

REPORT

STRUCTURAL BIOLOGY

Cryo-EM structures and atomic model of the HIV-1 strand transfer complex intasome

Dario Oliveira Passos,^{1*} Min Li,^{2*} Renbin Yang,² Stephanie V. Rebersburg,³ Rodolfo Ghirlando,² Youngmin Jeon,¹ Nikoloz Shkriabai,³ Mamuka Kvaratskhelia,³ Robert Craigie,² Dmitry Lyumkis^{1†}

Like all retroviruses, HIV-1 irreversibly inserts a viral DNA (vDNA) copy of its RNA genome into host target DNA (tDNA). The intasome, a higher-order nucleoprotein complex composed of viral integrase (IN) and the ends of linear vDNA, mediates integration. Productive integration into host chromatin results in the formation of the strand transfer complex (STC) containing catalytically joined vDNA and tDNA. HIV-1 intasomes have been refractory to high-resolution structural studies. We used a soluble IN fusion protein to facilitate structural studies, through which we present a high-resolution cryo-electron microscopy (cryo-EM) structure of the core tetrameric HIV-1 STC and a higher-order form that adopts carboxyl-terminal domain rearrangements. The distinct STC structures highlight how HIV-1 can use the common retroviral intasome core architecture to accommodate different IN domain modules for assembly.

Catalytic integration of a viral DNA (vDNA) copy of an RNA genome into host target DNA (tDNA) represents the hallmark characteristic of all retroviruses, including HIV-1. Integration establishes a permanent infection in host cells and enables the newly inserted provirus to be replicated and transcribed in parallel with other genes of the host organism (1). This critical step in the HIV-1 replication cycle represents one of the underlying difficulties in combating the HIV/AIDS pandemic. Integration is catalyzed by the viral integrase (IN) protein, which oligomerizes into a higher-order stable synaptic complex (SSC) containing the two vDNA ends. Following cleavage of the GT dinucleotide from both 3' vDNA ends and nuclear entry, cleaved SSCs engage tDNA and catalyze irreversible DNA strand transfer into host chromatin to form the strand transfer complex (STC) (2). Both retroviral SSCs and the post-catalytic STCs are collectively called intasomes. Clinically exploited HIV-1 IN strand transfer inhibitors (INSTIs) selectively bind cleaved SSCs and interfere with the formation of the STC. Therefore, high-resolution structures of key integration intermediate nucleoprotein complexes are required to further our understanding of the mechanisms of action of INSTIs and the evolution

of drug-resistant HIV-1 phenotypes (3). The mechanism of HIV-1 DNA integration has been extensively studied at the biochemical and

cellular level, but progress with structural studies of nucleoprotein reaction intermediates has been slow; only structures of domains of HIV-1 IN are currently available (4–8), although intasome structures have been determined for related retroviruses (9–12) and predicted for HIV-1 through homology modeling (13, 14).

Structural studies of HIV-1 intasomes have been challenging, owing to the tendency of HIV-1 IN protein and assembled intasomes to aggregate. Fusion of the DNA binding protein Sso7d to the N terminus of IN results in a protein that is hyperactive in vitro, has markedly improved solubility properties, and retains activity in vivo when incorporated into HIV-1 virions (15). We therefore used Sso7d-IN to assemble HIV-1 intasomes for structural studies. STC intasomes were assembled on branched DNA, mimicking the product of DNA integration (fig. S1A) using the strategy previously described for prototype foamy virus (PFV) (16) and Rous sarcoma virus (RSV) (12) intasomes. HIV-1 intasomes were first purified by Ni-affinity and anion-exchange chromatography (fig. S1B). Analytical ultracentrifugation after anion exchange chromatography indicated the presence of the tetrameric STC as well as larger discrete species (fig. S2), in agreement with previous studies (17–19). An additional gel-filtration step before cryo-electron microscopy (cryo-EM) structural analysis yielded a preparation that was mainly tetrameric but also included larger species, as evidenced by the broad and asymmetric peak shape (fig. S1C).

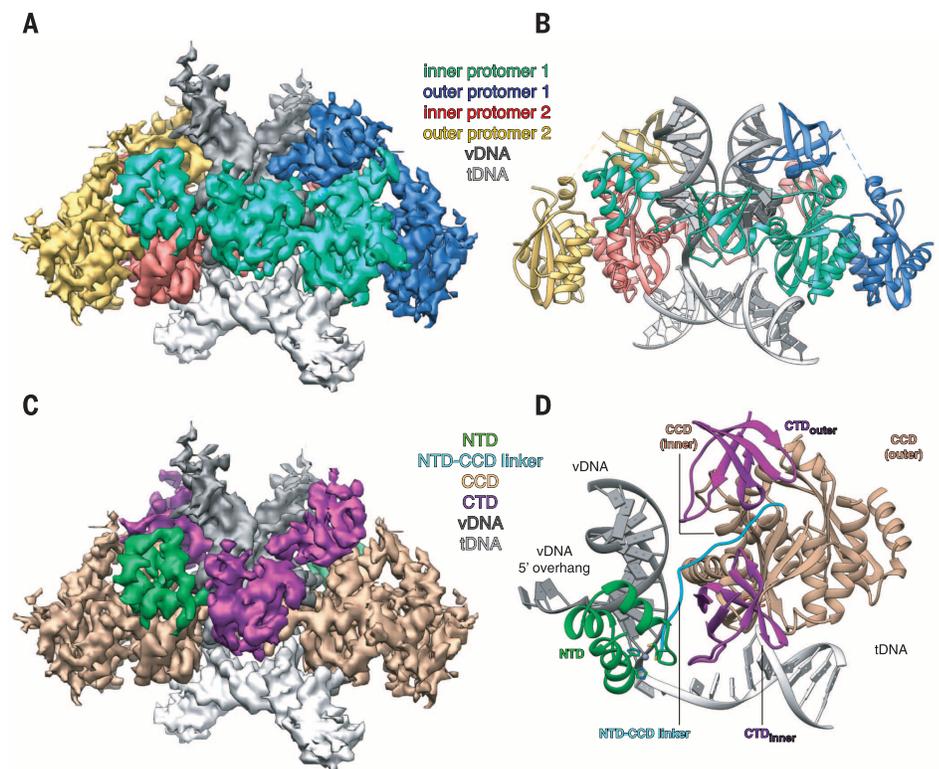


Fig. 1. HIV-1 STC intasome structure. (A) Cryo-EM reconstruction of the STC, segmented by IN protomers (red, green, yellow, and blue) and product DNA components (dark and light gray). (B) Atomic model derived from the cryo-EM density, colored as in (A). (C) Segmented cryo-EM density and (D) asymmetric subunit of the atomic model, colored by IN domain: NTD, green; CCD, beige; NTD-CCD linker, blue; CTD, purple.

¹Laboratory of Genetics and Helmsley Center for Genomic Medicine, The Salk Institute for Biological Studies, 10010 North Torrey Pines Road, La Jolla, CA 92037, USA. ²Laboratory of Molecular Biology, National Institute of Diabetes and Digestive and Kidney Diseases, National Institutes of Health, Bethesda, MD 20892, USA. ³Center for Retrovirus Research and College of Pharmacy, The Ohio State University, Columbus, OH 43210, USA. *These authors contributed equally to this work. †Corresponding author. Email: dlyumkis@salk.edu

Tetrameric HIV-1 STCs are relatively small by cryo-EM standards (~200 kDa) and require high salt and glycerol to prevent aggregation, factors that negatively affect image contrast of individual particles. To overcome these problems, we employed a high-dose imaging strategy and an exposure filter that accounts for the effects of radiation damage while maximizing low-frequency contrast (20). Single-particle classification and refinement of exposure-filtered images produced a density map resolved to ~3.5 to 4.5 Å, with the highest-resolution information characterizing the STC core, in and around the active site (figs. S3 and S4). This enabled derivation of a molecular model of an HIV-1 STC, which contained four IN protomers arranged with twofold symmetry around the product of DNA strand transfer (Fig. 1 and table S1).

The tetrameric HIV-1 STC intasome is a dimer of dimers with a similar overall architecture to PFV intasomes (Fig. 1, A and B). Each protomer contains an N-terminal domain (NTD), a catalytic core domain (CCD), and a C-terminal domain (CTD). The inner protomers wrap their three functionally relevant domains around a pair of vDNA ends and dock onto tDNA, bringing two vDNA 3'-OH groups into proximity to catalyze concerted integration and form the STC. The inner protomers also make most of the contacts with vDNA and tDNA. The outer IN CTD (CTD_{outer}) adopts a retracted configuration in the HIV-1 intasome, contributing partially to vDNA binding and positioning itself in proximity to the inner CTD (CTD_{inner}) (Fig. 1, C and D). The outer NTDs, as well as all Sso7d fusion domains, are disordered in the cryo-EM density.

Retroviral intasomes recognize and cut target sites with a characteristic 4- to 6-bp spacing, generating equivalently sized target-site duplications (TSDs) flanking either end of the integrated proviral DNA (2). To a large extent, the TSD sizes map onto the retroviral phylogenetic tree (21), although the precise TSD spacing can differ within an individual genus (22). In all of the available STC structures, the target DNA is substantially distorted from B form (fig. S5), resulting in a 4-bp TSD for PFV, which has the shortest distance between the active sites, and 5 and 6 bps for HIV-1 and RSV, respectively, which have a longer spacing.

The STC model substantiates and rationalizes much of the existing *in vitro* and *in vivo* data pertaining to HIV-1 IN residues involved in function, inhibitor binding, and mechanisms of drug resistance. In this regard, Fig. 2 and table S2 present a comprehensive analysis of predicted electrostatic protein-DNA and interdomain interactions within these core components. Notably, the table includes all of the residues that were experimentally shown to affect IN function, but it also provides additional details that were not captured by a homology model (13). Residues that interact with vDNA are distributed throughout the IN protein, whereas residues interacting with tDNA are largely clustered in the CCD (Fig. 2A). Multiple residues are also involved in interdomain interactions (Fig. 2B). Several specific DNA-binding residues (displayed in Fig. 2C and fig. S6) deserve particular mention. A cluster of basic residues—including K46, which was not identified by a homology model, and K156, K159, and K160—is

inserted into the vDNA minor groove next to the active site. The relevance of K46 is addressed below, whereas K156, K159, and K160 play various roles in vDNA binding, sequence specificity, and catalysis (13, 23, 24). R231, the only non-CCD residue that strongly interacts with tDNA (but also with vDNA through the other protomer), has previously been shown to affect nucleotide preferences within the target site, although the magnitude of this effect in HIV-1 IN mutants is considerably lower than analogous changes in PFV (25). The weaker interaction between R231 of HIV-1 with target-site nucleotides—as compared with R329 of PFV, which contains a longer loop that can accommodate subtle structural changes—helps to explain the phenotype. Substitutions of HIV-1 IN S119 (25), similarly to analogous changes in RSV and PFV INs (10, 26), alter target-site nucleotide specificity by perturbing interactions with tDNA. The model also provides important guidance for rationally improving clinically relevant inhibitors. Specifically, several residues around the vicinity of the active site, especially R231, are positioned differently in HIV-1 compared with PFV, which has been used as a model system to study mechanism of INSTI action (fig. S7). Slight differences in the active site can be exploited to facilitate rational inhibitor design. Collectively, the current model provides a composite platform for both understanding IN function and elucidating modes of action of INSTIs.

To gain a more thorough understanding of the heterogeneous STC data and improve regions of density outside of the core subunits, we employed a multistep classification approach that revealed

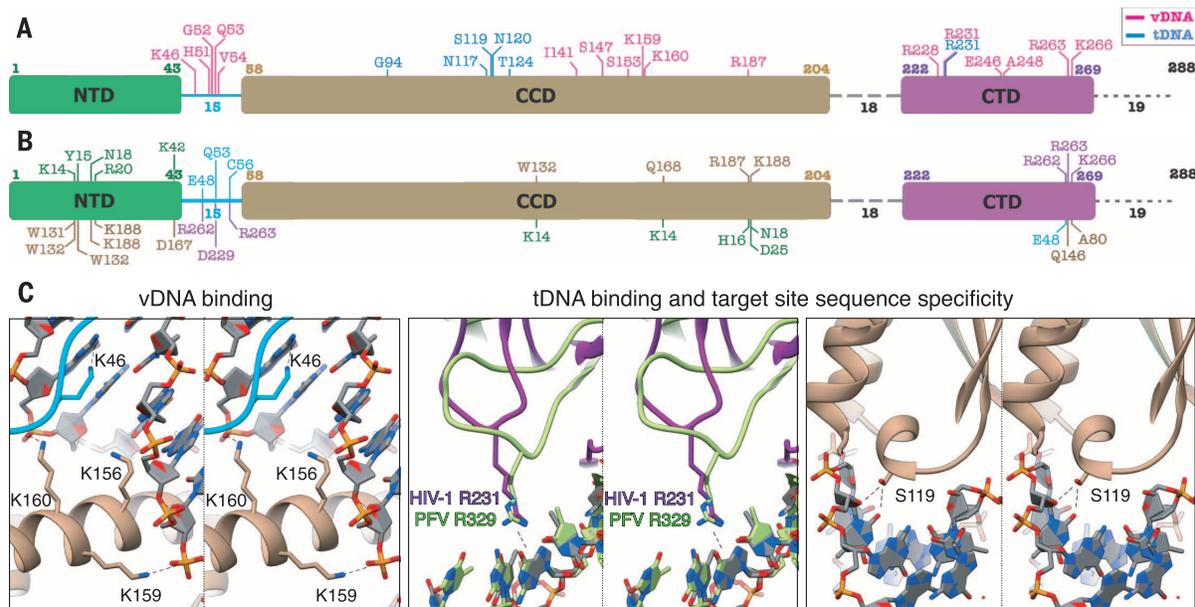


Fig. 2. Network of interactions within the HIV-1 STC intasome. (A) Map of IN residues predicted to be involved in electrostatic protein-DNA interactions within the STC intasome structure. All three domains and the NTD-CCD linker participate in interactions with vDNA (pink), whereas tDNA interactions (in blue) are mostly restricted to the CCD. Single-letter abbreviations for the amino acid residues are as follows: A, Ala; C, Cys; D, Asp; E, Glu; F, Phe; G, Gly; H, His; I, Ile; K, Lys; L, Leu; M, Met; N, Asn; P, Pro; Q, Gln;

R, Arg; S, Ser; T, Thr; V, Val; W, Trp; and Y, Tyr. (B) Predicted interdomain H-bond interactions within the STC. Residues designated below the domain schematic refer to the interacting domain and are colored accordingly. (C) Close-up views of selected regions involved in DNA interactions. For comparing HIV-1 R231 with prototype foamy virus (PFV) R329, the two structures were aligned to tDNA. For all panels, the protein color scheme is as in Fig. 1, C and D.

larger species containing flanking IN dimers (fig. S4) positioned in the trans configuration, similarly to RSV and mouse mammary tumor virus (MMTV) (11, 12). We then included the IN-binding domain (IBD) of LEDGF/p75 in the STC preparation, based on the rationale that IBD preferentially binds and stabilizes multimeric IN (27, 28), and performed a cryo-EM reconstruction of IBD-bound STCs (STC_{IBD}). The resulting data contained a

larger proportion of higher-order assemblies (fig. S8) but was also affected by substantial compositional heterogeneity; a cryo-EM reconstruction of the largest and best-resolved species clearly revealed 12 IN protomers within the map, with residual density contributed by a fraction of particles (Fig. 3, A and B, and figs. S9 to S11). IN can purify as tetramers from cells (29), and tetrameric INs constitute a portion of the higher-order

assemblies (Fig. 3C; see also below). It is therefore likely that the heterogeneous density corresponds to additional IN protomers that may collectively constitute a hexadecamer (or tetramer of tetramers). Possibly, the Sso7d fusion, which improves IN solubility (15), affects the assembly of fully formed higher-order species by mildly disrupting interprotomer associations (fig. S12). The higher-order assemblies utilize many of the principles

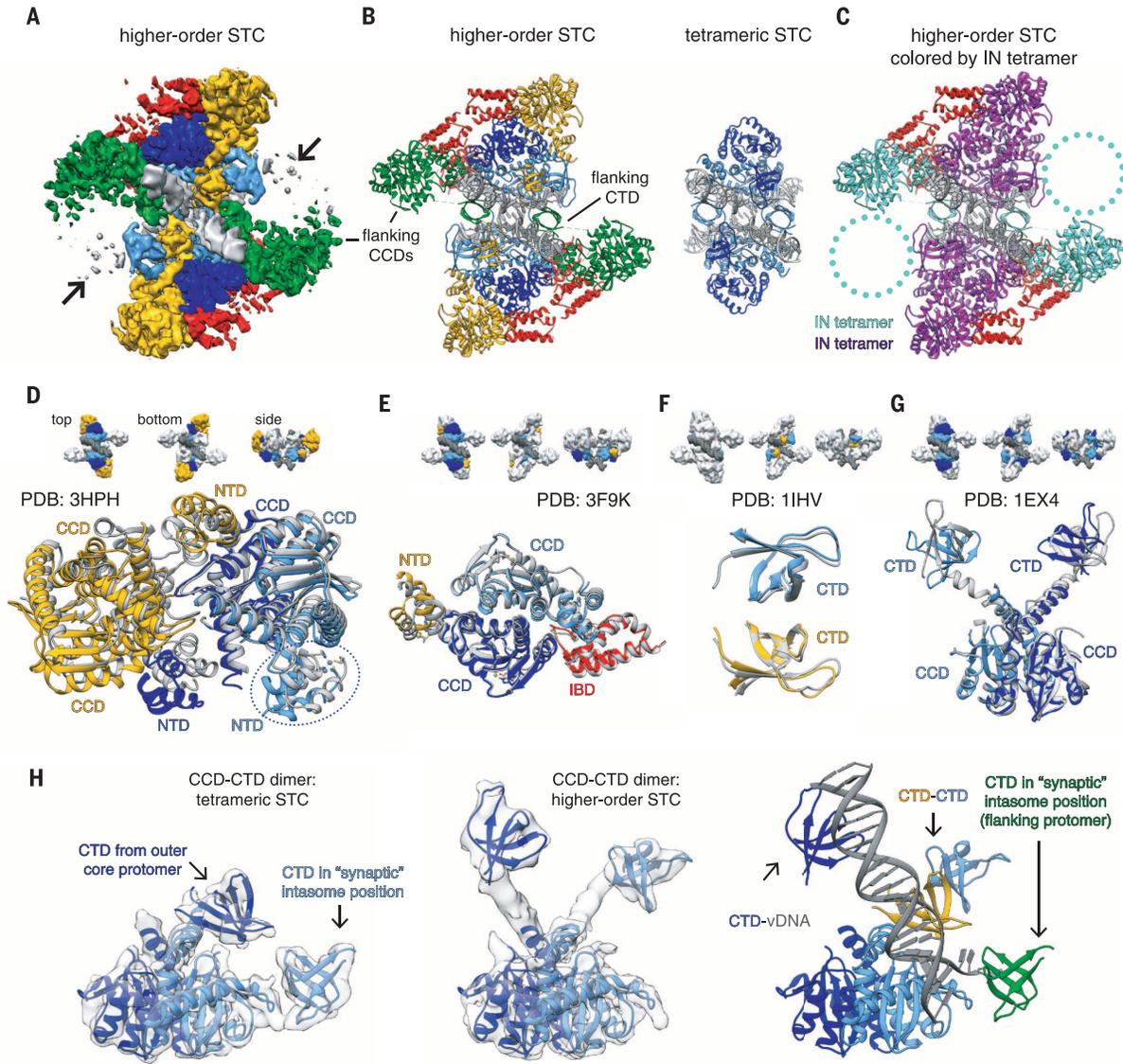


Fig. 3. HIV-1 STC intasomes form higher-order oligomers through distinct mechanisms of assembly. (A) Cryo-EM density map of IBD-bound STCs (STC_{IBD}). Densities are segmented either by IN protomers (inner core, light blue; outer core, dark blue) or IN dimers (yellow and green). The IBD is shown in red. (B) Higher-order STC model assembled by rigid-body-docking individual domain components, colored as in (A). The higher-order STC (left) is shown side by side with the tetrameric STC from Fig. 1 (right). (C) Model as in (B), colored by IN tetramers (28). The circled regions contain poorly resolved density that may harbor additional IN dimers. (D to G) Structural comparison of higher-order STCs assembled through rigid-body docking of individual domains with prior multidomain IN structures. The structural components of higher-order STCs are colored as in (A) and (B), whereas the Protein Data Bank (PDB) structures used for comparison are in gray.

Comparisons include: (D) MVV IN_{NTD-CCD} tetramer (PDB ID: 3HPH, IBD has been omitted for clarity; the circled NTD arises from an IN protomer on the opposite side of vDNA), (E) HIV-2 IN_{NTD-CCD} dimer bound to IBD (PDB ID: 3F9K), (F) HIV-1 CTD dimer (PDB ID: 1IHV), and (G) HIV-1 IN_{CCD-CTD} dimer (PDB ID: 1EX4). In all panels [(D) to (G)], structural schematics above highlight the corresponding location within visible dodecameric intasome density. (H) Conformational rearrangement within the core CCD-CTD dimer between (left) the tetrameric STC and (center) a higher-order STC, both overlaid on respective filtered experimental EM density. At right, the rearranged higher-order dimer is displayed in the context of additional CTDs and vDNA within the asymmetric unit. The “synaptic” position is required to form the conserved intasome core interface present in all retroviral intasomes.

underlying multimerization of IN protein in the absence of DNA (Fig. 3, D to G, and fig. S13). For example, the isolated tetramer from each asymmetric unit contains positionally conserved CCDs and NTDs that were previously observed within a two-domain HIV-1 NTD-CCD (IN_{NTD-CCD}) structure (7) and Maedi-Visna virus IN_{NTD-CCD} bound to IBD (Fig. 3D) (28). Individual dimers therein are also consistent with an HIV-2 IN_{NTD-CCD} structure (30) (Fig. 3E), whereas two of the CTDs interact in a manner identical to a nuclear magnetic resonance structure of a CTD dimer (Fig. 3F). Finally, the core CTDs adopt a configuration much like the two-domain IN_{CCD-CTD} (8) (Fig. 3G). The latter demonstrates an intriguing aspect of IN structure: Whereas tetrameric HIV intasomes adopt a domain configuration much like PFV, higher-order intasomes reorganize their CTDs, utilizing them to form an interprotomer CTD-CTD interface and to engage vDNA but replacing their respective positions with additional CTDs donated by outer IN protomers (Fig. 3H). These alternative domain arrangements preserve the positional integrity of the catalytically competent intasome and demonstrate the structural plasticity of HIV-1 IN.

HIV-1 intasomes assembled at lower protein and DNA concentrations than those used in our cryo-EM study were reported to be tetrameric (17–19). To test the relevance of the higher-order assemblies, we selected IN residues that were predicted to disrupt formation of these species but not the core tetramers. The most obvious candidates resided in the CTD-CTD interface—residues L242, I257, and V259—which are solvent-exposed within tetrameric intasomes. Several other residues—including K14, E35, K240, K244, and R269—were predicted to be more relevant in the context of higher-order oligomers, although we cannot completely exclude their involvement in tetrameric intasomes (fig. S14, A to E). The selected residues were substituted in the context of both Sso7d and WT (NL4-3) INs, and the mutant proteins were assayed for concerted integration activity. Similar results were observed in the presence and absence of the Sso7d fusion protein, suggesting that Sso7d does not alter the nature of functional complexes, although it may influence their relative abundance. The selected mutants, especially those in the CTD-CTD interface, affected strand transfer activity to various extents (fig. S14, F and G). Furthermore, mildly disrupting or deleting many of the residues within the CCD-CTD linker region (amino acids ~206 to 220), which is disordered in tetrameric intasomes (and thus would not be expected to play a major role) but is completely helical in higher-order assemblies, impaired catalytic activity (fig. S14H). In addition, we tested the importance of select residues that have not been examined previously (31–33) for virus replication (fig. S15). IN substitutions adversely affected virus replication, with >10-fold reductions observed for E212→K212 (E212K), K240E, and I257D mutations, and relatively less detrimental effects (twofold) seen for E35K. A mutation of K46, identified as a previously un-

identified vDNA binding residue, was also included in functional assays but was relevant for all oligomeric species. Whereas the K46A substitution did not detectably affect viral growth (34), the K46E substitution reduced virus replication by ~fivefold and substantially reduced strand transfer activity in vitro. These data suggest that the higher-order HIV-1 intasomes are functionally relevant for efficient catalysis. However, because point mutations can affect protein structures and/or the intasome assembly pathway in unexpected ways, further systematic studies will be required to delineate the (likely pleiotropic) effects of single-site substitutions. The simplest explanation for the distinct structures is that tetrameric intasomes, containing intact core domains, illustrate the minimal form upon which higher-order complexes are built, although their exact relevance is not clear. They may represent minimally active species or, alternatively, may serve as structural scaffolds for higher-order assembly within the pre-integration complex (PIC) or during PIC nuclear import. Further work, especially in the context of IN dynamics, will be required to unravel the role(s) of the tetrameric and higher-order forms in vivo.

Retroviruses are closely related evolutionarily and would be expected to utilize similar nucleoprotein structures for DNA integration. In this regard, the structure of the PFV intasome (9, 10) presented a conundrum. The length of the linkers between the domains of PFV IN is longer than in most retroviral INs, and many retroviral INs have linkers that are too short to form a tetrameric intasome that is analogous to the PFV structure (11). HIV-1 IN has linker lengths that are intermediate between PFV and MMTV or RSV (fig. S16A). The recent structures of MMTV (11) and RSV (12) intasomes show that these viruses overcome this problem by assembling intasomes with the same set of positionally conserved domains in contact with DNA, but for MMTV and RSV, two of the CTDs are contributed by an additional pair of flanking dimers in an octameric arrangement. Whereas PFV intasomes assemble tetramers, MMTV and RSV intasomes assemble octamers, and HIV-1 intasomes can apparently form a range of oligomeric configurations (fig. S16, B and C). The finding that HIV-1 IN can assemble intasomes in different ways while preserving the spatial arrangement of the key set of domains required for catalysis suggests that the evolutionary jump between retroviruses that assemble tetrameric and higher-order intasomes may not be as great as it initially appears.

The higher-order HIV-1 STC described here is very similar to a hexadecameric Maedi-visna virus intasome assembled using the LEDGF/p75 cofactor in an accompanying paper (35).

REFERENCES AND NOTES

1. R. Craigie, F. D. Bushman, *Cold Spring Harb. Perspect. Med.* **2**, a006890 (2012).
2. P. Lesbats, A. N. Engelman, P. Cherepanov, *Chem. Rev.* **116**, 12730–12757 (2016).
3. D. J. Hazuda, *Curr. Opin. HIV AIDS* **7**, 383–389 (2012).
4. F. Dydak et al., *Science* **266**, 1981–1986 (1994).

5. A. P. Eijkelenboom et al., *Nat. Struct. Biol.* **2**, 807–810 (1995).
6. M. Cai et al., *Nat. Struct. Biol.* **4**, 567–577 (1997).
7. J. Y. Wang, H. Ling, W. Yang, R. Craigie, *EMBO J.* **20**, 7333–7343 (2001).
8. J. C. H. Chen et al., *Proc. Natl. Acad. Sci. U.S.A.* **97**, 8233–8238 (2000).
9. S. Hare, S. S. Gupta, E. Valkov, A. Engelman, P. Cherepanov, *Nature* **464**, 232–236 (2010).
10. G. N. Maertens, S. Hare, P. Cherepanov, *Nature* **468**, 326–329 (2010).
11. A. Ballandras-Colas et al., *Nature* **530**, 358–361 (2016).
12. Z. Yin et al., *Nature* **530**, 362–366 (2016).
13. L. Krishnan et al., *Proc. Natl. Acad. Sci. U.S.A.* **107**, 15910–15915 (2010).
14. B. C. Johnson, M. Métifiot, A. Ferris, Y. Pommier, S. H. Hughes, *J. Mol. Biol.* **425**, 2133–2146 (2013).
15. M. Li, K. A. Jurado, S. Lin, A. Engelman, R. Craigie, *PLoS ONE* **9**, e105078 (2014).
16. Z. Yin, M. Lapkouski, W. Yang, R. Craigie, *Protein Sci.* **21**, 1849–1857 (2012).
17. M. Li, M. Mizuuchi, T. R. Burke Jr., R. Craigie, *EMBO J.* **25**, 1295–1304 (2006).
18. S. Kotova, M. Li, E. K. Dimitriadis, R. Craigie, *J. Mol. Biol.* **399**, 491–500 (2010).
19. S. Bera, K. K. Pandey, A. C. Vora, D. P. Grandgenett, *J. Mol. Biol.* **389**, 183–198 (2009).
20. T. Grant, N. Grigorieff, *eLife* **4**, e06980 (2015).
21. D. Dorse et al., *J. Virol.* **81**, 6731–6741 (2007).
22. A. Ballandras-Colas, H. Naraharisetty, X. Li, E. Serrao, A. Engelman, *PLoS ONE* **8**, e76638 (2013).
23. T. M. Jenkins, D. Esposito, A. Engelman, R. Craigie, *EMBO J.* **16**, 6849–6859 (1997).
24. A. Chen, I. T. Weber, R. W. Harrison, J. Leis, *J. Biol. Chem.* **281**, 4173–4182 (2006).
25. E. Serrao et al., *Nucleic Acids Res.* **42**, 5164–5176 (2014).
26. A. L. Harper, M. Sudol, M. Katzman, *J. Virol.* **77**, 3838–3845 (2003).
27. C. J. McKee et al., *J. Biol. Chem.* **283**, 31802–31812 (2008).
28. S. Hare et al., *PLoS Pathog.* **5**, e1000515 (2009).
29. P. Cherepanov et al., *J. Biol. Chem.* **278**, 372–381 (2003).
30. S. Hare et al., *PLoS Pathog.* **5**, e1000259 (2009).
31. R. Lu, H. Z. Ghory, A. Engelman, *J. Virol.* **79**, 10356–10368 (2005).
32. R. Lu, A. Limón, H. Z. Ghory, A. Engelman, *J. Virol.* **79**, 2493–2505 (2005).
33. S. J. Rihn, J. Hughes, S. J. Wilson, P. D. Bieniasz, *J. Virol.* **89**, 552–567 (2015).
34. R. Lu, N. Vandegraaff, P. Cherepanov, A. Engelman, *J. Virol.* **79**, 12584–12591 (2005).
35. A. Ballandras-Colas et al., *Science* **355**, 93–95 (2017).

ACKNOWLEDGMENTS

D.L. acknowledges support from NIH grant P50 GM103368 and the Leona M. and Harry B. Helmsley Charitable Trust grant 2012-PG-MED002. R.C. is supported by the Intramural Program of the National Institute of Diabetes and Digestive Diseases of the NIH and by the Intramural AIDS Targeted Antiviral Program of the Office of the Director of the NIH. These studies were also partly supported by NIH grant R01 AI062520 to M.K. Molecular graphics and analyses were performed with the University of California, San Francisco, Chimera package (supported by NIH grant P41 GM103331). We thank B. Anderson and J.-C. Ducom for help with EM data collection and network infrastructure, F. Dwyer for computational support, G. Lander and M. Herzik for help with ensemble refinements, and A. Engelman and M. Gellert for critical review of the manuscript. The data presented in this manuscript are tabulated in the main paper and in the supplementary materials. The EM maps of STC and STC_{IBD} are deposited into the Electron Microscopy Data Bank under accession codes EMD-8481 and EMD-8483, respectively. The STC model is deposited into the Protein Data Bank under ID 5U1C. The STC model ensemble and the composite model of the higher-order STC_{IBD} oligomers are available upon request.

SUPPLEMENTARY MATERIALS

www.sciencemag.org/content/355/6320/89/suppl/DC1

Materials and Methods
Supplementary Text
Figs. S1 to S16
Tables S1 and S2
References (36–52)

14 July 2016; accepted 2 December 2016
10.1126/science.aah5163



Supplementary Materials for

Cryo-EM structures and atomic model of the HIV-1 strand transfer complex intasome

Dario Oliveira Passos,* Min Li,* Renbin Yang, Stephanie V. Rebersburg, Rodolfo Ghirlando, Youngmin Jeon, Nikoloz Shkriabai, Mamuka Kvaratskhelia, Robert Craigie, Dmitry Lyumkis†

*These authors contributed equally to this work.

†Corresponding author. Email: dlyumkis@salk.edu

Published 6 January 2017, *Science* **355**, 89 (2017)

DOI: 10.1126/science.aah5163

This PDF file includes:

Materials and Methods
Supplementary Text
Figs. S1 to S16
Tables S1 and S2
References

Materials and Methods:

HIV-1 STC intasome assembly and purification. Sso7d-integrase active site mutant E152Q was over-expressed in *Escherichia coli* BL21 (DE3) and purified essentially as previously described (15) with minor modifications. Briefly, Sso7d-IN (E152Q) was purified by nickel-affinity chromatography and the His-tag was removed with thrombin. Aggregated protein was removed by gel filtration on a Superdex-200 column (GE Healthcare) equilibrated with 20 mM HEPES pH 7.5, 10% glycerol, 2 mM DTT, and 0.5 M NaCl. The protein was concentrated to 0.6 mg/ml using an Amicon centrifugal concentrator (EMD Millipore), flash-frozen in liquid nitrogen and stored at -80°C. The DNA substrate was made by annealing the following oligonucleotides:

5'-CAGTGTGGAAAATCTCTAGCAGTTACAGTCAGCGTAC-3'

5'-ACTGCTAGAGATTTTCCCACTG-3'

5'-GTACGCTGACT-3'

The HIV-1 STC intasome assembly strategy was similar to that previously described for PFV

(16) and RSV (12) intasomes. Large-scale HIV-1 intasome preparations were assembled by mixing 3.0 μM integrase with 1.0 μM DNA substrate in buffer containing 20 mM HEPES pH 7.5, 5 mM MgCl_2 , 8 mM 2-mercaptoethanol, 4 μM ZnCl_2 , 100 mM NaCl, 25% (w/v) glycerol and 50 mM 3-(Benzyldimethylammonio) propanesulfonate (NDSB-256), followed by incubation at 30°C for 90 min. The reaction was stopped by incubating on ice for 15 min. NaCl was then added to 500 mM, and after incubation at room temperature for 15 min the mixture was centrifuged at 15,000 g for 15 min to remove insoluble aggregate.

HIV-1 intasomes were purified by Ni-affinity chromatography and anion exchange chromatography prior to the size exclusion chromatography. The mixture was first loaded onto a HisTrap HP column (GE Healthcare) equilibrated with 20 mM Tris pH 8.0, 10 mM imidazole, 5 mM 2-mercaptoethanol, 0.5 M NaCl, 20% (w/v) glycerol, and then washed extensively with the same buffer. HIV-1 intasomes were eluted with a linear gradient of 10 mM to 500 mM imidazole. Intasome-containing fractions were combined and diluted with 20 mM Tris pH 8.0, 0.5 mM tris-(2-carboxyethyl)phosphine (TCEP), 20% glycerol to lower the NaCl concentration to 200 mM and then quickly loaded onto a Mono Q column (GE Healthcare) equilibrated with 20 mM Tris, pH 8.0, 0.5 mM TCEP, 200 mM NaCl, 20% glycerol. Intasomes were eluted with a linear gradient of NaCl from 200 mM to 1 M. Fractions were pooled and then separated on a Superdex-200 PC 3.2/30 column (GE Healthcare) equilibrated with 20 mM Tris pH 8.0, 0.5 mM TCEP, 500 mM NaCl and 6% (w/v) glycerol.

STC_{IBD} were prepared essentially the same as described for STCs, except 7.5 μM of His-IBD (347-429) was added after solubilization in the 0.5M NaCl buffer and incubation was continued overnight on ice. The subsequent purification steps differed only in that buffer for gel

filtration was 20 mM Tris pH 8.0, 0.5 mM TCEP, 650 mM NaCl and 6% (w/v) glycerol for the STC_{IBD}.

CryoEM specimen preparation. Sample containing HIV STCs or STC_{IBD} in SEC buffer was applied onto freshly plasma treated (6s, Gatan Solarus plasma cleaner) holey gold UltrAuFoil grids (Quantifoil), adsorbed for 30 sec and then plunged into liquid ethane using a manual cryo-plunger in an ambient environment of 4°C.

CryoEM data acquisition. Data were acquired using Legion software (36) installed on an FEI Titan Krios electron microscope operating at 300 kV. All data collection statistics for both STC and STC_{IBD} data are summarized in Table S1. The individual frames were gain corrected, aligned and summed with the application of an exposure filter using MotionCor2, according to the nominal dose rate.

CryoEM image analysis. Pre-processing operations prior to the refinement of the final models were performed using the Appion package (37). All specific parameters are described in Table S1. Single intasome particles were selected from the whole-frame aligned and summed micrographs and used to create an initial raw particle stack after removing regions of the micrographs containing gold or large areas of aggregation. 2D alignments and classifications were performed using the CL2D (38) and Relion (39) algorithms (Figure S2 and Figure S6), and an initial model was generated directly from the class averages using OptiMod (40). Following iterative rounds of 2D alignment and classification, 219,260 particles remained for 3D refinement and classification for STC data and 102,121 particles for STC_{IBD} data. 3D refinements

and classifications were initially performed within Relion (39, 41), after which the parameters were converted for use in Frealign (42, 43). The final map was refined in Frealign. Details of the classification are shown in Fig. S4 and Fig. S10.

Focused classification per asymmetric subunit. Global classifications in both Relion (39, 41) and Frealign (42, 43) showed presence of higher-order oligomers, with density in peripheral regions of the intasome (“above” the core) and a small amount of density in flanking regions (to the “side” of the core); the latter are analogous to MMTV (11) and RSV (12). Focused classifications were performed on each of these two regions, specifically for the purpose of excluding heterogeneity from other regions of the map that might bias classification results. Furthermore focused classification was performed on an asymmetric subunit basis, with the rationale that loosely associated peripheral, and especially flanking, regions may not abide by strict twofold symmetry. Thus, signal from each asymmetric subunit would contribute additively to the density, rather than diluting it. A similar concept has been previously proposed and implemented within Relion (44), whereas we performed this classification within Frealign. For the peripheral subunit, a mask with a radius of 32 Å, calculated to cover the entirety of an IN dimer, was positioned over the peripheral region. 100 cycles of classification using 4 classes and without adjusting orientation parameters (Euler angles and translations) were performed. The class containing the best density was selected and is displayed in Figure S3. The same strategy was performed with the flanking region. Notably, each of the 4 resulting classes had some amount of density, although this density was highly mobile and floppy. This is expected, because the CCD and CTD domains are connected by a flexible linker, and stabilizing interactions are absent in the structure. The presence of IBD dramatically stabilized the higher order oligomers

(Figure 4B), although flanking regions remained somewhat mobile.

Assembly of the atomic model and ensemble refinement. The model of the HIV-1 intasome was built and refined in a stepwise manner using Rosetta (45) and Phenix (46) starting with rigid-body fitted X-ray structures of individual domains as input. To optimally fit X-ray models into the EM density, we first independently refined each individual domain (NTD, CCD, and CTD) using multiple input starting seeds. NTD and CCD domains were seeded and refined using an independent rigid body-docked domain from the NTD-CCD X-ray structure (7) (PDB: 1K6Y). CTD_{inner} and CTD_{outer} were each seeded with 4 starting X-ray models: 2 CTD monomers from an NMR structure (5) (PDB: 1IHV) and 2 individual CTD domains from a two-domain CCD-CTD X-ray structure (8) (PDB: 1EX4). At least 1000 models were generated from each and the lowest energy model was selected for moving forward. Modeling quality was assessed by energy scores, structural similarity of the top scoring models by RMSD value, and visual inspection. *De novo* modeling was performed for NTD-CCD linker residues 45-57, for the CCD loop 139-149 containing the 3-10 helix, and for any minor regions that were not well represented by the procedure above. HIV-1 STC DNA was modeled based on the X-ray structure of the PFV intasome (PDB 3L2Q) and adjusted in Coot (47) and Phenix. Models were assembled into a composite asymmetric subunit containing an inner core protomer (NTD, NTD-CCD linker, CCD, and CTD), an outer core protomer (CCD and CTD), and product DNA. The complete intasome model was iteratively relaxed using two-fold symmetry with combinations of Rosetta and Phenix against one of the half-maps (working map) and inspected for self-consistency against the second half-map (free map), then adjusted manually using Coot (47). For the final ensemble modeling approach, we used the following procedure, using half maps for all aspects

of refinement and evaluation, respectively: first, we generated 500 models using the above procedure in Rosetta. From the top 100 scoring models by Rosetta energy, we selected the ten models that scored best by map-to-model FSC. These ten were then refined in real space using Phenix using secondary-structure restraints. Molprobit (48) was used throughout the refinement process. From the ensemble analysis, the highest scoring model (FSC map-to-model) is used for all figures throughout the text. Otherwise, values associated with ensemble modeling are reported as mean and standard deviation in Tables S1-2.

Size exclusion chromatography of WT and Sso7d IN. In parallel runs wild type IN and Sso7d IN fusion proteins (26 μ M) were analyzed with the Superdex 200 Increase 10/300 GL column (GE Healthcare) using 0.75 ml/min flow rate and the elution buffer containing 20 mM HEPES (pH 6.8), 1 M NaCl, 10 mM MgSO₄, 0.2 mM EDTA, 10% glycerol, and 5 mM β -mercaptoethanol. The column was calibrated with the Gel Filtration Standard (Bio-Rad) containing the mixture of the following proteins: Thyroglobulin (bovine) 670,000 Da, γ -globulin (bovine) 158,000 Da, Ovalbumin (chicken) 44,000 Da and Myoglobin (horse) 17,000 Da. Proteins were detected by absorbance at 280 nm.

Analysis of Sso7d IN oligomeric states by chemical crosslinking. BS³ (bis(sulfosuccinimidyl)suberate) (Thermo Fisher) was used to covalently crosslink interacting protein subunits. The crosslinking reactions were carried out in 20 mM HEPES pH 7.5, 0.6 M NaCl, 10% glycerol, 0.5 mM TCEP. In parallel reactions, 500 μ M BS³ was added to 8.4 μ g of Sso7d IN, STC or STC_{IBD}. The reactions were incubated for 1 h at 30°C and then quenched by trichloroacetic acid precipitation. The proteins were separated on a 4–12% Bis-Tris SDS-PAGE

gel and visualised by coomassie blue staining. Both low molecular weight (SeeBlue® Plus2 Thermo Fisher) and high molecular weight (HiMark™ Thermo Fisher) prestained protein standards were used for estimating the multimeric states of the crosslinked proteins.

Strand transfer activity assay of Sso7d IN. Assays were performed as previously reported (15). Briefly, 1 μ M Sso7d INs and 0.5 μ M 25 bp FAM labeled viral DNA and 300ng of target plasmid DNA pGEM-9zf substrate were incubated in 20 mM HEPES pH 7.5, 25% glycerol, 10 mM DTT, 5 mM MgCl₂, 4 mM ZnCl₂, and 100 mM NaCl in a 20 μ l reaction volume. The reaction was carried out at 37°C for 1 hr. The integration products were recovered by ethanol precipitation and subjected to electrophoresis in a 1.5% agarose gel in 1x TBE buffer. DNA was visualized by fluorescence using a Typhoon 8600 fluorescence scanner (GE Healthcare).

Strand transfer activity assay of NL4-3 IN. Assays were performed as previously reported (17). Briefly, 80 nM NL4-3 integrase was incubated on ice with 20 mM HEPES pH 7.5, 12% DMSO, 5 mM DTT, 10% PEG-6000, 10 mM MgCl₂, 20 μ M ZnCl₂, and 100 mM NaCl, followed by addition of 10 nM 1 kb P 32 labeled viral DNA substrate and 250 ng target plasmid DNA pBR322. The reaction was carried out at 37°C for 2 hrs. The reactions were stopped and deproteinized by addition of SDS, EDTA and proteinase K to 0.1%, 10 mM and 5 mg, respectively. 2.5 μ l of the reaction mixture was then subjected to electrophoresis in a 0.8% agarose gel in 1x TBE buffer. Gels were dried, exposed to imaging plates and visualized by a Typhoon 8600 fluorescence scanner (GE Healthcare).

Virus production and analysis of virus infectivity. pNL4-3 molecular clones containing E35K, K46E, E212K, K240E and I257D substitutions in the IN coding sequence were generated by site

directed mutagenesis and used for virus preparations. Transfections in HEK293T cells were performed with X-tremeGENE HP (Roche) and the desired plasmid DNA using the 1:3 ratio for DNA:reagent. 24 hr post-transfection, cells were washed and fresh media was added. Cell culture supernatants containing virions were harvested 72 hr post-transfection, and virus production was measured by Gag-p24 ELISA (ZeptoMetrix). Virus-containing supernatants with Gag p24 amounts equivalent to 5ng Gag p24 were then used to infect TZM-bl cells that were seeded the day before at a concentration of 4×10^5 cells per well in 6-well plates. These cells were harvested 48h after virus addition, cell lysates were prepared using 1x reporter lysis buffer (Promega) and luciferase activity was determined using a commercially available kit (Promega).

Figures S1-S16

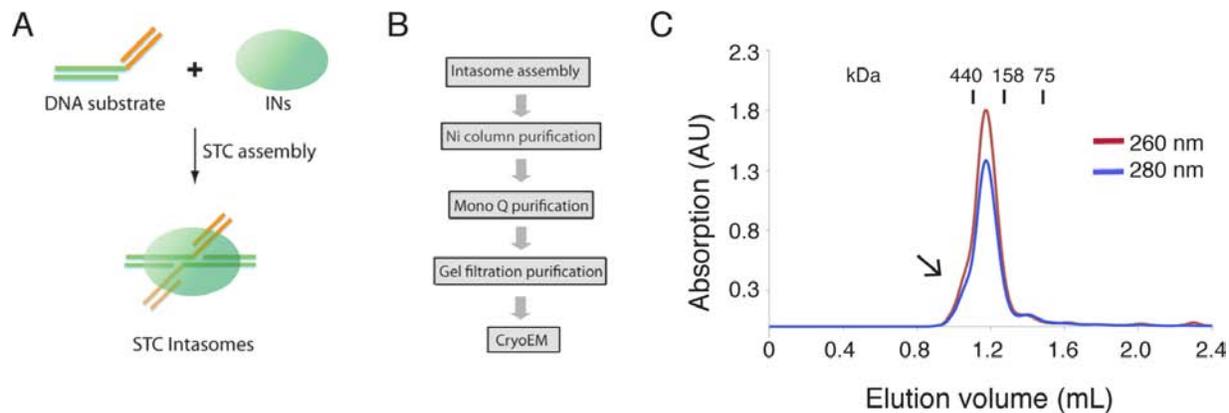


Fig. S1. HIV-1 STC assembly, purification, and characterization. (A) Schematic of STC intasome assembly on DNA substrate mimicking the product of the concerted integration reaction (orange, vDNA; green, tDNA). (B) Flowchart of STC preparation for CryoEM. (C) Gel filtration profile of STC intasome species for CryoEM. The intasome peak elutes at 1.2 ml and is ~200 kDa relative to protein standards, which is consistent with a tetrameric size of the STC. However, the broad and asymmetric peak indicates the presence of larger intasome species (arrow). Mass standards in kilodaltons (kDa) are indicated.

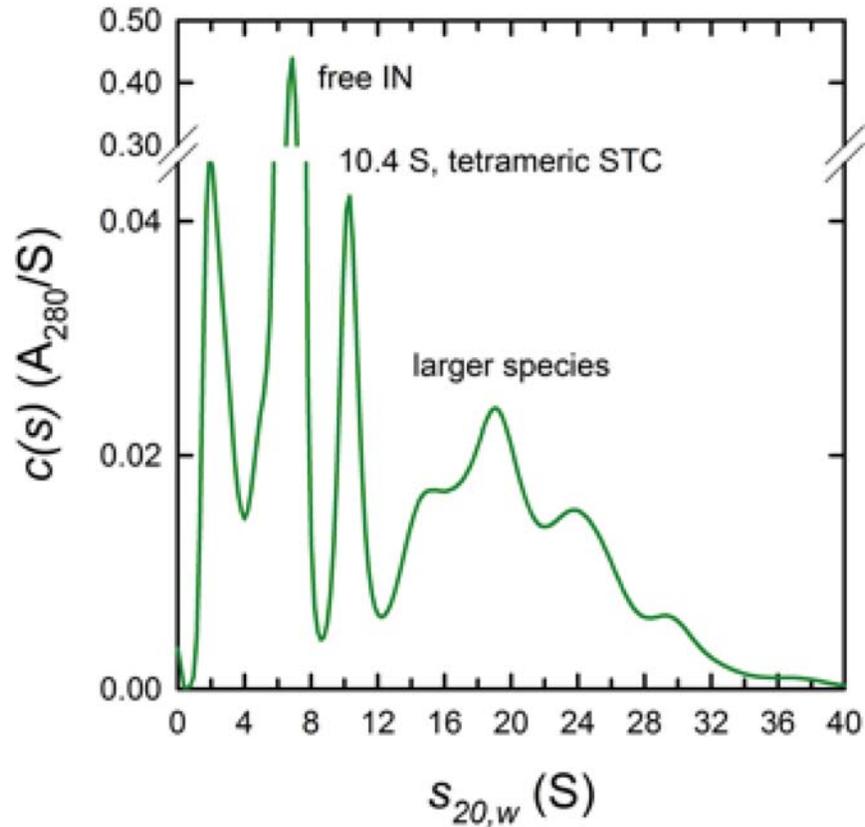


Fig. S2. Analytical ultracentrifugation absorbance $c(s)$ profile for a preparation of HIV intasome STCs after anion exchange chromatography. The majority of the material at 6.7 S represents free IN protein. The species with sedimentation coefficient 10.4 S has a calculated molar mass of 228 kDa, consistent with the expected size of a tetramer of IN complexed with the branched DNA substrate. The value of the molar mass is determined based on the best-fit frictional coefficient of 1.47 obtained for the whole $c(s)$ distribution. Larger species are indicated.

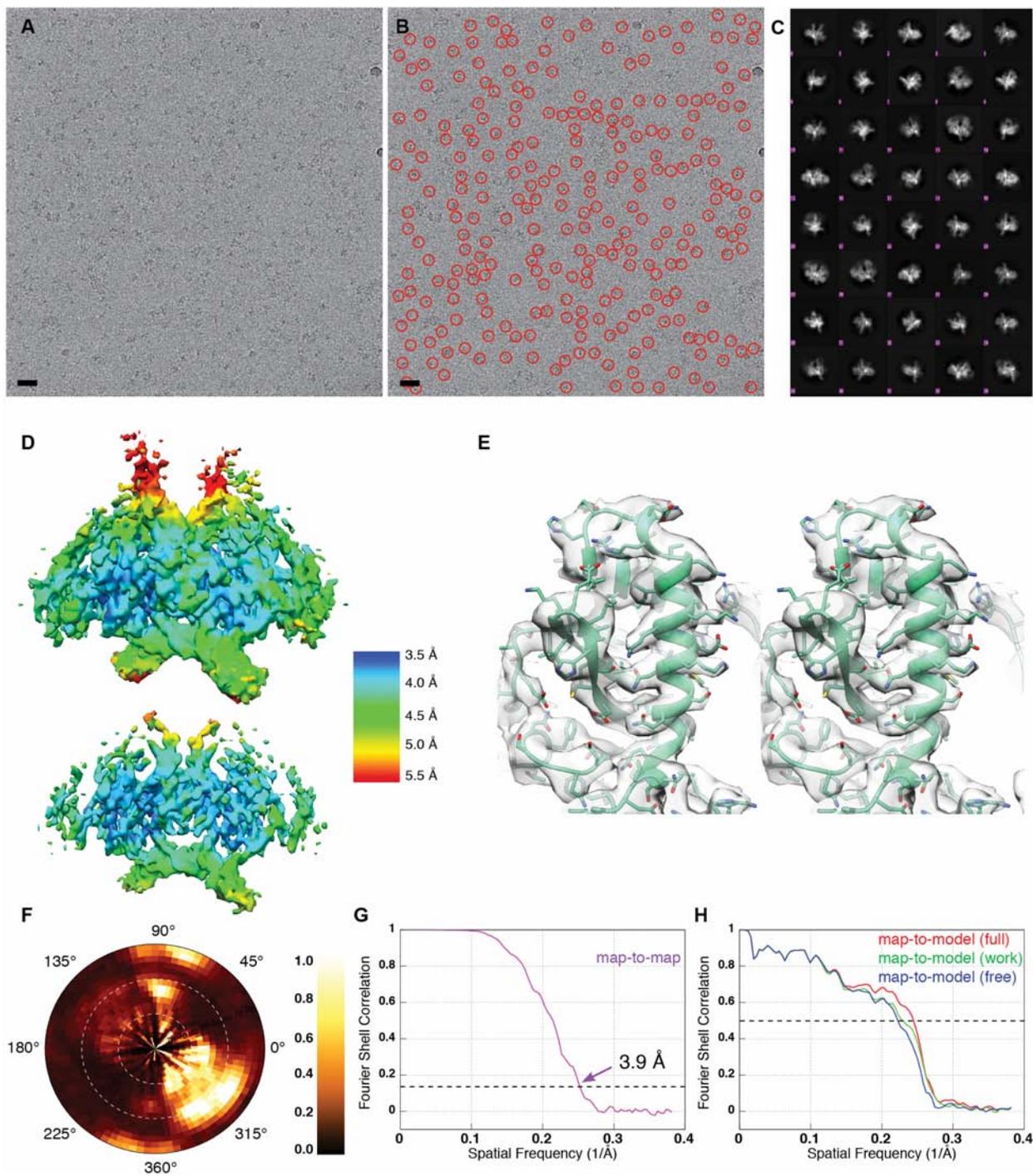


Fig. S3. CryoEM raw data and refinement of HIV-1 STC structure. (A) Representative cryoEM micrograph of HIV-1 STCs, taken at $-2.5 \mu\text{m}$ defocus. (B) Same as in panel (A), marked to show selected particles. (C) 2D class averages calculated using Relion (39). (D) Refined

cryoEM map shown at low (top) and high (bottom) contours, colored by local resolution using `sxlocres.py` that is implemented within Sparx (49). Regions at 3.5 Å are evident in the core of the STC, in and around the active site. Lower resolution regions surround the periphery of the STC. **(E)** Region of the CCD, displayed in stereo, showing features present within the map. **(F)** Euler angle distribution plot, shown as a heat map, describing the relative orientations of the refined map. **(G-H)** Fourier Shell Correlation (FSC) curves displaying **(G)** map-to-map (half-map) resolution of the derived density and **(H)** map-to-model resolution of one of the refined atomic models from ensemble analysis. The model was refined against a half-map (green, work-FSC), and the resolution was subsequently also obtained for the second half-map (blue, free-FSC), and the full map (red, full-FSC).

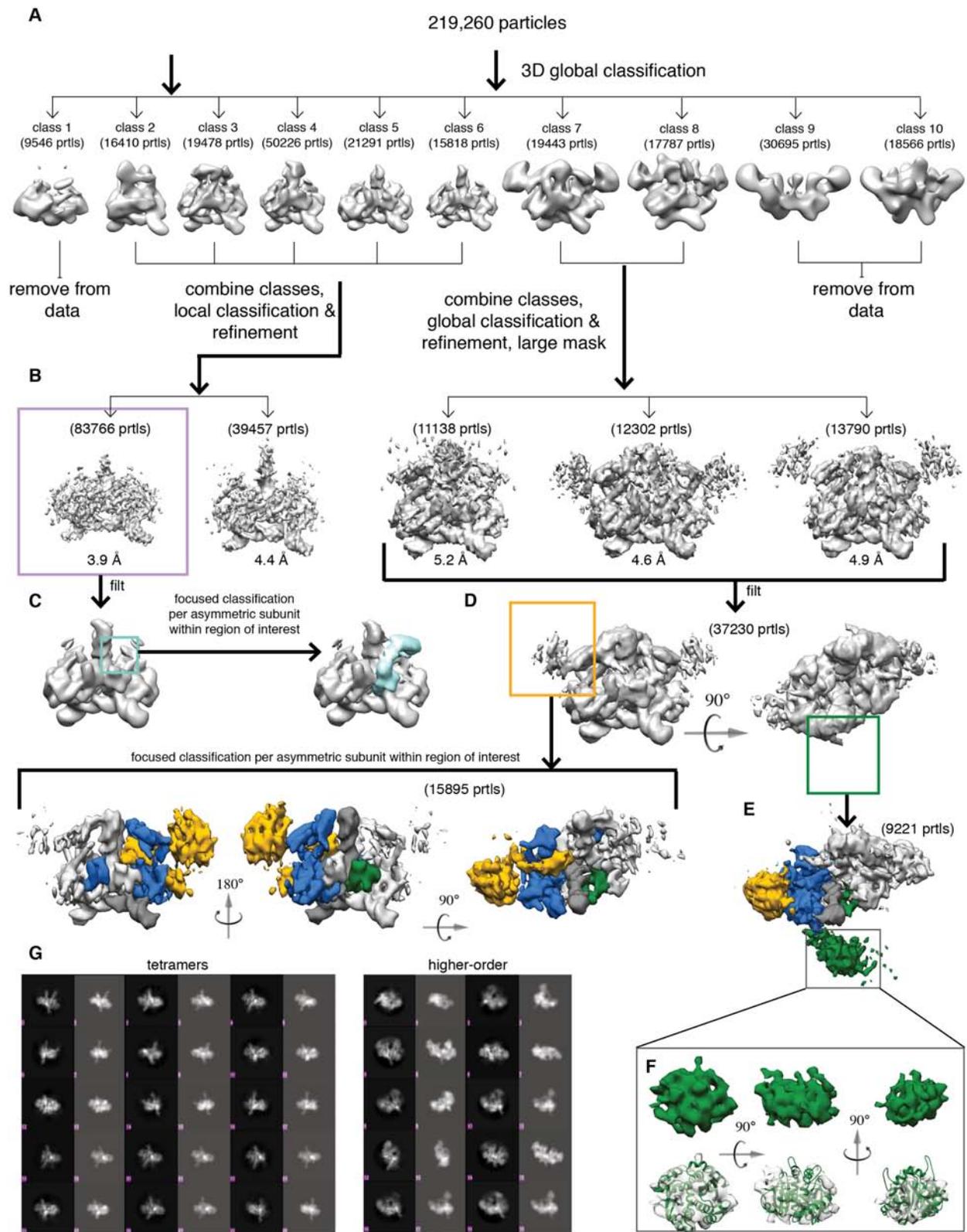


Fig. S4. 3D-classification flowchart of HIV-1 STC cryoEM data. (A) Results of global 3-D classification in Relion. Many of the particles that produced low quality maps (classes 1, 9, 10) were removed from subsequent analysis. Particles belonging to smaller oligomeric maps (classes 2, 3, 4, 5, 6) and larger oligomeric maps (classes 7, 8) were grouped. (B) Orientation parameters for the particles belonging to smaller oligomers were converted for classification and refinement in FREALIGN. Of the two refined maps, one was resolved to a global resolution of 3.9 Å (outlined in purple). This map was used to build the HIV-1 STC tetrameric model and is displayed in Figure 1. While some residual heterogeneity remains in this map, it disappears upon further classification (which also reduces resolution by decreasing particle count). (C-E) Focused classification on an asymmetric subunit basis was performed for several regions of the indicated densities in order to resolve (C) a CCD-CTD linker, which showed a similar configuration to PFV STCs (10) in the context of a tetramer, (D) an IN dimer “above” the core (yellow) and (E) an IN dimer “flanking” the core (green) (see Materials and Methods); the latter was particularly poorly resolved in the combined map, apparently due to large amounts of mobility. In all cases for panels (C-E), the resulting densities are segmented and displayed (and optionally filtered for clarity). (F) Fit of a CCD dimer (PDB 1K6Y) into refined density from the flanking asymmetric subunit is shown in all three axial orientations. The density is consistent with the size of a CCD dimer. (G) Comparison of original 2D class averages with re-projections from a density map along identical Euler angles. Re-projections were performed either with a tetrameric map (left) or with a higher-order map (right).

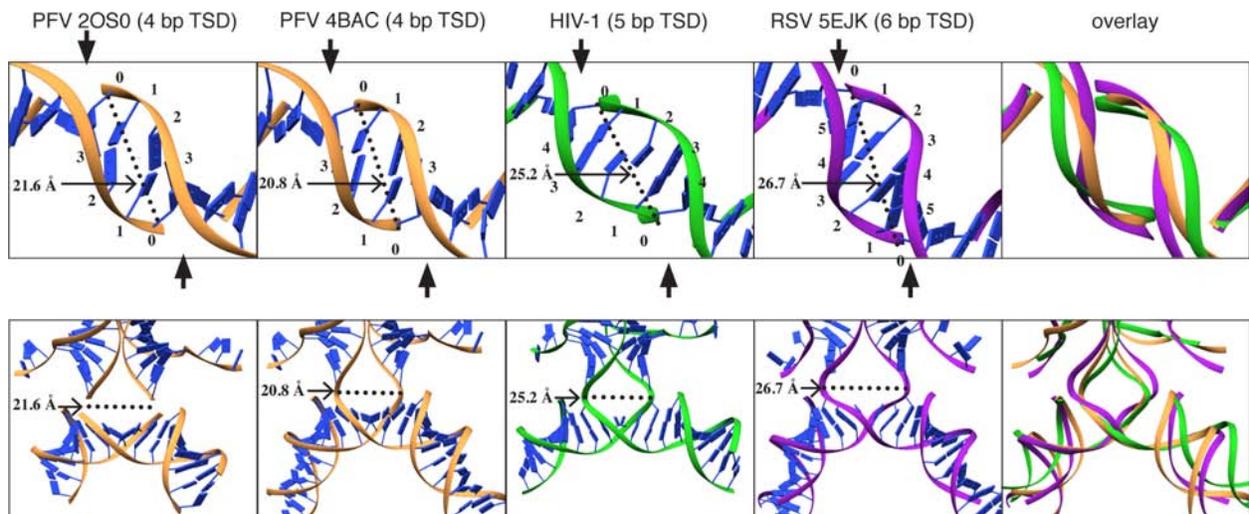


Fig. S5. Target site and tDNA of HIV-1 STCs. (A) target sites are shown for all three retroviral STCs available in the public domain, and overlaid with one another. An even number of nucleotides is present in PFV (4 bp) and RSV (6 bp) recognition sites, whereas an odd number of nucleotides is present in the HIV (5 bp) recognition site. Due to the presence of a T/T mismatch within the central bp of HIV-1 STCs (at position +2) that is inherent to sample preparation from a common product DNA, the two central dinucleotides apparently lose base pairing interactions. HIV/SIV 5 bp-recognition TSDs have been predicted to contain severe kinking within the central 3 bps using multiple experimental approaches (25, 50, 51), and STCs can be assembled from product DNA even without any bp matches within the tDNA region (16). Thus, this mismatch may not severely affect the overall structure, although local regions may slightly differ in comparison to the natural recognition sequence (15). (B) The overall configuration of tDNA is distorted in all three retroviruses. In both (A-B), distances are calculated between the P atoms of the acceptor phosphate. Similar relative patterns were observed between nearby atoms.

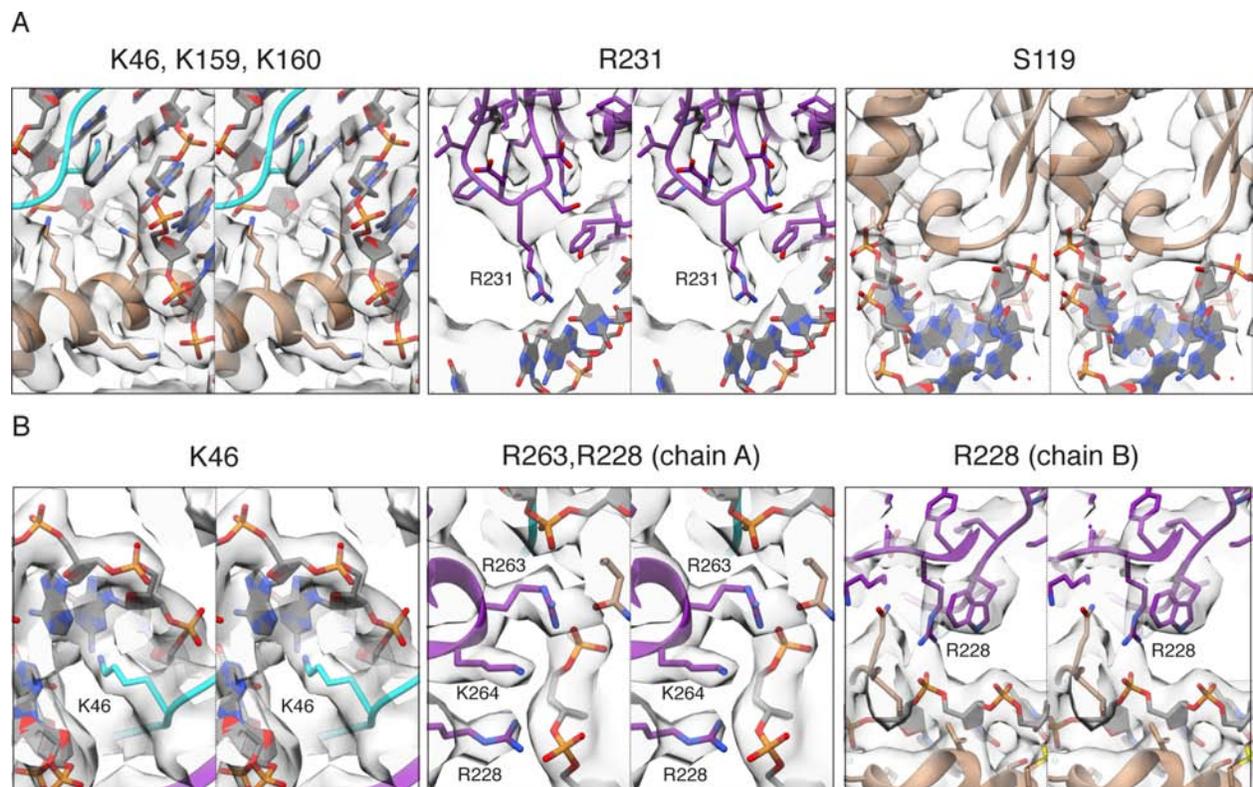


Fig. S6. CryoEM densities corresponding to areas of interest. (A) Panels for all interactions discussed in Figure 2 are overlaid with cryoEM density and displayed in stereo. (B) K46 is shown from a different angle, and several additional regions denoted in Table S2 are overlaid with cryoEM density. R263 is sandwiched between both vDNA strands. R263K mutants decreased integration in cell culture, but also conferred low-level resistance to an INSTI (52); the compromised catalytic activity of the R263K virus is likely due to the multiple relevant interactions made by R263. This same area is also occupied by the functionally relevant R262 and R228 (13). The latter makes an additional vDNA interaction with the outer protomer.

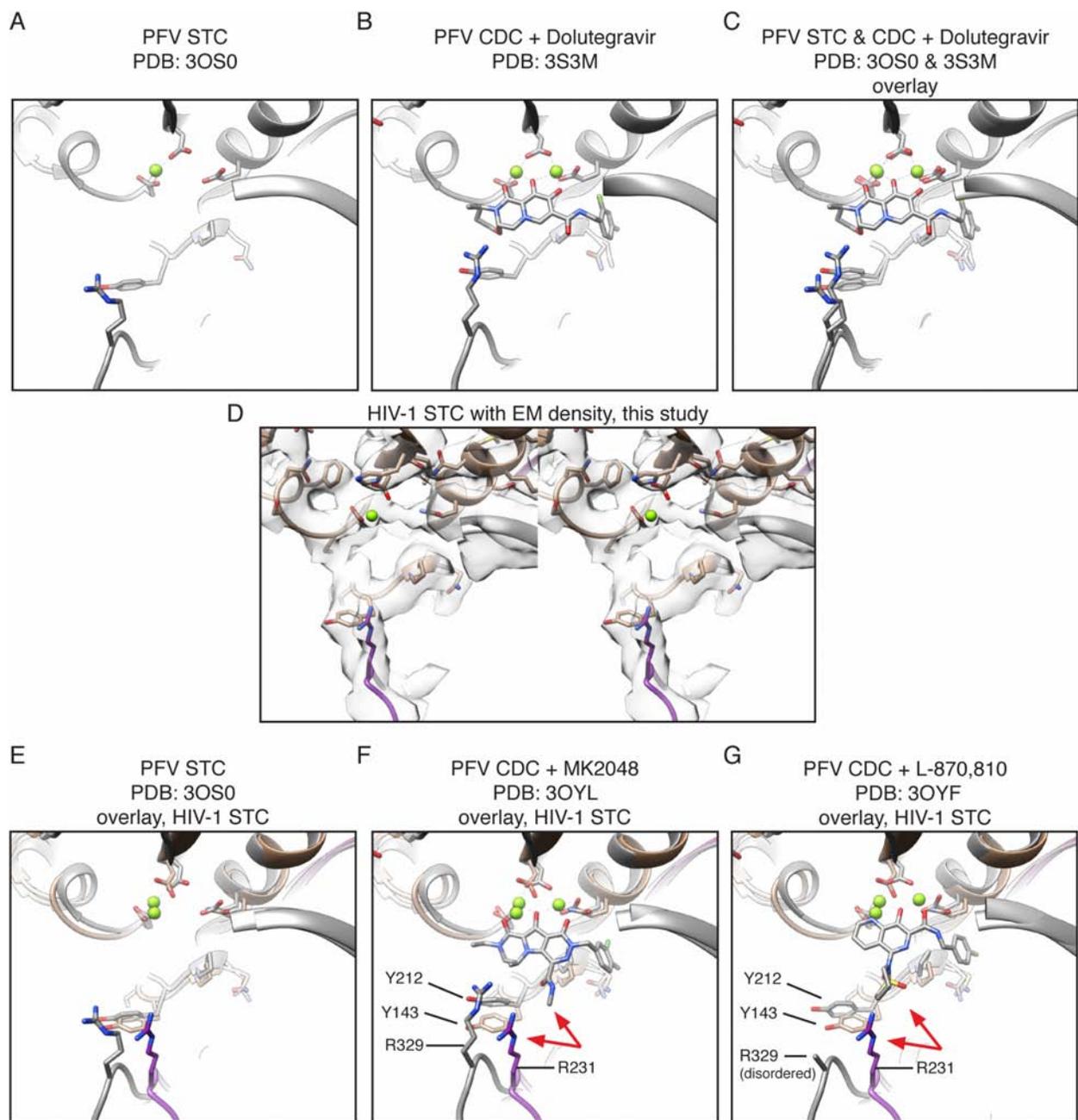


Fig. S7. Differences in the active sites of HIV-1 and PFV. (A) PFV STC (PDB 3OS0) (10), (B) PFV cleaved donor complex (CDC) bound to Dolutegravir (PDB 3S3M) (53), and (C) an overlay of both structures in (A-B). Panel (C) shows that, although INSTIs bind CDC and not STC, much of the active site outside of the terminal adenine (not shown for clarity), and in

particular the configuration of the neighboring CTD, remains intact. **(D)** Aligned HIV-1 STC active site from this study, displayed with cryoEM density, in stereo. **(E)** HIV-1 STC overlaid with PFV STC (PDB 3OS0). **(F-G)** HIV-1 STC overlaid with structures of **(F)** PFV CDC bound to MK2048 (PDB 3OYL) and **(G)** L-870-810 (PDB 3OYF) inhibitors (54). The active site Tyr (PFV Y212 / HIV-1 Y143), which exhibits slight differences in existing structures is denoted. Of note, red arrows point to the CTD-Arg (PFV R329 / HIV-1 R231) opposing the active site, which is rearranged in comparison to PFV, as well as a corresponding area within INSTIs that is often derivatized; R231, in particular, resides proximal to an area of INSTI derivatization that can potentially be exploited for improved inhibitor binding. In all panels containing PFV CDC, the terminal adenine nucleotide that interacts with INSTIs has been removed for clarity; likewise, the same adenine and downstream tDNA of HIV-1 STC have also been removed for clarity.

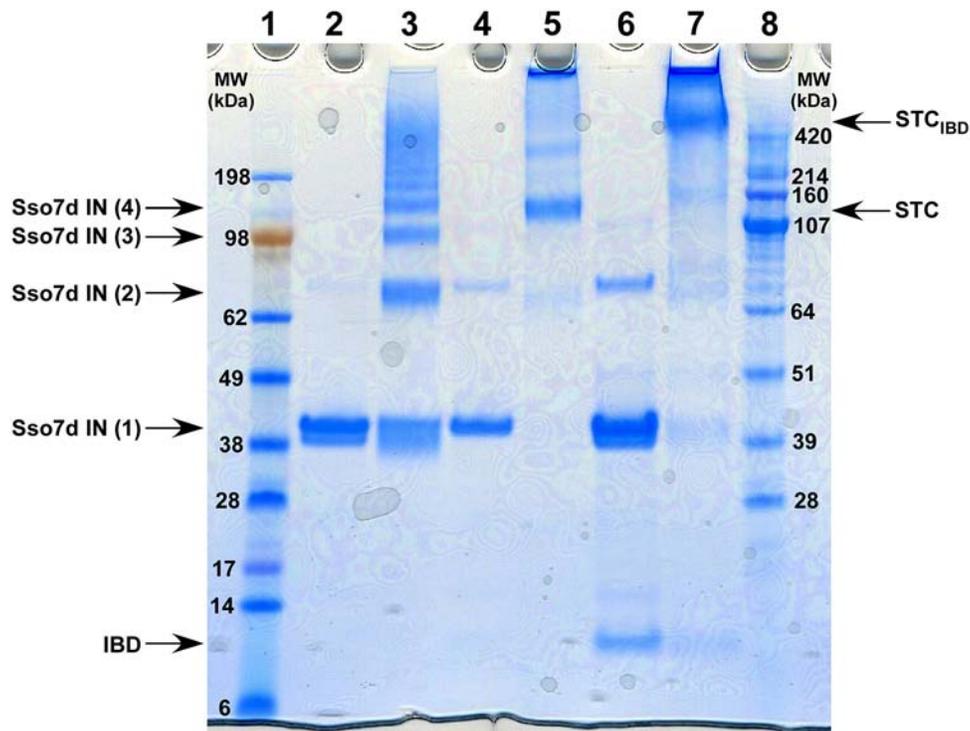


Fig. S8. BS^3 mediated crosslinking to examine multimeric states of Sso7d IN in STC and STC_{IBD} . A representative SDS-PAGE image of crosslinked species is shown. Lane 1: low molecular weight (MW) markers; Lane 2: Sso7d without BS^3 ; Lane 3: Sso7d + BS^3 ; Lane 4: STC without BS^3 ; Lane 5: STC + BS^3 ; Lane 6: STC_{IBD} without BS^3 ; Lane 7: STC_{IBD} + BS^3 . Lane 8; high MW markers. Analysis of unliganded Sso7d (Lane 3) reveals multiple species including monomer: Sso7d IN(1), dimer: Sso7d IN(2), trimer: Sso7d IN(3), tetramer Sso7d IN(4) and higher-order oligomers. In contrast, STC predominantly contains tetrameric Sso7d IN (compare lane 5 with lane 3) with relatively smaller amounts of higher-order oligomers being also observed (lane 5). The addition of IBD to STC shifts the oligomeric state of Sso7d IN to higher-order assemblies in the context of STC_{IBD} (compare lane 7 with lane 5).

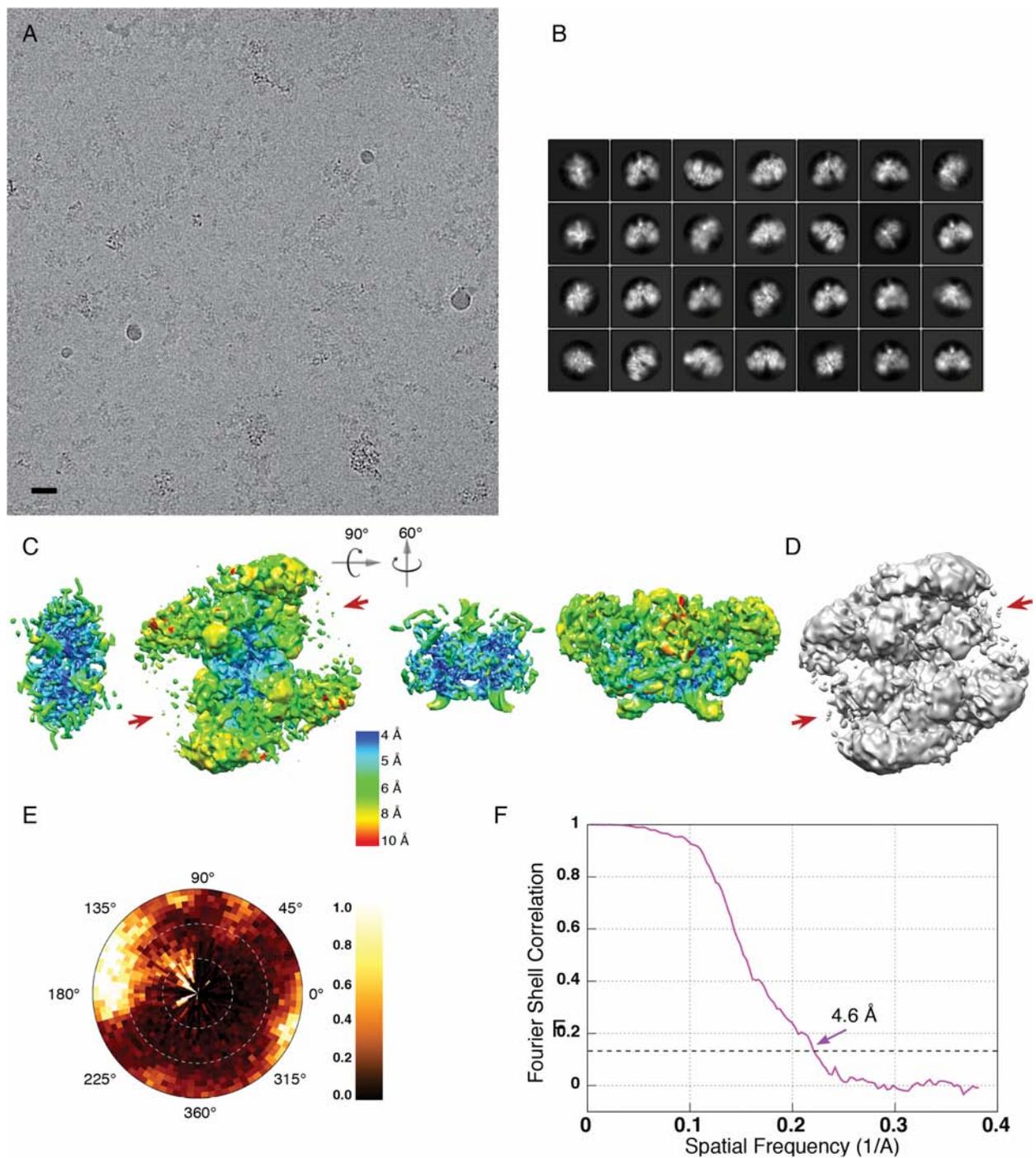


Fig. S9. CryoEM raw data and refinement of HIV-1 STC_{IBD}. (A) Representative cryoEM micrograph of HIV-1 STC_{IBD}, taken at $-2.8 \mu\text{m}$ defocus. (B) 2D class averages calculated using Relion (39). (C) Top and side views of the refined cryoEM map, displayed at two different

thresholds and colored by local resolution using `sxlocres.py` implemented within Sparx (49). Regions at ~ 4 Å are evident in the core of the STC_{IBD} , whereas significantly lower resolution is observed for the outer INs characteristic of the higher-order species. Red arrows indicate heterogeneous regions that are likely sub-stoichiometrically occupied by another IN dimer. **(D)** Top view of the STC_{IBD} reconstruction from (C), filtered to 12 Å and shown at a low contour level to emphasize sub-stoichiometric density likely accounting for additional INs that would make up a complete hexadecamer (16 protomers). **(E)** Euler angle distribution plot, shown as a heat map, describing the relative orientations of the refined map. **(F)** FSC curve describing the global resolution of the map, calculated using a soft-edge mask that occupies the entire density in (D).

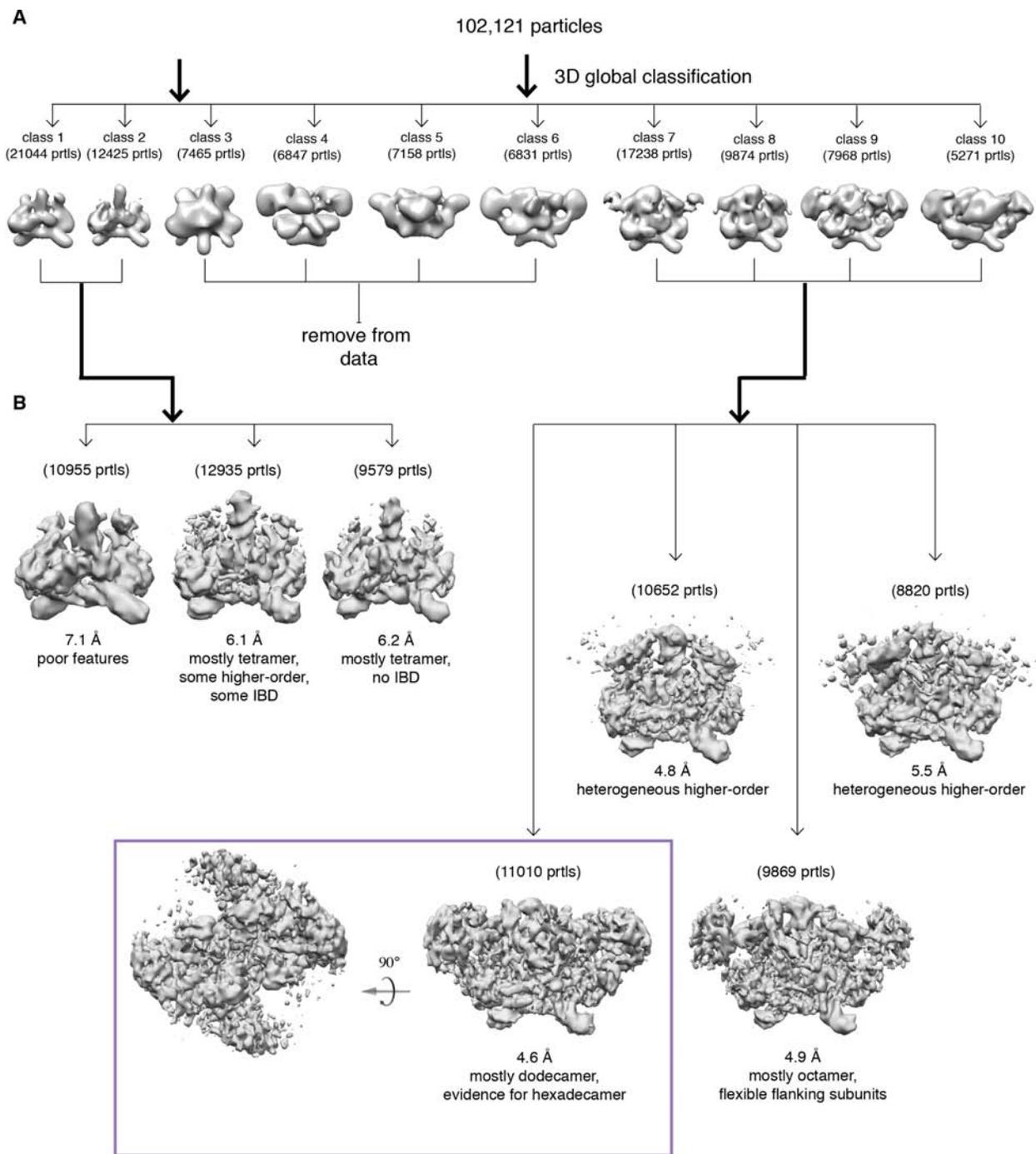
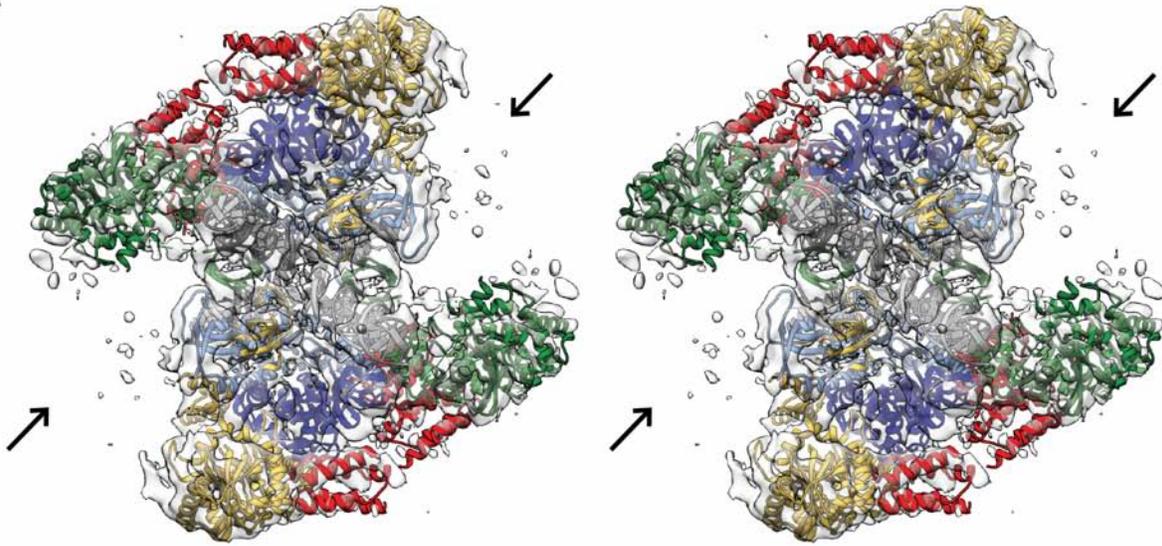


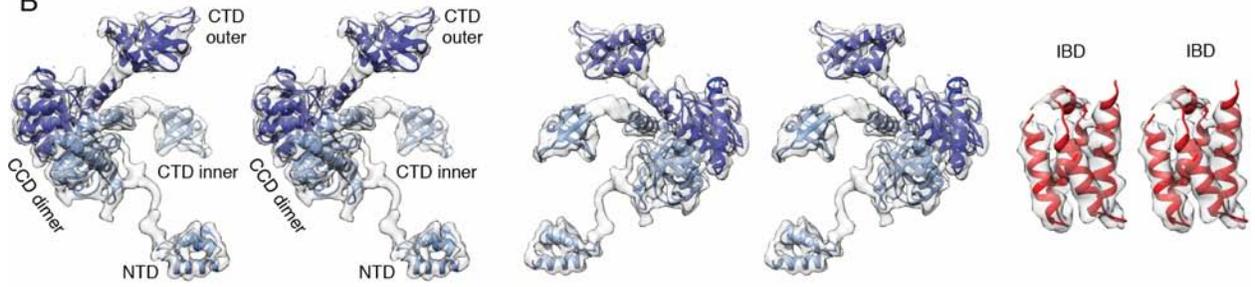
Fig. S10. 3D-classification flowchart of HIV-1 STC_{IBD} cryoEM data. (A) Results of global 3-D classification in Relion. Particles that produced low quality maps (classes 3, 4, 5, 6) were removed from subsequent analysis. Particles apparently belonging to smaller oligomeric maps

(classes 1, 2) and larger oligomeric maps (classes 7, 8, 9, 10) were grouped. **(B)** Orientation parameters for the particles were converted for classification and refinement in FREALIGN. For the smaller oligomers, reconstructions were obtained that were consistent with the size of a tetramer, with one of the maps containing clear density for IBD. Some higher-order species were apparent in this same map, indicating compositional (and likely conformational) heterogeneity. For the larger oligomers, reconstructions were obtained that showed different amounts of INs accumulating around the core. One map in particular contained full density for a dodecamer and additional residual density. We attempted to focus classify this residual density, but we were unable to distinguish reliable features. This map, resolved globally to 4.6 Å, is displayed in Figure 3 and Fig. S9, and was used for all subsequent analyses of STC_{IBD}.

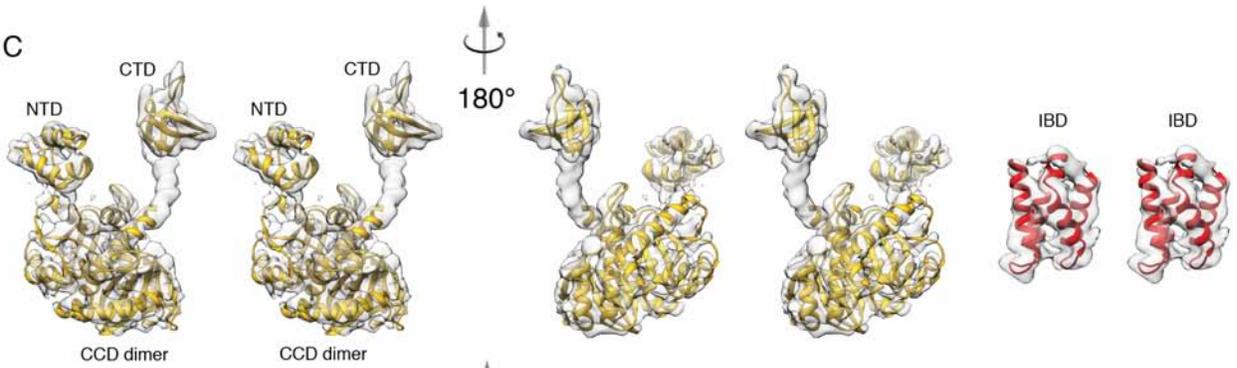
A



B



C



D

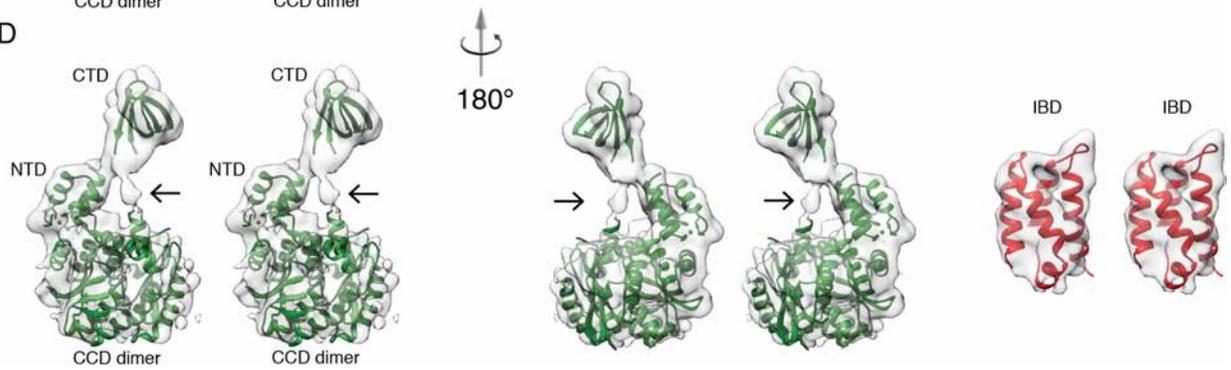


Fig. S11. Segmentation of STC_{IBD} cryoEM density and docking of individual domain components. (A) Composite model of STC_{IBD} rigid-body docked into filtered cryoEM density. Regions denoted by arrows likely represent additional IN dimers, but were very heterogeneous and could not be reliably interpreted, even with high filter values and at low density thresholds. (B). Model of core IN protomers (left, see also Figure 1) and of IBD bound to the core dimer (right) docked into STC_{IBD} cryoEM density, which was filtered to 8 Å. (C) As panel (B), but for the dimers “above” the core. (D) As panel (B), but for the flanking dimers, and with density filtered to 12 Å. Arrows point to the CCD-CTD linker, which contains partially broken density.

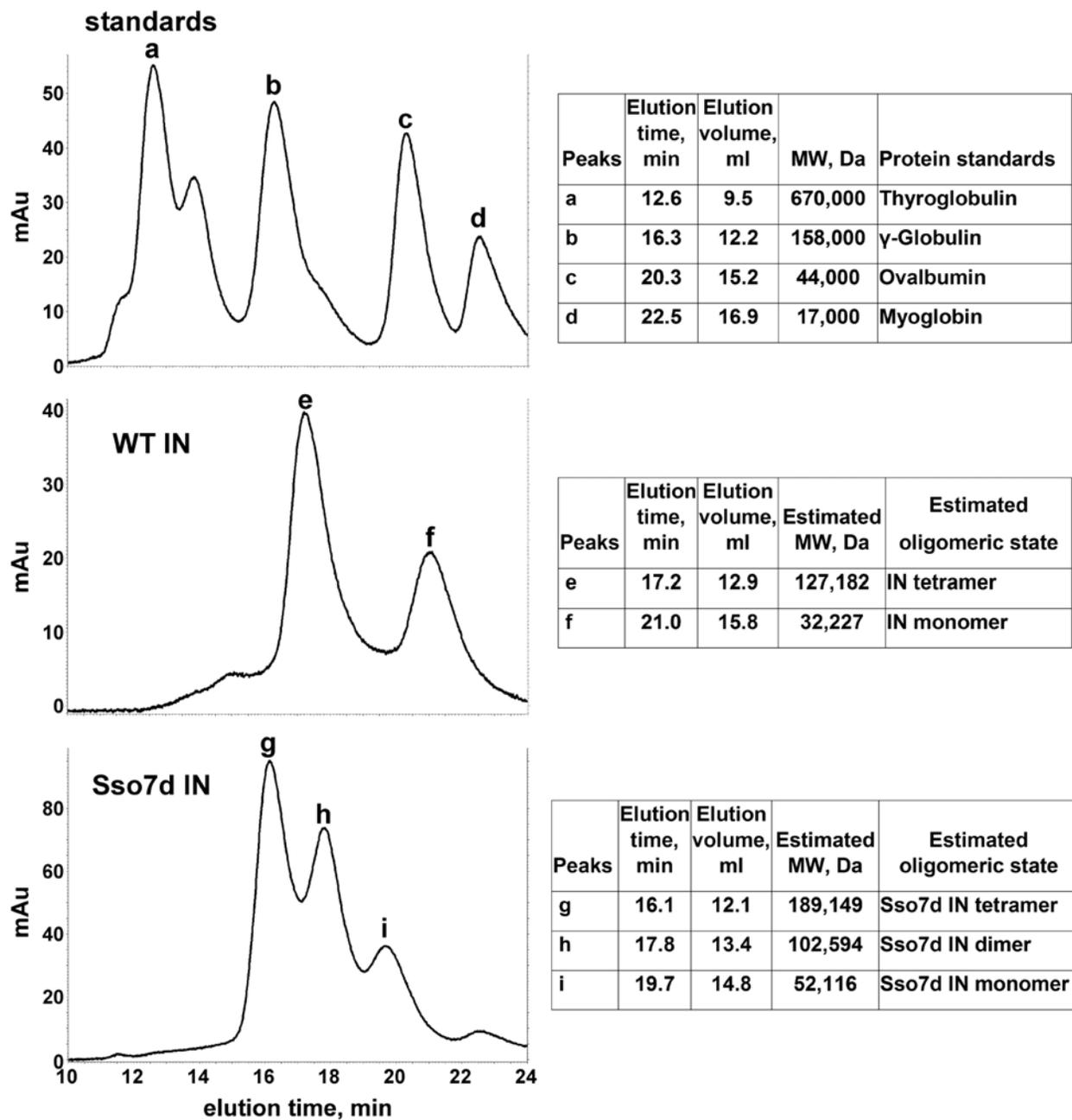


Fig. S12. SEC profiles of WT IN and Sso7d-IN. Upper, middle and lower panels show elution profiles of protein standards, WT IN and Sso7d IN, respectively. Estimated molecular weights (MWs) are indicated in respective tables. Elution times and milli-absorbance units (Au/s) are shown. The Gly-11 flexible linker connecting the NTD with the Sso7d fusion apparently disrupts

productive dimer-dimer associations, leading to a smaller proportion of tetramers and a greater population of dimers within Sso7d IN. This property may account for the large amounts of compositional heterogeneity that we observe within both STC and STC_{IBD} datasets (Fig. S4 and S10), additionally attributing to the sub-stoichiometric (dodecameric vs. hexadecameric) occupancy of flanking IN density within our higher-order cryoEM maps (Figure 3 and Fig. S9).

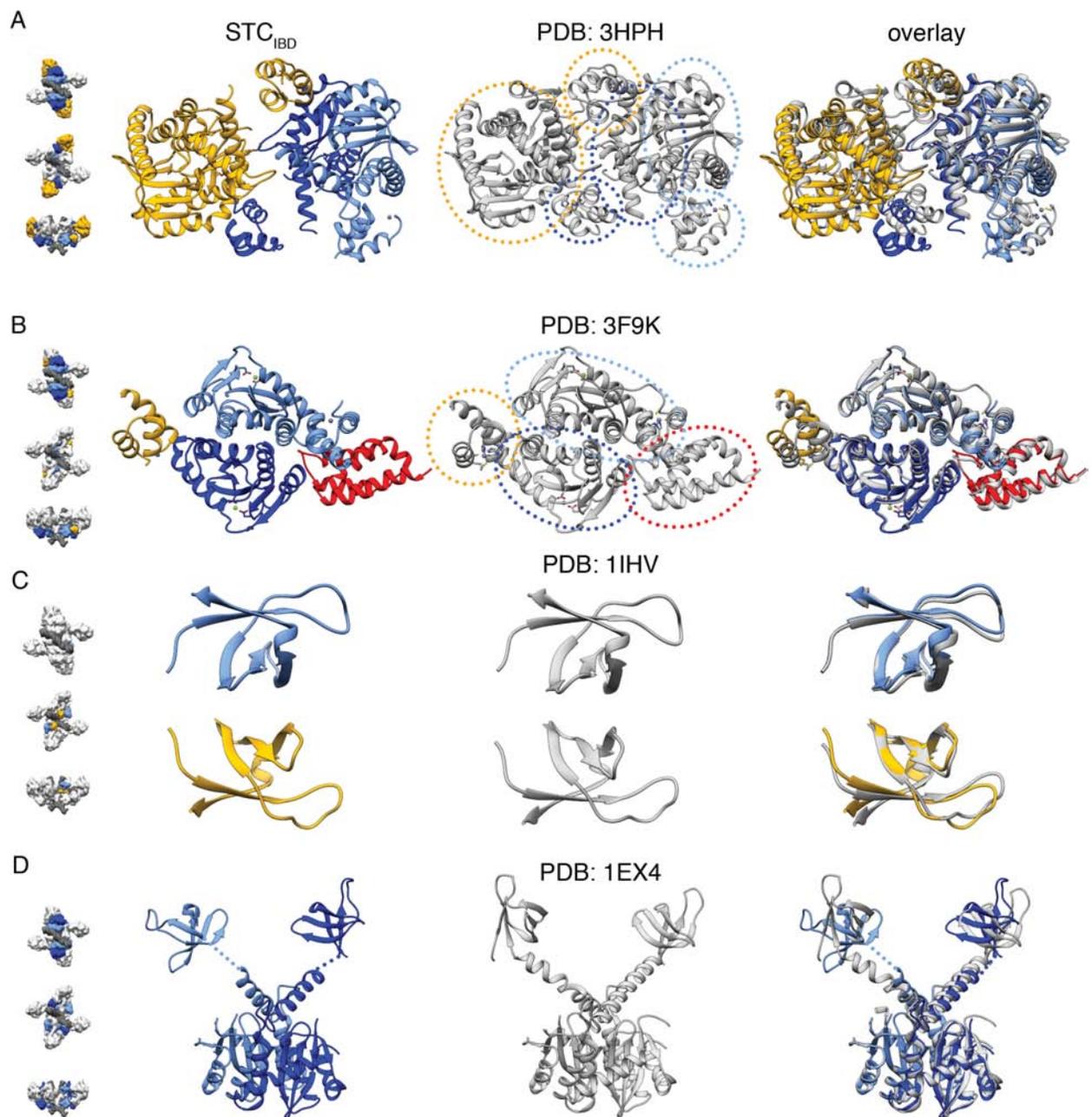


Fig. S13. Comparison of higher-order STC with prior IN structures. In all sections, the left column refers to IN domains docked into STC_{IBD} density and colored accordingly, the center columns refers to a structure in the PDB, and the right column is an overlay of both. Structural schematics at the far left highlight the corresponding location within visible dodecameric

intasome density. This figure expands upon the structures presented in Figure 3D-G. **(A)** Comparison with Maedi Visna Virus two-domain IN_{NTD-CCD} tetramer (PDB: 3HPH, representing Crystal Form 2) (28). IBD, as well as any additional chains that could not be reliably interpreted within the corresponding cryoEM density, are omitted for clarity. **(B)** Comparison with a two domain fragment of HIV-2 IN_{NTD-CCD} (PDB: 3F9K) in complex with IBD (30). **(C)** Comparison with a structure of an HIV-1 IN CTD dimer (PDB: 1IHV) solved by NMR (5). **(D)** Comparison with a two-domain fragment of HIV-1 IN_{CCD-CTD} (PDB: 1EX4).

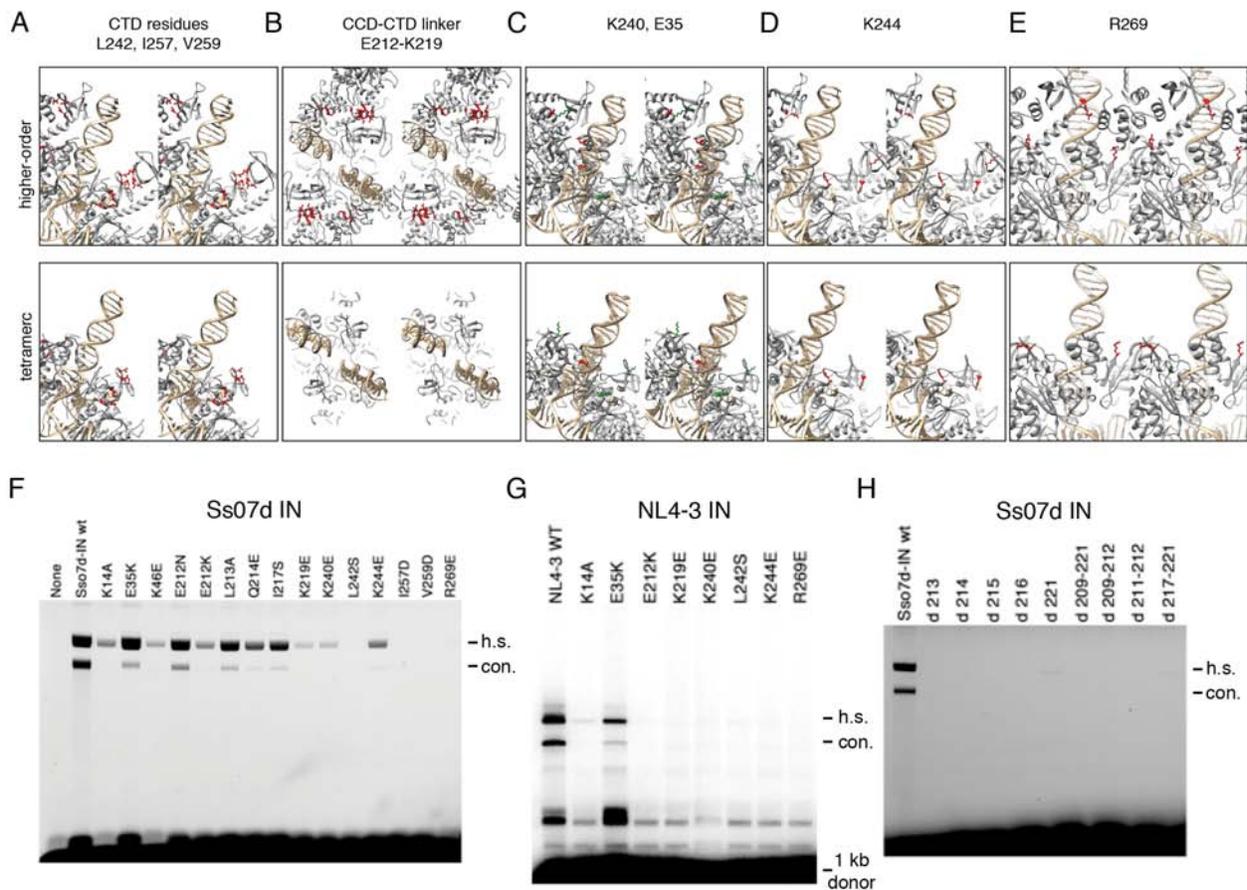


Fig. S14. Strand transfer activity assays of WT and Sso7d INs. (A-E) Residues that were presumed to be relevant within higher-order STCs (top), but not tetrameric STCs (bottom) were selected for strand transfer assays. In addition to those described below, the mutant K14A was selected based on the rationale that it is involved in IN tetramerization (27), and therefore formation of the core building block that would form the higher-order species. (A) Residues within the CTD are buried in several regions within higher-order assemblies, in particular surrounding the CTD-CTD interface in each protomer, but largely surface exposed within tetramers (see also (55) for L242A mutagenesis). (B) Residues within the CCD-CTD linker region are ordered and helical within higher-order STCs, but disordered (and thus not modeled)

within the tetramer. **(C)** K240 (green) is buried in a pocket with E35 (red) within higher-order STCs, but surface exposed within tetramers. E35 can make several possible interactions within higher-order assemblies, but is also largely exposed in tetramers, although we cannot exclude the possibility of its interaction with a flexible CCD-CTD linker. **(D)** K244 is predicted to make a strong salt bridge with vDNA within higher-order complexes in the context of an extended CCD-CTD, but resides at a distance of ~ 5.5 Å from vDNA within tetramers. It can potentially interact with the CCD-CTD linker in both oligomers, although such interactions would be more ordered within higher-order STCs. **(E)** R269 is buried in at least 2 regions of higher-order STCs, but is surface-exposed and/or flexible within tetramers. **(F)** Strand transfer activity assay of selected Sso7d IN mutants presumed to be involved in higher order intasome assembly. The K46E mutant identified within the core, but present in both oligomers, was additionally assayed. **(G)** Strand transfer activity assay of selected NL4-3 IN mutants presumed to be involved in higher order intasome assembly. **(H)** Strand transfer activity assay of Sso7d IN mutants with CCD-CTD linker region deletions. All strand transfer activity assays in panels (F-H) were independently repeated twice.

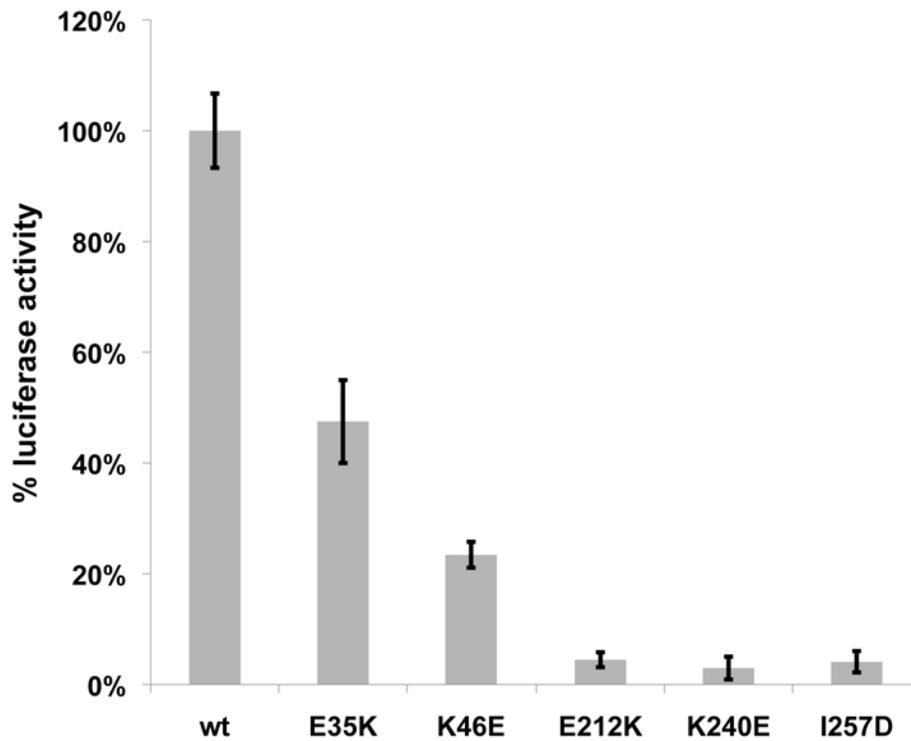
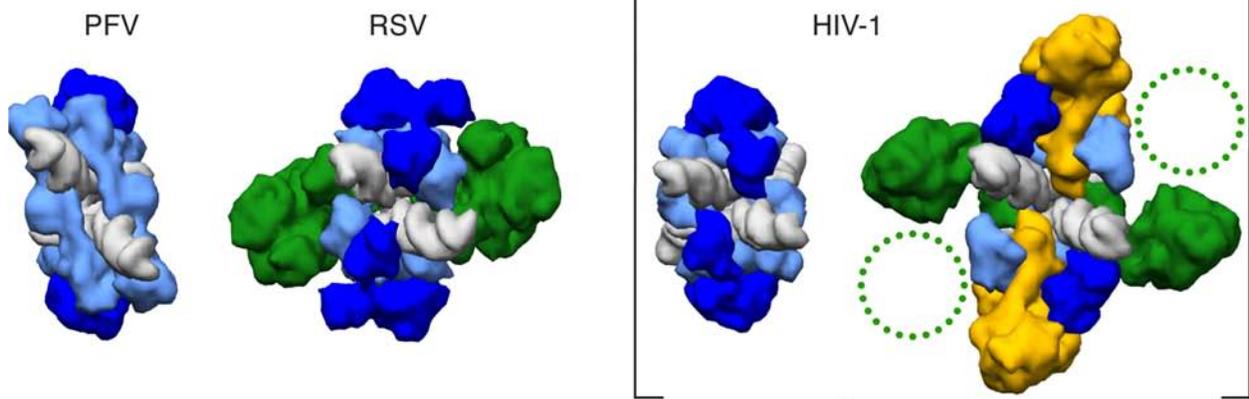


Fig. S15. Viral infectivity. Infectivity for WT HIV-1_{NL4-3} and mutant viruses containing indicated substitutions in the IN coding sequence. Effects of substitutions of K46 that is present within the STC core, as well as of additional select residues predicted to disrupt higher-order assemblies (see Fig. S14) that were not previously evaluated (31-33) have been examined. Luciferase activities in infected TZM-bl cells relative to WT virus are shown. Error bars represent standard deviations of three independent experiments.

A

	PDB code	residues	Distance between CCD and CTD (Å)	maximum length of linker (Å)
PFV	3OS0	272-322	52.4	175
PFV	4BAC	272-322	52.6	175
HIV-1	this study	203-223	43.9	
RSV	5EJK	214-223	46.2	31.5

B



C

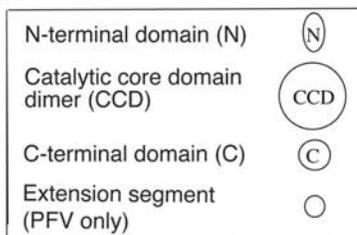
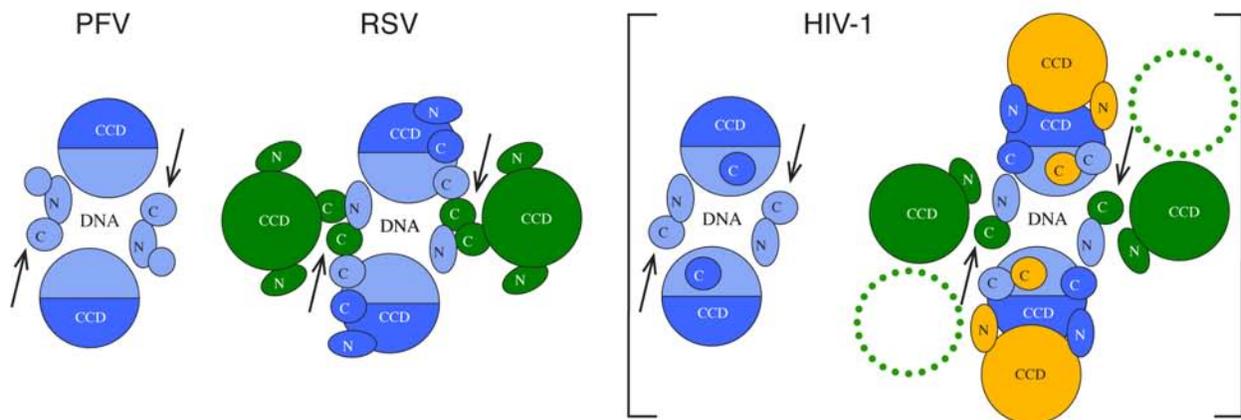


Fig. S16. Comparison of retroviral STCs. (A) Linker analyses for all three retroviral STCs for which structures are available. The distance between CCD and CTD is calculated for the C α atoms between the residues specified for the particular structure. The maximum linker length assumes a distance of 3.5 Å per residue, corresponding to a perfect β -strand (larger than the expected length of a flexible linker). The HIV-1 CCD-CTD linker is helical in the context of higher-order STCs or previous structures (e.g. PDB 1EX4), but disordered in STC tetramers and is just long enough to reach into the synaptic position. MMTV or RSV must employ flanking INs for proper CTD positioning. (B). Retroviral STCs from PFV (PDB: 3OS0) RSV (PDB: 5EJK) and HIV (modeled tetramer and higher-order dodecameric complex from this study) are filtered to 10 Å and displayed as molecular surfaces to show the relative orientations and sizes of experimentally derived structures. MMTV intasomes are omitted, because no STC structure is available. (C) Schematic of quaternary intasome assembly of different retroviruses. Positionally conserved CTDs are indicated with an arrow. Legend includes CCD, catalytic core domain dimer; C, C-terminal domain monomer; N, N-terminal domain monomer. CCD dimers surrounding DNA within the intasome core are colored light and dark blue for the inner and outer protomer, respectively. All other CCD dimers are colored in a single color. NTDs or CTDs that are disordered either in the X-ray structures or the EM structures are not displayed.

Construct	HIV-1 STC	HIV-1 STC _{IBD}
EM Image Processing		
Microscope	Titan Krios	Titan Krios
Voltage	300	300
Camera	Gatan K2 Summit	Gatan K2 Summit
Defocus range (μm)	1.5-4.0	1.5-3.5
Defocus mean ± std (μm)	2.8 ± 0.6	2.5 ± 0.5
Exposure time (s)	20	20
Dose rate (e-/pixel/s)	8.1	6.3
Total dose (e-/Å ²)	95	80
Pixel size (Å)	1.31	1.31
Number of micrographs	1225	1598
Number of particles (processed)	274,764	154,445
Number of particles (3D classification & refinement)	219,260	102,121
Number of particles (in final map)	83,766	11,099
Symmetry	C2	C2
Resolution (global) (Å)*	3.9	4.6
Resolution range (local) (Å)	3.5 – 4.5	4.0 – 10.0
Map sharpening B-factor (Å ²)	-180	-200
Model refinement		
Space group	P1	-
Cell dimensions		
<i>a</i> , <i>b</i> , <i>c</i> (Å)	326.1, 374.4, 340.4	-
<i>a</i> = <i>b</i> = <i>c</i> (°)	90	-
Number of atoms (modeled)	8596	-
Validation		
MolProbity score	1.7 ± 0.1 (89 th ± 2 percentile)	-
Clashscore, all atoms	5.6 ± 0.3 (88 th ± 2 percentile)	-
Protein		
Ramachandran favored (%)	93.4 ± 0.4	-
allowed (%)	6.3 ± 0.3	-
Disallowed (%)	0.3 ± 0.1	-
Good rotamers (%)	99.7 ± 0.2	-
Cβ deviations >0.25Å (%)	0	-
Cis Prolines (%)	0	-
r.m.s deviations		
Bond lengths (Å)	0.008 ± 0.001	-
Bond angles (°)	1.1 ± 0.1	-

*resolution assessment based on frequency-limited refinement using the 0.143 threshold for resolution analysis

Table S1. CryoEM data collection and model building statistics.

RESIDUE-vDNA INTERACTIONS						
Residue	Donor	Nucleotide	Acceptor	D..A dist	(SD)	Type
LYS	46.A NZ	T	17.F O2	2.8	0.0	sc - B
HIS	51.A ND1	G	18.I OP1	3.7	0.0	sc - bb
GLY	52.A N	T	17.I O3'	3.7	0.0	bb - bb
GLY	52.A N	G	18.I O5'	3.6	0.0	bb - bb
GLY	52.A N	G	18.I OP1	2.5	0.1	bb - bb
GLN	53.A N	T	17.I O2	3.6	0.2	bb - B
VAL	54.A N	G	18.I O3'	3.8	0.0	bb - bb
VAL	54.A N	C	19.I OP1	3.6	0.0	bb - bb
ILE	141.A N	T	17.I OP1	2.7	0.0	bb - bb
SER	147.A OG	T	17.I OP1	3.8	0.1	sc - bb
SER	153.A OG	T	20.I O2	3.4	0.2	sc - B
SER	153.A OG	T	20.I O4'	2.9	0.2	sc - bb
LYS	159.A NZ	A	21.J OP2	2.3	0.0	sc - bb
LYS	160.A NZ	G	22.I OP1	3.5	0.1	sc - bb
ARG	187.A NH1	G	22.I OP2	3.7	0.1	sc - bb
ARG	228.B NH1	C	19.I OP1	2.6	0.3	sc - bb
ARG	228.B NH2	C	19.I OP1	3.8	0.1	sc - bb
ARG	231.B NH1	G	24.E O3'	3.7	0.2	sc - bb
ARG	231.B NH1	A	25.E O5'	3.1	0.3	sc - bb
ARG	231.B NH1	A	25.E OP1	3.5	0.1	sc - bb
GLU	246.A N	C	16.I N3	3.2	0.0	bb - B
ALA	248.A N	C	16.I O2	2.7	0.1	bb - B
ARG	263.B NH1	T	17.I O4	3.2	0.2	sc - B
ARG	263.A NH1	A	18.F OP1	2.7	0.0	sc - bb
ARG	263.A NH2	G	18.I OP2	3.3	0.0	sc - bb
ARG	263.A NE	G	18.I OP1	3.6	0.1	sc - bb
LYS	266.A NZ	A	23.E O5'	3.6	0.4	sc - bb
LYS	266.A NZ	A	23.E OP1	3.4	0.0	sc - bb
LYS	266.A NZ	A	23.E OP2	3.7	0.0	sc - bb

RESIDUE-tDNA INTERACTIONS						
Residue	Donor	Nucleotide	Acceptor	D..A dist	(SD)	Type
GLY	94.A N	C	30.F OP1	3.3	0.3	bb - bb
ASN	117.A N	T	11.G OP1	3.7	0.0	bb - bb
SER	119.A OG	T	29.F O2	3.7	0.3	sc - B
SER	119.A OG	T	29.F O4'	3.5	0.3	sc - bb
ASN	120.A ND2	T	29.F O4'	3.9	0.0	sc - bb
THR	124.A N	A	31.F OP1	3.3	0.2	bb - bb
ARG	231.A NH1	G	22.J O6	3.8	0.3	sc - B
ARG	231.A NH1	T	23.J O4	3.9	0.0	sc - B
ARG	231.A NH2	G	22.J N7	3.3	0.2	sc - B

INTER-DOMAIN INTERACTIONS						
Residue_d	Donor	Residue_a	Acceptor	D..A dist	(SD)	Type
NTD						
LYS	14.A NZ	TRP	131.D O	3.4	0.3	sc - bb
LYS	14.A NZ	TRP	132.D O	3.8	0.1	sc - bb
TYR	15.A OH	TRP	132.D O	2.3	0.2	sc - bb
ASN	18.A ND2	LYS	188.C O	3.0	0.1	sc - bb
ARG	20.A NE	LYS	188.C O	2.9	0.2	sc - bb
ARG	20.A NH2	LYS	188.C O	2.8	0.1	sc - bb
LYS	42.A NZ	ASP	167.C OD1	3.6	0.0	sc - sc
LYS	42.A NZ	ASP	167.C OD2	2.6	0.1	sc - sc

Linker (NTD-CCD)						
GLU	48.A N	ARG	262.A O	3.6	0.0	bb - bb
GLN	53.A NE2	ASP	229.B O	2.5	0.0	sc - bb
CYS	56.A N	ARG	263.B O	3.2	0.3	bb - bb

CCD						
TRP	132.B NE1	LYS	14.C O	3.5	0.3	sc - bb
GLN	168.A NE2	LYS	14.C O	3.6	0.1	sc - bb
ARG	187.A NE	HIS	16.C O	3.3	0.3	sc - bb
ARG	187.A NH2	HIS	16.C O	2.6	0.1	sc - bb
LYS	188.A N	ASN	18.C OD1	3.9	0.0	bb - sc
LYS	188.A NZ	ASP	25.C OD1	2.9	0.0	sc - sc

CTD						
ARG	262.A NE	GLU	48.A OE1	2.6	0.0	sc - sc
ARG	262.A NE	GLU	48.A OE2	3.7	0.1	sc - sc
ARG	262.A NH2	GLU	48.A OE1	2.9	0.0	sc - sc
ARG	262.A NH2	GLU	48.A OE2	3.5	0.1	sc - sc
ARG	263.A NH2	GLN	146.A OE1	3.3	0.3	sc - sc
LYS	266.B NZ	ALA	80.A O	2.8	0.1	sc - sc

LEGEND		LEGEND	
NTD Domain	Linker (NTD-CCD)	D..A dist	- distance between donor-acceptor
CCD Domain	CTD Domain	SD	- standard deviation
		bb	- backbone
		sc	- side chain
		B	- nucleotide base

Table S2. Predicted hydrogen-bonding interactions within tetrameric core STC. H-bonds were computed within UCSF Chimera (56), and were considered only if they were identified in all 10 real-space refined models and contained corresponding density within the map. Non-H-bonding interactions were omitted from this analysis. Additional functionally relevant interactions likely exist, but may not be clearly identified at the nominal resolution of the map, or differed within the models. IN residues sharing H-bonds with either vDNA or tDNA are listed at left. The panels at right list predicted protein inter-domain interactions. Each domain is colored in agreement with the legend. The colored boxes above each panel represent the domains comprising the donor residues (Residues_d) while the colored shades underneath represent the

domains for the acceptor residues (Residue_a). Appropriate chains of the atomic model are indicated. Two CTD chains are present, including the inner CTD (chain A), and an outer CTD (chain B). The distance between donor and acceptor (D..A dist) is in Angstrom (\AA), and is listed as the average +/- standard deviation (SD) of ten independent models refined in Phenix. In instances when individual donor residues may make several possible interactions with acceptor residues/nucleotides that were not distinguishable in the map or models, both possibilities are listed.

Contributions: M.L. and R.C. discovered how to assemble HIV-1 STCs; M.L., R.C., and R.Y. purified STCs; R.G. performed analytical ultracentrifugation experiments; D.P. and Y.J. optimized cryoEM vitrification conditions and screened grids for data collection; D.P., Y.J., and D.L. collected cryoEM data; D.P. and D.L. processed cryoEM data. D.L. built and refined the atomic model; M.L. and R.Y. performed the in vitro DNA integration assays. S.V.R., N.S. and M.K. performed virology, crosslinking and size exclusion chromatography experiments; D.L., D.P., and R.C. wrote the manuscript; all authors contributed to manuscript editing.

References and Notes

1. R. Craigie, F. D. Bushman, HIV DNA integration. *Cold Spring Harb. Perspect. Med.* **2**, a006890 (2012). [doi:10.1101/cshperspect.a006890](https://doi.org/10.1101/cshperspect.a006890) [Medline](#)
2. P. Lesbats, A. N. Engelman, P. Cherepanov, Retroviral DNA integration. *Chem. Rev.* **116**, 12730–12757 (2016). [doi:10.1021/acs.chemrev.6b00125](https://doi.org/10.1021/acs.chemrev.6b00125) [Medline](#)
3. D. J. Hazuda, HIV integrase as a target for antiretroviral therapy. *Curr. Opin. HIV AIDS* **7**, 383–389 (2012). [doi:10.1097/COH.0b013e3283567309](https://doi.org/10.1097/COH.0b013e3283567309) [Medline](#)
4. F. Dyda, A. B. Hickman, T. M. Jenkins, A. Engelman, R. Craigie, D. R. Davies, Crystal structure of the catalytic domain of HIV-1 integrase: Similarity to other polynucleotidyl transferases. *Science* **266**, 1981–1986 (1994). [doi:10.1126/science.7801124](https://doi.org/10.1126/science.7801124) [Medline](#)
5. A. P. Eijkelenboom, R. A. Lutzke, R. Boelens, R. H. A. Plasterk, R. Kaptein, K. Hård, The DNA-binding domain of HIV-1 integrase has an SH3-like fold. *Nat. Struct. Biol.* **2**, 807–810 (1995). [doi:10.1038/nsb0995-807](https://doi.org/10.1038/nsb0995-807) [Medline](#)
6. M. Cai, R. Zheng, M. Caffrey, R. Craigie, G. M. Clore, A. M. Gronenborn, Solution structure of the N-terminal zinc binding domain of HIV-1 integrase. *Nat. Struct. Biol.* **4**, 567–577 (1997). [doi:10.1038/nsb0797-567](https://doi.org/10.1038/nsb0797-567) [Medline](#)
7. J. Y. Wang, H. Ling, W. Yang, R. Craigie, Structure of a two-domain fragment of HIV-1 integrase: Implications for domain organization in the intact protein. *EMBO J.* **20**, 7333–7343 (2001). [doi:10.1093/emboj/20.24.7333](https://doi.org/10.1093/emboj/20.24.7333) [Medline](#)
8. J. C. H. Chen, J. Krucinski, L. J. W. Miercke, J. S. Finer-Moore, A. H. Tang, A. D. Leavitt, R. M. Stroud, Crystal structure of the HIV-1 integrase catalytic core and C-terminal domains: A model for viral DNA binding. *Proc. Natl. Acad. Sci. U.S.A.* **97**, 8233–8238 (2000). [doi:10.1073/pnas.150220297](https://doi.org/10.1073/pnas.150220297) [Medline](#)
9. S. Hare, S. S. Gupta, E. Valkov, A. Engelman, P. Cherepanov, Retroviral intasome assembly and inhibition of DNA strand transfer. *Nature* **464**, 232–236 (2010). [doi:10.1038/nature08784](https://doi.org/10.1038/nature08784) [Medline](#)
10. G. N. Maertens, S. Hare, P. Cherepanov, The mechanism of retroviral integration from x-ray structures of its key intermediates. *Nature* **468**, 326–329 (2010). [doi:10.1038/nature09517](https://doi.org/10.1038/nature09517) [Medline](#)
11. A. Ballandras-Colas, M. Brown, N. J. Cook, T. G. Dewdney, B. Demeler, P. Cherepanov, D. Lyumkis, A. N. Engelman, Cryo-EM reveals a novel octameric integrase structure for betaretroviral intasome function. *Nature* **530**, 358–361 (2016). [doi:10.1038/nature16955](https://doi.org/10.1038/nature16955) [Medline](#)
12. Z. Yin, K. Shi, S. Banerjee, K. K. Pandey, S. Bera, D. P. Grandgenett, H. Aihara, Crystal structure of the Rous sarcoma virus intasome. *Nature* **530**, 362–366 (2016). [doi:10.1038/nature16950](https://doi.org/10.1038/nature16950) [Medline](#)
13. L. Krishnan, X. Li, H. L. Naraharisetty, S. Hare, P. Cherepanov, A. Engelman, Structure-based modeling of the functional HIV-1 intasome and its inhibition. *Proc. Natl. Acad. Sci. U.S.A.* **107**, 15910–15915 (2010). [doi:10.1073/pnas.1002346107](https://doi.org/10.1073/pnas.1002346107) [Medline](#)

14. B. C. Johnson, M. Métifiot, A. Ferris, Y. Pommier, S. H. Hughes, A homology model of HIV-1 integrase and analysis of mutations designed to test the model. *J. Mol. Biol.* **425**, 2133–2146 (2013). [doi:10.1016/j.jmb.2013.03.027](https://doi.org/10.1016/j.jmb.2013.03.027) [Medline](#)
15. M. Li, K. A. Jurado, S. Lin, A. Engelman, R. Craigie, Engineered hyperactive integrase for concerted HIV-1 DNA integration. *PLOS ONE* **9**, e105078 (2014). [doi:10.1371/journal.pone.0105078](https://doi.org/10.1371/journal.pone.0105078) [Medline](#)
16. Z. Yin, M. Lapkouski, W. Yang, R. Craigie, Assembly of prototype foamy virus strand transfer complexes on product DNA bypassing catalysis of integration. *Protein Sci.* **21**, 1849–1857 (2012). [doi:10.1002/pro.2166](https://doi.org/10.1002/pro.2166) [Medline](#)
17. M. Li, M. Mizuuchi, T. R. Burke Jr., R. Craigie, Retroviral DNA integration: Reaction pathway and critical intermediates. *EMBO J.* **25**, 1295–1304 (2006). [doi:10.1038/sj.emboj.7601005](https://doi.org/10.1038/sj.emboj.7601005) [Medline](#)
18. S. Kotova, M. Li, E. K. Dimitriadis, R. Craigie, Nucleoprotein intermediates in HIV-1 DNA integration visualized by atomic force microscopy. *J. Mol. Biol.* **399**, 491–500 (2010). [doi:10.1016/j.jmb.2010.04.026](https://doi.org/10.1016/j.jmb.2010.04.026) [Medline](#)
19. S. Bera, K. K. Pandey, A. C. Vora, D. P. Grandgenett, Molecular interactions between HIV-1 integrase and the two viral DNA ends within the synaptic complex that mediates concerted integration. *J. Mol. Biol.* **389**, 183–198 (2009). [doi:10.1016/j.jmb.2009.04.007](https://doi.org/10.1016/j.jmb.2009.04.007) [Medline](#)
20. T. Grant, N. Grigorieff, Measuring the optimal exposure for single particle cryo-EM using a 2.6 Å reconstruction of rotavirus VP6. *eLife* **4**, e06980 (2015). [doi:10.7554/eLife.06980](https://doi.org/10.7554/eLife.06980) [Medline](#)
21. D. Derse, B. Crise, Y. Li, G. Princler, N. Lum, C. Stewart, C. F. McGrath, S. H. Hughes, D. J. Munroe, X. Wu, 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). [doi:10.1128/JVI.02752-06](https://doi.org/10.1128/JVI.02752-06) [Medline](#)
22. A. Ballandras-Colas, H. Naraharisetty, X. Li, E. Serrao, A. Engelman, Biochemical characterization of novel retroviral integrase proteins. *PLOS ONE* **8**, e76638 (2013). [doi:10.1371/journal.pone.0076638](https://doi.org/10.1371/journal.pone.0076638) [Medline](#)
23. T. M. Jenkins, D. Esposito, A. Engelman, R. Craigie, Critical contacts between HIV-1 integrase and viral DNA identified by structure-based analysis and photo-crosslinking. *EMBO J.* **16**, 6849–6859 (1997). [doi:10.1093/emboj/16.22.6849](https://doi.org/10.1093/emboj/16.22.6849) [Medline](#)
24. A. Chen, I. T. Weber, R. W. Harrison, J. Leis, Identification of amino acids in HIV-1 and avian sarcoma virus integrase subsites required for specific recognition of the long terminal repeat ends. *J. Biol. Chem.* **281**, 4173–4182 (2006). [doi:10.1074/jbc.M510628200](https://doi.org/10.1074/jbc.M510628200) [Medline](#)
25. E. Serrao, L. Krishnan, M.-C. Shun, X. Li, P. Cherepanov, A. Engelman, G. N. Maertens, Integrase residues that determine nucleotide preferences at sites of HIV-1 integration: Implications for the mechanism of target DNA binding. *Nucleic Acids Res.* **42**, 5164–5176 (2014). [doi:10.1093/nar/gku136](https://doi.org/10.1093/nar/gku136) [Medline](#)

26. A. L. Harper, M. Sudol, M. Katzman, An amino acid in the central catalytic domain of three retroviral integrases that affects target site selection in nonviral DNA. *J. Virol.* **77**, 3838–3845 (2003). [doi:10.1128/JVI.77.6.3838-3845.2003](https://doi.org/10.1128/JVI.77.6.3838-3845.2003) [Medline](#)
27. C. J. McKee, J. J. Kessl, N. Shkriabai, M. J. Dar, A. Engelman, M. Kvaratskhelia, Dynamic modulation of HIV-1 integrase structure and function by cellular lens epithelium-derived growth factor (LEDGF) protein. *J. Biol. Chem.* **283**, 31802–31812 (2008). [doi:10.1074/jbc.M805843200](https://doi.org/10.1074/jbc.M805843200) [Medline](#)
28. S. Hare, F. Di Nunzio, A. Labeja, J. Wang, A. Engelman, P. Cherepanov, Structural basis for functional tetramerization of lentiviral integrase. *PLOS Pathog.* **5**, e1000515 (2009). [doi:10.1371/journal.ppat.1000515](https://doi.org/10.1371/journal.ppat.1000515) [Medline](#)
29. P. Cherepanov, G. Maertens, P. Proost, B. Devreese, J. Van Beeumen, Y. Engelborghs, E. De Clercq, Z. Debyser, HIV-1 integrase forms stable tetramers and associates with LEDGF/p75 protein in human cells. *J. Biol. Chem.* **278**, 372–381 (2003). [doi:10.1074/jbc.M209278200](https://doi.org/10.1074/jbc.M209278200) [Medline](#)
30. S. Hare, M.-C. Shun, S. S. Gupta, E. Valkov, A. Engelman, P. Cherepanov, 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). [doi:10.1371/journal.ppat.1000259](https://doi.org/10.1371/journal.ppat.1000259) [Medline](#)
31. R. Lu, H. Z. Ghory, A. Engelman, Genetic analyses of conserved residues in the carboxyl-terminal domain of human immunodeficiency virus type 1 integrase. *J. Virol.* **79**, 10356–10368 (2005). [doi:10.1128/JVI.79.16.10356-10368.2005](https://doi.org/10.1128/JVI.79.16.10356-10368.2005) [Medline](#)
32. R. Lu, A. Limón, H. Z. Ghory, A. Engelman, Genetic analyses of DNA-binding mutants in the catalytic core domain of human immunodeficiency virus type 1 integrase. *J. Virol.* **79**, 2493–2505 (2005). [doi:10.1128/JVI.79.4.2493-2505.2005](https://doi.org/10.1128/JVI.79.4.2493-2505.2005) [Medline](#)
33. S. J. Rihn, J. Hughes, S. J. Wilson, P. D. Bieniasz, Uneven genetic robustness of HIV-1 integrase. *J. Virol.* **89**, 552–567 (2015). [doi:10.1128/JVI.02451-14](https://doi.org/10.1128/JVI.02451-14) [Medline](#)
34. R. Lu, N. Vandegraaff, P. Cherepanov, A. Engelman, Lys-34, dispensable for integrase catalysis, is required for preintegration complex function and human immunodeficiency virus type 1 replication. *J. Virol.* **79**, 12584–12591 (2005). [doi:10.1128/JVI.79.19.12584-12591.2005](https://doi.org/10.1128/JVI.79.19.12584-12591.2005) [Medline](#)
35. A. Ballandras-Colas, D. P. Maskell, E. Serrao, J. Locke, P. Swuec, S. R. Jónsson, A. Kotecha, N. J. Cook, V. E. Pye, I. A. Taylor, V. Andrésdóttir, A. N. Engelman, A. Costa, P. Cherepanov, A supramolecular assembly mediates lentiviral DNA integration. *Science* **355**, 93–95 (2017).
36. C. Suloway, J. Pulokas, D. Fellmann, A. Cheng, F. Guerra, J. Quispe, S. Stagg, C. S. Potter, B. Carragher, Automated molecular microscopy: The new Leginon system. *J. Struct. Biol.* **151**, 41–60 (2005). [doi:10.1016/j.jsb.2005.03.010](https://doi.org/10.1016/j.jsb.2005.03.010) [Medline](#)
37. G. C. Lander, S. M. Stagg, N. R. Voss, A. Cheng, D. Fellmann, J. Pulokas, C. Yoshioka, C. Irving, A. Mulder, P.-W. Lau, D. Lyumkis, C. S. Potter, B. Carragher, Appion: An integrated, database-driven pipeline to facilitate EM image processing. *J. Struct. Biol.* **166**, 95–102 (2009). [doi:10.1016/j.jsb.2009.01.002](https://doi.org/10.1016/j.jsb.2009.01.002) [Medline](#)

38. C. O. S. Sorzano, J. R. Bilbao-Castro, Y. Shkolnisky, M. Alcorlo, R. Melero, G. Caffarena-Fernández, M. Li, G. Xu, R. Marabini, J. M. Carazo, A clustering approach to multireference alignment of single-particle projections in electron microscopy. *J. Struct. Biol.* **171**, 197–206 (2010). [doi:10.1016/j.jsb.2010.03.011](https://doi.org/10.1016/j.jsb.2010.03.011) [Medline](#)
39. S. H. W. Scheres, RELION: Implementation of a Bayesian approach to cryo-EM structure determination. *J. Struct. Biol.* **180**, 519–530 (2012). [doi:10.1016/j.jsb.2012.09.006](https://doi.org/10.1016/j.jsb.2012.09.006) [Medline](#)
40. D. Lyumkis, S. Vinterbo, C. S. Potter, B. Carragher, Optimod – An automated approach for constructing and optimizing initial models for single-particle electron microscopy. *J. Struct. Biol.* **184**, 417–426 (2013). [doi:10.1016/j.jsb.2013.10.009](https://doi.org/10.1016/j.jsb.2013.10.009) [Medline](#)
41. S. H. Scheres, A Bayesian view on cryo-EM structure determination. *J. Mol. Biol.* **415**, 406–418 (2012). [doi:10.1016/j.jmb.2011.11.010](https://doi.org/10.1016/j.jmb.2011.11.010) [Medline](#)
42. D. Lyumkis, A. F. Brilot, D. L. Theobald, N. Grigorieff, Likelihood-based classification of cryo-EM images using FREALIGN. *J. Struct. Biol.* **183**, 377–388 (2013). [doi:10.1016/j.jsb.2013.07.005](https://doi.org/10.1016/j.jsb.2013.07.005) [Medline](#)
43. N. Grigorieff, FREALIGN: High-resolution refinement of single particle structures. *J. Struct. Biol.* **157**, 117–125 (2007). [doi:10.1016/j.jsb.2006.05.004](https://doi.org/10.1016/j.jsb.2006.05.004) [Medline](#)
44. S. L. Ilca, A. Kotecha, X. Sun, M. M. Poranen, D. I. Stuart, J. T. Huiskonen, Localized reconstruction of subunits from electron cryomicroscopy images of macromolecular complexes. *Nat. Commun.* **6**, 8843 (2015). [doi:10.1038/ncomms9843](https://doi.org/10.1038/ncomms9843) [Medline](#)
45. F. DiMaio, Y. Song, X. Li, M. J. Brunner, C. Xu, V. Conticello, E. Egelman, T. C. Marlovits, Y. Cheng, D. Baker, Atomic-accuracy models from 4.5-Å cryo-electron microscopy data with density-guided iterative local refinement. *Nat. Methods* **12**, 361–365 (2015). [doi:10.1038/nmeth.3286](https://doi.org/10.1038/nmeth.3286) [Medline](#)
46. P. D. Adams, P. V. Afonine, G. Bunkóczi, V. B. Chen, I. W. Davis, N. Echols, J. J. Headd, L.-W. Hung, G. J. Kapral, R. W. Grosse-Kunstleve, A. J. McCoy, N. W. Moriarty, R. Oeffner, R. J. Read, D. C. Richardson, J. S. Richardson, T. C. Terwilliger, P. H. Zwart, PHENIX: A comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D* **66**, 213–221 (2010). [doi:10.1107/S0907444909052925](https://doi.org/10.1107/S0907444909052925) [Medline](#)
47. P. Emsley, B. Lohkamp, W. G. Scott, K. Cowtan, Features and development of Coot. *Acta Crystallogr. D* **66**, 486–501 (2010). [doi:10.1107/S0907444910007493](https://doi.org/10.1107/S0907444910007493) [Medline](#)
48. I. W. Davis, A. Leaver-Fay, V. B. Chen, J. N. Block, G. J. Kapral, X. Wang, L. W. Murray, W. B. Arendall III, J. Snoeyink, J. S. Richardson, D. C. Richardson, MolProbity: All-atom contacts and structure validation for proteins and nucleic acids. *Nucleic Acids Res.* **35** (suppl. 2), W375–W383 (2007). [doi:10.1093/nar/gkm216](https://doi.org/10.1093/nar/gkm216) [Medline](#)
49. M. Hohn, G. Tang, G. Goodyear, P. R. Baldwin, Z. Huang, P. A. Penczek, C. Yang, R. M. Glaeser, P. D. Adams, S. J. Ludtke, SPARX, a new environment for xryo-EM image processing. *J. Struct. Biol.* **157**, 47–55 (2007). [doi:10.1016/j.jsb.2006.07.003](https://doi.org/10.1016/j.jsb.2006.07.003) [Medline](#)
50. X. Wu, Y. Li, B. Crise, S. M. Burgess, D. J. Munroe, Weak palindromic consensus sequences are a common feature found at the integration target sites of many retroviruses. *J. Virol.* **79**, 5211–5214 (2005). [doi:10.1128/JVI.79.8.5211-5214.2005](https://doi.org/10.1128/JVI.79.8.5211-5214.2005) [Medline](#)

51. E. Serrao, A. Ballandras-Colas, P. Cherepanov, G. N. Maertens, A. N. Engelman, Key determinants of target DNA recognition by retroviral intasomes. *Retrovirology* **12**, 39 (2015). [doi:10.1186/s12977-015-0167-3](https://doi.org/10.1186/s12977-015-0167-3) [Medline](#)
52. P. K. Quashie, T. Mesplède, Y.-S. Han, M. Oliveira, D. N. Singhroy, T. Fujiwara, M. R. Underwood, M. A. Wainberg, Characterization of the R263K mutation in HIV-1 integrase that confers low-level resistance to the second-generation integrase strand transfer inhibitor dolutegravir. *J. Virol.* **86**, 2696–2705 (2012). [doi:10.1128/JVI.06591-11](https://doi.org/10.1128/JVI.06591-11) [Medline](#)
53. S. Hare, S. J. Smith, M. Métifiot, A. Jaxa-Chamiec, Y. Pommier, S. H. Hughes, P. Cherepanov, Structural and functional analyses of the second-generation integrase strand transfer inhibitor dolutegravir (S/GSK1349572). *Mol. Pharmacol.* **80**, 565–572 (2011). [doi:10.1124/mol.111.073189](https://doi.org/10.1124/mol.111.073189) [Medline](#)
54. S. Hare, A. M. Vos, R. F. Clayton, J. W. Thuring, M. D. Cummings, P. Cherepanov, Molecular mechanisms of retroviral integrase inhibition and the evolution of viral resistance. *Proc. Natl. Acad. Sci. U.S.A.* **107**, 20057–20062 (2010). [doi:10.1073/pnas.1010246107](https://doi.org/10.1073/pnas.1010246107) [Medline](#)
55. R. A. Lutzke, R. H. Plasterk, Structure-based mutational analysis of the C-terminal DNA-binding domain of human immunodeficiency virus type 1 integrase: Critical residues for protein oligomerization and DNA binding. *J. Virol.* **72**, 4841–4848 (1998). [Medline](#)
56. E. F. Pettersen, T. D. Goddard, C. C. Huang, G. S. Couch, D. M. Greenblatt, E. C. Meng, T. E. Ferrin, UCSF Chimera—A visualization system for exploratory research and analysis. *J. Comput. Chem.* **25**, 1605–1612 (2004). [doi:10.1002/jcc.20084](https://doi.org/10.1002/jcc.20084) [Medline](#)



Cryo-EM structures and atomic model of the HIV-1 strand transfer complex intasome

Dario Oliveira Passos, Min Li, Renbin Yang, Stephanie V. Rebensburg, Rodolfo Ghirlando, Youngmin Jeon, Nikoloz Shkriabai, Mamuka Kvaratskhelia, Robert Craigie and Dmitry Lyumkis (January 5, 2017)
Science **355** (6320), 89-92. [doi: 10.1126/science.aah5163]

Editor's Summary

High-resolution insights into the intasome

An essential step in the life cycle of lentiviruses such as HIV-1 is when viral DNA integrates into the host genome, establishing a permanent infection of the host cell. The viral integrase enzyme catalyzes this process and is a major drug target. During viral integration, integrase binds the ends of viral DNA, forming a higher-order structure called the intasome. Passos *et al.* and Ballandras-Colas *et al.* used cryo-electron microscopy to solve the structures of the intasomes from HIV-1 and maedi-visna virus (ovine lentivirus), respectively. These structures reveal how integrase self-associates to form a functional intasome and help resolve previous conflicting models of intasome assembly.

Science, this issue p. 89, p. 93

This copy is for your personal, non-commercial use only.

- Article Tools** Visit the online version of this article to access the personalization and article tools:
<http://science.sciencemag.org/content/355/6320/89>
- Permissions** Obtain information about reproducing this article:
<http://www.sciencemag.org/about/permissions.dtl>

Science (print ISSN 0036-8075; online ISSN 1095-9203) is published weekly, except the last week in December, by the American Association for the Advancement of Science, 1200 New York Avenue NW, Washington, DC 20005. Copyright 2016 by the American Association for the Advancement of Science; all rights reserved. The title *Science* is a registered trademark of AAAS.