






Structural basis of host protein hijacking in human T-cell leukemia virus integration

Veer Bhatt^{1,2,9}, Ke Shi^{2,3,4,9}, Daniel J. Salamango^{2,3,4,9}, Nicholas H. Moeller^{2,3,4}, Krishan K. Pandey⁵, Sibes Bera⁵, Heather O. Bohl^{2,3,4}, Fredy Kurniawan^{2,3,4}, Kayo Orellana^{2,3,4}, Wei Zhang^{2,4,6,7}, Duane P. Grandgenett⁵, Reuben S. Harris^{2,3,4,8}, Anna C. Sundborger-Lunna^{1,2} & Hideki Aihara^{2,3,4}

Integration of the reverse-transcribed viral DNA into host chromosomes is a critical step in the life-cycle of retroviruses, including an oncogenic delta(δ)-retrovirus human T-cell leukemia virus type-1 (HTLV-1). Retroviral integrase forms a higher order nucleoprotein assembly (intasome) to catalyze the integration reaction, in which the roles of host factors remain poorly understood. Here, we use cryo-electron microscopy to visualize the HTLV-1 intasome at 3.7-Å resolution. The structure together with functional analyses reveal that the B56 γ (B' γ) subunit of an essential host enzyme, protein phosphatase 2A (PP2A), is repurposed as an integral component of the intasome to mediate HTLV-1 integration. Our studies reveal a key host-virus interaction underlying the replication of an important human pathogen and highlight divergent integration strategies of retroviruses.

¹The Hormel Institute, University of Minnesota, 801 16th Avenue N.E., Austin, MN 55912, USA. ²Masonic Cancer Center, University of Minnesota, 2231 6th Street S.E., Minneapolis, MN 55455, USA. ³Department of Biochemistry, Molecular Biology and Biophysics, University of Minnesota, 321 Church Street S.E., Minneapolis, MN 55455, USA. ⁴Institute for Molecular Virology, University of Minnesota, 515 Delaware Street S.E., Minneapolis, MN 55455, USA. ⁵Department of Molecular Microbiology and Immunology, Saint Louis University, 1100 S. Grand Boulevard, St. Louis, MO 63104, USA. ⁶Department of Diagnostic and Biological Sciences, School of Dentistry, University of Minnesota, 515 Delaware Street S.E., Minneapolis, MN 55455, USA. ⁷Characterization Facility, College of Science and Engineering, University of Minnesota, 100 Union Street S.E., Minneapolis, MN 55455, USA. ⁸Howard Hughes Medical Institute, University of Minnesota, 2231 6th Street S.E., Minneapolis, MN 55455, USA. ⁹These authors contributed equally: Veer Bhatt, Ke Shi, Daniel J. Salamango. ✉email: asundbor@umn.edu; aihar001@umn.edu

Over 10 million people worldwide are infected with human T-cell leukemia virus type-1 (HTLV-1), an oncogenic delta(δ)-retrovirus related to HIV-1^{1,2}. HTLV-1 infection causes an aggressive CD4⁺ T-cell malignancy known as adult T-cell leukemia/lymphoma (ATL) after a latency period up to several decades^{3,4}. HTLV-1 infection can also cause a chronic inflammatory disease of the spinal cord known as HTLV-1-associated myelopathy/tropical spastic paraparesis (HAM/TSP)^{5,6}. However, despite the importance of HTLV-1 as the causative agent of these diseases, there is no therapeutic intervention against HTLV-1 infection or its diseases. A critical step and the hallmark of retroviral infection is the integration of a reverse-transcribed viral genome into host chromosomal DNA⁷. Integration is carried out by the virally encoded integrase (IN) enzyme, which forms a higher-order nucleoprotein assembly (intasome) and catalyzes the 3'-end resection of a linear reverse-transcribed viral DNA and the subsequent direct attack on a target cellular DNA backbone by the nascent viral DNA 3'-OH termini^{8–10}. Due to its essential role in the viral lifecycle, HIV-1 IN is the target of several clinically used antiviral drugs (INSTIs: IN strand-transfer inhibitors) that selectively inhibit the latter strand-transfer step and also inhibit INs from other retroviruses, including HTLV-1^{7,11–13}.

Structural studies have revealed a remarkable diversity of retroviral IN–DNA complex assemblies, ranging from tetrameric IN for a spumavirus prototype foamy virus (PFV)^{12,14}, octameric IN for an alpha-retrovirus Rous sarcoma virus (RSV)^{15,16} and a beta-retrovirus mouse mammary tumor virus (MMTV)¹⁷, to even hexadecameric IN for a lentivirus maedi-visna virus (MVV)¹⁸ (Supplementary Fig. 1a, c, d). IN from another lentivirus, and an important retroviral human pathogen, HIV-1, has been reported to form a heterogeneous mixture of tetrameric to dodecameric complexes¹⁹. In addition to this structural diversity, INs from different genera of retroviruses bind to distinct host co-factors^{20–23}, and allosteric IN inhibitors (ALLINs) that target the HIV-1 IN-LEDGF/p75 interface are being developed as novel antivirals for their capacity to modulate IN multimerization and inhibit late replication steps^{24–27}. Recent studies have identified a host serine/threonine phosphatase PP2A comprising the B56 regulatory subunit as the functional binding partner for IN from deltaretroviruses including HTLV-1²⁸. However, structural information is lacking for a deltaretroviral intasome, and it is unknown how the unique co-factor PP2A stimulates the concerted integration activity or regulates the integration-site selection of deltaretroviral INs. In this study, we use cryo-electron microscopy (cryo-EM), virus infectivity assays, and biochemical analyses to show that PP2A-B56 γ is an integral component of the HTLV-1 intasome that plays an important role in HTLV-1 infection.

Results

Structure determination of the HTLV-1 intasome. To address the knowledge gap described above, we determined the structure of the HTLV-1 intasome using cryo-EM and single particle analysis at 3.7-Å resolution (Supplementary Figs. 2 and 3; Table 1). We assembled a stable complex including HTLV-1 IN, a fragment of human B56 γ spanning residues 11–380²⁹, and a branched DNA molecule containing the viral U5 long terminal repeat (LTR) sequence^{30,31} and a target DNA, mimicking the product of the concerted strand-transfer reactions. We found the presence of B56 γ to be essential for a stable HTLV-1 IN–DNA complex formation, consistent with its reported strong stimulatory effect on deltaretroviral concerted integration reactions under certain biochemical conditions^{11,28}. SDS-PAGE analysis confirmed that B56 γ (11–380) is part of the size-exclusion chromatography (SEC)-isolated HTLV-1 intasome. The molecular

mass of the HTLV-1 intasome, or the strand-transfer complex (STC) thus formed, was estimated to be 345 and 305 kDa in solution by SEC-coupled multiangle light scattering (SEC-MALS) and mass photometry analyses, respectively (Supplementary Fig. 2). In accordance with these observations, the cryo-EM density map shows a complex with the total mass of 320 kDa, including an IN tetramer bound to a strand-transfer product DNA and two molecules of B56 γ (Fig. 1).

IN tetramer in the HTLV-1 intasome. HTLV-1 intasome has a twofold symmetric structure that shares the conserved intasome core (CIC)¹⁰ with the intasome assemblies of other genera of retroviruses. The IN tetramer consists of two inner catalytic and two outer non-catalytic subunits. Each inner IN comprises reciprocally swapped N-terminal domain (NTD) bound over the viral DNA major groove, a NTD-CCD linker that contacts both viral DNA ends, the catalytic core domain (CCD) engaging the viral/target DNA junction in the minor groove, and the C-terminal domain (CTD) that fits between the NTD and CCD of the same molecule in *cis* (Fig. 1b; Supplementary Figs. 4b and 5a–c). Both NTD-CCD and CCD-CTD linkers of the inner catalytic IN run across the synaptic interface, arranged antiparallel to each other and interacting with the 5' overhang of viral DNA non-transferred strand (Fig. 2b). CTD of the non-catalytic outer INs further contribute to the extensive viral DNA interaction, bridging between the two viral DNAs attached to opposing strands of the target DNA (Fig. 1a, c, f, Supplementary Figure 5e). The target DNA shows a kink at each of the viral/target DNA junctions 6 bp apart, resulting in a total bending of $\sim 80^\circ$ away from the intasome core (Fig. 1d). The configuration of viral and target DNA molecules is similar to that observed in RSV intasome¹⁵, which shares a 6-bp spacing between the strand-transfer points. This similarity includes a zigzagged trajectory of the target DNA with an offset of the helical axes in the direction perpendicular to that of the overall bending (Fig. 1b, e; Supplementary Fig. 1).

PP2A-B56 γ –IN interaction. Two molecules of the deltaretrovirus-specific host co-factor B56 γ are bound symmetrically to the core of the HTLV-1 intasome flanking the viral DNAs, as though to cradle the IN tetramer (Fig. 1a, c). Both inner and outer subunits of an IN dimer on each side of the intasome fit in the concave surface of B56 γ (Fig. 2; Supplementary Fig. 5d). CCD and CTD of the outer non-catalytic IN are bound toward either end of the banana boat-shaped B56 γ monomer²⁹ (Fig. 2b), while the inter-domain linker between CCD and CTD takes a U-shaped conformation and makes an anchoring interaction in the central peptide-binding cleft of B56 γ (Fig. 2a, c). The ²¹¹LQPIPE²¹⁶ sequence from the CCD–CTD linker, previously shown to be critical for the binding of HTLV-1 IN to PP2A-B56³², docks into a highly conserved binding pocket known to bind the “LxxIx E ” short linear motif found in a number of host proteins regulated by PP2A^{32–34}. Residues after the sharp U-turn, ²¹⁹SLSNK²²³, interact with charged amino acids on the B56 γ surface, including Arg197 (Fig. 2c; Supplementary Fig. 6). The CCD–CTD linker of the inner catalytic IN also traverses across the B56 γ surface, running normal to the axes of the pseudo-HEAT repeat α -helices (Fig. 2b). Consistent with the observed mode of interaction between B56 γ and IN, we found that the CCD–CTD 2-domain fragment of HTLV-1 is necessary and sufficient for forming a stable complex with B56 γ isolable by SEC, and this interaction is abolished by mutating ²¹¹LQPIPE²¹⁶ to ²¹¹AQPAPA²¹⁶ (Supplementary Fig. 7). B56 γ appears to stabilize each IN dimer, help organize the CCD–CTD linkers, and position CTDs for viral DNA interactions. The distinct conformations of the HTLV-1 IN CCD–CTD linkers mediating B56 γ interaction contrast those of the much longer CCD–CTD linker of PFV IN^{12,14}, extended

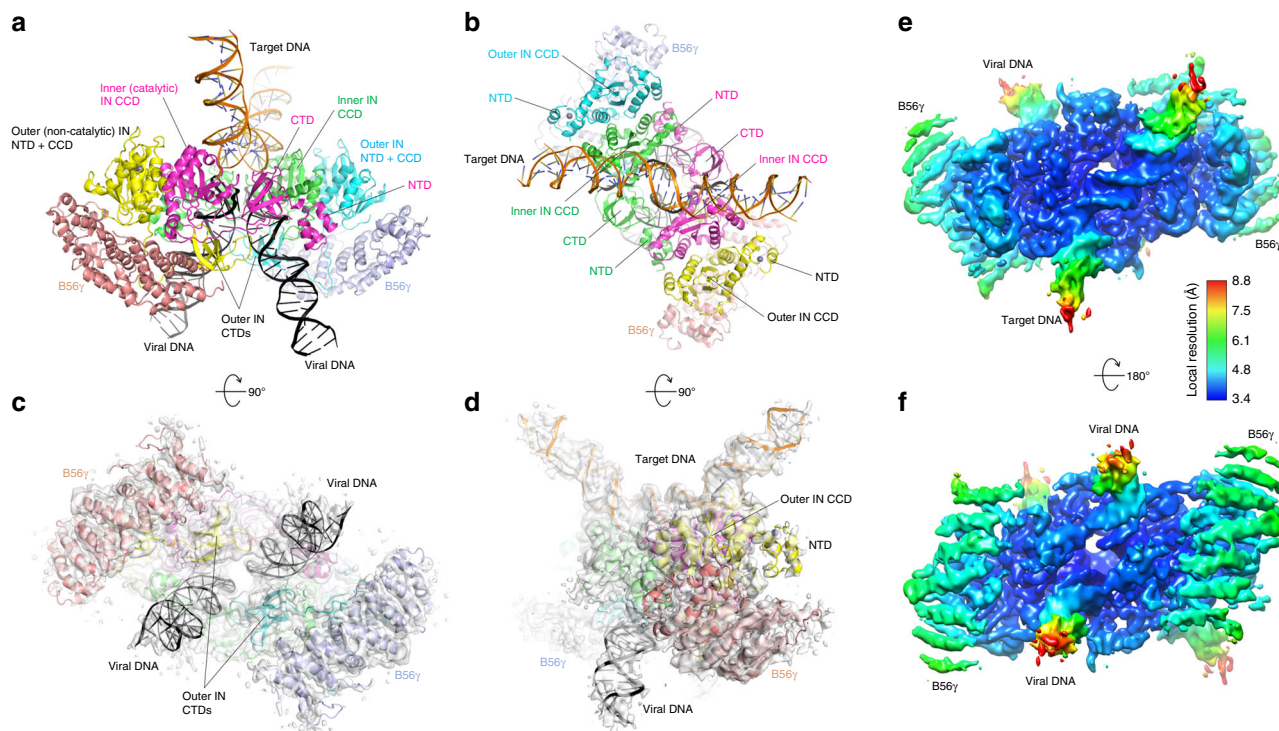


Fig. 1 Overall structure of the HTLV-1 intasome. **a** A view along a direction perpendicular to the overall twofold symmetry axis of the complex. Each protein chain is colored differently. Viral and target DNA are shown as black and orange ribbons, respectively. Zinc ions are shown as gray spheres. **b** A view along the twofold axis, from the target DNA side. **c, d** Views after 90° rotation of **a** and **b**, respectively, with the cryo-EM map in transparent surface overlaid on the atomic models. **e, f** Cryo-EM map in solid surfaces colored according to local resolution.

conformations of the short CCD–CTD linkers of RSV IN^{15,35,36} and MMTV IN¹⁷, and a crossed α -helical bundle structure assumed by the lentiviral IN CCD–CTD linkers^{18,19,37}.

B56γ is important for HTLV-1 integration in cells. Our structural data suggest that B56γ, which is a constitutively nuclear member of the PP2A B-subunit family, may play a key role in HTLV-1 integration as a scaffolding component or a regulator of the intasome assembly. To test whether B56γ is required for HTLV-1 integration in human cells, we performed HTLV-1 infectivity assays in the presence or absence of B56γ or the closely related cytoplasmic family member B56α (75% identity and 88% similarity within the core domain). As a control, HIV-1 infectivity was assessed in parallel to determine the requirement for B56γ in general retroviral genome integration. To perform HTLV-1 and HIV-1 infectivity assays, we used reverse-intron containing reporter vectors that only generate fluorescence upon successful integration of the viral genome into target cells³⁸ (Fig. 3a). Based on our structural data and subcellular localization, we predicted that shRNA-mediated knockdown of B56γ, but not B56α, would significantly impair HTLV-1 infectivity while HIV-1 infectivity would remain unchanged. As we predicted, HIV-1 infectivity showed no differences in control or knockdown cells; however, HTLV-1 infectivity was significantly impaired in cells stably expressing shRNA against B56γ, but not B56α (Fig. 3b). Of note, efficient and selective depletion of the targeted transcript in these shRNA-expressing cell lines was confirmed previously, and it was shown that neither B56γ nor B56α knockdown has discernable effect on the cell cycle³⁹.

To independently probe the requirement for B56γ in HTLV-1 integration, we assessed infectivity in cells overexpressing either wild-type B56γ or mutant variants that we predicted to have compromised interaction with HTLV-IN. As expected, all cell

lines overexpressing the indicated B56 variants had a minimal impact on HIV-1 infectivity (Fig. 3c). Interestingly, cells expressing either B56γ R197A or L194A/R197A had a significant impact on HTLV-1 infectivity, while the L194A variant alone had no effect (Fig. 3c). These findings are consistent with our structural data (Fig. 2c; Supplementary Fig. 6) and previous biochemical observations²⁸ that indicated that Arg197 plays a more important role in HTLV-IN binding to B56γ than Leu194. The dominant-negative effect of overexpressing the R197A variant of B56γ could be because of its residual interaction with IN and resulting interference with intasome formation. An alternative scenario is that HTLV-1 integration in cells actually requires the heterotrimeric PP2A holoenzyme containing B56γ (see “Discussion” below and Fig. 4) and that overexpressing the defective B56γ depleted the pool of PP2A holoenzyme containing the wild-type B56γ capable of supporting IN function.

We further examined the impact of B56γ on HTLV-1 integration by using a previously characterized high-affinity peptide inhibitor that has been shown to suppress Ebola virus infection by competitively inhibiting the nucleoprotein binding to B56 proteins⁴⁰. Co-expression of a plasmid containing four copies of the LxxIxEx peptide motif with HTLV-1 producing plasmids resulted in a significant decrease in HTLV-1 infectivity, while HIV-1 infectivity only displayed a modest decrease (Fig. 3d). However, when four copies of the control inhibitor (AxxAx) were expressed, we observed no discernable effect on HTLV-1 or HIV-1 infectivity. Taken together, these results support our structural data that B56γ is an important component of the HTLV-1 intasome.

Discussion

The observed mode of interaction with IN by PP2A-B56γ is distinct from those previously seen for the cellular co-factors of

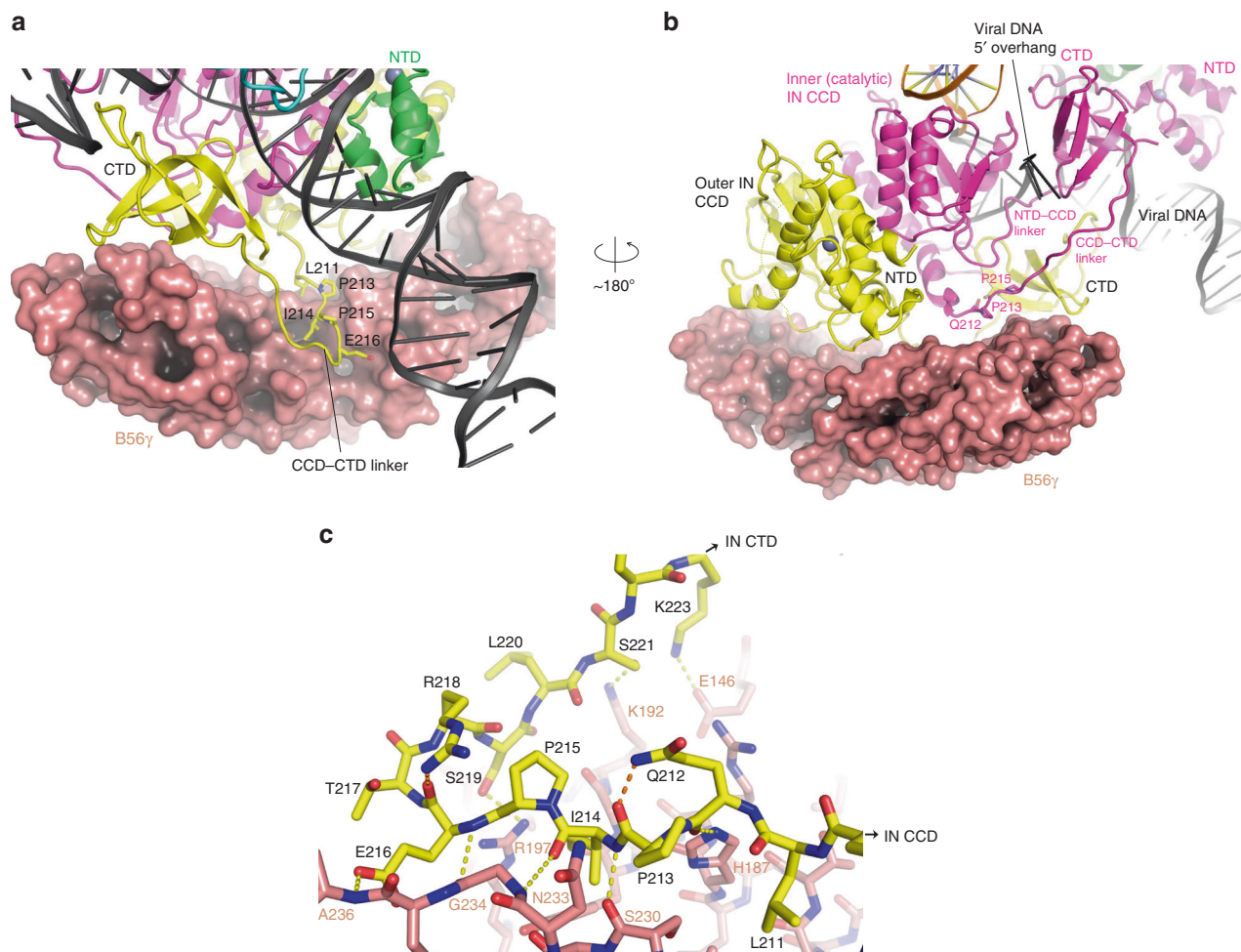


Fig. 2 B56 γ -IN interface. **a** A close-up view centered on the CCD-CTD linker of outer non-catalytic IN (yellow) containing the ²¹¹LQPIPE²¹⁶ short linear motif, which is docked in the central cleft of B56 γ . Molecular surface is shown for B56 γ . **b** A view from the opposite side of B56 γ . The CCD-CTD linker of inner catalytic IN (magenta) traverses the B56 γ surface. **c** A network of hydrogen-bonds and salt-bridges mediating the binding of IN CCD-CTD linker in a U-shaped conformation to B56 γ . Intermolecular and intramolecular contacts are highlighted by yellow and orange dotted lines, respectively.

retroviral IN from other genera, LEDGF/p75 for lentivirus^{41,42} and the BET family proteins for a gammaretrovirus murine leukemia virus (MLV)⁴³, both of which stimulate the concerted integration activity and dictate the unique integration-site preferences of cognate IN^{20,24,44–46}. The divergent IN–host factor interactions parallel the diverse strategies employed by these INs, including their different domain arrangements and oligomeric structures¹⁰, in achieving the stable CIC structure important for catalysis. A unique solution by deltaretroviruses is to hijack and re-purpose the nuclear localized subunit of a host enzyme PP2A for stabilizing the intasome assembly (Supplementary Fig. 1). PP2A is a major regulator of cell cycle and involved in numerous cellular signaling pathways, and as such, it is targeted or hijacked by a number of viruses including HIV-1^{39,40,47,48}. Notably, HTLV-1 trans-activator protein Tax has been reported to inhibit PP2A catalytic activity to manipulate cellular signaling, thereby achieving constitutive activation of NF- κ B⁴⁹. It remains to be further investigated whether the B56-IN interaction has any outcome that involves the PP2A phosphatase activity. Regardless, a simple modeling exercise suggests that PP2A holoenzyme, complete with all three subunits^{50,51}, can fit in the HTLV-1 intasome without significant steric clashes (Fig. 4), consistent with the reported association of PP2A holoenzyme comprising the B56 subunits with deltaretroviral IN²⁸. Thus, it is possible that the large scaffold (A) or the catalytic (C) subunit of the PP2A

holoenzyme plays a role in deltaretroviral integration either through its catalytic activity or mediating additional protein interactions. As previously suggested, PP2A may direct HTLV-1 integration into transcriptionally active regions via its interaction with certain chromatin-associated proteins^{28,52,53}. The HTLV-1 intasome structure reported here provides mechanistic insights into a critical host-pathogen interaction underlying the replication of an important human pathogen and affords a framework for the future development of novel therapeutic interventions against HTLV-1 infection or replication.

Methods

Protein purification. HTLV-1 IN suffers from poor solubility and is prone to aggregation, making structural studies difficult. To overcome this problem, we adapted the Sso7d-fusion strategy previously used successfully in the structural studies of HIV-1 IN¹⁹. Full-length HTLV-1 IN was expressed in *E. coli* strain BL21 (DE3) with a 6xHis-tag and a DNA-binding defective mutant of Sso7d (W24A/R43E) fused to its N-terminus and purified using nickel-affinity and gel-filtration chromatography. Purified Sso7d-IN in 20 mM HEPES-NaOH (pH 7.5), 1.0 M NaCl, 5% glycerol, 0.5 mM TCEP, was concentrated to ~ 20 mg ml⁻¹ by ultra-filtration. Sso7d (W24A/R43E)-HTLV-1 IN(wt) exhibited robust concerted integration activity, which was modestly stimulated by B56 γ and inhibited by dolutegravir (Supplementary Fig. 8) and showed the expected 6-bp spacing between the integration sites on opposing DNA strands (Supplementary Fig. 9). For the intasome assembly, we used a catalytically inactive E156Q mutant of HTLV-1 IN. B56 γ (11–380)²⁹ was expressed in *E. coli* BL21(DE3) as a 6xHis-Sumo-fusion protein and purified as above. For intasome assembly (Supplementary Fig. 2), the N-terminal 6xHis-Sumo tag was removed by Ulp1 protease treatment

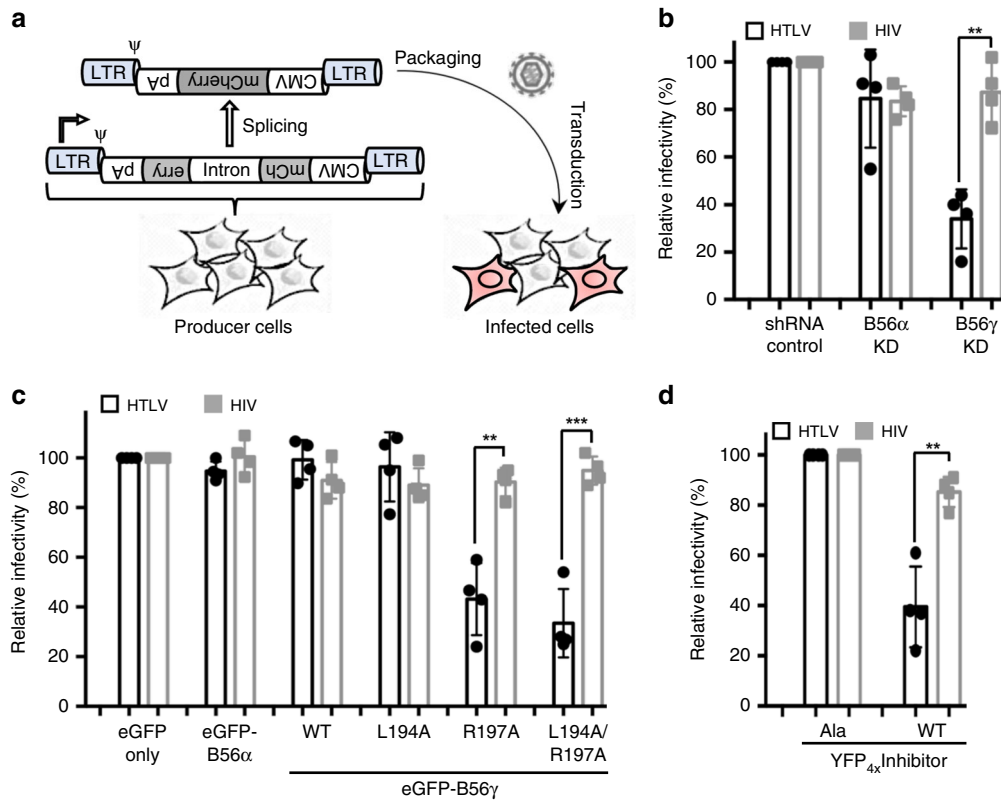


Fig. 3 B56 γ plays an important role in HTLV-1 infection. **a** A schematic diagram of the HTLV-1 and HIV-1 infection assay system. Fluorescence signal can only be generated following removal of the intron disrupting *mCherry* (producer cells) and subsequent reverse-transcription and integration of the viral genome (infected cells). **b** Relative infectivity of the indicated retroviral vectors in *B56 α* or *B56 γ* knockdown cells compared to controls. **c** Relative infectivity of the indicated retroviral vectors in cells overexpressing the indicated B56 construct compared to eGFP alone. **d** Relative infectivity of the indicated retroviral vectors in cells overexpressing four copies of a repeat peptide sequence containing **LPTIHE** (WT) compared to cells expressing four copies of the peptide sequence **APTAHA** (Ala). ** $P < 0.01$; *** $P < 0.001$ by the unpaired Student *t* test. Data shown are from four independent replicates.

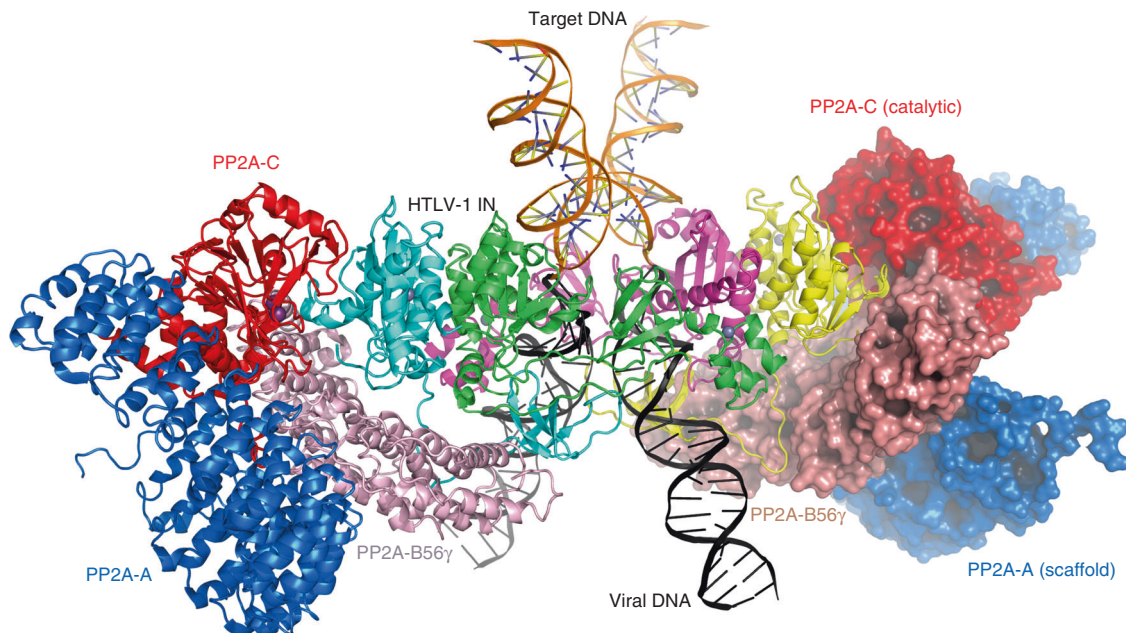


Fig. 4 A hypothetical model of HTLV-1 intasome containing PP2A holoenzymes. Two copies of PP2A holoenzyme⁵⁰ were docked on the HTLV-1 intasome by superposition of the B56 γ subunits, resulting in a snug fit between the catalytic C subunit of PP2A (red) and the catalytic domain of outer IN subunit (yellow and cyan).

during purification. Purified B56γ(11–380) in 20 mM Tris-HCl pH 7.4, 0.5 M NaCl, 5 mM β-mercaptoethanol was concentrated to ~40 mg ml⁻¹. For in vitro protein–protein-binding studies (Supplementary Fig. 7), various HTLV-1 IN fragments (NTD-CCD, CCD, CCD-CTD, CTD) and B56γ(11–380) were expressed and purified as Sumo-fusion proteins and used without removing the 6xHis-Sumo tag. The amino acid sequences of the protein constructs used in this study are shown in Supplementary Table 1. All purified proteins were flash-frozen in liquid nitrogen and stored at –80 °C until use.

HTLV-1 intasome preparation. A mixture containing 60 μM each of Sso7d-IN (E156Q) and B56γ(11–380), 50 μM each of the three pre-annealed oligonucleotides (U5-25T20, U5-nj25, Target20), 20 mM HEPES-NaOH (pH 7.5), 1.0 M NaCl, 10 mM DTT, 25% glycerol, and 10 mM MgCl₂ was dialyzed at room temperature for ~16 h against 45 mM Tris base, 45 mM boric acid, 0.1 M NaCl, 50 mM MgSO₄, 25% glycerol, and 0.5 mM TCEP. The mixture after dialysis was supplemented with 150 mM NaCl and left at room temperature for ~30 min, which helped to re-solubilize some of the precipitated IN–DNA complex. Following a brief centrifugation to remove insoluble aggregates, the protein/DNA mixture was either flash-frozen and stored for SEC-MALS analyses or injected into a Superdex200 Increase 10/300 SEC column equilibrated with the SEC buffer consisting of 20 mM Tris-HCl (pH 8.0), 0.5 M NaCl, 1 mM MgCl₂, and 0.5 mM TCEP, operating at room temperature. The IN-B56γ-DNA complex peak (Supplementary Fig. 2a, b) was used for EM grid preparation or mass photometry analysis. No IN–DNA complex was formed in the absence of B56γ(11–380).

Cryo-EM imaging and data processing. A 3.5 μL aliquot of SEC-purified HTLV-1 intasome was applied to Quantifoil R1.2/1.3 grids (Electron Microscopy Sciences) and vitrified in liquid ethane using a Mark IV Vitrobot (Thermo Fisher). Grids were imaged in a 300 kV Titan Krios electron microscope, and a total of 3326 micrographs (image stacks) were acquired with a Falcon III direct electron detector using EPU (Thermo Fisher) at a nominal magnification of 96,000×, corresponding to 0.89 Å/pixel. Motion correction was carried out using MotionCor2⁵⁴ on dose-weighted images, after deleting the first two frames. Contrast transfer function (CTF) estimation was carried out using Gctf⁵⁵ without dose weighting. All subsequent data processing and refinement steps were carried out in RELION3⁵⁶ (Supplementary Fig. 3). A small subset of micrographs was used for reference-free automated particle picking using a Laplacian of Gaussian filter, which generated a pool of 68,996 particles. 2D classification of this data set generated templates for automated particle picking from all micrographs and the resulting 1,184,769 particles were downsampled to 3.54 Å/pixel, and subjected to several iterative rounds of 2D classification to obtain a final pool of 181,688 good particles. These refined particles were re-extracted at the original pixel size of 0.89 Å/pixel, subjected to several rounds of iterative 3D classification and 3D-refinement. To improve map quality, C2 symmetry was imposed during subsequent rounds of 3D classification and refinement. Combining the half-data sets and flattening the solvent density resulted in a map of 4.1-Å resolution. The mask used to flatten the solvent density was generated in RELION3⁵⁶ at a threshold where one of the low-pass filtered half-maps stopped displaying any noise outside the reconstruction using Chimera to display the map⁵⁷. CTF refinement was employed to estimate per-particle defocus values and to refine beam tilt values. Beam-induced motion was corrected using Bayesian methods. Polished particles were subjected to one round of 3D classification leading to a final pool of 30,434 particles and a final round of 3D-refinement was carried out using a solvent mask and enabling phase-randomization based correction of Fourier shell correlation (FSC)⁵⁸ curves in every iteration of the refinement. The resolution of the final map was estimated to be 3.7 Å using the “gold standard” (FSC = 0.143)⁵⁹. Directional FSC plot (Supplementary Fig. 3l) was generated using 3DFSC⁶⁰. Local variation in resolution was estimated using RELION3⁵⁶ (Fig. 1e, f).

Model building and refinement. Homology models of HTLV-1 IN structural domains were generated using MODELLER⁶¹ and Phyre2⁶² based on the high-resolution crystal structures of RSV and HIV-1 IN domains^{35–37,63} and placed into the cryo-EM map. The linker segments and DNA molecules were built manually in COOT⁶⁴. B56γ crystal structure²⁹ was docked as a rigid body into the map. The preliminary model thus obtained was refined using PHENIX⁶⁵ real_space_refine against the cryo-EM density and a standard set of geometry/stereochemistry restraints. The resulting model closely matched an unbiased model generated de novo by PHENIX autobuild function, with the RMSD of protein main chain atoms of ~0.9 Å. A summary of the cryo-EM data collection/processing and model refinement statistics is shown in Table 1. Molecular graphics images were generated using PyMOL (www.pymol.org) or UCSF Chimera⁵⁷.

Size exclusion and light scattering. The SEC-MALS data were collected using a Superdex200 10/300 HR SEC column (GE Healthcare), connected to Agilent 1200 high performance liquid chromatography (HPLC) system, equipped with an autosampler. The elution from SEC was monitored by a photodiode array (PDA) UV/VIS detector (Agilent Technologies), differential refractometer (OPTI-Lab rEx Wyatt Technology), static and dynamic, multiangle laser light-scattering (LS) detector (HELEOS II with QELS capability, Wyatt Technology). The SEC-UV/LS/RI

Table 1 Cryo-EM data collection, refinement, and validation statistics.

	HTLV-1 intasome (EMDB-21301) (PDB: 6VOY)
<i>Data collection and processing</i>	
Magnification	96,000×
Voltage (kV)	300
Electron exposure (e ⁻ /Å ²)	30
Defocus range (μm)	1.0–2.0
Pixel size (Å)	0.8933
Symmetry imposed	C2
Initial particle images (no.)	1,184,769
Final particle images (no.)	30,434
Map resolution (Å)	3.7
FSC threshold	0.143
Map resolution range (Å)	3.4–7.1
<i>Refinement</i>	
Initial model used (PDB code)	5EJK
Model resolution (Å)	3.7
FSC threshold	0.143
Map sharpening B factor (Å ²)	–30
<i>Model composition</i>	
Non-hydrogen atoms	17,558
Protein residues	1724
DNA residues	180
Ligands (Zn ²⁺ , Mg ²⁺)	4,2
<i>B factors (Å²)</i>	
Protein	181.09
DNA	300.68
Ligand	203.04
<i>R.m.s. deviations</i>	
Bond lengths (Å)	0.010
Bond angles (°)	1.131
<i>Validation</i>	
MolProbity score	2.26
Clashscore	33.53
Poor rotamers (%)	0.66
<i>Ramachandran plot</i>	
Favored (%)	96.29
Allowed (%)	3.71
Disallowed (%)	0.00

system was equilibrated with 20 mM Tris-HCl (pH 8.0), 0.5 M NaCl, 1 mM MgCl₂, and 1 mM DTT at the flow rate of 0.5 ml min⁻¹. Two software packages were used for data collection and analysis; the Chemstation software (Agilent Technologies) controlled the HPLC operation and data collection from the multi-wavelength UV/VIS detector, while the ASTRA software (Wyatt Technology) collected data from the refractive index (RI) detector, the light-scattering detectors, and recorded the UV trace at 280 nm sent from the PDA detector. The weight average molar masses, *M_w*, were determined across the entire elution profile in the intervals of 2 s from static LS measurement using ASTRA software as previously described^{66,67}. During data analysis, a dn/dc value of 0.188 ml g⁻¹ was used as it proved satisfactory during analyses of protein standards analyzed before and after the samples of interest. The IN-B56γ-DNA complex eluted with a *M_w* of ~345 kDa; there were no changes in *M_w* when the complex was analyzed at a fivefold lower concentration (Supplementary Fig. 2c). Additional information about the stoichiometry of the protein–DNA complex was obtained from the UV/RI ratio, which is proportional to extinction coefficient of the molecule, measured individually for the proteins alone, DNA alone, and the complex samples. Since the UV/RI ratio for DNA was 7.5 times higher than the value observed for proteins, the parameter is very sensitive to the protein to DNA ratio present in the eluting complex. The observed UV/RI ratio for the eluting complex closely matched, with 1% deviation, the value calculated for the stoichiometry observed in the cryo-EM structure.

Plasmids and cloning for cell-based assays. The eGFP control and B56 expression vectors used in this study were cloned into the pQCXIH retroviral expression vector as described previously³⁹. B56γ point mutants were generated by PCR amplification using Phusion high-fidelity DNA polymerase (NEB, Ipswich, MA) and overlapping PCR to introduce the indicated mutations. To generate the wild-type and alanine YFP_ΔInhibitor vectors, cDNA sequences were obtained as gBlocks from Integrated DNA Technologies (IDT) and cloned into pcDNA5TO

expression vectors using *HindIII* and *NotI* restriction enzymes. The sequences used for generating the inhibitor vectors have been described previously⁴⁰. The control and knockdown shRNA constructs have also been described previously³⁹. All constructs were confirmed by restriction digestion and Sanger sequencing.

Cell lines and culture conditions. 293T HEK cells were maintained in DMEM (Hyclone, South Logan, UT) supplemented with 10% FBS (Gibco, Gaithersburg, MD) and 0.5% pen/strep (50 units). 293T cells were transfected with TransIT LTI (Mirus, Madison, WI) according to the manufacturer's protocol. To generate stable eGFP, B56 α , and B56 γ wild-type and mutant cell lines, viruses were produced from 293T cells transfected with the pQCXIH retroviral expression vectors described above, an MLV GagPol packaging vector, and a VSV-G expression vector. Media was harvested 48 h post transfection and frozen at -80°C for 4–6 h, thawed and centrifuged at $1500\times g$, and combined with fresh 293T cells. To generate pure cell populations, samples were treated with hygromycin B (Sigma, 200 $\mu\text{g}/\text{ml}$) 48 h post transduction. For generating stable shRNA knockdown/vector control lines, 293T cells were transfected with the shRNA vector, an HIV-1 Gag/Pol packaging construct, and a VSV-G expression vector. Media was harvested 48 h post transfection, and frozen at -80°C for 4–6 h, thawed and centrifuged at $1500\times g$, and combined with fresh 293T cells. Pure cell populations were generated by treating with puromycin for 48 h to produce a pure population (Sigma, 1 $\mu\text{g}/\text{ml}$).

HIV-1 and HTLV-1 infectivity assays. A one-step transfection/infection assay was performed in 293T cells using a 12-well culture plate. Roughly, 150,000 cells (either non-transduced 293T cells, cells stably expressing the indicated eGFP-B56 proteins, or cells stably expressing the indicated shRNA vector), were seeded into 12-well plates and allowed to adhere overnight. After 24 h, cells were transfected with either 0.6 μg of an HIV-1 packaging vector, 0.15 μg of VSV-G expression vector, and 0.9 μg of reporter plasmid, or, 0.6 μg of the full-length HTLV-1 pCMV HT1-M genome, and 0.9 μg of a reporter plasmid. The generation of these vectors has been described previously³⁸. To test the YFP₄Inhibitor vectors, the aforementioned HIV/HTLV transfection protocols were followed along with 0.4 μg of either wild-type or alanine inhibitor plasmids. Media was changed 16 h following transfection, and the cells were harvested 5 days later, and infectivity was assessed using flow cytometry (representative gating approaches are depicted in Supplementary Fig. 10). The data were collected from four independent experiments, and infectivity results were plotted using GraphPad Prism 6 software with error bars representing standard error mean (SEM).

Mass photometry. Microscope coverslips (High Precision coverslips, No. 1.5, 24 \times 50 mm, ThorLabs) were cleaned by sequential washing in 100% isopropanol and Milli-Q H₂O, followed by drying with a clean air stream. Silicon gaskets (Grace bio-labs, CultureWell™ Reusable Gaskets, CW-50R-1.0) were placed on the clean coverslips to create wells. Immediately prior to mass photometry measurements, protein or protein–DNA complex samples were diluted directly on the coverslip in the SEC buffer. Each sample was measured in a new well (i.e., each well was used once). To find focus, fresh buffer was first added into the well, the focal position was identified and secured in place with an autofocus system based on total internal reflection for the entire measurement. For each acquisition, 18–19 μL of buffer was introduced into the well and, following autofocus stabilization, 2–1 μL of sample was added then movies of 60-s duration were recorded. Data were acquired using a OneMP mass photometer (Refeyn Ltd, Oxford, UK). Data acquisition was performed using AcquireMP (Refeyn LTD, v2.0) using default settings. Mass photometry movies were processed and analyzed using DiscoverMP (v1.2.4) using default settings.

In vitro integration assay. The concerted integration activity of Sso7d(W24A/R43E)-HTLV-1 IN(wt) was tested using a 3'-OH recessed viral DNA substrate 25 R containing the HTLV-1 U5 LTR sequence, prepared by annealing two HPLC-purified oligonucleotides 5'-Cy5-CCAGGAGAGAAATTTAGTACACA-3' and 5'-ACTGTGTACTAAATTTCTCTCCTGG-3' (IDT). The reaction mix initially included 0.5 μM viral DNA substrate and 1.5 μM IN in 25 mM HEPES (pH 7.0), 100 mM NaCl, 10 mM MgCl₂, 10 mM ZnCl₂, 10 mM dithiothreitol (DTT), and 10% (v/v) dimethyl sulfoxide (DMSO). After an initial preincubation at 14°C for 15 min, the supercoiled target DNA, pBSKZeo (2.7 kb)⁴⁶, was added to a final concentration of 8 nM, and strand transfer was carried out at 37°C for 45 min. The reactions were stopped by adding EDTA to a final concentration of 25 mM, and samples were deproteinized with 0.5% SDS, 1 mg/ml proteinase K for 1 hr at 37°C . Strand-transfer products were separated on a 1.5% agarose gel and visualized by scanning for Cy5 fluorescence on Typhoon 9500 Laser Scanner (GE Healthcare Life Sciences). The gel was stained with SYBR Gold (Invitrogen) and analyzed by a Typhoon 9500 scanner to visualize the target DNA (shown on the left and right, respectively, in Supplementary Fig. 8).

HTLV-1 integration-site sequencing. The viral DNA–target junctions of the concerted integration products generated in vitro by Sso7d(W24A/R43E)-HTLV-1 IN(wt) were sequenced. The strand-transfer reactions were carried out as above, except with HTLV-1 U5 LTR DNA (39 CatRE: 5'- CCGTGC GAATTCGGATCC AGGAGAGAAATTTAGTACACA-3' and 41 Non-CatRE: 5'- ACTGTGTACTAA

ATTTCTCTCTGGATCCGAATTCGCACGG-3') for 20 min at 37°C . The concerted products were isolated from a 0.8% agarose gel and purified by electroelution. The products were treated with phi29 DNA polymerase (New England Biolabs) in the presence of 500 μM dNTPs followed by Klenow polymerase (Promega) treatment. The blunt-ended products were ligated into Zero Blunt PCR vector (Invitrogen), and the resulting DNA was used to transform TOP10F cells (Invitrogen). Recombinant clones were screened by restriction enzyme digestion using *EcoRI* and *HindIII* to confirm the presence of concerted products. In total, 24 clones having the correct size concerted products were sequenced using custom primers (KKPBlunt244: 5'-GGTGACGCGTTAGAATACTCAAGC-3', and ccd665-R: 5'-GCCCCGGCGTGTCAATAATATC-3') to analyze the LTR-target junction and host site duplications. The majority of the clones (22 out of 24) had the expected target DNA sequence duplication size of 6-bp. Sequence logos (Supplementary Fig. 9) were generated from 21 unique clones using WebLogo⁶⁸ and show similar integration target sequence preference to those previously reported for HTLV-1 IN^{69,70}. Two clones had deletions of 116-bp and 1026-bp, which could be due to multiple integration into a single target DNA.

Protein-binding analyses. Various fragments of HTLV-1 IN, fused to Sumo (yeast Smt3) on their N-terminus, were injected into a Superdex200 10/300 SEC column either by itself or after being mixed with an equimolar amount of Sumo-fused B56 γ (11–380). The column was operated at 4°C with a flow rate of 0.4 ml min⁻¹, and the elution buffer contained 20 mM Tris-HCl (pH 8.0), 0.5 M NaCl, 1 mM MgCl₂, and 0.5 mM TCEP. In total, 83.5 nanomoles of each protein or complex were brought up to a standardized volume of 242 μL with the running buffer, prior to each sample injection. Protein complex formation was assessed by monitoring the elution profiles and analyzing the collected fractions by SDS-PAGE.

Reporting summary. Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Data availability

Atomic coordinates and the cryo-EM density map have been deposited in the Protein Data Bank and the Electron Microscopy Data Bank (EMDB) under accession code 6VOY and EMD-21301, respectively. All other data are available from the authors upon request.

Received: 3 March 2020; Accepted: 3 June 2020;

Published online: 19 June 2020

References

- Gessain, A. & Cassar, O. Epidemiological aspects and world distribution of HTLV-1 infection. *Front. Microbiol.* **3**, 388 (2012).
- Proietti, F. A., Carneiro-Proietti, A. B., Catalan-Soares, B. C. & Murphy, E. L. Global epidemiology of HTLV-I infection and associated diseases. *Oncogene* **24**, 6058–6068 (2005).
- Mahieux, R. & Gessain, A. Adult T-cell leukemia/lymphoma and HTLV-1. *Curr. Hematol. Malig. Rep.* **2**, 257–264 (2007).
- Yoshida, M., Miyoshi, I. & Hinuma, Y. Isolation and characterization of retrovirus from cell lines of human adult T-cell leukemia and its implication in the disease. *Proc. Natl Acad. Sci. USA* **79**, 2031–2035 (1982).
- Gessain, A. & Mahieux, R. Tropical spastic paraparesis and HTLV-1 associated myelopathy: clinical, epidemiological, virological and therapeutic aspects. *Rev. Neurol.* **168**, 257–269 (2012).
- Matsuura, E. et al. HTLV-1 associated myelopathy/tropical spastic paraparesis (HAM/TSP): a comparative study to identify factors that influence disease progression. *J. Neurol. Sci.* **371**, 112–116 (2016).
- Lesbats, P., Engelman, A. N. & Cherepanov, P. Retroviral DNA integration. *Chem. Rev.* **116**, 12730–12757 (2016).
- Craigie, R. HIV integrase, a brief overview from chemistry to therapeutics. *J. Biol. Chem.* **276**, 23213–23216 (2001).
- Engelman, A., Mizuuchi, K. & Craigie, R. HIV-1 DNA integration: mechanism of viral DNA cleavage and DNA strand transfer. *Cell* **67**, 1211–1221 (1991).
- Engelman, A. N. & Cherepanov, P. Retroviral intasomes arising. *Curr. Opin. Struct. Biol.* **47**, 23–29 (2017).
- Barski, M. S., Minnell, J. J. & Maertens, G. N. Inhibition of HTLV-1 infection by HIV-1 first- and second-generation integrase strand transfer inhibitors. *Front. Microbiol.* **10**, 1877 (2019).
- Hare, S., Gupta, S. S., Valkov, E., Engelman, A. & Cherepanov, P. Retroviral intasome assembly and inhibition of DNA strand transfer. *Nature* **464**, 232–236 (2010).
- Hazuda, D. J. et al. Inhibitors of strand transfer that prevent integration and inhibit HIV-1 replication in cells. *Science* **287**, 646–650 (2000).

14. Maertens, G. N., Hare, S. & Cherepanov, P. The mechanism of retroviral integration from X-ray structures of its key intermediates. *Nature* **468**, 326–329 (2010).
15. Yin, Z. et al. Crystal structure of the Rous sarcoma virus intasome. *Nature* **530**, 362–366 (2016).
16. Bera, S., Pandey, K. K., Aihara, H. & Grandgenett, D. P. Differential assembly of Rous sarcoma virus tetrameric and octameric intasomes is regulated by the C-terminal domain and tail region of integrase. *J. Biol. Chem.* **293**, 16440–16452 (2018).
17. Ballandras-Colas, A. et al. Cryo-EM reveals a novel octameric integrase structure for betaretroviral intasome function. *Nature* **530**, 358–361 (2016).
18. Ballandras-Colas, A. et al. A supramolecular assembly mediates lentiviral DNA integration. *Science* **355**, 93–95 (2017).
19. Passos, D. O. et al. Cryo-EM structures and atomic model of the HIV-1 strand transfer complex intasome. *Science* **355**, 89–92 (2017).
20. Cherepanov, P. LEDGF/p75 interacts with divergent lentiviral integrases and modulates their enzymatic activity in vitro. *Nucleic Acids Res.* **35**, 113–124 (2007).
21. De Rijck, J. et al. The BET family of proteins targets moloney murine leukemia virus integration near transcription start sites. *Cell Rep.* **5**, 886–894 (2013).
22. Gupta, S. S. et al. Bromo- and extraterminal domain chromatin regulators serve as cofactors for murine leukemia virus integration. *J. Virol.* **87**, 12721–12736 (2013).
23. Llano, M. et al. An essential role for LEDGF/p75 in HIV integration. *Science* **314**, 461–464 (2006).
24. Debyser, Z., Christ, F., De Rijck, J. & Gijssbers, R. Host factors for retroviral integration site selection. *Trends Biochem. Sci.* **40**, 108–116 (2015).
25. Engelman, A., Kessl, J. J. & Kvaratskhelia, M. Allosteric inhibition of HIV-1 integrase activity. *Curr. Opin. Chem. Biol.* **17**, 339–345 (2013).
26. Gupta, K. et al. Allosteric inhibition of human immunodeficiency virus integrase: late block during viral replication and abnormal multimerization involving specific protein domains. *J. Biol. Chem.* **289**, 20477–20488 (2014).
27. Gupta, K. et al. Structural basis for inhibitor-induced aggregation of HIV integrase. *PLoS Biol.* **14**, e1002584 (2016).
28. Maertens, G. N. B'-protein phosphatase 2A is a functional binding partner of delta-retroviral integrase. *Nucleic Acids Res.* **44**, 364–376 (2016).
29. Magnusdottir, A. et al. The structure of the PP2A regulatory subunit B56 gamma: the remaining piece of the PP2A jigsaw puzzle. *Proteins* **74**, 212–221 (2009).
30. Mortreux, F. et al. Somatic mutation in human T-cell leukemia virus type 1 provirus and flanking cellular sequences during clonal expansion in vivo. *J. Natl Cancer Inst.* **93**, 367–377 (2001).
31. Seiki, M., Hattori, S., Hirayama, Y. & Yoshida, M. Human adult T-cell leukemia virus: complete nucleotide sequence of the provirus genome integrated in leukemia cell DNA. *Proc. Natl Acad. Sci. USA* **80**, 3618–3622 (1983).
32. Hertz, E. P. T. et al. A conserved motif provides binding specificity to the PP2A-B56 phosphatase. *Mol. Cell* **63**, 686–695 (2016).
33. Wang, J. et al. Crystal structure of a PP2A B56-BubR1 complex and its implications for PP2A substrate recruitment and localization. *Protein Cell* **7**, 516–526 (2016).
34. Wang, X., Bajaj, R., Bollen, M., Peti, W. & Page, R. Expanding the PP2A interactome by defining a B56-specific SLiM. *Structure* **24**, 2174–2181 (2016).
35. Shi, K. et al. A possible role for the asymmetric C-terminal domain dimer of Rous sarcoma virus integrase in viral DNA binding. *PLoS ONE* **8**, e56892 (2013).
36. Yang, Z. N., Mueser, T. C., Bushman, F. D. & Hyde, C. C. Crystal structure of an active two-domain derivative of Rous sarcoma virus integrase. *J. Mol. Biol.* **296**, 535–548 (2000).
37. Chen, J. C. et al. Crystal structure of the HIV-1 integrase catalytic core and C-terminal domains: a model for viral DNA binding. *Proc. Natl Acad. Sci. USA* **97**, 8233–8238 (2000).
38. Shunaeva, A. et al. Improvement of HIV-1 and human t cell lymphotropic virus type 1 replication-dependent vectors via optimization of reporter gene reconstitution and modification with intronic short hairpin RNA. *J. Virol.* **89**, 10591–10601 (2015).
39. Salamango, D. J. et al. HIV-1 Vif triggers cell cycle arrest by degrading cellular PPP2R5 phospho-regulators. *Cell Rep.* **29**, 1057–1065 e1054 (2019).
40. Kruse, T. et al. The Ebola virus nucleoprotein recruits the host PP2A-B56 phosphatase to activate transcriptional support activity of VP30. *Mol. Cell* **69**, 136–145 e136 (2018).
41. Hare, S. et al. Structural basis for functional tetramerization of lentiviral integrase. *PLoS Pathog.* **5**, e1000515 (2009).
42. Hare, S. et al. A novel co-crystal structure affords the design of gain-of-function lentiviral integrase mutants in the presence of modified PSIP1/LEDGF/p75. *PLoS Pathog.* **5**, e1000259 (2009).
43. Crowe, B. L. et al. Structure of the Brd4 ET domain bound to a C-terminal motif from gamma-retroviral integrases reveals a conserved mechanism of interaction. *Proc. Natl Acad. Sci. USA* **113**, 2086–2091 (2016).
44. Craigie, R. & Bushman, F. D. Host factors in retroviral integration and the selection of integration target sites. *Microbiol. Spectr.* **2**, <https://doi.org/10.1128/microbiolspec.MDNA3-0026-2014> (2014).
45. Kvaratskhelia, M., Sharma, A., Larue, R. C., Serrao, E. & Engelman, A. Molecular mechanisms of retroviral integration site selection. *Nucleic Acids Res.* **42**, 10209–10225 (2014).
46. Pandey, K. K., Sinha, S. & Grandgenett, D. P. Transcriptional coactivator LEDGF/p75 modulates human immunodeficiency virus type 1 integrase-mediated concerted integration. *J. Virol.* **81**, 3969–3979 (2007).
47. Guernon, J. et al. PP2A targeting by viral proteins: a widespread biological strategy from DNA/RNA tumor viruses to HIV-1. *Biochim. Biophys. Acta* **1812**, 1498–1507 (2011).
48. Wlodarchak, N. & Xing, Y. PP2A as a master regulator of the cell cycle. *Crit. Rev. Biochem. Mol. Biol.* **51**, 162–184 (2016).
49. Fu, D. X., Kuo, Y. L., Liu, B. Y., Jeang, K. T. & Giam, C. Z. Human T-lymphotropic virus type I tax activates I-kappa B kinase by inhibiting I-kappa B kinase-associated serine/threonine protein phosphatase 2A. *J. Biol. Chem.* **278**, 1487–1493 (2003).
50. Cho, U. S. & Xu, W. Crystal structure of a protein phosphatase 2A heterotrimeric holoenzyme. *Nature* **445**, 53–57 (2007).
51. Xu, Y. et al. Structure of the protein phosphatase 2A holoenzyme. *Cell* **127**, 1239–1251 (2006).
52. Meekings, K. N., Leipzig, J., Bushman, F. D., Taylor, G. P. & Bangham, C. R. HTLV-1 integration into transcriptionally active genomic regions is associated with proviral expression and with HAM/TSP. *PLoS Pathog.* **4**, e1000027 (2008).
53. Melamed, A. et al. Genome-wide determinants of proviral targeting, clonal abundance and expression in natural HTLV-1 infection. *PLoS Pathog.* **9**, e1003271 (2013).
54. Zheng, S. Q. et al. MotionCor2: anisotropic correction of beam-induced motion for improved cryo-electron microscopy. *Nat. Methods* **14**, 331–332 (2017).
55. Zhang, K. Gctf: real-time CTF determination and correction. *J. Struct. Biol.* **193**, 1–12 (2016).
56. Zivanov, J. et al. New tools for automated high-resolution cryo-EM structure determination in RELION-3. *eLife* **7**, <https://doi.org/10.7554/eLife.42166> (2018).
57. Pettersen, E. F. et al. UCSF Chimera-a visualization system for exploratory research and analysis. *J. Comput. Chem.* **25**, 1605–1612 (2004).
58. Chen, S. et al. High-resolution noise substitution to measure overfitting and validate resolution in 3D structure determination by single particle electron cryomicroscopy. *Ultramicroscopy* **135**, 24–35 (2013).
59. Scheres, S. H. & Chen, S. Prevention of overfitting in cryo-EM structure determination. *Nat. Methods* **9**, 853–854 (2012).
60. Tan, Y. Z. et al. Addressing preferred specimen orientation in single-particle cryo-EM through tilting. *Nat. Methods* **14**, 793–796 (2017).
61. Webb, B. & Sali, A. Comparative Protein Structure Modeling Using MODELLER. *Curr. Protoc. Bioinforma.* **54**, 5 6 1-5 6 37, <https://doi.org/10.1002/cpb1.3> (2016).
62. Kelley, L. A., Mezulis, S., Yates, C. M., Wass, M. N. & Sternberg, M. J. The Phyre2 web portal for protein modeling, prediction and analysis. *Nat. Protoc.* **10**, 845–858 (2015).
63. Wang, J. Y., Ling, H., Yang, W. & Craigie, R. Structure of a two-domain fragment of HIV-1 integrase: implications for domain organization in the intact protein. *EMBO J.* **20**, 7333–7343 (2001).
64. Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of Coot. *Acta Crystallogr. D. Biol. Crystallogr.* **66**, 486–501 (2010).
65. Liebschner, D. et al. Macromolecular structure determination using X-rays, neutrons and electrons: recent developments in Phenix. *Acta Crystallogr. D. Struct. Biol.* **75**, 861–877 (2019).
66. Foltá-Stogniew, E. Oligomeric states of proteins determined by size-exclusion chromatography coupled with light scattering, absorbance, and refractive index detectors. *Methods Mol. Biol.* **328**, 97–112 (2006).
67. Foltá-Stogniew, E. & Williams, K. R. Determination of molecular masses of proteins in solution: implementation of an HPLC size exclusion chromatography and laser light scattering service in a core laboratory. *J. Biomol. Tech.* **10**, 51–63 (1999).
68. Crooks, G. E., Hon, G., Chandonia, J. M. & Brenner, S. E. WebLogo: a sequence logo generator. *Genome Res.* **14**, 1188–1190 (2004).
69. Derse, D. et al. Human T-cell leukemia virus type 1 integration target sites in the human genome: comparison with those of other retroviruses. *J. Virol.* **81**, 6731–6741 (2007).
70. Serrao, E., Ballandras-Colas, A., Cherepanov, P., Maertens, G. N. & Engelman, A. N. Key determinants of target DNA recognition by retroviral intasomes. *Retrovirology* **12**, 39 (2015).

Acknowledgements

We thank Dmitriy Mazurov for the HT1-M packaging vector and replication-dependent HTLV-1 and HIV-1 reporter vectors, Ewa Folta-Stogniew for SEC-MALS analysis, Surajit Banerjee for X-ray diffraction experiments, Takahide Kouno for preliminary negative stain EM, Brenda Watt and Gabriella Kiss for assistance with mass photometry, Nadine Shaban for discussions. The SEC-LS/UV/RI instrumentation in Biophysics Resource of Keck Facility at Yale University was supported by NIH Award 1S10RR023748-01 (to E.F.-S.). This work was supported by NIGMS R35-GM118047 (to H.A.), NIAID R37 AI064046 (to R.S.H.), the Hormel Foundation (to A.C.S.-L.), and NIAID K99 AI147811 (to D.J.S.). R.S.H. is the Margaret Harvey Schering Land Grant Chair for Cancer Research, a Distinguished University McKnight Professor, and an Investigator of the Howard Hughes Medical Institute.

Author contributions

V.B. and A.C.S.-L. performed cryo-EM grid preparation, data acquisition, and processing. K.S. performed model building and refinement. D.J.S. and R.S.H. performed virus infectivity assays. N.H.M., H.O.B., F.K., K.O., K.S., and H.A. carried out intasome preparation and biochemical analyses. K.K.P., S.B., and D.P.G. examined in vitro integration activities. W.Z. assisted with preliminary cryo-EM analysis. H.A. conceived and managed the project, and put together the paper with contributions from all authors.

Competing interests

R.S.H. is a co-founder, shareholder, and consultant of ApoGen Biotechnologies Inc. H.A. is a consultant of ApoGen Biotechnologies Inc. Remaining authors declare no competing interests.

Additional information

Supplementary information is available for this paper at <https://doi.org/10.1038/s41467-020-16963-6>.

Correspondence and requests for materials should be addressed to A.C.S.-L. or H.A.

Peer review information *Nature Communications* thanks the anonymous reviewers for their contribution to the peer review of this work. Peer reviewer reports are available.

Reprints and permission information is available at <http://www.nature.com/reprints>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2020