.x
- Singh, K., Marchand, B., Rai, D. K., Sharma, B., Michailidis, E., Ryan, E. M., et al. (2012). Biochemical mechanism of HIV-1 resistance to rilpivirine. *J. Biol. Chem.* 287, 38110–38123. doi: 10.1074/jbc.M112.398180 PMID: PMC3488081 doi: 10.1074/jbc.M112.398180
- Takahata, T., Takeda, E., Tobiume, M., Tokunaga, K., Yokoyama, M., Huang, Y. L., et al. (2017). Critical contribution of Tyr15 in the HIV-1 integrase (IN) in facilitating IN assembly and nonenzymatic function through the IN precursor Form with reverse transcriptase. *J. Virol.* 91, e2003–e2016. doi: 10.1128/JVI.02003-16
- Thierry, E., Deprez, E., and Delelis, O. (2016). Different pathways leading to integrase inhibitors resistance. *Front. Microbiol.* 7:2165. doi: 10.3389/fmicb.2016.02165
- Tsiang, M., Jones, G. S., Goldsmith, J., Mulato, A., Hansen, D., Kan, E., et al. (2016). 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
- Vavro, C., Hasan, S., Madsen, H., Horton, J., Deanda, F., Martin-Carpenter, L., et al. (2013). Prevalent polymorphisms in wild-type HIV-1 integrase are unlikely to engender drug resistance to dolutegravir (S/GSK1349572). *Antimicrob. Agents Chemother.* 57, 1379–1384. doi: 10.1128/AAC.01791-12
- Wang, J. Y., Ling, H., Yang, W., and Craigie, R. (2001). Structure of a two-domain fragment of HIV-1 integrase: implications for domain organization in the intact protein. *EMBO J.* 20, 7333–7343. doi: 10.1093/emboj/20.24.7333
- Wensing, A. M., Calvez, V., Gunthard, H. F., Johnson, V. A., Paredes, R., Pillay, D., et al. (2017). 2017 Update of the drug resistance mutations in HIV-1. *Top. Antivir. Med.* 24, 132–133.
- Whitfield, T., Torkington, A., and Van Halsema, C. (2016). Profile of cabotegravir and its potential in the treatment and prevention of HIV-1 infection: evidence to date. *HIV AIDS* 8, 157–164. doi: 10.2147/HIV.S97920
- Wimley, W. C., Gawrisch, K., Creamer, T. P., and White, S. H. (1996). Direct measurement of salt-bridge solvation energies using a peptide model system: implications for protein stability. *Proc. Natl. Acad. Sci. U.S.A.* 93, 2985–2990. doi: 10.1073/pnas.93.7.2985
- Yin, Z., Shi, K., Banerjee, S., Pandey, K. K., Bera, S., Grandgenett, D. P., et al. (2016). Crystal structure of the Rous sarcoma virus intasome. *Nature* 530, 362–366. doi: 10.1038/nature16950

**Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2018 Rogers, Obasa, Jacobs, Sarafianos, Sönnnerborg, Neogi and Singh. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.