









Cryo-EM structure of the deltaretroviral intasome in complex with the PP2A regulatory subunit B56 γ

Michał S. Barski ¹, Jordan J. Minnell ¹, Zuzana Hodakova ², Valerie E. Pye ², Andrea Nans ³, Peter Cherepanov ^{1,2} & Goedele N. Maertens ¹ 

Human T-cell lymphotropic virus type 1 (HTLV-1) is a deltaretrovirus and the most oncogenic pathogen. Many of the ~20 million HTLV-1 infected people will develop severe leukaemia or an ALS-like motor disease, unless a therapy becomes available. A key step in the establishment of infection is the integration of viral genetic material into the host genome, catalysed by the retroviral integrase (IN) enzyme. Here, we use X-ray crystallography and single-particle cryo-electron microscopy to determine the structure of the functional deltaretroviral IN assembled on viral DNA ends and bound to the B56 γ subunit of its human host factor, protein phosphatase 2 A. The structure reveals a tetrameric IN assembly bound to two molecules of the phosphatase via a conserved short linear motif. Insight into the deltaretroviral intasome and its interaction with the host will be crucial for understanding the pattern of integration events in infected individuals and therefore bears important clinical implications.

¹Imperial College London, St Mary's Hospital, Department of Infectious Disease, Section of Virology, Norfolk Place, London W2 1PG, UK. ²Chromatin Structure and Mobile DNA Laboratory, The Francis Crick Institute, 1 Midland Road, London NW1 1AT, UK. ³Structural Biology Science Technology Platform, The Francis Crick Institute, 1 Midland Road, London NW1 1AT, UK. ✉email: g.maertens@imperial.ac.uk

Despite the severe, and sometimes fatal pathology caused by HTLV-1, including adult T-cell leukaemia/lymphoma (ATLL)¹, myelopathy (HAM/TSP)² and uveitis³, most aspects of deltaretrovirus biochemistry remain *terra incognita*. It has been exactly 40 years since the discovery of HTLV-1 as the first human retrovirus. The recent open letter to the WHO, by the co-discoverer of HTLV-1 and colleagues, stresses the importance of concentrating efforts on researching this pathogen and striving for its eradication⁴. Elucidation of mechanisms behind deltaretroviral integration is particularly important in order to address issues of much needed pharmacological intervention, and to understand how integration site targeting affects clonal expansion of malignant T-cells leading to ATLL.

Following entry of the viral core into the cytoplasm, reverse transcription of retroviral genomic RNA yields linear double-stranded viral DNA (vDNA) with a copy of long terminal repeat (LTR) at each end. A multimer of integrase (IN) binds and brings together the vDNA ends within the intasome nucleoprotein complex to insert them into host chromosomal DNA⁵. IN first catalyses 3'-processing of the vDNA ends, exposing reactive 3' OH nucleophile groups that it uses to attack a pair of phosphodiester bonds within target DNA (tDNA), resulting in strand transfer. Repair of single-stranded gaps flanking hemi-integrated vDNA by host cell enzymes completes establishment of a stable provirus. The structural mechanics of the IN-mediated reactions was revealed almost a decade ago with the example of a spumaviral intasome^{6–8}. Follow-up studies uncovered remarkable diversity of intasome architectures among the retroviral genera, revealing tetrameric (spumaviruses)⁶, octameric (betaretrovirus⁹ and alpharetrovirus¹⁰) and dodeca/hexadecameric (lentiviruses)^{11,12} IN assemblies. An added layer of complexity is the recruitment of IN-binding host proteins, which in many cases help guide retroviral integration to preferred genomic loci^{13–17}. In this work, we determine the three-dimensional structure of the deltaretroviral intasome and characterize its interactions with its host factor, the PP2A-B56γ subunit.

Results

STLV-1 IN forms stable, active intasomes. We found that IN from simian T-lymphotropic virus type 1 (STLV-1), which shares 83% amino acid sequence identity with its HTLV-1 counterpart (Supplementary Fig. 1), is competent for concerted strand-transfer activity *in vitro*. Akin to HTLV-1 IN, the enzyme readily utilizes short, double-stranded oligonucleotide mimics of vDNA ends for integration (Supplementary Fig. 2a, c)^{14,18}. Formation of stable intasomes *in vitro* can be technically challenging, and often requires the presence of host factors and/or hyperactivating mutations^{11,12,19}. In one approach, the positively charged N-terminal region of lens epithelium-derived growth factor (LEDGF/p75) was fused with HIV-1 IN, in order to promote formation of intasomes *in vitro*²⁰. The deltaretroviral IN host cofactor B56γ¹⁴, as part of the heterotrimeric protein phosphatase 2A (PP2A) holoenzyme, interacts with a number of chromatin-associated proteins^{21–24} and potentially stimulates concerted integration activity of deltaretroviral INs¹⁴. To aid stable intasome formation without altering IN, we constructed a LEDGF/ΔIBD-B56γ chimera containing the DNA-binding region of LEDGF (residues 1–324) and B56γ (residues 11–380) (Supplementary Fig. 2b). Electrophoretic mobility shift assays (EMSAs) confirmed that STLV-1 IN forms a stable nucleoprotein complex with vDNA in the presence of LEDGF/ΔIBD-B56γ (Supplementary Fig. 2d). Moreover, separation of the assembly reactions by size-exclusion chromatography revealed a high-molecular weight species that was competent for strand-transfer activity (Supplementary Fig. 3a, b). Negative-stain electron microscopy (EM) of

the peak fraction identified distinct particles measuring ~15 nm in the longest dimension, with prominent twofold symmetry (Supplementary Fig. 3c–e).

Architecture of the deltaretroviral intasome. To characterize the intasome at near-atomic resolution, we imaged the particles by cryogenic electron microscopy (cryo-EM, Supplementary Fig. 4). The nucleoprotein samples were vitrified on open hole grids as well as adsorbed onto graphene oxide film, resulting in two anisotropically sampled yet complementary datasets (Supplementary Figs. 5, 6). Merging the data allowed us to refine an isotropic 3D reconstruction to an overall resolution of 3.37 Å and 2.9 Å throughout the conserved intasome core (CIC) region (Supplementary Figs. 6, 7). To aid the interpretation of the cryo-EM map, we generated a high-quality homology model of the STLV-1 IN/NTD using the SWISS-MODEL server²⁵, and determined a series of X-ray crystal structures spanning the catalytic core domain (CCD) and the C-terminal domain (CTD) of HTLV-2 and -1 IN (Supplementary Figs. 8–11, Supplementary Tables 1 and 2). The IN/CTD structure was determined in isolation as well as in complex with B56γ (Supplementary Figs. 10–11 and Supplementary Table 2). The apo CTD crystal structure shows a canonical, small β-barrel SH3-like fold, with side-to-side orientation similar to that previously seen in a HIV-1 IN/CTD crystal structure (PDB ID 5TC2) (Supplementary Fig. 10c, d). In the co-crystal structure with B56γ, the short linear motif (SLiM) harboured by IN within the CCD-CTD linker is clearly resolved, bound to a groove in the centre of B56γ; the previously characterised binding site for endogenous substrates of PP2A (see below and Supplementary Fig. 11)^{26,27}.

Rigid-body docking of the IN and B56γ crystal structures into the cryo-EM map provided us with a reliable starting model, which we extended by building the remaining regions *ab initio* guided by the map (Supplementary Fig. 7b, e). The STLV-1 intasome structure revealed a tetrameric assembly of IN subunits organised around the vDNA ends, with all domains of the tetramer resolved in the cryo-EM map (Fig. 1). Flanking two sides of the intasome are two B56γ subunits, which resemble epaulettes. The LEDGF-derived portions of the chimeric host factor construct are not visible in the cryo-EM reconstruction. Thus, while the DNA-binding moiety helped to chaperone STLV-1 intasome assembly *in vitro*, it is not involved in stable interactions within the resulting nucleoprotein complex.

Interaction of the intasome with B56γ of PP2A. Although the LxxIxE-containing region in IN is predicted to be intrinsically disordered, it is stabilised by intimate interactions with B56γ within the structure of the intasome (Supplementary Fig. 12). IN residue Pro211 is highly conserved amongst HTLV/STLV isolates and caps the CCD domain, introducing a kink in the protein backbone and allowing the CCD-CTD linkers in both IN dimers to run perpendicular to one another supporting stable association with B56γ (Supplementary Fig. 12). All four SLiM regions in the IN tetramer participate in host factor binding, creating two distinct binding sites for each of the two intasome-recruited B56γ subunits. The previously characterised canonical PP2A SLiM-binding site²⁷, also resolved in our IN-B56γ crystal structure, involves IN residues Leu213, Ile216 and Glu218 from the outer IN protomer, and B56γ residues His187, Arg188, Tyr190, Arg197, Ile227 and Ile231 (Fig. 2a). We previously identified B56γ Arg197 to be critical for binding and stimulating the concerted integration activity of HTLV-1 and HTLV-2 INs¹⁴. In contrast, Arg188, which is important for binding to endogenous phosphorylated substrates²⁷, is dispensable for binding to deltaretroviral INs¹⁴.

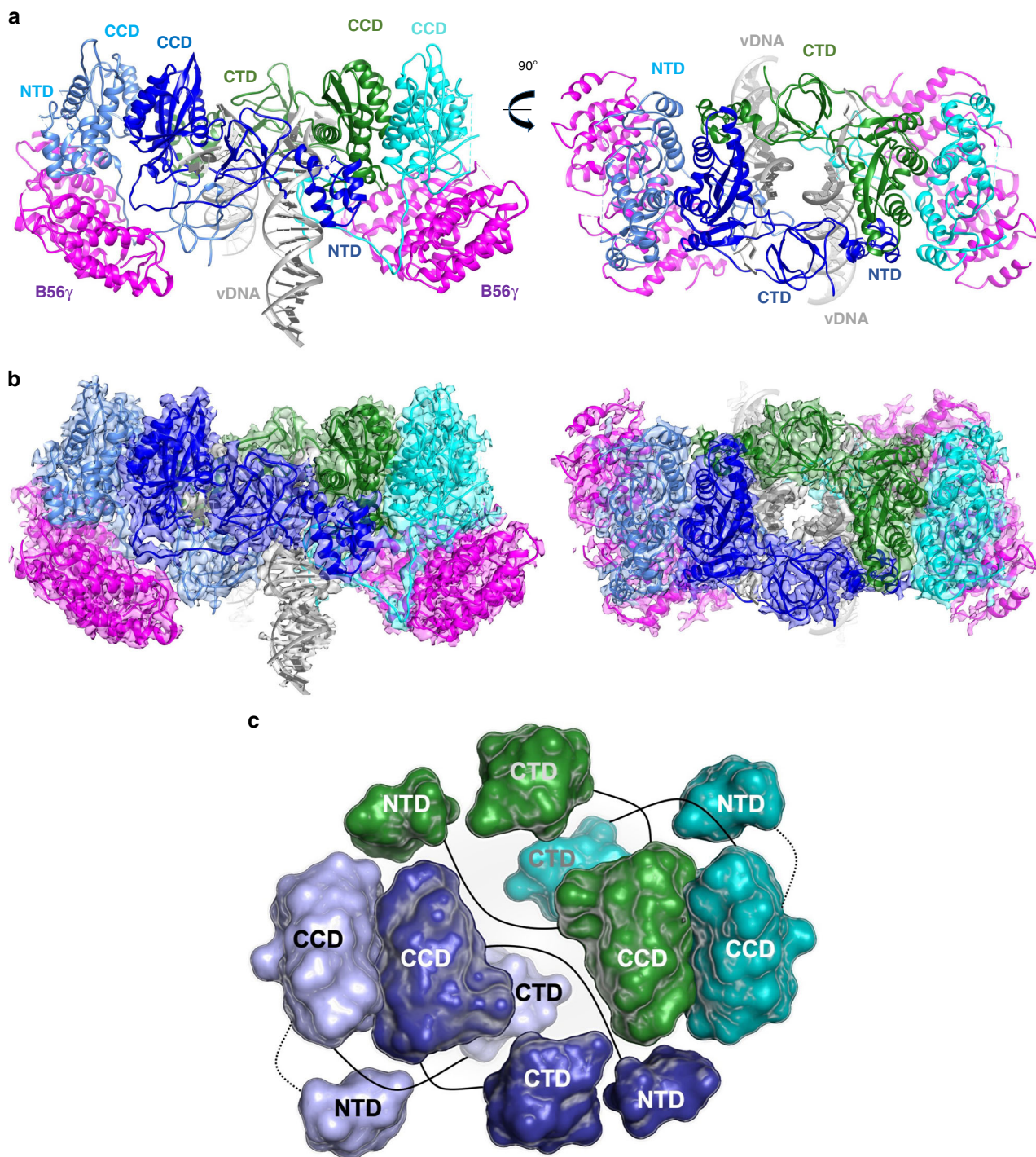
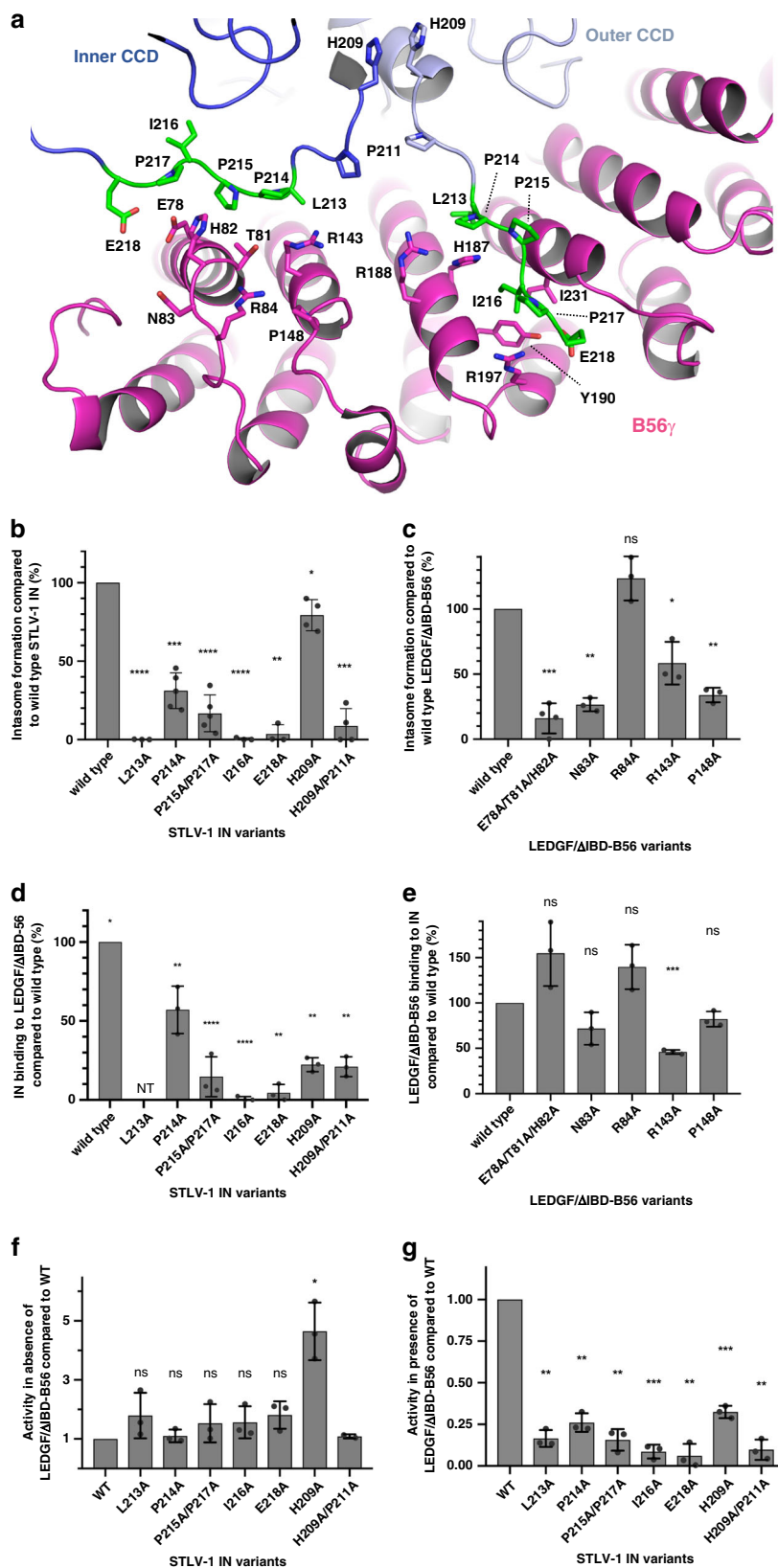


Fig. 1 Structure of the deltaretroviral intasome in complex with the B56 γ regulatory subunit of PP2A. **a** The intasome: B56 γ structure. Deltaretroviral intasome attains a tetrameric arrangement—IN chains are coloured in dark and light blue, their symmetrical counterparts in cyan and green. Viral DNA (vDNA) in grey is seen in the centre, complexed by extensive interactions with all IN domains. B56 γ (pink) flanks either side of the intasome, targeting all four available IN CCD-CTD linkers. Both side view (left) as well as top-down view (right) are shown. **b** 3.37 Å cryo-EM map reconstruction is overlaid on top of the model in orientations analogous to panel (b). **c** Exploded diagram of the lentiviral intasome showing the relative position and connectivity between the domains of the IN tetramer. IN chains are coloured as in panels (a, b). B56 γ and vDNA have been removed for clarity.

The cryo-EM structure revealed an additional novel binding site on B56 γ , accommodating the inner IN CCD-CTD linker that runs along the width of B56 γ and involves B56 γ residues Glu78, Thr81, His82 and Arg143 (Fig. 2a). Thus, the virus uses a remarkable strategy, exploiting the oligomeric assembly of the intasome, to bind a host factor at two separate sites by means of the same intrinsically disordered yet highly conserved region, located on

neighbouring IN protomers (Supplementary Fig. 12). The presence of a histidine residue, central to the binding in both B56 γ SLiM-recipient sites (His82 and His187), is also of note. Alanine substitutions of IN residues Leu213, Pro214, Pro214/Pro217, Ile216, Glu218 and His209/Pro211, as predicted, significantly reduced intasome assembly (Fig. 2b), binding to Δ IBD-B56 γ (Fig. 2d) and stimulation of concerted integrase activity



by the host factor (Fig. 2g). Although the Ala substitution of His209 only mildly affected intasome assembly (Fig. 2b), in contrast to the other IN mutants, the intrinsic concerted integration activity (in the absence of B56 γ) of the H209A mutant was elevated compared to WT IN (Fig. 2f). Mutations of B56 γ residues Glu78, Thr81, His82 and Arg143 to Ala abrogated

binding to IN and abolished stimulation of intasome assembly (Fig. 2c, e). In addition, B56 γ residues Asn83 and Pro148, that are in close proximity to the IN CTD, appear to only play a role specific to intasome assembly. Indeed, mutations of these B56 γ residues did not affect binding to free IN (Fig. 2e), while significantly reducing intasome assembly (Fig. 2c). Flag-

Fig. 2 Residues involved in B56 γ binding to the STLV-1 intasome. **a** Two distinct IN-binding sites on B56 γ are shown binding the conserved SLiM region contained on two IN chains constituting half of the twofold symmetrical IN dimer of dimers. The inner and outer CCD domains are highlighted in dark and bright blue, respectively, while B56 γ is visible in pink. Residues mutated in the following experiments are shown in sticks. A subset of these, belonging to the LxxIxE SLiM motif are coloured in green. **b, c** The effect of amino acid substitutions in STLV-1 IN **b** and LEDGF/ Δ IBD-B56 γ **c** on intasome assembly as measured by EMSA (a typical gel is shown in Supplementary Fig. 16). As the presence of LEDGF/ Δ IBD-B56 γ is crucial for STLV-1 intasome assembly in vitro, interface residues important for IN: B56 γ binding reduce or abrogate intasome formation. **d, e** Binding of His₆-STLV-1 IN with amino acid substitutions to wild-type (WT) LEDGF/ Δ IBD-B56 γ **d** and amino acid substitutions of LEDGF/ Δ IBD-B56 γ to WT IN by pull-down assays **e**. B56 γ residues comprising the novel IN SLiM-binding site do not affect binding to IN when not in the context of the supramolecular structure of the intasome. **f** Strand-transfer activity of STLV-1 IN WT and amino acid substitutions in the absence of LEDGF/ Δ IBD-B56 γ indicates all mutants are similarly or more (H2O9A) catalytically active compared to WT. **g** Strand-transfer catalytic activity of STLV-1 WT IN and amino acid substituted proteins stimulated in presence of LEDGF/ Δ IBD-B56 γ , measured by the accumulation of a concerted integration product. Stimulation of catalytic activity is impaired for IN mutants involved in both B56 γ interaction interfaces, and the P211A “kink”-inducing substitution—confirming results from EMSA (panel **b**). Bar graphs show mean values \pm SD. Individual data points of the biologically independent experiments are represented as dot plots superposed onto the bar graph. NT stands for not tested. Details pertaining to statistical significance are given in the Methods section. Source data are provided as a Source data file.

immunoprecipitation of wild type or mutant full-length B56 γ from human cells showed that Glu78, Thr81, His82, Asn83 and Arg84 are critical for interaction with endogenous PP2A binding partners BUBR1 and CHK2²¹, while not affecting holoenzyme formation (Supplementary Fig. 13).

Discussion

The PP2A regulatory subunit B56 γ interacts with and greatly stimulates the concerted strand-transfer activity of deltaretroviral INs, suggesting that the host factor helps templating the intasome assembly in vitro¹⁴. However, the stability of the active nucleoprotein complexes was insufficient to afford their purification for structural studies. Adding a non-specific DNA-binding domain to B56 γ allowed us to isolate and characterize the integration-competent species. A similar approach was used by Craigie and colleagues, who employed the archaeal Sso7d protein and fragments of LEDGF to improve solubility and activity of HIV-1 IN preparations^{19,20}. Remarkably, fusing either of these DNA-binding moieties to IN, allowed the structural determination of the HIV-1 strand-transfer complex^{12,28} and the CIC²⁰. The N-terminal region of LEDGF harbours a PWWP domain and an AT-hook, both of which display non-specific DNA-binding properties^{29,30}. Adding the AT-hook to B56 γ further stimulated concerted integration activity of STLV-1 IN, compared to B56 γ alone. However, while fusing the AT-hook of LEDGF to B56 γ improved intasome formation, the resulting nucleoprotein complexes were not sufficiently stable to allow purification for structural studies. Of note, the LEDGF-derived portion present in our B56 γ construct was not observed in our cryo-EM reconstructions. Thus, while the artificial DNA-binding moiety helped to stabilise and/or chaperone the intasome assembly, it is not involved in stable and defined interactions within the resulting nucleoprotein complex.

The tetrameric architecture of the deltaretroviral intasome closely resembles that of the prototype foamy virus (PFV), which also harbours a tetramer of IN⁶ (Supplementary Fig. 14). As was demonstrated by recent structures of lentiviral and betaretroviral intasomes the oligomeric state of IN within the intasome is dictated by the availability of the CTDs to reach their synaptic positions within the CIC^{9,11,12}. When the CCD-CTD linker length or topology does not allow for such positioning, the CTDs are provided in trans by additional IN subunits, yielding higher-order IN complexes. The STLV-1 IN CCD-CTD linker length, 19 amino acids, is similar to that of lentiviruses, which span 20–22 amino acids⁹ (Supplementary Table 3). However, while the lentiviral CCD-CTD linker adopts a compact α -helical conformation^{11,31}, the corresponding deltaretroviral region is intrinsically disordered (Supplementary Fig. 15). In the cryo-EM map, the STLV-1 IN CCD-CTD linker exists in an extended coil

conformation providing ample scope for synaptic CTD positioning in cis (Supplementary Fig. 15). While the canonical IN/CCD dimer observed in crystals could be docked directly into the cryo-EM map, the CTDs could only be docked as monomers. Thus, although four IN/CTD domains are resolved in the STLV-1 intasome structure, unlike in several other known intasome structures, they remain monomeric. We speculate that the dimers and trimers observed in IN/CTD crystals (Supplementary Fig. 10) may be relevant within viral particles prior to vDNA synthesis by reverse transcription.

B56 γ is recruited to the deltaretroviral intasome by the LxxIxE SLiM motif within the IN/CCD-CTD linker (Fig. 2a). The LxxIxE consensus sequence is known to provide a binding site for numerous endogenous PP2A interactors and substrates by targeting a conserved groove in the centre of the B56 γ subunit of the heterotrimeric PP2A^{21,26,27}. Deltaretroviruses may have acquired the LxxIxE motif in the course of their evolution to hijack this normal cellular function. Indeed, viruses often employ molecular mimicry to exploit host signalling, and usurping cellular PP2A is not uncommon³². For some pathogens, such as the Ebola virus, the interaction with PP2A became essential for viral replication³³.

It was recently shown that a subset of PP2A-B56 interactors harbour a positively charged motif, complementary to an acidic patch on B56 γ ³⁴. Deltaretroviral INs do not engage with this acidic patch but appear to have evolved to use an interface employed by other PP2A-B56 interactors, like BUBR1 and CHK2 (Supplementary Fig. 13). Future work will reveal whether BUBR1 and CHK2 interact with B56 γ in a fashion similar to deltaretroviral INs. Our data indicate at least some differences in the modes of binding, since Arg84, which is important for binding these endogenous PP2A-B56 partners, is dispensable for the association with IN (Supplementary Fig. 13). Intriguingly, B56 γ R188A appears to interact with a slightly faster-migrating species of CHK2 compared to WT B56 γ . Whether this is CHK2, in which Ser73 in the SLiM ₇₁LYSIPE₇₆ is dephosphorylated requires further investigation.

Not all PP2A-B56 substrates harbour LxxIxE motifs, and some of the SLiM-binding partners of PP2A-B56 serve to recruit the phosphatase for dephosphorylation of another macromolecule. For example, BUBR1 recruits PP2A-B56 to kinetochores to dephosphorylate KNL1 and allow mitotic progression^{35,36}, while the Ebola virus nucleoprotein NP recruits PP2A-B56 to dephosphorylate its viral transcription factor VP30³³. HTLV-1 IN forms complexes with the PP2A-B56 holoenzyme¹⁴. The structure reported here contains the regulatory subunit of PP2A and allows modelling of the supramolecular assembly with the entire heterotrimeric phosphatase (Fig. 3). Whether phosphatase activity per se plays a role in PP2A-B56 modulation of deltaretroviral infection is currently unknown. However, provided that the IN

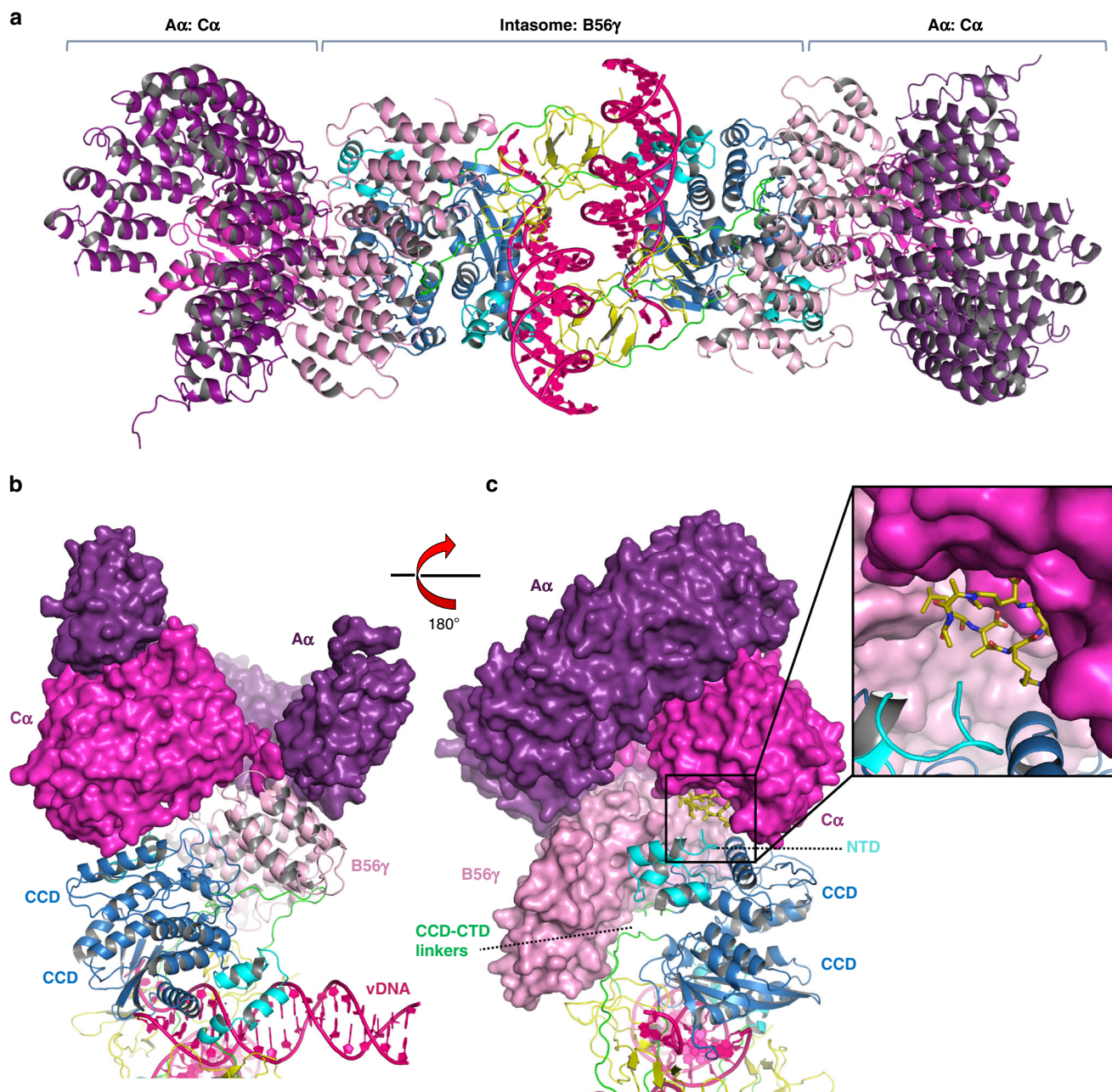


Fig. 3 Model of the complete HTLV-1 intasome: PP2A holoenzyme complex. **a** The model was created by superposition of the crystal structure of PP2A holoenzyme (PDB ID 2IAE)⁵⁹ and the cryo-EM structure of the HTLV-1 intasome: B56 γ complex using the B56 γ subunit as a common component of both structures. No inter- or intra-chain conformational changes were applied to either the intasome or PP2A structures. **b** The model reveals no clashes between atoms of the cryo-EM intasome: B56 complex solved here (shown as cartoons) and the PP2A A α and C α subunits (shown in surface representation). Interactions between A α and C α subunits of PP2A and IN may also take place in the context of the intasome: PP2A complex. **c** A rotation of the model shown in **(b)**. The holoenzyme is now shown in surface representation. A molecule of microcystin, a peptide mimetic, co-crystallised in the PP2A holoenzyme structure is illustrated in dark yellow sticks. The model suggests that in complex with the intasome, the active site of PP2A (shown in more detail in the inset) could remain open for substrate binding. The intasome is coloured by domain, with NTD in cyan, CCD in blue and CTD in yellow; vDNA is in pink and inter-domain linkers are in green. The holoenzyme is coloured in purple, magenta and salmon for the scaffold (A α), catalytic (C α) and B56 γ subunits, respectively.

NTD-CCD linker of the outer IN protomer does not occlude access to the active site, phosphorylated peptides could still be substrates of this supramolecular assembly. Recruitment of the holoenzyme greatly expands the surface of the nucleoprotein complex and may provide the deltaretroviral integration machinery an interface for chromatin-bound PP2A interaction partners^{21–24}. HTLV-1 gains access to chromatin during mitosis, when the bulk of chromatin is highly condensed. Thus, the

interaction with PP2A may allow the virus to locate chromatin loci that are bookmarked for expression soon after completion of cell division³⁷. Our structural work, presented here, will be instrumental to fully characterize the role of PP2A-B56 in HTLV-1 infection, and given that small changes in the active site between different retroviral INs significantly impact drug binding²⁸, forms the foundation to develop highly specific inhibitors of HTLV-1 integration. While this paper was under review,

Aihara and colleagues reported the cryo-EM structure of the HTLV-1 strand-transfer complex, representing the post-catalytic state of the deltaretroviral integration process. Using IN-Sso7d chimera, they obtained stable nucleoprotein complexes, which, as in our structures, are comprised of four IN and two B56 γ molecules³⁸.

Methods

Expression and purification of full-length STLV-1 IN. Expression was conducted in *E. coli* Rosetta-2 (DE3) pLacI cells (Novagen) in Terrific Broth (TB, Melford). Cells were grown to the OD₆₀₀ of 2.0 at 30 °C, followed by 30-min incubation at 18 °C and induction by 0.01% IPTG at 18 °C overnight. The pelleted cells were resuspended in 25 mM Tris-HCl pH 7.4, 1 M NaCl, 7.5 mM CHAPS, 1 mM PMSF, 10 mM imidazole and 20 μ g/mL lysozyme. Cells were then lysed by sonication and clarified by centrifugation at 50,000 \times g for 15 min at 4 °C. Following Ni-assisted IMAC, conducted as for HTLV-1 IN constructs (Supplementary Methods), cleavage of the SUMO solubility tag was conducted with either HRV 3C protease for the non-His₆-tagged product or Ulp-1 protease for the His₆-tagged product at 4 °C overnight, in presence of 5 mM DTT. IN concentration was kept below 2 mg/mL to prevent aggregation and precipitation. Ion-exchange chromatography was then conducted on a high-performance SP column (GE Healthcare, UK) following binding of the sample diluted with buffer without NaCl to achieve a final NaCl concentration of 250 mM NaCl. Peak fractions, eluted with a NaCl gradient, containing pure IN were then pooled and injected onto a Superdex 16/60 size-exclusion column (GE Healthcare), pre-equilibrated in 25 mM Tris-HCl pH 7.4, 7.5 mM CHAPS and 1 M NaCl. Fractions containing pure IN were then pooled and dialysed in a 10 K MWCO Snakeskin dialysis tubing (Life Technologies) against 20 mM BTP-HCl pH 6, 1 M NaCl, 2 mM DTT, at 4 °C overnight. Following completion of dialysis, the sample was recovered and concentrated in a 10 K MWCO ultrafiltration device (Vivaspin) to a concentration of 2 mg/mL or higher. For storage, glycerol was added to a final concentration of 10%, the sample was flash-frozen in liquid nitrogen and stored at -80 °C until needed. Note, the His₆-tagged IN proteins were less soluble than the untagged versions of the protein and yields of His₆-IN(L213A) were too low for use in binding assays.

Expression and purification of LEDGF/ Δ IBD-B56 γ . LOBSTR RIL cells³⁹ (Kerafast) were used for expression of LEDGF/ Δ IBD-B56 γ . Cells were grown in LB to an OD₆₀₀ of 0.6 and following induction with 0.01% IPTG were further incubated at 25 °C for 3 h. The temperature was then lowered to 16 °C and induction was continued overnight. Cells were resuspended in a solution containing 50 mM Tris-HCl pH 8, 1 M NaCl, 10 mM imidazole, 20 μ g/mL lysozyme and 1 mM PMSF. Following sonication to disrupt cells, the extract was clarified by centrifugation. IMAC was performed on a Ni-NTA column (GE Healthcare, UK). Thorough wash was performed in 50 mM Tris-HCl pH 8, 1 M NaCl, 10 mM imidazole. The last wash was performed in 50 mM Tris-HCl pH 8, 0.5 M NaCl, 10 mM imidazole, followed by elution with a buffer containing 25 mM Tris-HCl pH 8, 0.5 M NaCl, 200 mM imidazole. Cleavage was performed overnight with Ulp-1 SUMO protease in presence of 5 mM DTT. The protein was diluted with salt-free buffer to achieve NaCl concentration of 125 mM, injected into an HP Q column (GE Healthcare, UK) and eluted with a gradient of 0.15–0.5 M NaCl. Fractions containing LEDGF/ Δ IBD-B56 γ were pooled and the protein was polished by size-exclusion chromatography through a Superdex S200 16/60 gel-filtration column (GE Healthcare, UK) operated in 300 mM NaCl, 25 mM Tris-HCl pH 8. Fractions containing pure LEDGF/ Δ IBD-B56 γ were supplemented with 2 mM DTT and concentrated to 20 mg/mL using a 30-KDa MWCO ultrafiltration device (Vivaspin). Protein, supplemented with 10% glycerol, was flash-frozen in liquid nitrogen and stored at -80 °C until further use.

STLV-1 strand-transfer activity assays. Assays were conducted using purified recombinant STLV-1 IN and the vDNA LTR oligonucleotide mimics (Supplementary Table 4) were annealed in 100 mM Tris-HCl pH 7.4, 400 mM NaCl. The optimised reaction conditions were: 25 mM BTP-HCl pH 6, 0.8 μ M STLV-1 IN, 2 μ M vDNA, 60 mM NaCl, 13.28 mM DTT, 10 mM MgCl₂, 10 μ M ZnCl₂. After addition of vDNA, the reaction was incubated at 37 °C for 10 min. Where IN: LEDGF/ Δ IBD-B56 γ was used, 0.2 μ M IN was pre-incubated with 1:2 ratio of LEDGF/ Δ IBD-B56 γ to STLV-1 IN for 30 min at 4 °C prior to incubation with vDNA. Following a co-incubation with supercoiled target DNA (s.c. tDNA) for 30 min at 37 °C, samples were deproteinised by addition of 15 μ L of a solution containing 5% SDS and 250 mM EDTA followed by 1.5 μ L of 20 mg/mL proteinase K (Roche). Samples were then incubated at 37 °C for a further 30 min and DNA was precipitated by adding 1 μ L of glycogen (20 mg/mL, Roche) and 400 μ L of ice-cold ethanol. The resuspended DNA was then analysed on a 1.5% agarose gel stained with ethidium bromide.

Electrophoretic mobility shift assays (EMSA). Five μ L STLV-1 IN (1.6 mg/mL in 20 mM BTP-HCl pH 6, 1 M NaCl, 2 mM DTT) was mixed with 5 μ L 3.84 mg/mL LEDGF/ Δ IBD-B56 γ and diluted with 20 μ L buffer to yield a final NaCl concentration of 200 mM. Following incubation at 4 °C for 30 min, the samples were

supplemented with 0.5 μ L 20 μ M Atto680-labelled vDNA (Supplementary Table 4), and the reaction volume increased to yield a final NaCl concentration of 60 mM. The reaction was placed at 37 °C for 10 min, then NaCl concentration was increased to 1.2 M and allowed to equilibrate at room temperature. Samples, supplemented with 10 μ g/mL heparin, were separated on a 3% low melting point agarose gel containing 10 μ g/mL heparin¹⁹. Densitometry of bands corresponding for the intasome was carried out in ImageJ. Measurements from at least three independent experiments were taken, standard deviations and *p*-values were calculated in Prism 8.

Pull-down assays. Forty microlitres of His-select IMAC resin (Sigma) were washed with 500 μ L of ice-cold pull-down buffer containing: 25 mM Tris-HCl pH 7.4, 150 mM NaCl, 20 mM imidazole, 0.5% CHAPS. Resin was pelleted, the supernatant removed and replaced with 750 μ L of pull-down buffer supplemented with 10 μ g of BSA. Contents were mixed by inversion, and 10 μ g of 6xHis-tagged STLV-1 MarB43 IN (or mutants thereof) were added. Contents were mixed by inversion, and 20 μ g of LEDGF/ Δ IBD-B56 γ (or mutants thereof) were added. Samples were then incubated, tumbling, at 4 °C for 2 h. Resin was pelleted at 400 \times g for 5 min. Supernatant was replaced with 1 mL of ice-cold pull-down buffer, samples mixed by inversion and the latter wash was performed five times. The last centrifugation step was conducted at 400 g for 5 min. The supernatant was removed carefully with a gel-loading tip. Twenty microlitres of 1.5 \times concentrated SDS-PAGE loading dye (containing imidazole, urea and DTT) was added to each sample and the sample placed in 100 °C heating block for 2 min. Samples were centrifuged at 16,000 \times g for 1 min and 10 μ L of the supernatant was loaded onto an 11% SDS-PAGE gel.

Assembly and purification of STLV-1 intasome:LEDGF/ Δ IBD-B56 γ . The STLV-1 IN: LEDGF/ Δ IBD-B56 γ complex was first assembled by mixing equimolar (0.03 mmol) quantities of IN and LEDGF/ Δ IBD-B56 γ and dialysing overnight at 4 °C against 0.5 L of 25 mM Tris-HCl pH 7.4, 200 mM NaCl, 2 mM DTT. We have previously found (see section Crystallisation of HTLV-1 IN (200-297): B56 γ in Supplementary Methods) that this condition promotes IN: LEDGF/ Δ IBD-B56 γ complex formation. This sample was then concentrated at 4 °C, 1935 \times g in a 30-KDa MWCO ultrafiltration device (Vivaspin) to a concentration of 0.2 mM. A mixture containing 0.7 mL 20 mM BTP-HCl pH 6, 10 mM CaCl₂, 10 mM DTT, 10 μ M ZnCl₂ and 25 μ M STLV-1 U5 S30 double-stranded vDNA (Supplementary Table 4) was placed in a heat block set to 37 °C and incubated for 10 min. The previously prepared IN: LEDGF/ Δ IBD-B56 γ complex was then added, mixed by gently flicking the tube and the tube placed back in the heat block for 10 min. Upon addition of the protein complex and during the course of incubation dense, white precipitate appeared. Following the incubation, the precipitate was dissolved by addition of NaCl to a final concentration of 1.2 M, gentle up-and-down mixing, and a further 15 min incubation at room temperature. Increasing NaCl concentration allowed for complete dissolution of the precipitate and recovery of assembled nucleoprotein complex. The sample was then immediately loaded onto an S200 10/300 Increase size-exclusion column (GE Healthcare, UK). For samples prepared for negative-stain observations, the size-exclusion mobile phase was 20 mM BTP-HCl pH 6, 1.2 M NaCl. For cryo-EM preparations, the size-exclusion mobile phase was 20 mM BTP-HCl pH 6, 0.3 M NaCl. Peak fractions were pooled and tested for integration activity in the presence of 10 mM MgCl₂ and 300 ng target DNA (supercoiled pGEM-9Zf(-)), as well as by SDS-PAGE. Fractions corresponding to the highest strand-transfer activity were used for negative-stain and cryo-EM grid preparation.

Negative-stain imaging and data processing. Four-microlitre drops of freshly assembled and purified STLV-1 intasomes were spotted on carbon-coated 300-mesh copper grids (EM Resolutions, catalogue #C300Cu), which had been glow-discharged for 30 s at 45 mA using an Emitech K100X instrument (EMS) and allowed to bind for 1 min. Excess sample was blotted, and absorbed particles were stained with 2% uranyl acetate. Grids were imaged on a Tecnai G2 Spirit LaB6 transmission 120-kV electron microscope (Thermo Fisher Scientific) with an Ultrascan-1000 camera (Gatan) at \times 30,000 magnification, resulting in a magnified pixel size of 3.45 Å. A total of 152 micrographs were taken, from which 22,000 particles were picked using EMAN2 Boxer⁴⁰. 2D classification was done in Relion-2⁴¹ and 8790 particles were used for ab initio 3D reconstruction and homogenous refinement.

Cryo-EM grid preparation and data collection. C-flat holey carbon gold grids were obtained from Electron Microscopy Sciences (catalogue #CF-1.2/1.3-4Au). These were used within 6 months of purchase without glow discharging or plasma cleaning. UltraAuFoil R 1.2/1.3 grids⁴² (Electron Microscopy Sciences, catalogue #Q350AR13A) were freshly-coated with graphene oxide (Sigma-Aldrich, catalogue #763705)⁴³. Four microlitres of freshly prepared intasome (A₂₆₀ ~ 1.5, corresponding to ~2.3 μ M nucleoprotein complex) was applied on C-flat or graphene oxide-coated UltraAuFoil grids. The grids, incubated for 1 min at 22 °C and 95% humidity, were blotted for 2–3 s prior to plunge-freezing in liquid ethane using a VitroBot Mark IV instrument (Thermo Fisher Scientific). Data were collected on Titan Krios electron microscope operating at 300 kV with a Falcon III direct

electron detector in counting mode (Thermo Fisher Scientific). A pixel size of 1.09 Å and defocus range of -1.6 to -3.6 μm was used for the data collections. A total electron exposure of $34 \text{ e}/\text{Å}^2$ was fractionated across 30 movie frames over a 60 s exposure time. A total of 8088 and 8949 movies were recorded from open hole C-flat (OH dataset) and graphene oxide supported UltrAuFoil (GO dataset) grids, respectively, with EPU 1.9.0 software (Thermo Fisher Scientific).

Single-particle image processing and 3D reconstruction. Micrograph movie frames were aligned and summed with dose weighting applied as implemented in MotionCorr²⁴, and the contrast transfer function (CTF) parameters were estimated from the frame sums using Gctf-v1.06⁴⁵. Following removal of images with evidence of crystalline ice contamination and/or those lacking graphene oxide, 8022 (OH dataset) and 8049 (GO dataset) aligned micrographs were retained for particle picking and further image processing. A small subset of micrographs was picked manually with EMAN2 Boxer⁴⁰ and subjected to reference-free classification in Relion-2⁴¹ to generate initial 2D class averages (Supplementary Fig. 4). These were used as templates for picking the entire datasets using Gautomatch v0.56 (<http://www.mrc-lmb.cam.ac.uk/kzhang/>), resulting in the initial subsets of 2,198,454 (OH dataset) and 2,157,654 (GO dataset) particles. The particles extracted in Relion-3.0⁴⁶, binned by a factor of 2, were subjected to two rounds of reference-free 2D classification in CryoSPARC-2⁴⁷. Particles belonging to well-defined 2D classes (599,700 and 493,665 particles for OH and GO datasets, respectively) were subjected to 45 cycles of 3D classification into 17 (OH) or 13 (GO) classes in Relion-3.0 without imposing symmetry, with an initial model generated in CryoSPARC-2. The procedure yielded a single high-resolution class from each dataset. Particles belonging to the best 3D classes (94,517 and 67,397 from OH and GO dataset, respectively) were re-extracted as full-sized images. 3D reconstructions generated from the individual datasets resulted in highly anisotropic maps due to severe preferential orientations of the single particles (Supplementary Fig. 6). Since 3D-FSC analysis⁴⁸ indicated favourable complementarity of the data (Supplementary Fig. 6), the datasets were merged and refined as separate optics groups in Relion-3.1. 3D reconstruction, followed by Bayesian polishing, per-particle defocus and beam tilt refinement, as implemented in Relion-3.1, resulted in the final map with minimal anisotropy. 3D-FSC sphericity index of the final map was 0.967 (Supplementary Fig. 6). Gold-standard Fourier shell correlation (FSC) = 0.143 criterion^{49,50} was used to estimate resolutions of the 3D reconstructions (Supplementary Table 5). Local resolution of the cryo-EM map was estimated using Blocres from the Bsoft software package⁵¹.

Integrative model building and refinement. The quality of the cryo-EM map was marginally improved using Resolve density modification procedure⁵² implemented in Phenix 1.18-3845⁵³, which increased the estimated resolution of the reconstruction by 0.15 Å. Density modification was performed under default parameters, using half-maps and macromolecular sequence as inputs. Alternatively, cryo-EM map was sharpened using a global B factor (-143 Å^{-2} , determined automatically) or locally filtered using post-processing procedures implemented in Relion-3.1. Initially, X-ray crystal structures were docked into resulting cryo-EM maps as rigid bodies in Chimera⁵⁴. A homology model of the HTLV-1 IN/NTD was generated by SWISS-MODEL server²⁵. Ab initio building residues not present in docked models but resolved in the cryo-EM density and manual refitting of docked models was conducted in Coot⁵⁵. The globally and locally sharpened maps and density modified map were used to guide model building. The initial model, comprising chains A, B, C, K and L, was subjected to molecular dynamics structural fitting using Namdinator⁵⁶. This model was further adjusted in Coot before the model was duplicated to form chains D, E, F, M and N which were docked in place using Chimera and rigid body fitted in Coot. The model was manually checked again for clashes between the NCS chains before final real-space refinement using Phenix version 1.18-3845 and the density modified map, implementing secondary structure and base-pair/base stacking definitions based on the model, metal bond restraints and NCS constraints for the two halves of the symmetric nucleoprotein assembly. Quality of the final atomistic model was assessed with MolProbity⁵⁷ and EMRinger⁵⁸ (Supplementary Table 5).

For cloning of expression constructs used in this study, X-ray crystallographic analysis of IN and IN: B56 complex and Flag-immunoprecipitation methods, please see the Supplementary Methods.

Statistics and reproducibility. Statistical significance for Fig. 2b–g was calculated using the unpaired *t*-test with Welch's correction; *p*-values are two-sided. For data in Fig. 2b, the number of data points (*n*) and the calculated *p*-values (*p*) are from left to right: $n = 5$, $p = n/a$; $n = 3$, $p < 0.0001$; $n = 5$, $p = 0.0002$; $n = 5$, $p < 0.0001$; $n = 3$, $p < 0.0001$; $n = 3$, $p = 0.0013$; $n = 4$, $p = 0.025$; $n = 4$, $p = 0.0005$. For Fig. 2c these are: $n = 4$, $p = n/a$; $n = 4$, $p = 0.0007$; $n = 3$, $p = 0.0016$; $n = 4$, $p = 0.137$; $n = 3$, $p = 0.0481$; $n = 3$, $p = 0.0024$. For Fig. 2d these are: $n = 3$, $p = n/a$; $n = n/a$, $p = n/a$; $n = 3$, $p = 0.0384$; $n = 3$, $p = 0.0072$; $n = 3$, $p < 0.0001$; $n = 3$, $p = 0.0010$; $n = 3$, $p = 0.0011$; $n = 3$, $p = 0.0021$. For Fig. 2e these are: $n = 3$, $p = n/a$; $n = 3$, $p = 0.1198$; $n = 3$, $p = 0.1115$; $n = 3$, $p = 0.1069$; $n = 3$, $p = 0.0006$; $n = 3$, $p = 0.0668$. For Fig. 2f these are: $n = 3$, $p = n/a$; $n = 3$, $p = 0.2178$; $n = 3$, $p = 0.4927$; $n = 3$, $p = 0.2906$; $n = 3$, $p = 0.2147$; $n = 3$, $p = 0.0949$; $n = 3$, $p = 0.3369$; $n = 3$, $p = n/a$. For Fig. 2g these are: $n = 3$, $p = n/a$; $n = 3$, $p = 0.0012$; $n = 3$, $p = 0.0019$;

$n = 3$, $p = 0.0020$; $n = 3$, $p = 0.0007$; $n = 3$, $p = 0.0019$; $n = 3$, $p = 0.0010$; $n = 3$, $p = 0.0015$.

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

Data availability

The crystal structures have been deposited with the Protein Data Bank and are available under the following identifiers: HTLV-2/CCD-Mg²⁺ (dimeric form): 6QBV; HTLV-2/CCD-Mg²⁺ (trimeric form): 6QBT; HTLV-2/CCD-Ca²⁺ (dimeric form): 6QBW; HTLV-1/CTD: 6TJU; HTLV-1 IN(200-297)-B56y: 6TOQ. Raw diffraction images are available upon request. The cryo-EM structure has been deposited with the Protein Data Bank and EMDB and are available under the following identifiers 6Z2Y and EMD-11052. The authors declare that all other data supporting the findings of this study are available within the paper and its supplementary information files. Source data are provided with this paper.

Received: 21 May 2020; Accepted: 15 September 2020;

Published online: 07 October 2020

References

- Mahieux, R. & Gessain, A. Adult T-cell leukemia/lymphoma and HTLV-1. *Curr. Hematol. Malig. Rep.* **2**, 257–264 (2007).
- 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).
- Martin, F., Taylor, G. P. & Jacobson, S. Inflammatory manifestations of HTLV-1 and their therapeutic options. *Expert Rev. Clin. Immunol.* **10**, 1531–1546 (2014).
- Martin, F., Tagaya, Y. & Gallo, R. Time to eradicate HTLV-1: an open letter to WHO. *Lancet* **391**, 1893–1894 (2018).
- Lesbats, P., Engelman, A. N. & Cherepanov, P. Retroviral DNA integration. *Chem. Rev.* **116**, 12730–12757 (2016).
- 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).
- 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).
- Hare, S., Maertens, G. N. & Cherepanov, P. 3'-processing and strand transfer catalysed by retroviral integrase in crystallo. *EMBO J.* **31**, 3020–3028 (2012).
- Ballandras-Colas, A. et al. Cryo-EM reveals a novel octameric integrase structure for betaretroviral intasome function. *Nature* **530**, 358–361 (2016).
- Yin, Z. et al. Crystal structure of the Rous sarcoma virus intasome. *Nature* **530**, 362–366 (2016).
- Ballandras-Colas, A. et al. A supramolecular assembly mediates lentiviral DNA integration. *Science* **355**, 93–95 (2017).
- Passos, D. O. et al. Cryo-EM structures and atomic model of the HIV-1 strand transfer complex intasome. *Science* **355**, 89–92 (2017).
- Hare, S. & Cherepanov, P. The Interaction Between Lentiviral Integrase and LEDGF: Structural and Functional Insights. *Viruses* **1**, 780–801 (2009).
- Maertens, G. N. B'-protein phosphatase 2A is a functional binding partner of delta-retroviral integrase. *Nucleic Acids Res.* **44**, 364–376 (2016).
- Sharma, A. et al. BET proteins promote efficient murine leukemia virus integration at transcription start sites. *Proc. Natl Acad. Sci. USA* **110**, 12036–12041 (2013).
- 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).
- 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).
- 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).
- Li, M., Jurado, K. A., Lin, S., Engelman, A. & Craigie, R. Engineered hyperactive integrase for concerted HIV-1 DNA integration. *PLoS ONE* **9**, e105078 (2014).
- Li, M. et al. A peptide derived from lens epithelium-derived growth factor stimulates HIV-1 DNA integration and facilitates intasome structural studies. *J. Mol. Biol.* **432**, 2055–2066 (2020).
- Hertz, E. P. T. et al. A conserved motif provides binding specificity to the PP2A-B56 phosphatase. *Mol. Cell* **63**, 686–695 (2016).

22. Li, H. H., Cai, X., Shouse, G. P., Piluso, L. G. & Liu, X. A specific PP2A regulatory subunit, B56gamma, mediates DNA damage-induced dephosphorylation of p53 at Thr55. *EMBO J.* **26**, 402–411 (2007).
23. Li, X., Nan, A., Xiao, Y., Chen, Y. & Lai, Y. PP2A-B56 complex is involved in dephosphorylation of gamma-H2AX in the repair process of CPT-induced DNA double-strand breaks. *Toxicology* **331**, 57–65 (2015).
24. Wu, C. G. et al. PP2A-B' holoenzyme substrate recognition, regulation and role in cytokinesis. *Cell Disco.* **3**, 17027 (2017).
25. Biasini, M. et al. SWISS-MODEL: modelling protein tertiary and quaternary structure using evolutionary information. *Nucleic Acids Res.* **42**, W252–W258 (2014).
26. 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).
27. 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).
28. Passos, D. O. et al. Structural basis for strand-transfer inhibitor binding to HIV intasomes. *Science* **367**, 810–814 (2020).
29. Turlure, F., Maertens, G., Rahman, S., Cherepanov, P. & Engelman, A. A tripartite DNA-binding element, comprised of the nuclear localization signal and two AT-hook motifs, mediates the association of LEDGF/p75 with chromatin in vivo. *Nucleic Acids Res.* **34**, 1653–1665 (2006).
30. van Nuland, R. et al. Nucleosomal DNA binding drives the recognition of H3K36-methylated nucleosomes by the PSIP1-PWWP domain. *Epigenetics Chromatin* **6**, 12 (2013).
31. 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).
32. 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).
33. 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 (2018).
34. Wang, X. et al. A dynamic charge-charge interaction modulates PP2A: B56 substrate recruitment. *Elife* **9**, <https://doi.org/10.7554/eLife.55966> (2020).
35. Espert, A. et al. PP2A-B56 opposes Mps1 phosphorylation of Knl1 and thereby promotes spindle assembly checkpoint silencing. *J. Cell Biol.* **206**, 833–842 (2014).
36. Vallardi, G., Allan, L. A., Crozier, L. & Saurin, A. T. Division of labour between PP2A-B56 isoforms at the centromere and kinetochore. *Elife* **8**, <https://doi.org/10.7554/eLife.42619> (2019).
37. Xing, H., Vanderford, N. L. & Sarge, K. D. The TBP-PP2A mitotic complex bookmarks genes by preventing condensin action. *Nat. Cell Biol.* **10**, 1318–1323 (2008).
38. Bhatt, V. et al. Structural basis of host protein hijacking in human T-cell leukemia virus integration. *Nat. Commun.* **11**, 3121 (2020).
39. Andersen, K. R., Leksa, N. C. & Schwartz, T. U. Optimized E. coli expression strain LOBSTR eliminates common contaminants from His-tag purification. *Proteins* **81**, 1857–1861 (2013).
40. Ludtke, S. J. 3-D structures of macromolecules using single-particle analysis in EMAN. *Methods Mol. Biol.* **673**, 157–173 (2010).
41. Scheres, S. H. RELION: implementation of a Bayesian approach to cryo-EM structure determination. *J. Struct. Biol.* **180**, 519–530 (2012).
42. Russo, C. J. & Passmore, L. A. Ultrastable gold substrates: properties of a support for high-resolution electron cryomicroscopy of biological specimens. *J. Struct. Biol.* **193**, 33–44 (2016).
43. Lancey, C. et al. Structure of the processive human Pol delta holoenzyme. *Nat. Commun.* **11**, 1109 (2020).
44. Zheng, S. Q. et al. MotionCor2: anisotropic correction of beam-induced motion for improved cryo-electron microscopy. *Nat. Methods* **14**, 331–332 (2017).
45. Zhang, K. Gctf: Real-time CTF determination and correction. *J. Struct. Biol.* **193**, 1–12 (2016).
46. 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).
47. Punjani, A., Rubinstein, J. L., Fleet, D. J. & Brubaker, M. A. cryoSPARC: algorithms for rapid unsupervised cryo-EM structure determination. *Nat. Methods* **14**, 290–296 (2017).
48. Tan, Y. Z. et al. Addressing preferred specimen orientation in single-particle cryo-EM through tilting. *Nat. Methods* **14**, 793–796 (2017).
49. Rosenthal, P. B. & Henderson, R. Optimal determination of particle orientation, absolute hand, and contrast loss in single-particle electron cryomicroscopy. *J. Mol. Biol.* **333**, 721–745 (2003).
50. Scheres, S. H. & Chen, S. Prevention of overfitting in cryo-EM structure determination. *Nat. Methods* **9**, 853–854 (2012).
51. Heymann, J. B. Guidelines for using Bsoft for high resolution reconstruction and validation of biomolecular structures from electron micrographs. *Protein Sci.* **27**, 159–171 (2018).
52. Terwilliger, T. C., Ludtke, S. J., Read, R. J., Adams, P. D. & Afonine, P. V. Improvement of cryo-EM maps by density modification. *Nat. Methods* **9**, 923–927 (2012).
53. Adams, P. D. et al. PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D. Biol. Crystallogr.* **66**, 213–221 (2010).
54. Pettersen, E. F. et al. UCSF Chimera—a visualization system for exploratory research and analysis. *J. Comput. Chem.* **25**, 1605–1612 (2004).
55. Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallogr. D. Biol. Crystallogr.* **60**, 2126–2132 (2004).
56. Kidmose, R. T. et al. Namdinator—automatic molecular dynamics flexible fitting of structural models into cryo-EM and crystallography experimental maps. *IUCr* **6**, 526–531 (2019).
57. Williams, C. J. et al. MolProbity: more and better reference data for improved all-atom structure validation. *Protein Sci.* **27**, 293–315 (2018).
58. Barad, B. A. et al. EMRinger: side chain-directed model and map validation for 3D cryo-electron microscopy. *Nat. Methods* **12**, 943–946 (2015).
59. Cho, U. S. & Xu, W. Crystal structure of a protein phosphatase 2A heterotrimeric holoenzyme. *Nature* **445**, 53–57 (2007).

Acknowledgements

We thank Dr R. Carzaniga and L. Collinson for the maintenance of Vitrobot and Tecnai G2 microscope and user training; Drs A. Purkiss, P. Walker and M. Oliveira for computer and software support; Drs C. McAuley, P. Romano, D. Hall and J. Beale for their help and assistance at the Diamond Light Source, Dr M. Morgan (Imperial College London) for expert help with in-house crystallisation screening and data collection, N. Cook (Francis Crick Institute) for expert help with cryo-EM grid vitrification, and F. Martino (Francis Crick Institute) for generous advice on preparation of graphene oxide-coated grids. We are also grateful to Dr. A. Engelman (Dana-Farber Cancer Institute) for helpful comments and critical reading of the manuscript and Dr. A.L.B. Ambrosio (Laboratório Nacional de Biotecnologia, Brazil) for the generous gift of pET28a-SUMO. This work is supported by the Wellcome Trust (Investigator Award to G.N.M., 107005/Z/15Z) and the Royal Society (RG120032, to G.N.M.). Work in P.C. laboratory is supported by the Francis Crick Institute, which receives its core funding from Cancer Research UK (FC001061), the UK Medical Research Council (FC001061), and the Wellcome Trust (FC001061). This article is independent research funded by the National Institute for Health Research (NIHR) Imperial Biomedical Research Centre (BRC). The views expressed in this publication are those of the authors and not necessarily those of the NHS, the National Institute for Health Research or the Department of Health.

Author contributions

G.N.M. discovered how to assemble STL-1 intasomes, assembled, purified and prepared samples for negative-stain EM; G.N.M. and P.C. collected and analysed negative-stain EM data; G.N.M. conducted the Flag-IP experiments; M.S.B. purified all wild-type and mutant full-length proteins, assembled intasomes for biochemical analysis and cryo-EM data collection; M.S.B. purified, crystallised, solved and refined the X-ray structures of the individual IN domains; M.S.B. and P.C. analysed the cryo-EM data; M.S.B., P.C. and V.E.P. refined the atomistic model; J.J.M. optimized the assembly of the IN(200-297): B56γ(11–380) complex, conducted crystallisation screening, optimization, X-ray data collection; M.S.B., J.J.M. and G.N.M. refined the structure of the IN: B56γ complex; Z.H. prepared graphene oxide carbon and prepared the grids for cryo-EM data collection; A.N. collected the cryo-EM data; M.S.B., J.J.M., P.C. and G.N.M. prepared the manuscript with contributions of all authors.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information is available for this paper at <https://doi.org/10.1038/s41467-020-18874-y>.

Correspondence and requests for materials should be addressed to G.N.M.

Peer review information *Nature Communications* thanks Gert Weber and the other, anonymous, reviewer(s) 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