HIV-1 Mutants that Escape the Cytotoxic T-Lymphocytes are Defective in Viral DNA Integration

**Muthukumar Balasubramaniam^{1,2*}, Santosh Thapa^{1,2}, Benem-Orom Davids^{1,2,3}, Alex Bryer⁵,
Chaoyi Xu⁵, Jiong Shi⁶, Christopher Aiken⁶, Jui Pandhare^{1,3,4}, Juan R. Perilla⁵, and
Chandravanu Dash^{1,2*}**

¹The Center for AIDS Health Disparities Research

²Department of Biochemistry and Cancer Biology

³School of Graduate Studies and Research

⁴Department of Microbiology and Immunology
Meharry Medical College, Nashville, Tennessee.

⁵Department of Chemistry
University of Delaware, Newark, Delaware.

⁶Department of Microbiology, Pathology, and Immunology
Vanderbilt University Medical Center, Nashville, Tennessee.

***Correspondence:** cdash@mmc.edu and muthukumarb@mmc.edu

ABSTRACT

HIV-1 replication is durably controlled in certain untreated HIV-1-infected individuals expressing particular human leukocyte antigens (HLA). These HLAs tag infected cells for elimination by presenting specific viral epitopes to CD8⁺ cytotoxic T-lymphocytes (CTL). In individuals expressing HLA-B27, CTLs primarily target the capsid protein (CA)-derived KK10 epitope. Selection of CA mutation R264K helps HIV-1 escape the CTL response but severely diminishes virus infectivity. Here we report that the R264K mutation-associated infectivity defect arises primarily from impaired viral DNA integration. Strikingly, selection of the compensatory CA mutation S173A or depletion of host cyclophilin A largely rescues the R264K-associated integration and infectivity defects. Collectively, our study reveals novel mechanistic insights into the fitness defect incurred by an HIV-1 variant escaping a CA-directed CTL response.

KEY WORDS

Human immunodeficiency virus (HIV), Cytotoxic T lymphocytes (CTL), Capsid, Integration, Reverse transcription, Nuclear entry

INTRODUCTION

Over 35 million people have already died from the ongoing HIV/AIDS pandemic, and no curative therapy is available for the ~38 million people currently living with the virus. The standard of care is antiretroviral therapy (ART) involving combination of drugs targeting HIV-1-encoded enzymes¹. The ART, by controlling the viral load and preventing disease progression, has significantly reduced the mortality and morbidity associated with HIV-1 infections²⁻⁶. However, because ART does not eradicate the virus⁷, evolution and spread of drug-resistant virus strains significantly increase the risks of HIV-1 transmission and disease progression⁸⁻¹². Therefore, continued research directed towards novel therapeutic¹³⁻¹⁶ and preventive strategies are urgently needed to mitigate the devastating impact of the global HIV pandemic.

There is strong evidence that the host immune system can naturally control HIV and delay disease progression¹⁷⁻²¹. For instance, the virus replication is durably suppressed and disease progression is delayed in a small subset of untreated HIV-1-infected individuals known as “elite controllers”²²⁻²⁴. Such long-term asymptomatic infection in the absence of ART has been attributed to robust immunological control mechanisms coordinated by human leukocyte antigens (HLA)²⁵⁻²⁸ and CD8+ cytotoxic T-lymphocytes (CTLs)²⁹⁻³⁷. Notably, a strong association between HIV-1 control and presentation of viral epitopes by HLA-B57 or HLA-B27 to HIV-1-specific CTLs has been demonstrated^{18,26,38-45}. However, CTL-mediated HIV-1 control is compromised by escape mutations that diminish recognition of viral epitopes by the HLA and/or CTLs⁴⁶⁻⁵³.

CTL response against the HLA-B27-restricted KK10 epitope⁵⁴, derived from the capsid (CA) domain of the HIV-1 Gag protein, has been linked to robust virus control and delayed disease progression^{17,38,55-57}. The immunological constraint on virus replication eventually leads to the emergence and selection of an HIV-1 variant containing mutations in the KK10 epitope of CA^{55,58-63}; specifically, the prerequisite mutation L268M (LM) emerging early followed by the consequential escape mutation R264K (RK) selected late in the course of infection^{46,55,64}. The

CTL escape mutation RK inhibits the binding and presentation of the CA-derived mutant peptide by HLA-B27^{55,65} but also inflicts a significant fitness cost to the infectivity of the virus^{46,51,55,61,64-67}. This is unsurprising because CA is genetically the most fragile of all HIV-1 proteins^{68,69} and is a key player in multiple stages of viral replication⁷⁰⁻⁷⁸. For instance, the CA forms the conical capsid shell that houses the viral genome and other viral/cellular factors necessary for infection⁷⁹⁻⁸⁵. CA also provides critical binding interfaces for key host factors like cyclophilin A (CypA)⁸⁶⁻⁸⁸ and cleavage and polyadenylation-specific factor 6 (CPSF6)^{89,90} to promote early steps of HIV-1 infection⁹¹⁻⁹⁸. Remarkably, the infectivity defect of the RK mutant virus is largely restored upon selection of the compensatory CA mutation S173A (SA)⁶¹ or upon depletion of CypA^{61,66}. However, the mechanism(s) by which these CTL-escape CA mutations reduce or restore HIV-1 replication remains largely undefined.

Here we report the comprehensive biology of the CTL-escape RK mutant virus by using a multi-pronged approach that included structure-guided molecular dynamics (MD) simulations, binary protein interaction studies by yeast two hybrid (Y2H) assay, virus infectivity assays, biochemical characterization of isolated viral replication complexes, and genetic analysis of the viral genome. Collectively, our results have uncovered reduced HIV-1 integration as the principal mechanism underlying the infectivity defect of the RK variant. Notably, these findings also provide further evidence for a regulatory role of CA in post-nuclear entry steps of HIV-1 infection and reinforces the need of research focused on CA-targeting inhibitors and CTL-based vaccines.

RESULTS

CTL escape-associated CA mutations do not alter capsid structure

The HIV-1 capsid is assembled through specific intra- and inter-molecular interactions between ~1500 copies of CA monomers organized into a lattice of ~240 hexameric and 12 pentameric units^{79,80,99-101}. Since the structural integrity of the capsid is essential for infection¹⁰²,

we employed structure-based molecular modeling and MD simulations to probe whether the CTL escape CA mutations alter stability of the CA hexamers and CA tubular assemblies. Fig. 1A depicts the relative positions of the CTL escape-associated amino acids [R264(R132), L268(L136), and S173A(S41)] and the substitution changes [R264K, L268M, S173A] linked to the HLA-B27-restricted KK10 epitope in the CA. All three amino acids—R264, L268, and S173—are located in the NTD-NTD interface between two neighboring CA monomers and are positioned on the exterior of the CA hexamer (Fig. 1B) and pentamers (Fig. 1C). Notably, none of the three amino acids is in the proximity of CypA binding loop (BL) or the CPSF6 binding pocket.

MD simulations showed similar dynamic properties of the WT and CTL escape mutant hexamers (Fig. 2). Specifically, the C α root mean square deviations (RMSDs) of all the CA hexamer systems showed an increase at 20 ns and were plateaued at around 2Å after 50 ns (Fig. 2A). Importantly, the mean C α root mean square fluctuations (RMSFs) of all systems exhibited negligible differences, especially in the structured regions (Fig. 2A), and the structural flexibilities of all the hexamer systems were comparable (Fig. 2B). The inter-hexamer interactions of the CA residues R264, L268, and S173 in the WT hexamer are maintained in the mutant hexamers (Fig. 2C). For example, the top two contacts made by R264 (with residues I261 and I267), by L268 (with residues W265 and N271), and by S173 (with residues S176 and P170) are preserved in mutant hexamers. While the LM and SA substitutions preserve the electrostatic potential, the RK substitution renders the sidechain a relatively more positive potential (Fig. S1). Nevertheless, the overall electrostatic potentials of all the mutant hexamers are very similar to the WT hexamer (Fig. S1). Further, we probed for the presence of ions in the CA hexamers to better understand the interactions between the capsid and its native environment (Fig. S2). Specifically, chloride ions form an inner layer adjacent to the surface of the capsid, while sodium binds to the exterior of the capsid. Examination of the transfer rates for both ionic species present in the simulation reveal regions of high occupancy for chloride (cyan) and sodium (yellow) (Fig. S2). Comparison of ion

occupancy of WT hexamers to that of the RK, RKLM and RKLMSA hexamers, showed no significant difference in the interaction between the capsid and the ionic environment.

Furthermore, we constructed and simulated RK, RKLM and RKLMSA mutant CA tubes to test the effects of these mutations on the structure and dynamics of the tubular capsid assembly (Fig. 3). After 20 ns of MD equilibration, no structural differences between any of the CA tubes were observed. Tube radii and length values vary by no more than 0.1-0.2 nm for all mutant systems studied (Fig. 3A). Moreover, the conformational distributions of each mutation site were also quite similar (Fig. 3B), as well as the RMSF of capsid residues (Fig. 3C), indicating that mutations do not alter the dynamics of capsid lattice. These observations indicate the structural and dynamic similarity between these mutant CA tubes despite the presence of mutations.

The HIV-1 capsid has many binding interfaces for cellular co-factor recognitions^{72,103}. For example, CypA selectively bridges two neighboring CA hexamers using canonical and non-canonical binding sites^{88,104}, both located at the CypA binding loop (BL). The CPSF6⁹⁰ and Nucleoporin 153 (NUP153)¹⁰⁵ target the CA NTD-NTD interface. Other co-factor binding sites on capsid include the center pore of CA hexamer [to recognize IP6¹⁰⁶, nucleotide¹⁰⁷, and Fasciculation and Elongation Protein Zeta 1 (FEZ1)¹⁰⁸] and the tri-hexamer region [to recognize MX dynamin like GTPase 2 (MXB)¹⁰⁹]. However, the locations of the CTL escape RK mutations are far away from these co-factor binding sites on the capsid (Fig. 1B) and render almost identical structures and dynamics of the CA hexamer and tube, which are unlikely to affect these co-factor recognitions. Together, our *in silico* studies indicate that the CTL escape-associated CA amino acid changes linked to the HLA-B27-restricted KK10 epitope do not alter the structural integrity and dynamics of the CA hexamers and capsid lattice, as well as the capsid-cofactor interactions.

CA mutation RK does not perturb CA-CA and CA-host protein interactions

We next assessed the effect of the RK mutation on CA-CA and CA-host factor interactions by yeast GAL4-based two-hybrid (Y2H) assay¹¹⁰⁻¹¹². We focused on the RK mutation because it

is the primary determinant of the infectivity defect associated with HLA-B27 CTL escape⁶¹. First, we examined the well-defined dimerization of WT CA monomers. The yeast cells co-transformed with the plasmid pairs encoding the GAL4 DNA Binding Domain (BD) in fusion with the WT CA (BD-CA_{WT}) and the GAL4 Activation Domain (AD) in fusion with the WT CA (AD-CA_{WT}), exhibited histidine prototrophy and α -galactosidase activity (Fig. 4A, Top panel), thus indicating direct protein-protein interaction. Yeast cells co-transformed with plasmid pairs encoding BD-CA_{WT} and AD-mCherry or AD-CA_{WT} and BD-mCherry fusion proteins exhibited auxotrophic growth on selection media lacking leucine and tryptophan but did not exhibit histidine prototrophy and α -galactosidase activity. These results highlight specificity of the Y2H assay in detecting bona fide CA-associated binary protein interactions.

Importantly, we observed histidine prototrophy and α -galactosidase activity by the yeast co-expressing BD-CA_{RK} (BD in fusion with the RK CA) and AD-CA_{RK} (AD in fusion with the RK CA), but not by the yeast co-expressing BD-CA_{RK} and AD-mCherry or AD-CA_{RK} and BD-mCherry (Fig. 4A, Bottom panel). These results demonstrate that the CTL-escape RK mutation does not disrupt CA-CA binary interaction. Next, we tested whether the RK mutation affected interaction of CA with two key host factors-CypA and CPSF6 that facilitate post-entry steps of infection and are implicated in the infectivity defect of the RK variant virus. Interestingly, histidine prototrophy and α -galactosidase activity was observed by the yeast co-expressing BD-CA_{WT} or BD-CA_{RK} and AD-CypA (Fig. 4B) or AD-CPSF6 (Fig. 4C), but not by the yeast co-expressing one of these fusion proteins and the complementary negative control mCherry fusion protein. These results show that the RK mutation has no significant effect on the binary interactions of CA with either CypA or CPSF6. Collectively, the Y2H assay data strengthen our *in silico* studies and indicate that the CTL escape CA mutation RK does not compromise the structural integrity of CA or its functional interactions with key host factors.

RK mutation impairs HIV-1 infection but the compensatory mutation SA restores infectivity

HIV-1 escape from the KK10 epitope-targeted CTL response is functionally linked to specific CA mutations^{55,58,59,61,62}. The prerequisite LM mutation, independently, has no significant effect on HLA binding, CTL recognition, and viral infectivity^{61,113}. However, the RK mutation independently or in combination with LM (RKLM) significantly reduces the binding of the KK10 epitope to the HLA-B27 molecule and drastically impairs infectivity^{61-63,65}. Strikingly, the compensatory CA mutation S173A confers WT-level, or more, infectivity to the RK and RKLM mutant viruses^{61,67}. Therefore, to understand the precise mechanism by which these CTL-escape CA mutations affect HIV-1 infectivity, we compared the infectivity of the WT and mutants using replication-competent virus particles. The LM mutation alone does not significantly affect infectivity and hence was not included in the assessment.

For the TZM-bl-based infectivity assay, we inoculated TZM-bl cells¹¹⁴ with equivalent MOI of WT or the mutant virus and 48 hours post infection (hpi) assessed the luciferase activity in the cellular lysates. Compared to the WT virus, the infectivity of the RK and the RKLM mutant viruses were drastically reduced, by up to 10-fold (Fig. 5A). The comparable infectivity of the RK and the RKLM variants indicates that the RK, and not the LM mutation, is the principal determinant of the infectivity defect. Importantly, the addition of the compensatory SA mutation to the RKLM virus led to restoration of infectivity to the level of the WT virus (Fig. 5A and 5B). For the infectivity assays using Jurkat T-cells, we spinoculated cells with equivalent MOI of WT or any of the mutant viruses and assessed the HIV-1 Gag protein levels by flow cytometry. As shown in Fig. 5C-D, in comparison to the WT virus, the infectivity of the RK and RKLM mutant viruses were significantly reduced and the compensatory SA mutation led to significant restoration of its infectivity. Collectively, our infection assays in two different cell types establish that the CTL-escape CA mutation RK severely impairs HIV-1 infectivity and this defect is mostly rescued by the compensatory SA mutation.

RK-associated infectivity defect is not due to block in the late stages of virus replication

HIV-1 CA plays key roles during early and late events of virus replication. During the late stages, the CA-CA interaction is essential for hexameric Gag lattice formation and assembly of immature progeny viruses^{115,116}. Furthermore, cleavage of CA from the Gag precursor is critical for the formation of the conical capsid of the mature virions¹¹⁷. Therefore, the impaired infectivity of the RK mutant virus (Fig. 5A-D) could be a consequence of the CA mutation adversely affecting any of these late-stage processes. To test this possibility, we probed the effects of the CTL escape CA mutations on Gag processing and virus particle production.

HEK293T cells were transfected with the molecular clones of the WT or the CA mutants and cultured for 48 hours, after which the cells and the corresponding culture supernatants were collected. The virus proteins from the cells and from the pelleted virus particles were probed by western blotting (Fig. 5E). The Gag processing efficiency in the producer cell was not significantly altered by any of the CA mutations. Further, none of the CA mutations were found to reduce virus release efficiency—measured as the ratio of the amount of virus-associated CA to the total Gag (total amount of virus-associated CA, cell-associated Pr55Gag, cell-associated p41Gag and cell-associated CA) (Fig. 5F). Taken together, these results suggest that the CTL escape CA mutations do not significantly alter the late stages of HIV-1 replication and that the RK mutation-associated infectivity defect maps to one or more of the post-entry steps of the virus replication.

RK mutation has no significant effect on HIV-1 reverse transcription and nuclear entry

The CA mutations RK and RKLM are not positioned to significantly alter the structural integrity of the capsid structures and, by extension, the reverse transcription of the mutant viruses. Intriguingly, the infectivity defect imposed by the RK and RKLM mutations have previously been reported to arise from reduction in reverse transcription⁶¹. So, we quantified levels of reverse transcription in T-cells inoculated with these viruses. Jurkat or SupT1 cells spinoculated¹¹⁸ for 2 hours with equivalent MOI of the WT or the mutant viruses were cultured for 24 hours, after which

total DNA was isolated from the cell samples. The viral late RT products (i.e., viral DNA) in the samples was measured by qPCR and the copy numbers of the viral DNA were calculated by interpolation from a standard curve. The results revealed that the RK and the RKLM mutations did not impose significant alterations in the reverse transcription levels (Fig. 6A-B) that could be correlated with the magnitude of the impairment in virus infectivity (Fig. 5A-D). The addition of the compensatory SA mutation to the RKLM virus had no significant effect on the reverse transcription levels, as well. These results diverge from published work⁶¹ and suggest that the RK mutation-associated infectivity defect arises from a block downstream of the reverse transcription step.

HIV-1 nuclear import is the next step in the virus replication cycle and prior reports did not show the effect of RK mutation on viral nuclear import^{61,67}. So, we tested the nuclear import efficiency of the WT and the CA mutant viruses by quantitating the viral 2-LTR circles in infected cells. A canonical 2-LTR circle is generated in the nucleus when the two LTRs in an unintegrated viral DNA are joined by the host non-homologous DNA end-joining pathway^{119,120}. Therefore, the viral 2-LTR circles are the most commonly used surrogate marker for estimating the efficiency of the nuclear import of the viral DNA^{121,122}. The 2-LTR circles were measured by qPCR and their copy numbers were calculated by interpolation from a standard curve generated in parallel using known copy numbers of a plasmid containing a cloned 2-LTR sequence (Fig. 6C-D). The results show that the RK and the RKLM mutations did not impose significant reductions in the levels of nuclear import that could be correlated with the magnitude of the infectivity defect (Fig. 5A-D). Further, the RKLMSA virus did not exhibit any significant difference in the nuclear import levels. As the 2-LTR circles are derivatives of the late RT products and the RKLM virus showed a marginal reduction in reverse transcription (Fig. 6A-B) and nuclear entry (Fig. 6C-D), we calculated the ratio of 2-LTR circles to the late RT products to measure nuclear import efficiency (Fig. 6E). No significant difference in the nuclear import efficiency of the WT and CA mutant viruses was observed. These findings suggest that the CTL-associated CA mutations minimally affect the efficiency of reverse transcription and nuclear import of HIV-1.

RK mutation blocks proviral integration and SA compensatory mutation restores it

Our results in Fig. 6 indicated that a post nuclear import defect is most likely responsible for the infectivity defect caused by the RK mutation. To test this, we used the *Alu*-gag nested PCR assay to compare the levels of chromosomally integrated viral DNA (proviral DNA) in cells inoculated with WT or the mutant viruses. The *Alu*-gag nested PCR assay takes advantage of the widespread occurrence of the *Alu* sequence in the human genome^{123,124}, the reported preferential integration of HIV-1 near such *Alu* sequences¹²⁵⁻¹²⁷, and the selective exponential amplification of only the chromosomally integrated viral DNAs in the 1st round end-point PCR followed by their quantitation in the 2nd round qPCR^{128,129}.

Jurkat or SupT1 cells spinoculated for 2 hours with equivalent MOI of the WT or any of the mutant viruses were cultured for 24 hours, after which total DNA was isolated from the cell samples. The copy numbers of proviral DNA were calculated by interpolation from a standard curve generated in parallel using known copy numbers of pNL43 during the 2nd round qPCR (Fig. 7A-B). As the late RT products constitute the sole source of the proviral DNAs, the viral late RT products in the same samples were also measured by qPCR in parallel (Fig. 7C-D). The integration efficiency was calculated by plotting the data as the ratio of the copy numbers of proviral DNAs to the late RT products (Fig. 7E). Strikingly, the results revealed a significant reduction in the proviral DNA levels (Fig. 7A-B) and integration efficiency (Fig. 7E) of the RK and RKLM viruses when compared to the WT infection. Importantly, the addition of the compensatory SA mutation to the RKLM virus (RKLM_{SA}) led to significant restoration of its integration efficiency (Fig. 7E), which largely correlated with the magnitude of the restoration of the virus infectivity (Fig. 5A-D). These findings, together with our data of reverse transcription and nuclear import, establish that the infectivity defect of the RK and the RKLM viruses are primarily due to the reduction in viral DNA integration into the chromosomal DNA.

RK mutation has no significant effect on the integration activity of the cytoplasmic PICs

Integration of the HIV-1 DNA is carried out by the pre-integration complex (PIC)¹³⁰. Therefore, the significant reduction in proviral integration of the RK variant could be the consequence of a reduction in PIC-associated viral DNA integration activity. To test this, we compared the integration activity *in vitro* of PICs extracted from cells inoculated with the WT or RK virus. Due to the well-recognized technical challenges associated with isolating the nuclear PICs, we used the cytosolic extracts from Jurkat or SupT1 cells^{74,131,132}. The integration activity of these PICs is the measure of the number of viral DNA copies integrated into a heterologous target DNA and is quantified via a nested PCR assay. Intriguingly, compared to the WT PICs, there was no reduction in the integration activity of the RK mutant PICs (Fig. 8A-B and Fig. S3A-B). Notably, there was no reduction in the viral DNA content of the RK mutant PICs when compared to the WT PICs (Fig. 8C-D and Fig. S3C-D). Accordingly, the integration efficiency (i.e., the ratio of the integration activity to the viral DNA content) between the WT and the RK mutant PICs did not differ significantly (Fig 8E) or were comparable (Fig. S3E). Collectively, these results suggest that the integration activity of the RK mutant PICs is not defective and the RK mutation-associated reduction in proviral integration is not due to faulty assembly of PICs.

CypA depletion restores the integration efficiency of the RK mutant virus

Interaction between HIV-1 CA and CypA is essential for optimal HIV-1 infection and any perturbation of this interaction affects reverse transcription and nuclear import^{91-94,97,98}. Interestingly, absence of CypA in the target cell has been reported to rescue the infectivity defect of the RK mutant virus^{61,66}. The underlying mechanism behind this atypical restriction effect of CypA on the RK mutant virus infection remains unclear. Nevertheless, because our results indicated that the RK mutation impairs HIV-1 infection primarily at the integration step, we hypothesized that the rescue of the infectivity defect by the absence of CypA is a manifestation of restored integration. To test our hypothesis, we assessed reverse transcription, nuclear import

and integration efficiency of the WT and the RK mutant in the parental Jurkat (Jurkat_{CypA+/+}) and CypA knockout Jurkat (Jurkat_{CypA-/-}) cells¹³³. These cells were spinoculated for 2 hours with equivalent MOI of the WT or RK mutant virus and then cultured for 24 hours, after which total DNA was isolated. The copy numbers of late RT products, 2-LTR circles, and the proviral DNA levels in the DNA samples were measured by qPCR.

The RK mutation did not alter the levels of reverse transcription (Fig. 9A and 9B) and 2-LTR circles (Fig. 9C and 9D), or the nuclear import efficiency (Fig. 9E) in the Jurkat_{CypA+/+} cells, thereby corroborating the data presented in Fig. 6A-E. The ~2-fold reduction in the late RT product levels of the WT virus in the Jurkat_{CypA-/-} cells has previously been reported to be due to the lack of the positive effect of CypA on HIV-1 reverse transcription⁹⁴. Notably, our results showing ~2-fold reduction in the late RT product levels of the RK mutant virus in the Jurkat_{CypA-/-} cells indicate that the reverse transcription of the RK mutant virus is equally subject to the negative impact of the lack of CypA in the target cells. Interestingly, the proviral DNA copies (Fig. 9F and 9G) and the calculation of the virus integration efficiency by normalizing the proviral DNA levels to the corresponding late RT product levels and the plotting of the data as percentage relative to WT virus in parental Jurkat cells (Fig. 9H), revealed that the lack of CypA restored the proviral integration efficiency of the RK mutant virus to almost WT levels (Fig. 9F-H). These findings suggest that the rescue of the infectivity defect of the RK mutant in CypA depleted cells is primarily due to the restoration of the viral integration.

RK mutant viral DNA ends are appended with aberrant viral sequences

Our data showing that the RK mutation significantly inhibits viral DNA integration without affecting reverse transcription levels, nuclear import efficiency, and PIC function early during infection was intriguing. We hypothesized that the reduction in viral DNA integration into nuclear chromatin becomes evident only after the PIC enters the nucleus. Therefore, we tested whether the RK mutation compromises the integrity of the viral DNA ends after the nuclear entry step of

infection. The 2-LTR circles, which are generated when the open ends of the unintegrated linear viral DNAs are joined by host cell machinery^{119,120} represent an elegant experimental tool to probe the ends of the unintegrated viral DNAs¹³⁴⁻¹³⁶ and thereby assess their competency for integration. Thus, we examined the nucleotide sequences of the junctions of the 2-LTR circles in cells inoculated with the WT or the RK mutant virus. The junctions of the 2-LTR circles present in the total DNA samples from duplicate infections of each virus were first amplified by nested PCR and then cloned into a plasmid vector. For each virus type, at least 100 clones were derived from two independent PCR, and the sequences of the cloned 2-LTR junctions were determined by Sanger sequencing and analyzed by multiple alignment.

Results from this analysis revealed two interesting patterns (Table 1, Fig. S4A and S4B). First, compared to the WT virus, a higher proportion of the atypically appended aberrant viral sequences at the RK mutant viral DNA ends were extensively longer. Two, a significantly higher proportion of these aberrant sequences mapped to the viral poly-purine tract (PPT) and/or the tRNA/primer-binding site (PBS) sequences. It is possible that this atypical retention of the PPT (plus-strand primer) and tRNA (minus-strand primer) sequences at the RK mutant viral DNA ends is a contributory factor in its impaired integration. These results raised the possibility that the presence of the compensatory CA mutation SA may preclude the atypical retention of such aberrant PPT and/or PBS sequences at the ends of the RKLMSA viral DNAs. Remarkably, compared to the RK mutant viral DNAs, the proportion of the RKLMSA mutant viral DNAs with aberrant PPT and/or PBS sequences were significantly low and were even lesser than the WT sequences (Table 1, Fig. S4C). Together, these results suggest that the retention of the PPT and PBS sequences at the ends of the RK mutant viral DNAs may be contributing to their impaired integration and may also be indicative of an underlying qualitative glitch in reverse transcription (Figure 10).

DISCUSSION

The human immune response against the founder HIV-1 is executed by the CTLs in an HLA genotype-dependent manner. The CTL response against the HIV-1 KK10 epitope has been linked to robust virus control and delayed disease progression in certain HLA-B27-positive infected individuals. Selection of the CA mutation RK, associated with viral escape, disrupts epitope binding to HLA-B27 but also causes a severe infectivity defect. The compensatory CA mutation SA relieves the RK-associated virus infectivity defect and triggers loss of viremia control and progression to AIDS. In this study, we have identified impaired viral DNA integration as the principal molecular mechanism underlying the infectivity defect of the RK mutant virus. Our data also establishes that the RK mutation has limited impact on the efficiency of reverse transcription and nuclear import steps of virus replication and the SA compensatory mutation largely rescues the integration defect of the CTL-escape RK mutant virus.

Specific mutations in the genetically fragile CA compromise capsid integrity and disrupt early events of HIV-1 replication. While Schneidewind et al.⁶¹ reported that the RK mutation had no significant effect on capsid stability, work by Schommers et al.⁶⁷ revealed a modest increase in the stability of the RK mutant capsid. It is noteworthy that the RK mutation in the CA does not appear to overlap with the critical structural elements of the capsid. Accordingly, our molecular modeling and MD simulations demonstrated that the RK mutation neither altered the integrity of the hexameric, pentameric or the tubular structure of the capsid. Further, the RK mutation minimally affected interaction with CypA- a critical host factor known to modulate the capsid stability. Our Y2H assays corroborated these findings revealing that the binary interaction between the RK CA or the RK CA and CypA are largely preserved.

Our results confirmed that the RK mutation—independently or in combination with the LM mutation—causes a drastic loss of infectivity but could be rescued by the compensatory SA mutation. Interestingly, our infection assays carried out in an epithelial cell line (TZM-bl) and in a

T-cell line (Jurkat) with native envelope-containing viruses inform that the consequence of the RK, RKLM, and RKLMSA mutation on infectivity are unlikely to be dependent on cell types. Further, the infectivity defect of these variants irrespective of whether the virus contains the native HIV-1 envelope or the VSVg-pseudotyped envelope^{51,67} suggests that the mode of viral entry, i.e., fusion vs endocytosis, does not influence how these CA mutations impair infectivity. Our results of the Gag processing and virus release efficiency also preclude effect of these mutations on late stages of infection. Therefore, the RK-associated infectivity defect most likely arises from impairment of one or more of the early post-entry steps of infection.

Our findings that the RK and RKLM mutations are unlikely to perturb the integrity of the capsid and its interaction with CypA suggested that these CA mutations may not affect the reverse transcription efficiency. However, Schneidewind et al⁶¹ reported that the RK mutation, despite having no effect on the capsid stability, caused a severe reduction in HIV-1 reverse transcription levels. In contrast, our quantification of reverse transcription levels did not reveal any significant difference between the WT and the RK or RKLM mutant viruses. While Schneidewind et al.⁶¹ investigated reverse transcription using VSVg-pseudotyped virus that enters target cells by endocytosis, we measured the reverse transcription levels using native envelope-containing RK mutant virus that enters target cells by CD4-receptor mediated membrane fusion. This raises the possibility of a pseudotyping-associated effect on the RK mutant virus reverse transcription, since VSVg pseudotyping has been reported to alter the phenotype of certain CA mutants^{137,138}.

The lack of a significant effect of the RK mutation on HIV-1 reverse transcription levels in our study implies that this CA mutation disrupts one or more of the downstream steps in HIV-1 infection. Our quantification of the 2-LTR circles, widely considered a reliable surrogate marker for HIV-1 nuclear import, showed no significant differences in the levels of nuclear import of the WT or RK or RKLM mutant virus. It is increasingly becoming clear that the interaction between the HIV-1 CA and the host CPSF6 is essential for optimal nuclear import of the virus. Interestingly, results from our molecular modeling, MD simulations, and yeast two-hybrid assays showing no

significant effect of the RK or RKLM mutations on the interaction between CA and CPSF6 further supports our 2-LTR circle data and the improbability of a RK mutation-induced defect in viral nuclear import. Accordingly, measurement of the proviral copy numbers revealed that the CTL escape CA mutations RK and RKLM severely impaired the ability of the viral DNA to integrate into the host chromatin. Importantly, the compensatory SA mutation was able to significantly rescue the viral DNA integration defect exhibited by the RK and RKLM mutant viruses. Collectively, these results provide strong evidence that the RK mutation-associated reduction in infectivity is not due to alterations in reverse transcription and nuclear import but exclusively a consequence of impaired viral DNA integration.

In infected cells, integration of the viral DNA into the host chromosome is carried out by the HIV-1 PIC. Therefore, we hypothesized that a diminished integration activity of the RK PICs could be the underlying cause of impaired viral DNA integration of the mutant virus. Our results showed that the viral DNA levels in the isolated WT and RK mutant cytoplasmic PICs were comparable, thereby further corroborating that the RK mutation has no significant effect on viral reverse transcription levels. Unexpectedly, the integration activity of the WT and RK mutant cytoplasmic PICs did not vary significantly, in striking contrast to a significant reduction in the number of RK mutant proviral DNAs. There are key differences in our assays measuring *in vitro* integration and the proviral integration. These include contrasting nature of the target DNAs (naked DNA vs chromosomal DNA), distinct composition of the cytoplasmic PICs and nuclear PICs (e.g., reported reduction in PIC-associated integrase and CA levels upon nuclear entry), and spatial and temporal constraints in the involvement of cellular cofactors (e.g., nuclear CPSF6 and LEDGF). Although these differences may play a role, it should be noted that our measurements of reverse transcription are quantitative but not qualitative. Furthermore, emerging evidence suggest that HIV-1 reverse transcription may not be completed until the PIC enters the nucleus. Because proviral integration is dependent on precise viral DNA ends, we investigated whether the quality of the viral DNAs in the nucleus are incompetent for integration.

The sequence information of the junctions of 2-LTR circles, which are typically generated in the nucleus of HIV-1-infected cells by the end-to-end ligation of unintegrated viral DNA molecules, can be used to assess the integration competency. A key step leading up to HIV-1 integration is the 3'-end processing of viral DNAs containing precisely-defined termini. The generation of viral DNA with canonical termini depends on completion of strand transfer reactions and the proper removal of the tRNA and the PPT primers used for the minus strand and plus strand DNA synthesis, respectively. Any disruption of the above events results in viral DNA termini that are incompetent for integration and, consequently, are circularized by host machinery into 2-LTR circles. Accordingly, the 2-LTR circle junction sequences are typified by the presence of unprocessed viral DNA ends, deletions within the 3'U5 or 5'U3 region, and insertion (i.e., retention) of aberrant viral DNA mapping mostly to PPT and PBS sequences. Indeed, we observed all these modifications in our comparative analysis of the cloned junction sequences of 2-LTR circles from cells inoculated with the WT or RK mutant virus. However, compared to the WT, the insertions in RK 2-LTR circle junctions presented two distinctive patterns. One, a higher proportion of these insertions were significantly longer. Two, a significantly higher proportion of these insertions mapped to the viral PPT and/or the PBS sequences. This atypical and enhanced retention of the PPT and PBS sequences at the RK mutant viral DNA termini could be a major factor contributing to its impaired integration. Remarkably, compared to the RK mutant viral DNAs, the proportion of the RKLMSA mutant viral DNAs atypically appended with aberrant PPT and/or PBS sequences were significantly low and were even lesser when compared to the WT sequences. The atypical retention of the PPT and PBS sequences at viral DNA termini has previously been proposed to arise from non-optimal RNase H activity of the viral reverse transcriptase or from autointegration. It is unclear how the RK mutation that is positioned in the CA and did not cause any significant alterations in the reverse transcription levels can selectively alter the RNase H activity of the reverse transcriptase. Further, autointegration is unlikely to be the source of these viral DNA insertions for three reasons. One, primers used in the PCR-based

method for obtaining the 2-LTR circle junction sequences are complementary only to 2-LTR circle-specific sequences. Two, our use of 24-hpi sample as the source of template for PCR excludes the detection of autointegrants because they accumulate at early time points (~10 hpi)¹³⁹ and are rapidly depleted by 24 hpi. Three, requirement of viral DNA 3'-end processing for the integrase-mediated autointegration is incompatible with our sequencing data showing that majority of the viral DNA termini are unprocessed or partially processed. Nevertheless, these results suggest that the atypical retention of the PPT and PBS sequences at the ends of the RK mutant viral DNAs may significantly contribute to their drastically impaired integration into chromosomal DNA and suggest an underlying qualitative glitch in reverse transcription.

The rescue of the infectivity defect of the RK mutant virus in CypA-depleted cells and in cells treated with cyclosporin A indicates that an interaction between the RK mutant CA and CypA is a prerequisite for the manifestation of the virus infectivity defect^{61,67}. Accordingly, results from our virus infection assays carried out in CypA-depleted cells indicated that the infectivity of the RK mutant virus was comparable to that of the WT virus. Is CypA playing a causative role in the infectivity defect incurred by the RK mutant virus? Interestingly, supportive evidence comes from our data comparing the viral DNA integration levels of the RK mutant virus with that of the WT virus in the CypA knockout Jurkat cells. In line with the infectivity assays, the viral DNA integration level of the RK mutant virus was comparable to that of the WT virus in the CypA knockout Jurkat cells. Especially striking is the several fold increase in the integration levels of the RK mutant virus in the CypA knockout Jurkat cells relative to the control Jurkat cells. This indicates that CypA expression and its binding to the mutant CA is a prerequisite for the significant reduction in viral DNA integration by the RK mutant virus. Does this also imply that the reduction in virus infectivity by RK mutation is exclusively a consequence of the CypA-associated impaired viral DNA integration? In line with the findings from the control Jurkat cells, the reverse transcription and nuclear import levels of the RK mutant virus was comparable to that of the WT virus in the CypA

knockout Jurkat cells. This indicates that the CypA knockout-mediated restoration of the RK mutant virus infectivity is primarily an outcome of the rescue of the RK mutation-induced integration defect. It should be noted that the prior studies did not assess the effect of these CA mutations on HIV-1 integration^{61,67} and, instead, the log-fold reduction in infectivity of the RK and RKLM mutant viruses in single-cycle infectivity assays were attributed to the reduction in the levels of reverse transcription⁶¹.

This work also raises several important questions, including: (1) How does the RK mutation lead to the generation of imperfect viral DNA ends? (2) Why is CypA expression a requirement for the RK mutation-associated integration defect? and (3) How does the compensatory CA mutation SA restore the RK mutant virus infectivity without disrupting the interaction between the mutant CA and host CypA? The RK mutation appears to neither disrupt the interaction between CA and host CypA nor alter the quantitative aspects of reverse transcription. However, it is possible that the interaction between the RK mutant CA and the host CypA leads to an atypical and prolonged association between these proteins. This, in turn, by altering qualitative aspects of reverse transcription (e.g., altered RNase H activity) may lead to the generation of an increased proportion of viral DNA containing integration-incompetent termini. Conversely, the compensatory CA mutation SA or absence of CypA may preclude such an atypical and prolonged association between the RK mutant CA and CypA. However, the ability of the compensatory SA mutation to restore the RK mutant virus integration without imposing any loss of interaction between the mutant CA and host CypA, suggest the possible involvement of additional cellular co-factor(s). Accordingly, CPSF6 expression was recently reported as a requirement for the manifestation of the RK mutation-associated virus infectivity defect⁶⁶. Although there is no evidence linking the interaction between CA and CPSF6 to reverse transcription, a recent study reported that CypA binding to HIV-1 capsid prevents oligomerized cytoplasmic CPSF6 from prematurely binding and disrupting the capsid^{140,141}. An alternative possibility, from our in vitro integration assays showing that the cytoplasmic PICs are competent

for viral DNA integration, is that the RK mutation-associated aberrant viral DNA termini are generated only in the nucleus. For instance, an atypical association between RK mutant CA and CypA may delay viral integration during which the viral DNA termini may be subjected to host DNA polymerase-mediated copying and insertion reactions, thereby resulting in the generation of aberrant viral DNA termini in the nucleus.

It is important to note that ART only keeps HIV-1 replication in check and does not cure or eradicate the lifelong infection. One approach called “functional cure” entails controlling the virus to undetectable levels without ART and avert disease progression to AIDS^{142,143}. Remarkably, the elite controllers possess all these characteristics^{22,144,145} and are considered a suitable model for achieving functional cure^{24,146-148}. The elite controller phenotype has been linked to robust killing of the HIV-1-infected cells by CTLs¹⁴⁹ and has been shown to be enriched in individuals with certain HLA genotypes, including HLA-B27. Our work, showing impaired HIV-1 integration as the driver of the fitness cost of the clinically relevant CTL escape RK mutation (associated with HLA-B27), provides a novel mechanistic basis to advance CTL-based HIV cure strategies^{37,52,150-153} and to rationally design superior epitopes for T cell-based HIV vaccines.

MATERIALS AND METHODS

Cell culture

The human embryonic kidney cell line HEK293T and the T-cell line SupT1 were obtained from the American Type Culture Collection (Manassas, VA, USA). The following cell lines were obtained through the NIH AIDS Reagent Program, Division of AIDS, NIAID, NIH (USA): Jurkat (E6-1) and Jurkat CypA deficient (Jurkat CypA^{-/-}) T-cell lines from Douglas Braaten and Jeremy Luban, and the TZM-bl reporter cell line from John C. Kappes, Xiaoyun Wu, and Tranzyme, Inc. The HEK293T and TZM-bl cells were cultured in Dulbecco modified Eagle medium (DMEM; Gibco/Thermo Scientific, USA) supplemented with 10% heat-inactivated fetal bovine serum (FBS; Gibco/Thermo Scientific, USA), 2 mM glutamine, 100 U/mL penicillin, and 100 µg/mL

streptomycin. The SupT1, Jurkat, Jurkat CypA^{-/-} cell lines were cultured in Roswell Park Memorial Institute (RPMI) 1640 medium supplemented with 10% heat-inactivated FBS, 2 mM glutamine, 100 U/mL penicillin, and 100 µg/mL streptomycin. Cells were cultured at 37°C with 5% CO₂.

Proviral plasmids and cloning

The viruses used in this study were generated from the full-length wildtype HIV-1 molecular clone pNL43 and its mutant derivatives. Unless indicated otherwise, all PCRs were performed using the high-fidelity Phusion DNA polymerase (NEB, USA) and custom-made primers. Agarose gel-resolved DNAs were extracted using the Zymoclean Gel DNA Recovery Kit (Zymo Research, USA); plasmid DNA and PCR amplicons were digested with commercial restriction enzymes (NEB, USA); DNAs were ligated using T4 DNA ligase (NEB, USA); competent DH5α (Invitrogen, USA) or NEB-Stbl (NEB, USA) bacterial cells were used for bacterial transformation; plasmid DNAs were isolated using the Zyppy plasmid mini prep kit or ZymoPURE II plasmid maxiprep kit (Zymo Research, USA); recombinant plasmids were verified by restriction enzyme digestion and Sanger DNA sequencing; and site-specific mutations were introduced into DNA using the Q5 Site-Directed Mutagenesis Kit (NEB, USA) and custom-made primers. Site-specific mutations in the capsid (CA) coding region of pNL43 were first introduced into a 1310-bp DNA fragment—spanning the matrix (MA), capsid (CA), and part of the nucleocapsid (NC) coding regions and flanked by the BssHII and Apal restriction sites— of pNL43 cloned into the pMiniT 2.0 vector (NEB, USA). This pMiniT2-Gag/BssHII-Apal intermediary plasmid was constructed by cloning a PCR amplicon obtained by using custom-made forward primer harboring BssHII site (5'-TATAGCGCGCACGGCAAGAG-3') and reverse primer harboring Apal site (5'-TATAGGGCCCTGCAATTTTGGCTATG-3') of the corresponding DNA sequence from pNL43 into the linearized pMiniT 2.0 vector (NEB, USA) as per manufacturer-recommended protocol. The R264K/G791A mutation was introduced using a forward primer containing the desired mutation (G>A: 5'-ATCTATAAAAAATGGATAATCCTG-3') and a reverse primer (5'-

TTCTCCTACTGGGATAGG-3') to yield the pMiniT2-Gag-RK plasmid. The L268M/C802A mutation was introduced using a forward primer containing the desired mutation (C>A: 5'-ATGGATAATCATGGGATTAAATAAAATAG-3') and a reverse primer (5'-CTTTTATAGATTTCTCCTACTG-3') to derive the pMiniT2-Gag-LM plasmid. The R264K/G791A mutation was introduced into the pMiniT2-Gag-LM plasmid using a forward primer containing the desired mutations (G>A: 5'-AGGAGAAATCTATAAAAAATGGATAATCATGG-3') and a reverse primer (5'-ACTGGGATAGGTGGATTATGTG-3') to yield the pMiniT2-Gag-RKLM plasmid. The S173A/T517G mutation was introduced into the pMiniT2-Gag-RKLM plasmid using a forward primer containing the desired mutations (G>A: 5'-ACCCATGTTTGCAGCATTATC-3') and a reverse primer (5'-ATTACTTCTGGGCTGAAAG-3') to yield the pMiniT2-Gag-RKLMSA plasmid. The BssHII-ApaI DNA fragments from the pMiniT2-Gag-RK, pMiniT2-Gag-LM, pMiniT2-Gag-RKLM, and pMiniT2-Gag-RKLMSA plasmids were released by digestion with the BssHII and ApaI enzymes, and the gel-purified DNA fragments were then individually ligated to the BssHII+ApaI-cut pNL43 plasmid to yield pNL43-RK, pNL43-LM, pNL43-RKLM, and pNL43-RKLMSA plasmids, respectively.

GAL4-based yeast two-hybrid assay

The GAL4-based yeast two-hybrid (Y2H) system parental plasmids, pGBKT7 and pGADT7 (Clontech, USA), were used to generate the Y2H plasmids encoding the viral or host proteins. The DNA sequence encoding the CA ORF was amplified from the pNL43 plasmid by PCR using a forward primer harboring EcoR1 site (5'-TAAGCAGAATTCCTATAGTGCAGAACCTCCAGG-3') and a reverse primer harboring Sal1 site (5'-TCATTAGTCGACTATCACAAACTCTTGCTTTATGG-3') and GoTaq DNA polymerase (Promega, USA). The PCR amplicon was gel-purified using the Qiaquick gel purification kit (Qiagen, USA), digested with EcoR1 and Sal1 enzymes, and then ligated to the EcoR1+Sal1-cut pGBKT7 plasmid or EcoR1+Xho1-cut pGADT7 plasmid to yield the pGBKT7-CA and pGADT7-

CA plasmids, respectively. The DNA sequence encoding the mCherry ORF was amplified from mCherry-H2A-10 (Addgene plasmid# 55054) by PCR using a forward primer harboring EcoR1 site (5'-TGACAGAATTCATGGTGAGCAAGG-3') and a reverse primer harboring the Xho1 site (5'-TGACACTCGAGTTACTTGTACAGC-3'). The gel-purified PCR amplicon was digested with EcoR1 and Xho1 enzymes and then ligated to the EcoR1+Sal1-cut pGBKT7 plasmid or EcoR1+Xho1-cut pGADT7 plasmid to yield the pGBKT7-mCherry and pGADT7-mCherry, respectively. The DNA sequence encoding the CypA ORF was released from the pET21-CypA plasmid by digestion with Nde1 and Xho1 enzymes, and the gel-purified Nde1+Xho1-cut CypA ORF was ligated to Nde1+Xho1-cut pGADT7 plasmid to yield the pGADT7-CypA. Site-specific mutations in the DNA sequence encoding the CA ORF in the pGBKT7-CA and pGADT7-CA plasmids were introduced using custom-made primers. The R264K/G791A mutation was introduced using a forward primer containing the desired mutation (G>A: 5'-ATCTATAAAAAATGGATAATCCTG-3') and a reverse primer (5'-TTCTCCTACTGGGATAGG-3') to yield the pGBKT7-CA-RK and pGADT7-CA-RK plasmids, respectively. The DNA sequence encoding the CPSF6 ORF was amplified from the pTarget-CPSF6 by PCR using a forward primer harboring EcoR1 site (5'-TATAGAATTCATGGCGGACGGCGTGG-3') and a reverse primer harboring the Xho1 site (5'-TATACTCGAGCTAACGATGACGATATTCGCGCTC-3'). The gel-purified PCR amplicon was digested with EcoR1 and Xho1 enzymes and then ligated to the EcoR1+Xho1-cut pGADT7 plasmid.

Evaluation of protein interactions by the GAL4-based Y2H system was performed by following the manufacturer's recommendations for the Matchmaker GAL4 Two-hybrid System 3 (Clontech, USA). Competent AH109 yeast cells (Clontech, USA) were prepared and transformed with the Y2H plasmid constructs by using the Frozen-EZ Yeast Transformation II kit (Zymo Research, USA) and following the manufacturer-supplied protocol and recommendations. Briefly, competent AH109 cells were co-transformed with appropriate combinations of bait and prey Y2H plasmids and selected for growth on minimal synthetic dropout (SD) agar plates lacking

tryptophan (-T) and leucine (-L) [SD/-T/-L] at 30°C. The ability of the bait and prey proteins to interact was assayed by testing the yeast transformant's ability to grow on SD agar plates lacking tryptophan, leucine, and histidine (SD/-T/-L/-H) at 30°C, which is indicative of a positive protein interaction-driven expression of the reporter gene *HIS3*. Positive protein-protein interactions were subjected to another round of stringent selection by testing the yeast transformant's ability to express α -galactosidase from the *MEL1* reporter gene (blue/white screening) when selected on SD/-T/-L/-H agar plates supplemented with X- α -Gal (Sigma, USA) at 30°C.

Molecular Modelling Studies

Molecular modelling of the CA hexamer structure: The initial coordinates of the HIV-1 CA hexamer were generated by applying a six-fold symmetry operation onto a native full-length HIV-1 capsid protein (PDB accession code 4XFX). The two loops between residues 5 to 9 and residues 222 to 231, missing in the original structure, were built by Modeler¹⁵⁴. Three models of CA hexamer mutations, R132K (RK), R132KL136M (RKLM), and R132KL136MS41A (RKLMSA), were generated in VMD¹⁵⁵ Mutator Plugin from the 4XFX hexamer model. After placed an IP6 molecule¹⁰⁶ close to the Arg18 ring, sodium and chloride ions were introduced around these models, based on the local electrostatic potential, using CIONIZE plugin in VMD. Subsequently, all models were then solvated with CHARMM TIP3P water model¹⁵⁶ and the total NaCl concentrations were set to 150 mM, resulting charge-neutral systems of about 65 K atoms.

Molecular modelling of the CA pentamer structures: Details about the molecular modeling of two CA pentamer structures were described in Xu, et al. (2020)¹⁰⁷. Briefly, the coordinates of the disulfide stabilized pentamer were from crystal structure 3P05, after mutating the cysteine residues back to the HIV-1NL4-3 wildtype sequence. The coordinates of the missing residues in the PDB model were built by Modeler¹⁵⁴. Using the mutated 3P05 pentamer structure for the initial

coordinates, the 5MCY pentamer model was generated by running molecular dynamics flexible fitting (MDFF) of the initial coordinates into 5MCY cryo-EM density (EMBD: EMD-3466).

Simulation protocol: The solvated systems were then subjected to minimization in two stages, both using the conjugated gradient algorithm¹⁵⁷ with line search¹⁵⁸. Each stage consisted of 10,000 steps of energy minimization. During the first stage, only water molecules and ions were free to move, while the protein and IP6 molecule were fixed. In the second stage, the backbone atoms of the CA protein were applied with a harmonic restraint with a force constant of 10.0 kcal mol⁻¹ Å⁻². Convergence of the minimization procedure was confirmed once the variance of the gradient was below 0.1 kcal mol⁻¹ Å⁻¹. Following minimization, the systems were tempered from 50 K to 310 K in increments of 20 K over 1 ns. Subsequently, the systems were equilibrated at 310 K for 100,000 steps, while the protein backbone atoms were restrained. Then positional restraints were gradually released at a rate of 1.0 Kcal mol⁻¹ Å⁻² per 400 ps from 10.0 Kcal mol⁻¹ Å⁻² to 0.0 Kcal mol⁻¹ Å⁻².

All MD simulations in present study were performed with NAMD2.13¹⁵⁹ using CHARMM force fields^{160,161}. In present study, an internal time step of 2 fs was employed in the multi-step vRESPA integrator as implemented in NAMD, bonded interactions were evaluated every 2 fs. Temperature was held constant at 310 K using a Langevin thermostat with a coupling constant of 0.1 ps⁻¹. Pressure was controlled at 1 bar using a Nose-Hoover Langevin piston barostat with period and decay of 40 ps and 10 ps, respectively. The Shake algorithm was employed to constraint vibrations of all hydrogen atoms. Long range electrostatics was calculated using the particle-mesh-Ewald summation with a grid size of 1 Å and a cutoff for short-range electrostatics interactions of 12 Å. MD result analysis: The RMSD, RMSF, contact and ion occupancy analyses were performed in VMD¹⁵⁵. The APBS software¹⁶² was employed to compute the electrostatic potential surfaces. The sodium and chloride ions were added to all systems according to the local coulombic potential by CIONIZE in VMD¹⁵⁵.

Cell- and virus-associated viral protein expression in cells transfected with proviral plasmids.

HEK293T cells, seeded at 6×10^5 cells per well in a 6-well plate and cultured overnight, were transfected with 2 μ g of WT and mutant proviral plasmid DNAs using PEI. After 24 h or 48 h post transfection, the culture media containing the released virus particles (virus fraction) was collected, and the cells were lysed in TX-100 lysis buffer (300mM NaCl; 50mM Tris-HCl, pH 7.5; 0.5% Triton X-100; β -mercaptoethanol, Sigma protease inhibitor cocktail) for 10 min on ice. The virus fraction was centrifuged at 500x *g* for 3 min to remove cells/debris, passed through 0.45- μ m filter, and centrifuged in a SW 32.1 Ti rotor (Beckman Coulter) at 32,000 rpm for 45 min at 4°C. The supernatant was carefully removed, and the virus pellet was lysed in 0.1 mL of TX-100 lysis buffer for 10 min on ice. The total protein concentrations of the cell lysates were quantified by BCA protein assay (Pierce, USA). Equal amounts of the cell lysate protein and equal volume of the virus pellet protein were resolved by SDS-PAGE (NuPAGE 4-12% Bis-Tris gel; Invitrogen/Thermo Scientific) and transferred onto nitrocellulose membrane by semidry transfer. Membranes blocked in non-fat dry milk were first probed with mouse anti-CA monoclonal antibody (1:1000 dilution; 183-H12-5C; NIH AIDS Reagent Program, USA), followed by secondary HRP-conjugated goat anti-mouse IgG(H+L) (1:10,000 dilution; Bio-Rad, USA), and then developed using the enhanced chemiluminescence procedure (Pierce, USA). Virus release efficiency is calculated as the amount of virus-associated CA divided by the total Gag (virus-associated CA + cell-associated Pr55Gag + cell-associated p41Gag + cell-associated CA).

Preparation of virus stocks and determination of titer

Virus stocks were generated by PEI-mediated transient transfection of HEK293T cells with WT or mutant HIV-1 proviral plasmid constructs. Briefly, 2×10^6 HEK293T cells, seeded per 10-cm culture dish and cultured overnight, were transfected with 10 or 20 μ g of plasmid DNA. After 12-16 h, the cell culture media was removed, the cells washed once with 1X phosphate-buffered

saline (PBS), 6 mL of new growth medium was added to the cells, and the cells were cultured for additional 48 h. Then, the culture media containing the released virus particles was harvested, centrifuged at low-speed to remove the cell debris, passaged through 0.45-µm-pore-size syringe filters, and treated with DNase I (Calbiochem/MilliporeSigma, USA; 20 µg/mL of supernatant) in the presence of 10 mM magnesium chloride for 1 h at 37°C to eliminate any carryover plasmid DNA from transfection.

The titer of the virus preparations was determined by using the qPCR lentivirus titration kit (Applied Biological Materials, USA) as per the manufacturer-recommended protocol. Briefly, the viral RNA from the virus sample was extracted by mixing 2 µL of virus preparation with 18 µL of virus lysis buffer and incubating at room temperature for 3 min. The resulting virus lysate, and two lentiviral standards- STD1 and STD2, were used as templates, along with the reagent-mix, lentivirus 2X qPCR mastermix, and 1X iTaq Universal SYBR Green Supermix (Bio-Rad, USA), in qRT-PCRs carried out under following cycling conditions: initial incubations at 42°C for 20 min (reverse transcription) and 95°C for 10 min (enzyme activation), followed by 30 cycles of amplification and acquisition at 95°C for 15 s and 60°C for 1 min. The resulting Ct values were then used to calculate the viral titer of the virus preparations using the formula $5 \times 10^{7/2e3(Ctx - Ct1)/(Ct2 - Ct1)}$, where Ctx = average of 3 Ct values of the unknown sample, Ct1 = average of 3 Ct values of STD1, and Ct2 = average of 3 Ct values of STD2.

TZM-bl-based HIV-1 infectivity assay

TZM-bl cells, seeded at 4×10^4 cells per well in 24-well plate and cultured overnight, were inoculated with WT or mutant viruses (MOI of 0.05 or 1.0) in the presence of polybrene for 2 h at 37°C/5% CO₂. After sequential washes with 1X PBS and cell culture medium, the cells were replenished with new complete growth medium and cultured further for 48 h. The cells in each well were washed with 1X PBS, lysed in 200 µL of 1X GLO lysis buffer (Promega, USA), and

incubated with shaking at room temperature for 5 min. Luminescence activity in the resulting cell lysates was measured using the Luciferase Assay System (Promega, USA).

Assessment of HIV-1 infection in T-cells

Jurkat cells (2.5×10^5 cells per well) in 48-well plates were spinoculated with virus stocks (MOI of 1.0) at $480 \times g$ for 2 h at 25°C. After the spin, the spinoculated samples were individually transferred onto 1.7 mL microcentrifuge tubes, centrifuged at $300 \times g$ for 10 min at 25°C, and the supernatants were carefully removed and discarded. Each cell pellet was then thoroughly resuspended in 1 mL of RPMI complete medium, transferred onto new 48-well cell culture plate, and cultured for 24 h or 48 h at 37°C. The cells were then harvested by transferring the cell cultures to 1.7-mL microcentrifuge tubes and centrifuging at $500 \times g$ for 5 min at 25°C. The cell pellets were washed once with PBS, resuspended in 300 µL of BD Cytofix/Cytoperm (BD Biosciences, USA), and incubated for 20 min at 4°C. The cells were then centrifuged at $500 \times g$ for 5 min at 25°C and the supernatants were discarded. The cell pellets were washed once with 300 µL of BD Perm/Wash buffer and then resuspended in 300 µL of BD Perm/Wash buffer (BD Biosciences, USA) for storage at 4°C. The cell samples were centrifuged and the cell pellets were resuspended in 110 µL of BD Perm/Wash buffer. For staining Gag protein, 50 µL of each cell sample was mixed with 1 µL of KC57-FITC antibody that identifies the HIV-1 Gag and p24 (Beckman Coulter Life Sciences, USA) and incubated at 4°C for 20 min. The cell samples were centrifuged and the cell pellets were washed twice with 300 µL of BD Perm/Wash buffer. The cells were then resuspended in 100 µL of BD Perm/Wash buffer and the percentage of FITC-labelled cells, as a marker for HIV-1 infectivity, was determined using fluorescence-activated cell sorter (FACS) analysis on Guava easyCyte HT (Millipore Sigma, USA).

HIV-1 infection of target cells for quantitative assays

Jurkat or SupT1 cells (1×10^6 cells or 2×10^6 cells per well) in 24-well plates were spinoculated with virus stocks (MOI of 0.05 or 0.1) at $480 \times g$ for 2 h at 25°C . After the spin, the spinoculated samples were individually transferred onto 1.7 mL microcentrifuge tubes, centrifuged at $300 \times g$ for 10 min at 25°C , and the supernatants were carefully removed and discarded. Each cell pellet was then thoroughly resuspended in 1 mL of RPMI complete medium, transferred onto new 24-well cell culture plate, and cultured for 24 h at 37°C . The cells were then harvested by transferring the cell cultures to 1.7-mL microcentrifuge tubes and centrifuging at $1,500 \times g$ for 5 min at 25°C . The cell pellets were washed once with PBS and then resuspended in 150 μL of PBS or nuclease-free water, followed by isolation of total DNA using the Quick-DNA miniprep kit (Zymo Research, USA) according to the manufacturer's instructions.

Quantitation of reverse transcription products and 2-LTR circles by qPCR

A SYBR green-based qPCR was used to quantify the reverse transcription products, and a TaqMan probe-based qPCR was used to quantify the 2-LTR circles. The qPCRs were assembled in triplicates using 100 ng of sample DNA, $1 \times$ iTaq Universal SYBR Green Supermix or $1 \times$ iTaq Universal Probe Supermix (Bio-Rad, USA), 300 nM each of forward primer (late RT product: MH531/5'-TGTGTGCCCCGTCTGTTGTGT-3'; 2-LTR circle/5'-AACTAGGGAACCCACTGCTTAAG-3') and reverse primer (late RT product: MH532/5'-GAGTCCTGCGTCGAGAGAGC-3'; 2-LTR circle/5'-TCCACAGATCAAGGATATCTTGTC-3'), and, in the case of probe-based qPCR, 100 nM TaqMan probe (5'-[FAM] ACACTACTTGAAGCACTCAAGGCAAGCTTT-[TAMRA]-3'). The qPCRs were carried out under the following cycling conditions: late RT products- 95°C for 3 min, followed by 39 cycles of amplification and acquisition at 94°C for 15 s, 58°C for 30 s, and 72°C for 30 s; 2-LTR circles- 95°C for 3 min, followed by 39 cycles of amplification and acquisition at 94°C for 15 s, 58°C for 30 s, and 72°C for 30 s. The thermal profile for melting-curve analysis of SYBR green-based

qPCR was obtained by holding the sample at 65°C for 31 s, followed by a linear ramp in temperature from 65 to 95°C with a ramp rate of 0.5°C/s and acquisition at 0.5°C intervals. The CFX Manager software (Bio-Rad, USA) was used to determine the copy number of the target DNA in the qPCR samples by plotting the data against the respective standard curve that was generated in parallel using 10-fold serial dilutions of known copy numbers (10^0 to 10^8) of pNL43 (for late RT products) or the p2LTR plasmid containing the 2-LTR junction sequence (for 2-LTR circles).

Quantitation of proviral DNAs by nested PCR

To quantify the copy number of (chromosomally integrated) proviral DNA in HIV-1-infected cells, a nested PCR strategy wherein primers designed to amplify only the junctions of the chromosomal-integrated viral DNA but not any unintegrated viral DNA is used in the first-round end-point PCR, followed by the second round qPCR with primers that amplify viral LTR-specific sequences present in the first round PCR amplicons. The first-round PCR was assembled using 100 ng of sample DNA, 1× GoTaq reaction buffer (Promega, USA), deoxynucleoside triphosphate (dNTP) nucleotide mix containing 200 µM of each nucleotide (Promega, USA), 500 nM each of primers complementary to the target cell chromosomal *Alu* repeat sequence (5'-GCCTCCCAAAGTGCTGGGATTACAG-3') and HIV-1 Gag sequence (5'-GTTCTGCTATGTCACTTCC-3'), and 1.25 U of GoTaq DNA polymerase (Promega, USA), and the PCR cycling conditions comprised of an initial incubation at 95°C for 5 min, followed by 23 cycles of amplification at 94°C for 30 s, 50°C for 30 s, 72°C for 4 min; and a final extension at 72°C for 10 min. One-tenth volume of the first-round PCR products was used as the template DNA in the second-round qPCR, that also included 1× iTaq Universal Probe Supermix (Bio-Rad, USA), 300 nM each of the viral LTR-specific primers complementary to the LTR-R (5'-TCTGGCTAACTAGGGAACCCA-3') and the LTR-U5 (5'-CTGACTAAAAGGGTCTGAGG-3'), and 100 nM of TaqMan probe (5'-[6-FAM]-TTAAGCCTCAATAAAGCTTGCCTTGAGTGC-[TAMRA]-

3'), and the qPCR was carried out as follows: 95°C for 3 min, 39 cycles of amplification and acquisition at 94°C for 15 s, 58°C for 30 s, and 72°C for 30 s. The CFX Manager software (Bio-Rad, USA) was used to determine the copy number of the target DNA in the qPCR samples by plotting the data against the standard curve that was generated in parallel using 10-fold serial dilutions of known copy numbers (10^0 to 10^8) of pNL43.

Isolation of HIV-1 cytoplasmic PICs and determination of PIC-associated integration activity *in vitro*.

The HIV-1 PICs were isolated using a published protocol⁷⁴ with modifications. Briefly, target cells (10×10^6), distributed as 2×10^6 cells per well in 12-well plates, were spinoculated with virus stocks (MOI of 0.05) at $480 \times g$ for 2 h at 25°C. The spinoculated samples were individually transferred onto 1.7 mL microcentrifuge tubes, centrifuged at $300 \times g$ for 10 min at 25°C, and the supernatants were carefully removed and discarded. Each cell pellet was then thoroughly resuspended in 1 mL of RPMI complete medium, the spinoculated samples corresponding to each virus infection was pooled and distributed as 1 mL aliquots into a new 12-well plate and cultured for 5 h at 37°C. Unless indicated otherwise, all centrifugation steps were carried out at room temperature (~25°C). The cell samples corresponding to each virus infection was pooled, pelleted by centrifugation at $300 \times g$ for 10 min, washed twice with 1 mL of buffer $K^{-/-}$ (20 mM HEPES pH 7.6, 150 mM potassium chloride, 5 mM magnesium chloride), thoroughly resuspended in 0.5 mL of ice-cold buffer $K^{+/+}$ (20 mM HEPES pH 7.6, 150 mM potassium chloride, 5 mM magnesium chloride, 1 mM dithiothreitol [DTT], 20 µg/mL aprotinin, 0.025% [wt/vol] digitonin), incubated on a rocking platform shaker (60 to 80 rocking motions/min) for 10 min at room temperature, centrifuged at $1,500 \times g$ for 4 min at 4°C, and the supernatant was collected and centrifuged at $16,000 \times g$ for 1 min at 4°C. The resulting supernatant (cytoplasmic PIC) was treated with RNase A (Invitrogen/Thermo Scientific, USA) at a final concentration of 20 µg/mL for 30 min at room temperature, mixed with 60% sucrose (wt/vol) to a final concentration of 7%, and

aliquots were flash-frozen in liquid nitrogen and placed in a -80°C freezer for long-term storage. *In vitro* integration assays of cytoplasmic PICs were performed as described previously⁷⁴. Briefly, the *in vitro* integration reaction was carried out by mixing purified ΦX174 target DNA amplicon with 50 μL or 100 μL of PICs and then incubating the mixture at 37°C for 45 min. The integration reaction was stopped and deproteinized by adding SDS, EDTA, and proteinase K to final concentrations of 0.5%, 8 mM, and 0.5 mg/ml, respectively, and incubating overnight at 56°C . The DNA was purified from the sample by phenol chloroform extraction followed by ethanol precipitation and was used as the template DNA in the first round of the nested PCR. The first-round endpoint PCR, designed to amplify only the integrated viral DNA-target DNA junctions, was carried out in a final volume of 50 μL containing one-tenth volume of the purified DNA from the *in vitro* integration reaction, 500 nM each of primers targeting the viral LTR (5'-GTGCGCGCTTCAGCAAG-3') and the target DNA (5'-CACTGACCCTCAGCAATCTTA-3'), 1X GoTaq reaction buffer (Promega, USA), dNTP nucleotide mix containing 200 μM concentrations of each nucleotide (Promega, USA), and 1.25U of GoTaq DNA polymerase (Promega, USA), and under the following thermocycling conditions: initial denaturation at 95°C for 5 min; followed by 23 cycles at 94°C for 30 s, 55°C for 30 s, and 72°C for 4 min; and a final extension at 72°C for 10 min. The second-round qPCR, designed to amplify only the viral LTR-specific region, was assembled using one-tenth volume of the first-round PCR products as template DNA, 1X iTaq Universal Probe Supermix (Bio-Rad, USA), 300 nM each of the viral LTR-specific primers complementary to the LTR-R (5'-TCTGGCTAACTAGGGAACCCA-3') and the LTR-U5 (5'-CTGACTAAAAGGGTCTGAGG-3'), and 100 nM of TaqMan probe (5'-[6-FAM]-TTAAGCCTCAATAAAGCTTGCCCTTGAGTGC-[TAMRA]-3'), and the qPCR was carried out as follows: 95°C for 3 min, 39 cycles of amplification and acquisition at 94°C for 15 s, 58°C for 30 s, and 72°C for 30 s. The CFX Manager software (Bio-Rad, USA) was used to determine the copy number of the target DNA in the qPCR samples by plotting the data against the standard curve that was generated in parallel using 10-fold serial dilutions of known copy numbers (10^0 to 10^8) of

pNL43 plasmid. To determine the integration efficiency (i.e., ratio of integrated viral DNA copy numbers to corresponding PIC-associated viral DNA copy numbers), the PIC-associated viral DNA copy numbers were determined. Briefly, 50µL or 100µL of PICs was deproteinized, extracted with phenol-chloroform, and subjected to ethanol precipitation. One-tenth volume of the resulting purified DNA was directly used in qPCR designed to quantify the reverse transcription products. The CFX Manager software (Bio-Rad, USA) was used to determine the copy number of the target DNA in the qPCR samples by plotting the data against the standard curve that was generated in parallel.

Sequence analysis of junctions of 2-LTR circles from HIV-1-infected target cells

A nested PCR strategy was used to amplify the junction of the 2-LTR circles in HIV-1-infected cells. The first-round end-point PCR was assembled using 500 ng of total DNA from HIV-1-infected cells, 5× Phusion HF buffer (NEB, USA), a dNTP nucleotide mix containing 200 µM of each nucleotide (Promega), 500 nM each of the primers 2-LTR Forward-M (5'-AGCCTGGGAGCTCTCTGGCTAAC-3') and 2-LTR Reverse-M (5'-AGCCTTGTGTGTGGTAGATCCAC-3'), and 1 U of Phusion Hot Start Flex polymerase (NEB, USA), and the PCR was carried out using cycling conditions of 98°C for 2 min, 30 cycles of amplification at 98°C for 10 s, 71°C for 1 min, and 72°C for 15 s, and a final extension at 72°C for 10 min. The second-round end-point PCR was assembled using 1/10th volume of the first-round PCR products as template DNA, along with 5× Colorless GoTaq reaction buffer, 500 nM each of the primers MH535 (5'-AACTAGGGAACCCACTGCTTAAG-3') and MH536 (5'-TCCACAGATCAAGGATATCTTGTC-3'), and 1.25 U of GoTaq DNA polymerase, and the PCR was carried out under the following cycling conditions: 95°C for 5 min, 25 cycles of amplification at 95°C for 15 s, 58°C for 1 min, and 72°C for 30 s, and a final extension at 72°C for 10 min. The resulting PCR amplicon from the second-round PCR was gel-purified and ligated to the pGEMT-Easy vector (Promega, USA), as per the manufacturer-recommended protocol (Promega, USA).

Diagnostic restriction digestion was used to verify the presence of 2-LTR junction DNA insert in the recombinant plasmid DNA isolated from transformed bacterial colonies selected on X-Gal (5-bromo-4-chloro-3-indolyl- β -D-galactopyranoside)/IPTG (isopropyl- β -D-thiogalactopyranoside) indicator plates. The sequence of the 2-LTR junction DNA inserts was determined by Sanger DNA sequencing using primers flanking the cloning site, and multiple alignment was used to analyze the sequence data.

ACKNOWLEDGEMENTS:

This work was supported by the National Institutes of Health grants R01 AI136740, R01 DA 042348, R56 AI122960, R24 DA036420, and U54 MD007586 to C.D. and a Research Centers in Minority Institutions (RCMI) grant U54MD007586 to J.P. This work is also supported in part by the Meharry Translational Research Center (MeTRC) grant U54MD007593 and Tennessee CFAR grant P30 AI110527 from the National Institutes of Health to C.D.

REFERENCES

- 1 Arts, E. J. & Hazuda, D. J. HIV-1 antiretroviral drug therapy. *Cold Spring Harb Perspect Med* **2**, a007161, doi:10.1101/cshperspect.a007161 (2012).
- 2 Autran, B. *et al.* Positive effects of combined antiretroviral therapy on CD4+ T cell homeostasis and function in advanced HIV disease. *Science* **277**, 112-116, doi:10.1126/science.277.5322.112 (1997).
- 3 Lederman, M. M. *et al.* Immunologic responses associated with 12 weeks of combination antiretroviral therapy consisting of zidovudine, lamivudine, and zalcitabine: results of AIDS Clinical Trials Group Protocol 315. *J Infect Dis* **178**, 70-79, doi:10.1086/515591 (1998).
- 4 Autran, B. *et al.* Restoration of the immune system with anti-retroviral therapy. *Immunol Lett* **66**, 207-211, doi:10.1016/s0165-2478(98)00159-x (1999).
- 5 Walensky, R. P. *et al.* The survival benefits of AIDS treatment in the United States. *J Infect Dis* **194**, 11-19, doi:10.1086/505147 (2006).
- 6 Saag, M. S. *et al.* Antiretroviral Drugs for Treatment and Prevention of HIV Infection in Adults: 2020 Recommendations of the International Antiviral Society-USA Panel. *JAMA* **324**, 1651-1669, doi:10.1001/jama.2020.17025 (2020).
- 7 Finzi, D. *et al.* Latent infection of CD4+ T cells provides a mechanism for lifelong persistence of HIV-1, even in patients on effective combination therapy. *Nat Med* **5**, 512-517, doi:10.1038/8394 (1999).
- 8 Davey, R. T., Jr. *et al.* HIV-1 and T cell dynamics after interruption of highly active antiretroviral therapy (HAART) in patients with a history of sustained viral suppression. *Proc Natl Acad Sci U S A* **96**, 15109-15114, doi:10.1073/pnas.96.26.15109 (1999).
- 9 Blankson, J. N., Persaud, D. & Siliciano, R. F. The challenge of viral reservoirs in HIV-1 infection. *Annu Rev Med* **53**, 557-593, doi:10.1146/annurev.med.53.082901.104024 (2002).

- 10 Gupta, R. K. *et al.* HIV-1 drug resistance before initiation or re-initiation of first-line antiretroviral therapy in low-income and middle-income countries: a systematic review and meta-regression analysis. *Lancet Infect Dis* **18**, 346-355, doi:10.1016/S1473-3099(17)30702-8 (2018).
- 11 Siedner, M. J. *et al.* Reduced efficacy of HIV-1 integrase inhibitors in patients with drug resistance mutations in reverse transcriptase. *Nat Commun* **11**, 5922, doi:10.1038/s41467-020-19801-x (2020).
- 12 Bertagnolio, S. *et al.* Clinical Impact of Pretreatment Human Immunodeficiency Virus Drug Resistance in People Initiating Nonnucleoside Reverse Transcriptase Inhibitor-Containing Antiretroviral Therapy: A Systematic Review and Meta-analysis. *J Infect Dis* **224**, 377-388, doi:10.1093/infdis/jiaa683 (2021).
- 13 Churchill, M. J., Deeks, S. G., Margolis, D. M., Siliciano, R. F. & Swanstrom, R. HIV reservoirs: what, where and how to target them. *Nat Rev Microbiol* **14**, 55-60, doi:10.1038/nrmicro.2015.5 (2016).
- 14 Sengupta, S. & Siliciano, R. F. Targeting the Latent Reservoir for HIV-1. *Immunity* **48**, 872-895, doi:10.1016/j.immuni.2018.04.030 (2018).
- 15 Mylvaganam, G., Yanez, A. G., Maus, M. & Walker, B. D. Toward T Cell-Mediated Control or Elimination of HIV Reservoirs: Lessons From Cancer Immunology. *Front Immunol* **10**, 2109, doi:10.3389/fimmu.2019.02109 (2019).
- 16 Rogan, D. C. & Connors, M. Immunologic Control of HIV-1: What Have We Learned and Can We Induce It? *Curr HIV/AIDS Rep* **18**, 211-220, doi:10.1007/s11904-021-00545-2 (2021).
- 17 Altfeld, M. *et al.* HLA Alleles Associated with Delayed Progression to AIDS Contribute Strongly to the Initial CD8(+) T Cell Response against HIV-1. *PLoS Med* **3**, e403, doi:10.1371/journal.pmed.0030403 (2006).

- 18 Fellay, J. *et al.* A whole-genome association study of major determinants for host control of HIV-1. *Science* **317**, 944-947, doi:10.1126/science.1143767 (2007).
- 19 Saez-Cirion, A. & Pancino, G. HIV controllers: a genetically determined or inducible phenotype? *Immunol Rev* **254**, 281-294, doi:10.1111/imr.12076 (2013).
- 20 Balasubramaniam, M., Pandhare, J. & Dash, C. Immune Control of HIV. *J Life Sci (Westlake Village)* **1**, 4-37 (2019).
- 21 McLaren, P. J. & Fellay, J. HIV-1 and human genetic variation. *Nat Rev Genet*, doi:10.1038/s41576-021-00378-0 (2021).
- 22 Deeks, S. G. & Walker, B. D. Human immunodeficiency virus controllers: mechanisms of durable virus control in the absence of antiretroviral therapy. *Immunity* **27**, 406-416, doi:10.1016/j.immuni.2007.08.010 (2007).
- 23 Pereyra, F. *et al.* Genetic and immunologic heterogeneity among persons who control HIV infection in the absence of therapy. *J Infect Dis* **197**, 563-571, doi:10.1086/526786 (2008).
- 24 Jiang, C. *et al.* Distinct viral reservoirs in individuals with spontaneous control of HIV-1. *Nature* **585**, 261-267, doi:10.1038/s41586-020-2651-8 (2020).
- 25 Carrington, M. & O'Brien, S. J. The influence of HLA genotype on AIDS. *Annu Rev Med* **54**, 535-551, doi:10.1146/annurev.med.54.101601.152346 (2003).
- 26 International, H. I. V. C. S. *et al.* The major genetic determinants of HIV-1 control affect HLA class I peptide presentation. *Science* **330**, 1551-1557, doi:10.1126/science.1195271 (2010).
- 27 Carrington, M. & Walker, B. D. Immunogenetics of spontaneous control of HIV. *Annu Rev Med* **63**, 131-145, doi:10.1146/annurev-med-062909-130018 (2012).
- 28 McLaren, P. J. & Carrington, M. The impact of host genetic variation on infection with HIV-1. *Nat Immunol* **16**, 577-583, doi:10.1038/ni.3147 (2015).

- 29 Borrow, P. *et al.* Antiviral pressure exerted by HIV-1-specific cytotoxic T lymphocytes (CTLs) during primary infection demonstrated by rapid selection of CTL escape virus. *Nat Med* **3**, 205-211, doi:10.1038/nm0297-205 (1997).
- 30 Koup, R. A. *et al.* Temporal association of cellular immune responses with the initial control of viremia in primary human immunodeficiency virus type 1 syndrome. *J Virol* **68**, 4650-4655, doi:10.1128/JVI.68.7.4650-4655.1994 (1994).
- 31 Price, D. A. *et al.* Positive selection of HIV-1 cytotoxic T lymphocyte escape variants during primary infection. *Proc Natl Acad Sci U S A* **94**, 1890-1895, doi:10.1073/pnas.94.5.1890 (1997).
- 32 Ferre, A. L. *et al.* Mucosal immune responses to HIV-1 in elite controllers: a potential correlate of immune control. *Blood* **113**, 3978-3989, doi:10.1182/blood-2008-10-182709 (2009).
- 33 Walker, B. & McMichael, A. The T-cell response to HIV. *Cold Spring Harb Perspect Med* **2**, doi:10.1101/cshperspect.a007054 (2012).
- 34 Ndhlovu, Z. M. *et al.* Magnitude and Kinetics of CD8+ T Cell Activation during Hyperacute HIV Infection Impact Viral Set Point. *Immunity* **43**, 591-604, doi:10.1016/j.immuni.2015.08.012 (2015).
- 35 Pollack, R. A. *et al.* Defective HIV-1 Proviruses Are Expressed and Can Be Recognized by Cytotoxic T Lymphocytes, which Shape the Proviral Landscape. *Cell Host Microbe* **21**, 494-506 e494, doi:10.1016/j.chom.2017.03.008 (2017).
- 36 Halle, S., Halle, O. & Forster, R. Mechanisms and Dynamics of T Cell-Mediated Cytotoxicity In Vivo. *Trends Immunol* **38**, 432-443, doi:10.1016/j.it.2017.04.002 (2017).
- 37 Warren, J. A., Clutton, G. & Goonetilleke, N. Harnessing CD8(+) T Cells Under HIV Antiretroviral Therapy. *Front Immunol* **10**, 291, doi:10.3389/fimmu.2019.00291 (2019).

- 38 Kaslow, R. A. *et al.* Influence of combinations of human major histocompatibility complex genes on the course of HIV-1 infection. *Nat Med* **2**, 405-411, doi:10.1038/nm0496-405 (1996).
- 39 Goulder, P. J. *et al.* Novel, cross-restricted, conserved, and immunodominant cytotoxic T lymphocyte epitopes in slow progressors in HIV type 1 infection. *AIDS Res Hum Retroviruses* **12**, 1691-1698, doi:10.1089/aid.1996.12.1691 (1996).
- 40 Migueles, S. A. *et al.* HLA B*5701 is highly associated with restriction of virus replication in a subgroup of HIV-infected long term nonprogressors. *Proc Natl Acad Sci U S A* **97**, 2709-2714, doi:10.1073/pnas.050567397 (2000).
- 41 Moore, C. B. *et al.* Evidence of HIV-1 adaptation to HLA-restricted immune responses at a population level. *Science* **296**, 1439-1443, doi:10.1126/science.1069660 (2002).
- 42 Altfeld, M. *et al.* Influence of HLA-B57 on clinical presentation and viral control during acute HIV-1 infection. *AIDS* **17**, 2581-2591, doi:10.1097/00002030-200312050-00005 (2003).
- 43 Wang, Y. E. *et al.* Protective HLA class I alleles that restrict acute-phase CD8+ T-cell responses are associated with viral escape mutations located in highly conserved regions of human immunodeficiency virus type 1. *J Virol* **83**, 1845-1855, doi:10.1128/JVI.01061-08 (2009).
- 44 Fellay, J. *et al.* Common genetic variation and the control of HIV-1 in humans. *PLoS Genet* **5**, e1000791, doi:10.1371/journal.pgen.1000791 (2009).
- 45 Payne, R. P. *et al.* Efficacious early antiviral activity of HIV Gag- and Pol-specific HLA-B 2705-restricted CD8+ T cells. *J Virol* **84**, 10543-10557, doi:10.1128/JVI.00793-10 (2010).
- 46 Goulder, P. J. *et al.* Evolution and transmission of stable CTL escape mutations in HIV infection. *Nature* **412**, 334-338, doi:10.1038/35085576 (2001).

- 47 Cao, J., McNevin, J., Malhotra, U. & McElrath, M. J. Evolution of CD8+ T cell immunity and viral escape following acute HIV-1 infection. *J Immunol* **171**, 3837-3846, doi:10.4049/jimmunol.171.7.3837 (2003).
- 48 Allen, T. M. *et al.* Selective escape from CD8+ T-cell responses represents a major driving force of human immunodeficiency virus type 1 (HIV-1) sequence diversity and reveals constraints on HIV-1 evolution. *J Virol* **79**, 13239-13249, doi:10.1128/JVI.79.21.13239-13249.2005 (2005).
- 49 Leslie, A. *et al.* Transmission and accumulation of CTL escape variants drive negative associations between HIV polymorphisms and HLA. *J Exp Med* **201**, 891-902, doi:10.1084/jem.20041455 (2005).
- 50 Kawashima, Y. *et al.* Adaptation of HIV-1 to human leukocyte antigen class I. *Nature* **458**, 641-645, doi:10.1038/nature07746 (2009).
- 51 Schneidewind, A. *et al.* Transmission and long-term stability of compensated CD8 escape mutations. *J Virol* **83**, 3993-3997, doi:10.1128/JVI.01108-08 (2009).
- 52 Deng, K. *et al.* Broad CTL response is required to clear latent HIV-1 due to dominance of escape mutations. *Nature* **517**, 381-385, doi:10.1038/nature14053 (2015).
- 53 Carlson, J. M., Le, A. Q., Shahid, A. & Brumme, Z. L. HIV-1 adaptation to HLA: a window into virus-host immune interactions. *Trends Microbiol* **23**, 212-224, doi:10.1016/j.tim.2014.12.008 (2015).
- 54 Nixon, D. F. *et al.* HIV-1 gag-specific cytotoxic T lymphocytes defined with recombinant vaccinia virus and synthetic peptides. *Nature* **336**, 484-487, doi:10.1038/336484a0 (1988).
- 55 Goulder, P. J. *et al.* Late escape from an immunodominant cytotoxic T-lymphocyte response associated with progression to AIDS. *Nat Med* **3**, 212-217, doi:10.1038/nm0297-212 (1997).

- 56 Streeck, H. *et al.* Recognition of a defined region within p24 gag by CD8+ T cells during primary human immunodeficiency virus type 1 infection in individuals expressing protective HLA class I alleles. *J Virol* **81**, 7725-7731, doi:10.1128/JVI.00708-07 (2007).
- 57 Ammaranond, P. *et al.* HIV immune escape at an immunodominant epitope in HLA-B*27-positive individuals predicts viral load outcome. *J Immunol* **186**, 479-488, doi:10.4049/jimmunol.0903227 (2011).
- 58 Kelleher, A. D. *et al.* Clustered mutations in HIV-1 gag are consistently required for escape from HLA-B27-restricted cytotoxic T lymphocyte responses. *J Exp Med* **193**, 375-386, doi:10.1084/jem.193.3.375 (2001).
- 59 Feeney, M. E. *et al.* Immune escape precedes breakthrough human immunodeficiency virus type 1 viremia and broadening of the cytotoxic T-lymphocyte response in an HLA-B27-positive long-term-nonprogressing child. *J Virol* **78**, 8927-8930, doi:10.1128/JVI.78.16.8927-8930.2004 (2004).
- 60 Almeida, J. R. *et al.* Superior control of HIV-1 replication by CD8+ T cells is reflected by their avidity, polyfunctionality, and clonal turnover. *J Exp Med* **204**, 2473-2485, doi:10.1084/jem.20070784 (2007).
- 61 Schneidewind, A. *et al.* Escape from the dominant HLA-B27-restricted cytotoxic T-lymphocyte response in Gag is associated with a dramatic reduction in human immunodeficiency virus type 1 replication. *J Virol* **81**, 12382-12393, doi:10.1128/JVI.01543-07 (2007).
- 62 Brumme, Z. L. *et al.* Human leukocyte antigen-specific polymorphisms in HIV-1 Gag and their association with viral load in chronic untreated infection. *AIDS* **22**, 1277-1286, doi:10.1097/QAD.0b013e3283021a8c (2008).
- 63 Streeck, H. *et al.* Immune-driven recombination and loss of control after HIV superinfection. *J Exp Med* **205**, 1789-1796, doi:10.1084/jem.20080281 (2008).

- 64 Ammaranond, P. *et al.* A new variant cytotoxic T lymphocyte escape mutation in HLA-B27-positive individuals infected with HIV type 1. *AIDS Res Hum Retroviruses* **21**, 395-397, doi:10.1089/aid.2005.21.395 (2005).
- 65 Schneidewind, A. *et al.* Structural and functional constraints limit options for cytotoxic T-lymphocyte escape in the immunodominant HLA-B27-restricted epitope in human immunodeficiency virus type 1 capsid. *J Virol* **82**, 5594-5605, doi:10.1128/JVI.02356-07 (2008).
- 66 Henning, M. S., Dubose, B. N., Burse, M. J., Aiken, C. & Yamashita, M. In vivo functions of CPSF6 for HIV-1 as revealed by HIV-1 capsid evolution in HLA-B27-positive subjects. *PLoS Pathog* **10**, e1003868, doi:10.1371/journal.ppat.1003868 (2014).
- 67 Schommers, P. *et al.* Changes in HIV-1 Capsid Stability Induced by Common Cytotoxic-T-Lymphocyte-Driven Viral Sequence Mutations. *J Virol* **90**, 7579-7586, doi:10.1128/JVI.00867-16 (2016).
- 68 von Schwedler, U. K., Stray, K. M., Garrus, J. E. & Sundquist, W. I. Functional surfaces of the human immunodeficiency virus type 1 capsid protein. *J Virol* **77**, 5439-5450, doi:10.1128/jvi.77.9.5439-5450.2003 (2003).
- 69 Rihn, S. J. *et al.* Extreme genetic fragility of the HIV-1 capsid. *PLoS Pathog* **9**, e1003461, doi:10.1371/journal.ppat.1003461 (2013).
- 70 Ambrose, Z. & Aiken, C. HIV-1 uncoating: connection to nuclear entry and regulation by host proteins. *Virology* **454-455**, 371-379, doi:10.1016/j.virol.2014.02.004 (2014).
- 71 Campbell, E. M. & Hope, T. J. HIV-1 capsid: the multifaceted key player in HIV-1 infection. *Nat Rev Microbiol* **13**, 471-483, doi:10.1038/nrmicro3503 (2015).
- 72 Yamashita, M. & Engelman, A. N. Capsid-Dependent Host Factors in HIV-1 Infection. *Trends Microbiol* **25**, 741-755, doi:10.1016/j.tim.2017.04.004 (2017).

- 73 Francis, A. C. & Melikyan, G. B. Single HIV-1 Imaging Reveals Progression of Infection through CA-Dependent Steps of Docking at the Nuclear Pore, Uncoating, and Nuclear Transport. *Cell Host Microbe* **23**, 536-548 e536, doi:10.1016/j.chom.2018.03.009 (2018).
- 74 Balasubramaniam, M. *et al.* PF74 Inhibits HIV-1 Integration by Altering the Composition of the Preintegration Complex. *J Virol* **93**, doi:10.1128/JVI.01741-18 (2019).
- 75 AlBurtamani, N., Paul, A. & Fassati, A. The Role of Capsid in the Early Steps of HIV-1 Infection: New Insights into the Core of the Matter. *Viruses* **13**, doi:10.3390/v13061161 (2021).
- 76 Scoca, V. & Di Nunzio, F. The HIV-1 Capsid: From Structural Component to Key Factor for Host Nuclear Invasion. *Viruses* **13**, doi:10.3390/v13020273 (2021).
- 77 Toccafondi, E., Lener, D. & Negroni, M. HIV-1 Capsid Core: A Bullet to the Heart of the Target Cell. *Front Microbiol* **12**, 652486, doi:10.3389/fmicb.2021.652486 (2021).
- 78 Engelman, A. N. HIV Capsid and Integration Targeting. *Viruses* **13**, doi:10.3390/v13010125 (2021).
- 79 Briggs, J. A. & Krausslich, H. G. The molecular architecture of HIV. *J Mol Biol* **410**, 491-500, doi:10.1016/j.jmb.2011.04.021 (2011).
- 80 Zhao, G. *et al.* Mature HIV-1 capsid structure by cryo-electron microscopy and all-atom molecular dynamics. *Nature* **497**, 643-646, doi:10.1038/nature12162 (2013).
- 81 Mattei, S., Glass, B., Hagen, W. J., Krausslich, H. G. & Briggs, J. A. The structure and flexibility of conical HIV-1 capsids determined within intact virions. *Science* **354**, 1434-1437, doi:10.1126/science.aah4972 (2016).
- 82 Perilla, J. R. & Gronenborn, A. M. Molecular Architecture of the Retroviral Capsid. *Trends Biochem Sci* **41**, 410-420, doi:10.1016/j.tibs.2016.02.009 (2016).
- 83 James, L. C. & Jacques, D. A. The Human Immunodeficiency Virus Capsid Is More Than Just a Genome Package. *Annu Rev Virol* **5**, 209-225, doi:10.1146/annurev-virology-092917-043430 (2018).

- 84 Pornillos, O. & Ganser-Pornillos, B. K. Maturation of retroviruses. *Curr Opin Virol* **36**, 47-55, doi:10.1016/j.coviro.2019.05.004 (2019).
- 85 Christensen, D. E., Ganser-Pornillos, B. K., Johnson, J. S., Pornillos, O. & Sundquist, W. I. Reconstitution and visualization of HIV-1 capsid-dependent replication and integration in vitro. *Science* **370**, doi:10.1126/science.abc8420 (2020).
- 86 Luban, J., Bossolt, K. L., Franke, E. K., Kalpana, G. V. & Goff, S. P. Human immunodeficiency virus type 1 Gag protein binds to cyclophilins A and B. *Cell* **73**, 1067-1078, doi:10.1016/0092-8674(93)90637-6 (1993).
- 87 Gamble, T. R. *et al.* Crystal structure of human cyclophilin A bound to the amino-terminal domain of HIV-1 capsid. *Cell* **87**, 1285-1294, doi:10.1016/s0092-8674(00)81823-1 (1996).
- 88 Ni, T. *et al.* Intrinsic curvature of the HIV-1 CA hexamer underlies capsid topology and interaction with cyclophilin A. *Nat Struct Mol Biol* **27**, 855-862, doi:10.1038/s41594-020-0467-8 (2020).
- 89 Lee, K. *et al.* Flexible use of nuclear import pathways by HIV-1. *Cell Host Microbe* **7**, 221-233, doi:10.1016/j.chom.2010.02.007 (2010).
- 90 Price, A. J. *et al.* CPSF6 defines a conserved capsid interface that modulates HIV-1 replication. *PLoS Pathog* **8**, e1002896, doi:10.1371/journal.ppat.1002896 (2012).
- 91 Sokolskaja, E., Sayah, D. M. & Luban, J. Target cell cyclophilin A modulates human immunodeficiency virus type 1 infectivity. *J Virol* **78**, 12800-12808, doi:10.1128/JVI.78.23.12800-12808.2004 (2004).
- 92 Qi, M., Yang, R. & Aiken, C. Cyclophilin A-dependent restriction of human immunodeficiency virus type 1 capsid mutants for infection of nondividing cells. *J Virol* **82**, 12001-12008, doi:10.1128/JVI.01518-08 (2008).
- 93 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).

- 94 De Iaco, A. & Luban, J. Cyclophilin A promotes HIV-1 reverse transcription but its effect on transduction correlates best with its effect on nuclear entry of viral cDNA. *Retrovirology* **11**, 11, doi:10.1186/1742-4690-11-11 (2014).
- 95 Sowd, G. A. *et al.* A critical role for alternative polyadenylation factor CPSF6 in targeting HIV-1 integration to transcriptionally active chromatin. *Proc Natl Acad Sci U S A* **113**, E1054-1063, doi:10.1073/pnas.1524213113 (2016).
- 96 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).
- 97 Kim, K. *et al.* Cyclophilin A protects HIV-1 from restriction by human TRIM5alpha. *Nat Microbiol* **4**, 2044-2051, doi:10.1038/s41564-019-0592-5 (2019).
- 98 Selyutina, A. *et al.* Cyclophilin A Prevents HIV-1 Restriction in Lymphocytes by Blocking Human TRIM5alpha Binding to the Viral Core. *Cell Rep* **30**, 3766-3777 e3766, doi:10.1016/j.celrep.2020.02.100 (2020).
- 99 Ganser, B. K., Li, S., Klishko, V. Y., Finch, J. T. & Sundquist, W. I. Assembly and analysis of conical models for the HIV-1 core. *Science* **283**, 80-83, doi:10.1126/science.283.5398.80 (1999).
- 100 Briggs, J. A. *et al.* The stoichiometry of Gag protein in HIV-1. *Nat Struct Mol Biol* **11**, 672-675, doi:10.1038/nsmb785 (2004).
- 101 Ganser-Pornillos, B. K., Yeager, M. & Pornillos, O. Assembly and architecture of HIV. *Adv Exp Med Biol* **726**, 441-465, doi:10.1007/978-1-4614-0980-9_20 (2012).
- 102 Forshey, B. M., von Schwedler, U., Sundquist, W. I. & Aiken, C. Formation of a human immunodeficiency virus type 1 core of optimal stability is crucial for viral replication. *J Virol* **76**, 5667-5677, doi:10.1128/jvi.76.11.5667-5677.2002 (2002).

- 103 Price, A. J. *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).
- 104 Liu, C. *et al.* Cyclophilin A stabilizes the HIV-1 capsid through a novel non-canonical binding site. *Nat Commun* **7**, 10714, doi:10.1038/ncomms10714 (2016).
- 105 Matreyek, K. A., Yucel, S. S., 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).
- 106 Dick, R. A. *et al.* Inositol phosphates are assembly co-factors for HIV-1. *Nature* **560**, 509-512, doi:10.1038/s41586-018-0396-4 (2018).
- 107 Xu, C. *et al.* Permeability of the HIV-1 capsid to metabolites modulates viral DNA synthesis. *PLoS Biol* **18**, e3001015, doi:10.1371/journal.pbio.3001015 (2020).
- 108 Huang, P. T. *et al.* FEZ1 Is Recruited to a Conserved Cofactor Site on Capsid to Promote HIV-1 Trafficking. *Cell Rep* **28**, 2373-2385 e2377, doi:10.1016/j.celrep.2019.07.079 (2019).
- 109 Smaga, S. S. *et al.* MxB Restricts HIV-1 by Targeting the Tri-hexamer Interface of the Viral Capsid. *Structure* **27**, 1234-1245 e1235, doi:10.1016/j.str.2019.04.015 (2019).
- 110 Fields, S. & Sternglanz, R. The two-hybrid system: an assay for protein-protein interactions. *Trends Genet* **10**, 286-292, doi:10.1016/0168-9525(90)90012-u (1994).
- 111 Fields, S. & Song, O. A novel genetic system to detect protein-protein interactions. *Nature* **340**, 245-246, doi:10.1038/340245a0 (1989).
- 112 Chien, C. T., Bartel, P. L., Sternglanz, R. & Fields, S. The two-hybrid system: a method to identify and clone genes for proteins that interact with a protein of interest. *Proc Natl Acad Sci U S A* **88**, 9578-9582, doi:10.1073/pnas.88.21.9578 (1991).

- 113 Lichterfeld, M. *et al.* A viral CTL escape mutation leading to immunoglobulin-like transcript 4-mediated functional inhibition of myelomonocytic cells. *J Exp Med* **204**, 2813-2824, doi:10.1084/jem.20061865 (2007).
- 114 Wei, X. *et al.* Emergence of resistant human immunodeficiency virus type 1 in patients receiving fusion inhibitor (T-20) monotherapy. *Antimicrob Agents Chemother* **46**, 1896-1905, doi:10.1128/AAC.46.6.1896-1905.2002 (2002).
- 115 Sundquist, W. I. & Krausslich, H. G. HIV-1 assembly, budding, and maturation. *Cold Spring Harb Perspect Med* **2**, a006924, doi:10.1101/cshperspect.a006924 (2012).
- 116 Freed, E. O. HIV-1 assembly, release and maturation. *Nat Rev Microbiol* **13**, 484-496, doi:10.1038/nrmicro3490 (2015).
- 117 Kleinpeter, A. B. & Freed, E. O. HIV-1 Maturation: Lessons Learned from Inhibitors. *Viruses* **12**, doi:10.3390/v12090940 (2020).
- 118 O'Doherty, U., Swiggard, W. J. & Malim, M. H. Human immunodeficiency virus type 1 spinoculation enhances infection through virus binding. *J Virol* **74**, 10074-10080, doi:10.1128/jvi.74.21.10074-10080.2000 (2000).
- 119 Sloan, R. D. & Wainberg, M. A. The role of unintegrated DNA in HIV infection. *Retrovirology* **8**, 52, doi:10.1186/1742-4690-8-52 (2011).
- 120 Li, L. *et al.* Role of the non-homologous DNA end joining pathway in the early steps of retroviral infection. *EMBO J* **20**, 3272-3281, doi:10.1093/emboj/20.12.3272 (2001).
- 121 Butler, S. L., Hansen, M. S. & Bushman, F. D. A quantitative assay for HIV DNA integration in vivo. *Nat Med* **7**, 631-634, doi:10.1038/87979 (2001).
- 122 Bukrinsky, M. I. *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).
- 123 Lander, E. S. *et al.* Initial sequencing and analysis of the human genome. *Nature* **409**, 860-921, doi:10.1038/35057062 (2001).

- 124 Deininger, P. Alu elements: know the SINEs. *Genome Biol* **12**, 236, doi:10.1186/gb-2011-12-12-236 (2011).
- 125 Stevens, S. W. & Griffith, J. D. Sequence analysis of the human DNA flanking sites of human immunodeficiency virus type 1 integration. *J Virol* **70**, 6459-6462, doi:10.1128/JVI.70.9.6459-6462.1996 (1996).
- 126 Schroder, A. R. *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).
- 127 Cohn, L. B. *et al.* HIV-1 integration landscape during latent and active infection. *Cell* **160**, 420-432, doi:10.1016/j.cell.2015.01.020 (2015).
- 128 O'Doherty, U., Swiggard, W. J., Jeyakumar, D., McGain, D. & Malim, M. H. A sensitive, quantitative assay for human immunodeficiency virus type 1 integration. *J Virol* **76**, 10942-10950, doi:10.1128/jvi.76.21.10942-10950.2002 (2002).
- 129 Liszewski, M. K., Yu, J. J. & O'Doherty, U. Detecting HIV-1 integration by repetitive-sampling Alu-gag PCR. *Methods* **47**, 254-260, doi:10.1016/j.ymeth.2009.01.002 (2009).
- 130 Craigie, R. & Bushman, F. D. HIV DNA integration. *Cold Spring Harb Perspect Med* **2**, a006890, doi:10.1101/cshperspect.a006890 (2012).
- 131 Engelman, A. Isolation and analysis of HIV-1 preintegration complexes. *Methods Mol Biol* **485**, 135-149, doi:10.1007/978-1-59745-170-3_10 (2009).
- 132 Balasubramaniam, M., Davids, B., Addai, A. B., Pandhare, J. & Dash, C. Measurement of In Vitro Integration Activity of HIV-1 Preintegration Complexes. *J Vis Exp*, doi:10.3791/54581 (2017).
- 133 Braaten, D. & Luban, J. Cyclophilin A regulates HIV-1 infectivity, as demonstrated by gene targeting in human T cells. *EMBO J* **20**, 1300-1309, doi:10.1093/emboj/20.6.1300 (2001).
- 134 Mandal, D. & Prasad, V. R. Analysis of 2-LTR circle junctions of viral DNA in infected cells. *Methods Mol Biol* **485**, 73-85, doi:10.1007/978-1-59745-170-3_6 (2009).

- 135 Julias, J. G., McWilliams, M. J., Sarafianos, S. G., Arnold, E. & Hughes, S. H. Mutations in the RNase H domain of HIV-1 reverse transcriptase affect the initiation of DNA synthesis and the specificity of RNase H cleavage in vivo. *Proc Natl Acad Sci U S A* **99**, 9515-9520, doi:10.1073/pnas.142123199 (2002).
- 136 Julias, J. G., Boyer, P. L., McWilliams, M. J., Alvord, W. G. & Hughes, S. H. Mutations at position 184 of human immunodeficiency virus type-1 reverse transcriptase affect virus titer and viral DNA synthesis. *Virology* **322**, 13-21, doi:10.1016/j.virol.2004.01.004 (2004).
- 137 Jiang, J. *et al.* The interdomain linker region of HIV-1 capsid protein is a critical determinant of proper core assembly and stability. *Virology* **421**, 253-265, doi:10.1016/j.virol.2011.09.012 (2011).
- 138 Brun, S. *et al.* VSV-G pseudotyping rescues HIV-1 CA mutations that impair core assembly or stability. *Retrovirology* **5**, 57, doi:10.1186/1742-4690-5-57 (2008).
- 139 Yan, N., Cherepanov, P., Daigle, J. E., Engelman, A. & Lieberman, J. The SET complex acts as a barrier to autointegration of HIV-1. *PLoS Pathog* **5**, e1000327, doi:10.1371/journal.ppat.1000327 (2009).
- 140 Zhong, Z. *et al.* Cytoplasmic CPSF6 Regulates HIV-1 Capsid Trafficking and Infection in a Cyclophilin A-Dependent Manner. *mBio* **12**, doi:10.1128/mBio.03142-20 (2021).
- 141 Ning, J. *et al.* Truncated CPSF6 Forms Higher-Order Complexes That Bind and Disrupt HIV-1 Capsid. *J Virol* **92**, doi:10.1128/JVI.00368-18 (2018).
- 142 Walker, B. D. & Yu, X. G. Unravelling the mechanisms of durable control of HIV-1. *Nat Rev Immunol* **13**, 487-498, doi:10.1038/nri3478 (2013).
- 143 Davenport, M. P. *et al.* Functional cure of HIV: the scale of the challenge. *Nat Rev Immunol* **19**, 45-54, doi:10.1038/s41577-018-0085-4 (2019).
- 144 O'Connell, K. A., Bailey, J. R. & Blankson, J. N. Elucidating the elite: mechanisms of control in HIV-1 infection. *Trends Pharmacol Sci* **30**, 631-637, doi:10.1016/j.tips.2009.09.005 (2009).

- 145 Naranbhai, V. & Carrington, M. Host genetic variation and HIV disease: from mapping to mechanism. *Immunogenetics* **69**, 489-498, doi:10.1007/s00251-017-1000-z (2017).
- 146 Walker, B. D. Elite control of HIV Infection: implications for vaccines and treatment. *Top HIV Med* **15**, 134-136 (2007).
- 147 Tarancon-Diez, L., Dominguez-Molina, B., Viciana, P., Lopez-Cortes, L. & Ruiz-Mateos, E. Long-term Persistent Elite HIV-controllers: The Right Model of Functional Cure. *EBioMedicine* **28**, 15-16, doi:10.1016/j.ebiom.2018.01.013 (2018).
- 148 Li, J. Z. & Blankson, J. N. How elite controllers and posttreatment controllers inform our search for an HIV-1 cure. *J Clin Invest* **131**, doi:10.1172/JCI149414 (2021).
- 149 Saez-Cirion, A. *et al.* HIV controllers exhibit potent CD8 T cell capacity to suppress HIV infection ex vivo and peculiar cytotoxic T lymphocyte activation phenotype. *Proc Natl Acad Sci U S A* **104**, 6776-6781, doi:10.1073/pnas.0611244104 (2007).
- 150 Shan, L. *et al.* Stimulation of HIV-1-specific cytolytic T lymphocytes facilitates elimination of latent viral reservoir after virus reactivation. *Immunity* **36**, 491-501, doi:10.1016/j.immuni.2012.01.014 (2012).
- 151 Painter, M. M. *et al.* Concanamycin A counteracts HIV-1 Nef to enhance immune clearance of infected primary cells by cytotoxic T lymphocytes. *Proc Natl Acad Sci U S A* **117**, 23835-23846, doi:10.1073/pnas.2008615117 (2020).
- 152 Kaseke, C., Tano-Menka, R., Senjobe, F. & Gaiha, G. D. The Emerging Role for CTL Epitope Specificity in HIV Cure Efforts. *J Infect Dis* **223**, 32-37, doi:10.1093/infdis/jiaa333 (2021).
- 153 Collins, D. R., Gaiha, G. D. & Walker, B. D. CD8(+) T cells in HIV control, cure and prevention. *Nat Rev Immunol* **20**, 471-482, doi:10.1038/s41577-020-0274-9 (2020).
- 154 Eswar, N. *et al.* Comparative protein structure modeling using Modeller. *Curr Protoc Bioinformatics* **Chapter 5**, Unit-5 6, doi:10.1002/0471250953.bi0506s15 (2006).

- 155 Humphrey, W., Dalke, A. & Schulten, K. VMD: visual molecular dynamics. *J Mol Graph* **14**, 33-38, 27-38, doi:10.1016/0263-7855(96)00018-5 (1996).
- 156 Jorgensen, W. L., Chandrasekhar, J., Madura, J. D., Impey, R. W. & Klein, M. L. Comparison of Simple Potential Functions for Simulating Liquid Water. *J Chem Phys* **79**, 926-935, doi:Doi 10.1063/1.445869 (1983).
- 157 Fletcher, R. & Reeves, C. M. Function Minimization by Conjugate Gradients. *Comput J* **7**, 149-&, doi:DOI 10.1093/comjnl/7.2.149 (1964).
- 158 Sun, W. & Yuan, Y.-X. *Optimization theory and methods: nonlinear programming*. Vol. 1 (Springer Science & Business Media, 2006).
- 159 Phillips, J. C. *et al.* Scalable molecular dynamics with NAMD. *J Comput Chem* **26**, 1781-1802, doi:10.1002/jcc.20289 (2005).
- 160 Huang, J. *et al.* CHARMM36m: an improved force field for folded and intrinsically disordered proteins. *Nat Methods* **14**, 71-73, doi:10.1038/nmeth.4067 (2017).
- 161 Best, R. B. *et al.* Optimization of the additive CHARMM all-atom protein force field targeting improved sampling of the backbone ϕ , ψ and side-chain $\chi(1)$ and $\chi(2)$ dihedral angles. *Journal of chemical theory and computation* **8**, 3257-3273, doi:10.1021/ct300400x (2012).
- 162 Jurrus, E. *et al.* Improvements to the APBS biomolecular solvation software suite. *Protein Science* **27**, 112-128, doi:10.1002/pro.3280 (2018).

FIGURE LEGENDS

Figure 1. (A) Location of CTL-targeted epitopes in the HIV-1 capsid N-terminal domain and the CTL escape-associated amino acid changes linked to the HLA-B27-restricted KK10 epitope. The standard Gag numbering of the amino acids is followed, and the capsid numbering of the amino acids are in parentheses for reference. **(B)** Relative positions of three CA residues, R264(R132), L268(L136) and S173(S41) in the context of a WT CA monomer and hexamer. (Left) A full length native HIV-1 CA hexamer (PDBid: 4XFX) is represented by a surface model. The N-terminal domain (NTD) is colored in gold, while the C-terminal domain (CTD) is in blue. (Right) a CA monomer and the N-terminal domain from one of its neighboring monomers are highlighted in a ribbon representation. The positions of R264, L268 and S173 residues in the NTD-NTD interface are labelled, CypA binding loop (BL) is colored in red, and the CPSF6 binding pocket is depicted as a green triangle. **(C)** Relative positions of three CA residues, R264(R132), L268(L136) and S173(S41) in the context of a WT CA pentamer. (Left) Ribbon representation of a full-length native HIV-1 CA pentamer. The N-terminal domain (NTD) is colored in gold, while the C-terminal domain (CTD) is in blue. The positions of R264, L268 and S173 residues in the NTD-NTD interface are labelled and the CypA binding loop (BL) is colored in red. (Right) R264 and S173 residues on two CA pentamer models, PDB accession codes 3P05 and 5MCY.

Figure 2. (A) Structural stability of HIV-1 CA hexamer influenced by three CA residue substitutions, R264(R132), L268(L136) and S173(S41). (Left) The average C α root mean square fluctuations (RMSFs) of the HIV-1 WT, RK, RKLM, and RKLMSA hexamers derived from molecular dynamics simulations are shown. (Right) The C α root mean square deviations (RMSDs) of HIV-1 WT hexamer and 3 mutation hexamers from molecular dynamics simulations. **(B)** Structural flexibilities of the WT and three mutation hexamers, colored by the average C α

RMSF values, increasing from red to blue. **(C)** Inter-hexamer interactions of three CA residues, R264(R132), L268(L136) and S173(S41), maintain after substitutions. Top 5 inter-hexamer contacts involved in CA residue 132 (first row), 136 (middle row) and 41 (last row) are indicated. The corresponding contact occupancies are in the parenthesis after residue names. The residue names with occupancies greater than 60% are underlined. The formation of a residue contact is defined as the distance between sidechains from not neighboring residues within 3.0 Å.

Figure 3. (A) Effects of the CTL-escape mutation on structure of CA tube assembly. **(B)** Effects of the CTL-escape mutation on conformational distribution of CA monomers. Each mutation site, in each system, aligned to a single monomer, demonstrating the conformational distributions of these residues. **(C)** Effects of the CTL-escape mutation on the root-mean-square fluctuation (RMSF) of capsid residues. (Top Panel) Each monomer is shown aligned by backbone regions in helices in the top row, colored accordingly. (Bottom) RMSF of Carbon alpha atoms in each CA tube system. RMSF values are averaged for all monomers.

Figure 4. Effect of KK10-linked CTL escape mutation R264K on CA interactions. The effect of RK mutation on **(A)** CA:CA, **(B)** CA:CypA, and **(C)** CA:CPSF6 interactions were assessed by using GAL4-based yeast two-hybrid assays as described in Materials and Methods. Equal number of yeast cells co-transformed with the plasmid pairs encoding the GAL4 DNA Binding Domain (BD) in fusion with the bait protein and the GAL4 Activation Domain (AD) in fusion with the prey protein were spotted on synthetic dropout (SD) agar plates lacking leucine and tryptophan (-LT) or leucine, tryptophan, and histidine (-LTH) or leucine, tryptophan, and histidine but supplemented with X- α -D-galactoside (-LTH/+X α G) and were incubated at 30°C to assess for histidine prototrophy and α -galactosidase activity. (A-Top Panel) Growth of yeast co-expressing

BD-CA_{WT} and AD-CA_{WT}, BD-CA_{WT} and AD-mCherry (AD-mC), AD-CA_{WT} and BD-mC on selection media. (A-Bottom Panel) Growth of yeast co-expressing BD-CA_{RK} and AD-CA_{RK}, BD-CA_{RK} and AD-mC, and AD-CA_{RK} and BD-mC on selection media. **(B)** Growth of yeast co-expressing BD-CA_{WT} and AD-CypA, BD-CA_{RK} and AD-CypA, BD-CA_{WT} and AD-mC, BD-CA_{RK} and AD-mC, and BD-mC and AD-CypA on selection media. **(C)** Growth of yeast co-expressing BD-CA_{WT} and AD-CPSF6, BD-CA_{RK} and AD-CPSF6, BD-CA_{WT} and AD-mC, BD-CA_{RK} and AD-mC, and BD-mC and AD-CPSF6 on selection media.

Figure 5. Effect of KK10-linked CTL escape mutations on HIV-1 infectivity. Infectivity of WT and mutant viruses were determined using the **(A and B)** TZM-bl indicator cell line and **(C and D)** Jurkat T-cells as described in Materials and Methods. TZM-bl cells inoculated, in replicates, with the WT or mutant viruses for 2 hours were further cultured for 48 hours and then lysed. **(A)** Luciferase activity in the cell lysates, indicative of virus infectivity, was measured as luminescence and, after subtracting the luciferase activity in mock infected cells from the data, were plotted as relative light units (RLU). **(B)** Relative luciferase activity, indicative of percentage infection efficiency, was plotted relative to WT. Jurkat cells spinoculated, in replicates, with the WT or mutant viruses for 2 hours at 37°C were further cultured for 48 hours at 37°C, after which the cells were fixed and stained with FITC-labelled antibody that identifies viral Gag and p24. **(C)** The percentage of FITC-labelled cells (marker for HIV-1 infectivity) was determined using FACS analysis. **(D)** Relative Gag and p24 expression, indicative of percentage infection efficiency, was plotted relative to WT. Data shown are representative of at least three independent experiments, with error bars representing the SEMs.

Effect of KK10-linked CTL escape mutations on late stages of HIV-1 replication. HEK293T cells transfected with WT or the mutant proviral clones were cultured for 48 hours at 37°C and the viral particles in the culture media was collected by ultracentrifugation. **(E)** Cell (C) and virus (V) lysates

were resolved by SDS-PAGE and the virus proteins were probed by western blotting. **(F)** The virus release efficiency was determined as described in Materials and Methods.

Figure 6. Effect of KK10-linked CTL escape mutations on HIV-1 reverse transcription levels and nuclear import efficiency. Copy number of late RT products and 2-LTR circles present in the total DNA isolated from Jurkat cells spinoculated with the WT or mutant viruses for 2 hours at 25°C and cultured for 24 hours at 37°C were determined as described in Materials and Methods. **(A)** Copy number of late RT products of WT or the mutant viruses as measured by qPCR and calculated using a standard graph of known copy numbers of HIV-1 molecular clone (surrogate marker for RT product). **(B)** Percentage late RT products plotted relative to WT. **(C)** Copy number of 2-LTR circles as measured by qPCR and calculated using a standard graph of known copy numbers of 2-LTR-containing plasmid DNA. **(D)** Percentage copy number of the 2-LTR circles plotted relative to WT. **(E)** Calculated ratio of the copy number of 2-LTR circles to the respective late RT products plotted as percentage relative to WT. Data representative of three independent experiments, with error bars representing the SEMs.

Figure 7. Effect of KK10-linked CTL escape mutations on HIV-1 proviral integration. Copy number of chromosome-integrated viral DNAs and late RT products present in the total DNA isolated from SupT1 cells spinoculated with the WT or mutant viruses for 2 hours at 25°C and cultured for 24 hours at 37°C were determined as described in Materials and Methods. **(A)** Copy number of chromosome-integrated viral DNAs as measured by *Alu*-Gag nested PCR and calculated by using a standard graph of known copy numbers of HIV-1 molecular clone (surrogate marker for viral DNA). **(B)** Percentage copy number of chromosome-integrated viral DNA plotted relative to WT. **(C)** Copy number of late RT products of WT or the mutant viruses as measured by qPCR and

calculated using a standard graph of known copy numbers of HIV-1 molecular clone. **(D)** Percentage late RT products plotted relative to WT. **(E)** Calculated ratio of the copy number of chromosome-integrated viral DNAs to the respective late RT products plotted as percentage relative to WT. Data representative of three independent experiments, with error bars representing the SEMs.

Figure 8. Effect of KK10-linked CTL escape mutation R264K on the integration activity in vitro of cytoplasmic PICs. Integration activity in vitro and the viral DNA content of the WT and RK cytoplasmic PICs isolated from Jurkat cells spinoculated with the respective viruses for 2 hours at 25°C and cultured for 5 hours at 37°C were determined as described in Materials and Methods. **(A)** Integration activity of the cytoplasmic PICs, i.e., the copy number of viral DNAs that were integrated into the target DNA during the in vitro integration assay, as measured by qPCR and calculated using a standard graph of known copy numbers of HIV molecular clone. **(B)** Percentage integration activity plotted relative to WT. **(C)** Copy number of viral DNA content in the cytoplasmic PICs as measured by qPCR and calculated using a standard graph of known copy numbers of HIV molecular clone. **(D)** Percentage viral DNA content plotted relative to WT cytoplasmic PICs. **(E)** Ratio of the copy number of integrated viral DNAs to the corresponding PIC-associated viral DNAs plotted as percentage relative to WT. Data shown are representative of at least three independent experiments, with error bars representing the SEMs.

Figure 9. Assessment of the effect of KK10-linked CTL escape *mutation* R264K on viral DNA integration in cells lacking host cyclophilin A. Copy number of the reverse transcription (RT) products, 2-LTR circles, and the chromosome-integrated viral DNAs present in the total DNA isolated from the parental or CypA.KO Jurkat cells spinoculated with WT or the RK virus for 2

hours at 25°C and cultured for 24 hours at 37°C were determined as described in Materials and Methods. **(A)** Copy number of the late RT products in the parental or the CypA.KO cells as measured by qPCR and calculated by using a standard graph of known copy numbers of HIV-1 molecular clone. **(B)** Percentage late RT products plotted relative to WT virus in parental Jurkat cells. **(C)** Copy number of the 2-LTR circles as measured by qPCR and calculated by using a standard graph of known copy numbers of plasmid containing HIV-1 2-LTR. **(D)** Percentage 2-LTR circles plotted relative to WT virus in parental Jurkat cells. **(E)** Calculated ratio of the copy number of 2-LTR circles to the respective late RT products plotted as percentage relative to WT virus in parental Jurkat cells. **(F)** Copy number of chromosome-integrated viral DNAs as measured by *Alu*-Gag nested PCR and calculated by using a standard graph of known copy numbers of HIV-1 molecular clone. **(G)** Percentage copy number of chromosome-integrated viral DNA plotted relative to WT virus in parental Jurkat cells. **(H)** Ratio of the copy number of chromosome-integrated viral DNAs to the respective late RT products was calculated and the data plotted as percentage relative to WT virus in parental Jurkat cells. Data shown are representative of three independent experiments, with error bars representing the SEMs.

Table 1. Assessment of the effect of KK10-linked CTL escape mutations on the integrity of the termini of the reverse transcription products (viral DNAs). Junctions of the 2-LTR circles present in the total DNA isolated from Jurkat cells spinoculated with the WT or mutant viruses for 2 hours at 25°C and cultured for 24 hours at 37°C were amplified by a nested PCR strategy, cloned into a plasmid vector, and the DNA sequences determined by Sanger sequencing were aligned and analyzed as described in Materials and Methods. The total number of 2-LTR junction sequences analyzed for the WT and the mutant viruses is indicated at the top of the table. The total number of each type of 2-LTR junction sequence identified in the analyses is shown for the WT, RK, and RKLMSA virus-inoculated samples. Shown at the bottom of the table are the total number of 2-

LTR junctions containing aberrant insertion sequences and their percentage relative to the total sequences analyzed, and the total number of insertion sequences mapping to the PPT or PBS and their percentage relative to the total insertion sequences. See also Supplementary Figure S4.

Figure 10. Model for the presence of PPT and PBS sequences at HIV-1 DNA ends. Shown schematically is HIV-1 reverse transcription after the completion of the plus-strand strong stop DNA synthesis. In the case of wild type HIV-1, the RNase H removes the PPT/plus-strand primer (at the junction between PPT and plus-strand strong stop DNA) and the tRNA/minus-strand primer (one base from the junction between tRNA and minus-strand DNA). These RNase H-mediated removal of the PPT and tRNA primers, though not necessary for strand transfer, are essential for the formation of proper viral DNA ends that, after the synthesis of the complete double-stranded viral DNA, can function as competent substrates for the integration reaction. However, in the case of the RK mutant HIV-1, increased failure of RNase H to properly remove the PPT and tRNA primers can lead to the increased retention of these primer sequences at the respective viral DNA ends. These viral DNA ends are unsuitable substrates for the viral integrase-mediated integration reaction in the nucleus and therefore are ligated by host enzymes to form 2-LTR circles.

Figure 1A

Figure 1B

Figure 1C

Figure 2A

Figure 2B

Figure 2C

Figure 3A

Figure 3B

Figure 3C

Figure 4

Figure 5

Figure 5

Figure 6

Figure 7

Figure 8

Figure 9

Table 1

		WT	RK	RKLMSA
Total Number of Sequences Analyzed		104	106	166
2-LTR junctions without aberrant nucleotide insertions	Complete 3'-end processing of both LTR ends, with intact flanking regions	0	0	0
	Complete 3'-end processing of both LTR ends, accompanied by deletions in flanking regions	6	14	15
	Complete 3'-end processing of one LTR end (5'U3 or 3'U5)	1	7	14
	No 3'-end processing of both LTR ends	15	7	13
	Partial 3'-end processing of one LTR end	20	12	10
2-LTR junctions with aberrant nucleotide insertions	Complete 3'-end processing of both LTR ends, with intact flanking regions	0	1	1
	Complete 3'-end processing of both LTR ends, accompanied by deletions in flanking regions	2	1	8
	Complete 3'-end processing of one LTR end (5'U3 or 3'U5)	21	32	34
	No 3'-end processing of both LTR ends	35	26	59
	Partial 3'-end processing of one LTR end	5	6	22
Source of aberrant insertion sequences	Number of clones with insertions at the junction (% of total sequences)	63 (60.6%)	66 (62.3%)	119 (71.7%)
	Insertions mapping to PPT sequences (% of total insertion sequences)	4 (6.3%)	16 (24.2%)	7 (5.9%)
	Insertions mapping to PBS sequences (% of total insertion sequences)	8 (12.7%)	25 (37.9%)	9 (7.6%)

Figure 10