PNAS
www.pnas.org

Supplementary Information for

HIV-1 uncoats in the nucleus near sites of integration

Ryan C. Burdick, Chenglei Li, MohamedHusen Munshi, Jonathan Rawson, Kunio Nagashima, Wei-Shau Hu, and Vinay K. Pathak

Vinay K. Pathak
Email: vinay.pathak@nih.gov

This PDF file includes:

Other supplementary materials for this manuscript include the following:

www.pnas.org/cgi/doi/10.1073/pnas.1920631117

Materials and Methods

Cell lines and reagents. HeLa [American Type Culture Collection (ATCC CCL-2), TZM-bl cells [NIH AIDS Reagent Program, Division of AIDS, NIAID, NIH, from Dr. John C. Kappes and Dr. Xiaoyun Wu; Cat#8129; (1)], and Human embryonic kidney 293T cells (ATCC CRL-3216) were maintained as previously described (2, 3). CEM-SS cells (NIH AIDS Reagent Program, Division of AIDS, NIAID, NIH; gift from Dr. Peter L. Nara; Cat#776) and THP-1 cells (ATCC TIB-202) were maintained in Roswell Park Memorial Institute (RPMI) 1640 medium (CellGro, Manassas, VA) supplemented with 10% fetal calf serum (HyClone, Logan, UT) and 1% penicillin-streptomycin (penicillin 50 U/ml and streptomycin 50 µg/ml; Lonza, Walkersville, MD). THP-1 cells were differentiated with phorbol myristyl acetate treatment (Sigma-Aldrich, St. Louis, MO; 100 ng/ml) for 48 hours to generate THP-1-derived macrophages.

To generate the HeLa-Bgl cell line, pLVX-TRE3G-Bgl-mCherry, a lentiviral vector that expresses truncated bacterial protein BglG that is fused to mCherry at the C-terminus and contains a nuclear localization signal (Bgl-mCherry) from a doxycytidine-inducible promoter was first created. Bgl-mCherry, which specifically binds to Bgl RNA stem loops (4), was cloned into lentiviral vector pLVX-TRE3G (Clontech, Mountainview, CA) to generate pLVX-TRE3G-Bgl-mCherry. HeLa cells were transduced with pLVX-TRE3G-Bgl-mCherry and Tet-On 3G, which expresses the transactivator protein, and the transduced cells were maintained in complete media containing puromycin (1 µg/ml; Thermo Fisher Scientific) and G418 Sulfate (200 µg/ml; Sigma-Aldrich); a single cell clone expressing a low level of Bgl-mCherry was obtained by limiting dilution to generate the HeLa-Bgl cell line.

To generate the HeLa-Bgl:Tat-Rev cell line, a murine leukemia virus (MLV)-based vector that expressed HIV-1 Tat and Rev was first created by transferring a cassette that expresses a hygromycin b phosphotransferase gene (5) from an internal ribosomal entry site (IRES) from WH450, a spleen necrosis virus vector, into MLV-based vector pAR2 (6). Next, a cassette containing codon-optimized HIV-1 Tat and Rev separated by an in-frame self-cleaving peptide

from porcine teschovirus 2A [P2A; (7)] was inserted upstream of IRES-Hygro to generate MLV-Tat-P2A-Rev-IRES-Hygro. Infectious virions were produced by co-transfection of MLV-Tat-P2A-Rev-IRES-Hygro along with MLV *gag-pol* expression plasmid pLGPS (8) and amphotropic MLV envelope expression plasmid pSV-A-MLV-*env* [NIH AIDS Reagent Program, Division of AIDS, NIAID, NIH: from Dr. Nathaniel Landau and Dr. Dan Littman; Cat#1065 (9)]. HeLa-Bgl cells were transduced with MLV-Tat-P2A-Rev-IRES-Hygro to generate HeLa-Bgl:Tat-Rev cell line, which expresses Bgl-mCherry under the control of a doxycycline-inducible promoter and constitutively expresses HIV-1 Tat and Rev. The HeLa-Bgl:Tat-Rev cells were maintained in complete media containing hygromycin (200 µg/µl; Thermo Fisher Scientific), G418 Sulfate (200 µg/ml), and puromycin (1µg/ml). Doxycycline (Sigma-Aldrich; 1 µg/ml) was added for 24 hrs prior to imaging to induce Bgl-mCherry expression.

To generate the HeLa:mRuby-CPSF6 cell line, HeLa cells constitutively expressing an shRNA targeting CPSF6 were first generated by transduction with a lentiviral vector (pLKO.1 hygro; gift from Bob Weinberg; Addgene plasmid #24150), which was modified to express the shRNA targeting CPSF6 (5'-AGACCGTCATGACGATTATTA-3'). A control cell line that expresses a non-targeting control shRNA from the human U6 promoter was also constructed. The transduced cells were selected for resistance to hygromycin (200 µg/ml). Next, pLVX-EF1-mRuby-CPSF6-P2A-Puro, a lentiviral bicistronic vector that expresses mRuby-CPSF6 that is resistant to the CPSF6 shRNA under the control of the human EF1 promoter was created. Briefly, the IRES-Puro cassette in the lentiviral vector pLVX-IRES-Puro (Cat# 632183, Clontech) was replaced with mRuby, which was amplified by PCR from mRuby-LaminB1-10 (Addgene plasmid #55869), and the CMV promoter was replaced with the human EF1 promoter. Next, an shRNA-resistant CPSF6 (GenBank NM_007007.2; synonymous mutations in the nucleotide sequence targeted by CPSF6 shRNA were introduced) and an in-frame P2A fused to Puro were synthesized as multiple gBlocks (Integrated DNA Technologies, Inc., Coralville, Iowa). These fragments were assembled downstream of mRuby by Gibson Assembly (New England Biolabs, Inc., Ipswich, MA), generating the lentiviral vector pLVX-EF1-mRuby-CPSF6-P2A-Puro. The

HeLa cells stably expressing the CPSF6 shRNA were transduced with the lentiviral vector pLVX-EF1-mRuby-CPSF6-P2A-Puro and selected for puromycin resistance (1 µg/ml). The mRuby-CPSF6 fusion protein was expressed in HeLa cells stably expressing CPSF6 shRNA to reduce the amount of endogenous unlabeled CPSF6 and increase the mRuby-CPSF6 labeling efficiency. To validate the cell lines, endogenous CPSF6 and mRuby-CPSF6 were detected by SDS-PAGE and western blot analysis using an anti-CPSF6 antibody (Sigma-Aldrich; Cat#HPA039973) followed by goat anti-rabbit antibody (IRDye-800CW; LI-COR, Lincoln, NE); HSP90 was used as a loading control and detected using an anti-HSP90 antibody (Santa Cruz Biotechnology, Inc., Dallas, TX; Cat#sc-69703) followed by goat anti-mouse antibody (IRDye-680RD; LI-COR). Western blots were imaged and quantitated using the Odyssey infrared imaging system (LI-COR).

To generate the CEM-SS-mRuby-LaminB cell line, CEM-SS cells were transduced with the lentiviral vector pLVX-CMV-mRuby-LaminB-P2A-Puro, a bicistronic lentiviral vector that expresses a mRuby-LaminB fusion protein and Puro under the control of the CMV promoter. Briefly, the mRuby-CPSF6 cassette in pLVX-EF1-mRuby-CPSF6-P2A-Puro was replaced with mRuby-LaminB, which was amplified by PCR from mRuby-LaminB1-10 (Addgene plasmid #55869) and the EF1 promoter was replaced with the CMV promoter. Transduced cells were selected for resistance to puromycin (1 µg/ml). All cells were maintained in humidified 37º C incubators with 5% CO_2. Nevirapine (NVP) and raltegravir (RAL) were obtained through the NIH AIDS Reagent Program and were used at final concentrations of 5 µM and 10 µM, respectively. PF-3450074 (PF74; Sigma) was used at a final concentration of 10 µM.

Lentiviral vectors, protein expression vectors, and virus production. The HIV-1 based vector pHGFP-BglSL was generated by inserting a cassette containing 18 RNA stem loops [BglSL; (4)], which are specifically recognized by the bacterial Bgl protein[21], into pHGFP. The RNA stem loops were separated by short (~10 nucleotide) randomized linkers and were inserted into the Vpr-Vif region of the HIV-1 based vector pHGFP. pHGFP is an HIV-1 vector derived from

pHL (10) that contains a *gfp* reporter gene in place of *nef* and does not express *env*. The resulting HIV-1 vector expresses a *gfp* reporter gene and an RNA that contains 18 BglSL, but does not express Vif, Vpr, and Env. To generate the HIV-1 based vector, pHGFP-GFPCA-BglSL, a Y132I mutation in the protease cleavage site between GFP and CA, which prevents cleavage (11) was introduced into HIV Gag-iGFP [NIH AIDS Reagent Program, Division of AIDS, NIAID, NIH: gift from Dr. Benjamin Chen; Cat#12457; (12)] by site-directed mutagenesis. A cassette containing MA-GFP-CA (containing the Y132I mutation) was inserted in place of MA-CA in pHGFP-BglSL, generating a HIV-1 vector that is similar to pHGFP-BglSL, except that a GFP-CA fusion protein is generated after proteolytic processing. CA mutants N74D and A77V were generated by site-directed mutagenesis of pHGFP and cassettes containing CA with the mutations were inserted into pHGFP-GFPCA-BglSL and pHGFP-BglSL, generating pHGFP(N74D)-GFPCA-BglSL, pHGFP(N74D)-BglSL, pHGFP(A77V)-GFPCA-BglSL, and pHGFP(A77V)-BglSL, respectively. Infectious virions that were labeled with GFP-CA were prepared by co-transfection of 293T cells with pHGFP-GFPCA-BglSL and pHGFP-BglSL (or similar vectors containing CA mutation), which expresses wild-type gag-pol, at a 1:15 plasmid ratio, and pHCMV-G (13), which expresses the G glycoprotein of vesicular stomatitis virus (VSV-G). To generate infectious virions that are labeled with GFP-CA, but do not contain BglSL and express Vif and Vpr, a cassette containing MA-GFP-CA (containing the Y132I mutation) was inserted in place of MA-CA in pHGFP, generating a vector (pHGFP-GFPCA) that expresses a GFP-CA fusion protein after proteolytic processing that also expresses Vif and Vpr. Infectious virions that were labeled with GFP-CA and produced in the presence of Vif and Vpr were prepared by co-transfection of 293T cells with pHGFP-GFPCA and pHGFP at a 1:10 plasmid ratio, and pHCMV-G. Infectious virions containing only wild-type *gag-pol* were produced by co-transfection of 293T cells with pHGFP and pHCMV-G.

A plasmid expressing APOBEC3F (A3F) fused to red-red vine tomato (RRvT) fluorescent protein was generated by replacing YFP in A3F-YFP (3) with RRvT, which was amplified by PCR from pBad-HisB-RRvT [addgene plasmid #87364; (14)]. To generate infectious virions that were

labeled with GFP-CA and A3F-RRvT, a plasmid expressing an A3F-RRvT (1.25 µg) was co-transfected with pHGFP-GFPCA-BglSL, pHGFP-BglSL, and pHCMV-G. To generate single-labeled infectious virions, plasmids expressing either A3F-YFP, A3F-mNeonGreen (mNG; mNeonGreen was amplified by PCR from pmNeonGreen-N1 (Allele Biotechnology, San Diego, CA) and used to replace YFP in A3F-YFP), Vpr-Integrase-YFP (2) , or Vpr-Integrase-superfolder(sf)GFP [kindly provided by Dr. Greg Melikyan; (15)] were co-transfected with pHGFP-BglSL or pHGFPΔVifΔVpr, a pHGFP-derived vector in which the *vif* and *vpr* genes were deleted, and pHCMV-G.

To generate cell lines, the lentiviral vectors used to create the cell lines (*e.g.* pLVX-TRE3G-Bgl-mCherry) were co-transfected with pC-Help (16), which is an HIV-1 helper construct that expresses *gag-pol* but lacks several cis-acting elements needed for viral replication, and pHCMV-G. Supernatants from transfected 293T cells were filtered and the HIV-1 particles were concentrated by ultracentrifugation (100,000 × g) for 1.5 hrs at 4 ºC through a 20% sucrose cushion (wt/vol) in 1X phosphate buffered saline (PBS).

Virus infection. HeLa-based cell lines were seeded in ibiTreated µ-slides (30,000 cells/well; Ibidi, Gräfelfing, Germany) one day prior to infection. The CEM-SS:mRuby-LaminB cell line was seeded into µ-slides (100,000 cells/well) that were pretreated with poly-L-lysine (Sigma-Aldrich; Cat#P8920). THP-1 cells were seeded into µ-slides (50,000 cells/well) and treated with PMA (100 ng/ml) for 48-h prior to spinoculation. Cells were infected with viruses via spinoculation at 16ºC (17), which permitted virion binding to cell membranes but prevented virion endocytosis. For live-cell imaging experiments, infections with the GFP-CA-labeled virions were performed at a multiplicity of infection (MOI) of ≤0.1 GFP-expressing proviruses/cell and in the presence of aphidicolin (2 µg/ml) to prevent cell division during the long movies. After spinoculation, the media was replaced with prewarmed media to allow internalization of the virus (defined as the 0-h time point) and thereafter incubated at 37ºC. Time-lapse images of the infected cells were acquired by spinning disk confocal microscopy (described below) or the cells were fixed at

various time points post-infection with 4.0% (wt/vol) paraformaldehyde (PFA). The half maximal inhibitory concentrations (IC50) for NVP, PF74, and RAL were determined for GFP-CA-labeled and unlabeled virions in TZM-bl cells. Virus infectivity was determined by flow cytometry (LSRFortessa; BD Biosciences, San Jose, CA) 24-48 hpi or after infection of TZM-bl cells and measurement of luciferase activity using the britelite plus Reporter Gene Assay System (Perkin Elmer) 48 hpi.

Microscopy and image processing. Confocal images were acquired using Nikon Eclipse T*i*-E microscope equipped with a Yokogawa CSU-X1 spinning disk unit with a Plan-Apochromat 100x N.A. 1.49 oil objective, using 405-nm (DAPI/AF405), 488-nm (GFP), 514-nm (YFP), 561-nm (mRuby/RRvT), 594-nm (mCherry), and 647-nm (Cy5) lasers for illumination. Images were captured using a TwinCam system (Cairn, Faversham, UK) equipped with a 565-nm splitter and two iXon Ultra (Andor, Belfast, UK) cameras. A Tokai Hit microscope stage top incubator (Tokai, Japan) was used for all live-cell imaging experiments. To visualize the nuclear GFP-CA puncta, HIV-1 transcription sites, and *gfp* reporter expression, z-stacks of the infected cells (13 slices at 0.4 µm interval) were acquired every 20 min for 24 hrs starting at ~4 hrs after infection. Time-lapse images of the infected cells were examined using Nikon Elements or ImageJ (18). For display, a pixel-averaging filter was applied to the images and the contrast was adjusted; unmodified images were used for intensity analyses.

Live-cell imaging and image analysis. GFP-CA signals and HIV-1 TS were manually identified upon extensive analysis of the entire movie. GFP-CA intensities for individual particles were determined using a custom-written MATLAB program. Briefly, the 3D position of the GFP-CA signal in the z-stack was manually determined and then the local background-subtracted pixel intensities at location the GFP-CA signal were determined. Background intensities were determined by selecting random positions in the nuclei of infected cells and extracting pixel intensities. The GFP-CA-labeled nuclear complexes in CEM-SS:mRuby-LaminB cells were identified manually from z-stacks acquired between 6 and 10 hpi; GFP-CA intensities for these

complexes were determined as described above. Of the 116 HIV-1 TS observed (59 in HeLa-Bgl cells and 57 in HeLa-Bgl:Tat-Rev cells), 81 TS were the only TS detected in the cells. When cells had >1 TS, only the timing of the 1st detectable TS was considered. To visualize the nuclear import of GFP-CA or A3F-mNG particles, similar time-lapse images of the infected cells were acquired as described above, except the movies were started ~10 min after infection.

Distances between the GFP-CA-labeled viral complexes and associated HIV-1 TS were determined using a custom written MATLAB program. Briefly, the image of the nucleus in the last frame GFP-CA signal was detected prior to disappearance and the image of the nucleus in the first frame in which the associated HIV-1 TS was detected were aligned to account for cell movement in between GFP-CA loss and detection of HIV-1 TS. Next, the 3D positions of the GFP-CA particle and HIV-1 TS were determined using 3D gaussian fitting, which were then used to calculate distances. A similar method was used to determine the distance HIV-1 TS moved in ~1.5 hrs.

The distance between the HIV-1 TS and the nucleus boundary was determined using a custom-written MATLAB program. Bgl-mCherry contains a nuclear localization signal; thus, most of the Bgl-mCherry localized inside the nucleus and very low amounts of Bgl-mCherry localized with the NE. We identified nuclear pore complexes (NPCs) in HeLa cells expressing Bgl-mCherry by immunostaining with anti-NPC antibody Mab414 (Abcam, Cambridge, UK; Cat#ab24609) followed by AlexaFluor488-labeled secondary antibody (Thermo Fisher Scientific, Waltham, MA; Cat#A-11001) and compared the Bgl-mCherry signals localized to the NPCs and inside the nucleus (Fig. S3E-F). As expected, the Bgl-mCherry signals inside the nucleus were significantly higher than at the NPCs (414 and 145, respectively; Fig. S3F), indicating that the reduction in the Bgl-mCherry signal can be used to visualize the nucleus boundary. The position of the HIV-1 TS at the time of detection was determined using 3D gaussian fitting. Next, the nearest edge of the Bgl-mCherry signal to the HIV-1 TS was determined and then the 3D distance between the nucleus boundary and the HIV-1 TS was determined.

Fixed-cell imaging and image analysis. A custom-written MATLAB program was used to determine the colocalization of A3F-RRvT and GFP-CA signals inside the nuclei of infected cells. First, the A3F-RRvT signals were detected using Localize (19). Colocalization of the fluorescently-labeled virus particles with a mask of the nucleus interior (based on immunofluorescence staining using an anti-Lamin A/C antibody [Sigma-Aldrich; Cat#L1293 or Thermo Fisher Scientific; Cat#MA3-1000] followed by detection using a AlexaFluor405-labeled secondary antibody [Thermo Fisher Scientific, Cat#A-31553 or Cat#A-31556) was determined using a custom-written MATLAB program as previously described (2). Next, to determine background signals in the GFP channel, the intensity values in the GFP channel at random positions inside >50 nuclei were determined; the threshold was determined as the mean + 2 SD of the random intensities. The intensity values of the GFP channel at the position of each nuclear A3F-RRvT particle or at random positions in the nuclei were determined; A3F-RRvT particles co-localizing with GFP signals that were above the threshold intensity value were considered positive for GFP-CA. For some experiments, a mask of the NE was also created using the Lamin A/C signal. The percentage of GFP-CA, Integrase-YFP, or A3F-YFP particles that colocalized with the NE mask and the percentage of particles that colocalized with the nucleus mask were determined. To determine the percentage of GFP-CA- or Integrase-sfGFP-labeled viral complexes that were mRuby-CPSF6^{+} and the intensities of those complexes, the GFP-CA- or Integrase-sfGFP-labeled nuclear complexes in HeLa:mRuby-CPSF6 cells were identified manually from z-stacks acquired 6 hpi. Pixel intensities at the locations of the GFP-CA or Integrase-sfGFP particle signals were determined after subtracting the local background. Nuclear background intensities were determined by selecting 85 random positions from the nuclei of 5 cells and extracting pixel intensities in the mRuby channel; the threshold intensity was determined as the mean + 1 SD of the random intensities in the mRuby channel. GFP-CA or Integrase-sfGFP particles co-localizing with mRuby signals that were above the threshold value were considered positive for mRuby-CPSF6.

Single Virion Analysis (SVA). Fluorescently-labeled virus particles were centrifuged onto ibiTreated µ-slides (1,200 x g for 1 hr). CA was detected using an anti-CA antibody (AG3.0; NIH AIDS Reagent Program, Division of AIDS, NIAID, NIH: gift from Dr. Jonathan Allan; Cat#4121) followed by a Cy5-labeled secondary antibody (Thermo Fisher Scientific, Cat#A-10524), and then imaged by spinning disk confocal microscopy. The diffraction-limited spots were detected, and their positions were determined in each image using Localize. The positions of the spots were used to determine colocalization; spots were considered colocalized if the centers of the spots were within 3 pixels.

***In vitro* analysis of intact virions and viral cores.** GFP-CA-labeled virus particles were centrifuged onto ibiTreated µ-slides (1,200 × g for 1 hr). Intact virions bound to the slides were lysed by saponin treatment *in vitro* to remove viral membranes and free CA that was not incorporated into viral cores; an equal volume of PBS containing 2X saponin detergent (0.1% final concentration) was added to the wells containing virus particles between the 2nd and 3rd time points. Time-lapse images of the virus particles were acquired every 10 seconds for 5 minutes. A custom-written MATLAB program was used to identify GFP-CA particles and determine their signal intensities. Virus particles that lost ≥33% of the initial GFP-CA signal were considered to lose the viral membrane (typically 90% of particles) and were used for further analysis. Next, the virus particles were separated into three categories; first, some particles (4%) lost all GFP-CA signal at the time of saponin treatment; second, some particles (44%) lost the GFP-CA signal in two discrete steps, losing free GFP-CA that was not incorporated into viral cores at the time of saponin treatment and subsequently losing viral core-associated GFP-CA signal during the 5-min observation time (i.e. uncoated); third, some particles (52%) lost free GFP-CA that was not incorporated into viral cores and retained viral core-associated GFP-CA until the end of the 5-min observation time. The background-subtracted GFP-CA intensities of 4,441 intact virions (avg. 974; before saponin treatment) and 3,681 viral cores (avg. 291; ~1 min after saponin treatment) were determined. The detection limit in cells was determined using intensity values in the GFP channel at 120 random positions (5 random positions/nucleus inside the nuclei of 24 infected

cells). The threshold was determined as the mean + 3 SD of the random intensities and the 4,374 intact virions and 1,641 viral cores with GFP-CA signals above this threshold were considered detectable in cells. Thus, 98.5% of the GFP-CA-labeled intact virions (4,374/4,441 = 98.5%) and 45% of the GFP-CA-labeled viral cores (1,641/3,681 = 45%) would be detectable inside cells. The GFP-CA intensities for the 4,374 intact virions and 1,641 viral cores that would be detectable in cells were used to compare to the GFP-CA intensity of 116 nuclear complexes in Hela cells (59 from HeLa-Bgl and 57 from Hela-Bgl:Tat-Rev) and 223 nuclear complexes in CEM-SS:mRuby-LaminB cells (Fig. 3C).

Transmission Electron Microscope (TEM) analysis of virus pellets. Virus pellets were obtained by ultracentrifugation as described above, washed once in PBS, processed for thin-sectioned EM analysis as previously described (20, 21). Briefly, virus pellets were fixed in 2% (v/v) glutaraldehyde (Tousimis, Rockville, MD) in 0.1 M sodium cacodylate (pH 7.4; Electron Microscopy Sciences [EMS], Fort Washington, PA), followed by Osmium post fixation (1% Osmium tetroxide v/v in same buffer for 1 hr). The pellets were dehydrated in a series of ethanol solutions (e.g., 35%, 50%, 70%, 95%, 100%) and 100% propylene oxide. An infiltration was made in 1:1 mixture of propylene oxide and epoxy resin (EMS) overnight, and the pellets were embedded in pure resin. The resin was cured for 48 hrs in 55ºC and thin sections (60 to 70 nm) were made with an ultra-microtome equipped with a diamond knife and mounted on naked copper grid. The thin sections were stained in uranyl acetate (EMS) and lead citrate (Leica, Bannockburn, IL), stabilized by carbon evaporation. The sections were examined and images were captured by digital camera in Hitachi 7650 TEM (Hitachi, Tokyo, Japan) operated at 75kv (21). Virions exhibiting the mature and immature phenotype were scored manually.

Fractionation of viral cores using sucrose gradients. The fractionation of viral cores using sucrose gradients was performed as previously described, with slight modifications (22). Concentrated virions were prepared as described above and subjected to ultracentrifugation (100,000 × g for 16 hr at 4 ºC) through a layer of 1% Triton X-100 into a linear sucrose density

gradient (10 mls of 10 mM Tris-HCl [pH 7.4], 100 mM NaCL, 1mM EDTA containing 30% to 70% sucrose). Following ultracentrifugation, fractions (1 ml) were collected from the top of the gradient. CA and GFP-CA protein were detected in each fraction by western blot analysis using antibodies against CA (NIH AIDS Reagent Program, Division of AIDS, NIAID, NIH, HIV-1 p24 Gag Monoclonal (#24-3) from Dr. Michael H. Malim; Cat#6458) and GFP (Thermo Fisher Scientific; Cat#A-6455). The amount of CA in the peak fractions (fractions 8 and 9), which represent CA associated with viral cores, was divided by the total amount of CA in the fractions to determine the yield of cores.

Data analysis and statistics. The Welch's unpaired *t*-test and paired *t*-test were used to analyze parametric data. The Mann Whitney *U* test was used to analyze nonparametric data and the Kolmogorov-Smirnov test was used to analyze cumulative frequencies. A Fisher's exact test was used to analyze 2 x 2 contingency tables. All statistical tests were performed in Prism 8 (GraphPad Software, San Diego, CA). *P* values <0.05 were considered significant.

SI References

1. Wei X*, et al.* (2002) Emergence of resistant human immunodeficiency virus type 1 in patients receiving fusion inhibitor (T-20) monotherapy. *Antimicrob Agents Chemother* 46(6):1896-1905.
2. Burdick RC*, et al.* (2017) Dynamics and regulation of nuclear import and nuclear movements of HIV-1 complexes. *PLoS Pathog* 13(8):e1006570.
3. Burdick RC, Hu WS, & Pathak VK (2013) Nuclear import of APOBEC3F-labeled HIV-1 preintegration complexes. *Proc Natl Acad Sci U S A* 110(49):E4780-4789.
4. Chen J*, et al.* (2009) High efficiency of HIV-1 genomic RNA packaging and heterozygote formation revealed by single virion analysis. *Proc Natl Acad Sci U S A* 106(32):13535-13540.
5. Gritz L & Davies J (1983) Plasmid-encoded hygromycin B resistance: the sequence of hygromycin B phosphotransferase gene and its expression in Escherichia coli and Saccharomyces cerevisiae. *Gene* 25(2-3):179-188.
6. Yin PD, Pathak VK, Rowan AE, Teufel RJ, 2nd, & Hu WS (1997) Utilization of nonhomologous minus-strand DNA transfer to generate recombinant retroviruses. *J Virol* 71(3):2487-2494.
7. Kim JH*, et al.* (2011) High cleavage efficiency of a 2A peptide derived from porcine teschovirus-1 in human cell lines, zebrafish and mice. *PLoS One* 6(4):e18556.
8. Miller AD*, et al.* (1991) Construction and properties of retrovirus packaging cells based on gibbon ape leukemia virus. *J Virol* 65(5):2220-2224.
9. Landau NR, Page KA, & Littman DR (1991) Pseudotyping with human T-cell leukemia virus type I broadens the human immunodeficiency virus host range. *J Virol* 65(1):162-169.

10. Nikolenko GN*, et al.* (2007) Mutations in the connection domain of HIV-1 reverse transcriptase increase 3'-azido-3'-deoxythymidine resistance. *Proc Natl Acad Sci U S A* 104(1):317-322.
11. Lee SK, Harris J, & Swanstrom R (2009) A strongly transdominant mutation in the human immunodeficiency virus type 1 gag gene defines an Achilles heel in the virus life cycle. *J Virol* 83(17):8536-8543.
12. Hubner W*, et al.* (2007) Sequence of human immunodeficiency virus type 1 (HIV-1) Gag localization and oligomerization monitored with live confocal imaging of a replication-competent, fluorescently tagged HIV-1. *J Virol* 81(22):12596-12607.
13. Yee JK*, et al.* (1994) A general method for the generation of high-titer, pantropic retroviral vectors: highly efficient infection of primary hepatocytes. *Proc Natl Acad Sci U S A* 91(20):9564-9568.
14. Wiens MD*, et al.* (2016) A Tandem Green-Red Heterodimeric Fluorescent Protein with High FRET Efficiency. *Chembiochem* 17(24):2361-2367.
15. Francis AC & Melikyan GB (2018) Single HIV-1 Imaging Reveals Progression of Infection through CA-Dependent Steps of Docking at the Nuclear Pore, Uncoating, and Nuclear Transport. *Cell Host Microbe* 23(4):536-548 e536.
16. Mochizuki H, Schwartz JP, Tanaka K, Brady RO, & Reiser J (1998) High-titer human immunodeficiency virus type 1-based vector systems for gene delivery into nondividing cells. *J Virol* 72(11):8873-8883.
17. O'Doherty U, Swiggard WJ, & Malim MH (2000) Human immunodeficiency virus type 1 spinoculation enhances infection through virus binding. *J Virol* 74(21):10074-10080.
18. Schneider CA, Rasband WS, & Eliceiri KW (2012) NIH Image to ImageJ: 25 years of image analysis. *Nat Methods* 9(7):671-675.
19. Zenklusen D, Larson DR, & Singer RH (2008) Single-RNA counting reveals alternative modes of gene expression in yeast. *Nat Struct Mol Biol* 15(12):1263-1271.
20. Mariner JM, McMahon JB, O'Keefe BR, Nagashima K, & Boyd MR (1998) The HIV-inactivating protein, cyanovirin-N, does not block gp120-mediated virus-to-cell binding. *Biochem Biophys Res Commun* 248(3):841-845.
21. Tobin GJ, Nagashima K, & Gonda MA (1996) Immunologic and Ultrastructural Characterization of HIV Pseudovirions Containing Gag and Env Precursor Proteins Engineered in Insect Cells. *Methods* 10(2):208-218.
22. Forshey BM, von Schwedler U, Sundquist WI, & Aiken C (2002) Formation of a human immunodeficiency virus type 1 core of optimal stability is crucial for viral replication. *J Virol* 76(11):5667-5677.

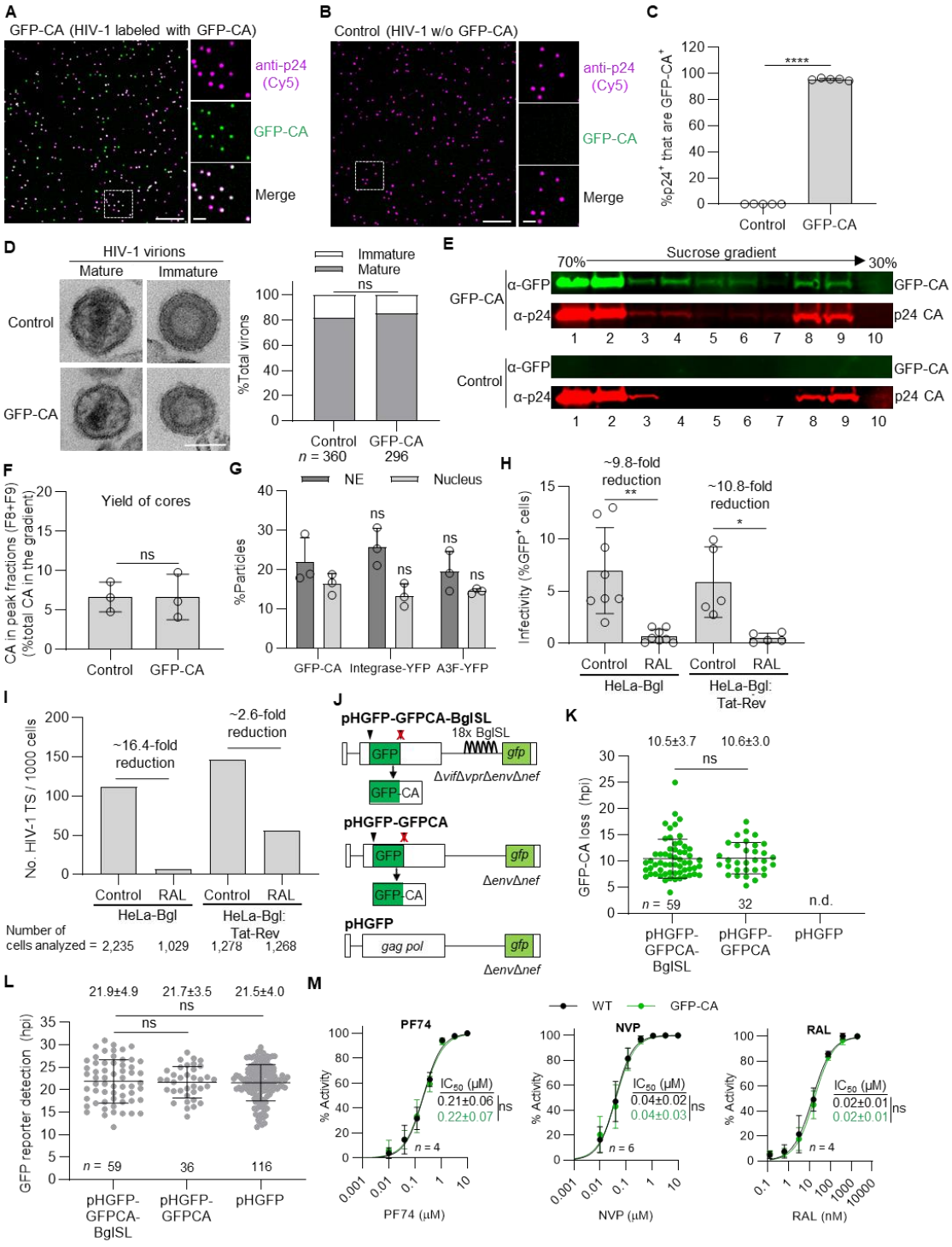

Fig. S1. Characterization of GFP-CA labeled virions and infectivity and detection of TS in HeLa-Bgl and HeLa-Bgl:Tat-Rev cells. (**A-B**) Single virion analysis of HIV-1 virions labeled with and

without GFP-CA. HIV-1 virions produced by co-transfection of 293T cells with pHGFP-GFPCA-BglSL and pHGFP-BglSL at a 1:15 ratio along with VSV-G expressing plasmid (**A**) or produced in the absence of pHGFP-GFPCA-BglSL (**B**) were centrifuged onto a slide, fixed, and immunostained with anti-CA antibody followed by Cy5-labeled secondary antibody. Scale bars, 10 μm; inset scale bars, 2 μm. (**C**) Quantitation of the proportion of p24 CA^+ virions that are GFP-CA^+ showed that 96% of the virions produced in the presence of pHGFP-GFPCA-BglSL were GFP-CA^+, whereas <0.1% of the virions produced in the absence of pHGFP-GFPCA-BglSL were GFP-CA^+. Data are mean ± s.d. from 5 independent experiments; *P* values are from paired *t*-tests. (**D**) Quantitative electron microscopy of HIV-1 virions shows that similar high proportions (>80%) of GFP-CA-labeled and control virions have a capsid core, indicative of a mature morphology. Scale bar, 100 nm. Statistical significance was determined using Fisher's exact test. (**E-F**) Sucrose gradient fractionation of HIV-1 virions with WT and GFP-CA Gag (1:15 ratio) and with only WT Gag (**E**) and the quantitation of the proportion of CA in fractions 8 and 9 of the sucrose gradient, which contain intact viral cores, relative to total CA in the gradient (**F**). The results show that similar proportions of CA associated with intact viral cores were recovered, indicating that GFP-CA labeling of viral cores did not influence their *in vitro* stability. Data are mean ± s.d. from 3 independent experiments; *P* values are from Welch's *t*-tests. (**G**) Comparison of the proportion of viral complexes in infected HeLa cells that stably associated with the NE and entered the nucleus at 6 hpi. The results show that similar proportions of virions labeled with GFP-CA, Integrase-YFP, or A3F-YFP were stably associated with the NE and entered the nucleus, indicating that GFP-CA labeling of viral cores did not significantly affect their association with the NE or their nuclear entry. Data are mean ± s.d. from 3 independent experiments; *P* values are from Welch's *t*-tests. (**H**) Infection of Hela-Bgl and HeLa-Bgl:Tat-Rev cells with GFP-CA-labeled virions at low MOIs resulted in GFP reporter expression in ~7% of the infected cells (MOIs ~0.05 – 0.1). Approximately 5- to 10-fold lower amounts of virions were used to infect HeLa:Tat-Rev cells to achieve a similar proportion of GFP^+ cells. Treatment of target cells with RAL significantly reduced GFP-expressing HeLa-Bgl cells by ~9.8-fold and HeLa-Bgl:Tat-Rev

cells by ~10.8-fold, indicating that integration was required for GFP reporter expression. Data are mean ± s.d. from ≥5 independent experiments; *P* values are from Welch's *t*-tests. (**I**) Number of HIV-1 TS/1,000 cells. Approximately 111 HIV-1 TS were detected per 1,000 infected Hela-Bgl cells and 147 TS were detected per 1,000 infected HeLa-Bgl:Tat-Rev cells. RAL treatment reduced HIV-1 TS in HeLa-Bgl cells ~16.4-fold and in HeLa-Bgl:Tat-Rev cells ~2.6-fold, indicating that the majority of HIV-1 TS represented transcription from integrated viral DNA. RAL treatment reduced HIV-1 TS to a lesser extent in HeLa-Bgl:Tat-Rev cells, indicating that constitutive expression of Tat increased detectable transcription from unintegrated DNA. (**J**) Structures of pHGFP-GFPCA-BglSL, pHGFP-GFPCA and pHGFP. pHGFP-GFPCA-BglSL does not express Vif, Vpr, Env, or Nef and contains 18 copies of stem loops that specifically bind to Bgl protein. pHGFP-GFPCA expresses Vif and Vpr but not Env or Nef and does not contain any Bgl stem loops. pHGFP expresses Vif and Vpr, but not Env or Nef, and does not contain any Bgl stem loops; in addition, it does not express a GFP-CA fusion protein. (**K**) GFP-CA loss occurs with the same kinetics in cells infected with pHGFP-GFPCA-BglSL and pHGFP-GFPCA, indicating that the presence of Bgl stem loops or absence of Vif and Vpr expression does not affect the kinetics of nuclear uncoating. (**L**) The kinetics of GFP reporter detection was not affected by the presence of Bgl stem loops, absence of Vif or Vpr expression, or GFP-CA labeling of viral complexes, indicating that labeling of virions with GFP-CA did not significantly affect the kinetics of GFP reporter expression. For (**K**) and (**L**), lines are mean ± s.d.; *P* values are from Welch's *t*-tests. n.d., not determined. (**M**) Sensitivity of GFP-CA-labeled and unlabeled virions to PF74 (left), NVP (middle), and RAL (right). The half maximal inhibitory concentrations (IC50) were determined for each experiment (n ≥ 4) and statistical significance was determined using Welch's *t*-tests. ****, $P < 0.0001$; **, $P < 0.01$; *, $P < 0.05$; ns, not significant ($P > 0.05$). Open circles in (**C**), (**F-H**) indicate independent experiments.

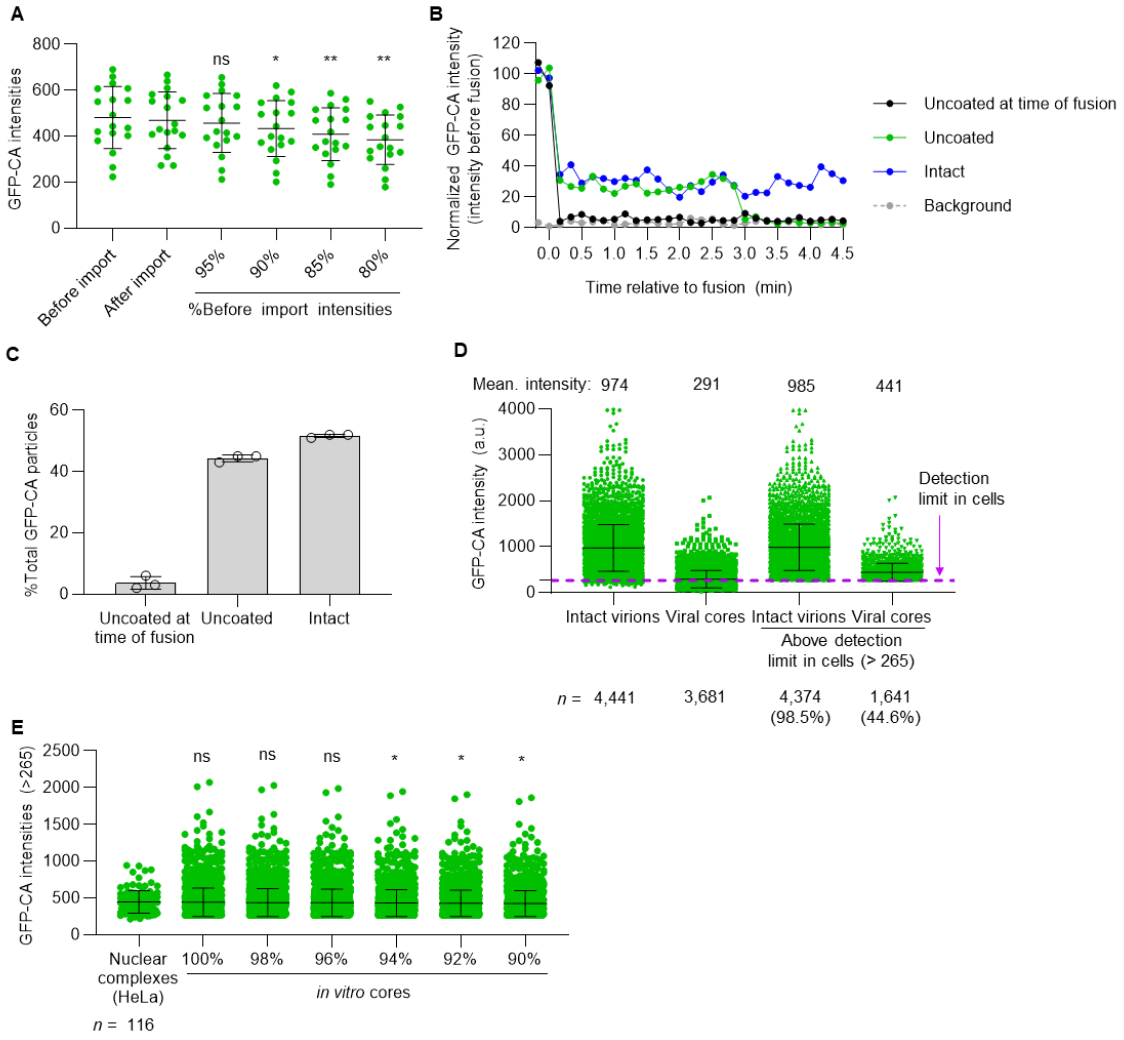

Fig. S2. Modeling GFP-CA loss to determine sensitivity of GFP-CA detection and determination of *in vitro* GFP-CA intensities of intact virions and viral cores. (**A**) Modeling GFP-CA loss during nuclear import. GFP-CA intensities of 18 viral complexes before nuclear import were adjusted to reflect 5-20% GFP-CA loss, particles with GFP-CA intensities below the limit of detection (<265) were removed, and the adjusted GFP-CA intensities were compared to the intensities of the 18 viral complexes after nuclear import. The results indicated that ≥10% loss of GFP-CA would be statistically significant ($P < 0.05$). (**B-C**) Representative graphs of GFP-CA intensities of viral complexes after 0.1% saponin treatment leading to disruption of the viral membrane, mimicking fusion of the viral and host membranes. A small proportion of viral complexes (~4%) lost all GFP-signal upon saponin treatment, indicative of unstable viral cores that disassembled upon fusion;

44% of the viral complexes exhibited a ≥33% loss of GFP-CA signal at the time of fusion, followed by loss of the remaining GFP-CA signal within 4.5 min after saponin treatment, indicative of uncoating; another 52% of the particles lost ≥33% of the GFP-CA signal at the time of fusion, and retained the remaining GFP-CA signal during the 4.5 min observation time. Open circles in (**C**) indicate independent experiments. (**D**) Comparison of the GFP-CA intensities (arbitrary units; a.u.) of intact virions and *in vitro* viral cores. Intact virions and *in vitro* viral cores with GFP-CA intensities below the detection limit in cells (<265 a.u.) were removed (right). The percentage of signals that were above the detection limit in cells is shown. Lines are mean ± s.d. (**E**) Modeling GFP-CA loss of nuclear viral complexes in comparison to *in vitro* cores. GFP-CA intensities of *in vitro* viral cores were adjusted to reflect 2-10% GFP-CA loss, viral cores with GFP-CA intensities below the limit of detection (<265) were removed, and the adjusted GFP-CA intensities were compared to the intensities of 116 GFP-CA labeled nuclear complexes in HeLa-Bgl and HeLa-Bgl:Tat-Rev cells (Fig. 3C). The results indicated that ≥6% loss of GFP-CA would be statistically significant ($P < 0.05$). Lines are mean ± s.d.; P values are from Welch's t-tests. **, $P < 0.01$; *, $P < 0.05$; ns, not significant ($P > 0.05$).

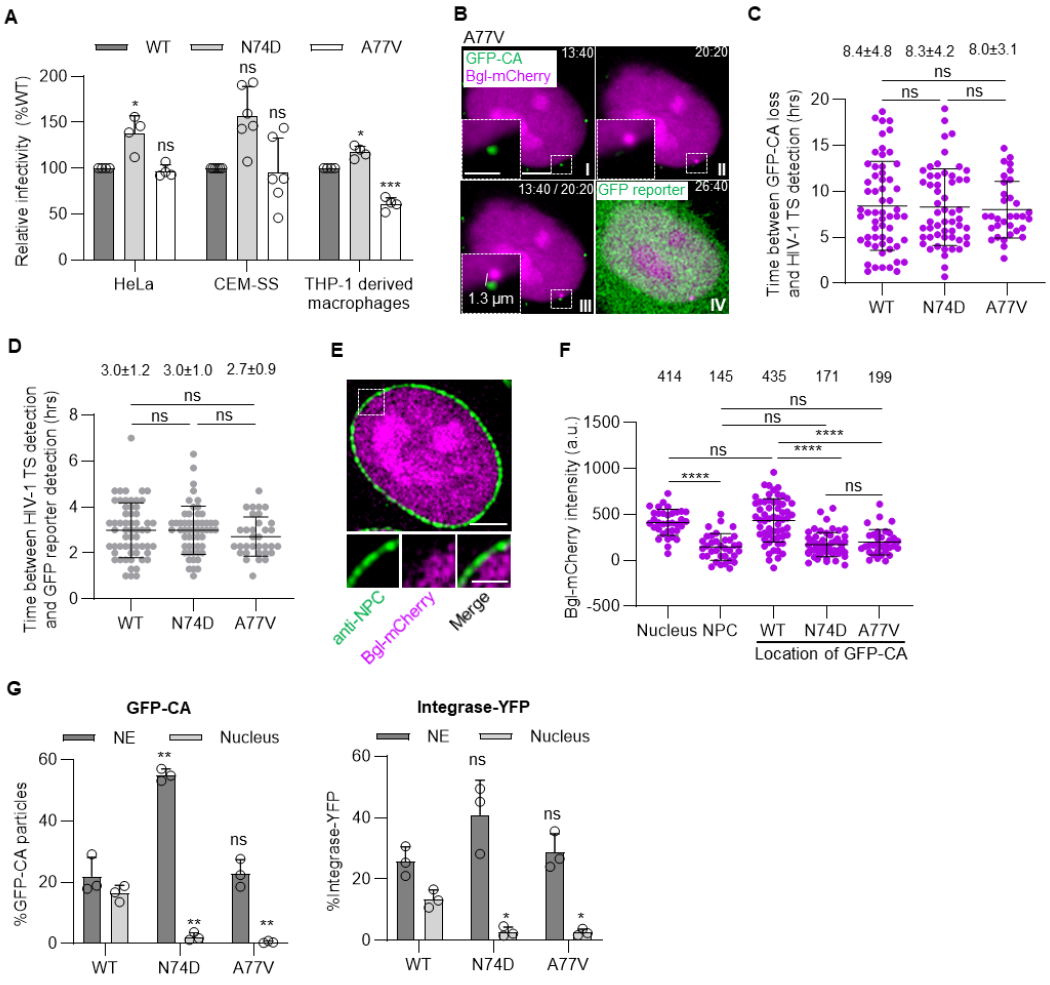

Fig. S3. Characterization of CA mutants N74D and A77V, which disrupt viral core interactions with CPSF6. (**A**) Relative infectivities of WT, N74D, and A77V mutants show that the CA mutations have <2-fold effect on virus infectivity in HeLa cells, CEM-SS cells, and THP-1-derived macrophages. Data are mean ± s.d. from ≥3 independent experiments; *P* values are from Welch's *t*-tests. (**B**) Live-cell microscopy (20 min/frame) images of Hela:Bgl-mCherry cells infected with GFP-CA-labeled virions of CA mutant A77V. A GFP-CA-labeled viral complex uncoated at the edge of the nuclear Bgl-mCherry signal, 13:40 hpi (I) and HIV-1 TS appeared near the site of GFP-CA disappearance 20:20 hpi (II). The HIV-1 TS appeared 1.3 µm from site of GFP-CA disappearance (III). GFP reporter expression was detected 26:40 hpi (IV). (**C**) Comparison of WT GFP-CA-labeled virions with N74D and A77V mutant GFP-labeled virions

indicates that the CA mutations did not influence the average time between GFP-CA loss and HIV-1 TS detection. Data for WT same as Fig. 1F. (**D**) Comparison of WT GFP-CA labeled virions with N74D and A77V mutant GFP-labeled- virions indicates that the CA mutations did not influence the average time between HIV-1 TS detection and GFP reporter detection. Data for WT same as Fig. 1G. (**E**) Immunofluorescence staining of HeLa:Bgl-mCherry cells with an anti-NPC antibody shows that the NE is located at the edge of the nuclear Bgl-mCherry signal. (**F**) Average Bgl-mCherry intensities in the nucleus were significantly higher than the Bgl-mCherry intensities at the NPCs (414 vs. 145 a.u.). The Bgl-mCherry intensities at the locations of WT GFP-CA complexes prior to uncoating were similar to those in the nucleus (435 vs. 414) indicating that the WT GFP-CA viral complexes are localized to the nucleus. The Bgl-mCherry intensities at the locations of the N74D and A77V GFP-CA-labeled viral complexes prior to uncoating were similar to those at the NPCs (171 and 199 vs. 145), indicating that the CA mutant GFP-CA viral complexes are localized at the NPCs. (**G**) Comparison of the proportion of WT, N74D and A77V viral complexes in infected HeLa cells that stably associated with the NE and entered the nucleus at 6 hpi. The results show that viral the N74D and A77V viral complexes stably associated with the NE, but a very low frequency of the mutant viral complexes could be detected in the nucleus compared to WT viral complexes. Both GFP-CA-labeled and Integrase-YFP-labeled N74D/A77V viral complexes were not detected in the nucleus, indicating that the low frequency of nuclear viral complexes was not a result of labeling viral complexes with GFP-CA. Scale bars, 5 μm; inset scale bars, 2 μm. For (**A**) and (**G**), data are mean ± s.d. from ≥3 independent experiments; *P* values are from Welch's *t*-tests. For (**C-D**), and (**F**), lines are mean ± s.d.; *P* values are from Welch's *t*-tests. ****, $P < 0.0001$; ***, $P < 0.001$; **, $P < 0.01$; *, $P < 0.05$; ns, not significant ($P > 0.05$). Open circles in (**A**) and (**G**) indicate independent experiments.

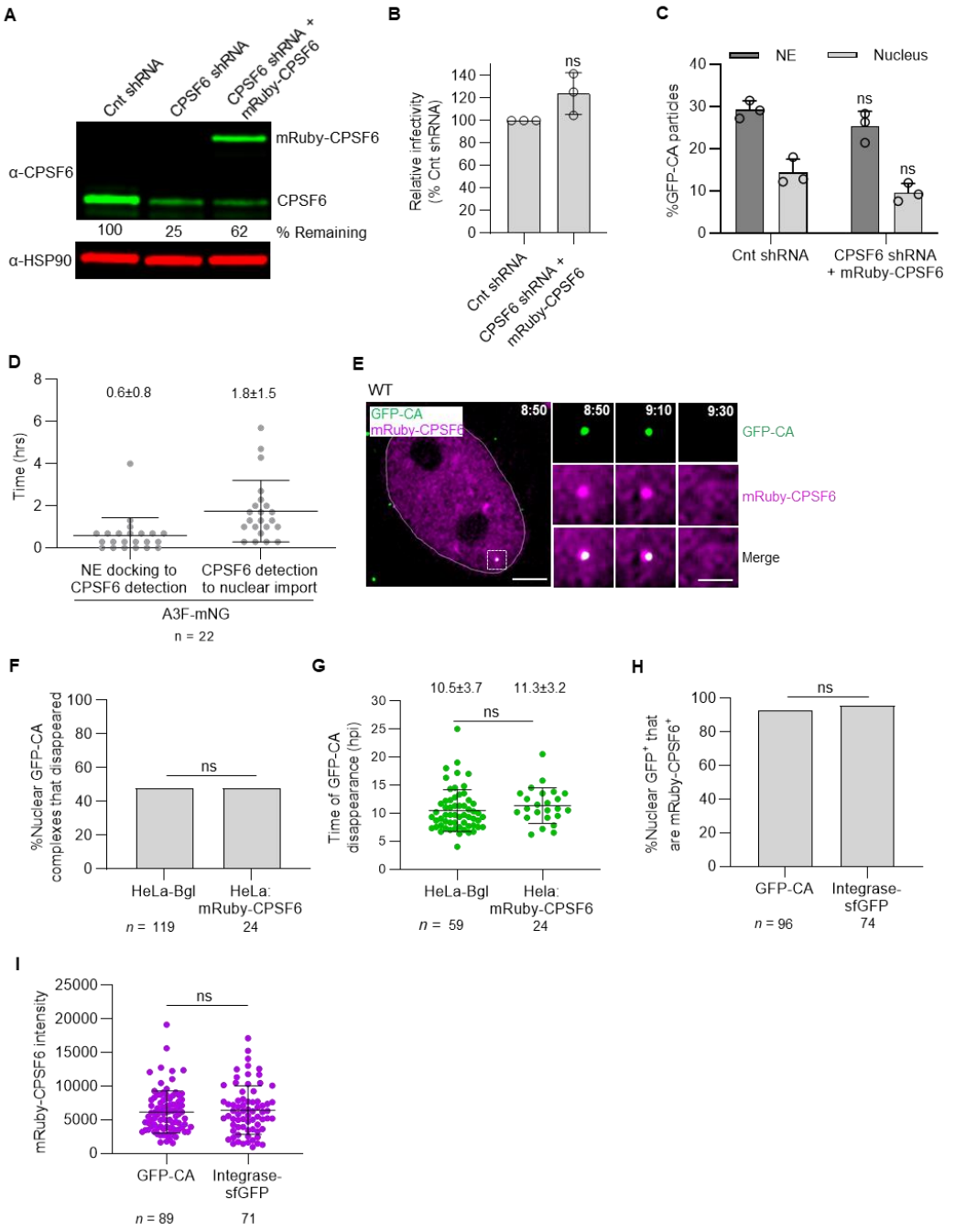

Fig. S4. Characterization of HeLa:mRuby-CPSF6 cells. (**A**) Western blot analysis of HeLa cells stably expressing control shRNA, CSPF6 shRNA, and CPSF6 shRNA + shRNA-resistant mRuby-CPSF6. The results show that stable expression of CPSF6 shRNA reduced endogenous CPSF6 expression to 25% of the control shRNA, and expression of the shRNA-resistant mRuby-CPSF6

restored the CPSF6 levels to 62% of the control shRNA. (**B-C**) CPSF6 shRNA + mRuby-CPSF6 expression did not influence virus infectivity (**B**), the NE-docking efficiency or nuclear import efficiency of GFP-CA-labeled viral complexes (**C**). Data are mean ± s.d. from 3 independent experiments; *P* values are from Welch's *t*-tests. (**D**) Analysis of 22 A3F-mNG-labeled viral complexes showed recruitment of mRuby-CPSF6 to the NE-associated viral complexes 0.6 ± 0.8 hrs after NE docking. In addition, nuclear import of the A3F-mNG + mRuby-CPSF6-labeled complexes occurred 1.8 ± 1.5 hrs after detection of CPSF6 recruitment to the NE. These values are not different than the values in Fig. 4g ($P > 0.05$). Lines are mean ± s.d. (**E**) Representative images of a nuclear GFP-CA-labeled viral complex that was labeled with mRuby-CPSF6. The images show simultaneous loss of both GFP-CA and mRuby-CPSF6 at the time of nuclear uncoating (9:30 hpi). Scale bar, 5 µm; inset scale bar, 2 µm. (**F-G**) CPSF6 shRNA + mRuby-CPSF6 expression did not influence the proportion of nuclear GFP-CA-labeled viral complexes that uncoated (**F**) or the time of GFP-CA disappearance after infection (**G**). (**H-I**) A comparison of the percentage of GFP-CA- or Integrase-sfGFP-labeled nuclear viral complexes that had detectable levels of mRuby-CPSF6 (**H**) and their associated mRuby-CPSF6 intensities (**I**) indicating that the levels of mRuby-CPSF6 associated with nuclear viral complexes are not altered with GFP-CA labeling. For (**F**) and (**H**), *P* values are from Fisher's exact test. For (**D**) and (**I**), lines are mean ± s.d.; *P* values are from Welch's *t*-tests. ns, not significant ($P > 0.05$). Open circles in (**B**) and (**C**) indicate independent experiments.

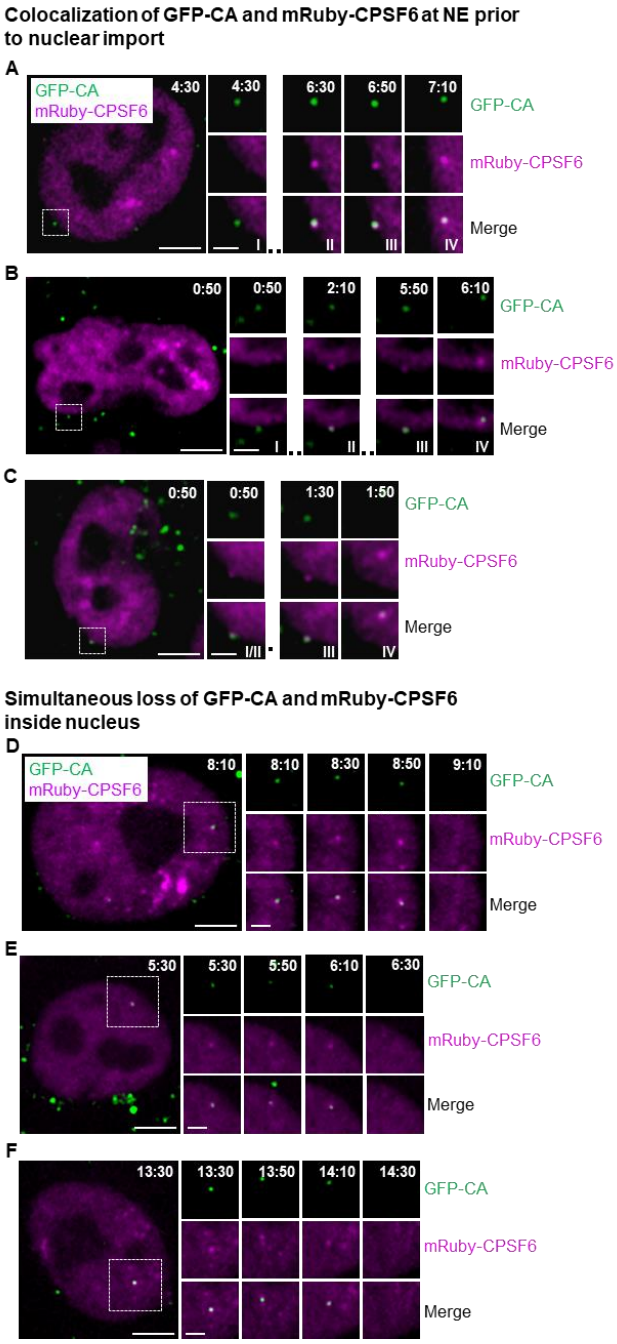

Fig. S5. Additional examples showing the recruitment of mRuby-CPSF6 to GFP-CA-labeled viral complexes at the NE prior to nuclear import and simultaneous disappearance of mRuby-CPSF6 and GFP-CA in the nucleus. (**A-C**) Representative images taken from movies showing mRuby-CPSF6 recruitment to GFP-CA-labeled viral complex. The first frame the GFP-CA-labeled viral complex appeared at the NE (I), first frame of mRuby-CPSF6 detection (II), frame immediately

prior to (III) and after (IV) nuclear import are shown. (**D-F**) Representative images taken from movies showing simultaneous loss of mRuby-CPSF6 and GFP-CA in the nucleus. The three frames prior to and one frame after GFP-CA disappearance are shown. Time scale, hours:minutes post-infection. Scale bar, 5 μm; inset scale bar, 2 μm.

Supplementary Movie Legends

Movie S1. Uncoating of an infectious HIV-1 complex. HeLa-Bgl cells were infected with GFP-CA-labeled virions. Time-lapse images of the cells were acquired every 20 min using spinning disk confocal microscopy. A nuclear GFP-CA-labeled viral complex uncoated and lost the GFP-CA signal 7:10 hpi, a HIV-1 transcription site appeared near the site of GFP-CA disappearance 21:50 hpi, and *gfp* reporter was detected 23:50 hpi. For display, the z-slice closest to the viral complex was extracted from the z-stack for each time point. Time scale, hours:minutes post-infection; scale bar, 5 μm.

Movie S2. Uncoating of an infectious HIV-1 complex in cells expressing exogenous Tat and Rev. HeLa-Bgl:Tat-Rev cells were infected with GFP-CA-labeled virions. Time-lapse images of the cells were acquired every 20 min using spinning disk confocal microscopy. A nuclear GFP-CA-labeled viral complex uncoated and lost the GFP-CA signal 6:20 hpi, a HIV-1 transcription site appeared near the site of GFP-CA disappearance 7:20 hpi, and *gfp* reporter was detected 11:40 hpi. For display, the z-slice closest to the viral complex was extracted from the z-stack for each time point. Time scale, hours:minutes post-infection; scale bar, 5 μm.

Movie S3. Disappearance of a nuclear GFP-CA-labeled viral complex <10 min after PF74 treatment. HeLa-Bgl cells were infected with GFP-CA-labeled virions. Starting at 4 hpi, time-

lapse images of the cells were acquired every 2 min for 1 hr using spinning disk confocal microscopy. After the second frame, the movie was paused, cells were treated with 10 µM PF74, and the imaging was resumed for the remaining time. For display, the z-slice closest to the viral complex was extracted from the z-stack for each time point. Time scale, hours:minutes relative to PF74 treatment; scale bar, 2 µm.

Movie S4. Uncoating of an infectious HIV-1 complex containing the N74D CA mutation. HeLa-Bgl cells were infected with GFP-CA-labeled virions containing the N74D CA mutation. Time-lapse images of the cells were acquired every 20 min using spinning disk confocal microscopy. A GFP-CA-labeled viral complex uncoated and lost the GFP-CA signal at or near the NE (7:20 hpi), a HIV-1 transcription site appeared near the site of GFP-CA disappearance 13:40 hpi, and *gfp* reporter was detected 17:00 hpi. For display, the z-slice closest to the viral complex was extracted from the z-stack for each time point. Time scale, hours:minutes post-infection; scale bar, 5 µm.

Movie S5. Visualizing the nuclear import of a GFP-CA-labeled viral complex in cells expressing mRuby-CPSF6. HeLa:mRuby-CPSF6 cells were infected with GFP-CA-labeled virions. Time-lapse images of the cells were acquired every 20 min using spinning disk confocal microscopy. A GFP-CA-labeled viral complex docked at the NE (5:10 hpi), colocalized with mRuby-CPSF6 (5:30 hpi), and entered the nucleus (7:10 hpi). For display, the z-slice closest to the viral complex was extracted from the z-stack for each time point. Time scale, hours:minutes post-infection; scale bar, 5 µm.

Movie S6. Visualizing the simultaneous loss of GFP-CA and associated mRuby-CPSF6 in the nucleus. HeLa:mRuby-CPSF6 cells were infected with GFP-CA-labeled virions. Time-lapse

images of the cells were acquired every 20 min using spinning disk confocal microscopy. A nuclear GFP-CA-labeled viral complex was visualized for ~4 hrs prior to simultaneous disappearance of GFP-CA and the associated mRuby-CPSF6 (9:30 hpi). For display, the z-slice closest to the viral complex was extracted from the z-stack for each time point. Time scale, hours:minutes post-infection; scale bar, 5 μm.