Supplementary Materials for

Reconstitution and visualization of HIV-1 capsid-dependent replication and integration in vitro

Devin E. Christensen[1*], Barbie K. Ganser-Pornillos[2*], Jarrod S. Johnson[1], Owen Pornillos[2†], Wesley I. Sundquist[1†]

correspondence to: opornillos@virginia.edu, wes@biochem.utah.edu

This PDF file includes:

Other Supplementary Materials for this manuscript includes the following:

Materials and Methods

Cell Culture, Virion Production, and Cell Lysates

Human 293T and HeLa cells were obtained from the American Type Culture Collection and grown in Dulbecco modified Eagle medium (ThermoFisher) supplemented with 10% fetal calf serum (Atlanta Biologicals) at 37C in 5% CO_2. Cell cultures were confirmed to be negative for mycoplasma in routine tests (Lonza) every six months. Viral particles were produced by Calcium Phosphate transfection (Takara) of 293T cells with the pSG3$^{\Delta env}$ plasmid, obtained from the NIH AIDS Reagent Program (Catalog # 11051). Virions produced from this construct lack the viral Env protein, but are fully infectious when pseudotyped with envelope glycoproteins from HIV-1 or VSV ((*13-15*) and confirmed for our construct). For virion production, 2.5×10^6 293T cells were seeded on 10 cm plates and transfected 24 h later. 18 h post transfection, media was removed and replaced with fresh media. 48 h post transfection, virion-containing media was removed and filtered through a 0.45 μM membrane, layered on a 20% sucrose cushion in HS buffer (10 mM Hepes pH 7.4 and 140 mM NaCl), and centrifuged for 2 h at 28,000 rpm in a SW32 Ti rotor (Beckman). After centrifugation, media was decanted and virion-containing pellets were resuspended in 100 μl HS buffer per 10 cm plate of transfected cells, aliquoted, flash frozen in liquid nitrogen, and stored at -80C. Virion CA concentrations were determined by HIV-1 p24 ELISA assay (XpressBio). Typical yields were 40-100 pmol CA per 10 cm plate.

HeLa cell lysates were prepared by seeding 2.8×10^6 cells on a 15 cm plate. After 48 h, media was removed and cells were washed with PBS, removed from the plate with a cell lifter (Sigma Aldrich), placed in a micro-centrifuge tube, and centrifuged for 1 min at 1500 rpm (4C). After centrifugation, excess PBS was removed and cells were resuspended in 200 μl of lysis buffer (50 mM Tris pH 8.0, 150 mM NaCl, 50 μg/ml melittin) per 15 cm plate and incubated on ice for 15 min. Lysates were then spun in a micro-centrifuge for 10 min at 17,000 rpm. The soluble fraction was collected, aliquoted, flash frozen in liquid nitrogen, and stored at -80C.

Lysates prepared in this fashion stimulated ERT (Fig. 1C) and were required for integration (Fig. 5B). Preliminary characterization of the extract activities indicates that flow throughs from lysates passed through Centricon filters with cutoffs of 3K, 50K, and 100K did not support the integration reaction, implying that the lysate is not simply supplying a small molecule like Mg^{2+} or IP_6. Integration activity was also lost when the lysates were heated to either 65C (20 min) or 95C (5 min), implying that at least one key factor is likely to be a structured biomolecule.

Endogenous Reverse Transcription (ERT) Assay

Prior to assaying for ERT, virion preparations were treated with cyanase (RiboSolutions) to digest any plasmid DNA carried-over from transfection. Briefly, virions were diluted to 20 nM CA in cyanase digestion buffer (50 mM Tris pH 8.0, 20 mM NaCl, and 12 mM $MgCl_2$) in a total volume of 800 μl. 1 μl cyanase was added and the reaction was incubated at 37C for 30 min. 60 μl of cyanase inactivation resin (RiboSolutions) was then added and incubated with constant inversion at 37C for 20 min. Virions were then centrifuged for 1 min at 5000×g and virion containing supernatant was removed and stored on ice. Background level measurements in the ERT and integration

assays were greatly reduced by cyanase treatment, demonstrating the importance of removing carry-over DNA from transfection.

ERT was initiated by combining 20 µl of virions with 20 µl of 2×ERT Buffer (50 mM Tris pH 7.4, 150 mM NaCl, 12 µM dCTP, 9.2 µM dGTP, 16 µM dTTP, 10.4 µM dATP, 50 µg/ml melittin, 80 µM IP_6, 360 µM CTP, 3.5 mM GTP, 1.4 mM UTP, 13.4 mM ATP, 6.67 mM $MgCl_2$, and 0.6 ng/µl of the pK184 target plasmid) and 5 µl of HeLa cell lysate or lysis buffer control, then incubated at 37C for 10 h (denoted "Standard Reaction Conditions" or "Standard Reaction Conditions -Lysate", respectively), or otherwise for time courses as indicated in figures. rNTPs and dNTPs were purchased from Promega and their concentrations were chosen to correspond to those estimated for T lymphocyte cytoplasms (*21, 22*). IP_6 was purchased from Sigma, and the optimal concentration was 40-80 µM, based upon yields of early (MSSS) DNA transcripts (e.g., see Fig. 2A). 40 µM IP_6 was chosen for our standard condition because this is the estimated intracellular concentration (*20*). Melittin was purchased from Sigma, titrated, and used at a concentration that balanced the optimal yields of early (MSSS) transcripts, late (SST) transcripts, 3' integration products and membrane-free cores. We found that Triton X-100 could work nearly as well as melittin in the ERT and integration assays, but the optimal concentration range was narrower and had greater variability in product yields. Mg^{2+} concentrations were also optimized, using early (MSSS) transcripts as the readout. Product formation was maximal in the range of 8-10 mM Mg^{2+}, and we chose a concentration (9.33 mM) that equalled the total rNTP concentration in the final ERT buffer conditions (i.e., rNTPs were added at 18.66 mM in 2× ERT buffer then diluted 1:1 with virions to 9.33 mM; $MgCl_2$ was added at 12 mM during Cyanase digestion and 6.67 mM in 2× ERT buffer, which when combined 1:1 resulted in a final concentration of 9.33 mM). 20 µl of 0.5 M EDTA was added to stop ERT. Final ERT DNA products were purified using a Qiaquick PCR purification kit (Qiagen) and eluted from the column in 120 µl elution buffer (Qiagen). Drugs at final concentrations were as follows; GS-CA1 (Gilead) in DMSO at 1 nM to 10 µM, raltegravir (NIH AIDS Reagent Program) in water at 1 nM to 100 µM, and efavirenz (NIH AIDS Reagent Program) in DMSO at 10 µM.

Two-Step PCR 3' Integration Assay

The plasmid pK184 (added to the 2×ERT Buffer, see above) was used as an integration target for the full-length viral cDNA produced in the ERT reaction. Purified reaction products (see above) were amplified by PCR using Phusion Polymerase (NEB), diluted 100-fold and measured by quantitative PCR (qPCR). Briefly, 5 µl of purified ERT product was combined with 10 µl Phusion HF Buffer, 200 µM each dNTP, 0.24 µM each of three primers (pK184-IN-FWD, pk184-IN-REV, and DC3014), 0.5 µl Phusion Polymerase, and water to a final volume of 50 µl. PCR cycles were as follows: Denaturation at 95C for 5 min; 23 cycles of 95C for 30 sec, 58C for 30 sec, and 72C for 90 sec; and a final extension at 72C for 5 min followed by a hold at 4C. Final PCR products were purified with a Qiaquick PCR purification kit (Qiagen) according to the manufacturer's instructions and eluted in 120 µl elution buffer and diluted 100-fold for qPCR analysis of 3' integration. Note that subsequent to these studies, re-optimization of the DC3014 primer further reduced the background PCR signal in the integration quantification assay by about 10-fold; the sequence of improved primer (designated DC3014c) is provided in Table S2.

To obtain 100 ng DNA for deep sequencing, 8 (for experimental repetitions 046 and 048) or 24 (for 047) identical PCR reactions were performed, combined, purified with a Qiaquick PCR purification kit (Qiagen), eluted in 50 µl elution buffer and submitted for deep sequencing analysis. Experimental repetitions 046 and 047 used 23 PCR amplification cycles, whereas 048 was amplified for 35 cycles. These samples were analyzed by deep sequencing to identify 3' integration target site sequences.

Quantitative PCR Analyses

Light Cycler 480 SYBR Green I Master (Roche) with a Light Cycler 480 (Roche) were used for all qPCR analyses. Briefly, 5 µl of either ERT product or diluted 3' Integration PCR product, 500 nM each of two primers (MSSS-FWD + MSSS-REV, FST-FWD + FST-REV, Late RT-FWD + Late RT-REV, or AE3013 + IN990-REV for integration) and 10 µl of SYBR Green I Master Mix were combined in a total volume of 20 µl and analyzed by qPCR according to the manufacturer's instructions. All qPCR samples were analyzed in triplicate. Standard curves for ERT products were generated by using a 10-fold dilution series of the $pSG3^{\Delta env}$ plasmid for quantitation of ERT products. The number of DNA copies per core was determined from the ratio of viral cDNA copies (determined by qPCR) to the number of input core particles in each reaction (typically $\sim 4\times10^7$). The number of HIV core particles in each reaction was determined from the input CA protein level (typically ~0.38 pmol, measured by ELISA, see above) and by assuming that there were 3000 Gag/CA molecules per virion and 1.33 cores per virion (*9, 25*). Yields were calculated by assuming a theoretical yield of one late RT product/core. In reality, RT may initiate on both of the packaged RNA strands, and it is not clear when the conversion to a single template occurs. Note also that the primers used to quantify early/minus strand strong stop (MSSS) and intermediate/first strand transfer (FST) products ultimately have two binding sites on the viral DNA template by the time the full dsDNA preintegration complex is formed. This effect is evident in the ERT time courses shown in Fig. 1A and 5A, where the qPCR products from these two primers rise late in the time course (e.g., compare the signals at 6 h and 8 h). These effects apparently explain why the qPCR measurements for the MSSS and FST primers can approach 2 copies/core in some experiments. Integration product levels were normalized relative to the qPCR signal observed for the t=0 time point. Three independent repetitions and analyses were performed, and representative examples are shown for the experiments in Figs. 1A, 1B, 2B, 3A, 5A, 5B and S7.

Negative Stain Transmission Electron Microscopy

ERT reaction mixtures (10 h, 37C, standard ERT conditions except that IP_6 was varied as described, 3.5 µl) were applied onto the carbon side of Formvar/Carbon-coated, 300-mesh copper grids (Electron Microscopy Sciences) for 90 s, rinsed for 90 s by flotation on a drop of 100 mM KCl, blotted dry, stained for 90 s by floating on a drop of 2% (*w*/*v*) solution of uranyl acetate, blotted dry, and allowed to air dry. Samples were viewed on an FEI Tecnai F20 transmission electron microscope operating at 120 kV. Core quantification was performed at a nominal magnification of ×11,500, where viral capsids could be distinguished from similar size debris. For each grid, three well-stained squares were selected at random and capsids were counted manually while sweeping the perimeter of each square, covering an area of ~900 μm^2. Quantification was done in three

independent experiments, and in two of these experiments the grids were randomized and the experimenter was blinded to their identities.

Cryo-Electron Tomography and Lattice Mapping

ERT reaction mixtures were mixed with an equal volume of 10-nm BSA Gold Tracer (Electron Microscopy Sciences), 3.5 µl aliquots were applied onto glow-discharged 2/2 carbon C-flat or Quantifoil grids (Electron Microscopy Science), blotted to near dryness manually, and then plunge-frozen into liquid ethane. Cryotomograms were acquired using an FEI Titan Krios electron microscope operating at 300 kV and equipped with either a Falcon III camera or a K3/GIF with a slit width of 20 eV. Tilt series were collected using the data collection software Tomography 4.0 (FEI) with an angular range of −60° to +60°, an angular increment of either 1 or 2°, defocus values of 5 to 10 µm, and a nominal magnification of ×29,000 (Falcon III) or ×33,000 (K3), which correspond to a pixel (px) size of 2.92 (Falcon III) or 2.69 (K3) Å. Tilt series were aligned by using imod (*58*). Weighted back-projection or SIRT was used to reconstruct tomograms in imod.

Lattice mapping was performed based on previously described protocols (*59, 60*) using as search template a cryoEM map of the HIV-1 CA hexamer (EMD-3465) (*8*) down sampled to 25 Å resolution. Each capsid surface was modeled and oversampled as a mesh by using surface modeling tools in the Dynamo software package (*58*). Two sets of sub-volumes centered at the same positions but of different box size (48×48×48 px^3 and 32×32×32 px^3) were extracted from 4× binned tomograms and assigned polar Euler angles in reference to the modeled capsid surface; azimuthal angles were randomized. For each box size set, four rounds of azimuthal angle search were performed, applying 6-fold symmetry, a low-pass filter of 30 Å and using a soft-edged cylindrical alignment mask in Dynamo. Overlapping and misaligned particles, as well as those with very low cross-correlation values were discarded. The two sets were then compared, and only particles common to both sets were retained. These were re-extracted (box size of 64×64×64 px^3) from 2× binned tomograms. To fill in gaps, the refined particle set was used to define a new surface model, which was oversampled and extracted (64×64×64 px^3) from 2× binned tomograms, retaining only those that are >10 px away from particles in the refined set. Four rounds of template alignment were then performed on the combined set of particles, using progressively narrower angular and positional search ranges and a low-pass filter set to 25 Å. After each round, overlapping and misaligned particles, as well as those with very low cross-correlation values were discarded. Final lattice maps were generated and visualized in Chimera (*61*) as described (*59, 60*). Software used in this study were curated by SBGrid (*62*).

Deep Sequencing and Analysis of 3’ Integration Events

100 ng of the first PCR product from the 3’ integration assay (above) was submitted for Ion Torrent sequencing, performed at the DNA Sequencing Core Facility, University of Utah. Reads were screened for sequences with at least 20/23 matches to the extreme end of the HIV-1 3’ LTR using Seal (BBTools suite, Joint Genome Institute). To identify reads with junctions aligning to the pK184 target plasmid, these Seal reads were then remapped onto linearized pK184 reference genomes using BWA-MEM (*63*), again requiring at least 20/23 matches to the reference. To identify reads that could represent

autointegration events, Seal reads with sequences matching the 3' LTR were remapped onto a modified HIV-1 reference genome using BWA-MEM. Because the alignment software cannot distinguish identical sequences, the reference genome was modified by substituting the 5' LTR with 633 Ns. Finally, subsets of reads from both types of alignments were checked manually to ensure that the alignment procedures identified integration sites and their orientations correctly. Alignments of integration events into pK184 and autointegration events into HIV-1 were visualized using the Integrative Genomics Viewer (*64*).

To score integration events at specific bases in the target plasmid, reads aligning to the HIV-1 3' LTR and pK184 were parsed using their CIGAR string to identify the pK184 integration positions and orientations (Python code archived at GitHub repository DOI: 10.5281/zenodo.3966034). Because the alignment and integration analyses were performed on linearized pK184 reference genomes, the analyses were performed twice on different linearized pK184 reference genomes in which the plasmid "ends" were shifted to very different positions. After correcting for this shift, natural log-transformed merged integration counts were plotted using the circlize package in R (*65*). Autointegration events in HIV-1 were similarly scored by counting junction sites in reads that aligned to the 3' LTR and to an additional site located between residues 1000 and 9000 of the HIV-1 genome. Autointegration counts were plotted using GraphPad Prism (*66*).

Cloning Products of Concerted Integration

99 independent plasmids carrying concerted viral integration sites from 12 different integration reactions were cloned in two different ways: 53 clones were obtained following "repair" of the integration sites (described below), and 46 clones were obtained by simply amplifying the backbone without repair, implying that at least in some cases at least one of the integration sites must have been repaired in the extract. The repair protocol was as follows: 1) To remove the 5' flap generated by integrase during integration, 10 µl of ERT product were incubated with 0.5 µl Thermostable FEN1 (NEB), 10 µl of Phusion HF Buffer, and 24.5 µl water at 65C for 5 min. 2) To fill in the 5 base integration target site gap, 200 µM each dNTP and 0.5 µl of Phusion Polymerase (NEB) was added, mixed, and incubated at 65C for 5 min. 3) To ligate the final phosphodiester bond and generate a complete double stranded DNA suitable for PCR amplification, T4 DNA ligase (NEB, 0.5 µl) was added and the sample was incubated at 37C for 30 min.

Whether or not samples were repaired, plasmid backbones that carried the two viral termini were amplified by adding 240 nM of each primer (SG3-U3-XhoI-REV and SG3-U5-XhoI-FWD) to the sample to PCR amplify the two ends of the HIV genome along with the entire pK184 plasmid backbone (see Fig. S9). PCR cycles were as follows: Denaturation at 95C for 5 min; 30 cycles of 95C for 30 sec, 58C for 30 sec, and 72C for 2 min; and a final extension at 72C for 5 min followed by a hold at 4C. Final PCR products were purified with a Qiaquick PCR purification kit (Qiagen), digested with XhoI, ligated with T4 DNA Ligase (NEB) and transformed in *E. coli* TOP10 cells. Plasmids were recovered from kanamycin-resistant colonies, and sequenced to verify concerted integration. In total, 151 plasmids recovered from 12 independent reactions were sequenced. Integration sites were judged to be necessarily independent if they did not match any of the other integration sites recovered from the same experiment (i.e., any duplicated integration sites from individual experiments were removed prior to analysis).

Of the 103 remaining independent concerted integration events, all showed transfer of the 3' $U3_{T15}$ to the 5' LTR, 99 showed the expected viral termini and 5' target site duplications, and four showed aberrant target site duplications of 22 bp (one case) and 65 bp (three cases, all at the same site).

Generation of Integration Consensus Sites

To generate sequence logos showing target base preferences from our deep sequencing analysess of 3' integration events, we first scaled integration counts to correct for PCR amplification bias using the gradients in read coverage around the pK184 target plasmid to guide the scaling. Read coverage depths, which decreased as a function of distance from the primer binding sites, were binned and plotted separately for forward- and reverse-oriented insertions into pK184. Forward and reverse coverage depth plots were used to generate LOWESS scaling curves and these curves were then applied to normalized integration count data from each independent experiment (using 50000 as a normalization multiplier for forward and reverse integrations from samples 46 and 47, and 15000 for forward and 25000 for reverse integrations from sample 48 – these normalization multipliers were chosen to keep the total number of sequence strings analyzed for each sample below 1 million). Integration frequencies at each base step between 200 and 2200 base pairs away from each primer were then analyzed (this 2000 bp segment was chosen to exclude regions of low coverage immediately adjacent or far away from the primers). The scaled sequence strings were submitted to WebLogo 3 (*67, 68*) to compute sequence logos. To generate the sequence logo for concerted integration, the sequences strings from the the 99 independent clones described above were submitted to WebLogo 3.

As shown in Figs. 5 and S9, the consensus integration sequence motifs arising from both of our different sets of experiments differ slightly and are not perfectly palindromic, but are nevertheless clearly related to one another and to those reported previously (*48-50*). We have not interpreted our data further because our DNA target plasmid was limited in size and was not chromatinized, and because we do not know whether our extracts contain any, or all, of the host factors that might favor site selection in cells. We note, however, that our cell free system should provide a well-controlled platform for analyzing the effects of candidate host factors on integration site selection.

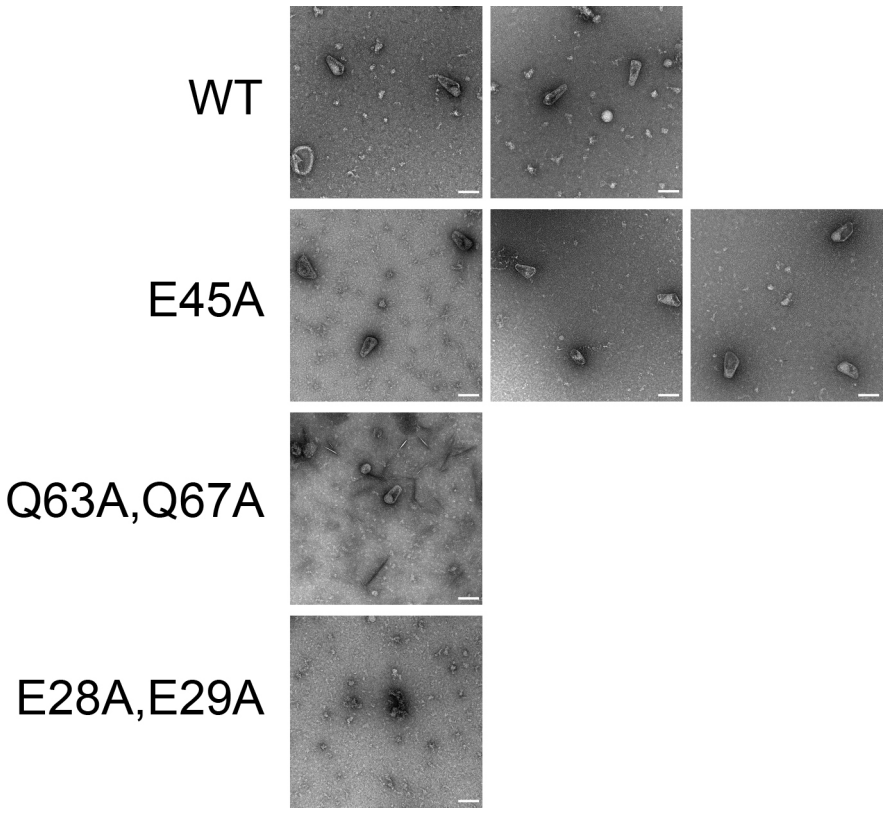

Fig. S1.
Negative Stain EM Images of HIV-1 Capsids after ERT (10 h, 37C, standard ERT conditions). At this magnification, the conical capsids were easily recognizable and could be distinguished from debris. Scale bars, 100 nm.

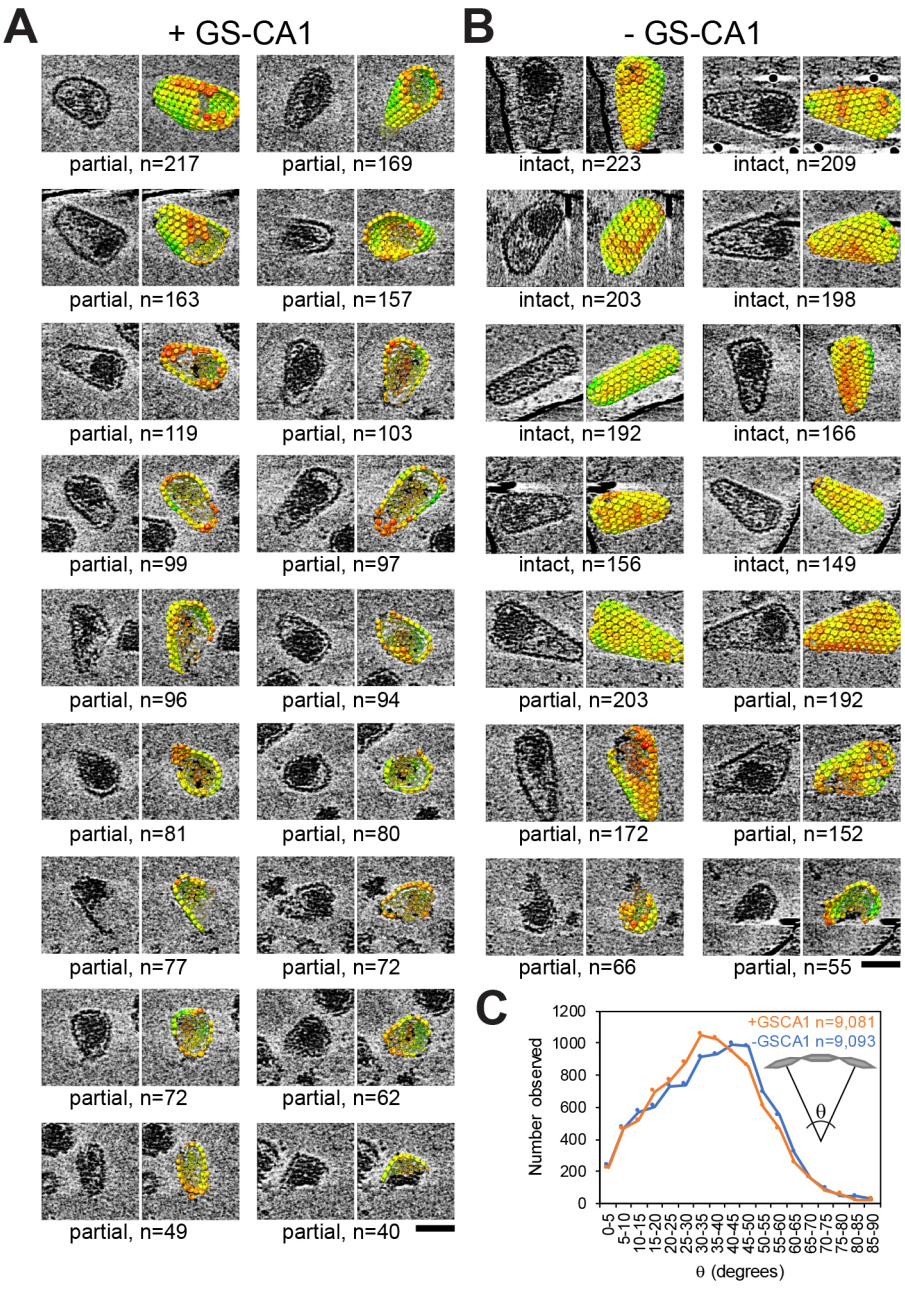
A
+ GS-CA1
B
- GS-CA1
partial, n=217
partial, n=169
intact, n=223
intact, n=209
partial, n=163
partial, n=157
intact, n=203
intact, n=198
partial, n=119
partial, n=103
intact, n=192
intact, n=166
partial, n=99
partial, n=97
intact, n=156
intact, n=149
partial, n=96
partial, n=94
partial, n=203
partial, n=192
partial, n=81
partial, n=80
partial, n=172
partial, n=152
partial, n=77
partial, n=72
partial, n=66
partial, n=55
partial, n=72
partial, n=62
partial, n=49
partial, n=40
C
+GSCA1 n=9,081
-GSCA1 n=9,093
θ
Number observed
1200
1000
800
600
400
200
0
0-5
5-10
10-15
15-20
20-25
25-30
30-35
35-40
40-45
45-50
50-55
55-60
60-65
65-70
70-75
75-80
80-85
85-90
θ (degrees)

Fig. S2.

Effect of GS-CA1 Binding on the Hexagonal Capsid Lattice. (A) Gallery of individual HIV-1 cores observed after 4 h ERT incubations in the presence of 100 nM GS-CA1. For each core, a central cryotomogram slice is shown, without (left) and with (right) the corresponding capsid lattice map as determined by subtomogram averaging, color coded according to cross-correlation values, with red denoting low correlations (*59, 60*). Each core is annotated as "intact" (i.e., having an intact or nearly intact capsid) or "partial" (i.e., having a ruptured or partially disassembled capsid); these phenotypes are defined in detail in Fig. S3. Scale bar, 50 nm. **(B)** Gallery of cores after 4 h ERT incubations in the absence of GS-CA1. Scale bar, 50 nm. **(C)** Histograms showing frequencies of different bite angles between alternating CA hexamers within partial capsids (with n values given within the figure). Note that GS-CA1 treatment reduced the most frequent bite angle by 15-20 degrees, which is indicative of local lattice flattening.

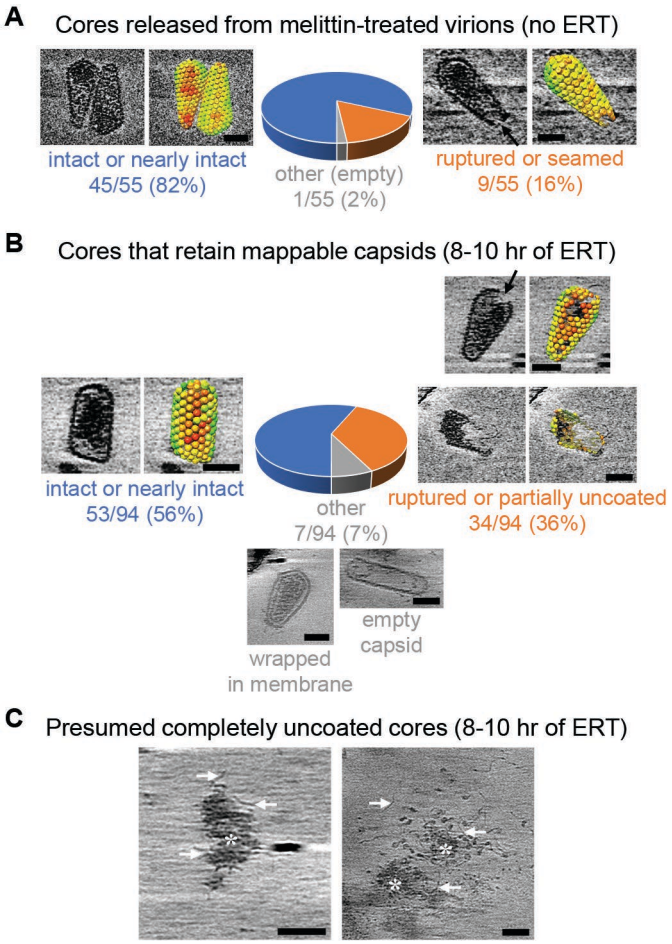

Fig. S3.

HIV-1 Capsid Disassembly. (A, B) Distributions of capsid morphologies under indicated conditions; no ERT control in (A) and standard ERT conditions in (B). Cores with mappable capsids fell into three major categories. Blue: Cores with intact capsids, in which all hexamers were mappable and all twelve pentamer positions could be accounted for by five-fold holes in the lattice map. Cores with nearly intact capsids, in which a small number of non-contiguous hexamers were missing from the lattice maps but holes were not evident in the tomograms are also included in this category. Orange: Cores with ruptured or partially disassembled capsids, in which clear gaps in the capsid shell were visible in tomogram slices (e.g., black arrows) and lattice maps showed corresponding missing patches of hexagonal lattice. Gray: Outliers include completely empty capsids (often cylindrical) and cores tightly wrapped in membranes, which we presume are artifacts arising from mellitin treatment and membrane remodeling. For each core, a central cryotomogram slice is shown. Corresponding capsid lattice maps, when shown, are color coded according to cross-correlation values, with red denoting low correlations. Scale bars, 50 nm. **(C)** A fourth major phenotype was observed in samples incubated under standard ERT conditions for 8-10 h. In these presumed core remnants, the capsids had either completely disassembled or had dissociated to the extent that any remaining lattice is too small to be mappable. Fibrous densities of nucleic acids (white arrows) are seen associated with amorphous aggregates (white asterisks). Scale bars, 50 nm.

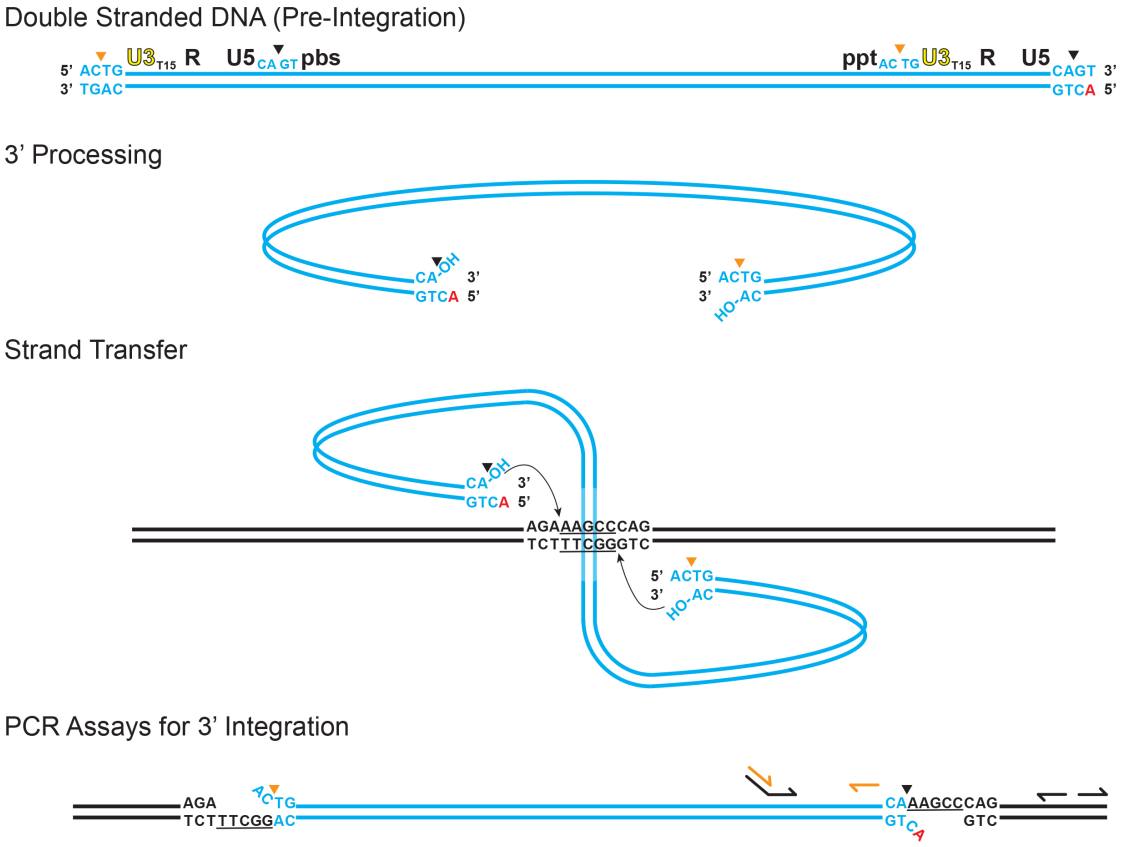

Fig. S4.

Schematic Illustration of Integration and the Quantification of 3' Integration Products. The integration reaction proceeds by removal of the two 3' terminal nucleotides of the double-stranded viral cDNA (blue, with a single terminal RNA base in red) (termed, "3' Processing"), followed by target site cleavage and concerted integration and ligation of the two viral 3' termini into the target DNA (black) (termed "Strand Transfer"). The reaction results in two mismatched base pairs and two gaps that are repaired by cellular FLAP endonuclease, polymerase and ligase enzymes, resulting in duplication of the integration target site. Integration of the 3' end of the viral DNA genome was detected by a PCR strategy in which the first pair of primers (black) amplify only integration events because one primer anneals to the viral genome and the other anneals to a site on the plasmid DNA. These products were then used either for deep sequencing reactions, or in a qPCR reaction used to quantify integration levels, in which the primers were directed against a unique sequence introduced by one of the first amplification primers and a sequence located near the 3' end of HIV-1 (orange primers). Note that: i) Sequences corresponding to original 5' and 3' termini of the viral RNA genome are denoted by orange and black arrowheads, respectively (see Fig. 1A), ii) A *pair* of bidirectional primers was used to amplify the target plasmid DNA (illustrated in Fig. S5, but not shown here), and iii) A single nucleotide polymorphism in the two viral U3 regions is shown explicitly (A15 in the 5' U3 element and U15 in the 3' U3 region, but converted to T15 in both U3 elements of the viral cDNA through the reverse transcription process, see Fig. 1).

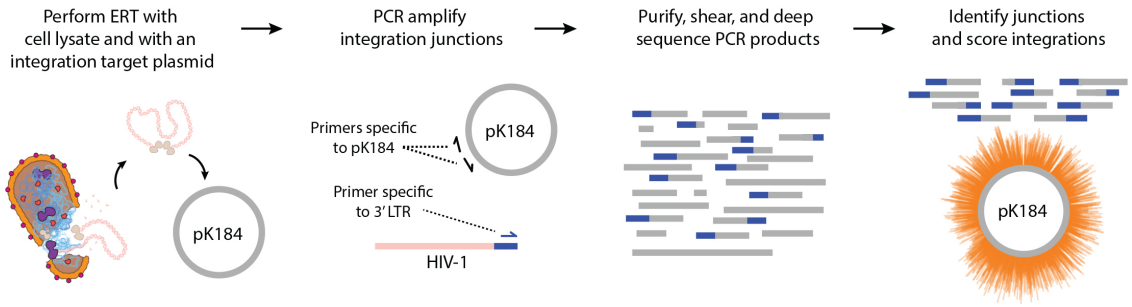

Fig. S5.

Schematic Illustrating the Different Steps in the Deep Sequencing and Mapping of Integration Sites.

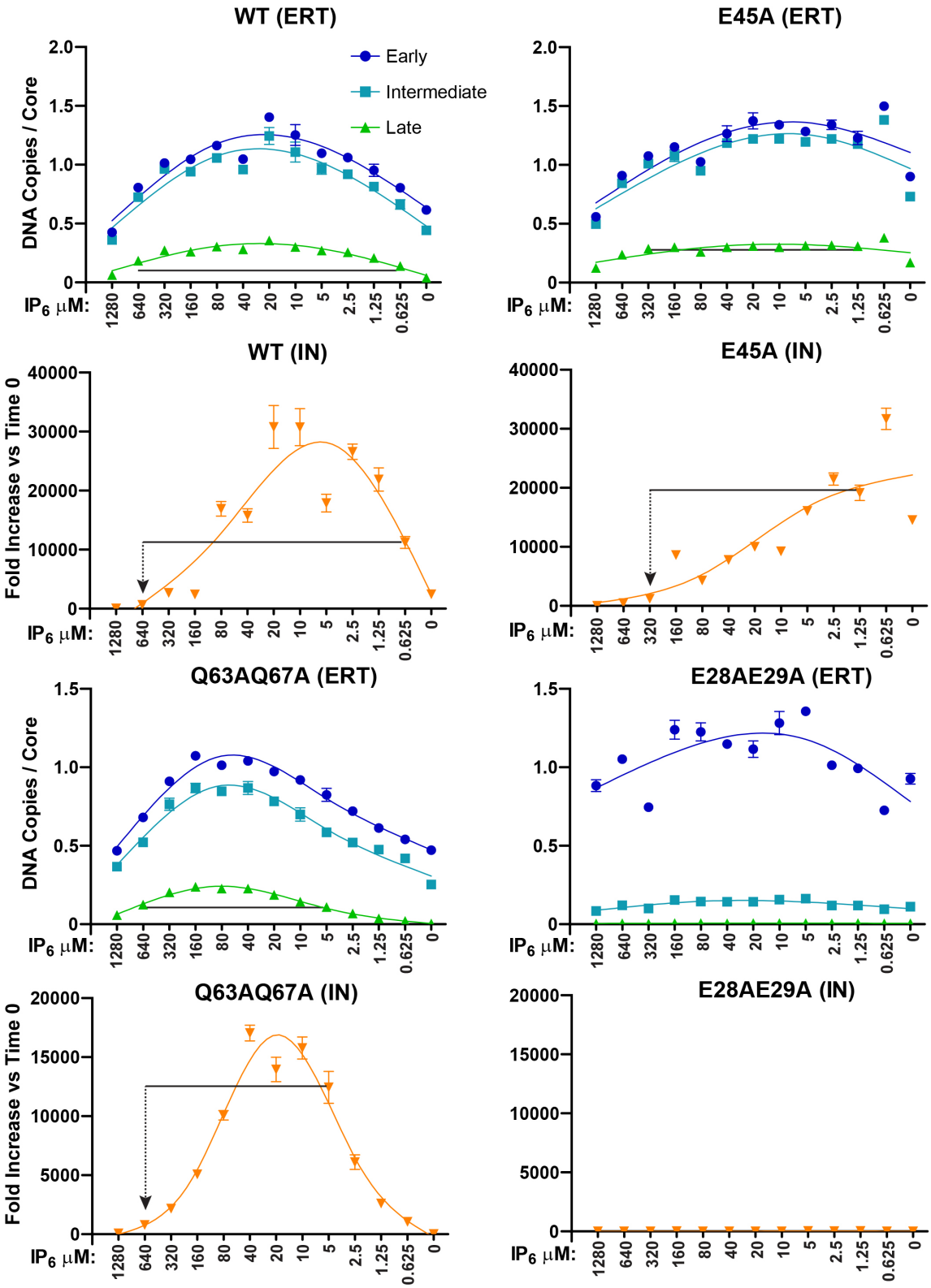
WT (ERT)
E45A (ERT)
WT (IN)
E45A (IN)
Q63AQ67A (ERT)
E28AE29A (ERT)
Q63AQ67A (IN)
E28AE29A (IN)
Early
Intermediate
Late
DNA Copies / Core
Fold Increase vs Time 0
IP6 μM:
1280
640
320
160
80
40
20
10
5
2.5
1.25
0.625
0

Fig. S6.

Effects of Capsid Stability on ERT and Integration Efficiency. Graphs showing quantification of ERT levels (upper panel) and integration products (lower panel) from wild type or mutant cores at different IP_6 concentrations. Black lines connect high and low IP_6 concentrations that produced equivalent levels of late RT products (upper panel) yet produced very different levels of integration products. This illustrates the point that high levels of IP_6 inhibit integration, suggesting that the viral capsid must disassemble completely to support integration, even when late reverse transcripts are present. Note that these reactions were performed under standard conditions (i.e., with added cell lysate), which was necessary in order to support integration. The actual IP_6 concentrations in the reactions were therefore higher than shown in the figure (and than those used in Fig. 2), because the concentrations shown here do not take into account any IP_6 present in the extract. The additional IP_6 accounts for the enhanced ERT seen at low levels of exogenous IP_6 (vs. the data shown in Fig. 2).

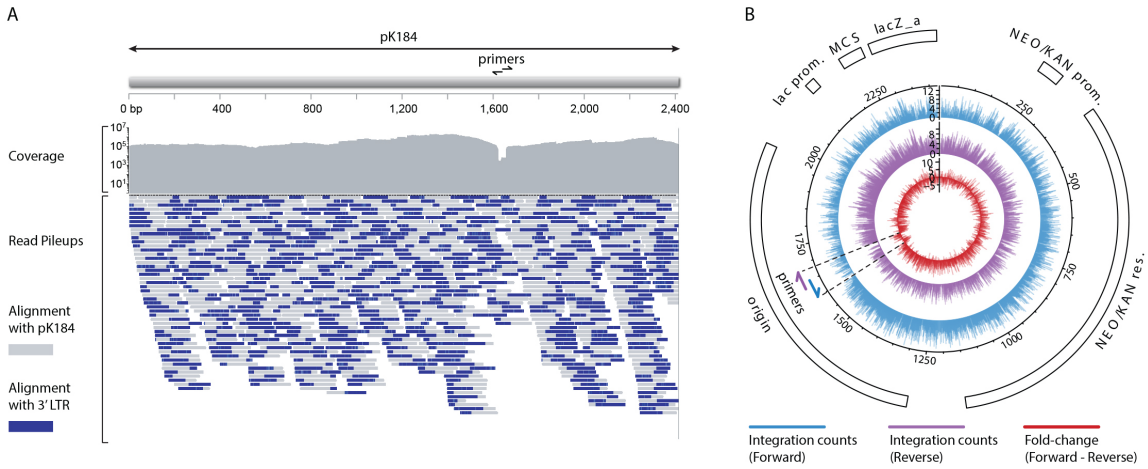

Fig. S7.

Mapping and Normalization of Bidirectional Integration into the Target Plasmid. (**A**) Coverage tracks and alignment pileups of chimeric deep sequencing reads that align to both the 3' LTR (blue) and to the pK184 target plasmid (gray), displayed as they align to pK184. Note that the pileups correspond to only a small, randomly selected fraction of the total reads, and that a drop in read coverage is observed at the primer binding sites (bases 1624-1654). (**B**) Circular pK184 plasmid map showing the position and frequency of reads corresponding to integration events (ln scale), determined by deep sequencing of read junctions between the 3' end of the viral genome and the target plasmid. The position of the HIV-1 LTR in the sequencing alignment was used to determine the directionality of integration as having either forward (blue) or reverse (purple) orientation. Strand-specific preference in each orientation at specific positions are plotted in the central rings (red: ln-transformed forward insertions – ln-transformed reverse insertions). As expected, integration counts trended downward as a function of distance from strand-specific primers (see Materials and Methods for further details). Figures show merged composite data from three independent experiments.

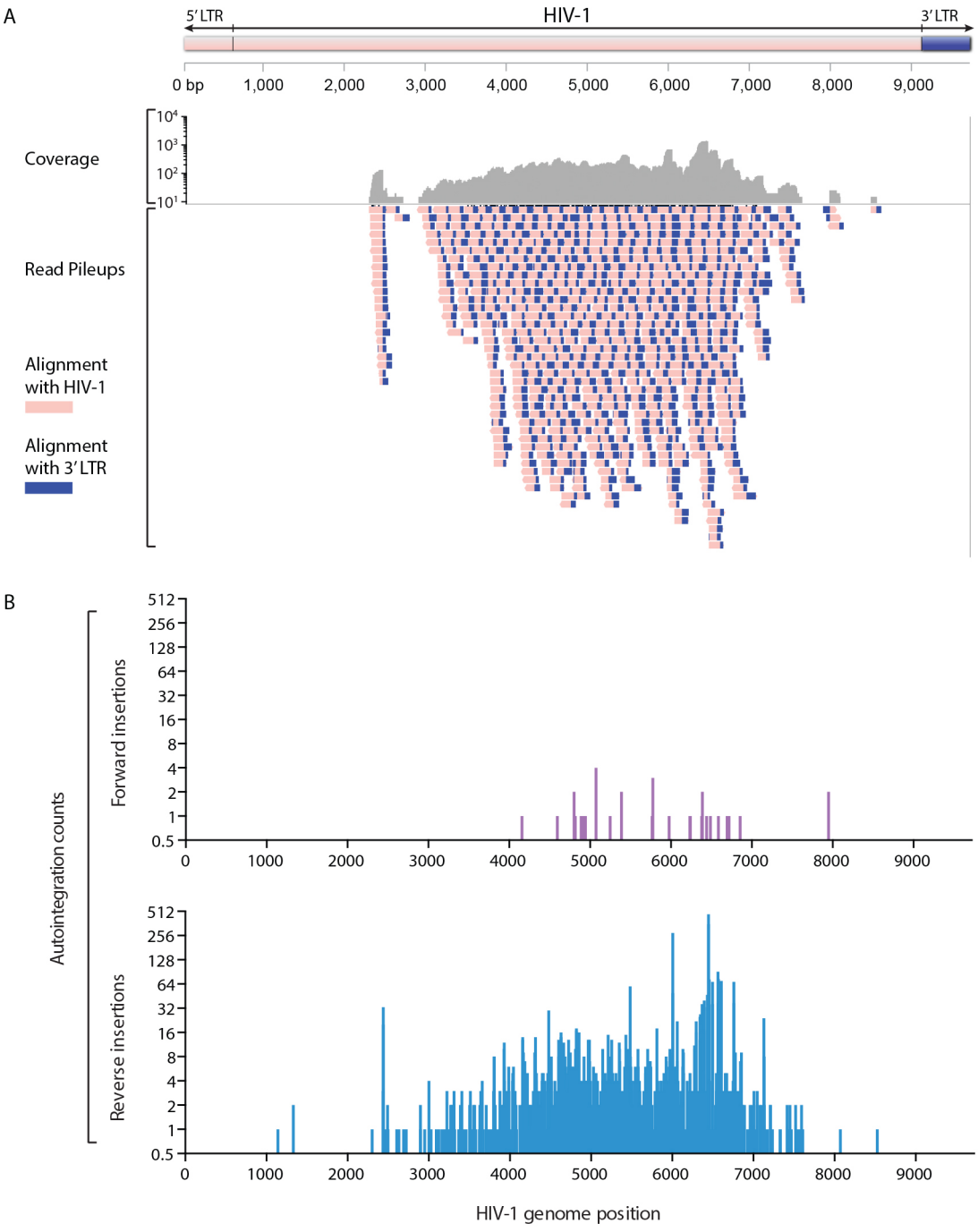
A
5′ LTR
HIV-1
3′ LTR
0 bp
1,000
2,000
3,000
4,000
5,000
6,000
7,000
8,000
9,000
Coverage
Read Pileups
Alignment with HIV-1
Alignment with 3′ LTR
B
Autointegration counts
Forward insertions
Reverse insertions
HIV-1 genome position

Fig. S8.

Autointegration Sites and Directionality. (A) Coverage tracks and a subset of the alignment pileups of chimeric deep sequencing reads that align to both the 3' LTR (blue) and to another site in the HIV-1 genome (pink), displayed as they align to HIV-1 bases 1000-9000. (**B**) Frequency of integration events corresponding to autointegration of the 3' end of the HIV-1 into upstream regions of the genome (from bp 1000 to 9000), separated by directionality. Note that "reverse insertions" correspond to the orientation in which concerted integration of both termini from the intasome into the viral genome could have occurred without twisting the intervening DNA. Data shown represent one of three experiments (see Table S1, sample 47).

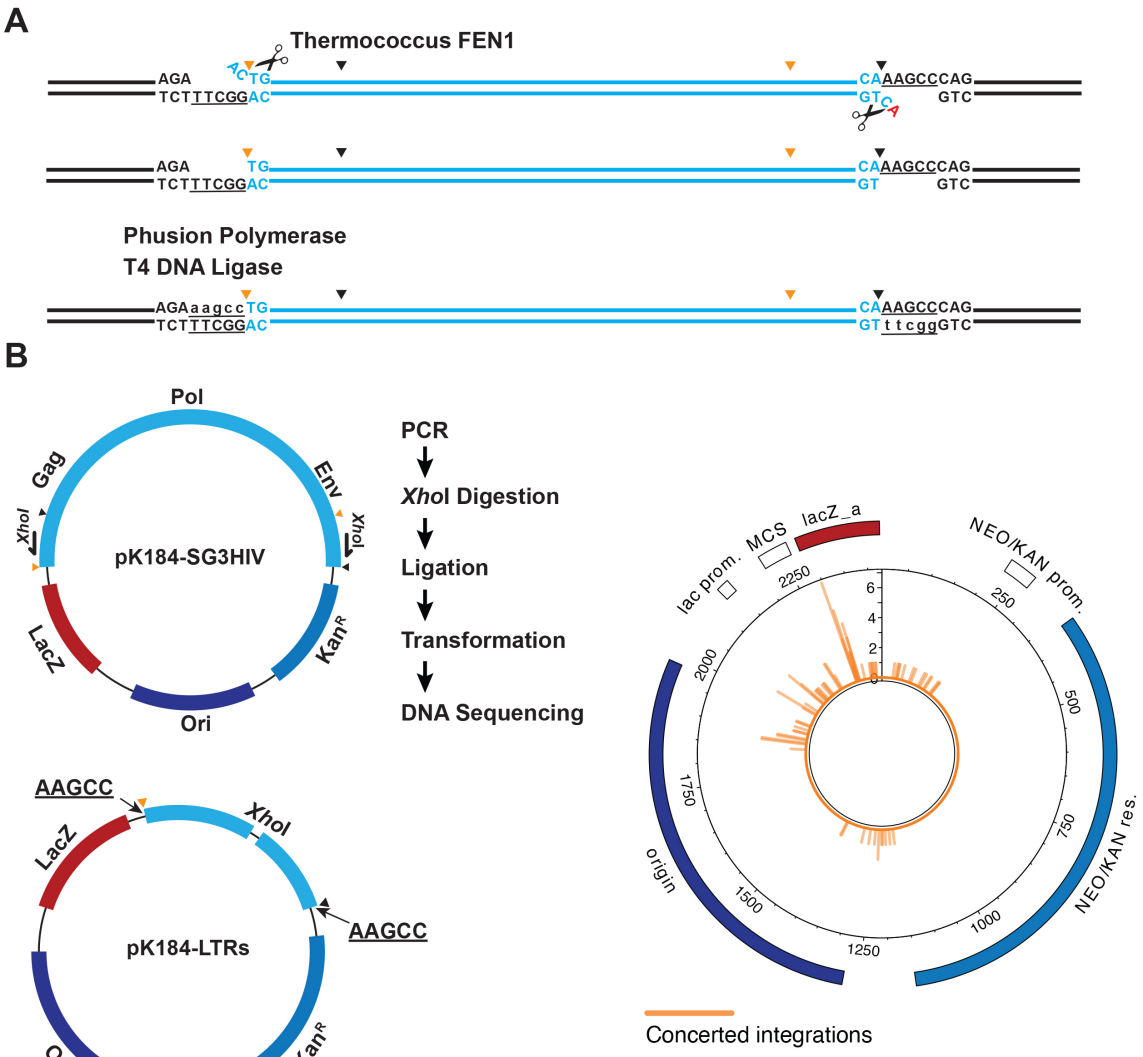
A
Thermococcus FEN1
AGA
TCTTTCGGAC
CAAAGCCCAG
GTC
Phusion Polymerase
T4 DNA Ligase
AGAaagccTG
TCTTTCGGAC
CAAAGCCCAG
GTttcggGTC
B
Pol
Gag
Env
XhoI
XhoI
pK184-SG3HIV
LacZ
Kan^R
Ori
PCR
XhoI Digestion
Ligation
Transformation
DNA Sequencing
AAGCC
XhoI
LacZ
AAGCC
pK184-LTRs
Ori
Kan^R
lac prom.
MCS
lacZ_a
NEO/KAN prom.
2250
250
2000
500
1750
750
origin
1500
1000
1250
NEO/KAN res.
Concerted integrations
C
bits
0.5
0.4
0.3
0.2
0.1
0.0
-5
0
5

Fig. S9.

Schematic Illustrations of Integration Site Repair and Cloning and Sequencing of Concerted Integration Products. (A) Enzymatic steps performed to repair mismatched flaps (Thermococcus FEN1 cleavage), DNA synthesis to fill in the gaps (Phusion Polymerase), and ligation of the 5' end of the genome to the target plasmid DNA (T4 DNA ligase) to repair integration sites and create continuous, clonable dsDNA products. (**B**) Steps in the cloning of repaired, concerted integration products (left panel) and sites of concerted integration (right panel). Plasmids carrying integrated viral cDNAs were amplified through the plasmid backbone region with primers that annealed to specific sites within each end of the integrated viral cDNA and contained *Xho*I restriction sites. Products of the appropriate size were gel purified, digested with *Xho*1, ligated to create closed circular DNAs, and transformed into bacteria. Kanamycin-resistant colonies were picked and the integration sites of the recovered plasmids were sequenced to confirm concerted integration and to map integration sites (arrows). We observed 103 unique integration events in 151 clones (the others were duplicates from within the same experiment), all clones showed transfer of the 3' $U3_{T15}$ to the 5' LTR, one clone had a 22-base pair target site duplication, and three clones had a 65-base pair target site duplication at the same site. (**C**) Consensus integration sequence derived from the 99 unique 5 base pair concerted integration sites sequenced. The Weblogo (*67, 68*) shows the elevated frequencies of different nucleotides surrounding the integration site. Integration sites (dotted lines) and 5 bp duplication sequence (horizontal line) are shown.

Table S1.

Summary of 3’ Integration Site Deep Sequencing Mapping Data

3' Integration Deep Sequencing Sample #	46	47	48	Pooled
Total reads	8,710,531	9,727,916	47,609,305	66,047,752
Seal reads (aligning to the viral 3' LTR)	1,066,394	1,155,319	8,293,463	10,515,176
Integrations (chimeric Seal reads aligning to the viral 3' LTR of HIV-1 and the pK184 target plasmid)	775,147	888,217	6,568,294	8,231,658
Forward integration read counts in pK184	384,044	437,373	5,529,333	6,350,750
Reverse integration read counts in pK184	391,103	450,844	1,038,961	1,880,908
Number of plasmid sites with zero integration read counts	153	88	37	7
Autointegrations (chimeric Seal reads aligning to the viral 3' LTR and bases 1000-9000 of HIV-1)	2909	5209	2633	10,751
Forward autointegration read counts	68	33	163	264
Reverse autointegration read counts	2841	5176	2470	10,487

Table S2.

Vectors, Plasmids, and Oligonucleotides Used in This Study

1A. Virus Production Vectors

Plasmid Name	Source[1]	Internal ID	Addgene Accession Number/ Vendor Catalog Number
SG3ΔENV	NIH AIDS Reagent Program	WISP20-3	11051
SG3ΔENV RT D185A		WISP20-4	145782
SG3ΔENV CA E45A		WISP20-5	145783
SG3ΔENV CA Q63A,Q67A		WISP20-6	145784
SG3ΔENV CA E28A,E29A		WISP20-7	145785
SG3ΔENV IN D116A		WISP20-8	145786
SG3ΔENV CA M66I		WISP20-10	149687

1B. Other Plasmids

Plasmid Name	Source[1]	Internal ID	
pK184	ATCC	WISP20-9	37766

1C. qPCR Oligonucleotides

Target	Sequence 5' to 3'
MSSS-FWD	AACCCACTGCTTAAGCCTCA
MSSS-REV	ACCAGAGTCACACAACAGACG
FST-FWD	AGCCGCCTAGCATTTCATCA
FST-REV	CCAGCGGAAAGTCCCTTGTA
Late RT-FWD	TGTGTGCCCGTCTGTTGTGT
Late RT-REV	CTTCAGCAAGCCGAGTCCTG
AE3013	ATGCCACGTAAGCGAAACTC
IN990-REV	CAACAGACGGGCACACACTA

1D. Other Oligonucleotides

Target	Sequence 5' to 3'
pK184-IN-FWD	GTGTCAGGCGTGGAATGAGA
pK184-IN-REV	TGGACTGTATGCACGAACCC
DC3014	ATGCCACGTAAGCGAAACTCTTGCCTGTACTGGGTCTCTC
DC3014c	ATGCCACGTAAGCGAAACTCGCATTTCATCACATGGCCCG
SG3-U3-XhoI-REV	ATGACTTGCTCGAGATGAAATGCTAGGCGGCTGT
SG3-U5-XhoI-FWD	TGTCGTATCTCGAGTTGCCTGTACTGGGTCTCTC

[1] Source refers to vendors for commercial vectors.

Movie S1-S3.

ECT Movies Showing HIV-1 Capsids during Replication and Uncoating. Movies S1-S3 correspond to the first three cores shown in Fig. 4, respectively. Each movie proceeds through tomographic *z* slices of the core and then traces in magenta the fibrous density features that we interpret as nucleic acid loops. Both ends of the loops are associated with less interpretable density within the core.